Percentage Fees in Thin Markets: An Optimal Pricing Perspective *

Simon Loertscher† Andras Niedermayer‡

December 26, 2016

Note: Online Appendices begin on page 62.

Abstract

Transaction fees and taxes fare prominently in public policy debates and antitrust cases. We derive transaction fees as the outcome of optimal pricing by intermediaries, show that extreme value theory implies asymptotic optimality of linear fees in thin markets and that, counterintuitively, more elastic demand may increase fees. We estimate a structural model based on our theory using real-estate data from Boston. Our counterfactual analyses show that, consistent with our theory of thin markets, percentage fees are nearly optimal, and that the effect of additional seller entry offsets almost exactly the price-decreasing effect of reductions in fees.

Keywords: brokerage, fee-setting, percentage fees, thin markets, Pareto distributions.

JEL-Classification: C72, C78, L13

*We want to thank Arnaud Costinot, Julio González-Díaz, David Genesove, Ali Hortacsu, Philipp Kircher, Stephan Lauermann, Aviv Nevo, Volker Nocke, François Ortalo-Magné, Marco Ottaviani, Harry Paarsch, Martin Peitz, Jean-Marc Robin, Bernard Salanié, Juan Santaló, Mark Satterthwaite, Philipp Schmidt-Dengler, Art Shneyerov, Dan Spulber, Adam Szeidl, Yuya Takahashi, Ernst-Ludwig von Thadden, Cédric Wasser, Asher Wolinsky, Abdullah Yavas, participants at the IIOC 2012 in Washington D.C., the SFB TR15 2012 meeting in Mannheim, the Swiss IO Day 2012 in Bern, EARIE 2012 in Rome, and seminar audiences at Bates White, Bristol, Columbia, Deakin, Georgetown, Hamburg, LSE, Mannheim, Melbourne, Michigan, Munich, UNSW, Northwestern, Sydney, WHU, Zurich, and the University of Pennsylvania for their comments and suggestions. We want to thank Mark Satterthwaite for stating the conjecture of convergence to linear fees as the opportunity costs of trading increase and for having encouraged us to prove this result and François Ortalo-Magné for highlighting to us the limitations of the principal-agent model in explaining the use of percentage fees by real-estate brokers. We are also grateful to David Genesove and Chris Mayer for giving us access to their data set and to Christian Michel for excellent research assistance. Financial support through a research grant by the Faculty of Business and Economics at the University of Melbourne is also gratefully acknowledged. The second author gratefully acknowledges financial support by the Deutsche Forschungsgesellschaft through SFB TR-15.

†Department of Economics, Level 4, FBE Building, University of Melbourne, 111 Barry St, Victoria 3010, Australia. Email: simonl@unimelb.edu.au.

‡Economics Department, University of Mannheim, L7, 3-5, D-68131 Mannheim, Germany. Email: aniederm@rumms.uni-mannheim.de.

1 Introduction

Internet platforms, brokers, and governments around the globe charge commission fees and taxes that are paid when a transaction occurs. Examples include the indirect taxes imposed by governments, the percentage fees charged by real-estate brokers, headhunters, and stock brokers, and the commission fees charged by auction houses and trading platforms such as Sotheby's, Christie's, eBay, iTunes, and Amazon. In the United States alone, the annual volume of trade via intermediaries who charge fees is in the order of trillions of USD.[1] Transaction fees fare prominently in public policy debates on issues as various as credit-card fees, allegations of collusive commission fee-setting by auction houses and real-estate agents, antitrust investigations into pricing by internet platforms, the – at times drastic – increases of value-added taxes in the wake of the global financial crisis, or the introduction of financial transaction taxes in the European Union (EU).[2] Little is known about the effects and determinants of the structure of such fees and as to when the use of such transaction fees dominates alternative trading mechanisms.

In this paper, we analyze the fee structure from an optimal pricing perspective. Our focus is on thin markets in which every seller owns a unique object and faces only a small number of buyers in every period. Assuming independent private values on both the seller's and on the buyers' side, we show that there is no loss of generality by focusing on fee-setting insofar as no mechanism fares better. We show the asymptotic optimality of linear fees in increasingly thin markets and relate it to asymptotic results from extreme value theory. There is a surprising, deep connection between linear fees and extreme value theory: as markets become increasingly thin, the distributions of participants' valuations converge to Generalized Pareto distributions, which in turn implies convergence to linear

[1] See, for example, Department of Justice (2007) for real-estate brokers and Business Wire, February 2014: "U.S. 2013 Credit and Debit Card Purchases Increased 8% Over 2012: Break $4 Trillion Barrier for First Time" for credit and debit cards.

[2] The European Commission published its revised proposal to introduce a financial transaction tax on February 14, 2013 (see Economist, February 23, 2013: "Europe's financial-transactions tax: Bin it"). A previous version of the proposal already had been approved by the European Parliament and the Council of the European Union. Notwithstanding the Tobin tax as the standard economics rationale for such a measure, the proposal for a financial-transactions tax has been primarily motivated by a desire or need to increase tax revenue. For Amazon, which uses linear fees in 33 of the 38 categories it uses for third-party sellers, revenue from the long tail of the distributions appears to be quantitatively important.

fees. Besides providing an explanation for the linear fees often used in practice, our asymptotic results allow for a surprisingly tractable analysis for an otherwise intractable problem. This allows us to obtain additional results. For example, more elastic demand may lead to higher optimal fees. While counterintuitive at first, these and other results can be explained with concepts from monopoly and monopsony pricing. The reason for the counterintuitive effect is that the seller's price responds endogenously to increases in the elasticity of demand. When this reaction is excessive, a fee increase may be called for to partly offset this price endogeneity effect. For the limiting linear fees the seller's and the intermediary's pricing incentives are perfectly aligned, so that the optimal fee is independent of the elasticity of demand. Further, allowing for free entry by sellers, the price-decreasing effect of a reduction of fees is in the limit exactly offset by the price-increasing effect of additional entry by high-cost sellers.

On the surface, linear fees in intermediated markets may appear similar to the more familiar concept of linear pricing in standard markets. However, linear fees are fundamentally different from linear pricing for a variety of reasons such as single-unit demand for indivisible objects and the linearity of the fee in *price* rather than quantity.[3] Because it would take a leap of faith to believe that explanations put forth for the optimality of linear pricing carry over directly to linear fees, our results on optimal linearity are novel.[4]

We supplement our theoretical results with an empirical analysis. Such an analysis

[3]The difference between linear pricing in quantity and linear fees in prices is fundamental and obvious, because production costs may plausibly be proportional to quantity, whereas costs of intermediation are unlikely to be proportional to the price. Further, we use the term "linear fee" to describe a fee that is possibly a fixed fee plus a percentage of the price, whereas "linear pricing" refers to pricing that is *proportional* to the quantity.

[4]One explanation for linear pricing is the possibility of arbitrage: if two units of a good are less than twice as expensive as one unit, a buyer might buy two units, consume one unit and resell the other. However, it should be clear that this does not apply to intermediaries: if the percentage fee were different for a house that sells at $200,000 than for a house that sells at $400,000, there is no possibility of arbitrage for the seller. Another explanation given for linear pricing is that sufficient competition makes price discrimination impossible. However, as it will become clear later, *linear fees* in contrast to *linear pricing* do not mean the absence of price discrimination. A further statement sometimes made about linear pricing is that there is not really any economic explanation for linear prices, they are just plain simpler. Leaving aside that such an explanation has little empirical content in general, for an intermediary's fees this is also not particularly convincing. For example, Amazon has an elaborate pricing scheme in which it charges different fees for 38 different categories of goods. 33 out of 38 categories have linear fees (i.e. a fixed fee plus a percentage of the transaction price).

is desirable for at least two reasons. First, asymptotic results in statistics are among the most commonly misunderstood concepts.[5] An empirical analysis serves as an illustration that helps avoiding such misunderstandings. Second, our approach of using extreme value theory in the context of linear fees, and mechanism design, is novel. Novelty raises another question: Are the assumptions needed for the applicability of extreme value theory plausible in real world settings? There are two definitions of plausibility. An assumption is plausible because it has often been made in theoretical papers, or, alternatively, because there is empirical evidence to back it up. Our empirical analysis provides plausibility of the latter kind.[6]

Using the real-estate data set from the Boston condominium market of Genesove and Mayer (1997, 2001), we construct a structural empirical model, show that it is non-parametrically identifiable and estimate it. Our first counterfactual analysis shows that the empirically observed 6 percent fees are nearly optimal and achieve more than 99 percent of what is achievable with the – otherwise unconstrained - optimal Bayesian mechanism that respects agents' incentive compatibility and individual rationality constraints. Our predictions based on extreme value theory are thus almost exactly borne out empirically. Our second counterfactual experiment consists of regulating percentage fees. Consistent with our theory, we show that the effect of fee decreases is almost completely offset by the additional entry of high-cost sellers. We also introduce a transfer tax, as imposed by governments in various jurisdictions, and show that quantitatively the most important factor on the welfare of buyers, sellers, and brokers is that sellers adjust their prices endogenously.

We have made a conscious modeling choice in this article: we are modeling optimal pricing behavior by intermediaries rather than the specifics of the services of intermediaries. There are good reasons for this choice. There is a wide variety of intermediaries

[5] A common misunderstanding is that asymptotic results are only applicable in one of two cases: either if one assumes very particular functional forms for distributions that are close to the limiting distribution or if one is exactly in the limit. As explained later, this is a false dichotomy.

[6] While we do not consider the former definition of plausibility particularly useful, it is still worth mentioning that extreme value theory has been successfully applied in a number of contexts, such as the distribution of income and wealth, the sizes of cities, and in many areas outside of economics, such as the sizes of the largest earthquakes.

charging fees, such as real-estate brokers, Amazon, eBay, and iTunes, offering a wide variety of services. If we add taxation by governments into the picture, there is even more heterogeneity. There are a variety of mutually non-exclusive explanations as to why the services of private intermediaries and governments are useful or necessary, such as for reducing transaction and search costs, certifying quality, improving matching, building reputation, providing infrastructure that facilitates trade, and enforcing contracts.[7] Rather than going through all the combinations of such explanations, which will be special for every industry (and even for a single industry, there is typically no consensus on which services are the most important), we provide a general model and focus on what is common to all intermediaries: that they raise revenues by charging transaction fees. This approach does come at a price: our theory remains silent on what is special about the services of one or another type of intermediary. However, we believe this price is worth paying, since we gain a deeper, more general understanding of the determinants of fees. There is precedent of such an approach paying off: industrial organization has developed a number of theoretical and structural estimation tools to deal with optimal pricing by one-sided (i.e. non-intermediary) firms. These general tools have turned out to be very useful, despite the fact that optimal pricing is done by very different firms offering very different products such as cereals, cars, and pharmaceutical products (see e.g. Berry, Levinsohn, and Pakes (1995) and the subsequent literature) and despite the fact that an optimal pricing approach remains silent about the question why consumers buy cereals, cars, pharmaceutical products, and so on.

Another, related question is whether the fees charged by real-estate brokers should alternatively be viewed as the solution of a principal-agent problem. Indeed, economic theorists often casually refer to real-estate brokerage as an example of a principal-agent problem. The (informal) theory typically looks like the following. A seller hires a real-estate agent and incentivizes him to drive up the sales price by offering a fee that increases with the price. While this theory is superficially appealing, it stands in stark contradiction with many empirical observations made in the real-estate brokerage market that we

[7] See Spulber (1999) and Salanié (2003) for an overview of the role of private intermediaries and governments, respectively.

will describe in detail later.[8] The many puzzles that appear when viewing real-estate brokerage fees from a principal-agent perspective all disappear if one is willing to take a novel approach and consider brokerage fees from an optimal pricing perspective. If one further has the ambition to have a theory that explains the fees not only of real-estate brokers, but also of Amazon, eBay, and possibly even the taxes levied by governments, then a principal-agent explanation becomes even less convincing: it is e.g. hard to believe that the reason why eBay's fees increase with the transaction price is that sellers are trying to incentivize eBay to exert effort to raise the price.

Related Literature. First and foremost, our paper contributes to the growing literature that applies insights and methods from mechanism design to pertinent questions in industrial organization. Recent and complementary contributions, such as Board (2008), Gomes (2014), Tirole (2016), Garrett (forthcoming), have applied multi-period mechanism design to intertemporal pricing and to optimal incentive schemes for platform participation. Our paper, and the predecessor (Loertscher and Niedermayer, 2007) it builds on and supersedes, is the first paper to connect fee-setting to optimal pricing in thin markets with two-sided private information as first studied by Myerson and Satterthwaite (1983).

This combination of market thinness and two-sided private information is also what sets the theoretical part of our paper apart from the existing theoretical literature on the transaction fees of profit maximizing intermediaries (Yavas (1992), Caillaud and Jullien (2003), Hagiu (2007), Matros and Zapechelnyuk (2008), Shy and Wang (2011), Niedermayer and Shneyerov (2014), Johnson (2014), and Wang and Wright (forthcoming)) and on the indirect taxes charged by governments (Salanié (2003, Chapter 3), Delipalla and Keen (1992), and Anderson, De Palma, and Kreider (2001a,b)).[9] Without this combi-

[8] To mention only one of the many observations we discuss later here, consider a transaction in which both the seller and the buyer have a broker. For such transactions, both brokers get 3% of the transaction price. This means that the buyer's broker gets *more* if the transaction price is *higher*, that is if the outcome is *less* favorable from the buyer's perspective, something that cannot be explained by a principal-agent theory.

[9] Yavas (1992), Caillaud and Jullien (2003), Shy and Wang (2011), Johnson (2014), and Wang and Wright (forthcoming), whose work is subsequent to ours, assume that the seller's cost is public information (or, equivalently, that either there is no uncertainty about the seller's cost or that there is perfect

nation, the theory would be silent about the functional form of the fee, that is, as to whether it is fixed, a percentage fee or a non-linear fee.[10,11] This highlights a robustness of linear fees: while in thin markets they are needed for optimality, in thick markets they do no harm, being equivalent to alternative ways of raising revenues. Moreover, our model predicts equilibrium price dispersion, which is consistent with the data but absent in most of the aforementioned models. Jullien and Mariotti (2006) assume two-sided private information in a static model with one broker and two buyers, but focus on fees that are a function of the reserve price rather than the transaction price without specifying the functional form of these fees.[12]

Our paper also contributes to the literature on dynamic random matching with search and matching frictions such as Wolinsky (1988), Rust and Hall (2003), Satterthwaite and Shneyerov (2007), Lauermann (2013), and Lauermann, Merzyn, and Virag (2012) by emphasizing the role market thinness and two-sided private information play in determining

competition between sellers). In Matros and Zapechelnyuk (2008) the seller's cost is sunk after he chooses to go to the intermediary. Therefore, the seller's private cost only matters for his participation decision, but not for anything that happens after he chooses to participate (in particular for the reserve and transaction price). In Niedermayer and Shneyerov (2014) there is a continuum of sellers and buyers, so that by the law of large numbers there is no uncertainty about the realized distribution of sellers' costs. None of these models can account for the counterintuitive effect of the elasticity of demand on the equilibrium fees. Salanié (2003, Chapter 3) provides an overview of the literature on indirect taxes in competitive (that is, thick) markets. Delipalla and Keen (1992); Anderson, De Palma, and Kreider (2001a,b) consider optimal taxation with imperfect competition and *public* information about the seller's cost. The lack of relevant two-sided private information is what leads to the finding in these articles that optimal fees or taxes are higher if demand is *less* elastic. Moreover, models that assume thick markets generate an irrelevance result concerning the functional form of the fee or tax (fixed, percentage, linear, or non-linear), because absent any uncertainty about he seller's cost, the optimal mechanism for the intermediary is to set the seller indifferent and choose optimal one-sided pricing for buyers.

[10] If the seller's cost is known to the intermediary, an optimal *unrestricted* mechanism for the intermediary is to cap the maximum price the seller can set at the monopoly price and charge a fee that is the difference between the seller's cost and the monopoly price, this fee can be any arbitrary linear or non-linear fee. Even if the seller's cost is not known to the intermediary, but known to other market participants, the same results hold because the intermediary can extract information about the seller's cost costlessly by a Cremer and McLean (1988) type of mechanism. Some of the above papers restrict the set of mechanisms the intermediary can choose in such a way that in the restricted set fees other than percentage fees are suboptimal.

[11] The setup with two-sided private information provides a coherent, internally consistent framework to analyze indirect taxation. Without private information about the seller's cost (and no fixed costs) and without imposing exogenously given constraints in policy instruments, the government could achieve first-best by forcing sellers to price at marginal costs. However, such large scale intervention across many different industries seems a daunting task for any government.

[12] A further difference, as mentioned before, is that we have multiple buyers and multiple period, and also structurally estimate the model.

the optimal fees.

Our paper provides new insights for the theory of optimal pricing, for public finance, and for competition policy. We show that results change fundamentally if there are not only buyers to consider, but both buyers and sellers. An example mentioned before is that the optimal fee may be higher if the elasticity of demand is higher.[13] While setups and questions differ, our paper has also similarities to the empirical work on auctions, by explicitly modeling an important aspect of many real world auctions: that many auctions are run by profit maximizing intermediaries.[14] In a wider sense, our paper relates to the recent literature on the role of intermediaries in international trade, see Antràs and Costinot (2011) and the references therein. Our application of extreme value theory relates to importance power laws (including the Pareto distribution) in a variety of economic contexts, which have been described as one of the most fundamental principles in economics by Gabaix (2016).

Last but not least, our paper contributes to the empirical literature on real-estate brokerage (Hsieh and Moretti (2003), Rutherford, Springer, and Yavas (2005), Levitt and Syverson (2008), Hendel, Nevo, and Ortalo-Magné (2009), Barwick, Pathak, and Wong (forthcoming), see also the survey of Han and Strange (2014)). This literature has

[13] If optimal fees are linear, then the fees are independent of the demand side. This result, and the counterintuitive result that more elastic demand can lead to higher fees cannot be accounted for by existing theories in industrial organization such as those based on the eminent contributions by Bulow and Pfleiderer (1983), Aguirre, Cowan, and Vickers (2010), Bulow and Klemperer (2012) and Weyl and Fabinger (2013). In the context of public finance, our theory makes the normative prescription that whether one wants to tax the inelastically demanded good depends on the curvature of the virtual cost function. Similarly, when there are concerns of anticompetitive behavior by fee-setting intermediaries such as auction houses or real-estate brokers, our theory prescribes that, as a first-order approximation, the researcher's focus should be on the supply side.

[14] For the empirical auctions literature see e.g. Donald and Paarsch (1993), Bajari (1997), Bajari and Hortaçsu (2003), Shneyerov (2006), and Balat, Haile, Hong, and Shum (2016). Of these papers, the most closely related are Bajari and Hortaçsu (2003) because the auctions they analyze are run by an intermediary – eBay – that charges a transaction fee. Further, this is to the best of our knowledge the first paper to structurally estimate an auction in which *the seller* has an informational advantage over buyers (see the Appendix). In Appendix E we develop and estimate a structural model based on Cai, Riley, and Ye (2007) with a common-value component, which we augment to allow for dynamics and percentage fees set by intermediaries. We show that empirically the common-value component is negligible. To the best of our knowledge, ours is the first paper to provide a structural estimation of the model of Cai, Riley, and Ye (2007). For a reduced-form estimation of another model of auctions with a common value component on the seller's side, see Niedermayer, Shneyerov, and Xu (2015), who extend Cai, Riley, and Ye (2007) in a different direction: to foreclosure auctions, where one bidder (the bank) sometimes acts as a seller and sometimes as a buyer.

documented a number of empirical observations which appear puzzling from a principal-agent perspective. As mentioned, an optimal pricing perspective resolves these puzzles. In a wider sense, this article also contributes to the literature on real-estate markets in general (Genesove and Mayer (1997, 2001), Ortalo-Magné and Rady (2006), Genesove and Han (2012), Ortalo-Magné, Merlo, and Rust (2015)), by estimating a structural search model of housing transaction with real-estate agents.[15,16]

The remainder of this paper is organized as follows. Section 2 sets up and analyzes the theoretical model. Section 3 contains the empirical analysis while the results from the counterfactual analyses are reported in Section 4. Section 6 concludes. Proofs and additional background material are in the Appendix.

2 Theory

2.1 Model

Setup Motivated by the widespread use of fee-setting in intermediated markets with both long-term and spot contracts, we set up and analyze a general infinite horizon model. Time is discrete and indexed by $t = 0, 1, ...$ The discount factor is $\delta \in [0, 1)$. This nests the static model as a special case. The basic analysis assumes that there is one intermediary and one seller. We provide extensions which relax this. The seller's *primitive* cost c_0 is the seller's private information and drawn from the primitive distribution G_0 with support $[\underline{c}_0, \overline{c}_0]$ and density $g_0(c_0) > 0$ for all $c_0 \in (\underline{c}_0, \overline{c}_0)$. His value of the outside option of not participating is zero. The cost c_0 can equivalently be thought of as the opportunity cost of selling or as a cost of production, both accruing to the seller in the period he sells.[17] The seller and the intermediary have the common discount factor

[15] See also the analysis by Ortalo-Magné, Merlo, and Rust (2015), who focus on a seller's optimal listing price strategy over time rather than brokerage fees in a structural estimation of a real-estate market in England.

[16] We also highlight the importance of heterogeneous valuations in a search model in real-estate markets: we show that there is considerable price dispersion even when considering a relatively homogeneous market only consisting of downtown condominiums, even after correcting for observable heterogeneity in quality using two different quality indices (hedonic pricing and price index corrected previous transaction prices), and even after using a structural model to take into account unobserved heterogeneity.

[17] For example, if the good is a real-estate property, the opportunity cost of selling is given by the discounted stream of income from renting the property or the discounted value of the flow utility from using the property.

$\delta \in [0,1)$, which may represent time preferences or the period-to-period probability that the seller stays in the market as in Satterthwaite and Shneyerov (2008), or a combination thereof.

We assume that in every period there is a fixed number of potential buyers $\overline{B}$, each of whom enters with the independent probability $\tilde{\pi}$, so that the probability π_B of having exactly $B \leq \overline{B}$ buyers is given by the probability mass function for the binomial $\pi_B = \begin{pmatrix} \overline{B} \\ B \end{pmatrix} \tilde{\pi}^B (1-\tilde{\pi})^{\overline{B}-B}$. Buyers who participate are sometimes also called bidders. Each bidder draws her primitive valuation v_0 independently from the (primitive) distribution F_0 with support $[\underline{v}_0, \overline{v}_0]$ and density $f_0(v_0) > 0$ for all $v_0 \in (\underline{v}_0, \overline{v}_0)$. The value of the outside option of not participating is zero for all buyers. All players – buyers, the seller, and the intermediary – are risk-neutral.

We call the buyer's valuation v_0 his *primitive valuation* as we assume that there are additionally transaction costs, such as shipping costs, the opportunity cost of buying later from another seller, and specifically for real estate – moving costs and the opportunity cost of renting rather than buying. The *effective valuation* $v := K^B + \hat{K}^B v_0$ takes into account these transaction costs, where K^B should be thought of as type independent transaction costs (such as shipping costs) and $\hat{K}^B$ as type dependent costs (such as the opportunity cost of renting). With the exception of Section 2.2, we will treat the cost parameters K^B and $\hat{K}^B$ as exogenous and fixed and simplify notation by dealing with the effective valuation v and its corresponding distribution F with support $[\underline{v}, \overline{v}]$. Analogously, on the seller's side denote the seller's *effective costs* as $c := K^S + \hat{K}^S c_0$ with the corresponding distribution G and support $[\underline{c}, \overline{c}]$.[18]

Denoting by f and g the densities of F and G, respectively, we assume that the functions

$$\Phi(v) := v - \frac{1-F(v)}{f(v)} \quad \text{and} \quad \Gamma(c) := c + \frac{G(c)}{g(c)}$$

are monotonically increasing in their arguments and continuously differentiable. Following Myerson (1981), $\Phi(v)$ is often called the virtual valuation while $\Gamma(c)$ can be thought

[18] Formally, the respective distributions and supports are given by $G(c) := G_0((c-K^S)/\hat{K}^S)$ with support $[\underline{c}, \overline{c}]$ and $F(v) := F_0((v-K^B)/\hat{K}^B)$ with support $[\underline{v}, \overline{v}]$, where $\underline{c} := \hat{K}^S \underline{c}_0 + K^S$, $\overline{c} := \hat{K}^S \overline{c}_0 + K^S$, $\underline{v} := \hat{K}^B \underline{v}_0 + K^B$, and $\overline{v} := \hat{K}^B \overline{v}_0 + K^B$.

of as a virtual cost function.[19] For most of the analysis, we simplify notation by assuming that $\underline{c} = \underline{v}$ and $\overline{c} = \overline{v}$, an assumption that turns out not to be restrictive.[20]

A sequence of fee functions $\boldsymbol{\omega} = (\omega_t)_{t=0}^{\infty}$ with $\omega_t(\breve{p})$ specifies the amount the seller has to pay to the intermediary when a transaction occurs in period t at the transaction price $\breve{p}$. To fix ideas, we assume that the transaction price is determined by an English auction with reserve price p_t set by the seller. It may be literally the case that the seller uses an English auction, as is for example the case on eBay and in some real-estate auctions, or the English auction may serve as a model for the way bargaining between the seller and buyers unfolds.[21] For example, in real-estate transactions bargaining is typically intermediated by the broker who keeps buyers informed about the highest standing offer, so that the ensuing bargaining game is equivalent to an English auction. The game ends in period t when a buyer bids higher than p_t.[22] The seller is not allowed to recall buyers after the period in which they arrived.[23] We assume full commitment throughout the paper.

The above specification is sufficiently general to include a number of setups that are of great applied interest. For auction platforms and auction houses (eBay, Sotheby's, Christie's), consider a one-shot setup ($\delta = 0$) and a binomial distribution of buyers ($\overline{B} > 1$) who literally participate in an English auction. For Amazon, a third-party

[19] Interpreting $G(p)$ and $1 - F(p)$ as expected quantities supplied and demanded, $\Phi(p)$ and $\Gamma(p)$ have the interpretation of marginal revenue and marginal cost functions; see Bulow and Roberts (1989).

[20] If one starts out with $\underline{c}_0 = \underline{v}_0$ and $\overline{c}_0 = \overline{v}_0$, then after taking into account the additional transaction costs, one ends up with $\underline{c} \leq \underline{v}$ and $\overline{c} \leq \overline{v}$. Sellers with costs $c > \overline{v}$ cannot find a buyer with whom they have positive gains from trade and hence can be ignored. Similarly, buyers with valuations $v < \underline{c}$ cannot find a seller with whom they have positive gains from trade and hence can be ignored. Therefore, we only need to consider sellers and buyers in the interval $[\underline{c}, \overline{v}]$ and truncate and rescale G and F to have this common support.

[21] There are many setups that are formally equivalent. For example, given that buyers have dominant strategies, it does not matter whether the reserve price is public or remains private information of the seller. The bargaining may also be such that the seller keeps rejecting bids that are below the maximum of his reserve and the highest standing bid by any buyer, allowing rejected buyers to revise their bids upwards. As briefly discussed in Section 2.5, it is also immaterial whether the auction format is an English auction or a first-price auction, provided fees are linear and the reserve price is known by the time buyers submit their bids. In our data set, these models are observationally equivalent. Importantly, however, none of our counterfactual analyses depend on the specifics of the setup.

[22] If $1 - \delta$ is interpreted as the probability that the seller drops out from one period to the next, the game can also end when the seller drops out.

[23] As shown by Riley and Zeckhauser (1983), this assumption is without loss of generality with a commonly known distribution F when the seller can commit to an optimal strategy and when one buyer enters in every period.

seller offers the good to a potential buyer at a fixed price (for $\overline{B} = 1$ the English auction with a reserve price reduces to a fixed price) in a single period ($\delta = 0$). For real-estate brokerage, a seller offers his house in multiple periods ($\delta > 0$) to a Poisson distributed random number of buyers that potentially arrive in every period (a Poisson arrival rate is the limit when $\overline{B} \to \infty$ and the expected number of buyers $\tilde{\pi}\overline{B}$ is kept constant) and bargaining is modeled as an English auction. In all these cases, intermediaries raise revenues by charging transaction fees.

While some of the economic insights from our model can also be obtained in a static setup ($\delta = 0$), there are a number of reasons why it is desirable to have a dynamic model. First, the fact that sellers offer their good for sale in multiple periods is an important feature of many real world markets. Second, some of the microfoundations for the transaction costs we will provide later are most naturally expressed in a dynamic environment. Third, for our empirical analysis of real estate brokerage fees, uses an important aspect – time on market – which can only be dealt with in a dynamic model.

For a given ω_t, the seller's expected net revenue $R_{\omega_t}(p_t)$ in period t conditional on a transaction occurring and given reserve p_t is

$$R_{\omega_t}(p_t) = \frac{(p_t - \omega_t(p_t))(F_{(2)}(p_t) - F_{(1)}(p_t)) + \int_{p_t}^{\overline{v}} (\breve{p} - \omega_t(\breve{p}))dF_{(2)}(\breve{p})}{1 - F_{(1)}(p_t)},$$

by standard arguments from auction theory (see e.g. Krishna, 2002), where $F_{(1)}(v) := \sum_{B=0}^{\infty} \pi_B F(v)^B$ and $F_{(2)}(v) := F_{(1)}(v) + (1 - F(v)) \sum_{B=1}^{\infty} \pi_B B F(v)^{B-1}$ are, respectively the unconditional distribution of the highest and the second-highest valuation.[24] Consequently, the maximization problem of a seller of type c given $\boldsymbol{\omega}$ is to choose a sequence of prices $\boldsymbol{p} = (p_t)_{t=0}^{\infty}$ to maximize his discounted expected profit

$$W_S(c, \boldsymbol{p}, \boldsymbol{\omega}) := \sum_{t=0}^{\infty} (R_{\omega_t}(p_t) - c)(1 - F_{(1)}(p_t)) \prod_{\tau=0}^{t-1} \delta F_{(1)}(p_\tau),$$

where we use the convention of setting $\prod_{\tau=x}^{x-1} y_\tau = 1$ for any sequence $(y_t)_{t=1}^{T}$. Let $\boldsymbol{P}(c) = (P_t(c))_{t=0}^{\infty}$ be the (or a) maximizer of $W_S(c, \boldsymbol{p}, \boldsymbol{\omega})$, whose dependence on the sequence of fees is kept implicit for ease of notation.

