Optimal Percentage Fees

Simon Loertscher† Andras Niedermayer‡

June 11, 2014

Abstract

Revenue seeking intermediaries and governments around the globe charge fees as a function of the transaction price. In a setup with few buyers, one seller and two-sided private and independent information about valuations, we derive the fee structure that maximizes a weighted average of revenue and seller welfare. In increasingly thin markets, such optimal fees converge to linear fees. The optimal fee decreases with the weight on welfare and with the elasticity of supply, but may increase with the elasticity of demand. We test the model's empirical predictions using data from the Boston condominium market in the 1990s and use the estimate to perform counterfactual analysis. We find that percentage fees achieve 99 percent of what can be obtained under the optimal Bayesian mechanism.

Keywords: brokerage, fee-setting, percentage fees, thin markets, real estate.

JEL-Classification: C72, C78, L13
1 Introduction

Revenue seeking intermediaries and governments around the globe charge transaction fees and taxes that are paid whenever 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. 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, the recent price-fixing verdict for ebooks of Apple, the - at times drastic - increases of value-added taxes in financially troubled countries, or the introduction of financial transaction taxes in the European Union (EU). As of now, 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 such as price posting.

In this paper, we analyze the structure of transaction fees 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. Information about valuations and costs is private, and sellers and buyers can only trade via the platform provided by the agent who chooses the trading mechanism (which can be a private intermediary or a government; for simplicity we only talk about intermediaries in the following). We

\[1\] 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.

Apple has been found guilty of participating in a price-fixing conspiracy for ebooks when it let publishers set prices and charged a commission fee upon a sale. Apple has defended its choice of trading mechanism as in the company’s independent business interest and is currently appealing the court ruling with the appeal trial scheduled to begin on July 14, 2014.

\[2\] As to why the services of private intermediaries and governments are useful or necessary, there are a variety of mutually non-exclusive explanations such as for reducing transaction and search costs, certifying quality, improving matching, building reputation, improving matching, building reputation, providing infrastructure that facilitates trade, and enforcing contracts; see [Spulber (1999)] and [Salanié (2003)] for overviews of the role of private intermediaries and governments, respectively. Rather than choosing a specific combination of these explanations, we focus on the optimal pricing problem, taking as given that governments impose indirect taxes and that buyers and sellers trade via intermediaries.
model the relationship between the intermediary and the seller as a long-term relationship with an exclusive contract. The intermediary announces and commits to a sequence of fees that specify, as a function of the transaction price, the commission the seller has to pay to the intermediary upon a sale in any given period. Observing this sequence of fees and his own cost, the seller chooses a reserve price in every period. Buyers arrive at random and live for one period. Upon drawing their values independently from the same distribution, they bid in a second-price auction. The seller’s cost and the buyers’ valuation are independently distributed, and distributions are common knowledge while realizations are the agents’ private information. We assume that the sequences of fees is chosen to maximize a weighted average of the intermediary’s revenue and the joint \textit{ex ante} expected surplus of the intermediary and the seller.

The dynamic model with an exclusive, long-run contract between the intermediary and the seller is motivated by the contracts used in real-estate brokerage, with the second-price auction – which is strategically equivalent to an English auction – representing the bargaining between the buyers and the seller.\footnote{If a broker has attracted multiple buyers for a given property, he will let the buyer with the highest standing bid know when another buyer makes a higher price offer, giving her the opportunity to increase her bid. Thus, this broker-intermediated bargaining is like an English auction.} Real estate is an important industry with an annual transaction volume in the order of \$1,000 billions in the United States alone. Moreover, the data we use in the empirical section and for the counterfactual analysis are from real-estate transactions in the Boston condominium market in the 1990s.\footnote{The data set is the one used by Genesove and Mayer (2001). We are grateful to the authors for having given us access to their data.} However, our model is also applicable to a variety of other industries and markets, and setups with short-term contracts or one-off interactions can obviously be accommodated by setting the discount factor to zero. Trading platforms like eBay and Amazon tend to take the structure of their fees seriously and choose, in some cases, very fine-grained fee structures. For example, eBay.com uses piecewise linear fees and a hybrid combination of a second-price and an English auction with reserve. The fees differ across a wide array of products. Another case in point is Amazon.com, which has 28 different categories of products which are subject to different fees, 25 of which are linear.\footnote{One of the three categories with non-linear fees are entertainment collectibles with \$1.44 and 20} Amazon’s, and eBay’s, sophisticated
fee-structures are hard to reconcile with the argument that linear (or percentage) fees are chosen because they are simple rather than because they are optimal, and impossible to explain by a principal-agent theory of intermediation.

We solve for the optimal structure of the transaction fee. Using methods from mechanism design and non-linear pricing, we show that transaction fees with second-price auctions are without loss of generality in this environment because the optimal mechanism can be implemented via an appropriately chosen sequence of fees.

Many of the empirically observed fees used in thin markets are linear (as mentioned most of Amazon’s fees are linear) or even proportional. For example, real-estate brokers in the United States almost universally charge 6 percent (see e.g. Hsieh and Moretti, 2003). Assuming that only the most motivated traders enter the market because of exogenously given transaction costs, we then provide a theory of the asymptotic optimality of linear fees. The assumption that only the most motivated traders enter the market is, for example, well-grounded for real-estate markets, in which at any given point typically less than 5 percent of all the properties are on the market. Our result states that, under mild regularity conditions, the optimal fees converge to linear fees as the fraction of active trades goes to zero. This result is strongly related to the statistical law in extreme value theory that truncated distributions converge to generalized Pareto distributions as the truncation threshold becomes larger (see e.g. Balkema and De Haan, 1974; Pickands, 1975; Falk, Hüsler, and Reiss, 2010). The efficacy of simple mechanisms relative to more elaborate, yet optimal mechanisms has recently been investigated under a variety of setups; see, for example, McAfee (2002), Rogerson (2003), Chu and Sappington (2007), and Chu, Leslie, and Sorensen (2011). The present paper contributes to this strand of literature by providing conditions under which simple mechanisms – that is, fee-setting with linear fees – are optimal, or converge to optimality. To the best of our knowledge, it is the first paper that relates the structure of commission fees to extreme value theory.6

percent of the price up to $100, 10 percent between $100 and $1,000 and 6 percent above $1,000.

6The motivation for this theory was an empirical regularity found in many situations: that the upper tail of a distribution is well approximated by a (Generalized) Pareto distribution. A prominent example is the distribution of the highest 20 percent of income and wealth in many countries, which was first observed by Vilfredo Pareto. Other examples include the distribution of the strength of earthquakes in historical data (which tend to contain only the most severe earthquakes); and for the discrete type
We further show that the distribution of reserve and transaction prices is independent of the weight on the intermediary’s profit when the optimal fee is linear. If percentage fees are optimal, then this also implies that the distributions of reserve and transaction prices do not vary with the percentage being charged. In fact, given demand the distribution of reserve prices pins down the distribution of the seller’s cost when the optimal fees are linear. Together with our theoretical arguments and empirical analysis (see below) that provide strong support for the optimality of linear fees, this result is likely to be useful for future empirical research. It also helps explain well-documented empirical regularities that have been found puzzling: Although the vast majority of real-estate transactions are intermediated by brokers who charge percentage fees, there are also for-sale-by-owner (FSBO) transactions in which no fees are charged. A natural conjecture is that, adjusting for quality, prices by such direct sellers will be lower essentially because of the double-marginalization inherent to intermediated trade. As shown by Hendel, Nevo, and Ortalo-Magne (2009), this is empirically not the case. In our model, an FSBO platform can be interpreted as the extreme case of our model in which the optimal fee is zero. The driving force behind these results is the additional entry of high cost sellers that occurs in the FSBO market and the invariance of (generalized mirrored) Pareto distributions to truncations that makes linear fees optimal.

The intuition for many of our other results is based on monopsony and Ramsey pricing. Nevertheless, our model makes a number of surprising predictions that cannot be accounted for by existing theories. For example, in our model an overall increase in the elasticity of demand can lead to an increase in the fee that maximizes the profit of the intermediary. Moreover, the asymptotically optimal fee in the one-shot setup is independent of the distribution from which the buyers’ valuations are drawn.

---

7Similar observations in related contexts were made by Rutherford, Springer, and Yavas (2005) and Levitt and Siverson (2008).

8As we explain in detail in the main text, important concepts like the curvature of the direct demand function, the curvature of the indirect demand function, the pass-through rate, and the markup/quantity weighted average pass-through, which are due to the contributions by Bulow and Pfeiderer (1983), Aguirre, Cowan, and Vickers (2010), Bulow and Klemperer (2012), and Weyl and Fabinger (2013) cannot account for these effects.
Our model makes a number of empirical predictions. We show that the underlying distributions in our model are non-parametrically identifiable and use the data set of Genesove and Mayer (2001) on the condominium market in Boston in the 1990s to estimate the distributions of the seller’s and buyers’ valuations (which can be interpreted as supply and demand functions). Assuming that the percentage fee was chosen to maximize a weighted average of the intermediary’s and the seller’s profit, we can back out which weight is consistent with the 6% fee that is being charged. The model and the estimated parameters lend themselves well to counterfactual simulations. As an illustration, we compare the surplus generated by simple 6% fees to the maximal theoretically attainable surplus using the optimal Bayesian mechanism. We find that the 6 percent fees achieve 99 percent or more of the maximum surplus. This result is explained by the almost linear functional form of the empirically estimated weighted virtual cost function of the seller. Another counterfactual is that a hypothetical monopolistic broker would charge a 22% fee, which achieves more than 99% of the theoretically achievable profit of a monopolistic broker.

Thin markets with two-sided private information are at the core of our analysis. The seminal paper by Myerson and Satterthwaite (1983) shows the importance of two-sided private information in thin markets: bargaining may break down even with positive gains from trade. A subsequent literature shows the relevance of two-sided private information in markets that with search and matching frictions (Wolinsky (1988), Satterthwaite and Shneyerov (2007), Lauermann (2013), and Lauermann, Merzyn, and Virag (2012)). Our article stresses the importance of market thinness and two-sided private information for a question that is very different from those studied in this literature so far: for transaction fees and taxes. It turns out that the distribution of private types pins down the fee structure and is also the driving force for the surprising effect of the elasticities on the optimal fee. Further, market thinness drives optimal fees towards linearity.

The combination of market thinness and two-sided private information is what sets the theoretical part of our paper apart from the existing literature on the transaction fees of profit maximizing intermediaries (Yavas (1992), Caillaud and Jullien (2003), Hagiu
1 INTRODUCTION

This characterization is at the heart of our analysis: without the combination of market thinness and two-sided private information, the theory would be silent about the functional form of the fee (fixed, percentage, non-linear). Assuming two-sided private information, Jullien and Mariotti (2006) consider a static model with one broker and two buyers. Their focus is on fees that are a function of the reserve price rather than the transaction price.

Our paper also contributes to the literature on real estate and real-estate brokerage, which has witnessed an upsurge of interest over the past decade; see, for example, Genesove and Mayer (1997, 2001), Hsieh and Moretti (2003), Hendel, Nevo, and Ortalo-Magné (2009), and Genesove and Han (2012). While our question and setup are quite different, our paper has similarities to the empirical work on auctions, such as Donald and Paarsch (1993), Bajari (1997), Bajari and Hortacsu (2003), and Shneyerov (2006) because we model the bargaining procedure as an auction. The similarity to Bajari and Hortacsu (2003) goes even further: the auctions for which they estimate the optimal reserve prices are run by a profit maximizing intermediary – eBay – that also charges a transaction fee. Some of our methods should be applicable in a modified way to the analysis of the fee-setting behavior of intermediaries in setups beyond real-estate brokerage.

The remainder of this paper is structured as follows. Section 2 introduces and analyzes the theoretical model. In Section 3, we perform the empirical analysis. Counterfac-

---

9 Yavas (1992), Caillaud and Jullien (2003), and Shy and Wang (2011) assume that the seller’s cost is public information. 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 (2012) 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. 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).
tuals are analyzed in Section 4. Section 5 concludes. Proofs and supplementary material are in the Appendix.

2 Model

In this section, we first introduce the model and derive optimal seller behavior and the optimal fees. Then we show that the optimal fees converge to linear fees when only the keenest sellers participate, provide intuition for the theoretical results, and discuss a number of applications and possible extensions.

2.1 Setup

We study an infinite horizon model with discrete time indexed $t = 0, 1, \ldots$. There is one intermediary and one seller who either owns a single indivisible good or has the capacity to produce one good. The seller’s cost $c$ is his private information and drawn from the distribution $G$ with support $[c, \overline{c}]$ and density $g(c) > 0$ for all $c \in (c, \overline{c})$. His value of the outside option of not participating is zero. The cost $c$ 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. 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. The seller and the intermediary have the common discount factor $\delta \in [0, 1)$, 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.

