Essays on Housing and Land Economics

by

Gabriel M. Ehrlich

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2012

Doctoral Committee:

Professor Robert B. Barsky, Chair
Professor Dennis R. Capozza
Associate Professor Christopher L. House
Assistant Professor David Yves Albouy

ACKNOWLEDGEMENTS

I thank my committee, Bob Barsky, Dennis Capozza, Chris House, and David Albouy, for their guidance, encouragement, and advice.

I thank Ruediger Bachmann and Stephan Lauerman for their extensive comments and guidance during the early stages of this project.

I thank Josh Montes, Ryan Nunn, and David Ratner for completing so many problem sets and for studying so many hours with me, as well as for many helpful discussions and insights.

I thank the Research Seminar in Quantitative Economics for providing such a supportive and stimulating environment in which to finish writing my dissertation.

I thank my parents, Charles Ehrlich and Susan Morris, my sister, Rebecca Ehrlich, my brother-in-law, Curtis Lacy, and all of my family and friends for their encouragement and support. I know it seemed like I might never finish school but you never told me so.

Finally, I thank Meredith Pedde for her love, companionship, and support during this process.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

ABSTRACT

Essays on Housing and Land Economics

by

Gabriel M. Ehrlich

Chair: Robert B. Barsky

This dissertation comprises three chapters examining the economics of the housing and land markets. In "A Model of Sales, Prices, and Liquidity in the Housing Market," I use a search and matching model of the housing market to address three main questions. First, what model of search and price determination best describes the housing market? Second, can a general equilibrium model generate the observed correlations between housing market variables? Third, what shocks have driven the historical behavior of the housing market? A model of competitive search is more likely than a model featuring random search and bargaining. Simulated data from the model qualitatively matches the correlations between housing market time series. Finally, the recent housing boom and bust were associated with an increase and subsequent decrease in the pool of eligible buyers, in addition to unrealized expectations of higher future productivity, the estimated size of which suggests a role for unwarranted optimism on the part of housing market participants.

In "Metropolitan Land Values and Housing Productivity," David Albouy and I present the first nationwide index of directly-measured land values and investigate

their relationship with housing prices. Construction prices and geographic and regulatory constraints increase the cost of housing relative to land. On average, one-third of housing costs are due to land, and the elasticity of substitution between land and other inputs is one-half. Conditional on input prices, housing productivity is low in larger cities. The increase in housing costs associated with greater regulation appears to outweigh any benefits from improved quality-of-life.

In “Price and Time to Sale Dynamics in the Housing Market: the Role of Incomplete Information,” I propose a model of the house-selling process in which sellers possess incomplete information regarding the state of the market. This model generates a negative correlation between house prices and time on market, a result that can persist even when realtors possess complete information. I construct an empirical measure of homeowner misperceptions regarding the state of the housing market, and show that sales volumes are negatively correlated with an increase in homeowners' perceptions of house prices relative to actual market conditions.

CHAPTER I

Introduction

This dissertation comprises three chapters, which study the economics of the housing and land markets from three different perspectives. The first chapter presents a general equilibrium macroeconomic model of the housing market and uses it to study several interesting features of the housing market time series data, as well as to infer what shocks have driven the behavior of the housing market over the past thirty years. The second chapter presents a locational equilibrium model for a system of cities, along with an index of land values for United States metropolitan areas constructed from a database of vacant land sales, which together provide estimates of the cost function for producing housing services. The third chapter presents a partial equilibrium model of the home selling process in which sellers possess incomplete information about the state of the housing market, and demonstrates that empirically, misperceptions about housing market conditions have economically meaningful effects on market outcomes. I summarize each chapter in more detail below.

In the first chapter, "A Model of Sales, Prices, and Liquidity in the Housing Market," I embed a search and matching model of the housing market into a dynamic stochastic general equilibrium framework and use it to address three main questions. First, what model of search and price determination best describes the housing market? Second, can a general equilibrium model generate the observed correlations between housing market variables? Third, what shocks have driven the historical behavior of the housing market, especially the recent boom and bust in house prices?

I calibrate the parameters that determine the steady state of the model and estimate the parameters of the shock process using the Metropolis-Hastings algorithm, a standard Bayesian method. The search and matching process in my model gives rise to a 'match surplus' between buyers and sellers when a match is formed: both parties can benefit from transacting immediately rather than continuing to search. I show that a model of price determination in which the division of the match surplus between the two parties depends on 'market tightness', or the ratio of buyers to sellers in the market, matches the data better than the more traditional model of price determination in which the division of the match surplus is fixed over time. Simulated data from the model qualitatively matches the correlations between key housing market time series, although quantitatively the fit is disappointing for some time series. Finally, the recent housing boom and bust were associated with a large increase and subsequent decrease in the pool of eligible buyers, in addition to expectations of higher future productivity that turned out not to be realized. However, the estimated size of the anticipated increase in productivity during the housing boom suggests a role for unwarranted optimism on the part of housing market participants.

In the second chapter, "Metropolitan Land Values and Housing Productivity," David Albouy and I present the first nationwide index of directly-measured land values by metropolitan area and investigate their relationship with housing prices. We show that construction prices and geographic and regulatory constraints increase the cost of housing relative to land. On average, approximately one-third of housing costs are due to land. The elasticity of substitution between land and other inputs into housing is about one-half, implying that the cost share of land ranges from 11 to 49 percent in our sample. Conditional on land and construction prices, housing productivity is relatively low in larger cities. Housing productivity is generally low on the East and West coasts and high in the interior of the country. The increase in housing costs associated with greater regulation appears to outweigh any benefits from improved quality-of-life.

In the third chapter, “Price and Time to Sale Dynamics in the Housing Market: the Role of Incomplete Information,” I propose a stylized model of the house-selling process in which sellers possess incomplete information regarding the state of the housing market. The model generates a negative correlation between house prices and time on the market as observed in the time series data. This result can persist in the presence of realtors with complete information as long as sellers’ and realtors’ incentives are imperfectly aligned. I construct empirical measures of homeowner perceptions of house prices at the state and metropolitan area levels using data from the American Community Survey. I compare these measures to indices of house prices based on market transactions to construct a ‘misperceptions index’ of house prices. Empirically, sales volumes are negatively correlated with an increase in homeowners’ perceptions of house prices relative to actual market conditions: homeowner misperceptions appear to explain between one fifth and one quarter of the variation in sales volumes.

CHAPTER II

A Model of Sales, Prices, and Liquidity in the Housing Market

2.1 Introduction

Historically, house prices have been positively correlated with sales volumes and starts, which have been negatively correlated with housing market liquidity and the vacancy rate. However, despite a growing literature examining the housing market through the lens of search and matching models, it remains unclear how well a dynamic general equilibrium model featuring search and matching frictions in the housing market can replicate these patterns. Ideally, a model that is able to replicate housing market dynamics successfully would also improve our inference concerning what shocks drive the housing market, and more specifically, what accounts for the recent boom and bust cycle in the U.S. housing market.

Figures 2.1 and 2.2 illustrate some of the key housing market time series using U.S. data. Figure 2.1 shows detrended GDP, house prices, and home sales, while Figure 2.2 shows housing starts, months' supply (the inventory of homes listed for sale divided by the number of sales in a month), and the homeowner vacancy rate over the period 1982Q3-2010Q4.[1] Several empirical regularities, summarized in Table 2.1,

[1] All series are in logs and have been linearly detrended and seasonally adjusted. Please see section 3.2, Observed Data Series, for details of the series' construction.

emerge from the figure. First, house prices, sales volumes, and vacancies are much more volatile than GDP, while housing starts and months' supply are more volatile still. Second, GDP, prices, sales, and starts all comove positively, while months' supply and vacancies comove negatively with sales and starts. Ideally, a theory of search and matching in the housing market would account for many of these empirical regularities.

A substantial literature has examined search frictions in the housing market both empirically and theoretically. Rosen and Topel (1988) find that time to sale has a large effect on new construction in the U.S. over the period 1963-1983. Following Wheaton's (1990) seminal model, a number of papers have taken a search and matching approach to the housing market. Recently, Diaz and Jerez (2009) and Head et al. (2010) have examined different models of search and price determination in this context. Aqeel (2009) and Magnus (2010) both illustrate the difficulty of matching the joint dynamics of key housing market time series such as prices, starts, and vacancies.

Identifying what shocks drive the housing market has been a topic of significant research interest recently, but no consensus has emerged. Some authors, such as Iacoviello and Neri (2010) and Wheaton and Nechayev (2008) offer fundamentals-based explanations. Others, such as Kahn (2008) and Lambertini et al. (2010) emphasize households' learning process concerning the economy. Finally, many authors, such as Case and Shiller (2003) and Piazzesi and Schneider (2009) argue that unrealistic expectations of future price appreciation were a key driver in the housing boom, implying that the recent boom and bust in prices was an irrational bubble.

Thus, there appear to be a number of open questions in the area of housing search. First, can a search and matching model of the housing market match the observed behavior of housing prices, sales, and construction over the business cycle? Second, what model of search and price determination best describes the housing market? Third, what types of shocks drive fluctuations in the housing sector?

I embed a search and matching model of the housing market into a DSGE framework and use it to address these questions. I consider both random search and bargaining and competitive search models, and compare the two statistically. I consider the role of several shocks in the model: housing productivity shocks, consumption productivity shocks, and "eligible buyer" shocks meant to represent changes in financing conditions. Crucially, I allow for "news shocks" to consumption productivity, in which agents anticipate changes in productivity ahead of time. I estimate the parameters of the shock process using the Metropolis-Hastings algorithm and the historical shocks hitting the housing market using the Kalman filter. Finally, I simulate data from the estimated model and compare it to the historical data.

My model replicates most of the key patterns in the U.S. housing market qualitatively. Price, sales, and starts are positively correlated, while prices comove strongly with GDP. Furthermore, starts are negatively correlated with months' supply of housing. Contrary to the data, starts are negatively correlated with GDP, although this correlation is not statistically significant. Data simulated from the competitive search model matches the historical correlations better than data from the random search model. The competitive search model performs better because the division of the match surplus depends on the state of the housing market: when more buyers enter the market, sellers receive a higher share of the surplus, increasing prices. Competitive search thus helps to generate the positive co-movement between prices and sales volumes that is prominent in the data.

I compare the random and competitive search models in terms of their posterior marginal densities. The higher likelihood of the competitive model constitutes "decisive evidence" in its favor according to the guidelines of DeJong and Dave (2007). Therefore, I use the competitive search model as my baseline specification, although I report estimation results for both models.

I estimate the historical shocks that hit the economy over the last thirty years using the Kalman filter. The estimates imply that the housing boom from 1997 to 2006 had two primary causes, a large increase in the fraction of eligible buyers, and expectations that productivity in the consumption sector would rise quickly in the future. These patterns reversed in the subsequent housing bust, as the fraction of eligible buyers fell and agents became more pessimistic about future productivity. Furthermore, the anticipated improvements to productivity turned out not to be realized over the period of the housing bust.

2.2 Model

I consider an economy in which households value a perishable consumption good and a durable housing good. Households invest the consumption good to create two types of physical capital, consumption capital and housing capital. Two representative, perfectly competitive firms rent capital and employ labor to produce the consumption and housing goods. Trade in the consumption good occurs on a frictionless spot market, but trade in the housing good is subject to search and matching frictions.

2.2.1 Households

Following Merz (1995), I assume there is a measure-one continuum of households, each comprising a 'large family' with a continuum of members. The members of the family pool income and consumption, but live in different dwellings[2]. Figure 2.3 illustrates the household structure graphically. At any given time some members of the household are satisfied with their dwelling, while others are dissatisfied. Household i

[2]This assumption is equivalent to assuming a complete markets allocation.

maximizes the objective function

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Big(\underbrace{(x\tilde{C}_{it}^{\lambda} + (1-x)\tilde{N}_{it}^{\lambda})^{\frac{1}{\lambda}}}_{\substack{\text{flow utility from}\\ \text{consumption}\\ \text{and housing}}} - \underbrace{\frac{\sigma}{2} E_{it}^2 \gamma_t \tilde{B}_{it}}_{\substack{\text{disutility}\\ \text{from}\\ \text{search}\\ \text{effort}}} - \underbrace{\rho \frac{\tilde{L}_{it}^{1+\frac{1}{\mu}}}{1+\frac{1}{\mu}}}_{\substack{\text{disutility}\\ \text{from}\\ \text{labor}\\ \text{effort}}} \Big) \tag{2.1}$$

where x is a preference parameters that governs the household's relative taste for consumption and housing, ρ is a parameter that governs the households distaste for labor, C_{it} is household i's consumption in period t, N_{it} is the fraction of family members who live in a well-matched house in period t ('Non-traders'), E_{it} is the household's search effort in period t, B_{it} is the fraction of family members who do not live in a well-matched house in period t ('Buyers'), and γ_t is the fraction of Buyers who are able to search in a given period ('Active Buyers'). λ is a parameter governing the elasticity of substitution between consumption and housing, and μ is the Frisch elasticity of labor supply. The basic idea of the households' flow utility function is that households value consumption and housing, but dislike labor and expending effort searching for housing; the disutility from search effort is proportional to the number of buyers but convex in the effort exerted per active buyer. There is population growth in the model; variables marked with a ˜ have been divided by the number of agents per household. In effect, the household's preferences are over per-agent variables rather than household totals.

The household maximizes its utility subject to the constraints:

$$
\begin{aligned}
C_{it} = W_t L_{it} + R_t^C K_{it}^C + R_t^H K_{it}^H - I_{it}^C - I_{it}^H + \Pi_{it} \\
+ \underbrace{q_t P_{it}^R S_{it}}_{\substack{\text{revenue} \\ \text{from} \\ \text{selling} \\ \text{houses}}} - \underbrace{P_t^W Y_{it}^H}_{\substack{\text{money} \\ \text{spent} \\ \text{on new} \\ \text{houses}}} - \underbrace{f_{it} P_{it}^R \gamma_t B_{it}}_{\substack{\text{money} \\ \text{spent on} \\ \text{existing} \\ \text{houses}}}
\end{aligned} \tag{2.2}
$$

$$
K_{it+1}^C = (1-\delta)K_{it}^C + I_t^C - \frac{\chi^C}{2}\Big(\frac{K_{it+1}^C - (1+g)K_{it}^C}{(1+g)K_{it}^C}\Big)^2 K_{it}^C \tag{2.3}
$$

$$
K_{it+1}^H = (1-\delta)K_{it}^H + I_t^H - \frac{\chi^H}{2}\Big(\frac{K_{it+1}^H - (1+g)K_{it}^H}{(1+g)K_{it}^H}\Big)^2 K_{it}^H \tag{2.4}
$$

Equation 2.2 is the household's budget constraint. There are two types of capital in the economy, consumption sector capital, K^C, and housing sector capital, K^H, which command rental rates R^C and R^H respectively. The households hold the economy's productive capital stock directly; investment in consumption capital is denoted I^C and investment in housing capital is denoted I^H. Π is profits from firms, which are owned by the households. P^W is the price at which the household buys newly built houses (the 'wholesale' price) and Y^H is the quantity of new houses purchased, so $P_{it}^W Y_{it}^H$ is the amount spent on newly built housing. P^R is the price of already existing housing (the 'retail' price), S is the number of existing houses the household markets for sale ('Sellers'), and $q(\theta_t, E_t)$ is the probability that a seller will meet a buyer given θ_t, the ratio of buyers to sellers in the housing market, and E_t, the average search effort expended by buyers. Later I will impose an assumption that a sale will occur any time a buyer meets a seller, so that $q(\theta_t, E_t)P_t^R S_{it}$ is the amount the household spends on already existing housing. Finally, $\gamma_t B_{it}$ is the number of buyers the household has on the market and $f(\theta_t, E_t, E_{it})$ is the probability that a buyer will meet a seller, so $f(\theta_t, E_t, E_{it})P_t^R \gamma_t B_{it}$ is the household's proceeds from selling existing houses. Taken together, equation 2.2 says that the household's consumption in period t equals its wage income plus its capital income and proceeds from selling its stock of existing houses, minus its investment, its spending on newly built houses, and its spending on existing houses.

Equations 2.3 and 2.4 are the accumulation equations for capital in the consumption and housing sectors. They feature quadratic costs of adjustment to the level of the capital stock, the severity of which is parameterized by χ^C and χ^H.

The household also faces equations of motion for its numbers of buyers, houses for sale, and satisfied homeowners. Because the equations of motion are specific to the process for trade I assume in the housing market, I defer these equations to the subsection describing the Housing Market. Therefore, I will also defer discussion of the first order conditions (FOCs) of the household's problem.

2.2.2 Firms

There are two perfectly competitive firms in the economy, one of which produces the consumption good and one of which produces houses. Both firms rent capital and hire labor from the households, and the construction firm also rents land from the households. The consumption firm produces output according to the production function:

$$Y_t^C = Z_t^C (K_t^C)^{\nu_C} (L_t^C)^{1-\nu_C} \tag{2.5}$$

where Y_t^C is the production of the consumption good in period t, Z_t^C is total factor productivity in the consumption sector in period t, K_t^C is the capital stock employed in the consumption sector in period t, and L_t^C is the amount of labor employed in the consumption sector in period t. The price of output in the consumption sector is normalized to one. Therefore the consumption firm's problem is

$$\max_{K^C,L^C} Z_t^C (K_t^C)^{\nu_C} (L_t^C)^{1-\nu_C} - R_t^C K_t^C - W_t L_t^C \tag{2.6}$$

The FOCs of this problem are

$$R_t^C = \nu_C Z_t^C (K_t^C)^{\nu_C - 1} (L_t^C)^{1-\nu_C} \tag{2.7}$$

$$W_t = (1-\nu_C) Z_t^C (K_t^C)^{\nu_C} (L_t^C)^{-\nu_C} \tag{2.8}$$

The housing firm produces houses according to the production function:

$$Y_t^H = Z_t^H (K_t^H)^{\nu_H} (L_t^H)^{\rho_H} \tag{2.9}$$

where Y_t^H is the production of houses in period t, Z_t^H is total factor productivity in the housing sector in period t, K_t^H is the capital stock employed in the housing sector in period t, and L_t^H is the amount of labor employed in the housing sector in period t. I assume labor is freely mobile, so that the wage is the same across sectors. The housing firm's problem is

$$\max_{K^H,L^H} P_t^W Z_t^H (K_t^H)^{\nu_H} (L_t^H)^{\rho_H} - R_t^H K_t^H - W_t L_t^H \tag{2.10}$$

where, as noted above, P^W is the 'wholesale' price at which the housing firm sells newly built houses to the households as unmatched units. The FOCs of the housing firm's problem are

$$R_t^H = \nu_H P_t^W Z_t^H (K_t^H)^{\nu_H - 1} (L_t^H)^{\rho_H} \tag{2.11}$$

$$W_t = \rho_H P_t^W Z_t^H (K_t^H)^{\nu_H} (L_t^H)^{\rho_H - 1} \tag{2.12}$$

$$\tag{2.13}$$

In my calibration, I assume decreasing returns to scale in the housing sector. This assumption is meant to capture the idea that land acts as a fixed factor in the production of houses without explicitly modeling the details of land supply and demand. With no trend growth in productivity in the housing sector, the assumption of decreasing returns would lead to house prices steadily increasing with population. Therefore, I assume that Z_t^H grows at rate $(1+g)^{1-\nu_H-\rho_H}$ over time. This assumption implies that the price of housing is constant on the balanced growth path.

2.2.3 The Housing Market

The housing market of this model exhibits search frictions. Buyers and sellers in the market for well-matched houses cannot transact on a frictionless spot market but instead form pairs according to a matching function that relates the numbers of active buyers and sellers to the number of successful matches. I assume that when buyers and sellers form a match, the seller's house is always a good match for the buyer. The implicit concept is that housing units and household preferences are heterogeneous, so that many houses a buyer visits will be ill-suited to their taste. A "match" between a buyer and a seller occurs when a buyer finds an appropriate house. The matching function is a reduced form way to capture the time consuming nature of search and matching in the housing market without specifying the microeconomic process by which matches are formed.

I consider two different matching functions in the model. The first is the Cobb-Douglas matching function that is standard in much of the literature (e.g., Wheaton 1990):

$$M(E_t, \gamma_t B_t, S_t) = A(E_t \gamma_t B_t)^{\phi} S_t^{1-\phi} \tag{2.14}$$

The second is the generalized urn-ball matching function of Head et al. (2010):

$$M(E_t, \gamma_t B_t, S_t) = AS_t(1 - e^{-\zeta \frac{E_t \gamma_t B_t}{S_t}}) \tag{2.15}$$

where M is the number of matches in a period. As is standard, I denote the 'tightness' of the housing market as

$$\theta_t = \frac{\gamma_t B_t}{S_t} \tag{2.16}$$

For convenience, I define the probability that a buyer exerting effort E_{it} meets a seller in a given period as

$$f(\theta, E_t, E_{it}) = \frac{E_{it} M(E_t, B_t, S_t)}{E_t \gamma_t B_t} = \begin{cases} AE_{it}(E_t\theta_t)^{\phi-1} & \text{w/Cobb-Douglas matching} \\ \frac{E_{it}A}{E_t\theta_t}(1 - e^{-E_t\theta_t\zeta}) & \text{w/ urn-ball matching} \end{cases} \tag{2.17}$$

At the cost of anticipating the equilibrium concept, in a symmetric equilibrium, in which all households choose the same search effort, $E_{it} = E_t$, the probability that a buyer will meet a seller will be

$$f(\theta_t, E_t) = f(\theta_t, E_t, E_t) = \frac{E_t M(E_t, B_t, S_t, \gamma_t)}{E_t \gamma_t B_t} = \begin{cases} AE_t^{\phi}\theta_t^{\phi-1} & \text{w/Cobb-Douglas matching} \\ \frac{A}{\theta_t}(1 - e^{-E_t\theta_t\zeta}) & \text{w/ urn-ball matching} \end{cases} \tag{2.18}$$

Similarly, I define the probability that a seller meets a buyer as

$$q(\theta_t, E_t) = \frac{M(E_t, B_t, S_t)}{S} = \begin{cases} A(E_t\theta_t)^{\phi} & \text{w/Cobb-Douglas matching} \\ A(1 - e^{-E_t\theta_t\zeta}) & \text{w/ urn-ball matching} \end{cases} \tag{2.19}$$

In the remainder of the text, I will suppress the dependence of f and q on their arguments for simplicity.

In the model, individuals remain in their current house so long as it remains 'well-matched', but each period well-matched homeowners face probability α of becoming mismatched. Once an individual becomes mismatched, the parent household no longer receives any utility from owning the house. Therefore, the parent household immediately puts the house for sale, and the individual begins looking for a new house to purchase. An individual who becomes mismatched therefore becomes simultaneously a seller and a potential buyer of housing. I assume that individuals who do not own a well-matched house live with other members of their household. The population of each household grows each period at rate g, with the new members being born as poorly-matched households. Finally, I assume that only a fraction γ_t of poorly matched individuals are able to search for a home each period. Implicitly, one could imagine financing or other constraints that prevent some households from searching in a given period.

The timing of each period is as follows:

1. Period starts.
2. Aggregate shocks realized.
3. Household receives flow utility from satisfied homeowners.
4. Housing market occurs
   - Matches formed
   - Bargaining and sales occur
5. Relocation shock hits
6. Production of goods and new houses
7. Housing company sells homeowners 'unmatched' new houses
8. Capital depreciates
9. Factors are paid and consumption occurs
10. New agents born
11. Period ends

Because matching is costly and time consuming, there is a surplus value associated with each match, which I will define in the section on Price Determination. I assume that buyers and sellers always exploit potential gains from trade, so that every match results in the sale of a house.

These assumptions give rise to the equations of motion for buyers (B_t), houses for sale (S_t), and non-traders (N_t):

$$B_{it+1} = B_{it} - \underbrace{(1-\alpha) f_{it} \gamma_t B_{it}}_{\substack{\text{buyers who} \\ \text{become} \\ \text{non-traders}}} + \alpha N_{it} + \underbrace{g(B_{it} + N_{it})}_{\substack{\text{population} \\ \text{growth}}} \tag{2.20}$$

$$S_{it+1} = (1 - q_{it}) S_{it} + \underbrace{\alpha(N_{it} + f_{it}\gamma_t B_{it})}_{\substack{\text{agents hit by move} \\ \text{shock}}} + Y^H_{it-2} \tag{2.21}$$

$$N_{it+1} = (1-\alpha)(N_{it} + f_{it}\gamma_t B_{it}) \tag{2.22}$$

The first equation says that the number of buyers next period equals the number this period, less the buyers who purchase a house and do not receive a relocation shock, plus non-traders who receive a relocation shock and new agents. The second equation says that the number of homes for sale next period equals the number this period minus those sold, plus the homes for sale posted by newly mis-matched homeowners and new construction from two periods ago. The lag of two periods represents a time to build in the construction sector of approximately three quarters. The third equation says that the number of satisfied homeowners equals the number from last period, less those hit by a relocation shock, plus the number of buyers who successfully purchase a well-matched home.

2.2.4 Recursive Formulation of the Household's Problem

Given this structure for trade in the housing market, the household's problem can be re-formulated as a recursive problem with the following Bellman equation, where I have detrended the budget constraint and equations of motion to adjust for population growth:

$$\begin{aligned}
V(\tilde{B}_{it}, \tilde{S}_{it}, \tilde{N}_{it}, \tilde{K^C}_{it}, \tilde{K^H}_{it}) = \max_{\tilde{L}_{it}, E_{it}, \tilde{K}^C_{it+1}, \tilde{K}^H_{it+1}, \tilde{Y}^H_{it}} &\Big\{ (x\tilde{C}_{it}^{\lambda} + (1-x)\tilde{N}_{it}^{\lambda})^{\frac{1}{\lambda}} - \frac{\sigma}{2} E_{it}^2 \gamma_t \tilde{B}_{it} \\
&- \rho \frac{\tilde{L}_{it}^{1+\frac{1}{\mu}}}{1+\frac{1}{\mu}} + \beta \mathbb{E}_t \big[V(\tilde{B}_{it+1}, \tilde{S}_{it+1}, \tilde{N}_{it+1}, \tilde{K^C}_{it+1}, \tilde{K^H}_{it+1}) \big] \Big\}
\end{aligned}$$

s.t.

$$\begin{aligned}
\tilde{C}_{it} &= W_t \tilde{L}_{it} + R_t^C \tilde{K^C}_{it} + R_t^H \tilde{K^H}_{it} - \tilde{I^C}_{it} - \tilde{I^H}_{it} + \Pi_{it} \\
&\quad + q_t P_{it}^R \tilde{S}_{it} - P_t^W \tilde{Y^H}_{it} - f_{it} P_{it}^R \gamma_t \tilde{B}_{it} \\
\tilde{K^C}_{it+1} &= (1-\delta)\tilde{K^C}_{it} + \tilde{I^C}_{it} \\
\tilde{K^H}_{it+1} &= (1-\delta)\tilde{K^H}_{it} + \tilde{I^H}_{it} \\
(1+g)\tilde{B}_{it+1} &= \tilde{B}_{it} - (1-\alpha) f_{it} \gamma_t \tilde{B}_{it} + \alpha \tilde{N}_{it} + g \\
(1+g)\tilde{S}_{it+1} &= (1-q_t)\tilde{S}_{it} + \alpha(\tilde{N}_{it} + f_{it}\gamma_t \tilde{B}_{it}) + \tilde{Y^H}_{it-2} \\
(1+g)\tilde{N}_{it+1} &= (1-\alpha)(\tilde{N}_{it} + f_{it}\gamma_t \tilde{B}_{it})
\end{aligned}$$

For convenience, I define the marginal values to the household of having an additional buyer (V^B), seller (V^S), and non-trader (V^N):

$$V_{it}^{B} = -\underbrace{U_{\tilde{C},it} P_t^R f_{it}\gamma_t}_{\substack{\text{purchase}\\ \text{prob. times}\\ \text{price in}\\ \text{utility}}} - \underbrace{\frac{\sigma}{2} E_{it}^2 \gamma_t}_{\substack{\text{disutility}\\ \text{of}\\ \text{search}\\ \text{effort}}} + \beta \frac{1}{1+g} \mathbb{E}_t\Big[\big(1-(1-\alpha) f_{it}\gamma_t\big) V_{it+1}^{B} + \alpha f_{it}\gamma_t V_{it+1}^{S} + (1-\alpha) f_{it}\gamma_t V_{it+1}^{N}\Big] \tag{2.23}$$

$$V_{it}^{S} = \underbrace{U_{\tilde{C},it} P_t^R q_t}_{\substack{\text{sale prob.}\\ \text{times price}\\ \text{in utility}}} + (1-q_t)\beta \frac{1}{1+g} \mathbb{E}_t\big[V_{it}^{S}\big] \tag{2.24}$$

$$V_{it}^{N} = U_{\tilde{N},it} + \beta \frac{1}{1+g} \mathbb{E}_t\big[\alpha(V_{it+1}^{B} + V_{it+1}^{S}) + (1-\alpha) V_{it+1}^{N}\big] \tag{2.25}$$

where $U_{\tilde{C},it}$ denotes the marginal utility of consumption and $U_{\tilde{N},it}$ denotes the marginal flow utility of a well-matched house in period t. The household's Bellman equation features five control variables: K_{it+1}^{C}, K_{it+1}^{H}, Y_{it}^{H}, E_{it}, and $\tilde{L}_{it}$. The FOCs for these variables are:

$$\Big(1+\chi^C\frac{\tilde{K^C}_{it+1}-\tilde{K^C}_{it}}{\tilde{K^C}_{it}}\Big)U_{\tilde{C},it} = \beta\mathbb{E}_t\Big[U_{\tilde{C},it+1}\Big((1+R^C_{t+1}-\delta)+\frac{\chi^C}{2}\Big(\frac{\tilde{K^C}_{it+2}-\tilde{K^C}_{it+1}}{\tilde{K^C}_{it+1}}\Big)^2 + \chi^C\frac{\tilde{K^C}_{it+2}-\tilde{K^C}_{it+1}}{\tilde{K^C}_{it+1}}\Big)\Big] \tag{2.26}$$

$$\Big(1+\chi^H\frac{\tilde{K^H}_{it+1}-\tilde{K^H}_{it}}{\tilde{K^H}_{it}}\Big)U_{\tilde{C},it} = \beta\mathbb{E}_t\Big[U_{\tilde{C},it+1}\Big((1+R^H_{t+1}-\delta)+\frac{\chi^H}{2}\Big(\frac{\tilde{K^H}_{it+2}-\tilde{K^H}_{it+1}}{\tilde{K^H}_{it+1}}\Big)^2 + \chi^H\frac{\tilde{K^H}_{it+2}-\tilde{K^H}_{it+1}}{\tilde{K^H}_{it+1}}\Big)\Big] \tag{2.27}$$

$$U_{\tilde{C},it}P^W_t = \beta^3\mathbb{E}_t[V^S_{it+3}] \tag{2.28}$$

$$\sigma E_{it} = \frac{\partial f(\theta, E_t, E_{it})}{\partial E_{it}}\Big(-U_{\tilde{C},it}P^R_t+\beta\frac{1}{1+g}\times \mathbb{E}_t\big[(1-\alpha)(V^N_{it+1}-V^B_{it+1})\alpha V^S_{t+1}\big]\Big) \tag{2.29}$$

$$\tilde{L}_{it} = \big(U_{\tilde{C},it}\frac{W_{it}}{\rho}\big)^\mu \tag{2.30}$$

Equations (2.26) and (2.27) are the household's Euler equations for consumption and housing capital, and reflect the quadratic costs of adjustment to both capital stocks. Equation (2.28) states that the wholesale price at which the household purchases unmatched houses from the construction firm, times the marginal utility of consumption, equals the discounted value of having an additional house for sale in three quarters. Equation (2.29) specifies the household's optimal search effort, at which the marginal disutility of additional search equals the marginal improvement in the probability of forming a match times the buyer's surplus from forming a match.

In a symmetric equilibrium in which all households choose the same search effort, equation (2.29), simplifies to

$$E_t^{2-\phi} = \frac{A(\theta_t)^{\phi-1}}{\sigma}\Big(-U_{\tilde{C},it}P^R_t+\beta\frac{1}{1+g}\mathbb{E}_t\big[(1-\alpha)(V^N_{t+1}-V^B_{t+1})+\alpha V^S_{t+1}\big]\Big) \tag{2.31}$$

in the case of Cobb-Douglas matching, and to

$$E_t^2 = \frac{A(1-e^{-E_t\theta_t\zeta})}{\sigma\theta_t}\Big(- U_{\tilde{C},it}P_t^R + \beta\frac{1}{1+g}\mathbb{E}_t\big[(1-\alpha)(V_{t+1}^N - V_{t+1}^B) + \alpha V_{t+1}^S\big]\Big) \tag{2.32}$$

in the case of urn-ball matching.

2.2.5 Price Determination

The buyer's surplus from purchasing a house at price P_t^R is the difference in utility from buying the house versus continuing to search, less the price of the house in utility terms:

$$\text{Buyer's Surplus} = \beta\frac{1}{1+g}\mathbb{E}_t\big[(1-\alpha)V_{t+1}^N + \alpha V_{t+1}^S - V_{t+1}^B\big] - U_{\tilde{C},it}P_t^R \tag{2.33}$$

For the seller, selling a house at price P_t^R gives a payoff in utility terms of $U_{\tilde{C},it}P_t^R$, while the payoff from not selling is the continuation value of keeping the house on the market, $\beta\frac{1}{1+g}\mathbb{E}_t\big[V_{t+1}^S\big]$. The seller's surplus is the difference:

$$\text{Seller's Surplus} = U_{\tilde{C},it}P_t^R - \beta\frac{1}{1+g}\mathbb{E}_t\big[V_{t+1}^S\big] \tag{2.34}$$

The match surplus is the total surplus to both parties from completing the transaction rather than parting ways, so it is the sum of the buyer's surplus and the seller's surplus:

$$\text{Match Surplus} = \beta\frac{1}{1+g}\mathbb{E}_t\big[(1-\alpha)(V_{t+1}^N - V_{t+1}^B) + \alpha(V_{t+1}^S - V_{t+1}^B) - V_{t+1}^S\big] \tag{2.35}$$

The purchase price of the house divides the match surplus between the buyer and the seller. I will denote the share of the match surplus accruing to the buyer as η_t. Setting the buyer's surplus equal to η_t times the total match surplus gives the following rule for price determination:

$$P_t^R = \frac{1}{U_{\tilde{C},it}} \beta \frac{1}{1+g} \mathbb{E}_t \Big[(1-\eta_t) \Big((1-\alpha) V_{t+1}^N - V_{t+1}^B + \alpha V_{t+1}^S \Big) + \eta_t V_{t+1}^S \Big] \tag{2.36}$$

Traditionally in the housing search and matching literature, it is assumed that the buyer receives a fixed share of the match surplus: $\eta_t = \eta$. This sharing rule for the surplus is motivated as the result of an asymmetric Nash bargain between the buyer and seller. I will call the model with this sharing rule the "random search" model.

More recently, Diaz and Jerez (2009) and Head et al. (2010) have explored search and matching models of the housing market under the competitive search framework introduced by Moen (1997) in the context of labor search. In the competitive search environment, sellers post list prices for their houses, and can credibly commit not to bargain with buyers over the price after a match forms. Moen shows that in a competitive search equilibrium, the share of the match surplus going to each party equals the elasticity of the matching function with respect to that party's side of the market. In such an environment, when the housing market "heats up", so that there are more buyers relative to sellers, the share of the match surplus going to the sellers will rise, while when the market "cools down", the share of the match surplus going to buyers will rise. Therefore, the competitive search framework has the potential to add volatility to house prices relative to the random search framework, in which the share of the match surplus accruing to the buyer remains constant.

Diaz and Jerez and Head et al. show that in the housing market, a competitive search equilibrium can be implemented similarly to a standard search equilibrium by adding an equation to endogenize η_t, the share of the match surplus going to the buyer, to be equal to the elasticity of the matching function with respect to buyers (in the case of my model, this will be effective buyers, or effort times buyers):

$$\eta_t = \frac{\partial M(E_t B_t, S_t)}{\partial E_t B_t} \frac{E_t B_t}{M(E_t B_t, S_t)} \tag{2.37}$$

With a Cobb-Douglas matching function, the elasticity of the matching function with respect to both sides of the market is constant. In this case, the competitive search model is equivalent to the random search model if the buyer's share of the match surplus is equal to the elasticity of the matching function with respect to buyers, i.e. if $\eta = \phi$. This equality, known as the Hosios condition after Hosios (1990), is a necessary condition for efficiency in the random search model. Therefore, as in much of the literature, I impose this condition in my calibration, implying that the two methods of determining the sharing rule are equivalent with the Cobb-Douglas matching function.

To allow the competitive search model to generate different results than the random search model, I follow Diaz and Jerez and Head et al. in using the urn-ball matching function described above. In this case, which I will call the "competitive search" model, the buyer's share of the match surplus equals the elasticity of the matching function with respect to search effort times active buyers:

$$\eta_t = \frac{\partial M(E_t \gamma_t B_t, S_t)}{\partial E_t \gamma_t B_t} \frac{E_t \gamma_t B_t}{M(E_t \gamma_t B_t, S_t)} = \frac{E_t \theta_t \zeta}{e^{E_t \theta_t \zeta} - 1} \tag{2.38}$$

2.2.6 Sources of Stochastic Variation

I include four aggregate shocks in the model: anticipated and unanticipated shocks to productivity in the consumption sector, an unanticipated shock to productivity in the housing sector, and an unanticipated shock to the fraction of buyers who are eligible to search. I assume Z_t^C, Z_t^H, and γ_t follow the $AR(1)$ processes:

$$\ln(Z_t^C) = \psi_C \ln(Z_{t-1}^C) + (1-\psi_C)\ln(\overline{Z^C}) + \epsilon_t^C + \epsilon_{t-20}^A \tag{2.39}$$

$$\epsilon_t^C \sim N(0, \sigma_C^2) \tag{2.40}$$

$$\epsilon_t^A \sim N(0, \sigma_A^2) \tag{2.41}$$

$$\ln(Z_t^H) = \psi_H \ln(Z_{t-1}^H) + (1-\psi_H)\ln(\overline{Z^H}) + \epsilon_t^H \tag{2.42}$$

$$\epsilon_t^H \sim N(0, \sigma_H^2) \tag{2.43}$$

$$\ln(\gamma_t) = \psi_\gamma \ln(\gamma_{t-1}) + (1-\psi_\gamma)\ln(\overline{\gamma}) + \epsilon_t^\gamma \tag{2.44}$$

$$\epsilon_t^\gamma \sim N(0, \sigma_\gamma^2) \tag{2.45}$$

where bars above variables represent their steady state values. I assume the shocks are independently distributed.

I will call the shock ϵ_t^A an anticipation shock, as it represents an anticipated movement in the future level of consumption productivity. Because I take the model period to be one quarter, anticipation shocks concern productivity changes 5 years in the future. These anticipation shocks are on average "correct" in the sense that the actual technology level in the housing sector equals the anticipated level in expectation. However, the presence of a contemporaneous or unanticipated shock to housing technology allows me to study the case of an unrealized expectation of a change in future productivity, which would correspond to ϵ_t^C being exactly equal to $-\epsilon_{t-20}^A$. This scenario is in the spirit of Beaudry and Portier's (2004) "Pigou Cycles".

2.2.7 Equilibrium

I define an equilibrium of the model as follows:

Definition A *symmetric recursive search equilibrium* of this model is a set of policy functions for the households and firms, equations of motion for the stocks of capital, buyers, sellers, and non-traders, and prices for factors of production, new houses, and existing houses, such that:

1. Households maximize their utility taking factor prices and the price of new houses as given;

2. Firms maximize their profits taking all prices as given;

3. The consumption good, new housing, and factor markets clear;

4. Every household chooses the same search effort;

5. The number of sales in of existing houses is consistent with the matching function; and

6. The price of existing houses is determined according to the appropriate sharing rule.

Please see Appendix 2.6.1 for a complete list of equations characterizing equilibrium both in the random search model and in the competitive search model.

2.3 Empirics

2.3.1 State Space Representation

The model above can be linearized around its steady state equilibrium to give the following state space representation:

$$Z_t = \overline{Z}(\Theta) + \mathbf{B}(\Theta) Z_{t-1} + \mathbf{G}(\Theta) e_t \tag{2.46}$$

$$Y_t = \mathbf{H} Z_t \tag{2.47}$$

$$E[e_t e_t'] = \mathbf{V}(\Theta) \tag{2.48}$$

In the transition equation, (2.46), Θ is a vector of the structural parameters of the model, $\overline{Z}(\Theta)$ is a vector of the steady state values of the model variables (which are functions of Θ), and Z_t is a vector of the deviations of the model variables from their steady state values. $\mathbf{B}(\Theta)$ is a system matrix that relates this period's deviations from steady state to last period's. Finally, $\mathbf{G}(\Theta)$ is a policy function matrix and e_t is the vector of structural shocks to the economy.

In the observation equation, (2.47), Y_t is a vector of the observable variables I will use to estimate the model. $\mathbf{H}$ is a matrix of ones and zeros that selects the variables to be observed. In equation (2.48), $\mathbf{V}(\Theta)$ is the variance-covariance matrix of the shock process; I impose that the shocks are i.i.d., so $\mathbf{V}(\Theta)$ is a diagonal matrix.

Together, equations (2.46) and (2.47) form a system of Kalman filter equations. I use the Kalman filter recursions to evaluate the log-likelihood of the model conditional on the structural parameters Θ and the observed data series Y_t.

2.3.2 Observed Data Series

I use GDP, the price of existing homes, sales volumes, and starts as my observable data series. Because I have four shocks in the model, using four observable data series allows me to avoid using observation errors in the observation equation (2.47). The model makes different predictions for how these series will react to each of the four shocks in the model, so using these series in the estimation should allow for successful identification of the parameters of the shock process and the historical shocks.

I use the following procedure to match the data series I am using as closely as possible to the conceptual variables in my model. For all variables that are available on a monthly basis, I take simple averages to construct quarterly values. Next I convert all nominal variables to 2010 dollars using the CPI-U. Because the unit of analysis in my model is the household, I construct the GDP, Sales, and Starts series on a per household basis. To calculate the number of households, I divide the quarterly total population, which I construct as the average of the monthly population over the quarter, by the average household size provided by the Census. Average household size is only provided annually, so I use a cubic spline to interpolate quarterly values. I then take logs of each series and regress the log values on a linear time trend, and for the not seasonally adjusted series, a set of quarter dummies. Finally I add the deviations from the linear time trend to the average value for each series over the sample period, 1982q3-2010q4.

The house price series I use is the CoreLogic Single Family Detached House Price Index (HPI). Because the CoreLogic HPI is not expressed in dollar terms, I normalize it to \$194,592, the value of the FHFA U.S. single family detached HPI, in 2000q1. For my sales volume series, I take the sum of new single family houses sold from the Census Bureau and existing single family houses sold from the National Association of Realtors (NAR). Using the total of new and existing home sales is conceptually appropriate because in the model, all new houses are immediately sold on a frictionless spot market to the households, who then market them for sale on the frictional housing market along with previously built homes. For my starts series, I use single family starts from the Census. I take the GDP series from the Bureau of Economic Analysis. The resulting series are shown in Figures 2.1 and 2.2.[3]

2.3.3 Calibrated Parameters and Steady State

I calibrate the parameters that determine the steady state of the model, which is equivalent to imposing a degenerate prior distribution for their values in the estimation procedure. Table 2.2 shows the calibrated parameters. I take the model period to be one quarter. The only parameters that differ between the random and compet-

[3]Figure 2.2 shows two additional series that I do not use in the estimation procedure, months' supply and the vacancy rate. For my months' supply series, I add the inventory of existing single family homes published by the NAR to the number of newly constructed homes for sale published by the Census, and divide by the sum of existing single family home sales reported by the NAR and new single family houses sold reported by the Census. For the vacancy rate, I use the homeowners' vacancy rate from the Census Bureau's Housing Vacancy Survey.

itive search models concern the matching function and disutility of search effort. For these parameters, I calibrate the random search model to have the same steady state values for the buyer's share of the match surplus, η, search effort, E, and months' supply of housing as the random search model.

I also impose unit roots in the technology shocks in the model. I have experimented with estimating the persistence of these shocks; the results do not change appreciably. I also calibrate the level of capital adjustment costs in both sectors, χ_C and χ_H, to be zero[4].

One key parameter that affects the model results is the elasticity of substitution between housing and the consumption good in the utility function. Unfortunately, there is not a consensus in the literature regarding the value of this parameter. Typically, cross-sectional studies such as Rappaport (2008), find values for this parameter of approximately 0.5. In contrast, time series studies such as Rupert et al. (1995) typically find values of 1 or greater. I calibrate the elasticity of substitution to be 0.5 in my baseline case, but I discuss briefly how the results differ when the elasticity is set to 1 in the Analysis section.

Some key steady state values implied by these parameters are described in Table 2.3. Some steady state values bear discussion because they deviate from values in the data. The steady state proportion of the labor force, $\overline{L^H}$, is too low at 1.6%. This is chiefly because I calibrate my model using only single family home construction, whereas the set of all construction workers includes those working in multi-family and non-residential construction. Similarly, the number of sales per household is 1.1% per quarter. This implies an unrealistically long period between moves. This discrepancy results from the exclusion of multi-family dwellings from the sales figures I use to calibrate the model, as well as the absence of renters from the model.

[4] I may re-introduce capital adjustment costs into the estimation procedure in the future

2.4 Analysis

2.4.1 Estimation Results

I use the random walk Metropolis Hastings Algorithm to estimate the standard deviations of the shocks and the persistence of the eligible buyers shock. Table 2.4 displays the prior distributions I specify for these parameters.

I run the sampler for 40,000 iterations and drop the first 45% before conducting posterior simulations. Figure 2.4 illustrates the prior (in gray) and posterior (in black) distributions for each of the parameters to be estimated. The posterior means and standard deviations of the estimated parameters are listed in Table 2.5, and are illustrated in Figure 2.4. The posterior distributions for most parameters are similar for the random and competitive search models. In both models, the parameters are tightly identified by the estimation procedure. Furthermore, the priors I specify do not appear to constrain the posterior distributions I obtain.

The Laplace approximation of the log marginal density is 745.7 for the competitive search model and 715.5 for the random search model. According to DeJong and Dave, the implied posterior odds ratio constitutes "decisive evidence" in favor of the competitive search model.

2.4.2 Impulse Responses

I linearize the system of equations around the steady state to find impulse responses to the shocks in the model. Figures 2.5 through 2.8 show the impulse response functions for each of the shocks, which are normalized to be one standard deviation in size. The plotted values in all impulse responses are proportional deviations from the variable's steady state (i.e. a value of 0.01 is a 1% deviation from the steady state value). The time period covered is 240 quarters.

In response to an anticipated increase in consumption productivity, GDP falls slightly on impact while construction and prices rise and sales rise a bit. When the shock is realized, construction falls sharply and GDP rises[5]. Prices remain at their new, higher level after the productivity level rises, while construction falls sharply as the cost of producing housing relative to the consumption good jumps. Construction levels eventually settle at very slightly above their original steady state levels.

In response to an unanticipated consumption productivity shock, GDP, prices, and sales all rise on impact. Construction exhibits a hump-shaped response, while months' supply initially falls before rising above its steady state level.

The random and competitive search models respond differently to the eligible buyers shock. In both cases sales rise sharply on impact and the market becomes much tighter, as represented by lower months' supply. In the random search model the fixed sharing rule for dividing the match surplus mutes the effect on prices, and therefore on construction. In the competitive search model, the tighter market gives the sellers greater bargaining power and a greater share of the match surplus, so the rise in prices is much more pronounced. The higher prices cause construction to rise sharply as well. Again, GDP is essentially flat in response to the shock.

In response to a housing productivity shock, GDP rises by a small amount, construction rises and house prices fall. Sales and months' supply both rise. The impulse responses are quite similar between the random and competitive search models.

2.4.3 Estimated Shocks

Figures 2.9 and 2.10 illustrates the smoothed shocks from the estimation procedure for the competitive and random search models. Here, I will focus on the shocks from the competitive search model. Several patterns in emerge from the figure. First, there were a series of positive shocks to anticipated consumption productivity in the early 2000s. From 2001q1 to 2004q4, the anticipated productivity shock was

[5] The technology shocks in the model have a unit root, so the shocks will generally have permanent effects.

positive in all quarters but one, for a cumulative increase of 41%. In the ensuing years this pattern reverses sharply, with overwhelmingly negative shocks to anticipated consumption productivity. Furthermore, in the period 2006q1-2009q4, a series of negative shocks to unanticipated productivity almost perfectly counteracts the earlier positive shocks to anticipated productivity (in fact, the cumulative size of these shocks is -48%, larger than the positive productivity shocks). There is also a large series of positive shocks to the fraction of eligible buyers beginning in the mid-1990s and intensifying in the 2000s, which reverses sharply in 2007. Finally, a number of negative shocks to housing productivity in the 2000s contributed to rising prices during the housing boom. Figure 2.11 illustrates the historical decomposition of the change in house prices over the sample period into the changes due to each shock, while table 2.6 displays the conditional variance decomposition of the four series used in the estimation procedure at two time horizons, 4 quarters and 40 quarters.

