Quality Regulation in Intermediation Markets: Evidence from the Real Estate Brokerage Industry*

Santiago Varela Seoane†
Stanford University

November 14, 2026

Updated frequently. Please click **here** for the latest version.

Abstract

In intermediation markets, the quality of agent services is a key determinant of consumer welfare. However, intermediation markets are subject to quality underprovision due to two factors. First, search frictions limit consumers' ability to observe agent quality. Second, agents' incentives to invest in quality are attenuated when quality differences cannot be observed. Understanding the relative contribution of each factor can inform the design of remediation policies. In this paper, I study how heterogeneity in agent quality and consumer demand for quality shape welfare in intermediation markets. I answer this question by developing a model of buyer agent intermediation with heterogeneity in agent services, which I estimate using data on the US real estate brokerage market. In the model, agent recommendations maximize a weighted sum of buyer utility and agent earnings from commissions. Agents differ in the weights assigned to each, and in their ability to obtain lower prices for a given product. Using property transactions data from the Phoenix metro area, I find that, despite notable dispersion across agents in both measures of heterogeneity, demand for agent services is almost insensitive to these differences. Comparative statics exercises show that increasing demand sensitivity can sort consumers to high-quality agents, providing benefits in terms of housing consumption. Direct improvements to agent quality of service can achieve comparable effects if they target a sufficiently high fraction of the agent market.

*I am especially grateful to my advisors Liran Einav, Hunt Allcott, and José Ignacio Cuesta for their guidance, encouragement, and support. I also thank Nick Bloom, Greg Buchak, Sebastián Bauer, Lautaro Chittaro, Evan Flack, Matthew Gentzkow, Allan Hsiao, Ravi Jagadeesan, Neale Mahoney, Cecilia Moreira, Elena Pastorino, Steve Redding, Greg Rosston, Janet Stefanov, Amit Seru, Isaac Sorkin, Tess Snyder, Sam Thau, Shoshana Vasserman, Alessandra Voena, Ali Yurukoglu, and seminar participants at Stanford University for their helpful comments and suggestions. I gratefully acknowledge financial support from the Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship through a grant to the Stanford Institute for Economic Policy Research. This research has been approved by the Stanford IRB (protocol #89750). Some of the computing for this project was performed on the Sherlock cluster. I thank Stanford University and Stanford Research Computing for providing computational resources and support that contributed to these research results. All errors are my own.

†Email: svseoane@stanford.edu.

1 Introduction

Intermediaries, or agents, are prominent in markets with search frictions. Agents help buyers and sellers complete transactions by providing expert advice and reducing search costs. Because of their central role, the quality of service provided by agents has important implications for consumer welfare.

However, intermediation markets are subject to quality underprovision due to two reasons. First, the same frictions that give rise to intermediation also limit consumers' ability to observe agent quality (Egan *et al.*, 2024). Second, agents' incentives to invest in quality may be diminished if quality improvements cannot be observed. Understanding the relative contribution of supply- and demand-side determinants of quality provision can inform the design of remediation policies.

This paper studies how heterogeneity in agent quality, and consumer demand for quality, shape welfare in intermediation markets. I answer this question by developing an empirical model of intermediation with heterogeneous agent quality, which I estimate using data on the United States real estate brokerage industry. The model allows me to define the two primitives that determine the equilibrium allocation of quality. The first primitive is the population distribution of the quality of the advice provided by agents. The second primitive is the sensitivity of consumer demand to differences in quality of advice. I quantify these primitives using data on property transactions from the Phoenix metro area. Using the estimated results, I perform comparative statics exercises to assess the effects on consumer welfare of changes to the primitives of the model.

I start my analysis by providing a conceptual framework, which introduces the model that guides my work. I extend a canonical empirical model of intermediation (Robles-García, 2020; Kim, 2025) to allow for heterogeneity on how agents affect recommendations and transaction outcomes. The model has two stages. In the first stage, consumers choose an agent to assist them in purchasing a product. In the second stage, the consumer-agent pair chooses a product to maximize a weighted sum of their utilities. Consumer utility is a function of product characteristics and price. Agent utility is a function of product-specific commission income received upon completion of a transaction.

In the model, agents are heterogeneous with respect to two dimensions. First, agents can differ in the weights they assign to consumer utility and own earnings when making their recommendations. Agents with a higher weight on buyer utility are less likely to steer consumers to high-commission products and more likely to recommend products more aligned with buyer preferences. Second, conditional on a match with a given product, agents differ in their ability

to obtain lower transaction prices. These dimensions are motivated by my empirical setting. According to home buyer surveys, recommendations about what property to purchase and assistance in price negotiations are the two elements of advice predominantly valued by consumers (NAR, 2025).

The combination of the two heterogeneity dimensions defines the first economic primitive of my analysis. The model delivers a closed-form expression for the inclusive value of agent advice, which collapses the impact of both heterogeneity dimensions in a single expression. The empirical distribution of this expression allows me to evaluate how much consumers stand to gain from working with different agents. The second primitive of my analysis, consumer sensitivity to quality differences, comes from the agent choice model. This model allows me to quantify the relationship between the value of advice and consumers' choices of agents, holding other agent characteristics fixed.

I quantify these primitives for the US real estate brokerage industry. Agent intermediation in the real estate industry is significant. More than 4.5 million homes were transacted in the United States in 2025,[1] and about 90% of home buyers and sellers use agent services (Tracey, 2025). Home purchases are typically one of the largest financial transactions, if not the largest, that households perform in their lifetimes. Because of this, low quality of advice can be particularly costly in this setting. Concerns about low quality of service in the real estate industry have been expressed by industry stakeholders and observers (Swanepoel, 2015; Brobeck, 2023; White, 2026).

I estimate the pricing ability dimension of agent heterogeneity using a value added model. Value added models have been used to estimate individual measures of performance in multiple contexts (e.g. Ahn *et al.*, 2025; Abaluck *et al.*, 2021; Einav *et al.*, 2025; Limodio, 2021; Fenizia, 2022; Muñoz and Prem, 2024; Cunningham *et al.*, 2022, among others). I leverage information on more than 15 years of property transactions in the Phoenix metro area to estimate this model. This allows me to account for multiple sources of selection concerns and property unobservables. My results indicate that, conditional on initial list prices and several determinants of property values, the standard deviation of agent effects on final sale prices is 0.8 and 0.9 percentage points of list prices for seller and buyer agents, respectively. This dispersion is about three times as large as the standard deviation of commission rates in Phoenix in the same period. Higher-quality seller agents are associated with a price premium and higher-quality buyer agents are associated with a discount on final prices.

[1] The National Association of Realtors estimated 4.06 million sales of existing homes in 2025 (Olick, 2026). In addition, according to the U.S. Census Bureau and HUD's Survey of Construction, 678 thousand new homes were sold in 2025 (U.S. Census Bureau and U.S. Department of Housing and Urban Development, 2026).

I estimate agent heterogeneity in incentive alignment weights using a discrete choice model of agent-intermediated property transactions. Differences in incentive alignment weights across agents are identified from variation in property characteristics and commission income of the properties they help purchase. I account for buyer-specific heterogeneity in preferences for property characteristics to recover how agent recommendations vary when representing different consumer segments.

The main identification concern to estimate the property choice model is the endogeneity of prices and commission rates. Price endogeneity reflects the standard concern that prices can be correlated with unobserved product quality. During the period of my sample, commission rates for buyer agents, posted by sellers as a listing attribute, is another potential source of endogeneity.[2] In my setting, commission rates and prices are co-determined, with sellers earning net-of-commissions sale prices. Because commission rates affect sellers' proceeds from a sale, they are subject to similar endogeneity concerns as prices. For instance, sellers that are more concerned about their final earnings might both invest in unobserved attributes that increase sale prices and be more willing to offer low commission rates to buyer agents.

To instrument for prices, I construct a shift-share instrument that exploits variation in national growth rates of demographic groups with different tastes for property attributes, following Cook *et al.* (2025). My instrument for commission rates leverages a number of institutional features. Commission earnings from a transaction are typically split between the agent and their brokerage company according to pre-specified rules. These rules vary across brokerage companies, and the distribution of commission rates exhibits variation across cities (Banerjee and Paciorek, 2025). I exploit these two institutional features to construct a Hausman-type instrument for commission rates.

My estimates suggest considerable dispersion in incentive alignment weights. Property choices can be rationalized as if agents place relative weights on their own earnings ranging from 3.9% to 22.6%. Holding property characteristics and prices fixed, agents with the lowest weight on buyer utility are seven times less likely to recommend properties with low commission rates than their high-weight counterparts.

By combining the two heterogeneity dimensions and my estimated model of property choice, I construct the first primitive of my model. The distribution of the value of agent advice allows me to assess how much consumers stand to gain from the interaction of pricing ability and incentive alignment weights. The combination of the two heterogeneity dimensions implies

[2]The 2024 NAR settlement, by which the National Association of Realtors agreed to remove offers of compensation for buyer agents in MLS listings (Han, 2024), is outside my sample. Thus, I model the industry following pre-settlement rules.

that, for the median home buyer, the standard deviation of the value of agent advice equals 1.6% of average sale prices, or five times the standard deviation of buyer agent commission rates. This dispersion is particularly salient given that differences in advice quality are largely unpriced. During my sample period, more than 95 percent of commission rate offers are either 3 percent or 2.5 percent of sale prices.

After estimating the aggregate distribution of agent quality, I turn to estimating consumer sensitivity to quality differences. I estimate consumer preferences for agent characteristics using a discrete choice model of agent demand. Consumers of agent services choose among agents based on the value of agent advice and observable agent characteristics. I find that consumers are largely insensitive to quality differences. For the median home buyer, a one standard deviation increase in the value of agent advice increases choice probabilities by 2.8%. This result is consistent with prior estimates and qualitative evidence of limited search for agent services.

Using the model estimates, I perform comparative statics exercises to shed light on the incidence of agent quality differences and demand for quality on consumer welfare from housing consumption. I evaluate the relative impact of consumer demand sensitivity by applying a multiplier to the estimated demand parameter on agent advice. The range of multipliers I consider is consistent with empirical evidence on the impact of informational interventions in the financial advising industry (Zhao, 2023; Egan *et al.*, 2026a) and other contexts (Hastings *et al.*, 2007; Kling *et al.*, 2012). I find that doubling consumer sensitivity to differences in agent advice increases average consumer surplus from housing consumption by the equivalent of 0.91% of average sale prices.

I compare the demand sensitivity results against the effects of targeting the supply of advice, by increasing incentive alignment weights for low-alignment agents. I evaluate these effects with alternative scenarios, in which agents below a given threshold have their alignment weights raised to the threshold. My estimates suggest that, to achieve comparable effects as those from the demand side, improvements in the value of advice need to be large and target a large fraction of the agent market. Concretely, I find that doubling demand sensitivity achieves similar effects to increasing incentive alignment weights for agents below the median alignment weight distribution.

Consumer sorting patterns inform the interpretation of the comparative statics results. The estimated agent demand parameters suggest that agent choices are relatively insensitive to quality differences. When demand sensitivity increases, consumers reallocate their demand from low- to high-quality agents. This benefits consumers through better recommendations and lower transaction prices on average. In contrast, direct improvements to the quality of

advice primarily benefit consumers that were previously using the targeted agents. These results underscore the role of consumer demand responses in shaping the effectiveness of quality regulation interventions.

1.1 Related Literature

This paper contributes to four bodies of literature. First, a recent stream of papers in empirical IO has studied intermediation industries (Robles-García, 2020; Salz, 2022; Boehm, 2025; Grennan *et al.*, 2025). This literature has developed empirical tools to study how intermediaries affect consumer choices, and to quantify the welfare costs borne by consumers when incentives between agents and consumers are misaligned.

Two papers in this literature, in particular, are closely related to my research. First, Kim (2025) studies the welfare effects of decoupling commission rates between buyer and seller agents in the real estate industry using an empirical model of intermediated transactions. I extend his model of buyer agent services to allow for and estimate heterogeneity in quality attributes across agents. This allows me to assess how changes to the distribution of agent quality or to consumer sensitivity shape consumer welfare. This discussion is relevant given the history of policy discussion about quality regulation in the real estate brokerage industry (Swanepoel, 2015; Barwick and Wong, 2019; Brobeck, 2023; White, 2026). Second, Bhattacharya *et al.* (2025) study the consequences of fiduciary duty in the market for financial advisers. I provide insights about the consequences of low quality of advice in the real estate industry, and highlight the relative roles of quality provision and demand for advice.

Second, I contribute to a large body of research analyzing the role of agents in the real estate industry (Hsieh and Moretti, 2003; Levitt and Syverson, 2008; Hendel *et al.*, 2009; Bernheim and Meer, 2013; Barwick and Pathak, 2015; Barwick *et al.*, 2017; Buchak *et al.*, 2024, 2025; Barry *et al.*, 2024; Kim, 2025; Kenney, 2026). Recent research has analyzed how agent heterogeneity shapes consumer outcomes. Gilbukh and Goldsmith-Pinkham (2024) and Gilbukh (2026) analyze how performance of seller agents depends on experience. Cunningham *et al.* (2022) estimate agent effects on sale prices and time on market using value added models. I build upon this literature by studying individual agent effects on close/list price ratios using recent methods developed in the teacher value added literature (Ahn *et al.*, 2025). Moreover, I integrate these estimates with a model of property choice to study the implications of differences in pricing ability on consumer welfare. Additionally, I further estimate heterogeneity in the relative weights placed by agents on private commission earnings compared to buyer preferences. Finally, I speak to the interaction between heterogeneity in agent performance and consumer demand for agent quality.

Third, this paper connects to research in finance documenting low quality of advice, particularly misconduct, in the financial advising industry, and the effectiveness of policy interventions in mitigating misconduct rates (Egan, 2019; Egan *et al.*, 2019, 2022, 2024, 2026a,b; Kowaleski *et al.*, 2020; Zhao, 2023; Reuter and Schoar, 2024; Honigsberg *et al.*, 2025). I extend insights of this literature to the market for real estate agent services, where, as with financial advisers, intermediaries advise households on high-stakes transactions and quality is difficult to observe. My comparative statics are related in spirit to the analyses of policy interventions in this literature, although I do not interpret my exercises as corresponding to concrete policies.

Finally, I contribute to a body of research that investigates the effects of quality regulation in markets with informational asymmetries (Dranove and Jin, 2010; Kleiner and Soltas, 2023; Farronato *et al.*, 2024; Atal *et al.*, 2025a,b; Vatter, 2025; Kepler *et al.*, 2026). I analyze the role of heterogeneity in and demand for quality in the real estate brokerage industry through a unified framework, which allows me to assess the relative contribution of each in shaping consumer welfare.

2 Model

In this section, I introduce a model of a market for intermediation services to shed light on the economic primitives that motivate my work. The goal of this section is to define the key objects that I will recover in my estimation strategy. To this end, I abstract from some flexibility that will later show up in the empirical sections.

