Selling Fast or Selling Junk: Is iBuying Sustainable?

So Hye Yoon *†

Oct 8, 2025

This is a preliminary draft. Please do not distribute without the permission of the author.

Most recent version of this draft

Abstract

This paper examines challenges in algorithmic intermediation and proposes a framework to mitigate adverse selection when private information about product quality is intertwined with private information about preferences. I examine these issues in the context of iBuyers—firms that offer instant home purchases using big-data-driven pricing models—and analyze why they have struggled to achieve sustainable profitability. I develop a model in which home sellers choose between selling to an iBuyer and listing on the open market based on two dimensions of private information: unobserved house quality and the hassle costs of traditional selling. Sellers may select an iBuyer either to avoid the time and effort of listing or because the iBuyer' s offer exceeds their expected market price, with the latter case generating adverse selection against the iBuyer. Using detailed transaction and listing data, I estimate the joint distribution of these factors, identified from repeated sales and seller choice following iBuyer entry. Counterfactual analyses show that a revenue-sharing contract mitigates adverse selection by improving selection incentives, while incorporating an LLM-based text score derived from unstructured listings further reduces informational frictions by providing a standardized signal of unobserved house quality. Together, these mechanisms enhance the viability of algorithmic transaction markets.

JEL Codes: D82, L15, L85, R31

Keywords: Adverse selection, algorithmic intermediation, housing markets, contract design, artificial intelligence, large language models, iBuyer

*Princeton University, Department of Economics. (email: sohye.yoon@princeton.edu)

†I would like to thank my advisors, Jakub Kastl and Sylvain Chassang for their generous help and support. Comments from Adam Kapor, Bo Honoré, Alessandro Lizzeri, Jeff Gortmaker, Kate Ho, Martin Schmalz, Nicholas Buchholz, Quan Le, Simon Park, Soon Hyeok Choi, and participants at the IO workshop led to further improvements in the manuscript. All remaining errors are mine.

1 Introduction

Housing is a cornerstone of household wealth in the United States, representing approximately 60–70% of the median household' s net worth (U.S. Census Bureau [2022]). Yet, selling a home remains a slow, costly, and uncertain process, limiting liquidity in one of the nation' s largest asset classes. In response, technology-enabled real estate firms such as Opendoor, Offerpad, Zillow Offers, and RedfinNow have adopted the "iBuyer" (instant buyer) model, which streamlines home sales through rapid, data-driven purchase offers, paralleling the rise of algorithmic business models in fintech and insurtech. Firms operating under this model—commonly referred to as iBuyers—make direct offers to homeowners using pricing algorithms, providing a largely online, hassle-free process, which aims to minimize the time and effort required to sell. In exchange for speed and convenience, they purchase homes at a discount and charge service fees comparable to traditional agent commissions.[1]

Despite significant technological capabilities and access to extensive market data, however, many iBuyers have struggled to achieve sustainable profitability, with some discontinuing their programs entirely (RubyHome [2022]). This paper examines the factors underlying these challenges and explores potential paths toward viability for the iBuyer model.

In this paper, I examine why big-data-driven pricing models often fall short in the iBuyer housing market. I focus on two types of seller-side information that iBuyers struggle to price or contract upon using only observable house and market characteristics, both of which also motivate the decision to sell to an iBuyer: (1) unobserved house quality, such as the subjective "mood" or condition of a home that is apparent only through an in-person walkthrough and that, when likely to be valued less in the open market, may prompt sellers to prefer iBuyer offers; and (2) hassle costs, which capture the seller' s personal time, logistical, financial, and emotional burdens associated with the traditional sales process. Both factors are well known to the individual seller but difficult for iBuyers to observe or incorporate into algorithmic pricing. Mismatch in house quality gives rise to adverse selection, since iBuyers ultimately resell homes to individual buyers who can physically inspect them; as a result, sellers of lower-quality homes—even with only moderate hassle costs—are disproportionately more likely to accept iBuyer offers. Using a novel dataset of U.S. housing transactions and listings, I quantify the distribution of these two unobserved factors and propose two strategies

[1]Opendoor, currently the largest active iBuyer, notes in its 10K report (Opendoor Technologies Inc. [2023]) that the discount at which they purchase a home directly affects the percentage of homeowners who accept an offer. Narrower price spreads raise conversion rates, but they also reduce potential profit margins.

to mitigate adverse selection and improve iBuyer performance: (1) a redesigned contract that enables better selection through revenue sharing, and (2) an enhanced pricing model that incorporates unstructured listing text via a large language model.

More broadly, these challenges reflect a general problem in algorithmic intermediation: when private information about product quality and seller preferences jointly determine participation and pricing, data-driven intermediaries may find it difficult to separate the two. This entanglement of information—where quality affects resale value and preferences influence willingness to transact—creates an environment prone to adverse selection, even for intermediaries with extensive data and predictive algorithms.

This paper makes four contributions across distinct literatures. First, I contribute to the empirical Industrial Organization literature on asymmetric information (Cardon and Hendel [2001]; Chiappori and Salanie [2000]; Cohen and Einav [2007]; Finkelstein and McGarry [2006]) by jointly estimating two dimensions of seller-side private information—unobserved house quality and heterogeneous hassle costs—that are not captured by standard pricing inputs. To the best of my knowledge, empirical models capturing more than one dimension of asymmetric information are rare, while my setting requires modeling both dimensions jointly. This is essential because if only unobserved house quality varied, selling to an iBuyer would imply that those sellers systematically have lower-quality homes. If only hassle costs varied, selection into iBuyers would be unrelated to property quality, so iBuyers' resale margins would not systematically differ from the broader market. Observed outcomes suggest both forces are at play, necessitating a joint model of the two dimensions. I show how adverse selection arises when iBuyers cannot observe certain house attributes or account for variation in sellers' willingness to accept discounted offers due to hassle costs.

Second, I propose and evaluate a revenue-sharing contract with a guaranteed minimum that improves selection without simply raising offers, showing that better market design can increase profits. This mechanism parallels the cream-skimming logic of Rothschild and Stiglitz [1976], but adapts it to a multi-dimensional private-information environment in which competition between firms is not explicitly modeled and the indifference condition is otherwise difficult to implement. Relatedly, Caplin et al. [2007] analyze buyer-side adverse selection in housing markets through shared-equity mortgage contracts, while my focus is on seller-side adverse selection and a resale-based revenue-sharing mechanism. The counterfactual contract also shows that improving pricing models alone is insufficient in my sample, and contract design plays a critical role.

Third, I evaluate a counterfactual iBuyer pricing algorithm that incorporates unstructured listing text through a large language model (LLM). While prior work has examined algorithmic pricing in real estate (Fu et al. [2022]; Buchak et al. [2020]; Calder-Wang [2021];

Raymond [2023]) and the use of unstructured data for demand estimation (Compiani et al. [2025]; Quan and Williams [2019]; Lee [2025]), little is known about whether listing text can uncover unobserved quality signals in the housing market. My setting is particularly well-suited for this, as real estate listing descriptions convey nuanced, high-dimensional property features. The LLM approach abstracts from strategic listing descriptions and therefore represents an upper bound on potential gains, as strategic or selective wording may limit how much unobserved quality can be inferred.

Finally, these combined counterfactuals improve the viability of iBuyers in the market, with broader implications for enabling iBuyers to function as financial intermediaries that reduce transaction burdens and enhance liquidity in the housing market. By combining insights from market design and advances in data-driven prediction, the paper illustrates how both market design and data innovation can help mitigate adverse selection in algorithmic transaction platforms. This paper also connects more broadly to research on housing market behavior (Anenberg [2016]; Anenberg and Bayer [2020]; Bayer et al. [2004]; Bayer et al. [2021]).

To address the research questions underlying these contributions, I use a rich dataset covering both the pre- and post-entry periods of iBuyers. The dataset combines CoreLogic Deeds records—which include buyer/seller names, transaction dates, and prices—with CoreLogic MLS records detailing listing dates and house characteristics. By exploiting the legal entity names of buyers, I can systematically identify iBuyer purchases.

Because iBuyers trade price for convenience, I examine both iBuyer offer pricing and the role of hassle costs in homeowners' selling decisions. My estimates show that iBuyers typically purchase homes at an average discount of about 5%, which is broadly consistent with prior findings (Buchak et al. [2020]) and aligns with the pricing incentives described by Opendoor.[2] While this discount may deter some sellers, it attracts those who place a high value on convenience. Indeed, iBuyer marketing emphasizes the "hassle-free" nature of the transaction, and I find supporting empirical evidence: homeowners who relocate to a different state—a move associated with higher logistical complexity—are approximately 1 percentage point more likely to sell to an iBuyer. This effect is economically meaningful, given that iBuyers account for only 2–6% of transactions across cities during the sample period, and about 3.5% in the subsample used for this analysis. The result is consistent with the interpretation that sellers facing higher private hassle costs are more likely to accept iBuyer offers.

Although hassle costs are central to sellers' decisions, private information about house quality is often overlooked in discussions of iBuyer performance. Because iBuyers rely on

[2] Opendoor Technologies Inc. [2023]

algorithmic pricing and conduct limited in-person inspections—typically only after a client accepts an offer (e.g., Opendoor)—they may struggle to assess features that are hard to encode in structured data, such as natural lighting, layout flow, or exposure to ambient noise. These subtle characteristics contribute to the overall attractiveness of a home, which is inherently subjective and difficult to contract upon, yet easily evaluated by individual buyers through direct inspection. I refer to these attributes as "unobserved house quality," since they are visible to individual buyers and sellers in person but not captured by data-driven models.

Hence, following Buchak et al. [2020], I assess the explanatory power of hedonic regression models for iBuyer purchases versus traditional individual-to-individual transactions. I find that the model fits substantially better for iBuyer transactions: the R^2 for the iBuyer regression is seven percentage points higher than that for individual transactions. This implies that structured house characteristics explain more of the variation in iBuyer pricing, whereas individual buyers rely more heavily on unobserved factors. In this sense, iBuyers remain partially "in the dark" about key features of the homes they purchase.

To examine whether unobserved house quality leads to adverse selection, I compare the subsequent resale margins of iBuyers and individual buyers. Across multiple specifications—including both absolute and relative margin definitions, and using the full sample as well as a short-term resale subsample—I consistently find that iBuyers earn lower average margins. The gap is typically around 10–15 percentage points, suggesting that iBuyers systematically overpay for lower-quality homes or face limitations in screening property characteristics that are unobservable at the time of purchase.

To formalize how private information shapes selection into iBuyer sales, I develop a model of seller choice under asymmetric information. I estimate a discrete choice model of sellers to jointly identify the distributions of hassle costs and unobserved house quality. In the model, the iBuyer makes a take-it-or-leave-it offer, and the seller compares it with their expected market payoff, which depends on observable characteristics and private information about home quality. Identification relies on two features: unobserved house quality is house-specific and persistent across sales, while hassle costs are seller–transaction–specific and may vary with timing, liquidity needs, or personal circumstances. I use pre-entry market transactions to estimate the unconditional distribution of unobserved quality through a hedonic pricing model with property-level random effects, exploiting repeated sales of the same home to separate persistent quality from idiosyncratic price shocks. Once iBuyer entry introduces a new selling channel, observed choices between iBuyer and open-market sales reveal the distribution of hassle costs and its correlation with unobserved quality. In general, when both quality and preferences are privately known to the counterparty, and transaction prices are observed in markets with and without selection, prices identify the dimension driving adverse selection, while the discrete choice model identifies the preference distribution.

I jointly estimate the distributions of hassle costs and unobserved house quality, allowing for correlation between them. The hedonic pricing model for open-market transactions includes a property-level random effect capturing unobserved quality, identified from repeated sales that separate persistent quality from idiosyncratic price shocks. The iBuyer' s pricing model, by contrast, depends only on observable attributes. Sellers' discrete choices between iBuyers and the open market—based on the price gap and their private hassle costs—complete the estimation. The results show that unobserved house quality accounts for substantial variation in transaction prices, and that modeling both dimensions of private information is essential for understanding selection and improving iBuyer performance.

Turning to the counterfactual analysis, I develop both a theoretical and numerical framework to explore how contract design can mitigate adverse selection. Building on the classic cream-skimming argument of Rothschild and Stiglitz [1976], I first demonstrate theoretically that iBuyers can selectively acquire higher-quality homes by offering sellers a lower upfront payment coupled with a conditional revenue-sharing mechanism upon resale. When the idiosyncratic price error outside of unobserved quality approaches zero, this contract structure always improves selection by discouraging only lower-quality homeowners.

Building on this theoretical foundation, I evaluate the model numerically using estimated parameters. First, I show that even under highly favorable assumptions—where iBuyers can resell properties immediately at market prices, without dynamic concerns such as resale timing, inventory costs, or holding risk—they still earn negative profits under the existing contract structure. While iBuyers' recent struggles are often attributed to macroeconomic conditions, resale timing, or inventory management, this result suggests that static adverse selection may also play a significant role.

Second, I simulate alternative revenue-sharing contracts by varying the upfront payment and the revenue-sharing ratio. Contracts with upfront payments of 60% or less, combined with conditional revenue sharing, can generate positive profits for iBuyers while remaining feasible for homeowners. For example, a property valued at $420,000[3] with an outstanding mortgage balance of $150,000 would yield an upfront payment of $252,000. After retiring the mortgage, the homeowner would retain $102,000 in equity—well above the 20% down-payment threshold ($84,000) required to purchase a similarly priced home, thereby avoiding PMI. These results suggest that contract design may constitute a more effective and sustain-

[3] Median existing-home sales price ($428,500 in July 2025) from National Association of Realtors [2025]; median debt secured by primary residence ($155,600) from Board of Governors of the Federal Reserve System [2023]; 20% down-payment benchmark from Bank of America Corporation [2025].

able lever for improving iBuyer performance than efforts focused solely on dynamic resale risk. Moreover, by reducing selection on unobserved quality at the time of purchase, such contracts may also limit subsequent exposure to inventory-related losses.

As a further counterfactual, I augment the iBuyer' s pricing model by incorporating unstructured text from housing listings, specifically the "Public Remarks" field. Standard text representations such as TF-IDF or Sentence Transformer embeddings are high-dimensional—requiring more than 50 principal components to capture 80 percent of the variance—making it challenging to integrate text into pricing models without overfitting or losing interpretability.

To address this, I propose a one-dimensional projection using a large language model (LLM), which generates a scalar "text score" incorporated into the pricing algorithm through three approaches: prompt engineering, parameter-efficient fine-tuning (LoRA), and feature extraction with subsequent regression. Incorporating this score raises the minimum upfront payment ratio for positive expected profits from about 60% to 75%, in a way that better aligns with sellers' hassle cost considerations.

The LLM-based text score may partially capture subjective or latent home characteristics—such as natural lighting, layout flow, or aesthetic appeal—that are difficult to encode using structured data alone. Incorporating this score into the pricing algorithm improves expected profitability across both current and counterfactual contract designs and is especially valuable in settings where contract-based cream-skimming is less effective. These findings suggest that richer data inputs, when combined with flexible contract structures, can help mitigate adverse selection in algorithmic housing markets.

The remainder of the paper is organized as follows. Section 2 provides a brief background on the iBuyer industry. Sections 3 illustrates the model of hassle costs and unobserved house quality. Section 4 and 5 present the data and descriptive evidence on hassle costs, unobserved house quality, and iBuyers' adverse selection. Sections 6 and 7 then describe the identification and estimation strategy, and parameter estimates, respectively. Section 8 presents counterfactual simulations based on contract design and extends the analysis by incorporating unstructured listing text into the pricing model. Finally, Section 9 concludes.

2 Institutional Background

This section outlines the iBuyer model and examines its core attributes, business model, transaction patterns, and operational challenges in key metropolitan markets where these firms operate.

2.1 iBuyers in Home Selling: Role and Process

iBuyers, short for 'instant buyers,' are firms that acquire residential properties directly from individual sellers in a streamlined manner. These companies utilize algorithmic pricing models to determine purchase offers, drawing on seller-provided information and broader market data to estimate property values. However, because these pricing algorithms do not incorporate in-person inspections at the time of offer generation, they may overlook certain qualitative property characteristics that are only observable through direct evaluation. This limitation can lead to pricing inefficiencies and expose iBuyers to adverse selection risks, as sellers with properties of lower unobserved quality may have a greater incentive to accept offers.

In the United States, selling a house is typically a complex and time-consuming process, requiring the involvement of both buying and selling brokers, as well as navigating extended listing and closing procedures. In Table 1, Zillow Group Inc. [2024] outlines the key steps in the home-selling process, highlighting the numerous hurdles sellers must navigate. Even after deciding to list their house, homeowners face multiple intricate steps that demand considerable time and effort. To address these challenges, iBuyers position themselves as a more efficient alternative, offering sellers rapid, algorithm-driven purchase offers that simplify the traditional transaction process.

Stage	**Traditional Home Sale Process**	**iBuyer Transaction Process**
Pre-Sale	Hire a real estate agent Prepare and stage the home Set an asking price	Submit property details online
Sale	Home is cleaned and staged List the home on the market Host showings for prospective buyers Negotiate contract terms	Accept the offer
Post-Sale	Complete a home inspection Plan the move Close the sale and pay costs	iBuyer conducts post-offer inspection Finalizes the transaction

Table 1: Comparison of Traditional Home Sales and iBuyer Transactions
Note: This table summarizes the key differences between the conventional home-selling process and the streamlined approach offered by iBuyers. It emphasizes how iBuyers reduce the seller' s effort in preparation, listing, and overall transaction management.

Table 1 also illustrates the typical selling process as of 2018, when iBuyers were becoming relatively active in my sample data. I use web archive data from Opendoor and Offerpad to examine their procedures in 2018, as Zillow and Redfin have exited the market, making it challenging to retrieve archival records of their now-defunct iBuyer operations. More details are provided in Appendix A.

As described above, iBuyers provide rapid purchase offers based on online home details without requiring an in-person showing. Once a contract is accepted, an inspection follows to verify the property's condition and determine any necessary repair costs before resale.

There are several rationales for this ex-post inspection. First, as shown in Table 1, inspections are an integral part of the traditional home-selling process. If an iBuyer were to conduct an inspection before finalizing a contract, individual sellers might exploit this service to obtain a free home assessment, regardless of whether they proceed with an iBuyer sale. This could result in substantial costs for iBuyers.

Second, conducting an in-person inspection before making an offer introduces subjectivity into the pricing process. Certain home characteristics—such as the overall mood or atmosphere—are difficult to index or contract upon, making it challenging for new market entrants like iBuyers to provide a trustworthy and credible offer. Instead, iBuyers generate price offers based on observable, verifiable data supplied by the seller. This data-driven approach helps establish transparency and credibility, both of which are essential for competing with traditional real estate models. By shifting inspections to after the offer stage, iBuyers can use them to verify repair costs without casting doubt on the objectivity of the offer price. In contrast, ex-ante inspections may lead sellers to suspect that pricing reflects subjective judgments rather than data, thereby undermining trust in the pricing mechanism and reducing the appeal of the iBuyer model.

Lastly, avoiding in-person visits aligns with iBuyers' value proposition of reducing hassle for sellers. Traditional home sales often involve multiple showings, which can be burdensome. By streamlining the process, iBuyers aim to offer a more convenient alternative to conventional transactions.

According to Buchak et al. [2020], iBuyers experienced significant growth starting in 2015 in Phoenix, followed by expansions between 2016 and 2018 in Orlando and other markets. Major iBuyer companies included Opendoor, Offerpad, Redfin, and Zillow. Notably, Redfin and Zillow, already prominent players in real estate services, expanded into the iBuyer market by launching RedfinNow and Zillow Offers, respectively. However, after incurring substantial losses, both RedfinNow and Zillow Offers eventually exited the market.

2.2 iBuyer business model

The primary source of iBuyer revenue stems from purchasing houses directly from individual sellers and reselling them to buyers. According to Opendoor's 10-K filing (Opendoor

Technologies Inc. [2023]), the company defines its spread as the total discount to its internal home valuation at the time of offer, minus a 5% service fee. They emphasize that smaller spreads are associated with higher seller conversion, suggesting that offering relatively modest discounts plays a crucial role in persuading individual sellers to choose iBuyers over listing on the open market.[4]

The evidence that iBuyers purchase properties at prices below market value is also observable in my data. Table 2 shows that the buying price of an iBuyer is approximately 5% lower than the individual buying price, after controlling for house and market characteristics.

	Log iBuyer price
iBuyer dummy	-0.05 (0.00)***
Log living area square feet	1.00 (0.00)***
Bedroom number	-0.12 (0.00)***
Bathroom number	0.02 (0.00)***
Building age	0.00 (0.00)***
Garage dummy	0.19 (0.00)***
Heating dummy	0.42 (0.01)***
seasonalFE: 2nd quarter	0.03 (0.00)***
seasonalFE: 3rd quarter	0.02 (0.00)***
seasonalFE: 4th quarter	-0.00 (0.00)
30-year mortgage rate	-0.01 (0.00)***
Federal funds rate	0.03 (0.00)***
Log CPI index	-0.90 (0.01)***
Log CS index	1.08 (0.00)***
R^2	0.85
Adj. R^2	0.85
Num. obs.	976,883

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Log iBuyer price estimation
Note: This table presents regression results estimating the iBuyer discount on home purchases. The dependent variable is the log of the iBuyer home purchase price. The transaction price is expressed in units of $100,000, and the living area is measured in units of 1,000 square feet. Standard errors are reported in parentheses.

On the other hand, iBuyers do not generate revenue from renting inventory, as their core business model focuses on buying and reselling properties. This is corroborated not only by their websites, which indicate no additional revenue sources from renting, but also by my raw data, which contains no observations of Offerpad, Opendoor, Redfin, or Zillow listing properties for rent. Since inventory is not utilized for other purposes, as mentioned in

[4] Opendoor Technologies Inc. [2023]

Opendoor's 2024 10-K, holding costs are incurred for each unit of inventory, and inefficient management of inventory can negatively impact their financial performance.[5] Table 3 also documents the time homes remain on the market after listing, comparing how long iBuyers take to resell properties with how long individual owners take to sell theirs.[6]

Finally, individuals make three additional types of payments to iBuyers. A detailed breakdown—based on information from Opendoor's official website archived in 2019 (Internet Archive [2019])—is provided in Appendix B, Figure 9. The first category is service fees. According to Opendoor's 10-K filing (Opendoor Technologies Inc. [2023]), its official website (Opendoor Technologies Inc. [2025]), and Offerpad's website (Offerpad LLC [2025]), both companies charge a 5% fee, which is comparable to the traditional real estate agent commission (typically around 5∼6%). These service costs also cover expenses related to holding and reselling, including property taxes, insurance, marketing, and other associated costs. Additional detail from Offerpad is in Appendix B.

The second category is repair costs, which are charged after an in-person inspection once the contract is accepted, as explained in the previous section. These costs do not imply that iBuyers renovate the property to enhance its market value and subsequently charge sellers for improvements. Instead, repair costs cover the necessary fixes required before the property can be listed for resale, similar to the preparatory steps in the traditional home-selling process, as illustrated in Table 1.

The final category is closing costs, which encompass the same expenses as those incurred in the traditional home-selling process, as illustrated in Table 1. These costs typically include title insurance, escrow fees, and other administrative expenses necessary to finalize the transaction.

2.3 iBuyers transaction patterns

The iBuyer business model is a relatively recent innovation, grounded in algorithmic home pricing and rapid transaction execution. Following Buchak et al. [2020], the analysis focuses on markets where iBuyers were most active during their early expansion. In addition, the selection of cities reflects data availability and constraints of the dataset. The final sample includes six metropolitan areas: Phoenix, Atlanta, Jacksonville, Orlando, Tampa, and Charlotte.

[5] *"Efficiently turning our inventory, inclusive of repairing, listing, and reselling the home, is important to our financial performance, as we bear holding costs (including utilities, property taxes, maintenance and insurance) and financing costs during our ownership period."* (Opendoor Technologies Inc. [2023])

[6] Table 3 is generated and explained in greater detail in a later section, but note that it is conditioned on houses that were subsequently transacted at least once more.

Figure 1 illustrates the growth in iBuyers' transaction volumes since 2015. Their buying and selling volumes account between 2 to 6% percent of total transactions in each city.

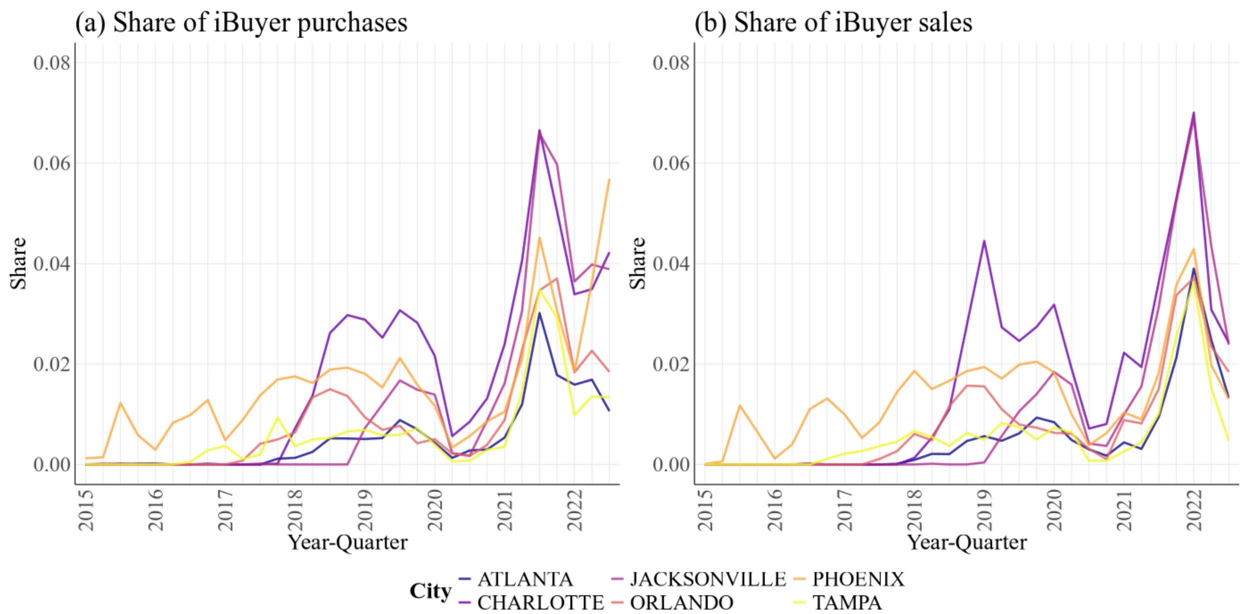

Figure 1: Share of iBuyer transactions by city
Note: Panel (a) shows the share of home purchases made by iBuyers in each quarter across six metropolitan areas. Panel (b) displays the corresponding share of homes sold by iBuyers. The sample includes Phoenix, Atlanta, Jacksonville, Orlando, Tampa, and Charlotte. Shares are calculated as the number of iBuyer transactions divided by the number of entire transactions in each city-quarter.