To deal with the stochastic arrival process of buyers, 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 $\pi_B$ of having exactly $B \leq \overline{B}$ buyers is given by the probability mass function for the binomial $\pi_B = \left( \begin{array}{c} \overline{B} \\ B \end{array} \right) \tilde{\pi}^B (1 - \tilde{\pi})^{\overline{B} - B}$. Important special cases are those with a deterministic number of buyers $\overline{B}$, which corresponds to assuming $\tilde{\pi} = 1$, so that $\pi_{\overline{B}} = 1$, and the Poisson distribution, for which

$$\pi_B = \frac{\zeta^B e^{-\zeta}}{B!},$$
which is obtained as the limit of the binomial when $B$ goes to infinity, keeping $B \pi$ constant at $\zeta$. Buyers who participate are sometimes also called *bidders*. Each bidder draws her valuation $v$ independently from the distribution $F(v)$ with support $[\underline{v}, \overline{v}]$ and density $f(v) > 0$ for all $v \in (\underline{v}, \overline{v})$. We assume that $c = \underline{v}$ and $\overline{c} = \overline{v}$. 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. Throughout the paper, we use the standard convention of denoting cumulative distribution functions by upper case letters and the corresponding probability densities by lower case letters.

For a given number of buyers $B$, the distribution of the highest valuation is $F(v)^B$. Accordingly, the unconditional distribution of the highest valuation is

$$F(1)(v) := \sum_{B=0}^{\infty} \pi_B F(v)^B,$$

and the unconditional distribution of the second-highest valuation is

$$F(2)(v) := \sum_{B=1}^{\infty} \pi_B BF(v)^{B-1}(1 - F(v)) + F(1)(v).$$

A useful fact is \[ F(2)(v) - F(1)(v) f(1)(v) = 1 - F(v) f(v). \tag{1} \]

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 as a virtual cost function. In what follows, we will derive other distributions of buyers’ types that are relevant for the analysis, which will be denoted $\tilde{F}(v), F_\infty(v)$ and $\overline{F}(v)$.

The corresponding virtual valuation functions will be denoted $\tilde{\Phi}(v), \Phi_\infty(v)$ and $\overline{\Phi}(v)$, respectively.

The intermediary chooses a sequence of fee functions $\omega = (\omega_t)_{t=0}^{\infty}$, with $\omega_t(\hat{p})$ specifying the amount the seller has to pay to the intermediary when a transaction occurs in

\[ \text{This is easily seen to be true once one notes that } f(1)(v) \text{ can be written as } f(1)(v) = \sum_{B=1}^{\infty} \pi_B BF(v)^{B-1} f(v). \]
period \( t \) at the transaction price \( \bar{p} \). The transaction price is determined by a second-price auction (or the strategically equivalent English auction) with reserve price \( p_t \) set by the seller. This may literally be a second-price auction, or the second-price auction may serve as a plausible description of the bargaining protocol. For example, in real-estate transaction 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 \( p_t \). Observe that the seller is not allowed to recall buyers after the period in which they arrived.

For a given \( \omega_t \), 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

\[
R_{\omega_t}(p_t) = \frac{(p_t - \omega_t(p_t))(F(2)(p_t) - F(1)(p_t)) + \int_{p_t}^{\bar{p}} (\bar{p} - \omega_t(\bar{p}))dF(2)(\bar{p})}{1 - F(1)(p_t)},
\]

by standard arguments from auction theory (see e.g. Krishna, 2002). Consequently, the maximization problem of a seller of type \( c \) given \( \omega \) is to choose a sequence of prices \( p = (p_t)_{t=0}^{\infty} \) to maximize

\[
W_S(c, p, \omega) := \sum_{t=0}^{\infty} (R_{\omega_t}(p_t) - c)(1 - F(1)(p_t)) \prod_{\tau=0}^{t-1} \delta F(1)(p_\tau).
\]

(Here and throughout the remainder of the paper, we use the convention of setting \( \prod_{\tau=z}^{T} y_\tau = 1 \) for any sequence \( (y_\tau)_{\tau=1}^{T} \).) Let \( P(c) = (P_t(c))_{t=0}^{\infty} \) be the (or a) maximizer of \( W_S(c, p, \omega) \), making explicit its dependence on \( c \) while its dependence on the sequence of fees is kept implicit for ease of notation.

Given \( \omega_t \) and \( p_t \), the intermediary’s expected revenue in period \( t \) when facing a seller of type \( c \) is

\[
\omega_t(p_t)(F(2)(p_t) - F(1)(p_t)) + \int_{p_t}^{\bar{p}} \omega_t(\bar{p})dF(2)(\bar{p}).
\]

\(^{11}\)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.

\(^{12}\)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.

\(^{13}\)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 \( \bar{p} \sim F(2) \).
Therefore, the intermediary’s discounted expected profit from a seller of type $c$ given $p$ and $\omega$ is

$$W_I(c, p, \omega) := \sum_{t=0}^{\infty} \left( \omega_t(p_t)(F(2)(p_t) - F(1)(p_t)) + \int_{p_t}^{\overline{p}} \omega_t(\tilde{p}) dF(2)(\tilde{p}) \right) \prod_{\tau=0}^{t-1} \delta F(1)(p_\tau).$$

We assume that the fees $\omega$ are chosen to maximize the expectation of $\alpha W_S + (1 - \alpha)(W_S + W_I)$, where $\alpha \in [0, 1]$ is the weight on the intermediary’s profit. It can 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, 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.

Formally, let

$$W(\alpha, \omega) := E_c[\alpha W_I(c, P(c), \omega) + (1 - \alpha)(W_I(c, P(c), \omega) + W_S(c, P(c), \omega))],$$

denote the expected value of the weighted sum of intermediary profit and joint surplus of the intermediary and the seller. Then $\omega$ is chosen to maximize the weighted sum

$$\max_{\omega} W(\alpha, \omega).$$

Notice that the objective function in (2) depends on $\omega$ directly and indirectly via the seller’s pricing behavior $P(c)$, which depends on $\omega$.\(^{14}\)

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. Below we will show that our assumptions regarding the menu of mechanisms that can be chosen are without loss of generality insofar as 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 conditions under which the results extend to first-price auctions. In the independent private values setup we study in this paper, second-price auctions are equivalent to English auctions. They serve as a plausible description of

\(^{14}\)Observe also that maximizing $W(\alpha, \omega)$ is equivalent to maximizing the weighted sum $E_c[\alpha_0 W_I(c, P(c), \omega) + (1 - \alpha_0)W_S(c, P(c), \omega)]$ with $\alpha_0 \in [1/2, 1]$.\)
bargaining intermediated by a broker as brokers keep buyers informed about the highest standing offer to allow them to increase their bid. 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. At the end of Section 2 we discuss what happens when we relax these assumptions.

### 2.2 Optimal Seller Behavior

Given a sequence of fee functions $\omega = (\omega_t)_{t=0}^{\infty}$ the seller will choose a sequence of reserve prices $p = (p_t)_{t=0}^{\infty}$ such that it maximizes the expected net present value of his profits. In general, this maximization problem will be complex, since the fees could be non-stationary and the implied profit function of the seller could not be quasi-concave, so that the first order condition will not be sufficient. However, we will later show that even with an arbitrary non-stationary mechanism one could not do better than 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, i.e. $\omega_t = \omega$ and $p_t = p$ for all $t$, and use the first-order condition for maximization.\footnote{By using standard techniques it is possible to extend the analysis to non-optimal fees which imply a non-stationary non-quasi-concave problem.}

Given stationary fees $\omega$ and stationary prices $p$, the seller’s utility becomes

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

where

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

is the probability of ever selling.\footnote{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 seller chooses the reserve $p$ such that it maximizes $W_S$. The following proposition gives the solution to this maximization problem.
Proposition 1. Given a stationary fee $\omega$ that implies a monotone net virtual valuation

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

the optimal price set by a seller with cost $c$ is $P(c) = \Phi^{-1}_\omega(c)$ in every period, where $\Phi_\omega$ is the monotone function

$$\Phi_\omega(p) := \bar{v} - \int_p^\bar{v} \frac{1 - \delta F(1)(v)}{1 - \delta} \Phi'_\omega(v)dv.$$ 

As mentioned before, we will show that the optimal fee is such that the seller’s profit function is quasi-concave, which is equivalent to an increasing $\Phi_\omega$. It will be helpful for the empirical analysis to observe that for a linear fee $\omega$ with slope less than 1 and an increasing $\Phi$, $\Phi_\omega$ (and hence also $\Phi_\omega$) can be shown to be increasing as well. For notational ease, we let $\Phi(v) \equiv \Phi_0(v)$ and $R(p) \equiv R_0(p)$.

Note that $\Phi_\omega$ can be interpreted as the net dynamic virtual valuation function. In a static setup ($\delta = 0$), $\Phi_\omega$ simplifies to the net virtual valuation $\Phi_\omega$. With a zero fee ($\omega(p) = 0$), it further simplifies to the virtual valuation function $\Phi$.

2.3 Optimal Fees

The maximization problem [2] looks tedious. Moreover, even if a solution is obtained, it is not clear how much value is lost because of the restriction to fee-setting. To circumvent both issues, we now employ a mechanism design approach, which allows us to derive the optimal sequence of fees indirectly and to show that the restriction to fee-setting is without loss of generality.

Mechanisms 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 $\delta$ as the pure per period survival probability of the seller, or as a discount factor that reflects pure and common time preferences, or as a combination of the survival probability and time preferences. However, the interpretation of many concepts used in the mechanism design framework
is most straightforward if one interprets $\delta$ 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 type. 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 $v_b \in [v, \overline{v}]$ to the mechanism. The realization of the valuations of buyers who do not enter are set to $v_b = -\infty$. Let $\mathbf{v}_t = (v_t^b)_{b=1}^{B}$ be a vector of such reports by buyers in period $t$ with buyers label $b = 1, \ldots, B$ and let $\mathbf{v} = (\mathbf{v}_t)_{t=0}^{\infty}$ be a sequence of such reports.

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

Feasibility further requires

$$
\sum_{b=1}^{\overline{B}} Q_t^b(\mathbf{v}_t, c) \leq Q_t^S(\mathbf{v}_t, c) \tag{3}
$$

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

Let $Q_B^t(\mathbf{v}_t, c) = (Q_1^t(\mathbf{v}_t, c), \ldots, Q_{\overline{B}}^t(\mathbf{v}_t, c))$ and $M_B^t(\mathbf{v}_t, c) = (M_1^t(\mathbf{v}_t, c), \ldots, M_{\overline{B}}^t(\mathbf{v}_t, c))$. For a given $(\mathbf{v}, c)$, let

$$
Q_S(\mathbf{v}, c) = (Q_t^S(\mathbf{v}_t, c))_{t=0}^{\infty} \quad \text{and} \quad Q_B(\mathbf{v}, c) = (Q_t^B(\mathbf{v}_t, c))_{t=0}^{\infty}
$$

and

$$
M_S(\mathbf{v}, c) = (M_t^S(\mathbf{v}_t, c))_{t=0}^{\infty} \quad \text{and} \quad M_B(\mathbf{v}, c) = (M_t^B(\mathbf{v}_t, c))_{t=0}^{\infty}.
$$

Letting $Q$ and $M$ be, respectively, collections $\{Q_S(\mathbf{v}, c), Q_B(\mathbf{v}, c)\}$ and $\{M_S(\mathbf{v}, c), M_B(\mathbf{v}, c)\}$ for all possible $(\mathbf{v}, c)$, a direct mechanism is summarized by $(Q, M)$ where $Q$ satisfies (3). 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, the ultimate discounted probability of trade for a seller who reports type $c$ is

$$q_S(c) := E_v \left[ \sum_{t=0}^{\infty} Q^t_S(v_t, c) \prod_{\tau=0}^{t-1} \delta(1 - Q^\tau_S(v_\tau, c)) \right]$$

was introduced in Satterthwaite and Shneyerov (2008). We introduce a second, novel concept, the ultimate conditional expected revenue, which we will describe later.\footnote{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.}

The seller’s expected discounted payment is

$$m_S(c) := E_v \left[ \sum_{t=0}^{\infty} M^t_S(v_t, c) \prod_{\tau=0}^{t-1} \delta(1 - Q^\tau_S(v_\tau, c)) \right].$$

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

$$W_S(c) = m_S(c) - q_S(c)c,$$

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

$$W_I(c) = E_v \left[ \sum_{t=0}^{\infty} \left( \sum_{b=1}^{B} M^t_b(v_t, c) \right) \prod_{\tau=0}^{t-1} \delta(1 - Q^\tau_S(v_\tau, c)) \right] - m_S(c),$$
where the notation $W_i(c)$ for $i = I, S$ emphasizes that we are referring to payoffs in a direct mechanism as opposed to fee-setting as defined in Section 2.1. The natural extension of the objective function $(2)$ to the general mechanism design setup is then

$$\max_{\langle Q, M \rangle} E_c [\alpha W_I(c) + (1 - \alpha)(W_I(c) + W_S(c))]$$

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 $(2)$, which is confined to fee-setting. However, as we will show, the objective in $(4)$ can be maximized with an appropriately chosen sequence of transaction fees $\omega$. 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 $\alpha$ we study.\[18\]

**Analysis** 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

$$m_S(c) = q_S(c)c + \int_c^\tau q_S(x)dc + W_S(\tau)$$

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 $\tau$, that is if $W_S(\tau) \geq 0$ (and if the seller’s incentive constraint is satisfied). Because $W_S(\tau)$ enters the objective function as the constant $-\alpha W_S(\tau)$, it will be optimal to set $W_S(\tau) = 0$ for any $\alpha \in [0, 1]$.

We say that the good is auctioned off in period $t$ with reserve $p_t$ if $Q_t^i(v_t, c) = 1$ if $v_b = \max\{v_t\}$ and $v_b \geq p_t$ and $Q_t^i(v_t, c) = 0$ for all $i = 1, \ldots, T$ otherwise.\[19\]

---

18 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).