[24] The above expression can be obtained by observing that $1 - F_{(1)}(p_t)$ is the probability that a transaction occurs, the probability that the transaction price is equal to the reserve is $F_{(2)}(p_t) - F_{(1)}(p_t)$, and the distribution of transaction prices above the reserve is $\breve{p} \sim F_{(2)}$.

Given ω_t, the intermediary's expected revenue in period t when facing a seller of type c who sets the reserve price $p_t = P_t(c)$ is $\omega_t(p_t)(F_{(2)}(p_t) - F_{(1)}(p_t)) + \int_{p_t}^{\overline{v}} \omega_t(\breve{p}) dF_{(2)}(\breve{p})$. Therefore, the intermediary's discounted expected profit from a seller of type c given $\boldsymbol{p}$ and $\boldsymbol{\omega}$ is

$$W_I(c, \boldsymbol{p}, \boldsymbol{\omega}) := \sum_{t=0}^{\infty} \left(\omega_t(P_t(c))(F_{(2)}(P_t(c)) - F_{(1)}(P_t(c))) + \int_{P_t(c)}^{\overline{v}} \omega_t(\breve{p}) dF_{(2)}(\breve{p}) \right) \prod_{\tau=0}^{t-1} \delta F_{(1)}(P_\tau(c)).$$

We assume that the fees $\boldsymbol{\omega}$ are chosen to maximize a weighted average of the intermediary's profit W_I and the joint surplus of the intermediary and the seller $W_I + W_S$

$$W(\alpha, \boldsymbol{\omega}) := E_{c \sim G}[\alpha W_I(c, \boldsymbol{P}(c), \boldsymbol{\omega}) + (1-\alpha)(W_I(c, \boldsymbol{P}(c), \boldsymbol{\omega}) + W_S(c, \boldsymbol{P}(c), \boldsymbol{\omega}))], \quad (1)$$

where $\alpha \in [0, 1]$ is a parameter measuring the intermediary's bargaining power. It can also be interpreted as a measure of competition between brokers for sellers, with $\alpha = 0$ corresponding to perfect competition and $\alpha = 1$ corresponding to monopoly power (or perfect collusion by intermediaries), and the resulting fee structure as the outcome of bargaining between the intermediary and the seller. As shown below, $\alpha = 0$ implies that the fees are 0 for all prices. Observe also that the objective function in (1) depends on $\boldsymbol{\omega}$ directly and also indirectly via the seller's pricing behavior $\boldsymbol{P}(c)$, which depends on $\boldsymbol{\omega}$.[25]

The assumption that the intermediary and the seller bargain over the division (and size) of their joint surplus captures the notion that in many markets of interest, in particular in real-estate markets, sellers typically sign long-term exclusive dealership contracts with brokers. According to our modeling choice, sellers who are more patient than others would be characterized by larger opportunity costs of selling. Although space constraints refrain us from so doing, the setup can be extended to allow for sellers who have heterogenous deadlines and discount in non-stationary ways. Below we will show that our assumptions regarding the menu of mechanisms that can be chosen are without loss of generality. We will show that it is optimal to choose a second-price auction with a reserve price set by the seller with an appropriately chosen fee structure, and we discuss

[25] Maximizing $W(\alpha, \boldsymbol{\omega})$ is equivalent to maximizing the weighted sum $E_c[\alpha_0 W_I(c, \boldsymbol{P}(c), \boldsymbol{\omega}) + (1 - \alpha_0) W_S(c, \boldsymbol{P}(c), \boldsymbol{\omega})]$ with $\alpha_0 = 1/(2-\alpha)$.

conditions under which the results extend to first-price auctions. The assumptions that the environment is stationary and that F and G have the same support and exhibit monotone virtual valuation and cost functions are imposed for expositional simplicity as they do not affect the key insights from our analysis.[26]

It is important to keep in mind that optimality in our context means optimal pricing by an intermediary to extract rents (for $\alpha > 0$). It does not imply that the fees charged are socially optimal: as usual when dealing with optimal pricing by a firm that has market power and extracts rents, the pricing is not socially optimal.[27]

Seller Behavior Given a sequence of fee functions $\boldsymbol{\omega} = (\omega_t)_{t=0}^{\infty}$ the seller will choose a sequence of reserve prices $\mathbf{p} = (p_t)_{t=0}^{\infty}$ to maximize the expected net present value of his profits. In general, this maximization problem will be complex because the fees could be non-stationary and the implied profit function of the seller could fail to be quasi-concave, so that the first-order condition would not be sufficient. However, we will later show that even with an arbitrary non-stationary mechanism one could not do better than one can by charging fees which are stationary and imply a quasi-concave profit function for the seller. Therefore, to reduce the notational burden, we will focus on stationary fees and reserve prices, that is $\omega_t = \omega$ and $p_t = p$ for all t, and use the first-order condition for maximization.[28]

Given stationary fees $\boldsymbol{\omega}$ and stationary prices $\boldsymbol{p}$, the seller's utility becomes

$$W_S(c, \boldsymbol{p}, \boldsymbol{\omega}) = (R_\omega(p) - c)(1 - F_\infty(p)),$$

[26]In a previous version of our paper (Loertscher and Niedermayer, 2012), we have derived results for the cases when these assumptions do not hold. The results gave essentially the same economic insights, but the notation was far more tedious.

[27]We do not deal with the question of social optimality, because of different controversial aspects of many intermediaries which are orthogonal to the research question (optimal pricing) of this paper. Intermediaries may extract rents to cover fixed costs of operation, which is second-best efficient if one does not want government subsidized (or even government run) intermediaries. There is some controversy surrounding private intermediaries, e.g. the International Labor Organization demanded a ban of private fee-charging labor market intermediaries in its C96 convention (Fee-Charging Employment Agencies Convention (Revised), 1949). Instead, they demanded public employment agencies. Even if one agrees on having private intermediaries, one may be skeptical of an intermediary's ability of extracting rents, since an intermediary's market power may be due to collusion. For example, there is an allegation of collusion for real estate agents and a conviction for collusion of Sotheby's and Christie's.

[28]By using standard techniques it is possible to extend the analysis to non-optimal fees which imply a non-stationary non-quasi-concave problem.

where

$$1 - F_\infty(p) := (1 - F_{(1)}(p)) \left(\sum_{t=0}^{\infty} \delta^t F_{(1)}(p)^t \right) = \frac{1 - F_{(1)}(p)}{1 - \delta F_{(1)}(p)} \tag{2}$$

is the *ultimate probability of selling.*[29] Let

$$\Phi_\omega(p) := p - \omega(p) - (1 - \omega'(p)) \frac{1 - F(p)}{f(p)}$$

be the net virtual valuation associated with the stationary fee ω, and define

$$\tilde{\Phi}_\omega(p) := \overline{v} - \int_p^{\overline{v}} \frac{1 - \delta F_{(1)}(v)}{1 - \delta} \Phi'_\omega(v) dv.$$

The function $\tilde{\Phi}_\omega(p)$ is monotone and thus invertible if $\Phi_\omega(p)$ is monotone.

The seller chooses the reserve p to maximize W_S. The following proposition gives the solution to this maximization problem.

Proposition 1. *Given a stationary fee ω that implies a monotone net virtual valuation $\Phi_\omega(p)$, the optimal price set by a seller with cost c is $P(c) = \tilde{\Phi}_\omega^{-1}(c)$ in every period.*

As mentioned, we will show that the optimal fee is such that the seller's profit function is quasi-concave, which is equivalent to an increasing Φ_ω. One can show that Φ_ω (and hence also $\tilde{\Phi}_\omega$) is increasing, provided Φ is increasing and the fee ω is linear with slope less than 1, which proves helpful in the empirical application. For notational ease, we let $\tilde{\Phi}(v) := \tilde{\Phi}_0(v)$ and $R(p) := R_0(p)$. Note that $\tilde{\Phi}_\omega$ can be interpreted as the net dynamic virtual valuation function and satisfies $\tilde{\Phi}_\omega(\overline{v}) = \overline{v}$. In a static setup ($\delta = 0$), $\tilde{\Phi}_\omega$ simplifies to the net virtual valuation Φ_ω. If the fee is zero ($\omega(p) = 0$ for all p), it further simplifies to the virtual valuation function Φ.

Proposition 1 relates the seller's optimal reserve price $P(c)$ to the fee function ω. Once we know ω, we know $P(c)$ provided ω is monotone. Because the good will be sold to the buyer with the highest value in the first period in which this value exceeds $P(c)$, the search for the optimal fee function can be separated into two steps. First, find the pricing function that is jointly optimal for the intermediary and the seller. Second, find

[29] If one interprets the discount factor as the probability of drop-out, $1 - F_\infty$ is the probability of selling taking into account that one might die with probability $1 - \delta$ in every period. Satterthwaite and Shneyerov (2007, 2008) use a similar notion which they call the "ultimate discounted probability of trade".

the fee function that induces the seller to set the optimal price. A priori it is not clear whether such a fee function exists, but a key insight of our paper is to show that it does.

Optimal Fees The problem of maximizing $W(\alpha, \boldsymbol{\omega})$ over $\boldsymbol{\omega}$ is tedious and does not address whether the use of fee setting is without loss of generality. In Appendix A.2 we set up and solve the general mechanism design problem for our model, without imposing any constraints on the mechanism other than incentive compatibility[30] and individual rationality[31]. Mechanisms that solve this problem are called optimal. We show there that the focus on direct mechanisms – these are mechanisms that ask each agent to report his type upon arrival, provided the seller is still in the game – is without loss of generality and that revenue equivalence holds. That is, once the allocation rule is determined, the interim expected payoff of every agent of every type is determined by the allocation rule up to an additive constant, which in the optimal mechanism is set equal to zero because the individual rationality constraints will optimally bind.

Let $\Gamma_\alpha(c) := \alpha\Gamma(c) + (1-\alpha)c$ be the weighted average of the seller's virtual cost $\Gamma(c)$ and his type c. The key result from the mechanism design analysis is the following:

Lemma 1. *In any optimal mechanism, the good is sold, to the buyer with the highest valuation present in that period, in the earliest period t for which*

$$\max_{b_t} \tilde{\Phi}(v_{b_t}) \geq \Gamma_\alpha(c),$$

and the expected payoff of every buyer of type $\underline{v}$ and of the seller of type $\overline{c}$ is 0.

While the result is intuitive, the proof is surprisingly involved. The reason why one cannot use standard mechanism design techniques is that potential future buyers have not yet arrived, so they cannot be asked to reveal their types in the beginning. Further, it is not obvious how to compare probabilities of transactions and revenues from transactions today with probabilities and revenues in the future, because of discounting. To get around these difficulties, we use the concept of the *ultimate (discounted) probability of trade* introduced in Satterthwaite and Shneyerov (2007, 2008). Since on top of

[30]No participant has an incentive to choose an option that is meant for a participant of another type.
[31]Participants are willing to choose the option offered to them rather than the outside option.

the dynamic bargaining game with private information considered in Satterthwaite and Shneyerov (2007, 2008) we also have a mechanism design problem, we need to introduce an additional concept, the *ultimate conditional expected revenue.* These concepts are described in more detail in the proof in the Appendix.[32]

Lemma 1 generalizes Theorem 3 of Myerson and Satterthwaite (1983) to our dynamic setting with multiple buyers using the concept of the dynamic virtual valuation. It is based on the insight that in any optimal mechanism, the good goes to the buyer with the highest value in any given period if this value is above some threshold, and stays with the seller otherwise. Lemma 1 adds to this the insight that the good goes to the buyer with the highest dynamic virtual valuation, appropriately defined, provided it exceeds $\Gamma_\alpha(c)$. Intuitively, Myerson and Satterthwaite (1983) find in a one-buyer-one-seller-one-period setup that the good is transferred whenever $\Phi(v)$ (which can be interpreted as marginal revenue) is larger than $\Gamma(c)$ (which can be interpreted as marginal cost). In our setup, the dynamic virtual valuation $\tilde{\Phi}(v)$ has to be used because it adjusts for the option value of future trade and the weighted virtual cost $\Gamma_\alpha(c)$, which accounts for the weight on the seller's utility.

In light of the remarks after Proposition 1, Lemma 1 answers the first question: it derives the optimal allocation rule, which can be implemented via fee-setting if a seller of type c can be induced to set the reserve price

$$P^*(c) := \tilde{\Phi}^{-1}(\Gamma_\alpha(c)) \tag{3}$$

in every period in which he is active. Bidding in the second-price auction will ensure that the object goes to the buyer with the highest virtual valuation while the reserve price $P^*(c)$ insures that trade only takes place if this virtual value $\tilde{\Phi}$ exceeds $\Gamma_\alpha(c)$. The discounted probability that a seller of type c who always sets the price $P^*(c)$ ever sells is $1 - F_\infty(P^*(c))$. A seller with cost $\Gamma_\alpha^{-1}(\overline{v})$ should optimally set the price $\overline{v}$ and never trade. By a standard revenue equivalence argument,[33] the expected discounted payoff $V(c)$ of a seller of type $c \in [\underline{c}, \Gamma_\alpha^{-1}(\overline{v})]$ who always sets the price $P^*(c)$ is pinned down by

[32] The ultimate conditional expected revenue sounds similar to, but is distinctively different from the expected net present value of the revenue.

[33] See e.g. Myerson (1981).

the allocation rule and given by

$$V(c) = \int_c^{\Gamma_\alpha^{-1}(\overline{v})} (1 - F_\infty(P^*(y)))dy.$$

With this in hand, we can now describe the optimal transaction fees $\omega_t(\breve{p})$.

Proposition 2. *The optimal transaction fees that implement the optimal mechanism described in Lemma 1 are such that for all* $t = 0, 1, ..$

$$\omega_t(p) = \omega(p) := p - \frac{\int_p^{\overline{v}} \left[\Gamma_\alpha^{-1}(\tilde{\Phi}(v)) + \delta V(\Gamma_\alpha^{-1}(\tilde{\Phi}(v)))\right] f(v)dv}{1 - F(p)}. \quad (4)$$

The proof that this fee induces the seller of type c to set the price $P^*(c)$ in every period is surprisingly simple. By the one-period-deviation principle, we can confine attention to a deviation by the seller in the present period to some reserve price p and assume that the seller sets the price $P^*(c)$ in every period after that, whereby he gets $\delta V(c)$. The expected payoff from so doing given the fee $\omega(p)$ defined in (4) is $(R_{\omega_t}(p) - c)(1 - F_1(p))$ in the period of deviation and $F_1(p)\delta V(c)$ afterwards, which can be rearranged to

$$(p - \omega(p))[F_{(2)}(p) - F_{(1)}(p)] + \int_p^{\overline{v}} [y - \omega(y) f_{(2)}(y)dy + F_{(1)}(p)[c + \delta V(c)].$$

The first-order condition for a maximum is

$$0 = f_{(1)}(p)\left[-\Gamma_\alpha^{-1}(\tilde{\Phi}(p)) - \delta V(\Gamma_\alpha^{-1}(\tilde{\Phi}(p))) + c + \delta V(c)\right],$$

which follows after cancelling terms (in particular, using the fact that $F_{(2)}(p) - F_{(1)}(p) = f_{(1)}(p)(1 - F(p))/f(p))$. The first-order condition is satisfied at $p = P^*(c)$ and because the term in brackets decreases in p, it follows that the objective function is quasi-concave, implying that the first-order condition is sufficient for a maximum. Importantly, Proposition 2 implies that fee-setting with the fee given in (4) is optimal in the domain of all incentive compatible, individually rational mechanisms.

As an illustration, consider the static setup by setting $\delta = 0$, which has been studied extensively. With $\pi_1 = 1$, we have one buyer with certainty, and if we set $\alpha = 1$, our fee-setting mechanism implements the broker-optimal mechanism derived by Myerson and Satterthwaite (1983). For the example in Myerson and Satterthwaite (1983) (F and G

uniform on $[0,1]$), the broker-optimal fee is 50%.[34] It is easy to check that the broker's expected revenue is 1/24, which is the same as for the direct mechanism derived in Myerson and Satterthwaite (1983). For any fixed number of buyers $\overline{B}$ ($\pi_{\overline{B}} = 1$) and any distribution F satisfying regularity, our fee-setting mechanism specializes to the optimal auction with reserve $P^*(c) = \Phi^{-1}(c)$ of Myerson (1981) if all the weight is on the seller's welfare ($\alpha = 0$).

2.2 Thin Markets and Extreme Value Theory

In principle, the optimal fee schedule can be a complicated non-linear function. Empirically, however, fee-setting with simple linear fees is often used. Linear fees are particularly prevalent in thin markets such as real-estate markets and high-skill labor markets, where typically only a small percentage of potential sellers is active in the market. As an example, less than 5% of home owners offer their property for sale at a given point of time. Amazon's fees for third-party sellers of most types of goods (including books, consumer electronics, and personal computers) are another case in point. We now show how additional transaction costs whose presence induces only the most motivated traders to participate imply that the optimal fees will be asymptotically linear. We do so by applying results from extreme value theory to markets with fee-setting. There are various possible and mutually non-exclusive sources of such costs. The cost of physical relocation – of moving or shipping – is one that is due to exogenous costs. In dynamic models, such transaction costs may also arise endogenously from the agents' opportunity costs of future trade, which in any given period makes agents less inclined to trade. To fix ideas, we will focus on the case of exogenous transaction costs, and we will assume that after the realization of their types, and knowing the transaction costs, agents can decide whether they want to participate in the market. Later on, we will discuss in more detail microfoundations for such transaction costs.

[34] The uniform is a special case of a mirrored Generalized Pareto distribution $G(c) = c^\sigma$ for $c \in [0,1]$ with $\sigma > 0$, which yields as the broker-optimal fee of $\omega(p) = p/(1+\sigma)$ for the static setup for any F and $\tilde{\pi}$.

Convergence to Linear Fees as Transaction Costs Increase To capture the notion that only a small fraction of potential traders are active, we introduce increasing transaction costs. For simplicity, we normalize the supports of the primitive distributions F_0 and G_0 from which buyers and sellers draw their primitive types v_0 and c_0 to $[0,1]$. We will study a sequence of economies characterized by transaction costs $\mathbf{K}_j := (K_j^S, \hat{K}_j^S, K_j^B, \hat{K}_j^B)$ and focus on the limit of this sequence, indexed by $j \geq 0$, as the cost becomes large with $\mathbf{K}_0 = (0,1,0,1)$. The distribution of effective costs $c = K_j^S + \hat{K}_j^S c_0$ is $G_j(c) = G_0((c - K_j^S)/\hat{K}_j^S)$ and the distribution of effective valuation $v = K_j^B + \hat{K}_j^B v_0$ is $F_j((v - K_j^B)/\hat{K}_j^B)$. Our previous analysis directly applies by replacing F by F_j and G by G_j.

There are many different ways in which transaction costs may reduce that fraction of active traders. One example are moving costs for a buyer of a property, which are additive ($K_j^B > 0$, $\hat{K}_j^B = 0$). Another example is an option value x for the buyer of real estate, which may be due to the possibility of buying another property or the possibility of renting, such that the buyer's willingness to pay is $v = \lambda_j x + (1 - \lambda_j) v_0$, where λ_j is a weight put on the option value that will be discussed later. For this example $K_j^B = \lambda_j x$ and $\hat{K}_j^B = 1 - \lambda_j$. The same applies for the seller's transaction costs K_j^S and $\hat{K}_j^S$. There are many different combinations of changes of K_j^B, $\hat{K}_j^B$, K_j^S, and $\hat{K}_j^S$ that lead to a decrease of the fraction of active traders. However, it is not necessary to go through all combinations. Instead, one can greatly simplify the analysis by introducing two variables u_j^B and u_j^S whose decrease implies a decrease of the fraction of active traders. Therefore, we defer providing microfoundations of different changes of $\mathbf{K}_j$ to Section 2.3 and turn to the variables u_j^B and u_j^S in the following.

Denote the implied supports with $[\underline{c}_j, \overline{c}_j]$ and $[\underline{v}_j, \overline{v}_j]$, respectively. In the following, it will be useful to think of the *relevant range* $[\underline{c}_j, \overline{v}_j]$ in which the two supports overlap. It is also useful to define the ratio of the length of the relevant range to the length of the seller's support $u_j^S := (\overline{v}_j - \underline{c}_j)/(\overline{c}_j - \underline{c}_j)$. Analogously, define $u_j^B := (\overline{v}_j - \underline{c}_j)/(\overline{v}_j - \underline{v}_j)$ for the buyer. Since there is a one-to-one mapping between the set of parameters $(\underline{c}_j, u_j^S, \overline{v}_j, u_j^B)$ and $\boldsymbol{K}_j$, we can write the following analysis in terms of $(\underline{c}_j, u_j^S, \overline{v}_j, u_j^B)$.

Since sellers with $c > \overline{v}_j$ trade with probability 0, a seller participates if and only if

$c \leq \overline{v}_j$, which is equivalent to $c_0 \leq u_j^S$. Therefore, the mass of active sellers is $G_0(u_j^S)$. Analogously, the mass of active buyers is $1 - F_0(1 - u_j^B)$.

The analysis simplifies by normalizing $\tilde{c} := (c_0 - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$, $\tilde{v} := (v_0 - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$, and $\tilde{p} := (p - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$. The distributions of the normalized effective cost $\tilde{c}$ and the normalized effective valuation $\tilde{v}$, truncated to $[\underline{c}_j, \overline{v}_j]$ and denoted, respectively, $\tilde{G}_j$ and $\tilde{F}_j$, are then given as

$$\tilde{G}_j(\tilde{c}) := \frac{G_0(u_j^S \tilde{c})}{G_0(u_j^S)} \quad \text{and} \quad \tilde{F}_j(\tilde{v}) := 1 - \frac{1 - F_0(1 - u_j^B(1 - \tilde{v}))}{1 - F_0(1 - u_j^B)},$$

with the normalized fee defined as $\tilde{\omega}_j(\tilde{p}) = \omega(p)/(\overline{v}_j - \underline{c}_j)$.[35]

The following Proposition relies on Extreme Value Theory, which states that the upper tail of any distribution converges to a Generalized Pareto distribution as one moves the truncation point closer to the upper bound of the support, as long as the distribution satisfies some weak regularity assumptions (see Appendix F for more details on extreme value theory and also for a version of the theory with an infinite upper bound of the support). These regularity assumptions can be shown to be satisfied in our setup. Further, analogous mirror image results hold with regards to the lower bound of the support. The proposition also shows the relation between Extreme Value Theory and linear fees.

Proposition 3. *Let the shifting constants $\underline{c}_j$, $\overline{v}_j$ be arbitrary sequences satisfying $\underline{c}_j < \overline{v}_j$ for all j. Let the ratios of the relevant ranges u_j^S and u_j^B be sequences that go to 0 as j goes to infinity. Then, as $j \to \infty$,*

(i) the buyers' and the seller's normalized distributions converge to Generalized Pareto and mirrored Generalized Pareto distributions, respectively: $\lim_{j\to\infty} \tilde{F}_j(\tilde{v}) = \tilde{F}^*(\tilde{v}) := 1 - (1 - \tilde{v})^\beta$ *and* $\lim_{j\to\infty} \tilde{G}_j(\tilde{c}) = \tilde{G}^*(\tilde{c}) := \tilde{c}^\sigma$.

(ii) the normalized fee $\tilde{\omega}_j(\tilde{p})$ converges to $\alpha\tilde{p}/(\alpha + \sigma)$, that is:

$$\lim_{j\to\infty} \tilde{\omega}_j(\tilde{p}) = \frac{\alpha}{\alpha + \sigma}\tilde{p}. \tag{5}$$

[35] Despite notational similarities, the distribution $\tilde{F}_j$ has no relation to the dynamic virtual valuation $\tilde{\Phi}$ introduced after Proposition 1 above. We will not use virtual valuations associated with $\tilde{F}_j$.

We first provide an intuition for part (i) of the Proposition. Convergence of the distribution is immediate when the primitive distributions G_0 and F_0 are (mirrored) Generalized Pareto distributions on $[0, 1]$, that is if $G_0(c_0) = c_0^\sigma$ for $\sigma > 0$ and $F_0(v_0) = 1 - (1 - v_0)^\beta$ with $\beta > 0$, for this implies $\tilde{G}_j(\tilde{c}) = G_0(u_j^S \tilde{c})/G_0(u_j^S) = \tilde{c}^\sigma$ and $1 - \tilde{F}_j(\tilde{v}) = (1 - F_0(1 - u_j^B(1 - \tilde{v}))/(1 - F_0(1 - u_j^B)) = (1 - \tilde{v})^\beta$ for all j. In other words, $\tilde{G}_j$ and $\tilde{F}_j$ do not change with j. This is, of course, the well-known property of Pareto distributions that they are invariant to truncation.

Next, let us discuss the correct interpretation of the asymptotic results if we start away from the limiting distribution. This is important, since asymptotic results in statistics are among the most often misunderstood concepts. A common misunderstanding is that asymptotic results are only applicable in one of two cases: either if one assumes very particular functional forms for distributions that are close to the limiting distribution or if one is exactly in the limit. While for often used asymptotic results such as the central limit theorem such misunderstandings (mostly based on a false dichotomy[36]) are seldom, for other asymptotic results, such as extreme value theory, they are quite common.[37]

We take two strategies to clarify such potential misunderstandings. First, we provide numerical results that illustrate that even when starting with a distribution quite different from the limiting distribution and when not going too close to the limit, the results of Extreme Value Theory are already a good approximation. Second, we provide an empirical analysis, showing that for the empirically estimated distributions, Extreme Value Theory is a quite good approximation (according to a metric we will be more precise on in the empirical section).

Consider the following numerical example – in which G_0 is quite far away from a

[36] For the central limit theorem, such a misinterpretation would mean the following. The central limit theorem is *supposedly* only applicable in one of two cases: (i) the distribution of a variable has a peculiar functional form that is very close to a normal distribution to start with or (ii) one has to be (almost) exactly in the limit, which means taking the average of an infinite number of random draws. One would then *falsely* believe that either one has to make an overly restrictive assumption on functional forms (case (i)) or that the variance of the average is (almost exactly) zero (since we are taking the average of an infinite number of random draws, case (ii)). However, the distinction of cases (i) and (ii) is a false dichotomy: the applicability of the central limit theorem is due to the middle ground between cases (i) and (ii).

[37] For extreme value theory, the misinterpretation is that asymptotic results apply in only one of two cases (i) the distribution is close to Generalized Pareto to begin with or (ii) the mass in the tail of the truncated distribution is (close to) zero. Again, a false dichotomy.

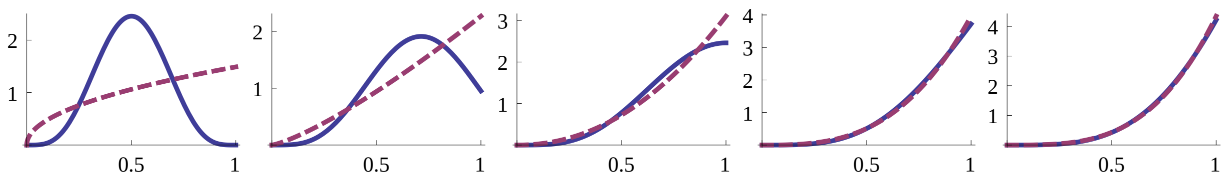


Figure 1: Density of truncated, rescaled distribution $G_u(c) = G_0(\underline{c}+u(c-\underline{c}))/G_0(\underline{c}+u(\bar{c}-\underline{c}))$ for $u \in \{1, 0.7, 0.5, 0.3, 0.2\}$ for a Beta distribution with support $[0, 1]$ and density $g_0(c) \propto c^4(1-c)^4$ (solid line) compared to an approximating mirrored Generalized Pareto density with support $[0, 1]$ (dashed). Masses in the relevant range are (a) $G(1) = 1$, (b) $G_0(0.7) = 0.9$, (c) $G_0(0.5) = 0.5$, (d) $G_0(0.3) = 0.1$, (e) $G_0(0.2) = 0.02$. As the mass decreases, the distribution converges to the approximating Pareto distribution and the approximating Pareto distribution converges to the limiting Pareto distribution.

mirrored Generalized Pareto. Figure 1 shows the density $g_0(c_0) \propto c_0^4(1 - c_0)^4$ of a Beta-distribution whose support is $[0, 1]$. The figure shows the distribution conditional on $c_0 \in [0, u]$ for $u \in \{1, 0.7, 0.5, 0.3, 0.2\}$. Moving the truncation point u downwards brings the density of the conditional distribution closer to the density of a mirrored Generalized Pareto distribution. Of course, the statement of Extreme Value Theory holds in the limit as is always the case for asymptotic results. However, as usual in statistics, one should interpret asymptotically founded results as providing good approximations away from the limit. For example, panel (d) in Figure 1 depicts the case when still ten percent of all seller types are active. The distribution is already very well approximated by a mirrored Generalized Pareto distribution. This is even more so in panel (e), when two percent of sellers are active. In this case, the overlap is almost perfect. One may wonder how close we are at the limit in practice. This question can only be answered empirically, which we will do later on.