In most cases, iBuyers buy and sell to individuals. Table 3 presents the buying and selling behavior between iBuyers and individuals. Conditional on sales, Table 3 includes recent listing information (2001–2022) associated with different buyers and sellers. The data is filtered to exclude transactions where the buyer and seller are the same person and those under $1,000 to reduce noise in the data. The data originates from MLS listing records, and to merge it with the transaction data, I include only houses transacted more than once.[7]

When iBuyers sell to individual buyers, the majority of transactions occur through MLS listings, with a smaller share through their own websites or via syndication on real estate portals. Conversely, when individuals sell to iBuyers, only a small portion choose to list their properties first and then later sell to iBuyers. This is because listing a property itself

[7]The reason for using houses transacted more than once is that MLS data only contains listing information and lacks details about buyers and sellers, which are included in the transaction data. Therefore, I match the listing information to transactions. If a house is purchased and then immediately listed for sale, and the next transaction involves buyer B, I associate the listing information with the transaction involving buyer B. The names of the transaction participants are then used to construct the table.

creates significant hassle, and sellers choosing iBuyers typically want to avoid this. Therefore, individuals who prioritize convenience over maximizing sale price have little incentive to list their homes before approaching an iBuyer.

Seller Type	Buyer Type	Mean	Q1	Median	Q3	Number Sold via Listings	Total Sales
iBuyer	iBuyer	137	46	48	139	4	5
iBuyer	Individual	56	8	28	70	7642	8854
Individual	iBuyer	86	6	37	111	637	3720
Individual	Individual	103	9	38	113	186450	267655

Table 3: Days Listed on Market by Seller and Buyer Types (2001-2022)
Note: This table reports the number of days properties were listed on the market (DOM), conditional on a subsequent sale occurring, for homes transacted more than once between 2001 and 2022. Only transactions above $1,000 and those involving different buyer and seller entities are included. DOM is calculated using MLS listing records matched to transaction data. "Number Sold via Listings" refers to sales where the property was actively listed on MLS before the transaction. Because MLS listings lack buyer/seller info, DOM is associated with the next transaction, which may result in differences from unconditional summary statistics.

2.4 iBuyer exits and difficulties

The significant downturns for iBuyers occurred in 2021 and 2022. According to an article published by the real estate brokerage RubyHome (RubyHome [2022]), both Opendoor and Offerpad reported substantial financial losses in 2022. While Opendoor and Offerpad continued to bear the risks and remained in the market during this period, Zillow and Redfin exited the iBuying business on November 2, 2021 (Zillow Offers) and November 9, 2022 (RedfinNow), respectively.

Figure 2 (a) presents data from SEC 10-Q filings, using keywords related to "Net Income" and "Net Loss," to illustrate the financial hardships reported by RubyHome. As shown in the figure, both Opendoor and Offerpad generally reported net losses throughout the observed period. Offerpad recorded relatively small positive net income in the first quarter of 2022, along with very marginal gains in the second quarters of 2021 and 2022; however, these gains were minor and did not reverse the overall trend of financial losses. In contrast, Opendoor experienced larger and more volatile losses, with a particularly sharp decline in the second half of 2022. These trends support the notion that, although Opendoor and Offerpad remained active in the iBuying market, both companies faced significant financial challenges during this period.

Figure 2 (b) examines project-level profit data for Zillow, as reported in its SEC filings. Unlike the company-wide financial disclosures, the project-level data was less detailed. As a result, I was able to observe only the Contribution Profit before Interest for Zillow Offers. Even before accounting for interest expenses, Zillow Offers reported negative contribution profit, indicating that financial hardships were present prior to Zillow' s decision to shut down the Zillow Offers division.

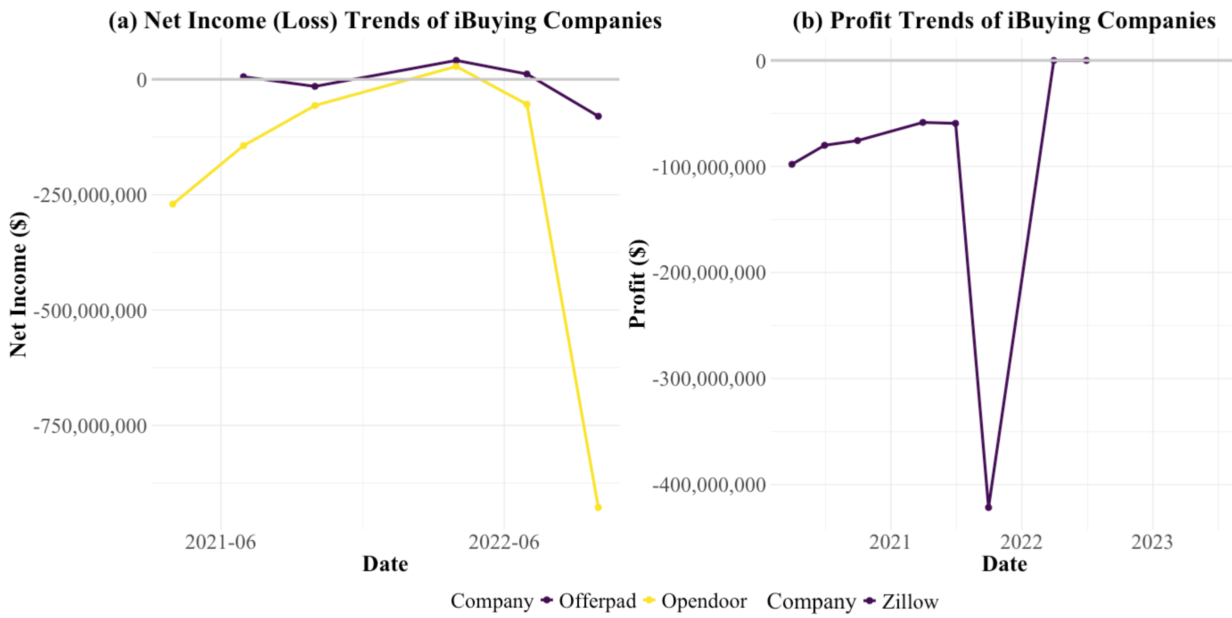

Figure 2: Financial Performance of Major iBuyers
Note: Panel (a) presents company-wide net income (loss) reported by Opendoor and Offerpad, based on SEC 10-Q filings. Panel (b) shows project-level contribution profit before interest for Zillow Offers, also based on SEC 10-Q disclosures. While Offerpad occasionally reported modest profits, all three companies generally experienced sustained losses, underscoring the financial headwinds faced by iBuyers.

3 Model

In this section, I present a structural model of individual sellers' decision-making, aimed at estimating the distribution of hassle costs and unobserved house quality. Sellers decide whether to transact with an iBuyer or sell on the open market based on their private information about houses' qualities and the burden associated with the sales process.

The model incorporates three key determinants of the seller's choice: the hassle cost, the offer received from the iBuyer, and the expected proceeds from listing the house on the open market. Strategic pricing, involving a trade-off between a higher price and a lower probability of selling, is simplified by using the hedonic pricing model for the expected market price. This assumption is justified since setting a house price with a full understanding of the selling probability is complex for individuals. They often rely on agents' advice or adhere to fair market prices instead of determining house prices from scratch.

3.1 Individual seller's problem

I begin by outlining the individual seller' s decision-making process, which is modeled as a two-stage decision. At time t, an individual seller decides whether to sell house h to the iBuyer or not, based on the following criteria.[8]

In Stage 1, a representative iBuyer presents a take-it-or-leave-it offer, and the individual seller's payoff from accepting this offer is given by:

$$p^i_{ht}.$$

If the seller rejects the offer, the property proceeds to Stage 2, where it is sold on the open market at price p^l_{ht}. However, the seller also incurs a hassle cost, c_{ht}, which enters multiplicatively. The resulting payoff from this alternative is:

$$\frac{p^l_{ht}}{c_{ht}}.$$

Hence, an individual seller compares the Stage 1 payoff, p^i_{ht}, with the expected Stage 2 payoff, given by:

$$\frac{\mathbb{E}[p^l_{ht}|X_{ht}, \xi_h]}{c_{ht}}.$$

Here, X_{ht} and ξ_h constitute the seller' s information set at Stage 1: the vector X_{ht} captures observable characteristics of the house and market at time t, while ξ_h represents unobserved house quality.

I define unobserved quality as attributes that cannot be identified through data analysis alone, without physically inspecting the property. However, individual buyers and sellers who visit the property are aware of these features. Examples include factors such as natural lighting or noise levels. Thus, ξ_h is known to the seller and included in their information set.

3.2 Hassle costs

Hassle cost refers to the time, logistical, financial, and emotional burdens associated with selling a home on the market. As often highlighted in iBuyer advertisements, examples of hassle costs include accommodating random visitors during the selling process, discussing with real estate agents to list and sell the house, and enduring extended waiting periods

[8]This modeling assumption is supported by transaction data indicating that a large majority—83%—of homeowners who sold to an iBuyer bypassed the MLS listing. As shown in Table 3, this statistic is derived from a sample of properties that transacted multiple times between 2001 and 2022 and were matched to MLS listing records. Because the MLS does not include buyer and seller identities, listing durations are attributed to the next observed transaction between distinct parties.

before the home is sold. Since hassle cost in this model is one-dimensional, it potentially represents a projection of various burdens in the market, such as concerns about the time constraints of purchasing a new house.

In this model, the individual seller of house h at time t draws a hassle cost c_{ht} from a mixed distribution. I allow for correlation between hassle costs and unobserved house quality, ξ_h, which captures aspects of the property that cannot be evaluated solely through data analysis. A positive correlation implies that sellers with higher hassle costs tend to own homes with higher unobserved quality. Since sellers facing greater hassle costs are more inclined to choose the iBuyer, this correlation mitigates the extent of adverse selection faced by the iBuyer, as higher-quality homes are more likely to enter its purchase pool. Conversely, a negative correlation intensifies adverse selection, as sellers who prefer the iBuyer due to higher hassle costs tend to own homes of lower unobserved quality.

With probability a, the seller experiences no hassle cost, represented by $\log c_{ht} = -\infty$. With the remaining probability $1-a$, the pair $(\log c_{ht}, \xi_h)$ is jointly normally distributed:

$$\begin{pmatrix} \log c_{ht} \\ \xi_h \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\log c} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\log c} & \rho\sigma_{\log c}\sigma_\xi \\ \rho\sigma_{\log c}\sigma_\xi & \sigma^2_\xi \end{pmatrix}\right).$$

c_{ht} is private information of each individual but the distribution is common knowledge.

Note that $-\infty$ represents the possibility that some individuals are either unaware of or uncomfortable interacting with iBuyers. For example, they may be tech-illiterate and unfamiliar with using a website to sell their houses. Given that the iBuyer business model is relatively new and traditional home-selling methods still dominate the market, the probability a reflects that iBuyers are not yet widely recognized.

4 Data

This paper utilizes two datasets from CoreLogic: property transactions and MLS recordings. The analysis focuses on six cities where iBuyers were most active: Phoenix, Atlanta, Jacksonville, Orlando, Tampa, and Charlotte. The data span from 1990 through the third quarter of 2022, excluding the period from 2006 to 2012. The years 2006 to 2012 represent a pre-iBuyer era (with the earliest iBuyer entry in my sample occurring in 2015) and a period heavily affected by the financial crisis. For example, the surge in foreclosures during and after the crisis not only depressed prices of foreclosed properties but also distorted broader housing market prices. To avoid any potential linear and nonlinear distortions and enhance the accuracy of expected market price estimates, I restrict the analysis to the recent period unaffected by the financial crisis.

To ensure a conservative approach, I apply several additional data restrictions. Within the CoreLogic data, I exclude foreclosure transactions and retain only recordings for single-family homes and condominiums, as iBuyers primarily target these two property types. I also remove extreme outliers in log house prices, log living area (square footage), and the number of bedrooms and bathrooms, since iBuyers typically avoid transacting in extreme property types and instead focus on median-type homes (Buchak et al. [2020]). Finally, although corporate ownership names can sometimes represent individuals seeking privacy rather than institutional buyers, I exclude transactions involving corporate-named buyers and sellers, except for those involving iBuyers, to conservatively avoid including institutional investors.

CoreLogic transaction recordings offer detailed transaction-level data, primarily including transaction dates, sales prices, and buyer and seller names. The buyer and seller names are recorded as legal transaction names, which I use to identify whether participants are iBuyers. Following Buchak et al. [2020], I use a detector to capture regular expressions related to iBuyer companies such as Opendoor, Offerpad, Redfin, and Zillow. For instance, Opendoor's regular expressions include "`OPENDOOR`", "`OPEN DOOR`" and "`\\\\ OD [A-Z].* LLC$`".[9]

CoreLogic MLS recordings provide listing-level data, including information on listing dates, Days on Market, and house characteristics. To combine house characteristics with transaction records, I merge all transaction and listing records using the composite property linkage key.[10]

Additional data includes macroeconomic variables to assess market conditions. CPI, Federal funds rate, 30-year mortgage rate, and the Case-Shiller Index are used. Except for CPI (sourced from the World Bank), all other variables are sourced from FRED. All macroeconomic variables are at the nationwide monthly level.

Table 4 provides a summary of house prices and characteristics for properties that iBuyers purchase from individuals, as well as the types of inventory iBuyers sell to individuals compared to market transactions between individuals.

[9]See Buchak et al. [2020] for more details.

[10]I also double-checked that the composite property linkage key and clip (from both the transaction record file and the listing record file) were one-to-one mappings in my sample.

Variable	Mean	SD	Q1	Median	Q3	N
Seller: iBuyer → Buyer: Individual						
Sales price	335,085.31	107,873.79	253,288.84	317,781.52	403,567.61	8,945
Living area square feet	1,748.57	526.70	1,367.00	1,659.00	2,033.00	8,945
Bedroom number	3.17	0.70	3.00	3.00	4.00	8,945
Bathroom number	2.41	0.57	2.00	2.00	3.00	8,945
Garage dummy	0.97	0.16	1.00	1.00	1.00	8,945
Heating dummy	1.00	0.01	1.00	1.00	1.00	8,945
Building age	28.78	17.00	15.00	25.00	40.00	8,945
Seller: Individual → Buyer: iBuyer						
Sales price	329,663.86	111,184.89	245,149.03	312,369.31	398,798.17	10,911
Living area square feet	1,752.85	512.03	1,380.50	1,670.00	2,030.50	10,911
Bedroom number	3.20	0.69	3.00	3.00	4.00	10,911
Bathroom number	2.43	0.57	2.00	2.00	3.00	10,911
Garage dummy	0.97	0.16	1.00	1.00	1.00	10,911
Heating dummy	1.00	0.03	1.00	1.00	1.00	10,911
Building age	27.75	16.82	15.00	24.00	38.00	10,911
Seller: Individual → Buyer: Individual						
Sales price	278,085.36	151,648.18	171,734.86	244,015.20	348,712.21	957,023
Living area square feet	1,698.20	562.64	1,300.00	1,590.00	1,988.00	957,023
Bedroom number	3.07	0.76	3.00	3.00	4.00	957,023
Bathroom number	2.29	0.62	2.00	2.00	3.00	957,023
Garage dummy	0.99	0.10	1.00	1.00	1.00	957,023
Heating dummy	1.00	0.05	1.00	1.00	1.00	957,023
Building age	28.36	22.07	12.00	22.00	42.00	957,023

Table 4: Summary Statistics (1990–2022)
Note: Sales prices are adjusted to March 2023 dollars using the Consumer Price Index (CPI). The table reports summary statistics for three transaction types: (i) iBuyers selling to individuals, (ii) iBuyers purchasing from individuals, and (iii) individual-to-individual market transactions. Garage and heating are binary indicators; building age is calculated as the difference between sale year and year built. Outliers in price, living area, and bedroom/bathroom counts have been removed.

5 Descriptive evidence

Before turning to structural estimation, this section presents reduced-form evidence supporting the existence of two key forces: hassle costs and unobserved quality. The analysis yields two main takeaways: (i) individual sellers possess private information along two dimensions—hassle costs associated with selling on the open market and unobserved house quality, and (ii) iBuyers face adverse selection when purchasing homes from individuals.

First, I will demonstrate that individual sellers encounter varying hassle costs in the market, which differ across individuals. Next, I will illustrate that some house characteristics cannot be identified without physical inspection. Lastly, I will show that these unobserved characteristics contributing to unobserved quality can cause iBuyers to face adverse selection, potentially reducing their profitability.

5.1 Hassle costs

As defined in the previous section, hassle cost refers to the various burdens—time, logistical, financial, and emotional—associated with selling a home on the market. As shown in Table 1, this process involves several steps, such as preparing the home and hosting open houses. Zillow Group Inc. [2024] reports that the average time from listing to closing is approximately 90 days, indicating that the stress of selling is significant. Each individual may experience different levels of hassle costs each time they decide to sell a house.

Examining the front pages of popular iBuyer websites, such as Offerpad and Opendoor, it is evident that they specifically target individuals with high hassle costs. Screenshots of these websites are provided in Appendix C, Figures 12 and 13, to illustrate their marketing strategies.

Because hassle costs are difficult to measure directly without survey data, I instead compare two groups of sellers that plausibly differ in their hassle costs: those who relocate within the state and those who move out of state. Tables 5 and 6 confirm that individuals with higher hassle costs are more likely to sell their homes to iBuyers, consistent with iBuyers' strategy of targeting time-constrained or convenience-seeking sellers.

To classify sellers, I define relocation within the state in two ways. The broader definition includes individuals who purchase another home in the same state within one year before or after the sale. The stricter version includes only those who buy another home within one year after the sale. To accurately link transactions to individuals and avoid mistakenly grouping different people who share the same name, I restrict the analysis to sellers with relatively uncommon legal names. This restriction retains 58% of the full dataset.

Using this classification, I estimate a logit regression where the dependent variable is a dummy indicating whether a seller transacted with an iBuyer. The key independent variable is an indicator for across-state movers. The regression specification for individual i at transaction t is:

$$1_{\text{Sell to iBuyers},it} = \beta_0 \cdot 1_{\text{Across-State Mover},it} + X_{it}\beta_1 + \varepsilon_{it}$$

where X_{it} includes controls for the decile of the home's sale price, year fixed effects, and city fixed effects. The sale price decile is included to account for heterogeneity in seller behavior based on home value. This variable may proxy for income level or reflect the extent to which sellers engage in more deliberate decision-making for higher-value properties.

Sellers relocating across state lines are more likely to face logistical challenges and time constraints, making them more susceptible to the appeal of iBuyers. The results, reported in Tables 5 and 6 as average marginal effects, show that being in the across-state mover group increases the probability of selling to an iBuyer by about 1 percentage point. Considering that iBuyer market share across cities ranges from 2% to 6%, and averages around 3.5% among sellers with uncommon names, this increase is economically meaningful.

	Sell to iBuyers (*Marginal Effects*)				
	Without FE (1)	Without FE(2)	With FE (1)	With FE (2)	Mean of Indep. Var.
Across-State Mover	0.011***		0.010***		0.765
	(0.001)		(0.001)		
Across-State Mover (Strict)		0.010***		0.008***	0.842
		(0.001)		(0.001)	
Mean of Dep. Var.		0.0354			
Year FE	No	No	Yes	Yes	
City FE	No	No	Yes	Yes	
AIC	89463.256	89536.652	81706.747	81791.499	
BIC	89484.433	89557.829	81865.572	81950.324	
Log Likelihood	-44729.628	-44766.326	-40838.373	-40880.749	
Deviance	89459.256	89532.652	81676.747	81761.499	
Num. obs.	293121	293121	293121	293121	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: Logit Regression Results (Marginal Effects Only)
Note: This table reports average marginal effects from logit regressions where the dependent variable is an indicator for selling to an iBuyer. The key independent variable is an indicator for across-state movers, defined either broadly (columns 1 and 3) or strictly (columns 2 and 4). Columns 3–4 add year and city fixed effects. The sample is restricted to sellers with uncommon names to improve matching accuracy.

	Sell to iBuyers (*Marginal Effects*)				
	Without FE (1)	Without FE(2)	With FE (1)	With FE (2)	Mean of Indep. Var.
Across-State Mover	0.011***		0.010***		0.765
	(0.001)		(0.001)		
Across-State Mover (Strict)		0.010***		0.008***	0.842
		(0.001)		(0.001)	
Sale amount (decile)	-0.002***	-0.002***	-0.002***	-0.002***	
	(0.000)	(0.000)	(0.000)	(0.000)	
Mean of Dep. Var.	0.0354				
Year FE	No	No	Yes	Yes	
City FE	No	No	Yes	Yes	
AIC	89279.338	89346.317	81481.410	81560.959	
BIC	89311.103	89378.082	81650.824	81730.372	
Log Likelihood	-44636.669	-44670.158	-40724.705	-40764.479	
Deviance	89273.338	89340.317	81449.410	81528.959	
Num. Observations	293,121	293,121	293,121	293,121	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 6: Logit Regression Results (Marginal Effects Only)
Note: This table extends the baseline specification by including the sale price decile as an additional control. The marginal effect on the decile variable is negative and statistically significant, suggesting that higher-value homes are slightly less likely to be sold to iBuyers. The interpretation of the across-state mover indicators remains consistent. All specifications use the same sample of sellers with uncommon names.

5.2 Unobserved quality

As previously defined, unobserved quality refers to characteristics of a house that are not captured in standard datasets but can be recognized by individuals who physically inspect the property. These include factors such as natural lighting, noise levels, and other subtle features that affect a home's desirability. While these attributes are typically invisible to algorithmic valuation models, they are apparent to individual buyers and sellers through in-person visits.

As a result, unobserved quality is more likely to be reflected in transaction prices when properties are sold between individuals. Buyers often conduct thorough on-site inspections, and sellers—as residents—are intimately familiar with the property's attributes. In contrast, iBuyers formulate offers based primarily on observable data and do not incorporate such in-person insights into their pricing models. As noted by Buchak et al. [2020], "Non-iBuyer real estate buyers use other inputs to determine prices that do not seem to be captured in the iBuyer algorithm. Such information can either arise from other participants using difficult-to-encode information that is available or information acquired through a thorough and lengthy inspection, which iBuyers do not conduct because they offer a speedy closure."

This distinction is empirically illustrated in Table 7, which presents hedonic pricing model results for both individual purchase prices and iBuyer purchase prices, using data from 2015 —the earliest period of iBuyer entry in my sample. The R-squared from a regression of iBuyer purchase prices on basic observable characteristics, time, and regional fixed effects is significantly higher than that from the corresponding model for individual buyers. This finding aligns with Buchak et al. [2020] and supports the interpretation that iBuyer pricing relies more heavily on observable inputs, while prices in individual-to-individual transactions incorporate both observed and unobserved home qualities.

	Log iBuyer price	Log individual price
Intercept	7.04 (1.30)***	−5.92 (0.35)***
Log living area square feet	0.66 (0.01)***	0.93 (0.00)***
Bedroom number	0.01 (0.00)*	−0.12 (0.00)***
Bathroom number	−0.08 (0.01)***	0.02 (0.00)***
Building age	0.00 (0.00)***	0.00 (0.00)***
Garage dummy	0.10 (0.02)***	0.23 (0.01)***
Heating dummy	0.11 (0.08)	0.51 (0.02)***
Seasonal FE: 2nd quarter	0.06 (0.01)***	0.03 (0.00)***
Seasonal FE: 3rd quarter	0.08 (0.01)***	0.02 (0.00)***
Seasonal FE: 4th quarter	0.02 (0.01)**	0.00 (0.00)*
30-year mortgage rate	0.03 (0.01)***	−0.03 (0.00)***
Federal funds rate	−0.03 (0.01)***	0.01 (0.00)***
Log CPI index	−3.99 (0.48)***	0.54 (0.13)***
Log CS index	2.26 (0.18)***	0.68 (0.04)***
R^2	0.96	0.89
Adj. R^2	0.96	0.89
Num. obs.	10,911	380,650

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: R^2 comparison (2015-2022)
Note: This table presents regression results comparing the explanatory power (R^2) of models predicting iBuyer and individual transaction prices. The dependent variable is the log of the transaction price—iBuyer-to-individual in Column 1 and individual-to-individual in Column 2—expressed in $100,000s. Living area is measured in 1,000 square feet. Standard errors are reported in parentheses. The models use data from 2015, the earliest period of iBuyer entry in the sample.

To ensure that this result is not driven by differences across the cities where iBuyers entered, I conduct an additional hedonic regression using the same set of variables, but now interacting all housing characteristics and market characteristics with city fixed effects. The detailed results of this robustness check are presented in Appendix D. The R-squared comparison results remain similar, with iBuyer purchase prices exhibiting higher explanatory power. However, the increase in R-squared for both regressions is modest and not significantly different from those reported in Table 7.

5.3 iBuyers' adverse selection on house quality

In a descriptive comparison of resale margins between individuals and iBuyers, iBuyers generally experience greater losses than individuals. This pattern may indicate that iBuyers purchase lower-quality properties, potentially due to adverse selection. The finding that iBuyers incur larger losses persists even when the sample is restricted to short-term resales only.

Margins are defined in two ways: (1) the difference between resale and purchase prices, adjusted to 2023 levels using the Consumer Price Index (CPI), expressed in absolute terms; and (2) the ratio of this difference to the original purchase price. Since iBuyers often acquire properties at a discount, it is meaningful to examine both absolute and relative (ratio) margins.

To visualize how margins evolve over time for each buyer type, I apply nonparametric smoothing using Generalized Additive Models (GAMs). This technique flexibly estimates non-linear trends without imposing a specific functional form, enabling clearer insight into temporal patterns.

To analyze these differences in greater detail, Figure 3 compares the resale margins of individuals and iBuyers for transactions conducted since 2015. Additional figures—presenting absolute margins and short-term resale analyses—are provided in the Appendix E (Figures 14, 15 and 16). To account for potential time effects, Appendix Figures 15 and 16 restrict the sample to transactions completed within one year of purchase. Focusing on short-term resales is particularly meaningful—not only because it mitigates the influence of market-wide time trends, but also because iBuyers intend to quickly flip properties for profit. Unlike individual buyers, iBuyers have no alternative motives, such as residential occupancy or rental income, that would justify holding properties for extended periods.

Across all four visualizations, iBuyers consistently earn lower resale margins than individual buyers, reinforcing the conclusion that they face greater resale losses—or reduced gains—on average. In average terms, iBuyers earn approximately \$15,000–\$25,000 less than individual buyers in absolute dollar margins, and about 10–15 percentage points less in relative (percentage) terms prior to the end of 2022. The apparent improvement in iBuyer margins in 2022 should be interpreted with caution: it may reflect a lower frequency of iBuyer sales in the recent period, or the fact that many recently purchased, underperforming homes have not yet been resold and are thus excluded from the margin calculations. These trends remain consistent even after controlling for holding period by restricting the sample to resales completed within one year. Overall, these trends underscore the difficulty iBuyers face in achieving profitable resales—challenges consistent with adverse selection arising from private information about house quality, as discussed in earlier sections.