19 This definition neglects the possibility of ties at the highest value, which have probability 0. If one
Lemma 1. A mechanism \((Q, M)\) is optimal only if the good is auctioned off in every period \(t\) at some reserve \(p_t\).

Lemma 1 implies that the choice set for allocation rules \(Q\) can be narrowed down to sequences of reserves \(P(c) = (p_t(c))_{t=0}^{\infty}\), 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 \(P(c)\) is equivalent to choosing a sequence \(k(c) = (k_t(c))_{t=0}^{\infty}\) of expected transaction prices (conditional on a transaction occurring).

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

\[
q_t(k) := (1 - F(1)(R^{-1}(k_t))) \prod_{\tau=0}^{t-1} \delta F(1)(R^{-1}(k_\tau))
\]

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

\[
k = \frac{\sum_{t=0}^{\infty} q_t(k) k_t}{\sum_{t=0}^{\infty} q_t(k)}
\]

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 wants to account for such ties explicitly, one can arbitrarily set \(Q_b^k(v_t, c) = 1\) for the buyer \(b\) with the highest valuation and, say, the highest index \(b\) amongst all buyers with the highest value.
simplifies calculations.

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

In the proof of the following lemma (Lemma 2), we show that for an arbitrary \( k \in [v, \overline{v}] \), the maximized ultimate discounted probability of trade is

\[
1 - F(k) := 1 - F_{\infty}(R^{-1}(k)).
\]

Therefore, for a given \( c \) and \( k \), we now have \( W_I(c) = k(c)(1 - F(k(c))) - m_S(c) \) and \( W_S(c) = m_S(c) - q_S(c)c \). Using incentive compatibility \( (5) \) and \( W_S(\overline{c}) = 0 \) by individual rationality, the objective given \( c \) becomes

\[
\alpha W_I(c) + (1 - \alpha)(W_I(c) + W_S(c)) = k(1 - F(k)) - cq_S(c) - \alpha \int_c^{\overline{c}} q_S(x)dx.
\]

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

\[
\max_{k(c)} \int_c^{\overline{c}} \left[ k(c)(1 - F(k(c))) - (1 - F(k(c)))\Gamma_\alpha(c) \right] g(c)dc
\]

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

\[
0 = -F(k(c)) \left[ F(k(c)) - \Gamma_\alpha(c) \right],
\]

which is equivalent to \( k(c) = \Phi^{-1}(\Gamma_\alpha(c)) \). 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 \( \Phi(v) \) is monotone.

---

20One 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.
Lemma 2. 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
\[ \max_{b_t} \Phi(v_{b_t}) \geq \Gamma_{\alpha}(c), \]
and the expected payoff of every buyer of type $v$ and of the seller of type $c$ is 0.

Relation to Transaction Fees We now relate these findings to transaction fees. We first derive the expectational fees $\omega(k)$ that are levied on the expected transaction price $k$, and then relate these expectational fees to the sequence of transaction fees $\omega = (\omega_t(\bar{p}))_{t=0}^{\infty}$ that are levied on the transaction price $\bar{p}$. We will show that the optimal sequence of fees is stationary, that is, $\omega_t(\bar{p}) = \omega(\bar{p})$ for all $t$. The purpose of analyzing the hypothetical expectational fees $\omega(k)$ first is two-fold. First, it will allow us to derive $\omega$ straightforwardly. Second, in the special case when the optimal $\omega(k)$ is linear, the optimal $\omega(\bar{p})$ will be linear, which is useful for the analysis on convergence to linear fees in Section 2.4 below.

Assume temporarily that the seller chooses an expected transaction price $k$ in period 0 and is charged the $\omega(k)$ with the ultimate discounted probability $1 - F(k)$ with which a trade occurs.

Proposition 2. The expectational transaction fees that implement the optimal mechanism described in Lemma 2 is
\[ \omega(k) = k - \int_k^{\infty} \frac{\Gamma_{\alpha}(v) f(v) dv}{1 - F(k)}. \]

We now derive the optimal transaction fee. The problem of a seller of type $c$ who faces a transaction fee $\omega_t(.)$ and a sequence of optimal transaction fees $((\omega_t(.)))_{\tau=t+1}^{\infty}$ subsequently is to maximize
\[ \max_{p} (p - \omega(p))(F(2)(p) - F(1)(p)) + \int_p^{\infty} [\bar{p} - \omega(\bar{p})] \bar{p} d\bar{p} - (1 - F(1)(p))[c + \delta V(c)], \] (9)
where $\delta V(c)$ is the discounted option value of future trade. Because the subsequence of transaction fees $((\omega_t(.)))_{\tau=t+1}^{\infty}$ is optimal, it will induce the same allocation rule and the same expected discounted payoff as the optimal expectational fee $\omega(k)$. Consequently,
\[ V(c) = (1 - F(k(c)))(k(c) - \omega(k(c)) - c), \] (10)
where \( \omega(k) \) is as in Proposition \( \text{[2]} \) and \( k(c) := \Phi^{-1}(\Gamma_{\alpha}(c)) \). Observe that \( V'(c) = -(1 - F'(k(c))) > -1 \) by the envelope theorem. Therefore, \( c + \delta V(c) \) is an increasing function.

**Proposition 3.** The optimal transaction fees that implement the optimal mechanism described in Lemma \( \text{[2]} \) are such that for all \( t = 0, 1, \ldots \)

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

To see that the individual rationality constraint of the seller is satisfied \textit{ex post}, it suffices to note that the seller never sets a reserve price \( p \) that is such that \( p - \omega(p) \) is below his cost \( c + \delta V(c) \). Observe also that the fees \( \omega(p) \) incentivize the seller of type \( c \) to set the reserve \( p \) that implements the allocation rule in Lemma \( \text{[2]} \). Therefore, the seller’s incentive constraints are indeed satisfied in every period as claimed above (and not only at date 0). Lastly, the equilibrium is essentially unique in the following sense. The second-price auction with a reserve \( p > \bar{v} \) is well-known to have a unique Bayes Nash equilibrium \cite{Blume and Heidhues, 2004}. Therefore, the equilibrium in every buyers’ subgame is unique. Given this and the fee \( \omega(p) \), there is a uniquely optimal price for every seller who trades with positive probability. The only actions that are not uniquely pinned down in equilibrium are the prices sellers with costs above \( \Gamma_{-1/\alpha}(\bar{v}) \) set because any of price above \( \bar{v} \) will be optimal and consistent with an equilibrium.

As an illustration, consider the special case with \( \delta = 0 \) and bilateral trade by setting \( \pi_1 = 1 \) and \( F \) and \( G \) uniform on \([0, 1]\). This implies \( F(1) = F \) and \( \Gamma_{\alpha}(c) = (1 + \alpha)c \) for \( c \in [0, 1] \) and

\[
\omega(p) = \frac{\alpha}{\alpha + 1} p.
\]

If the broker has all the bargaining power, we have \( \alpha = 1 \) and \( \omega(p) = p/2 \), which implements the bilateral one-period broker optimal direct mechanism derived by \cite{Myerson and Satterthwaite, 1983, Section V}. If the seller has all the bargaining power, we have \( \alpha = 0 \) and \( \omega(p) = 0 \) for all \( p \). This corresponds to an optimal auction as in \cite{Myerson, 1981}.

### 2.4 Convergence to Linear Fees

The optimal fee schedule derived above is potentially a complicated non-linear function of the transaction price. Interestingly, fee-setting with simple linear fees is often used by
intermediaries in thin markets. Examples include real-estate brokers, head hunters and Amazon’s fees for third party sellers of most types of goods (including books, consumer electronics, and personal computers). We next provide a theory of the asymptotic optimality of linear fees in thin markets. This theory rests on the notion that only the most motivated traders are active in equilibrium. This assumption is a good approximation for many markets that are thin. For example, in real estate markets typically less than 5 percent of homeowners offer their property for sale at any given time. Similarly, in labor markets at any given point in time only a small fraction of the working population will be actively looking for a job.

2.4.1 Transaction Costs and Participation by the Most Motivated

We now show how adding transaction costs to buyers’ and sellers’ initial valuations and costs affects the fee structure. Assume the initial distributions $F$ and $G$ have support $[0, 1]$. This assumption is without loss of generality, as will become evident shortly. We assume that the additional transaction costs are additive, multiplicative (multiplying $v$ and $c$), or a combination of both. We model this by considering a sequence of linear transformations of $c$ and $v$, indexed by $j$. Let the seller’s gross cost $c_j$ and a buyer’s gross valuation $v_j$ be

$$c_j = \frac{\overline{v}_j - c_j}{u_j^S} c + c_j \quad \text{and} \quad v_j = \overline{v}_j - (1 - v) \frac{\overline{v}_j - c_j}{u_j^B},$$

with respective supports $[c_j, \overline{v}_j + (\overline{v}_j - c_j)/u_j^S]$ and $[\overline{v}_j - (\overline{v}_j - c_j)/u_j^B, \overline{v}_j]$. The parameters $c_j$ and $\overline{v}_j$ shift the supports. $u_j^S$ can be interpreted as follows. Sellers with costs outside the relevant range $[c_j, \overline{v}_j]$ trade with probability zero. The ratio of the length of the relevant range $\overline{v}_j - c_j$ to the length of $c_j$’s support $[(\overline{v}_j - c_j)/u_j^S + c_j] - c_j$ is $u_j^S$. Therefore, the mass of sellers in the relevant range is $G(u_j^S)$. Similarly, for buyers, the ratio of the relevant range is $u_j^B$ and the mass of buyers in the relevant range is $1 - F(1 - u_j^B)$. The advantage of a parametrization of the linear transformation of $c_j$ and $v_j$ is that it isolates the effects of $u_j^S$ and $u_j^B$.

For an example of a purely additive transaction cost, $c_j = c + x_j$, $v_j = v - y_j$, our parametrization is $c_j = x_j$, $\overline{v}_j = 1 - y_j$, $u_j^S = u_j^B = 1 - x_j - y_j$. For an example
of a purely multiplicative transaction cost that stretches the sellers’ support to \([0, x_j]\)
(that is, to length \(x_j\)) and the buyer’s support to \([- (1 - y_j), 1]\) (that is, length \(y_j\)), our
parametrization is \(\zeta_j = 0, \tau_j = 1, u^S_j = 1/x_j, u^B_j = 1/y_j\).

There are many economic reasons why entry occurs only by the most motivated. A
plausible and simple one are exogenous transaction and search costs. Such costs are
for example due to moving in and out of a house, transporting a good sold through an
auction, the legal costs of signing sales contracts, the time needed to visit houses and to
show houses to prospective buyers. Therefore, even if a buyer’s valuation is higher than
the seller’s opportunity costs when ignoring additional transaction costs, it may not be
desirable for them to trade when taking into account transaction costs.

One can also think of endogenous transaction costs. One example are the fees or
taxes charged by the intermediary. If a buyer’s valuation for a real estate property is
only 5 percent higher than the seller’s, but the brokerage fee is 6 percent, the buyer
and the seller will not be able to agree to a transaction. Another endogenous additional
opportunity cost is the option value of future trade in a dynamic model: a buyer that
does not like the property he is currently visiting has the option to continue visiting
other properties with the hope of finding a property which he values more or for which
the price is lower.

In this section, we will develop the analytically most clean version of the force at
work: exogenous transaction costs that cause a clear cutoff. We defer the discussion of
endogenous costs, of non-clear cutoffs, and also of the issues that need to be kept in mind
for the empirical implications of our theory to later sections.

2.4.2 Convergence to Linear Fees

The analysis is simplified by introducing a normalization with respect to the relevant
range \([\zeta_j, \tau_j]\). Define the normalized (gross) cost \(\tilde{c} = (c - \zeta_j)/(\tau_j - \zeta_j)\), the normalized
(gross) valuation \(\tilde{v} = (c - \zeta_j)/(\tau_j - \zeta_j)\), and the normalized price \(\tilde{p} = (p - \zeta_j)/(\tau_j - \zeta_j)\).
The distributions of the normalized gross cost \(\tilde{c}\) and the normalized gross valuation \(\tilde{v}\),
truncated to \([c_j, v_j]\) and denoted, respectively, \(\tilde{G}_j\) and \(\tilde{F}_j\), are then given as

\[
\tilde{G}_j(\tilde{c}) = \frac{G(u_j^S \tilde{c})}{G(u_j^S)} \quad \text{and} \quad \tilde{F}_j(\tilde{v}) = 1 - \frac{1 - F(1 - u_j^B (1 - \tilde{v}))}{1 - F(1 - u_j^B)}.
\] (11)

The normalized fee is defined as \(\tilde{\omega}_j(\tilde{p}) = \omega(p)/\left( v_j - c_j \right)\). Similarly, the normalized expectational fee is \(\tilde{\omega}_j(\tilde{p}) = \overline{\omega}(p)/\left( v_j - c_j \right)\).

The following proposition shows that increasing transaction costs (that is entry by the most motivated becoming more selective) makes optimal fees closer to linear fees. We show this in the sense of asymptotic results: as transaction costs increase, fees become linear in the limit. We restrict \(F\) and \(G\) to have finite boundaries and discuss infinite boundaries later.