The intuition for part (ii) of Proposition 3 is most easily gleaned by specializing to a static setup (i.e. $\delta = 0$) and assuming that G_0 is a mirrored Generalized Pareto distribution, that is $G_0(c_0) = c_0^\sigma$ for $c_0 \in [0, 1]$. This implies that the virtual cost function is linear, that is, $\Gamma_{\alpha,0}(c_0) := c_0 + \alpha G_0(c_0)/g_0(c_0) = c_0(1 + \alpha/\sigma)$.[38] The optimal fee can

[38] By adjusting the support appropriately, for any element j in the sequence a similar analysis applies and delivers completely analogous results if G_0 is a mirrored Generalized Pareto distribution because of the truncation invariance of the virtual cost and of these distributions.

thus be written as

$$\omega(p) = p - E_{v_0 \sim F_0}[\Gamma_{\alpha,0}^{-1}(\Phi_0(v_0))|v_0 \geq p] = p - \Gamma_{\alpha,0}^{-1}(E_{v_0 \sim F_0}[\Phi_0(v_0)|v_0 \geq p]). \tag{6}$$

Because it is linear, one can pull $\Gamma_{\alpha,0}^{-1}$ outside the expectation, and because $E_{v_0 \sim F_0}[\Phi_0(v_0)|v_0 \geq p] = p$, one obtains $\omega(p) = p - \Gamma_{\alpha,0}^{-1}(p) = p\alpha/(\alpha+\sigma)$.[39]

The reasoning in the above paragraph and part (i) of Proposition 3 do not yet constitute a proof of part (ii) of the Proposition, since one still has to establish that the transformation from distributions to fees is continuous. The proof is quite lengthy, which may not come as a surprise given the complexity of expression is (4) and the continuation value $V(\cdot)$. We relegate the proof to Appendix A and provide an illustration here, based on the same numerical example used for Figure 1. Figure 2 shows that the optimal fee moves closer to a linear fee as the transaction costs increase. The mass of traders $G_0(u)$ and $1 - F_0(1-u)$ does not have to be very close to 0 for the optimal fees to be close to linear. In Figure 2 (d) and (e) the optimal fee is already well approximated by a linear fee with the masses of traders being ten percent and two percent, respectively, and the length of the support u is 30 percent and 20 percent of the length of the original support, respectively. This numerical example should hopefully clarify a misunderstanding of asymptotic results not only for convergence of the distributions, but also for the convergence of the fee.

To analyze and interpret the true – that is, denormalized – fee $\omega(p)$ that is approximately optimal away from but sufficiently close to the limit, it is also insightful and useful to consider a j such that u_j^S and u_j^B are "sufficiently small" and to work with the denormalized limiting Pareto distributions

$$G^*(c) := \left(\frac{c - \underline{c}}{\overline{c} - \underline{c}}\right)^\sigma \quad \text{and} \quad 1 - F^*(v) := \left(\frac{\overline{v} - v}{\overline{v} - \underline{v}}\right)^\beta,$$

where $\underline{c} = \underline{v} = \underline{c}_j$ and $\overline{v} = \overline{c} = \overline{v}_j$ for notational simplicity. Observe that $G^*(c)$ has the linear virtual type function $\Gamma_\alpha^*(c) = c(1+\alpha/\sigma) - \alpha\underline{c}/\sigma$, which implies that the fee is

$$\omega(p) = p - \Gamma_\alpha^{*-1}(p), \tag{7}$$

[39]Sufficiency of Generalized mirror Pareto distributions for the optimality of linear fees follows from the argument in the text. Necessity was shown by Loertscher and Niedermayer (2007), a working paper superseded by the present paper. Building on this work, Niazadeh, Yuan, and Kleinberg (2014) extend the analysis of take-it-or-leave-it offers to linear fees that are close to optimal.

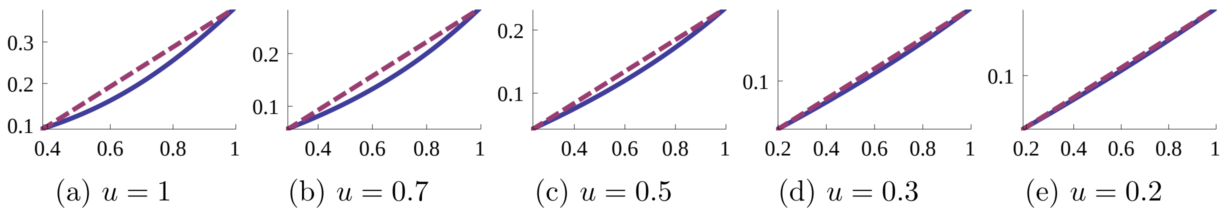


Figure 2: Optimal fee $\omega(\cdot)$ for truncated, rescaled distributions $F_u(v) = 1 - [1 - F_0(\overline{v} - u(\overline{v} - v))]/[1 - F_0(\overline{v} - u(\overline{v} - \underline{v}))]$, and $G_u(c) = G_0(\underline{c} + u(c - \underline{c}))/G_0(\underline{c} + u(\overline{c} - \underline{c}))$ for the same setup as in Figure 1; that is, for $u \in \{1, 0.7, 0.5, 0.3, 0.2\}$ for Beta distributions with support $[0, 1]$ and density $f(x) = g(x) \propto x^4(1 - x)^4$ (solid line) compared to an approximating linear fee (dashed).

which is linear in p.

The expression in (7) offers a neat interpretation and simple comparative statics: A "Ramsey monopsonist" with valuation v who faces the supply $G^*(\hat{p})$ would set the price $\Gamma_\alpha^{*-1}(v)$.[40] A more elastic supply means a higher monopsony price and hence a lower fee. As an illustration, $\tilde{G}(\tilde{c}) = \tilde{c}^\sigma$ can be seen as isoelastic supply with elasticity σ. As the elasticity σ increases, the fee $\tilde{\omega}(\tilde{p}) = \alpha\tilde{p}/(\sigma + \alpha)$ decreases. Surprisingly, and interestingly, the fee is independent of distribution of valuations, which is a point we will return to below.

There is another, potential misunderstanding worth clarifying. One may be tempted to think that the linearity of the optimal fees in the limit stems from the fact that the range of the relevant, overlapping interval $[\underline{c}, \overline{v}]$ goes to zero with the standard reasoning that functions are close to linear over short intervals. However, this is not so. The conditional distribution $\tilde{G}_j(\tilde{c})$ does, in general, not converge to a linear function (see Figure 1) nor does the optimal price function $\tilde{\Phi}^{-1}(\Gamma_\alpha(c))$ when $\delta > 0$. As shown in (7), the linearity of the optimal fee derives from the linearity of $\Gamma_\alpha(c)$, which in turn stems from the convergence of the distribution to (mirrored) Generalized Pareto distributions.

[40] A Ramsey monopsonist cares about his valuation v and a weighted average of the price p he pays and the seller's expected utility $E[c|c \leq p]$ and therefore maximizes $(v - (\alpha p + (1 - \alpha)E[c|c \leq p]))G^*(p)$, which implies the first-order condition $(v - \Gamma_\alpha^*(p))g^*(p) = 0$.

2.3 Microfoundations for the Transaction Costs

We now return to the microfoundation of transaction costs we alluded to previously. There are various ways transaction costs may occur. Some types of transaction costs are simply exogenously given, such as transportation and moving costs. For additive transportation and moving costs K_j^B and K_j^S, the asymptotic results hold if $K_j^B + K_j^S \to \overline{v}_0 - \underline{c}_0$, that is, if the total transportation costs are close to the maximum possible gains from trade.

Other types of transaction costs can be viewed as the endogenous outcome of a larger model, as we will describe in the following. No matter what the exact source of transaction costs, such transaction costs provide an explanation for why at any given point of time only a small percentage of real-estate properties are on the market or why Amazon considers the "tail of the distribution" (i.e. goods in thin markets, that are seldom traded) as one of its main sources of revenue.

Competing Direct Market Consider a competing direct market in the sense that the good can be traded at some price p, but this market is inefficient due to frictions: the probability of finding a trading partner is less than one and traders may have to spend potentially significant amount of time finding a trading partner. The intermediary's service is to enable trade with less frictions. So traders face the trade-off of either going to the intermediary and paying fees or going to the direct market and bare the costs of frictions. Examples would be buyers and sellers having the option to trade through other channels than eBay or Amazon. Another example is the real estate market, where buyers and sellers have the option to rent rather than sell/buy, in which case p should be viewed as the net present value of rents.

Formally, consider a competing market that opens at regular intervals of length μ, with trade occurring at a price p. In the spirit of Satterthwaite and Shneyerov (2008), the rematching frequency μ can be viewed as a measure of efficiency of this market. $\mu \to \infty$ should be viewed as infinitely large frictions in the direct market, which is equivalent to the direct market not existing as an outside option. $\mu \to 0$ should be viewed as frictions in the direct market vanishing. Define a buyer's probability of trade at an instance when the competing market opens as $\tilde{\beta} = e^{-\bar{\beta}\mu}$. Analogously, define a seller's probability of trade $\tilde{\gamma} = e^{-\bar{\gamma}\mu}$. Let the discount factor between two reopenings of the competing market (or, equivalently, the probability of not dropping out), be $\tilde{\delta} = e^{-\bar{\delta}\mu}$. The option value of trading in the competing market for a buyer with primitive valuation v_0 is $\sum_{t=0}^{\infty}(v_0 - p)\tilde{\beta}(1-\tilde{\beta})^t\tilde{\delta}^t = (v_0 - p)\beta$, where $\beta = \tilde{\beta}/[1-(1-\tilde{\beta}\tilde{\delta})]$ is the ultimate probability of trade in the competing market. Taking into account the option value of trading through the competing market, a buyer's effective valuation is $v = v_0 - \beta(v_0 - p) = \beta p + (1-\beta)v_0$, so that $K^B = \beta p$ and $\hat{K}^B = 1-\beta$. The upper bound of the buyer's effective valuation $\overline{v} = \beta p + (1-\beta)\overline{v}_0$ converges to p as μ converges to 0, that is, as the competing market becomes increasingly efficient. By a similar logic, the seller's effective cost is $c = c_0 + \gamma(p - c_0) = \gamma p + (1-\gamma)c_0$, where $\gamma = \tilde{\gamma}/[1-(1-\tilde{\gamma})\tilde{\delta}]$ is the seller's ultimate discounted probability of trade in the competing market. The lower bound of the seller's effective cost $\underline{c} = \gamma p + (1-\gamma)\underline{c}_0$ goes to p as $\mu \to 0$. Putting the pieces together, as the competing market becomes more efficient ($\mu \to 0$), the overlap of the seller's and the buyers' support $[\underline{c}, \overline{v}]$ shrinks towards p, which is what we need for our asymptotic results. Therefore, as frictions in the competing bilateral exchange decrease, an intermediary (whose service is to offer trade with less frictions) is forced to offer fees that are closer to linearity.

Dynamic Random Matching Model In most of our analysis, we are assuming that the buyer's valuation v and its distribution F are exogenously given. However, one can view our model as being embedded in a larger model, in which v and F are endogenous. Consider a dynamic random matching model in the spirit of Satterthwaite and Shneyerov (2008) (extended to include intermediaries): mass 1 of seller-intermediary pairs and mass 1 of buyers enter a market in every period. Buyers randomly choose a seller-intermediary pair to visit, which results in a Poisson distribution of the number of buyers a seller meets in every period. Sellers run an English auction with a reserve. If at least one bid is above the reserve, the seller-intermediary pair and the trading buyer leave the market. All participants who do not trade stay in the market. A steady-state equilibrium is one in which the mass and distribution of entering buyers equals the mass and distribution of exiting buyers, the same holding for sellers. The buyer's effective valuation is now endogenously given by $v = v_0 - \delta V_B(v_0)$, where $V_B(v_0)$ is the option value of future trade. One has to be additionally careful about the fact that buyers that trade with a lower probability stay longer in the market and are hence overrepresented in the pool of buyers in the market compared to entering buyers. While this endogeneity renders the model analytically intractable, in a previous version of this paper we solved this model numerically (with somewhat different assumptions)[41], see Loertscher and Niedermayer (2012). We found that the linear fee approaches the optimal fee as the market becomes more efficient (which is modeled as δ increasing). The fundamental driving force is that an increase of δ reduces the upper bound of the buyer's effective valuation $\overline{v} = \overline{v}_0 - \delta V_B(\overline{v}_0)$ and hence reduces the overlap of supports $[\underline{c}, \overline{v}]$, which in turn makes linear fees more optimal. It turns out that the effect identified in the previous version of our paper is more general than the opportunity cost of future trade in dynamic random matching models: any transaction cost reducing gains from trade for the current transaction makes optimal fees "more linear".

2.4 Comparative Statics

Our asymptotic results are interesting on their own, since they provide an explanation for the prevalent use of linear fees and also additional empirically testable implications, which we will discuss in detail in the empirical section. But our asymptotic results also have a number of additional benefits. First, we discuss the benefits for comparative statics, which become much simpler in the limit. Using the results in the limit as a starting point, we can then derive comparative statics results away from the limit.

Comparative statics are most easily performed, and intuition for these developed, if one assumes a static setup with one buyer, that is, $\pi_1 = 1$ and $\delta = 0$. Replacing the primitive distributions and (virtual) types in (6) by F and G and v and c (and Φ and Γ_α^{-1}), the optimal fee with $\delta = 0$ is

$$\omega(p) = p - E[\Gamma_\alpha^{-1}(\Phi(v))|v \geq p].$$

[41] We assumed that seller-intermediary pair write myopic contracts.

Interpreting $G(p)$ and $1 - F(p)$ as quantities supplied and demanded at price p (see Bulow and Roberts, 1989), one can define the price elasticity of demand at v as $\eta_d(v) := vf(v)/(1 - F(v))$ and the price elasticity of supply at c as $\eta_s(c) := cg(c)/G(c)$.

We first start with the simpler comparative statics: for η_s and α results are as expected: a global increase in $\eta_s(c)$ leads to lower fees and an increase in α leads to higher fees because $\Gamma_\alpha(c) = c(1 + \alpha/\eta_s(c))$ increases in α and decreases in $\eta_s(c)$.

The effect of the elasticity of demand η_d (and equivalently of $\Phi(v) = v(1-1/\eta_d(v)))$ is clearly more complicated. It is useful to first consider the limit, in which G^* is Generalized Pareto, so that both Γ^*_α and the optimal fee ω are linear. Equation (7) shows that in the limit, the elasticity of demand η_d does not play any role for the fee whatsoever! This suggests that the way Γ differs from a linear function determines how η_d affects the optimal fee. As shown below, this is indeed the case and there is also an intuitive economic interpretation of the effects of the elasticity of demand.

Proposition 4. *Using the Taylor expansion of $\Gamma_\alpha^{-1}(x)$ around $\overline{v}$, the net price received by the seller is*

$$\begin{aligned} p - \omega(p) = \; & \overbrace{\Gamma_\alpha^{-1}(p)}^{\textit{Ramsey monopsony price}} + \overbrace{\frac{[\Gamma_\alpha^{-1}(\overline{v})]''}{2} Var_{v\sim F}[\Phi(v) - \overline{v} | v \geq p]}^{\textit{second-order price endogeneity effect}} \\ & + \underbrace{\sum_{n=3}^{\infty} \frac{[\Gamma_\alpha(\overline{v})]^{(n)}}{n!} \{E_{v\sim F}[(\Phi(v) - \overline{v})^n | v \geq p] - E_{v\sim F}[\Phi(v) - \overline{v} | v \geq p]^n\}}_{\textit{higher-order price endogeneity effects}}, \end{aligned} \tag{8}$$

where $[\Gamma_\alpha^{-1}(v)]^{(n)}$ denotes the n-th derivative of $\Gamma_\alpha^{-1}(v)$.

Recall that a Ramsey monopsonist with value x would set the price $\Gamma_\alpha^{-1}(x)$. Naturally, the intermediary's valuation for the good is p, so that absent any other effects the seller's net price should be $\Gamma_\alpha^{-1}(p)$, which is exactly what the seller receives when Γ_α is linear. However, for Γ_α nonlinear the price p is determined endogenously, which requires the optimal net price the seller receives to be adjusted.

According to Proposition 4, η_d has no first-order effect on the optimal fee. Indeed, as seen above when G is a mirrored Generalized Pareto distribution (which is equivalent to a linear Γ_α), $\omega(p)$ is independent of F. However, the second- and higher-order effects can go either way. To see this, assume $\alpha = 1$ and that $\Gamma_1^{-1}(x)$ is quadratic with a curvature $[\Gamma_1^{-1}(x)]'' = \overline{\gamma}_2$.[42] A quadratic form shuts down the higher-order effects and allows us to focus on the second-order price endogeneity effect. The fee is $\omega(p) = p - \Gamma^{-1}(p) - (\overline{\gamma}_2/2)\mathrm{Var}[\Phi(v) - \overline{v}|v \geq p]$. For Γ_1^{-1} concave ($\overline{\gamma}_2 < 0$), an overall increase of the elasticity of demand can be shown to lead to an overall increase of the fee.[43]

These results are surprising and counterintuitive at first as one would expect more elastic demand to lead to lower fees. However, the intuition is that a more elastic demand causes the seller to lower the price excessively from the intermediary's point of view. To compensate for this, the optimal fee increases. If Γ_1^{-1} is convex, the converse occurs.

Important contributions by Bulow and Pfleiderer (1983), Aguirre, Cowan, and Vickers (2010), Bulow and Klemperer (2012), and Weyl and Fabinger (2013) have identified a number of properties of the demand function such as its curvature, the curvature of the inverse demand function, the pass-through rate, and the markup- or quantity-weighted average pass-through, which prove useful in a variety of contexts in industrial organization. Naturally, one may then wonder whether the counterintuitive result that optimal fees are sometimes higher for a higher elasticity of demand may be explained by alternative properties of the demand function. The example with Γ_1^{-1} quadratic shows that the answer is no. Any change of any property of the demand function F that leads to higher fees for $\overline{\gamma}_2 < 0$ will lead to lower fees for $\overline{\gamma}_2 > 0$ and will have no effect on the fees for $\overline{\gamma}_2 = 0$.

Our results are also relevant for public finance in the context of indirect taxation.[44]

[42] It can be checked that a distribution G exists that generates a quadratic Γ_1^{-1} by inverting Γ^{-1} to get Γ and then solving the differential equation $\Gamma(c) = c + G(c)/g(c)$ with initial condition $G(\overline{c}) = 1$ for G. The closed-form solution for G is somewhat lengthy and hence not reported here.

[43] To see this, take distributions $\hat{F}$ and F with elasticities $\hat{\eta}_d(v)$ and $\eta_d(v)$ satisfying $\hat{\eta}_d(v) > \eta_d(v)$ for all $v < \overline{v}$. Because $\Phi(v) = v(1 - 1/\eta_d(v))$, this implies $\hat{\Phi}(v) > \Phi(v)$, which in turn implies $(\hat{\Phi}(v) - \overline{v})^2 < (\Phi(v) - \overline{v})^2$ for all $v < \overline{v}$. Further, F hazard rate dominates $\hat{F}$ because $\hat{f}(v)/(1 - \hat{F}(v)) = \hat{\eta}_d(v)/v > \eta_d(v)/v = f(v)/(1 - F(v))$. This implies $E[\hat{v}|\hat{v} \geq p] \leq E[v|v \geq p]$ for all p. Together with $(\hat{\Phi}(v) - \overline{v})^2 < (\Phi(v) - \overline{v})^2$, this implies $E[(\hat{\Phi}(\hat{v}) - \overline{v})^2|\hat{v} \geq p] \leq E[(\Phi(v) - \overline{v})^2|v \geq p]$. Therefore, fees are higher with $\hat{F}$ than with F, since $\gamma_2 < 0$ and $\mathrm{Var}[\Phi(v) - \overline{v}|v \geq p] = E[(\Phi(v) - \overline{v})^2|v \geq p] - E[\Phi(v) - \overline{v}|v \geq p]^2 = E[(\Phi(v) - \overline{v})^2|v \geq p] - (p - \overline{v})^2$.

[44] Indirect taxes are often different for different product categories. As an example, the EU financial transaction tax levies 0.1% on share transactions and 0.01% on transactions involving derivatives. Value added taxes and sales taxes in many countries differ across products, with some goods being exempt from indirect taxes altogether.

In thick markets, it is well known that less elastically demanded goods should be taxed more heavily (see Salanié, 2003). For thin markets, our results imply that the elasticity of supply is key. Depending on the curvature of Γ_α^{-1}, one should either tax the good whose demand is more elastic or the one whose demand is less elastic. Our results are also of relevance for the practice of competition policy. When there is suspicion that fee-setting intermediaries, such as auction houses and platforms or real-estate brokers, collude, the standard approach would be to estimate the demand function and to then evaluate how closely prices are to the monopoly price implied by the estimates. Our analysis suggests that in thin markets with intermediaries, the first look should be at the elasticity of *supply* rather than the elasticity of demand. As a first-order approximation, the elasticity of demand does not matter for the fees of a profit maximizing intermediary. Instead, colluding intermediaries should be expected to leave a net price to the seller which corresponds to the price set by a monopsonist whose valuation is the gross price (again, as a first-order approximation).

Seller Entry and Invariance of Price Distribution in Thin Markets While the vast majority of real estate is sold via brokers, some properties – so called for-sale-by-owner (FSBO) properties – are also sold directly from sellers to buyers. As documented by Hendel, Nevo, and Ortalo-Magné (2009), the average prices charged on such FSBO platforms are not significantly different from the gross prices in intermediated trade although the sellers bear the broker's fees in the latter case but not on FSBO platforms. Similarly, Barwick, Pathak, and Wong (forthcoming) find that average sales prices do not vary with commission fees.[45,46] In light of the double marginalization that occurs in

[45] As noted before, there is very little variation in commission fees. However, there is a small fraction of brokers that offer lower fees.

[46] Rutherford, Springer, and Yavas (2005) and Levitt and Syverson (2008) make a different, but somewhat related comparison: they compare real estate brokers selling their own houses with real estate brokers selling other people's houses. While the results are difficult to directly compare, since real estate brokers are likely to be a different demographic group of sellers than the general population, these papers also find that the asking price is not lower for brokers selling their owner properties. It is actually somewhat higher. A way to reconcile the findings of these articles with those of Hendel, Nevo, and Ortalo-Magné (2009) is that real estate brokers are a different demographic group than average sellers and have higher opportunity costs of selling (i.e. their c is higher on average). Note that we will show later on that the distribution of opportunity costs G is non-parametrically identified in our setup, so that one could in principle test whether the distribution is different for different demographic groups.

intermediated trade, this finding is puzzling at first sight.[47]

Note that the question at hand is more complicated than standard double marginalization: we are not simply dealing with one upstream and one downstream monopolist setting markups which result in a single price, but rather with a distribution of heterogeneous sellers' costs and an implied distribution of heterogeneous prices. Thus, in general, one has to compare average prices for FSBO sellers and sellers selling through an intermediary. However, it turns out that with the limiting Generalized Pareto distribution, the analysis becomes tractable. Moreover, it provides a natural explanation for this otherwise puzzling empirical observation.

In the following we show that with a small adjustment of our model, these invariance results are an implication of our theory. For this purpose, assume now that there are two market segments or platforms – the standard intermediated market where sellers face a percentage fee $\omega(p) = bp$ with $b \in [0, 1]$ and a FSBO-market where sellers face no fees – and assume that buyers multi-home, so that demand is the same across the two platforms. Observe that the marginal seller type who is active in the sense of setting a reserve price that at least one buyer type would accept is given by $c_b^* := (1 - b)\overline{v}$ in the intermediated market and by $c_{FSBO}^* = \overline{v}$ on the FSBO platform. Let $G_{p,b}(p)$ denote the distribution of reserve prices set by sellers who are active in the intermediated market with percentage fee b and denote by $G_{FSBO,p}(p)$ the distribution of reserve prices set by active sellers on the FSBO platform.

Proposition 5. *Assume that the sellers' cost distribution is a mirrored Generalized Pareto distribution of the form $G(c) = (c/\overline{c})^\sigma$ for both the intermediated market and the FSBO platform and the buyers' multi-home. Then we have*

$$G_{p,b}(p) = G_{FSBO,p}(p).$$

Proposition 5 provides a parsimonious explanation for a puzzling empirical observation documented by Hendel, Nevo, and Ortalo-Magné (2009) and Barwick, Pathak,

[47] In intermediated trade, a seller of type c optimally sets the price $\tilde{\Phi}^{-1}(\Gamma_\alpha(c)) > \tilde{\Phi}^{-1}(c)$, where $\tilde{\Phi}^{-1}(c)$ is the price set by the same seller when selling directly.

and Wong (forthcoming).[48] Intuitively, mirrored Generalized Pareto distributions are invariant to truncation from the right and to re-scaling. Such re-scaling is exactly what happens with additional seller entry as linear fees decrease. While we have stated the above results for $\underline{c} = 0$ because this assumption is consistent with the use of percentage fees, which are observed empirically for example in real-estate, these results generalize to arbitrary values of $\underline{c}$ with the adjustment that the optimal fee will be linear with a non-zero fixed component when $\underline{c}$ is different from zero. The distribution $G(c) = (c/\overline{c})^\sigma$ is a simple way of modeling isoelastic supply and does not literally mean that there are sellers with costs close to zero.[49]

It can also be shown that the average price in the FSBO-market may be higher or lower than in the intermediated market, depending on whether Γ is convex or concave.

2.5 Extensions

The setup we study is amenable to a variety of interesting and natural extensions. The limiting Pareto distribution turns out to have interesting implications in these extensions. Due to space constraints, we will only sketch what we consider to be the most valuable ones.

Non-Stationarity Let us first briefly explain how our analysis can be extended to account for non-stationary environments at what is essentially a cost in notation. Assuming that the sequences of time varying discount factors δ_t, distributions F_t and random arrival processes π_B^t are known, one can proceed in analogy to the way we proceeded under the assumption of stationarity. Although the optimal transaction fee ω_t in period t will in general vary over time because of non-stationarity, a simple argument based on what we call "expectational fees" (which are defined and derived in Lemma 2 in the Appendix) shows that, in the limit as markets become thinner, the optimal (normalized) transaction

[48] As mentioned in footnote 46, a comparison with Rutherford, Springer, and Yavas (2005) and Levitt and Syverson (2008) is more difficult. But our results can be interpreted as the price increasing effect of fees being offset by the price decreasing effect of less entry by high cost sellers, so that the price differences described in Rutherford, Springer, and Yavas (2005) and Levitt and Syverson (2008) would be explained by different distributions of c in different demographic groups.

[49] As we will see in the empirical section, our estimates suggest that the elasticity of supply is close to isoelastic, yet less than 0.5% of sellers have a cost less than $0.2\overline{c}$.

fees will be linear and stationary in the limit, too. The limit results also hold in a setup in which the distribution of buyers' valuations changes stochastically over time. Appendix D contains more details. The stationarity of the optimal limiting fees is particularly remarkable because the optimal reserve price path chosen by the seller will in general be non-stationary.

Linear Fees, First-Price Auctions, and the Informed Principal Problem Given linear transaction fees ω, the payoff of the seller upon selling at some price $\breve{p}$ is linear in the transaction price. Consequently, linear fees correspond to the case where the seller is a risk-neutral agent with a linear Von-Neuman-Morgenstern utility function. As is well-known, with risk-neutral agents the revenue equivalence theorem applies.[50] This implies that, given linear fees, using a first-price auction in which the seller sets the reserve is equivalent to the second-price auction we have assumed thus far. Moreover, due to the results for the informed principal problem in linear environments with independent private values of Mylovanov and Tröger (2014), keeping fixed the linear fee the expected payoffs conditional on type would be unchanged if the seller could choose the trading mechanism after having learned his type (in any strongly neologism-proof perfect Bayesian equilibrium). Thus, with linear fees the broker could even delegate the choice of the mechanism to the seller. This has the important implication that the intermediary could raise his revenues in a rather decentralized way: he simply sets the linear fees and can leave the choice of the details of the bargaining protocol to the seller.

3 Empirical Analysis

We now turn to our empirical analysis of real-estate transactions. There are a number of reasons for analyzing real-estate data. Real estate is a quantitatively important industry, with property often representing the main asset individual households own. Given the importance of real estate there is a sizable empirical literature investigating this market with a large amount of high quality data, going back several decades. This makes the

[50] The only additional assumption with a random number of bidders is that bidders be symmetrically informed about the number of other bidders participating (see e.g. Krishna, 2002, chapter 3).

real estate brokerage market particularly suitable to gain insights on intermediation that should extend to intermediaries in other industries. Moreover, the empirical literature on real estate brokerage has uncovered many stylized facts that seem puzzling as detailed below.

3.1 Real Estate Market

Economists often and perhaps casually refer to the fee structure of real-estate brokers (i.e. how brokers' fees depend on the transaction price) as the incentive scheme that solves a principal-agent problem, with the fees incentivizing the broker to exert effort to achieve a higher price, or so the theory goes. This theory may well be referred to as the zombie theory of real-estate brokerage fees because the theory keeps rising from the dead no matter how often new evidence arrives that seemingly kills it.