The model has two stages and follows standard formulations in the intermediation literature (Robles-García, 2020; Kim, 2025). In stage 1, consumers of intermediation services choose among prospective agents given observable agent characteristics and the expected value of advice. In stage 2, consumers choose a product with the assistance of agents that differ in their value of advice. I start by defining the main elements of the model, before describing each stage. I describe how the inclusive value of the product choice model defines a measure of the expected value of agent advice, which is an input for the agent choice model.

Preliminaries. Let i denote consumers and a denote agents. Consumers choose among products j to maximize consumption utility u_i. Consumption utility depends on the price of the product p_j and on product characteristics, defined by the vector X_j. I define consumer utility over products j as

$$u_{ij} = -\alpha p_j + \beta X_j + \epsilon_{ij} \tag{1}$$

where ϵ_{ij} is a Type 1 extreme value shock.

Agents obtain earnings π_j from assisting consumers in purchasing product j. Motivated by the characteristics of my setting, I assume that agent earnings are product-specific but common across all agents conditional on a product. I postpone a discussion of how property-specific earnings are determined in the real estate industry to section 3.3.

Stage 2: Product choice. In the second stage of the model, consumers, with the assistance of agents, choose among the set of alternative products in the market. As is standard in the intermediation literature, I model this choice as a weighted maximization problem of consumer utility and agent earnings (Robles-García, 2020; Kim, 2025). Differences in relative weights placed on consumer utility, relative to agent earnings, determine agents' steering incentives to high-commission products. Agents can further affect transaction prices conditional on a match.

The buyer-agent pair chooses among products j to maximize

$$\hat{V}_{iaj} = (1 - \theta_a)(-\alpha p(p^l, \phi_a) + \beta X + \epsilon_{ij}) + \theta_a(\pi_j) \tag{2}$$

where θ_a indicates the weight placed on agent earnings. When $\theta_a = 0$, choices are insensitive to agent earnings. As θ_a increases, agent earnings have a greater relative weight in the joint decision. I allow for heterogeneity in these weights across agents by having θ_a depend on agent identity a.

The second source of agent heterogeneity in the model is agents' impact on prices, ϕ_a. Agents can influence the final sale of the product, p_j, which depends on the initial list price p_j^l and agents' pricing quality term ϕ_a, according to the function $p_j = p(p_j^l, \phi_a)$, where p is a continuous and differentiable function such that $p_{p^l} > 0$, $p_\phi < 0$. Conditional on a match with a product j, agents with higher pricing quality provide value to consumers by delivering lower final prices p_j.

Inclusive value of agent advice. The Type 1 Extreme Value assumption on ϵ_{ij} allows to derive a monetary metric for the value of the advice that the agent provides to the consumer. First, dividing equation (2) by $1 - \theta_a$ yields

$$\frac{\hat{V}_{iaj}}{1 - \theta_a} = -\alpha p(p^l, \phi_a) + \beta X + \tilde{\theta}_a(\pi_j) + \epsilon_{ij} \tag{3}$$

with $\tilde{\theta}_a = \frac{\theta_a}{1-\theta_a}$. The distributional assumption on ϵ_{ij} implies that the probability that the buyer

i-agent a pair chooses product j is

$$s_{iaj} = \frac{\exp(V_{iaj})}{\sum_k \exp(V_{iak})}.$$

with $V_{iaj} = (1-\theta_a)(-\alpha p(p^l, \phi_a) + \beta X) + \theta_a(\pi_j)$. Following standard results in the literature of intermediation (Dubois *et al.*, 2018; Grennan *et al.*, 2025; Kim, 2025), the inclusive value of agent advice can be defined as[3]

$$I_{ia} = \frac{1}{\alpha} \ln \left(\sum_k \exp(V_{iak}) \right) - \frac{1}{\alpha} \sum_k s_{iak} \theta_a(\pi_k) \tag{4}$$

The first block of equation (4) is the standard Small and Rosen (1981) consumer surplus formula. The second block of equation (4) adjusts for distortions in product choice driven by the agent's private incentives that arise from earnings π_{aj} associated with the choice of product j.

Stage 1: Agent choice. In the first stage, consumers of intermediation services choose among prospective agents a as a function of the advice value I_{ia} and agent characteristics x_a^A, according to the function

$$V_{ia}^A = \lambda I_{ia} + \beta^A x_a^A + \nu_{ia} \tag{5}$$

where ν_{ia} is a Type 1 Extreme Value shock. The coefficient λ measures the sensitivity of consumer demand for the value of advice.

3 Characteristics of real estate brokerage industry

In this section I outline the characteristics of the real estate brokerage industry that guide my research question and empirical analysis. I start by describing the role played by agents in real estate transactions, and the aspects of agent advice that consumers value most. This motivates the dimensions of agent heterogeneity I consider. I next introduce the institutional features that are most relevant for my question. The large majority of housing transactions occur through a centralized platform, from which I can construct quality measures for the entire agent market. The rules governing transactions in this platform inform my property choice model.

[3] A derivation of this result is provided in Appendix A.1.

3.1 Role of real estate agents

Real estate agents may assist their customers across multiple dimensions of the home purchase process, depending on the transaction side they represent. Assistance in particular transaction outcomes is especially valued by consumers. Figure A1 shows the results of a survey by the National Association of Realtors (NAR) in 2025 (NAR, 2025). Results show that buyers especially value being matched to the best property for them and assistance in pricing. This motivates matching and pricing as the two most important determinants of agent quality. As introduced in the general model in section 2, the dimensions I study for the real estate brokerage market speak to these determinants.

The majority of home buyers and sellers in the United States seek advice from real estate agents. According to a 2025 survey by the NAR, 91% of sellers and 88% of buyers completed transactions with the assistance of agents (Tracey, 2025). This contrasts with other countries, where only sellers are typically represented by agents. For example, estimates suggest that 20% of home buyers in Netherlands and less than 5% in the United Kingdom use agents (Dagher, 2023). Most agents take part in both roles, with only a fraction specializing on either the seller side or the buyer side.[4]

3.2 Institutional setting

Real estate transactions within local markets are primarily mediated by centralized platforms called Multiple Listing Services (MLS). Originated in the early 20th century, the term first referred to written publications issued by local real estate boards with information about listings for sale. Since its inception, the MLS has become the main source of listing information for buyer agents. Despite listings posted in the MLS are frequently replicated in online public aggregators such as Zillow or Redfin, some fields are exclusive to the MLS. Additionally, MLS platforms are claimed to have superior functionality (Barry *et al.*, 2024).

Market features generate strong incentives for agents to affiliate with local boards to obtain MLS access and post listings in a timely manner. MLS access has been historically restricted to members of the local board that administers it.[5] NAR affiliation implies that MLSs must abide by a set of common rules. In particular, agents were required to post listings they were repre-

[4] Across all agents ever active in Phoenix between 2003 and 2019 in my data, 70% of them completed at least one transaction as seller agents and one as buyer agents. When I restrict to agents with at least 10 transactions, this figure grows to 98%.

[5] In 2025, the NAR eliminated the affiliation mandate from the MLS Handbook, which regulates MLS activity across all platforms. Decisions on whether to allow access to non-affiliated parties are now at the discretion of each MLS (Indursky, 2025).

senting as a seller agent in the MLS, unless explicit written consent from the seller (NAR 2014 Handbook). These characteristics imply that MLS activity covers the majority of transactions in a market. Hence, MLS data allows to measure agent performance for a large fraction of a local market in a comprehensive way.

3.3 Commission rate setting

NAR rules also regulated agent compensation. Specifically, historical NAR-imposed MLS regulations required that listings must include offers of compensation for buyer agents contingent on the closure of the sale. These offers were typically denominated as a fraction of sale prices. Although this practice has been subject of recent antitrust litigation and formally abandoned in 2024 (Han, 2024), it is a prevalent feature of my sample period. Because of this, my model of MLS transactions incorporates commission rates offers as a listing attribute. Differential commission rates affect agents' incentives to recommend properties, which has to be accounted for when deriving quality measures.

The distribution of commission rates offered across markets has two peculiarities. First, within a market, there is a discrete set of common commission rate offers. Second, the prevailing rate or rates differ across markets. Figure A2 shows the distribution of commission rates for the period 2014-2018 for four metro areas: Phoenix; Las Vegas; Portland, OR; and Tampa. In all these metro areas, the majority of commission rate offers are concentrated in two price points. However, the specific common rates and its relative frequency vary across metro areas. Previous research has noted that the factors determining local modal rates across markets can be idiosyncratic (Banerjee and Paciorek, 2025), and can be related to historical commission rate schedules designed by local real estate boards. Despite having been repealed many decades ago, path dependence to initial standards has been documented (Barry *et al.*, 2024; FTC, 1983). These features will be important when I construct an instrument for commission rates in section 6.

4 Data

I combine property listings and transactions data with data on buyers' demographic characteristics. These sources allow me to derive measures of agents' quality of advice and estimate determinants of agent choice.

4.1 Property listings and transactions data

Data on property listings and transactions comes from real estate software provider Corelogic. My main source of information is the Corelogic MLS dataset, which includes listing-level data for a sample of MLSs across the US. I observe listing-level data from more than 100 multiple listing services. The time coverage of the data varies across MLSs, because of the different times at which local boards contract Corelogic's software. For some markets, the data goes as far back as 1992, and ends in August 2024. Due to the effects of the COVID pandemic on the housing market, I use transactions up to 2019.

The data provides me with three important pieces of information. First, I observe a wide range of property characteristics as well as listing conditions, such as the list price. Second, I observe information on listing outcomes, including whether the property sold, the closing price, and the time from the initial listing until sale. Third, the dataset identifies the listing agent and, where applicable, the buyer agent. This allows me to measure agent performance by leveraging outcomes of agent-intermediated listings, after controlling for a rich set of property characteristics.

I focus my empirical analysis on the Phoenix metropolitan area. Between 2003 and 2019, my MLS data covers more than 50% of total market activity in each year, and in most cases more than 70%.[6] Data for a large time window allows me to observe repeat sales for multiple properties. This enables me to control for property unobservables, which are a main identification concern for my value added estimation. I apply a series of data cleaning steps that are standard in the literature (Cunningham *et al.*, 2022; DeFusco *et al.*, 2022; Gilbukh and Goldsmith-Pinkham, 2024). A complete description of this procedure is provided in Appendix C.1.

4.2 Buyer characteristics data

Information about homebuyers' demographic characteristics comes from publicly available data on mortgage issuances. This information allows me to account for covariates that influence demand for property attributes. Measuring this heterogeneity is key for my study, since the value of an agent for homebuyers depends on how likely she is to match buyers to their preferred properties, as highlighted in section 3.1.

Information on mortgages associated with MLS transactions comes from the Corelogic Mort-

[6]Individual MLSs are heterogeneous in geographical coverage, from a metro area to multiple counties or even multiple states. I cannot observe statutory areas corresponding to each MLS. Because of this, I define metro areas as my relevant unit of observation. To check metro area-level coverage of the MLS data, I compare the number of listings in the MLS data with the number of transactions present in the Corelogic Deeds data for each metro area-year for the period 2003-2019.

gages dataset. For each mortgage, I observe the date of issuance, loan attributes, and the property associated with the mortgage. I apply the same sample restrictions as in the MLS data to filter properties.

I then match this sample of mortgages with publicly available, anonymized data on mortgage issuances from HMDA. HMDA is a mortgage-level dataset that includes information on loan attributes and borrower demographic characteristics. I fuzzily match loans in the Corelogic Mortgages data with HMDA loans with similar attributes, following a standard procedure in the literature (Bayer *et al.*, 2025; Huang, 2025). I describe the methodology and results in Appendix C.3. From this procedure, I obtain the likely income of the mortgage applicant, as well as whether it had a co-applicant, which I use as proxy for marital status.

5 Agent Pricing Ability

In this section, I estimate agents' pricing ability conditional on a property match, ϕ. I leverage transaction outcomes in the MLS to compute estimates of the impact of an agent on final transaction prices. I start this section by describing my estimation procedure. I then outline my identification strategy and assumptions. I conclude the section by providing descriptive facts about agent pricing ability.

5.1 Methodology

I estimate agent-level pricing ability using an adaptation of the value-added estimation methodology developed by Ahn *et al.* (2025) in the teacher context.[7] The main difference in my setting has to do with the bilateral nature of property transactions. Unlike other value added applications, outcomes in real estate transactions are jointly shaped by agents on both sides of the transaction. Therefore, my estimation strategy jointly estimates value added terms for seller and buyer agents from each transaction.

My outcome of interest to measure pricing ability is agents' impact on the ratio of closing to list prices. This measure is directly related to consumers' interest in help on terms of sale as an important factor that determines demand for agents, as mentioned in section 3.1. I can observe this ratio for all closed transactions.

My estimation procedure follows otherwise the canonical structure of value added estimation

[7] Value added models to estimate real estate agent productivity have been previously used by Cunningham *et al.* (2022) to estimate effects on close prices and time on market using data from Charlotte, Minneapolis, and Houston.

routines. In the first step, I obtain prior agent level measures of value added using fixed effects models. In the second step, I postulate a population distribution for the agent value added terms and use the distribution of prior estimates to recover the population parameters by maximum likelihood. Finally, I use the recovered posterior distribution to construct Empirical Bayes estimates of the agent-level effects.

5.1.1 Step 1: Fixed Effects Regression and Prior VA Estimates

Denote by i a listing event at time t. Let s and b index the seller and buyer agent associated with the closing of listing i, respectively. I model the determinants of the close/list ratio Y_{isbt} as a linear combination of property characteristics and agent effects as follows:

$$Y_{isbt} = \underbrace{\beta X_{it}}_{\text{Property characteristics}} + \underbrace{\overbrace{\psi^S X_s + \psi^B X_b + \psi^{SB} X_{sb}}^{\text{General agent effects}} + \overbrace{\phi_s + \phi_b}^{\text{Individual agent effects}}}_{\text{Agent effects}} + \epsilon_{isbt} \tag{6}$$

The first block of equation (6), X_{it}, includes property characteristics that affect listing outcomes regardless of the identity of the agents involved in the transaction. I control for a set of property characteristics that are standard in the hedonic pricing literature (Sirmans *et al.*, 2005; Rutherford *et al.*, 2005; Levitt and Syverson, 2008; Barwick *et al.*, 2017; Cunningham *et al.*, 2022). I list my control property characteristics in Table 1. Furthermore, I focus my sample on properties that sell multiple times during my sample period, allowing me to include property fixed effects. These account for all unobservable property-specific attributes that might affect its desirability. In addition, property fixed effects allow me to account for potential non-random sorting of agents into properties with particular unobservable attributes.

The second block of equation (6), corresponding to agent effects, contains two components. The first component, corresponding to general agent effects, includes agent-level covariates that enter the outcome equation linearly and have a uniform effect across all agents. I distinguish two sets of variables in this component. First, I include covariates corresponding to each agent side in X_s and X_b. Second, I include covariates corresponding to the interaction between both agents in $X_{s,b}$. The second component of the agent effects block corresponds to the individual agent effects. These terms capture any added value in the transaction corresponding to the agents involved in the transaction that cannot be explained by observable agent attributes.