**Proposition 4.** Assume \(F\) and \(G\) have bounded support and \(\Phi\) and \(\Gamma\) are continuously differentiable. Let the shifting constants \(c_j\), \(v_j\) be arbitrary sequences satisfying \(c_j < v_j\) for all \(k\). Let the ratios of the relevant ranges \(u_j^S\) and \(u_j^B\) be sequences that go to 0 as \(k\) goes to infinity. Then, as \(k \to \infty\),

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

(ii) the normalized expectational fee and the normalized fee converges to the same linear fee:

\[
\lim_{k \to \infty} \tilde{\omega}_j(\tilde{p}) = \tilde{\omega}^*(\tilde{p}), \quad \lim_{k \to \infty} \tilde{\omega}_j(\tilde{p}) = \tilde{\omega}^*(\tilde{p}), \quad \tilde{\omega}^*(\tilde{p}) = \frac{\alpha}{\alpha + \sigma} \tilde{p}.
\] (12)

Part (i) of the 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 B 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.
Figure 1: Density of truncated, rescaled distribution \( G_u(c) = G(c + u(c - c)) / G(c + u(c - c)) \) for \( u \in \{1, 0.7, 0.5, 0.3, 0.2\} \) for a Beta distribution with support \([0, 1]\) and density \( g(c) \propto c^4(1 - c)^4 \) (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) \( G(0.7) = 0.9 \), (c) \( G(0.5) = 0.5 \), (d) \( G(0.3) = 0.1 \), (e) \( G(0.2) = 0.02 \). As the mass decreases, the distribution converges to the approximating Pareto distribution and the approximating Pareto distribution converges to the limiting Pareto distribution.

2.4.3 Intuition and Illustration

To develop an intuition for these results take a setup with one buyer, one seller, and one period, and first assume that we start with (mirrored) Generalized Pareto distributions \( G(c) = c^\sigma \) and \( F(v) = 1 - (1 - v)^\beta \). It is easy to see that convergence is immediate: \( \tilde{G}_j(\tilde{c}) = G(u^S_i \tilde{c}) / G(u^S_i) = \tilde{c}^\sigma \), that is, \( \tilde{G}_j \) does not change with \( k \). The same holds for \( \tilde{F}_j \).

We provide a numerical example for convergence if one starts with a non-Pareto distribution. Figure 1 shows the density of a distribution of which only the lower tail is taken. In particular, it shows only the distribution conditional on being on \([0, u]\) rather than the full support \([0, 1]\). As depicted in the figure, moving the truncation point \( u \) downwards brings the density of the conditional distribution closer to the density of a mirrored Generalized Pareto distribution. As usual for asymptotic results, the statement of Extreme Value Theory holds in the limit. However, in many empirical settings, Pareto distributions are a good approximation far away from the limit. We defer the discussion of empirics and only note that for the numerical example considered in Figure 1, lower tails consisting of the lowest 10 percent (Figure 1 (d)) and the lowest 2 percent (Figure 1 (e)) of the distributions are already very close to a Pareto distribution.

An intuition for part (ii) of the proposition can be gained by starting with a mirrored Generalized Pareto distribution \( G \). Observe that this implies a linear virtual cost...
Figure 2: Optimal fee $\omega(\cdot)$ for truncated, rescaled distributions $F_u(v) = 1 - \left[1 - F(\overline{v} - u(\overline{v} - v))\right]/\left[1 - F(\overline{v} - u(\overline{v} - v))\right]$, and $G_u(c) = G(c + u(c - \underline{c}))/G(c + u(\overline{c} - \underline{c}))$ 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).

$\Gamma_\alpha(c) = c(1 + \alpha/\sigma)$. The fee can be written as $\omega(p) = p - E[\Gamma_\alpha^{-1}(\Phi(v))|v \geq p] = p - \Gamma_\alpha^{-1}(E[\Phi(v)|v \geq p])$ by the linearity of $\Gamma_\alpha$. Further, observe that $E[\Phi(v)|v \geq p] = p$. Therefore, we get the linear fee $\omega(p) = p - \Gamma_\alpha^{-1}(p) = p\alpha/(\alpha + \sigma)$. The above discussion trivially extends to expectational fees $\overline{\omega}(p)$ by replacing $F$ with $\overline{F}$.

Of course, this by itself does not prove that fees converge to linear fees. The additional steps of the proof are in the Appendix. Here we provide an illustration of the convergence of fees which is based on the same numerical example as Figure 1. Figure 2 shows that the optimal fee moves closer to a linear fee as the transaction costs increase. Again, the mass of traders $G(u)$ and $1 - F(1 - u)$ 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, the masses of traders being 10 percent and 2 percent, respectively.

Next, we can turn our attention to denormalized fees. When doing so, the usual care has to be taken when considering the practical interpretation of asymptotic results. For practical purposes, one should not think of $u_j^S = u_j^B = 0$ being necessary to get linear fees, but $u_j^S$ and $u_j^B$ being sufficiently small (which can be still quite large) to have (approximately) linear fees. This is analogous to the practical interpretation of other asymptotic results in statistics. For example, the central limit theorem should not be viewed as the standard deviation being equal to zero in order to get a normal distribution. Instead, a distribution with a (potentially) large standard deviation can
be well approximated by a normal distribution, provided that the underlying random
type is the average of a sufficiently large (but finite) number of random draws.

Take a $k$ for which $u_j^S$ and $u_j^B$ are “sufficiently small”. For the sake of clarity, denote
the lower bound of the relevant range $c_j$ as $c = v = c_j$ and the upper bound $v_j$ as $v = v_j$. The denormalized limiting Pareto distributions are $G^*(c) = [(c - \xi)/(\bar{c} - \xi)]^\sigma$ and $F^*(v) = 1 - [(\bar{v} - v)/(\bar{v} - v)]^\beta$. The denormalized fee is
\begin{equation}
\omega(p) = p - \Gamma_\alpha^{-1}(p).
\end{equation}

This expression allows for an intuitive interpretation of fees, since $\Gamma_\alpha^{-1}(v)$ is the price
set by a “Ramsey monopsonist” with valuation $v$ who face the supply $G(\hat{p})$, that is,
$\Gamma_\alpha^{-1}(v) = \arg \max_{\hat{p}} [\alpha(v - \hat{p}) + (1 - \alpha)(v - c)]G(\hat{p})|_{c=\hat{p}}$. A more elastic supply means
a higher monopsony price and hence a lower fee. As an illustration, $\tilde{G}(\hat{c}) = \hat{c}^\sigma$ can
be seen as isoelastic supply with elasticity $\sigma$. As the elasticity $\sigma$ increases, the fee
$\tilde{\omega}(\hat{p}) = \alpha\hat{p}/(\sigma + \alpha)$ decreases.

What is surprising is that the buyer’s distribution does not matter for the fee. We
will provide an intuition for this in Section 2.5 where we draw an analogy between our
results on optimal fees and the theory on optimal monopoly (and monopsony) pricing.
A careful investigation of this analogy provides an intuition for our results, shows how
the intuition carries over beyond the class of the asymptotic Pareto distributions, and
also allows us to identify the driving forces that determine the level of the fee. In some
sense, the analysis of the asymptotic Pareto distribution can be viewed as an analysis of
first-order effects as it will become clear later.

One may be tempted to misinterpret the above asymptotic results in the following
way: in the limit, both the length of the relevant range and the mass of buyers and
sellers in the relevant range are zero. Hence, extreme value theory and our results on
linear fees only apply to markets in which the length of the relevant range and the mass
of traders is (almost) zero. Figures 1 (d) and 2 (d) illustrate that such an interpretation
of asymptotic results is misleading: in the example provided, the length of the relevant
range is 30% of the original length and the mass of sellers is 10% of the original mass.
Yet Pareto distributions and linear fees are already a reasonable approximation.\footnote{The example in the figures is in line with the findings of a large body of empirical literature in a variety of areas. As an example, Pareto distributions are a good approximation of the distribution of incomes already for the top 20\% of incomes.} The correct interpretation is analogous to that of other asymptotic results, such as the central limit theorem.\footnote{The central limit theorem states that the distribution of the average of $N$ random draws converges to a normal distribution as $N$ goes to infinity. As $N$ goes to infinity, the variance of the average goes to zero. However, this does not mean that normal distributions are only a reasonable approximation if the variance is (almost) zero.} One may also be tempted to think that the convergence of fees to linearity simply stems from the relevant range becoming shorter. That linearity does not simply stem from the length of the relevant range can be seen from the fact that $G$ does not converge to a linear function (see Figure 1). Neither does the optimal price $\Phi^{-1}(\Gamma_\alpha(c))$ converge to a linear function in a dynamic setup ($\delta > 0$).

### 2.5 Intuition

We now provide an intuition for our results based on an analogy to monopoly/monopsony pricing and describe the driving forces that determine fees.

**Ramsey Monopsony and Price Endogeneity Effects** To develop an intuition for our results, consider a one-shot setup ($\delta = 0$) with only one buyer ($\pi_B = 1$). In this setup, the probability of selling is simply $1 - F(p)$ for price $p$, which can be interpreted as a demand function. Similarly, the probability that the seller is willing to trade if he gets price $p$ is $G(p)$, which can be interpreted as a supply function. As noticed by Bulow and Roberts (1989), treating $1 - F(p)$ as quantity demanded and $G(p)$ as quantity supplied the virtual valuation function has the interpretation of marginal revenue and the virtual cost function that of marginal costs.\footnote{(Alternatively, one can think of a setup with multiple periods and multiple buyers and think of demand being $1 - F(k)$ rather than $1 - F(p)$.} The fee provided in Proposition 3 simplifies to

$$\omega(p) = p - E_v \left[ \Gamma_\alpha^{-1}(\Phi(v)) \left| v \geq p \right. \right]$$ (14)

Note that $\Gamma_\alpha^{-1}(v)$ is the procurement price a Ramsey monopsonist with valuation $v$
would set, thereby maximizing a weighted average of profits and revenue $(1-\alpha)W+\alpha\Pi$.\footnote{A monopolist with cost $c$ maximizes $\Pi = (p-c)(1-F(p))$ with respect to the price $p$, which implies the first-order condition $(c-\Phi(p))f(p) = 0$. A Ramsey monopsonist cares about his valuation $v$ and a weighted average of the price $p$ he pays and the seller’s expected utility $E[c|c \leq p]$. He maximizes $(v-(\alpha p + (1-\alpha)E[c|c \leq p]))G(p)$, which implies the first-order condition $(v-\Gamma_\alpha(p))g(p) = 0$.
}

For the special case $\alpha = 1$ the Ramsey monopsonist is simply a standard monopsonist. It is also useful to define the elasticity of demand as $\eta_d(v) = \vert -vf(v)/(1-F(v))\vert$. Similarly, the elasticity of supply is $\eta_s(c) = cg(c)/G(c)$. Observe that $\Gamma_\alpha(c) = c(1+\alpha/\eta_s(c))$ increases for all $c$ as $\alpha$ increases and decreases as $\eta_s$ increases.

Given (14), the comparative statics with respect to $\alpha$ and $\eta_s$ are straightforward. An overall increase of the elasticity of supply leads to lower fees, since $\Gamma^{-1}_\alpha$ increases. Similarly, an increase of $\alpha$ leads to higher fees. The effect of a change of $\eta_d$ is more complicated and will be discussed after we provide a price-theoretic interpretation in the following proposition.

**Proposition 5.** Using the Taylor expansion of $\Gamma^{-1}_\alpha(x)$ around $\bar{v}$, the net price received by the seller is

$$p - \omega_\alpha(p) = \Gamma^{-1}_\alpha(p) + \frac{[\Gamma^{-1}_\alpha(\bar{v})]''}{2} \text{Var}[\Phi(v) - \bar{v}|v \geq p] + \sum_{n=3}^{\infty} \frac{[\Gamma_\alpha(\bar{v})]^{(n)}}{n!} \{E[(\Phi(v) - \bar{v})^n|v \geq p] - E[\Phi(v) - \bar{v}|v \geq p]^n\},$$

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

The net price offered to the seller $p - \omega_\alpha(p)$ can be seen as follows. First, recall that a Ramsey monopsonist with valuation $x$ will offer the price $\Gamma^{-1}_\alpha(x)$ to the seller. The intermediary’s valuation for the good is $x = p$. However, this Ramsey monopsony price ignores that the transaction price $p$ is endogenously chosen by the seller rather than exogenously given. This is taken care of by the second- and higher-order price endogeneity effects.

For $G$ mirrored Generalized Pareto, $\Gamma^{-1}_\alpha$ is linear, which implies that the second- and higher-order price endogeneity effects disappear. Observe that a change of the elasticity
of demand has no first order effect on the optimal fee (e.g. for $G$ mirrored Generalized Pareto the fee is independent of the distribution of the buyer).

Note that for $p$ close $\bar{v}$, the second and higher order price endogeneity effects become negligible. For such prices, the seller walks a fine line between never selling ($p > \bar{v}$) and making a loss ($p-\omega(p) < c$). For the highest price $p = \bar{v}$, the net price $p - \omega(p) = \Gamma_{\alpha}^{-1}(\bar{v})$ is exactly equal to his cost $P^{-1}(p) = \Gamma_{\alpha}^{-1}(\bar{v})$.