Among the many pieces of evidence contradicting the principal-agent explanation the most damning is that when the brokerage fee is split between a seller's agent and a buyer's agent, both agents get 3%. This means that the buyer's agent is paid *more* if the outcome is *less* favorable to the buyer, that is, the price is higher. This is despite the fact that buyers' agents are contractually obliged to act in the buyer's fiduciary interests.[51] Therefore, a principal agent theory cannot explain why a buyer would write such a contract with his agent. Fees also cannot simply reflect the costs of brokerage for the following reason. The empirical literature has found a very puzzling stylized fact: relative fees have essentially been constant at 6% over time and across regions (e.g. Hsieh and Moretti, 2003).[52] This means that in a city in which real estate prices doubled absolute fees also doubled. It is hard to believe that brokerage costs doubled for the same type of property within a decade, especially since wages changed far less and the

[51]Since the 1990s, buyer agency has become common, that is, agents on the buyer's side are contractually obliged to act in the buyer's interest. It should be noted that before the 1990s, agents on the buyer's side were typically so called cooperating agents, who worked with the buyer, but were obliged to act in the seller's interest. Both buyer's agents and cooperating agents typically get a 3% fee. See Han and Strange (2014, p. 873ff) for an overview of the literature on buyer agency.

[52]Hsieh and Moretti (2003) analyze data from 282 metropolitan areas from 1980 to 1990 and find that while real estate prices varied a lot across cities and over time (in some cities, real estate prices more than doubled), relative fees remained around 6%. The invariance of brokerage fees has also been reported in Department of Justice (2007). For a summary of the literature, see Han and Strange (2014).

changes of wages were found not to be correlated with changes of house prices.[53] There is a host of additional evidence contradicting a principal-agent theory of brokerage fees, which one could try to explain away by repeatedly modifying the theory, which, however, has always lead to contradictions to other pieces of evidence.[54,55] Empirical researchers have recognized the difficulties of the principal-agent theory in explaining real estate brokerage fees. But since no comprehensive alternative theory has been provided yet, the best approach for empirical research has been to keep documenting the puzzles that appear when viewing real estate brokerage fees from a principal agent perspective.[56]

It is not the goal of this article to take another shot at the moving target of the principal-agent theory of real-estate brokerage fees. Instead, we present a parsimonious theory that naturally predicts the above empirical observations[57] or at least does not

[53]This is especially unconvincing since brokerage fees have remained roughly 6% also after sharp drops of real-estate prices, which implies a sharp drop of absolute fees.

[54]The theory is contradicted by the findings of Hendel, Nevo, and Ortalo-Magné (2009) that sellers selling directly through a for-sale-by-owner platform get on average the same price as the gross price sellers get who sell through a broker. The theory could be defended by the claim that this is due to the 6% fee contract not giving sufficient incentives to the broker to get a higher price. However, this leads to the question why brokers do not get more high powered incentives – a marginal fee close to 100% and possibly an inframarginal fee below 6%. The usual explanation in a principal agent setting is that incentives are insufficiently steep because the principal provides insurance to the agent. However, the larger brokerage firms have hundreds of transactions per year (even individual brokers have half a dozen transactions per year, see Hsieh and Moretti (2003)), whereas an individual seller has a sale every couple of years. This would suggest that the seller should be more rather than less risk averse than the broker and hence not provide insurance to him.

[55]The cost based explanation of brokerage fees is also inconsistent with the observation that the relative fee for a $100,000 studio and a $1,000,000 mansion is 6% in both cases, making it unlikely that the costs are $6,000 in one case and $60,000 in the other. Even if one were to believe that brokerage costs vary that much between different types of property, this would still not explain Hsieh and Moretti (2003)'s findings.

[56]We thank François Ortalo-Magné for pointing out to us some of the difficulties of the principal-agent explanations of real-estate brokerage fees and for providing anecdotes of real-estate brokers who consider their value-added in making transactions more convenient (reduce search frictions, legal uncertainty, etc.) rather than driving up the price. Brokers seem to view it as important to get both the seller and the *buyer* on board, a reputation of driving up the price is hence viewed bad for business, especially for buyers' brokers. See also the discussion in Levitt and Syverson (2008, p. 610) about the puzzles created by a principal-agent explanation of brokerage fees and Han and Strange (2014) for an overview of the findings of the empirical literature on real estate brokerage fees.

[57]The findings of Hendel, Nevo, and Ortalo-Magné (2009) mentioned in footnote 54 come out naturally from Proposition 5. The optimal marginal fee need not be 100% if the fee is not meant to incentivize the broker, but rather to extract rents from buyers and sellers. That the relative fee remained constant while real-estate prices fluctuated wildly, can be explained by the observation that the optimal percentage fee is independent of the buyer's distribution (see Proposition 3). Even the seller's distribution need not remain constant: for the percentage fee to remain constant it is sufficient that the Pareto tail index does not change. (While we have no clear-cut explanation for the invariance of the Pareto tail index,

contradict them[58]: the theory of real-estate brokerage fees as optimal pricing, combined with extreme value theory. An added benefit of this approach is that the theory can also explain the fee structure of other intermediaries, such as Amazon and eBay, and the structure of transaction taxes. By Occam's razor, a parsimonious general model for all types of intermediaries is preferable to a number of theories, each tailor-made to a particular type of intermediary.

3.2 Data

The data set we use is the one constructed by Genesove and Mayer (2001). In order to save space and given that Genesove and Mayer (2001) is one of the most cited empirical articles on real-estate with one of the highest quality data sets, we provide a summary of the description of the data here and refer the reader to Genesove and Mayer (2001) for more details, rather than repeating a detailed description of the data here.

The data set includes property listings from January 6, 1990 to December 28, 1997 and property delistings (due to sale or withdrawal) from May 10, 1990 to March 16, 1998 with a total of 5792 observations from the Multiple Listing Service. The data track individual properties in the condominium market in downtown Boston and contain the date of the entry and exit of a property, listing price, and, if applicable, sale price, and property characteristics.

---

invariant Pareto tail indices have been observed in a variety of context, such as for income and wealth distributions.)

[58]That the buyer's fee and the seller's fee are split in 3% plus 3% can be thought of as the buyer's and the seller's agent jointly choosing optimal pricing to avoid double marginalization. From this perspective, it is not a contradiction that the buyer's broker's fee increases with the price, since the fee does not serve to incentivize the broker, but to extract rents. It should be noted that there is an observation that needs some explanation, since it seemingly contradicts the empirical findings in Hendel, Nevo, and Ortalo-Magné (2009): Levitt and Syverson (2008) and Rutherford, Springer, and Yavas (2005) find that brokers selling their own houses get a higher price than brokers selling other people's houses. An explanation may be that brokers are a different demographic from average sellers, so that their opportunity cost of selling c is drawn from a different distribution than average sellers', which leads to different prices. The comparison in Hendel, Nevo, and Ortalo-Magné (2009), on the other hand, is between *non-brokers* selling through a for-sale-by-owner platform or through a broker, which could mean a similar distribution G for the two types of sales and hence (by Prop. 5) similar prices. This may sound like explaining observables (prices) with unobservables (the distribution G), but it is not: we will show in the following how to estimate the distribution G. While we do not have data on brokers vs non-brokers selling their houses, the following analysis could well be applied to a data set containing such a distinction of sellers to test whether G differs for different groups.

A challenge when working with real-estate data is the heterogeneity of properties. The data set of Genesove and Mayer (2001) is particularly appealing since it has three features which help to deal with heterogeneity. First, all properties considered are condominiums in downtown Boston, which by itself already means a certain level of homogeneity. Second, Genesove and Mayer (2001) construct a quality index from property characteristics using a standard hedonic pricing approach, that is, they run a regression of the form $\ln y_i = Z_i\beta + \epsilon_i$, where y is the transaction price of a property, Z_i a vector of characteristics of the property, and ϵ_i an error term. $\exp(Z_i\beta)$ serves as a quality index. Third, besides the standard approach of hedonic pricing, Genesove and Mayer (2001) also merge the Multiple Listing Services data set with data from the registrar's office. The registrar's office contains data on the previous transaction price of a property, when it was sold several years before. The previous transaction price – adjusted by inflation and the change of the real-estate price index – has the advantage that unobserved heterogeneity of the property plays a lesser role. They report an R^2 in the order of 0.85 across the board.

We construct quality adjusted prices by taking the residual of the price that cannot be explained by the quality index. In the main part of this article we use the previous transaction price based quality index. As a robustness check we have done the same analysis with the quality index based on hedonic pricing and obtained essentially the same results. We report robustness checks in the appendix.

Given that the novelty and the focus of our approach is in the cross-sectional variation of seller's opportunity costs of selling, we avoid excessive complexity of our structural model by using a stationary model for structural estimation. To make sure that the intertemporal variation in our data is small compared to the cross-sectional variation, we perform the structural estimation for each year separately. Further, we restrict attention to the period between April 1, 1993 to April 1, 1996, which exhibit only a modest change of real-estate prices. In Appendix B we show that stationarity is a reasonable – even if not perfect – approximation for the years chosen for the estimation. In particular, as Table 4 there shows, the inter-temporal variation in quality adjusted prices is less than 2% of the cross-sectional variation for the years considered. Excluding data for the years before 1993 and after 1995 has the additional advantage of avoiding truncation issues, which would occur for the first two and last two years in our data set.[59]

Of the 2455 observations between April 1, 1993 and April 1, 1996, we further exclude data with a quality-adjusted price larger than two and less than half as well as properties that were on the market for more than two years. This applies to 5.0% of the observations and results in 2333 remaining observations. A property that is offered at less than half of or more than twice the previous transaction price (adjusted by the movement of the real-estate price index) is likely to have undergone significant changes in quality or to constitute an error within the data set. Similarly, a property on the market for more than two years was probably not seriously marketed. Table 1 contains descriptive statistics of the included data. The average ratio of transaction price over list price is remarkably similar to the one found by Merlo and Ortalo-Magné (2004, Table 1), who use data from two regions in the United Kingdom.

Descriptive Statistics			
Variables	All Houses	Sold Houses	Unsold Houses
Observations	2333	1522	811
Listing Price	\$223,077 (\$177,736)	\$231973 (\$172,861)	\$206,383 (\$185,501)
Quality Adjusted Listing Price	1.139 (0.219)	1.125 (0.220)	1.165 (0.215)
$100\frac{\text{Transaction Price}}{\text{Listing Price}}$		92%	
Time on Market	148 (134) days	130 (123) days	182 (147) days

Table 1: Sample Means (Standard Deviations) of Descriptive Statistics

3.3 Data Generating Process

The data generating process is essentially given by the theoretical model of Section 2.

For the data generating process we make sure not to assume that the 6 percent fees charged by brokers are optimal, but take them as given.[60] We assume that there is a large number of intermediaries and that a large number of buyers and sellers enter in

[59]As a robustness check, we have also estimated the model for all the other years. We find little variation in parameter estimates.

[60]See e.g. Hsieh and Moretti (2003) and Department of Justice (2007) for evidence on the almost universal use 6 percent fees.

every period. Each seller draws his cost c independently from the distribution G and is randomly matched to an intermediary who offers a fee-setting contract with a fee of 6 percent. These contracts offer exclusive dealership to the intermediary, so a seller who accepts a contract is matched to the intermediary as long as he stays in the game. Buyers, in contrast, visit the house in one period and if they cannot buy in that period, leave and do not come back. A buyer's valuation already incorporates the option value of visiting another property and is drawn from a distribution F.[61]

At any intermediary, we model the arrival process of buyers, π_B, as a Poisson process with arrival rate ξ, which implies for the distribution of the highest valuation

$$F_{(1)}(p) = e^{-\xi(1-F(p))}. \tag{9}$$

Letting τ denote the length of a period (say, the number of days), the time on market t of a property has a geometric distribution with the distribution function $1-(\delta F_{(1)}(p))^{t/\tau}$ because of the assumption that the environment in a given year is stationary. The expected time on market $T(p)$ is

$$T(p) = \frac{\tau}{1-\delta F_{(1)}(p)}, \tag{10}$$

and the probability that a property sells in period t is $(1-F_{(1)}(p))(\delta F_{(1)}(p))^{t/\tau}$, leading to the ultimate probability of selling $1-F_{\infty}(p)$ defined in (2).

Observable and Unobservable Heterogeneity To account for heterogeneity, we use the multiplicative quality index constructed by Genesove and Mayer (2001) based on previous transaction prices. A multiplicative quality index is the standard approach to correct for heterogeneity, see e.g. Genesove and Mayer (1997); Levitt and Syverson (2008); Hendel, Nevo, and Ortalo-Magné (2009).

The recent auction econometrics literature highlights the importance of taking into account unobservable heterogeneity additionally to observable heterogeneity. For example, Krasnokutskaya (2011) shows that for Michigan highway procurement auctions, 66% of the variation in estimated costs is explained by unobserved heterogeneity and 34%

[61] In a larger model, one would think of F being endogenous and being composed of a fundamental valuation of the buyer and an additional option value of buying from other sellers in the future.

by private values. Therefore, despite the strengths of the quality index constructed by Genesove and Mayer (2001), we additionally take into account unobserved heterogeneity in quality besides the observable heterogeneity covered by the quality index.

Define the true quality index ϑ_i for each property i. This multiplicative quality index captures heterogeneity of properties, that is, a buyer's valuation is $\vartheta_i v$ and a seller's opportunity cost $\vartheta_i c$. All of our above theoretical analysis holds with the only modification that everything is multiplied by ϑ_i. The seller's reserve price is $\vartheta_i p$ and the transaction price $\vartheta_i \breve{p}$.

If the quality index ϑ_i were observed without noise, one could simply divide everything by ϑ_i to correct for heterogeneity. However, one should expect any observed quality index to be a noisy measure of the true quality index. For the quality measure used in the main text – the previous transaction price adjusted by the change of the real-estate price index – the noise might be due to the price of an individual property changing differently than the overall price index. For the quality index used as a robustness check in the appendix – the index constructed from a hedonic regression – there may be characteristics of a property observable by market participants but unobservable by the econometrician.[62]

We make the standard assumption of multiplicative noise ϵ_i^Q, such that the observed quality index is $\hat{\vartheta}_i = \vartheta_i / \epsilon_i^Q$. Hence, when dividing the observed transaction price $\hat{\breve{P}}_i = \vartheta_i \breve{p}_i$ by the observed quality index $\hat{\vartheta}_i$ we get the *observed quality-adjusted transaction price* $\breve{P}_i = \hat{\breve{P}}_i / \hat{\vartheta}_i = \breve{p}_i \epsilon_i^Q$, which is a noisy measure of the true quality adjusted transaction price $\breve{p}_i$. For the quality index based on the previous transaction price, P_i should be viewed as the change of price not explained by a change of the real-estate price index. For the hedonic pricing based quality index, P_i should be viewed as the residual when projecting property prices on observable attributes.[63]

[62] To be more specific, the main (previous transaction price based) quality index constructed by Genesove and Mayer (2001) is $\hat{\vartheta}_i = \frac{P_q^{\text{index}}}{P_{q'}^{\text{index}}} \hat{P}_{iq'}$, where q' and q denote two different quarters and P_q^{index} is the real-estate price index in quarter q constructed by Genesove and Mayer (2001). The alternative hedonic pricing based quality index is $\hat{\vartheta}_i = \exp(Z_i \beta)$. Z_i are the characteristics of the property and β are the coefficients from the hedonic regression $\ln \hat{\breve{P}}_i = Z_i \beta + \epsilon_i$ with the $\hat{\breve{P}}_i$ being the observed transaction price. See Genesove and Mayer (2001) for more details.

[63] This could be referred to as a "homogenized price" in analogy to the "homogenized bids" in Balat, Haile, Hong, and Shum (2016).

For the reserve price, there is an additional complication: we observe the listing price the seller announces, which is a noisy measure of the reserve price below which the seller is not willing to sell, even if one were to correct for the true quality index. Taking into account this "discount noise" ϵ_i^D, the listing price is $\hat{P}_i = \vartheta_i p_i \epsilon_i^D$. When dividing the listing price $\hat{P}_i$ by the observed quality index $\hat{\vartheta}_i$, we get the *observed quality-adjusted listing price* $P_i = p_i \epsilon_i^D \epsilon_i^Q$. Our analysis will be simplified by using the "reserve noise" $\epsilon_i^P := \epsilon_i^D \epsilon_i^Q$ instead of the quality noise ϵ_i^Q in the following. This simplification is no restriction, since knowing the distribution of ϵ_i^P and ϵ_i^D, the distribution of ϵ_i^Q is only one deconvolution away.

We also assume that the true time on market t_i is observed with an error, denoted ϵ_i^T and that the observed time on the market of object i, denoted T_i, satisfies $T_i = t_i + \epsilon_i^T$. The error ϵ_i^T may arise because a broker starts to show the property some time after it has been listed or because a property is delayed in being delisted after the buyer and the seller agreed on a deal. In the data set we use, most properties are listed and delisted on a Sunday. Thus, we essentially have weekly data and delay happens at least until the end of the week. A typical observation is then given by $\boldsymbol{X}_i = (T_i, P_i, S_i, \breve{P}_i)$, where $S_i = 1$ if the object has been sold and $S_i = 0$ denotes that the object has never been sold.

The Likelihood Function It is useful to first write down the likelihood function ignoring "noise" $(\epsilon_i^P, \epsilon_i^D, \epsilon_i^T)$, $\hat{l}(t, p, S, \breve{p}|\boldsymbol{\theta}) = h_{tpS}(t, p, S) h_{\breve{p}}(\breve{p}|p, S)$, where $\boldsymbol{\theta}$ is a vector of parameters determining the underlying distributions. The density h_{tpS} of the joint distribution of t, p, and S is given by G, the pricing function $p = \tilde{\Phi}^{-1}(c/0.94)$, and the geometric distribution of t given the probability of sale per period $1 - F_{(1)}(p)$ and the probability of dropping out $1 - \delta$. The density $h_{\breve{p}}$ of the distribution of the transaction price $\breve{p}$ conditional on p and S is given by standard auction theoretical reasoning. We provide a formal derivation in Appendix B.1.1.

When taking into account noise, the likelihood $l(\boldsymbol{X}_i, \boldsymbol{\theta})$ of an observation $\boldsymbol{X}_i$ given parameters $\boldsymbol{\theta}$ is obtained by integrating and adding over the noise variables:

$$l(\boldsymbol{X}_i|\boldsymbol{\theta}) = \sum_{k=1}^{T_i/\tau} \int_0^\infty \int_0^\infty \hat{l}\left(T_i - k\tau, \frac{P_i}{\epsilon^P}, S_i, \frac{\epsilon^D \breve{P}_i}{\epsilon^P} \middle| \boldsymbol{\theta}\right) h_t(k\tau) h_P(\epsilon^P) h_D(\epsilon^D) d\epsilon^D d\epsilon^P, \quad (11)$$

where h_D and h_P are the densities of ϵ^D and ϵ^P, respectively, k is the summation variable of $\epsilon^T = k\tau$, and h_T is the probability mass function of ϵ^T.

3.4 Identification

Even though our estimation is based on a Bayesian approach, it is worth noting that our model is non-parametrically identifiable. This is easiest to see for the case without noise, that is when ϵ^Q and ϵ^D are constant, and ϵ^T is zero. The basic idea of the identification is the following. We observe two functions empirically, the probability of ever selling $F_\infty(p)$ and the expected time on market $T(p)$. Evaluating one of the functions at all values of p and the other function at two values $p \in \{p_1, p_2\}$, we can identify the distribution of the highest valuation $F_{(1)}$, the rematching time τ, and the dropout probability $1 - \delta$. From this, we can construct the optimal price $p = \tilde{\Phi}^{-1}(c/0.94)$ a seller of type c would set. Using the inverse pricing function $c = 0.94\tilde{\Phi}(p)$ and the empirical distribution of prices, one can back out the distribution G of costs c.

When allowing for noise (ϵ^P, ϵ^D, and ϵ^T), the reasoning is more involved and makes use of the entire joint distribution of p and t rather than only the expected time on market conditional on price $T(p)$ to identify the distributions of ϵ^P, ϵ^D, and ϵ^T. We provide formal results for identification in Appendix B.2.

3.5 Estimation

We parameterize the virtual type Φ and Γ as Chebyshev polynomials rather than taking the usual approach of parameterizing the distributions F and G. This has a number of advantages for our purposes. Most importantly, this class is a strict superset of (mirrored) Generalized Pareto distributions, which correspond to linear virtual types functions. It also permits us to impose Myerson's regularity condition with a simple parameter restriction, that is, monotonicity of Φ and Γ. Note also that there is a one-to-one mapping between Φ and F and between Γ and G and that closed form solutions can be obtained for F and G.[64]

Like in the theoretical model, we impose the restriction $\underline{v} = \underline{c}$ and $\overline{v} = \overline{c}$ and only use the parameters $\underline{c}$ and $\overline{v}$ in the following. To reduce complexity of the estimation procedure we use a parametric Bayesian estimation.[65] As we model virtual type functions as Chebyshev polynomials, there is in principle little loss in flexibility as one can simply increase the number of polynomial terms considered for estimation at the cost of reducing the degrees of freedom and increasing the computational complexity. Specifically, let $y_B := \frac{2v-\underline{v}-\overline{c}}{\overline{v}-\underline{c}}$ and $y_S := \frac{2c-\underline{c}-\overline{v}}{\overline{v}-\underline{c}}$, be the valuation and cost that are normalized to the range $[-1, 1]$. The Chebyshev polynomials that parameterize the virtual type functions are then given $\Phi(v) = \sum_{i=0}^{N} \phi_i \mathcal{T}_i(y_B)$ and $\Gamma(c) = \sum_{i=0}^{N} \gamma_i \mathcal{T}_i(y_S)$, where $\mathcal{T}_i(x)$ is the degree i Chebyshev polynomial and N is the degree of polynomial approximation.[66] Calligraphic font $\mathcal{T}$ is used to distinguish Chebyshev polynomials from time on market T.

The parameters ϕ_0 and γ_0 in the polynomial parametrization of Φ and Γ are pinned down by the constraints $\Phi(\overline{v}) = \overline{v}$ and $\Gamma(\overline{c}) = \overline{c}$.[67] The measurement error in the time on market ϵ^T has a geometric distribution with survival probability β_T. The reserve price noise ϵ^P and the discount noise ϵ^D are lognormally distributed with $\ln \epsilon^P \sim N(\mu_D, \sigma_P)$ and $\ln \epsilon^D \sim N(\mu_D, \sigma_D)$ with the restriction $\sigma_P > \sigma_D$.[68]

Given observations $\boldsymbol{X} = (\boldsymbol{X}_i)_{i=1}^n$ with $\boldsymbol{X}_i = (T_i, P_i, S_i, \breve{P}_i)$, the likelihood function $l(\boldsymbol{X}|\boldsymbol{\theta})$ is $l(\boldsymbol{X}|\boldsymbol{\theta}) = \prod_{i=1}^n l(\boldsymbol{X}_i|\boldsymbol{\theta})$, where $l(\boldsymbol{X}_i|\boldsymbol{\theta})$ is as given in (11) and $\boldsymbol{\theta} = ((\phi_i)_{i=0}^N, (\gamma_i)_{i=0}^N, \beta_T, \sigma_p, \delta, \underline{v}, \overline{v}, \underline{c}, \overline{c}, \xi, \mu_D, \sigma_D)$ is the vector of parameters over which we take Bayesian expectations. We run a Bayesian estimation with an uninformative prior (i.e. $\pi(\boldsymbol{\theta})$ constant), which means that by Bayes' Law our posterior

$$\pi(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{l(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{l(\boldsymbol{X})}$$

[64] F can be obtained by solving the differential equation $\Phi(v) = v - (1 - F(v))/f(v)$ with initial condition $F(\underline{v}) = 0$. Similarly, G can be obtained by solving $\Gamma(c) = c + G(c)/g(c)$ with initial condition $G(\overline{c}) = 1$.

[65] A Maximum Likelihood Estimator would suffer from errors in variables.

[66] The sequence of Chebyshev polynomials starts with $\{\mathcal{T}_i(y)\}_{i=0} = \{1, y, 2y^2 - 1, 4y^3 - 3y, 4y^4 - 8y^2 + 1, ...\}$.

[67] The explicit expressions are $\phi_0 = \overline{v} - \sum_{i=1}^N \phi_i$ and $\gamma_0 = \underline{c} - \sum_{i=1}^N \gamma_i(-1)^i$.

[68] It is computationally more convenient to work with ϵ^P rather than ϵ^Q, but otherwise equivalent.

is proportional to the likelihood function. We compute Bayesian estimates using Markov Chain Monte Carlo simulations. Since the computation is quite time intensive, we use a number of numerical techniques to improve the speed of the algorithm.[69] The Julia[70] code is available upon request from the authors.

Estimated Parameter Values						
Parameters	1993		1994		1995	
ϕ_1	1.15	(0.107)	1.11	(0.124)	1.06	(0.106)
ϕ_2	-0.0601	(0.0421)	-0.0177	(0.0469)	-0.0200	(0.0330)
γ_1	1.27	(0.0328)	1.25	(0.0272)	1.20	(0.0200)
γ_2	0.134	(0.0130)	0.138	(0.0109)	0.140	(0.00612)
$\underline{x}$	0.0165	(0.0159)	0.0151	(0.0145)	0.0122	(0.0119)
$\overline{x}$	1.30	(0.0266)	1.29	(0.0218)	1.23	(0.0133)
δ	0.945	(0.0112)	0.965	(0.0135)	0.984	(0.00113)
ξ	0.136	(0.0973)	0.321	(0.168)	0.175	(0.0914)
μ_D	0.0953	(0.00367)	0.0833	(0.00335)	0.0735	(0.00235)
σ_D	0.0747	(0.00262)	0.0738	(0.00234)	0.0572	(0.00167)
σ_P	0.243	(0.00724)	0.201	(0.00553)	0.191	(0.00514)
β_T	0.0701	(0.0170)	0.142	(0.0491)	0.385	(0.0585)
# Observations	736		840		793	

Table 2: Estimated Parameter Values for 1993 to 1995. Table entries read: Mean (Standard Deviation).

Parameter Estimates Table 2 contains the posterior mean and standard deviations of the parameter estimates for 1993, 1994, and 1995. There is remarkably little variation in parameter estimates over time. Although the coefficient γ_2 is positive and about 0.13 for each year, which means that the virtual cost function is not exactly linear, it is only roughly one tenth of the coefficient γ_1 for the linear term in the Chebyshev polynomial.[71]

[69] We approximate functions with polynomials which allows us to work with closed forms of polynomials for some calculations. Since we have parameterized the virtual type functions as polynomials, it is fast to numerically check that whether the virtual type functions are increasing everywhere on the support by computing the roots of the derivatives of Φ and Γ. It further helps that the Just-In-Time compiler of the Julia programming language emits code with a speed within a factor 2 of Fortran.

[70] See `http://julialang.org/`.

[71] Because the extrema of all Chebyshev polynomials are either -1 or 1, such comparisons of the size of coefficients are meaningful.

Goodness of Fit Figure 3 compares the predictions of our structural model for quality adjusted listing prices and time on market with reduced form estimates. The figure suggests that the predictions of the model fit the data reasonably well. This suggests that our parametric restrictions are not too strong. Note that we could in principle fit any (analytical) functions F and G arbitrarily well by increasing the degree of the polynomials Φ and Γ, since there is a one-to-one mapping between Φ and F and between G and Γ. This also means that we could in principle fit the price distribution and the time on market as a function of price arbitrarily well. However, to avoid an excessive computational burden, we have chosen second-degree polynomials for Φ and Γ.

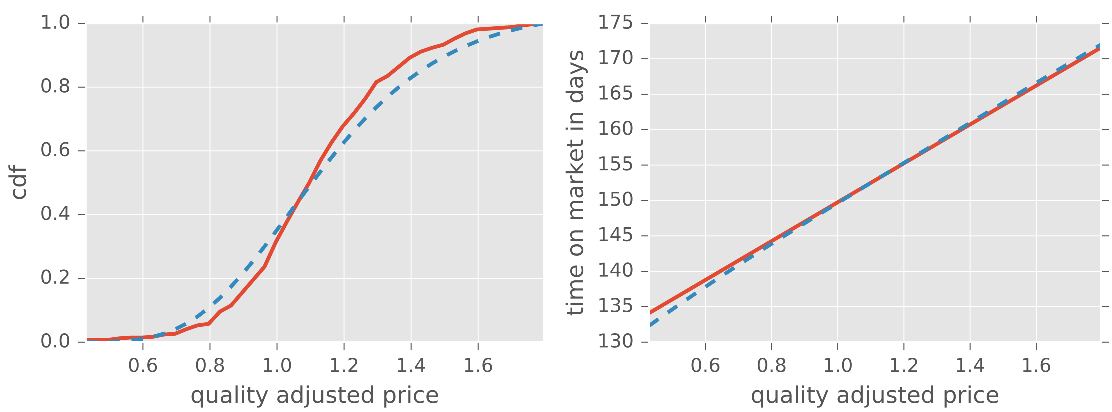
One may wonder what role noise (i.e. unobservable heterogeneity) plays in fitting the data. One way to check this is the following. In one of the robustness checks in the Appendix, we use an alternative quality index (hedonic regression rather than previous transaction prices). While the noise for the two quality indices is quite different, the estimated parameters of F and G are almost the same (see B.4). This suggests that we are indeed picking up deep structural parameters rather than fitting data with noise.[72]

Robustness Checks Appendix B.4 shows a number of robustness checks: (i) using a hedonic regression rather than previous transaction prices for the quality index, (ii) an estimation assuming that the empirically chosen fee by brokers' is 5% or 5.5% rather than 6%, and (iii) using the excluded years (we only use the years 1993-1995 in the main text, since there is little price variation over time in these years). Robustness checks (i) and (ii) provide remarkably similar parameter estimates. Even for robustness check (iii) the estimates are relatively close to the main estimation results, despite the the fact that

[72] One might also ask how well we could fit some additional aspects of the data besides the distribution of prices and the time on market (such as the probability of sale as a function of price) without allowing for noise ϵ^P, ϵ^D, and ϵ^T. While this could be done (by simply running the estimation while restricting the distributions of the noise variables to be degenerate), we do not think that would the proper way to deal with this issue: we know from the auction econometrics literature (see e.g. Krasnokutskaya (2011)) that unobserved heterogeneity is an aspect of real world data that should not be ignored. The fact that part of the data is explained by unobserved heterogeneity should not be viewed as a problem, but rather that we seem to be on a middle ground between two extreme views on price dispersion. One extreme view is that a quality index perfectly corrects for heterogeneity and all observed price dispersion is true price dispersion. The other extreme view is that there is no true price dispersion and that all observed price dispersion is due to unobserved heterogeneity.

the additional years showed a considerable amount of price changes over time.