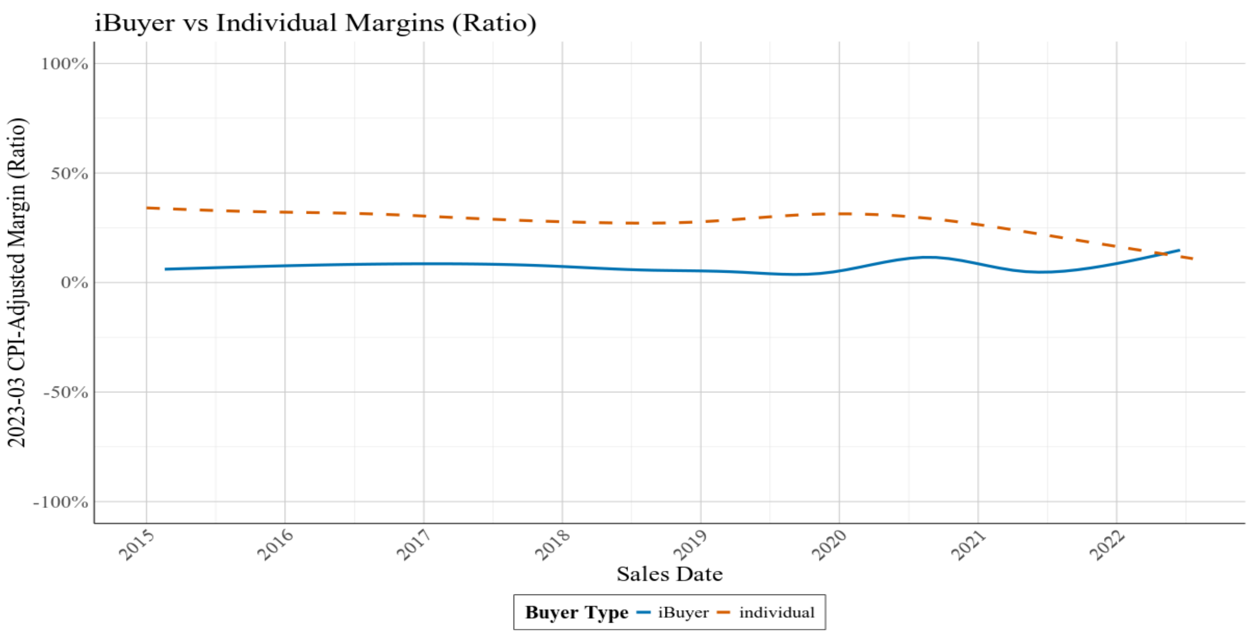

Figure 3: Comparing ratio margins
Note: This figure shows the average ratio margin (resale price minus purchase price, divided by purchase price) for iBuyers and individuals over time. Values are adjusted to 2023 dollars using the Consumer Price Index. Smoothed trends are estimated using Generalized Additive Models. Additional supporting figures appear in the Appendix.

6 Identification and Estimation

In this section, I present the identification and estimation strategies for the demand parameters, the distribution parameters of hassle costs, and the distribution parameters of unobserved quality, as outlined in the model described in the previous section. I explain how variations in the data serve to identify each parameter of interest. The estimation process is organized into three parts: the market pricing model, the iBuyer pricing model, and the choice model based on hassle costs and unobserved house quality. In principle, it would be most efficient to estimate all three components jointly. However, due to computational burden, I estimate them sequentially in the order outlined above, accepting some loss in statistical efficiency for tractability.

Demand parameters are derived from each pricing model. With these parameters, the unobserved house quality is identified, at most, in a distributional sense, primarily using pre-iBuyer entry period prices. Once demand parameters and unobserved quality distribution are determined, hassle costs are identified under distributional assumptions through the individual seller's choice model.

6.1 Demand parameters

Hedonic regression uses observed market transaction prices and iBuyer offer prices to identify the demand parameters β^l and β^i, based on variation in house and market characteristics. For a given house h at time t, the market transaction price is denoted by p^l_{ht} (where the superscript l refers to listings), and the iBuyer offer price is denoted by p^i_{ht} (where the superscript i refers to iBuyer offers).

The market pricing model is specified as a random effects model:

$$\log p^l_{ht} = X_{ht}^\top \beta^l + \xi_h + \varepsilon^l_{ht}$$

and the iBuyer pricing model is a standard linear model:

$$\log p^i_{ht} = X_{ht}^\top \beta^i + \varepsilon^i_{ht}$$

where X_{ht} denotes a set of house and market characteristics, and ξ_h is a house-specific random effect interpreted as unobserved quality. I start with a common set of characteristics but allow the set to be narrowed differently for each pricing model if a near-multicollinearity problem arises.

Formally, the identification of the demand parameters in the market pricing model and the iBuyer pricing model is based on standard assumptions about unobserved heterogeneity, error structure, and variation in observables. These conditions ensure that the coefficients can be consistently estimated using standard regression techniques.

Assumption 1 (Exogeneity)

$$\mathbb{E}\left[\xi_h | X_{ht}\right] = 0, \quad \mathbb{E}\left[\varepsilon^l_{ht} | X_{ht}, \xi_h\right] = 0, \quad \mathbb{E}\left[\varepsilon^i_{ht} | X_{ht}\right] = 0$$

Assumption 2 (Full rank condition) *Let $X_{ht} \in \mathbb{R}^k$ denote the $k \times 1$ vector of regressors for house h at time t.*

(i) *For the market pricing model: Let $\tilde{X}_{ht} = X_{ht} - \theta \bar{X}_h$ denote the quasi-demeaned regressors used in the random effects transformation of the market pricing model, where $\bar{X}_h =$*

$\frac{1}{T_h}\sum_t X_{ht}$ *is the within-house average and* $\theta \in [0, 1)$ *is the quasi-demeaning parameter.*[11] *Let* $\tilde{\mathbf{X}} = [\tilde{X}_{11}^\top; \ldots; \tilde{X}_{hT_h}^\top]$ *be the stacked matrix of quasi-demeaned regressors. Then:*

$$\text{rank}(\tilde{\mathbf{X}}) = k.$$

(*ii*) *For the iBuyer pricing model: Let* $\mathbf{X} = [X_{11}^\top; \ldots; X_{hT_h}^\top]$ *be the stacked regressor matrix. Then:*

$$\text{rank}(\mathbf{X}) = k$$

Assumption 3 is not required for identification of the parameters in the market or iBuyer pricing models. However, it is imposed if the market and iBuyer pricing models are estimated via maximum likelihood estimation (MLE). In this sense, normality is an auxiliary assumption rather than a core identification condition.

Assumption 3 (Normality for Likelihood-Based Estimation)

$$\xi_h \sim N(0, \sigma_\xi^2), \quad \varepsilon_{ht}^l \sim N(0, \sigma_{\varepsilon^l}^2), \quad \varepsilon_{ht}^i \sim N(0, \sigma_{\varepsilon^i}^2)$$

Proposition 1 (Identification of β^l and β^i) *Under the linear random effects specification for the market pricing model and the linear specification for the iBuyer pricing model, and given Assumptions 1 and 2, the demand parameters* β^l *and* β^i *are identified from the observed data* $\{\log p_{ht}^l, \log p_{ht}^i, X_{ht}\}_{h,t}$.

6.2 Unobserved house quality distribution

Following the definition of the previous section, an unobserved house quality is only unobserved by people who use data to evaluate the house without actually visiting the house like iBuyer or Econometrician. In econometrics, these unobserved qualities are typically modeled using either fixed effects or random effects. This paper employs the random effects model because some houses in the dataset are traded only once during the estimation period. In each city sample, houses that are traded more than once account for between 40% and 75%. By employing the random effects model, I aim to include rarely traded houses, under the assumption that their unobserved qualities come from a common distribution. This approach allows us to leverage the identification power provided by houses that have been traded multiple times.

I use the pre-entry period of iBuyers in the market pricing model to ensure that the distribution of ξ_h is observed unconditionally, meaning it is not further partitioned into iBuyer versus market transactions.

[11] If all regressors lack variation across houses h (i.e., they are constant across h), then the quasi-demeaning parameter $\theta = 1$. I rule out this degenerate case.

The identification of the ξ_h distribution is based on the assumption that ξ_h has an unconditional mean of zero and an unconditional variance σ_ξ^2. Consistent with the previous subsection, standard regression assumptions underpin the identification of this distributional parameter.

Proposition 2 (Identification of σ_ξ^2) *Under the linear random effects specification of the market pricing model, and given Assumptions 1 and 2, the variance of the random effect, σ_ξ^2, is identified from the observed data $\{\log p_{ht}^l, X_{ht}\}_{h,t}$.*

6.3 Hassle cost distribution

Assuming that each seller's hassle cost is derived from a common distribution, the identification is based on the individual' s choice data regarding whether they opt to sell to iBuyer or not.

An individual seller decides to sell to iBuyer if

$$p_{ht}^i > \frac{\mathbb{E}[p_{ht}^l | X_{ht}, \xi_h]}{c_{jt}}$$

and will not sell to iBuyer if

$$p_{ht}^i \leq \frac{\mathbb{E}[p_{ht}^l | X_{ht}, \xi_h]}{c_{jt}}.$$

The estimation of hassle cost distribution is based on information obtained from pricing regressions. When estimating demand parameters and the distribution of unobserved house quality, I can observe either the market price or the iBuyer offer price for each house, but not both at the same time. If a house is sold on the market, the iBuyer offer price remains unknown; conversely, if it is sold to an iBuyer, the market price is not available. As a result, recovering the counterfactual iBuyer offer price (in the first scenario) or the market price (in the second scenario) requires estimating demand parameters and the distribution of unobserved house quality. Once I recover these prices, the variation in the price difference ($\mathbb{E}[p_{ht}^l | X_{ht}, \xi_h]/p_{ht}^i$ for the house h at time t) can help me determine the variation in choice probability for each transaction.

I currently assume a joint mixed normal distribution to capture the possibility that some individual sellers may be unaware of or distrustful toward iBuyers, given that it is a relatively new business model. This structure also aligns with the empirical observation that iBuyers capture a relatively small share of the market (2%–6%).

With probability a, $\log c_{ht} = -\infty$. With the remaining probability $1 - a$, the pair

$(\log c_{ht}, \xi_h)$ follows a joint normal distribution:

$$\begin{pmatrix} \log c_{ht} \\ \xi_h \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\log c} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\log c} & \rho\sigma_{\log c}\sigma_\xi \\ \rho\sigma_{\log c}\sigma_\xi & \sigma^2_\xi \end{pmatrix} \right).$$

The following assumption formalizes the independence structure across observations, and the subsequent proposition establishes the identification of the key distributional parameters governing hassle costs under this mixed normal setup.

Assumption 4 (Independent pair) *The pair* $(\log c_{ht}, \xi_h)$ *is independently and identically distributed* (*i.i.d.*) *across houses* h*. Dependence between* $\log c_{ht}$ *and* ξ_h *within a given house is allowed.*

Proposition 3 (Identification of $\mu_{\log c}, \sigma_{\log c}, \rho, a$) *Given the demand parameters* β^l, β^i*, the distribution parameter* σ^2_ξ*, and the joint mixed normal structure of* $(\log c_{ht}, \xi_h)$*, under Assumptions 2 and 4, the hassle cost distribution parameters* $(\mu_{\log c}, \sigma_{\log c}, \rho, a)$ *are identified from the observed choice data* $\{1_{\log c_{ht} > \Delta_{ht}(\xi_h)}\}$*, where* $\Delta_{ht}(\xi_h) = \log \mathbb{E}[p^l_{ht} | X_{ht}, \xi_h] - \log p^i_{ht}$*.*

6.4 Estimation strategy

To reduce the computational burden, I estimated the model in three sequential steps. First, I estimated the market pricing model using a linear panel specification with random effects. Second, I estimated the iBuyer pricing model using a standard linear regression. Finally, I estimated the individual seller's choice model using a simulated maximum likelihood approach, incorporating the previously estimated demand parameters and a distribution parameter capturing unobserved house quality. In this third step, I estimate the mean, variance, and mixture probability a of the hassle cost distribution, as well as its correlation with unobserved house quality. Details about the software and estimation tools are provided in Online Appendix A, and the mathematical formulation of the simulated maximum likelihood function used for the choice model is provided in Appendix G.

7 Estimates

I conducted a city-level analysis, accounting for varying distributions of hassle costs and unobserved quality in each market. The primary focus is on the distribution parameters associated with unobserved quality and hassle costs. Although the estimated parameters differ slightly across the six cities, the main conclusions from the counterfactual analysis remain consistent. Therefore, I focus on one city, Charlotte, as a representative example in this section. For brevity, the results for the remaining five cities are reported in Appendix J.

7.1 Demand parameters and unobserved house quality distribution

Table 8 and 9 provide a summary of the results from the pricing estimations.[12]

The primary focus of Table 8 is the magnitude of the random effect variance. In comparison to the variance of the error term, the random effect variance is 2 times greater. This indicates that the unobserved quality can vary significantly. Even after addressing the precision of the estimation (reflected in the error term variance), there could still be a considerable risk of adverse selection. Table 9 adjusted R^2 indicates the iBuyer offer price can be approximated by a simple linear model with high precision.

	Log price (Charlotte)
Intercept	-0.240 (0.081)**
Log living area square feet	1.103 (0.006)***
Bedroom number	-0.091 (0.002)***
Bathroom number	0.003 (0.002)
Building age	0.002 (0.000)***
Garage dummy	0.040 (0.008)***
Heating dummy	0.194 (0.026)***
seasonalFE_2nd quarter	0.015 (0.002)***
seasonalFE_3rd quarter	0.005 (0.002)*
seasonalFE_4th quarter	0.004 (0.002)
30-year mortgage rates	0.003 (0.002)
Federal fund rate	0.013 (0.001)***
Log CPI index	-0.163 (0.018)***
Log CS index	0.239 (0.006)***
Error variance	0.033
Random effect variance	0.065
Num. obs.	118604

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: Log market price estimation
Note: This table presents regression results for individual market transaction prices. The dependent variable is the log of the transaction price, expressed in \$100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. Standard errors are reported in parentheses. The model includes a random effect to capture unobserved heterogeneity.

[12] Prices are reported in units of \$100,000 and adjusted to the March 2023 CPI level. Living area is measured in units of 1,000 square feet.

	Log price (Charlotte)
Intercept	4.520 (1.886)*
Log living area square feet	0.712 (0.019)***
Bedroom number	−0.001 (0.007)
Bathroom number	−0.013 (0.008)
Building age	0.002 (0.000)***
Garage dummy	0.108 (0.029)***
seasonalFE_2nd quarter	0.044 (0.011)***
seasonalFE_3rd quarter	0.031 (0.011)**
seasonalFE_4th quarter	0.018 (0.011)
30-year mortgage rate	−0.018 (0.009)*
Federal funds rate	0.028 (0.012)*
Log CPI index	−3.841 (0.709)***
Log CS index	2.588 (0.273)***
R^2	0.972
Adj. R^2	0.972
Num. obs.	2877

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 9: Log iBuyer price estimation
Note: This table presents regression results for iBuyer offer prices. The dependent variable is the log of the iBuyer offer price, expressed in \$100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. Standard errors are reported in parentheses. The model is estimated using a linear regression framework.

7.2 Hassle cost distribution

Table 10 presents the estimation results of the hassle cost distribution. The confidence interval and standard deviation are computed using bootstrap methods.

Recall that the estimation is based on the model, which assumes joint normality between the log of hassle costs ($\log c$) and the unobserved house quality (ξ) as follows.

Either $\log c = -\infty$ with probability a or

$$\begin{pmatrix} \log c \\ \xi \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\log c} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\log c} & \rho\sigma_{\log c}\sigma_\xi \\ \rho\sigma_{\log c}\sigma_\xi & \sigma^2_\xi \end{pmatrix}\right)$$

with probability $1 - a$.

Table 10 shows that in Charlotte, the correlation between hassle costs and unobserved house quality is statistically insignificant. At first glance, this may give the impression that targeting sellers based on hassle cost does not introduce a quality-related selection issue. However, the variance of unobserved quality in the log market price estimation is 0.065, while the variance in log hassle costs is 0.116. Drawing on the insights from Finkelstein and McGarry [2006], selection problems can arise even when private information dimensions are uncorrelated, as each dimension may independently influence participation. This implies that an iBuyer strategy focusing solely on hassle costs may still face selection concerns.

Parameter	Estimate	Pivot CI Lower	Pivot CI Upper	Bootstrap mean	Bootstrap standard deviation
$\sigma^2_{\log c}$	0.116^{***}	0.087	0.145	0.115	0.016
$\mu_{\log c}$	-0.193^{**}	-0.321	-0.128	-0.222	0.050
a	0.828^{***}	0.777	0.851	0.819	0.019
ρ	0.031	-0.094	0.253	0.056	0.089

Table 10: Hassle cost estimation results
Note: This table presents parameter estimates from the joint mixed normal model of log hassle cost ($\log c$) and unobserved house quality (ξ). The model assumes that $\log c = -\infty$ with probability a, and follows a bivariate normal distribution with probability $1 - a$. Estimates are based on maximum likelihood and evaluated using 1000 bootstrap replications. Pivot confidence intervals, bootstrap means, and bootstrap standard deviations are reported. Standard errors are computed via the bootstrap.

8 Counterfactual Analysis

In this section, I propose counterfactual strategies for iBuyers designed to mitigate adverse selection in unobserved house quality. I begin by examining the performance of the current contract structure through a sanity check, which shows that iBuyers often earn low or negative profits. I then outline a counterfactual contract variable that exploits insights from the cream-skimming framework of Rothschild and Stiglitz [1976] and demonstrate how alternative contract designs can improve profitability.

Building on this framework, I further enhance counterfactual pricing design by incorporating unstructured data from listing text descriptions of houses. Leveraging approaches from large language models (LLMs), I use these data to better capture aspects of unobserved quality—such as mood, natural light, or perceived desirability—that are difficult to quantify using structured data alone. Combining improved contract design with more precise pricing mitigates adverse selection more effectively and increases iBuyer profitability.

8.1 Contract Design

The core idea of the counterfactual contract is to ensure that the individual seller's payoff remains partially linked to market prices, even when selling to the iBuyer. In the original arrangement, the iBuyer provides a lump sum payment when purchasing from the individual seller. This arrangement is characterized as a 100 percent upfront payment with no revenue sharing. In contrast, the counterfactual contract introduces flexibility in determining the proportions of upfront payment and conditional revenue sharing.

To focus on the issue of adverse selection, I abstract away from dynamic concerns such as resale timing, inventory costs, and the probability of resale. Instead, I adopt the simplifying assumption that the iBuyer resells the property at the market price immediately after purchase.

As a first step, I conduct a sanity check on the profitability of the current iBuyer contract. Using the counterfactual framework developed later in this section, I find that the iBuyer earns negative expected profits under the status quo arrangement of full upfront payment and no revenue sharing. This finding suggests that—even in the absence of dynamic considerations—adverse selection alone poses a significant challenge to the sustainability of the iBuyer model.[13]

Formally, when an individual seller sells house h at time t, the current contract gives the seller a payoff of p^i_{ht}, and yields per-transaction-revenue to the iBuyer equal to $p^l_{ht} - p^i_{ht}$, where p^i_{ht} is the payment made by the iBuyer to the seller, and p^l_{ht} is the market resale price of the house.

Under the counterfactual contract, $\delta \in [0, 1]$ denotes the proportion of the upfront payment. The seller receives an upfront payment of

$$\delta p^i_{ht},$$

and the remaining portion is subject to revenue sharing based on the realized resale price, as follows:[14]

$$\begin{cases} p^i_{ht} - \delta p^i_{ht} & \text{if } p^l_{ht} > p^i_{ht}, \\ p^l_{ht} - \delta p^i_{ht} & \text{if } p^i_{ht} > p^l_{ht} > \delta p^i_{ht}, \\ 0 & \text{if } p^l_{ht} \leq \delta p^i_{ht}. \end{cases}$$

Figure 4 provides a graphical illustration of the seller' s payoff schedule under the counterfactual contract.

[13]If data on iBuyers' resale strategies become available, future research could extend the model to incorporate strategic decisions such as resale timing and inventory management. At present, there is limited empirical evidence on the resale and inventory management strategies employed by iBuyers.

[14]In Appendix H.3, I extend the contract to allow for partial revenue sharing, i.e., the seller receives a fraction $\alpha(p^l_{ht} - \delta p^i_{ht})$. In my evaluation, setting $\alpha = 1$ (full sharing) always yields the highest revenue.

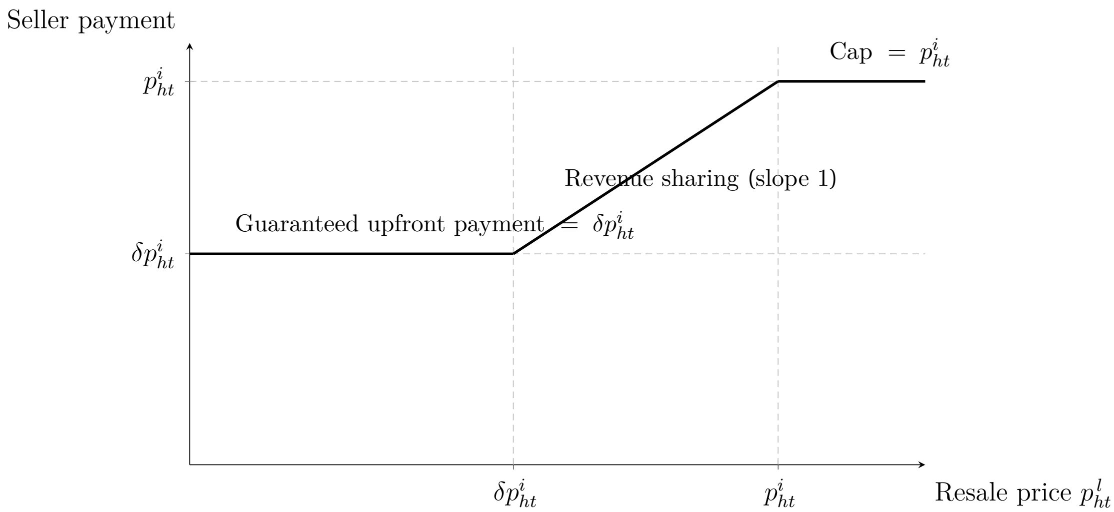

Figure 4: Seller payment schedule under the counterfactual contract
Note: The x-axis plots the resale price p^l_{ht}; the y-axis plots the seller' s payment. For $p^l_{ht} \leq \delta p^i_{ht}$, the seller receives the guaranteed upfront payment δp^i_{ht}. For $\delta p^i_{ht} < p^l_{ht} < p^i_{ht}$, the seller' s payment increases one-for-one with the resale price (revenue sharing). For $p^l_{ht} \geq p^i_{ht}$, the payment is capped at p^i_{ht}. The figure illustrates the piecewise schedule corresponding to the contract in the text; $\delta \in [0, 1]$ denotes the upfront-payment share and p^i_{ht} the iBuyer' s original offer.

Then, the iBuyer' s revenue per transaction under the counterfactual contract is given by:

$$\begin{cases} p^l_{ht} - p^i_{ht} & \text{if } p^l_{ht} > p^i_{ht}, \\ 0 & \text{if } p^i_{ht} > p^l_{ht} > \delta p^i_{ht}, \\ p^l_{ht} - \delta p^i_{ht} & \text{if } p^l_{ht} \leq \delta p^i_{ht}. \end{cases}$$

8.1.1 Individual Seller Payoffs Under the Counterfactual Contract

Following the payment structure specified above, the actual payoff to individual sellers depends on how hassle costs are interpreted. In the main text, I focus on the case where hassle costs arise entirely from time constraints (e.g., liquidity needs or urgency). Under this interpretation, any payment that is delayed until after the resale incurs the full hassle cost c_{ht}. The seller' s payoff is given by:

$$\delta p^i_{ht} + \begin{cases} \frac{p^i_{ht} - \delta p^i_{ht}}{c_{ht}} & \text{if } p^l_{ht} > p^i_{ht}, \\ \frac{p^l_{ht} - \delta p^i_{ht}}{c_{ht}} & \text{if } p^i_{ht} > p^l_{ht} > \delta p^i_{ht}, \\ 0 & \text{if } p^l_{ht} \leq \delta p^i_{ht}. \end{cases}$$

Here, the post-resale payment is discounted by the hassle cost, as it mimics the burden experienced when selling on the open market.

This interpretation presents the most pessimistic scenario for expected iBuyer profits, as all delayed payments incur hassle costs. An alternative, more optimistic interpretation —where hassle costs are psychological burdens unrelated to timing—is provided in Appendix H.2. Any realized seller payoff will lie between these two extremes.

In the next subsection, I theoretically demonstrate cream-skimming. I then complement the theoretical results with a quantitative evaluation of profitability, calibrated to the estimated variance of the idiosyncratic market price error.

8.1.2 Theory of cream skimming

This subsection examines cream-skimming in a stylized environment where sellers discount any delayed payments and have full information about resale prices. These assumptions isolate the role of selection and enable a clear theoretical characterization of profitability under the counterfactual contract. For robustness, I also derive parallel cream-skimming results under an alternative interpretation—where hassle costs are purely psychological and sellers face no discounting of delayed payments—in Appendix H.2.

Building on the framework of Rothschild and Stiglitz [1976], I show that cream-skimming is achievable in the limiting case where the only source of uncertainty in the market price is unobserved house quality, and the idiosyncratic pricing error approaches zero. As emphasized by Rothschild and Stiglitz [1976], sustaining profitability under such mechanisms requires knowledge of the distribution of private information.

Assumption 5 (No idiosyncratic market pricing error) $\varepsilon^l_{ht} \to 0$

This assumption implies that individual sellers know the resale price with certainty. The logic of this section remains applicable as long as the market pricing error is sufficiently small.

I begin by introducing the per-transaction profit function. This profit depends on the realized resale price, the structure of the payment contract, and the unobserved quality of the house. The corresponding expressions for the iBuyer's ex-ante expected profit, information set, and selling probability are provided in Appendix H.1.

iBuyer Per-transaction Profit. The iBuyer profit from transaction (h, t) under the counterfactual contract is:

$$\pi(\delta, \xi_h) = p^l_{ht} - \left(\delta p^i_{ht} + \mathbf{1}_{\{p^i_{ht} > p^l_{ht} > \delta p^i_{ht}\}} (p^l_{ht} - \delta p^i_{ht}) + \mathbf{1}_{\{p^l_{ht} > p^i_{ht}\}} (1-\delta) p^i_{ht} \right),$$

where $\delta \in [0, 1]$ is the proportion of the upfront payment, and $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.

Proposition 4 (Cream-Skimming) *Under Assumption 5, the derivative of the selling probability with respect to the upfront payment share δ satisfies:*

$$\frac{\partial \mathbb{P}(\textit{Sell to iBuyer} \mid \xi_h)}{\partial \delta} > 0 \quad \textit{if and only if} \quad \delta p^i_{ht} > p^l_{ht},$$

$$\frac{\partial \mathbb{P}(\textit{Sell to iBuyer} \mid \xi_h)}{\partial \delta} = 0 \quad \textit{if and only if} \quad \delta p^i_{ht} \leq p^l_{ht}.$$

Decreasing δ reduces the attractiveness of the contract for lower-quality homes (which are more likely to have low resale value), thereby screening out sellers of lower-quality houses. This improves the selection of homes sold to the iBuyer.

8.1.3 Quantitative Evaluation of Contract Counterfactuals

To assess iBuyer profitability using the data, I use the estimated demand parameters, along with the parameters of the unobserved quality and hassle cost distributions. Let $\mathcal{H}$ denote the set of all houses transacted in the market after the iBuyer' s entry, and $\mathcal{T}$ the set of corresponding time periods. I simulate expected profits under the counterfactual contract for each value of $\delta \in [0, 1]$, building on the iBuyer Per-transaction Profit function introduced above, but incorporating idiosyncratic pricing error ($\varepsilon^l_{ht} > 0$).

Figure 5 presents expected iBuyer profits across different values of δ. In the case of Charlotte, I simulate expected profits across 43,911 housing transactions that occurred after the iBuyer' s market entry, computing counterfactual profits on an evenly spaced grid of $\delta \in [0, 1]$.