Agent-specific covariates X_s and X_b account for general factors that can generally affect agent performance, regardless of the specific agent involved in the transaction. I include this component to capture average market-wide determinants of performance, that I isolate from the

individual agent effects measured in ϕ_s and ϕ_b. I then assess the relative impact of general covariates compared to individual effects. For each agent side, I include in X_s and X_b the log number of transactions completed by the seller (buyer) in the previous year, across all Phoenix and in the zipcode of the current transaction.

I further control for factors corresponding to the interaction between seller and buyer agent. In theory, agents could complete transactions with other agents in a non-random way. Failing to account for potential sources of non-randomness in agents' bilateral pairings can bias my estimated value added coefficients. The source of this bias is ambiguous. On the one hand, non-random sorting could be associated with exploiting synergies that make trade more efficient, benefiting consumers. On the other hand, non-random sorting could be associated to steering concerns, which have been a first-order concern in the real estate industry and source of antitrust litigation.[8]

As factors of potential non-random sorting between agents, $X_{s,b}$ includes an indicator for when the seller agent is also the buyer agent, and when the seller and the buyer agent work for the same brokerage company. In Arizona, dual agency, meaning that the same agent represents both seller and buyer, is legal.[9] I also control for an indicator of whether seller and buyer agent work for the same brokerage company.

Small agent grouping. To deal with the high uncertainty in fixed effect estimates for units with a small number of observations, and ensure connectivity of the graph of agent transactions, I group all agents with at most 10 transactions in a given side (seller or buyer) in a single term. I estimate an agent-specific term in all other cases. Thus, if, for example, an agent has 20 transactions as seller agent but 5 transactions as buyer agent, I estimate an agent-specific seller value added term for this agent, whereas its buyer value added corresponds to the small agent group.

Results. I estimate equation (6) by ordinary least squares. To account for potential correlation between contemporaneous listing outcomes across regions, I cluster my standard errors by zipcode and listing year-month. Table 2 shows results. My results suggest that market-wide agent covariates have little impact on the close/list ratio. I find that the number of previous transactions for seller agents has a statistically insignificant effect on close/list ratios. For buyer agents, doubling the yearly number of transactions only increases final sale prices, as a function

[8] Sitzer v. National Association of Realtors, 4:19-cv-00332, (W.D. Mo.)

[9] Arizona Administrative Code § R4-28-1101 subsection F states that a real estate agent"shall not accept compensation from or represent more than one party to a transaction without the prior written consent of all parties" (Arizona Administrative Code, 2024).

of the list price, by 0.02 percentage points. These results are consistent with prior literature (Gilbukh and Goldsmith-Pinkham, 2024). Doubling the yearly number of transactions by an agent pair increases the close-list ratio by 0.11 percentage points. Properties where the same agent represents both sides of the transactions tend to sell at a lower final price, conditional on the list price. I estimate that properties transacted with a dual agent experience a decrease in their close/list ratio of 0.55 percentage points.

5.1.2 Step 2: Estimation of Population Distribution Parameters

In the second step, I estimate the parameters Ψ of the distribution of agent-level effects $f(\phi|\Psi)$. Following common practice, I model f as a multivariate Normal distribution: $f(\phi|\Psi) = \mathcal{N}(\eta, \Delta)$.

I leverage the relationship between the true parameters ϕ_a and the estimated parameters $\hat{\phi}_a^{OLS}$ to obtain an estimating equation for Ψ. Specifically, the log-likelihood function of Ψ given the estimated $\hat{\phi}^{OLS}$ and the population distribution of ϕ is

$$\ell(\Psi|\hat{\phi}) = \sum_{j=1}^{N} \ln \left[\int \mathcal{N}\left(\hat{\phi}_j|\phi_j, \hat{\Sigma}_j\right) \mathcal{N}\left(\phi_j|\eta_c, \Delta_c\right) d\phi \right] \tag{7}$$

In Appendix D.1, I show that the likelihood function in equation (7) can be expressed exclusively in terms of the OLS coefficients $\hat{\phi}$, $\hat{\Sigma}$ and the population parameters η and Δ:

$$\ell(\Psi) = \sum_{j=1}^{N} \ln \left(\mathcal{N}\left(\hat{\phi}_j|\eta, \hat{\Sigma}_j + \Delta\right) \right) \tag{8}$$

I estimate the population parameters η and Δ by maximum likelihood. Following Ahn *et al.* (2025), I maximize equation (8) using the quasi-Newton algorithm proposed by Jamshidian and Jennrich (1997). After I recover the population distribution, I perform an Empirical Bayes procedure to obtain posterior estimates of the individual value added effects. I follow Ahn *et al.* (2025) and standard practice by shrinking the point estimates $\hat{\phi}_a^{k,OLS}$ of individual value added terms towards the population mean, as a function of the signal-to-noise ratio driven by the population variance and the variance of $\hat{\phi}_a^{K,OLS}$.

5.2 Identification

Before presenting the main descriptive facts regarding agent pricing ability, in this section I discuss identification of the value added coefficients. The main identification concern to causally

interpret ϕ_j as individual measures of agent quality is selection bias. Three non-exclusive forms of selection are possible in my setting. First, agents may non-randomly sort into intermediating in transactions of properties with unobserved attributes that affect their final sale price. Second, for a given agent, the counterpart agents with which they close transactions might be non-random. Third, client-related unobservables correlated with listing outcomes (for example, seller patience) can bias the estimated quality effects if such unobservables also affect agent choice. I discuss each of these selection concerns in more detail, and how I address them, in turn.

As described in Section 5.1, I address the potential non-random sorting of agents into properties by focusing my estimation on properties that sell more than once during my sample period. This allows me to include in equation (6) property-level fixed effects, which account for all unobservables associated with a specific property that affect its outcomes in the market. By adding property fixed effects, I remove from my estimated $\hat{\phi}^{OLS}$ all unobserved property attributes that could affect my first-stage estimation of agent quality.

As a second concern, non-random sorting of agent pairs could bias the estimated value added coefficients. As discussed in section 5.1.1, the sign of this bias is ex ante unclear. To address this concern, $X_{s,b}$ in equation (6) includes factors that could affect the propensity of specific agent side pairs to work together. Specifically, $X_{s,b}$ includes a dual agency indicator (the same agent representing the buyer and the seller), an indicator for agents of the same brokerage firm representing each side, and the number of transactions closed by the pair in the prior year. This alleviates the concern under the assumption that agent pair sorting under other observable or unobservable characteristics is either absent or uncorrelated with agent quality.

Finally, non-random sorting of clients into agents could bias the estimated coefficients. For example, clients with more knowledge about the housing market could exert more effort in finding high-quality agents and could also exert actions that improve their outcomes in a transaction (for example, incur in value-increasing maintenance prior to selling). If so, this could bias my estimated terms $\hat{\phi}^{OLS}$.

To assess this concern, I exploit the repeat-sales nature of my transactions panel. Except for the first seller and the last buyer of a property, I observe the remaining clients transacting the property in two instances: when they initially buy the property and when they sell it. Under the assumption that client-specific factors that affect agent choice are time- and side-invariant, clients of agent services should select an agent with similar price effects for both transactions. In Appendix Figure A3, I correlate the buyer price effect on the first transaction with the seller price effect on the second transaction. I find that quality measures are negatively

correlated, meaning that clients do not systematically select agents of a certain quality. This gives suggestive evidence that client-driven selection may not be a significant concern in this setting.

5.3 Results

Heterogeneity in pricing ability. Figure 1 shows the distribution of the estimated agent pricing ability terms after the Bayes shrinkage correction. All value added measures are centered in the "small agent" composite. All indicators are labeled such that higher quality of service for consumers correspond to a higher value added measure; specifically, greater pricing ability terms for buyer agents correspond to *lower* purchase prices, conditional on a list price.

The dispersion in price effects is sizable across sides. The standard deviation of the seller agent distribution is 0.8 percentage points. The standard deviation of the buyer agent distribution is 0.9 percentage points. When compared to the lack of variation in prices obtained as a function of experience, baseline differences in agent-specific pricing ability are significant. This variation is similar in magnitude to the spread of total commission rates in Phoenix.[10]

Relationship between pricing ability and market activity. Figure 2 shows the relationship between pricing ability estimates and market activity, measured as the number of transactions completed across the entire sample. Mirroring the first stage results, I find that transaction volume is essentially flat with respect to pricing ability. At any point in the distribution of market activity, the distribution of pricing ability estimates is symmetric around 0, meaning that high pricing ability agents do not face greater market shares.

6 Agent Incentive Alignment Weights

In this section, I estimate agent incentive alignment weights θ. As introduced in section 2, this parameter determines how aligned agent recommendations are with respect to buyers' preferences for property characteristics. I begin by writing my empirical model of agent-intermediated property choice. The model has the same structure as the stylized model in stage 2 of section 2, where properties are chosen to maximize a weighted sum of buyer utility and agent earnings. In particular, θ measures the relative weight agents place on their own earnings in this decision. I then explain how I estimate and identify the model. I discuss how I

[10] Assuming equal rates offered to both buyer and seller agent, typical total commissions in Phoenix are either 5% or 6%.

address the main identification concern of endogeneity of prices and commission rates. I finally show results, highlighting the estimated variation in incentive alignment weights.

6.1 Preliminaries

My model of agent-intermediated property choice combines standard elements of the literatures on demand estimation for housing and intermediation services (Bayer *et al.*, 2007; Robles-García, 2020; Kim, 2025). Buyers and agents choose among the set of properties available in the market to maximize a weighted sum of their respective utilities. Buyer utility depends on property characteristics and transaction prices. Agent earnings equal their commission income obtained if the transaction is complete. Motivated by the institutional setting, these earnings are the product of the final sale prices and a commission rate pre-established by the seller.

To estimate the property choice model, I apply some sample restrictions due to data availability constraints and identification considerations. Because data to construct the instruments detailed in section 6.2.1 is only available for the later part of the sample, I estimate the model using data on property transactions between 2015 and 2019. To avoid mechanical correlation between my two estimates of agent quality, I further restrict to transactions of properties that only sold once between 2003 and 2019, excluding the repeat sale properties used to estimate the model in section 5.

6.2 Setting

Let i denote buyers and b denote buyer agents. Consumers and brokers jointly choose a property j listed by seller agent s in period (quarter) t to maximize joint utility:

$$\begin{aligned} u_{ibjt} = & \underbrace{(1-\theta_b)}_{\text{Weight on buyer utility}} \times \underbrace{[-\alpha_i p^e_{jt} + \beta_i X_j + \xi_{jt} + \epsilon_{ijt}]}_{\text{Buyer utility}} \\ & + \underbrace{\theta_b}_{\text{Weight on agent utility}} \times \underbrace{(c_{jt} p^e_{jt})}_{\text{Agent utility}} \\ & + \underbrace{\gamma^s X^s_{st} + \gamma^{sb} X^{sb}_{sbt} + \gamma^{jsb} X^{jsb}_{jsbt}}_{\text{Utility from agent network}} \end{aligned} \tag{9}$$

As introduced in section 2, buyers' utility depends on expected closing prices p^e_{jt}. Closing prices are determined by the interaction of the list price and the pricing ability terms estimated in section 5. Specifically, $p^e_{jt} = (1 + \phi_s - \phi_b) p^l_{jt}$, where ϕ_s and ϕ_b are the estimated pricing abil-

ity terms from section 5 and p^l_{jt} is the list price. Buyers have also preferences about property attributes X_j. I assume that preferences over prices and property characteristics can be heterogeneous across buyers i. Buyer preferences are further influenced by an unobserved quality term ξ_{jt} and an idiosyncratic, type I extreme value shock ϵ_{ijt}.

From the agent side, expected earnings from closing a transaction are a function of the expected closing prices p^e_{jt} times the commission rate corresponding to the buyer agent, c_{jt}. Motivated by the institutional setting described in section 3, the commission rate is set by the seller side upon origination of the listing and non-negotiable. Agent b puts a relative weight of θ_b on their own earnings and $1 - \theta_b$ on buyer utility.

When taking the model to the data, I allow choices to depend on the characteristics of the agent network. This allows me to account for factors that might influence agents' awareness of properties available in the market, and incentives other than buyer-side earnings in recommending such properties to consumers. I parametrize the value provided by the agent network as a linear function of three terms. These terms echo the formulation in equation (6) by accounting for general agent experience factors and sources of non-random sorting in agent counterparts. In principle, the effects on buyers of receiving these recommendations are ambiguous, and depend on how correlated the characteristics of these properties are with respect to buyers' own preferences (Kim, 2025).

6.2.1 Identification

The two main identification concerns for estimating equation (9) relate to the endogeneity of list prices p^l_{jt} and commission rates c_{jt}. Price endogeneity reflects the standard concern that prices can be correlated with unobserved property quality (Bayer *et al.*, 2007). In turn, commission rates c_{jt} can be correlated with latent seller attributes that affect a property's unobserved condition. For example, sellers with better information about the real estate market might be more likely to successfully negotiate a lower commission rate for their listing. If these sellers are also more likely to undertake unobserved investments that increase the property's selling price, this introduces bias in my estimate of θ_b.

I instrument for prices by constructing a shift-share instrument that exploits national shifts in the relative size of demographic groups with preferences for certain property attributes, following Cook *et al.* (2025). Concretely, the shifts are national population trends for different demographic groups, and shares are the likelihood that such demographic groups are observed living in properties of each type in the five years prior to the estimation sample (2010-2014). This instrument mirrors Waldfogel instruments: if some demographic groups become rela-

tively more popular, and these groups have preferences for particular property types, prices for these properties tend to increase due to increased demand and the inelasticity of housing supply (Saiz, 2010; Baum-Snow and Han, 2024; Cook *et al.*, 2025). Details on how I construct this instrument are provided in Appendix E.1.

My motivation to construct an instrument for commission rates exploits seller agent-level frictions in their ability to negotiate down commission rates with a particular seller that are plausibly uncorrelated with unobserved property or seller attributes. The main institutional feature motivating this approach is the structure of the contracts between agents and their brokerage company. These contracts typically specify a commission split that the agent must share with their brokerage company.[11] These contracts can influence the ability of seller agents to offer lower rates, for example, due to revenue requirements by the brokerage company or incentives driven by agent-specific contractual caps.

I instrument for commission rates c_{jt} by using a Hausman instrument. Specifically, my instrument is the average commission rate offered in listings: (i) by the same brokerage company; (ii) at the same time t, (iii) in metro areas other than Phoenix, and (iv) for properties of similar characteristics to j. I motivate this instrument on correlation in agent-level contract designs for the same company across metro areas, and idiosyncratic determinants of commission rates across metro areas as described in section 3.3.