The most involved robustness check is provided in Appendix E. In Appendix E we relax the independent private values assumption and allow the seller to have private information about the quality of the property. In particular, we assume that the buyer's valuation is $v(x, c) = \lambda c + (1 - \lambda)x$, where $x \sim F$ is the buyer's private signal about his idiosyncratic valuation and $\lambda \in [0, 1)$ is the weight of the common value component. This specializes to an independent private values setup for $\lambda = 0$. In order to structurally estimate such a model, we extend the informed seller auction model (where the seller's reserve serves as a signal) of Cai, Riley, and Ye (2007) by percentage fees and dynamics. To the best of our knowledge, this is the first paper to structurally estimate an informed seller auction model.[73] We find that independent private values is a good approximation (λ is close to zero) and therefore relegate this robustness check to the Appendix to save space.



(a) Price Dispersion (b) Time on Market

Figure 3: The structural model's predictions on the distribution of quality adjusted listing prices and time on market (dashed line). As a comparison, reduced-form estimates (solid lines) are provided: the empirical cdf of quality-adjusted listing prices (left panel) and an Ordinary Least Squares estimate of the relation between quality adjusted listing price and time on market (right panel).

[73]See Niedermayer, Shneyerov, and Xu (2015) for a reduced form estimation of reserve price signaling in foreclosure auctions, a setup very different from the current setup.

Backing-out the Bargaining Parameter α In slight abuse of notation, let $W(\alpha, b)$ denote the value of the objective when a percentage fee $\omega(p) = bp$ is used. Assuming that the empirically observed 6 percent fees are chosen to maximize $W(\alpha, b)$ over b, we can back-out α^* by solving

$$\frac{\partial W(\alpha^*, b)}{\partial b}|_{b=0.06} = \alpha W_I'(b) + (1-\alpha)(W_I'(b) + W_S'(b))|_{b=0.06} = 0,$$

for α, where $W_I(b)$ and $W_S(b)$ are the broker's and the seller's expected welfare given the percentage fee b. More details about the underlying procedure are provided in Appendix B.3. The results are displayed in Table 3.

Price Dispersion Before turning to the counterfactuals, we can look at a more immediate result following from our estimation: price dispersion. It is an old question in Economics to what extent prices are dispersed. Stigler (1961) argues that price dispersion is important in real world markets, but notes that it is difficult to disentangle "true price dispersion" from unobserved heterogeneity in quality (which is observable to market participants but not to the econometrician).[74] Our structural estimates of the parameters $\underline{x}$ and $\overline{x}$ (and the parameters ϕ_1, ϕ_2, γ_1, γ_2) are in line with Stigler's claim that there is indeed true price dispersion: the underlying distributions of v and c and hence also of true quality adjusted prices are clearly non-degenerate.

4 Counterfactual Analyses

Our theoretical model and the empirical estimates of demand and supply lend themselves to a variety of insightful counterfactual exercises. In this section, we perform three such analyses. In turn, we compare the performance of 6 percent fees against the benchmark of the Bayesian optimal mechanism, analyze fee-regulation, and the effects of introducing a transfer tax. These counterfactual exercises are based on a partial equilibrium analysis.

[74] Stigler (1961, p. 214) writes: "Dispersion is a biased measure of ignorance because there is never absolute homogeneity in the commodity. Thus, some automobile dealers might perform more service, or carry a larger range of varieties in stock, and a portion of the observed dispersion is presumably attributable to such differences. But it would be methaphysical, and fruitless, to assert that all dispersion is due to heterogeneity." See also Pennerstorfer, Schmidt-Dengler, Schutz, Weiss, and Yontcheva (2014) for a paper that uses commuter data to estimate true price dispersion in the Austrian gasoline market.

In particular, we abstract from any effects our experiments have on the arrival rate of buyers ξ and thereby on the distribution $F_{(1)}$. We also maintain the assumption that α does not vary with our policy experiments.

4.1 Almost Optimal Percentage Fees

Denote by $\omega_{\alpha^*}^{opt}$ the optimal fee given α^*, the backed-out bargaining weight. Our first counterfactual experiment consists of comparing the weighted joint surplus for 6 percent fees, $W(\alpha^*, 0.06)$, with the weighted joint surplus for the optimal fee schedule, $W(\alpha^*, \omega_{\alpha^*}^{opt})$. As Table 3 shows, for each of the three years the 6% fee achieves more than 99 percent of the joint surplus that the seller and an intermediary can obtain given α^*.

An explanation sometimes given for linear pricing is that competition makes price discrimination impossible. One may wonder whether this is also an explanation for linear fees. Leaving aside the issue that linear fees are fundamentally different from linear prices,[75] we can take a closer look at the effect of decreasing the bargaining parameter α^* of the intermediary, which can be seen to capture competitive pressure in a reduced form.

There are two effects moving linear fees closer to optimality (by moving $\Gamma_{\alpha^*}(c) = \alpha^*\Gamma(c)+(1-\alpha^*)c$ closer to linearity): Γ being close to linear and α^* being less than 1. To make the point that the former effect is important, we make an additional counterfactual with a hypothetical monopolist (i.e. $\alpha = 1$ and $\Gamma_\alpha(c) = \Gamma(c)$). In particular, we compute the optimal percentage fee b_1^* a monopoly would set for each of years considered, and the ratio of monopoly profit with the optimal percentage fee over the monopoly profit with the optimal fee ω_1^{opt}, $W(1, b_1^*)/W(1, \omega_1^{opt})$. As shown in Table 3, for each year this ratio is 0.996. Because the linearity ω_1^{opt} is driven by the linearity of $\Gamma(c)$, these results demonstrate that the remarkable performance of percentage fees is indeed driven by the near linearity of Γ.[76] Numerical simulations, which we do not display because of

[75] As mentioned, linear pricing refers to a *linear transformation*, i.e., the price increasing proportionally with the quantity bought. Contrary to this, *linear fees* refer to fees that are an *affine transformation* of the transaction price, i.e., they are the sum of a fixed transaction fee and a fee that is proportional to the transaction price.

[76] We know of no clear-cut sensible metric other than how well linear (or proportional) fees fare relative to the optimal fees to assess how close Γ is to linearity.

space constraints, further show that percentage fees are robust mechanisms insofar as the respective objective functions vary only little in the neighborhood of 1 or 2 percentage points of the maximizer.

Bargaining Parameter and Counterfactual Fees and Welfare			
Variable	1993	1994	1995
α^*	0.089	0.083	0.077
Performance of 6% Fee: $\frac{W(\alpha^*,0.06)}{W(\alpha^*,\omega_{\alpha^*}^{opt})}$	0.992	0.995	0.997
Optimal Percentage Fee for a Monopoly: b_1^*	35.6%	36.0%	36.2%
Performance of b_1^*: $\frac{W(1,b_1^*)}{W(1,\omega_1^{opt})}$	0.996	0.996	0.996

Table 3: α^* and counterfactual fees and welfare for the three years 1993 to 1995.

One way to interpret the parameters in Tables 2 and 3 is to first consider a linear approximation of Γ with $\gamma_2 = 0$ and $\underline{x} = 0$. For the corresponding approximating Pareto distribution $G(c) = (c/\overline{x})^{\gamma_1}$, a percentage fee $\omega(p) = \alpha/(\alpha + \gamma_1)$ is exactly optimal. For the estimated values of γ_1 and α, the implied percentage fee is close to 6%. How far away a percentage fee is from the optimal fee is driven by how much γ_2 and $\underline{x}$ differ from 0.

4.2 Regulated Fees

One way of intervening in markets that are perceived as insufficiently competitive is by regulating prices, which in our setup is equivalent to regulating fees. To model such fee-regulation, we assume that the government can directly determine the fee a broker can charge, in particular that the government can cap fees. This is e.g. the policy currently in effect in Austria.[77] The effect of a fee cap can be decomposed into three effects: the direct effect (assuming market participants do not change their behavior), the indirect inframarginal seller effect (sellers active in the market change their reserve prices), and a marginal seller effect (additional sellers enter the market). We provide formal results and detailed numerical calculations in Appendix C.1 and only provide a brief summary here. A cap on fees has only a small effect on prices (a 1% decrease in fees reduces the average

[77] The regulation of real-estate brokerage fees ("Immobilienmaklerverordnung") in Austria caps brokerage fees at 6%, excluding a VAT of an additional 1.2%.

price by less than 0.05%), which is consistent with the findings of Proposition 5 (stated in terms of approximations rather than exact statements): for a distribution close to a mirrored Generalized Pareto distribution, the average price changes little as fees change, since the three effects approximately cancel out. In terms of welfare, sellers benefit the most (a 1% decrease of the fee changes the seller's welfare by 2.33% as a fraction of total welfare before the change), intermediaries lose nearly as much as sellers gain (2.25% of total welfare), and buyers are least affected (a gain of 0.25% of total welfare).

4.3 Transfer Taxes

Transaction taxes are an important source of revenue for governments around the globe, generating at times controversial policy debates as, for example, in the case of the financial transaction tax in the European Union. Transfer taxes on real-estate transactions vary across countries and, in some cases, across states and other jurisdictions within a country. For example, in the United States the marginal transfer fees range from zero in Alaska to more than 2.6 percent in New York City according to the National Conference of State Legislatures.[78] We provide an analysis of the equilibrium effects of such transfer taxes on consumer surplus, and the welfare of intermediaries and sellers in Appendix C.2 and a brief summary here. For transfer taxes, there is an additional effect besides the three effects described above (direct effect, inframarginal seller effect, marginal seller effect): the adjustment of fees charged by intermediaries.

Using our estimated parameters, we find that an increase of the transfer tax by 1% decreases intermediaries' welfare by 0.4% as a fraction of total welfare before taxes and decreases the welfare of buyers by 0.2%. Sellers are hit most severely (a decrease of welfare by 1.4%). We also find that the first-order effect of the endogenous adjustment of fees to the changes of taxes is a redistribution from intermediaries to sellers.

[78]See http://www.ncsl.org/research/fiscal-policy/real-estate-transfer-taxes.aspx

5 Discussion

Our analysis has been about how brokers extract revenues by choosing the *fee structure*, that is how the fee depends on the transaction price. There are other aspects of estate brokerage contracts which are orthogonal to the fee structure such as exclusivity clauses, whether the seller or the buyer pays the fee (for the rental market: the landlord versus the tenant paying), or in-house transactions.[79] Moreover, regardless of their exact structure and magnitude, transaction fees provide incentives to exert effort because they are only paid in case a transaction occurs.

As mentioned before, we are deliberately agnostic about the various (competing) explanations that have been provided for why the services of real estate brokers are demanded.[80] Another issue in the real estate brokerage industry are allegations of collusion. This is a difficult question and a challenge the empirical literature has faced is that there is no theoretical model of collusion in transaction fees yet (see Han and Strange, 2014, p. 862). We will not attempt to give a complete answer to this question here, but want to point out that one can view our bargaining weight α as capturing the degree of competitiveness in a reduced form: $\alpha = 0$ for perfect competition, $\alpha = 1$ for a monopolist, and $\alpha \in (0, 1)$ for imperfect competition (either due to service differentiation/geographical differentiation or due to collusion).[81]

[79]See Bar-Isaac and Gavazza (2015) for an intriguing analysis of the latter two points for the rental market in New York, where there is more variation of contractual forms than for the sales market. See also Han and Hong (forthcoming) for an analysis of in-house transactions of brokerage firms. We hope that future work can combine insights on such additional features of brokerage contracts with insights on transaction fees as optimal pricing.

[80]The duties of a real estate agent include "list[ing] the house on the MLS, assist[ing] sellers in staging and marketing the house, advis[ing] sellers on the listing price, help[ing] sellers evaluate offers and formulate counteroffers, help[ing] negotiate directly with the buyer or the buyers agent, and provid[ing] assistance in closing a transaction" (Han and Strange, 2014, p. 851). Agents on the buyer's side " attempt to find houses that match buyers tastes, show buyers prospective homes, advise them in making offers, and provide assistance in the negotiation process" (Han and Strange, 2014, p. 851). Various other reasons have also been provided for why real estate brokers' services are demanded. See Han and Strange (2014) for an overview.

[81]The problem is actually even more difficult than one might initially think: it is not only that there is so far no way to structurally estimate collusion, there is not even a way to structurally estimate competition. This is because transaction fees are very similar to non-linear pricing and it is well known that the structural estimation of competitive non-linear pricing (without making strong parametric assumptions on endogenous price schedules) is an unsolved problem in the literature. A difficulty to start with is that even theoretically, non-linear pricing with product differentiation is quite involved (see Rochet and Stole (2002)). We think that the current paper can serve as a starting point for an

As an illustration, consider the roughly 2% real estate brokerage fees reported for the UK. From the perspective of our theory, an explanation for the difference between fees in the UK and the US can be seen by considering a Pareto approximation of the distribution (which is equivalent to taking a first-degree approximation of the Chebyshev polynomial Γ) $G(c) = c^{\gamma_1}$, which implies a percentage fee $\omega(p)/p = \alpha/(\alpha + \gamma_1)$. Our estimates ($\alpha^{US} \approx 0.08$ and $\gamma_1^{US} \approx 1.25$) correspond to a percentage fee of 6%. The lower fees in the UK could be explained by the real estate market in the UK being more competitive ($\alpha^{UK} \approx 0.0255$ and $\gamma_1^{UK} = \gamma_1^{US}$). The reason for the US market being less competitive ($\alpha^{US} > \alpha^{UK}$) may be either collusion by real estate brokers or services by US real estate brokers being more differentiated. An alternative explanation – that we find less plausible – is that seller's supply is more elastic in the UK ($\sigma^{UK} \approx 3.9$ and $\alpha^{UK} = \alpha^{US}$).

Our analysis can be seen as complementary to Barwick, Pathak, and Wong (forthcoming). Barwick, Pathak, and Wong (forthcoming) find that the splitting of brokerage fees between the seller's and the buyer's agent (in the cases where two agents are involved) facilitates collusion, because buyers' brokers would steer away buyers from sellers' brokers who undercut fees. Barwick, Pathak, and Wong (forthcoming) consider the optimal (potentially collusive) fee as exogenously given and focus on a possible method of sustaining collusion. Our paper focuses on endogenizing the optimal fee by a broker with market power (either due to collusion or imperfect competition) and does not consider how potential collusion may be sustained. Our optimal pricing explanation is orthogonal to how the fee is split between the buyer's and the seller's broker, provided that the two brokers can find an agreement to overcome the double marginalization problem. As mentioned before, the alternative principal-agent explanation of brokerage fees is not orthogonal to the issue of fees being split between the buyer's and the seller's broker, since this explanation would predict that the buyer's broker's fee should *decrease* with the transaction price.

analysis of collusion, but several additional steps are needed, including dealing with issues that are so far unsolved problems in the literature.

6 Conclusions

We provide a parsimonious theory of optimal transaction fees in thin markets. As markets become increasingly thin, the optimal fees converge to linear fees. We show empirically that linear fees are nearly optimal. Moreover, our theory predicts that in thin markets average prices do not vary with the percentage fee charged. This prediction is also almost exactly borne out in the data. Our counterfactual analyses show further that the first-order effect for changes in agents' welfare from changes in fees or from the imposition of a transfer tax resides in the endogenous adjustment of the reserve prices sellers set.

Our theory assumes optimizing behavior by economic agents, which can be justified on the usual grounds that such a theory is robust to the Lucas-critique and that (approximately) optimal behavior may be the result of an evolutionary trial-and-error process.

While the main purpose of this article is to develop a general model of transaction fees as optimal pricing, a positive side effect of having such a theory is that it resolves many of the puzzling observations documented in the empirical literature on real estate brokerage fees by providing an alternative to the principal-agent view.

An aspect of our model that deserves emphasis is that in thin markets optimal fees vary little with the underlying environment. In real-estate brokerage, the invariance of the 6% fees across times and markets is a well-documented stylized fact (see e.g. Hsieh and Moretti, 2003). According to our theory, the asymptotically optimal fee in thin markets is linear and independent of demand-side factors. The asymptotic optimal fee depends on the Pareto tail index σ. The invariance of fees is hence ultimately related to the invariance of the Pareto tail index. The invariance of Pareto tail indices has been observed in a number of empirical settings, such as for income and wealth distributions, the sizes of cities, and the strengths of earthquakes. See Gabaix (2016) for a general discussion of power laws in Economics.

The theoretical and empirical analyses in our paper imposed only mild regularity conditions on distributions and no restrictions on the mechanisms brokers could choose. Given the near-optimality of percentage fees that we found in this paper, future research that models competition between brokers, taking percentage fees as given, or that estimates the demand and supply side taking as given that distributions are (mirrored) Generalized Pareto distributions seems particularly valuable and promising.

References

AGUIRRE, I., S. COWAN, AND J. VICKERS (2010): "Monopoly Price Discrimination and Demand Curvature.," *American Economic Review*, 100(4), 1601 – 1615.

ANDERSON, S., A. DE PALMA, AND B. KREIDER (2001a): "Tax Incidence in Differentiated Product Oligopoly," *Journal of Public Economics*, 81, 173–192.

——— (2001b): "The Efficiency of Indirect Taxes under Imperfect Competition," *Journal of Public Economics*, 81, 231–251.

ANTRÀS, P., AND A. COSTINOT (2011): "Intermediated Trade," *The Quarterly Journal of Economics*, 126(3), 1319–1374.

BAJARI, P. (1997): "The First Price Sealed Bid Auction with Asymmetric Bidders: Theory and Applications," *University of Minnesota, unpublished Ph. D. dissertation.*

BAJARI, P., AND A. HORTAÇSU (2003): "The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions," *RAND Journal of Economics*, 34(2), 329–355.

BALAT, J., P. A. HAILE, H. HONG, AND M. SHUM (2016): "Nonparametric Tests for Common Values In First-Price Sealed-Bid Auctions," .

BALKEMA, A., AND L. DE HAAN (1974): "Residual Life Time at Great Age," *Annals of Probability*, 2(5), 792–804.

BAR-ISAAC, H., AND A. GAVAZZA (2015): "Brokers? contractual arrangements in the Manhattan residential rental market," *Journal of Urban Economics*, 86, 73–82.

BARVINEK, E., I. DALER, AND J. FRANCU (1991): "Convergence of Sequences of Inverse Functions," *Archivum Mathematicum*, 27(3-4), 201–204.

Barwick, P. J., P. A. Pathak, and M. Wong (forthcoming): "Conflicts of Interest and Steering in Residential Brokerage," *American Economic Journal: Applied Economics*.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, pp. 841–890.

Board, S. (2008): "Durable-Goods Monopoly with Varying Demand," *Review of Economic Studies*, 75(2), 391–413.

Bulow, J., and P. Klemperer (2012): "Regulated Prices, Rent Seeking, and Consumer Surplus.," *Journal of Political Economy*, 120(1), 160 – 186.

Bulow, J., and J. Roberts (1989): "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, 97(5), 1060–90.

Bulow, J. I., and P. Pfleiderer (1983): "A Note on the Effect of Cost Changes on Prices.," *Journal of Political Economy*, 91(1), 182 – 185.

Cai, H., J. Riley, and L. Ye (2007): "Reserve Price Signaling.," *Journal of Economic Theory*, 135(1), 253 – 268.

Caillaud, B., and B. Jullien (2003): "Chicken & Egg: Competition among Intermediation Service Providers," *RAND Journal of Economics*, 34(2), 309–328.

Cremer, J., and R. P. McLean (1988): "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica*, pp. 1247–1257.

De Haan, L., and A. Ferreira (2006): *Extreme Value Theory: An Introduction*. Springer Verlag.

Delipalla, S., and M. Keen (1992): "The comparison between ad valorem and specific taxation under imperfect competition," *Journal of Public Economics*, 49(3), 351 – 367.

Department of Justice (2007): "Competition in the Real Estate Brokerage Industry," *Report by the Federal Trade Commission and the U.S. Department of Justice.*

Donald, S., and H. Paarsch (1993): "Piecewise pseudo-maximum likelihood estimation in empirical models of auctions," *International Economic Review*, pp. 121–148.

Einav, L., D. Knoepfle, J. Levin, and N. Sundaresan (2014): "Sales Taxes and Internet Commerce," *American Economic Review*, 104(1), 1 – 26.

Falk, M., J. Hüsler, and R. Reiss (2010): *Laws of Small Numbers: Extremes and Rare Events*. Springer Verlag.

Gabaix, X. (2016): "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 30(1), 185–206.

Garrett, D. (forthcoming): "Intertemporal price discrimination: dynamic arrivals and changing values," *American Economic Review.*

Genesove, D., and L. Han (2012): "Search and matching in the housing market," *Journal of Urban Economics*, 72(1), 31–45.

Genesove, D., and C. J. Mayer (1997): "Equity and Time to Sale in the Real Estate Market," *American Economic Review*, 87(3), 255–269.

——— (2001): "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 116(4), 1233–1260.

Glower, M., D. Haurin, and P. Hendershott (1998): "Selling Time and Selling Price: The Influence of Seller Motivation.," *Real Estate Economics*, 26(4), 719–721.

Gomes, R. (2014): "Optimal Auction Design in Two-Sided Markets," *RAND Journal of Economics*, 45, 248–272.

Gresik, T. (1991): "Ex Ante Efficient, Ex Post Individually Rational Trade," *Journal of Economic Theory*, 53, 131–145.

HAGIU, A. (2007): "Merchant or Two-Sided Platform?," *Review of Network Economics*, 6(2), 115–33.

HAN, L., AND S.-H. HONG (forthcoming): "Understanding In-House Transactions in the Real Estate Brokerage Industry," *RAND Journal of Economics*.

HAN, L., AND W. STRANGE (2014): "The microstructure of housing markets: Search, bargaining, and brokerage," *Handbook of Regional and Urban Economics*, 5.

HENDEL, I., A. NEVO, AND F. ORTALO-MAGNÉ (2009): "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com," *American Economic Review*, 99(5), 1878 – 98.

HSIEH, C.-T., AND E. MORETTI (2003): "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry," *Journal of Political Economy*, 111(5), 1076–1122.

JOHNSON, J. P. (2014): "The agency model and MFN Clauses," *Available at SSRN 2217849*.

JULLIEN, B., AND T. MARIOTTI (2006): "Auction and the Informed Seller Problem," *Games and Economic Behavior*, (56), 225258.

KANG, H. B., AND M. J. GARDNER (1989): "Selling Price and Marketing Time in the Residential Real Estate Market," *Journal of Real Estate Research*, 4(1), 21–35.

KRASNOKUTSKAYA, E. (2011): "Identification and estimation of auction models with unobserved heterogeneity," *Review of Economic Studies*, 78(1), 293–327.

KRISHNA, V. (2002): *Auction Theory*. Elsevier Science, Academic Press.

LAMY, L. (2010): "Corrigendum to 'Reserve price signaling'," *Journal of Economic Theory*, (145), 2498 – 2504.

LAUERMANN, S. (2013): "Dynamic Matching and Bargaining Games: A General Approach," *American Economic Review*, 103(2), 663–89.

Lauermann, S., W. Merzyn, and G. Virag (2012): "Learning and Price Discovery in a Search Model," *Working Paper.*

Levitt, S. D., and C. Syverson (2008): "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions," *Review of Economics and Statistics*, 90(4), 599–611.

Loertscher, S., and A. Niedermayer (2007): "When is Seller Price Setting with Linear Fees Optimal for Intermediaries?," *Working Paper, University of Bern.*

——— (2012): "Fee Setting Intermediaries: On Real Estate Agents, Stock Brokers, and Auction Houses," *Working Paper.*

Matros, A., and A. Zapechelnyuk (2008): "Optimal Fees in Internet Auctions," *Review of Economic Design*, 12(3), 155–63.

Merlo, A., and F. Ortalo-Magné (2004): "Bargaining over residential real estate: evidence from England," *Journal of Urban Economics*, 56(2), 192–216.

Myerson, R. (1981): "Optimal Auction Design," *Mathematical Operations Research*, 6(1), 58–73.

Myerson, R., and M. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29(2), 265–281.

Mylovanov, T., and T. Tröger (2014): "Mechanism Design by an Informed Principle," *Working Paper.*

Niazadeh, R., Y. Yuan, and R. Kleinberg (2014): "Simple and Near-Optimal Mechanisms For Market Intermediation," http://arxiv.org/abs/1409.2597v1.

Niedermayer, A., and A. Shneyerov (2014): "For-Profit Search Platforms," *International Economic Review*, 55(3), 765–789.

Niedermayer, A., A. Shneyerov, and P. Xu (2015): "Foreclosure Auctions," Mimeo.

Ortalo-Magné, F., A. Merlo, and J. Rust (2015): "The Home Selling Problem: Theory and Evidence," *International Economic Review*, 56, 457 – 484.

Ortalo-Magné, F., and S. Rady (2006): "Housing market dynamics: On the contribution of income shocks and credit constraints," *Review of Economic Studies*, 73(2), 459–485.

Pennerstorfer, D., P. Schmidt-Dengler, N. Schutz, C. Weiss, and B. Yontcheva (2014): "Information and price dispersion: Evidence from retail gasoline," *Vienna University of Economics and Business*.

Pickands, J. (1975): "Statistical Inference Using Extreme Order Statistics," *Annals of Statistics*, pp. 119–131.

Quigley, J. M. (1999): "Real Estate Prices and Economic Cycles," *International Real Estate Review*, 2(1), 1–20.

Riley, J., and R. Zeckhauser (1983): "Optimal Selling Strategies: When to Haggle, When to Hold Firm ," *Quarterly Journal of Economics*, pp. 267–89.

Rochet, J.-C., and L. A. Stole (2002): "Nonlinear pricing with random participation," *The Review of Economic Studies*, 69(1), 277–311.

Rust, J., and G. Hall (2003): "Middlemen versus Market Makers: A theory of Competitive Exchange," *Journal of Political Economy*, 111(2), 353–403.

Rutherford, R., T. Springer, and A. Yavas (2005): "Conflicts between Principals and Agents: Evidence from Residential Brokerage," *Journal of Financial Economics*, 76, 627–65.

Salanié, B. (2003): *The Economics of Taxation*. MIT Press.

Satterthwaite, M., and A. Shneyerov (2007): "Dynamic Matching, Two-Sided Information, and Participation Costs: Existence and Convergence to Perfect Competition," *Econometrica*, 75(1), 155–200.

Satterthwaite, M., and A. Shneyerov (2008): "Convergence to Perfect Competition of a Dynamic Matching and Bargaining Market with Two-sided Incomplete Information and Exogenous Exit Rate.," *Games and Economic Behavior*, 63(2), 435–467.

Shneyerov, A. (2006): "An empirical study of auction revenue rankings: the case of municipal bonds," *RAND Journal of Economics*, 37(4), 1005–1022.

Shy, O., and Z. Wang (2011): "Why Do Payment Card Networks Charge Proportional Fees?," *American Economic Review*, 101(4), 1575–1590.

Spulber, D. F. (1999): *Market Microstructure: Intermediaries and the Theory of the Firm*. Cambridge University Press, Cambridge.

Stigler, G. J. (1961): "The economics of information," *Journal of Political Economy*, pp. 213–225.

Tirole, J. (2016): "From Bottom of the Barrel to Cream of the Crop: Sequential Screening With Positive Selection," *Econometrica*, 84(4), 1291–1343.

Wang, Z., and J. Wright (forthcoming): "Ad-valorem platform fees and efficient price discrimination," *RAND Journal of Economics*.

Weyl, G., and M. Fabinger (2013): "Pass-Through as an Economic Tool: Principle of Incidence under Imperfect Competition," *Journal of Political Economy*, 121(3), 528 – 583.

Wheaton, W. (1990): "Vacancy, search, and prices in a housing market matching model," *Journal of Political Economy*, pp. 1270–1292.

Wolinsky, A. (1988): "Dynamic Markets with Competitive Bidding," *Review of Economic Studies*, 55(1), 71–84.

Yavas, A. (1992): "Marketmakers versus Matchmakers," *Journal of Financial Intermediation*, 2, 33–58.

Online Appendix

A Proofs

A.1 Propositions 1, and 3 to 5

Proof of Proposition 1. The first order condition for the seller's maximization problem is

$$[(R_\omega(p) - c)(1 - F_\infty(p))]' = -[\tilde{\Phi}_\omega(p) - c]f_\infty(p)$$

with

$$\tilde{\Phi}_\omega(p) := R_\omega(p) - R'_\omega(p)\frac{1 - F_\infty(p)}{f_\infty(p)}. \tag{12}$$

We will show that the expression for $\tilde{\Phi}_\omega$ in (12) is the same as the one in the proposition.