Note that while increasing the elasticity of demand has no first-order effect on fees, the second- and higher-order effects can have either a positive or negative sign. As an illustration, assume $\alpha = 1$ and let $\Gamma_{1}^{-1}$ be quadratic, that is, $\Gamma_{1}^{-1}(x) = \gamma_{0} + \gamma_{1}(x - \bar{v}) + (\gamma_{2}/2)(x - \bar{v})^{2}$. This shuts down the higher-order effects and allows us to focus on the second-order price endogeneity effect. The fee is $\omega(p) = p - \Gamma_{1}^{-1}(p) - (\gamma_{2}/2)\text{Var}[\Phi(v) - \bar{v}|v \geq p]$. For $\Gamma_{1}^{-1}$ concave ($\gamma_{2} < 0$), an overall increase of the elasticity of demand can be shown to lead to an overall increase of the fee. This surprising result is counterintuitive at first sight: one would expect more elasticity of demand to lead to lower fees. 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 intermediary increases the fee. For $\Gamma_{1}^{-1}$ convex, an overall increase of the elasticity of demand leads to an overall decrease of the fee. Here, the intuition is that the seller lowers the price insufficiently as the elasticity of demand increases. The intermediary hence lowers fees to induce the seller to lower the price.

The 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 that prove useful in different contexts in

$^{25}$The distribution $G$ corresponding to a quadratic $\Gamma_{1}^{-1}$ is immaterial to our analysis, but can be easily reverse engineered: invert $\Gamma_{1}^{-1}()$ and solve the differential equation $\Gamma(c) = c + G(c)/g(c)$ with initial condition $G(\bar{v}) = 1$ for $G$. Of course, for $\gamma_{2} = 0$, the somewhat lengthy expression for $G$ simplifies to a mirrored Generalized Pareto distribution.

$^{26}$To see this, take a distribution $\hat{F}$ which has an overall higher elasticity than another distribution $F$, that is, $\hat{\eta}_{d}(v) > \eta_{d}(v)$ for all $v$. This implies $\hat{\Phi}(v) > \Phi(v)$, since $\Phi(v) = v(1 - 1/\hat{\eta}_{d}(v))$. This in turn implies $(\hat{\Phi}(v) - \bar{v})^{2} < (\Phi(v) - \bar{v})^{2}$ for all $v$, since $\Phi(v) < \bar{v}$. Further, $F$ hazard rate dominates $\hat{F}$, since $f(v)/(1 - \hat{F}(v)) = |v\hat{f}(v)/(1 - \hat{F}(v))|/v = \hat{\eta}_{d}(v)/v > \eta_{d}(v)/v = f(v)/(1 - F(v))$. This implies $E_{v}[v|v \geq p] \leq E_{v}[v|v \geq p]$ for all $p$. Together with $(\hat{\Phi}(v) - \bar{v})^{2} < (\Phi(v) - \bar{v})^{2}$, this implies $E_{v}[(\hat{\Phi}(v) - \bar{v})^{2}|v \geq p] \leq E_{v}[(\Phi(v) - \bar{v})^{2}|v \geq p]$. Therefore, fees are higher with $\hat{F}$ than with $F$, since $\gamma_{2} < 0$ and $\text{Var}[\Phi(v) - \bar{v}|v \geq p] = E[(\Phi(v) - \bar{v})^{2}|v \geq p] - E[\Phi(v) - \bar{v}|v \geq p]^{2} = E[(\Phi(v) - \bar{v})^{2}|v \geq p] - (p - \bar{v})^{2}$.
industrial organization. These include the curvature of the direct demand function, the curvature of the indirect demand function, the pass-through rate, and the markup- or quantity-weighted average pass-through. It is then natural to wonder whether the seemingly counterintuitive result that optimal fees are sometimes higher for a higher elasticity of demand may be explained by alternative properties of the demand function $F$. A look at the quadratic $\Gamma^{-1}$ example provided above shows that this is not the case: any change of any property of the demand function $F$ that leads to higher fees for $\gamma_2 < 0$ will lead to lower fees for $\gamma_2 > 0$ and to no change of the fees for $\gamma_2 = 0$.

Our results are highly relevant from a policy perspective: 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. It is therefore crucial which goods should be taxed more and which less heavily. In thick markets, it is well known that less elastically demanded goods should be taxed more heavily (see Salanie, 2003). Our results imply that for thin markets it is the elasticity of supply that matters. Depending on the curvature of $\Gamma^{-1}$, one should tax the more or less elastic good.

Our results also matter for competition policy. A first natural step when dealing with the suspicion of collusion (for example, in the case of Sotheby’s and Christie’s and in the case of the inconclusive investigation of real estate brokerage by the Department of Justice in 1983 and 2007) is to look at the elasticity of demand. If the price charged is close to the monopoly price implied by the elasticity of demand, then this could (potentially) indicate collusion. Our results suggest that in thin markets with intermediaries, the first look should be at the elasticity of supply of sellers 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, we should expect colluding intermediaries 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 first-order approximation).

It should be noted that the above observations about the elasticity of demand are about the fee (or tax) charged. When considering the gross price paid by the buyer, the
standard result that the buyer should pay a lower gross price if the good is less elastically demanded still holds. That is, the optimal gross price $P(c) = \Phi^{-1}(\Gamma_\alpha(c))$ decreases with the elasticity of demand. Further, the gross price decreases with the elasticity of supply and increases with the importance of revenue $\alpha$, just as the fee $\omega$.

### 2.6 Extensions

**Mechanism design with two-sided private information** The literature on Bayesian mechanism design with two-sided private information has studied a variety of settings that are related to our problem when we set $\delta = 0$. Myerson and Satterthwaite (1983)’s seminal paper analyzes a problem with one buyer and one, which corresponds to our setup with $\pi_1 = 1$.

Like Gresik and Satterthwaite (1989) and Gresik (1991), their analysis can be interpreted as analyzing incentive compatible, individual rational trading mechanisms that, for $\beta \in [0, 1]$, maximize the weighted sum

$$\beta W_I + (1 - \beta)(W_I + \sum_s W_s + \sum_b W_b),$$

where $W_I$ is the expected payoff of the intermediary (or mechanism designer), $W_s$ the expected payoff of seller $s$ and $W_b$ the expected payoff of buyer $b$. The case of ex post efficiency corresponds to $\beta = 0$ and broker optimality to $\beta = 1$. It is straightforward to extend our setup to account for these specifications as well. Let $\Phi_\beta(v) := v - \beta \frac{1-F(v)}{f(v)}$.

For $B \geq 1$ buyers and one seller the fee that implements with an essentially unique equilibrium the allocation rule of the optimal Bayesian mechanism given any $\beta \in [0, 1]$ is

$$\omega(p) = p - \frac{\int p \Gamma_\beta^{-1}(\Phi_\beta(v)) f(v) dv}{1 - F(p)}.$$

This encompasses, for example, ex post efficient mechanisms (for $\beta = 0$), the broker optimal mechanisms studied by Myerson and Satterthwaite (1983, Section V) (and, for one seller, by Baliga and Vohra (2003)), and the “second-best” mechanisms studied by Myerson and Satterthwaite (1983) do not impose the constraint of identical supports. However, this constraint can easily be relaxed if one wants to expand the analysis. Focusing on ex post efficiency, Makowski and Mezzetti (1993) show that such ex post efficient trade may be possible without a deficit when there are multiple buyers and $\overline{\tau} > \overline{\tau}$ holds. In what follows, we maintain the identical support assumption.
Myerson and Satterthwaite (1983, Section IV) and Gresik (1991) (and, for the case of one seller, those studied by Gresik and Satterthwaite (1989)), which maximize the expected welfare of buyers and sellers, subject to incentive and individual rationality constraints and subject to $W_I \geq 0$.\footnote{Fee-setting can also be extended to multiple sellers. However, the homogenous goods assumptions underlying the mechanism design approach will often be a stretch. Because of this and because of space constraints, we do not do so here.}

As noted by Myerson and Satterthwaite (1983), for the bilateral trade problem with $F$ and $G$ uniform on $[0, 1]$ the second-best mechanism can be implemented in the linear equilibrium of the double-auction auction of Chatterjee and Samuelson (1983). With fee-setting, it can be implemented with the fee $\omega(p) = p/2 - 1/4$.

**Optimal Linear Fees and Pricing by Direct and Indirect Sellers** We have seen that optimal fees are linear if the seller’s cost $c$ is drawn from generalized Pareto distribution, that is, if $G(c) = ((c - \underline{c})/(\overline{c} - \underline{c}))^\sigma$. Let $G^\alpha_p(p)$ denote the distribution of reserve prices when the weight on the intermediary’s profit is $\alpha$. We now state a result that has a number of interesting and useful implications.

**Proposition 6.** The distributions of reserve prices $G^\alpha_p(p)$ and the distribution and the mean of transaction prices do not vary with $\alpha$ if optimal fees are linear.

The intuition for Proposition 6 is that mirrored Generalized Pareto distributions are invariant to truncation from the right and to re-scaling, which is exactly what happens with additional seller entry as linear fees decrease.

Proposition 6 also has a number of implications that are relevant in the context of empirical research on real estate. 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 FSBO platforms. In light of the double marginalization that
occurs in intermediated trade, this finding is puzzling at first sight. However, as shown in Proposition 6 it is an immediate implication of our theory if one assumes that the demand side is the same on FSBO platforms as in intermediated trade and that there is no differential selection of sellers into the two market segments (other than the additional entry in the FSBO market). Proposition 6 further implies that if proportional fees $\omega(p) = bp$ are optimal, then the distribution of reserve and transaction prices does not vary with $b$.

**Non-stationary environments** We have assumed a stationary environment as it simplifies and clarifies the exposition. However, as we explain now, nothing of substance hinges on this simplifying assumption. To that end, let us assume now that the environment is described by known sequences $\delta = (\delta_t)_{t=0}^{\infty}$ and $F = (F_t)_{t=0}^{\infty}$, where $\delta_t$ is the discount factor in period $t$ and $F_t$ is the distribution from which buyers’ types are drawn in period $t$ and that for all $t$, $\Phi_t(v) = v - \frac{1 - F_t(v)}{F_t(v)}$ is monotone in $v$. One example is the exponential discounting (or constant drop out probability) $\delta_\tau = \delta$ considered so far. Another is exponential discounting up to a deadline $T$ after which the seller leaves the market for sure ($\delta_\tau = \delta$ for $\tau \leq T$ and $\delta_\tau = 0$ for $\tau > T$). Let $\pi_B$ describe the arrival process of buyers in period $t$ and denote by $F_{(1),t}(v)$ and $F_{(2),t}(v)$ the distributions of the highest and second-highest draw in $t$. Expected revenue given reserve $p$ in period $t$, conditional on trade in period $t$, is then given as $R_t(p) = \int_0^p \Phi_t(v) dF_{(1),t}(v) 1 - F_{(1),t}(p)$.

Given a sequence $k$ of expected transaction prices conditional on trade, the seller’s discounted ultimate probability of trade is still given as $\sum_{t=0}^{\infty} q_t(k)$, where

$$q_t(k) := (1 - F_{(1),t}(R_t^{-1}(k_t))) \prod_{\tau=0}^{t-1} \delta_\tau F_{(1),\tau}(R_\tau^{-1}(k_\tau)).$$

Next define $1 - F(k) := \lim_{T \to \infty} 1 - F_T(k)$, where $1 - F_T(k)$ is the maximum of $\sum_{t=1}^{T} q_t(k)$ subject to the constraint $\sum_{t=0}^{T} q_t(k) k_t / (\sum_{t=0}^{T} q_t(k)) = k$, as defined in the proof of Proposition 3. At date 0, the objective function that accounts for incentive compatibility provided the pointwise maximizer $k(c)$ of the integrand is monotone is then still given

29 In intermediated trade, a seller of type $c$ optimally sets the price $\tilde{\Phi}^{-1}(\Gamma_c(c)) > \Phi^{-1}(c)$, where $\Phi^{-1}(c)$ is the price set by the same seller when selling directly.
by (8), yielding the allocation rule allocation rule (31). Consequently, the functional
form of the expectational fees \( \omega(k) \) under non-stationarity will be the same as under
stationarity. Hence, it is as given in Proposition 2. This also implies that in the limit,
as \( G \) converges to a generalized Pareto distribution, the optimal expectational fee will
be linear as in the stationary case.

Letting \( \tilde{\Phi}_t(p) = R_t(p) - (R_t(p) - \Phi_t(p)) \frac{1 - \delta_t}{1 - \delta_t} \) and denoting by \( \delta_t V_t(c) \) the option
value future trade for a seller of type \( c \), discounted to period \( t \), the optimal transaction
fee in period \( t \) is then given as

\[
\omega_t(p) = p - \frac{\int_{\tilde{p}} [\Gamma^{-1} \tilde{\Phi}_t(v)] + \delta_t V_t(\Gamma^{-1} \tilde{\Phi}_t(v))] f_t(v) dv}{1 - F_t(p)}.
\]

Although \( \omega_t(p) \) will in general vary over time because the environment is non-stationary,
the linearity of the expectational fees \( \omega \) in the limit implies that the optimal (normalized)
transaction fees will be linear in the limit too.

**Linear Fees, First-Price Auctions and the Informed Principal Problem**

Given linear transaction fees \( \omega \), the payoff of the seller upon selling at some price \( \tilde{p} \)
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. 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 in-
dependent 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).

\(^{30}\)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).
3 Empirical Analysis

3.1 Data

The data set we use is the one constructed and used by Genesove and Mayer (2001). These 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. Importantly, where available the data contain the sale price of previous transactions, which Genesove and Mayer (2001) also used to account for unobserved property heterogeneity in constructing a quality-adjusted price as discussed below. 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. We have 5792 observations in total.