The anticipated shocks to consumption productivity during the housing boom are quite large. However, when I set the elasticity of substitution between housing and the consumption good to one, to reflect Cobb-Douglas utility, the estimated anticipation shocks become even larger, around 70 percent in the boom period. Furthermore, reducing the Frisch elasticity of labor supply to one-quarter from one does not reduce the size of the estimated anticipation shocks substantially. While there are many interpretations of the large size of the estimated anticipation shocks during the housing boom, my preferred interpretation is that the mechanisms that generate variation in house prices in this paper omit important aspects of reality. Specifically, I believe the estimation results suggest that unwarranted optimism regarding future market conditions on the part of participants in the housing market was an important part of the recent housing boom and bust.

2.4.4 Simulations

Because the model features unit root shocks, I use Monte Carlo simulations to assess the model's dynamic behavior. Tables 2.7 and 2.8 show the results of simulating the competitive and random search models using the mean estimated parameter values. For each model, I simulate 114 quarters of data (the same number as in my observed sample) 500 times. The tables show the same statistics for the simulated data that Table 2.1 shows for the actual data. Therefore, comparison with Table 2.1 should help in evaluating the model's empirical performance.

Qualitatively, the model is able to generate most of the observed correlations between key housing market variables. Prices, sales, and starts are positively correlated, and prices are comove positively with GDP. Starts are negatively correlated with months' supply and the vacancy rate. However, there is a slight negative correlation between starts and GDP, contrary to what is observed in the data. The competitive search model does a better job matching the observed correlations than the random search model, although neither model is able to match the strength of the observed correlations between prices, sales, and starts quantitatively.

2.5 Conclusion

This paper presents a search and matching model of the housing market embedded in a DSGE framework. Conditional on the observed data, a model with competitive search is more likely than a model with random search and bargaining. The model reproduces many of the stylized facts of the housing market, most notably the positive co-movement of prices, sales, and starts, and the negative co-movement of starts and months' supply. The estimation results imply that the recent housing boom was driven by a large increase in the fraction of eligible buyers and anticipated increases in future productivity in the consumption sector, while the ensuing bust was caused by a sharp reversal of these trends in conjunction with a series of unanticipated negative shocks to consumption productivity. I interpret this pattern as suggesting that the housing boom and bust were driven in part by expectations of above trend productivity growth that later turned out to be unfounded; however, the large size of these shocks during the housing boom suggests a role for irrational or unwarranted optimism on the part of housing market participants in this period.

2.6 Appendix

2.6.1 Equilibrium Equations

The following system of equations characterizes a recursive Nash equilibrium of this model (the i subscripts have been dropped because households are identical):

$$(1+g)\tilde{K}^C_{t+1} = (1-\delta)\tilde{K}^C_t + \tilde{I}^C_t - \frac{\chi^C}{2}\Big(\frac{\tilde{K}^C_{t+1}-\tilde{K}^C_t}{\tilde{K}^C_t}\Big)^2\tilde{K}^C_t \tag{2.49}$$

$$(1+g)\tilde{K}^H_{t+1} = (1-\delta)\tilde{K}^H_t + \tilde{I}^H_t - \frac{\chi^H}{2}\Big(\frac{\tilde{K}^H_{t+1}-\tilde{K}^H_t}{\tilde{K}^H_t}\Big)^2\tilde{K}^H_t \tag{2.50}$$

$$(1+g)\tilde{B}_{t+1} = B_t - (1-\alpha)f(\theta_t,E_t)\gamma_t\tilde{B}_t + \alpha\tilde{N}_t \tag{2.51}$$

$$(1+g)\tilde{S}_{t+1} = (1-q(\theta_t,E_t))\tilde{S}_t + \alpha(\tilde{N}_t + f(\theta_t,E_t)\gamma_t\tilde{B}_t) + \tilde{Y}^H_{t-2} \tag{2.52}$$

$$(1+g)\tilde{N}_{t+1} = (1-\alpha)(\tilde{N}_t + f(\theta_t,E_t)\gamma_t\tilde{B}_t) \tag{2.53}$$

$$\begin{aligned} U_{\tilde{C},t} = \beta\mathbb{E}_t\Big[U_{\tilde{C},t+1}\Big((1+R^C_{t+1}-\delta) + \frac{\chi^C}{2}(\frac{K^C_{t+2}-K^C_{t+1}}{K^C_{t+1}})^2 \\ + \chi^C\frac{K^C_{t+2}-K^C_{t+1}}{K^C_{t+1}}\Big)\Big]\Big(1+\chi^C\frac{K^C_{t+1}-K^C_t}{K^C_t}\Big)^{-1} \end{aligned} \tag{2.54}$$

$$\begin{aligned} U_{\tilde{C},t} = \beta\mathbb{E}_t\Big[U_{\tilde{C},t+1}\Big((1+R^H_{t+1}-\delta) + \frac{\chi^H}{2}(\frac{K^H_{t+2}-K^H_{t+1}}{K^H_{t+1}})^2 \\ + \chi^H\frac{K^H_{t+2}-K^H_{t+1}}{K^H_{t+1}}\Big)\Big]\Big(1+\chi^H\frac{K^H_{t+1}-K^H_t}{K^H_t}\Big)^{-1} \end{aligned} \tag{2.55}$$

$$U_{\tilde{C},t}P^W_t = \beta^3\mathbb{E}_t[V^S_{t+3}] \tag{2.56}$$

$$E_t^{2-\phi} = \frac{A(\theta_t)^{\phi-1}}{\sigma}\Big(-U_{\tilde{C},t}P_t^R + \beta\frac{1}{1+g}\mathbb{E}_t\big[(1-\alpha)(V_{t+1}^N - V_{t+1}^B) + \alpha V_{t+1}^S\big]\Big) \tag{2.57}$$

$$\begin{aligned} V_t^B &= -U_{\tilde{C},t}P_t^R f(\theta, E_t) - \frac{\sigma}{2}E_t^2 + \beta\frac{1}{1+g}\mathbb{E}_t\big[\big(1-(1-\alpha)f(\theta, E_t)\gamma_t\big)V_{t+1}^B \\ &\qquad + \alpha f(\theta, E_t)\gamma_t V_{t+1}^S + (1-\alpha)f(\theta, E_t)\gamma_t V_{t+1}^N\big] \end{aligned} \tag{2.58}$$

$$V_t^S = U_{\tilde{C},t}P_t^R q(\theta_t, E_t) + (1-q(\theta_t, E_t))\beta\frac{1}{1+g}\mathbb{E}_t\big[V_{t+1}^S\big] \tag{2.59}$$

$$V_t^N = U_{\tilde{N},t} + \beta\frac{1}{1+g}\mathbb{E}_t\big[(\alpha V_{it}^B + (1-\alpha)V_{t+1}^N\big] \tag{2.60}$$

$$\tilde{Y}_t^C = Z_t^C(\tilde{K}_t^C)^{\nu_C}(\tilde{L}_t^C)^{1-\nu_C} \tag{2.61}$$

$$\tilde{Y}_t^H = \tilde{Z}_t^H(\tilde{K}_t^H)^{\nu_H}(\tilde{L}_t^H)^{\rho_H} \tag{2.62}$$

$$R_t^C = \nu_C Z_t^C(\tilde{K}_t^C)^{\nu_C-1}(\tilde{L}_t^C)^{1-\nu_C} \tag{2.63}$$

$$W_t = (1-\nu_C)Z_t^C(\tilde{K}_t^C)^{\nu_C}(\tilde{L}_t^C)^{-\nu_C} \tag{2.64}$$

$$R_t^H = \nu_C \tilde{Z}_t^H(\tilde{K}_t^C)^{\nu_H-1}(\tilde{L}_t^C)^{\rho_H} \tag{2.65}$$

$$W_t = \rho_H P_t^W \tilde{Z}_t^H(\tilde{K}_t^H)^{\nu_H}(\tilde{L}_t^H)^{\rho_H-1} \tag{2.66}$$

$$P_t^R = \frac{1}{U_{\tilde{C},t}}\beta\frac{1}{1+g}\mathbb{E}_t\Big[(1-\eta_t)\Big((1-\alpha)V_{t+1}^N - V_{t+1}^B + \alpha V_{t+1}^S\Big) + \eta_t V_{t+1}^S\Big] \tag{2.67}$$

$$M(E_t, \gamma_t\tilde{B}_t, \tilde{S}_t) = A(E_t\gamma_t\tilde{B}_t)^{\phi}\tilde{S}_t^{1-\phi} \tag{2.68}$$

$$\theta_t = \frac{\gamma_t\tilde{B}_t}{\tilde{S}_t} \tag{2.69}$$

$$f(\theta, E_t, E_{it}) = AE_{it}(E_t\theta_t)^{\phi-1} \tag{2.70}$$

$$q(\theta_t, E_t) = A(E_t\theta_t)^{\phi} \tag{2.71}$$

$$\tilde{Y}_t^C = \tilde{C}_t + \tilde{I}_t^C + \tilde{I}_t^H \tag{2.72}$$

$$\tilde{L}_t = \tilde{L}_t^C + \tilde{L}_t^H \tag{2.73}$$

$$\tilde{L}_t = \big(U_{\tilde{C},t}\frac{W_t}{\rho}\big)^{\mu} \tag{2.74}$$

$$\ln(\tilde{Z}_t^H) = \psi_H\ln(\tilde{Z}_{t-1}^H) + (1-\psi_H)\ln(\overline{Z^H}) + \epsilon_t^H \tag{2.75}$$

$$\ln(Z_t^C) = \psi_C\ln(Z_{t-1}^C) + (1-\psi_C)\ln(\overline{Z^C}) + \epsilon_t^C + \epsilon_{t-20}^A \tag{2.76}$$

$$\ln(\gamma_t) = \psi_\gamma \ln(\gamma_{t-1}) + (1-\psi_\gamma)\ln(\overline{\gamma}) + \epsilon_t^\gamma \tag{2.77}$$

$$G\tilde{D}P_t = \tilde{Y}_t^C + \overline{P^W}\tilde{Y}_t^H \tag{2.78}$$

$$TOM_t = \frac{1}{q(\theta_t, E_t)} \tag{2.79}$$

$$GDP_{obs,t} = \ln(GDP) \tag{2.80}$$

$$TOM_{obs,t} = \ln(3TOM_t) \tag{2.81}$$

$$PR_{obs,t} = \ln(PR_t) \tag{2.82}$$

$$Sales_{obs,t} = \ln(Sales_t) \tag{2.83}$$

$$Starts_{obs,t} = \ln(Y_t^H) \tag{2.84}$$

$$S_{obs,t} = \ln(S) \tag{2.85}$$

$$U_{\tilde{C},t} = x\tilde{C}_t^{\lambda-1}(x\tilde{C}_t^\lambda + (1-x)\tilde{N}_t^\lambda)^{\frac{1}{\lambda}-1} \tag{2.86}$$

$$U_{\tilde{N},t} = (1-x)\tilde{N}_t^{\lambda-1}(x\tilde{C}_t^\lambda + (1-x)\tilde{N}_t^\lambda)^{\frac{1}{\lambda}-1} \tag{2.87}$$

As in Iacoviello and Neri (2010) and Lambertini et al. (2010), I calculate GDP using the steady state price of housing. In the last six equations, I calculate the simulated data series in logs, which is how they are expressed in my estimation procedure, Figures 2.1 and 2.2, and Tables 2.1, 2.8, and 2.7. Note that I equate time on the market with months' supply in the data; I multiply the model's time on the market by 3 because the model period is quarterly, not monthly. Finally, I equate S_t, the number of sellers, with the homeowner vacancy rate from the Census Housing Vacancy Survey, whereas in reality many homes for sale remain occupied.

In the competitive search model, I replace equations 2.68, 2.70, 2.71, and 2.57 with the following equations, respectively (the only change in the price setting equation is that the buyer's share of the match surplus, η, is time varying):

$$M(E_t, \gamma_t\tilde{B}_t, \tilde{S}_t) = AS_t(1 - e^{-\zeta\frac{E_t\gamma_t\tilde{B}_t}{\tilde{S}_t}}) \tag{2.88}$$

$$f(\theta_t, E_t) = \frac{E_{it}A}{E_t\theta_t}(1 - e^{-E_t\theta_t\zeta}) \tag{2.89}$$

$$q(\theta_t, E_t) = A(1 - e^{-E_t\theta_t\zeta}) \tag{2.90}$$

$$E_t^2 = \frac{A}{\sigma\theta_t}(1 - e^{-E_t\theta_t\zeta})\Big(- U_{\tilde{C},t}P_t^R + \tag{2.91}$$

$$\beta\frac{1}{1+g}\mathbb{E}_t\big[(1-\alpha)(V_{t+1}^N - V_{t+1}^B) + \alpha V_{t+1}^S\big]\Big) \tag{2.92}$$

I also add an equation for the determination of the buyer's share of the match surplus:

$$\eta_t = \frac{E_t\theta_t\zeta}{e^{E_t\theta_t\zeta} - 1} \tag{2.93}$$

Table 2.1: Housing Market Time Series

	Standard Deviation	Correlation Coefficient with GDP	Prices	Sales	Starts	Months' Supply	Vacancy Rate
GDP	0.024	1.000					
		(1.000)					
Prices	0.130	0.510	1.000				
		(0.000)	(1.000)				
Sales	0.152	0.668	0.474	1.000			
		(0.000)	(0.000)	(1.000)			
Starts	0.314	0.618	0.467	0.894	1.000		
		(0.000)	(0.000)	(0.000)	(1.000)		
Months' Supply	0.277	-0.378	-0.010	-0.724	-0.800	1.000	
		(0.000)	(0.915)	(0.000)	(0.000)	(1.000)	
Vacancy Rate	0.143	-0.100	0.204	-0.422	-0.582	0.860	1.000
		(0.288)	(0.030)	(0.000)	(0.000)	(0.000)	(1.000)

Series are seasonally adjusted log values from 1982Q3-2010Q4, expressed as deviations from linear trend plus average value over sample period. p-values for correlation coefficients are in parentheses. GDP, sales, and starts expressed on a per household basis. Prices are from CoreLogic House Price Index for single family detached homes. Sales, sales, starts, and months' supply are for single family homes only. Vacancy rate is for home owners only. Please see Estimation section for details on series construction.

Table 2.2: Calibrated Parameters

Parameter	Symbol	Value	Reason
Competitive Search Model			
Discount Factor	β	0.995	0.98 annually
Depreciation Rate of Productive Capital	δ	0.026	0.1 annually
Relocation Probability	α	.009	Home sales per household $\approx$ .0114
Population Growth Rate	g	.0026	Starts per household $\approx$ .0027
SS Consumption Productivity	$\overline{Z^C}$	324.9	Quarterly GDP $\approx$ \$27,092
SS Housing Productivity	$\overline{Z^H}$	0.0059	House Price $\approx$ \$200,750
CES Utility Parameter	λ	-1	Elasticity of Substitution $= \frac{1}{2}$
Preference for housing vs. consumption	x	0.99998	Quarterly Consumption $\approx$17,950
Frisch Elasticity of Labor Supply	μ	1	Standard Value
Disutility of Labor	ρ	1005	Labor Supply = 1
SS fraction of eligible buyers	$\overline{\gamma}$	0.7	$\theta \approx 0.7$
Efficiency of Matching Function	A	1.31	Months' Supply $\approx$ 7 months
Disutility of Search Effort	σ	24608	Finding Rate $\approx$ .6
Capital's Share in Consumption Production Function	ν_C	.35	Standard Value
Capital's Share in Housing Production Function	ν_H	.19	Albouy and Ehrlich 2011
Labor's Share in Housing Production Function	ρ_H	.56	Albouy and Ehrlich 2011
Persistence of Consumption Productivity Shock	ψ_C	1	Unit Root Technology Shock
Persistence of Housing Productivity Shock	ψ_C	1	Unit Root Technology Shock
Consumption Capital Adjustment Costs	χ_C	0	Might drop this parameter
Housing Capital Adjustment Costs	χ_H	0	Might drop this parameter
Urn-ball Generalization Parameter	ζ	1.08	Steady State $\eta = 0.81$
Random Search Model (where different)			
Efficiency of Matching Function	A	1	Months' Supply $\approx$ 7
Buyers' Exponent in Matching Function	ϕ	.79	Genesove and Han 2010
Disutility of Search Effort	σ	24610	Finding Rate $\approx$ .6
Buyer's Share of Match Surplus	η	.81	Hosios Condition

Table 2.3: Steady State Values

Variable	Symbol	Value
Total Labor Supply	L	1
Proportion of Labor Force in Housing	L^H	0.016
Wholesale Housing Price	P^W	\$194,323
Retail Housing Price	P^R	\$200,750
Number of Eligible Buyers	γB	0.019
Houses for Sale	S	0.026
Market Tightness	θ	0.74
Probability of Sale	$q(E,\theta)$	0.443
Probability of Purchase	$f(E,\theta)$	0.598
Search Effort	E	0.517
Months' Supply	TOM	6.77
Sales	$M(E,\theta,\gamma)$	0.011

Table 2.4: Prior Distributions for Estimated Parameters

Variable	Symbol	Distribution
Persistence of Eligible Buyers Shock	ψ_γ	Uniform(0.001,0.999)
Standard Error of Anticipated Consumption Productivity Shock	σ_A	Uniform(0,0.1)
Standard Error of Unanticipated Consumption Productivity Shock	σ_C	Uniform(0, 0.1)
Standard Error of Eligible Buyers Shock	σ_γ	Uniform(0, 0.5)
Standard Error of Housing Productivity Shock	σ_H	Uniform(0, 0.2)

Table 2.5: Posterior Values for Estimated Parameters

Variable	Symbol	Posterior Mean (S.D.) Competitive Search	Posterior Mean (S.D.) Random Search
Persistence of Eligible Buyers Shock	ψ_γ	0.9945 (0.0032)	0.9669 (0.0098)
Standard Error of Anticipated Consumption Productivity Shock	σ_A	0.0264 (0.0019)	0.0321 (0.0022)
Standard Error of Unanticipated Consumption Productivity Shock	σ_C	0.0259 (0.0019)	0.0319 (0.0021)
Standard Error of Eligible Buyers Shock	σ_γ	0.2027 (0.0137)	0.1721 (0.0108)
Standard Error of Housing Productivity Shock	σ_H	0.0192 (0.0013)	0.0177 (0.0013)

Table 2.6: Conditional Variance Decomposition - Competitive Search Model

Observable Series	ϵ_t^A	ϵ_t^C	ϵ_t^γ	ϵ_t^H
4 quarter horizon				
Prices	0.03	0.51	0.05	0.41
Starts	0.36	0.01	0.48	0.15
Sales	0.00	0.02	0.98	0.00
GDP	0.00	1.00	0.00	0.00
40 quarter horizon				
Prices	0.17	0.53	0.01	0.29
Starts	0.64	0.15	0.17	0.04
Sales	0.03	0.05	0.92	0.00
GDP	0.18	0.82	0.00	0.00

Table shows the proportion of each series' variance accounted for by each shock at the specified horizons.

Table 2.7: Simulations of Competitive Search Model

	Standard Deviation of First Difference	Correlation Coefficient with					
		GDP	Prices	Sales	Starts	Months' Supply	Vacancy Rate
GDP	0.050	1.000 (1.000)					
Prices	0.033	0.771 (0.014)	1.000 (1.000)				
Sales	0.085	0.123 (0.240)	0.181 (0.212)	1.000 (1.000)			
Starts	0.069	-0.017 (0.166)	0.168 (0.155)	0.253 (0.117)	1.000 (1.000)		
Months' Supply	0.076	-0.120 (0.106)	-0.252 (0.083)	-0.450 (0.010)	-0.397 (0.053)	1.000 (1.000)	
Vacancy Rate	0.041	-0.101 (0.099)	-0.229 (0.090)	-0.261 (0.092)	-0.346 (0.084)	0.975 (0.000)	1.000 (1.000)

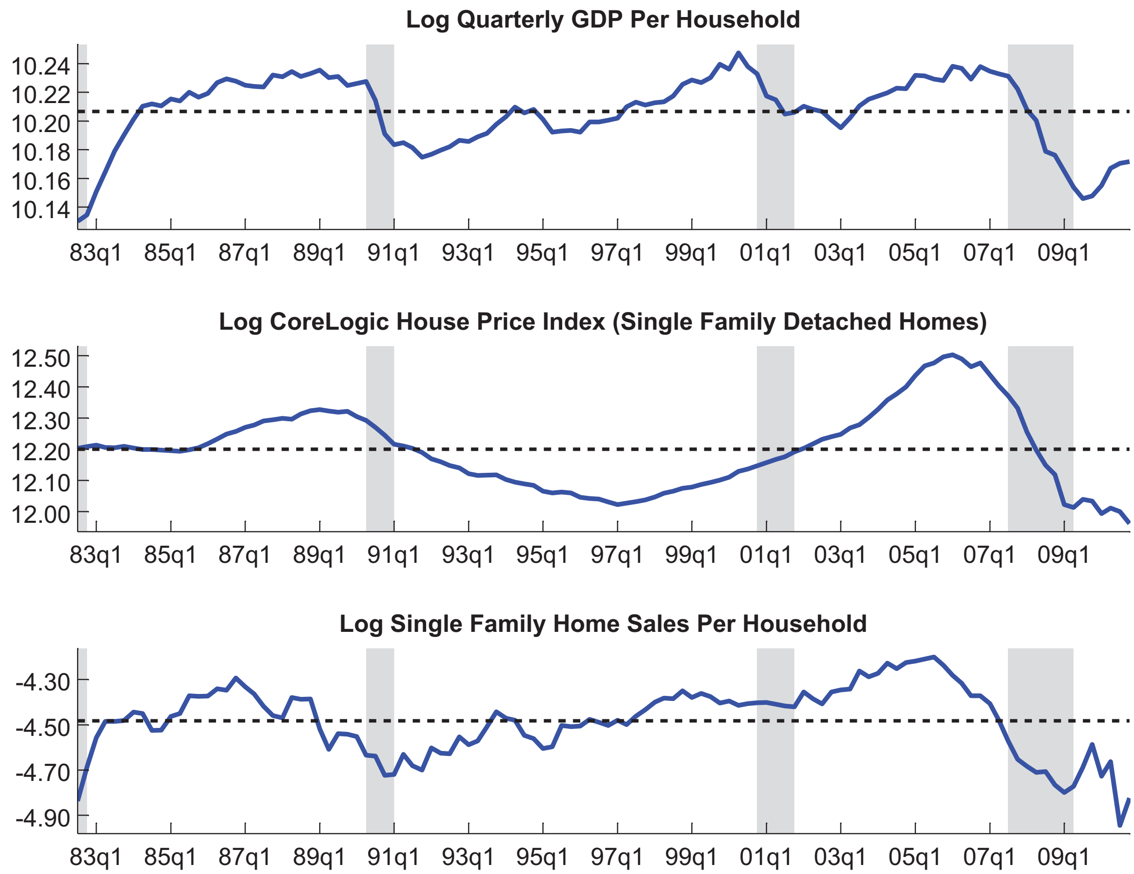
Average values from 500 model simulations of 114 quarters each. Numbers in parentheses are average p-values of correlation coefficients over each simulation.

Table 2.8: Simulations of Random Search Model

	Standard Deviation of First Difference	Correlation Coefficient with GDP	Prices	Sales	Starts	Months' Supply	Vacancy Rate
GDP	0.061	1.000					
		(1.000)					
Prices	0.032	0.843	1.000				
		(0.007)	(1.000)				
Sales	0.088	0.144	0.119	1.000			
		(0.258)	(0.263)	(1.000)			
Starts	0.081	-0.063	0.127	0.033	1.000		
		(0.158)	(0.145)	(0.317)	(1.000)		
Months' Supply	0.078	-0.176	-0.228	-0.339	-0.090	1.000	
		(0.094)	(0.084)	(0.040)	(0.195)	(1.000)	
Vacancy Rate	0.042	-0.153	-0.205	-0.129	-0.046	0.971	1.000
		(0.102)	(0.090)	(0.258)	(0.207)	(0.000)	(1.000)

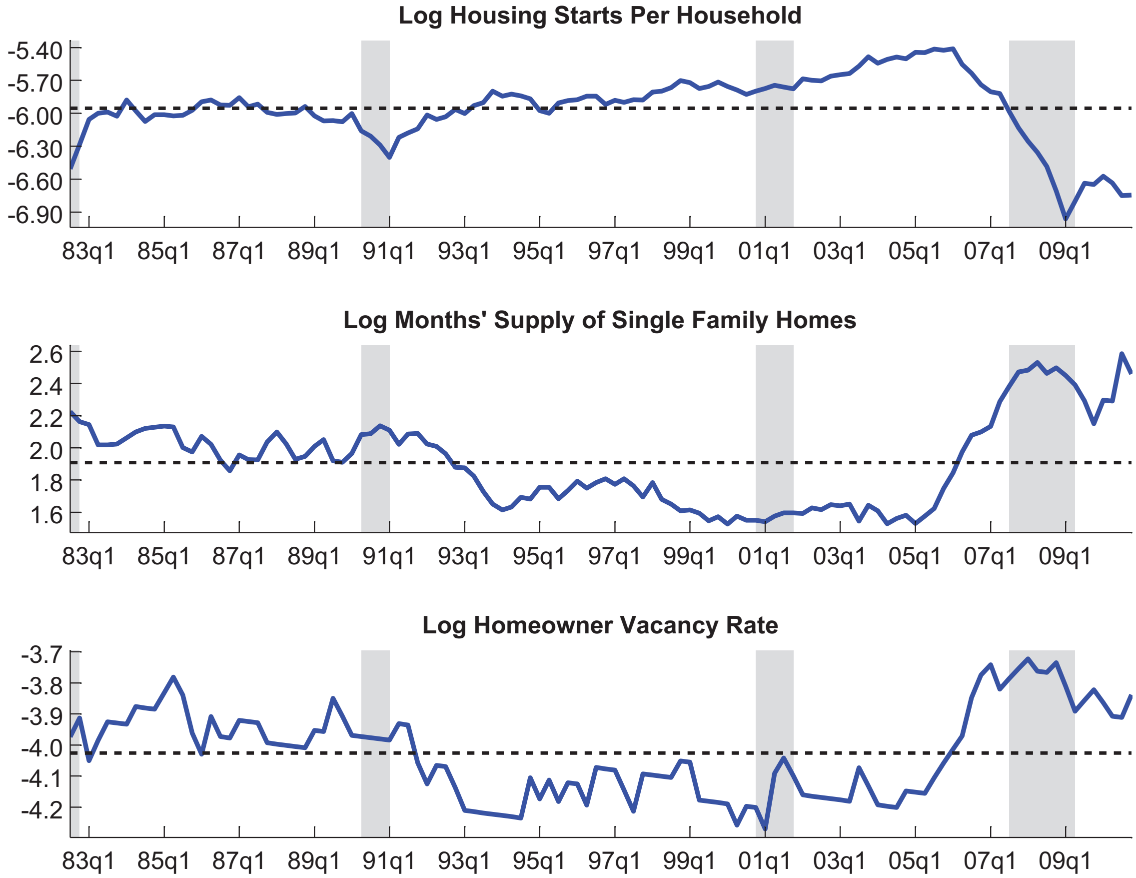
Average values from 500 model simulations of 114 quarters each. Numbers in parentheses are average p-values of correlation coefficients over each simulation.

Figure 2.1: Housing Market Time Series 1983q1-2010q4



Series are seasonally adjusted log values from 1982q3-2010q4, expressed as deviations from linear trend plus average value over sample period. Shaded areas are NBER recession dates. Please see Observed Data Series section for details on series construction.

Figure 2.2: Housing Market Time Series 1983q1-2010q4 (cont'd.)



Series are seasonally adjusted log values from 1982q3-2010q4, expressed as deviations from linear trend plus average value over sample period. Shaded areas are NBER recession dates. Please see Observed Data Series section for details on series construction.

Figure 2.3: The Structure of Households

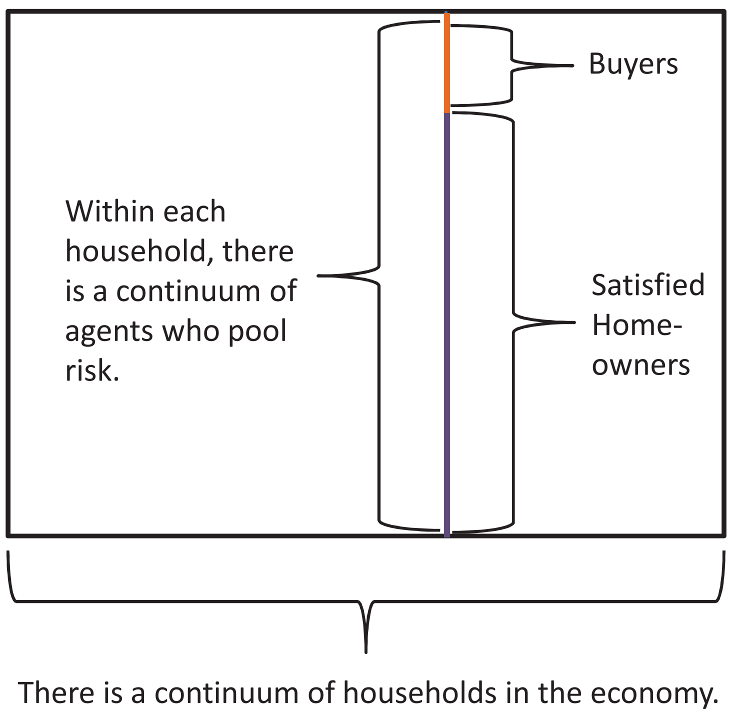
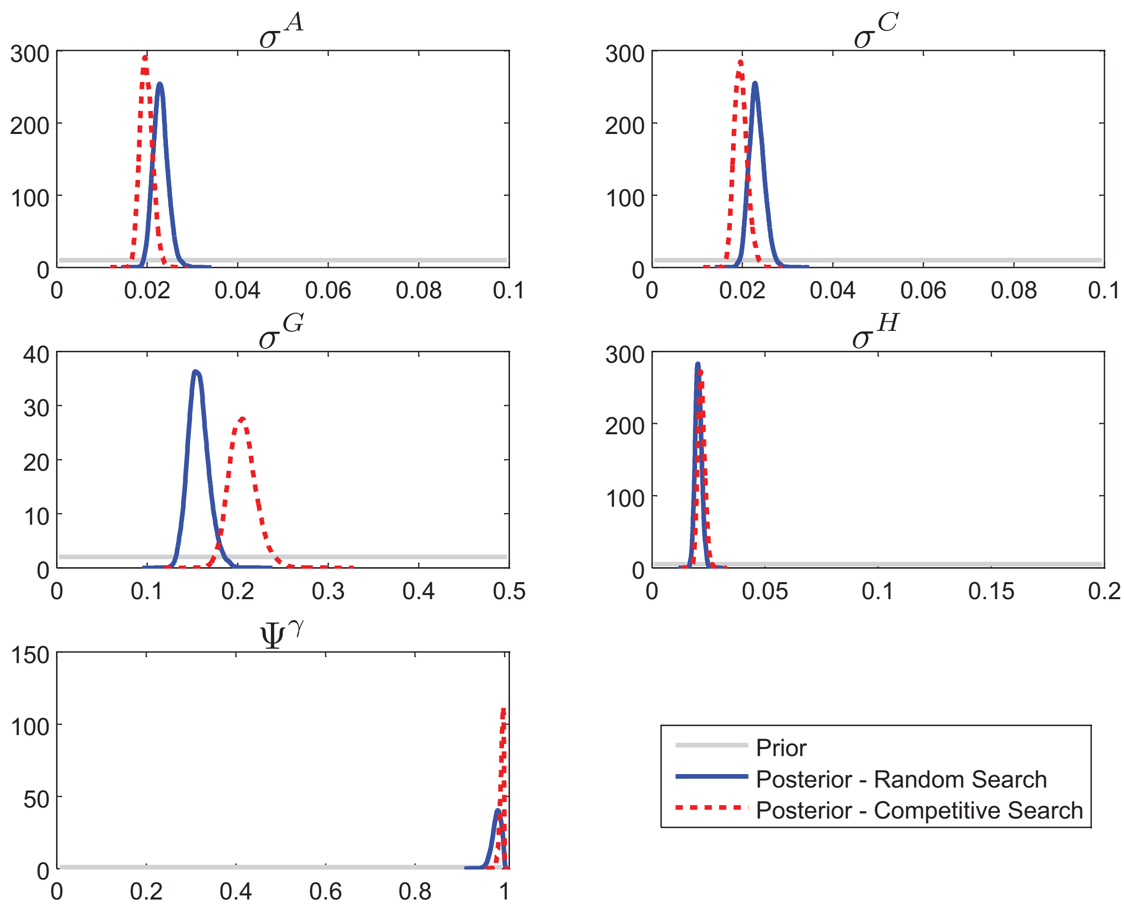
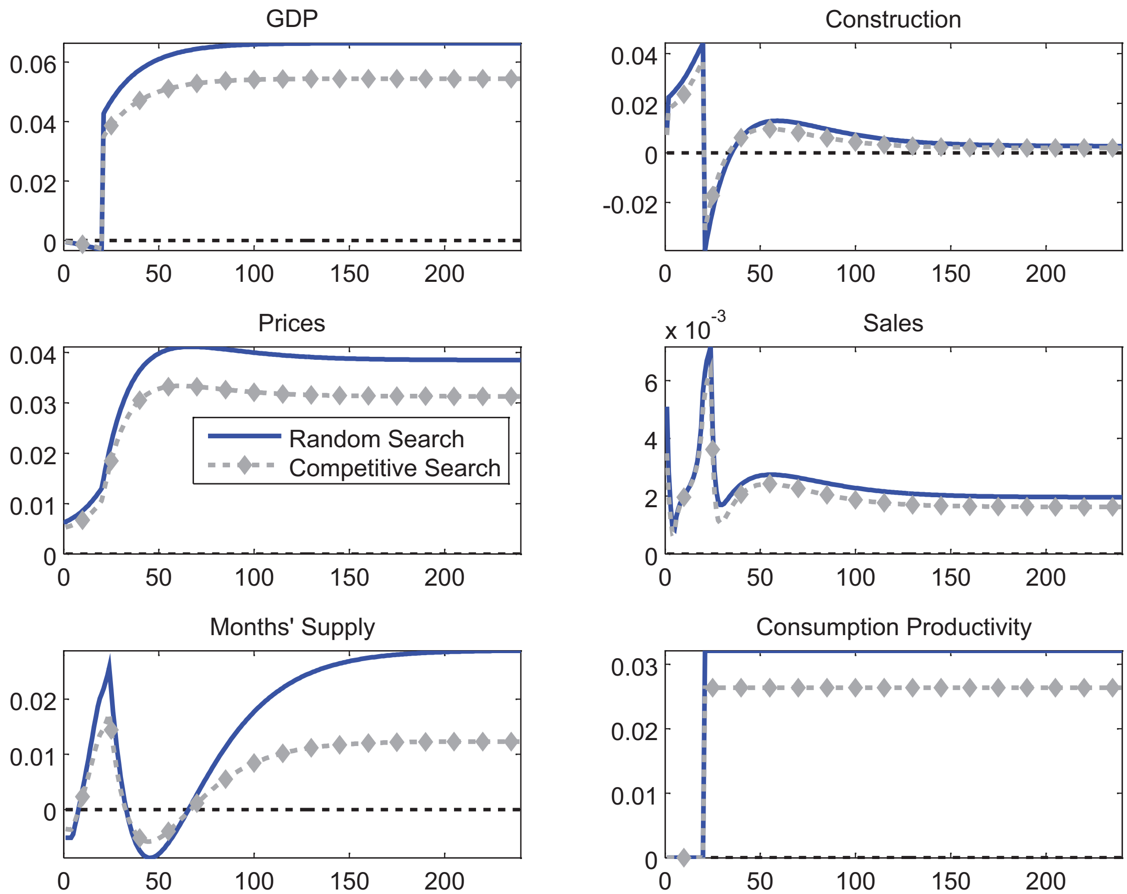


Figure 2.4: Priors and Posteriors for Estimated Parameters



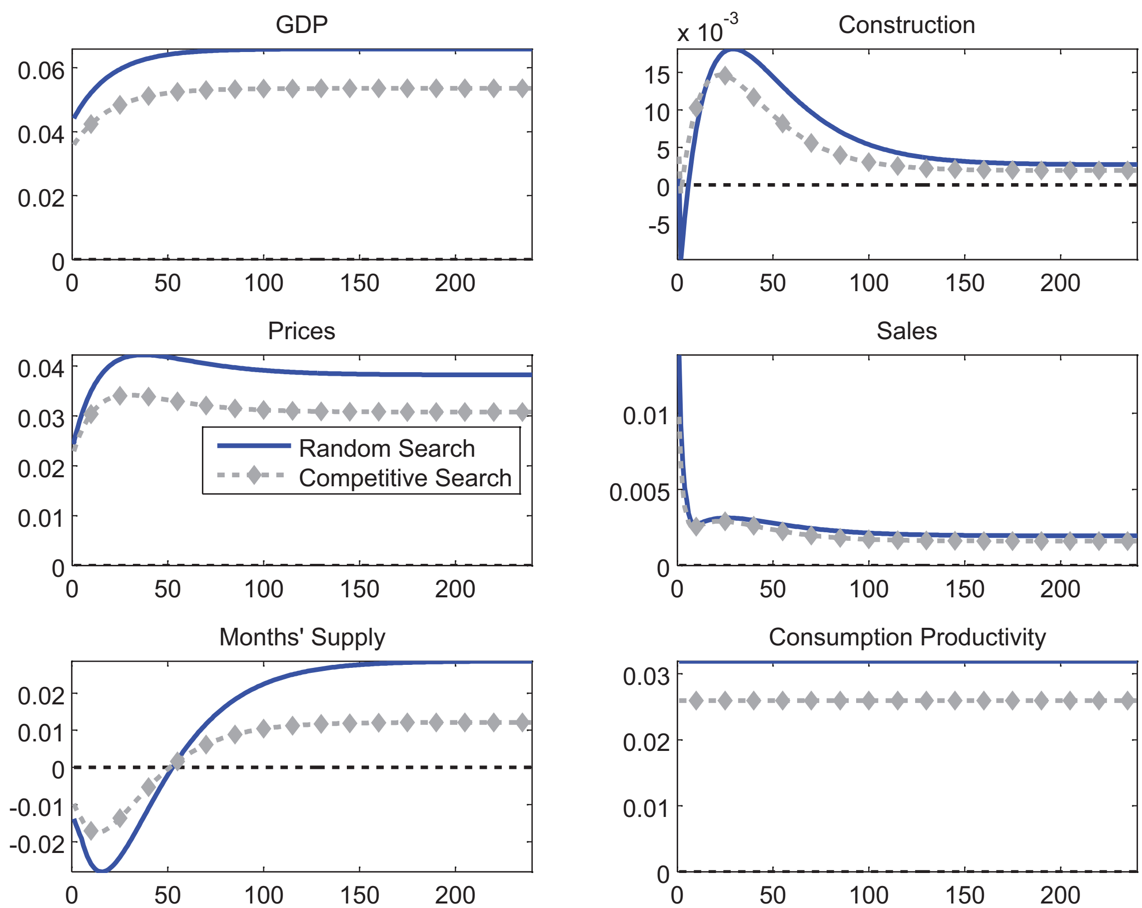
Posteriors obtained from random walk Metropolis Hastings Algorithm.

Figure 2.5: Impulse Responses to an Anticipated Consumption Productivity Shock



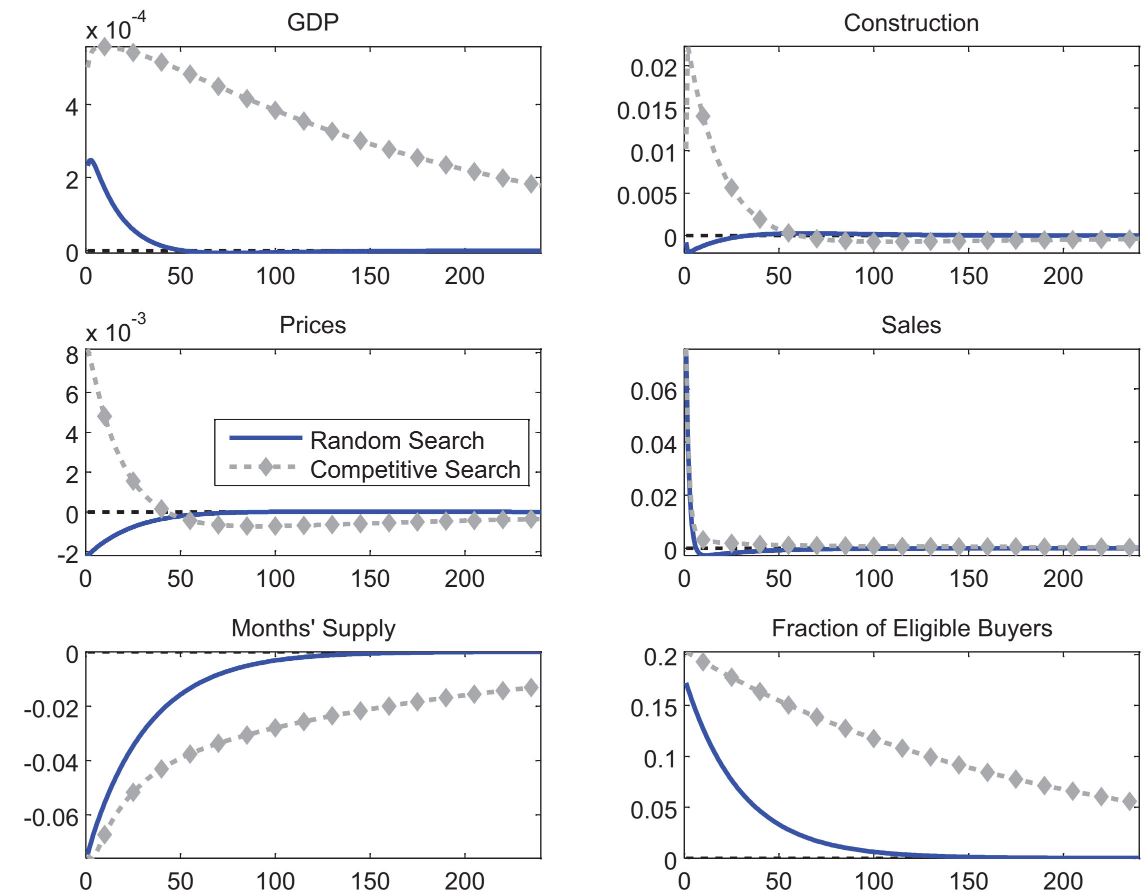
Time period is quarterly. Values shown are log deviations from steady state values.

Figure 2.6: Impulse Responses to an Unanticipated Consumption Productivity Shock



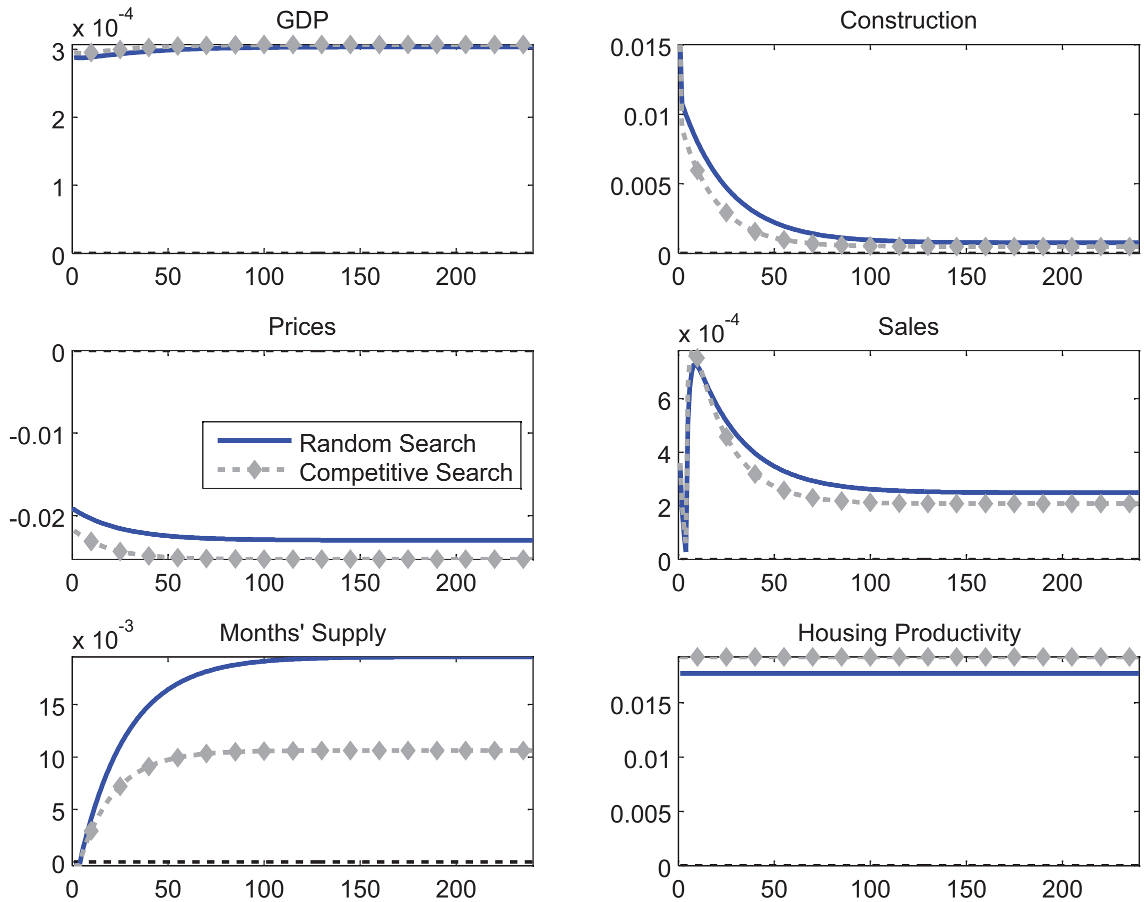
Time period is quarterly. Values shown are log deviations from steady state values.

Figure 2.7: Impulse Responses to an Eligible Buyers Shock



Time period is quarterly. Values shown are log deviations from steady state values.

Figure 2.8: Impulse Responses to an Unanticipated Housing Productivity Shock



Time period is quarterly. Values shown are log deviations from steady state values.