First, observe that

$$R_\omega(p) = \frac{(p - \omega(p))(F_{(2)}(p) - F_{(1)}(p)) + \int_p^{\overline{v}}(v - \omega(v))dF_{(2)}(v)}{1 - F_{(1)}(p)}$$

can be rewritten as

$$R_\omega(p) = \frac{\int_p^{\overline{v}} \Phi_\omega(v)dF_{(1)}(v)}{1 - F_{(1)}(p)}$$

where

$$\Phi_\omega(p) := p - \omega(p) - (1 - \omega'(p))\frac{1 - F(p)}{f(p)}$$

That the two expressions for R_ω are equal can be checked by observing that $R_\omega(\overline{v}) = \overline{v}$ for both expressions and that the derivatives $[R_\omega(p)(1 - F_{(1)}(p))]'$ can be shown to be equal for both expression for R_ω with some algebra and by using the fact[82]

$$\frac{F_{(2)}(p) - F_{(1)}(p)}{f_{(1)}(p)} = \frac{1 - F(p)}{f(p)}.$$

One can also show with some algebra that

$$R'_\omega(p) = \frac{f_{(1)}(p)}{1 - F_{(1)}(p)}(R_\omega(p) - \Phi_\omega(p)) \tag{13}$$

and that

$$\frac{f_{(1)}(p)}{1 - F_{(1)}(p)}\frac{1 - F_\infty(p)}{f_\infty(p)} = \frac{1 - \delta F_{(1)}(p)}{1 - \delta} \tag{14}$$

[82]This is easily seen to be true once one notes that $f_{(1)}(v)$ can be written as $f_{(1)}(v) = f(v)\sum_{B=1}^{\infty}\pi_B BF(v)^{B-1}$ and by noticing that $F_{(2)}(v) - F_{(1)}(v) = (1 - F(v))\sum_{B=1}^{\infty}\pi_B BF(v)^{B-1}$.

Plugging (13) and (14) into (12) yields

$$\tilde{\Phi}_\omega(p) = R_\omega(p) - (R_\omega(p) - \Phi_\omega(p))\frac{1-\delta F_{(1)}(p)}{1-\delta}$$

the derivative of which can be rearranged to

$$\tilde{\Phi}'_\omega(p) = \frac{1-\delta F_{(1)}(p)}{1-\delta}\Phi'_\omega(p)$$

Since this expression for $\tilde{\Phi}'_\omega$ is equal to the expression for $\tilde{\Phi}'_\omega$ in the proposition and since the two expressions for $\tilde{\Phi}_\omega(p)$ are equal to $\overline{v}$ for $p = \overline{v}$, Φ_ω is the same in the proposition as in (12).

Therefore, the seller's first-order condition can be written as $-(\tilde{\Phi}_\omega(p) - c)f_\infty(p) = 0$, which implies the optimal price $\tilde{\Phi}_\omega^{-1}(c)$ for a seller with cost c as stated in the proposition.

□

Proof of Proposition 3. The proof of part (i) relies on Extreme Value Theory, which we summarize in Appendix F. Φ continuously differentiable implies that, for some constant $\overline{\beta}$,

$$\lim_{v\to\overline{v}} \frac{d}{dv}\left[\frac{1-F(v)}{f(v)}\right] = \lim_{v\to\overline{v}} \frac{d}{dv}\left[v - \Phi(v)\right] = \overline{\beta}. \tag{15}$$

Equation (15) is the von Mises condition as stated in Theorem 2 in Appendix F. By Theorem 2, this implies that F is in the domain of attraction of an extreme value distribution (see Definition 1). By the Pickands-Balkema-de Haan Theorem (see Theorem 1), this in turn implies that F has a Generalized Pareto upper tail as defined in (47). This implies uniform convergence of the normalized distribution $\tilde{F}_j$ to $\tilde{F}^*(\tilde{v}) = 1 - (1-\tilde{v})^\beta$, because of the definition of the normalized variable $\tilde{v}$. Analogous reasoning applies for the convergence of $\tilde{G}_j$.

Proof of part (ii): First, define $\bar{\beta} := \lim_{v\to\overline{v}} 1 - \Phi'(v)$, $\bar{\sigma} := \lim_{c\to\underline{c}} \Gamma'(c) - 1$, $\beta := -1/\overline{\beta}$, and $\sigma := 1/\overline{\sigma}$. Observe that by l'Hôpital's rule

$$\lim_{v\to\overline{v}} \frac{(\overline{v}-v)f(v)}{1-F(v)} = \lim_{v\to\overline{v}} \frac{\overline{v}-v}{v-\Phi(v)} = \lim_{v\to\overline{v}} \frac{-1}{1-\Phi'(v)} = \beta.$$

The following two constructs are used in the remainder of the proof: First, instead of setting a reserve price p that leads to expected revenue $k = R(p)$ conditional on trade in this period, one can alternatively and hypothetically assume that the seller sets an expected transaction price k, conditional on trade ever occurring, that leads to trade with probability $1 - \overline{F}(k) := 1 - F_\infty(R^{-1}(k))$. Second, the intermediary can levy an "expectational fee" $\varpi(k)$ on the expected transaction price k. The following lemma derives the expectational fee $\varpi(k)$ that implements the allocation rule derived in Lemma 1.

Lemma 2. *The expectational transaction fees that implement the optimal mechanism described in Lemma 1 are*

$$\varpi(k) = k - \frac{\int_k^{\overline{v}} \Gamma_\alpha^{-1}(\overline{\Phi}(v))\overline{f}(v)dv}{1 - \overline{F}(k)}.$$

Proof of Lemma 2. The expected profit of a seller with cost who faces a fee ϖ is

$$(1 - \overline{F}(k))(k - \varpi(k) - c).$$

Substituting $\varpi(k)$ by the expression in Proposition 2, the maximization problem becomes

$$\max_k \int_k^{\overline{v}} \Gamma_\alpha^{-1}(\overline{\Phi}(v))\overline{f}(v)dv - (1 - \overline{F}(k))c.$$

The first-order condition is

$$0 = -\overline{f}(k(c)) \left[\Gamma_\alpha^{-1}(\overline{\Phi}(k)) - c\right],$$

which is equivalent to $\overline{\Phi}(k) = \Gamma_\alpha(c)$. This is equivalent to the allocation rule in Lemma 1 (see its proof for details). The second-order condition is satisfied whenever the first-order condition is satisfied if $\overline{\Phi}(v)$ is monotone. □

The remainder of the proof now proceeds in four steps: we show that (a) for the limiting distributions $\tilde{F}^*$ and $\tilde{G}^*$ the expectational fee ϖ^* is equal to the limiting fee $\frac{\alpha}{\alpha+\sigma}\tilde{p}$, (b) the expectational fee converges to the limiting fee, (c) the transaction fee $\tilde{\omega}$ is equal to the limiting fee for the limiting distributions, and (d) the transaction fee converges to $\frac{\alpha}{\alpha+\sigma}\tilde{p}$.

Step (a): First, we show that linearity of fees holds for the denormalized limiting distributions F^* and G^*. For simplicity, denote the supports of the denormalized limiting distributions as $[\underline{v}, \overline{v}]$ and $[\underline{c}, \overline{c}]$. The distributions are hence $F^*(v) = 1 - [(\overline{v} - v)/(\overline{v} - \underline{v})]^\beta$ and $G^*(c) = [(c - \underline{c})/(\overline{c} - \underline{c})]^\sigma$. The virtual cost function is linear: $\Gamma^*_\alpha(c) = c + (c - \underline{c})\alpha/\sigma$. The optimal expectational fees given in Lemma 2 can be rearranged to yield

$$\overline{\omega}^*(p) = p - E_v[\Gamma^{*-1}_\alpha(\overline{\Phi}^*(v))|v \geq p] = p - \Gamma^{*-1}_\alpha(E_v[\overline{\Phi}^*(v)|v \geq p]) = p - \Gamma^{*-1}_\alpha(p),$$

where the second equality stems from the linearity of Γ^*_α and the third from the well-known fact that for any p and any distribution $\overline{F}$ with virtual value $\overline{\Phi}$, $E_v[\overline{\Phi}(v)|v \geq p] = p$. Plugging in the functional form for Γ^*_α yields

$$\overline{\omega}^*(p) = (p - \underline{c}) \left[\frac{\alpha}{\alpha + \sigma}\right].$$

This implies that the equation for $\overline{\omega}^*$ in (5) holds for the limiting distributions F^* and G^*, because of the definitions of $\tilde{\omega}$ and $\tilde{p}$.

Step (b): Next, we show convergence to linearity. For this, it is useful to consider a linear transformation of the original problem, such that the length of the support is 1 for both F and G, and the lower bound is 0. This can be done without loss of generality. Formally, the support of the seller's distribution $[\underline{c}_j, (\overline{v}_j - \underline{c}_j)/u^S_j + \underline{c}_j]$ is transformed to $[0, 1]$ and the support of the buyer's distribution becomes $[\overline{v}_j - (\overline{v}_j - \underline{c}_j)/u^B_j, \overline{v}_j]$ to $[u_j - 1, u_j]$ with some $u_j > 0$. Note that as $j \to \infty$, $u_j \to 0$. In part of the following analysis, we will drop the subscript j and simply write $u \to 0$.

This has the advantage that the transformed distributions are only shifted and not stretched compared to F and G. Call these transformed distributions $\hat{F}_j$ and $\hat{G}_j$, with $\hat{G}_j(\hat{c}) = G(\hat{c})$ and $\hat{F}_j(\hat{v}) = F(\hat{v} + (1 - u))$. The transformed fee is

$$\hat{\overline{\omega}}(\hat{p}) = u\hat{p} - \frac{\int_{\hat{p}}^1 \hat{\Gamma}^{-1}_\alpha(\hat{\overline{\Phi}}(u\hat{v}))d\hat{\overline{F}}(u\hat{v})}{1 - \hat{\overline{F}}(u\hat{p})} \tag{16}$$

where the expression comes from plugging in $u\hat{p}$ for p in the expression in Lemma 2.

We need to show that the expression in the integral uniformly converges to its limit, which implies convergence of the integral and also convergence of the whole expression for $\hat{\overline{\omega}}$.

By the definition of β we have

$$\lim_{u \to 0} \frac{\partial}{\partial(u\hat{v})} \left[\frac{1 - \hat{F}(u\hat{v})}{\hat{f}(u\hat{v})}\right] = \lim_{v' \to 1} \left[\frac{1 - F(v')}{f(v')}\right]' = \frac{1}{\beta}.$$

This implies that

$$\frac{1}{u}\left[\frac{1-\hat{F}(u\hat{v})}{f(u\hat{v})}\right] \overset{u\to 0}{\rightrightarrows} \frac{\hat{v}}{\beta}$$

and hence

$$\frac{1}{u}\hat{\Phi}(u\hat{v}) \overset{u\to 0}{\rightrightarrows} \hat{v} - \frac{1-\hat{v}}{\beta}$$

where the double arrow $\rightrightarrows$ stands for uniform convergence. By a similar logic

$$\frac{1}{u}\hat{\Gamma}_{\alpha}(u\hat{c}) \overset{u\to 0}{\rightrightarrows} \hat{c}\left(1+\frac{\alpha}{\sigma}\right)$$

and hence

$$\frac{1}{u}\hat{\Gamma}_{\alpha}^{-1}(ux) \overset{u\to 0}{\rightrightarrows} \frac{x}{1+\alpha/\sigma},$$

because uniform convergence of a function implies uniform convergence of its inverse (see for example Barvinek, Daler, and Francu (1991)).

Observe that

$$\hat{\overline{F}}(k) = \hat{\overline{F}}_{\infty}(\hat{R}^{-1}(k)), \qquad \hat{F}_{\infty}(p) = \frac{1-\hat{F}_{(1)}(p)}{1-\delta\hat{F}_{(1)}(p)}, \qquad \hat{F}_{(1)}(p) = \sum_{B=0}^{\infty}\pi_B\hat{F}(p)^B \tag{17}$$

and let

$$\hat{R}_j(p) = \frac{\int_p^{u_j}\hat{\Phi}(v)d\hat{F}_{(1)}(v)}{1-\hat{F}_{(1)}(p)}. \tag{18}$$

By Theorem 1 the expressions in (17) uniformly converge to their respective limits if $\hat{R}_j^{-1}$ uniformly converges. $\hat{R}_j^{-1}$ converges uniformly if $\hat{R}_j$ converges uniformly. So we are left to show that $\hat{R}_j$ converges uniformly in order to show uniform convergence of the integrand in (16) and hence convergence of $\hat{\overline{\omega}}$.

Since the integrand in the integral in $\hat{R}$ converges uniformly, $\hat{R}$ converges pointwise to its limit. Further, observe that the sequence $\hat{R}_j(p)$ with

$$\hat{R}_j(p) = \frac{\int_p^{u_j}\hat{\Phi}(v)d\hat{F}_{(1)}(v)}{1-\hat{F}_{(1)}(p)} = \frac{\int_{p+1-u_j}^{1}(\Phi(y)-(1-u_j))f_{(1)}(y)dy}{1-F_{(1)}(p+1-u_j)}$$

monotonically increases as j goes to infinity (and thus u_j goes to zero). Pointwise convergence and monotonicity of the sequence imply uniform convergence of $\hat{R}_j$ by Dini's theorem. Putting this together implies that $\hat{\overline{\omega}}$ converges to $\overline{\omega}^*$.

Step (c): Observe that if expectational fees are linear, the transaction fees are equal to expectational fees, since a linear function can be taken into an expectation. Hence the transaction fee $\omega(p) = \overline{\omega}(p)$ is also linear in the limit.

Step (d): Next, we turn to convergence of the normalized transaction fee $\hat{\omega}$.

From Proposition 2 and the arguments that precede it, we know that the optimal transaction fee $\omega(p)$ is given by

$$\omega(p) = p - \frac{\int_p^{\overline{v}}[\Gamma_\alpha^{-1}(\tilde{\Phi}(v)) + \delta V(\Gamma_\alpha^{-1}(\tilde{\Phi}(v)))]dF(v)}{1-F(p)},$$

where $V(c) = \int_c^{\Gamma_\alpha^{-1}(\overline{v})}(1 - F_\infty(\tilde{\Phi}^{-1}(\Gamma_\alpha(y))))dy$. Defining $B := \int_p^{\overline{v}} \Gamma_\alpha^{-1}(\tilde{\Phi}(v))dF(v)$ and $A := \int_p^{\overline{v}} \int_{\Gamma_\alpha^{-1}(\tilde{\Phi}(p))}^{\Gamma_\alpha^{-1}(\overline{v})} (1 - F_\infty(\tilde{\Phi}^{-1}(\Gamma_\alpha(y))))dydF(v)$, we can thus write $\omega(p)$ as $\omega(p) = p - (B+\delta A)/(1-F(p))$. Using $\tilde{\Phi}(p) = \overline{\Phi}(R(p))$ it is clear that the integrand in B converges uniformly and hence B converges. Reversing the order of integration in A and integrating we obtain

$$\begin{aligned} A &= \int_{\Gamma_\alpha^{-1}(p)}^{\Gamma_\alpha^{-1}(\overline{v})} \int_p^{\tilde{\Phi}(\Gamma_\alpha(y))} dF(v)(1 - F_\infty(\tilde{\Phi}^{-1}(\Gamma_\alpha(y))))dy \\ &= \int_{\Gamma_\alpha^{-1}(p)}^{\Gamma_\alpha^{-1}(\overline{v})} (F(\tilde{\Phi}(\Gamma_\alpha(y))) - F(p))(1 - F_\infty(\tilde{\Phi}^{-1}(\Gamma_\alpha(y))))dy. \end{aligned}$$

The integrand in the last expression is a combination of functions which we have shown to converge uniformly, hence we get convergence of A. Putting this together, we get that $\hat{\omega}$ converges to the limit given by substituting in the limiting distributions for F and G. □

Proof of Proposition 4. The Taylor expansion of $\Gamma_\alpha^{-1}(x)$ around $\overline{v}$ is

$$\Gamma_\alpha^{-1}(x) = \Gamma_\alpha^{-1}(\overline{v}) + [\Gamma_\alpha^{-1}(\overline{v})]'(x-\overline{v}) + \frac{[\Gamma_\alpha^{-1}(\overline{v})]''}{2}(x-\overline{v})^2 + \sum_{n=3}^{\infty} \frac{[\Gamma_\alpha^{-1}(\overline{v})]^{(n)}}{n!}(x-\overline{v})^n.$$

Denote the nth derivative at $\overline{v}$ as $\overline{\gamma}_n := [\Gamma_\alpha^{-1}(\overline{v})]^{(n)}$. We further use the shorthand

$\varphi := \Phi(v)$. The net price received by the seller can be rearranged as

$$\begin{aligned}
p - \omega(p) =& E[\Gamma_\alpha^{-1}(\varphi)|v \geq p] \\
=& \sum_{n=0}^{\infty} \frac{\overline{\gamma}_n}{n!} E[(\varphi - \overline{v})^n | v \geq p] \\
=& \sum_{n=0}^{\infty} \frac{\overline{\gamma}_n}{n!} \left\{ (E[\varphi|v \geq p] - \overline{v})^n - (E[\varphi|v \geq p] - \overline{v})^n + E[(\varphi - \overline{v})^n | v \geq p] \right\} \\
=& \Gamma_\alpha^{-1}(E[\varphi|v \geq p]) + \sum_{n=0}^{\infty} \frac{\overline{\gamma}_n}{n!} \left\{ E[(\varphi - \overline{v})^n | v \geq p] - (E[\varphi|v \geq p] - \overline{v})^n \right\} \\
=& \Gamma_\alpha^{-1}(p) + \sum_{n=0}^{\infty} \frac{\overline{\gamma}_n}{n!} \left\{ E[(\varphi - \overline{v})^n | v \geq p] - (E[\varphi|v \geq p] - \overline{v})^n \right\},
\end{aligned}$$

where the second equality stems from the Taylor expansion, the fourth from reversing a Taylor expansion, and the fifth from the fact that $E[\Phi(v)|v \geq p] = p$. Note that for $n = 0$ and $n = 1$, the expressions in curly braces cancel out in the last expression. For $n = 2$, the expression in curly braces is the conditional variance $\text{Var}[\varphi - \overline{v}|v \geq p] = E[(\varphi - \overline{v})^2|v \geq p] - (E[\varphi|v \geq p] - \overline{v})^2$. This completes the proof. □

Proof of Proposition 5. Because demand is the same across the two platforms and because $G_{FSBO,p}(p) \equiv G_{0,p}(p)$, to prove the proposition it suffices to show that $G_{p,b}(p)$ does not vary with b when b has no effect on demand. This is what we do in the following. Recall that given a distribution $G(c) = c^\sigma$, the optimal fee given α is $\omega(p) = \frac{\alpha}{\alpha+\sigma}p$. We are now going to show that $G_p^\alpha(p) := G_{p,\alpha/(\alpha+\sigma)}(p)$ does not vary with α.

To see that $G_p^\alpha(p)$ does not vary with α, notice first that a seller can only sell with positive probability if his cost is less than $\Gamma_\alpha^{-1}(\overline{v})$. Since $\tilde{\Phi}^{-1}(\overline{v}) = \overline{v}$ and $\Gamma_\alpha(\underline{c}) = \underline{c}$, it follows that $[\tilde{\Phi}^{-1}(\underline{c}), \overline{v}]$ is the support of reserve prices for any $\alpha \in [0, 1]$. For any p in the support, we then have $G_p^\alpha(p) = \frac{G(\Gamma_\alpha^{-1}(\tilde{\Phi}(p)))}{G(\Gamma_\alpha^{-1}(\overline{c}))}$. The result then follows if we can show that for any $p \in [\tilde{\Phi}^{-1}(\underline{c}), \overline{v}]$, $\frac{G(\Gamma_\alpha^{-1}(p))}{G(\Gamma_\alpha^{-1}(\overline{c}))}$ is independent of α. We now show that in fact

$$\frac{G(\Gamma_\alpha^{-1}(p))}{G(\Gamma_\alpha^{-1}(\overline{c}))} = G(p). \tag{19}$$

To see that this is true, observe first that $\Gamma_\alpha^{-1}(p) = \frac{\sigma}{\sigma+\alpha}p + \frac{\alpha}{\sigma+\alpha}\underline{c}$. Therefore,

$$G(\Gamma_\alpha^{-1}(p)) = \left(\frac{\sigma}{\sigma + \alpha}\right)^\sigma \left(\frac{p - \underline{c}}{\overline{c} - \underline{c}}\right)^\sigma.$$

Consequently, $G(\Gamma_\alpha^{-1}(\overline{c})) = \left(\frac{\sigma}{\sigma+\alpha}\right)^\sigma$, whence (19) follows. □

A.2 Lemma 1

The proof Lemma 1 relies on mechanism design. We say that a *mechanism* is active in period t if the seller has not exited prior to t, which can happen because a transaction has occurred or because of the exogenously given probability $1-\delta$ of dropping out from one period to the next. As mentioned, one can alternatively and equivalently interpret δ as the pure per period survival probability of the seller, or as a discount factor that reflects pure and common time preferences, or a as a combination of the survival probability and time preferences. However, the interpretation of many concepts used in the mechanism design framework is most straight forward if one interprets δ as a pure survival probability. After the seller exits, no good is left to be traded and the mechanism shuts down. The following, therefore, applies only to active mechanisms.

A mechanism is said to be a *direct mechanism* if it asks all agents who participate in the mechanism to report their types. For the seller, who is present at date 0, this simply means that he reports his cost c. A direct mechanism then asks all buyers who enter in period t to report their valuations $v_b \in [\underline{v}, \overline{v}]$ to the mechanism. The realization of the valuations of buyers who do not enter are set to $v_b = -\infty$. Let $\boldsymbol{v}_t = (v_b^t)_{b=1}^{\overline{B}}$ be a vector of such reports by buyers in period t with buyers label $b = 1, .., \overline{B}$ and let $\boldsymbol{v} = (\boldsymbol{v}_t)_{t=0}^{\infty}$ be a sequence of such reports.

A direct mechanism specifies the probability $Q_S^t(\boldsymbol{v}_t, c)$ that the seller sells in period t and the probability $Q_b^t(\boldsymbol{v}_t, c)$ that buyer b receives the good and the payment $M_S^t(\boldsymbol{v}_t, c)$ made from the mechanism to the seller and the payments made by buyers b to mechanism $M_b^t(\boldsymbol{v}_t, c)$, given reports $(\boldsymbol{v}_t, c)$ and given that the mechanism is still active.

Feasibility further requires

$$\sum_{b=1}^{\overline{B}} Q_b^t(\boldsymbol{v}_t, c) \leq Q_S^t(\boldsymbol{v}_t, c) \tag{20}$$

for all t and all $(\boldsymbol{v}_t, c)$. Accordingly, the mechanism ceases to be active in period t with probability $Q_S^t(\boldsymbol{v}_t, c)$, and it proceeds to period $t+1$ with probability $(1-\delta)(1-Q_S^t(\boldsymbol{v}_t, c))$.

Let $\boldsymbol{Q}_B^t(\boldsymbol{v}_t, c) = (Q_1^t(\boldsymbol{v}_t, c), .., Q_{\overline{B}}^t(\boldsymbol{v}_t, c))$ and $\boldsymbol{M}_B^t(\boldsymbol{v}_t, c) = (M_1^t(\boldsymbol{v}_t, c), .., M_{\overline{B}}^t(\boldsymbol{v}_t, c))$.

For a given $(\boldsymbol{v}, c)$, let

$$\boldsymbol{Q}_S(\boldsymbol{v}, c) = \left(\boldsymbol{Q}_S^t(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty} \quad \text{and} \quad \boldsymbol{Q}_B(\boldsymbol{v}, c) = \left(\boldsymbol{Q}_B^t(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty}$$

and

$$\boldsymbol{M}_S(\boldsymbol{v}, c) = \left(M_S^t(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty} \quad \text{and} \quad \boldsymbol{M}_B(\boldsymbol{v}, c) = \left(\boldsymbol{M}_B^t(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty}.$$

Letting $\boldsymbol{Q}$ and $\boldsymbol{M}$ be, respectively, collections $\{\boldsymbol{Q}_S(\boldsymbol{v}, c), \boldsymbol{Q}_B(\boldsymbol{v}, c)\}$ and $\{\boldsymbol{M}_S(\boldsymbol{v}, c), \boldsymbol{M}_B(\boldsymbol{v}, c)\}$ for all possible $(\boldsymbol{v}, c)$, a direct mechanism is summarized by $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ where $\boldsymbol{Q}$ satisfies (20). It is said to satisfy interim individual rationality and incentive compatibility if it satisfies these constraints for every possible type of every agent who participates at the period the agent first participates in the mechanism. For buyers, the latter condition is vacuously satisfied because they participate in the mechanism in one period only, if they participate at all. For the seller it means that these constraints have to be satisfied at date 0 only.

The focus on direct mechanisms is now easily seen to be without loss of generality: In every period t, no mechanism that respects buyers' incentive and interim individual rationality constraints can do better than a direct mechanism that respects these constraints (see e.g. Krishna, 2002). Applied iteratively, this then implies that no incentive compatible and interim individually rational mechanism can do better than an incentive compatible and interim individually rational mechanism that asks the seller to report his type in period 0.

The analysis is greatly simplified by using two concepts. The first is the *ultimate probability of selling* for a seller who reports type c

$$q_S(c) := E_{\boldsymbol{v}} \left[\sum_{t=0}^{\infty} Q_S^t(\boldsymbol{v}_t, c) \prod_{\tau=0}^{t-1} \delta(1 - Q_S^\tau(\boldsymbol{v}_\tau, c)) \right],$$

which was introduced in Satterthwaite and Shneyerov (2008). We introduce a second, novel concept, the *ultimate conditional expected revenue*, which we will describe later.[83]

The seller's expected discounted payment is

$$m_S(c) := E_{\boldsymbol{v}} \left[\sum_{t=0}^{\infty} M_S^t(\boldsymbol{v}_t, c) \prod_{\tau=0}^{t-1} \delta(1 - Q_S^\tau(\boldsymbol{v}_\tau, c)) \right].$$

[83]Satterthwaite and Shneyerov (2008) did not need this second concept, since their analysis is mostly about a full-trade equilibrium, in which all sellers trade with probability 1.

In a direct mechanism, the seller of type c who reports truthfully has thus an expected discounted payoff of

$$\mathcal{W}_S(c) = m_S(c) - q_S(c)c,$$

while the intermediary's expected discounted payoff when facing a seller who reports to be of type c is

$$\mathcal{W}_I(c) = E_{\boldsymbol{v}}\left[\sum_{t=0}^{\infty}\left(\sum_{b=1}^{\overline{B}} M_b^t(\boldsymbol{v}_t, c)\right)\prod_{\tau=0}^{t-1}\delta(1 - Q_S^\tau(\boldsymbol{v}_\tau, c))\right] - m_S(c),$$

where the notation $\mathcal{W}_i(c)$ for $i = I, S$ emphasizes that we are referring to payoffs in a direct mechanism as opposed to fee-setting as defined in Section 2. The natural extension of the objective function (1) to the general mechanism design setup is then

$$\max_{\langle \boldsymbol{Q},\boldsymbol{M}\rangle} E_c[\alpha\mathcal{W}_I(c) + (1-\alpha)(\mathcal{W}_I(c) + \mathcal{W}_S(c))] \tag{21}$$

subject to incentive compatibility and interim individual rationality constraints of buyers and the seller. As there is no other restriction on the mechanisms used, this objective is more general than (1), which is confined to fee-setting. However, as we will show, the objective in (21) can be maximized with an appropriately chosen sequence of transaction fees $\boldsymbol{\omega}$. Moreover, we will show that individual rationality constraints are not only satisfied in the interim stage but also *ex post* and that the seller's incentive constraint can be satisfied period by period. While the results concerning *ex post* individual rationality of buyers is immediate because of the nature of second-price auctions, it is far from obvious *a priori* that such a mechanism exists in the dynamic setup with two-sided private information and arbitrary α we study.[84]

Standard arguments imply that a direct mechanism is incentive compatible for the seller if and only if it such that $q_S(c)$ is monotone in c and that in any direct, incentive compatible mechanism

$$m_S(c) = q_S(c)c + \int_c^{\overline{c}} q_S(x)dc + \mathcal{W}_S(\overline{c}) \tag{22}$$

[84] The direct mechanism problem that we set up here is thus a relaxed problem, and we will show that the additional constraints are not binding. For an analysis of *ex post* individual rationality of a bilateral trade problem where the intermediary makes zero profit, see Gresik (1991).

holds (see e.g. Krishna, 2002). Monotonicity of $q_S(c)$ implies that the interim individual rationality constraint will be satisfied if it is satisfied for the seller of type $\bar{c}$, that is if $\mathcal{W}_S(\bar{c}) \geq 0$ (and if the seller's incentive constraint is satisfied). Because $\mathcal{W}_S(\bar{c})$ enters the objective function as the constant $-\alpha\mathcal{W}_S(\bar{c})$, it will be optimal to set $\mathcal{W}_S(\bar{c}) = 0$ for any $\alpha \in [0, 1]$.

We say that the good is auctioned off in period t with reserve p_t if $Q_b^t(\boldsymbol{v}_t, c) = 1$ if $v_b = \max\{\boldsymbol{v}_t\}$ and $v_b \geq p_t$ and $Q_i^t(\boldsymbol{v}_t, c) = 0$ for all $i = 1, .., \overline{B}$ otherwise.[85]

Lemma 3. *A mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ is optimal only if the good is auctioned off in every period t at some reserve p_t.*

Proof of Lemma 3. Suppose to the contrary that the optimal mechanism, denoted $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$, does not auction off the good at some reserve in period t and in some states $(\boldsymbol{v}_t, c)$. This implies that with positive probability the good is sold in period t to a buyer whose valuation is not the highest amongst all the buyers present. Consider then an alternative mechanism that coincides with $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ except for the states in period t in which $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ does not auction off the good. Let the alternative mechanism sell the good to the highest value buyer in all those instances for which $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ sells it to some other buyer. This alternative mechanism will increase the broker's payoff $\mathcal{W}_I(c)$ by increasing the revenue it raises while leaving the seller's payoff $\mathcal{W}_S(c)$ unaffected. Therefore, the mechanism $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ cannot be optimal. □

Lemma 3 implies that the choice set for allocation rules $\boldsymbol{Q}$ can be narrowed down to sequences of reserves $\boldsymbol{p}(c) = (p_t(c))_{t=0}^{\infty}$, one sequence for every seller type c, with the understanding that, provided the mechanism is still active in period t, the good will be sold to the buyer with the highest valuation present in that period, provided this valuation is no less than $p_t(c)$. Letting $k_t := R(p_t)$ denote the expected transaction price in period t, conditional on a transaction occurring in period t, choosing a sequence of reserves $\boldsymbol{p}(c)$ is equivalent to choosing a sequence $\boldsymbol{k}(c) = (k_t(c))_{t=0}^{\infty}$ of expected transaction prices (conditional on a transaction occurring).