6.2.2 Estimation

To estimate the model, following standard practice (Robles-García, 2020; Kim, 2025), I first divide Equation (9) and rearrange terms to obtain:

$$\tilde{u}_{ibjt} = \underbrace{-\alpha_i p^e_{jt} + \beta_i X_j + \xi_{jt} + \tilde{\theta}_b(c_{jt}p^e_{jt}) + (\tilde{\gamma}^s X^s_{st} + \tilde{\gamma}^{sb} X^{sb}_{sbt} + \tilde{\gamma}^{jsb} X^{jsb}_{jsbt})}_{U_{ibj}} + \epsilon_{ijt} \tag{10}$$

where $\tilde{\theta}_b = \frac{\theta_b}{1-\theta_b}$ and $\tilde{\gamma} = \frac{\gamma}{1-\theta_b}$. Since ϵ_{ijt} is type I extreme value, choice probabilities are given by

$$s^H_{ibj^*t} = \frac{\exp\left(-\alpha_i p^e_{j^*t} + \beta_i X_{j^*} + \xi_{j^*t} + \tilde{\theta}_b(c_{j^*t}p^e_{j^*t}) + [\tilde{\gamma}^s X^s_{st} + \tilde{\gamma}^{sb} X^{sb}_{sbt} + \tilde{\gamma}^{jsb} X^{jsb}_{j^*sbt}]\right)}{\sum_{j\in\mathcal{J}} \exp\left(-\alpha_i p^e_{jt} + \beta_i X_j + \xi_{jt} + \tilde{\theta}_b(c_{jt}p^e_{jt}) + [\tilde{\gamma}^s X^s_{st} + \tilde{\gamma}^{sb} X^{sb}_{sbt} + \tilde{\gamma}^{jsb} X^{jsb}_{jsbt}]\right)} \tag{11}$$

[11] For example, eXp Realty operates in an 80%-20% commission split, under which the agent retains 80% of the commission, while the remaining 20% goes to the company. After reaching a $16,000 annual cap, agents keep 100% of the commission for the remainder of the year. Source: https://www.exprealty.com/income

Finally, I parametrize the unobserved heterogeneity as

$$\zeta_{jt} = \zeta_{jt}^{P} + \zeta_{jt}^{C} \tag{12}$$

where ζ_{jt}^{P} and ζ_{jt}^{C} are the terms correlated with price and commission rate, respectively.

I estimate the model by Maximum Likelihood Estimation, using a control function approach (Petrin and Train, 2010; Kim, 2025). I estimate the IV regressions for price and commission rate on both instruments and property characteristics on a first stage, and then include the residuals of these regressions as an additional term in the second stage. In terms of consumer heterogeneity, I allow the price coefficient to vary with income. I follow Miravete *et al.* (2026) and Kim (2025) by parametrizing the price coefficient as $\alpha_i = \frac{\alpha}{y_i^{1-\rho}}$.[12] This formulation allows price sensitivity to depend on income, with ρ governing the rate at which price sensitivity decreases for higher-income households. Following Kim (2025), I calibrate $\rho = 0.79$. In terms of property attributes, I interact preferences for number of bedrooms using income quartile by marital status bins.

My main object of interest is to estimate heterogeneity in agent matching quality, measured by the θ_b parameters. To reduce computational burden, while at the same time being able to pick up meaningful heterogeneity in incentive alignment weights, I implement a version of the grouped fixed effects methodology developed by Bonhomme *et al.* (2022). I first estimate a simplified version of the model in equation (10) as a function of list prices, without agent heterogeneity terms and agent network effects. Using the estimated parameters, I calculate buyers' utility from each property in their choice set, excluding commission income. I then calculate, for each choice instance, the percentile rank of the chosen property with respect to buyers' utility. Intuitively, choices of properties that rank higher in buyers' preferences reflect greater value of agent advice for buyers, consistent with buyers' stated preference for agents that help them "find the right properties" as shown in section 3.

I use the empirical distributions of buyers' property ranks to group agents in terms of their incentive alignment weights. Intuitively, agents who tend to choose the properties ranking higher in buyers' preferences provide greater value to consumers. Following Bonhomme *et al.* (2022), I slice the distribution of chosen buyer ranks in ventiles, and compute, for each agent, the empirical CDF of the ventile ranks chosen. This procedure yields a 20-dimensional vector for each agent. I then perform k-means clustering to group the agents in four groups.[13] I index

[12] Miravete *et al.* (2026) motivate this price sensitivity formulation as a first-order Taylor approximation at $p = 0$ of a Box-Cox transformation of disposable income $y_i - p$, with Box-Cox parameter ρ.

[13] I exclude from this classification all agents that have been assigned to the small agent composite group for

these groups so that group 1 corresponds to agents with worse choices for buyers on average, while group 4 corresponds to better choices on average. In the full model, I estimate group-specific θ_b coefficients for these four groups.

Choice set construction. In the MLS data, I do not observe which properties were on buyers and/or agents' choice sets. I define choice sets by leveraging recent research on homebuyers' search behavior patterns (Piazzesi *et al.*, 2020). I start by dividing the Phoenix market in segments. Following Piazzesi *et al.* (2020), I divide each zipcode's housing market in eight groups, composed of four zipcode-specific price quartiles and whether the property has multiple bathrooms, which proxies for size. I then define a segment S as a zipcode-size-price band triplet. I then include in the choice set all properties in the segment S_{j^*} of the chosen property j^* listed during the quarter t where the transaction is completed. Let $\mathcal{S}_{j^*}$ denote the number of properties in this segment. I then augment the choice set by randomly sampling, from the set of properties listed at time t, properties in segments other than S_{j^*} that share a characteristic with j^*. Specifically, I randomly sample, with reposition, $\mathcal{S}_{j^*}$ properties from each of the following groups: (i) properties in the same zipcode as j^*, (ii) properties in the same price band as j^*, (iii) properties in the same zipcode as j^*, listed by agent b as seller agent, and (iv) properties in the same price band as j^*, listed by agent b as seller agent. The latter two groups ensure that I accurately measure agents' incentives to recommend their own listings, which are likely to be in the agent's consideration set, as well as differences in agent inventories.

This definition of choice sets ensures that buyers can, in principle, consider a set of properties that is considerably broader, both in number and scope, than the specific market segment in which they end up transacting. Indeed, using search alerts data from Trulia, Piazzesi *et al.* (2020) shows that the median searcher considers an inventory multiple times larger than any individual segment. Indeed, my choice set construction strategy implies that buyers will consider an inventory that is at most four times larger than their chosen segment, which is consistent with prior evidence.

6.2.3 Results

Table 3 shows the results of the first stage estimation. My estimates imply that list prices are positive in the shift-share instrument: as competitive pressures increase, prices tend to increase. Regarding the commission rate instrument, an increase in the share of low-commission listings in other markets is associated with a decrease in commissions for the focal properties. This is

pricing quality estimation in Section 5. I estimate a separate matching quality term for this group. Differences across groups in terms of their matching quality estimates are then identified from variation in the correlation between their choices and buyers' preferences across property attributes and expected gains from pricing quality.

also consistent with intuition: when low commission rates become more common for a given property type across markets, commissions for the focal property are more likely to be low as well. In both regressions, the F-test of joint significance of the instrument is satisfied above conventional levels.

Table 4 shows the results of the estimation of the demand model in equation (10). My model picks up heterogeneity in buyer agents' incentive alignment weights. For the lowest incentive alignment weights group, I estimate a normalized weight of $\tilde{\theta}_1 = 0.292$, corresponding to a 22.6% relative weight on commission earnings.[14] At the average sale price in the data of $242,500, this weight implies that, holding property characteristics fixed, reducing the commission rate from 3% to 2.5% is associated with a decrease in choice probabilities of 5.3 percentage points.

In contrast, for the highest incentive alignment weight group, I estimate a normalized weight of 0.041, corresponding to a relative weight of 3.9% on commission earnings. For this group, reducing the commission rate of a listing from 3% to 2.5%, at the average sample price and holding property characteristics fixed, is associated with a decrease in choice probabilities of 0.7 percentage points. In other words, when the commission rate decreases from the high to the low rate, low incentive alignment weight agents become more than 7 times less likely to recommend the property than high incentive alignment weight agents. The wedge in relative weights also has implications for whether the value of the agent network is weighed more heavily with respect to buyer utility or agent earnings.

These results suggest considerable differences in performance in terms of incentive alignment weights and property recommendations. While the distribution shows a heavy right tail for the high buyer weight group, differences with other groups are also important. For the group in the third quartile, I estimate a relative weight on commission earnings of 16.2%, which implies that choice probabilities decrease by 3.5 percentage points when commissions decrease to the low rate. Compared to the lowest buyer weight group, group 3 is 33% more likely to keep recommending the property. This shows that, while not on the same magnitudes as the difference between the extremes, groups with intermediate incentive alignment weights can still offer non-trivial benefits to consumers of agent services.

[14] I back out this weight as $\theta_b = \frac{\tilde{\theta}_b}{1+\tilde{\theta}_b}$.

7 Demand for Agent Services

After estimating the dimensions of agent heterogeneity, I now turn to estimate consumers' sensitivity with respect to quality differences. I start by defining my aggregate measure of advice that combines benefits to buyers from the two heterogeneity indicators ϕ and θ. I then formalize my demand model, where the key parameter is consumers' sensitivity with respect to the aggregate advice measure. I finally show the estimated results, which show that consumer choices are insensitive with respect to differences in the value of advice.

7.1 Inclusive Value of Agent Advice

As defined in section 2, I define the value provided by agent representation to consumers as the inclusive value in equation (9) experienced by buyer i from the advice of buyer agent b. As derived in section 2, the expression of the inclusive value in this context is

$$I_{ibt} = \frac{1}{\alpha_i} \ln \left(\sum_j \exp U_{ibjt} \right) - \frac{1}{\alpha_i} \sum_j s^H_{ibjt} \tilde{\theta}_b (c_{jt} p_{jt} + \tilde{\gamma}_b X_{bjt}) \tag{13}$$

The first term on this expression is the standard consumer surplus formula in discrete choice models (Small and Rosen, 1981) from the joint buyer-broker decision. The second term substracts the component of joint utility related to brokers' preference for high-commission listings, weighted by the choice probability of each listing. The value of agent advice corresponds then to the "experienced" utility induced by the buyer-agent pair choice probabilities, net of distortions introduced in the "decision" utility from agents' private incentives (Dubois *et al.*, 2018). The distribution of the estimated inclusive values is shown in Figure 3. I find that, for the median buyer, the distribution in the value of advice, considering all agents with at least one transaction in their zipcode-year, is 1.6% of average sale prices.

7.2 Parametrization and Estimation

I define buyer i's utility from seeking representation from agent b as

$$u^B_{ibt} = \lambda I_{ibt} + \beta^B X^B_{bt} + \xi_b + \epsilon'_{ibt} \tag{14}$$

where ϵ_{ibt} is a type I extreme value shock, X^B_{bt} are agent b's observable characteristics that could potentially affect agent choice, such as agent experience, and ξ_b are agent fixed effects.

The buyer's choice set $\mathcal{B}$ includes all agents that have completed at least one transaction in year $y(t)$ in the same zipcode as the property chosen by buyer i, plus the small agent composite.

7.3 Estimation and Identification

By the type I extreme value assumption on ϵ'_{ibt}, agent choice probabilities take the form

$$p_{ib^*t} = \frac{\exp(\lambda I_{ib^*t} + \beta^B X^B_{b^*t})}{\sum_{b \in \mathcal{B}} \exp(\lambda I_{ibt} + \beta^B X^B_{bt})} \tag{15}$$

The estimated coefficient on I_{ibt}, λ, can potentially be biased due to measurement error in my pricing quality estimates (Chen *et al.*, 2025). The variation in the data that identifies λ are repeated choices by agents as a function of property attributes and commission earnings. To leverage this source of variation for identification of λ, I follow a similar approach to Kim (2025) by constructing a proxy version of I_{ibt}. I compute this proxy from an estimation of equation (9) without agent quality terms. Let $\tilde{I}_{ibt}$ denote this proxy measure. I then use $\tilde{I}_{ibt}$ as an instrument for I_{ibt}.

The estimation procedure is analogous to the estimation of the property choice parameters in section 6.2.2. I estimate the parameters of equation (14) by Maximum Likelihood, using the control function approach from Petrin and Train (2010) to control for the potential measurement error in I_{ibt}. Formally, I instrument I_{ibt} as a function of $\tilde{I}_{ibt}$ and the remaining included variables in a first stage. I then compute the residuals ν_I of this first stage regression, which I add as a further covariate in equation (14). I then estimate the agent choice model parameters by maximum likelihood.

7.4 Results

Table 5 shows the results of the first stage estimation, indicating that the proxy variable predicts the variation of the full advice value term. Table 6 shows the estimated parameters of the discrete choice model. I find that agent choices are virtually insensitive with respect to the estimated quality differences. My estimates imply that, for the median home buyer, a one standard deviation increase in the value of advice increases choice probabilities by 2.8%. This result is consistent with prior estimates (e.g. Kim, 2025) and with a large body of anecdotal evidence suggesting that consumers exert little search when looking for an agent. A 2025 report by the National Association of Realtors found that between 75% and 80% of buyers and sellers only interview with one agent before hiring them (Tracey, 2025). In the financial advising services industry, where previous research has suggested benefits from increased consumer

awareness (Egan *et al.*, 2024), 57% of clients have reported contacting only one advisor (Dynasty Financial Partners, 2023).

8 Comparative Statics

After estimating the model primitives, I assess the welfare effects of comparative statics changes to the primitives of the model. I evaluate changes that target the two key primitives: the sensitivity of consumer demand to agent advice and the agent-specific determinants of advice.

I first define my notion of surplus. I measure consumer surplus as the expected consumer surplus from the housing market (net of agent incentives), weighted by property choice probabilities induced by agent recommendations. I next describe and evaluate my two comparative statics exercises.

8.1 Definition of Surplus

As standard in the intermediation literature (Robles-García, 2020; Kim, 2025), I measure consumer surplus as the expected product consumption utility (net of agent distortions) weighted by agent choice probabilities:

$$CS_i = \sum_{b \in \mathcal{B}} p_{ibt} \left[\frac{1}{\alpha_i} \ln \left(\sum_l \exp U_{ibl} \right) - \frac{1}{\alpha_i} \sum_l s^{ibl} (\tilde{\theta}_b \pi_{bl} + \tilde{\gamma}_b X_b) \right] \tag{16}$$

where p_{ibt} is defined as in equation (15) and U_{ibt} is defined as in equation (10). This expression highlights that expected surplus is a function of the expected housing consumption benefits provided by buyer agent b, weighted by the probability of choosing agent b. Compared to I_{ibt}, this expression only considers benefits from housing consumption as part of consumer welfare, abstracting from the network value accrued to the buyer.