Figure 2.9: Smoothed Shocks - Competitive Search Model

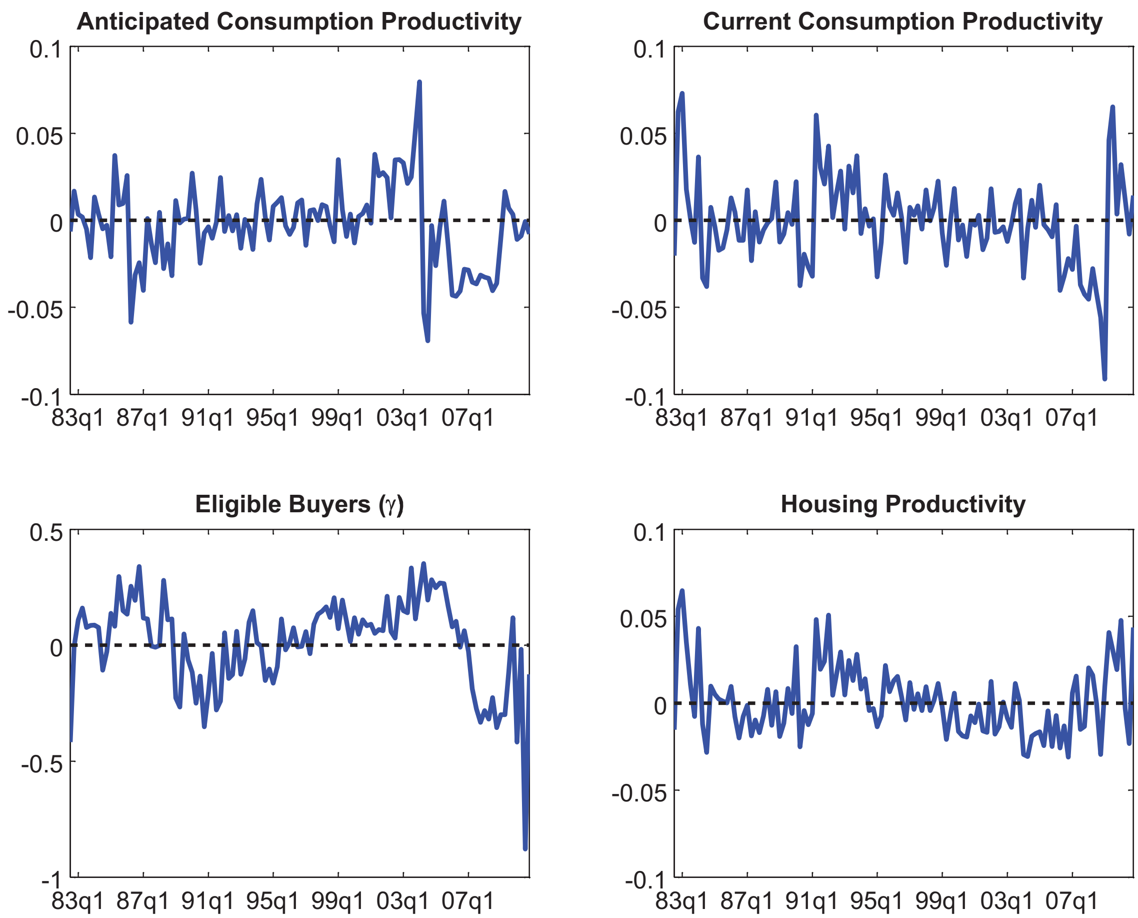


Figure 2.10: Smoothed Shocks - Random Search Model

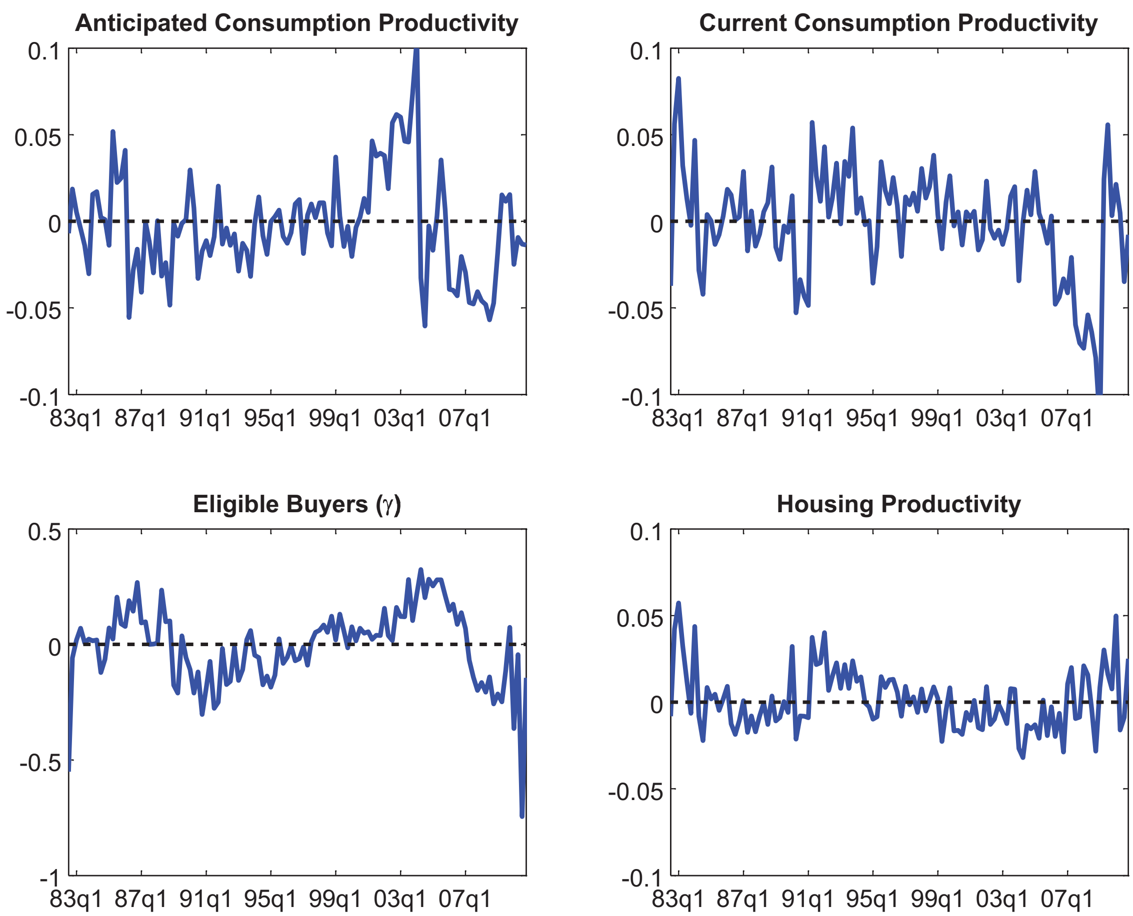
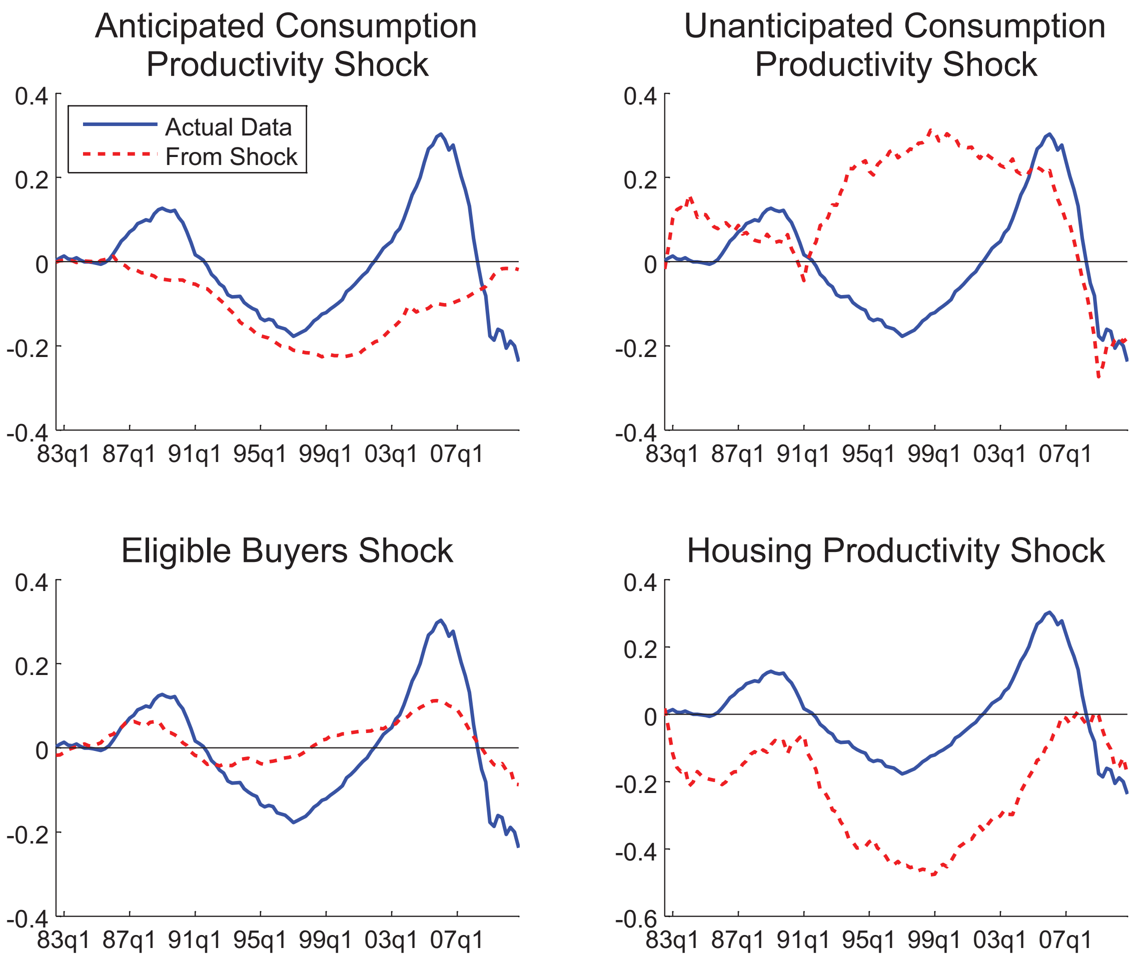


Figure 2.11: Historical Decomposition - Competitive Search Model



CHAPTER III

Metropolitan Land Values and Housing Productivity[1]

3.1 Introduction

Households spend more on housing than any other good, and the value of housing depends fundamentally on the land upon which it is built. Land values can vary tremendously, reflecting the scarcity of the many heterogeneous amenities and labor-market opportunities to which land provides access. They also reflect opportunities for development, as land that cannot be built on generally has little private value. Land values are quite possibly the most fundamental prices examined in spatial and urban economics.

Accurate data on land values have been notoriously piecemeal, although data on housing values are widespread. Housing values can differ considerably from land values, partly because of the labor and material costs of producing housing structures. The topographical nature of a land parcel's terrain can also influence the quantities of inputs needed to produce housing structures. Restrictions and regulations on land use can raise expensive barriers to building, which can lower the efficiency with which housing services are provided to occupants, creating what is often referred

[1]This chapter was co-authored with David Albouy.

to as a "regulatory tax." While these regulations may be costly, they could also provide benefits to local residents by promoting positive neighborhood externalities, or curtailing negative ones. Whether land-use regulations are welfare improving is perhaps the most hotly debated issue in the microeconomics of housing.

Here, we provide the first inter-metropolitan index of directly-observed land values that covers a large number of American metropolitan areas, using recent data from CoStar, a commercial real estate company. This index varies far more than a similarly constructed index of housing values; the two indices are strongly but imperfectly correlated, with potentially informative deviations. We use duality methods (Fuss and McFadden 1978) to estimate the cost relationship between housing output and input prices using these land and housing-value indices, together with indices on non-land input prices and other measures. This supply-side approach to valuing housing strongly complements the demand-side approach to studying housing prices, based on housing's proximity to local amenities and labor-market opportunities.

Our analysis provides a new measure of local productivity in the housing sector, which we infer from the difference between the observed price of housing and the cost predicted by land and other input prices. This productivity metric is a summary indicator of how efficiently housing inputs are transformed into valuable housing services within a metropolitan area. It is also a novel indicator of local productivity in sectors that produce goods not traded across cities. This measure may be contrasted with measures of productivity in tradeables sectors, such as in Beeson and Eberts (1989) and Albouy (2009). Using recent measures by Gyourko, Saiz, and Summers (2008) and Saiz (2010), we investigate how local housing productivity is influenced by natural and artificial constraints to development arising from geography and regulation.

We find that, on average, approximately one-third of housing costs are due to land: this share ranges from 11 to 48 percent in low to high-value areas, implying an elasticity of substitution between land and other inputs of about 0.5 in our baseline specification. Consistent estimation of these parameters requires controlling for regulatory and geographic constraints: a standard deviation increase in aggregate measures of these constraints is associated with 8 to 9 percent higher housing costs. We also examine disaggregated measures of regulation and geography and find that approval delays, supply restrictions, local political pressure, and state court involvement predict the lowest productivity levels, although our estimates are imprecise.

Overall, housing productivity differences across metro areas are large, with a standard deviation equal to 23 percent of total costs, with 22 percent of the variance explained by observed regulations. Contrary to assumptions in the literature (e.g. Shapiro 2006 and Rappaport 2007) that productivity in tradeables and housing are the same, we find the two are negatively correlated. For example, the San Francisco Bay Area is extremely efficient in tradeable industries, and extremely inefficient in producing housing, largely because of its regulations and geography. In general, we find housing productivity to be decreasing, rather than increasing, in city size, suggesting that there are urban diseconomies of scale in housing production. Additionally, we find that lower housing productivity associated with land-use regulation is correlated with a higher quality of life, suggesting that households may value the neighborhood effects these regulations promote. However, the welfare costs of lower housing productivity appear to outweigh these benefits.

Our transaction-based measure differs from common measures of land values based on the difference between a property's entire value and the estimated value of its structure only. Davis and Palumbo (2007) employ this "residual" method successfully across metro areas, although as the authors note, "using several formulas, different sources of data, and a few assumptions about unobserved quantities, none of which is likely to be exactly right." Moreover, the residual method attributes higher costs due to inefficiencies in factor usage – possibly from geographic and regulatory constraints – to higher land values. This may explain why Davis and Palumbo often find higher costs shares of land than we do.[2]

A number of studies have examined data on both housing and land values. Rose (1992) acquires data across 27 cities in Japan and finds greater geographic land availability is associated with lower land and housing values. Ihlanfeldt (2007) takes measures of assessed land values from tax rolls in 25 Florida counties, and finds that land-use regulations are associated with higher housing prices but lower land values. Glaeser and Gyourko (2003) use an augmented residual method to compare housing and inferred land values across the United States, and find that the two differ most in heavily regulated environments. Glaeser, Gyourko, and Saks (2005b) find that the price of units in Manhattan multi-story buildings exceeds the marginal cost of producing them, attributing the difference to regulation. They find the cost of this regulatory tax is larger than the externality benefits they consider, mainly from preserving views.[3]

The econometric approach we use differs in that it explicitly incorporates a cost function, which models land as a variable input to housing production. This approach has similarities to Epple, Gordon, and Sieg (2010), who use separately assessed land and structure values for houses in Alleghany County, PA, and find land's cost share to be 14 percent. We focus on variation across, rather than within cities, which allows us to identify the cost structure from variation in construction prices, geography, and a wide array of regulations. Unlike Epple et al. and Thorsnes (1997), who uses data from Portland, our estimated elasticity of substitution between land and non-land inputs is less than one, which is consistent with much of the older literature – see McDonald (1981) for a survey – based on within-city variation in housing values.

[2]Although hedonic methods can theoretically provide estimates of land values from housing values, these estimates can be questioned. Using an augmented residual method based on hedonics, Glaeser and Ward (2009) estimate a value of $16,000 per acre of land in the Greater Boston area, while presenting evidence that the market price of an acre is approximately $300,000 if new housing can be built on it. They attribute this discrepancy to zoning regulations.

[3]Other works of note that consider the relationship between land-use regulations, land values, and housing values include Ohls et al. (1974), Courant (1976), and Katz and Rosen (1987).

Three recent papers also make use of the CoStar COMPS data to construct land-value indices. Haughwout, Orr, and Bedoll (2008) construct a land price index for the period 1999-2006 within the New York metro area, documenting many sales within the densest areas of Manhattan, as well as in outlying areas. Kok, Monkkonen, and Quigley (2010) also document land sales throughout the San Francisco Bay Area, and relate the sales prices to the topographical, demographic, and regulatory features of the site. Nichols, Oliner, and Mulhall (2010) construct a panel of land-value indices for 23 metro areas from the 1990s through 2009. They demonstrate that land values vary more across time than housing values, much as our analysis demonstrates is true across space.

Section 3.2 presents our cost-function approach for modeling housing prices and relates it to an econometric model. It also provides a general-equilibrium model for the full determination of land values. Section 3.3 discusses our data and explains how we use them to construct indices of land values, housing prices, construction prices, geography, and regulation across metro areas. Section 3.4 presents our estimates of the housing-cost function and how housing productivity is influenced by geographic and regulatory constraints. Section 3.5 considers how housing productivity varies across cities and is related to measures of urban productivity in tradeables and quality of life.

3.2 Model of Land Values and Housing Production

Our econometric model uses a cost function for housing production within a system-of-cities model, proposed by Roback (1982), and developed by Albouy (2009). The national economy contains many cities indexed by j, which produce a numeraire good, X, traded across cities, and housing, Y, which is not traded across cities, and has a local price, p_j. Cities differ in their productivity in the housing sector, A_Y^j.

3.2.1 Cost Function for Housing

We begin with a two-factor model in which firms produce housing, Y_j, using land L and materials M according to the production function

$$Y_j = F^Y(L, M; A_j^Y), \tag{3.1}$$

where F_j^Y is concave and exhibits constant returns to scale (CRS) in L and M at the firm level. Housing productivity, A_j^Y, is a city-level characteristic that may be fixed or determined endogenously by city characteristics, such as population size. Land is paid a city-specific price, r_j, while materials are paid price v_j. In our empirical work, we operationalize M as the installed structure component of housing, so v_j is conceptualized as an index of construction input prices, possibly an aggregate of local labor and mobile capital. Unit costs in the housing sector, equal to marginal and average costs, are $c^Y(r_j, v_j; A_j^Y) \equiv \min_{L,M}\{r_j L + v_j M : F_Y(L, M; A_j^Y) = 1\}$.

The use of a single function to model the production of a heterogeneous housing stock is well established in the literature, beginning with Muth (1960) and Olsen (1969). In the words of Epple et al. (2010, p. 906)

> The production function for housing entails a powerful abstraction. Houses are viewed as differing only in the quantity of services they provide, with housing services being homogeneous and divisible. Thus, a grand house and a modest house differ only in the number of homogeneous service units they contain.

This abstraction also implies that a highly capital-intensive form of housing, e.g., an apartment building, can substitute in consumption for a highly land-intensive form of housing, e.g., single-story detached houses.[4]

[4] Our analysis uses data from owner-occupied properties, accounting for 67% of homes, of which 82% are single-family and detached.

Assuming the housing market in city j is perfectly competitive[5], then in equilibrium housing price equals the unit cost in cities with positive production:

$$c^Y(r_j, v_j; A_j^Y) = p_j. \tag{3.2}$$

Our methodology of estimating housing productivity is illustrated in figure 3.1 panel A, holding v_j constant. The thick solid curve represents the cost function of housing for cities with average productivity. As land values rise from Denver to New York, housing prices rise, albeit at a diminishing rate, as housing producers substitute away from land as a factor input. The higher, thinner curve represents the cost function for a city with lower productivity, such as San Francisco. The lower productivity level is identified by how much higher the housing price in San Francisco is relative to a city with the same factor costs, such as in New York.

The first-order log-linear approximation of equation (3.2) around the national average expresses how housing prices should vary with input prices and productivity, $\hat{p}_j = \phi^L \hat{r}_j + (1-\phi^L)\hat{v}_j - \hat{A}_j^Y$. $\hat{z}^j$ represents, for any z, city j's log deviation from the national average, $\bar{z}$, i.e. $\hat{z}^j = \ln z^j - \ln \bar{z}$. ϕ^L is the cost share of land in housing at the average, and A_Y^j is normalized so that a one-point increase in $\hat{A}_j^Y$ corresponds to a one-point reduction in log costs.[6] Rearranged, this equation infers home-productivity from how high land and material costs are relative to housing costs:

$$\hat{A}_Y^j = \phi^L \hat{r}_j + (1-\phi^L)\hat{v}_j - \hat{p}_j. \tag{3.3}$$

[5]Although this assumption may seem stringent, the empirical evidence is consistent with perfect competition in the construction sector. Considering evidence from the 1997 Economic Census, Glaeser et al. (2005b) report that "...all the available evidence suggests that the housing production industry is highly competitive." Basu et al. (2006) calculate returns to scale in the construction industry (average cost divided by marginal cost) as 1.00, indicating firms in the construction industry having no market power. This seems sensible as new homes must compete with the stock of existing homes. If markets are imperfectly competitive, then A_j^Y will vary inversely with the mark-up on housing prices above marginal costs.

[6]This requires that productivity at the national average obeys $\bar{A}^Y = -\bar{p}/[\partial c^Y(\bar{r}, \bar{m}, \bar{A}^Y)/\partial A]$.

If housing productivity is factor neutral, i.e., $F^Y(L, M; A_j^Y) = A_j^Y F^Y(L, M; 1)$, then the second-order log-linear approximation of (3.2), drawn in figure 3.1 panel B, is

$$\hat{p}_j = \phi^L \hat{r}_j + (1 - \phi^L)\hat{v}_j + \frac{1}{2}\phi^L(1 - \phi^L)(1 - \sigma^Y)(\hat{r}_j - \hat{v}_j)^2 - \hat{A}_j^Y, \tag{3.4}$$

where σ^Y is the elasticity of substitution between land and non-land inputs. This elasticity of substitution is less than one if costs increase in the square of the factor-price difference, $(\hat{r}_j - \hat{v}_j)^2$. The actual cost share is not constant across cities, but is approximated by

$$\phi_j^L = \phi^L + \phi^L(1 - \phi^L)(1 - \sigma^Y)(\hat{r}_j - \hat{v}_j), \tag{3.5}$$

and thus is increasing with $\hat{r}_j - \hat{v}_j$ when $\sigma^Y < 1$. Our estimates of $\hat{A}_j^Y$ assume that a single elasticity of substitution describes production in all cities. If this elasticity varies, then our estimates will conflate a lower elasticity with lower productivity. This is seen in figure 3.1 panels A and B, which compare $\sigma^Y = 1$ in the solid curves, with $\sigma^Y = 1$ in the dashed curves. When production has low substitutability, the cost curve is flatter, as housing does not use less land in higher-value cities. This has the same observable consequence of increasing housing prices, although theoretically the concepts are different.[7]

If housing productivity is not factor neutral, then as derived in Appendix 3.7.1, equation (3.4) contains additional terms to account for the productivity of land relative to materials, A_j^{YL}/A_j^{YM}:

$$-\phi^L(1 - \phi^L)(1 - \sigma^Y)(\hat{r}_j - \hat{v}_j)(\hat{A}_j^{YL} - \hat{A}_j^{YM}). \tag{3.6}$$

If $\sigma^Y < 1$, then cities where land is expensive relative to materials, i.e., $\hat{r}_j > \hat{v}_j$, see

[7]Housing supply, as a quantity, is less responsive to price increases when substitutability is low, rather than when productivity is low. While it would be desirable to distinguish the two, this would be significantly more challenging and require much additional data, and so we leave it for future work.

greater cost reductions where the relative productivity level, A_j^{YL}/A_j^{YM}, is higher.

3.2.2 Econometric Model

As a starting point, we estimate housing prices using an unrestricted translog cost function (Christensen et al. 1973) in terms of land and non-land factor prices:

$$\hat{p}_j = \beta_1 \hat{r}_j + \beta_2 \hat{v}_j + \beta_3 (\hat{r}_j)^2 + \beta_4 (\hat{v}_j)^2 + \beta_5 (\hat{r}_j \hat{v}_j) + Z^j \gamma + \varepsilon_j, \tag{3.7}$$

where Z^j is a vector of city-level observable attributes that may affect housing prices. This specification is equivalent to the second-order approximation of the cost function (see, e.g., Binswager 1974, Fuss and McFadden 1978) under the restrictions imposed by CRS

$$\beta_1 = 1 - \beta_2, \beta_3 = \beta_4 = -\beta_5/2, \tag{3.8}$$

where $\phi^L = \beta_1$ and, with factor-neutral productivity, $\sigma^Y = 1 - 2\beta_3/\left[\beta_1(1-\beta_1)\right]$. Housing productivity is determined by attributes in Z^j and unobservable attributes in the residual, ε_j:

$$\hat{A}_Y^j = Z^j(-\gamma) + \hat{A}_{0Y}^j, \ \hat{A}_{0Y}^j = -\varepsilon_j. \tag{3.9}$$

The second-order approximation of the cost function (i.e. the translog) is not a constant-elasticity form. Hence, the elasticity of substitution we estimate is evaluated at the sample mean parameter values (see Griliches and Ringstad 1971). The assumption of Cobb-Douglas (CD) production technology imposes the restriction $\sigma^Y = 1$, which in equation (3.7) amounts to the three restrictions:

$$\beta_3 = \beta_4 = \beta_5 = 0. \tag{3.10}$$

Without additional data, non-neutral productivity differences are impossible to detect unless we know what may shift A_j^{YL}/A_j^{YM}. In the context, it seems reasonable to interact productivity shifters Z_j with the difference in input prices ($\hat{r}_j - \hat{\nu}_j$) in equation (3.7). The reduced-form model allowing for non-neutral productivity shifts, imposing the CRS restrictions may be written as:

$$\hat{p}_j - \hat{v}_j = \beta_1(\hat{r}_j - \hat{v}_j) + \beta_3\left[(\hat{r}_j)^2 + (\hat{v}_j)^2 - 2(\hat{r}_j\hat{v}_j)\right] + \gamma_1 Z^j + \gamma_2 Z^j(\hat{r}_j - \hat{v}_j) + \varepsilon_j \quad (3.11)$$

As shown in Appendix 3.7.1, $\gamma_2 Z^j/2\beta_3 = (\hat{A}_j^{YM} - \hat{A}_j^{YL}) - (\hat{A}_{0j}^{YM} - \hat{A}_{0j}^{YL})$ identifies observable differences in factor-biased technical differences. If $\sigma_Y < 1$, then $\gamma_2 > 0$ implies that the shifter Z lowers the productivity of land relative to the non-land input.[8]

3.2.3 Full Determination of Land Values

In this section, we determine land values and local-wage levels in a model of location demand based on amenities to households, bundled as quality of life, Q_j, and to firms in the tradeable sector, bundled as trade productivity, A_j^X. Casual readers may skip this section without loss of intuition. We posit two types of mobile workers, $k = X, Y$, where type-Y workers work in the housing sector. Preferences are modeled by the utility function $U^k(x, y; Q_j^k)$, which is quasi-concave over consumption x and y, and increases in Q_j^k, which may vary by type. The household expenditure function is $e^k(p, u; Q) \equiv \min_{x,y}\{x + py : U^k(x, y; Q) \geq u\}$. Each household supplies a single unit of labor and is paid $w_{j,}^k$, which with non-labor income, I, makes up total income $m_j^k = w_j^k + I$, out of which federal taxes, $\tau(m_j^k)$, are paid. We assume households are mobile and that both types occupy each city. Equilibrium requires that households receive the same utility in all cities, so that higher prices or lower quality-of-life must be compensated with greater after-tax income, $e^k(p_j, \bar{u}^k; Q_j^k) = m_j^k - \tau(m_j^k)$, $k = X, Y$, where $\bar{u}^k$ is the level of utility attained nationally by type k. Log-linearizing this

[8] In equation (3.11), non-neutral productivity implies $\beta_1 = \phi_L + \beta_3(\hat{A}_{0j}^{YM} - \hat{A}_{0j}^{YL})$ and $\varepsilon^j = -[\phi^L \hat{A}_j^{YL} + (1-\phi^L)\hat{A}_j^{YM}] + (1/2)\phi^L(1-\phi^L)(1-\sigma^Y)(\hat{A}_j^{YL} - \hat{A}_j^{YM})^2$

condition around the national average

$$\hat{Q}_j^k = s_y^k \hat{p}_j - (1-\tau^k) s_w^k \hat{w}_j^k, \; k = X, Y. \tag{3.12}$$

Q_j^k is normalized $\hat{Q}_j^k$ is equivalent to a one-percent drop in total consumption, s_y^k is the average expenditure share on housing, and τ^k is the average marginal tax rate, and s_w^k is the share of income from labor. Define the aggregate quality-of-life differential $\hat{Q}_j \equiv \mu^X \hat{Q}_j^X + \mu^Y \hat{Q}_j^Y$, where μ^k is the share of income earned by type k households, and let $s_y \equiv \mu^X s_y^X + \mu^Y s_y^Y$, and $(1-\tau)\, s_w \hat{w} \equiv \mu^X (1-\tau^X) s_w^X \hat{w}_j^X + \mu^Y (1-\tau^Y) s_w^Y \hat{w}_j^Y$.

Unlike housing output, tradeable output has a uniform price across all cities, and is produced through the CRS and CD function, $X_j = F^X(L^X, N^X, K^X; A_j^X)$, where N^X is labor and K^X is mobile capital, which also has the uniform price, i, everywhere. We also assume that land commands the same price, r_j, within a city in all sectors. A derivation similar to the one for (3.3) yields the measure of tradeable productivity:

$$\hat{A}_j^X = \theta^L \hat{r}_j + \theta^N \hat{w}_j^X, \tag{3.13}$$

where $\theta^L$and θ^N are the average cost-shares of land and labor in the tradeable sector.

To complete the model, let non-land inputs be produced through the CRS and CD function $M_j = (N^Y)^a (K^Y)^{1-a}$, which implies $\hat{v}_j = a\hat{w}_j^Y$, where a is the cost-share of labor in non-land inputs. Defining $\phi^N = a(1-\phi^L)$, we can derive an alternative measure of housing productivity based on wages:

$$\hat{A}_j^Y = \phi^L \hat{r}_j + \phi^N \hat{w}_j^Y - \hat{p}_j. \tag{3.14}$$

Combining the productivity in both sectors, the total-productivity differential of a city is

$$\hat{A}_j \equiv s_x \hat{A}_j^X + s_y \hat{A}_j^Y, \tag{3.15}$$

where s_x is the average expenditure share on tradeables.

Combining the first-order approximation equations (3.12), (3.13), (3.14), and (3.15), we get that the land-value differential times the average income share of land, $s_R = s_x\theta_L + s_y\phi_L$, equals the total productivity differential plus the quality-of-life differential, minus the tax differential to the federal government, $\tau s_w \hat{w}_j$:

$$s_R\hat{r}_j = s_x\hat{A}_j^X + s_y\hat{A}_j^Y + \hat{Q}_j - \tau s_w \hat{w}_j. \tag{3.16}$$

In other words, land fully capitalizes the value of local amenities minus federal tax payments.

Proper identification of the model requires that the observed determinants of land values, $\hat{r}_j$, $\hat{w}_j$, and Z_j are uncorrelated with unobserved determinants of A_j^Y in the residual, ε_j. To some extent, this is inevitable if the vector of characteristics Z_j is incomplete and $\hat{A}_j^{Y0} = -\varepsilon_j \neq 0$, as $\hat{r}_j$ is determined by $\hat{A}_j^Y$ in (3.16). We have considered modeling the simultaneous determination of $\hat{r}_j$ by $\hat{A}_j^{Y0}$, but this requires knowing the covariance structure between $\hat{A}_j^X$,$\hat{A}_j^Y$, and $\hat{Q}_j$. A more promising approach is to find instrumental variables (IVs) that influence $\hat{A}_j^X$ or $\hat{Q}^j$ but are unrelated to $\hat{A}_j^Y$. Below, we consider two instruments for land and non-land input prices that we think are plausible, although certainly not unassailable. The first is the inverse distance to the nearest salt-water coast. The second is average winter temperature. We find the IV estimates are consistent with, but less precise than, our ordinary least square (OLS) results, and thus focus on the latter. The geographic constraints are predetermined, so we treat them as exogenous. We have not found a plausible strong instrument for regulatory constraints.

3.3 Data and Metropolitan Indicators

3.3.1 Land Values

We calculate our land-value index from transactions prices reported in the CoStar COMPS database. The CoStar Group provides commercial real estate information and claims to have the industry's largest research organization, with researchers making over 10,000 calls a day to commercial real estate professionals. The COMPS database includes transaction details for all types of commercial real estate, including what they term "land." In this study, we take every land sale in the COMPS database provided by CoStar University, which is provided for free to academic researchers.

Our sample includes transactions that occurred between 2005 and 2010 in a Metropolitan Statistical Area (MSA).[9] It excludes all transactions CoStar has marked as non-arms length or without complete information for lot size, sales price, county, and date, or that appear to feature a structure. Finally, we drop observations we could not geocode successfully, leaving us with 68,757 observed land sales.[10]

CoStar provides a field describing the "proposed use" of each property, useful for our analysis. We use 12 of the most common categories of "proposed use," which are neither mutually exclusive nor collectively exhaustive. Properties can have multiple proposed uses or none at all. Thus, we also use an indicator for no proposed use.

The median price per acre in our sample is $272,838, while the mean is $1,536,374; the median lot size is 3.5 acres while the mean is 26.4. Land sales occur more frequently in the beginning of our sample period, with 21.7% of our sample from 2005,

[9]We use the June 30, 1999 definitions provided by the Office of Management and Budget. The data are organized by Primary Metropolitan Statistical Areas (PMSAs) within larger Consolidated Metropolitan Statistical Areas (CMSAs).

[10]We consider an observation to feature a structure when the transaction record includes the fields for "Bldg Type", "Year Built", "Age", or the phrase "Business Value Included" in the field "Sale Conditions." We geocoded using the Stata module "geocode" described in Ozimek and Miles (2011). In addition, we drop outlier observations that we calculate as farther than 75 miles from the city center or that have a predicted density greater than 50,000 housing units per square mile using the weighting scheme described below. We also exclude outlier observations with a listed price of less than $100 per acre or a lot size over 5,000 acres.

and 11.4% from 2010. The frequencies of proposed uses are reported in table 3.1: 17.6% is for residential, including 10.7% is for single-family homes, 3.3% for multi-family; and 3.6% for apartments; industrial, office, retail, medical, parking, and commercial uses together account for 24.1%. 23.4% is being held for development or investment, and 15.9% of the sample had no proposed use.

We calculate the metropolitan index of land values by regressing the log price per acre of each sale, $\ln \tilde{r}_{ijt}$ on a set of a vector of controls, X_{ijt}, and a set of indicator variables for each year-MSA interaction, ψ_{jt} in the equation $\ln \tilde{r}_{ijt} = X_{ijt}\beta + \psi_{jt} + e_{ijt}$. In our regression tables we use land-value indices, $\hat{r}_{jt}$, based on estimates of ψ_{jt} by year and MSA, normalized to have a national average of zero, weighting by number of housing units; in our summary statistics and figures, we report land-value indices, $\hat{r}_j$, aggregated across years. Furthermore, because of our limited sample size, land-value indices derived from metro areas with fewer land sales may exhibit excess dispersion because of sampling error. We correct for this using shrinkage methods described in Kane and Staiger (2008), accounting for yearly as well as metropolitan variation in the estimated $\hat{\psi}_{jt}$. The shrinkage effects are generally small, but do appear to correct for mild amounts of attenuation bias in our subsequent analysis.

Table 3.1 reports the results for four successive land-value regressions. The first regression has no controls. In column 2, we control for log lot size in acres, which improves the R^2 substantially from 0.30 to 0.70. The coefficient on lot size is -0.66, illustrating the "plattage effect," documented by Colwell and Sirmans (1980, 1993). According to these authors, when there are costs to subdividing parcels (e.g. because of zoning restrictions), large lots contain more land than is optimal for their intended use, thus lowering their value per acre. Another possible explanation for this effect is that large lots are located in less desirable areas. In column 3, we add controls for intended use raising the R^2 to 0.71. These intended uses help control for various characteristics of the land parcels, although ultimately their inclusion has little impact on our land-value index.

The sample of land parcels in our data set is not a random sample of all lots, which raises the concern of sample selection bias. As discussed in Nichols et al. (2010), it is impossible to correct for this selection bias because we do not observe prices for unsold lots.[11] One especially relevant source of selection bias is that the geographic distribution of sales may differ systematically from the overall distribution of land. For instance, we may be more likely to observe land sales on the urban fringe, where development activity is more intense. Such land will more closely reflect agricultural land values, thus attenuating land-value differences across cities.

To handle sample selection, we re-weight our land observations to reflect the distribution of housing units in the metro area. For each MSA, we pinpoint the metropolitan center using Google Maps.[12] Then, we regress the log number of housing units per square mile at the census-tract level on the North-South and East-West distances between the tract center and the city center, and the squares and product of these distances. We calculate the predicted density of each observed land sale using the city-specific coefficients from this regression, and use this predicted density in column 4, which we take as our preferred specification. The un-weighted and weighted indices are highly correlated (the correlation coefficient is above 0.99), although the latter are more dispersed, as predicted.

Because our focus is on residential housing, we were initially concerned about using land sales with non-residential proposed uses. Ultimately, we find that indices constructed only from land sales with a proposed residential use do not differ systematically from our preferred index, except that they are less precise. Nonetheless, when we conduct our analysis below using residential-only indices, our chief results are largely unaffected, although we lose some MSAs from our sample.

[11] There is a modest literature that attempts to control for selection bias in commercial real estate and land prices, for example, Colwell and Munneke (1997), Fisher et al. (2007), and Munneke and Slade (2000, 2001). Sample selection generally appears to be weak in this context.

[12] These centers are generally within a few blocks of the city hall of the MSA's central city.

Our preferred land-value index is based on the shrunken and weighted estimators based on all land sales, as described above. To illustrate the impact of these choices, figure 3.7 contrasts the differences between shrunken and unshrunken indices; figure 3.8, between weighted and un-weighted indices; and figure 3.9, between using all land and land only for residential uses. While there are some differences between these indices, their overall patterns are rather similar.

Land values for a selected group of metropolitan areas are reported in table 3.2, together with averages by metropolitan population size. These values are very dispersed, with a weighted standard deviation of 0.76. The highest land values in the sample are around New York City, San Francisco, and Los Angeles; the lowest are in Saginaw, Utica, and Rochester, which has land values 1/35th those of New York City. In general, large, coastal cities have the highest land values, while smaller cities in the South and Midwest have lower values.

3.3.2 Housing Prices, Wages, and Construction Prices

We calculate housing-price and wage indices for each year from 2005 to 2010 using the 1% samples from the American Community Survey. Our method, described in detail in Appendix 3.7.2, mimics that for land values. For each year, we regress housing prices of owner-occupied units on a set of indicators for each MSA, controlling flexibly for observed housing characteristics, including age and type of building structure, number of rooms and bedrooms interacted, and kitchen and plumbing facilities. The coefficients on these metro indicators, normalized to have a weighted average of zero, provide our index of housing prices, $\hat{p}_{jt}$, which we aggregate across years for display.

We estimate wage levels in a similar fashion, controlling for worker skills and characteristics. We estimate indices for all workers, $\hat{w}_j$, and for the purpose of our cost estimates, workers in the construction industry only, $\hat{w}_j^Y$. As seen in figure 3.10,

$\hat{w}_j^Y$ is similar to, but more dispersed than, overall wages, $\hat{w}_j$.[13]

Our primary price index for construction inputs is calculated from the Building Construction Cost data from the RS Means company, widely used in the literature, e.g., Davis and Palumbo (2007), and Glaeser et al. (2005b). For each city in their sample, RS Means reports construction costs for a composite of nine common structure types. The index reflects the costs of labor, materials, and equipment rental, but not cost variations from regulatory restrictions, restrictive union practices, or regional differences in building codes. We re-normalize this index as a $z-$score with an average value of zero and a standard deviation of one across cities.[14]

The model of housing equilibrium requires that equation (3.2) be satisfied, so that the replacement cost of a housing unit equals its market price. Because housing is durable, this condition may not bind in cities where housing demand is so weak that there is effectively no new supply (Glaeser and Gyourko 2005). In this case, replacement costs will be above market prices, biasing the estimate of A_j^Y upwards. Technically, there is new housing supply in all of the MSAs in our sample, as measured by building permits. However, we suspect that the equilibrium condition may not bind throughout metro areas where population growth has been low. To indicate MSAs with weak growth, we mark with an asterisk ($*$), MSAs where the population growth between 1980 and 2010 is in the lowest decile of our sample, weighted by 2010 population. These include metros such as Pittsburgh, Buffalo, and Detroit. In Appendix 3.7.3, we find that the results are relatively unchanged when we exclude these areas, although we report their estimates of housing productivity with caution.

The housing-price, construction-wage, and construction-cost indices, reported in

[13] We estimate wage levels at the CMSA level to account for commuting behavior across PMSAs.

[14] The RS Means index is based on cities as defined by three-digit zip code locations, and as such there is not necessarily a one-to-one correspondence between metropolitan areas and RS Means cities, but in most cases the correspondence is clear. If an MSA contains more than one RS Means city we use the construction cost index of the city in the MSA that also has an entry in RS Means. If a PMSA is separately defined in RS Means we use the cost index for that PMSA; otherwise we use the cost index for the principal city of the parent CMSA. We only have 2010 edition of the RS Means index.

columns 2, 3, and 4 of table 3.2, are strongly related to city size and positively correlated with land values. They also exhibit considerably less dispersion. The highest housing prices are in San Francisco, which are 9 times the lowest housing prices, in McAllen, TX. The highest construction prices are in New York City, 1.9 times the lowest, in Rocky Mount, NC.

3.3.3 Regulatory and Geographic Constraints

Our index of regulatory constraints is provided by the Wharton Residential Land Use Regulatory Index (WRLURI), described in Gyourko, Saiz, and Summers (2008). The index is constructed from the survey responses of municipal planning officials regarding the regulatory process. These responses form the basis of 11 subindices, coded so that higher scores correspond to greater regulatory stringency: the approval delay index (ADI), the local political pressure index (LPPI), the state political involvement index (SPII), the open space index (OSI), the exactions index (EI), the local project approval index (LPAI), the local assembly index (LAI), the density restrictions index (DRI), the supply restriction index (SRI), the state court involvement index (SCII), and the local zoning approval index (LZAI). The authors construct a single aggregate WRLURI index through factor analysis: we consider both the aggregate index and the subindices in our analysis, each of which we r-enormalize as $z-$scores, with a mean of zero and standard deviation one, as weighted by the housing units in our sample. Typically, the WRLURI subindices are positively correlated, but not always; for instance, the SCII is negatively correlated with five of the other subindices.

Our index of geographic constraints is provided by Saiz (2010), who uses satellite imagery to calculate land scarcity in metropolitan areas. The index measures the fraction of undevelopable land within a 50 km radius of the city center, where land is undevelopable if it is i) covered by water or wetlands, or ii) has a slope of 15 degrees or steeper. While this land is not actually built on, it serves as a proxy for geographic features that may lower housing productivity. We consider both Saiz's aggregate index and his separate indices based on solid and flat land, each of which is renormalized as a $z-$score.

According to the aggregate indices, reported in columns 5 and 6, the most regulated land is in Boulder, CO, and the least regulated is in Glens Falls, NY; the most geographically constrained is in Santa Barbara, CA, and the least is in Lubbock, TX.

3.4 Cost-Function Estimates

Below, we use the indices from section 3.3 to test and estimate the cost function presented in section 3.2, and examine how it is influenced by geography and regulation using both aggregated and disaggregated measures. We restrict our analysis to MSAs with at least 10 land-sale observations, and years with at least 5. For our main estimates, the MSAs must also have available WRLURI, Saiz and construction-price indices, leaving 206 MSAs and 856 MSA-years.

3.4.1 Estimates and Tests of the Model

Figure 3.2 plots metropolitan housing prices against land values. The simple regression line, weighted by the number of housing units in our sample, has a slope of 0.59; if there were no other cost or productivity differences across cities, this number would estimate the cost share of land, ϕ_L, assuming CD production. The convex curvature in the quadratic regression yields an imprecise estimate of the elasticity of substitution of 0.18.[15] Of course, this regression is biased, as land values are positively correlated with construction prices and geographic and regulatory constraints. This figure illustrates how housing productivity is inferred by the vertical distance between a marker and the regression line. Accordingly, San Francisco has low housing

[15]In levels, the cost curve must be weakly concave, but the log-linearized cost curve is convex if $\sigma^Y < 1$, although the convexity is limited as $\sigma^Y \geq 0$ implies $\beta_3 \leq 0.5\beta_1(1-\beta_1)$.

productivity and Las Vegas has high housing productivity.

To illustrate differences in construction prices, we plot them against land values in figure 3.3. We use these data to estimate a cost surface shown in figure 3.4 without controls. As in figure 3.2, cities with housing prices above this surface are inferred to have lower housing productivity. Figure 3.3 plots the level curves for the surface in 3B, which correspond to the zero-profit conditions (ZPCs) for housing producers, seen in equation (3.4). These curves correspond to fixed sums of housing prices and productivities, $\hat{p}_j + \hat{A}_j^Y$, with curves further to the upper-right corresponding to higher sums. With the log-linearization, the slope of the ZPC is the ratio of land cost shares to non-land cost shares, $-\phi_j^L/(1-\phi_j^L)$. In the CD case, this slope is constant, as illustrated by the solid line; with an elasticity, σ^Y, of less than one, the slope of the ZPC increases with land values, as the land-cost share rises with land prices, as illustrated by the dashed curves.

Columns 1 and 2 of table 3.3 present cost-function estimates using the aggregate geographic and regulatory indices, assuming CD production, as in (3.10); column 2 imposes the restriction of CRS in (3.8), which is barely rejected at the 5% level. The CRS restriction is not rejected in the more flexible translog equation, presented in columns 3 and 4. The restricted regression in column 4 estimates the elasticity of substitution σ^Y to be 0.37. While we cannot reject the CD restriction (3.10) jointly with the CRS restriction (3.8), our interpretation of the evidence is that the restricted translog equation in column 4 describes the data best and provides fairly good evidence that σ^Y is less than one.

The OLS estimates in columns 1 through 4 produce stable values of 0.37 for the cost-share of land parameter, ϕ_L. Furthermore, we find that a one standard deviation increase in the geographic and regulatory indices predict a 9- and 8-percent increase in housing costs, respectively. These effects are consistent with our theory of housing productivity and the belief that geographic and regulatory constraints impede the production of housing services.

Columns 5 and 6 present our IV estimates, which use inverse distance to a salt-water coast and average winter temperature as instruments for the differentials $(\hat{r}-\hat{v})$ and $(\hat{r}-\hat{v})^2$. in the restricted equation (3.11) with $\gamma_2 = 0$. Column 5 imposes the CD restriction, $\beta_3 = 0$ and only uses the coastal instrument. Estimates of the first-stage, presented in table 3.8, reveal that these instruments are strong, with F-statistics of 64 in column 5, and 15 and 17 in column 6. The IV estimates are largely consistent with our OLS estimates, but less precise. The last row of table 3.3 reports the Chi-squared test of regressor endogeneity, in the spirit of Hausman (1978): these tests do not reject the null of regressor exogeneity at any standard size. The consistency of the IV estimates requires that distance-to-coast and winter temperature are uncorrelated with housing productivity, conditional on measures of geography and regulation. This assumption may be violated, as it may be difficult to build housing in extreme temperatures. We believe our IVs are much more strongly related to quality of life and trade productivity than to housing productivity, and should produce mostly exogenous variation in land values, as expressed in (3.16). The similarity of our OLS and IV estimates is reassuring and so we proceed under the assumption that the OLS estimates are consistent.

We test the assumption that the productivity shifters are factor neutral in column 7. This allows γ_2 to be non-zero in equation (3.11) by interacting the differential $(\hat{r}-\hat{v})$ with the geographic and regulatory indices. This interaction does not produce significant estimates of γ_2 and does not change our other estimates significantly. While this test of factor bias is imperfect, the evidence suggests that factor neutrality is not strongly at odds with the data.

Finally, in column 8, we use an alternative measure of non-land input prices, namely wage levels in the construction industry. The results in column 8 are quite similar to those in column 4. We perform a number of additional robustness checks in table 3.9. We split the sample into two periods: a ”housing-boom” period, from 2005 to 2007, and a ”housing-bust” period, from 2008 to 2010. We also use alternative land-value indices, one using only residential land, a second not controlling for proposed use or lot size, and another not shrinking the land-value index. The last two robustness checks drop observations in our low-growth areas. The results of these robustness checks, discussed in Appendix 3.7.3, reveal that the regression parameters are surprisingly stable over these specifications.

3.4.2 Disaggregating the Regulatory and Geographic Indices

As discussed above, the WRLURI regulatory index aggregates 11 subindices, while the Saiz index aggregates two. The factor loading of each of the WRLURI subindices in the aggregate index is reported in column 1 of table 3.4, ordered according to its factor load. Alongside, in column 2, are coefficient estimates from a regression of the aggregate WRLURI $z-$score on the $z-$scores for the subindices. These coefficients differ slightly from the factor loads because of differences in samples and weights. Column 3 presents similar estimates for the Saiz subindices. The coefficients on these measures are negative because the subindices indicate land that may be available for development.

The specification in column 4 is identical to the specification in column 4 of table 3.3, but with the disaggregated regulatory and geographic subindices. The results indicate that approval delays, local political pressure, state political involvement, supply restrictions, and state court involvement are all associated with economically significant reductions in housing productivity, ranging between 3- to 7- percent for a one-standard deviation increase. All five subindices are statistically significant at the 10-percent level, although only the last three are significant at 5 percent: these tests may lack precision because of the high degree of correlation between the subindices. None of the subindices has a significantly negative coefficient. The first three subindices are roughly consistent with the factor loading; the last two, for supply restrictions and state court involvement, appear to be of greater importance than a single-factor model captures.

Both of the Saiz subindices have statistically and economically significant negative coefficients. The estimates imply that a one standard-deviation increase in the share of flat or solid land is associated with a 7- to 9-percent reduction in housing costs.

Overall, the results of these regressions are encouraging. The estimated cost share of land and the elasticity of substitution between land and other inputs into housing production in our regressions are quite plausible, and the coefficients on the regulatory and geographic variables have the predicted signs and reasonable magnitudes. The tight fit of the cost-function specification, as measured by the R^2 values approaching 90 percent, implies that even our imperfect measures of input prices and observable constraints explain the variation in housing prices across metro areas quite well.

As our favored specification, we take the one from column 4 of table 3.4 – with CRS, factor-neutrality, non-unitary σ^Y, and disaggregated subindices – and use it for our subsequent analysis. It provides a value of $\phi_L = 0.33$ and $\sigma^Y = 0.49$. Using formula (3.5), this implies that the cost share of land ranges from 11 percent in Rochester to 48 percent in New York City.