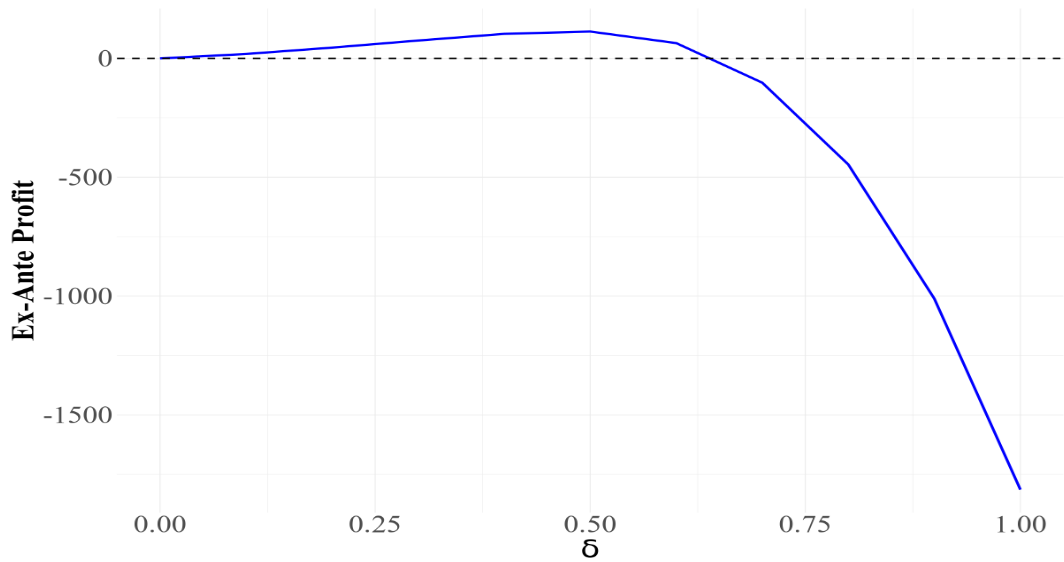

Figure 5: Expected Profit (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of $100,000) under a counterfactual contract with varying δ, the fraction of the original iBuyer offer price p_{ht}^{i} paid upfront ($\delta = 1$ corresponds to the current contract structure). Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte. This evaluation interprets hassle costs in the least favorable way for the iBuyer; the most optimistic case is shown in the appendix.

The first notable finding is that the iBuyer earns negative expected profits at $\delta = 1$, which corresponds to the current contract structure (i.e., full upfront payment and no conditional revenue sharing), assuming away any dynamic concerns such as inventory costs or resale timing.

In contrast, by reducing the upfront payment and introducing conditional revenue sharing (i.e., lowering δ), the iBuyer can achieve significantly higher expected profits. This is consistent with the cream-skimming mechanism identified in the theory: lower values of δ selectively attract higher-quality sellers, improving the average resale margin. There exists a range of δ values for which expected profit becomes strictly positive, suggesting that alternative contract structures can mitigate adverse selection and make the iBuyer model financially sustainable in static settings. These results are based on the least favorable interpretation of hassle costs for the iBuyer. In Appendix H.2, I show that under more optimistic assumptions about hassle costs, the qualitative conclusions remain unchanged.

In Appendix H.3, I extend the numerical simulations by relaxing the assumption of full conditional revenue sharing for the median-type house. In this extension, only an α fraction of the resale gain is shared with the seller, with $\alpha \in [0, 1]$. This adjustment reduces the seller's payout when the resale price lies between δp_{ht}^{i} and p_{ht}^{i}. Although this lowers payments in these intermediate cases, the associated decline in seller participation more than offsets the savings. As a result, the iBuyer' s expected profit decreases as α falls. The central conclusion from this section—that cream-skimming can mitigate adverse selection—remains unchanged. Full derivations, simulation details, and results for various interpretations of hassle costs are provided in the appendix.

8.2 Price Design with Unstructured Data

As an alternative or addition to contract design, pricing can be improved by incorporating additional data sources. I incorporate unstructured data from housing listings—specifically, textual descriptions—to increase the precision of the iBuyer pricing algorithm. Unobserved quality, as defined previously, encompasses latent attributes such as mood, natural light, and perceived desirability, which are difficult to capture using structured data alone. Unstructured listing text can partially reflect these subjective attributes, thereby helping to capture variation in unobserved quality.

Leveraging recent advances in large language models (LLMs), I develop a simple approach to incorporating unstructured housing data and demonstrate its potential to enhance profitability. First, I use an LLM to construct a one-dimensional projection of unstructured housing data, facilitating ease of use. Second, I show that including this projected variable as a covariate improves the precision of the iBuyer pricing model. Third, I evaluate the resulting gains and incorporate the enhanced pricing algorithm into the counterfactual contract design introduced earlier, assessing how the combination increases expected profit. In this evaluation, the content of listing descriptions is treated as exogenous, meaning that sellers are assumed not to strategically alter their wording in response to iBuyer pricing. Under this assumption, text data reveal information about unobserved housing quality without feedback effects. Consequently, the results would be interpreted as an upper bound on the potential value of incorporating listing text.

Additionally, incorporating LLM-based measures offers two advantages over using lagged transaction prices as explanatory variables. Even abstracting from the endogeneity concern that arises when directly including past prices, the LLM-based approach provides an additional benefit by projecting historical prices into the semantic space of housing descriptions. This projection filters out noise in historical prices that reflect market fluctuations rather than intrinsic property quality, particularly when market and property conditions are not jointly observed. The comparative performance of models using the LLM score versus lagged prices is detailed in Appendix H.7. Furthermore, unstructured listing data offer broader coverage than lagged prices. Historical prices are available only for properties with completed past transactions, whereas listing descriptions are observed for a wider set of homes—including those sold, pending, or delisted shortly after posting. Consequently, the LLM approach can exploit a richer information set, capturing signals from listings that would otherwise be excluded from traditional price-based models.

To maintain consistency with the earlier analysis in Section 8.1, I adopt the same assumptions: the iBuyer resells each property at the market price immediately after purchase, abstracting away from dynamic factors such as resale timing, inventory costs, and resale probability. This static framework isolates the impact of pricing precision and enables a clean comparison of profits under different pricing algorithms. I also retain the interpretation of seller payoffs established in Section 8.1.1, evaluating outcomes under both timing-related and non-timing-related interpretations of hassle costs. This consistency ensures that any observed changes in profitability are attributable solely to the improved pricing precision from incorporating unstructured data via the LLM-enhanced model.

8.2.1 Leveraging Large Language Models to Process Unstructured Housing Data

I focus on the "Public Remarks", which is a text field containing publicly available descriptions intended for prospective buyers. For example, a listing might state: "Experience serene living in one of the area's most beautiful properties, featuring a private swimming pool and breathtaking views. Nestled on 5.2 wooded acres, this peaceful retreat offers the perfect blend of privacy and natural beauty—a truly exceptional lifestyle opportunity."[15]

Due to substantial variation in listing descriptions, projecting the "Public Remarks" text into a compact, low-dimensional space is challenging. For both TF-IDF embeddings (restricted to 100 features) and Sentence Transformer embeddings, the first few principal components explain only a small share of the total variance, underscoring the difficulty of capturing meaningful structure in a compact, low-dimensional representation. Cumulative explained variance plots for both embedding methods are provided in Appendix H.4.

Instead, I generate a one-dimensional projection of this information using the Llama-3.2-1B-Instruct model developed by Meta AI (Meta, 2024). I explore three approaches: (i) prompt engineering, (ii) LoRA fine-tuning, and (iii) feature extraction with subsequent regression.

For prompt engineering, I use the instruction-tuned version of the model with a deterministic temperature setting to produce factual, consistently formatted outputs.[16] The prompt

[15]This public remark is a fabricated example used for illustrative purposes to preserve privacy.

[16]For instance, given the instruction "You are a helpful AI that rates how likely a statement is true from 0 (definitely false) to 1 (definitely true), including intermediate values like 0.5 for uncertain statements,"

is: "You are a real estate investment assistant. Read the following listing description and rate from 0 to 1 how likely this is a good opportunity to buy a house. Rate 1 for definitely a good deal, 0 for a bad deal, and values in between for uncertain or mixed listings." The resulting score serves as a compact proxy for subjective desirability.

For LoRA fine-tuning, I adapt the same Llama-3.2-1B-Instruct model to predict transaction prices. The input consists of the "Public Remarks" text and the city name, and the target output is the sale price (inflation-adjusted to March 2023 CPI levels). LoRA fine-tuning is chosen for computational efficiency while retaining most of the base model' s capabilities.

For feature extraction with regression, instead of fine-tuning, I pass the same input text and city information through the Llama model and extract the second-to-last hidden layer as a high-dimensional semantic representation. These features are then used in an Elastic Net regression—a regularized linear model combining Lasso and Ridge penalties—to predict the adjusted sale price.

For both the fine-tuning and feature-extraction approaches, I generate LLM-based price predictions using only the "Public Remarks" and city. I then convert these predicted prices into city-specific deciles and treat these deciles as a pseudo-score of the property.

In all three approaches—prompt engineering, LoRA fine-tuning, and feature extraction —the resulting LLM-derived score is added as an additional covariate to the iBuyer pricing model. I then evaluate whether this improves predictive accuracy and expected profit. The improvements in profitability from the fine-tuning and feature-extraction methods are broadly similar to those from the simpler prompt-engineering approach. For brevity, I present detailed counterfactual simulation results for the prompt-engineering method in the main text and relegate the results for the other two methods to the Appendix H.5 and H.6.

Using the historical average of the Public Remarks score for each house in Charlotte, I train a linear model on data from the pre-iBuyer entry period. Table 11 presents a comparison between models estimated with and without the inclusion of the LLM-based text score. Since some houses lack Public Remarks, approximately 300 out of 118,604 transactions are excluded from the model.

Table 11 shows that including the LLM-based text score improves the model' s explanatory power, increasing the R^2 from 0.909 to 0.913 while leaving the coefficients on other covariates largely stable. The coefficient on the LLM-based text score is 0.43, implying that, all else equal, a house with the highest possible score (1) is associated with a 43% higher

and the statement "The Earth orbits the Sun," the model returns a score of 0.95—close to 1. This slightly conservative score reflects the model' s tendency to hedge predictions and avoid absolute certainty, even for well-established facts. Such cautious behavior is common in instruction-tuned models.

transaction price compared to a house with the lowest possible score (0).

	Without Score (Pre-iBuyer Period)	With Score (Pre-iBuyer Period)
Intercept	0.029 (0.100)	0.182 (0.098)
Log living area square feet	1.106 (0.005)***	1.037 (0.005)***
Bedroom number	−0.093 (0.002)***	−0.082 (0.002)***
Bathroom number	0.006 (0.002)**	0.005 (0.002)**
Building age	0.002 (0.000)***	0.002 (0.000)***
Garage dummy	0.045 (0.007)***	0.060 (0.007)***
Heating dummy	0.209 (0.023)***	0.147 (0.022)***
SeasonalFE_2nd quarter	0.021 (0.003)***	0.020 (0.003)***
SeasonalFE_3rd quarter	0.001 (0.003)	0.001 (0.003)
SeasonalFE_4th quarter	−0.002 (0.003)	−0.002 (0.003)
30-year mortgage rate	0.004 (0.002)	0.004 (0.002)
Federal funds rate	0.009 (0.001)***	0.007 (0.001)***
Log CPI index	−0.214 (0.023)***	−0.299 (0.023)***
Log CS index	0.229 (0.008)***	0.243 (0.008)***
LLM-based text score	–	0.430 (0.007)***
R^2	0.909	0.913
Adj. R^2	0.909	0.913
Num. obs.	118297	118297

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 11: Effect of LLM-based text score on log transaction price estimates
Note: This table presents regression results for transaction prices in the pre-iBuyer entry period in Charlotte. The dependent variable is the log of the transaction price, expressed in $100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. The first model excludes the AI-generated score, while the second includes it as an additional predictor. Standard errors are reported in parentheses.

8.2.2 Quantitative Evaluation of Combined Price and Contract Counterfactuals

This section builds on the previous counterfactual framework, retaining the same structure for computing expected profits. The key difference lies in the pricing algorithm: I augment the iBuyer pricing model by incorporating the LLM-based text score, using estimates reported in the second column of Table 11. This score, derived from unstructured listing descriptions, may partially capture dimensions of unobserved quality not reflected in structured variables. I refer to this augmented specification as the counterfactual pricing algorithm.

Using this updated model, I simulate expected profits across a sample of 43,911 housing transactions in Charlotte that occurred after the iBuyer' s market entry—the same sample used in the prior analysis—to ensure a direct comparison of results. As before, I evaluate expected profits across values of $\delta \in [0, 1]$, combining the updated pricing rule with the previously defined contract counterfactuals and following the iBuyer Per-transaction Profit function.

Figures 6 plots the resulting profit curves for each value of δ. The solid lines show expected profits under the counterfactual pricing algorithm, while the dotted lines reproduce the expected profits from the original iBuyer pricing model, as reported in the previous section.

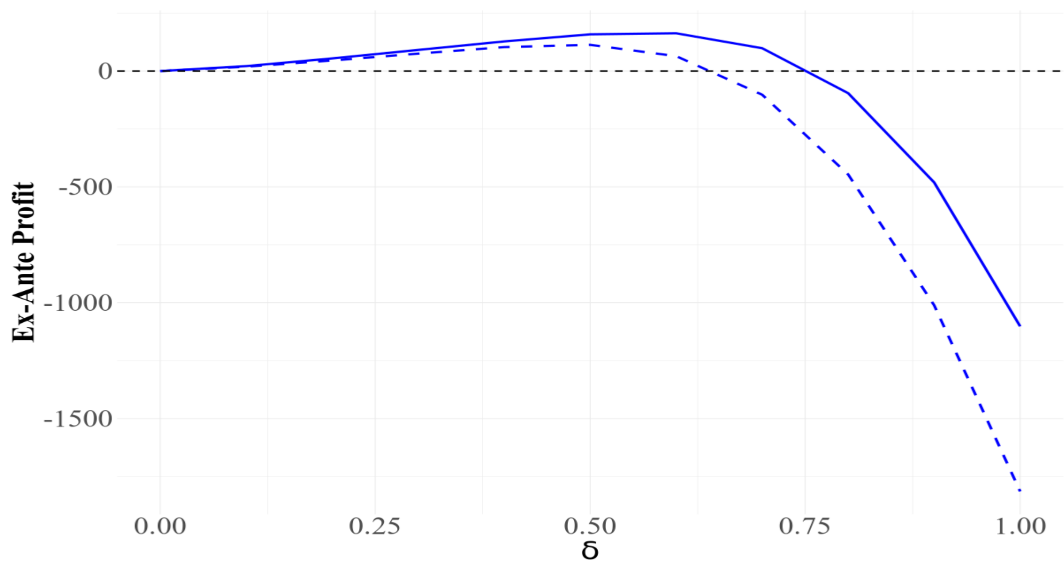

Figure 6: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of $100,000) under a counterfactual contract with varying δ, the fraction of the original iBuyer offer price p^i_{ht} paid upfront ($\delta = 1$ corresponds to the current contract structure). The solid line represents profits simulated using a counterfactual LLM-based pricing algorithm; the dotted line reproduces profits from the original iBuyer pricing model, as reported earlier. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte. This evaluation interprets hassle costs in the least favorable way for the iBuyer; the most optimistic case is shown in the appendix.

The current iBuyer contract, in which the full payment is made upfront, corresponds to the case $\delta = 1$. Incorporating unstructured listing data through the LLM-based text score in the counterfactual pricing algorithm increases expected profits.

Another notable pattern in Figures 6 is that the counterfactual pricing algorithm consistently outperforms the original iBuyer pricing model in terms of expected profits for all $\delta \in (0, 1]$. The profit gains are especially pronounced as δ approaches 1. This pattern reflects the cream-skimming mechanism discussed in Section 8: when δ is low, the contract tends to attract sellers of higher-quality homes, reducing the incremental benefit of incorporating the LLM-based text score. In contrast, when δ is high and adverse selection is more severe, the unstructured text score adds more value by helping recover unobserved quality, thereby improving pricing precision and expected profits.

Similar to the previous section, these results are based on the least favorable interpretation of hassle costs for the iBuyer. The most favorable case is presented in Appendix H.2, and the qualitative conclusions remain unchanged.

9 Conclusion

This paper investigates why iBuyers—firms that offer instant home purchases using big-data-driven pricing algorithms—struggle with profitability despite access to extensive housing data and computational tools. I identify two key sources of seller-side private information that iBuyers cannot fully observe or contract on: the unobserved quality of the home (e.g., features like natural lighting or layout flow that are difficult to encode in data) and the seller' s hassle costs associated with traditional home selling. Asymmetric information along these dimensions distorts seller selection into the iBuyer channel, reducing profit margins. By integrating structural estimation with insights from the adverse selection literature and by incorporating large language models to leverage unstructured listing data, I show both the limits of algorithmic pricing under private information and the potential of contract design and richer data to mitigate these frictions—offering guidance for researchers and practitioners in tech-driven marketplaces. I substantiate these arguments with descriptive evidence and structural analysis.

Descriptive evidence shows that sellers with higher hassle costs—proxied by long-distance moves—are more likely to sell to iBuyers, suggesting some success in attracting that segment. However, iBuyers also face persistent difficulties in pricing unobserved home characteristics. I show that individual-to-individual transactions rely more heavily on unobservables, and that iBuyers earn consistently lower resale margins—even after controlling for holding period—suggesting that they may be overpaying for lower-quality homes.

To formalize these observations, I estimate a structural model that captures the joint distribution of unobserved house quality and hassle costs. The results show that unobserved quality can significantly erode iBuyer margins, whereas higher hassle costs continue to influence sellers to choose the iBuyer option. Building on adverse selection theory, I propose a counterfactual contract where iBuyers lower the upfront payment but share resale revenue, thereby cream-skimming higher-quality homes. Numerical simulations based on the estimated distributions confirm that a carefully calibrated revenue-sharing mechanism can improve iBuyer profitability by limiting overpayment for subpar properties. As a complementary strategy, I enhance the iBuyer' s pricing model by incorporating a one-dimensional projection of unstructured listing text, constructed using a large language model (LLM). This additional input improves pricing precision and further increases expected profitability

—particularly in settings where contract-based cream-skimming is less effective.

These results have broader implications beyond housing. As algorithmic pricing becomes increasingly prevalent in sectors like insurtech, fintech, and e-commerce, markets must account for private information that remains difficult to observe—even with large datasets. Failure to do so may lead to selection problems that erode the performance of automated pricing models.

A Appendix: Steps to Selling a Home to iBuyers

Using the Wayback Machine, I examine the procedure for selling to an iBuyer in 2018 through their archived websites, when iBuyers began actively operating in my sample. This process is illustrated for Opendoor and Offerpad in Figures 7 (Internet Archive [2018b]) and 8 (Internet Archive [2018a]). iBuyer strategies may have evolved over time and could differ from those in 2025, which falls outside my sample period.

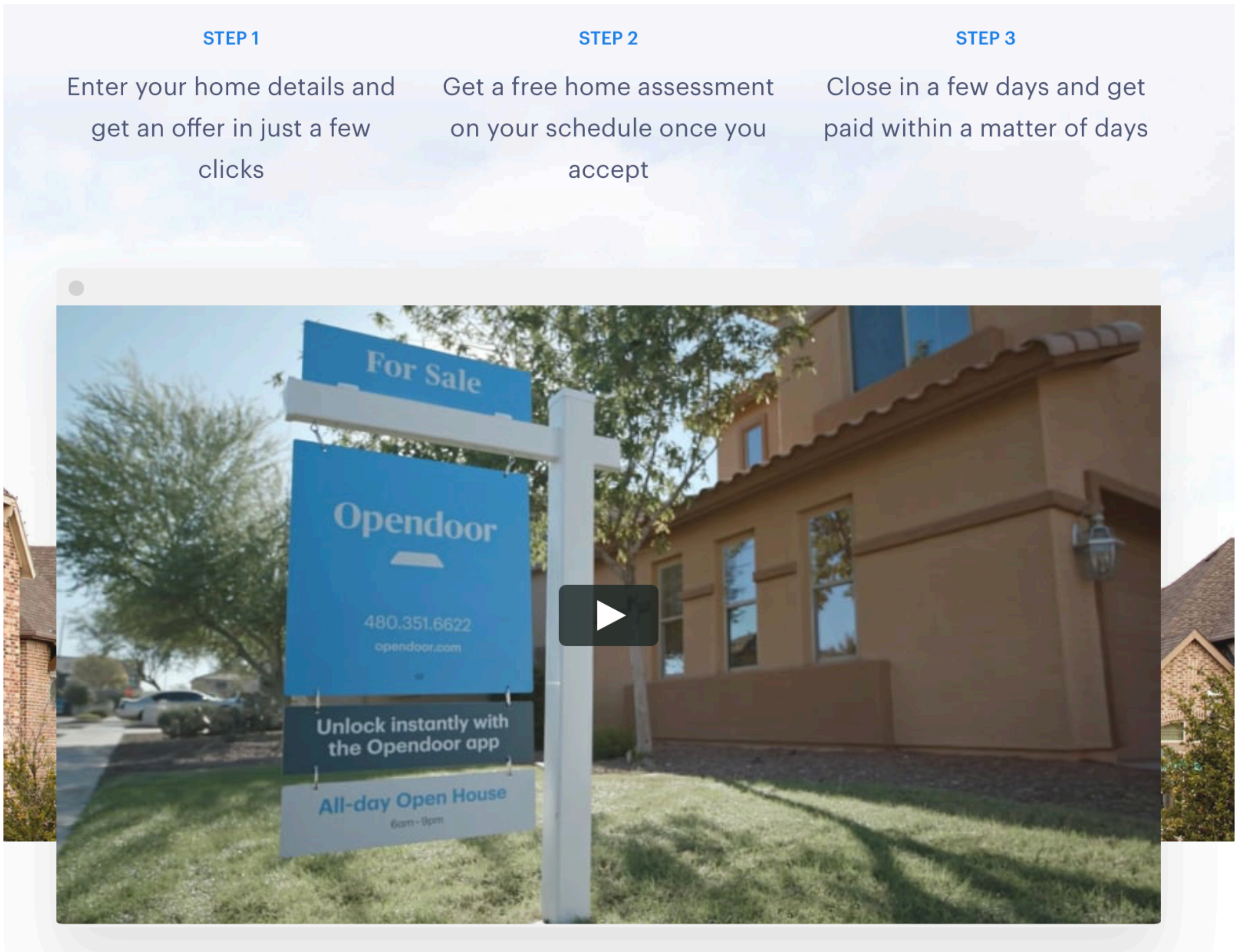

Figure 7: Selling a House to Opendoor
Note: This figure shows Opendoor's step-by-step selling process as it appeared in 2018. The procedure emphasizes speed, simplicity, and limited seller effort—core components of minimizing hassle costs.

Figure 8: Selling a House to Offerpad
Note: Offerpad' s 2018 process outlines key steps including online valuation, home inspection, and direct offer acceptance. These features reflect the company' s early positioning as a hassle-reducing alternative to traditional home sales.

B Appendix: Payments to iBuyers

Figures 9, 10, and 11 illustrate the additional types of payments sellers make to iBuyers, based on information from Opendoor and Offerpad websites. The first figure is derived from Opendoor’ s archived 2019 website (Internet Archive [2019]), while the latter two are based on current website disclosures as of 2025 (Opendoor Technologies Inc. [2025], Offerpad LLC [2025]).

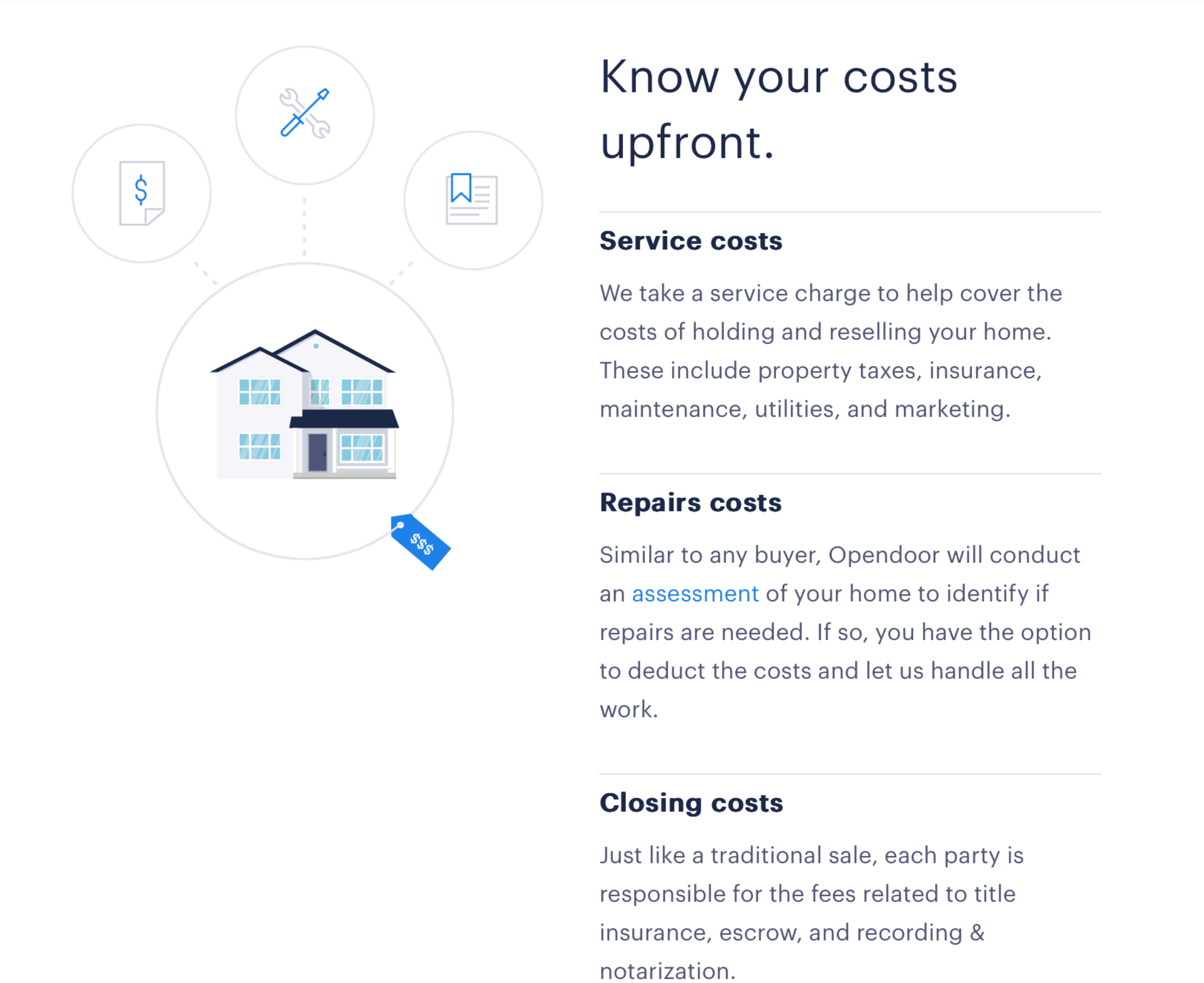

Figure 9: Breakdown of Payments to iBuyers (Opendoor, 2019)
Note: This figure summarizes seller charges reported by Opendoor, based on its website archived in 2019, with additional references to more recent disclosures in Opendoor’ s 10-K filing (Opendoor Technologies Inc. [2023]). The service fee (typically 5%) is comparable to traditional agent commissions (5–6%) and covers holding and resale-related costs such as property taxes, insurance, and marketing. Repair costs reflect required pre-listing fixes. Closing costs include standard administrative items such as escrow and title fees.