8.2 Increases to Demand Sensitivity

I first evaluate the effects of changing consumer sensitivity with respect to differences in the inclusive value of agent quality I_{ibt}. Formally, I operationalize this counterfactual by multiplying the estimated λ from equation (14) by a parameter ω:

$$\tilde{u}^B_{ibt} = (\omega \hat{\lambda}) I_{ibt} + \hat{\beta}^B X^B_{bt} \tag{17}$$

I evaluate multipliers ω between 1.1 and 2. This interval covers the estimated impact of consumer awareness campaigns in the financial advising literature. For instance, Zhao (2023) and Egan *et al.* (2026a) study the impacts of a national advertising campaign ran by the Financial Industry Regulatory Authority (FINRA) to promote BrokerCheck, a public-facing website that discloses information about financial advisors, including experience, certifications, and sanctions. Zhao (2023) and Egan *et al.* (2026a) find that, after the campaign, Google searches of the term "BrokerCheck" increased between 50 and 60 percent.[15] I estimate the effects of consumer surplus of increasing $\hat{\lambda}$ in 10% increments up to 100%.

Figure 4 shows results. I find that effects on surplus are increasing in ω. Doubling consumer sensitivity to differences in advice increases surplus for the average consumer by the equivalent of 0.91% of average sale prices.

8.3 Increases to Incentive Alignment Weights

I next evaluate the effects of increasing alignment weights for low alignment agents. I simulate three scenarios. In Scenario 1, agents of the lowest incentive alignment weight quartile (Q1) increase their weight on buyer utility to the level of agents in group 2. Given the estimated incentive alignment weights from Table 4, this scenario implies that weights on own earnings for the bottom quartile increase by 10 percent. In Scenario 2, both Q1 and Q2 agents increase buyer incentive alignment weights to the level of the composite agent, implying that the bottom quartile reduces their weight on own earnings by 18 percent. In Scenario 3, Q1, Q2, and composite group agents increase buyer incentive alignment weights to the level of group 3. This scenario implies that agents in the bottom quartile reduce their weight on own earnings by 33 percent.

I benchmark these numbers by comparing them to interventions aimed at improving quality of service in the financial advising industry. Kowaleski *et al.* (2020) analyze the effects of a 2010 regulatory change that reduced the weight of rules and ethics questions in a qualification exam for financial advisers. They find that advisers passing the old exam are associated with 25 percent lower misconduct rates, from a mean misconduct rate at the individual-firm-year level of 0.86%.

Table 7 shows results. In all three scenarios, consumer surplus increases. Surplus gains in the first two scenarios are 0.27% and 0.97% of average sale prices, comparable to increasing the demand sensitivity parameter by 20% and 100%, respectively. Scenario 3, where the agents

[15] Additionally, informational interventions have been found to double consumer awareness about the disclosed characteristics in some contexts, including information on school average test scores regarding school choice and differences in out-of-pocket costs regarding health plan choice (Hastings *et al.*, 2007; Kling *et al.*, 2012).

below the median and the small agent composite increase buyer weights to the third quartile, achieves gains of 2.52%. These results suggest that gains from comparative statics changes to the supply of advice depend both on the incremental weight on buyer surplus for affected agents, and on the fraction of agents that are subject to this increase. Indeed, consumer surplus effects experience a relatively large jump when the low-volume agent group is subject to an increase in buyer weights, despite the increase that I test is fairly monotonic in magnitude across scenarios.

These results highlight two lessons. First, because the agent demand results suggest that agent choices are relatively fixed, surplus gains from targeting the provision of advice directly fall mostly on the current consumers of the agents whose quality parameters increase, whereas gains from reallocation are limited. In contrast, shifts to demand sensitivity induce greater sorting to high-quality agents. Second, because of this, shifts to the quality of advice need to achieve improvements in service for a large population of the agent market.

9 Conclusion

In this paper, I study how heterogeneity in agent quality and consumer demand for quality shape welfare in intermediation markets, with an application to the U.S. real estate brokerage industry. My conceptual framework highlights that the relative contribution of each margin depends on the distribution of agent quality of advice and the sensitivity of consumer demand to these differences. In the real estate brokerage industry, I find that consumer demand is relatively insensitive to the value of agent advice. My comparative statics exercises suggest that increases in demand sensitivity can increase consumer surplus from housing consumption, and that similar results can be achieved through the supply side if interventions that increase quality of advice directly target a sufficiently high fraction of agents.

Understanding what types of interventions can affect consumer decisions is a promising area for future research. For instance, information about agents' past transactions is publicly available in real estate web aggregators (Brobeck, 2020). Whether and how consumers use this information remains an open question. Beyond the real estate brokerage industry, the conceptual framework of this paper can be broadly applicable to other markets for intermediation services.

References

ABALUCK, J., CACERES BRAVO, M., HULL, P. and STARC, A. (2021). Mortality effects and choice across private health insurance plans. *Quarterly Journal of Economics*, **136** (3), 1557–1610.

AHN, T., AUCEJO, E. M. and JAMES, J. (2025). The importance of student-teacher matching: A multidimensional value-added approach. *Review of Economics and Statistics*, pp. 1–45.

ARIZONA ADMINISTRATIVE CODE (2024). Arizona Administrative Code, Title 4, Chapter 28: State Real Estate Department, R4-28-1101 (Licensee Duties). Ariz. Admin. Code R4-28-1101, https://apps.azsos.gov/public_services/Title_04/4-28.pdf.

ATAL, J. P., CUESTA, J. I. and SÆTHRE, M. (2025a). Quality Regulation and Competition: Evidence from Pharmaceutical Markets. *Working Paper*.

—, LARROUCAU, T., MUNOZ, P., OTERO, C. and FREEDMAN, S. (2025b). Physicians' Occupational Licensing and the Quantity-Quality Trade-Off. *Working Paper*.

BANERJEE, R. and PACIOREK, A. (2025). Commissions and Omissions: Trends in Real Estate Broker Compensation. *FEDS Notes*, (2025-05), 12.

BARRY, J. M., FRIED, W. and HATFIELD, J. W. (2024). Et Tu, agent? Commission-based steering in residential real estate. *Iowa L. Rev.*, **110**, 1473.

BARWICK, P. J., LI, S., WAXMAN, A., WU, J. and XIA, T. (2024). Efficiency and equity impacts of urban transportation policies with equilibrium sorting. *American Economic Review*, **114** (10), 3161–3205.

— and PATHAK, P. A. (2015). The costs of free entry: an empirical study of real estate agents in Greater Boston. *The RAND Journal of Economics*, **46** (1), 103–145.

—, — and WONG, M. (2017). Conflicts of interest and steering in residential brokerage. *American Economic Journal: Applied Economics*, **9** (3), 191–222.

— and WONG, M. (2019). Competition in the real estate brokerage industry: A critical review. *Brookings Institution Publication*.

BAUM-SNOW, N. and HAN, L. (2024). The microgeography of housing supply. *Journal of Political Economy*, **132** (6), 1897–1946.

BAYER, P., CASEY, M., MCCARTNEY, W. B., ORELLANA-LI, J. and ZHANG, C. (2025). Distinguishing causes of neighborhood racial change: A nearest-neighbor design. *American Economic Review*, **115** (11), 3999–4039.

—, FERREIRA, F. and MCMILLAN, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of political economy*, **115** (4), 588–638.

BERNHEIM, B. D. and MEER, J. (2013). Do real estate brokers add value when listing services are unbundled? *Economic Inquiry*, **51** (2), 1166–1182.

BHATTACHARYA, V., ILLANES, G. and PADI, M. (2025). Fiduciary duty and the market for financial advice. *Econometrica*, **93** (4), 1449–1480.

BOEHM, E. (2025). Intermediation, choice frictions, and adverse selection: Evidence from the Chilean pension market. *Working Paper*.

BONHOMME, S., LAMADON, T. and MANRESA, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, **90** (2), 625–643.

BROBECK, S. (2020). *Choosing a Real Estate Agent: An Evaluation of Information Sources about Quality of Service*. Report, Consumer Federation of America, Washington, DC.

— (2023). *A Surfeit of Real Estate Agents 2: Is Entry Too Easy?* Report, Consumer Federation of America, Washington, DC.

BUCHAK, G., MATVOS, G., PISKORSKI, T. and SERU, A. (2024). Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy. *Journal of Political Economy*, **132** (2), 616–693.

—, —, — and — (2025). NAR settlement, house prices, and consumer welfare. *NBER Working Paper 32855*.

CHEN, J., GU, J. and KWON, S. (2025). Empirical Bayes shrinkage (mostly) does not correct the measurement error in regression. *arXiv preprint arXiv:2503.19095*.

COOK, C., LI, P. Z. and BINDER, A. J. (2025). Where to Build Affordable Housing? Evaluating the Tradeoffs of Location. *Working Paper*.

CUNNINGHAM, C. R., GERARDI, K. and SHEN, L. (2022). The good, the bad, and the ordinary: Estimating agent value-added using real estate transactions. *Working Paper*.

DAGHER, V. (2023). Almost No One Pays a 6% Real-Estate Commission—Except Americans. *The Wall Street Journal*, https://www.wsj.com/personal-finance/real-estate-buying-home-charts-6dc40caa.

DEFUSCO, A. A., NATHANSON, C. G. and ZWICK, E. (2022). Speculative dynamics of prices and volume. *Journal of Financial Economics*, **146** (1), 205–229.

DRANOVE, D. and JIN, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, **48** (4), 935–963.

DUBOIS, P., GRIFFITH, R. and O'CONNELL, M. (2018). The effects of banning advertising in junk food markets. *Review of Economic Studies*, **85** (1), 396–436.

DYNASTY FINANCIAL PARTNERS (2023). Surprising Due Diligence Gap Uncovered: Most High-Net-Worth Investors Hire First Advisor They Speak With, According to New Dynasty Connect Survey. Https://www.businesswire.com/news/home/20230712308488/en/Surprising-Due-Diligence-Gap-Uncovered-Most-High-Net-Worth-Investors-Hire-First-Advisor-They-Speak-With-According-to-New-Dynasty-Connect-Survey.

EGAN, M. (2019). Brokers versus retail investors: Conflicting interests and dominated products. *Journal of Finance*, **74** (3), 1217–1260.

—, MATVOS, G. and SERU, A. (2019). The market for financial adviser misconduct. *Journal of Political Economy*, **127** (1), 233–295.

—, — and — (2022). When harry fired sally: The double standard in punishing misconduct. *Journal of Political Economy*, **130** (5), 1184–1248.

—, — and — (2024). The Problem of Good Conduct among Financial Advisers. *Journal of Economic Perspectives*, **38** (4), 193–210.

—, — and — (2026a). Understanding financial advisor misconduct. *Annual Review of Financial Economics*, **18**.

—, —, — and YANG, H. (2026b). What do certifications certify? *Working Paper*.

EINAV, L., FINKELSTEIN, A. and MAHONEY, N. (2025). Producing Health: Measuring Value Added of Nursing Homes. *Econometrica*, **93** (4), 1225–1264.

FARRONATO, C., FRADKIN, A., LARSEN, B. J. and BRYNJOLFSSON, E. (2024). Consumer protection in an online world: An analysis of occupational licensing. *American Economic Journal: Applied Economics*, **16** (3), 549–579.

FENIZIA, A. (2022). Managers and productivity in the public sector. *Econometrica*, **90** (3), 1063–1084.

FTC (1983). *The Residential Real Estate Brokerage Industry*. Tech. rep., Federal Trade Commission, Washington, DC, https://www.ftc.gov/system/files/ftc_gov/pdf/The-Residential-Real-Estate-Brokerage-Report–Butters-Report.pdf.

GILBUKH, S. (2026). More Experienced Intermediaries – More Principal-Agent Problems? *Working Paper*.

— and GOLDSMITH-PINKHAM, P. (2024). Heterogeneous real estate agents and the housing cycle. *Review of Financial Studies*, **37** (11), 3431–3489.

GRENNAN, M., MYERS, K. R., SWANSON, A. and CHATTERJI, A. (2025). No free lunch? Welfare analysis of firms selling through expert intermediaries. *Review of Economic Studies*, **92** (4), 2537–2577.

HAN, B. (2024). NAR's commission lawsuit settlement receives final approval. *HousingWire*, https://www.housingwire.com/articles/nar-commission-lawsuit-settlement-approved-doj-homeservices-keller-williams-3/.

HASTINGS, J. S., VAN WEELDEN, R. and WEINSTEIN, J. M. (2007). Preferences, information, and parental choice behavior in public school choice. *NBER Working Paper No. 12995*.

HENDEL, I., NEVO, A. and ORTALO-MAGNÉ, F. (2009). The relative performance of real estate marketing platforms: MLS versus FSBOMadison.com. *American Economic Review*, **99** (5), 1878–1898.

HONIGSBERG, C., HU, E. and JACKSON JR, R. J. (2025). Regulatory leakage among financial advisors: Evidence from finra regulation of "bad" brokers. *Journal of Financial Economics*, **174**, 104170.

HSIEH, C.-T. and MORETTI, E. (2003). Can free entry be inefficient? Fixed commissions and social waste in the real estate industry. *Journal of Political Economy*, **111** (5), 1076–1122.

HUANG, Z. (2025). The Redistributive and Efficiency Effects of Property Taxation. *Working Paper*.

INDURSKY, J. (2025). NAR Drops Membership Requirement for MLS Access. https://therealdeal.com/national/2025/11/18/nar-drops-membership-requirement-for-mls-access/.

JAMSHIDIAN, M. and JENNRICH, R. I. (1997). Acceleration of the EM algorithm by using quasi-Newton methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59** (3), 569–587.

KENNEY, J. (2026). What's the Deal with Brokers? Evidence from Regulatory Change in the U.S. Real Estate Market. *Working Paper*.

KEPLER, J. D., NIKOLAEV, V. V., SCOTT-HEARN, N. and STEWART, C. R. (2026). Quality transparency and healthcare competition. *Working Paper*.

KIM, G. H. (2025). The Equilibrium Impacts of Broker Incentives in the Real Estate Market. *Working paper*.

KLEINER, M. M. and SOLTAS, E. J. (2023). A welfare analysis of occupational licensing in US states. *Review of Economic Studies*, **90** (5), 2481–2516.

KLING, J. R., MULLAINATHAN, S., SHAFIR, E., VERMEULEN, L. C. and WROBEL, M. V. (2012). Comparison friction: Experimental evidence from Medicare drug plans. *Quarterly Journal of Economics*, **127** (1), 199–235.

KOWALESKI, Z. T., SUTHERLAND, A. G. and VETTER, F. W. (2020). Can ethics be taught? Evidence from securities exams and investment adviser misconduct. *Journal of Financial Economics*, **138** (1), 159–175.

LEVITT, S. D. and SYVERSON, C. (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics*, **90** (4), 599–611.

LIMODIO, N. (2021). Bureaucrat allocation in the public sector: Evidence from the World Bank. *The Economic Journal*, **131** (639), 3012–3040.

MIRAVETE, E., SEIM, K. and THURK, J. (2026). Elasticity and Curvature of Discrete Choice Demand Models. *Working Paper*.

MUÑOZ, P. and PREM, M. (2024). Managers' productivity and recruitment in the public sector. *American Economic Journal: Economic Policy*, **16** (4), 223–253.

NAR (2025). 2025 Home Buyers and Sellers Generational Trends Report. *National Association of REALTORS*.