As we are adding complexity to our model by solving for the optimal mechanism, we prefer to reduce complexity in other dimensions. In particular, we use a model that assumes a stationary environment. For stationarity to be a plausible assumption, we only use data from April 1, 1993 to April 1, 1996 because the changes in the real estate price index were relatively modest during this period. As an additional measure to reduce the effect of price changes over time, we run separate estimations for individual years, that is the year starting from April 1, 1993; from April 1, 1994; and from April 1, 1995. There is a trade-off when choosing the time length of the interval for which one does a separate estimation: a shorter interval reduces the effect of price changes over time and increases the weight of the effect of cross-sectional price variation. However, it also leads to fewer observations per estimation.\footnote{Another effect is that some property is listed at the end of one period and sold at the beginning of another. As the results of the counterfactual analysis performed for different years are similar, we expect this effect to be small.} A calendar year also 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. 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 1 shows the standard deviation of the price index within the time interval, the standard deviation of quality adjusted prices, the change of the price index, and the slope of time on market in price. Fig. 3 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-sectional. 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 dataset.

There are 2455 observations between April 1, 1993 and April 1, 1996. We 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. The reason for the exclusion is that a property that is offered at less than half of or more than two times 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 dataset. Similarly, a property on the market for more than two years was probably not seriously marketed. This exclusion does not change the estimation results qualitatively as shown in the robustness checks reported in the appendix.

Table 2 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.

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 Hendershot (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).
### Measures of Stationarity

<table>
<thead>
<tr>
<th>Year</th>
<th>Index Std/Mean</th>
<th>Price Std/Mean</th>
<th>Relative ΔIndex</th>
<th>Slope Time on Market – Price Coeff. ( Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>0.110</td>
<td>0.199</td>
<td>-0.241</td>
<td>-87.6 (28.2)</td>
</tr>
<tr>
<td>1991</td>
<td>0.021</td>
<td>0.204</td>
<td>-0.030</td>
<td>39.0 (20.9)</td>
</tr>
<tr>
<td>1992</td>
<td>0.051</td>
<td>0.214</td>
<td>0.131</td>
<td>67.0 (21.5)</td>
</tr>
<tr>
<td>1993</td>
<td>0.016</td>
<td>0.203</td>
<td>0.006</td>
<td>21.3 (22.0)</td>
</tr>
<tr>
<td>1994</td>
<td>0.020</td>
<td>0.185</td>
<td>-0.003</td>
<td>19.2 (22.0)</td>
</tr>
<tr>
<td>1995</td>
<td>0.026</td>
<td>0.190</td>
<td>0.067</td>
<td>40.1 (20.2)</td>
</tr>
<tr>
<td>1996</td>
<td>0.041</td>
<td>0.204</td>
<td>0.097</td>
<td>-7.6 (15.6)</td>
</tr>
<tr>
<td>1997</td>
<td>0.020</td>
<td>0.201</td>
<td>0.044</td>
<td>7.2 (9.8)</td>
</tr>
</tbody>
</table>

Table 1: Standard deviation of the real estate index divided by its mean in a given time interval, standard deviation of the quality adjusted price divided by its mean in a given time interval, change of the index divided by the index at the beginning of the interval, and slope time on market – quality adjusted price. The slope is the coefficient $\beta_1$ in the regression $T = \beta_0 + \beta_1 P + \epsilon$, where $T$ is the time on market and $P$ the quality adjusted price.

(2004, Table 1), who use data from two regions in the United Kingdom (UK). Unsold houses stayed longer on the market than houses that did sell. If a house is delisted and relisted within four weeks, it is considered to be the same transaction. For further details about the data, see Genesove and Mayer (2001).

### Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>All Houses</th>
<th>Sold Houses</th>
<th>Unsold Houses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>2333</td>
<td>1522</td>
<td>811</td>
</tr>
<tr>
<td>Listing Price</td>
<td>$223,077 ($177,736)</td>
<td>$231973 ($172,861)</td>
<td>$206,383 ($185,501)</td>
</tr>
<tr>
<td>Quality Adjusted Listing Price</td>
<td>1.139 (0.219)</td>
<td>1.125 (0.220)</td>
<td>1.165 (0.215)</td>
</tr>
<tr>
<td>Transaction Price Listing Price</td>
<td>92%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time on Market</td>
<td>148 (134) days</td>
<td>130 (123) days</td>
<td>182 (147) days</td>
</tr>
</tbody>
</table>

Table 2: Sample Means (Standard Deviations) of Descriptive Statistics
Figure 3: Development of the real estate price index in Boston. Data between the two dashed lines were used for estimations.

3.2 Data Generating Process

The data generating process is essentially given by the theoretical model introduced and analyzed in Section 2 above. We make minimal adjustments to the model by assuming that the arrival process of buyers is a Poisson process by allowing for both observable and unobservable heterogeneity of real estate, and by taking into account that time on market is observed with an error. We also need to make an assumption about how sellers enter and exit the market. Our discussion begins with the latter.

Market in Steady-State The assumption underlying our empirical analysis is that the market is in a steady state. That is, in every period the distribution and mass of sellers who exit are the same as the distribution and mass of sellers who enter. However, the behavior we observe and draw inferences from is given by the steady-state distribution of sellers. As noticed in the theoretical analysis, provided \( \delta > 0 \) there is a difference between a seller’s “static” type – that is, a seller’s opportunity cost of selling in a one-
shot game – and his “dynamic” opportunity cost of selling in a given period, which will
depend both on his static type and factors that are endogenous to the model, such as the
broker’s fee. In the following, we lay out the detail of how we back out the distributions
of entrants’ dynamic types and the distributions of static costs of both entrants and in
the pool of sellers who are active in the steady state.

Denote by $c$ a seller’s static, exogenously given cost and by $c_D$ his dynamic cost. The
distribution of entrants’ static costs is denoted $G(c)$ and the steady-state distribution of
static costs is denoted $G_S(c)$. The distribution of entrants’ dynamic costs is denoted
$G_D(c_D)$ and the steady-state distribution of dynamic costs is denoted $G_{SD}(c_D)$, whose
support will be denoted $[c_D, \bar{c}_D]$. This $G_{SD}(c_D)$ is the distribution we are going to
estimate. At the end of this section, we explain in detail how to back out $G(c)$.

Poisson Process and Time on Market For the empirical analysis, we assume that
$\pi_B$ is given by a Poisson process with arrival rate $\lambda$. This means that the probability
$\pi_B$ of there being $B$ buyers with $B = 0, 1, 2, ..$ in a given period is $\pi_B = e^{-\lambda B} / (B!)$. Accordingly, the probability that a seller who sets the reserve price $p$ does not sell in a
given period is

$$F(1)(p) := \sum_{B=0}^{\infty} \pi_B F(p)^B = e^{-\lambda(1-F(p))}. \quad (16)$$

Note that (16) implies that $F(p) = 1 + \ln(F(1)(p))/\lambda$. This allows us to express the
virtual valuation $\Phi(v)$ without $\lambda$ and $F$ and in terms of $F(1)$ only:

$$\Phi(v) = v + \frac{F(1)(v)}{f(1)(v)} \ln(F(1)(v)),$$

where $f(1)$ is the density of $F(1)$.

The time on market $t$ of a property has a geometric distribution with the distribution
function $1 - (\delta F(1)(p))^{t/\tau}$ and mean

$$T(p) = \frac{\tau}{1 - \delta F(1)(p)}. \quad (17)$$

\footnote{The reason that we are estimating $G_{SD}$ rather than $G$ is on one hand that it is simpler, on the other hand that we already estimated $G_{SD}$ in a previous version of the paper with a different counterfactual. This previous counterfactual was based on the assumption that the seller and the broker write a myopic contract.}
The probability that a property sells in period $t$ is $(1 - F_{(1)}(p))(\delta F_{(1)}(p))^{t/\tau}$. Taking the sum of this geometric series over $t$ from 0 to infinity gives us the probability $1 - F_{\infty}(p)$ that a house that is offered at price $p$ is ever sold:

$$1 - F_{\infty}(p) = \frac{1 - F_{(1)}(p)}{1 - \delta F_{(1)}(p)}. \quad (18)$$

**Accounting for Observable and Unobservable Heterogeneity and Measurement Errors** When taking the model to the data, we have to control for the differential quality of properties, which is only partially observable by the econometrician, and for measurement errors in the listing price, which is different from the reserve price and in the time on the market. We now describe in turn how we do this.

We use the quality index constructed by Genesove and Mayer (2001) which is based on previous transaction prices and a quarterly real estate price index $P_{q}^{\text{index}}$. Formally, in quarter $q$ the measured “objective” quality index $\hat{\vartheta}_i$ of house $i$ which was previously traded in quarter $q' < q$ for the price $\hat{P}_{iq'}$ is

$$\hat{\vartheta}_i = \frac{P_{q}^{\text{index}}}{P_{q'}^{\text{index}}} \hat{P}_{iq'}. $$

We assume that the true quality index $\vartheta_i$ is measured with a multiplicative error $\epsilon_{Q}^{i}$, which may stem from an imperfect measurement of the quality index or changes in the quality of a property, that is, the measured quality is $\hat{\vartheta}_i = \vartheta_i \epsilon_{Q}^{i}$.

While the price index $P_{q}^{\text{index}}$ is adjusted every quarter, our analysis is based on annual data. In what follows we therefore drop the time-index for individual observations.

Although it would be interesting to have a complex empirical model that takes into account many effects besides those of our model, it is also interesting and informative to analyze how well our baseline model can explain observations if we add two minimal modifications to this model.

First, the listing price is an imperfect proxy of the seller’s reserve price as a sizable number of properties sell below the listing price. We account for this by assuming that

---

34 This measure of quality adjustment is appropriate if all house prices move in proportion from one year to another. It neglects the impact of the seller specific type at time $q'$ and of structural changes in demand from $q'$ to time $q$. 

---
the true quality-adjusted reserve price of seller $i$, denoted by $p_i$, involves a multiplicatively
error term $\epsilon_i^D$, which may be thought of as a “discount” on the listing price. If $P_i$ is
the observed listing price of object $i$, the observed quality- and discount-adjusted reserve
price $p_i$ satisfies $P_i = p_i \epsilon_i^D \hat{\vartheta}_i$, or equivalently,
\[ p_i = \frac{P_i}{\epsilon_i^D \hat{\vartheta}_i} = \frac{P_i}{\epsilon_i^T \vartheta_i}, \]
where the error of the proxy is $\epsilon_i^T := \epsilon_i^Q \epsilon_i^D$. We denote the density of $\epsilon_i^T$ as $h_p(\epsilon_i^T)$ and assume that it is log-normal.

Second, the true time on market $t_i$ is also observed with an error, denoted $\epsilon_i^T$, so that
\[ T_i = t_i + \epsilon_i^T, \]
where $T_i$ is the observed time on the market. The error $\epsilon_i^T$ 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 (described above), 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.

Denote the joint density of prices and times on market as predicted by our baseline
model as $h_{tps}(p,t,s)$ and the density of the error term for the time on market $\epsilon_i^T$ as $h_t(\epsilon_i^T)$. Let $s_i$ be 1 if the house was sold and 0 otherwise.

First, we describe the prediction of the baseline model without error terms. Note that
there are two ways of calculating the optimal price set by the seller. The first is based
on the static costs of the seller and is $\Phi^{-1}(c/0.94)$ as described in the previous analysis.
The second, equivalent way is based on dynamic costs $c_D = c + \delta V(c)$: the seller sets
the optimal reserve in a one-shot auction, but uses his dynamic rather than static costs,
that is, the price is $\Phi^{-1}(c_D)$. In the following we will use this second way. The empirical
inverse price function is denoted by $P_I(p) = 0.94\Phi(p)$. This function gives us the cost $c_D$
of a seller who will optimally set the price $p = \Phi^{-1}(c_D/0.94)$. As sellers’ dynamic costs
have the steady-state density $g_{SD}(c_D)$, the steady state density of prices is proportional
to $g_{SD}(P_I(p))P_I'(p) =: g_p(p)$. Sellers that spend a long time on the market are over
represented in steady state compared to the entrant population. Hence, to obtain the
entrant distribution of prices, we have to divide by the average time on market, denoted $T(p)$. The entran
t density of prices is therefore given as $\sigma g_{SD}(P_1(p))P'_1(p)/T(p) =: g_{p0}(p)$, where $\sigma$ is a constant that ensures that the density adds up to 1. 

Since rematching occurs every $\tau$ periods, and a house drops out of the market with probability $1 - \delta$ and is sold with probability $1 - F_{(1)}(p)$, the probability that a house is still on the market after $t$ periods is $(\delta F_{(1)}(p))^{t/\tau}$. Hence the joint distribution of $p$, $t$, and $s$ has the density

$$h_{tps}(t, p, s) = \begin{cases} (1 - F_{(1)}(p))(\delta F_{(1)}(p))^{t}g_{p0}(p) & \text{if } s = 1, \\ (1 - \delta)F_{(1)}(p)(\delta F_{(1)}(p))^{t}g_{p0}(p) & \text{if } s = 0. \end{cases}$$

As the quality-adjusted reserve price $p$ and time on market $t$ are observed with noise $\epsilon^P$ and $\epsilon^T$, the likelihood of an observation $X_i = (T_i, P_i, s_i)$ given the parameter vector $\theta$ as specified below is

$$l(X_i|\theta) = \sum_{k=1}^{T_i/\tau} \int_{-\infty}^{\infty} h_{tps}(T_i - k\tau, P_i/\epsilon^P, s_i) h_t(k\tau) h_p(\epsilon^P) d\epsilon^P$$

where $k\tau$ is the summation variable representing the error term $\epsilon^T$.