3.5 Housing Productivity across Metropolitan Areas

3.5.1 Productivity in Housing and Tradeables

In column 1 of table 3.5 we list our inferred measures of housing productivity from the favored specification, using both observed and unobserved components of housing productivity, i.e., $\hat{A}_j^Y = Z_j(-\hat{\gamma}) - \hat{\varepsilon}_j$; column 2 reports only the value of productivity predicted by the regulatory subindices, Z_j^R, i.e., $\hat{A}_j^{YR} = -\hat{\gamma}_1^R Z_j^R$. The cities with the most and least productive housing sectors are McAllen, TX and San Luis Obispo, CA. Among large metros, with over one million inhabitants the top five, excluding our low-growth sample, are Houston, Indianapolis, Kansas City, Fort Worth, and Columbus; the bottom five are San Francisco, San Jose Oakland, Los Angeles, and Orange County, all on California's coast. Along the East Coast, Bergen-Passaic and Boston are notably unproductive. Cities with average productivity include Phoenix, Chicago, and Miami. Somewhat surprisingly, New York City is in this group. Although work by Glaeser et al. (2005b) suggests this is not true of Manhattan, the New York PMSA includes all five boroughs and Westchester county, and houses nearly 10 million people.[16]

In addition, we provide estimates of trade productivity $\hat{A}_j^X$ and quality-of-life $\hat{Q}_j$ in columns 3 and 4, using formulas (3.13) and (3.12), calibrated with parameter values taken from Albouy (2009).[17] Housing productivity is plotted against trade productivity in figure 3.5. This figure draws level curves for total productivity averaged across the housing and tradeables sectors, weighted by their expenditure shares, according to formula (3.15).[18]

Our estimates of trade-productivity, based primarily on overall wage levels, are largely consistent with the previous literature.[19] Interestingly, trade productivity and

[16] See Table 3.10 for the values of the major indices and measures for all of the MSAs in our sample.

[17] These calibrated values are $\theta^L = 0.025$, $s_w = 0.75$, $\tau = 0.32$, $s_x = 0.64$. θ^N is set at 0.8 so that it is consistent with s_w. For the estimates of $\hat{Q}_j$, we account for price variation in both housing and non-housing goods. We measure cost differences in housing goods using the expenditure-share of housing, 0.18, times the housing-price differential $\hat{p}_j$. To account for non-housing goods, we use the share of 0.18 times the predicted value of housing net of productivity differences, setting $\hat{A}_Y^j = 0$, i.e., $\hat{p}_j - \hat{A}_Y^j = \phi_L \hat{r}_j + \phi_N \hat{w}_j$, the price of non-tradeable goods predicted by factor prices alone. Furthermore, we subtract a sixth of housing-price costs to account for the tax-benefits of owner-occupied housing. This procedure yields a cost-of-living index roughly consistent with that of Albouy (2009). Our method of accounting for non-housing costs helps to avoid problems of division bias in subsequent analysis, where we regress measures of quality of life, inferred from high housing prices, with measures of housing productivity, inferred from low housing prices.

[18] The estimated productivities are positively related to the housing supply elasticities provided by Saiz (2010): a 1-point increase in productivity predicts a 1.94-point (s.e. = 0.24) increase in the supply elasticity ($R^2 = 0.41$).

[19] Also shown in Figure 3.5 is a line which depicts the bias to trade productivity estimates if land values are proxied with housing values, assuming housing productivities are uniform across cities (see Albouy 2009). Cities along this line would be inferred to have the same trade productivity, as cities with higher housing productivity have housing values low relative to land values, leading to

housing productivity are negatively, rather than positively, correlated. According to the regression line, a 1-point increase in trade-productivity predicts a 1.7-point decrease in housing productivity. For instance, cities in the San Francisco Bay Area have among the highest levels of trade productivity and the lowest levels of housing productivity. On the other hand, Houston, Fort Worth, and Atlanta are relatively more productive in housing than in tradeables. The large metro area with the greatest overall productivity is New York; that with the least is Tucson.

The negative relationship between trade and housing productivity estimates may stem from differing scale economies at the city level. While trade productivity is known to increase with city size (e.g., Rosenthal and Strange, 2004), it is possible that economies of scale in housing may be decreasing, possibly because of negative externalities in production from congestion, regulation, or other sources. It may be more difficult for producers to build new housing in already crowded environments, such as on a lot surrounded by other structures. New construction may impose negative externalities in consumption on incumbent residents, e.g., by blocking views or increasing traffic. Aware of this, residents may seek to constrain housing development to limit these externalities through regulation, lowering housing productivity.

We explore this hypothesis in table 3.6, which examines the relationship of productivity with population levels, aggregated at the consolidated metropolitan (CMSA) level, in panel A, or population density, in panel B. In column 1, the positive elasticities of trade productivity with respect to population of roughly 6 percent are consistent with those in the literature. The results in column 2 reveal negative elasticities, nearly 8 percent in magnitude. According to the results in column 3, which uses only the housing productivity component predicted by the regulatory subindices, about half of this relationship results from greater regulation. Overall productivity, examined in column 4, increases with population, but much more weakly than trade

lower inferred measures of trade productivity.

productivity. The results in column 5 suggest that this relationship would be stronger if the greater regulation associated with higher populations were held constant. As we explore in the next section, holding the regulatory environment constant could have negative consequences for urban quality of life.

3.5.2 Housing Productivity and Quality of Life

The model of section 3.2 predicts that if the sole effects of regulations were to reduce housing productivity, then they would increase housing prices while reducing land values, unambiguously reducing welfare (Albouy 2009). Ostensibly, the purpose of land-use regulations is to raise housing values by "recogniz[ing] local externalities, providing amenities that make communities more attractive," (Quigley and Rosenthal 2005) i.e., by raising demand, rather than by limiting supply, giving rise to terms such as "externality zoning." To our knowledge, there are only a few, limited estimates of the benefits of these regulations, e.g. Cheshire and Sheppard (2002) and Glaeser et al. (2005b), both of which suggest that the welfare costs of regulation outweigh the benefits.

To examine this hypothesis we relate our quality-of-life and housing-productivity estimates, shown in figure 3.6. The regression line in this figure suggests that a one-point decrease in housing productivity is associated with a 0.1-point increase in quality of life. If we accept the relationship as causal, the net welfare benefit of this trade-off, measured as a fraction of total consumption, equals this 0.1-point increase, minus the one-point decrease multiplied by the expenditure share of housing, which we calibrate as 0.18. Thus, a one-point decrease in housing productivity results in a net welfare loss of 0.08-percent of consumption. These results help to rationalize the existence of welfare-reducing regulations, if the benefits accrue to incumbent residents, who control the political process, while the costs are borne by potential residents, who do not have a local political voice.[20]

We explore this relationship further in table 3.7, which controls for possible confounding factors and isolates housing productivity predicted by regulation. The odd numbered columns include controls for natural amenities, such as climate, adjacency to the coast, and the geographic constraint index; the even numbered columns add controls for artificial amenities, such as the population level, density, education, crime rates, and number of eating and drinking establishments. In columns 1 and 2, these controls undo the relationship, as geographic amenities are related negatively to productivity and positively to quality of life. When we focus on productivity predicted by regulation, in columns 3 and 4, the original relationship is restored, although it is slightly weaker. As before, if these results are interpreted causally, the impact of land-use regulations is on net welfare-reducing.

Non-causal explanations for the relationship in table 3.7 are also plausible. For instance, residents in areas with unobserved amenities may simply elect to regulate land-use for reasons unrelated to urban quality of life. Alternatively, with preference heterogeneity, the quality-of-life measure represents the willingness-to-pay of the marginal resident. In cities with low-housing productivity, the supply of housing is effectively constrained, raising the willingness-to-pay of the marginal resident, much as in the "Superstar City" hypothesis of Gyourko, Mayer, and Sinai (2006). However, the negative relationship between productivity and quality of life appears to hold for more than a small subset of superstar cities.

[20]The net welfare loss from regulations implies that land should lose value while housing gains value. While property owners should in the long run seek to maximize the value of their land, frictions, due to moving costs and the immobility of housing capital, may cause most owners to maximize the value of their housing stock over their voting time horizons.

3.6 Conclusion

Our novel index of land values seems to contain important information not captured by typical indices of housing prices. As theory would predict, the variation of land values is greater than that of housing, and ultimately implies an average cost share of land of approximately one-third. The housing-cost model performs surprisingly well at explaining housing prices despite the disparate data sources. Our empirical model is consistent with constant returns to scale at the firm level, with an elasticity of substitution between land and non-land inputs of roughly one-half. This implies that the cost of share land rises from as low as 11 percent in low-value areas to 48 percent in high-value areas.

The housing-cost function modeled above provides the previously untested hypothesis that geographic and regulatory constraints will increase the wedge between the prices of housing and its inputs. The data strongly support this hypothesis and may provide guidance as to which regulations have the greatest impact on housing costs at the metropolitan level. Furthermore, our parsimonious model explains nearly 90 percent of the variation in metropolitan housing prices and our instrumental variable estimates provide reassurance that our ordinary least squares estimates are likely consistent. In general, the plausibility of the indices and the reasonableness of the empirical results are mutually reinforcing.

The pattern of housing productivity across metropolitan areas is also illuminating. Cities that are productive in tradeables sectors tend to be less productive in housing as the two appear to subject to opposite economies of scale. Larger cities have lower housing productivity, much of which seems attributable to greater regulation. These regulatory costs are associated cross-sectionally with a higher quality of life for residents, although this relationship is weak, suggesting that land-use regulations lead to net welfare costs for the economy as a whole.

3.7 Appendix

3.7.1 Factor-Specific Productivity Biases

When housing productivity is factor specific we may write the production function for housing as $Y_j = F^Y(L, M; A_j^Y) = F^Y(A_j^{YL} L, A_j^{YM} M; 1)$. The first-order log-linear approximation of the production function around the national average is

$$\hat{p}_j = \phi^L \hat{r}_j + (1-\phi^L)\hat{v}_j - [\phi^L \hat{A}_j^{YL} + (1-\phi^L)\hat{A}_j^{YM}]$$

As both $\hat{A}_j^{YL}$ and $\hat{A}_j^{YM}$ are only in the residual, it is difficult to identify them separately. The second-order log-linear approximation of the production function is

$$\hat{p}_j = \phi^L(\hat{r}_j - \hat{A}_j^{YL}) + (1-\phi^L)(\hat{v}_j - \hat{A}_j^{YM}) + (1/2)\phi^L(1-\phi^L)(1-\sigma^Y)(\hat{r}_j - \hat{A}_j^{YL} - \hat{v}_j + \hat{A}_j^{YM})^2 \tag{3.17}$$

$$= \phi^L \hat{r}_j + (1-\phi^L)\hat{v}_j + (1/2)\phi^L(1-\phi^L)(1-\sigma^Y)(\hat{r}_j - \hat{v}_j)^2$$
$$+ \phi^L(1-\phi^L)(1-\sigma^Y)(\hat{r}_j - \hat{v}_j)(\hat{A}_j^{YM} - \hat{A}_j^{YL})$$
$$- [\phi^L \hat{A}_j^{YL} + (1-\phi^L)\hat{A}_j^{YM}] + (1/2)\phi^L(1-\phi^L)(1-\sigma^Y)(\hat{A}_j^{YL} - \hat{A}_j^{YM})^2$$

The terms on the second-to-last line demonstrate that if $\sigma^Y < 1$, then productivity improvements that affect land more will exhibit a negative interaction with the rent variable and a positive interaction with the material price, while productivity improvements that affect material use more, will exhibit the opposite. Therefore, if a productivity shifter Z_j, biases productivity so that $(\hat{A}_j^{YM} - \hat{A}_j^{YL}) = Z_j\zeta$,we may identify factor-specific productivity biases with the following reduced-form equation:

$$\hat{p}_j = \beta_1 \hat{r}_j + \beta_2 \hat{v}_j + \beta_3(\hat{r}_j)^2 + \beta_4(\hat{v}_j)^2 + \beta_5(\hat{r}_j\hat{v}_j) + \gamma_1 Z_j + \gamma_2 Z_j \hat{r}_j + \gamma_3 Z_j \hat{v}_j + \varepsilon_j \tag{3.18}$$

The model embodied in (3.17) imposes the restriction that $\gamma_2 = -\gamma_3 = \zeta\phi^L(1 - \phi^L)(1 - \sigma^Y)$.

3.7.2 Wage and Housing Price Indices

The wage and housing price indices are estimated from the 2005 to 2010 American Community Survey, which samples 1% of the United States population every year. The indices are estimated with separate regressions for each year. For the wage regressions, we include all workers who live in an MSA and were employed in the last year, and reported positive wage and salary income. We calculate hours worked as average weekly hours times the midpoint of one of six bins for weeks worked in the past year. We then divide wage and salary income for the year by our calculated hours worked variable to find an hourly wage. We regress the log hourly wage on a set of MSA dummies and a number of individual covariates, each of which is interacted with gender:

- 12 indicators of educational attainment;
- a quartic in potential experience and potential experience interacted with years of education;
- age and age squared;
- 9 indicators of industry at the one-digit level (1950 classification);
- 9 indicators of employment at the one-digit level (1950 classification);
- 5 indicators of marital status (married with spouse present, married with spouse absent, divorced, widowed, separated);
- an indicator for veteran status, and veteran status interacted with age;

- 5 indicators of minority status (Black, Hispanic, Asian, Native American, and other);
- an indicator of immigrant status, years since immigration, and immigrant status interacted with black, Hispanic, Asian, and other;
- 2 indicators for English proficiency (none or poor).

This regression is first run using census-person weights. From the regressions a predicted wage is calculated using individual characteristics alone, controlling for MSA, to form a new weight equal to the predicted wage times the census-person weight. These new income-adjusted weights allow us to weight workers by their income share. The new weights are then used in a second regression, which is used to calculate the city-wage indices from the MSA indicator variables, renormalized to have a national average of zero every year. In practice, this weighting procedure has only a small effect on the estimated wage differentials. All of the wage regressions are at the CMSA level rather than the PMSA level to reflect the ability of workers to commute relatively easily to jobs throughout a CMSA.

To calculate construction wage differentials, we drop all non-construction workers and follow the same procedure as above. We define the construction sector as occupation codes 620 through 676 in the ACS 2000-2007 occupation codes. In our sample, 4.5% of all workers are in the construction sector.

The housing price index of an MSA is calculated in a manner similar to the differential wage, by regressing housing prices on a set of covariates. The covariates used in the regression for the adjusted housing cost differential are:

- survey year dummies;
- 9 indicators of building size;

- 9 indicators for the number of rooms, 5 indicators for the number of bedrooms, and number of rooms interacted with number of bedrooms;
- 3 indicators for lot size;
- 13 indicators for when the building was built;
- 2 indicators for complete plumbing and kitchen facilities;
- an indicator for commercial use;
- an indicator for condominium status (owned units only).

A regression of housing values on housing characteristics and MSA indicator variables is first run weighting by census-housing weights. A new value-adjusted weight is calculated by multiplying the census-housing weights by the predicted value from this first regression using housing characteristics alone, controlling for MSA. A second regression is run using these new weights on the housing characteristics, along with the MSA indicators. The housing-price indices are taken from the MSA indicator variables in this second regression, renormalized to have a national average of zero every year. As with the wage differentials, this adjusted weighting method has only a small impact on the measured price differentials. In contrast to the wage regressions, the housing price regressions were run at the PMSA level rather than the CMSA level to achieve a better geographic match between the housing stock and the underlying land.

3.7.3 Estimate Stability

We conduct several exercises in order to gauge the stability of our estimates; the results of these exercises are reported in table 3.9. First, we split the sample into two periods: a "housing-boom" period, from 2005 to 2007, and a "housing-bust" period, from 2008 to 2010. As seen in columns 2 and 3, the regression results for the split samples are not statistically different from those in the pooled sample, in column 1. Comparing the two split samples, the latter period does appear to have a somewhat lower elasticity of substitution and weaker effects of geographic and regulatory constraints. Whether this is a product of sampling error or secular changes in housing production remains to be seen.

Second, we report results for the same regressions using three alternative land-value indices: i) residential land values only, ii) "raw" land-value indices, iii) unshrunken land-value indices. Land is defined as residential if its proposed use is listed as single-family, multi-family, or apartments. Raw land-value indices are procured by regressing log price per acre on a set of MSA indicators without any additional covariates, such as proposed use or lot size, and are not reweighted by location, corresponding to the regression in column 1 of table 3.1. The unshrunken indices are derived directly from the regression in column 4 of table 3.1, without applying the Kane and Staiger (2008) shrinkage technique. The results for the residential land values in column 4 are nearly identical to those in column 1. In columns 5 and 6, the estimated land share is lower as we see more dispersion in the land index, which appears to cause attenuation effects: the first, due to noise introduced by not controlling for observable characteristics; the second, from sampling error.

The results in column 7 drop observations that we deemed to have low growth, i.e. metro areas with population growth from 1980 to 2010 in the bottom decile. The estimated cost share of land and the elasticity of substitution using this sample is slightly lower, albeit not significantly. However, in a regression using our favored specification, with all of the regulatory and geographic subindices, not shown, the results are more similar. If, as in column 8, we instead define our low-growth sample using the bottom decile of MSAs in terms of the building permits issued from 2005 to 2010 relative to the size of the housing stock, the results are quite close to our base specification.

Table 3.1: Land Value Index Regressions

	Fraction of Sample	Dependent Variable: Log Price per Acre			
	(0)	(1)	(2)	(3)	(4)
Log lot size (acres)			-0.660	-0.647	-0.597
			(0.002)	(0.002)	(0.003)
No proposed use	15.9%			-0.198	-0.332
				(0.012)	(0.014)
Proposed use: commercial	0.3%			-0.369	-0.252
				(0.064)	(0.077)
Proposed use: industrial	7.5%			-0.316	-0.522
				(0.015)	(0.019)
Proposed use: retail	8.1%			0.260	0.211
				(0.014)	(0.017)
Proposed use: single-family	10.7%			-0.020	-0.188
				(0.014)	(0.020)
Proposed use: multi-family	3.3%			-0.071	-0.174
				(0.021)	(0.019)
Proposed use: office	6.3%			0.072	0.185
				(0.016)	(0.020)
Proposed use: apartment	3.6%			0.468	0.366
				(0.021)	(0.016)
Proposed use: hold for development	19.2%			-0.069	-0.085
				(0.013)	(0.013)
Proposed use: hold for investment	4.3%			-0.358	-0.287
				(0.020)	(0.027)
Proposed use: mixed use	1.7%			0.373	0.407
				(0.028)	(0.027)
Proposed use: medical	1.0%			0.162	-0.038
				(0.038)	(0.051)
Proposed use: parking	0.9%			0.181	0.253
				(0.039)	(0.033)
Number of Observations		68,757	68,757	68,757	68,757
Adjusted R-squared		0.301	0.699	0.711	0.762
Weighted by Predicted Density		No	No	No	Yes

Robust standard errors, clustered by MSA/PMSA, reported in parentheses. Land-value data from CoStar COMPS database for years 2005 to 2010. All specifications include a full set of interacted MSA and year-of-sale indicator (not shown). Predicted density is number of land sales predicted by a geographical model of housing units relative to city center; please see section 3.1, Land Values, for a full description.

Table 3.2: Measures for Selected Metropolitan Areas, Ranked by Land-Value Differential: 2005-2010

Name of Area	Population	Observed No. of Land Sales	Land Value	Housing Price	Wages (Const. Only)	Const. Price Index	Regulation Index (z-score)	Geo Avail. Index (z-score)	Land Value Rank
Metropolitan Areas:									
New York, NY PMSA	9,747,281	1,603	1.67	0.84	0.25	0.31	0.66	0.56	1
San Francisco, CA PMSA	1,785,097	152	1.49	1.29	0.21	0.23	0.77	2.17	3
San Jose, CA PMSA	1,784,642	217	1.31	1.08	0.21	0.18	-0.02	1.71	4
Orange County, CA PMSA	3,026,786	233	1.24	0.93	0.12	0.10	0.20	1.15	5
Los Angeles-Long Beach, CA PMSA	9,848,011	1,760	1.00	0.86	0.12	0.10	0.38	1.15	7
Washington, DC-MD-VA-WV PMSA	5,650,154	1,840	0.71	0.39	0.18	0.01	0.30	-0.74	16
Boston, MA-NH PMSA	3,552,421	122	0.52	0.62	0.10	0.18	2.18	0.24	21
Chicago, IL PMSA	8,710,824	3,511	0.24	0.14	0.06	0.17	-0.32	0.54	35
Phoenix-Mesa, AZ MSA	4,364,094	5,946	0.23	-0.03	-0.01	-0.10	0.64	-0.74	37
Philadelphia, PA-NJ PMSA	5,332,822	859	0.12	0.02	0.05	0.16	1.31	-0.93	43
Riverside-San Bernardino, CA PMSA	4,143,113	2,452	0.02	0.22	0.12	0.07	0.47	0.44	51
Atlanta, GA MSA	5,315,841	5,229	-0.16	-0.32	0.03	-0.10	-0.29	-1.23	74
Houston, TX PMSA	5,219,317	1,143	-0.37	-0.54	0.04	-0.12	-0.92	-1.01	107
Dallas, TX PMSA	4,399,895	811	-0.40	-0.46	0.00	-0.14	-0.69	-0.98	114
Detroit, MI PMSA*	4,373,040	679	-0.45	-0.35	-0.04	0.05	-0.27	-0.22	118
Saginaw-Bay City-Midland, MI MSA*	390,032	41	-1.74	-0.63	-0.16	-0.03	-0.35	-0.62	213
Utica-Rome, NY MSA*	293,280	15	-1.82	-0.58	-0.27	-0.05	-1.20	-0.56	214
Rochester, NY MSA*	1,093,434	110	-1.89	-0.54	-0.07	0.01	-0.42	0.07	215
Metropolitan Population:									
Less than 500,000	30,837,205	5,017	-0.53	-0.22	-0.08	-0.39	-0.03	-0.05	4
500,000 to 1,500,000	55,777,644	13,942	-0.42	-0.20	-0.07	-0.26	-0.16	-0.06	3
1,500,000 to 5,000,000	89,173,333	32,032	0.15	0.07	0.01	0.11	0.17	0.01	2
5,000,000+	49,824,250	15,945	0.61	0.31	0.11	0.20	0.01	0.10	1
Standard Deviations (pop. wtd.)			0.76	0.51	0.16	0.96	1.01	0.14	
Correlation with land values (pop. wtd.)			1.00	0.89	0.66	0.61	0.52	0.57	

Land-value data from CoStar COMPS database for years 2005 to 2010. Wage and housing-price data from 2005 to 2010 American Community Survey 1-percent samples. Wage differentials based on the average logarithm of hourly wages. Housing-price differentials based on the average logarithm of prices of owner-occupied units. Regulation Index is the Wharton Residential Land Use Regulatory Index (WRLURI) from Gyourko et al. (2008). Geographic Availability Index is the Land Unavailability Index from Saiz (2010). Construction-price index from R.S. Means. MSAs with asterisks after their names are in the weighted bottom 10% of our sample in population growth from 1980-2010.

Table 3.3: Model of Housing Costs with Aggregate Geographic and Regulatory Indices

Specification	Basic Cobb-Douglas (1)	Restricted Cobb-Douglas (2)	Translog (3)	Restricted Translog (4)	Restricted Cobb-Douglas (Instr. Var.) (5)	Restricted Translog (Instr. Var.) (6)	Non-neutral Productivity Translog (7)	Restricted Translog w/ Constr Wages (8)
Land-Value Differential	0.365	0.372	0.348	0.374	0.427	0.342	0.373	0.352
	(0.040)	(0.039)	(0.042)	(0.042)	(0.081)	(0.097)	(0.041)	(0.038)
Construction-Price Differential	0.965	0.628	0.887	0.626	0.573	0.658	0.627	0.648
	(0.170)	(0.039)	(0.177)	(0.042)	(0.081)	(0.097)	(0.041)	(0.038)
Land-Value Differential Squared			0.034	0.074		0.091	0.062	0.072
			(0.036)	(0.040)		(0.113)	(0.042)	(0.036)
Construction-Price Differential Squared			-1.095	0.074		0.091	0.062	0.072
			(1.145)	(0.040)		(0.113)	(0.042)	(0.036)
Land-Value Differential X Construction-Price Differential			0.384	-0.148		-0.182	-0.124	-0.144
			(0.367)	(0.080)		(0.226)	(1.077)	(0.072)
Geographic Constraint Index: z-score	0.091	0.099	0.091	0.089	0.083	0.095	0.088	0.107
	(0.027)	(0.028)	(0.026)	(0.029)	(0.042)	(0.035)	(0.025)	(0.027)
Regulatory Index: z-score	0.060	0.076	0.075	0.082	0.065	0.090	0.085	0.078
	(0.018)	(0.015)	(0.014)	(0.015)	(0.023)	(0.028)	(0.016)	(0.015)
Geographic Constraint Index times Land Value Differential minus Construction							0.007	
							(0.050)	
Regulatory Index times Land Value Differential minus Construction Price							0.022	
							(0.026)	
Constant	0.001	0.001	-0.016	-0.032	0.001	-0.032	-0.035	-0.031
	(0.024)	(0.025)	(0.039)	(0.029)	(0.024)	(0.010)	(0.028)	(0.026)
Number of Observations	856	856	856	856	206	856	856	888
Number of MSAs	206	206	206	206	206	206	206	215
Adjusted R-squared	0.854	0.848	0.863	0.857	0.799	0.856	0.793	0.847
p -value for CRS restrictions		0.046		0.511				0.019
p -value for CD restrictions	0.552	0.149						
p- value for all restrictions		0.126						
Elasticity of Substitution	1.000	1.000		0.367	1.000	0.189	0.470	0.371
				(0.321)		(1.074)	(0.347)	(0.296)
p-value from Chi-squared test of regressor endogeneity					0.571	0.674		

Dependent variable in all regressions is the housing price index. Robust standard errors, clustered by CMSA, reported in parentheses. Data sources are described in Table 2. Restricted model specifications require that the production function exhibits constant returns to scale (CRS). Cobb-Douglas (CD) restrictions impose that the squared and interacted differential coefficients equal zero (the elasticity of substitution between factors equals 1). Instrumental variables are the inverse mean distance from the sea and average winter temperature; first-stage regressions are reported in table A1.

Table 3.4: Model of Housing Costs with Disaggregated Geographic and Regulatory Indices

Specification	Regulatory Index Factor Loading			Restricted Translog w Cons Price
Dependent Variable		Reg Index	Geo Index	Hous. Price
	(1)	(2)	(3)	(4)
Land-Value Differential				0.326
				(0.029)
Land-Value Differential Squared				0.056
				(0.024)
Approval Delay: z-score	0.29	0.489		0.074
		(0.034)		(0.044)
Local Political Pressure: z-score	0.22	0.200		0.055
		(0.061)		(0.033)
State Political Involvement: z-score	0.22	0.373		0.040
		(0.023)		(0.020)
Open Space: z-score	0.18	0.014		-0.001
		(0.007)		(0.003)
Exactions: z-score	0.15	-0.039		0.032
		(0.066)		(0.054)
Local Project Approval: z-score	0.15	0.218		0.004
		(0.017)		(0.018)
Local Assembly: z-score	0.14	0.147		-0.022
		(0.041)		(0.024)
Density Restrictions: z-score	0.09	0.107		-0.049
		(0.065)		(0.040)
Supply Restrictions: z-score	0.02	0.132		0.028
		(0.009)		(0.010)
State Court Involvement: z-score	-0.03	-0.130		0.049
		(0.018)		(0.018)
Local Zoning Approval: z-score	-0.04	-0.093		-0.025
		(0.062)		(0.037)
Flat Land Share: z-score			-0.493	-0.088
			(0.035)	(0.023)
Solid Land Share: z-score			-0.787	-0.069
			(0.059)	(0.022)
Constant		0.000	0.000	-0.024
		(0.020)	(0.042)	(0.023)
Number of Observations		890	890	856
Adjusted R-squared		0.948	0.846	0.894
Elasticity of Substitution				0.488
				(0.207)

Robust standard errors, clustered by CMSA, reported in parentheses. Data sources are described in table 2; constituent components of Wharton Residential Land Use Regulatory Index (WRLURI) are from Gyourko et al (2008). Constituent components of geographical index are from Saiz (2010).

Table 3.5: Inferred Indices of Selected Metropolitan Areas, Ranked by Total Amenity Value

	Housing Productivity				
	Total (Including Indices)	Predicted by Regulation Subindices	Trade Productivity	Quality of Life	Total Amenity Value
	(1)	(2)	(3)	(4)	(5)
Metropolitan Areas:					
New York, NY PMSA	-0.002	-0.058	0.148	0.090	0.184
San Francisco, CA PMSA	-0.578	-0.171	0.200	0.083	0.107
San Jose, CA PMSA	-0.484	-0.047	0.196	0.063	0.101
Orange County, CA PMSA	-0.411	-0.061	0.095	0.089	0.076
Los Angeles-Long Beach, CA PMSA	-0.447	-0.120	0.089	0.072	0.048
Washington, DC-MD-VA-WV PMSA	-0.138	-0.086	0.115	0.019	0.068
Boston, MA-NH PMSA	-0.342	-0.335	0.085	0.030	0.023
Chicago, IL PMSA	0.029	0.090	0.048	0.004	0.040
Phoenix-Mesa, AZ MSA	0.021	-0.092	-0.004	0.016	0.017
Philadelphia, PA-NJ PMSA	0.113	-0.022	0.054	-0.011	0.044
Riverside-San Bernardino, CA PMSA	-0.188	-0.094	0.065	-0.016	-0.008
Atlanta, GA MSA	0.184	0.006	-0.015	-0.022	0.001
Houston, TX PMSA	0.318	0.084	-0.002	-0.051	0.005
Dallas, TX PMSA	0.217	0.104	-0.023	-0.041	-0.017
Detroit, MI PMSA*	0.227	0.005	-0.015	-0.040	-0.008
Saginaw-Bay City-Midland, MI MSA*	0.184	0.009	-0.130	-0.095	-0.145
Utica-Rome, NY MSA*	0.105	0.225	-0.082	-0.108	-0.142
Rochester, NY MSA*	0.108	0.140	-0.121	-0.108	-0.166
Metropolitan Population:					
Less than 500,000	0.013	0.034	-0.061	-0.021	-0.059
500,000 to 1,500,000	0.031	0.029	-0.051	-0.017	-0.045
1,500,000 to 5,000,000	-0.020	-0.019	0.015	0.008	0.014
5,000,000+	-0.034	-0.022	0.072	0.026	0.066
United States	0.234	0.131	0.087	0.045	0.067
		standard deviations (population weighted)			

Housing productivity, in column 1 is calculated from the specification in column 4 of table 4, as the negative of the sum of the regression residual plus the housing price predicted by the WRLURI and Saiz subindices. Housing productivity predicted by regulation is based upon the projection of housing prices on the WRLURI subindices. Trade productivity is calculated as 0.8 times the overall wage differential plus 0.025 times the land-value differential. Refer to section 5 of the text for the calculation of quality-of-life estimates . Quality of life and total amenity value are expressed as a fraction of average pre-tax household income.

Table 3.6: Productivity in Tradeable and Housing Sectors According to Metropolitan Population and Density

	Dependent Variable				
	Trade Productivity (1)	Housing Productivity (2)	Hous. Prod. Predicted by Regulation (3)	Total Productivity (4)	Total Productivity No Reg. (5)
Panel A: Population					
Log of Population	0.055	-0.075	-0.034	0.022	0.028
	(0.004)	(0.023)	(0.007)	(0.004)	(0.004)
Number of Observations	206	206	206	206	206
Adjusted R-squared	0.646	0.166	0.156	0.436	0.548
Panel B: Population Density					
Weighted Density Differential	0.063	-0.078	-0.042	0.026	0.034
	(0.004)	(0.031)	(0.011)	(0.005)	(0.004)
Number of Observations	206	206	206	206	206
Adjusted R-squared	0.44	0.091	0.121	0.333	0.421

Robust standard errors, clustered by CMSA, reported in parentheses. Trade and housing productivity differentials are calculated as in table 5. Total productivity is calculated as 0.18 times housing productivity plus 0.64 times trade productivity. Weighted density differential is calculated as the population density at the census-tract level, weighted by population.

Table 3.7: Quality of Life and Housing Productivity

	Dependent Variable: Quality of Life			
Housing Productivity Measure:	Total Housing Productivity		Housing Productivity Predicted by Regulation	
	(1)	(2)	(3)	(4)
Housing Productivity	0.001	0.035	-0.087	-0.049
	(0.029)	(0.023)	(0.027)	(0.029)
Natural Controls	X	X	X	X
Artificial Controls		X		X
Number of Observations	201	201	201	201
Adjusted R-squared	0.54	0.73	0.57	0.73

Robust standard errors, clustered by CMSA, in parentheses. Quality of life is calculated as in table 6. Housing productivity predicted by regulation is calculated as in table 5. Natural controls: heating and cooling degree days, July humidity, annual sunshine, annual precipitation, adjacency to coast, geographic constraint index. Artificial controls include metropolitan population, density, eating and drinking establishments, violent crime rate, and fractions with a college degree, some college, and high-school degree. Both sets of controls are from Albouy et al. (2012).

Table 3.8: Instrumental Variables Estimates, First-Stage Regressions

Dependent Variable	Land Rent minus Construction Price (1)	Land Rent minus Construction Price (2)	Land Rent minus Construction Price Squared (3)
Geographic Constraint Index: z-score	0.191	0.129	0.078
	(0.057)	(0.054)	(0.065)
Regulatory Index: z-score	0.143	0.168	-0.151
	(0.044)	(0.040)	(0.037)
Inverse of Mean Distance from Sea: z-score	0.237	0.225	0.226
	(0.030)	(0.050)	(0.052)
Mean Winter Temperature: z-score		0.181	-0.092
		(0.045)	(0.060)
Constant	0.003	0.004	0.402
	(0.053)	(0.045)	(0.048)
Number of Observations	206	206	206
Adjusted R-squared	0.486	0.555	0.184
F-statistic of Excluded Instruments	64.0	15.1	17.1
First Stage Regression for	Table 3 Column 5	Table 3 Column 6	Table 3 Column 6

Robust standard errors, clustered by CMSA, in parentheses. Inverse of mean distance from sea and mean winter temperature are from Albouy et al. (2012).

Table 3.9: Model of Housing Costs, Robustness Analysis

Specification Dependent Variable	Base Specification Hous. Price (1)	2005-2007 Sample Hous. Price (2)	2008-2010 Sample Hous. Price (3)	Residential Land Sample Hous. Price (4)	Raw Land Values Hous. Price (5)	Unshrunken Land Values Hous. Price (6)	High Population Growth Sample Hous. Price (7)	High Building Permits Sample Hous. Price (8)
Land-Value Differential	0.374	0.366	0.380	0.382	0.270	0.284	0.317	0.381
	(0.042)	(0.047)	(0.041)	(0.020)	(0.014)	(0.040)	(0.039)	(0.036)
Construction-Price Differential	0.626	0.634	0.620	0.618	0.730	0.716	0.683	0.619
	(0.042)	(0.047)	(0.041)	(0.020)	(0.014)	(0.040)	(0.039)	(0.036)
Land-Value Differential Squared	0.074	0.051	0.097	0.131	-0.031	0.056	0.106	0.078
	(0.040)	(0.044)	(0.041)	(.038)	(.019)	(.023)	(0.050)	(0.053)
Construction-Price Differential Squared	0.074	0.051	0.097	0.131	-0.031	0.056	0.106	0.078
	(0.040)	(0.044)	(0.041)	(.038)	(.019)	(.023)	(0.050)	(0.053)
Land-Value Differential X Construction-Price Differential	-0.148	-0.102	-0.194	-0.262	0.062	-0.112	-0.212	-0.156
	(.08)	(.088)	(.082)	(.076)	(.038)	(.046)	(0.100)	(0.106)
Geographic Constraint Index: z-score	0.089	0.115	0.065	0.082	0.094	0.117	0.104	0.085
	(.029)	(.035)	(.025)	(0.011)	(0.011)	(0.034)	(0.031)	(0.029)
Regulatory Index: z-score	0.082	0.098	0.070	0.081	0.094	0.090	0.081	0.078
	(.015)	(.018)	(.015)	(0.008)	(0.007)	(0.016)	(0.014)	(0.015)
Constant	-0.031	-0.041	-0.030	-0.011	0.015	-0.032	-0.026	-0.037
	(0.029)	(0.034)	(0.028)	(0.010)	(0.011)	(0.032)	(0.030)	(0.030)
Number of Observations	853	337	516	678	853	853	754	731
Adjusted R-squared	0.793	0.786	0.812	0.856	0.845	0.744	0.786	0.795
Elasticity of Subsitution	0.367	0.558	0.175	0.447	1.156	0.449	0.019	0.338
	(0.321)	(0.366)	(0.326)	(0.343)	(0.216)	(0.198)	(0.437)	(0.438)

Robust standard errors, clustered by CMSA, reported in parentheses. Regressions correspond to column 4 of Table 3. See appendix C for discussion.

Table 3.10: List of Metropolitan Indices Ranked by Land Price Differential, 2005-2010

				Adjusted Differentials					*Raw Differentials*			*Productivity*		
Full Name	Population	Census Division	Obs. Land Sales	Land Value	Land Value (No Wts.)	Housing Price	Wages (All)	Wages (Const. Only)	Reg. Index (z-score)	Geo Avail. Index (z-score)	Const. Price Index	Housing	Tradeables	Land Value Rank
Metropolitan Areas:														
New York, NY PMSA	9,747,281	2	1,603	1.673	1.546	0.842	0.152	0.247	0.661	0.561	0.307	-0.002	0.148	1
Jersey City, NJ PMSA	597,924	2	43	1.588	1.457	0.643	0.159	0.262	0.103	0.236	0.118	0.054	0.150	2
San Francisco, CA PMSA	1,785,097	9	152	1.488	1.416	1.287	0.207	0.214	0.769	2.169	0.234	-0.578	0.200	3
San Jose, CA PMSA	1,784,642	9	217	1.307	1.273	1.081	0.207	0.214	-0.018	1.710	0.178	-0.484	0.196	4
Orange County, CA PMSA	3,026,786	9	233	1.235	1.192	0.930	0.087	0.115	0.196	1.153	0.097	-0.411	0.095	5
Stamford-Norwalk, CT PMSA	361,024	1	19	1.103	1.020	0.972	0.157	0.243	-0.326	0.561	0.134	-0.493	0.139	6
Los Angeles-Long Beach, CA PMSA	9,848,011	9	1,760	1.000	0.961	0.862	0.087	0.115	0.384	1.153	0.097	-0.447	0.089	7
Miami, FL PMSA	2,500,625	5	1,233	0.924	0.909	0.288	-0.057	-0.081	1.104	2.341	-0.075	-0.002	-0.018	8
Seattle-Bellevue-Everett, WA PMSA	2,692,066	9	1,626	0.884	0.848	0.420	0.051	0.035	1.013	0.719	0.065	-0.071	0.065	9
Fort Lauderdale, FL PMSA	1,766,476	5	741	0.871	0.863	0.179	-0.057	-0.082	0.763	2.295	-0.099	0.068	-0.020	10
San Diego, CA MSA	3,053,793	9	957	0.869	0.878	0.737	0.061	0.093	0.276	1.691	0.065	-0.395	0.065	11
Oakland, CA PMSA	2,532,756	9	132	0.857	0.781	0.850	0.192	0.186	0.608	1.605	0.171	-0.448	0.175	12
Santa Barbara-Santa Maria-Lompoc, CA MSA	407,057	9	29	0.836	0.746	0.945	0.073	-0.066	0.994	2.802	0.075	-0.612	0.100	13
Ventura, CA PMSA	802,983	9	131	0.737	0.757	0.803	0.086	0.115	1.503	2.489	0.080	-0.506	0.083	14
Las Vegas, NV-AZ MSA	2,141,893	8	2,553	0.732	0.756	-0.014	0.044	-0.063	-1.392	0.150	-0.110	0.199	0.070	15
Washington, DC-MD-VA-WV PMSA	5,650,154	5	1,840	0.714	0.728	0.387	0.132	0.177	0.305	-0.741	0.008	-0.138	0.115	16
Bergen-Passaic, NJ PMSA	1,387,028	2	79	0.692	0.652	0.684	0.152	0.247	0.684	0.561	0.118	-0.382	0.123	17
Nassau-Suffolk, NY PMSA	2,875,904	2	396	0.687	0.663	0.663	0.152	0.247	0.738	0.561	0.307	-0.246	0.123	18
Santa Rosa, CA PMSA	472,102	9	153	0.567	0.584	0.816	0.207	0.214	1.683	1.671	0.133	-0.552	0.177	19
Naples, FL MSA	318,537	5	78	0.547	0.572	0.490	0.008	-0.213	0.104	2.291			0.053	20
Boston, MA-NH PMSA	3,552,421	1	122	0.517	0.497	0.619	0.093	0.101	2.177	0.241	0.179	-0.342	0.085	21
West Palm Beach-Boca Raton, FL MSA	1,279,950	5	321	0.513	0.419	0.133	-0.033	0.046	0.097	1.720	-0.126	-0.051	-0.026	22
Vallejo-Fairfield-Napa, CA PMSA	541,884	9	146	0.437	0.465	0.529	0.207	0.214	1.134	0.990	0.120	-0.320	0.174	23
Tacoma, WA PMSA	796,836	9	539	0.414	0.414	0.080	0.050	0.033	1.822	0.377	0.042	0.070	0.053	24
Newark, NJ PMSA*	2,045,344	2	142	0.401	0.357	0.545	0.159	0.262	0.698	0.073	0.132	-0.345	0.120	25
Sarasota-Bradenton, FL MSA	688,126	5	601	0.398	0.420	0.128	-0.103	-0.061	1.069	1.849	-0.095	-0.070	-0.078	26
Portland-Vancouver, OR-WA PMSA	2,230,947	9	1,191	0.378	0.361	0.141	-0.049	-0.071	0.071	0.419	0.012	-0.025	-0.026	27
San Luis Obispo-Atascadero-Paso Robles, CA MSA	266,971	9	43	0.361	0.267	0.786	0.003	-0.112	1.364	1.810	0.038	-0.659	0.029	28
Orlando, FL MSA	2,082,421	5	1,612	0.347	0.382	-0.116	-0.088	-0.121	0.146	0.350	-0.092	0.156	-0.056	29
Salinas, CA MSA	410,370	9	12	0.306	0.238	0.644	0.113	-0.035	0.098	1.824	0.096	-0.501	0.120	30
Provo-Orem, UT MSA	545,307	8	47	0.294	0.265	-0.201	-0.128	-0.217	-0.019	1.503	-0.140	0.189	-0.081	31
Middlesex-Somerset-Hunterdon, NJ PMSA	1,247,641	2	101	0.289	0.277	0.444	0.152	0.247	1.377	0.561	0.118	-0.288	0.113	32
Tampa-St. Petersburg-Clearwater, FL MSA	2,747,272	5	1,220	0.271	0.288	-0.116	-0.091	-0.140	-0.671	0.621	-0.061	0.146	-0.058	33
Olympia, WA PMSA	250,979	9	250	0.252	0.245	0.095	0.060	0.054	0.539	0.466	0.031	-0.011	0.055	34
Chicago, IL PMSA	8,710,824	3	3,511	0.239	0.259	0.140	0.054	0.058	-0.316	0.541	0.169	0.029	0.048	35
Reading, PA MSA	407,125	2	36	0.230	0.246	-0.259	-0.050	-0.046	0.504	-0.616	0.015	0.323	-0.035	36
Phoenix-Mesa, AZ MSA	4,364,094	8	5,946	0.226	0.296	-0.030	-0.013	-0.014	0.637	-0.741	-0.099	0.021	-0.004	37
Trenton, NJ PMSA	366,222	2	35	0.225	0.258	0.323	0.159	0.262	2.338	-0.847	0.114	-0.193	0.116	38
Baltimore, MD PMSA	2,690,886	5	802	0.203	0.212	0.165	0.131	0.177	2.087	-0.351	-0.061	-0.157	0.102	39
Melbourne-Titusville-Palm Bay, FL MSA	536,357	5	420	0.152	0.189	-0.169	-0.112	-0.065	0.456	1.732	-0.074	0.149	-0.092	40
Stockton-Lodi, CA MSA	674,860	9	163	0.140	0.128	0.224	0.067	0.165	0.563	-0.834	0.079	-0.146	0.041	41
Salt Lake City-Ogden, UT MSA	1,567,650	8	145	0.124	0.104	-0.116	-0.078	-0.159	-0.383	2.113	-0.125	0.053	-0.046	42
Philadelphia, PA-NJ PMSA	5,332,822	2	859	0.124	0.146	0.016	0.062	0.045	1.306	-0.927	0.163	0.113	0.054	43
Fort Myers-Cape Coral, FL MSA	586,908	5	294	0.120	0.140	0.001	-0.078	-0.071	-0.565	1.187	-0.119	-0.062	-0.060	44
Jacksonville, FL MSA	1,301,808	5	793	0.110	0.139	-0.196	-0.076	-0.130	-0.371	0.900	-0.154	0.110	-0.050	45
Grand Junction, CO MSA	146,093	8	21	0.099	0.045	-0.008	-0.146	-0.060	0.367	0.702	-0.091	-0.042	-0.126	46
Reno, NV MSA	414,820	8	57	0.074	0.087	0.182	0.003	-0.178	-0.803	1.328	-0.022	-0.195	0.032	47
Albuquerque, NM MSA	841,428	8	114	0.041	0.020	-0.129	-0.087	-0.195	0.233	-0.854	-0.098	0.054	-0.051	48
Allentown-Bethlehem-Easton, PA MSA	706,374	2	85	0.035	0.037	-0.132	-0.040	0.079	-0.306	-0.401	0.056	0.159	-0.049	49
Wilmington-Newark, DE-MD PMSA	635,430	5	107	0.021	0.051	-0.030	0.062	0.045	0.377	-0.705	0.059	0.057	0.052	50
Riverside-San Bernardino, CA PMSA	4,143,113	9	2,452	0.016	0.054	0.219	0.087	0.115	0.465	0.437	0.072	-0.188	0.065	51
Minneapolis-St. Paul, MN-WI MSA	3,269,814	4	846	0.014	0.012	-0.042	0.024	-0.011	-0.074	-0.481	0.129	0.112	0.025	52
Lawrence, MA-NH PMSA	413,626	1	29	0.012	-0.028	0.338	0.099	0.121	1.920	0.241	0.135	-0.266	0.075	53
Denver, CO PMSA	2,445,781	8	2,015	0.006	0.000	-0.002	-0.010	0.003	0.968	-0.604	-0.038	-0.043	-0.009	54
Kenosha, WI PMSA	165,382	3	58	0.001	0.041	-0.125	0.059	0.068	0.962	0.929	0.011	0.110	0.046	55
Fort Pierce-Port St. Lucie, FL MSA	406,296	5	71	-0.012	-0.017	-0.084	-0.075	-0.191	0.343	1.765			-0.042	56
Boulder-Longmont, CO PMSA	311,786	8	183	-0.013	0.001	0.204	-0.010	0.003	3.626	0.695	-0.088	-0.288	-0.010	57
Modesto, CA MSA	510,385	9	142	-0.020	-0.028	0.120	0.041	0.044	0.040	-0.724	0.080	-0.095	0.031	58
Daytona Beach, FL MSA	587,512	5	93	-0.036	-0.022	-0.143	-0.135	-0.303	0.393	1.549	-0.106	0.038	-0.082	59

Table 2.10: List of Metropolitan Indices Ranked by Land Price Differential, 2005-2010
page 2 of 4