OLICK, D. (2026). Existing Home Sales End 2025 with a Strong Beat, as Prices Ease Further. https://www.cnbc.com/2026/01/14/existing-home-sales-end-2025-stronger-than-expected.html.

PETRIN, A. and TRAIN, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, **47** (1), 3–13.

PIAZZESI, M., SCHNEIDER, M. and STROEBEL, J. (2020). Segmented housing search. *American Economic Review*, **110** (3), 720–759.

REUTER, J. and SCHOAR, A. (2024). Demand-side and supply-side constraints in the market for financial advice. *Annual Review of Financial Economics*, **16**, 391–411.

ROBLES-GARCÍA, C. (2020). Competition and Incentives in Mortgage Markets: The Role of Brokers. *Working Paper*.

RUTHERFORD, R. C., SPRINGER, T. M. and YAVAS, A. (2005). Conflicts between principals and agents: evidence from residential brokerage. *Journal of Financial Economics*, **76** (3), 627–665.

SAIZ, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics*, **125** (3), 1253–1296.

SALZ, T. (2022). Intermediation and competition in search markets: An empirical case study. *Journal of Political Economy*, **130** (2), 310–345.

SIRMANS, S., MACPHERSON, D. and ZIETZ, E. (2005). The composition of hedonic pricing models. *Journal of Real Estate Literature*, **13** (1), 1–44.

SMALL, K. A. and ROSEN, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica*, pp. 105–130.

SWANEPOEL, S. (2015). *D.A.N.G.E.R. Report: Definitive Analysis of Negative Game Changers Emerging in Real Estate*. Industry report, National Association of REALTORS, https://www.dangerreport.com/danger-report.pdf.

TRACEY, M. D. (2025). NAR 2025 Profile of Home Buyers, Sellers Reveals Market Extremes. National Association of REALTORS, REALTOR News, https://www.nar.realtor/news/real-estate-news/nar-2025-profile-of-home-buyers-sellers-reveals-market-extremes.

TRAIN, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edn.

U.S. CENSUS BUREAU AND U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT (2026). New Privately-Owned Houses Sold. https://www.census.gov/construction/nrs/data/series.html.

VATTER, B. (2025). Quality disclosure and regulation: Scoring design in Medicare Advantage. *Econometrica*, **93** (3), 959–1001.

WHITE, N. (2026). For real estate to be a true profession, we must first raise the bar. *Real Estate News*, https://www.realestatenews.com/2026/06/20/for-real-estate-to-be-a-true-profession-we-must-first-raise-the-bar.

ZHAO, L. (2023). Out of sight, out of mind: Search frictions and financial adviser misconduct. *Working Paper*.

Tables and Figures

Table 1: Controls included in value added first stage estimation

Variable	Description
Close/List ratio	Ratio between the transaction's closing and listing prices.
Sq. Ft.	Living area of the property.
# Bedrooms	Number of bedrooms of the property.
# Bathrooms	Number of bathrooms of the property.
View	Indicator taking the value 1 if the property has a "quality view" (Cunningham *et al.*, 2022): e.g. city view or mountain view.
Gated Community	Indicator taking the value 1 if the property is in a gated community.
New Construction	Indicator taking the value 1 if the listing corresponds to a new construction.
Waterfront	Indicator taking the value 1 if the property is located on a waterfront, such as a lake.
Pool	Indicator taking the value 1 if the property has a pool.
Need renovation	Indicator taking the value 1 if the property is in need of a renovation, as defined in Appendix C.2.
Post renovation	Indicator taking the value 1 if the property has undertaken major renovation prior to listing, as defined in Appendix C.2.
Green energy investments	Indicator taking the value 1 if the property features any green energy investment.
Probate	Indicator taking the value 1 if the listing corresponds to a probate sale.
# Yearly transactions	Number of transactions completed by the agent in the year prior to the listing close date.
Same agent	Indicator taking the value 1 if the seller and buyer agent are the same.
Same brokerage	Indicator taking the value 1 if seller and buyer agent work for the same brokerage company.

Table 2: Value Added First Stage Estimates

	(Close Price/List Price) x 100 (1)
ln(Sq. Ft.)	-0.6305**
	(0.2919)
# Bedrooms	0.0368
	(0.0452)
# Bathrooms	0.1371*
	(0.0815)
View	-0.0658*
	(0.0339)
Gated Community	-0.0871
	(0.0624)
New Construction	0.0709
	(0.0804)
Waterfront	0.0980
	(0.2441)
Pool	0.1074
	(0.1089)
Need renovation	0.1585*
	(0.0914)
Post renovation	0.2209
	(0.1409)
Green energy investments	-0.0278
	(0.0250)
Probate	-0.6914***
	(0.1863)
Same agent	-0.5977***
	(0.0572)
Same brokerage	-0.0506
	(0.0395)
ln(1 + # Yearly transactions) (Seller, total)	0.0189
	(0.0156)
ln(1 + # Yearly transactions) (Buyer, total)	0.0300**
	(0.0150)
ln(1 + # Yearly transactions) (Seller, zipcode)	-0.0156
	(0.0151)
ln(1 + # Yearly transactions) (Buyer, zipcode)	-0.0346*
	(0.0189)
ln(1 + # Yearly transactions) (Seller-Buyer pair)	0.1563***
	(0.0347)
Property FE	✓
Seller Agent FE	✓
Buyer Agent FE	✓
Observations	451,557
R^2	0.5645
Dependent variable mean	98.3601

Note: This table shows the result of the first stage regression to estimate the prior pricing ability terms. The regression model is presented in equation (6). Standard errors are clustered at the zipcode-listing year/month level.

Table 3: Property choice model: Control function estimates

	ln(List Price)	Buyer agent commission rate
	(1)	(2)
Price instrument	0.154***	0.004
	(0.008)	(0.004)
Commission rate instrument	-0.006	-0.052***
	(0.004)	(0.005)
Property covariates	✓	✓
Zipcode-Listing Year/Month FE	✓	✓
Observations	133,659	133,659
R^2	0.884	0.098
Dependent variable mean	5.412	2.908
F-test of joint significance of instruments	183.27	46.36

Note: This table shows the results from a regression to estimate the impact of the instruments on the endogenous variables of the property choice model. The price and commission rate instruments are defined in sections 6.2.1 and Appendix E.1. Standard errors are clustered at the zipcode-listing year level.

Table 4: Property choice model: Parameter estimates

Variable	Estimate	Std. Err.
Price	-0.027	(0.001)
ln(Sq. Ft.)	0.368	(0.051)
# Bedrooms (Income Q1, Single)	-0.058	(0.014)
# Bedrooms (Income Q2, Single)	0.050	(0.016)
# Bedrooms (Income Q3, Single)	0.145	(0.020)
# Bedrooms (Income Q4, Single)	0.168	(0.025)
# Bedrooms (Income Q1, Married)	0.008	(0.020)
# Bedrooms (Income Q2, Married)	0.114	(0.015)
# Bedrooms (Income Q3, Married)	0.198	(0.017)
# Bedrooms (Income Q4, Married)	0.186	(0.020)
New Construction	-0.403	(0.028)
View	0.039	(0.017)
Waterfront	-0.061	(0.051)
Gated Community	-0.110	(0.019)
Pool	0.357	(0.187)
Spa	0.017	(0.014)
Needs Renovation	-0.280	(0.036)
Probate	-0.227	(0.058)
Green Energy Efficient	0.030	(0.009)
Price CF Residual	0.781	(0.057)
Commission Rate CF Residual	-0.396	(0.032)
Commission Income (Composite agent group)	0.239	(0.015)
Commission Income (Q1)	0.292	(0.016)
Commission Income (Q2)	0.261	(0.016)
Commission Income (Q3)	0.193	(0.015)
Commission Income (Q4)	0.041	(0.016)
1(Same agent)	1.477	(0.042)
1(Same brokerage)	0.531	(0.016)
ln(1 + # Yearly transactions) (Seller)	-0.022	(0.005)
ln(1 + # Yearly transactions) (Seller-Buyer Pair)	0.146	(0.039)
Number of choice events	66,203	

Note: This table shows the parameter estimates and standard errors of the property demand model in equation (10). Standard errors are clustered at the zipcode-listing year level.

Table 5: Agent choice model: Control function estimates

	Advice value
	(1)
Advice value instrument	1.152***
	(0.015)
Property covariates	✓
Buyer Agent FE	✓
Observations	23,058,932
R^2	0.929
Dependent variable mean	473.720
F-test of significance of instruments	5981.18

Note: This table shows the results from a regression to estimate the impact of the instruments on the endogenous variables of the agent choice model. The advice value instrument is defined in sections 7.3. Standard errors are clustered at the zipcode-closing year level.

Table 6: Agent choice model: Parameter estimates

Variable	Estimate	Std. Err.
Advice Value	0.0071	(0.0004)
Advice Value Residual	-0.0081	(0.0003)
ln(1 + # Transactions), zipcode	0.2355	(0.0092)
ln(1 + # Transactions), total	-0.0437	(0.0056)
Agent fixed effects	✓	
Number of choice events	66,203	

Note: This table shows the parameter estimates and standard errors of the agent demand model in equation (14). Standard errors are clustered at the buyer agent-year level.

Table 7: Comparative statics: Increases to buyer incentive alignment weights

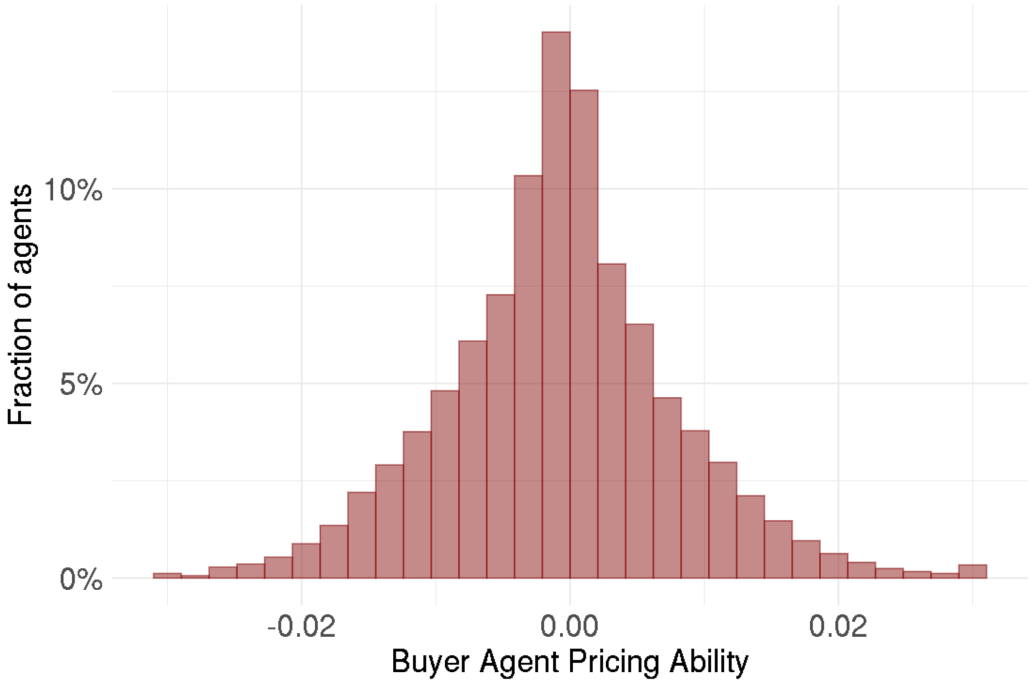
	Scenario 1	Scenario 2	Scenario 3
Δ CS / avg. sale price	0.27%	0.97%	2.52%

Note: This table shows the effects on consumer surplus, as a percentage of average sale prices, of increasing weights on buyer utility of the lower-weight groups. Consumer surplus is defined as welfare derived from housing consumption, computed using the expression in equation (16). Weight parameters are defined by the heterogeneous parameters on commission income estimated in Table 4. Scenario 1 sets buyer weights for Q1 agents to the level of Q2 agents. Scenario 2 sets buyer weights for Q1 and Q2 agents to the level of the composite agent group. Scenario 3 sets buyer weights for Q1, Q2, and the composite group agents to the level of Q3 agents.

Figure 1: Distribution of pricing ability estimates



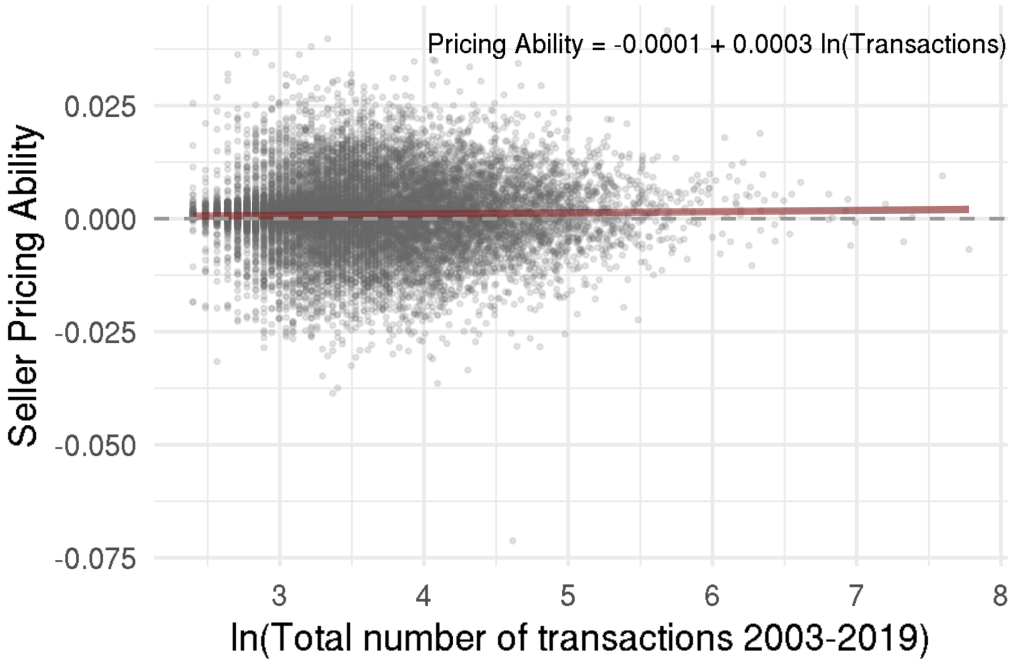
(a) Seller agents



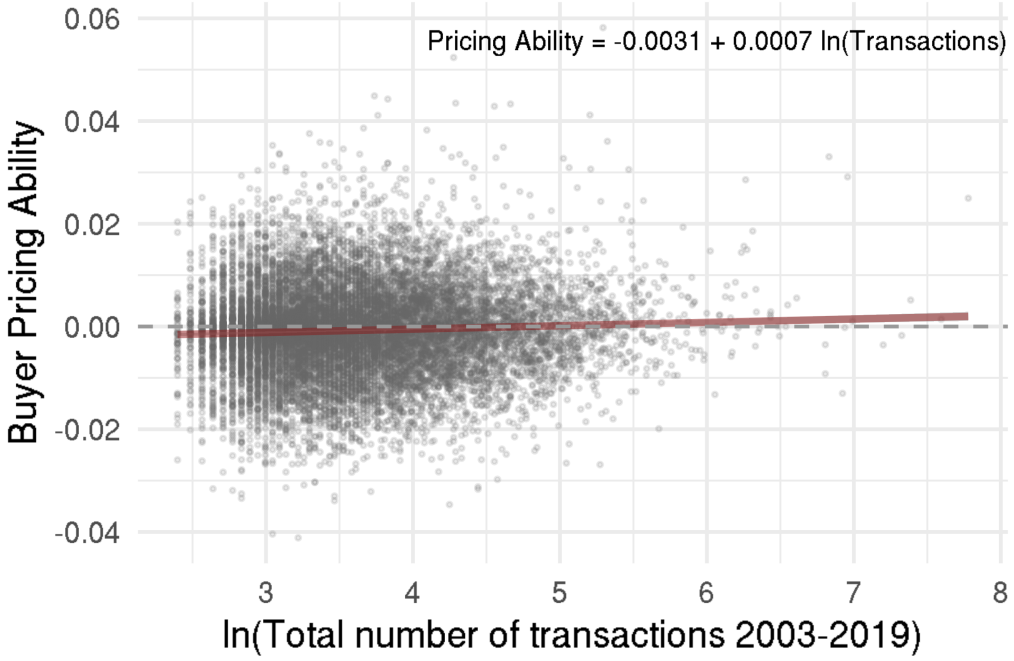
(b) Buyer agents

Note: This figure shows agent-level pricing ability estimates, estimated using the methodology in Section 5.1 and after an Empirical Bayes adjustment.