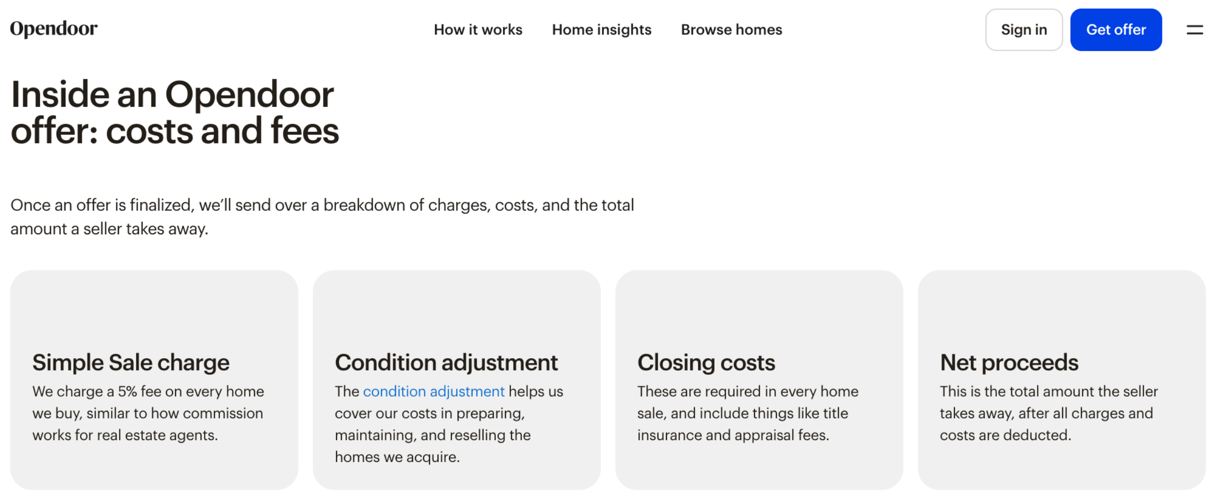

Figure 10: Payments to Opendoor (2025)
Note: This figure is based on seller cost breakdowns described on Opendoor' s official website as of 2025. Categories remain consistent with earlier disclosures: a 5% service fee, post-inspection repair costs, and standard closing costs.

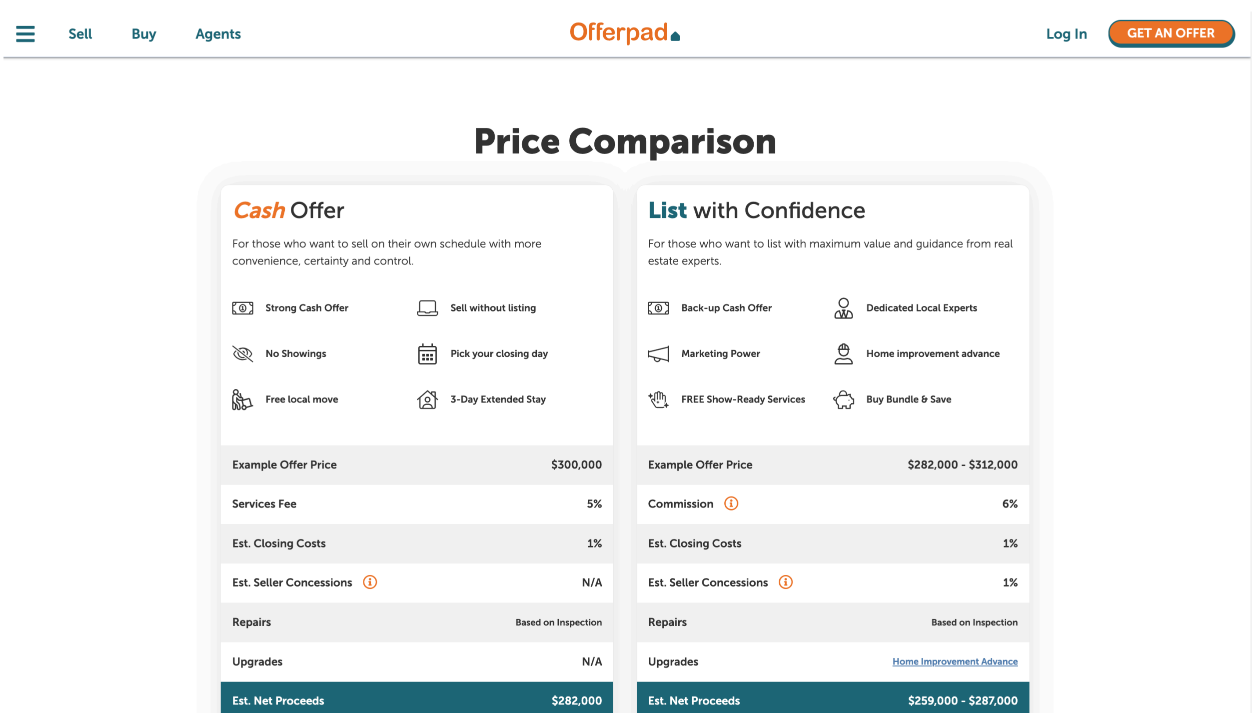

Figure 11: Payments to Offerpad (2025)
Note: Based on seller-facing cost information provided on Offerpad' s website as of 2025. The fee structure similarly includes a 5% service charge, estimated repair costs after home inspection, and closing costs. These components are aligned with those in traditional home sales but bundled under the iBuyer transaction model.

C Appendix: iBuyer Website Advertisements

This appendix provides screenshots from the homepages of major iBuyers, Offerpad and Opendoor, as accessed in January 2025. These advertisements are intended to highlight how iBuyers explicitly market their services as a way to avoid the hassle and uncertainty of selling on the traditional market.

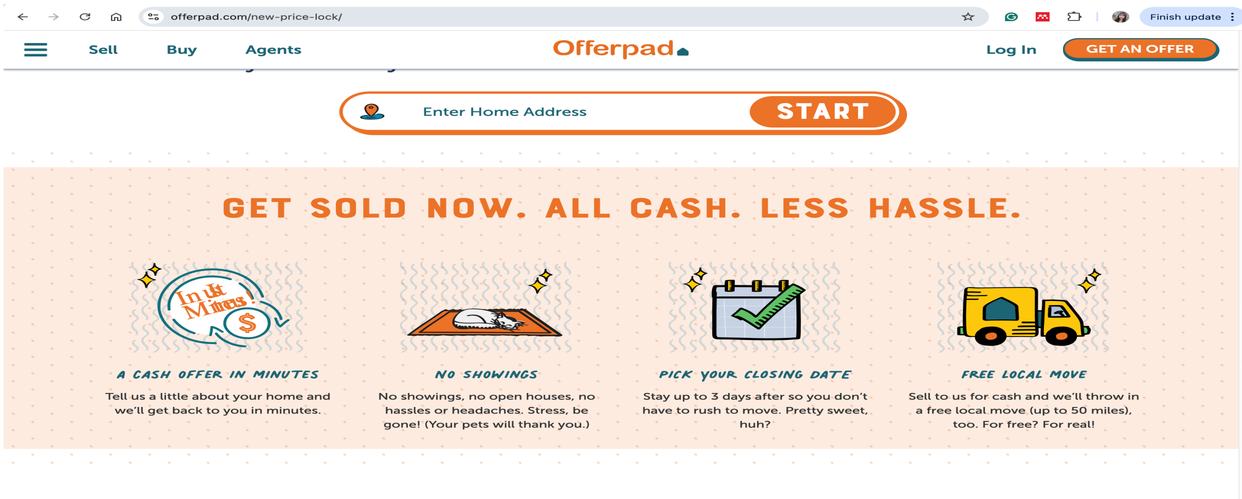

Figure 12: Offerpad's Website Homepage (Accessed January 2025)
Note: The website emphasizes convenience and speed, appealing to sellers seeking to avoid time-consuming or emotionally taxing steps in the traditional home-selling process.

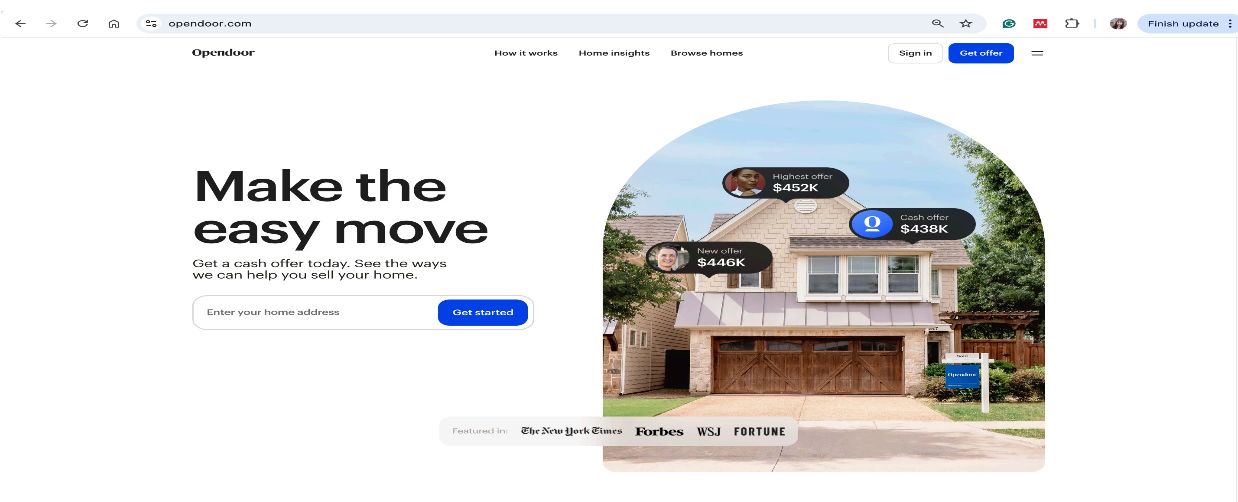

Figure 13: Opendoor's Website Homepage (Accessed January 2025)
Note: Opendoor markets its service by highlighting simplicity and certainty, consistent with targeting sellers with high hassle costs.

D Appendix: R^2 Comparison with City Fixed Effect Interactions

This appendix presents robust R^2 results from models that include interaction terms between city fixed effects and all structural housing characteristics, as well as market characteristics. These specifications allow the effects of housing attributes and market condition to vary by city, improving model fit and capturing local market heterogeneity.

	Log iBuyer buying price	Log individual buying price
City FE $\times$ House Characteristics	Yes	Yes
City FE $\times$ Market Characteristics	Yes	Yes
R^2	0.97	0.91
Adj. R^2	0.97	0.91
Num. obs.	$10,911$	$380,650$

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 12: R^2 comparison (2015 - 2022)
Note: This table presents regression results including interaction terms between city fixed effects and both structural housing and market characteristics. These interactions allow the influence of housing and market variables to vary across cities, capturing local market heterogeneity and improving model fit. The dependent variables are the log of iBuyer and individual transaction prices, expressed in $100,000s.

E Appendix: Resale Margins

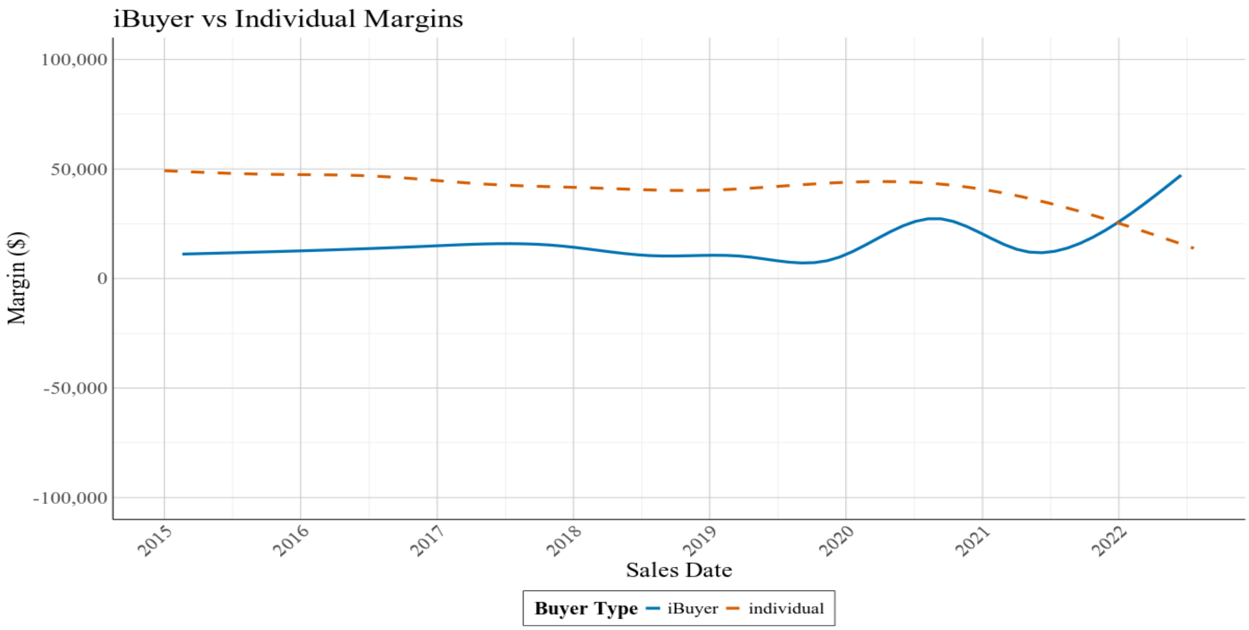

Figure 14: Comparing absolute margins
Note: This figure presents the average absolute resale margin in dollars, adjusted to 2023 CPI. Smoothed trends are estimated using Generalized Additive Models.

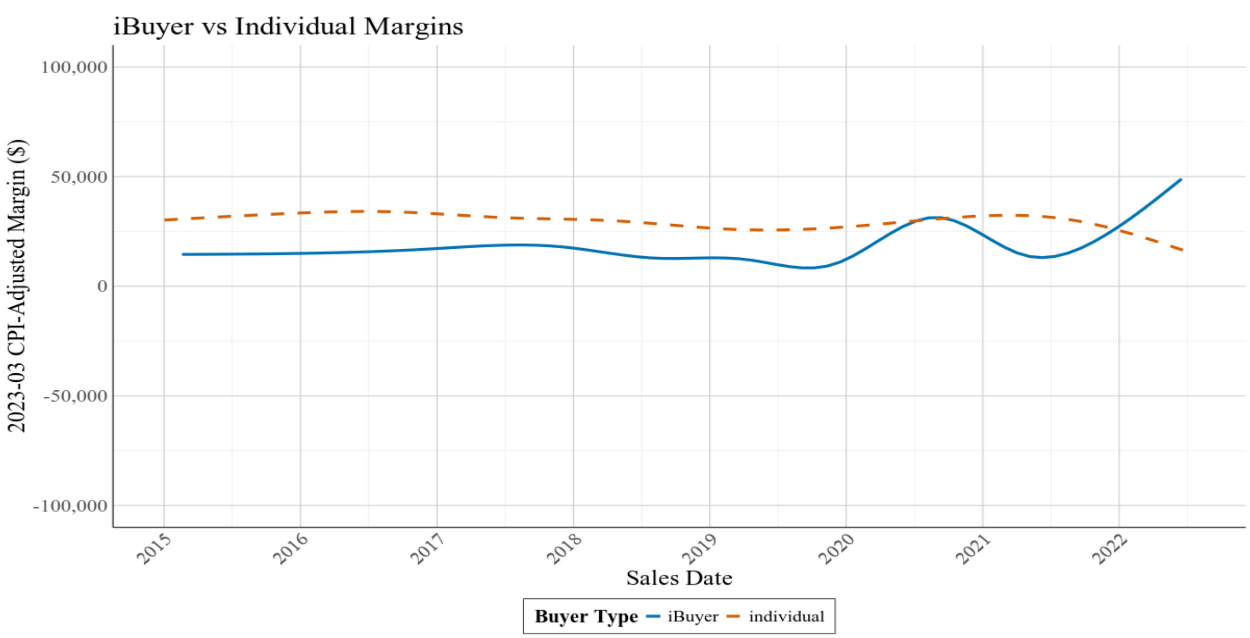

Figure 15: Comparing absolute margins (Short-term transaction)
Note: This figure restricts to properties resold within one year. Absolute resale margins are in 2023 CPI-adjusted dollars. Trends are smoothed using Generalized Additive Models.

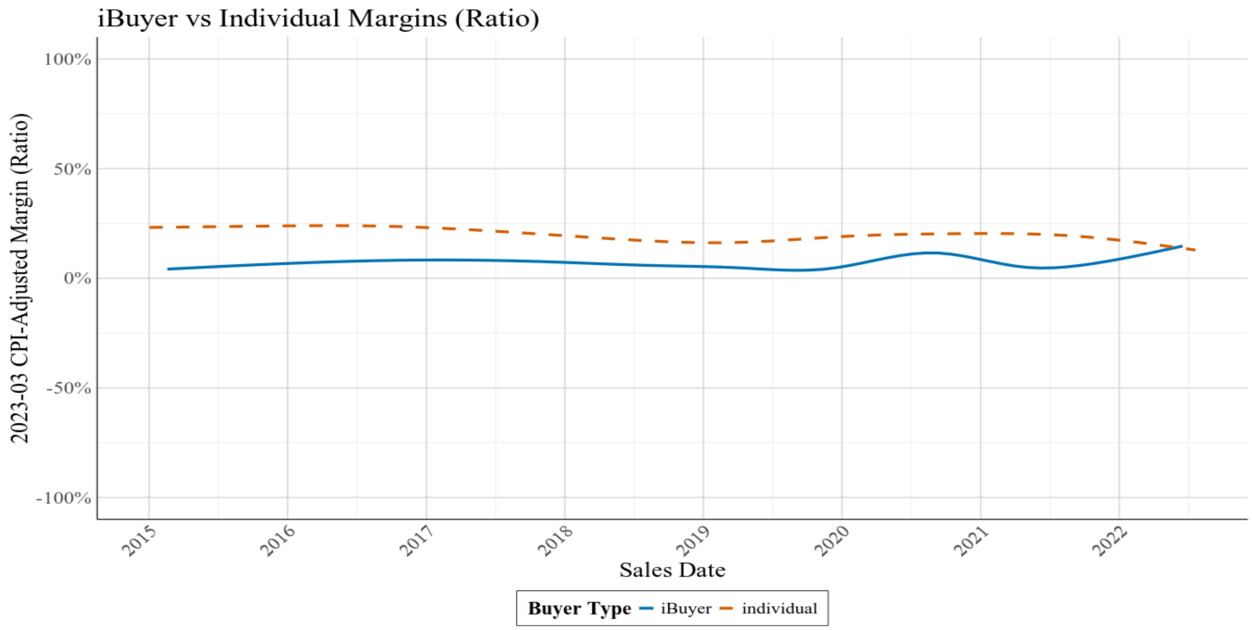

Figure 16: Comparing ratio margins (Short-term transaction)
Note: Ratio margins for properties resold within one year of purchase. All prices are adjusted to 2023 CPI. Smoothed trends use Generalized Additive Models.

F Appendix: Proofs

Proof of Proposition 1. For β^l, identification follows from standard random effects estimation via Generalized Least Squares or Maximum Likelihood Estimation. Given $\mathbb{E}\left[\xi_h \mid X_{ht}\right] = 0$ and $\mathbb{E}\left[\varepsilon^l_{ht} \mid X_{ht}, \xi_h\right] = 0$, β^l can be consistently recovered from the partial correlation between X_{ht} and $\log p^l_{ht}$, provided that X_{ht} for market transaction data is not collinear across or within houses.

For β^i, identification follows from standard linear model estimation via Ordinary Least Squares or Maximum Likelihood. Given $\mathbb{E}\left[\varepsilon^i_{ht} \mid X_{ht}\right] = 0$, β^i can be consistently recovered from the partial correlation between X_{ht} and $\log p^i_{ht}$, assuming non-collinearity of X_{ht} for iBuyer transaction data. ■

Proof of Proposition 2. From the exogeneity assumptions, the conditional variance of the residual is $\mathrm{Var}(\log p^l_{ht} - X_{ht}\beta^l | X_{ht}) = \mathrm{Var}(\xi_h + \varepsilon^l_{ht} | X_{ht}) = \sigma^2_\xi + \sigma^2_{\varepsilon^l}$. Because ξ_h is constant over time for a given house h, the within-house covariance of residuals across time identifies σ^2_ξ. Therefore, if there are repeated observations for some houses in the market transaction data, σ^2_ξ is identified. ■

Proof of Proposition 3. The decision to sell to an iBuyer implies that $\log c_{ht} - \xi_h > \tilde{\Delta}_{ht}$, where

$$\tilde{\Delta}_{ht} \equiv \Delta_{ht}(\xi_h) - \xi_h = X_{ht}\beta^l + 0.5\sigma^2_{\varepsilon^l} - \log p^i_{ht}.$$

Hence, the observed choice data $\{1_{\log c_{ht} > \Delta_{ht}(\xi_h)}\}$ arises from a latent threshold model based on $\log c_{ht} - \xi_h$.

Assume that $(\log c_{ht}, \xi_h)$ is jointly normal. (The extension to the mixture case with parameter a is straightforward.) Under this assumption, the difference $\log c_{ht} - \xi_h$ is normally distributed, with its mean and variance depending on the parameters $(\mu_{\log c}, \sigma_{\log c}, \rho)$.

Variation in $\tilde{\Delta}_{ht}$ across houses and time identifies how the threshold for $\log c_{ht} - \xi_h > \tilde{\Delta}_{ht}$ shifts. If there is sufficient variation in $\tilde{\Delta}_{ht}$ driven by X_{ht} in the sample, then I can recover the mean and variance of $\log c_{ht} - \xi_h$, which in turn allows identification of the distribution parameters $(\mu_{\log c}, \sigma_{\log c}, \rho, a)$. ■

Proof of Proposition 4. Recall that the probability of selling to the iBuyer is given by:

$$\mathbb{P}(\text{Sell to iBuyer}|\xi_h) = (1-a)\cdot\left(1 - F_{c|\xi}\left(\frac{\delta p^i_{ht}}{p^l_{ht} - \mathbf{1}_{\{p^i_{ht} > p^l_{ht} > \delta p^i_{ht}\}}(p^l_{ht} - \delta p^i_{ht}) - \mathbf{1}_{\{p^l_{ht} > p^i_{ht}\}}(1-\delta)p^i_{ht}}\right)\right),$$

where the net payment is a function of δ, depending on the resale outcome.

In the limiting case $\varepsilon^l_{ht} \to 0$, the resale price p^l_{ht} becomes deterministic from the seller' s perspective. As such, the only scenario in which δ affects the seller' s utility is when the seller receives only the upfront payment — that is, when $p^l_{ht} \le \delta p^i_{ht}$, and the revenue-sharing component is zero.

In this case, increasing δ raises the upfront payment, increasing the total payoff and making the seller more likely to accept the iBuyer offer. Hence, the derivative with respect to δ is strictly positive.

In contrast, when $\delta p^i_{ht} < p^l_{ht}$, the seller anticipates receiving the full payoff (either p^i_{ht} or p^l_{ht}), and the total payment is invariant to δ. Thus, $\frac{\partial \mathbb{P}(\text{Sell to iBuyer}|\xi_h)}{\partial \delta} = 0$.

■

G Appendix: Simulated Maximum Likelihood Function

The details of the third step are outlined below. The key distributional assumption is the joint normality of $\log c$ and ξ. (The extension to the mixture case with parameter a is straightforward.)

$$\begin{pmatrix} \log c \\ \xi \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\log c} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\log c} & \rho\sigma_{\log c}\sigma_\xi \\ \rho\sigma_{\log c}\sigma_\xi & \sigma^2_\xi \end{pmatrix}\right)$$

Therefore, the conditional distribution of $z \equiv \log c - \xi$ given ξ is

$$z|\xi = \log c - \xi|\xi \sim N(\mathbb{E}\left[\log c - \xi|\xi\right], Var\left[\log c - \xi|\xi\right])$$

where

$$\mu_z \equiv \mathbb{E}\left[\log c - \xi|\xi\right] = \mu_{\log c} + \rho\frac{\sigma_{\log c}}{\sigma_\xi}(\xi - \mu_\xi) - \xi = \mu_{\log c} + \rho\frac{\sigma_{\log c}}{\sigma_\xi}\xi - \xi$$

and

$$\sigma^2_z \equiv Var\left[\log c - \xi|\xi\right] = \sigma^2_{\log c}(1 - \rho^2).$$

The conditional probability of selling house h to iBuyer at time t assuming joint normality, is given by the following equation:

$$q^z_{ht}(\xi_h) = 1 - \Phi\left(\frac{\left(X^\top_{ht}\beta^l + \xi_h + 0.5\sigma^2_{\varepsilon_l} - X^\top_{ht}\beta^i - 0.5\sigma^2_{\varepsilon_i}\right) - \mu_z}{\sigma_z}\right)$$

To incorporate the possibility that some sellers do not consider the iBuyer option at all, I extend the model to a mixture framework. In this case, the probability becomes:

$$q^z_{ht}(\xi_h) = (1 - a)\left(1 - \Phi\left(\frac{\left(X^\top_{ht}\beta^l + \xi_h + 0.5\sigma^2_{\varepsilon_l} - X^\top_{ht}\beta^i - 0.5\sigma^2_{\varepsilon_i}\right) - \mu_z}{\sigma_z}\right)\right)$$

where Φ is a standard normal distribution cdf.

Let $\mathbb{1}^{iBuyer}_{ht}$ represent an indicator function for selling house h to an iBuyer at time t. The equation is given by

$$L_{ht}(\xi_h) = q^z_{ht}(\xi_h)^{\mathbb{1}^{iBuyer}_{ht}}(1 - q^z_{ht}(\xi_h)^{1-\mathbb{1}^{iBuyer}_{ht}})$$

The simulated maximum likelihood model integrates $L_{ht}(\xi_h)$ over ξ_h, utilizing the estimated distribution parameter of ξ_h obtained in the first step.

H Appendix: Counterfactual Analysis supplementary materials

This appendix provides supplementary materials related to the counterfactual analysis. It contains additional derivations, model details, robustness exercises, and extensions (including an optimistic benchmark for the iBuyer), as well as simulation results that were omitted from the main text for brevity.

H.1 Formal Model Details

This appendix collects the formal definitions and derivations underlying the iBuyer' s profit function and seller decision problem.

iBuyer Ex-Ante Expected Profit. The ex-ante expected profit across all houses $\mathcal{H}$ and time periods $\mathcal{T}$ is given by:

$$\Pi(\delta) = \sum_{t\in\mathcal{T}} \sum_{h\in\mathcal{H}} \mathbb{E}_{\Omega_{iBuyer}} \left[\mathbb{P}(\text{Sell to iBuyer}|\xi_h) \cdot \pi(\delta, \xi_h)\right].$$

iBuyer Information Set. Ω_{iBuyer} denotes the iBuyer' s information set prior to the realization of the resale price. Formally:

$$\Omega_{iBuyer} = \{\nu_{ht}^l, p_{ht}^i, F_{c,\xi}\},$$

where $\nu_{ht}^l \equiv p_{ht}^l - \xi_h$ is the observable component of the resale price, p_{ht}^i is the offer price, and $F_{c,\xi}$ is the joint distribution of hassle cost and unobserved house quality.