				Adjusted Differentials					Raw Differentials			Productivity		
Full Name	Population	Census Division	Obs. Land Sales	Land Value	Land Value (No Wts.)	Housing Price	Wages (All)	Wages (Const. Only)	Reg. Index (z-score)	Geo Avail. Index (z-score)	Const. Price Index	Housing	Tradeables	Land Value Rank
Sioux Falls, SD MSA	224,266	4	17	-0.053	-0.146	-0.556	-0.182	0.181	-1.804	-1.260	-0.188	0.389	-0.202	60
Visalia-Tulare-Porterville, CA MSA	429,668	9	32	-0.061	-0.162	-0.038	-0.016	-0.019	0.362	-0.471			-0.014	61
Raleigh-Durham-Chapel Hill, NC MSA	1,589,388	5	782	-0.077	-0.060	-0.205	-0.042	-0.033	0.427	-1.027	-0.231	0.004	-0.037	62
New Haven-Meriden, CT PMSA	558,692	1	43	-0.100	-0.135	0.291	0.159	0.262	-0.330	0.786	0.108	-0.271	0.108	63
Eugene-Springfield, OR MSA	351,109	9	36	-0.108	-0.173	0.105	-0.154	-0.239	0.187	1.647	0.000	-0.163	-0.112	64
Norfolk-Virginia Beach-Newport News, VA- MSA	1,667,410	5	392	-0.109	-0.092	0.029	-0.078	-0.053	-0.154	1.512	-0.117	-0.164	-0.068	65
Monmouth-Ocean, NJ PMSA	1,217,783	2	124	-0.117	-0.122	0.369	0.152	0.247	2.078	0.561	0.307	-0.211	0.103	66
Huntsville, AL MSA	406,316	6	29	-0.120	-0.115	-0.437	-0.055	-0.177	-2.258	-0.230	-0.147	0.275	-0.028	67
Salem, OR PMSA	396,103	9	54	-0.130	-0.182	-0.050	-0.042	-0.063	0.281	0.199	0.004	-0.009	-0.034	68
Newburgh, NY-PA PMSA	444,061	2	54	-0.132	-0.120	0.169	0.166	0.265	-0.407	0.047	0.163	-0.120	0.113	69
Lakeland-Winter Haven, FL MSA	583,403	5	561	-0.138	-0.080	-0.359	-0.138	-0.201	-0.067	0.156	-0.070	0.246	-0.103	70
Savannah, GA MSA	343,092	5	64	-0.139	-0.144	-0.206	-0.102	-0.168	-1.136	1.528	-0.178	0.018	-0.074	71
Bridgeport, CT PMSA	470,094	1	26	-0.156	-0.253	0.473	0.157	0.254	0.392	0.561	0.107	-0.469	0.105	72
Merced, CA MSA	245,321	9	64	-0.157	-0.139	-0.013	-0.005	0.255	0.636	-0.927			-0.047	73
Atlanta, GA MSA	5,315,841	5	5,229	-0.162	-0.113	-0.325	-0.007	0.033	-0.288	-1.225	-0.098	0.184	-0.015	74
Madison, WI MSA	491,357	3	239	-0.164	-0.150	-0.033	-0.070	-0.180	0.280	-0.869	-0.002	-0.041	-0.042	75
Brockton, MA PMSA	268,092	1	22	-0.166	-0.170	0.329	0.097	0.096	3.044	0.241	0.121	-0.319	0.073	76
Colorado Springs, CO MSA	604,542	8	892	-0.170	-0.198	-0.231	-0.134	-0.150	0.987	-0.331	-0.070	0.110	-0.108	77
Janesville-Beloit, WI MSA	160,155	3	15	-0.184	-0.255	-0.382	-0.172	-0.046	-0.758	-1.190	-0.017	0.288	-0.160	78
Hamilton-Middletown, OH PMSA	363,184	3	151	-0.203	-0.172	-0.466	-0.052	-0.040	-0.301	-1.083	-0.088	0.319	-0.048	79
Medford-Ashland, OR MSA	201,286	9	12	-0.205	-0.338	0.286	-0.164	-0.046	0.957	2.003	0.004	-0.372	-0.153	80
Lincoln, NE MSA	281,531	4	24	-0.217	-0.264	-0.444	-0.212	-0.189	0.824	-1.348	-0.115	0.272	-0.177	81
Bremerton, WA PMSA	240,862	9	18	-0.217	-0.249	0.238	0.060	0.063	0.536	1.124	0.065	-0.285	0.042	82
Fargo-Moorhead, ND-MN MSA	200,102	4	13	-0.227	-0.378	-0.515	-0.188	-0.340	-2.275	-1.281	-0.164	0.307	-0.131	83
Tucson, AZ MSA	1,020,200	8	1,749	-0.228	-0.230	-0.071	-0.123	-0.178	1.978	-0.292	-0.134	-0.115	-0.095	84
Fort Collins-Loveland, CO MSA	298,382	8	344	-0.231	-0.216	-0.079	-0.116	-0.216	0.920	0.110	-0.085	-0.073	-0.083	85
Champaign-Urbana, IL MSA	195,671	3	22	-0.237	-0.273	-0.387	-0.130	-0.289	-0.934	-1.355	0.046	0.321	-0.084	86
Rocky Mount, NC MSA	146,596	5	12	-0.244	-0.382	-0.643	-0.100	-0.275	-1.123	-0.519	-0.314	0.328	-0.058	87
Lexington, KY MSA	554,107	6	29	-0.249	-0.303	-0.421	-0.156	-0.100	-0.052	-1.136	-0.119	0.238	-0.138	88
New Orleans, LA MSA*	1,211,035	7	66	-0.249	-0.292	-0.194	-0.057	-0.156	-2.227	2.255	-0.110	0.018	-0.036	89
Danbury, CT PMSA	223,095	1	23	-0.273	-0.336	0.511	0.149	0.255	-0.600	0.561	0.307	-0.396	0.095	90
Nashville, TN MSA	1,495,419	6	455	-0.275	-0.252	-0.267	-0.066	-0.095	-0.956	-0.795	-0.118	0.078	-0.055	91
Atlantic-Cape May, NJ PMSA	367,803	2	37	-0.286	-0.307	0.246	0.074	0.055	0.713	1.778	0.102	-0.284	0.055	92
Asheville, NC MSA	251,894	5	41	-0.287	-0.302	-0.021	-0.174	-0.229	-1.268	1.891	-0.263	-0.273	-0.136	93
Lafayette, IN MSA	202,331	3	13	-0.289	-0.297	-0.571	-0.207	-0.454	-0.862	-0.147	-0.089	0.395	-0.133	94
Springfield, MA MSA*	609,993	1	28	-0.302	-0.362	0.058	-0.011	-0.007	0.702	-0.095	0.052	-0.138	-0.017	95
South Bend, IN MSA	267,613	3	12	-0.313	-0.410	-0.641	-0.182	0.057	-2.228	-0.907	-0.079	0.464	-0.189	96
Fresno, CA MSA	1,063,899	9	137	-0.337	-0.304	0.074	-0.013	-0.034	1.048	-0.793	0.082	-0.141	-0.016	97
Billings, MT MSA	144,797	8	25	-0.345	-0.383	-0.261	-0.198	-0.441	-0.632	-0.868	-0.094	0.065	-0.128	98
Austin-San Marcos, TX MSA	1,705,075	7	384	-0.347	-0.295	-0.303	-0.049	-0.069	-0.769	-1.241	-0.208	0.029	-0.045	99
Dutchess County, NY PMSA	293,562	2	33	-0.348	-0.304	0.325	0.170	0.262	0.205	0.561	0.163	-0.338	0.112	100
Worcester, MA-CT PMSA	547,274	1	56	-0.351	-0.359	0.191	0.093	0.101	3.051	0.241	0.111	-0.240	0.063	101
Columbus, OH MSA	1,718,303	3	671	-0.357	-0.319	-0.395	-0.058	-0.040	0.060	-1.303	-0.045	0.231	-0.058	102
Richland-Kennewick-Pasco, WA MSA	245,649	9	27	-0.357	-0.397	-0.318	0.022	0.146	0.866	-0.823	-0.036	0.160	-0.010	103
Cleveland-Lorain-Elyria, OH PMSA*	2,192,053	3	416	-0.358	-0.340	-0.417	-0.084	-0.115	-0.576	0.565	0.010	0.291	-0.070	104
Greeley, CO PMSA	254,759	8	320	-0.361	-0.303	-0.305	-0.010	0.002	-0.431	-0.930	-0.153	0.068	-0.019	105
Charleston-North Charleston, SC MSA	659,191	5	214	-0.367	-0.352	-0.085	-0.111	-0.107	-1.566	1.545	-0.188	-0.182	-0.098	106
Houston, TX PMSA	5,219,317	7	1,143	-0.367	-0.336	-0.539	0.014	0.036	-0.924	-1.013	-0.120	0.318	-0.002	107
Chattanooga, TN-GA MSA	510,388	6	51	-0.372	-0.373	-0.399	-0.141	-0.219	-1.436	-0.157	-0.147	0.158	-0.109	108
Richmond-Petersburg, VA MSA	1,119,459	5	399	-0.377	-0.354	-0.126	-0.012	-0.072	-0.913	-0.993	-0.122	-0.096	-0.010	109
La Crosse, WI-MN MSA	132,923	3	21	-0.388	-0.435	-0.334	-0.154	-0.258	0.293	0.333	-0.050	0.156	-0.116	110
Cincinnati, OH-KY-IN PMSA	1,776,911	3	637	-0.391	-0.366	-0.395	-0.052	-0.040	-1.262	-0.920	-0.073	0.203	-0.053	111
Gary, IN PMSA*	657,809	3	111	-0.394	-0.412	-0.425	0.060	0.068	-1.396	0.124	0.035	0.306	0.036	112
York, PA MSA	428,937	2	47	-0.396	-0.383	-0.217	-0.022	-0.107	0.991	-0.831	-0.020	0.057	-0.014	113
Dallas, TX PMSA	4,399,895	7	811	-0.403	-0.381	-0.461	-0.014	0.000	-0.692	-0.976	-0.140	0.217	-0.023	114
McAllen-Edinburg-Mission, TX MSA	741,152	7	61	-0.405	-0.409	-0.959	-0.205	-0.136	-1.021	-1.380	-0.261	0.629	-0.183	115
Hagerstown, MD PMSA	145,910	5	28	-0.415	-0.403	-0.033	0.136	0.171	0.171	-0.505	-0.270	-0.305	0.092	116
Fayetteville-Springdale-Rogers, AR MSA	425,685	7	43	-0.426	-0.410	-0.361	-0.118	-0.152	-0.953	-0.004	-0.271	0.017	-0.099	117
Detroit, MI PMSA*	4,373,040	3	679	-0.455	-0.442	-0.347	-0.010	-0.037	-0.265	-0.221	0.053	0.227	-0.015	118
Lancaster, PA MSA	507,766	2	57	-0.463	-0.451	-0.186	-0.083	-0.154	0.104	-0.841	-0.061	-0.020	-0.066	119
Boise City, ID MSA	571,271	8	106	-0.466	-0.461	-0.222	-0.136	-0.186	-1.042	0.361	-0.111	-0.019	-0.112	120

Table 2.10: List of Metropolitan Indices Ranked by Land Price Differential, 2005-2010

				Adjusted Differentials					Raw Differentials			Productivity		
Full Name	Population	Census Division	Obs. Land Sales	Land Value	Land Value (No Wts.)	Housing Price	Wages (All)	Wages (Const. Only)	Reg. Index (z-score)	Geo Avail. Index (z-score)	Const. Price Index	Housing	Tradeables	Land Value Rank
Fort Walton Beach, FL MSA	178,473	5	14	-0.467	-0.595	-0.201	-0.152	0.283	-0.446	1.457			-0.199	121
Milwaukee-Waukesha, WI PMSA	1,559,667	3	399	-0.468	-0.445	-0.076	-0.031	-0.015	0.360	0.628	0.052	-0.049	-0.039	122
Portsmouth-Rochester, NH-ME PMSA	262,128	1	13	-0.473	-0.545	0.185	0.095	0.171	1.448	0.241	-0.037	-0.377	0.052	123
Myrtle Beach, SC MSA	263,868	5	84	-0.474	-0.493	-0.211	-0.185	-0.128	-1.604	1.614			-0.168	124
Manchester, NH PMSA	212,326	1	23	-0.476	-0.537	0.132	0.095	0.171	1.854	0.241	-0.036	-0.325	0.052	125
Harrisburg-Lebanon-Carlisle, PA MSA	667,425	2	89	-0.478	-0.451	-0.344	-0.068	-0.006	0.440	-0.245	-0.010	0.171	-0.075	126
Galveston-Texas City, TX PMSA	286,814	7	39	-0.488	-0.498	-0.498	0.032	0.063	0.662	2.266	-0.139	0.229	0.009	127
Indianapolis, IN MSA	1,823,690	3	193	-0.514	-0.514	-0.500	-0.065	-0.119	-1.380	-1.355	-0.059	0.282	-0.056	128
Spokane, WA MSA	468,684	9	55	-0.538	-0.578	-0.203	-0.118	-0.148	0.711	-0.082	-0.052	-0.017	-0.102	129
Baton Rouge, LA MSA	685,419	7	99	-0.543	-0.548	-0.350	-0.069	-0.044	-1.577	0.222	-0.148	0.059	-0.072	130
Tulsa, OK MSA	873,304	7	245	-0.547	-0.527	-0.518	-0.129	-0.069	-1.474	-1.117	-0.218	0.177	-0.125	131
Akron, OH PMSA*	699,935	3	169	-0.550	-0.523	-0.444	-0.084	-0.114	-0.234	-1.109	-0.024	0.241	-0.075	132
Bakersfield, CA MSA	807,407	9	64	-0.578	-0.604	-0.083	0.044	-0.102	0.266	-0.236	0.065	-0.062	0.043	133
Louisville, KY-IN MSA	1,099,588	6	126	-0.579	-0.592	-0.366	-0.101	-0.141	-1.046	-0.802	-0.078	0.115	-0.088	134
Fort Worth-Arlington, TX PMSA	2,113,278	7	506	-0.583	-0.572	-0.569	-0.014	0.000	-0.728	-1.185	-0.171	0.252	-0.028	135
Roanoke, VA MSA	243,506	5	23	-0.591	-0.641	-0.267	-0.105	-0.138	-1.182	0.513	-0.159	-0.046	-0.093	136
Johnson City-Kingsport-Bristol, TN-VA MSA	503,010	6	28	-0.598	-0.634	-0.552	-0.217	-0.206	-1.674	1.292	-0.220	0.193	-0.189	137
Dayton-Springfield, OH MSA*	933,312	3	116	-0.599	-0.613	-0.510	-0.119	-0.193	-1.038	-1.375	-0.091	0.244	-0.098	138
Yuba City, CA MSA	165,539	9	13	-0.600	-0.735	-0.217	0.012	-0.066	-0.030	-0.743			0.007	139
Greensboro--Winston Salem--High Point, NC MSA	1,416,374	5	438	-0.602	-0.590	-0.408	-0.124	-0.190	-0.907	-1.273	-0.240	0.037	-0.104	140
Omaha, NE-IA MSA	799,130	4	118	-0.603	-0.618	-0.522	-0.126	-0.065	-1.189	-1.262	-0.092	0.257	-0.124	141
Gainesville, FL MSA	243,574	5	34	-0.605	-0.637	-0.128	-0.148	-0.215	-0.227	-0.669	-0.128	-0.166	-0.122	142
Racine, WI PMSA	200,601	3	80	-0.624	-0.597	-0.177	-0.028	0.024	-0.661	1.233	0.015	-0.014	-0.045	143
Springfield, MO MSA	383,637	4	43	-0.626	-0.643	-0.561	-0.201	-0.260	-1.483	-1.100	-0.097	0.284	-0.166	144
Rockford, IL MSA	409,058	3	104	-0.630	-0.564	-0.544	-0.089	-0.011	-1.145	-1.319	0.104	0.415	-0.099	145
St. Louis, MO-IL MSA	2,733,694	4	364	-0.655	-0.638	-0.314	-0.039	-0.126	-1.340	-0.881	0.044	0.135	-0.034	146
Hartford, CT MSA	1,231,125	1	101	-0.665	-0.664	0.199	0.090	0.091	0.389	-0.281	0.104	-0.336	0.054	147
El Paso, TX MSA	751,296	7	94	-0.670	-0.699	-0.659	-0.215	-0.182	0.773	-1.174	-0.233	0.270	-0.192	148
St. Cloud, MN MSA	189,148	4	17	-0.673	-0.818	-0.370	-0.113	-0.313	-0.512	-0.414	0.103	0.230	-0.076	149
Kansas City, MO-KS MSA	2,005,888	4	477	-0.683	-0.657	-0.458	-0.068	-0.086	-1.546	-1.140	0.049	0.277	-0.069	150
Amarillo, TX MSA	238,299	7	27	-0.685	-0.749	-0.556	-0.176	-0.414	-0.945	-1.254	-0.180	0.202	-0.120	151
Green Bay, WI MSA	247,319	3	49	-0.691	-0.700	-0.327	-0.076	-0.005	0.538	-0.282	-0.029	0.083	-0.089	152
Canton-Massillon, OH MSA*	408,005	3	40	-0.697	-0.735	-0.526	-0.122	0.048	-1.568	-0.808	-0.065	0.254	-0.140	153
Hickory-Morganton-Lenoir, NC MSA	365,364	5	88	-0.716	-0.687	-0.433	-0.201	-0.171	-1.216	-0.395	-0.303	-0.017	-0.181	154
Biloxi-Gulfport-Pascagoula, MS MSA	355,075	6	30	-0.720	-0.779	-0.449	-0.127	-0.019	-1.849	1.134	-0.177	0.088	-0.135	155
Brazoria, TX PMSA	309,208	7	62	-0.738	-0.730	-0.544	0.035	0.056	-1.127	-1.013	-0.120	0.221	0.006	156
Memphis, TN-AR-MS MSA	1,230,253	6	173	-0.745	-0.726	-0.529	-0.050	-0.075	1.464	-0.827	-0.136	0.194	-0.055	157
Knoxville, TN MSA	785,490	6	193	-0.801	-0.779	-0.410	-0.156	-0.175	-0.904	0.468	-0.200	0.011	-0.140	158
Ann Arbor, MI PMSA	630,518	3	136	-0.803	-0.762	-0.231	-0.010	-0.036	0.209	-0.949	0.019	-0.003	-0.024	159
Portland, ME MSA	256,178	1	25	-0.806	-0.881	0.135	-0.079	-0.156	1.658	1.005	-0.083	-0.448	-0.071	160
San Antonio, TX MSA	1,928,154	7	348	-0.809	-0.800	-0.592	-0.126	-0.136	-0.662	-1.270	-0.187	0.202	-0.118	161
Wichita, KS MSA	589,195	4	54	-0.828	-0.889	-0.643	-0.112	-0.183	-2.148	-1.345	-0.176	0.255	-0.099	162
Lake Charles, LA MSA	187,554	7	14	-0.830	-0.925	-0.606	-0.079	0.268	-1.936	0.980	-0.189	0.207	-0.136	163
Little Rock-North Little Rock, AR MSA	657,416	7	110	-0.846	-0.849	-0.459	-0.116	-0.181	-1.631	-0.752	-0.155	0.083	-0.103	164
Des Moines, IA MSA	536,664	4	99	-0.862	-0.855	-0.370	-0.079	-0.036	-1.618	-1.123	-0.111	0.023	-0.090	165
Pittsburgh, PA MSA*	2,287,106	2	240	-0.873	-0.844	-0.543	-0.097	-0.118	-0.187	0.050	0.016	0.295	-0.096	166
Fort Wayne, IN MSA	528,408	3	39	-0.874	-0.871	-0.687	-0.142	-0.142	-2.197	-1.300	-0.104	0.342	-0.134	167
Cedar Rapids, IA MSA	209,226	4	33	-0.881	-0.946	-0.490	-0.124	-0.159	-0.965	-1.252	-0.087	0.155	-0.115	168
Oklahoma City, OK MSA	1,213,704	7	395	-0.895	-0.905	-0.552	-0.154	-0.276	-0.857	-1.305	-0.174	0.155	-0.125	169
Jackson, MS MSA	483,852	6	43	-0.924	-0.940	-0.525	-0.091	-0.138	-1.450	-0.869	-0.152	0.132	-0.088	170
Birmingham, AL MSA	997,770	6	148	-0.928	-0.943	-0.401	-0.071	-0.141	-0.695	-0.720	-0.112	0.047	-0.069	171
Brownsville-Harlingen-San Benito, TX MSA	396,371	7	52	-0.928	-0.954	-0.957	-0.239	-0.280	-1.795	-0.068			-0.206	172
Wausau, WI MSA	131,612	3	16	-0.931	-1.013	-0.448	-0.132	-0.040	-0.134	-0.843	-0.059	0.123	-0.142	173
Greenville-Spartanburg-Anderson, SC MSA	1,096,009	5	507	-0.946	-0.923	-0.482	-0.126	-0.174	-1.763	-0.794	-0.251	0.008	-0.116	174
Scranton--Wilkes-Barre--Hazleton, PA MSA*	614,565	2	27	-0.956	-1.006	-0.505	-0.147	-0.257	-0.325	-0.011	0.018	0.235	-0.124	175
Augusta-Aiken, GA-SC MSA	516,357	5	66	-0.956	-0.971	-0.500	-0.104	-0.102	-2.002	-0.913	-0.166	0.089	-0.107	176
Fayetteville, NC MSA	315,207	5	25	-0.960	-1.000	-0.577	-0.191	-0.259	-1.112	-0.663	-0.228	0.118	-0.165	177
Pensacola, FL MSA	455,102	5	102	-0.962	-0.980	-0.338	-0.173	-0.162	-1.644	1.158	-0.136	-0.051	-0.163	178
Tyler, TX MSA	204,665	7	13	-0.969	-1.057	-0.543	-0.146	-0.862	-0.099	-0.930	-0.203	0.098	-0.030	179
Macon, GA MSA	356,873	5	20	-0.971	-0.989	-0.607	-0.076	-0.097	-1.617	-1.038	-0.171	0.187	-0.081	180
Davenport-Moline-Rock Island, IA-IL MSA*	362,790	4	28	-0.974	-1.018	-0.534	-0.111	0.184	-1.729	-1.201	-0.050	0.207	-0.158	181

Table 2.10: List of Metropolitan Indices Ranked by Land Price Differential, 2005-2010
page 4 of 4

				Adjusted Differentials					Raw Differentials			Productivity		
Full Name	Population	Census Division	Obs. Land Sales	Land Value	Land Value (No Wts.)	Housing Price	Wages (All)	Wages (Const. Only)	Reg. Index (z-score)	Geo Avail. Index (z-score)	Const. Price Index	Housing	Tradeables	Land Value Rank
Columbia, SC MSA	627,630	5	139	-0.977	-0.983	-0.433	-0.136	-0.141	-1.503	-0.678	-0.234	-0.032	-0.131	182
Lynchburg, VA MSA	232,895	5	13	-0.990	-1.045	-0.407	-0.168	-0.147	-1.596	-0.329	-0.150	-0.002	-0.161	183
Duluth-Superior, MN-WI MSA*	242,041	4	22	-0.993	-1.046	-0.481	-0.149	-0.058	-0.896	0.266	0.075	0.248	-0.157	184
Benton Harbor, MI MSA*	160,472	3	12	-1.011	-1.120	-0.337	-0.142	0.075	-0.979	1.040			-0.171	185
Corpus Christi, TX MSA	391,269	7	74	-1.018	-1.047	-0.634	-0.154	-0.155	-0.711	0.443	-0.229	0.160	-0.147	186
Montgomery, AL MSA	354,108	6	33	-1.031	-1.086	-0.528	-0.136	-0.138	-1.849	-0.897	-0.203	0.071	-0.134	187
Lubbock, TX MSA	270,550	7	45	-1.035	-1.039	-0.629	-0.185	-0.207	-1.811	-1.404	-0.208	0.167	-0.169	188
Bryan-College Station, TX MSA	179,992	7	34	-1.037	-1.058	-0.437	-0.178	-0.634	0.249	-1.110	-0.195	-0.017	-0.097	189
Buffalo-Niagara Falls, NY MSA*	1,123,804	2	104	-1.064	-1.054	-0.595	-0.076	-0.067	-0.646	-0.490	0.034	0.315	-0.089	190
Lafayette, LA MSA	415,592	7	15	-1.082	-1.169	-0.572	-0.118	-0.065	-1.472	-1.327	-0.189	0.113	-0.129	191
Grand Rapids-Muskegon-Holland, MI MSA	1,157,672	3	121	-1.122	-1.144	-0.450	-0.113	-0.176	-0.549	-0.970	-0.121	0.035	-0.108	192
Beaumont-Port Arthur, TX MSA*	378,477	7	60	-1.148	-1.167	-0.733	-0.051	-0.259	-1.309	-0.499	-0.176	0.271	-0.037	193
Killeen-Temple, TX MSA	358,316	7	32	-1.161	-1.187	-0.693	-0.180	-0.219	-1.876	-1.262	-0.259	0.161	-0.166	194
Lansing-East Lansing, MI MSA	453,603	3	40	-1.162	-1.164	-0.453	-0.112	-0.083	-0.043	-1.089	-0.010	0.118	-0.122	195
St. Joseph, MO MSA*	106,908	4	12	-1.163	-1.251	-0.621	-0.112	0.000	-2.636	-1.119	-0.049	0.254	-0.029	196
Kalamazoo-Battle Creek, MI MSA	462,250	3	31	-1.171	-1.175	-0.519	-0.124	-0.110	-0.210	-0.941	-0.061	0.141	-0.129	197
Mobile, AL MSA	591,599	6	135	-1.184	-1.186	-0.465	-0.147	-0.272	-1.860	0.015	-0.154	0.012	-0.127	198
Flint, MI PMSA*	424,043	3	85	-1.187	-1.152	-0.682	-0.010	-0.037	-0.818	-0.955	-0.009	0.344	-0.033	199
Longview-Marshall, TX MSA	222,489	7	14	-1.209	-1.245	-0.621	-0.123	-0.377	-2.295	-0.903	-0.289	0.055	-0.089	200
Erie, PA MSA*	280,291	2	29	-1.304	-1.354	-0.613	-0.163	-0.211	-1.292	1.080	-0.041	0.226	-0.155	201
Appleton-Oshkosh-Neenah, WI MSA	385,264	3	79	-1.311	-1.294	-0.366	-0.090	-0.058	-0.602	-0.544	-0.067	-0.044	-0.109	202
Toledo, OH MSA*	631,275	3	107	-1.325	-1.322	-0.542	-0.096	-0.217	-1.201	-0.494	-0.009	0.179	-0.091	203
Fort Smith, AR-OK MSA	225,132	7	18	-1.344	-1.409	-0.652	-0.222	-0.200	-1.928	-0.454	-0.204	0.125	-0.214	204
Albany-Schenectady-Troy, NY MSA	906,208	2	120	-1.389	-1.369	-0.175	-0.026	-0.067	-0.457	-0.279	-0.003	-0.196	-0.049	205
Peoria-Pekin, IL MSA*	357,144	3	25	-1.425	-1.447	-0.515	-0.060	0.003	-0.913	-1.181	0.045	0.180	-0.092	206
Syracuse, NY MSA*	725,610	2	65	-1.460	-1.456	-0.597	-0.091	-0.040	-1.231	-0.549	-0.015	0.208	-0.116	207
Binghamton, NY MSA*	244,694	2	16	-1.502	-1.571	-0.701	-0.104	0.088	-1.090	0.262	-0.034	0.285	-0.150	208
Youngstown-Warren, OH MSA*	554,614	3	49	-1.515	-1.526	-0.741	-0.169	-0.232	-0.920	-0.909	-0.038	0.321	-0.163	209
Evansville-Henderson, IN-KY MSA	305,455	3	33	-1.539	-1.667	-0.633	-0.136	-0.347	-1.932	-1.000	-0.070	0.187	-0.114	210
Sioux City, IA-NE MSA*	123,482	4	17	-1.678	-1.770	-0.824	-0.227	-0.590	-2.041	-1.276	-0.148	0.284	-0.166	211
Glens Falls, NY MSA	128,774	2	21	-1.707	-1.681	-0.293	-0.131	-0.292	-2.784	0.583	-0.068	-0.182	-0.122	212
Saginaw-Bay City-Midland, MI MSA*	390,032	3	41	-1.744	-1.763	-0.631	-0.118	-0.165	-0.354	-0.620	-0.034	0.184	-0.130	213
Utica-Rome, NY MSA*	293,280	2	15	-1.817	-1.889	-0.577	-0.082	-0.267	-1.199	-0.556	-0.047	0.105	-0.082	214
Rochester, NY MSA*	1,093,434	2	110	-1.891	-1.896	-0.540	-0.089	-0.071	-0.424	0.071	0.007	0.108	-0.121	215
Census Divisions:														
New England	8,966,068	1	530	0.042	0.010	0.421	0.092	0.117	1.435	0.240	0.126	-0.330	0.070	5
Middle Atlantic	36,338,768	2	4,567	0.315	0.278	0.260	0.072	0.121	0.537	0.080	0.165	-0.014	0.057	2
East North Central	34,261,505	3	8,520	-0.411	-0.402	-0.272	-0.036	-0.047	-0.529	-0.294	0.031	0.158	-0.037	6
West North Central	12,257,516	4	2,184	-0.501	-0.510	-0.337	-0.061	-0.083	-1.058	-0.897	0.019	0.185	-0.057	7
South Atlantic	41,912,174	5	19,501	0.097	0.109	-0.052	-0.031	-0.032	-0.042	0.177	-0.102	0.011	-0.022	4
East South Central	9,366,975	6	1,473	-0.637	-0.642	-0.421	-0.105	-0.138	-0.834	-0.418	-0.140	0.116	-0.094	9
West South Central	25,989,458	7	4,804	-0.571	-0.565	-0.522	-0.066	-0.081	-0.949	-0.794	-0.164	0.208	-0.064	8
Mountain	15,672,803	8	14,517	0.140	0.157	-0.058	-0.040	-0.084	0.300	-0.046	-0.095	0.026	-0.021	3
Pacific	40,847,165	9	10,840	0.669	0.643	0.604	0.080	0.089	0.514	0.988	0.089	-0.321	0.079	1
Metropolitan Population:														
Less than 500,000	30,837,205		5,017	-0.526	-0.556	-0.221	-0.063	-0.077	-0.389	-0.026	-0.054	0.013	-0.061	4
500,000 to 1,500,000	55,777,644		13,942	-0.419	-0.423	-0.204	-0.053	-0.066	-0.261	-0.159	-0.059	0.031	-0.051	3
1,500,000 to 5,000,000	89,173,333		32,032	0.155	0.159	0.072	0.013	0.009	0.106	0.174	0.006	-0.020	0.015	2
5,000,000+	49,824,250		15,945	0.605	0.589	0.314	0.078	0.113	0.197	0.012	0.104	-0.034	0.072	1

Figure 3.1: The effect of low productivity or low substitutability on housing prices

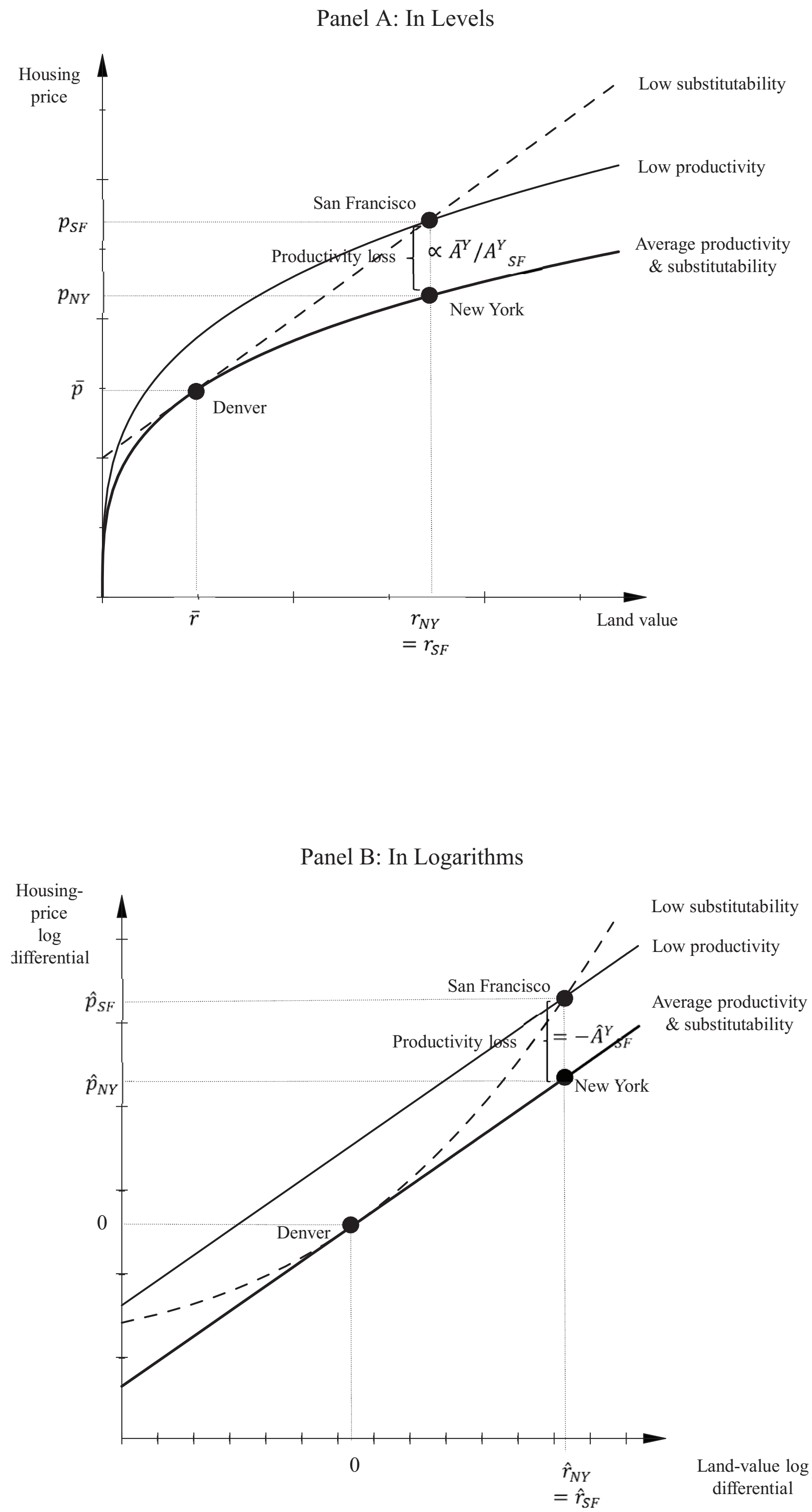


Figure 3.2: Housing Prices vs. Land Values

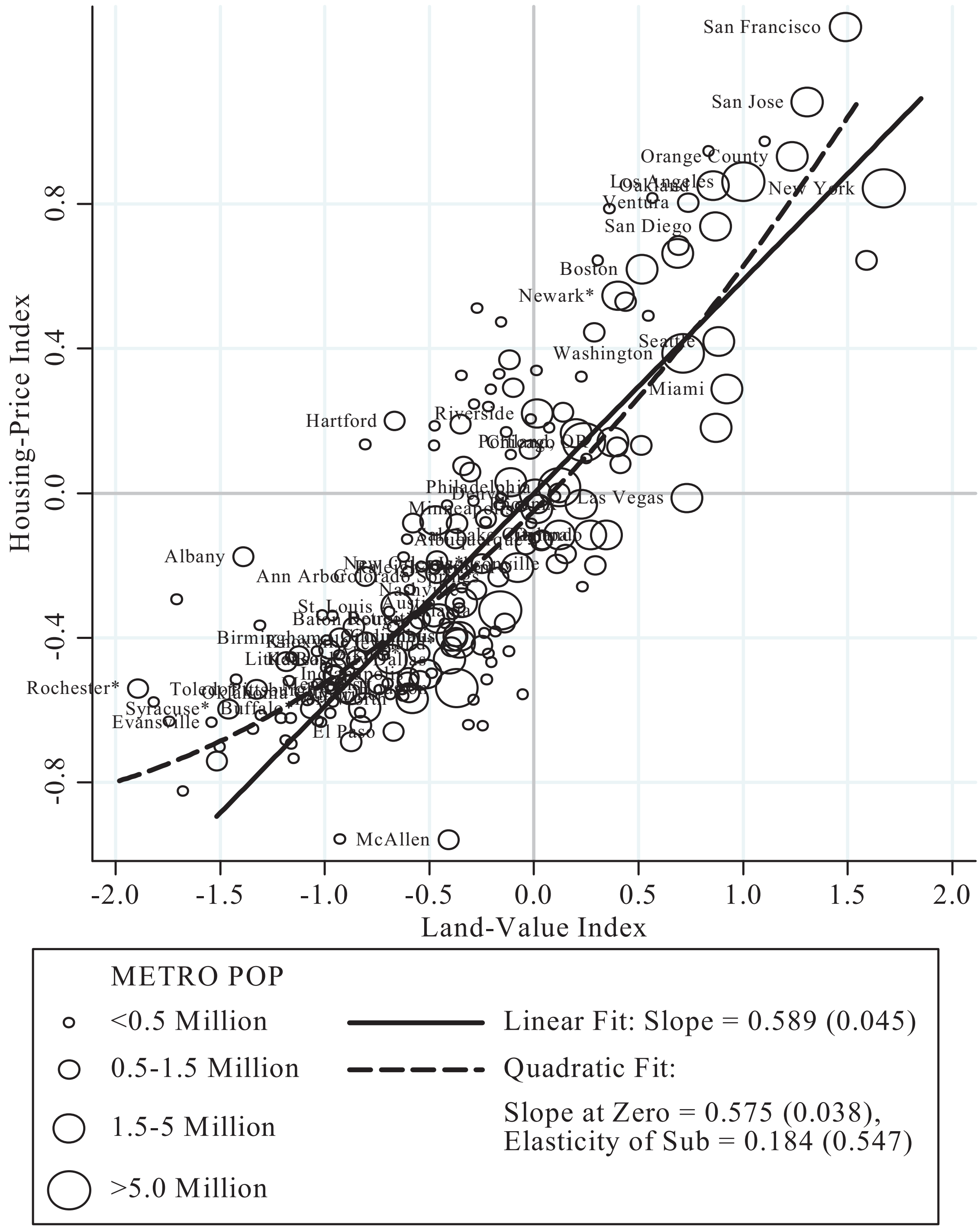


Figure 3.3: Construction Prices vs. Land Values

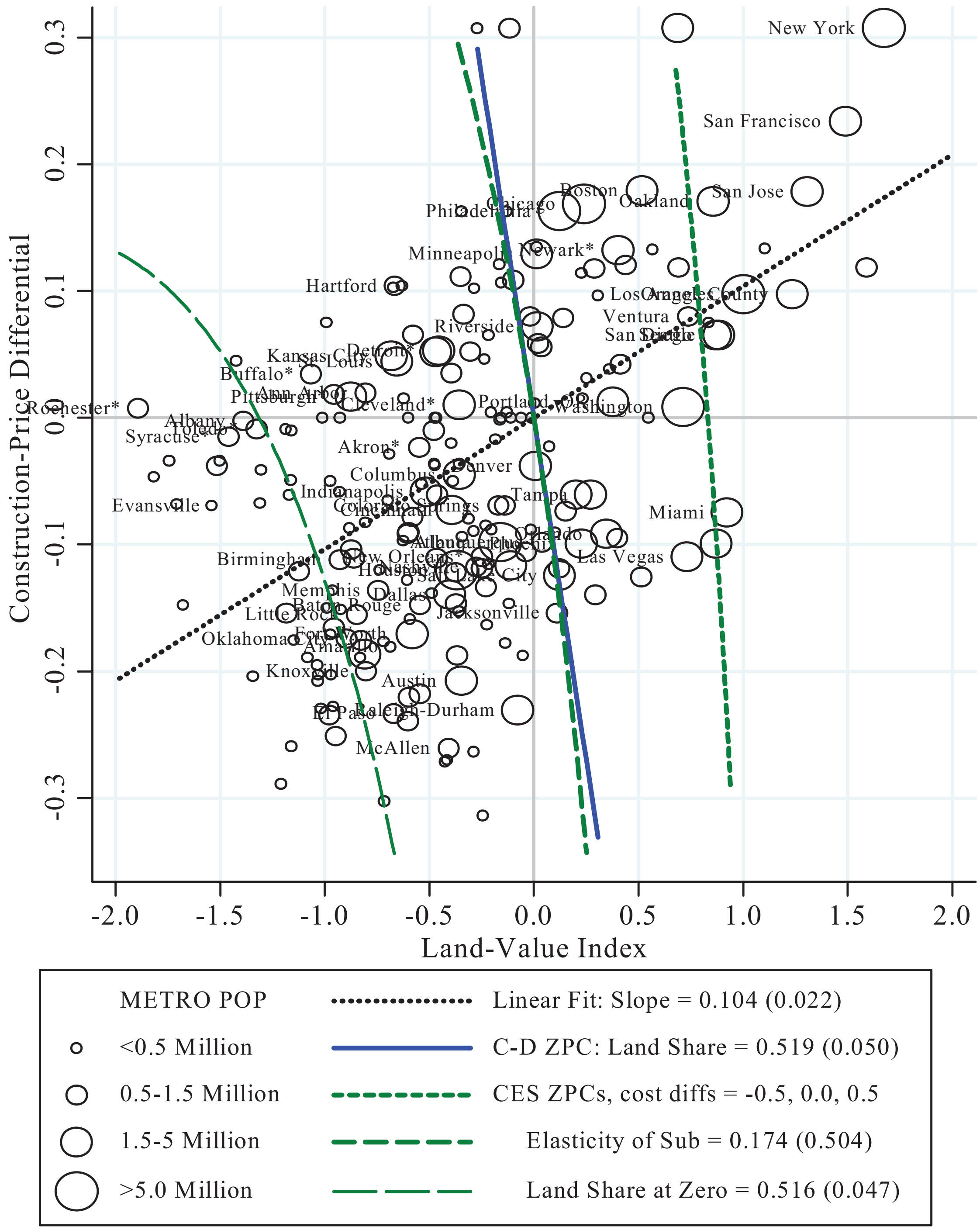


Figure 3.4: Three-Dimensional Cost Curve, corresponding to ZPC Curves in Figure 3.3, Estimated from Data, No Covariates