Figure 2: Relationship between pricing ability and market activity



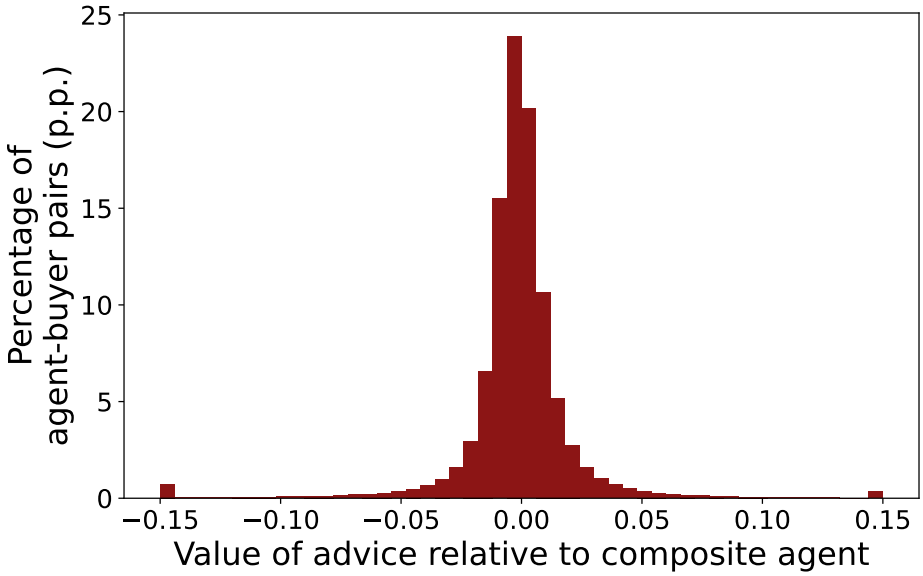
(a) Seller agents



(b) Buyer agents

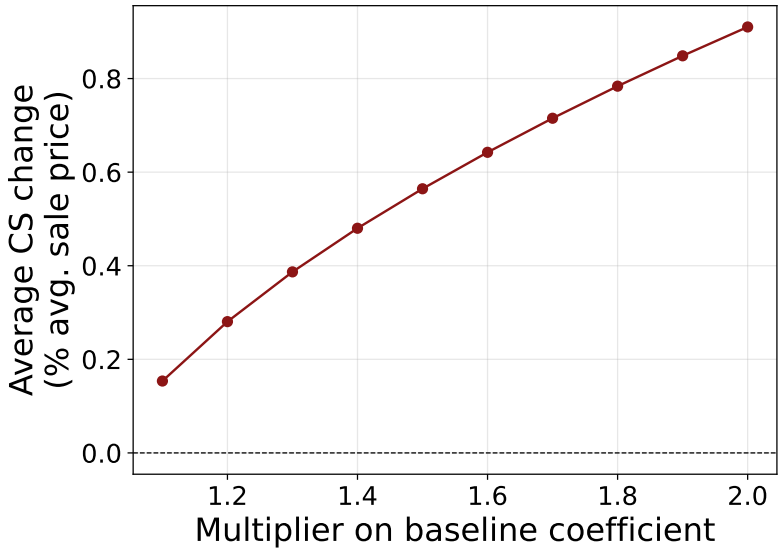
Note: This figure shows the relationship between log total number of transactions and pricing ability estimates, at the agent level. Each dot represents one agent. The line is the best linear fit.

Figure 3: Distribution of agent advice values



Note: This figure shows the distribution of the estimated advice values, computed after estimating the property choice model following equation (13). Each observation corresponds to an agent-buyer pair. Advice values are measured in percentage terms relative to the advice value of the composite agent for the corresponding buyer.

Figure 4: Comparative statics: Increases to demand sensitivity parameter



Note: This figure shows the effects on consumer surplus, as a percentage of average sale prices, of increasing the demand sensitivity parameter on the value of agent advice. Consumer surplus is defined as welfare derived from housing consumption, computed using the expression in equation (16).

Appendix

A Model Appendix

A.1 Proof of inclusive value formula

The proof follows Dubois *et al.* (2018). The utility accrued to the buyer in the (rescaled) joint decision problem is

$$-\alpha p + \beta X + \epsilon,$$

which is equal to

$$V - \tilde{\theta}\pi.$$

Expected utility from product choice is (Train, 2009)

$$\mathbb{E}[\max_j\{-\alpha p + \beta X + \epsilon\}] = \mathbb{E}[\max_j V] - \tilde{\theta}(\pi + \gamma X') \tag{18}$$

By Small and Rosen (1981), and transforming to monetary units by dividing by α, it follows that

$$I_{ia} = \frac{1}{\alpha}\ln\left(\sum_k \exp(V_{iak})\right) - \frac{1}{\alpha}\sum_k s_{iak}\tilde{\theta}_a\pi_k \tag{19}$$

which completes the proof.

A.2 Impact of welfare weights on product recommendations

Let $V^0_{iaj} = (1-\theta_a)(-\alpha p_j + \beta X_j + \epsilon_{iaj})$ be the value derived exclusively from buyer preferences. By the same arguments as in appendix A.1, expected surplus from buyer preferences is $\ln(\sum_j \exp V^0_j)$. Now consider that we add in agent non-pecuniary terms $(1-\theta)\gamma X'_{aj}$ for only one choice j^*. This implies that the expected surplus becomes $\ln(\sum_{j\neq j^*} \exp V^0_j + \exp V'_{j*})$, where $V'_{iaj} = (1-\theta_a)(-\alpha p_j + \beta X_j + \gamma X'_{aj} + \epsilon_{iaj})$. Since the derivative of $\ln(\sum_j \exp(x_j))$ is

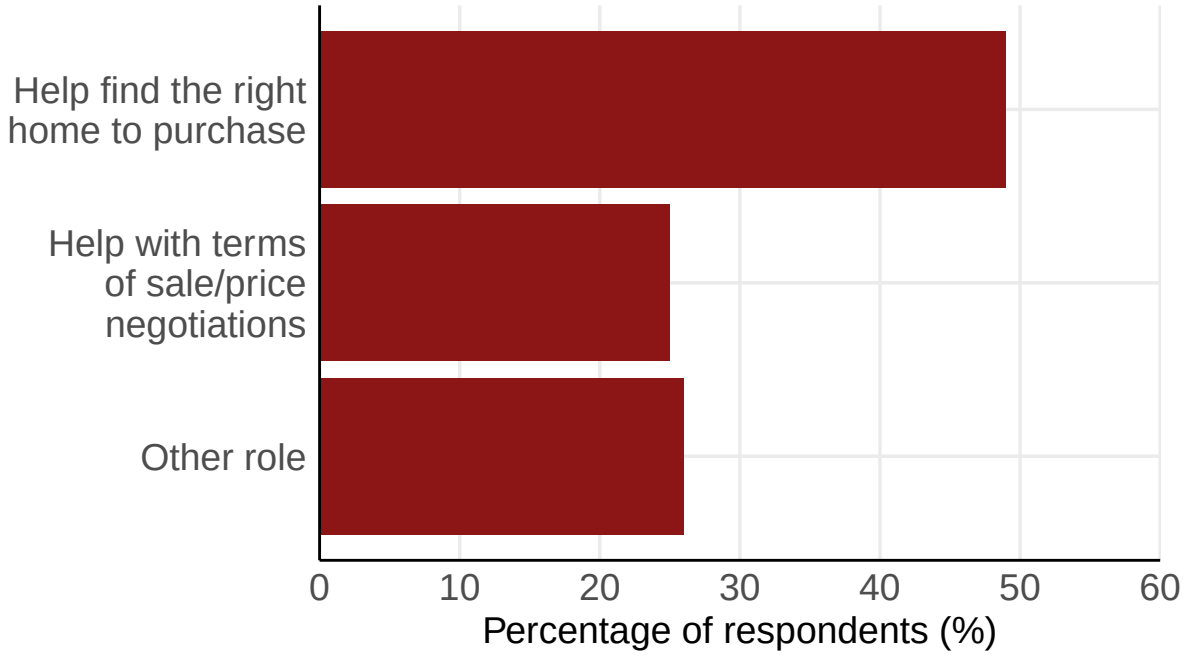
$$\frac{\exp(x_j)}{\ln(\sum_j \exp(x_j))} = s_j,$$

a first-order approximation of the impact of non-pecuniary incentives is $s^0_j(1-\theta)X'_j$. This implies that, when $\theta < 1$, consumers benefit from the private impact of agent recommendations when such impact X'_j is correlated with buyers' preferences for properties, given by the buyer-specific choice probabilities s^0_j.

B Institutional Setting Appendix

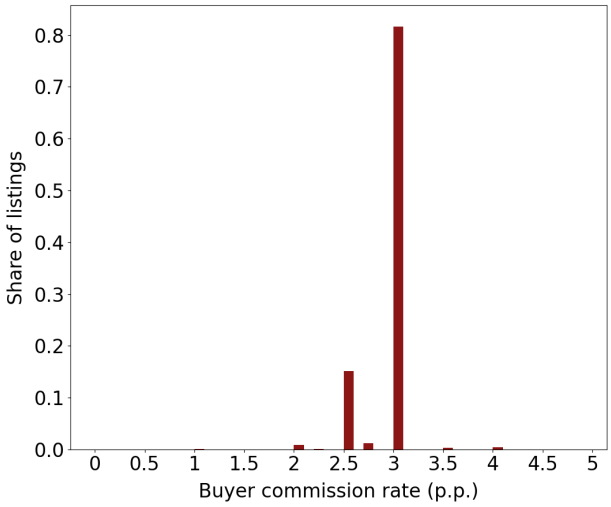
B.1 Agent roles most valued by home buyers

Figure A1: Agent roles most valued by home buyers

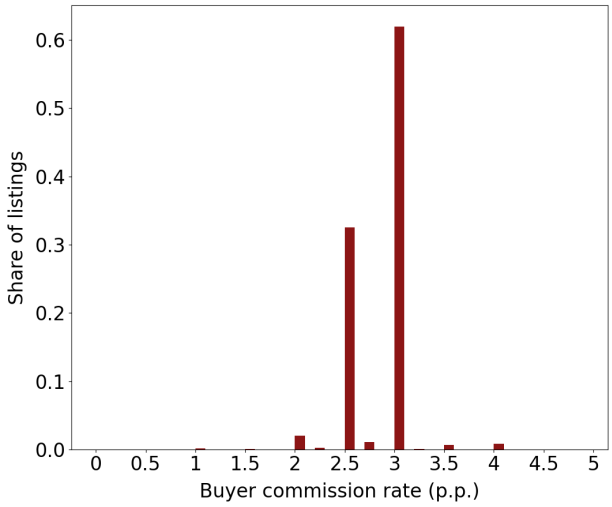


Note: This figure shows the roles most valued by home buyers about real estate agents, according to a 2025 survey by the National Association of Realtors (NAR, 2025).

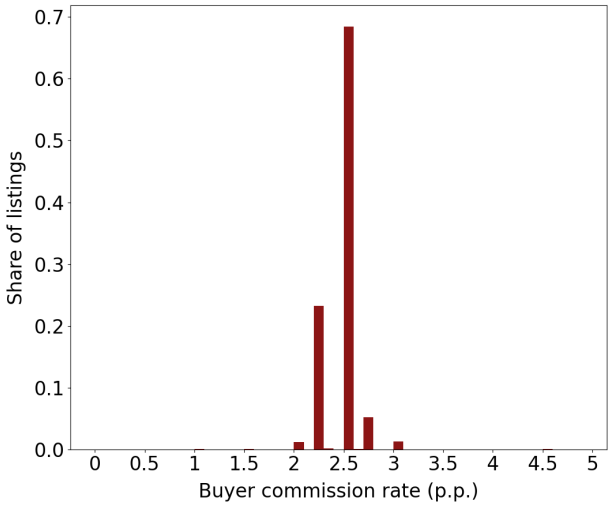
B.2 Distribution of commission rates in select metro areas



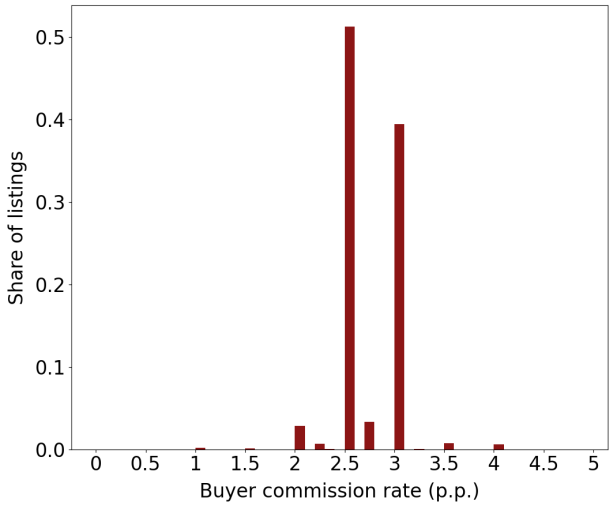
(a) Phoenix



(b) Las Vegas



(c) Portland, OR



(d) Tampa

Figure A2: Distribution of commission rates offered, 2014-2018

Note: This figure shows the distribution of commission rates offered to buyer agents in four metropolitan areas between 2014 and 2018. Each observation is a listing. The vertical axis indicates the shared of posted listings offering each commission rate level.