Selling Probability to iBuyer. The probability that an individual seller of house h at time t chooses to sell to the iBuyer, conditional on unobserved house quality ξ_h, is given by:

$$\mathbb{P}(\text{Sell to iBuyer}|\xi_h) = (1-a)\cdot\Big(1-F_{c|\xi}\Big(\frac{\delta p_{ht}^i}{p_{ht}^l - \mathbf{1}_{\{p_{ht}^i > p_{ht}^l > \delta p_{ht}^i\}}(p_{ht}^l - \delta p_{ht}^i) - \mathbf{1}_{\{p_{ht}^l > p_{ht}^i\}}(1-\delta)p_{ht}^i}\Big)\Big),$$

where $F_{c|\xi}$ is the cumulative distribution function of the hassle cost c_{ht} conditional on house quality ξ_h, and $a \in [0, 1]$ is the probability that the seller does not consider the iBuyer option.

H.2 Alternative Interpretation of Hassle Costs: Psychological Burden

For completeness, I also present the most optimistic case for expected iBuyer profits. In this case, hassle costs are interpreted as psychological or emotional burdens—such as the stress of dealing with agents, staging, or uncertainty—rather than as arising from time constraints. Under this interpretation, the timing of payments is irrelevant, and the seller' s payoff is:

$$\delta p^i_{ht} + \begin{cases} p^i_{ht} - \delta p^i_{ht} & \text{if } p^l_{ht} > p^i_{ht}, \\ p^l_{ht} - \delta p^i_{ht} & \text{if } p^i_{ht} > p^l_{ht} > \delta p^i_{ht}, \\ 0 & \text{if } p^l_{ht} \leq \delta p^i_{ht}. \end{cases}$$

Here, the iBuyer fully alleviates the seller' s hassle, regardless of payment timing, since the seller avoids the psychological cost of handling the sale process. This interpretation therefore represents the most optimistic bound on expected profitability.

With this interpretation, the cream-skimming logic in the limiting case parallels Section 8.1.2. Note that Assumption 5 and the iBuyer profit function remain unchanged.

The probability of an individual seller to sell to the iBuyer of house h at time t conditional on unobserved house quality ξ_h is

$$\mathbb{P}(\text{Sell to iBuyer} \mid \xi_h) = (1-a)\cdot\Big(1-F_{c|\xi}\Big(\frac{\delta p^i_{ht} + \mathbf{1}_{\{p^i_{ht} > p^l_{ht} > \delta p^i_{ht}\}}(p^l_{ht} - \delta p^i_{ht}) + \mathbf{1}_{\{p^l_{ht} > p^i_{ht}\}}(1-\delta)p^i_{ht}}{p^l_{ht}}\Big)\Big).$$

Proposition 5 (Cream-skimming under the case of psychological hassle costs) *Under Assumption 5, the selling probability' s derivative with respect to δ satisfies:*

$$\frac{\partial \mathbb{P}(\textit{Sell to iBuyer} \mid \xi_h)}{\partial \delta} > 0 \quad \textit{iff} \quad \delta p^i_{ht} > p^l_{ht}$$

$$\frac{\partial \mathbb{P}(\textit{Sell to iBuyer} \mid \xi_h)}{\partial \delta} = 0 \quad \textit{iff} \quad \delta p^i_{ht} \leq p^l_{ht}.$$

Hence, the conclusion that lowering δ reduces participation by sellers of lower-quality homes continues to hold under this interpretation.

Proof. Recall that the probability of selling to the iBuyer is given by:

$$\mathbb{P}(\text{Sell to iBuyer}|\xi_h) = (1-a)\cdot\Big(1-F_{c|\xi}\Big(\frac{\delta p^i_{ht} + \mathbf{1}_{\{p^i_{ht} > p^l_{ht} > \delta p^i_{ht}\}}(p^l_{ht} - \delta p^i_{ht}) + \mathbf{1}_{\{p^l_{ht} > p^i_{ht}\}}(1-\delta)p^i_{ht}}{p^l_{ht}}\Big)\Big),$$

where the net payment is a function of δ, depending on the resale outcome.

In the limiting case $\varepsilon^l_{ht} \to 0$, the resale price p^l_{ht} becomes deterministic from the seller' s perspective. As such, the only scenario in which δ affects the seller' s utility is when the seller receives only the upfront payment — that is, when $p^l_{ht} \leq \delta p^i_{ht}$, and the revenue-sharing component is zero.

In this case, increasing δ raises the upfront payment, increasing the total payoff and making the seller more likely to accept the iBuyer offer. Hence, the derivative with respect to δ is strictly positive.

In contrast, when $\delta p^i_{ht} < p^l_{ht}$, the seller anticipates receiving the full payoff (either p^i_{ht} or p^l_{ht}), and the total payment is invariant to δ. Thus, $\frac{\partial \mathbb{P}(\text{Sell to iBuyer}|\xi_h)}{\partial \delta} = 0$.

■

To illustrate the quantitative implications of this alternative interpretation, Figure 17 shows the expected iBuyer profit per transaction across values of δ, while Figure 18 reports the additional improvement from the pricing design. The corresponding evaluation results, parallel to those in the main text, are provided below.

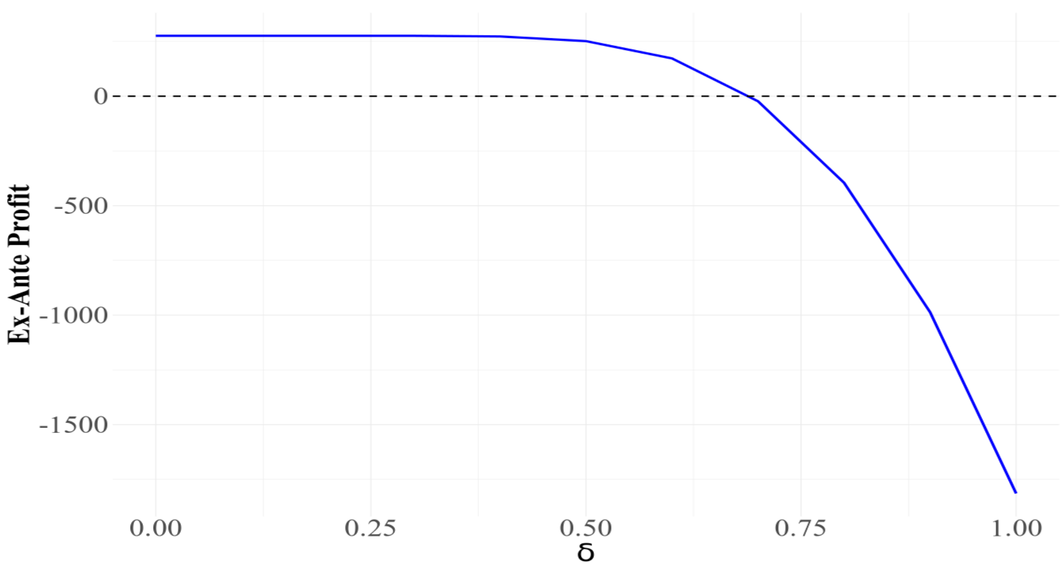

Figure 17: Expected Profit (Unit: 100,000 dollars)
Note: This figure reports the expected iBuyer profit per transaction (in units of $100,000) under a counterfactual contract with varying δ, the fraction of the original iBuyer offer price p^i_{ht} paid upfront. A value of $\delta = 1$ corresponds to the current contract structure. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte. The evaluation interprets hassle costs in the most favorable way for the iBuyer; the main text presents the most pessimistic case.

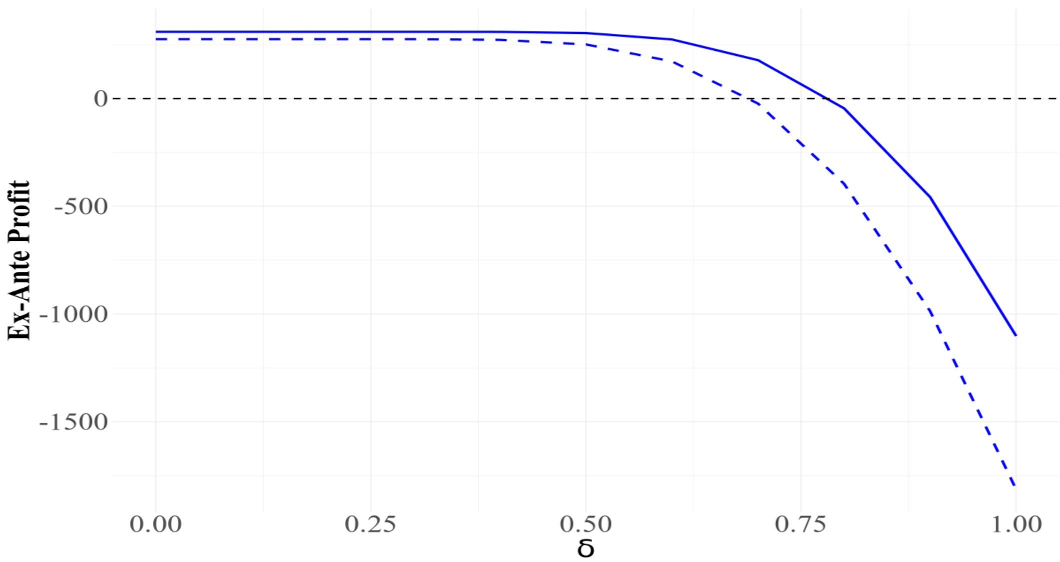

Figure 18: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of $100,000) under a counterfactual contract with varying δ, the fraction of the original iBuyer offer price p^i_{ht} paid upfront ($\delta = 1$ corresponds to the current contract structure). The solid line represents profits simulated using a counterfactual LLM-based pricing algorithm; the dotted line reproduces profits from the original iBuyer pricing model, as reported earlier. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte. This evaluation interprets hassle costs in the most favorable way for the iBuyer; the most pessimistic case is shown in the main text.

H.3 Extension — Contract Design with Partial Revenue Sharing

This appendix extends the numerical simulations by relaxing the assumption of full conditional revenue sharing for the median-type house—defined by the price range $\delta p^i_{ht} < p^l_{ht} < p^i_{ht}$—and instead allow only an α fraction of the resale gain to be shared with the seller.

Accordingly, the revised payment structure becomes:

$$\delta p^i_{ht} + \begin{cases} p^i_{ht} - \delta p^i_{ht} & \text{if } p^l_{ht} > p^i_{ht}, \\ \alpha(p^l_{ht} - \delta p^i_{ht}) & \text{if } p^i_{ht} > p^l_{ht} > \delta p^i_{ht}, \\ 0 & \text{if } p^l_{ht} \leq \delta p^i_{ht}. \end{cases}$$

Following the logic in Subsection 8.1.1, the realized seller payoff under this extension should continue to lie between the two extreme benchmark cases, depending on the value of α and the interpretation of hassle costs.

Case 1: Hassle Costs Fully Loaded on Time Constraints

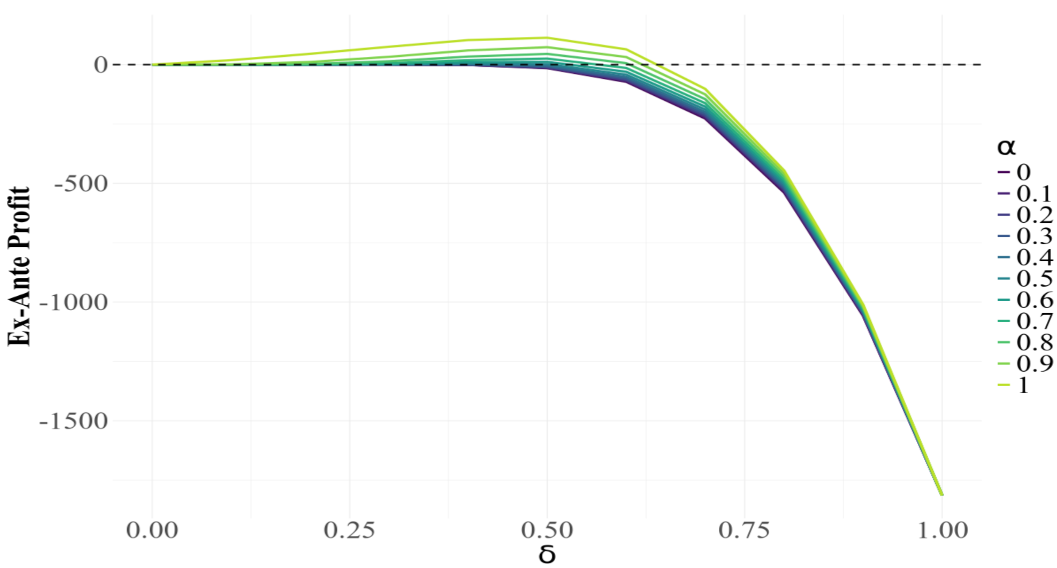

Figure 19: Expected Profit under Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure extends the counterfactual contract analysis by introducing α (the share of the iBuyer' s resale revenue passed back to the seller). Each line corresponds to a different value of $\alpha \in 0, 0.1, \ldots, 1$, with $\alpha = 1$ reproducing the original counterfactual design. δ (on the x-axis) remains the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront. Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte. This figure represents one of two polar cases of interpreting hassle costs; intermediate interpretations would yield results between the two extremes.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

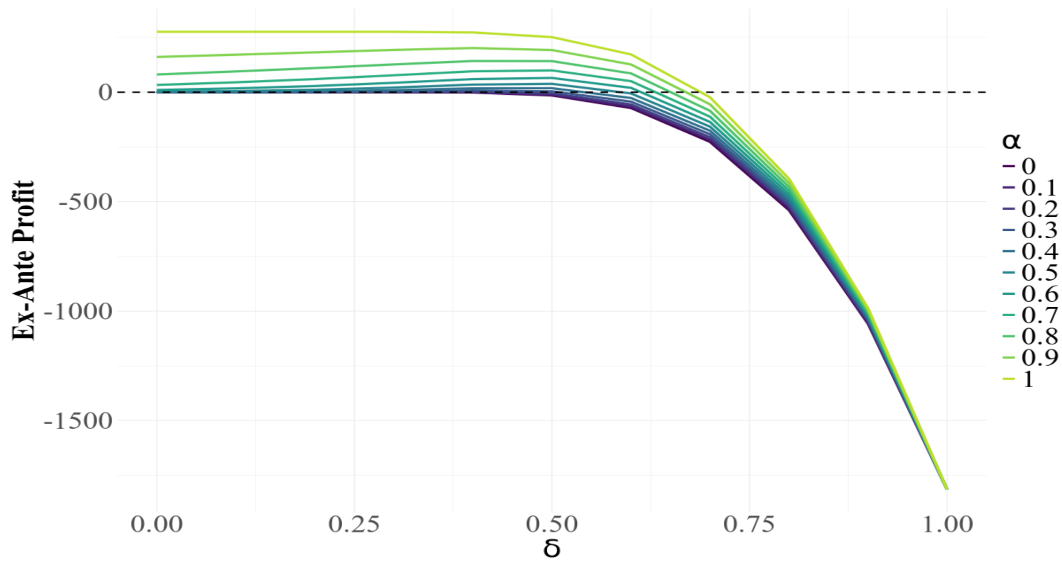

Figure 20: Expected Profit under Non-Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure presents the same extended counterfactual contract analysis as Figure 19, but under the assumption that hassle costs are unrelated to timing (e.g., psychological or logistical burden). α denotes the share of resale revenue returned to the seller, with each curve corresponding to a different $\alpha \in 0, 0.1, \ldots, 1$. δ (x-axis) is the fraction of the iBuyer' s offer, p^i_{ht}, paid upfront. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte. This is the second of two polar interpretations of hassle costs; actual outcomes are likely to lie between these bounding cases.

The key insight from Figures 19 and 20 is that, when $\alpha < 1$, the reduction in seller participation (i.e., the demand-side effect) dominates the benefit from reduced payments to sellers. As a result, the iBuyer' s expected profit declines with α.

Therefore, the core takeaway remains consistent with the $\alpha = 1$ case discussed in the previous subsection: cream-skimming mitigates adverse selection and can potentially render the iBuyer model profitable.

H.4 Cumulative Explained Variance for "Public Remarks" Embeddings

Figures 21 and 22 report the cumulative explained variance from applying PCA to text embeddings derived from the full set of "Public Remarks" in the CoreLogic MLS data across cities.

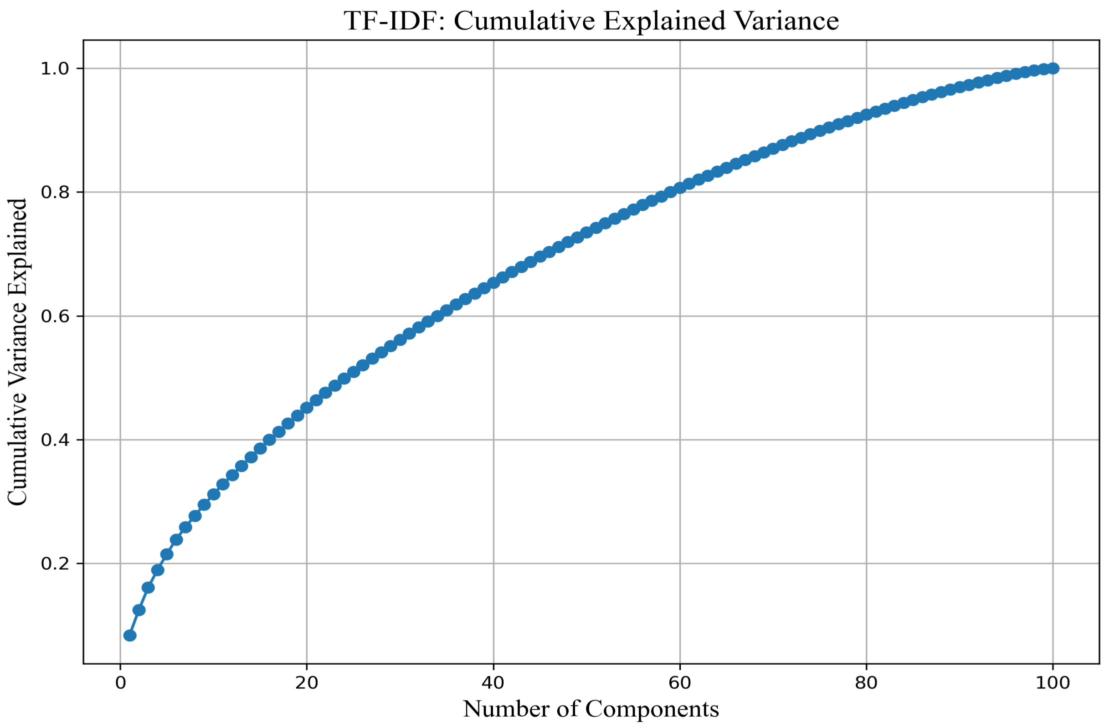

Figure 21: Cumulative explained variance from PCA on TF-IDF embeddings of "Public Remarks"
Note: PCA applied to TF-IDF vectors (restricted to the top 100 terms) computed from all listing descriptions in the CoreLogic MLS dataset across cities.

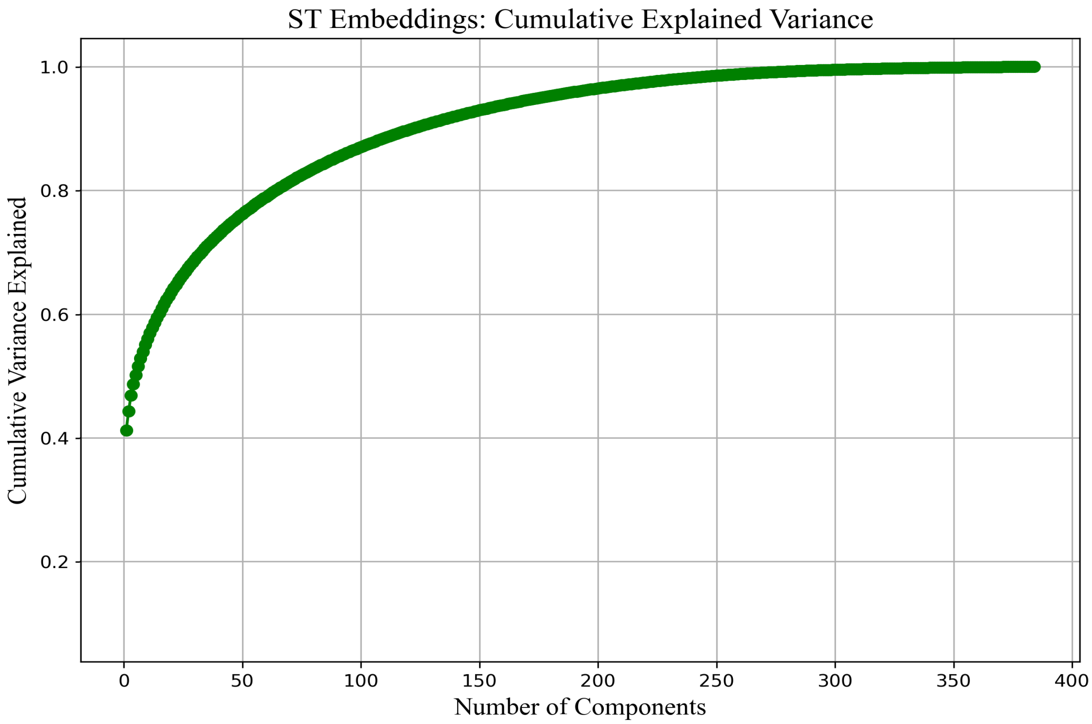

Figure 22: Cumulative explained variance from PCA on Sentence Transformer embeddings of "Public Remarks"

Note: PCA applied to 384-dimensional embeddings from a Sentence Transformer model computed from all listing descriptions in the CoreLogic MLS dataset across cities.

H.5 LoRA Fine-Tuning

Table 13 reports regression results for the pre-iBuyer entry period in Charlotte. I use the historical average of the *Public Remarks* score for each house. Because some properties lack *Public Remarks* or matching closing date information, approximately 3500 out of 118,604 transactions are excluded from the analysis. For fairness, I apply the same filtering to all three models in the table, ensuring that they are estimated on identical samples.

The first column in the table corresponds to the baseline model, which does not include an LLM-based text score. The second and third columns include the LLM-based text score: the second column is obtained via the prompt-engineering approach, while the third column uses the LoRA fine-tuning approach.

	Without Score	Prompt engineering approach	LoRA fine-tuning approach
House Characteristics	Yes	Yes	Yes
Market Characteristics	Yes	Yes	Yes
LLM-based text score	–	0.434 (0.007)***	0.519 (0.004)***
R^2	0.910	0.913	0.920
Adj. R^2	0.910	0.913	0.920
Num. obs.	115085	115085	115085

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 13: Effect of LLM-based text score on log transaction price estimates
Note: This table presents regression results for transaction prices in the pre-iBuyer entry period in Charlotte. The dependent variable is the log of the transaction price, expressed in $100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. The first model excludes the LLM-derived score, while the second and third include scores from the prompt engineering and LoRA fine-tuning approaches, respectively. Standard errors are reported in parentheses.

The counterfactual pricing results using the LoRA fine-tuning approach are shown in Figures 23 and 24. In both figures, the dotted lines represent simulated profits from the original iBuyer pricing model. The dark blue solid lines correspond to the prompt engineering approach, while the light blue solid lines correspond to the LoRA fine-tuning approach.

Case 1: Hassle Costs Fully Loaded on Time Constraints

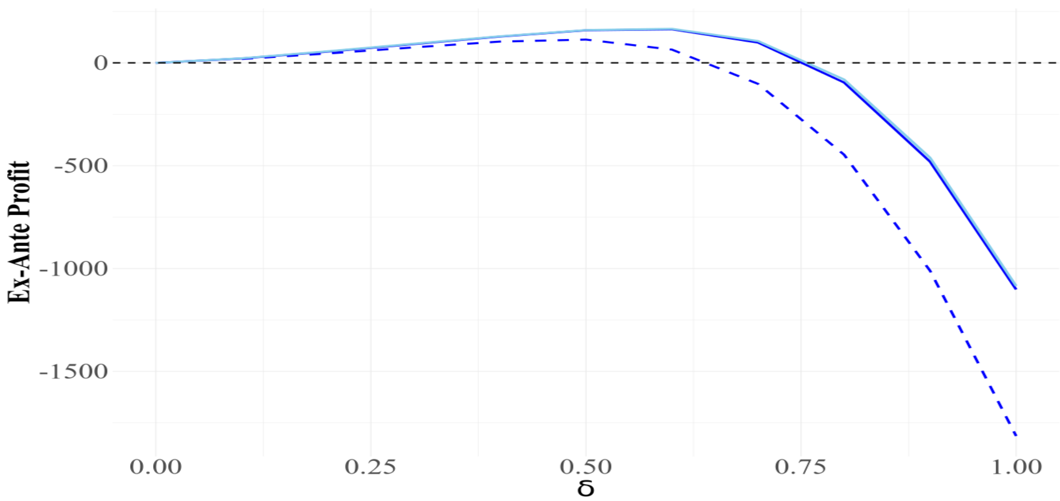

Figure 23: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt engineering approach, the light blue solid line represents profits from the LoRA fine-tuning approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using post-entry housing transactions in Charlotte.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

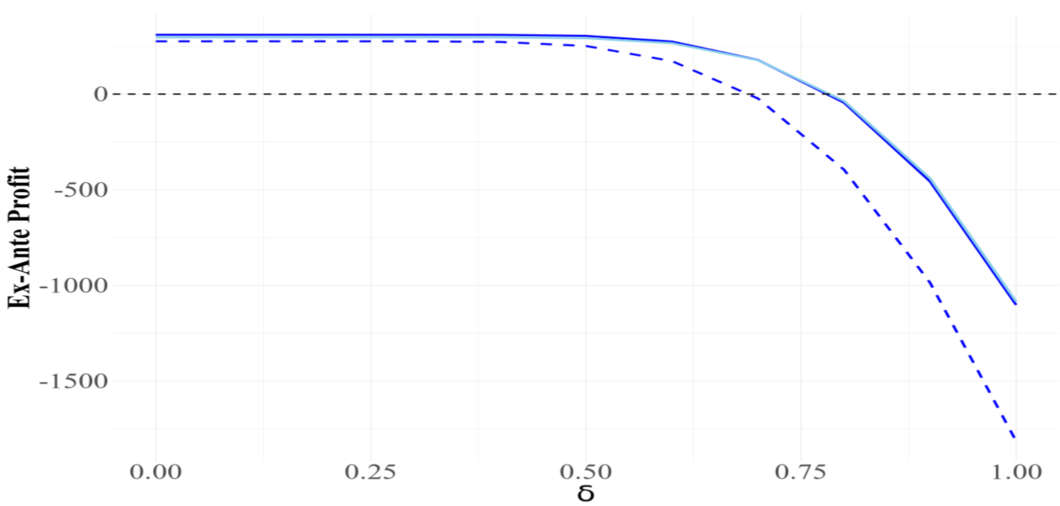

Figure 24: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt-engineering approach, the light blue solid line represents profits from the LoRA fine-tuning approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to be unrelated to timing. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte.

Using Residualized Prices. I extend the analysis by employing residualized prices obtained from a linear regression on observable house and market characteristics. This approach isolates the portion of price variation not explained by structured attributes, thereby capturing variation more likely associated with information in the *Public Remarks*.