### 3.3 Identification

**Identification without measurement errors** If the quality-adjusted price and time on market were observable without the errors $\epsilon^P$ and $\epsilon^T$, it would be easy to see that our model is non-parameterically identifiable given observations of quality-adjusted price, time on market and whether a house was sold: Rearranging (17) and (18), 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

$$1 - F_{(1)}(p) = \frac{1 - F_{\infty}(p)}{T(p)/\tau},$$

$$\delta = \frac{T(p_2) - T(p_1)}{T(p_2)(1 - F_{\infty}(p_1)) - T(p_1)(1 - F_{\infty}(p_1))},$$

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

\[35\text{In the matching model, } \sigma \text{ is the mass or stocks of sellers in the market.}\]
where $p_1$ and $p_2$ are two arbitrary prices (or – with some modification of the equations – price segment)\textsuperscript{36}. This makes $F(1), \epsilon$, and $\tau$ non-parametrically identifiable. Knowing $F(1)$, the distribution of sellers’ costs $G_{SD}$ is non-parametrically identifiable by the distribution of prices $G_p$ via the relationship $G_{SD}(c_D) = G_p(\Phi^{-1}(c_D/0.94))$ because a seller with cost $c$ sets the optimal reserve price $\Phi^{-1}(c_D/0.94)$. We will discuss later on how to back out the entrant static cost distribution $G$ from the steady-state dynamic distribution $G_{SD}$. Given $F(1)$ and $G$, our theory provides the unique best-response fee-setting mechanism of the broker.

**Identification with measurement errors** With errors in the measurement of the quality adjusted discounted reserve price $\epsilon^P$ and of time on market $\epsilon^T$, 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 $P_i/\vartheta_i$ 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 $P_i/\vartheta_i$ and $\epsilon_i^P$ is $\text{Prob}(s = 1|P_i/\vartheta_i, \epsilon_i^P) = (1 - F(1)(P_i/\vartheta_i, \epsilon_i^P))/\delta s = 0|P_i/\vartheta_i, \epsilon_i^P)) = (1 - \delta)F(1)(P_i/\vartheta_i, \epsilon_i^P))/\delta s = 0|P_i/\vartheta_i, \epsilon_i^P)).$ Given the unconditional density $h_p(\epsilon^P)$, the conditional densities are $h_p(\epsilon^P|P_i/\vartheta_i, s = 1) \propto h_p(\epsilon^P)/\text{Prob}(s = 1|P_i/\vartheta_i, \epsilon_i^P)$ and $h_p(\epsilon^P|P_i/\vartheta_i, s = 0) \propto h_p(\epsilon^P)/\text{Prob}(s = 0|P_i/\vartheta_i, \epsilon_i^P)$ by Bayes’ Law. This gives us

$$1 - F_\infty(P_i/\vartheta_i) = E_{\epsilon_i^P \sim h_p} \left[ \frac{1 - F(1)(P_i/\vartheta_i, \epsilon_i^P))}{1 - \delta F(1)(P_i/\vartheta_i, \epsilon_i^P))} \right],$$

$$\hat{T}^s(P_i/\vartheta_i) = E_{\epsilon_i^P \sim h_p} \left[ \tau \frac{1 - \delta F(1)(P_i/\vartheta_i, \epsilon_i^P))}{1 - \delta F(1)(P_i/\vartheta_i, \epsilon_i^P))} \right] + E[\epsilon^T],$$

$$\hat{T}^u(P_i/\vartheta_i) = E_{\epsilon_i^P \sim h_p} \left[ \tau \frac{1 - \delta F(1)(P_i/\vartheta_i, \epsilon_i^P))}{1 - \delta F(1)(P_i/\vartheta_i, \epsilon_i^P))} \right] + E[\epsilon^T].$$

\textsuperscript{36}The simplest example would be a price that never leads to trade and a price that leads to instantaneous trade, $p_2 = v$ and $p_1 = \vartheta$. This simplifies the expressions to $\tau = T(v)$ and $\epsilon = (T(\vartheta) - T(v))/T(\vartheta)$. In practice, one would want to take two different price segments and expectations over them, rather than two prices.
Note that (22) and $\hat{T}^u(P_i/\vartheta_i) - \hat{T}^s(P_i/\vartheta_i)$ do not require any knowledge about the distributions of $\epsilon^T$ and $c$ and identify $F_{(1)}$ and $H_p$ for given $\epsilon$ and $\tau$. The density of time on market $t$ conditional on a particular price $P_i/\vartheta_i$, $E_{\epsilon^P \sim H_p, \epsilon^T \sim H_t}[(\delta F_{(1)}(P_i/(\vartheta_i \epsilon^P)))^{t/\tau - \epsilon^T}|\epsilon^T \leq t/\tau]$ identifies $H_t$. This uses only one price $P_i/\vartheta_i$. The different densities of $t$ for two additional prices $P_j/\vartheta_j$ and $P_l/\vartheta_l$ pin down $\epsilon$ and $\tau$. Only the seller's distribution $G_{SD}$ remains to be identified. It is non-parametrically identifiable in the same way as without measurement errors $\epsilon^P$ and $\epsilon^T$ because $G_{SD}(c) = G_p(\Phi^{-1}(c/0.94))$.

3.4 Estimation

While our model is non-parametrically identifiable in principle, our estimation is based on a parametric specification of the model and on Bayesian estimation methods. The main reason is that our main hypothesis is – loosely speaking – that if one was to pick distributions $F$, or equivalently $F_{(1)}$, and $G_{SD}$ randomly, then “most of the time” (or “on average”) linear fees would perform close to optimally. Drawing $F_{(1)}$ and $G_{SD}$ from the Bayesian posteriors distribution is a natural choice and also provides a clearer meaning to “on average”. Further, the standard maximum likelihood estimator is biased because of the nuisance parameters and possible discontinuities in the likelihood function. Using a Bayesian estimator avoids the problems arising for the maximum likelihood estimator.

For our estimation we make the following functional form assumptions and take the following parametrization. We assume that $F_{(1)}(v)$ and $G_{SD}(c)$ are Beta distributions in the sense that $(v - v_i)/(\overline{v} - v_i)$ and $(c - c_i)/(\overline{c} - c_i)$ are Beta-distributed with respective parameters $(\alpha_F, \beta_F)$ and $(\alpha_G, \beta_G)$, where the density of the Beta distribution for $x_i$ is proportional to $x_i^{\alpha_i-1}(1-x_i)^{\beta_i-1}$ with $i = F, G$ and $x_F = (v - v)/(/\overline{v} - v)$ and $x_G = (c - \overline{c})/(c - \overline{c})$. The error in time on market $\epsilon^T$ follows a geometric distribution with parameter $\beta_T$, whose probability mass function is proportional to $e^{-\epsilon^T/\beta_T}$. Finally, the error in the quality-adjusted price $\epsilon^P$ is assumed to be normally distributed with mean 0 and variance $\sigma^2_p$. The advantage of using Beta distributions is that they are flexible in shape and specialize to linear virtual cost and valuation functions for $\beta_G = 1$ and $\alpha_F = 1$, respectively. The vector of parameters is thus $\theta = (\alpha_F, \beta_F, \alpha_G, \beta_G, \beta_T, \sigma_p, \epsilon, v, \overline{v}, c, \overline{c}, \lambda)$.

Note that $\theta$ does not include the period length $\tau$, which is the only parameter that we cannot
We set \( c = v \) to simplify the computations and \( \bar{c} = (1 - \xi_{\text{empirical}}) \bar{v} \) and the empirical percentage fee \( \xi_{\text{empirical}} \) to 0.06.

Given observations \( X = \{X_i\}_{i=1}^N = \{(T_i, P_i, s_i)\}_{i=1}^N \) and the parameter vector \( \theta \), the likelihood function for the \( N \) observations is \( l(X|\theta) := \prod_i l(X_i|\theta) \), where \( l(X|\theta) \) is the probability of observing \( X \) given \( \theta \). The unconditional probability of observing \( X \) is denoted by \( l(X) \). We are searching for the posterior beliefs about the parameters \( \theta \) given \( X \), \( \pi(\theta|X) \) \( ^{38} \) By Bayes’ Law \( \pi(\theta|X) = l(X|\theta)\pi(\theta)/l(X) \), where \( \pi(\theta) \) is the prior about \( \theta \). Assuming a uniform prior \( \pi(\theta) \), we obtain the proportionality \( \pi(\theta|X) \propto l(X|\theta) \) as \( l(X) \) does not depend on \( \theta \) \( ^{39} \).

We are seeking to find Bayesian estimates of the mean and variance of several functions \( y(\theta) \), that is, \( E_{\theta \sim \pi(\theta|X)}[y(\theta)] \) and \( \text{Var}_{\theta \sim \pi(\theta|X)}[y(\theta)] \). These functions \( y(\theta) \) are the ratio the weight \( \alpha(\theta) \) that rationalizes the choice of a 6% fee in the class of percentage fees, the ratio of the objective function for 6% fees compared to the optimal mechanism, and the fee a hypothetical monopolistic broker would set. In the current version of the paper, we only evaluate these functions for the Bayesian mean, i.e. \( y(E_{\theta \sim \pi(\theta|X)}[\theta]) \), for the sake of computational simplicity. We are currently working on computing \( E_{\theta \sim \pi(\theta|X)}[y(\theta)] \) and \( \text{Var}_{\theta \sim \pi(\theta|X)}[y(\theta)] \).

Computing the expectations requires computing a 10-dimensional integral. This is numerically challenging as a simple approach would lead to a computational time of several years \( ^{40} \) We use several numerical improvements to reduce computational time to a few hours.

While our main interest is the Bayesian posterior beliefs about the different variables, estimate. This is partly because we use a discrete time model, so the distribution of the error in time on market \( \epsilon^T \) cannot be compared across different \( \tau \). Hence comparisons of the likelihood function would not make sense. A remedy would be to use continuous time errors \( \epsilon^T \). We chose to run robustness checks with alternative values of \( \tau \).

\(^{38}\)We use lower case \( \pi \) to denote beliefs and, as above, upper case \( \Pi \) to denote expected profits.

\(^{39}\)To be precise, we constrain the uniform prior \( \pi(\theta) \) to be the same positive constant wherever the constraints that we impose to avoid numerical problems are satisfied and to be 0 wherever they are not. These constraints are: (a) virtual valuation/cost functions must be increasing (to avoid the need for ironing), (b) \( \alpha_F, \beta_F, \alpha_G, \beta_G \geq 1 \) (to avoid infinite densities at endpoints of \( f \) and \( g \)), (c) \( \bar{\tau} - \bar{\mu} \geq 0.2 \) (to avoid numerical problems with nearly degenerate distributions that arise when \( \bar{\tau} \approx \bar{\mu} \)).

\(^{40}\)This estimate is based on extrapolation of the time needed to solve part of the problem. In addition, standard Markov Chain Monte Carlo techniques for multidimensional integration do not work in our setup.
for illustration purposes (for example, to plot the predicted distribution of prices or the relation of price and time on market) we also want a pointwise estimator for $\theta$. For this we take the Maximum A Posteriori Probability estimator $\theta_{MAP} = \arg \max_\theta \pi(\theta|X)$, that is, the mode of the Bayesian posterior distribution.\textsuperscript{41}

Parameter Estimates  In Table 3 we report the Bayesian estimates for the parameter vector $\theta$ under the assumption that there are seven matchings per week (that is, $\tau = 1/365$). We report the Bayesian mean (for example $E_{\theta \sim \pi(\theta|X)}[\alpha_F]$ in the first row), the Bayesian standard deviation (for example $\sqrt{\text{Var}_{\theta \sim \pi(\theta|X)}[\alpha_F]}$ in parentheses in the first row), and the computational error (in square brackets). The computational error is as reported by the algorithm in Hahn (2005). It provides the 99% confidence interval for the results of the (quasi-)Monte Carlo integration used when computing the Bayesian estimates.\textsuperscript{42}

Figure 4(a) provides a graphical illustration of parameter estimates for the Maximum A Posteriori Probability estimates of $\theta$ for 1993. In particular, it displays the estimated endogenous densities $f(v)$ (dashed line) and $g_{SD}(c)$ (solid line). Panels (b) to (f) illustrate how well predictions of the model fit the observed data. Panels (c), (d) and (f) show, respectively, the relationships between time on the market, and quality-adjusted price, density of quality adjusted prices, and the probability of selling. In each case, the empirically observed relationship is displayed as a dashed line while the relationship that our model implies when evaluated at the estimated parameter values is shown as a solid line. The densities of time on the market for sold and unsold houses are displayed, respectively, with solid and dashed lines, in panel (b) as implied by our model and the

\textsuperscript{41}Alternatively, one could use the Bayesian mean $\theta^* = E[\theta]$. However, some of the constraints imposed on $\theta$ (for example, increasing virtual valuations) by setting $\pi(\theta|X) = 0$ where constraints are violated, may not be satisfied at $\theta^*$. They hold for $\theta_{MAP}$. Note that for uniform priors, the Maximum A Posteriori Probability estimator coincides with the Maximum Likelihood Estimator.