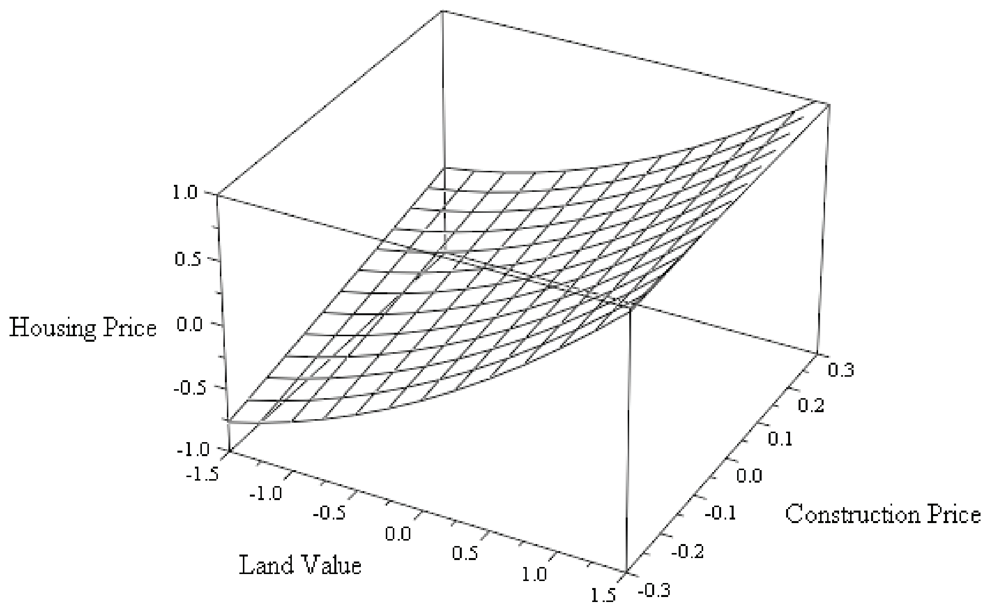


Figure 3.5: Productivity in the Tradeable and Housing Sectors

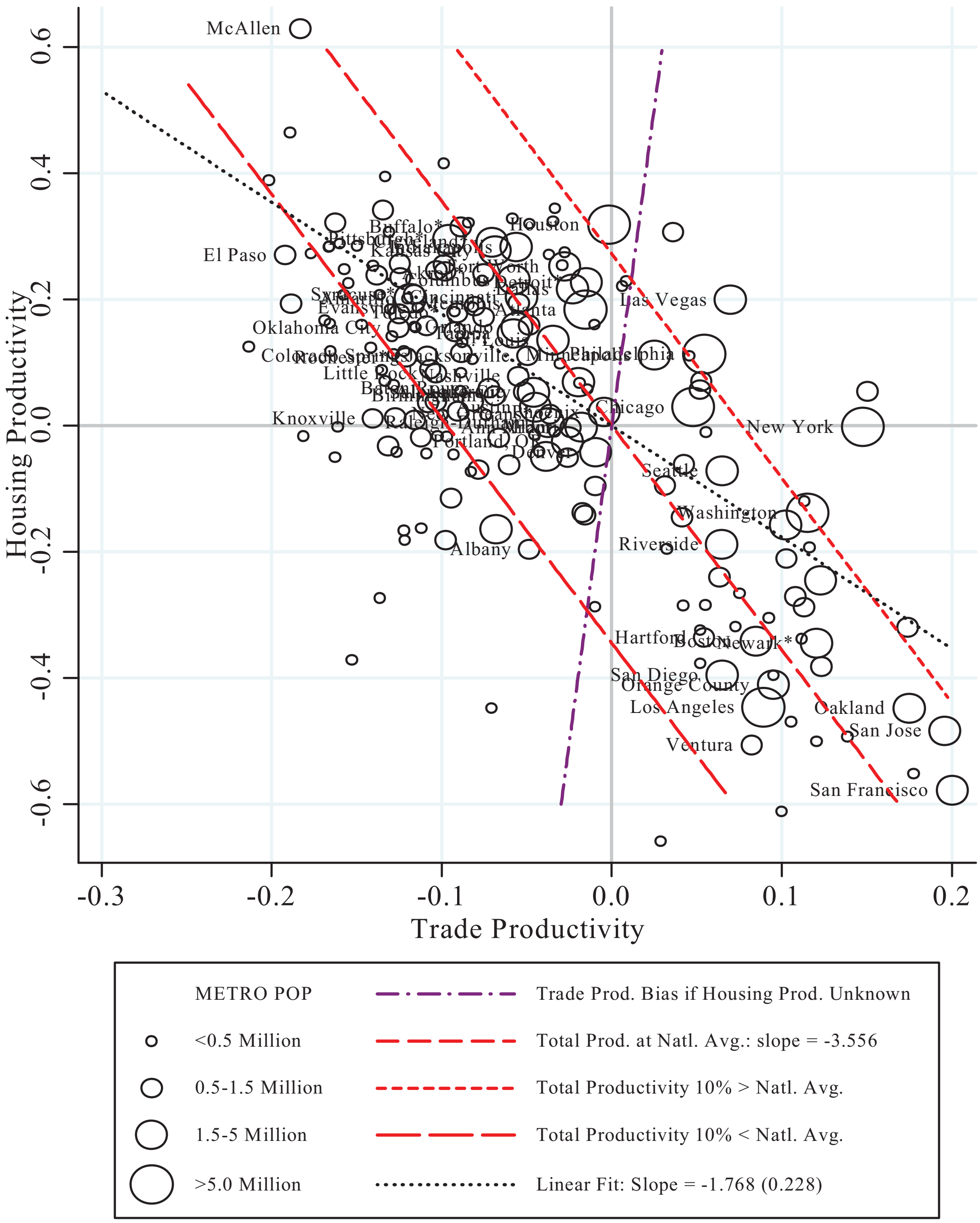


Figure 3.6: Quality of Life vs. Housing Productivity

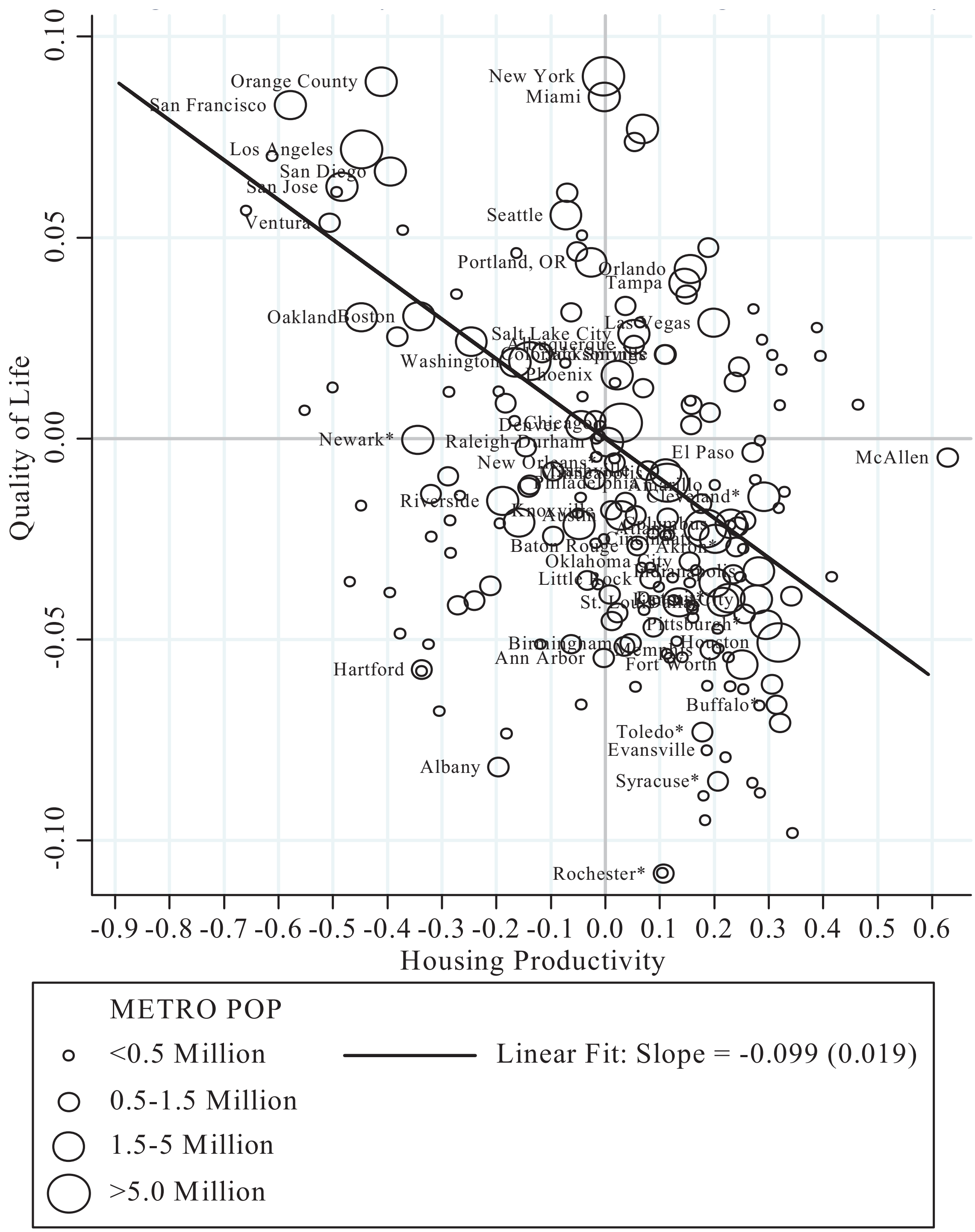


Figure 3.7: Shrunken vs. Unshrunken Land Values

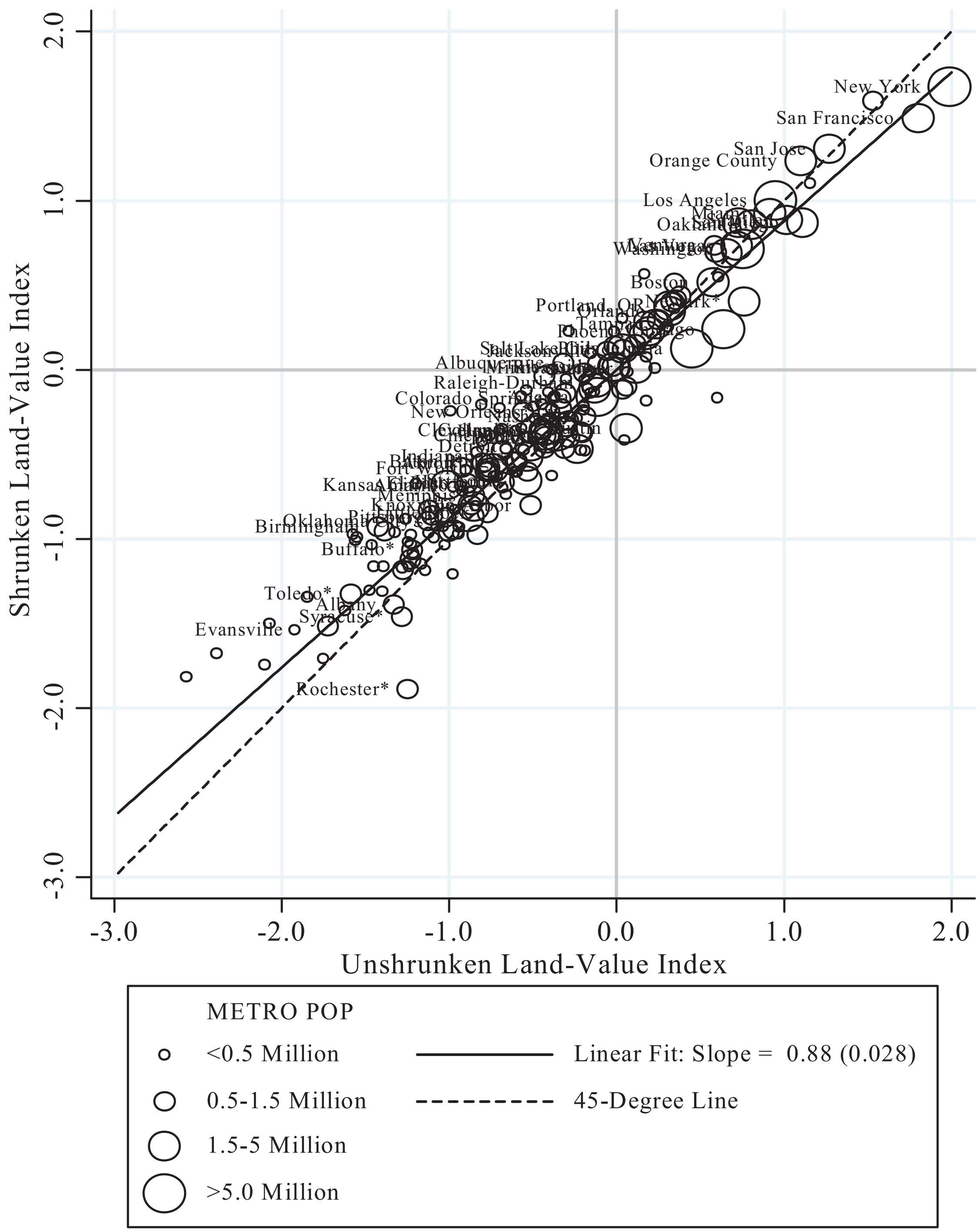


Figure 3.8: Weighted vs. Unweighted Land Values

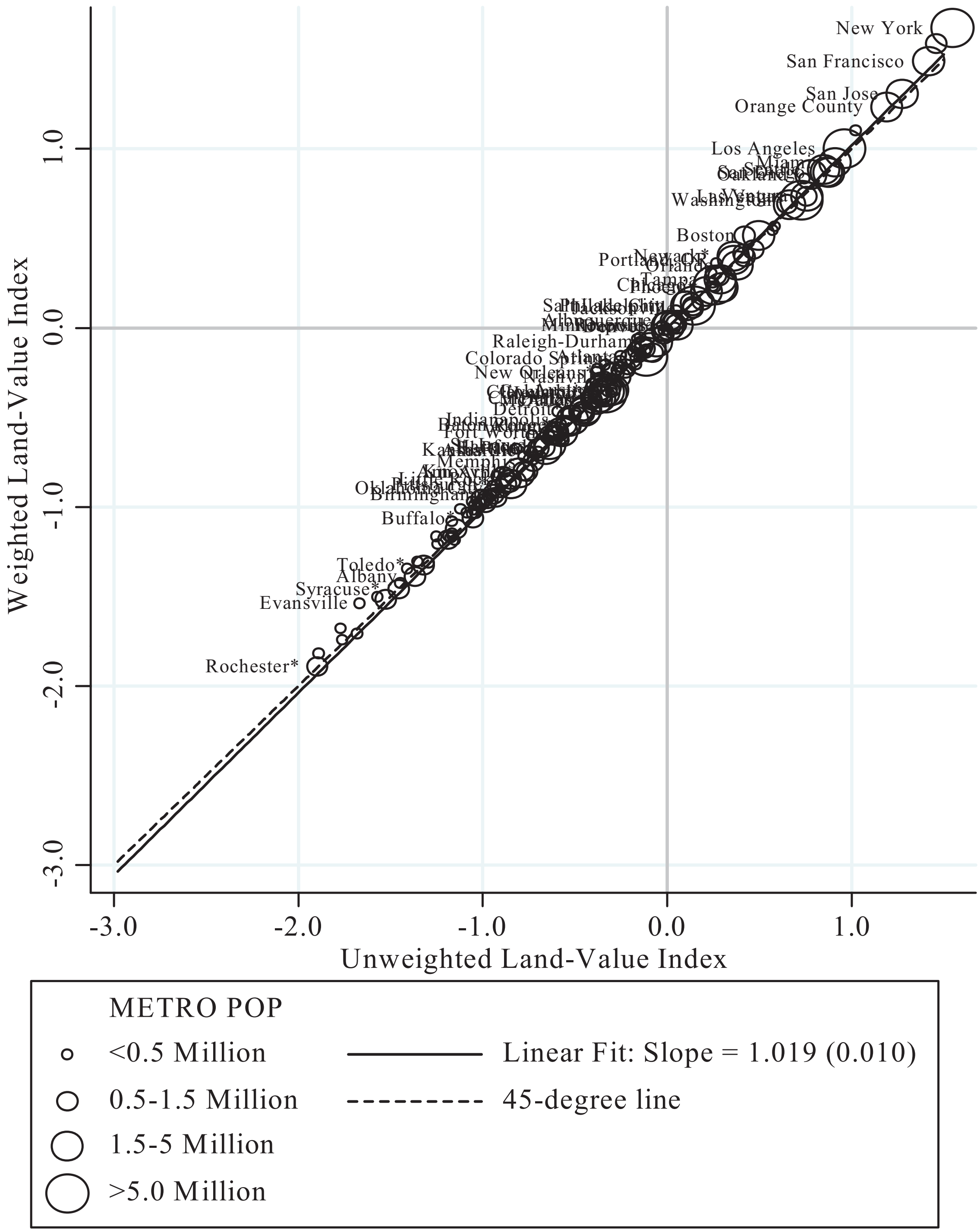


Figure 3.9: Residential vs. Overall Land Values

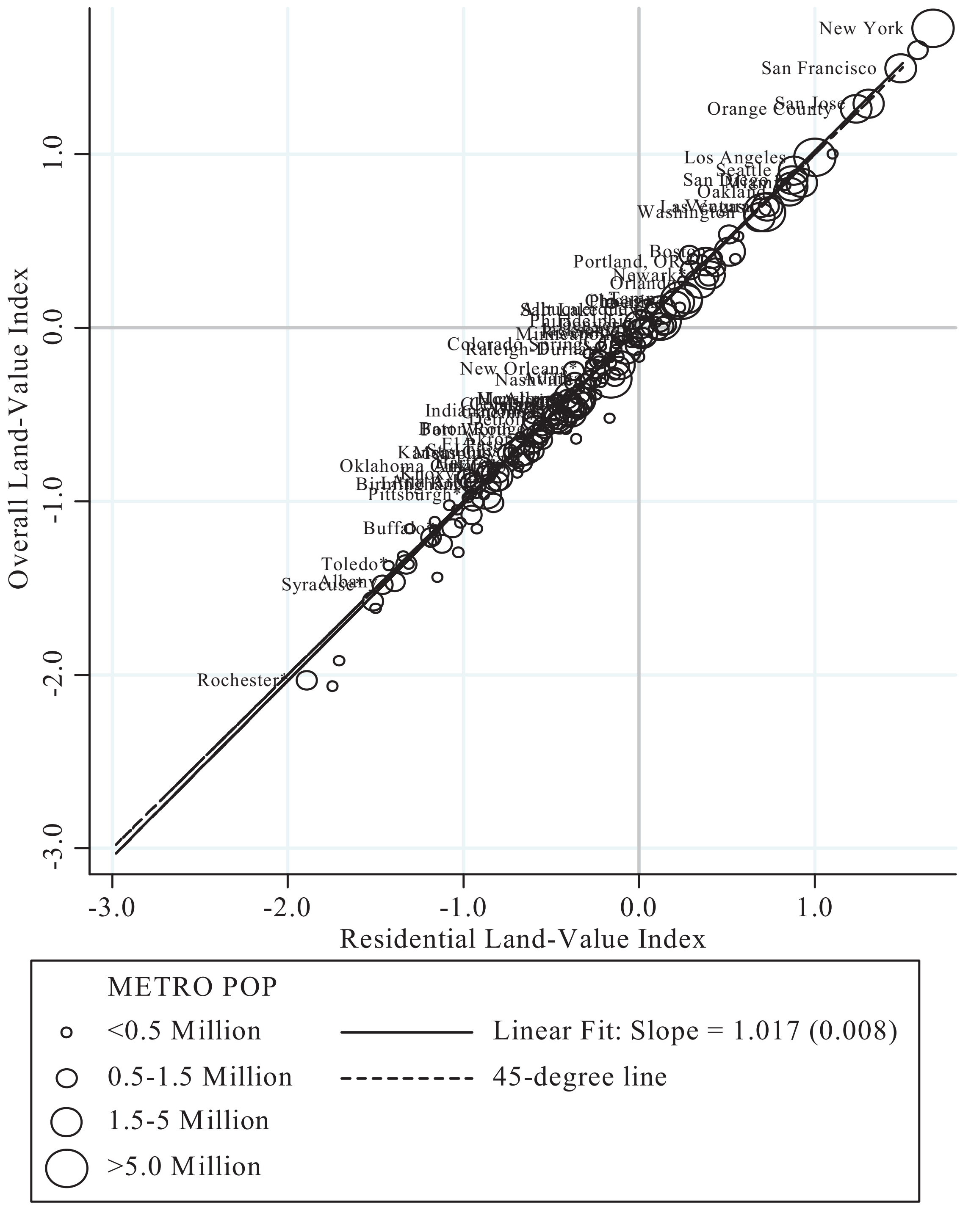


Figure 3.10: Construction Wages vs. Overall Wages

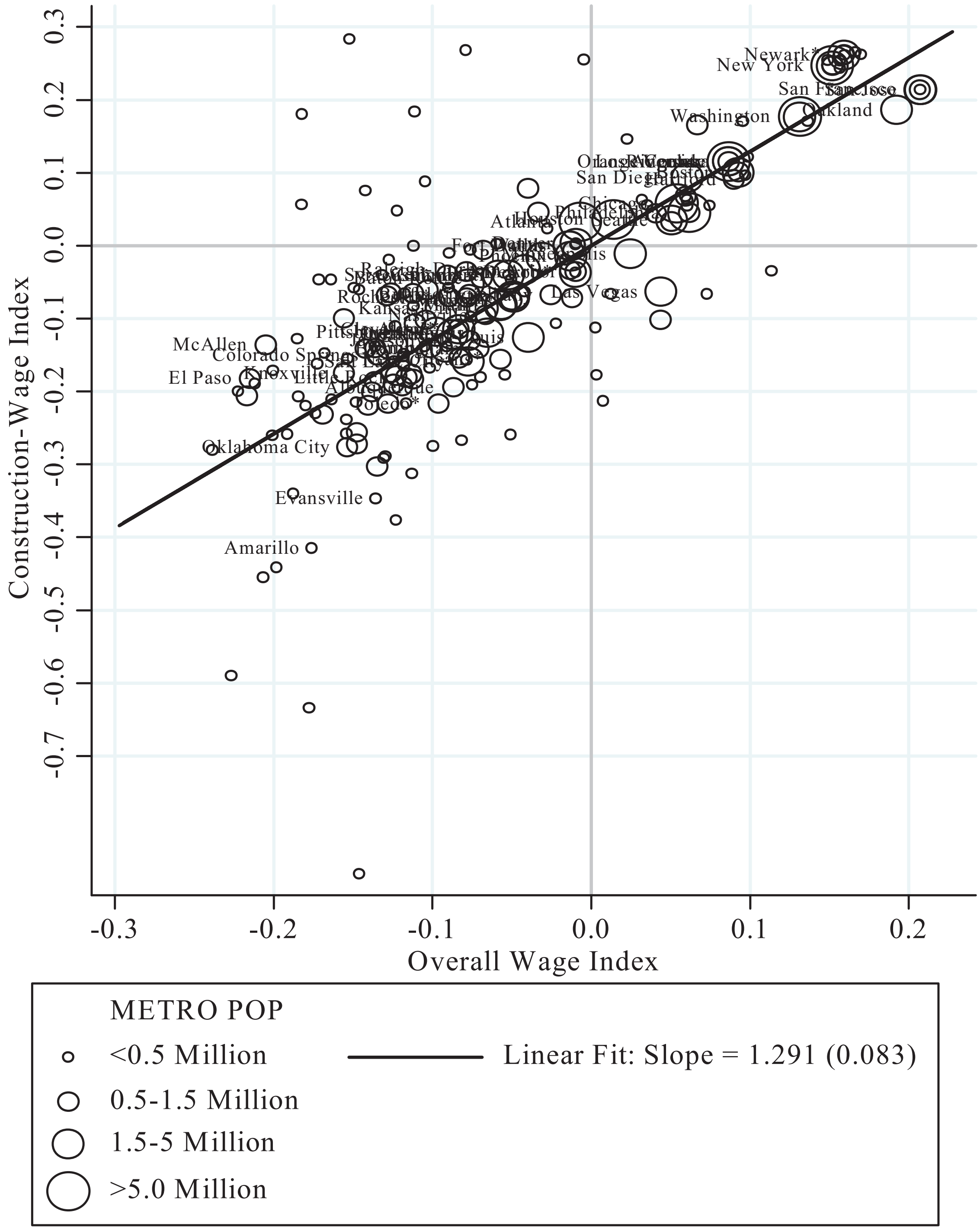
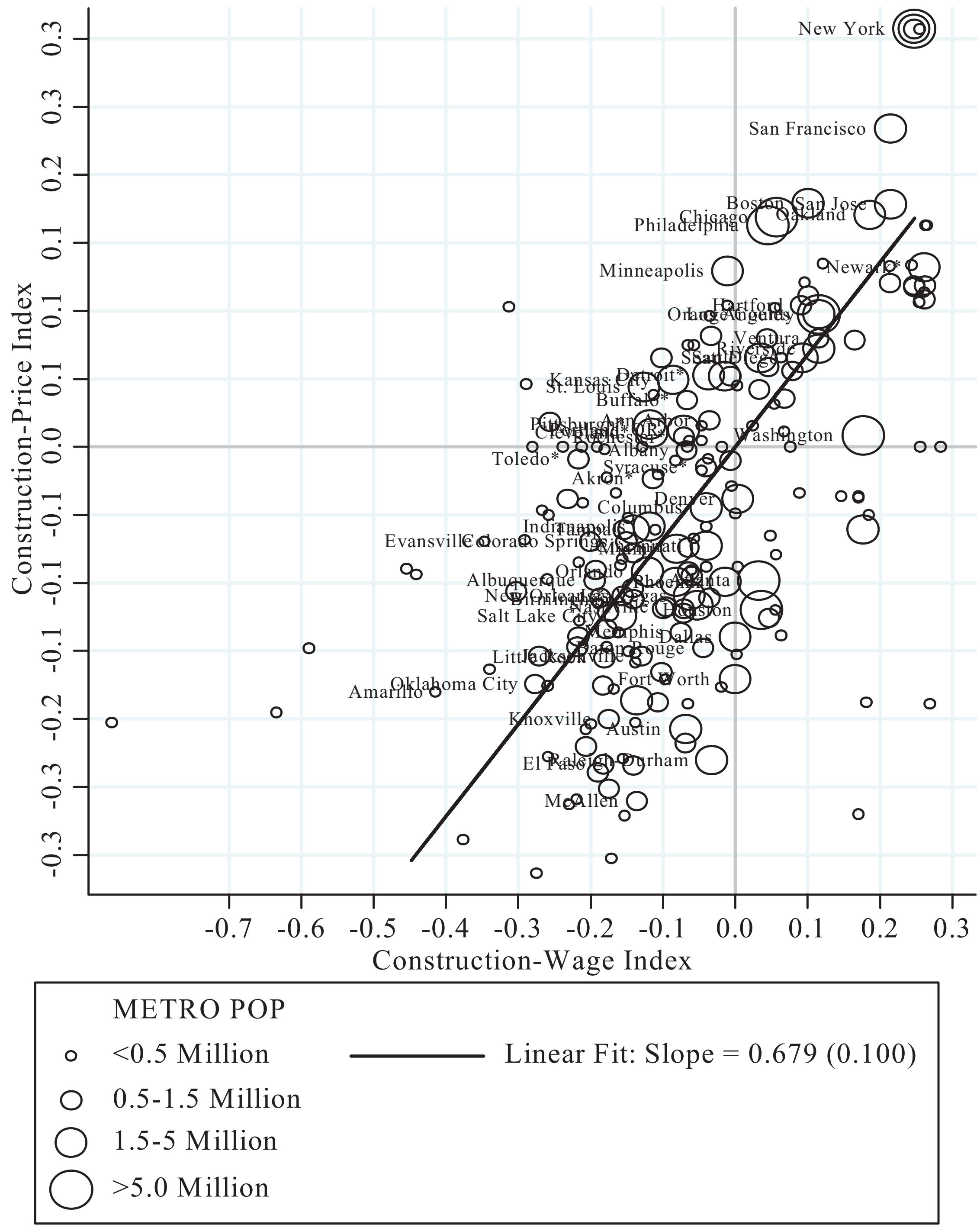


Figure 3.11: Construction Prices vs. Construction Wages



CHAPTER IV

Price and Time to Sale Dynamics in the Housing Market: the Role of Incomplete Information

4.1 Introduction

It is a stylized fact of the market for existing homes that there is a strong positive correlation between sales prices and sales volumes and there is a negative correlation between sales prices and the average time houses take to sell. Figure 1 illustrates this pattern using United States data from 1983q1 to 2010q4.[1] The top panel shows the log of the CoreLogic House Price Index for single family detached homes, the middle panel illustrates log single family home sales per household, and the bottom panel displays months' supply of single family homes for sale (hereafter months' supply).[2] Since 2000, the correlation coefficient between prices and sales volumes has been 0.73 and the correlation coefficient between prices and months' supply has been -0.34; both correlations are statistically significant at the 5-percent level.[3]

[1]Sales volume and industry data are from the National Association of Realtors and price data is from CoreLogic.

[2]Months' supply is the ratio of the number of homes listed as being for sale at the end of the month divided by the number of sales that month. It is used as a proxy for the average time on market because nationally representative data for time on market is unavailable. All series have had a linear time trend removed, and prices were deflated using the CPI.

[3]The correlation between prices and sales volumes remains strong over the entire sampling period, but the correlation between prices and months' supply becomes much weaker. However, studies at the city level, such as Genesove and Mayer (2001) and Miller and Sklarz (1986), find the same relationship between prices and time on market.

Although direct data on the time it takes houses to sell (hereafter time on market or time to sale) is not available at a national level, studies on the city level find a similar pattern when considering time on market directly. For instance, Genesove and Mayer (2001) document that in the Boston condominium market, fewer than 30 percent of listed units sold within 180 days during the trough year of 1992, while in 1997, after the market had recovered, more than 60 percent of new listings sold within 180 days. Miller and Sklarz (1986) document similar trends in Hawaii and Salt Lake City. Regarding the correlation between prices sales volumes, Stein (1995) shows that between 1986 and 1992, a 10 percent drop in prices reduces sales volumes by approximately 1.6 million units in a time span with average volume of three to four million units. Ortalo-Magne and Rady (1998) show the same qualitative pattern holds in the U.K. The robustness of this pattern leads Genesove and Mayer (1997) to describe it as "one of the most distinctive and puzzling macro features of the market for existing homes."

This paper offers a stylized model to explain this pattern in which home sellers have incomplete information about conditions in the housing market. This assumption seems natural in light of most households' infrequent participation in the housing market. Anily et al. (1999) report that the expected total time for a household to reside in an owner-occupied unit is 13 years. Case and Shiller (2003) note the relative absence of professional traders in the housing market: "Buyers and sellers in the housing market are overwhelmingly amateurs, who have little experience with trading. High transactions costs, moral hazard problems, and government subsidization of owner-occupied homes have kept professional speculators out of the market." Therefore many sellers may begin the sales process with limited or outdated information regarding conditions in the real estate market.

Another central assumption in the model is that sellers face idiosyncratic variation in the offers they receive. Although in the model I present the offer distribution is exogenous, this variation may come from differences in the match quality between the potential buyer and the house, variations in buyers' eagerness to transact quickly, or other factors. Although there is limited data concerning the distributions of offers sellers receive, Merlo and Ortalo-Magne (2004) present evidence both that offer amounts vary within a negotiation between a buyer and a seller, and that later bidders for a house tend to make higher bids than earlier bidders. The existence of idiosyncratic variation in offers seems well-accepted in the theoretical literature. For instance, Haurin discusses the optimal decision rule for sellers facing particular offer distributions.

Finally, the model considers the interaction between sellers with incomplete information and real estate agents with complete information. Recent research by Levitt and Syverson (2002,2008) has shown that because realtors' typical compensation structure leads to misalignment between realtors' and sellers' incentives, realtors may not be able to convey their knowledge of the real estate market to sellers efficiently. They show that empirically, realtors seem to encourage their clients to sell their homes more quickly than would be optimal for a fully informed seller. The model demonstrates that in such an environment, expected time to sale remains negatively correlated with house prices even in the presence of fully informed realtors.

I construct a measure of homeowner misperceptions of housing market conditions using data from the American Community Survey and use it to test one of the model's key predictions. I find that elevated misperceptions of house prices are associated with lower sales volumes at both the Metropolitan Statistical Area (MSA) and state levels. A simple fixed effects regression of sales volumes on the misperceptions index explains 21% of the variation in sales at the MSA level and 24% at the state level. Furthermore, I show that variation in the misperceptions index is related to volatility in prices and measures of heterogeneity in the housing stock.

A substantial literature attempts to explain the stylized fact. Magnus (2010) and

Ehrlich (2012) show that the strong correlations between prices, sales, and time on market are difficult to replicate in a general equilibrium search and matching model of the housing market. Stein (1995) models the role of downpayment requirements and credit constraints in generating these correlations, which Genesove and Mayer (1997) document empirically. Genesove and Mayer (2001) argue that prospect theory can explain the correlations if sellers use the price they paid for their house as a reference point to evaluate offers. They provide evidence that this reference point influences seller behavior. Albrecht et al. (2007) model a process in which sellers become more anxious to sell the longer the sales process takes, and therefore accept lower offers when a house has sat on the market for a long time. Finally, Lazear (2010) argues that consistent with the theory of monopoly pricing, sellers find it optimal to accept a lower probability of sale when demand, and therefore prices, fall.

The model in the present paper departs from the previous literature in assuming that sellers have imperfect information regarding the state of the housing market. It demonstrates that relaxing the assumption of perfect information can generate the observed patterns in the data, even when some agents in the model do have full information. I show that empirically, homeowners' misperceptions regarding housing values are correlated with changes in sales volumes, with an elasticity of sales with respect to misperceptions is about 1.7 in my baseline specification.

4.2 The Model

The model I consider is a partial equilibrium model in the sense that the distribution of offers that sellers receive is exogenous, and sellers receive exactly one offer per period. There is no negotiation in the model, leaving sellers with a choice only about whether to accept or reject a given offer. Once a seller has rejected an offer, they cannot recall it. I consider the situation in which sellers have no outside assistance in making their decisions, and the situation in which they are assisted by a real estate agent. I assume the agent has complete information regarding the state of the housing market but has incentives that are not entirely aligned with the seller's. In the body of the paper I consider a two-period model; in the appendix I consider extending the model to multiple periods. In all cases, I assume that in the last period, the seller must sell the house for the final offered price.

4.2.1 The Offer Distribution

Offers are the sum of an 'aggregate demand' component, z, which is constant across periods, and an idiosyncratic component, x_t, which is distributed i.i.d. across periods. z is distributed $U[z_L, z_H]$, and x_t is distributed $U[x_L, x_H]$. z and x_t are independently distributed. Denoting the period t offer as ψ_t,

$$\psi_t = z + x_t \tag{4.1}$$

ψ_t is distributed according to a 'modified triangular distribution'.[4]

The assumption that the offers a seller receive are a function both of aggregate market conditions and the buyer's idiosyncratic taste for the property is essential to the results of the model, so it is worth discussing further. As mentioned previously, there is limited empirical evidence on the distribution of offers that home sellers receive, although Merlo and Ortalo-Magne (2004), using data from England, document that there is economically meaningful variation in the bids a seller receives on a given property. One issue of particular relevance is to what extent buyers are aware of aggregate market conditions, such that their offers reflect these conditions. In the context of the housing market, it seems likely that buyers are more aware of overall market conditions than are sellers. As Genesove and Han (2009) argue, the institutional framework of the housing market is such that homes for sale are prominently advertised and it is straightforward for buyers to locate homes to examine. In

[4]The p.d.f. is defined in the appendix.

contrast, buyers do not advertise themselves and are not "searchable" to sellers, implying that it is easier for buyers than for sellers to acquire information on the state of the market. For instance, Baryla et al. (2000), document that broker-assisted buyers reported visiting an average of 3.81 homes per week. Genesove and Han show that the empirical data concerning buyer search behavior and seller time on market are consistent with sellers adjusting to demand shocks in the housing market more slowly than buyers do. These results reassure me that the key features of the offer distribution I assume are broadly consistent with the realities of the housing market.

4.2.2 The Model with no Realtor

First I consider a two-period model with no realtor. As outlined above, sellers receive one offer each period, and if they do not accept the first offer they must accept the second. I assume sellers are risk-neutral, perfectly patient, and do not bear any flow cost of leaving their homes on the market. Therefore, a seller's goal is simply to obtain the highest possible price for their home. Accordingly, a seller will accept the period 1 offer, ψ_1, if and only if it is greater than or equal to the expectation of the period 2 offer, $E[\psi_2]$. Sellers cannot observe the state of market demand z or the idiosyncratic component of the offer x_t directly, so they must infer z using Bayes' Theorem:

$$f_Z(z|\Psi = \psi) = \frac{f_{Z,\Psi}(z, \psi)}{f_\Psi(\psi)} = \frac{f_\Psi(\psi|z)f_Z(z)}{f_\Psi(\psi)}$$

Define $\tilde{z}_{L,1} = \max(z_L, \psi_1 - x_H)$ and $\tilde{z}_{H,1} = \min(z_H, \psi_1 - x_L)$. Then the seller's posterior belief about the distribution of z conditional on ψ_1 is $z \sim U[\tilde{z}_{L,1}, \tilde{z}_{H,1}]$.[5] Knowing the seller's belief about z as a function of ψ_1 allows us to calculate the seller's expectation of ψ_2. The seller's conditional expectation function, $E[\psi_2|\psi_1]$, is plotted in Figure 4.2 for two cases, the first in which $x_H - x_L \geq z_H - z_L$, and the

[5] I demonstrate this result along with proofs of the propositions in the appendix.

second in which $z_H - z_L > x_H - x_L$, and is defined explicitly in the appendix. As the figure indicates, the seller's expectation of ψ_2 is a weakly increasing function of psi_1. In the first case, $E[\psi_2|\psi_1] \geq \psi_1$ when $\psi_1 \leq \overline{\psi}_1$; in the second case , $E[\psi_2|\psi_1] \geq \psi_1$ when $\psi_1 \leq x_H + z_L$.

This allows us to state the main result of the model with no realtor.

Proposition 1: When the seller follows the optimal policy, the expected time to sale in the model with no realtor is weakly decreasing in the state of aggregate demand, z, while the expected sales price is strictly increasing in z.

Figure 4.3 illustrates the expected time to sale and expected sales price in the model with no realtor for the two cases. If the variance of the idiosyncratic component of the offer is greater than the variance of aggregate demand, then the expected time to sale will be falling for all values of z. If the variance of aggregate demand is greater, however, the expected time to sale will be constant for very low and very high values of aggregate demand, while for intermediate values expected time to sale will be falling in z. Therefore, the model implies that expected time to sale will be negatively correlated with expected prices.

Although this model is quite simple, it illustrates the essential mechanism by which incomplete information generates a negative correlation between the strength of housing demand and the expected time to sale. When sellers are uncertain about the state of aggregate demand, they follow a reservation price strategy in period 1, whereas if they could observe z, they would ignore the aggregate demand component of the offer, which is stable over time, and make their decision based solely on the idiosyncratic component, x_1. With incomplete information, offers with middling values of x_1 will be above the seller's reservation price when z is high and below the reservation price when z is low. Therefore, the probability that the seller accepts the first period offer is an increasing function of z.

4.2.3 A Two-Period Model with a Realtor

I now introduce a realtor to the two-period model. I assume the realtor can observe aggregate housing demand z directly. However, in the model the realtor's and the seller's incentives are only partially aligned, so the realtor will not always be able to signal his knowledge of z to the seller in a credible way. In the model, the only possible contract between sellers and realtors is one in which the realtor receives a fixed fraction α of the sales price. Furthermore, I assume the seller must employ a realtor when selling the house. If there is no sale in period 1, the realtor must pay flow cost c_R at the beginning of period 2 in order to market the house.[6]

The realtor can only communicate with the seller by advising him on whether to accept an offer after it has been received. Although this assumption seems restrictive, as Levitt and Syverson (2002) argue, if the realtor is constrained to advising the seller on each offer only after it has been received, there is no way for the realtor to report the intensity of his preferences credibly. Therefore it is not overly restrictive to limit the realtor to recommending either 'accept' or 'reject' after each offer is received.

To fix terminology,denote the realtor's recommendation about the time t offer as ξ_t, with $\xi_t = 0$ if the realtor recommends 'reject' and $\xi_t = 1$ if the realtor recommends 'accept'. Let $\tilde{f}(\psi_2|\psi_1, \xi_1)$ be the seller's posterior belief about the distribution of ψ_2 conditional on the first period offer ψ_1 and the realtor's recommendation ξ_1. Finally, call the seller's period 1 policy function $\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, \xi_1))$, with $\gamma_1 = 0$ indicating that the seller rejects the period 1 offer and $\gamma_1 = 1$ indicating that the seller accepts the period 1 offer. Then the seller's period 1 and period 2 value functions can be

[6] Implicitly, one could imagine that c_R is a flow cost the realtor must pay at the beginning of each period to market the house, but the cost paid at the beginning of period 1 is sunk and does not affect the realtor's maximization problem.

written as:

$$V_S(\psi_1, \xi_1, 1) = \overset{\max}{\gamma_1}(1-\alpha)\{\psi_1, E[\psi_2|\tilde{f}(\psi_2|\psi_1, \xi_1)]\}$$
$$V_S(\psi_2, \xi_2, 2) = (1-\alpha)\psi_2$$

The realtor's value functions can be written:

$$\begin{aligned} V_R(\psi_1, z, 1) = \overset{\max}{\xi_1} &\{\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, 0)\alpha\psi_1 + (1-\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, 0)))(E[V_R(\psi_2, z, 2)]), \\ &\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, 1)\alpha\psi_1 + (1-\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, 1)))(E[V_R(\psi_2, z, 2)])\} \\ V_R(\psi_2, z, 2) = -c_R &+ \alpha\psi_2 \end{aligned}$$

It is then natural to define a Bayesian Nash equilibrium of the game between realtors and sellers as a policy function $\xi_1(\psi_1, z)$ for the realtor, a policy function $\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, \xi_1))$ for the seller, and a belief updating strategy $\tilde{f}(\psi_2|\psi_1, \xi_1)$ for the seller such that:

1. $\xi_1(\psi_1, z)$ maximizes the realtor's value function for all (ψ_1, z), taking $\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, \xi_1))$ as given;

2. $\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, \xi_1))$ maximizes the seller's value function for all ψ_1, ξ_1 taking $\xi_1(\psi_1, z)$ as given; and

3. $\tilde{f}(\psi_2|\psi_1, \xi_1)$ is consistent with $\xi_1(\psi_1, z)$.

Consider what the realtor would choose if he could decide which offers the seller would accept and which he would reject. After some algebra, the realtor's value function in period 1 can be re-written

$$V_R(\psi_1, z, 1) = \alpha z + \max\{\alpha x_1, \alpha\overline{x} - c_r\}$$

Therefore, in period 1 the realtor would prefer that the seller accept any offer ψ_1 such that $\alpha x_1 \geq \alpha \overline{x} - c_r$, or equivalently, $x_1 \geq \overline{x} - \frac{c_r}{\alpha}$. Denote $\overline{x} - \frac{c_r}{\alpha}$ as $\hat{x}_1$, which represents the realtor's cutoff value of x_1, i.e. the realtor would like the seller to accept all offers with an idiosyncratic component x_1 above $\hat{x}_1$ and reject all others.[7] Note that $\frac{c_r}{\alpha}$ measures the degree of misalignment between the seller's and the realtor's incentives. If $\frac{c_r}{\alpha}$ were zero, the realtor's and seller's incentives would be perfectly aligned.

The following is a Bayesian Nash equilibrium of the game between the realtor and the seller. The realtor reports 'accept' ($\xi_1 = 1$) for any offer such that $x_1 \geq \hat{x}_1$ and 'reject' ($\xi_1 = 0$) for any offer such that $x_1 < \hat{x}_1$. Define $\tilde{x}_{L,1}$ as x_L if the realtor recommends reject and $\hat{x}_1$ if the realtor recommends accept, and define $\tilde{x}_{H,1}$ as $\hat{x}_1$ if the realtor recommends reject and x_H if the realtor recommends accept. Further, define $\tilde{z}_{L,1} = \max(z_L, \psi_1 - \tilde{x}_{H,1})$ and $\tilde{z}_{H,1} = \min(z_H, \psi_1 - \tilde{x}_{L,1})$. Then the seller's posterior belief is that $z \sim U[\tilde{z}_{L,1}, \tilde{z}_{H,1}]$.[8]

Figure 4.4 displays the seller's expectation of ψ_2 as a function of psi_1 and the realtor's recommendation ξ_1. The red lines show the expectation when the realtor recommends reject and the green lines show the expectation when the realtor recommends accept; the dashed portions of the red and green lines show off-equilibrium path recommendations. As before, I show two cases, differentiated by whether z or x has the higher variance conditional on the realtor's recommendation. The figure illustrates that the seller's expectation of ψ_2 is always greater than ψ_1 when the realtor recommends 'reject'. Therefore, the seller will always reject the first offer when the realtor recommends it. However, the opposite is not true: the seller will sometimes reject an offer when the realtor recommends 'accept'. These cases arise when the seller's expectation of ψ_2 is greater than psi_1 despite the realtor's recommendation to accept the offer. They are illustrated in the figure as values of ψ_1 for which the green line is above the dashed blue line.

[7] I assume that α and c_R are such that $\hat{x}_1 \geq x_L$.

[8] I do not currently offer a proof but it would follow the proof of the seller's belief updating strategy with no realtor very closely.

To verify that reporting his own preference truthfully is a best response for the realtor, note that in equilibrium, the realtor's recommendation weakly increases the chance that the seller will take the realtor's preferred action. For very high and very low values of ψ_1, the realtor's recommendation will not affect the seller's decision, so any policy the realtor follows is a best response. However, for medium values of ψ_1, the expected value of ψ_2 when the realtor recommends 'accept' is below ψ_1, but the expected value of ψ_2 when the realtor recommends 'reject' is above ψ_1. For these offers, the realtor's recommendation will be decisive - the seller will follow the realtor's recommendation. Therefore, for this range of offers it must also be a best response for the realtor to report his own preference truthfully. This allows us to state Proposition 2, which shows that the main result from the model with no realtor continues to hold when the realtor is added to the model.

Proposition 2: In the equilibrium I consider, the expected time to sale in the model with a realtor is weakly decreasing in the state of aggregate demand, z, while the expected sales price is strictly increasing in z.

Figure 4.5 illustrates the expected time to sale and sales price as a function z. The expected sales price is a strictly increasing function of z, but the expected time on market is weakly decreasing with z. Specifically, when the variance of z is large, there will be a range in which the expected time to sale does not depend on z.

4.3 Empirics

A key prediction of the model presented above is that the correlations between prices, time on market, and sales volumes are driven by seller confusion over the true state of the housing market. To test this prediction, I construct a measure of homeowner perceptions of housing values using data from the American Community

Survey. I then compare this measure to an index of home values provided by Zillow.com to construct a measure of homeowner misperceptions of housing values both at the MSA and state levels. Finally, I regress sales volumes on the misperceptions index to test whether homeowner misperceptions of the state of the housing market influence sales levels. I conclude by exploring the determinants of homeowner misperceptions about the housing market.

4.3.1 Constructing the Misperceptions Indices

I begin by constructing time series house value indices at the state and MSA levels. I use data from the American Community Survey 1% national samples, in which homeowners were asked to self-report the dollar value of their homes, as well as to report a rich set of physical characteristics of their homes. The data is available yearly from 2000 to 2010 at the state level, but MSA level identifiers are only available from 2005-2010. At the state level, there are approximately 6.5 million home value observations in the data set, just over 100,000 in year 2000, between 300,000 and 400,000 per year between 2001 and 2004, and about 840,000 per year thereafter. Approximately 680,000 of these observations per year are located in MSAs from 2005 to 2010.

To construct my housing value indices, within each geographical area I regress log self-reported house values on the following set of covariates[9]: survey year dummies, 9 indicators of building size, 9 indicators for the number of rooms, 5 indicators for the number of bedrooms, the number of rooms interacted with the number of bedrooms, 2 indicators for lot size, 13 indicators for when the building was built, 2 indicators for complete plumbing and kitchen facilities, an indicator for commercial use, and an indicator for condominium status. I take the coefficients on the year dummies to be

[9]To maintain comparability with the Zillow Home Value Index, which measures median home values, I use quantile regression. To aid convergence in the state level regressions, I drop the most and least expensive 5% of houses each year.

my perceived housing value time series indices.

To construct my misperceptions indices, I compare the perceived housing value indices to the Zillow Home Value Indices (hereafter, ZHVI) published by Zillow.com. These indices measure the value of the median non-distressed owner-occupied home in a given geographical area over time. The ZHVI lines up well with other indices of housing values such as the Case-Shiller index and the Federal Housing Finance Authority index. I restrict my data set to MSAs and states with complete ZHVI and sales volume data; because some states with very low sales volumes show an implausible amount of volatility in the ZHVI, at the state level I also require that Zillow reports average sales volumes of at least 2,000 houses a month. This leaves me with 23 states and 113 MSAs.

Figure 9 displays the perceived and Zillow home value indices for the 23 states in my data set, with values in 2000 normalized to zero. Figure 11 displays the same series at the MSA level for 19 of the 20 MSAs in the Case-Shiller 20-city house price index, with values in 2005 normalized to zero. I construct the misperceptions index as the difference between these series, i.e. the misperceptions index is the vertical distance between the lines in Figures 9 and 11. Figures 10 and 12 display the misperceptions indices along with the log change in home sales volumes from the base year at the state and MSA levels, respectively.[10]

4.3.2 Misperceptions and Home Sales

Table 1 shows fixed effects regressions at the state level of log home sale volumes on the misperceptions index. The time period is 2000 to 2010. Columns 1 through 4 are specified in levels, while columns 5 through 8 are specified in first differences; columns 2, 4, 6, and 8 include the ZHVI, while columns 3, 4, 7, and 8 include year fixed effects. Consistent with the theory, the coefficient on the misperceptions index is

[10] I show the change in sales volumes rather than the level in order to better illustrate the variation, but I use levels in the regression analysis.

negative in all specifications, significantly so in all columns except for column 7. The coefficient on the misperceptions index becomes somewhat weaker when year fixed effects are included or the model is specified in first differences. An attractive feature of the state-level data is that it includes a period of broadly rising prices from 2000 to 2006, as well as a period of broadly falling prices from 2006 to 2010, as opposed to the MSA-level data, which cover a period of mostly falling prices. The coefficient on the ZHVI is not significant in the state level specifications without the year fixed effects, but is negative and statistically significant in the specifications with the year fixed effects.

Table 2 shows similar regressions at the MSA level. The time period included is 2005 to 2010. As predicted by the model, the coefficient on the misperceptions index is negative in all specifications, although it is statistically insignificant in columns 3 and 7. As in table 1, including year fixed effects and specifying the model in first differences weakens the effect of misperceptions. Interestingly, the effect of the ZHVI is significantly negative in all specifications, in contrast to the pattern we observe when sales are not conditioned on homeowner perceptions of the housing market.

Taking column 2 of table 1 as my baseline specification, these results suggest an elasticity of sales volumes with respect to misperceptions of approximately 1.7. The within-R^2 of the simple fixed effects regression of sales volumes on the misperceptions index is 0.24 at the state level and 0.21 at the MSA level, suggesting that homeowner misperceptions alone can explain over a fifth of the time variation in sales volumes.

4.3.3 Determinants of Misperceptions

In table 3, I briefly explore the empirical determinants of homeowner misperceptions of housing market conditions. The dependent variable in both specifications is the standard deviation of the misperceptions index, which I take to be an indicator of how much perceptions tend to deviate from actual market prices. Column 1 explores misperceptions at the MSA level, while column 2 examines the state level. At both levels, variation in market prices is the strongest predictor of misperceptions, accounting for roughly half of the variation in misperceptions in a simple regression. As additional covariates, I examine two different measures of heterogeneity in the housing stock. At the MSA level, I consider the ratio of the ZHVI of the most expensive third of houses to the ZHVI of the least expensive third. At the state level, I consider the median error of the ZHVI, that is the median difference between Zillow's predicted house value and a home's actual sales price.[11]

Both measures of housing stock heterogeneity are predictive of a higher level of misperceptions. At the MSA level, a one standard deviation increase in the ratio of top to bottom ZHVIs is associated with an approximately one-quarter standard deviation increase in the standard deviation of misperceptions. At the state level, a one standard deviation increase in the median error of the ZHVI is associated with a nearly one-half standard deviation increase in the standard deviation of misperceptions. These results suggest that increasing heterogeneity of the housing stock across metro areas is associated with greater homeowner misperceptions of the state of the housing market.

4.4 Conclusion

This paper demonstrates that astylized model of the house-selling process in which sellers have incomplete information regarding the state of the housing market can generate the negative correlation between prices and time on market observed in the data. Importantly, this effect can persist if sellers employ realtors with complete information about the state of housing demand as long as the incentives of realtors and sellers are not perfectly aligned. Empirically, an increase in homeowners' percep-

[11] I consider an increase in both of these variables to represent increasing heterogeneity in an area's housing stock. I do not consider the ratio of high to low ZHVIs at the state level because this ratio at the state level is likely to reflect variation in urbanization rates as much as heterogeniety in the housing stock. I do not consider the median error of the ZHVI at the MSA level because Zillow only publishes it for a very small number of large MSAs.

tions of house prices relative to true market conditions predicts a decrease in sales volumes. Homeowner misperceptions of housing market conditions appear to account for between one-fifth and one-quarter of the variation in sales volumes.

4.5 Appendix

4.5.1 Proofs of Propositions

4.5.1.1 The p.d.f. of ψ

Let $\psi_L = x_L + z_L$ and $\psi_H = x_H + z_H$. Then the p.d.f. of ψ is:

$$f(\psi) = \begin{cases} \frac{\psi - \psi_L}{(x_H - x_L)(z_H - z_L)} & \text{if } \psi_L \le \psi \le \psi_L + \min(x_H - x_L, z_H - z_L) \\ \frac{1}{\max(x_H - x_L, z_H - z_L)} & \begin{array}{l}\text{if } \psi_L + \min(x_H - x_L, z_H - z_L) \le \psi \\ \le \psi_L + \max(x_H - x_L, z_H - z_L)\end{array} \\ \frac{\psi_H - \psi}{(x_H - x_L)(z_H - z_L)} & \text{if } \psi_L + \max(x_H - x_L, z_H - z_L) \le \psi \le \psi_H \end{cases}$$

4.5.1.2 Seller's posterior belief about z

First note

$$f_\Psi(\psi|z) = \begin{cases} \frac{1}{(x_H - x_L)} & \text{if } z + x_L \le \psi \le z + x_H \\ 0 & \text{otherwise} \end{cases}$$

There are multiple cases to consider to verify that the seller's posterior distribution for z is $z \sim U[\tilde{z}_L, \tilde{z}_H]$.

Case 1: $x_H - x_L \ge z_H - z_L$

Case 1a: $\psi \le x_L + z_H$. In this range, $f(\psi) = \frac{\psi - \psi_L}{(x_H - x_L)(z_H - z_L)}$. If $z < z_L$ or $z > z_H$, $f(z) = 0$. If $z < \psi - x_H$ or $z > \psi - x_L$, $f(\psi|z) = 0$. Therefore, $f(z|\psi) = 0$ if $z < \max(z_L, \psi - x_H) \equiv \tilde{z}_L$ or if $z > \min(z_H, \psi - x_L) \equiv \tilde{z}_H$. In the range $[\tilde{z}_L, \tilde{z}_H]$,

$$f_Z(z|\Psi = \psi) = \frac{f_\Psi(\psi|z) f_Z(z)}{f_\Psi(\psi)} = \frac{\frac{1}{x_H - x_L} \cdot \frac{1}{z_H - z_L}}{\frac{\psi - \psi_L}{(x_H - x_L)(z_H - z_L)}} = \frac{1}{\psi - \psi_L}$$

To see that $f_Z(z|\Psi = \psi)$ is a proper density, note that in this case $\tilde{z}_L = z_L$ and $\tilde{z}_H = \psi - x_L$, so that $\tilde{z}_H - \tilde{z}_L = \psi - x_L - z_L = \psi - \psi_L$.

Case 1b: $x_L + z_H \leq \psi \leq x_H + z_L$. In this case, $f(\psi) = \frac{1}{x_H - x_L}$, $f(z) = \frac{1}{z_H - z_L}$, $f(\psi|z) = \frac{1}{x_H - x_L}$, $\tilde{z}_L = z_L$, and $\tilde{z}_H = z_H$. Then in the range $[\tilde{z}_L, \tilde{z}_H]$,

$$f_Z(z|\Psi = \psi) = \frac{f_\Psi(\psi|z) f_Z(z)}{f_\Psi(\psi)} = \frac{\frac{1}{x_H - x_L} \cdot \frac{1}{z_H - z_L}}{\frac{1}{(x_H - x_L)}} = \frac{1}{z_H - z_L}$$

and $f_Z(z|\Psi = \psi) = 0$ elsewhere.