A practical limitation arises because the Llama 3.2 1B Instruction-tuned model has difficulty processing numerical strings containing commas or decimal points. As a result, the model fails to generate predictions for 76,123 of the 118,604 properties, leaving 35 percent of the sample with valid predicted scores.[17] To ensure comparability with the baseline iBuyer pricing model, I use its predicted prices for the 65 percent of observations without LLM-based predictions and apply the LoRA-fine-tuned prices to the remaining 35 percent. The results from this exercise are presented below.

Figures 25 and 26 present the counterfactual pricing outcomes based on residualized prices. As in the main specification, the dotted lines depict simulated profits from the baseline iBuyer pricing model, while the dark and light blue solid lines correspond to the prompt-engineering and LoRA fine-tuning approaches, respectively. The apparent improvement is modest, which is expected given that only 35 percent of the properties received updated prices.

Case 1: Hassle Costs Fully Loaded on Time Constraints

[17] This limitation could be mitigated by scaling the residual prices to remove digits or by using a model less sensitive to numerical formatting. Updates are forthcoming.

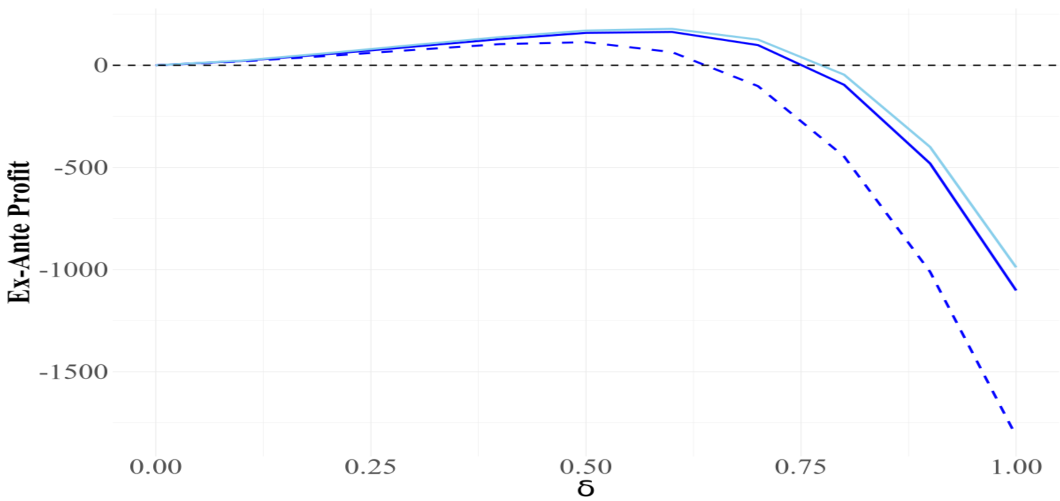

Figure 25: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt engineering approach, the light blue solid line represents profits from the LoRA fine-tuning approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using post-entry housing transactions in Charlotte.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

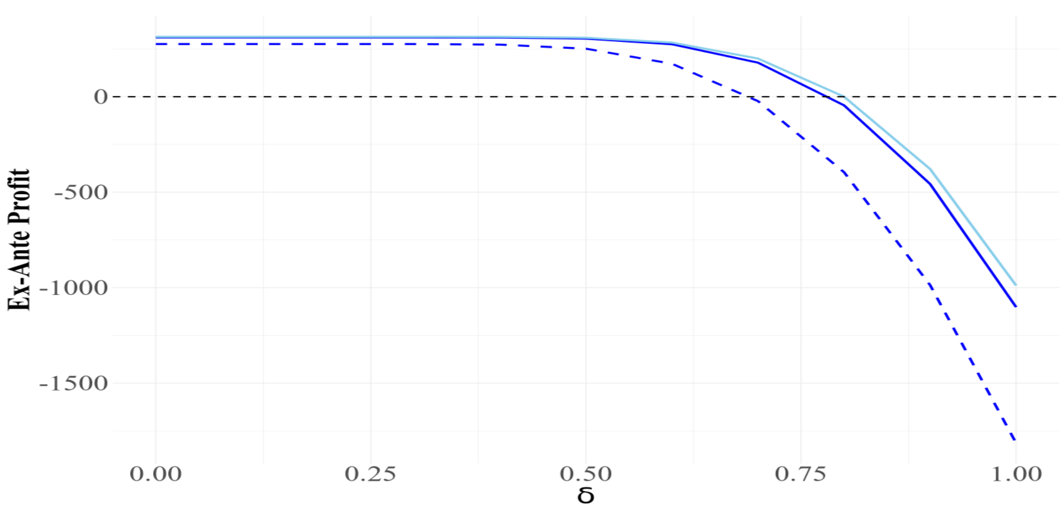

Figure 26: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt engineering approach, the light blue solid line represents profits from the LoRA fine-tuning approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to be unrelated to timing. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte.

H.6 Feature Extraction

Table 14 reports regression results for the pre-iBuyer entry period in Charlotte. I use the historical average of the *Public Remarks* score for each house. Because some properties lack *Public Remarks* or matching closing date information, approximately 3500 out of 118,604 transactions are excluded from the analysis. For fairness, I apply the same filtering to all three models in the table, ensuring that they are estimated on identical samples.

The first column in the table corresponds to the baseline model, which does not include an LLM-based text score. The second and third columns include the LLM-based text score: the second column is obtained via the prompt-engineering approach, while the third column uses the feature extraction approach.

	Without Score	Prompt engineering approach	Feature extraction approach
House Characteristics	Yes	Yes	Yes
Market Characteristics	Yes	Yes	Yes
LLM-based text score	–	0.434 (0.007)***	0.290 (0.005)***
R^2	0.910	0.913	0.913
Adj. R^2	0.910	0.913	0.913
Num. obs.	115085	115085	115085

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 14: Effect of LLM-based text score on log transaction price estimates
Note: This table presents regression results for transaction prices in the pre-iBuyer entry period in Charlotte. The dependent variable is the log of the transaction price, expressed in \$100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. The first model excludes the LLM-derived score, while the second and third include scores from the prompt engineering and feature extraction approaches, respectively. Standard errors are reported in parentheses.

The counterfactual pricing results using the feature extraction approach are shown in Figures 27 and 28. In both figures, the dotted lines represent simulated profits from the original iBuyer pricing model. The dark blue solid lines correspond to the prompt engineering approach, while the light blue solid lines correspond to the feature extraction approach.

Case 1: Hassle Costs Fully Loaded on Time Constraints

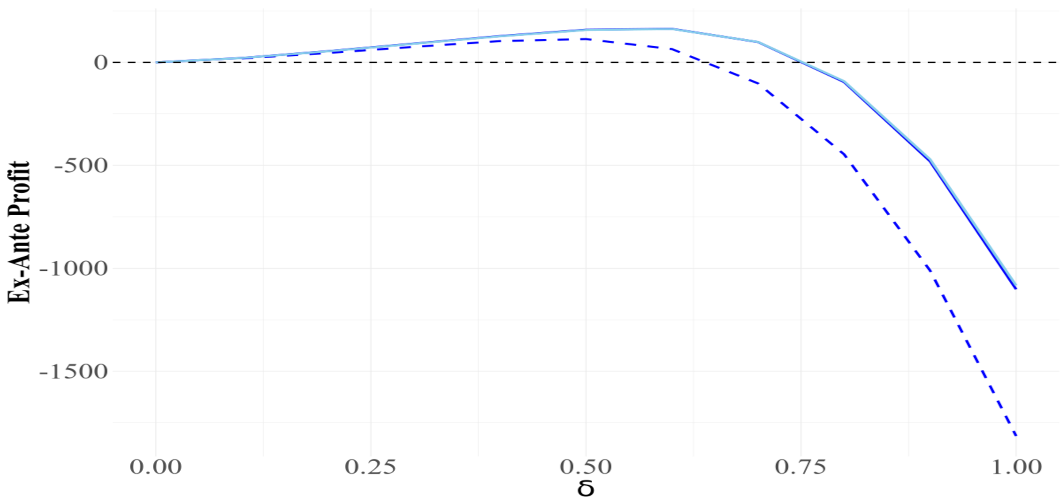

Figure 27: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt engineering approach, the light blue solid line represents profits from the feature extraction approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using post-entry housing transactions in Charlotte.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

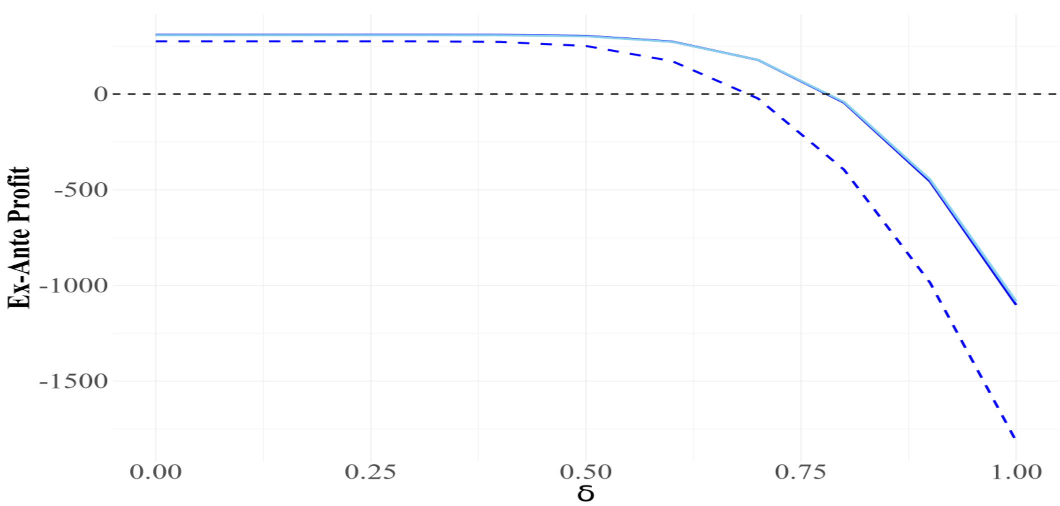

Figure 28: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt engineering approach, the light blue solid line represents profits from the feature extraction approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to be unrelated to timing. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte.

Using Residualized Prices. Additionally, I extend the analysis by using residualized prices obtained from a linear regression on observables. This residualization isolates the portion of price variation less likely to be explained by structured attributes and more likely attributable to information in the *Public Remarks*.

The counterfactual pricing results with residualized prices are shown in Figures 29 and 30. As before, the dotted lines represent simulated profits from the original iBuyer pricing model, the dark blue solid lines correspond to the prompt-engineering approach, and the light blue solid lines correspond to the feature extraction approach.

Case 1: Hassle Costs Fully Loaded on Time Constraints

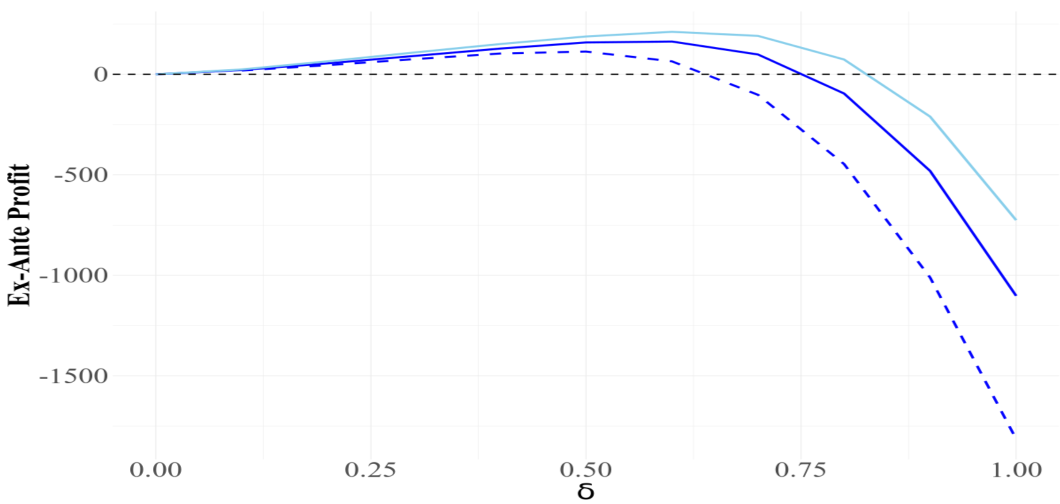

Figure 29: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt engineering approach, the light blue solid line represents profits from the feature extraction approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using post-entry housing transactions in Charlotte.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

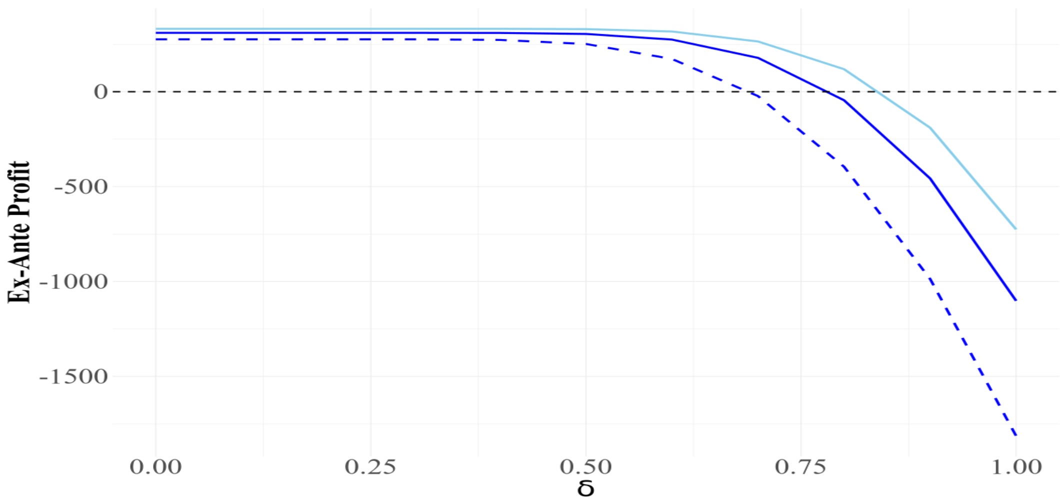

Figure 30: Expected Profit with LLM-Based Pricing (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). The dark blue solid line represents simulated profits from the prompt-engineering approach, the light blue solid line represents profits from the feature extraction approach, and the dotted line reproduces profits from the original iBuyer pricing model. Hassle costs are assumed to be unrelated to timing. Profits are simulated using numerical expectations based on post-entry housing transactions in Charlotte.

H.7 Comparing LLM and Lagged Price Predictive Performance

Another potential proxy for unobserved house quality is the lagged sale price of the same property, as it embeds a house-specific random effect. Although the lagged price also contains idiosyncratic noise unrelated to quality, employing an LLM to interpret and standardize the *Public Remarks* offers an alternative way to capture the semantic information conveyed in listing descriptions. Comparing the predictive performance of models that include either the LLM-based text score or the lagged price as an additional covariate reveals how much of the variation in prices the semantic information helps explain. This comparison illustrates the trade-off between how effectively textual features filter out irrelevant noise in price variation and how much they may omit relevant quality information.

	Prompt engineering	Prompt engineering	Fine-tuning (residualized)	Fine-tuning (residualized)	Feature extraction (residualized)	Feature extraction (residualized)
Dependent variable	Log price	Log price	Log price	Log price	Log price	Log price
House Char.	Yes	Yes	Yes	Yes	Yes	Yes
Market Char.	Yes	Yes	Yes	Yes	Yes	Yes
Log lagged price	0.31 (0.00)***		0.32 (0.00)***		0.31 (0.00)***	
LLM-based text score		0.46 (0.01)***		0.91 (0.01)***		0.78 (0.00)***
R^2	0.933	0.916	0.933	0.947	0.933	0.938
Adj. R^2	0.933	0.916	0.933	0.947	0.933	0.938
Num. obs.	109281	109281	39264	39264	109106	109106

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 15: LLM and Lagged Price Predictive Performance
Note: This table presents regression results for transaction prices in the pre-iBuyer entry period in Charlotte. The dependent variable is the log transaction price obtained after regressing prices on observable house and market characteristics. The log price is expressed in $100,000s, and living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. Standard errors are reported in parentheses. For each LLM approach, the sample is restricted to observations for which both lagged price and valid LLM-based text scores are available to ensure consistent comparison for each specification.

Table 15 reports regression results for the pre-iBuyer entry period in Charlotte and compares the performance of each LLM approach against a specification that includes the lagged price as an additional covariate. To ensure a valid comparison, the sample is restricted to observations with both lagged prices and valid LLM-based text scores. The results indicate that while the prompt-engineering approach performs slightly worse than the lagged price —likely because it abstracts away some information related to property quality—the fine-tuning and feature-extraction approaches perform better when the residualized log price is used as the target variable. Results based on the non-residualized log price are omitted for brevity, as their explanatory power is generally lower than in the residualized cases.

I Appendix: Lasso and Lagged Variable Checks for Model Fit

Tables 16 and 17 indicate that including additional independent variables and a lagged transaction price does not significantly improve model fit. To assess this, I expanded the set of housing characteristics and applied LASSO regression for variable selection. Additionally, I introduced the most recent historical market transaction price (excluding the sale to the iBuyer) as a lagged variable for both individual and iBuyer pricing.

The iBuyer pricing model is relatively straightforward, as I can simply include the lagged market transaction price and apply LASSO for variable selection.

In contrast, the market pricing model presents two econometric challenges. First, including a lagged dependent variable directly in a random effects model can introduce bias. To address this, I implement a two-step approach, where I first estimate the lagged variable using a separate regression and then use the fitted values in the random effects model.

Second, while the ideal method for variable selection would be to apply LASSO within the random effects framework, this approach is computationally intensive and fails to converge in a reasonable time. As a practical alternative, I perform LASSO without accounting for random effects to select variables, and then estimate the final model using a random effects regression with the selected variables.

The variable selection process proceeds as follows. First, I run a LASSO regression including the lagged market transaction price to identify a set of predictive variables—referred to as Variable Set 1. Second, I run another LASSO regression where the dependent variable is the lagged log market transaction price and the independent variables are the housing characteristics. This yields Variable Set 2. I then take the union of Variable Sets 1 and 2 (excluding the lagged log market transaction price itself) as the final set of covariates for the random effects model.

In the estimation stage, I first regress the lagged log market transaction price on the selected variables to obtain fitted values. These fitted values are then used in the final random effects regression, where I regress the log market transaction price on the fitted lagged values and the selected covariates.

	Log price (Charlotte)
House characteristics	Yes
Market characteristics	Yes
Lagged market price	Yes
Error variance	0.028
Random effect variance	0.065
Num. obs.	109551

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 16: Log market price estimation
Note: This table presents results from a robustness check using a two-step estimation procedure with a LASSO-selected variable set and a fitted lagged market price. The lagged market transaction price is instrumented using housing and market characteristics selected via LASSO. Prices are expressed in units of $100,000 and adjusted to the March 2023 CPI level. Living area is measured in units of 1,000 square feet. Estimates are obtained from a random effects model.

	Log price (Charlotte)
House characteristics	Yes
Market characteristics	Yes
Lagged market price	Yes
R^2	0.973
Adj. R^2	0.973
Num. obs.	2323

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 17: Log iBuyer price estimation
Note: This table presents results from a robustness check using a LASSO-selected set of housing and market characteristics, including the most recent lagged market transaction price. The model is estimated using linear regression. Prices are expressed in units of $100,000 and adjusted to the March 2023 CPI level. Living area is measured in units of 1,000 square feet.

J Appendix: Results for Other Cities

The following are the estimation results and numerical simulation outcomes for counterfactual contracts in five cities other than Charlotte. Tables 18 and 19 present the hedonic regression results. The variance of the random effect ranges from 0.05 to 0.17. Among these, Orlando, Jacksonville, and Phoenix exhibit variances comparable to Charlotte's, while Atlanta and Tampa show relatively higher variances. The model fit for the iBuyer pricing regression is high across all cities—including Charlotte, as reported in the main text—suggesting that a simple linear model can effectively approximate iBuyer pricing behavior.

	Atlanta	Orlando	Jacksonville	Phoenix	Tampa
Intercept	-2.252 (0.112)***	0.726 (0.126)***	-0.762 (0.107)***	-1.416 (0.206)***	-1.549 (0.144)***
Log living area square feet	0.800 (0.008)***	0.980 (0.008)***	1.164 (0.007)***	1.043 (0.004)***	1.018 (0.011)***
Bedroom number	-0.149 (0.003)***	-0.046 (0.003)***	-0.044 (0.003)***	-0.079 (0.002)***	-0.057 (0.004)***
Bathroom number	0.106 (0.004)***	0.027 (0.004)***	-0.043 (0.003)***	0.014 (0.003)***	0.022 (0.005)***
Building age	0.002 (0.000)***	-0.001 (0.000)***	-0.004 (0.000)***	-0.005 (0.000)***	-0.002 (0.000)***
Garage dummy	0.181 (0.035)***	0.171 (0.009)***	0.676 (0.054)***	0.179 (0.187)	0.197 (0.016)***
Heating dummy	0.338 (0.035)***	0.213 (0.025)***	0.342 (0.015)***	0.107 (0.077)	0.281 (0.031)***
seasonalFE_2nd quarter	0.018 (0.003)***	0.035 (0.004)***	0.041 (0.003)***	0.012 (0.001)***	0.042 (0.004)***
seasonalFE_3rd quarter	0.017 (0.003)***	0.044 (0.004)***	0.037 (0.003)***	0.011 (0.001)***	0.032 (0.004)***
seasonalFE_4th quarter	-0.001 (0.003)	0.018 (0.004)***	0.014 (0.003)***	0.006 (0.001)***	0.015 (0.004)***
30-year mortgage rates	0.022 (0.003)***	-0.020 (0.003)***	0.004 (0.002)	-0.019 (0.001)***	-0.007 (0.003)*
Federal fund rate	0.022 (0.001)***	0.023 (0.001)***	0.014 (0.001)***	0.037 (0.000)***	0.014 (0.001)***
Log CPI index	-0.793 (0.025)***	-1.696 (0.029)***	-1.446 (0.023)***	-0.979 (0.010)***	-1.173 (0.035)***
Log CS index	1.193 (0.010)***	1.387 (0.010)***	1.300 (0.009)***	1.262 (0.004)***	1.356 (0.012)***
Error variance	0.060	0.067	0.096	0.019	0.096
Random effect variance	0.170	0.060	0.067	0.050	0.110
Num. obs.	119786	66896	130604	190590	65639

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 18: Log market price estimation
Note: This table presents regression results for individual market transactions across five cities. The dependent variable is the log of the transaction price—individual-to-individual—expressed in \$100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. Standard errors are reported in parentheses. The models include random effects to account for unobserved housing quality.

	Atlanta	Orlando	Jacksonville	Phoenix	Tampa
Intercept	19.934 (5.423)***	9.190 (4.009)*	8.156 (3.682)*	−0.855 (1.685)	5.777 (6.405)
Log living area square feet	0.635 (0.040)***	0.477 (0.027)***	0.747 (0.025)***	0.712 (0.018)***	0.458 (0.047)***
Bedroom number	−0.062 (0.014)***	0.105 (0.012)***	0.025 (0.009)**	−0.004 (0.006)	0.181 (0.020)***
Bathroom number	0.052 (0.016)**	0.009 (0.013)	−0.086 (0.011)***	−0.065 (0.011)***	−0.063 (0.021)**
Building age	0.004 (0.001)***	−0.002 (0.000)***	−0.002 (0.000)***	−0.000 (0.000)	−0.002 (0.001)*
Garage dummy	0.272 (0.070)***	0.059 (0.018)**			0.131 (0.033)***
seasonalFE_2nd quarter	0.143 (0.028)***	0.057 (0.020)**	0.047 (0.013)***	0.040 (0.010)***	0.069 (0.037)
seasonalFE_3rd quarter	0.159 (0.027)***	0.033 (0.019)	0.054 (0.015)***	0.032 (0.010)***	0.080 (0.034)*
seasonalFE_4th quarter	0.043 (0.028)	0.039 (0.020)*	0.014 (0.014)	−0.011 (0.010)	0.016 (0.033)
30-year mortgage rate	0.025 (0.025)	0.018 (0.016)	−0.015 (0.019)	−0.017 (0.007)*	−0.020 (0.025)
Federal funds rate	0.019 (0.032)	0.037 (0.024)	0.062 (0.020)**	−0.022 (0.006)***	0.049 (0.034)
Log CPI index	−9.466 (1.996)***	−5.115 (1.530)***	−4.892 (1.320)***	−1.799 (0.620)**	−4.484 (2.429)
Log CS index	4.545 (0.748)***	2.804 (0.590)***	2.856 (0.476)***	1.917 (0.225)***	2.884 (0.925)**
R^2	0.941	0.978	0.972	0.976	0.962
Adj. R^2	0.941	0.978	0.972	0.975	0.962
Num. obs.	1102	951	1711	3744	575

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 19: Log iBuyer price estimation
Note: This table presents regression results for iBuyer offer prices across five cities. The dependent variable is the log of the offer price, expressed in $100,000s. Living area is measured in 1,000 square feet. Prices are adjusted to the March 2023 CPI level. Standard errors are reported in parentheses. The models are estimated using linear regression.

Table 20 presents the estimated parameters of the hassle cost distribution, along with bootstrap standard deviations. These results are preliminary, based on only 100 bootstrap draws, and will be updated to match the 1,000-draw bootstrap estimates used in the main text. While the limited number of draws makes the current estimates statistically imprecise, a potentially important pattern emerges: Atlanta and Jacksonville exhibit a positive correlation between hassle cost and unobserved quality. This suggests that sellers facing higher hassle costs may also own higher-quality homes. Such a relationship is favorable to the iBuyer model, as it implies that targeting high–hassle cost sellers naturally leads to acquiring better properties. This mechanism helps explain why only Atlanta and Jacksonville yield positive expected profits under the current contract design in the numerical simulations.

Parameter	Atlanta	Orlando	Jacksonville	Phoenix	Tampa
$\sigma^2_{\log c}$	0.779 (2.440)	0.151 (0.560)	1.814 (0.934)	8.036 (7.691)	1.801 (1.790)
$\mu_{\log c}$	-1.987 (0.946)	-0.038 (0.716)	-2.005 (0.507)	-0.734 (0.709)	-2.648 (1.031)
a	0.283 (0.342)	0.946 (0.244)	0.279 (0.126)	0.888 (0.035)	0.073 (0.288)
ρ	0.832 (0.540)	-0.208 (0.449)	0.056 (0.422)	-0.602 (0.537)	-0.289 (0.620)

Table 20: Hassle cost estimation results by city
Note: This table presents preliminary parameter estimates of the hassle cost distribution for five cities. $\sigma^2_{\log c}$ denotes the variance of log hassle costs, $\mu_{\log c}$ is the mean, a is the mass point at $-\infty$, and ρ is the correlation between log hassle cost and unobserved house quality. Estimates are based on 100 bootstrap replications; standard deviations are reported in parentheses. These results will be updated in a future revision using 1,000 bootstrap draws, consistent with the main text.

The counterfactual simulations yield two notable findings, consistent with the main text results. First, even under the favorable assumption that an iBuyer faces no dynamic concerns —such as resale timing or inventory costs—all cities except Atlanta and Jacksonville exhibit negative expected profits at $\delta = 1$, which corresponds to the current contract design. Second, when δ is reduced and revenue sharing is allowed, there exists a conditional revenue-sharing contract that yields positive expected profits.