[85] This definition neglects the possibility of ties at the highest value, which have probability 0. If one wants to account for such ties explicitly, one can arbitrarily set $Q_b^t(\boldsymbol{v}_t, c) = 1$ for the buyer b with the highest valuation and, say, the highest index b amongst all buyers with the highest value.

The key observations are the following. For any sequence of expected transaction prices $\boldsymbol{k} = (k_t)_{t=0}^{\infty}$, let

$$q_t(\boldsymbol{k}) := \big(1 - F_{(1)}(R^{-1}(k_t))\big) \prod_{\tau=0}^{t-1} \delta F_{(1)}(R^{-1}(k_\tau))$$

denote the discount factor, adjusted for the probability of prior sale (and for the probability of prior exit if δ is interpreted as probability of survival), for a transaction occurring in period t. The number $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t$ is then the expected discounted transaction price, or average price for short, given $\boldsymbol{k}$ while $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})$ is the ultimate probability of selling. The number

$$k = \frac{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t}{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})} \tag{23}$$

has then the interpretation of an *ultimate conditional expected revenue*. The simplest interpretation for k can be provided if one interprets $1-\delta$ purely as the probability of dropping out from one period to the next (without any impatience on top of that). With this interpretation, k is the expected revenue conditional on trading and not dropping out. If impatience stems from a time preference rather than a drop-out probability, k cannot be simply interpreted as a conditional expectation, but should be viewed as an abstract mathematical concept that helps to unite probabilities and discounting and simplifies calculations.[86]

A mechanism can only be optimal if it maximizes the ultimate conditional expected revenue $(\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t)/(\sum_{t=0}^{\infty} q_t(\boldsymbol{k}))$ for a given ultimate probability of selling $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})$. By a duality argument, it also holds that a mechanism can only be optimal if it maximizes the ultimate probability of selling for a given ultimate conditional expected revenue.

Proof of Lemma 1. For $k \in [\underline{v}, \overline{v}]$ let

$$1 - \overline{F}_T(k) := \max_{(k_t)_{t=0}^{T}} \left\{ \sum_{t=0}^{T} q_t(\boldsymbol{k}) \right\} \quad \text{s.t.} \quad \frac{\sum_{t=0}^{T} q_t(\boldsymbol{k})k_t}{\sum_{t=0}^{T} q_t(\boldsymbol{k})} = k,$$

[86] One of the advantages of using the ultimate conditional expected revenue is that it avoids the problem of a standard conditional expected revenue with a time preference interpretation of discounting: for any positive constant per period probability of sale, the seller eventually trades with probability 1, so that conditioning on trade occurring would not be a useful concept.

and define

$$1-\overline{F}(k) := \lim_{T\to\infty} 1-\overline{F}_T(k).$$

Let $(k_t^*(k))_{t=0}^T$ be a maximizer of $\max_{(k_t)_{t=0}^T}\left\{\sum_{t=0}^T q_t(\boldsymbol{k})\right\}$ s.t. $\frac{\sum_{t=0}^T q_t(\boldsymbol{k})k_t}{\sum_{t=0}^T q_t(\boldsymbol{k})} = k$. Under stationarity, we have $k_t^*(k)=k$ for all t. Therefore,

$$\begin{aligned}
1-\overline{F}(k) &= \lim_{T\to\infty}\sum_{t=0}^T\left(\prod_{\tau=0}^{t-1}\delta F_{(1)}(R^{-1}(k))\right)\left(1-F_{(1)}(R^{-1}(k))\right)\\
&= \lim_{T\to\infty}\frac{1-\delta^{T+1}F_{(1)}(R^{-1}(k))^{T+1}}{1-\delta F_{(1)}(R^{-1}(k))}\left(1-F_{(1)}(R^{-1}(k))\right)\\
&= \frac{1-F_{(1)}(R^{-1}(k))}{1-\delta F_{(1)}(R^{-1}(k))} = 1-F_\infty(R^{-1}(k)).
\end{aligned}$$

Therefore, for a given c and k, we now have $\mathcal{W}_I(c) = k(c)(1-\overline{F}(k(c))) - m_S(c)$ and $\mathcal{W}_S(c) = m_S(c) - q_S(c)c$. Using incentive compatibility (22) and $\mathcal{W}_S(\overline{c}) = 0$ by individual rationality, the objective given c becomes

$$\alpha\mathcal{W}_I(c) + (1-\alpha)(\mathcal{W}_I(c)+\mathcal{W}_S(c)) = k(1-\overline{F}(k)) - cq_S(c) - \alpha\int_c^{\overline{c}} q_S(x)dx.$$

Substituting $q_S(k) = 1-\overline{F}(k)$ and integrating after reversing the order of integration in the double-integral then yields the objective function

$$\max_{k(c)}\int_{\underline{c}}^{\overline{c}}\left[k(c)-\Gamma_\alpha(c)\right](1-\overline{F}(k(c)))g(c)dc \tag{24}$$

with

$$\Gamma_\alpha(c) := c + \alpha\frac{G(c)}{g(c)}.$$

Observe that monotonicity of $\Gamma(c)$ implies monotonicity of $\Gamma_\alpha(c)$. The integral can be maximized pointwise by choosing k such that

$$0 = -\overline{f}(k(c))\left[\overline{\Phi}(k(c)) - \Gamma_\alpha(c)\right],$$

which is equivalent to $k(c) = \overline{\Phi}^{-1}(\Gamma_\alpha(c))$. This is a monotone function and thus incentive compatible. Moreover, the second-order condition for a maximum is satisfied whenever the first-order condition is satisfied if $\overline{\Phi}(v)$ is monotone.

This means that the optimal allocation rule is such that trade takes place as soon as

$$\overline{\Phi}(k) \geq \Gamma_\alpha(c). \tag{25}$$

Let $k^*(c) := \overline{\Phi}^{-1}(\Gamma_\alpha(c))$. Since $\overline{F}(k) = F_\infty(R^{-1}(k))$, $\overline{\Phi}(k) = \tilde{\Phi}(R^{-1}(k))$. According to (25) trade should take place as soon as $v \geq R^{-1}(k^*(c))$, which because of the afore-noted equalities and the monotonicity of $\tilde{\Phi}$, is equivalent to $\tilde{\Phi}(v) \geq \Gamma_\alpha(c)$ as claimed in the lemma. □

B Background for Empirical Analysis

B.1 Using Annual Data from April 1993 to April 1996

A calendar year appears to be the appropriate choice for the time interval on the ground that the standard deviation of the price index within a year between 1993 and 1995 is much smaller than the standard deviation of the quality adjusted price. A further indicator is a measure stemming from a widely documented stylized fact in real-estate markets: the correlation between price and time on market is weakly positive in cross-sectional data and negative in longitudinal data.[87] An intuitive explanation for the former observation is that expensive houses need more time to sell. An explanation for the latter is that in times of booms, houses sell faster and at higher prices. Time on market increases in quality adjusted price for the years we include in our estimation. Further, the relative change of the real-estate price index is generally small for the years considered. Table 4 shows the variance of the price index divided by its mean squared, V_I, within the time interval, the variance of quality adjusted prices, V_P, divided by the square of the mean in a given time interval, the ratio of these mean-adjusted variances V_I/V_P, the relative change of the price index Δ, and the slope of a regression of the time on market on the quality adjusted price. Figure 4 displays the movement of the real-estate price index. The table and the figure suggest that for the time period 1993 to 1995, a time interval of one year appears to be sufficiently short to be considered cross-

[87] Kang and Gardner (1989) provide empirical evidence that time on market increases with price in cross-sectional data – both based on their own dataset and on a review of other empirical work. Similar findings are reported in Glower, Haurin, and Hendershott (1998) and Genesove and Mayer (1997, 2001). The empirical literature typically finds a negative correlation between prices and vacancies – the latter can be seen as a proxy for time-on-market. Quigley (1999) investigates the effect of economic cycles on the housing market using international data on housing. He finds a negative correlation between vacancies and prices. See also the overview about empirical findings on the relation of vacancies and prices provided in Wheaton (1990).

sectional. Excluding data for the years before 1993 and after 1995 has the additional advantage of avoiding truncation issues.

Cross-sectional and intertemporal variation					
Year	V_I/mean^2 Price Index	V_P/mean^2 QAP	$\frac{V_I/mean^2}{V_P/mean^2}$	Relative Δ_I	Slope of Time on Market Qual. adj. Price Coeff. (Std. Err.)
1990	0.012	0.0396	0.31	-0.241	-87.6 (28.2)
1991	0.00044	0.0416	0.011	-0.030	39.0 (20.9)
1992	0.0026	0.0458	0.057	0.131	67.0 (21.5)
1993	0.00026	0.0412	0.0062	0.006	21.3 (22.0)
1994	0.00040	0.0342	0.012	-0.003	19.2 (22.0)
1995	0.00068	0.0361	0.019	0.067	40.1 (20.2)
1996	0.0017	0.0416	0.040	0.097	-7.6 (15.6)
1997	0.00040	0.0404	0.0099	0.044	7.2 (9.8)

Table 4: Variance V_I over squared mean of the real-estate price index in a given time interval, variance V_P over mean squared of the quality adjusted price in a given time interval, the ratio V_I/V_P, the change of the price index divided by the price index at the beginning of the interval Δ_I, and the slope β_1 of time on market T on quality adjusted price P in the regression $T = \beta_0 + \beta_1 P + \epsilon$.

The variation in quality adjusted prices is a measure of cross-sectional variation while the changes in the price index are related to intertemporal variation because the index uses quarterly data. For the three years considered, the cross-sectional variation exceeds the intertemporal variation by a factor of roughly fifty.

B.1.1 Formal Derivation of the Likelihood Function

We denote by $g_p(p) := g(P^{-1}(p))[P^{-1}(p)]'$ the density of reserve prices, where g is the density associated with the sellers' cost distribution $G(c)$ and $P^{-1}(p)$ is the inverse of the optimal pricing function $P^*(c) = \tilde{\Phi}^{-1}(c/0.94)$ derived in Section 2. Given reserve p the probability mass function for time on the market t with $t = 0, .., \infty$ is then given, if the object is sold, as $[\delta F_{(1)}(p)]^t(1 - F_{(1)}(p))$ and for objects that do not sell as $[\delta F_{(1)}(p)]^t(1 - \delta)F_{(1)}(p)$. We denote by $h_{tpS}(t, p, S)$ the likelihood function for observing (t, p, S). This

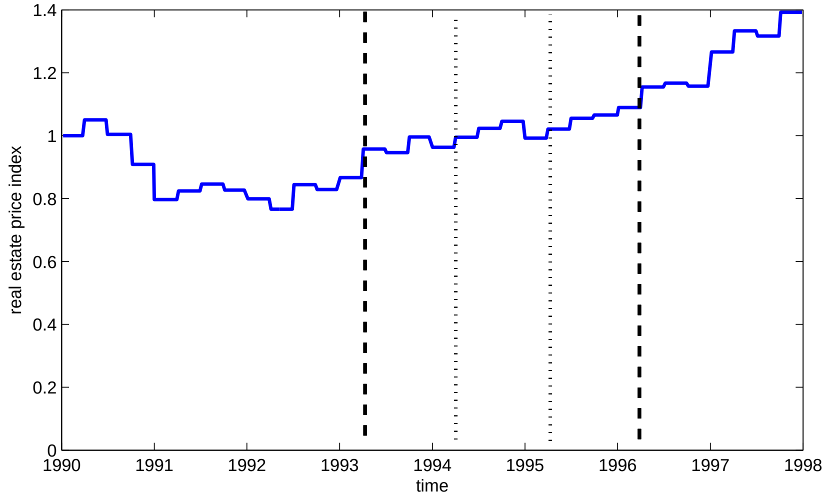


Figure 4: Development of the real-estate price index in Boston. Data between the two dashed lines were used for estimations.

function is given as

$$h_{tpS}(t,p,S) = \begin{cases} [\delta F_{(1)}(p)]^t(1-F_{(1)}(p))g_p(p) & \text{if } S=1 \\ [\delta F_{(1)}(p)]^t(1-\delta)F_{(1)}(p)g_p(p) & \text{if } S=0 \end{cases}.$$

Next we consider transaction prices $\breve{p}$. Denote by $\breve{f}(\breve{p}|p)$ the density of the transaction price $\breve{p}$ given the reserve price p. This density can be written as

$$\breve{f}(\breve{p}|p) = \begin{cases} 0 & \text{if } \breve{p} < p \\ \frac{F_{(2)}(p)-F_{(1)}(p)}{1-F_{(1)}(p)}\Delta(\breve{p}-p) & \text{if } \breve{p} = p \\ \frac{f_{(2)}(\breve{p})}{1-F_{(1)}(p)} & \text{if } \breve{p} > p \end{cases},$$

where $\Delta(.)$ is the Dirac delta-function. Denote by $h_{\breve{p}}(\breve{p}|p,S)$ the likelihood of observing $\breve{p}$ given (p,S) and set $\breve{p}=0$ if $S=0$. Under stationarity $h_{\breve{p}}(\breve{p}|p,S)$ is independent of t. By the previous arguments,

$$h_{\breve{p}}(\breve{p}|p,S) = \begin{cases} \breve{f}(\breve{p}|p) & \text{if } S=1 \\ \Delta(\breve{p}) & \text{if } S=0 \end{cases}.$$

For a given observation $\boldsymbol{X}_i = (t_i, p_i, S_i, \breve{p}_i)$, the likelihood function absent unobservable heterogeneity in quality, the discount noise, and measurement error in time on the market

$l(\boldsymbol{X}_i|\boldsymbol{\theta})$ would be given as

$$l(\boldsymbol{X}_i|\boldsymbol{\theta}) = h_{tpS}(t_i, p_i, S_i)h_{\breve{p}}(\breve{p}_i|p_i, S_i),$$

where $\boldsymbol{\theta}$ is the vector of parameters determining the shapes of h_{tpS} and $h_{\breve{p}S}$.

The likelihood function $l(\boldsymbol{X}_i|\boldsymbol{\theta})$ for an observation $\boldsymbol{X}_i = (T_i, P_i, S_i, \breve{P}_i)$ that accounts for unobservable heterogeneity and measurement error in time on the market is then given as

$$l(\boldsymbol{X}_i|\boldsymbol{\theta}) = \sum_{k=1}^{T_i/\tau} \int_0^\infty \int_0^\infty h_{tpS}(T_i - k\tau, P_i/\epsilon^P, S_i)h_{\breve{p}}(\breve{P}_i/\epsilon^Q|P_i/\epsilon^P, S_i)h_t(k\tau)h_Q(\epsilon^Q)h_D(\epsilon^D)d\epsilon^Q d\epsilon^D, \tag{26}$$

where $\breve{P}_i$ is the constructed transaction price and $h_j(\epsilon^j)$ is the density of the error term ϵ^j with $j \in \{Q, D\}$ and where k is the summation variable and $\epsilon^T = k\tau$. It will be convenient for our estimation to rewrite (26) in terms of ϵ^D, ϵ^P using $\epsilon^Q = \epsilon^P/\epsilon^D$:

$$l(\boldsymbol{X}_i|\boldsymbol{\theta}) = \sum_{k=1}^{T_i/\tau} \int_0^\infty \int_0^\infty h_{tpS}(T_i - k\tau, P_i/\epsilon^P, S_i)h_{\breve{p}}(\breve{P}_i/(\epsilon^P/\epsilon^D)|P_i/\epsilon^P, S_i)h_t(k\tau)h_P(\epsilon^P)h_D(\epsilon^D)d\epsilon^D d\epsilon^P,$$

where $h_P(\epsilon^P)$ is the density of ϵ^P.

B.2 Identification

First, consider the case without unobserved heterogeneity, discount noise, and measurement error in the time on market, that is when ϵ^Q and ϵ^D are constant, and ϵ^T is zero. Rearranging (10) and (2), the expressions for time on market as a function of the quality-adjusted price $T(p)$ and for the probability of ever selling $1 - F_\infty(p)$, we obtain

$$F_{(1)}(p) = 1 - \frac{1 - F_\infty(p)}{T(p)/\tau} \quad \text{and} \quad \delta = \frac{T(p_2) - T(p_1)}{T(p_2)(1 - F_\infty(p_1)) - T(p_1)(1 - F_\infty(p_2))},$$

and

$$\tau = \frac{T(p_1)F_\infty(p_2) - T(p_2)F_\infty(p_1)}{F_\infty(p_2) - F_\infty(p_1)},$$

where p_1 and p_2 are two arbitrary prices (or – with some modification of the equations – price segments). This makes $F_{(1)}(p)$, the discount factor δ, and the period length τ non-parametrically identifiable over the range of prices that are set in equilibrium. Given

$F_{(1)}(p)$ and the observable distribution of reserve prices $G_p(p)$, the sellers' cost distribution is non-parametrically identifiable via the relationship $G(c) = G_p(\tilde{\Phi}^{-1}(c/0.94))$, where $\tilde{\Phi}^{-1}(c/0.94)$ is the optimal price a seller of type c sets when facing a 6 percent fee.[88]

With unobservable heterogeneity and measurement errors in the time on market ϵ^T, the argument is more involved, but identification is still possible. Redefine the observed probability of ever selling $1 - F_\infty(P_i/\vartheta_i)$ and the time on market of sold and unsold houses $T^s(P_i/\vartheta_i)$ and $T^u(P_i/\vartheta_i)$ as functions of the observed quality adjusted listing price P_i/ϑ_i rather than the true (quality- and discount-adjusted) reserve price p_i. Further, let $\hat{T}^s(P_i/\vartheta_i) := T^s(P_i/\vartheta_i) + E[\epsilon^T]$ and $\hat{T}^u(P_i/\vartheta_i) := T^u(P_i/\vartheta_i) + E[\epsilon^T]$ be the observed average times on market. The probability of ever selling given P_i/ϑ_i and ϵ_i^P is $\text{Prob}(s = 1|P_i/\vartheta_i, \epsilon_i^P) = (1-F_{(1)}(P_i/(\vartheta_i\epsilon_i^P)))/(1-\delta F_{(1)}(P_i/(\vartheta_i\epsilon_i^P)))$ and the probability of never selling $\text{Prob}(s = 0|P_i/\vartheta_i, \epsilon_i^P) = (1-\delta)F_{(1)}(P_i/(\vartheta_i\epsilon_i^P))/(1-\delta F_{(1)}(P_i/(\vartheta_i\epsilon_i^P)))$. Given the unconditional density $h_p(\epsilon^P)$, the conditional densities are $h_p(\epsilon^P|P_i/\vartheta_i, s = 1) \propto h_p(\epsilon^P)/\text{Prob}(s = 1|P_i/\vartheta_i, \epsilon^P)$ and $h_p(\epsilon^P|P_i/\vartheta_i, s = 0) \propto h_p(\epsilon^P)/\text{Prob}(s = 0|P_i/\vartheta_i, \epsilon^P)$ by Bayes' Law. This gives us

$$
\begin{aligned}
1 - F_\infty(P_i/\vartheta_i) &= E_{\epsilon_i^P \sim H_p}\left[\frac{1 - F_{(1)}(P_i/(\vartheta_i\epsilon_i^P))}{1 - \delta F_{(1)}(P_i/(\vartheta_i\epsilon_i^P))}\right], \\
\hat{T}^s(P_i/\vartheta_i) &= E_{\epsilon_i^P \sim H_p(\cdot|P_i/\vartheta_i, s=1)}\left[\frac{\tau}{1 - \delta F_{(1)}(P_i/(\vartheta_i\epsilon_i^P))}\right] + E[\epsilon^T], \\
\hat{T}^u(P_i/\vartheta_i) &= E_{\epsilon_i^P \sim H_p(\cdot|P_i/\vartheta_i, s=0)}\left[\frac{\tau}{1 - \delta F_{(1)}(P_i/(\vartheta_i\epsilon_i^P))}\right] + E[\epsilon^T]
\end{aligned}
\tag{27}
$$

Note that (27) and $\hat{T}^u(P_i/\vartheta_i) - \hat{T}^s(P_i/\vartheta_i)$ do not require any knowledge about the distributions of ϵ^T and c and identify $F_{(1)}$ and H_p for given ϵ and τ. The density of time on market t conditional on a particular price P_i/ϑ_i, $E_{\epsilon^P \sim H_p, \epsilon^T \sim H_t}[(\delta F_{(1)}(P_i/(\vartheta_i\epsilon^P)))^{t/\tau - \epsilon^T}|\epsilon^T \leq t/\tau]$ identifies H_t. This uses only one price P_i/ϑ_i. The different densities of t for two additional prices P_j/ϑ_j and P_l/ϑ_l pin down ϵ and τ. The sellers' distribution $G(c)$ is then non-parametrically identifiable in the same way as discussed in the text via the relationship $G(c) = G_p(\tilde{\Phi}^{-1}(c/0.94))$.

[88]In Appendix B.4, we show that our estimation is robust to the alternative assumptions that brokers' fees are, maybe because of unobserved discounts, 5.5% or 5%.

Equation (9) means that we can express $F(p)$ as $F(p) = 1 + \ln(F_{(1)}(p))/\xi$, which allows us to write the virtual valuation function as

$$\Phi(v) = v + \frac{F_{(1)}(v)}{f_{(1)}(v)} \ln(F_{(1)}(v)).$$

This relates $F_{(1)}$ to Φ without knowledge of ξ, which means that we cannot identify ξ from Φ. Intuitively, by observing the distribution of the highest bid $F_{(1)}$ one cannot tell the extent to which it is driven by F or by the arrival rate ξ. Note that this is not a problem, it turns out that all the counterfactuals also only depend on $F_{(1)}$ and not on ξ and F separately. Further note that in our estimation ξ and F will be *parametrically* identified.

It is worthwhile mentioning that while the above reasoning is the most elegant identification proof we could find, it is not necessarily the best estimation technique, since it essentially relies on the fact that even if we "throw away" data, the underlying distributions are still identifiable. We throw away data e.g. by using the expected time on market $T(p)$ rather than the distribution of the time on market conditional on p and by not using the distribution of transaction prices $\breve{p}$. For finite sample sizes, it is preferable to make use of all the data, which we do with our Bayesian estimation.

B.3 Backing-Out the Bargaining Parameter α

In the following, we describe the procedure to back-out α. Letting

$$\begin{aligned} W_I(b) &= bE_c[\tilde{\Phi}^{-1}(c/(1-b))(F_{(2)}(\tilde{\Phi}^{-1}(c/(1-b))) - F_{(1)}(\tilde{\Phi}^{-1}(c/(1-b)))) \\ &+ \int_{(\tilde{\Phi}^{-1}(c/(1-b)))}^{\overline{v}} y f_{(2)}(y) dy] \end{aligned}$$

be the expected payoff to an intermediary given the percentage fee b,

$$\begin{aligned} W_S(b) &= (1-b)E_c[\tilde{\Phi}^{-1}(c/(1-b))(F_{(2)}(\tilde{\Phi}^{-1}(c/(1-b))) - F_{(1)}(\tilde{\Phi}^{-1}(c/(1-b)))) \\ &+ \int_{(\tilde{\Phi}^{-1}(c/(1-b)))}^{\overline{v}} y f_{(2)}(y) dy - c/(1-b)(F_{(1)}(\tilde{\Phi}^{-1}(c/(1-b))))] \end{aligned}$$

be the expected profit of the seller and $W(\alpha, b) = \alpha W_I(b) + (1-\alpha)(W_I(b) + W_S(b))$, α can be obtained by solving $\alpha W_I'(b) + (1-\alpha)(W_I'(b) + W_S'(b)) = 0$, for α yielding

$$\alpha(b) = \frac{W_I'(b) + W_S'(b)}{W_S'(b)}$$

as the value of α consistent with the fee b

B.4 Robustness Checks

Using Physical Characteristics for Estimation As a robustness check of our estimation using previous transaction prices as a proxy for the quality adjusted price, we also estimate the model using physical characteristics as a proxy for quality. Table 5 shows that the parameters estimates are basically the same.

Estimated Parameter Values						
Parameters	1993		1994		1995	
ϕ_1	1.15	(0.106)	1.11	(0.126)	1.07	(0.104)
ϕ_2	-0.0596	(0.0420)	-0.0171	(0.0468)	-0.0209	(0.0328)
γ_1	1.27	(0.0326)	1.25	(0.0274)	1.20	(0.0201)
γ_2	0.134	(0.0130)	0.138	(0.0109)	0.140	(0.00609)
$\underline{x}$	0.0168	(0.0162)	0.0155	(0.0149)	0.0123	(0.0120)
$\overline{x}$	1.30	(0.0265)	1.29	(0.0216)	1.23	(0.0134)
δ	0.945	(0.0113)	0.966	(0.0133)	0.984	(0.00112)
ξ	0.136	(0.0978)	0.323	(0.169)	0.175	(0.0896)
μ_D	0.0952	(0.00363)	0.0833	(0.00331)	0.0735	(0.00235)
σ_D	0.0746	(0.00258)	0.0738	(0.00231)	0.0572	(0.00166)
σ_P	0.243	(0.00721)	0.201	(0.00550)	0.191	(0.00512)
β_T	0.0705	(0.0182)	0.143	(0.0486)	0.385	(0.0575)
# Observations	736		840		793	

Table 5: Estimated Parameter Values for 1993 to 1995 using Physical Characteristics. Table entries read: Mean (Standard Deviation).

Estimation Using 5% and 5.5% Fees In the main text, we estimate the distribution of sellers' costs using the observed distribution of reserve prices $G_p(p)$ and assuming $p = \tilde{\Phi}^{-1}(c/0.94)$, which is the optimal price for a seller of type c to set if the brokers employ 6% fees. While there is strong evidence that 6% are the empirically relevant fees (see e.g. Hsieh and Moretti (2003)), it is also important to check the robustness of our estimates if we assumed that brokers use fees with 5% or 5.5%, perhaps because of unobserved discounts. In this appendix, we show that our estimation is robust in these regards by estimating $G(c)$ using $G_p(p)$ and assuming $p = \tilde{\Phi}^{-1}(c/0.945)$ and $p = \tilde{\Phi}^{-1}(c/0.95)$, respectively. Tables 6 and 7 give the parameter estimates for the private values model when fees are 5.5% and 5%, respectively, and unobserved heterogeneity is controlled for using previous transaction prices. The estimates are almost the same as for the model with 6% fees in the main text (see Tables 6 and 7).

Estimated Parameter Values						
Parameters	1993		1994		1995	
ϕ_1	1.14	(0.0980)	1.11	(0.114)	1.02	(0.0995)
ϕ_2	-0.0724	(0.0395)	-0.0346	(0.0463)	-0.0207	(0.0315)
γ_1	1.24	(0.0353)	1.23	(0.0318)	1.15	(0.0208)
γ_2	0.133	(0.0125)	0.135	(0.0108)	0.134	(0.00588)
$\underline{x}$	0.0181	(0.0176)	0.0167	(0.0161)	0.0123	(0.0120)
$\overline{x}$	1.28	(0.0281)	1.26	(0.0260)	1.18	(0.0144)
δ	0.946	(0.0108)	0.956	(0.0168)	0.984	(0.00114)
ξ	0.110	(0.0891)	0.244	(0.176)	0.170	(0.0880)
μ_D	0.0984	(0.00499)	0.0841	(0.00340)	0.0736	(0.00234)
σ_D	0.103	(0.00355)	0.0739	(0.00238)	0.0572	(0.00168)
σ_P	0.263	(0.00793)	0.236	(0.00650)	0.232	(0.00632)
β_T	0.0693	(0.0158)	0.116	(0.0517)	0.380	(0.0576)
# Observations	727		830		787	

Table 6: Estimated Parameter Values for 1993 to 1995 assuming 5.5% fees. Table entries read: Mean (Standard Deviation).

Other Years As an additional robustness check, we also estimated the model for the other years (using previous transaction prices and 6 percent fees as in the main model). The parameter estimates appear to be fairly robust in this regards as well, as shown by the results in Table 8. As show in Figure 4, real-estate prices declined substantively in 1990 and 1991 and experienced a strong increase in the last two years for which we have data. Because the estimated parameter values are roughly the same for all years, this suggests that there is little if any bias arising from non-stationarity. The notable variation in the price discount μ_D, and the way it differs from the years 1993 to 1995, is as one would expect because the discount is likely to be overestimated (underestimated)

Estimated Parameter Values						
Parameters	1993		1994		1995	
ϕ_1	1.16	(0.0980)	1.10	(0.119)	1.02	(0.101)
ϕ_2	-0.0839	(0.0437)	-0.0333	(0.0474)	-0.0206	(0.0317)
γ_1	1.25	(0.0405)	1.22	(0.0325)	1.15	(0.0213)
γ_2	0.134	(0.0129)	0.135	(0.0108)	0.134	(0.00603)
$\underline{x}$	0.0195	(0.0188)	0.0183	(0.0177)	0.0128	(0.0126)
$\overline{x}$	1.29	(0.0332)	1.26	(0.0258)	1.18	(0.0145)
δ	0.945	(0.00980)	0.956	(0.0168)	0.984	(0.00114)
ξ	0.129	(0.103)	0.241	(0.177)	0.171	(0.0891)
μ_D	0.0983	(0.00496)	0.0841	(0.00342)	0.0736	(0.00236)
σ_D	0.103	(0.00350)	0.0739	(0.00235)	0.0573	(0.00168)
σ_P	0.263	(0.00791)	0.236	(0.00662)	0.232	(0.00630)
β_T	0.0677	(0.0119)	0.117	(0.0519)	0.379	(0.0576)
# Observations	727		830		787	

Table 7: Estimated Parameter Values for 1993 to 1995 using 5% fees. Table entries read: Mean (Standard Deviation).

when the overall real-estate price index is decreasing (increasing).