Case 1c: $x_H + z_L < \psi$. In this case $f(\psi) = \frac{\psi_H - \psi}{(x_H - x_L)(z_H - z_L)}$. $f(\psi|z) = \frac{1}{x_H - x_L}$, $f(z) = \frac{1}{z_H - z_L}$, $\tilde{z}_L = \psi - x_H$, and $\tilde{z}_H = z_H$. Then in the range $[\tilde{z}_L, \tilde{z}_H]$,

$$f_Z(z|\Psi = \psi) = \frac{f_\Psi(\psi|z) f_Z(z)}{f_\Psi(\psi)} = \frac{\frac{1}{x_H - x_L} \cdot \frac{1}{z_H - z_L}}{\frac{\psi_H - \psi}{(x_H - x_L)(z_H - z_L}} = \frac{1}{\psi_H - \psi}$$

and $f_Z(z|\Psi = \psi) = 0$ elsewhere. Because $\tilde{z}_H - \tilde{z}_L = x_H + z_H - \psi = \psi_H - \psi$, the posterior distribution is a proper density.

Case 2: $z_H - z_L > x_H - x_L$.

Case 2a: $\psi < x_H + z_L$. In this case the proof is the same as in case 1a.

Case 2b: $x_H + z_L \leq \psi \leq x_L + z_H$. In this case $f(\psi) = \frac{1}{z_H - z_L}$, $f(z) = \frac{1}{z_H - z_L}$, $f(\psi|z) = \frac{1}{x_H - x_L}$, $\tilde{z}_L = \psi - x_H$, and $\tilde{z}_H = \psi - x_L$. Then in the range $[\tilde{z}_L, \tilde{z}_H]$,

$$f_Z(z|\Psi = \psi) = \frac{f_\Psi(\psi|z) f_Z(z)}{f_\Psi(\psi)} = \frac{\frac{1}{x_H - x_L} \cdot \frac{1}{z_H - z_L}}{\frac{1}{(z_H - z_L)}} = \frac{1}{x_H - x_L}$$

and $f_Z(z|\Psi = \psi) = 0$ elsewhere. Because $\tilde{z}_H - \tilde{z}_L = \psi - x_L - (\psi - x_H) = x_H - x_L$, the posterior distribution is a proper density.

Case 2c: $x_L + z_H < \psi$. In this case the proof is the same as in case 1c.

4.5.1.3 The seller's expectation of ψ_2 conditional on ψ_1

Let $\overline{x} = \frac{x_L + x_H}{2}$ and $\tilde{\tilde{z}}_1 = \frac{\tilde{z}_{L,1} + \tilde{z}_{H,1}}{2}$. Then $E[\psi_2|\psi_1] = \overline{x} + \tilde{\tilde{z}}_1$. If we further define $\overline{z} = \frac{z_L + z_H}{2}$, we can write the unconditional expectation of ψ as $E[\psi] = \overline{\psi} = \overline{x} + \overline{z}$. If $x_H - x_L \geq z_H - z_L$, we can write:

$$E[\psi_2|\psi_1] = \begin{cases} \frac{\psi_1 + x_H + z_L}{2} & \text{if } \psi_L \leq \psi_1 < x_L + z_H \\ \overline{\psi} & \text{if } x_L + z_H \leq \psi_1 \leq x_H + z_L \\ \frac{\psi_1 + x_L + z_H}{2} & \text{if } x_H + z_L < \psi_1 \leq \psi_H \end{cases}$$

Then for all $\psi_1 < \overline{\psi}$, $\psi_1 < E[\psi_2|\psi_1]$, while for all $\psi_1 \geq \overline{\psi}$, $\psi_1 \geq E[\psi_2|\psi_1]$.
If $z_H - z_L > x_H - x_L$:

$$E[\psi_2|\psi_1] = \begin{cases} \frac{\psi_1 + x_H + z_L}{2} & \text{if } \psi_L \leq \psi_1 < x_H + z_L \\ \psi_1 & \text{if } x_H + z_L \leq \psi_1 \leq x_L + z_H \\ \frac{\psi_1 + x_L + z_H}{2} & \text{if } x_L + z_H < \psi_1 \leq \psi_H \end{cases}$$

Then for all $\psi_1 < x_H + z_L$, $\psi_1 < E[\psi_2|\psi_1]$, while for all $\psi_1 \geq x_H + z_L$, $\psi_1 \geq E[\psi_2|\psi_1]$.

4.5.1.4 Proof of Proposition 1

We assume the seller will accept any offer $\psi_1 \geq E[\psi_2|\psi_1]$. Let $\bar{t}(z)$ denote the expected number of periods the seller leaves his house on the market. Then if $x_H -$

$x_L \geq z_H - z_L$:

$$\begin{aligned}
\bar{t}(z) &= Pr(\psi_1 \geq \overline{\psi}) + 2Pr(\psi_1 < \overline{\psi}) \\
&= Pr(x_1 \geq \overline{\psi} - z) + 2Pr(x_1 < \overline{\psi} - z) \\
&= 1 - \frac{\overline{\psi} - x_L - z}{x_H - x_L} + 2\frac{\overline{\psi} - x_L - z}{x_H - x_L} \\
&= 1 + \frac{\overline{\psi} - x_L - z}{x_H - x_L}
\end{aligned}$$

Then

$$\frac{\partial \bar{t}}{\partial z} = \frac{-1}{x_H - x_L} < 0$$

Now consider the case in which $z_H - z_L > x_H - x_L$:

$$\begin{aligned}
\bar{t}(z) &= Pr(\psi_1 \geq x_H + z_L) + 2Pr(\psi_1 < x_H + z_L) \\
&= Pr(x_1 \geq x_H + z_L - z) + 2Pr(x_1 < x_H + z_L - z)
\end{aligned}$$

If $z > x_H - x_L + z_L$, $Pr(x_1 \geq x_H + z_L - z) = 1$, so $\bar{t} = 1 + 0 = 1$. If $z \leq x_H - x_L + z_L$,

$$\begin{aligned}
\bar{t}(z) &= 1 - \frac{x_H - x_L + z_L - z}{x_H - x_L} + 2\frac{x_H - x_L + z_L - z}{x_H - x_L} \\
&= 1 + \frac{x_H - x_L + z_L - z}{x_H - x_L} \\
&= 2 - \frac{z - z_L}{x_H - x_L}
\end{aligned}$$

Then when $z_H - z_L > x_H - x_L$,

$$\bar{t}(z) = \begin{cases} 2 - \frac{z - z_L}{x_H - x_L} & \text{if } z_L \leq z \leq x_H - x_L + z_L \\ 1 & \text{if } z > x_H - x_L + z_L \end{cases}$$

and

$$\frac{\partial \bar{t}}{\partial z} = \begin{cases} \frac{-1}{x_H - x_L} & \text{if } z_L \leq z \leq x_H - x_L + z_L \\ 0 & \text{if } x_H - x_L + z_L \leq z \leq z_H \end{cases}$$

Therefore, $\frac{\partial \bar{t}}{\partial z}$ must always be weakly negative.

Let $\overline{p}(z)$ denote the expected sales price for the house. When $x_H - x_L \geq z_H - z_L$,

$$\begin{aligned}
\overline{p}(z) &= Pr(\psi_1 \geq \overline{x} + \overline{z})E[\psi_1 | \psi_1 \geq \overline{x} + \overline{z}] + Pr(\psi_1 < \overline{x} + \overline{z})E[\psi_2] \\
&= Pr(x_1 \geq \overline{x} + \overline{z} - z)E[\psi_1 | \psi_1 \geq \overline{x} + \overline{z}] + Pr(x_1 < \overline{x} + \overline{z} - z)E[\psi_2] \\
&= (1 - \frac{\overline{x} - x_L + \overline{z} - z}{x_H - x_L})(\frac{\overline{x} + x_H + \overline{z} + z}{2}) + (\frac{\overline{x} - x_L + \overline{z} - z}{x_H - x_L})(\overline{x} + z) \\
&= \frac{\overline{x} + x_H + \overline{z} + z}{2} + (\frac{\overline{x} - x_L + \overline{z} - z}{x_H - x_L})(\frac{\overline{x} - x_H - \overline{z} + z}{2})
\end{aligned}$$

This implies $\frac{\partial \overline{p}}{\partial z} = 1 + \frac{\overline{z} - z}{x_H - x_L}$. To see that this must always be positive, consider the case $z = z_H$. Then $\frac{\partial \overline{p}}{\partial z} = 1 + \frac{\overline{z} - z_H}{x_H - x_L} = \frac{x_H - x_L - (z_H - \overline{z})}{x_H - x_L}$. $\overline{z} > z_L$, so $z_H - \overline{z} < z_H - z_L$. By assumption, $x_H - x_L \geq z_H - z_L$. Therefore the numerator of this expression is positive, and $\frac{\partial \overline{p}}{\partial z} > 0$ when $x_H - x_L \geq z_H - z_L$.

When $z_H - z_L > x_H - x_L$,

$$\begin{aligned}
\overline{p}(z) &= Pr(\psi_1 \geq x_H + z_L)E[\psi_1 | \psi_1 \geq x_H + z_L] + Pr(\psi_1 < x_H + z_L)E[\psi_2] \\
&= Pr(x_1 \geq x_H + z_L - z)E[\psi_1 | \psi_1 \geq x_H + z_L] + Pr(x_1 < x_H + z_L - z)E[\psi_2]
\end{aligned}$$

If $z \geq x_H - x_L + z_L$, $Pr(\psi_1 \geq x_H + z_L) = 1$, and $E[\psi_1 | \psi_1 \geq x_H + z_L] = E[\psi_1]$. Then $\overline{p} = \overline{x} + z$, so $\frac{\partial \overline{p}}{\partial z} = 1$. If $z < x_H - x_L + z_L$, $Pr(\psi_1 \geq x_H + z_L) = \frac{z - z_L}{x_H - x_L}$. Then

$$\begin{aligned}
\overline{p} &= (\frac{z - z_L}{x_H - x_L})(x_H + \frac{z_L + z}{2}) + (1 - \frac{z - z_L}{x_H - x_L})(\overline{x} + z) \\
&= \overline{x} + z + (\frac{z - z_L}{x_H - x_L})(\frac{x_H - x_L + z_L - z}{2})
\end{aligned}$$

implying

$$\frac{\partial \overline{p}}{\partial z} = 1 + \frac{x_H - x_L + z_L - 2z + z_L}{2(x_H - x_L)}$$
$$= \frac{3}{2} - \frac{z - z_L}{x_H - x_L}$$

Therefore when $z_H - z_L > x_H - x_L$,

$$\overline{p}(z) = \begin{cases} \overline{x} + z + (\frac{z-z_L}{x_H-x_L})(\frac{x_H-x_L+z_L-z}{2}) & \text{if } z_L \leq z \leq x_H - x_L + z_L \\ \overline{x} + z & \text{if } z > x_H - x_L + z_L \end{cases}$$

and

$$\frac{\partial \overline{p}}{\partial z} = \begin{cases} \frac{3}{2} - \frac{z-z_L}{x_H-x_L} & \text{if } z_L \leq z \leq x_H - x_L + z_L \\ 1 & \text{if } x_H - x_L + z_L \leq z \leq z_H \end{cases}$$

By assumption $z < x_H - x_L + z_L$, so $\frac{z-z_L}{x_H-x_L} < 1$. Therefore $\frac{\partial \overline{p}}{\partial z} > 0$ when $z_H - z_L > x_H - x_L$ and $z < x_H - x_L + z_L$, implying that $\frac{\partial \overline{p}}{\partial z} > 0$ in all cases. ■

4.5.1.5 Proof of Proposition 2

Let $\tilde{\psi}_L = \tilde{z}_{L,1} + x_L$ and $\tilde{\psi}_L = \tilde{z}_{L,1} + x_L$. The seller's posterior belief about the distribution of ψ_2 is then

$$\tilde{f}(\psi_2|\psi_1,\xi_1) = \begin{cases} \frac{\psi-\tilde{\psi}_L}{(x_H-x_L)(\tilde{z}_{H,1}-\tilde{z}_{L,1})} & \text{if } \tilde{\psi}_L \leq \psi \leq \tilde{\psi}_L + \min(x_H - x_L, \tilde{z}_{H,1} - \tilde{z}_{L,1}) \\ \frac{1}{\max(x_H-x_L,\tilde{z}_{H,1}-\tilde{z}_{L,1})} & \text{if } \tilde{\psi}_L + \min(x_H - x_L, \tilde{z}_{H,1} - \tilde{z}_{L,1}) \leq \psi \\ & \leq \tilde{\psi}_L + \max(x_H - x_L, \tilde{z}_{H,1} - \tilde{z}_{L,1}) \\ \frac{\tilde{\psi}_H-\psi}{(x_H-x_L)(\tilde{z}_{H,1}-\tilde{z}_{L,1})} & \text{if } \tilde{\psi}_L + \max(x_H - x_L, \tilde{z}_{H,1} - \tilde{z}_{L,1}) \leq \psi \leq \tilde{\psi}_H \end{cases}$$

Consider $E[\psi_2|\psi_1, \xi_1 = 1]$. If $x_H - \hat{x}_1 \geq z_H - z_L$ (call this case 1),

$$E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)] = \begin{cases} \overline{x} + \frac{\psi_1 - \hat{x}_1 + z_L}{2} & \text{if } \hat{x}_1 + z_L \leq \psi_1 \leq \hat{x}_1 + z_H \\ \overline{x} + \overline{z} & \text{if } \hat{x}_1 + z_H \leq \psi_1 \leq x_H + z_L \\ \overline{x} + \overline{z} + \frac{\psi_1 - x_H - z_L}{2} & \text{if } x_H + z_L \leq \psi_1 \leq \psi_H \end{cases}$$

If $z_H - z_L > x_H - \hat{x}_1$ (case 2),

$$E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)] = \begin{cases} \overline{x} + \frac{\psi_1 - \hat{x}_1 + z_L}{2} & \text{if } \hat{x}_1 + z_L \leq \psi_1 \leq x_H + z_L \\ \psi_1 - \frac{\hat{x}_1 - x_L}{2} & \text{if } x_H + z_L \leq \psi_1 \leq \hat{x}_1 + z_H \\ \frac{\psi_1 + x_L + z_H}{2} & \text{if } \hat{x}_1 + z_H \leq \psi_1 \leq \psi_H \end{cases}$$

In case 1, if $\hat{x}_1 + z_H \leq \overline{x} + \overline{z}$ (call this case 1a), $\psi_1 \geq E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)]$ when $\psi_1 \geq \overline{x} + \overline{z}$ and $\psi_1 < E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)]$ otherwise. If $\hat{x}_1 + z_H > \overline{x} + \overline{z}$ (case 1b), $\psi_1 \geq E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)]$ when $\psi_1 \geq x_L + x_H - \hat{x}_1 + z_L$ and $\psi_1 < E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)]$ otherwise. In case 2 $\psi_1 \geq E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)]$ when $\psi_1 \geq x_L + x_H - \hat{x}_1 + z_L$ and $\psi_1 < E[\psi_2|\tilde{f}(\psi_2|\psi_1, 1)]$ otherwise.

Again, we assume the seller accepts any offer $\psi_1 \geq E[\psi_2|\psi_1, \xi_1]$. Recall that the seller always rejects an offer when the realtor recommends 'reject'. Then in case 1a ($x_H - \hat{x}_1 \geq z_H - z_L$ and $\hat{x}_1 + z_H \leq \overline{x} + \overline{z}$),

$$\begin{aligned}
\overline{t}(z) &= 2Pr(\xi_1 = 0) + Pr(\psi_1 \geq \overline{x} + \overline{z} \cap \xi_1 = 1) + 2Pr(\psi_1 < \overline{x} + \overline{z} \cap \xi_1 = 1) \\
&= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(Pr(x_1 \geq \overline{x} + \overline{z} - z|\xi_1 = 1) + 2Pr(x_1 < \overline{x} + \overline{z} - z|\xi_1 = 1)) \\
&= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(1 - \frac{\overline{x} - \hat{x}_1 + \overline{z} - z}{x_H - \hat{x}_1} + 2(\frac{\overline{x} - \hat{x}_1 + \overline{z} - z}{x_H - \hat{x}_1})) \\
&= 1 + \frac{\hat{x}_1 - x_L}{x_H - x_L} + (\frac{x_H - \hat{x}_1}{x_H - x_L})(\frac{\overline{x} - \hat{x}_1 + \overline{z} - z}{x_H - \hat{x}_1}) \\
&= 1 + \frac{\overline{x} - x_L + \overline{z} - z}{x_H - x_L}
\end{aligned}$$

Therefore in case 1a,

$$\frac{\partial \bar{t}}{\partial z} = \frac{-1}{x_H - x_L} < 0$$

In cases 1b and 2 ($\hat{x}_1 + z_H > \overline{x} + \overline{z}$),

$$\begin{aligned}
\bar{t}(z) &= 2Pr(\xi_1 = 0) + Pr(\psi_1 \geq x_L + x_H - \hat{x}_1 + z_L \cap \xi_1 = 1) \\
&\qquad + 2Pr(\psi_1 < x_L + x_H - \hat{x}_1 + z_L \cap \xi_1 = 1) \\
&= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(Pr(x_1 \geq x_L + x_H - \hat{x}_1 + z_L - z|\xi_1 = 1) \\
&+ 2Pr(x_1 < x_L + x_H - \hat{x}_1 + z_L - z|\xi_1 = 1))
\end{aligned}$$

If $z > x_L + x_H - 2\hat{x}_1 + z_L$, $Pr(x_1 \geq x_L + x_H - \hat{x}_1 + z_L - z|\xi_1 = 1) = 1$. In that case $\bar{t} = 2Pr(\xi_1 = 0) + Pr(\xi_1 = 1) = 2\frac{(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L}) = 1 + \frac{\hat{x}_1 - x_L}{x_H - x_L}$. If $z < x_L + x_H - 2\hat{x}_1 + z_L$,

$$\begin{aligned}
\bar{t}(z) &= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L}) \\
&\quad \times (1 - \frac{x_L + x_H - 2\hat{x}_1 + z_L - z}{x_H - \hat{x}_1} + 2(\frac{x_L + x_H - 2\hat{x}_1 + z_L - z}{x_H - \hat{x}_1})) \\
&= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(1 + \frac{x_L + x_H - 2\hat{x}_1 + z_L - z}{x_H - \hat{x}_1}) \\
&= 1 + \frac{\hat{x}_1 - x_L}{x_H - x_L} + (\frac{x_H - \hat{x}_1}{x_H - x_L})(\frac{x_L + x_H - 2\hat{x}_1 + z_L - z}{x_H - \hat{x}_1}) \\
&= 1 + \frac{x_H - \hat{x}_1 + z_L - z}{x_H - x_L}
\end{aligned}$$

Then in cases 1b and 2,

$$\bar{t}(z) = \begin{cases} 1 + \frac{x_H - \hat{x}_1 + z_L - z}{x_H - x_L} & \text{if } z_L \leq z < x_L + x_H - 2\hat{x}_1 + z_L \\ 1 + \frac{\hat{x}_1 - x_L}{x_H - x_L} & \text{if } x_L + x_H - 2\hat{x}_1 + z_L \leq z \leq z_H \end{cases}$$

and

$$\frac{\partial \bar{t}}{\partial z} = \begin{cases} \frac{-1}{x_H - x_L} & \text{if } z_L \leq z < x_L + x_H - 2\hat{x}_1 + z_L \\ 0 & \text{if } x_L + x_H - 2\hat{x}_1 + z_L \leq z \leq z_H \end{cases}$$

In case 2 ($z_H - z_L > x_H - \hat{x}_1$),

$$\begin{aligned} \bar{t}(z) &= 2Pr(\xi_1 = 0) + Pr(\psi_1 \geq x_L + x_H - \hat{x}_1 + z_L \cap \xi_1 = 1) \\ &\quad + 2Pr(\psi_1 < x_L + x_H - \hat{x}_1 + z_L \cap \xi_1 = 1) \\ &= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(Pr(x_1 \geq x_L + x_H - \hat{x}_1 + z_L - z | \xi_1 = 1) \\ &+ 2Pr(x_1 < x_L + x_H - \hat{x}_1 + z_L - z | \xi_1 = 1)) \end{aligned}$$

If $z > x_L + x_H - 2\hat{x}_1 + z_L$, $Pr(x_1 \geq x_L + x_H - \hat{x}_1 + z_L - z | \xi_1 = 1) = 1$. In that case $\bar{t} = 2Pr(\xi_1 = 0) + Pr(\xi_1 = 1) = 2\frac{(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L}) = 1 + \frac{\hat{x}_1 - x_L}{x_H - x_L}$. Therefore $\frac{\partial \bar{t}}{\partial z} = 0$. If $z < x_L + x_H - 2\hat{x}_1 + z_L$,

$$\begin{aligned} \bar{t}(z) &= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L}) \\ &\quad \times (1 - \frac{x_L + x_H - \hat{x}_1 + z_L - z}{x_H - \hat{x}_1} + 2(\frac{x_L + x_H - \hat{x}_1 + z_L - z}{x_H - \hat{x}_1})) \\ &= \frac{2(\hat{x}_1 - x_L)}{x_H - x_L} + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(1 + \frac{x_L + x_H - \hat{x}_1 + z_L - z}{x_H - \hat{x}_1}) \end{aligned}$$

In this case,

$$\frac{\partial \bar{t}}{\partial z} = (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})\frac{-1}{x_H - \hat{x}_1} < 0$$

Therefore, in case 2

$$\frac{\partial \bar{t}}{\partial z} = \begin{cases} (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})\frac{-1}{x_H - \hat{x}_1} & \text{if } z_L \leq z < x_L + x_H - 2\hat{x}_1 + z_L \\ 0 & \text{if } x_L + x_H - 2\hat{x}_1 + z_L \leq z \leq z_H \end{cases}$$

Thus, $\frac{\partial \bar{t}}{\partial z}$ must always be weakly negative.

In case 1a,

$$\overline{p}(z) = Pr(\xi_1 = 0)E[\psi_2] + Pr(\xi_1 = 1)\left[Pr(\psi_1 \geq \overline{x} + \overline{z}|\xi_1 = 1)E[\psi_1|\psi_1 \geq \overline{x} + \overline{z}, \xi_1 = 1]\right.$$

$$\left. + Pr(\psi_1 < \overline{x} + \overline{z}|\xi_1 = 1)E[\psi_2]\right)]$$

$$= (\frac{\hat{x}_1 - x_L}{x_H - x_L})(\overline{x} + z) + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})\left[(1 - \frac{\overline{x} - \hat{x}_1 + \overline{z} - z}{x_H - \hat{x}_1})(z + \frac{\overline{x} + x_H + \overline{z} - z}{2})\right.$$

$$\left. + (\frac{\overline{x} - \hat{x}_1 + \overline{z} - z}{x_H - \hat{x}_1})(\overline{x} + z))\right]$$

$$= \frac{\frac{1}{2}(\overline{x}^2 + x_H^2 - \overline{z}^2 - z^2) - x_L\overline{x} + (x_H - x_L + \overline{z})z}{x_H - x_L}$$

Therefore in case 1a,

$$\frac{\partial \overline{p}}{\partial z} = \frac{x_H - x_L + \overline{z} - z}{x_H - x_L}$$

To see that this must be positive, consider the case in which $z = z_H$. Then

$$\frac{\partial \overline{p}}{\partial z} = \frac{x_H - x_L + \overline{z} - z_H}{x_H - x_L} = \frac{x_H - x_L - \frac{z_H - z_L}{2}}{x_H - x_L}$$

By assumption $x_H - x_L \geq x_H - \hat{x}_1 \geq z_H - z_L$, so $\frac{\partial \overline{p}}{\partial z} \geq 0$ in case 1a. In cases 1b and 2, if $z > x_L + x_H - 2\hat{x}_1 + z_L$ (so that the seller will always accept the period 1 offer if the realtor recommends accept),

$$\overline{p} = Pr(\xi_1 = 0)E[\psi_2] + Pr(\xi_1 = 1)E[\psi_1|x_1 \geq \hat{x}_1]$$

$$= (\frac{\hat{x}_1 - x_L}{x_H - x_L})(\overline{x} + z) + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})(z + \frac{\hat{x}_1 + x_H}{2})$$

$$= \frac{(\hat{x}_1 - x_L)(x_L + x_H) + (x_H - \hat{x}_1)(\hat{x}_1 + x_H)}{2(x_H - x_L)} + z$$

If $z < x_L + x_H - 2\hat{x}_1 + z_L$,

$$\begin{aligned}
\overline{p}(z) &= Pr(\xi_1 = 0)E[\psi_2] + Pr(\xi_1 = 1)\left(Pr(\psi_1 \geq x_L + x_H - \hat{x}_1 + z_L | x_1 \geq \hat{x}_1)\right. \\
&\left.\times E[\psi_1 | \psi_1 \geq x_L + x_H - \hat{x}_1 + z_L, x_1 \geq \hat{x}_1] + Pr(\psi_1 < x_L + x_H - \hat{x}_1 + z_L | x_1 \geq \hat{x}_1)E[\psi_2]\right) \\
&= (\frac{\hat{x}_1 - x_L}{x_H - x_L})(\overline{x} + z) + (1 - \frac{\hat{x}_1 - x_L}{x_H - x_L})\Big[(1 - \frac{x_L + x_H - 2\hat{x}_1 + z_L - z}{x_H - \hat{x}_1}) \\
&\times (z + \frac{x_L + x_H - \hat{x}_1 + z_L - z + x_H}{2}) + (\frac{x_L + x_H - 2\hat{x}_1 + z_L - z}{x_H - \hat{x}_1})(\overline{x} + z)\Big] \\
&= (\frac{1}{x_H - x_L})\Big(-\frac{1}{2}(x_L^2 + \hat{x}_1^2 + z_L^2 + z^2) + (\overline{x} + z_L)\hat{x}_1 + (x_H + z_L)\overline{x} \\
&- (x_H + z_L + z_H)x_L + (2x_H - \overline{x} - \hat{x}_1 - x_L + z_L)z\Big)
\end{aligned}$$

Then in case 1b,

$$\overline{p}(z) = \begin{cases} \dfrac{-\frac{1}{2}(x_L^2+\hat{x}_1^2+z_L^2+z^2)+(\overline{x}+z_L)\hat{x}_1+(x_H+z_L)\overline{x} \quad -(x_H+z_L+z_H)x_L+(2x_H-\overline{x}-\hat{x}_1-x_L+z_L)z}{x_H-x_L} & \text{if } z_L \leq z < x_L + x_H - 2\hat{x}_1 + z_L \\ \dfrac{(\hat{x}_1-x_L)(x_L+x_H)+(x_H-\hat{x}_1)(\hat{x}_1+x_H)}{2(x_H-x_L)} + z & \text{if } x_L + x_H - 2\hat{x}_1 + z_L \leq z \leq z_H \end{cases}$$

and

$$\frac{\partial \overline{p}}{\partial z} = \begin{cases} \frac{2x_H - \overline{x} - \hat{x}_1 - x_L + z_L - z}{x_H - x_L} & \text{if } z_L \leq z < x_L + x_H - 2\hat{x}_1 + z_L \\ 1 & \text{if } x_L + x_H - 2\hat{x}_1 + z_L \leq z \leq z_H \end{cases}$$

To see that $\frac{\partial \overline{p}}{\partial z} > 0$ when $z \leq x_L + x_H - 2\hat{x}_1 + z_L$, note that $2x_H - \overline{x} - \hat{x}_1 + z_L \geq x_L + x_H - 2\hat{x}_1 + z_L$ is equivalent to $x_H - x_L \geq \overline{x} - \hat{x}_1$, which must be true because $x_H \geq \overline{x}$ and $\hat{x}_1 \geq x_L$. Therefore $\frac{\partial \overline{p}}{\partial z} > 0$ in cases 1b and 2. ■

4.5.2 Extending the Model to Multiple Periods

This section of the paper will relax one of the more restrictive assumptions in the previous sections, that the seller must accept the second period offer if he rejects the first period offer. Although the present paper extends the model only to allow for a third period, hopefully this extension will illustrate the key differences between a two period and multiple period model and present an equilibrium concept that is compatible with any finite number of periods. There are two main differences between the equilibria of the two period model and the three period model. First, in the first period of the three period model, the realtor's cutoff rule for recommending 'accept' will be a function of z. Second, in equilibrium there will be ranges of first-period offers for which the realtor will "babble": his recommendation will contain no information regarding the state of demand, and consequently the seller will ignore it when updating his beliefs. Without this feature of equilibrium, the realtor will sometimes have an incentive to misreport his own preference regarding the seller's decision in period 1 in order to manipulate the seller's beliefs regarding the state of demand.

The notation of the three-period model will follow the notation of the two-period model closely. The realtor's recommendation in period t shall be denoted ξ_t and the seller's decision in period t shall be denoted γ_t, where in both cases a value of 1 indicates 'accept' and a value of 0 indicates 'reject'. Let $\widehat{x}_1(z)$ and $\widehat{x}_2$ denote the realtor's cutoff value for reporting 'accept' in periods 1 and 2, respectively. As discussed, in equilibrium there will be some values for ψ_1 such that the realtor will babble; for all other values of ψ_1 the realtor will employ a cutoff rule in x_1 to determine his recommendation, but the cutoff will be a function of z. Let $\widetilde{z}_{L,t}$ and $\widetilde{z}_{H,t}$ denote the seller's beliefs about the lowest and highest possible values of z after receiving the period t offer and recommendation. Then the realtor's value function can be written:

$$
\begin{aligned}
V_R(\psi_1, z, 1) = \overset{\max}{\xi_1} \{ & \gamma_1(\psi_1, \tilde{f}(\psi|\psi_1, 0)\alpha\psi_1 + (1 - \gamma_1(\psi_1, \tilde{f}(\psi|\psi_1, 0))) \\
& \times (E[V_R(\psi_2, z, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}, 2)|\xi_1 = 0]), \\
& \gamma_2(\psi_1, \tilde{f}(\psi|\psi_1, 1)\alpha\psi_1 + (1 - \gamma_1(\psi_1, \tilde{f}(\psi|\psi_1, 1))) \\
& \times (E[V_R(\psi_2, z, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}, 2)|\xi_1 = 1])\}
\end{aligned}
$$

$$
\begin{aligned}
V_R(\psi_2, z, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}, 2) = \overset{\max}{\xi_2} \{ & \gamma_2(\psi_2, \tilde{f}(\psi|\psi_2, 0, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})\alpha\psi_2 + \\
& (1 - \gamma_2(\psi_2, \tilde{f}(\psi_2|\psi_2, 0, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})))(E[V_R(\psi_3, z, \widetilde{z}_{L,2}, \widetilde{z}_{H,2}, 3)|\xi_2 = 0]), \\
& \gamma_2(\psi_2, \tilde{f}(\psi|\psi_1, 1, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})\alpha\psi_2 + \\
& (1 - \gamma_2(\psi_2, \tilde{f}(\psi_2|\psi_1, 1, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})))(E[V_R(\psi_2, z, \widetilde{z}_{L,2}, \widetilde{z}_{H,2}, 2)|\xi_2 = 1])\}
\end{aligned}
$$

$$V_R(\psi_3, z, \widetilde{z}_{L,2}, \widetilde{z}_{H,2}, 3) = -c_R + \alpha\psi_3$$

The seller's value function can be written:

$$V_S(\psi_1, \xi_1, 1) = \overset{\max}{\gamma_1}(1 - \alpha)\{\psi_1, E[V_S(\psi_2, \xi_2, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}, 2)|\tilde{f}(\psi_2|\psi_1, \xi_1)]\}$$

$$V_S(\psi_2, \xi_2, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}, 2) = \overset{\max}{\gamma_2}(1 - \alpha)\{\psi_2, E[\psi_3|\tilde{f}(\psi_3|\psi_1, \xi_2,)]\}$$

$$V_S(\psi_3, \xi_3, \widetilde{z}_{L,2}, \widetilde{z}_{H,2}, 3) = (1 - \alpha)\psi_3$$

Then we can update our equilibrium definition as follows.

Definition: a **Bayesian Nash equilibrium** of the game between realtors and sellers is a set of policy functions $\xi_1(\psi_1, z)$ and $\xi_2(\psi 2, z, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})$ for the realtor, a set of policy functions $\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, \xi_1))$ and $\gamma_2(\psi_2, \tilde{f}(\psi_3|\psi_2, \xi_2, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}))$ for the seller, and a set of belief updating strategies $\tilde{f}(\psi_2|\psi_1, \xi_1)$ and $\tilde{f}(\psi_3|\psi_2, \xi_2, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})$

for the seller such that:

1. $\xi_1(\psi_1, z)$ and $\xi_2(\psi 2, z, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})$ solve the realtor's problem taking the seller's policy functions and belief updating strategies as given;
2. $\gamma_1(\psi_1, \tilde{f}(\psi_2|\psi_1, \xi_1))$ and $\gamma_2(\psi_2, \tilde{f}(\psi_3|\psi_2, \xi_2, \widetilde{z}_{L,1}, \widetilde{z}_{H,1}))$ solve the seller's problem taking the realtor's policy functions as given; and
3. $\tilde{f}(\psi_2|\psi_1, \xi_1)$ and $\tilde{f}(\psi_3|\psi_2, \xi_2, \widetilde{z}_{L,1}, \widetilde{z}_{H,1})$ are consistent with the realtor's policy functions.

As discussed, the first main difference between the equilibria of three-period model and the two-period model is that in the first period of the three-period model, the realtor's optimal cutoff strategy is a function of z. This is because when z is low, it is more likely that ψ_2 will be so low enough that the seller rejects the offer regardless of the realtor's recommendation, and this loss of control lowers the realtor's payoff. To determine the realtor's optimal cutoff rule in period 1, consider the realtor's payoff minus αz. If the seller accepts an offer in period t, this value will be $\alpha(x_t - c_R(t-1))$. Call this value the idiosyncratic component of the realtor's payoff. The realtor will receive the αz portion of his payoff no matter which offer the seller accepts, but the idiosyncratic portion of his payoff depends on the seller's decisions and on the realizations of x_t. If the seller accepts the first period offer, the idiosyncratic portion of the realtor's payoff is αx_1. Therefore the realtor would prefer that the seller accept any period 1 offers such that $\alpha x_1 \geq E[\alpha(x_t - c_R(t-1))]$ and to reject all others. Recall that the realtor's preference in the penultimate period is for the seller to accept any offers such that $x_2 \geq \overline{x} - \frac{c_R}{\alpha}$ and reject all others. Call this value $\widehat{x}_2$. Then the realtor's optimal cutoff rule in period 1 is defined implicitly by the following mapping (where $\widehat{x}_1(z)$ is written $\widehat{x}_1$ on the right-hand side for simplicity):

$$
\widehat{x}_1(z) = \begin{cases} \dfrac{\frac{1}{2}(x_H^2 - \overline{x}^2 - \widetilde{\widetilde{z}}_1(\widehat{x}_1)^2 - z^2) + (\overline{x} - \widehat{x}_2 + \widetilde{\widetilde{z}}_1(\widehat{x}_1))z + (\overline{x} - x_L + \widetilde{\widetilde{z}}_1(\widehat{x}_1))\widehat{x}_2 - \widetilde{\widetilde{z}}_1(\widehat{x}_1)\overline{x}}{x_H - x_L} & \text{if } \widehat{x}_2 + \widetilde{z}_{H,1}(\widehat{x}_1) \leq \overline{x} + \widetilde{\widetilde{z}}_1(\widehat{x}_1) \\ \frac{\widehat{x}_2 + x_H}{2} + \left(\frac{\widehat{x}_2 - x_L}{x_H - x_L}\right)\left(\frac{\widehat{x}_2 - x_H}{2}\right) & \text{if } \widehat{x}_2 + \widetilde{z}_{H,1}(\widehat{x}_1) \geq \overline{x} + \widetilde{\widetilde{z}}_1(\widehat{x}_1) \\ & \text{and } z \geq x_L + x_H - 2\widehat{x}_2 + \widetilde{z}_{L,1}(\widehat{x}_1) \\ \dfrac{(x_H - \widehat{x}_2 + \widetilde{z}_{L,1}(\widehat{x}_1) - z)\widehat{x}_2 + \frac{1}{2}(\widehat{x}_2 - x_L - \widetilde{z}_{L,1}(\widehat{x}_1) + z) \times (x_L + 2x_H - \widehat{x}_2 + \widetilde{z}_{L,1}(\widehat{x}_1) - z)}{(x_H - x_L)} & \text{if } \widehat{x}_2 + \widetilde{z}_{H,1}(\widehat{x}_1) \geq \overline{x} + \widetilde{\widetilde{z}}_1(\widehat{x}_1) \\ & \text{and } z < x_L + x_H - 2\widehat{x}_2 + \widetilde{z}_{L,1}(\widehat{x}_1) \end{cases}
$$

Because $\widehat{x}_1(z)$ is the fixed point of this functional equation, I have estimated it numerically. A description of the estimation algorithm is included at the end of this section. Figure 6 illustrates the realtor's optimal period 1 cutoff rule as a function of z.

Given the realtor's cutoff rule, the seller will update his beliefs concerning z as follows. Define $\widetilde{\widetilde{z}}$ as the value of z such that $\psi_1 - z = \widehat{x}_1(z)$. Further define

$$
\widetilde{z}_{L,1} = \begin{cases} \max(z_L, \widetilde{\widetilde{z}}) & \text{if } \gamma_1 = 0 \\ \max(z_L, \psi_1 - x_H) & \text{if } \gamma_1 = 1 \end{cases} \quad \text{and } \widetilde{z}_{H,1} = \begin{cases} \min(z_H, \psi_1 - x_L) & \text{if } \gamma_1 = 0 \\ \min(\widetilde{\widetilde{z}}, \psi_1 - x_L) & \text{if } \gamma_1 = 1 \end{cases}
$$

Then the seller's posterior belief about the distribution of z will again be that $z \sim U[\widetilde{z}_{L,1}, \widetilde{z}_{H,1}]$.[12] In period 2, the seller's belief updating strategy will be the same as the one outlined in the two-period model.

The second main difference between the equilibria of the two and three period models is the presence of 'babbling regions' in period 1 of the three period model. To see that these regions are a necessary feature of equilibrium, consider a hypothetical

[12] Again I omit the proof, but it will follow the proof of Lemma 1 closely.

equilibrium in which the realtor always reports his own preference truthfully to the seller in period 1, and in which the realtor's and seller's behavior in the second period is the same as their behavior in the first period of the two-period model. Then the seller's expected value of rejecting the first offer would look as in the top panel of figure 7. Define $\psi_1^\star$ as the fixed point of the seller's expected value of waiting for period 2 conditional on the realtor recommending 'accept', and $\psi_1^\dagger$ as the fixed point of the seller's expected value of waiting for period 2 conditional on the realtor recommending 'reject'. Then for $\psi_1 < \psi_1^\star$, the expected value of waiting for period 2 is higher than ψ_1 whether the the realtor recommends 'accept' or 'reject'–the seller's action will not depend on the realtor's recommendation in this range. Similarly, when $\psi_1 > \psi_1^\dagger$, the expected value of waiting for period 2 will be below ψ_1 no matter what the realtor recommends. Therefore, the seller will always reject offers $\psi_1 < \psi_1^\star$ and will always accept offers $\psi_1 > \psi_1^\dagger$.

Consider the realtor's best response when $\psi_1 < \psi_1^\star$ and $x_1 > \widehat{x}_1(z)$ in the hypothetical equilibrium. No matter what the realtor recommends, the seller will reject the offer. However, because the seller expects the realtor to report his own preference regarding the offer truthfully, the seller will have a higher expectation of future offers if the realtor recommends 'reject' (thus indicating that z is high) than if the realtor recommends 'accept' (indicating that z is low). Intuitively, this state of affairs cannot be optimal for the realtor: we have shown in the two-period model that in the second-to-last period, the seller will always reject an offer that the realtor recommends rejecting. However, there are second-period offers the realtor recommends accepting that the seller will not accept. Therefore, the realtor will always prefer that the seller be more pessimistic (have a lower expectation of ψ) in the second period: a pessimistic seller is more likely to accept offers the realtor would like him to accept, but will always reject offers the realtor would like him to reject. Therefore, the realtor has a unilateral incentive to deviate from his proposed strategy in the hypothetical equilibrium. For any offer $\psi_1 < \psi_1^\star$, the realtor should recommend 'reject'.

Babbling regions solve the problem of the realtor's incentive to misreport his preferences in the first period. As we have seen, if the realtor's recommendation in period 1 changes the seller's expectation of future offers but does not change the seller's action, the realtor will always choose to send the message that will make the seller more pessimistic. Then in equilibrium, it cannot be the case that the realtor's first period recommendation changes the seller's beliefs when it does not change the seller's first period action. Therefore, in the equilibrium I will consider, the realtor will babble in the first period when his recommendation will not change the seller's action, and will report his own preference truthfully when his recommendation is decisive for the seller's action. The bottom panel of figure 7 illustrates the period 1 equilibrium, assuming that in the second period the seller and the realtor play the same strategies as they did in the first period of the two-period model. The second period equilibrium will then look like it does in Figure 4.

Because $\widehat{x}_1(z)$ must be estimated numerically, the expected sales price and expected time to sale must be simulated as well. Figure 8 shows the results from such a simulation. The general pattern from the two-period model remains intact: as aggregate demand rises, the expected sales price rises and the expected time on market mostly falls. However, there is a slight bump in the expected time to sale for high levels of z. This is due to the upward-sloping portion of $\widehat{x}_1(z)$, which causes the realtor to recommend rejecting a higher percentage of offers when z is high. Overall, however, the correlation between expected sales price and expected time to sale is negative, and the simulated results of the three period model are consistent with the stylized facts observed in the data.

I used the following algorithm for finding the realtor's cutoff rule $\widehat{x}_1(z)$ in the three period model:

1. Pick a candidate schedule for $\widehat{x}_1(z)$. In practice I chose $\widehat{x}_1(z) = \overline{x}$ for all z.

2. On a fine grid of points for z:

    (a) Go through a fine grid of points for all values of x to create a grid of all values of ψ consistent with each value of z.

    (b) Calculate $\widetilde{z}_{L,1}$ and $\widetilde{z}_{H,1}$ for each value of ψ conditional on the realtor recommending 'reject'.

    (c) For each value of ψ, find the expected value of the idiosyncratic component of the realtor's payoff if the seller rejects the first period offer. Denote this value $\overline{x}_R$. For a fixed z, this gives $\overline{x}_R$ as a function of x.

    (d) Find the fixed point of $\overline{x}_R(x)$; use this value as the new candidate for $\widehat{x}_1(z)$.

3. Repeat this procedure using the new schedule for $\widehat{x}_1(z)$ until the maximum distance between the old and new schedules is below a specified tolerance level.

Table 4.1: House Price Misperceptions and Home Sales at the State Level

Dependent Variable	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold
Specification	Levels	Levels	Levels	Levels	First Differences	First Differences	First Differences	First Differences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yearly Misperceptions Index	-1.786	-1.691	-0.516	-0.996	-1.597	-1.453	-0.359	-0.752
	(0.213)	(0.223)	(0.204)	(0.203)	(0.253)	(0.296)	(0.231)	(0.242)
Zillow Home Value Index		0.097		-0.616		0.110		-0.567
		(0.065)		(0.097)		(0.119)		(0.134)
Constant	11.615	11.587	11.541	11.751	0.006	0.002	0.016	-0.006
	(0.152)	(0.147)	(0.155)	(0.152)	(0.010)	(0.011)	(0.023)	(0.023)
Year Fixed Effects?	No	No	Yes	Yes	No	No	Yes	Yes
Number of States	23	23	23	23	23	23	23	23
Number of Observations	253	253	253	253	230	230	230	230
R-squared within	0.237	0.243	0.565	0.637	0.161	0.166	0.558	0.588

Notes: Fixed Effects estimates using yearly state-level data from 2000-2010. Homes sold and Zillow Home Value Index data is from Zillow.com. The misperceptions index is the difference between the change in housing values from 2000 reported in the American Community Survey and the change in housing values from 2000 according to the Zillow Home Value Index.

Table 4.2: House Price Misperceptions and Home Sales at the MSA Level

Dependent Variable	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold	Log Homes Sold
Specification	Levels	Levels	Levels	Levels	First Differences	First Differences	First Differences	First Differences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cumulative Misperceptions Index	-1.881	-2.877	-0.130	-1.794	-0.512	-1.068	-0.257	-1.152
	(0.161)	(0.203)	(0.150)	(0.151)	(0.174)	(0.164)	(0.178)	(0.178)
Zillow Home Value Index		-0.572		-1.058		-0.841		-1.076
		(0.076)		(0.059)		(0.072)		(0.094)
Constant	9.694	9.723	9.421	9.369	-0.089	-0.111	-0.040	-0.082
	(0.104)	(0.105)	(0.106)	(0.105)	(0.008)	(0.008)	(0.016)	(0.015)
Year Fixed Effects?	No	No	Yes	Yes	No	No	Yes	Yes
Number of MSAs	113	113	113	113	113	113	113	113
Number of Observations	678	678	678	678	565	565	565	565
R-squared within	0.205	0.278	0.546	0.712	0.017	0.211	0.117	0.312

Notes: Fixed Effects estimates using yearly metropolitan-level data from 2005-2010. Homes sold and Zillow Home Value Index data is from Zillow.com. The misperceptions index is the difference between the change in housing values from 2005 reported in the American Community Survey and the change in housing values from 2005 according to the Zillow Home Value Index.

Table 4.3: Determinants of House Price Misperceptions

Dependent Variable	Std. Dev. of Misperceptions Index (State Level)	Std. Dev. of Misperceptions Index (MSA Level)
Specification	Levels	Levels
	(1)	(2)
Std. Dev. Of Zillow Home Value Index	0.227	0.213
	(0.035)	(0.018)
Ratio of Top Third to Bottom Third Home Prices		0.012
		(0.003)
Median Error of Zillow Home Value Index	0.660	
	(0.183)	
Constant	-0.047	-0.013
	(0.020)	(0.009)
Number of Observations	23	113
Adjusted R-squared	0.669	0.613

Notes: MSA-level regression uses data from 2005-2010; state-level regression uses data from 2000-2010. Ratio of top third to bottom third home prices and median error of Zillow Home Value Index data is from Zillow.com. The cumulative misperceptions index is as defined in Tables 2 and 3.