On the other hand, Atlanta and Jacksonville do not experience negative profits, which may reflect a less severe adverse selection problem—likely due to the positive correlation between hassle cost and unobserved quality, as discussed above. When adverse selection is less pronounced, reducing the upfront payment may be more harmful, as it weakens individual sellers' incentives to sell to an iBuyer. In such cases, revenue sharing may contribute less to profitability, since its role in mitigating adverse selection becomes less critical.

J.1 Atlanta

Case 1: Hassle Costs Fully Loaded on Time Constraints

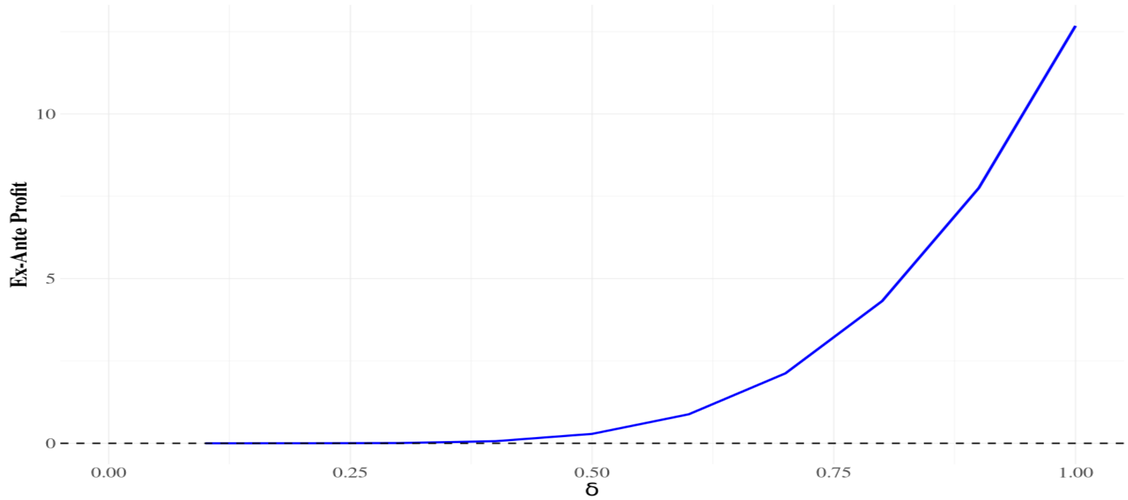

Figure 31: Expected Profit under Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using numerical expectations based on post-entry housing transactions in Atlanta.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

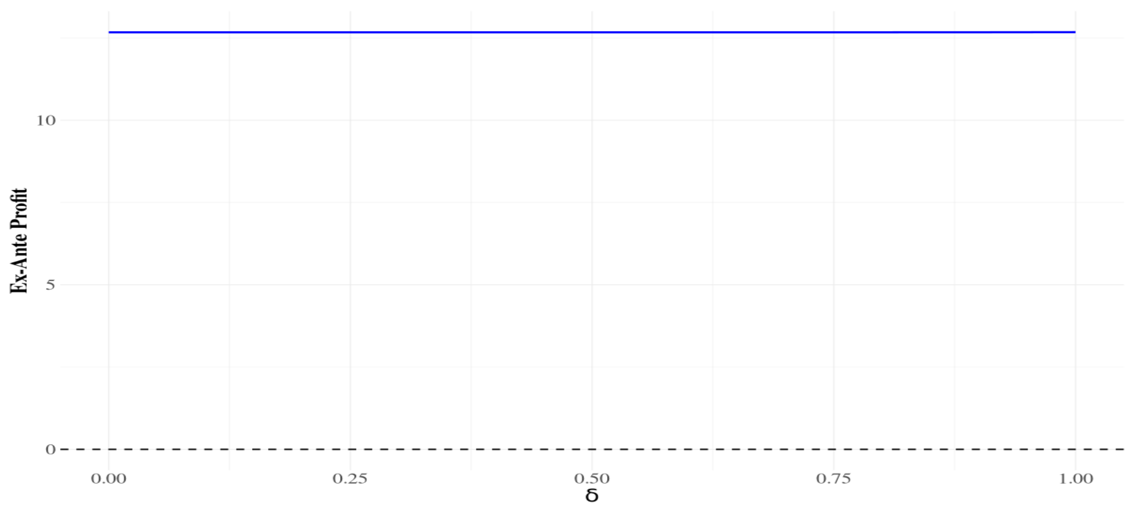

Figure 32: Expected Profit under Non-Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to be unrelated to timing. Simulations are based on numerical expectations from post-entry housing transactions in Atlanta.

J.2 Orlando

Case 1: Hassle Costs Fully Loaded on Time Constraints

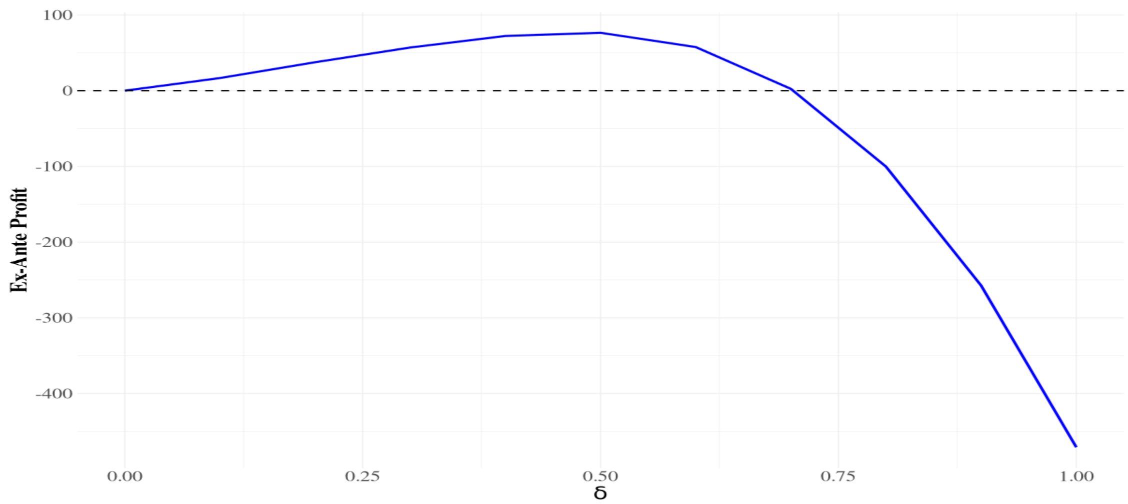

Figure 33: Expected Profit under Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using numerical expectations based on post-entry housing transactions in Orlando.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

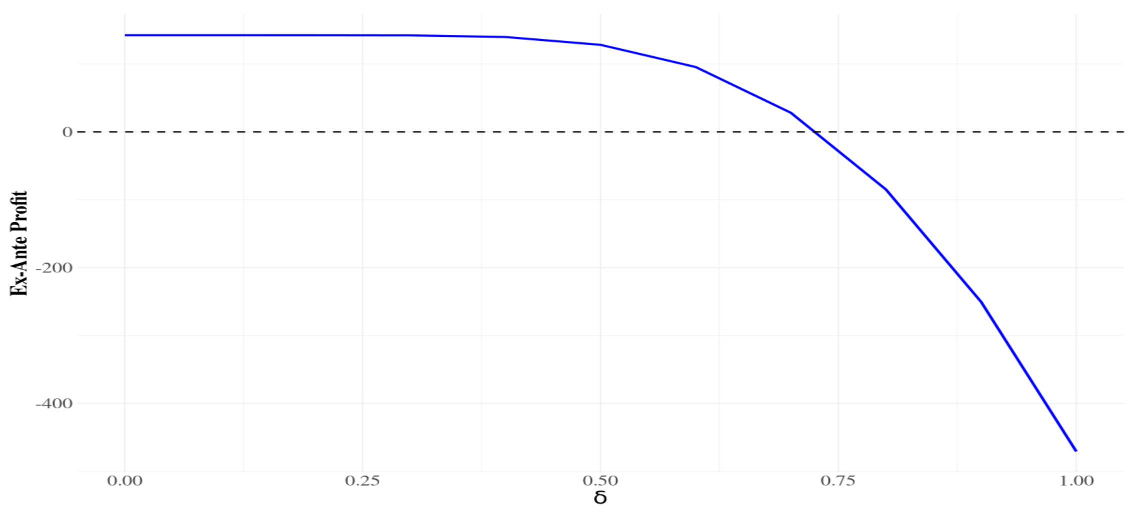

Figure 34: Expected Profit under Non-Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to be unrelated to timing. Simulations are based on numerical expectations from post-entry housing transactions in Orlando.

J.3 Jacksonville

Case 1: Hassle Costs Fully Loaded on Time Constraints

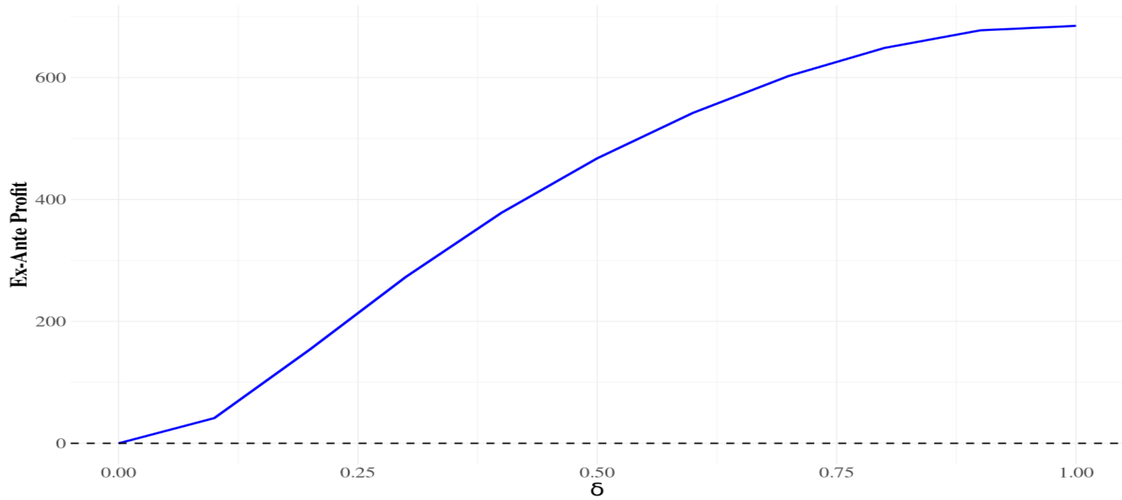

Figure 35: Expected Profit under Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using numerical expectations based on post-entry housing transactions in Jacksonville.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

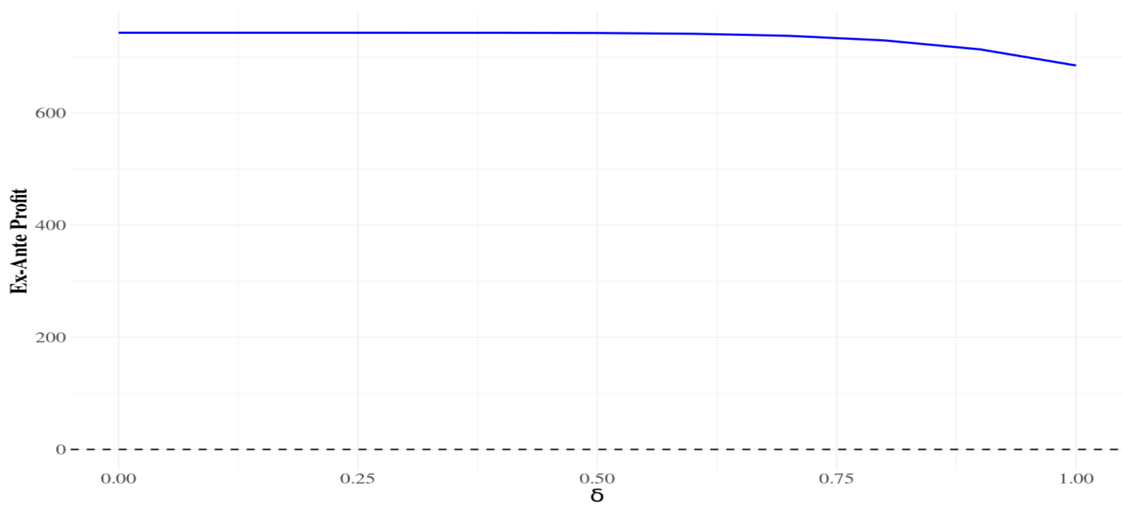

Figure 36: Expected Profit under Non-Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to be unrelated to timing. Simulations are based on numerical expectations from post-entry housing transactions in Jacksonville.

J.4 Phoenix

Case 1: Hassle Costs Fully Loaded on Time Constraints

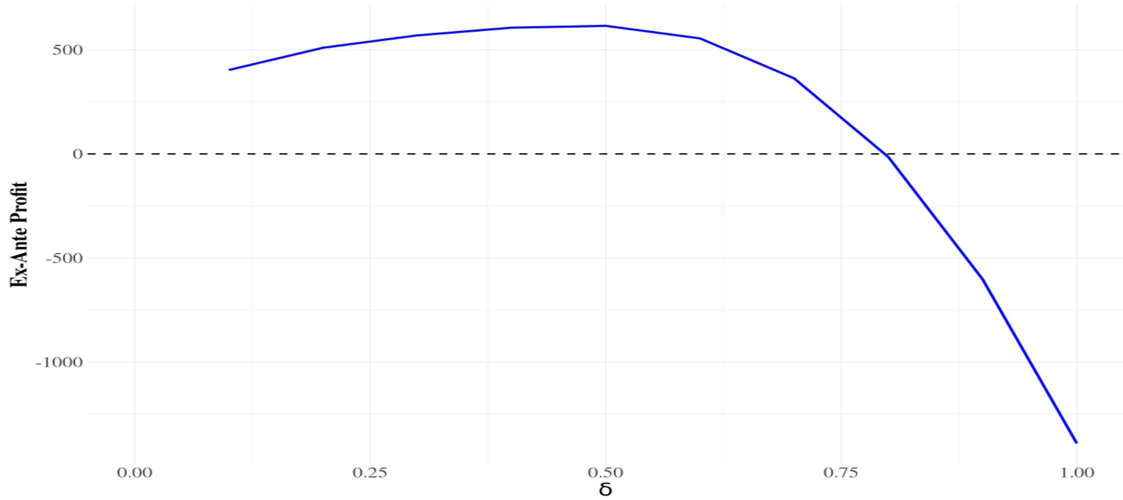

Figure 37: Expected Profit under Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using numerical expectations based on post-entry housing transactions in Phoenix.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

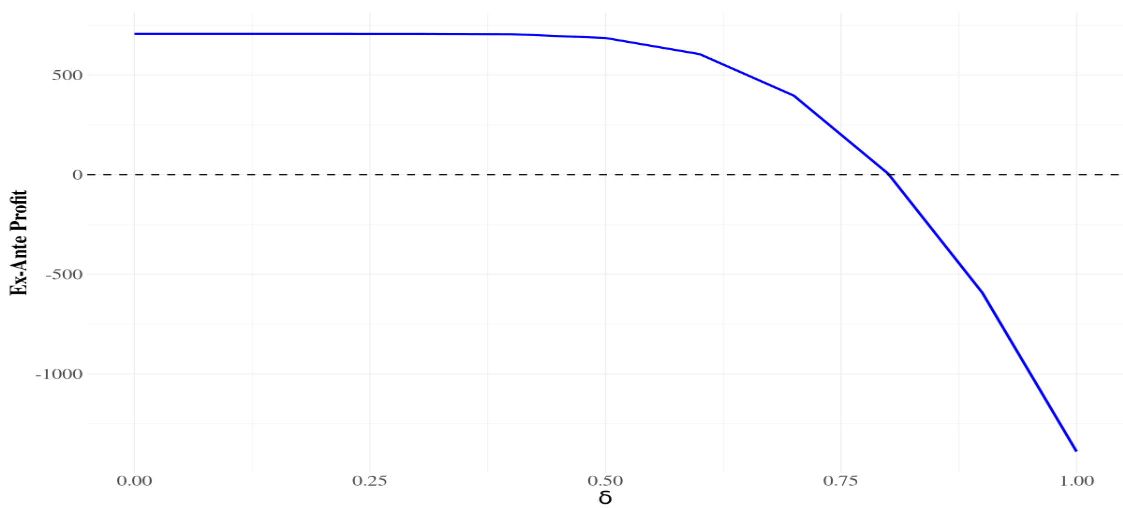

Figure 38: Expected Profit under Non-Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of \$100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to be unrelated to timing. Simulations are based on numerical expectations from post-entry housing transactions in Phoenix.

J.5 Tampa

Case 1: Hassle Costs Fully Loaded on Time Constraints

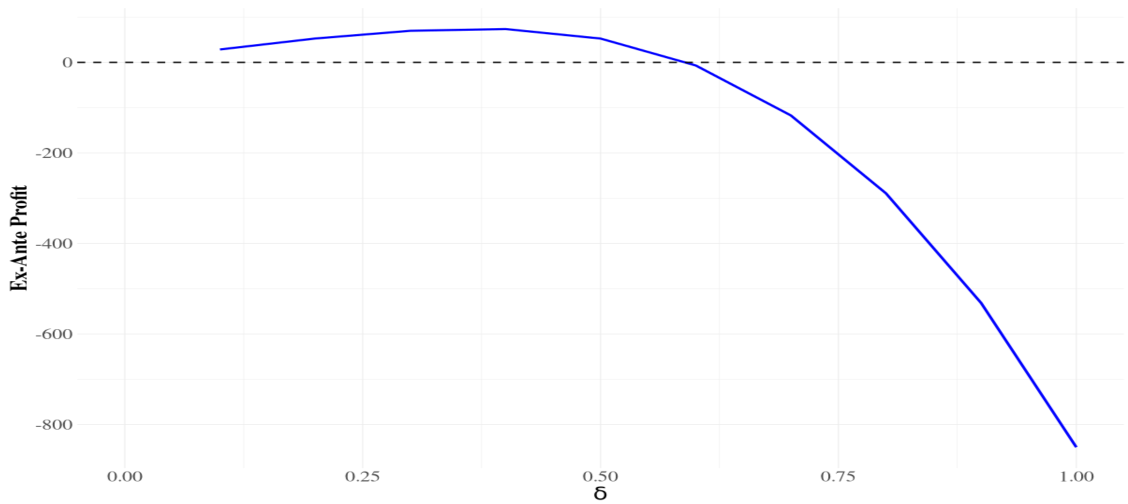

Figure 39: Expected Profit under Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of $100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to arise entirely from time constraints. Profits are simulated using numerical expectations based on post-entry housing transactions in Tampa.

Case 2: Hassle Costs Unrelated to Timing (e.g., Psychological Burden)

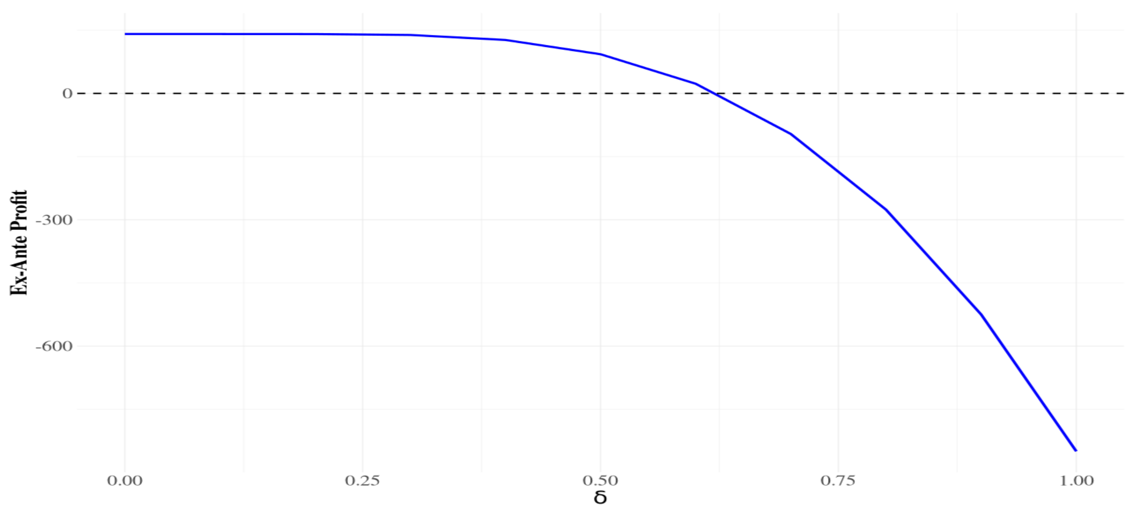

Figure 40: Expected Profit under Non-Time-Constrained Hassle Costs (Unit: 100,000 dollars)
Note: This figure shows the expected iBuyer profit per transaction (in units of $100,000) under a counterfactual contract with varying δ (the fraction of the original iBuyer offer price, p^i_{ht}, paid upfront). Hassle costs are assumed to be unrelated to timing. Simulations are based on numerical expectations from post-entry housing transactions in Tampa.

A Online Appendix: Software and Estimation Details

The market pricing model is estimated using the plm package in R, which implements linear panel data models with random effects. The iBuyer pricing model is estimated using the base R lm() function for ordinary least squares regression. I verified that the results from both pricing models are consistent with those obtained from a manually coded Generalized Method of Moments (GMM) estimator. The simulated maximum likelihood estimation for the seller's choice model is conducted using R' s optim() function from the base optimization package.

References

E. Anenberg. Information frictions and housing market dynamics. *International Economic Review*, 57(4):1449–1479, 2016.

E. Anenberg and P. Bayer. Endogenous sources of volatility in housing markets: The joint buyer–seller problem. *International Economic Review*, 61(3):1195–1228, 2020.

Bank of America Corporation. Understanding Down Payments, 2025. URL https://www.bankofamerica.com/mortgage/learn/mortgage-down-payment/?subCampCode=98969&sourceCd=18168. Accessed on July 14, 2025.

P. Bayer, R. McMillan, and K. Rueben. An equilibrium model of sorting in an urban housing market, 2004.

P. Bayer, K. Mangum, and J. W. Roberts. Speculative fever: Investor contagion in the housing bubble. *American Economic Review*, 111(2):609–651, 2021.

Board of Governors of the Federal Reserve System. Changes in U.S. Family Finances from 2019 to 2022: Evidence from the Survey of Consumer Finances, 2023. URL https://www.federalreserve.gov/publications/files/scf23.pdf. Accessed on September 17, 2025.

G. Buchak, G. Matvos, T. Piskorski, and A. Seru. Why is intermediating houses so difficult? evidence from ibuyers. Technical report, National Bureau of Economic Research, 2020.

S. Calder-Wang. The distributional impact of the sharing economy on the housing market. *Available at SSRN 3908062*, 2021.

A. Caplin, J. H. Carr, F. Pollock, Z. Yi Tong, K. M. Tan, and T. Thampy. Shared-equity mortgages, housing affordability, and homeownership. *Housing Policy Debate*, 18(1):209–242, 2007.

J. H. Cardon and I. Hendel. Asymmetric information in health insurance: evidence from the national medical expenditure survey. *RAND Journal of Economics*, pages 408–427, 2001.

P.-A. Chiappori and B. Salanie. Testing for asymmetric information in insurance markets. *Journal of political Economy*, 108(1):56–78, 2000.

A. Cohen and L. Einav. Estimating risk preferences from deductible choice. *American economic review*, 97(3):745–788, 2007.

G. Compiani, I. Morozov, and S. Seiler. Demand estimation with text and image data. *arXiv preprint arXiv:2503.20711*, 2025.

CoreLogic. CoreLogic Transaction and MLS Recordings. CoreLogic, Inc., 2024. Proprietary dataset. Accessed in September 2024.

Federal Reserve Economic Data (FRED). S&P/Case-Shiller U.S. National Home Price Index (CSUSHPISA). Federal Reserve Economic Data (FRED), Economic Research Division, Federal Reserve Bank of St. Louis, 2024a. Available at: https://fred.stlouisfed.org, Accessed in February 2024.

Federal Reserve Economic Data (FRED). Federal Funds Rate (FEDFUNDS). Federal Reserve Economic Data (FRED), Economic Research Division, Federal Reserve Bank of St. Louis, 2024b. Available at: https://fred.stlouisfed.org, Accessed in February 2024.

Federal Reserve Economic Data (FRED). 30-Year Fixed Rate Mortgage Average in the United States (MORTGAGE30US). Federal Reserve Economic Data (FRED), Economic Research Division, Federal Reserve Bank of St. Louis, 2024c. Available at: https://fred.stlouisfed.org, Accessed in February 2024.

A. Finkelstein and K. McGarry. Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, 96(4):938–958, 2006.

R. Fu, G. Z. Jin, and M. Liu. Does human-algorithm feedback loop lead to error propagation? evidence from zillow' s zestimate. 2022.

Internet Archive. Archived Offerpad Website, 2018a. URL https://web.archive.org/web/20180627000000/https://www.offerpad.com. Archived on June 27, 2018, via the Wayback Machine.

Internet Archive. Archived Opendoor Website, 2018b. URL https://web.archive.org/web/20180429000000/https://www.opendoor.com. Archived on April 29, 2018, via the Wayback Machine.

Internet Archive. Archived Opendoor Website, 2019. URL https://web.archive.org/web/20190331000000/https://www.opendoor.com. Archived on March 31, 2019, via the Wayback Machine.

K. Lee. Generative brand choice. 2025.

Meta AI. Llama-3.2-1B-Instruct. https://huggingface.co/meta-llama/Llama-3.2-1B-Instruct, 2024. Accessed: 2025-05-07.

National Association of Realtors. Existing-Home Sales Report Shows 2.0% Increase in July, 2025. URL https://www.nar.realtor/newsroom/nar-existing-home-sales-report-shows-2-0-increase-in-july. Accessed on September 17, 2025.

Offerpad LLC. Offerpad Official Website, 2025. URL https://www.offerpad.com. Accessed on February 18, 2025.

Opendoor Technologies Inc. Form 10-k annual report. U.S. Securities and Exchange Commission (SEC) Filing, 2023. Available at: https://investor.opendoor.com/financials-filings/quarterly-reports, Accessed on January 26, 2025.

Opendoor Technologies Inc. Opendoor Official Website, 2025. URL https://www.opendoor.com. Accessed on February 18, 2025.

T. W. Quan and K. R. Williams. Extracting characteristics from product images and its application to demand estimation. *University of Georgia, Department of Economics*, 2019.

L. Raymond. The market effects of algorithms. Technical report, Working Paper, 2023.

M. Rothschild and J. Stiglitz. Imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649, 1976.

RubyHome. ibuyer stats: ibuying companies and market trends, 2022. Available at: https://www.rubyhome.com/blog/ibuyer-stats, Accessed on January 27, 2025.

U.S. Census Bureau. Wealth and asset ownership, 2022. https://www.census.gov/data/tables/2022/demo/wealth/wealth-asset-ownership.html, Accessed on April 30, 2025.

World Bank. World Bank High Frequency Inflation Database (HCPI). World Bank Group, Research Department, 2024. Available at: https://www.worldbank.org/en/research/brief/inflation-database, Accessed in February 2024.

Zillow Group Inc. Steps to Selling a House, 2024. URL https://www.zillow.com/learn/steps-to-selling-a-house/. Accessed February 19, 2025.