C Additional Counterfactuals

C.1 Regulated Fees

We assume that the government can directly determine the percentage fee $\bar{b}$ brokers can charge, with the natural focus being on regulated fees $\bar{b} \leq 0.06$.[89] In the following we will conduct a counterfactual analysis based on a partial equilibrium analysis: we assume that buyers' valuation distributions do not change as the policy changes. This is certainly an imperfect approximation of reality: if e.g. the distribution of house prices changes, then buyers' option value of future trade is likely to change as well. There are

[89]Besides being the natural counterfactual, it also means that the optimal prices of all sellers are well-defined after a fee decrease even in scenarios that do not allow for entry (or exit). That said, when allowing for entry, we need to extrapolate because in the data we do not observe sellers with higher costs.

Estimated Parameter Values					
Parameters	1990	1991	1992	1996	1997
ϕ_1	1.17 (0.170)	1.24 (0.162)	1.23 (0.131)	0.992 (0.0983)	0.985 (0.106)
ϕ_2	0.0611 (0.0630)	−0.0192 (0.0616)	−0.0579 (0.0534)	−0.0163 (0.0313)	0.00198 (0.0370)
γ_1	1.39 (0.0453)	1.44 (0.0437)	1.38 (0.0441)	1.12 (0.0214)	1.11 (0.0283)
γ_2	0.141 (0.0212)	0.148 (0.0194)	0.141 (0.0175)	0.128 (0.00700)	0.119 (0.0113)
$\underline{x}$	0.0292 (0.0275)	0.0224 (0.0215)	0.0208 (0.0209)	0.0122 (0.0121)	0.0179 (0.0172)
$\overline{x}$	1.44 (0.0313)	1.49 (0.0355)	1.42 (0.0365)	1.15 (0.0153)	1.15 (0.0191)
δ	0.956 (0.00484)	0.950 (0.00766)	0.934 (0.0129)	0.979 (0.00151)	0.953 (0.00388)
ξ	0.755 (0.107)	0.476 (0.146)	0.233 (0.118)	0.169 (0.0845)	0.209 (0.0936)
μ_D	0.191 (0.0115)	0.169 (0.00916)	0.128 (0.00640)	0.0605 (0.00234)	0.0546 (0.00267)
σ_D	0.153 (0.00818)	0.161 (0.00657)	0.118 (0.00454)	0.0532 (0.00169)	0.0491 (0.00190)
σ_P	0.233 (0.00903)	0.283 (0.00932)	0.288 (0.00929)	0.248 (0.00723)	0.231 (0.00775)
β_T	0.0637 (0.00621)	0.0643 (0.00758)	0.0602 (0.0108)	0.657 (0.114)	0.650 (0.127)
# Observations	497	668	671	704	543

Table 8: Estimated Parameter Values for the years other than 1993 to 1995. Table entries read: Mean (Standard Deviation).

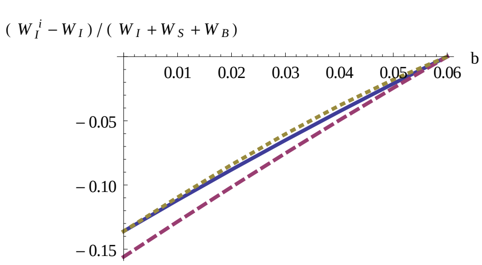
a number of reasons why we think that these counterfactuals simulations provide useful insights nonetheless. First, it turns out that the distribution of prices changes only by a small amount (which is in line with Proposition 5 and the arguments in its proof), so that the option value of future trade of buyers should only change moderately. Second, the seller's cost distribution seems to be relatively close to a mirrored Generalized Pareto distribution. For G mirrored Generalized Pareto, the optimal fee is independent of the buyers' distribution. Therefore, even if the buyers' distribution changed, the effect of this change on the optimal fee structure should be small. Third, one can interpret our counterfactual analysis as a change of regulation in a small (open) jurisdiction, in which buyers' option values are mostly determined by trade opportunities outside of the jurisdiction.

When so doing, it is instructive to decompose the effect of a decrease in fees into three effects: (1) a *direct effect* consisting of the "mechanical" price-adjustments that occur if sellers' net prices are kept the same, (2) an indirect *inframarginal seller effect*, which occurs because sellers who are active before and after the reduction in fees change their prices, and (3) an indirect *marginal seller effect*, which stems from the fact that additional sellers with high costs will enter the market. To distinguish between these three effects, we calculate average prices and welfare for three counterfactual scenarios that correspond to this decomposition.

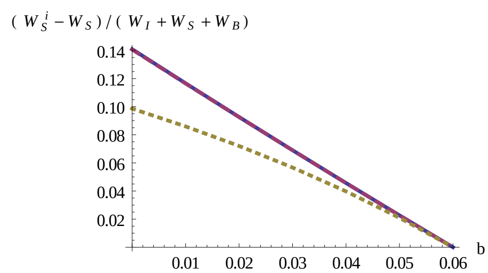
Let $p^i(c)$ be the reserve price a seller of type c sets in scenario $i \in \{1, 2, 3\}$ and let $\overline{c}^i$ be the least efficient seller type who enters the market in this scenario. Expected welfare of buyers, sellers, and intermediaries in scenario i is denoted, respectively, by W_B^i, W_S^i and W_I^i. The detailed derivations of these expressions are in Appendix C.3. The assumption underlying W_B^i is that buyers are short-lived and participate in one period only, so that changes in fees have no effect on their future payoffs.

The average reserve price EP^i in scenario i is given as $EP^i = \int_{\underline{c}}^{\overline{c}^i} p^i(c) \frac{dG(c)}{G(\overline{c}^i)}$. An analogous conditional expectation underlies the estimation of average prices of active sellers in the analysis Hendel, Nevo, and Ortalo-Magné (2009). Hendel, Nevo, and Ortalo-Magné (2009) compare average prices on a for-sale-by-owner platform that virtually has no fees to prices in the brokerage market. Quite surprisingly, they find that the average price in the market without fees is roughly the same as in the market with fees.[90] This appears puzzling, since one would expect higher fees to increase prices.

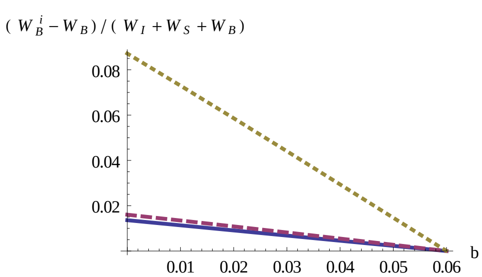
We write W_B, W_I, W_S and EP to denote the variables absent fee-regulation, that is, when $\overline{b} = 0.06$. Figure 5 plots the results, expressed as percentages, for 1993 for $\overline{b}$ between 0 and 0.06.



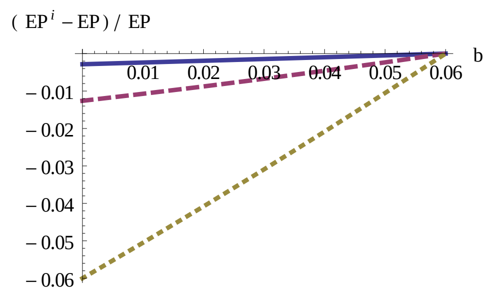
(a) Intermediary's profit W_I



(b) Seller's utility W_S



(c) Buyer's utility W_B



(d) Average price EP

Figure 5: The percentage change of different variables $(y^i - y)/y$ as a function of the regulated fee $\overline{b} \in [0, 0.06]$ for scenarios $i = 1$ (dotted), $i = 2$ (dashed), and $i = 3$ (solid) for $y \in \{W_I, W_S, W_B, EP\}$ for 1993.

[90] Hendel, Nevo, and Ortalo-Magné (2009) even find prices in the market without fees to be higher, even though the difference is not significant. See also Rutherford, Springer, and Yavas (2005) and Levitt and Syverson (2008) for two papers that compare houses of third-party sellers sold through brokers with brokers selling their own houses. A direct comparison with the findings of Hendel, Nevo, and Ortalo-Magné (2009) is difficult, since brokers selling their own houses are likely to be a different demographic group with a different distribution G. But still, these papers find that the prices obtained by brokers selling their own houses are not lower than for other houses. Of course, all three papers correct for a selection effect in terms of *the quality of properties*.

From Proposition 5 we know that if the sellers' costs are drawn from a mirrored Generalized Pareto distributions with $\underline{c} = 0$, then the direct effect (1) and the indirect inframarginal seller effect (2) on EP^3 are exactly offset by the marginal seller effect (3). Because empirically the distribution of the sellers' cost is not exactly mirrored Generalized Pareto, the extent to which these effects wash out empirically is an open question. Panel (d) in Figure 5 shows that a fee-decrease to $\bar{b} = 0$ would, by construction, reduce average prices by 6 percent if only the direct effect (1) is accounted for. Taking into account effects (1) and (2), the decrease in average prices is only about 1.25 percent. Accounting for all three effects, the total effect is that the average prices decrease by only 0.28 percent. This result provides an explanation for the puzzling observation made by Hendel, Nevo, and Ortalo-Magné (2009): the price increasing effect of higher fees is almost completely offset by the additional entry of high cost sellers.[91] In contrast, in panels (b) and (c) the difference between the dashed and solid lines, that is between scenarios 2 and 3, is almost negligible as far as agents' welfare is concerned. The difference between the dotted and the dashed line, that is between scenarios 1 and 2, is pronounced and shows a large and negative impact of the price-endogeneity effect on buyers' welfare and a positive effect of similar size on the sellers' welfare.

C.2 Transfer Taxes

We now provide an analysis of the equilibrium effects of such transfer taxes on consumer surplus, and the welfare of intermediaries and sellers, assuming that a small local government imposes a percentage transfer tax $\kappa \geq 0$.[92]

Again it proves useful to distinguish various layers through which the policy experiment can affect outcomes. Because we look at the effect of imposing a tax $\kappa > 0$ when

[91] Hendel, Nevo, and Ortalo-Magné (2009) (and also Rutherford, Springer, and Yavas (2005) and Levitt and Syverson (2008)) correct for a selection effect with respect to the characteristics of properties, but quite naturally, cannot correct for selection effects in the idiosyncratic valuations of sellers.

[92] The assumption of a percentage tax is imposed for analytical tractability and because of its practical relevance. For example, in the United States roughly one third of the states impose a percentage transfer tax.

there was none before, high cost sellers that absent the tax barely broke even would now make a loss if they did not adjust their prices. To circumnavigate this problem, the first scenario we consider is the exit of sellers. Before the tax was introduced, sellers with costs $c \leq (1-b)\overline{v}$ entered the market. After the tax is introduced, sellers enter only if their cost is below $\overline{c}^0 := (1-b)\overline{v}/(1+\kappa)$, which we refer to as scenario (0). Scenarios (1) and (2) are then analogous to (1) and (2) under fee regulation with the exception that scenario (2) is now subdivided into (2.a) – fees do not change – and (2.b) fees adjust optimally to the tax. In scenario (1), which corresponds to the analysis in Einav, Knoepfle, Levin, and Sundaresan (2014), sellers with costs $c \leq \overline{c}^0$ set the same reserve price $\tilde{\Phi}^{-1}(c/(1-b))$ as they did before the tax was imposed. Accordingly, the reserve price $p^1(c)$ bidders face in this scenario given a seller type c satisfies $p^1(c) = (1+\kappa)\tilde{\Phi}^{-1}(c/(1-b))$. Of course, the least efficient seller type who is active in scenario (1) has the same cost $\overline{c}^1 = (1-b)\overline{v}/(1+\kappa) = \overline{c}^0$ as in scenario (0).

In scenario (2.a), the fee is still b but now for given b and κ all seller-types set the optimal after-tax-pre-fee reserve price $p^{2.a}_{\kappa}(c) = \tilde{\Phi}^{-1}\left((1+\kappa)c/(1-b)\right)/(1+\kappa)$. Accordingly, the reserve price $p^{2.a}(c)$ bidders face with a seller of type c is $p^{2.a}(c) = \tilde{\Phi}^{-1}\left((1+\kappa)c/(1-b)\right)$ while the least efficient active seller type has the cost $\overline{c}^{2.a} = (1-b)\overline{v}/(1+\kappa) = \overline{c}^1 = \overline{c}^0$. Lastly, in scenario (2.b), the fee adjusts optimally to κ to induce a seller of type c to set the $p^{2.b}_{\kappa}(c) = \tilde{\Phi}^{-1}((1+\kappa)\Gamma_{\alpha}(c))/(1+\kappa)$ because the optimal allocation rule given a transfer tax κ is such that the seller of type c sells to the bidder with the highest virtual value if and only if this buyer's value exceeds $\tilde{\Phi}^{-1}\left((1+\kappa)\Gamma_{\alpha}(c)\right)$. Accordingly, the reserve price faced by buyers who are matched to a seller of type c is $p^{2.b}(c) = \tilde{\Phi}^{-1}\left((1+\kappa)\Gamma_{\alpha}(c)\right)$ and the least efficient active seller has the cost $\overline{c}^{2.b} = \Gamma_{\alpha}^{-1}(\overline{v}/(1+\kappa))$.

In all scenarios other than (2.b) the fees ω_i^{κ} are simple percentage fees with $b = 0.06$. In scenario (2.b) the fee $\omega^{\kappa}_{2.b}(p_{\kappa})$ is given by the expression in (31) in Appendix C.3, which also contains details on the other derivations. As in the analysis of regulated fees, we assume that buyers are short-lived and denote expected buyer surplus in scenario i by W_B^i the expression for which is identically to the one under regulated fees. As before, expected seller and intermediary welfare in scenario i are denoted W_S^i and W_I^i while

K^i denotes tax revenue in scenario i. These expression are derived in Appendix C.3. Variables without superscripts i denote the benchmark scenario when $\kappa = 0$. .

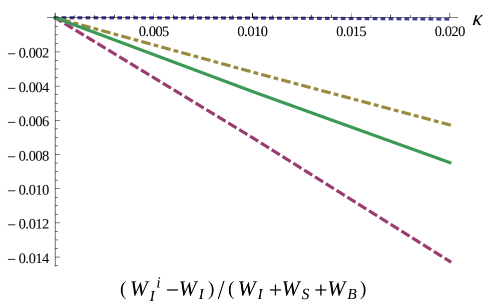
Figure 6 depicts the results for 1993. All agents – buyers, sellers, and intermediaries – are adversely affected by the introduction of the transfer tax in the small local jurisdiction as shown by the solid lines. Remarkably, sellers are most severely hit by the transfer tax and incur a decrease in welfare that is roughly ten times as large as are the corresponding decreases for buyers and intermediaries. All agents benefit from the additional adjustments that occur because sellers and intermediaries adjust their choice variables, as is shown in the difference between the solid line (total effect) and the dotted line (exit and mechanical price adjustment). Interestingly, for buyers (and the local government) the additional adjustment in fees (difference between scenarios 2a and 2b) has almost zero effect on their welfare while its effect on intermediaries and sellers is substantive with opposite signs. Thus, the first-order effect of the adjustment in fees is how the pie is split between the intermediaries and the sellers while the first-order effect of the sellers' adjustment of prices is to increase the size of the pie to be shared between the buyers, sellers, intermediaries, and the government.[93]

C.3 Background for Counterfactuals

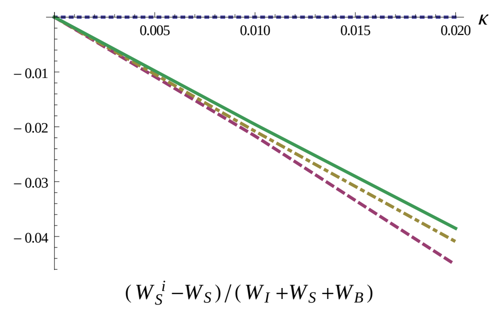
Regulated Fees Because $p(c) = \tilde{\Phi}^{-1}(c/0.94)$ is the reserve price sellers of type c set with a 6 percent fee, the reserve price the set in scenario 1, denoted $p^1(c)$, is $p^1(c) = 0.94\tilde{\Phi}^{-1}(c/0.94)/(1-\bar{b})$. In scenarios 2 and 3, sellers of type c set the price $p^i(c) = \tilde{\Phi}^{-1}(c/(1-\bar{b}))$ with $i \in \{2,3\}$. In scenarios 1 and 2, the least efficient active seller type has a cost equal to $\bar{c}^i = 0.94\bar{v}$ with $i = 1, 2$ while in scenario 3 the least efficient active seller has a cost of $\bar{c}^3 = \bar{v}/(1-\bar{b})$.

Assuming that buyers are short-lived and participate only for one period, the expected buyer surplus in any given period a seller of type c who sets the reserve price $p^i(c)$

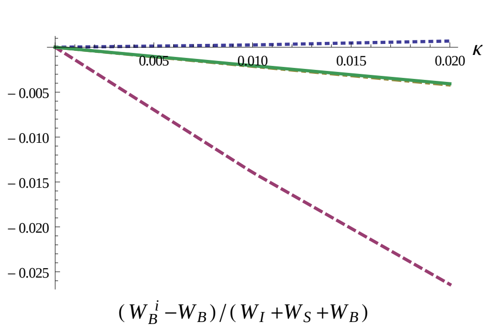
[93] That the adjustment of the fee – scenario 2b (solid) versus scenario 2a (dash-dotted) – is to the detriment of the intermediaries and to the benefit of the sellers is due to the fact that $\alpha^* \approx 0.08$ puts little weight on the intermediaries' welfare.



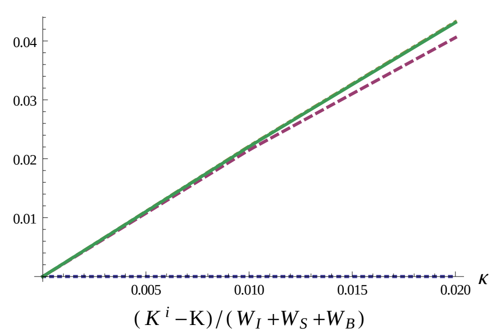
(a) Intermediary's profit W_I



(b) Seller's utility W_S



(c) Buyer's utility W_B



(d) Tax revenue K

Figure 6: The percentage change of different variables $(y^i - y)/y$ as a function of the transaction tax κ for scenarios $i = 0$ (dotted), $i = 1$ (dashed), $i = 2a$ (dash-dotted), and $i = 2b$ (solid) for $y \in \{W_I, W_S, W_B\}$. For variable K (tax revenues), the absolute normalized change is given, since the starting value of K is 0.

generates is

$$\int_{p^i(c)}^{\overline{v}} v[f_{(1)}(v) - f_{(2)}(v)]dv - \left(F_{(2)}(p^i(c)) - F_{(1)}(p^i(c))\right) p^i(c).$$

Because such a seller fails to trade in any given period with probability $F_{(1)}(p^i(c))$ and enters the subsequent period with probability $\delta F_{(1)}(p^i(c))$, the expected discounted buyer surplus such a seller generates is

$$\begin{aligned} &\sum_{t=0}^{\infty}(\delta F_{(1)}(p^i(c)))^t \left[\int_{p^i(c)}^{\overline{v}} v[f_{(1)}(v) - f_{(2)}(v)]dv - \left(F_{(2)}(p^i(c)) - F_{(1)}(p^i(c))\right) p^i(c)\right] \\ = \; & \frac{\int_{p^i(c)}^{\overline{v}} v[f_{(1)}(v) - f_{(2)}(v)]dv - \left(F_{(2)}(p^i(c)) - F_{(1)}(p^i(c))\right) p^i(c)}{1 - \delta F_{(1)}(p^i(c))}. \end{aligned}$$

Consequently, taking the expectation over the relevant seller types, the expected discounted buyer surplus in scenarios $i \in \{1, 2, 3\}$ gives

$$W_B^i = \int_{\underline{c}}^{\overline{c}^i} \frac{\left[\int_{p^i(c)}^{\overline{v}} v[f_{(1)}(v) - f_{(2)}(v)]dv - \left(F_{(2)}(p^i(c)) - F_{(1)}(p^i(c))\right) p^i(c)\right]}{1 - \delta F_{(1)}(p^i(c))} dG(c).$$

The expected intermediary surplus W_I^i in scenario i is

$$W_I^i = \overline{b} \int_{\underline{c}}^{\overline{c}^i} \frac{p^i(c)(F_{(2)}(p^i(c)) - F_{(1)}(p^i(c))) + \int_{p^i(c)}^{\overline{v}} y f_{(2)}(y)dy}{1 - \delta F_{(1)}(p^i(c))} dG(c).$$

Expected sellers' surplus in scenario i, denoted W_S^i, can be written as

$$W_S^i \;=\; (1-\overline{b}) \int_{\underline{c}}^{\overline{c}^i} \frac{(p^i(c) - \frac{c}{1-\overline{b}})(F_{(2)}(p^i(c)) - F_{(1)}(p^i(c))) + \int_{p^i(c)}^{\overline{v}} \left[y - \frac{c}{1-\overline{b}}\right] f_{(2)}(y)dy}{1 - \delta F_{(1)}(p^i(c))} dG(c).$$

Transfer Taxes We assume that the government imposes a percentage transfer tax $\kappa \geq 0$. We denote by $p_\kappa(c)$ the "after-tax pre-fee" reserve price the seller of type c sets. That is, if the intermediary charges a percentage fee b and the government collects a transfer tax κ, the seller's net reserve price is $p_\kappa(c)(1-b)$. Accordingly, if a bidder buys at the reserve price, she pays $p := (1+\kappa)p_\kappa$ and the intermediary receives bp_κ. Given p_κ and κ, the probability that a given bidder is willing to buy the object is $F^\kappa(p_\kappa) := F(p_\kappa(1+\kappa))$ with density $f^\kappa(p_\kappa) := f(p_\kappa(1+\kappa))(1+\kappa)$. Accordingly, for $h = 1, 2$ we let $F_{(h)}^\kappa(p_\kappa) := F_{(h)}(p_\kappa(1+\kappa))$ and $f_{(h)}^\kappa(p_\kappa) := f_{(h)}(p_\kappa(1+\kappa))(1+\kappa)$. Denoting by $v^\kappa := v/(1+\kappa)$ the transaction-relevant valuation and by $\Phi^\kappa(v^\kappa) := v^\kappa - \frac{1-F^\kappa(v^\kappa)}{f^\kappa(v^\kappa)}$ the associated virtual valuation, we get $\Phi^\kappa(v^\kappa) = \Phi(v)/(1+\kappa)$ as the virtual valuation relevant for the static mechanism design problem with $v = (1+\kappa)v^\kappa$. Analogously, the virtual valuation relevant for the dynamic setup given transfer tax κ, denoted $\tilde{\Phi}^\kappa(v^\kappa)$, satisfies

$$\tilde{\Phi}^\kappa(v^\kappa) = \frac{\tilde{\Phi}(v)}{1+\kappa},$$

where $\tilde{\Phi}(v)$ is given by the expression in Proposition 1 evaluated at $\omega = 0$.

The expected intermediary surplus in scenario i is denoted W_I^i and given as

$$W_I^i = \frac{\int_{\underline{c}}^{\overline{c}^i} \omega_i^\kappa(p_\kappa^i(c))(F_{(2)}^\kappa(p_\kappa^i(c)) - F_{(1)}^\kappa(p_\kappa^i(c))) + \int_{p_\kappa^i(c)}^{\overline{v}^\kappa} \omega_i^\kappa(y) dF_{(2)}^\kappa(y)}{1-\delta F_{(1)}^\kappa(p_\kappa^i(c))} dG(c),$$

where $F_{(i)}^\kappa(y) := F_{(i)}(y(1+\kappa))$ for $i = 1, 2$.

Expected sellers' surplus in scenario i is

$$W_S^i = \frac{\int_{\underline{c}}^{\overline{c}^i} [p_\kappa^i(c) - \omega_i^\kappa(p_\kappa^i(c)) - c](F_{(2)}^\kappa(p_\kappa^i(c)) - F_{(1)}^\kappa(p_\kappa^i(c))) + \int_{p_\kappa^i(c)}^{\overline{v}^\kappa} [y - \omega_i^\kappa(y) - c] dF_{(2)}^\kappa(y)}{1-\delta F_{(1)}^\kappa(p_\kappa^i(c))} dG(c),$$

and government's expected tax revenue in scenario i, denoted K^i, is

$$K^i = \kappa \int_{\underline{c}}^{\overline{c}^i} \frac{p_\kappa^i(c)(F_{(2)}^\kappa(p_\kappa^i(c)) - F_{(1)}^\kappa(p_\kappa^i(c))) + \int_{p_\kappa^i(c)}^{\overline{v}^\kappa} y dF_{(2)}^\kappa(y)}{1-\delta F_{(1)}^\kappa(p^i(c))} dG(c).$$

Next we derive the optimal fee given κ. Let

$$\begin{aligned}
v^\kappa &:= v/(1+\kappa), \quad \overline{v}^\kappa := \overline{v}/(1+\kappa) \quad \text{and} \quad \underline{v}^\kappa := \underline{v}/(1+\kappa) \\
F^\kappa(v^\kappa) &:= F(v^\kappa(1+\kappa)) \quad \text{and} \quad F_{(1)}^\kappa(v^\kappa) := F_{(1)}(v^\kappa(1+\kappa)) \\
F_{(2)}^\kappa(v^\kappa) &:= F_{(2)}(v^\kappa(1+\kappa)) \quad \text{with support} \quad [\underline{v}^\kappa, \overline{v}^\kappa] \\
1 - F_\infty^\kappa(v^\kappa) &:= \frac{1-F_{(1)}^\kappa(v^\kappa)}{1-\delta F_{(1)}^\kappa(v^\kappa)} \\
R^\kappa(p) &:= \frac{p[F_{(2)}^\kappa(p) - F_{(1)}^\kappa(p)] + \int_p^{\overline{v}^\kappa} y dF_{(2)}^\kappa(y)}{1-F_{(1)}^\kappa(p)} \\
1 - \overline{F}^\kappa(k) &:= 1 - F_\infty^\kappa(R^{\kappa-1}(k)) \\
f^\kappa(v^\kappa) &:= [F^\kappa(v^\kappa)]', \quad f_{(1)}^\kappa(v^\kappa) := [F_{(1)}^\kappa(v^\kappa)]', \quad \text{and} \quad \overline{f}^\kappa(v^\kappa) := [\overline{F}^\kappa(v^\kappa)]' \\
\Phi^\kappa(v^\kappa) &:= v^\kappa - \frac{1-F^\kappa(v^\kappa)}{f^\kappa(v^\kappa)} \quad \text{and} \quad \tilde{\Phi}^\kappa(p) := \overline{v}^\kappa - \int_p^{\overline{v}^\kappa} \frac{1-\delta F_{(1)}^\kappa(y)}{1-\delta} [\Phi^\kappa(y)]' dy \\
\overline{\Phi}^\kappa(v^\kappa) &:= v^\kappa - \frac{1-\overline{F}^\kappa(v^\kappa)}{\overline{f}^\kappa(v^\kappa)}.
\end{aligned}$$

A direct application of Lemma 1 implies that the optimal allocation rule is such that there is trade in period t from the seller to the buyer b_t with the highest virtual valuation $\tilde{\Phi}^\kappa(v_{bt})$ if $\tilde{\Phi}^\kappa(v_{bt}) \geq \Gamma_\alpha(c)$. Otherwise, there is no trade in period t. This allocation rule can be implemented with fee-setting if a seller of type c can be induced to set the (after-tax pre-fee) reserve price

$$p_\kappa^{2.b}(c) := \tilde{\Phi}^{\kappa-1}(\Gamma_\alpha(c)). \tag{28}$$

Next, let

$$\overline{\omega}^\kappa(k) := k - \frac{\int_p^{\overline{v}^\kappa} \Gamma_\alpha^{-1}(\overline{\Phi}^\kappa(y))\overline{f}^\kappa(y))dy}{1 - \overline{F}^\kappa(p)} \tag{29}$$

and

$$V^\kappa(c) := (1 - \overline{F}^\kappa(k^\kappa(c)))(k^\kappa(c) - \overline{\omega}^\kappa(k^\kappa(c)) - c), \tag{30}$$

where

$$k^\kappa(c) := \overline{\Phi}^{\kappa-1}(\Gamma_\alpha(c)).$$

In a stationary environment, the optimal transaction fee $\omega^\kappa(p_\kappa)$ is then given by

$$\omega^\kappa(p_\kappa) := p_\kappa - \frac{\int_{p_\kappa}^{\overline{v}^\kappa} \left[\Gamma_\alpha^{-1}(\tilde{\Phi}^\kappa(y)) + \delta V^\kappa(\Gamma_\alpha^{-1}(\tilde{\Phi}^\kappa(y)))\right] f^\kappa(y)dy}{1 - F^\kappa(p_\kappa)}, \tag{31}$$

which follows as a direct implication of Proposition 2 and an application of the various definitions.

Online Supplement

Appendices D, E, and F

Appendices D, E, and F are provided in the online supplement, see `http://andras.niedermayer.ch/research/`.