Figure 4.1: Housing Market Time Series 1983q1-2010q4

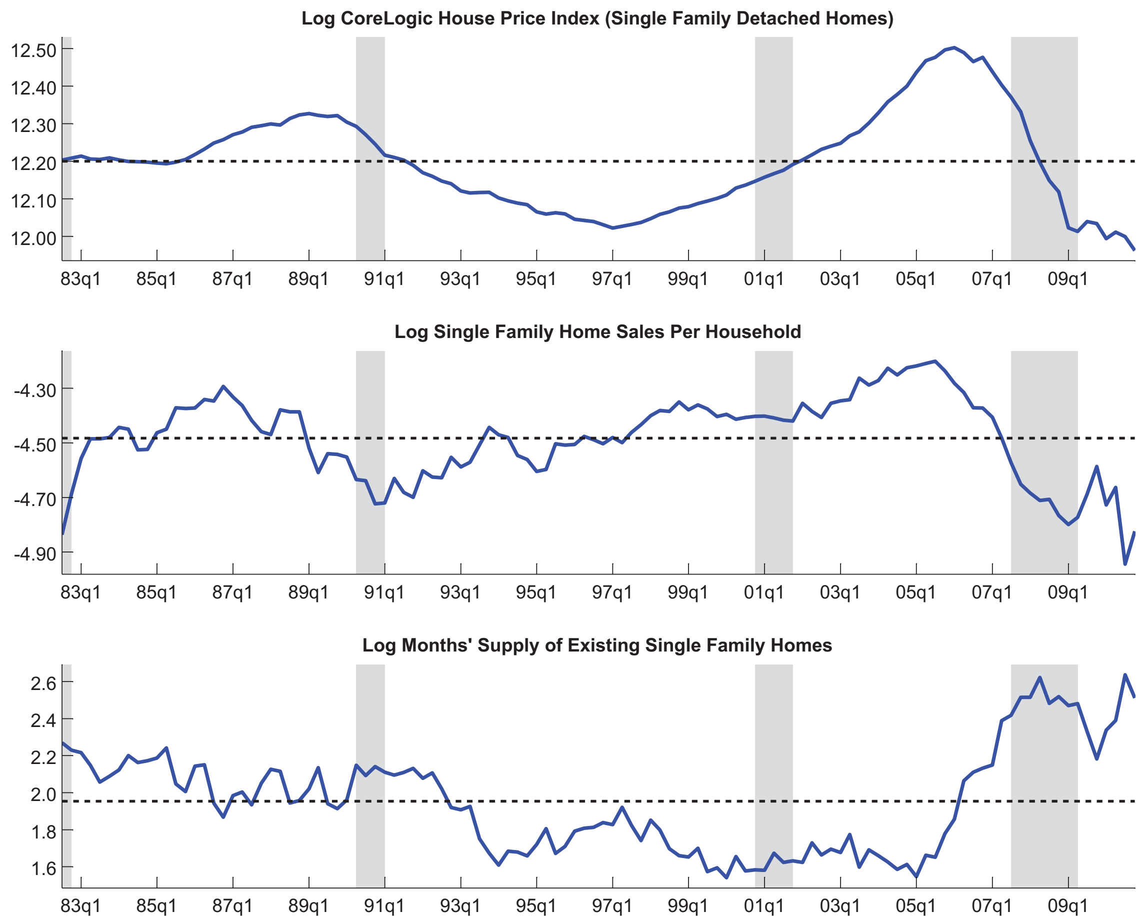
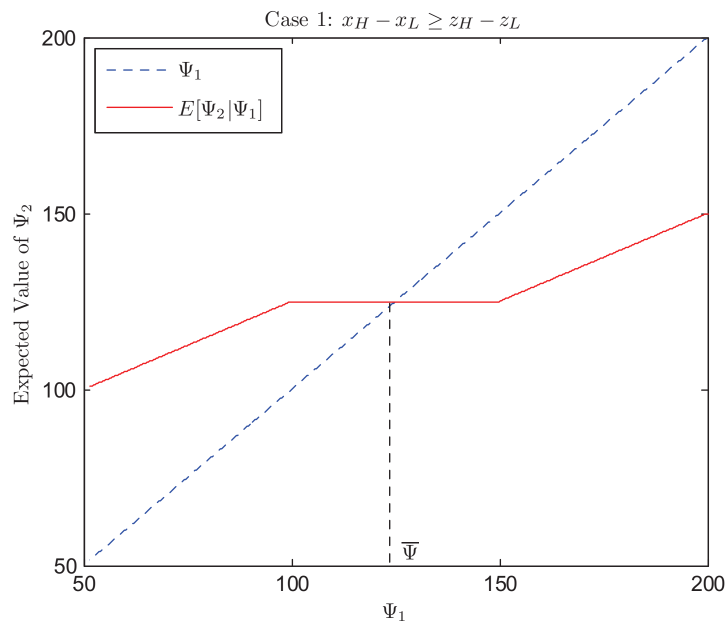


Figure 4.2: Expected Values of ψ_2 as a function of ψ_1



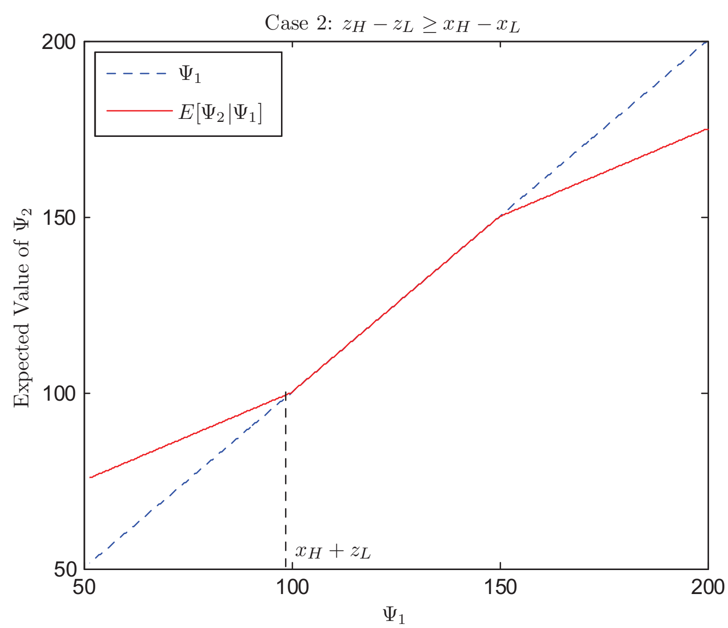


Figure 4.3: Expected Time to Sale and Sales Price without Realtor

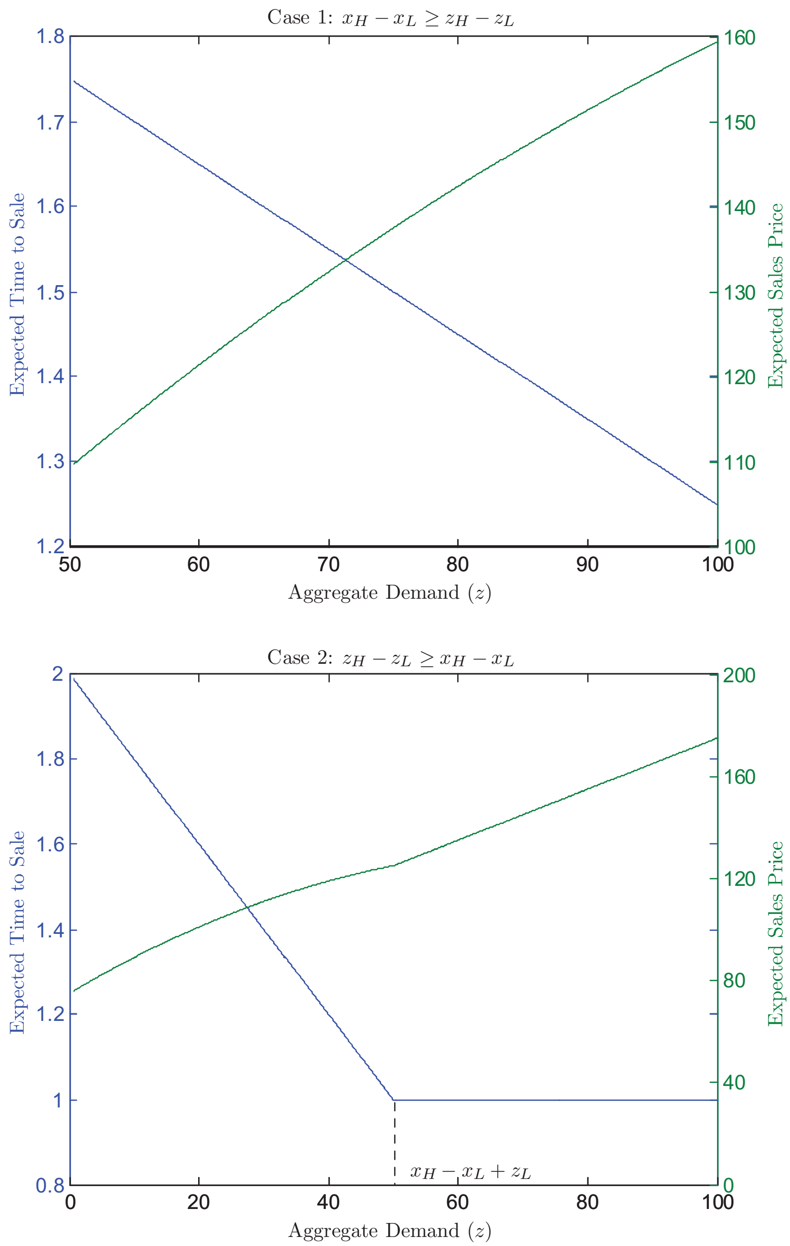
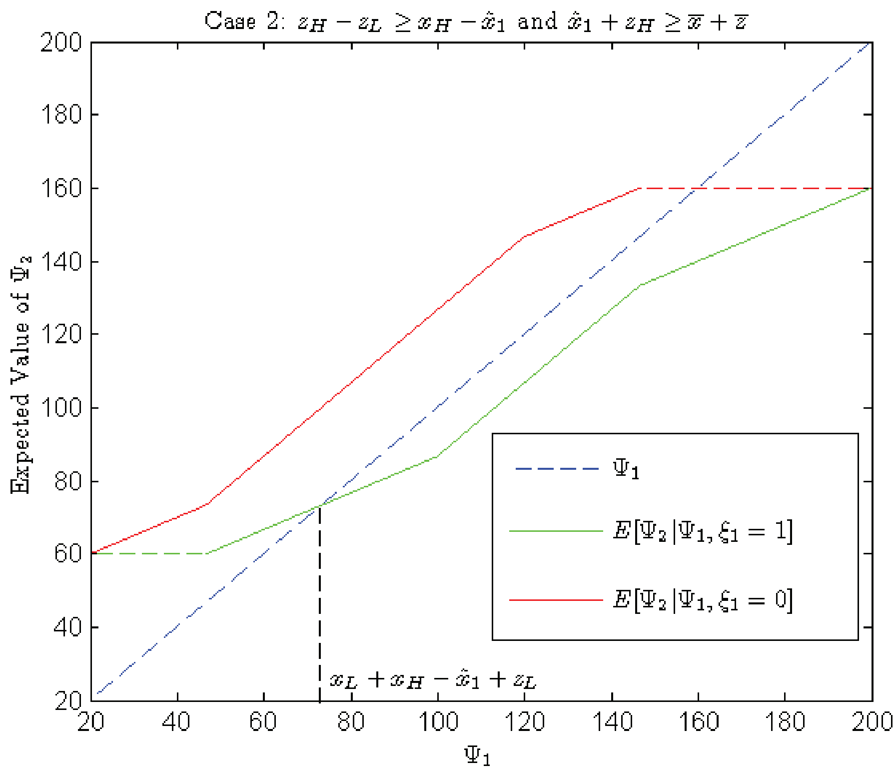


Figure 4.4: Expected Values of ψ_2 as a function of ψ_1



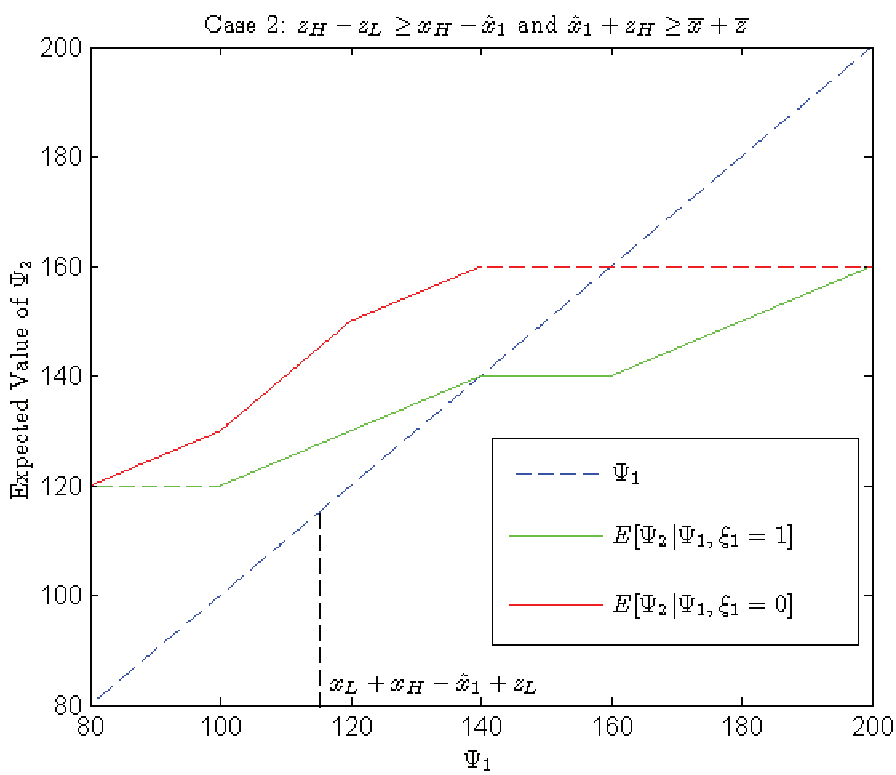
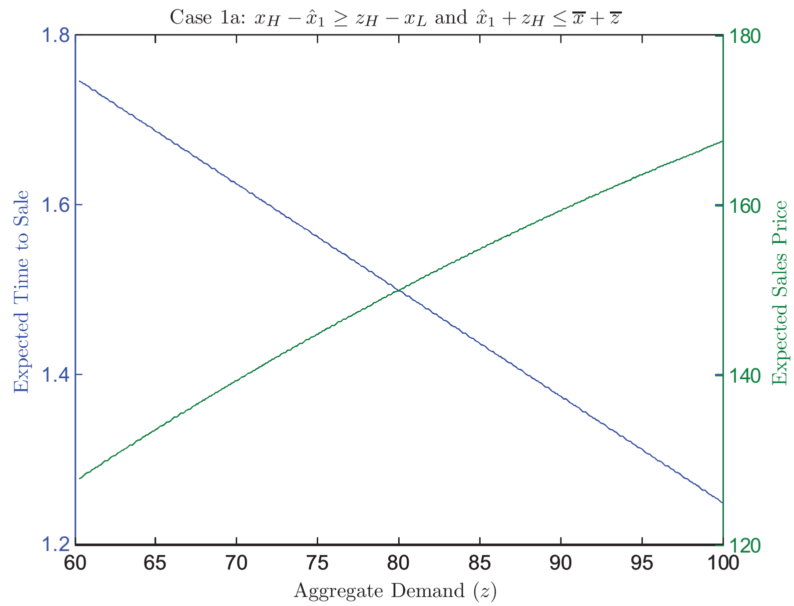


Figure 4.5: Expected Time to Sale and Sales Price with Realtor



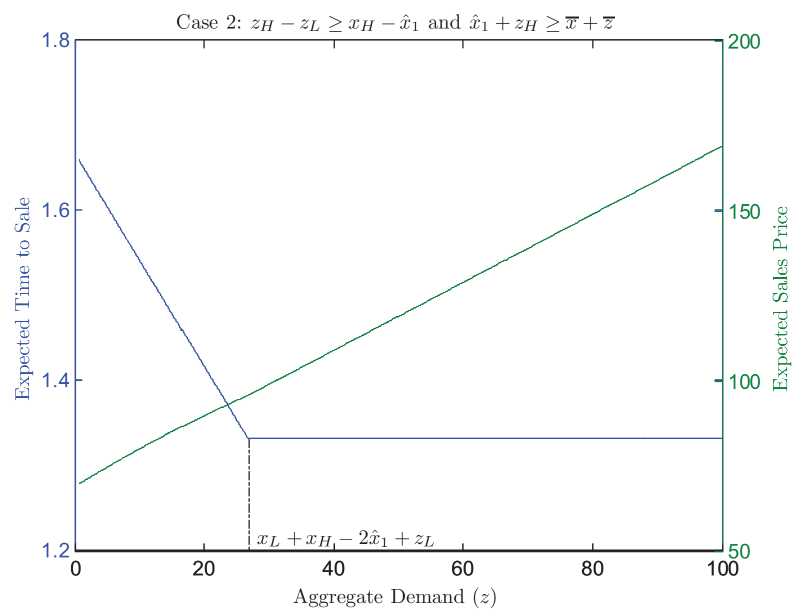


Figure 4.6: Realtor's Optimal Cutoff Rule in First Period of Three-Period Model

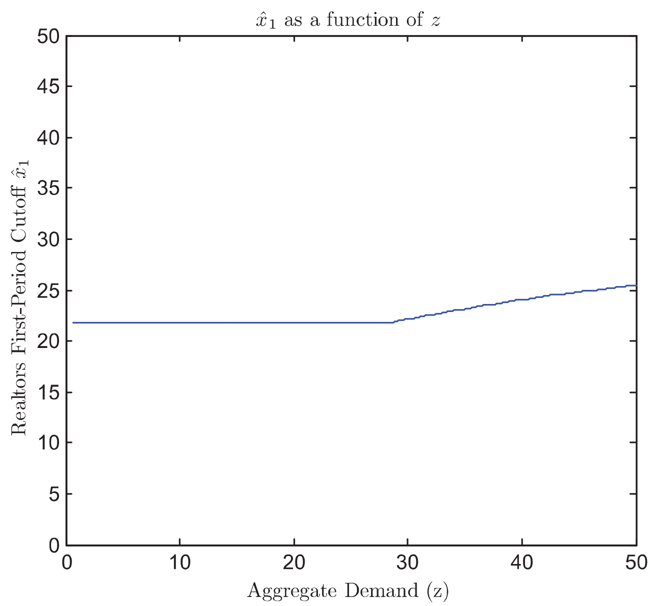


Figure 4.7: Seller's Expected Value of waiting for Period 2 in 3-period Model

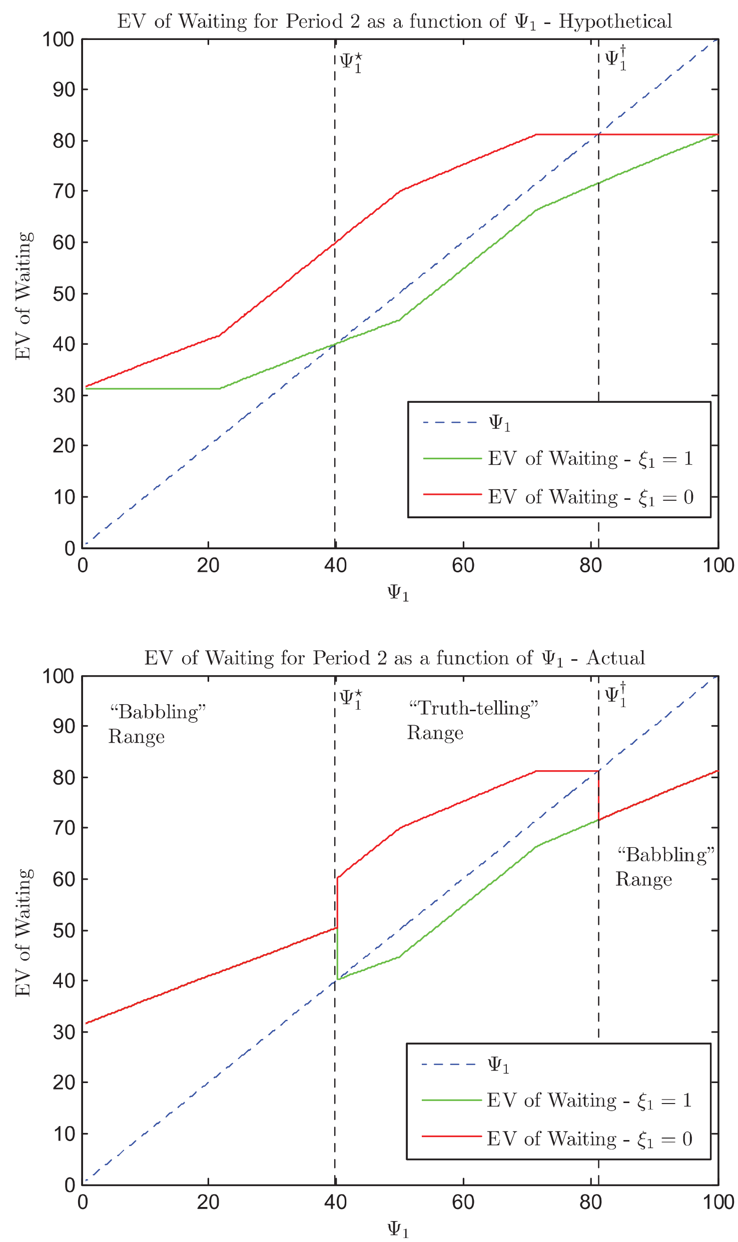


Figure 4.8: Expected Time to Sale and Sales Price in 3-period Model

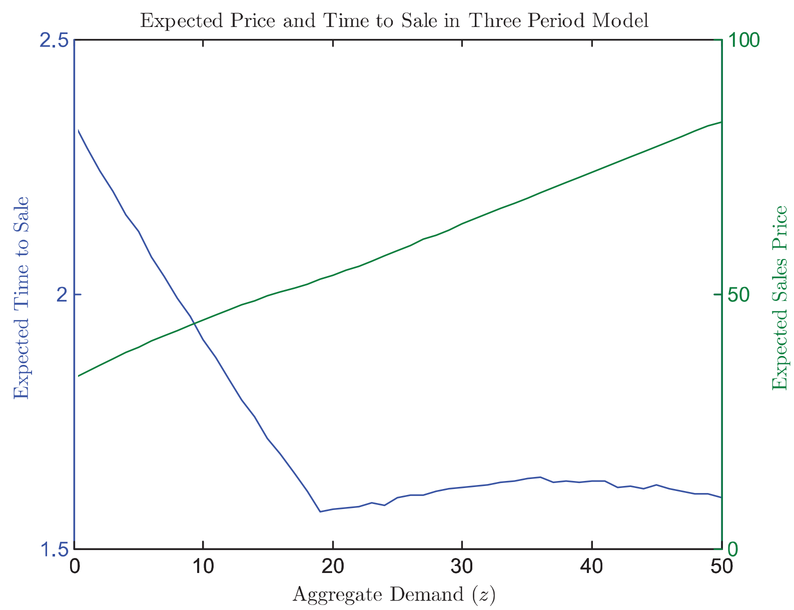


Figure 4.9: Percent Price Change from 2000: Perceived and Zillow (States)

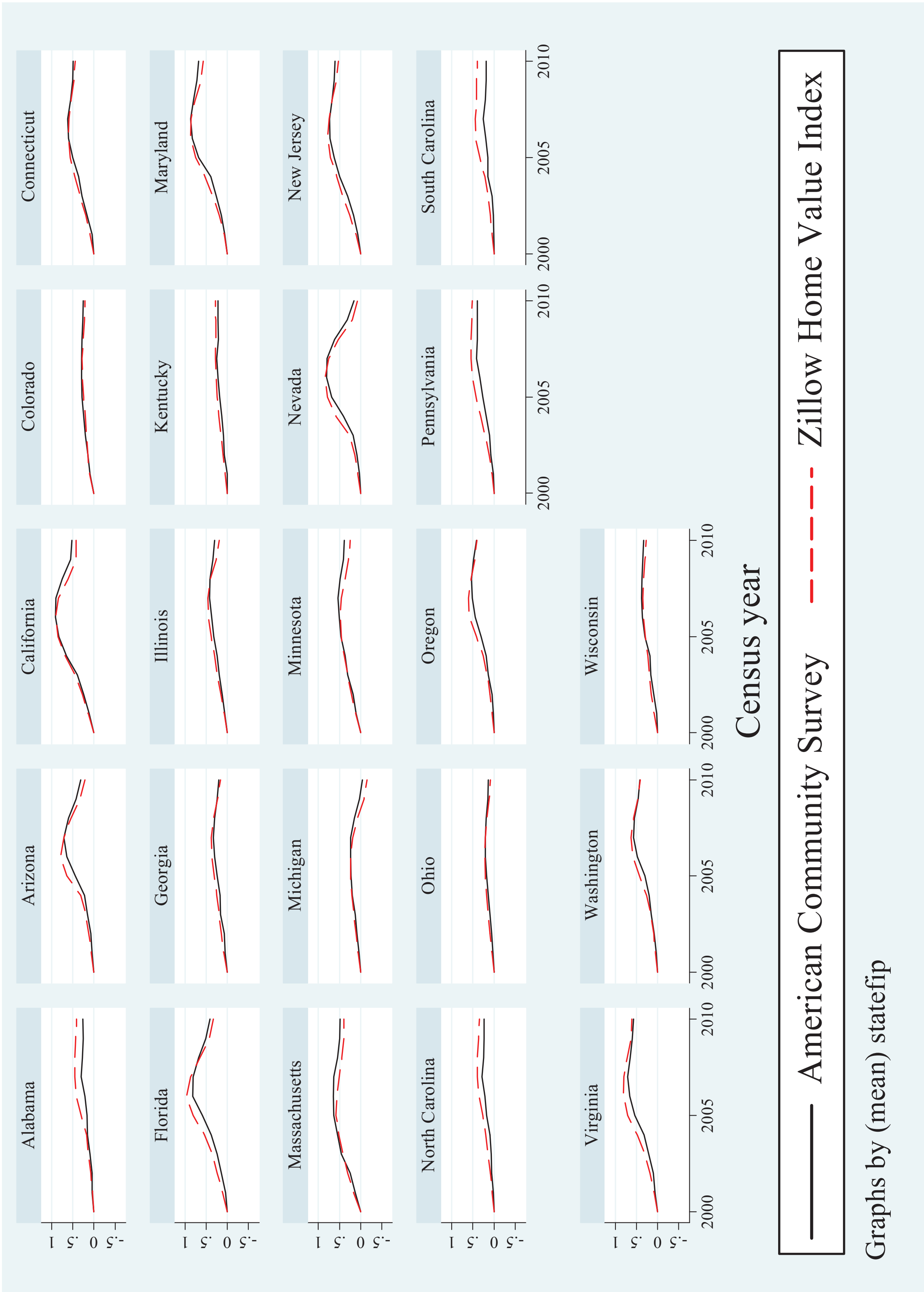


Figure 4.10: Misperceptions Index and Home Sales (States)

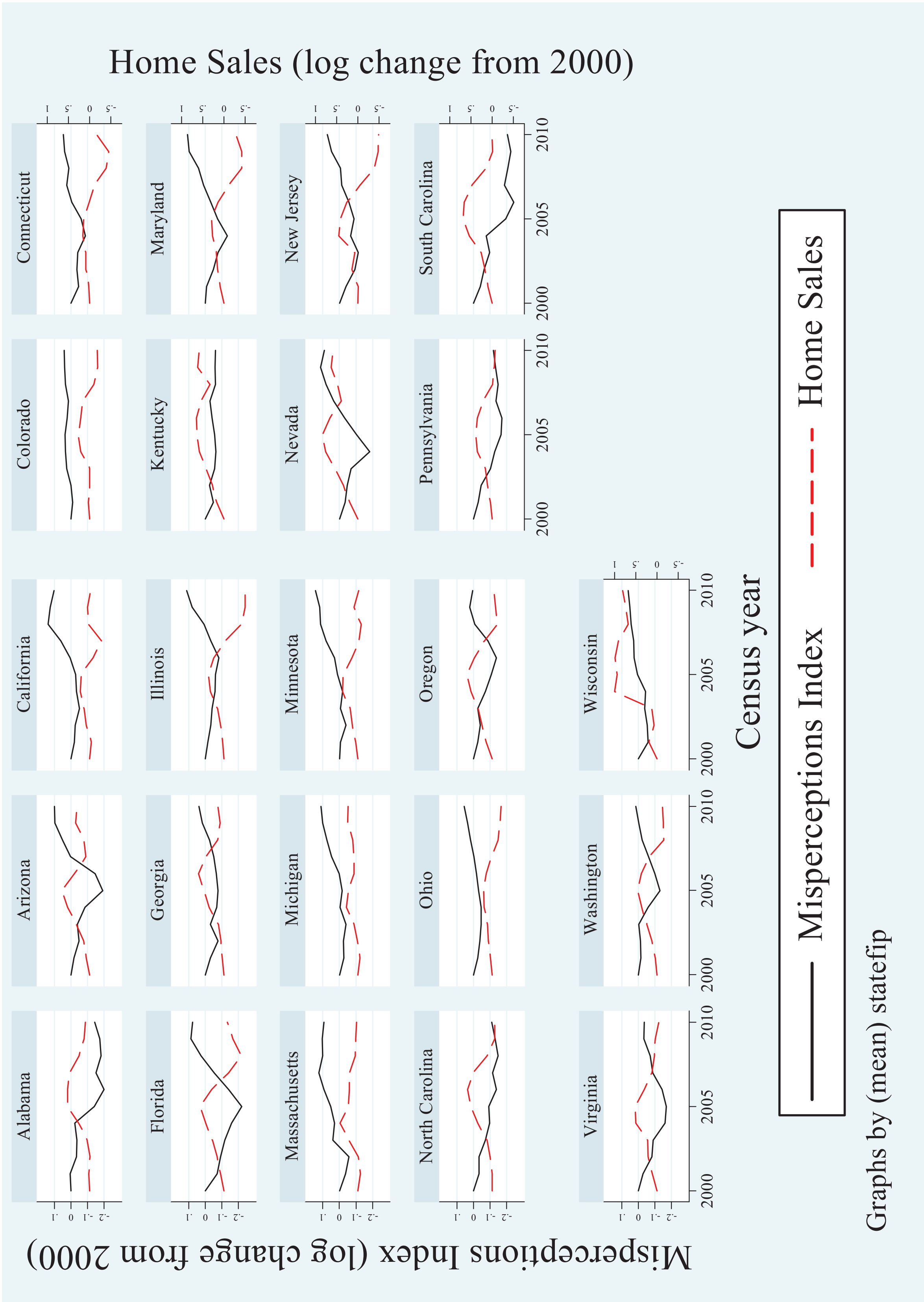


Figure 4.11: Percent Price Change from 2000: Perceived and Zillow (MSAs)

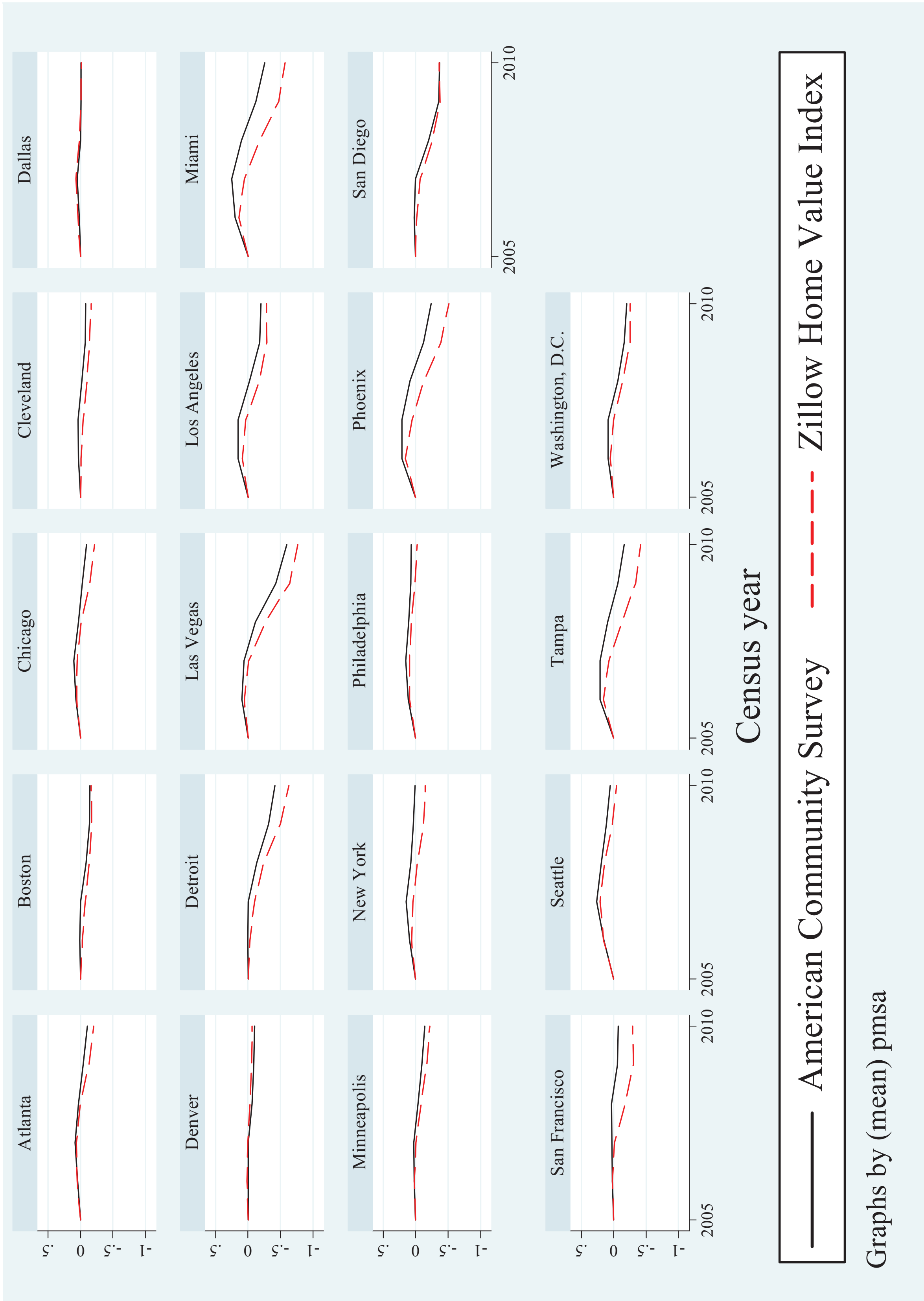
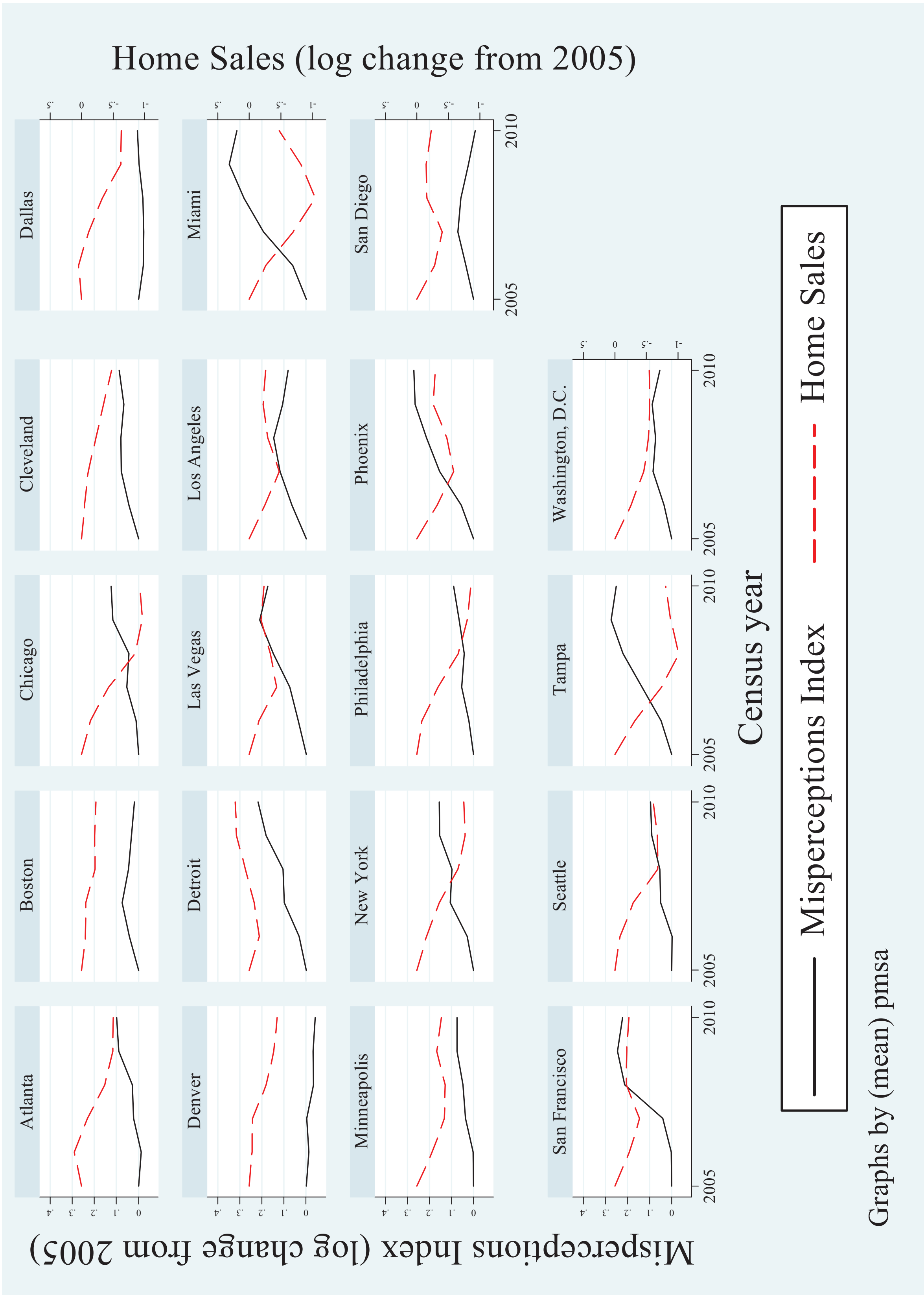


Figure 4.12: Misperceptions Index and Home Sales (MSAs)



BIBLIOGRAPHY

BIBLIOGRAPHY

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life Across Metropolitan Ares," 2008. NBER Working Paper No. 14472.

_ , "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values," 2009. NBER Working Paper No. 14981.

_ **and Gabriel Ehrlich**, "Metropolitan Land Values and Housing Productivity," 2012.

_ **, Walter Graf, Hendrik Wolff, and Ryan Kellogg**, "Extreme Temperature, Climate Change, and American Quality of Life," 2012.

Albrecht, James, Axel Anderson, Eric Smith, and Susan Vroman, "Opportunistic Matching in the Housing Market," *International Economic Review*, 2007, *48*, 641–664.

Aniliy, Shoshana, Jacob Hornik, and Miron Israeli, "Inferring the Distribution of Households' Duration of Residence from Data on Current Residence Time," *Journal of Business & Economic Statistics*, 1999, *17*, 373–381.

Aqeel, Aneeqa, "House Prices and Vacancies after Hurricane Katrina: Empirical Analysis of a Search and Matching Model," 2009.

Bajari, Patrick, Phoebe Chan, Dirk Krueger, and Daniel Miller, "A Dynamic Model of Housing Demand: Estimation and Policy Implications," 2010. NBER Working Paper 15955.

Baryla, Edward, Leonard Zumpano, and Harold Elder, "An Investigation of Buyer Search in the Residential Real Estate Market Under Different Market Conditions," *Journal of Real Estate Research*, 2000, *20*, 75–91.

Basu, Susanto, John Fernald, and Miles Kimball, "Are Technology Improvements Contractionary?," *American Economic Review*, 2006, *96*, 1418–1438.

Baum, C.F., M. E. Schaffer, and S. Stillman, "ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML, and k-class regression," 2007. Boston College Department of Economics, Statistical Software Components S425401. Downloadable from http://ideas.repec.org/c/boc/bocode/s425401.html.

Beaudry, Paul and Franck Portier, "An Exploration into Pigou's Theory of Cycles," *Journal of Monetary Economics*, 2004, *51*, 1183–1216.

Beeson, Patricia E. and Randall W. Eberts, "Identifying Productivity and Amenity Effects in Interurban Wage Differentials," *Review of Economics and Statistics*, 1989, *71*, 443–452.

Binswanger, Hans P., "The Measurement of Technical Change Biases with Many Factors of Production," *American Economic Review*, 1974, *64*, 964–976.

Case, Karl and Robert Shiller, "Is there a Bubble in the Housing Market?," *Brookings Papers on Economic Activity*, 2003, *2*, 299–362.

Case, Karl E., "The Value of Land in the United States: 1975 to 2005," in Gregory K. Ingram and Yu-Hung Hong, eds., *Proceedings of the 2006 Land Policy Conference: Land Policies and Their Outcomes*, Lincoln Institute of Land Policy Press Cambridge, MA 2007.

Cassimatis, Peter J., *Economics of the Construction Industry*, New York: The National Industrial Conference Board, 1969.

Cheshire, Paul and Stephen Sheppard, "The Welfare Economics of Land Use Planning," *Journal of Urban Economics*, 2002, *52*, 242–269.

Colwell, Peter and C.F. Sirmans, "A Comment on Zoning, Returns to Scale, and the Value of Undeveloped Land," *Review of Economics and Statistics*, 1993, *75*, 783–786.

_ **and Henry Munneke**, "The Structure of Urban Land Prices," *Journal of Urban Economics*, 1997, *41*, 321–336.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "The Costs of Agglomeration: Land Prices in French Cities," 2011.

Courant, Paul, "On the Effect of Fiscal Zoning on Land and Housing Values," *Journal of Urban Economics*, 1976, *3*, 88–94.

Davis, Morris and Jonathan Heathcote, "Housing and the Business Cycle," *International Economic Review*, 2005, *46*, 751–784.

_ **and** _ , "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*, 2007, *54*, 2595–2620.

_ **and Michael Palumbo**, "The Price of Residential Land in Large U.S. Cities," *Journal of Urban Economics*, 2007, *63*, 352–384.

Dejong, David and Chetan Dave, *Structural Macroeconometrics*, Princeton, NJ: Princeton University Press, 2007.

Diaz, Antonia and Belen Jerez, "House Prices, Sales and Time on the Market: a Search-Theoretic Framework," 2009. Working Paper 09-25, Economic Series (15). Universidad Carlos III de Madrid.

Epple, Dennis, Brett Gordon, and Holger Sieg, "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, 2010, *100*, 905–924.

Fisher, Jeff, David Geltner, and Henry Pollakowski, "A Quarterly Transactions-based Index of Institutional Real Estate Investment Performance and Movements in Supply and Demand," *Journal of Real Estate Finance and Economics*, 2007, *34*, 5–33.

Fuss, Melvyn and Daniel McFadden, eds, *Production Economics: A Dual Approach to Theory and Applications.*, New York: North Holland, 1978.

Genesove, David and Christopher Mayer, "Equity and Time to Sale in the Real Estate Market," *American Economic Review*, 1997, *87*, 255–269.

_ **and** _ , "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 2001, *116*, 1233–1260.

_ **and Lu Han**, "Search and Matching in the Housing Market," 2010.

Glaeser, Edward L. and Bryce A Ward, "The causes and consequences of land use regulation: Evidence from Greater Boston," *Journal of Urban Economics*, 2009, *65*, 265–278.

Glaeser, Edward L and Joseph Gyourko, "The Impact of Building Restrictions on Housing Affordability," *Federal Reserve Bank of New York Economic Policy Review*, 2003, *9*, 21–29.

Glaeser, Edward L. and Joseph Gyourko, "Urban Decline and Durable Housing," *Journal of Political Economy*, 2005, *113*, 345–375.

_ **and Joshua Gottlieb**, "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, pp. 155–239.

_ **and Kristina Tobio**, "The Rise of the Sunbelt," *Southern Economic Journal*, 2008, *74*, 610–643.

_ **, Joseph Gyourko, and Albert Saiz**, "Housing Supply and Housing Bubbles," *Journal of Urban Economics*, 2008, *64*, 198–217.

_ **,** _ **, and Raven Saks**, "Urban Growth and Housing Supply," *Journal of Economic Geography*, 2005, *6*, 71–89.

_ **,** _ **, and** _ , "Why is Manhattan so Expensive? Regulation and the Rise in Housing Prices," *Journal of Law and Economics*, 2005, *48*, 331–369.

Griliches, Zvi and Vidar Ringstad, *Economies of Scale and the Form of the Production Function: an Econometric Study of Norwegian Manufacturing Establishment Data*, Amsterdam: North Holland Publishing Company, 1971.

Gyourko, Joseph, Albert Saiz, and Anita Summers, "New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 2008, *45*, 693–729.

_ **, Christopher Mayer, and Todd Sinai**, "Superstar Cities," 2006. NBER Working Paper No. 12355.

Haurin, Donald, "The Duration of Marketing Time of Residential Housing," *AREUEA Journal*, 1988, *16*, 396–410.

Hausman, Jerry, "Specification Tests in Econometrics," *Econometrica*, 1978, *46*, 1251–1271.

Head, Allen, Huw Lloyd-Ellis, and Amy Sun, "Search and the dynamics of house prices and construction," 2010.

Hosios, Arthur, "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 1990, *57*, 279–298.

Iacoviello, Matteo and Stefano Neri, "Housing Market Spillovers from an Estimated DSGE Model," *American Economic Journal: Macroeconomics*, 2010, *2*, 125–164.

Ihlanfeldt, Keith R., "The Effect of Land Use Regulation on Housing and Land Prices," *Journal of Urban Economics*, 1984, *16*, 1–12.

Kahn, James, "What Drives Housing Prices?," 2008. Federal Reserve Bank of New York Staff Report no. 345.

Kane, Thomas and Douglas Staiger, "Estimating Teacher Impacts on Student Achievement: an Experimental Evaluation," 2008. NBER Working Paper No. 14607.

Katz, Lawrence and Kenneth T. Rosen, "The Interjurisdictional Effects of Growth Controls on Housing Prices," *Journal of Law and Economics*, 1987, *30*, 149–160.

Kok, Nils, Paavo Monkkonen, and John Quigley, "Economic Geography, Jobs, and Regulations: The Value of Land and Housing," 2010. Working Paper No. W10-005. University of California.

Krainer, John, "A Theory of Liquidity in Residential Real Estate Markets," *Journal of Urban Economics*, 2001, *49*, 32–53.

Lambertini, Luisa, Caterina Mendicino, and Maria-Teresa Punzi, "Expectations-Driven Cycles in the Housing Market," 2010. Working Paper No. 1021. Bank of Spain.

Lazear, Edward, "Why Should Inventories Rise when Demand Falls in Housing and Other Markets?," 2010. NBER Working Paper No. 15878.

Leung, Charla Ka Yui and Jun Zhang, "Housing Markets with Competitive Search," 2007.

Levitt, Steven and Chad Syverson, "Market Distortions when Agents are Better Informed: A Theoretical and Empirical Exploration of the Value of Information in Real Estate Transactions," 2002. University of Chicago Working Paper.

— **and** — , "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions," *Review of Economics and Statistics*, 2008, *90*, 599–611.

Magnus, Gideon, "The dynamics of prices, liquidity, and vacancies in the housing market," 2010.

Malpezzi, Stephen, Gregory H. Chun, and Richard K. Green, "New Place-to-Place Housing Price Indexes for U.S. Metropolitan Areas, and Their Determinants," *Real Estate Economics*, 1998, *26*, 235–274.

Mayer, Christopher J. and C. Tsuriel Somerville, "Land Use Regulation and New Construction," *Regional Science and Urban Economics*, 2000, *30*, 639–662.

McDonald, J.F., "Capital-Land Substitution in Urban Housing: A Survey of Empirical Estimates," *Journal of Urban Economics*, 1981, *9*, 190–211.

Merlo, Antonio and Francois Ortalo-Magne, "Bargaining over residential real estate: evidence from England," *Journal of Urban Economics*, 2004, *56*, 192–216.

Merz, Monika, "Search in the labor market and the real business cycle," *Journal of Monetary Economics*, 1995, *36*, 269–300.

Miller, Norman and Michael Sklarz, "A Note on Leading Indicators of Housing Price Trends," *Journal of Real Estate Research*, 1986, *1*, 99–109.

Moen, Espen, "Competitive Search Equilibrium," *Journal of Political Economy*, 1997, *105*, 385–411.

Munneke, Henry and Barrett Slade, "An Empirical Study of Sample-Selection Bias in Indices of Commercial Real Estate," *Journal of Real Estate Finance and Economics*, 2000, *21*, 45–64.

— **and** — , "A Metropolitan Transaction-Based Commercial Price Index: A Time-Varying Parameter Approach," *Real Estate Economics*, 2001, *29*, 55–84.

Muth, Richard, *The Demand for Durable Goods*, Chicago: University of Chicago Press, 1960.

Nichols, Joseph, Stephen Oliner, and Michael Mulhall, "Commercial and Residential Land Prices Across the United States," 2010.

Ohls, James C, Richard Chadbourn Weisberg, and Michelle J. White, "The Effect of Zoning on Land Value," *Journal of Urban Economics*, 1974, *1*, 428–444.

Olsen, Edgar, "A Competitive Theory of the Housing Market," *American Economic Review*, 1969, *59*, 612–622.

Ortalo-Magne, Francois and Sven Rady, "Housing Market Fluctuations in a Life-Cycle Economy with Credit Constraints," 1998. Stanford University Working Paper No. 1501.

Ozimek, Adam and Daniel Miles, "Stata utilities for geocoding and generating travel time and travel distance information," *The Stata Journal*, 2011, *11*, 106–119.

Peters, Michael, "Ex Ante Price Offers in Matching Games Non-Steady States," *Econometrica*, 1991, *59*, 1425–1454.

Piazzesi, Moinka and Martin Schneider, "Momentum Traders in the Housing Market: survey evidence and a search model," *American Economic Review*, 2009, *99*, 406–411.

Quigley, John and Larry Rosenthal, "The Effects of Land Use Regulation on the Price of Housing: What Do We Know? What Can We Learn?," *Cityscape: A Journal of Policy Development and Research*, 2005, *8*, 69–137.

_ **and Stephen Raphael**, "Regulation and the High Cost of Housing in California," *American Economic Review*, 2005, *95*, 323–329.

Rappaport, Jordan, "Consumption amenities and city population density," *Regional Science and Urban Economics*, 2008, *38*, 533–552.

_ , "A Productivity Model of City Crowdedness," *Journal of Urban Economics*, 2008, *65*, 715–722.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90*, 1257–1278.

Rose, Louis A., "Land Values and Housing Rents in Urban Japan," *Journal of Urban Economics*, 1992, *31*, 230–251.

Rosenthal, Stuart S. and William C. Strange, *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: North Holland,

RSMeans, *Building Construction Cost Data 2010*, Kingston, MA: Reed Construction Data, 2009.

Rupert, Peter, Richard Rogerson, and Randall Wright, "Estimating substitution elasticities in household production models," *Economic Theory*, 1995, *6*, 179–193.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 2010, *125*, 1253–1296.

Shapiro, Jesse M., "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital," *Review of Economics and Statistics*, 2006, *88*, 324–335.

Stein, Jeremy, "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects," *American Economic Review*, 1995, *110*, 379–406.

Thorsnes, Paul, "Consistent Estimates of the Elasticity of Substitution between Land and Non-Land Inputs in the Production of Housing," *Journal of Urban Economics*, 1997, *42*, 98–108.

Topel, Robert and Sherwin Rosen, "Housing Investment in the United States," *Journal of Political Economy*, 1988, *96*, 718–740.

Wheaton, William, "Vacancy, Search, and Prices in a Housing Market Matching Model," *Journal of Political Economy*, 1990, *98*, 1270–1292.

_ **and Gleb Nechayev**, "The 1998-2005 Housing 'Bubble' and the Current 'Correction': What's Different this Time?," *Journal of Real Estate Research*, 2008, *30*, 1–26.