C Data Appendix

C.1 Property Transactions Data Cleanup

I apply a series of standard data cleaning steps. Following Cunningham *et al.* (2022), I focus my analysis on single-family, agent-intermediated sales. This excludes foreclosures, short sales, and real estate-owned and bank-owned sales. I restrict to agent-intermediated transactions, excluding sales with missing agent IDs. I further exclude listings without an Assessor's Parcel Number (APN) or a county code.

I then exclude listings with extreme prices or property attributes. Following Cunningham *et al.* (2022), I restrict to listings with the following characteristics: square footage between 800 and 5000 square feet, properties with at least one bathroom and at most 10 bathrooms and bedrooms, at most 10 acres of lot size, listing price between $20,000 and $4,000,000, and between 1 and 500 days listed. I also focus on percentage-denominated commissions, excluding flat-fee commissions. I also exclude listings with negative commission entries or buyer-side commissions greater than 5%.

I have access to two MLS datasets provided by Corelogic: Quick Search (QS) and Quick History (QH). Broadly speaking, the QH dataset includes one observation for every modification of any listing variable. This includes list price as well as updated information on property attributes or description in the public remarks field. The QS dataset includes, for each listing, the information corresponding to the latest update. My main source of information is the QS dataset. I use the QH dataset to obtain information on listing-level attributes that might potentially be missing in some QS entries. Occasionally, some listings could feature in the QH dataset but not in the QS one. I include such listings in my main analysis. For each dataset, I separately apply the sample filters described above.

C.2 Property Transactions Variables Construction

- *AgentOwned*: Indicator that equals 1 if the field Disclosures contains any of the following (case-insensitive substring match): `Agent Owned`, `Agent/Owner`, `Broker-Owner`, `Broker/`, `Owner Agent`, `Owner Broker`, `Owner Has R.E. License`, `Owner Has Real Estate License`, `Owner Is A Broker`, `Owner Is A Licensed Broker`, `Owner Is A Licensed Real Estate Agent`, `Owner Is A Licensed Real Estate Broker`, `Owner Is A Licensed Realtor`, `Owner Is A Real Estate Agent`, `Owner Is A Real Estate Broker`, `Owner Is A Realtor`, `Owner Is Agent`, `Owner Is An Agent`, `Owner Is Broker`, `Owner Is Lic Real Est Agt`, `Owner Is Licensed`

Agent,Owner Is Licensed Broker,Owner Is Licensed RE Broker,Owner Is Licensed Real Estate Agent,Owner Is Licensed Real Estate Broker,Owner Is Licensed Realtor,Owner Is List Agent,Owner Is Listing Agent,Owner Is Listing Broker, Owner Is Real Estate Agent, Owner Is Real Estate Broker,Owner Is Realtor,Owner Is Related To Agent,Owner Is Related To Broker, Owner Is The Broker, Owner Is The Listing Agent, Owner Is The Listing Broker,Owner Related To Agent,Owner Related To Broker, Owner-Agent, Owner-Broker, Owner/Agent, Owner/Broker, Owner / Broker, Owner/Realtor,Ownership Interest

- *ToRenew*: Indicator that takes the value 1 if the listing needs significant renewal. Specifically, *ToRenew* equals one: if the field PublicRemarks contains any of the following: "as is"; as is condition; as is, where is; as-is; do it yourself; need major repair; needs major repair; sold as is, or the field SpecialListingConditions contains any of the following: As Is; Fix Up.
- *PostRenovation*: Indicator that equals 1 if the field PublicRemarks contains any of the following: complete renovation; completely renovated; new renovation; newly renovated; revamped.
- *Probate*: Indicator that equals 1 if the field SpecialListingConditions contains the term Probate.

C.3 Corelogic-HMDA merge

To match mortgages between Corelogic and HMDA, I follow the standard procedure in the literature (Bayer *et al.*, 2025; Huang, 2025). I use the following loan attributes in the HMDA data as matching variables: census tract, year of application, loan purpose (home purchase or refinance), loan type (conventional, FHA, or VA loans), loan amount, and lender name. I construct equivalent variables in the Corelogic data using the same definitions as HMDA. This includes using the same loan amount rounding criteria as in HMDA (applying the correct definitions up to 2017 and from 2018 onwards), and same variable definitions. Following Huang (2025), to match lender names I only use the first word in the name.

Following Bayer *et al.* (2025) and Huang (2025), my matching strategy consists of four rounds. In the first round, I merge the datasets on the six matching variables. I store all mortgages in Corelogic with a unique match in the HMDA database, and I discard all Corelogic mortgages that match to more than one HMDA mortgage. I keep all unmatched mortgages to keep matching sequentially on a smaller set of variables.

In the second step, I match on census tract, year of application, loan amount, and first word of lender name. In the third round, I match on census tract, year of application, loan purpose, loan type, and loan amount. In the fourth step, I match on census tract, year of application, and loan amount.

I apply this process separately for the dataset of mortgages up to 2014 (for the population instrument) and the mortgages post 2014 (for demand estimation). The process yields total match rates of 40.7% and 50.9%, respectively. Out of these, 73.5% and 85.4% occur in the first step of the matching process.

D Value Added Estimation Appendix

D.1 Likelihood Function Derivation

By properties of ordinary least squares, $\hat{\phi}_j^{OLS} \sim \mathcal{N}(\phi_j, \hat{\Sigma}_j)$. If ϕ_j was known, one could use the distribution of $\hat{\phi}_j^{OLS}$ to write the likelihood function of Ψ given ϕ_j:

$$\mathcal{L}(\Psi|\phi) = \prod_{i=1}^{N} f(\phi|\Psi) = \prod_{i=1}^{N} \left[\sum_{c=1}^{2} \pi_c \mathcal{N}(\phi_j|\eta_c, \Delta_c) \right] \tag{20}$$

However, since ϕ_j is not known, we must construct the likelihood function using $\hat{\phi}_j^{OLS}$, whose distribution depends on the true ϕ_j. Since ϕ_j is unknown, we must integrate over its possible values according to the distribution $f(\phi|\Psi) = \sum_{c=1}^{2} \pi_c \mathcal{N}(\phi_j|\eta_c, \Delta_c)$:

$$\mathcal{L}(\Psi) = \prod_{i=1}^{N} \int \mathcal{N}(\hat{\phi}|\phi, \hat{\Sigma}) \left[\sum_{c=1}^{2} \pi_c \mathcal{N}(\phi_j|\eta_c, \Delta_c) \right] d\phi_j \tag{21}$$

with log-likelihood

$$\ell(\Psi) = \sum_{j=1}^{J} \ln \left[\int N(\hat{\phi}_j|\phi_j, \hat{\Sigma}_j) \left(\sum_{c=1}^{2} \pi_c \mathcal{N}(\phi_j|\eta_c, \Delta_c) \right) d\eta_j \right] = \tag{22}$$

Interchanging the integral and summation:

$$\ell(\Psi) = \sum_{j=1}^{J} \ln \left[\sum_{c=1}^{2} \pi_c \int \mathcal{N}(\hat{\phi}_j|\phi_j, \hat{\Sigma}_j) \mathcal{N}(\phi_j|\eta_c, \Delta_c) d\phi_j \right] \tag{23}$$

Now, since $\hat{\phi}_j \sim \mathcal{N}(\phi_j, \hat{\Sigma}_j)$, we can express $\hat{\phi}_j$ as $\hat{\phi}_j = \phi_j + \epsilon_j$ with $\epsilon_j \sim \mathcal{N}(0, \hat{\Sigma}_j)$. Hence,

$$\ell(\Psi) = \sum_{j=1}^{J} \ln \left[\sum_{c=1}^{2} \pi_c \int \mathcal{N}(\phi_j + \epsilon_j|\phi_j, \hat{\Sigma}_j) \mathcal{N}(\phi_j|\eta_c, \Delta_c) d\phi_j \right] \tag{24}$$

Now, since ϵ_j is symmetric at 0, $\mathcal{N}(\phi_j + \epsilon_j|\phi_j, \hat{\Sigma}_j) = mathcal N(\phi_j - \epsilon_j|\phi_j, \hat{\Sigma}_j)$. Hence,

$$\ell(\Psi) = \sum_{j=1}^{J} \ln \left[\sum_{c=1}^{2} \pi_c \int \mathcal{N}(\phi_j - \epsilon_j | \phi_j, \hat{\Sigma}_j) \mathcal{N}(\phi_j | \eta_c, \Delta_c) d\phi_j \right] \tag{25}$$

By definition, the term inside the integral is a convolution. Moreover, since, conditional on ϕ_j, the realization of $\hat{\phi}_j$ is independent of ϕ_j, it is a convolution of independent variables. By properties of convolutions, the convolution of two independent random variables is equal to the sum of its densities. Hence,

$$\int \mathcal{N}(\hat{\phi}_j | \phi_j, \hat{\Sigma}_j) \mathcal{N}(\phi_j | \eta_c, \Delta_c) d\phi_j = \mathcal{N}(\hat{\phi}_j | \eta_c, \Delta_c + \hat{\Sigma}_j) \tag{26}$$

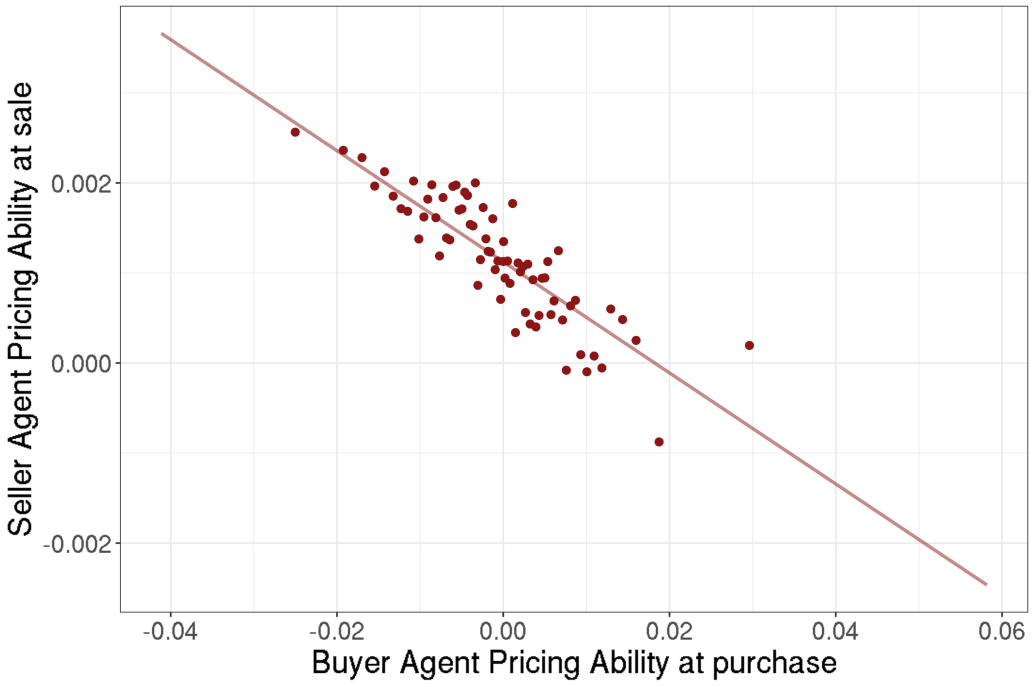
Plugging equation (26) in (27):

$$\ell(\Psi) = \sum_{j=1}^{J} \ln \left[\sum_{c=1}^{2} \pi_c \mathcal{N}(\hat{\phi}_j | \eta_c, \Delta_c + \hat{\Sigma}_j) \right] \tag{27}$$

which completes the proof.

D.2 Correlation between client-specific agent effects

Figure A3: Correlation between client-specific agent effects



Note: This figure shows the correlation between the pricing ability of the buyer agent when a property was purchased, and the pricing ability of the seller agent when the same property was sold. Each dot represents the average estimates within one of 100 equal-sized bins of repeat sales events. The line is the best linear fit estimated on the underlying repeat sales data.

E Demand Estimation Appendix

E.1 Price Instrument Construction

To construct my price instrument, I begin by estimating baseline population levels in Phoenix per demographic group N_d, in the year before my property choice model data starts (2015), by combining the ACS-1 and ACS-5 datasets. From the 2010-2014 ACS-5 dataset, I obtain the number of owner-occupied households across zipcodes by race, marital status, (binned) income, age group, and presence of children. I define a demographic group d as a unique combination of these five characteristics. For these characteristics, the ACS-5 typically reports characteristic-specific marginal distributions, but not their joint distribution. For example, the data contains the total number of single owner-occupied households, but does not decompose this number by income, for example. I estimate these joint distributions using data for Phoenix residents in the 2014 ACS-1 dataset. I then combine these joint distributions with the population counts from the ACS-5 to estimate the number of owner-occupied households across demographic groups in 2014 per zipcode. I further construct growth rates per demographic group and year, g_{dt}, from the ACS-1 datasets.

I construct population baseline shares per demographic group and property type, $\hat{P}_{jd}$, by combining the Corelogic and HMDA datasets. I estimate the probability of observing each demographic group in each property type from home purchases between 2009 and 2014. To ensure I have the most comprehensive measure of the population possible, I use both on- and off-MLS transactions. I start with the sample of all mortgages issued for home purchases in Phoenix between 2009 and 2014 from the Corelogic Mortgages dataset. I link these mortgages with property characteristics in the Corelogic Historical Property dataset. This dataset has a coarser set of property attributes than the MLS data. I define property type groups j as a triplet of zipcode, decade built, and number of bathrooms. I then link these mortgages to HMDA using the procedure described in section C.3. The HMDA dataset does not include information on presence of children, and age groups are only available for 2018 and 2019. To circumvent this, for the prediction analysis I use instead in d the previous zipcode of residence of the buyer, which I obtain from the Corelogic Deeds dataset.

The resulting dataset links mortgage-financed purchase events linking property types j with demographic groups d. Following Cook *et al.* (2025), I estimate a softmax model to predict the likelihood of observing each demographic group in each property type. I restrict the choice set for each purchase event via choice-based sampling, to a 1% random sample of all the available property types (Barwick *et al.*, 2024). This procedure yields predicted market shares $\hat{P}_{jd}$.

After computing the three objects mentioned above, I construct my price instrument for each property type j and demographic group d as

$$z_{jt} = \frac{\sum_d g_{dt} N_d \hat{P}_{jd}}{N_d \hat{P}_{jd}}$$

To compute this instrument, I need to transform the groups from the aggregate definition (using previous zipcode of residence) to the preferred definition (using age groups and presence of children). To do this, I estimate the joint distribution of age groups and presence of children per zipcode using the ACS-1 and ACS-5 datasets. This allows me to compute the expected probabilities of property type choice for each age group-presence of children pair within a zipcode. I then aggregate across all previous zipcodes of residence to express probabilities in age group-presence of children bins.