\textsuperscript{42}For example, for the first column in the first row, if the correct value of the integral $\int \alpha_F \pi(\theta|X) d\theta$ was above 2.706 or below 2.702, then the null hypothesis that it is a correct calculation would be rejected with a probability of at least 99%. Note the difference to the Bayesian standard deviation of the estimate, which is determined by the (fixed) number of observations. The computational error depends on the number of computational steps. The algorithm increases the number of computational steps until the length of the confidence interval is below the desired level. While it is common not to report the computational error separately, it is useful to do so as it is an indicator of the reliability of the computational method.
## Estimated Parameter Values

<table>
<thead>
<tr>
<th>Parameters</th>
<th>1993</th>
<th>1994</th>
<th>1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_F$</td>
<td>$2.704 (0.356)$</td>
<td>$6.622 (0.855)$</td>
<td>$6.697 (1.119)$</td>
</tr>
<tr>
<td></td>
<td>[0.002]</td>
<td>[0.003]</td>
<td>[0.006]</td>
</tr>
<tr>
<td>$\beta_F$</td>
<td>$3.696 (0.958)$</td>
<td>$8.563 (0.344)$</td>
<td>$8.510 (0.460)$</td>
</tr>
<tr>
<td></td>
<td>[0.003]</td>
<td>[0.004]</td>
<td>[0.007]</td>
</tr>
<tr>
<td>$\alpha_G$</td>
<td>$8.240 (0.615)$</td>
<td>$5.381 (1.141)$</td>
<td>$5.238 (1.709)$</td>
</tr>
<tr>
<td></td>
<td>[0.006]</td>
<td>[0.003]</td>
<td>[0.003]</td>
</tr>
<tr>
<td>$\beta_G$</td>
<td>$6.373 (1.998)$</td>
<td>$8.633 (2.017)$</td>
<td>$7.893 (2.049)$</td>
</tr>
<tr>
<td></td>
<td>[0.006]</td>
<td>[0.004]</td>
<td>[0.005]</td>
</tr>
<tr>
<td>$\beta_T$</td>
<td>$13.96 (0.64)$</td>
<td>$11.676 (0.657)$</td>
<td>$12.204 (0.677)$</td>
</tr>
<tr>
<td></td>
<td>[0.1]</td>
<td>[0.006]</td>
<td>[0.01]</td>
</tr>
<tr>
<td>$\sigma_p$</td>
<td>$0.2060 (0.0028)$</td>
<td>$0.18586 (0.00226)$</td>
<td>$0.1894 (0.0025)$</td>
</tr>
<tr>
<td></td>
<td>[0.0001]</td>
<td>[9.e-05]</td>
<td>[0.0001]</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>$0.9683 (0.0016)$</td>
<td>$0.9707 (0.0015)$</td>
<td>$0.9750 (0.0019)$</td>
</tr>
<tr>
<td></td>
<td>[0.0006]</td>
<td>[0.0005]</td>
<td>[0.0008]</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$0.7605 (0.0497)$</td>
<td>$0.9490 (0.0193)$</td>
<td>$0.9105 (0.0316)$</td>
</tr>
<tr>
<td></td>
<td>[0.0005]</td>
<td>[0.0005]</td>
<td>[0.0007]</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$1.3492 (0.0458)$</td>
<td>$1.3383 (0.0293)$</td>
<td>$1.351 (0.024)$</td>
</tr>
<tr>
<td></td>
<td>[0.0009]</td>
<td>[0.0006]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$0.5743 (0.0764)$</td>
<td>$0.3098 (0.1032)$</td>
<td>$0.5331 (0.1218)$</td>
</tr>
<tr>
<td></td>
<td>[0.0004]</td>
<td>[0.0002]</td>
<td>[0.0003]</td>
</tr>
<tr>
<td># Observations</td>
<td>720</td>
<td>831</td>
<td>782</td>
</tr>
</tbody>
</table>

Table 3: Bayesian estimates of parameter values for 1993 - 1995 for seven matchings per week. Table entries read: Mean (Standard Deviation) [Computational Error].
Figure 4: Theory and Empirics (1993, 7 matchings per week). The graphs are based on maximum a posteriori probability estimates of the parameters. (a) Endogenous densities $f(v)$ (dashed) and $g_{SD}(c)$ (solid), (b) model prediction for densities time on market for sold (solid) and unsold (dashed) houses, (c) time on market as a function of quality-adjusted price, empirical (kernel regression, dashed) and predictions of model (solid), (d) density of quality-adjusted price, empirical (kernel density estimator, dashed) and model prediction (solid), (e) empirical densities time on market (kernel density estimator) for sold (solid) and unsold (dashed) houses, (f) probability of ever selling, empirical (kernel regression, dashed) and model predictions (solid).
parameter estimates and in panel (e) for what is observed in the data. In both cases the modes are positive, implying that the densities of time on the market are not exponential and suggesting that time on the market is measured with some error, as one would expect. Panels (b) to (f) in Figure 4 indicate that our model is a fairly good fit to the data.

**Backining Out the Distributions of “Static” Costs** We have estimated the steady-state distribution of dynamic costs $G_{SD}(c_D)$, and now we need to back out the distribution of entrants’ static costs $G(c)$. We know that the following relationships hold:

$$c_D = c + \delta V(c) := S(c) \quad (23)$$
$$g_S(c) \propto T(P(c))g(c) \quad (24)$$
$$g_{SD}(c_D) \propto T(P_D(c_D))g_D(c_D), \quad (25)$$

where $V(c)$ is given by (10), $T(P(c)) = \tau/(1 - F(1)(P(c)))$ is the function defined in (17) and where

$$P_D(c_D) := \Phi^{-1}(c_D/0.94).$$

The relation (25) is obviously equivalent to $g_D(c) \propto g_{SD}(c_D)/T(P_D(c_D))$. Therefore, $G_D(c_D)$ is given by the solution to the differential equation

$$G'_D(c) = G'_{SD}(c_D)/T(P_D(c_D)) \quad (26)$$

with the boundary condition $G_D(\underline{c}_D) = 0$.

Next, given $G_D(c_D)$, we want to compute $G(c)$, which is given as

$$G(c) = \frac{G_D(S(c))}{G_D(\overline{c}_D)},$$

with the boundaries of support given as $\overline{c} = \overline{c}_D$ and $\underline{c} = \delta V(\underline{c}) = \underline{c}_D$. Because $P_D(c_D) = \Phi^{-1}(c_D/0.94)$ and $P(c) = \tilde{\Phi}^{-1}(c/0.94)$, the equality $P_D(c_D) = P(c)$ implies

$$c_D = P_D^{-1}(P(c)) = 0.94\Phi(\tilde{\Phi}^{-1}(c/0.94)) =: S(c). \quad (27)$$

Thus, $\underline{c}_D = 0.94\Phi(\tilde{\Phi}^{-1}(\underline{c}/0.94))$. Consequently, we can recover $G$ from $G_D$ using (26), (27), $\overline{c} = \overline{c}_D$ and $\underline{c}_D = 0.94\Phi(\tilde{\Phi}^{-1}(\underline{c}/0.94))$.

Figure 5 illustrates the densities $g(c)$, $g_D(c_D)$, and $g_{SD}(c_D)$ for 1993 (assuming 7 matching per week). That there are sellers with negative costs should not be much of a concern. Their mass is 0.035% and
4 Counterfactual Analysis

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 one such analysis by investigating how much, if any, surplus is lost by the use of 6 percent fees instead of the fee (or Bayesian mechanism) that would be optimal given the parameter estimates.

4.1 The Counterfactual Experiment

The parameters estimated in Section 3 imply an optimal fee, which is stationary and denoted $\omega_{\alpha}^{opt}$. As shown in Section 2, the focus on fees is without loss of generality. In slight abuse of notation, denote by $W(\alpha, \omega_{\alpha}^{opt})$ the value of the objective function when the optimal fee is used for a given $\alpha$. In similar abuse of notation, let $W(\alpha, b)$ denote the value of the objective when a percentage fee $\omega(p) = bp$ is used. We will determine $\alpha$ by stipulating that the empirically observed 6 percent fees are chosen to maximize $W(\alpha, b)$. That is, $\alpha$ will be determined by the equation

$$\frac{\partial W(\alpha, b)}{\partial b}|_{b=0.06} = 0.$$ 

is only an artifact of the parametrization of $G$. We are currently working on a different parametrization. Observe that even the mass of sellers with costs below 0.25 is only 0.14%.
\(\alpha\) can be obtained by using
\[W(\alpha, b) = \alpha W_I(b) + (1 - \alpha)(W_I(b) + W_S(b))\]
and solving
\[\alpha W'_I(b) + (1 - \alpha)(W'_I(b) + W'_S(b)) = 0,\]
for \(\alpha\). This yields
\[
\alpha(b) = \frac{W'_I(b) + W'_S(b)}{W'_S(b)} = \frac{b}{1-b} \int_{\Omega} (1-b)^{\mathbb{1}} \frac{\partial k}{\partial b}(k(c))dG(c)
\]
\[\int_{\Omega} (1-b)^{\mathbb{1}} k(c)(1 - F(k(c)))dG(c),\]
where the second equality can be obtained by substituting in the definitions of \(W_I\) and \(W_S\). Letting \(\alpha^* = \alpha(0.06)\), our counterfactual experiment then consists of comparing \(W(\alpha^*, 0.06)\) with \(W(\alpha^*, \omega^{opt}_{\alpha^*})\).

### 4.2 (Almost) Optimal Percentage Fees

![Figure 6](image-url)

Figure 6: The objective \(W(\alpha^*, b) = \alpha W_I(b) + (1 - \alpha)(W_I(b) + W_S(b))\) as a function of the percentage fee \(b\) (solid line), the intermediary’s profits \(W_I(b)\) (dashed line), and the seller’s profit \(W_S(b)\) (dash-dotted line). \(W(\alpha^*, b)\) is maximized at \(b = 0.06\), \(W_I(b)\) is maximized at \(b = 0.219\). Calculations based on parameter estimates for the year 1993, 7 matchings per week.

The results do not vary substantively with the year of observation or with respect to the frequency of matchings we impose. The following counterfactual analysis is based on the first column in Table 3, that is, the year 1993 and seven matchings per week. For these parameters, we obtain \(\alpha^* = 0.112\) and
\[
\frac{W(\alpha^*, 0.06)}{W(\alpha^*, \omega^{opt}_{\alpha^*})} = 0.996.
\]
That is, 6 percent fees generate more than 99 percent of the maximum surplus that the intermediary and the seller can generate. Figure 6 illustrates the objective function at $\alpha^* = 0.112$ for percentage fees as a function of the percentage $b$. The figure also shows the broker’s (dashed line) and the seller’s (dash-dotted line) profit. The broker’s profit is maximized at $b = 0.219$, that is, a monopolistic broker restricted to charging percentage fees would set a fee of 21.9% in order to maximize $W(b) = W(1,b)$. A monopolistic broker that could charge an arbitrary non-linear fee could attain profits $W(1,\omega_{opt}^1)$. The ratio is

$$\frac{W(1,0.219)}{W(1,\omega_{opt}^1)} = 0.999.$$ 

Figure 7 illustrates that the virtual cost function $\Gamma_{SD,\alpha^*}(c_D)$ is almost linear, except for very high costs. But sellers with very high costs trade with a probability close to zero and hence have little effect on the maximization problem.

## 5 Conclusions

We provide a theoretical analysis of transaction fees from an optimal non-linear pricing perspective. We show that in increasingly thin markets, optimal fees are asymptotically linear. If percentage fees are optimal, then the distribution of reserve prices does not vary with the percentage being charged. We show that the optimal (potentially non-linear) fees are pinned down by the underlying distributions of the seller’s cost and the
buyers’ valuations. We also show that these distributions are non-parametrically identifiable. Using the data set of Genesove and Mayer (2001) on real-estate transactions in the Boston condominium market in the 1990s, we find that the empirically observed 6 percent fees generate 99 percent or more of the maximum achievable surplus of the intermediary and the seller.

The following are issues we have not addressed in the present version but are currently working on. In the empirical analysis, we are first estimating the endogenous dynamic cost distribution of the seller and then backing out the exogenous static cost distribution needed for our counterfactual analysis. We are working on directly estimating the exogenous distribution. In the empirical model, it would also be valuable to replace the parametric assumption of Beta distributions to estimate demand and supply by a specification in which virtual valuation and cost functions are Chebyshev polynomials. The advantage of using polynomials is that the functional form chosen for the distributions is less restrictive: by increasing the degree of the polynomial one can arbitrarily well approximate any analytical function. Parameterizing the virtual cost function as a polynomial has the advantage that one has a clearer notion of how close the estimated distribution $G$ is to a mirrored Generalized Pareto distribution (which has a linear virtual cost function).

Our analysis focuses on commission fees in thin markets. In order to improve economists’ understanding of the determinants of such fees, it is important to expand the analysis increasingly less thin markets. We hope that our paper provides a useful first step in this direction.
Appendix

A Proofs

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

\[ \tilde{\Phi}_\omega(p) := R_\omega(p) - R'_\omega(p) \frac{1 - F_\infty(p)}{f_\infty(p)}. \] (28)

We will show that the expression for \( \tilde{\Phi}_\omega \) in (28) is the same as the one in the proposition.

First, observe that

\[ R_\omega(p) = \frac{(p - \omega(p))(F_\infty(2)(p) - F_\infty(1)(p)) + \int_p^\pi (v - \omega(v))dF_\infty(2)(v)}{1 - F_\infty(1)(p)} \]

can be rewritten as

\[ R_\omega(p) = \frac{\int_p^\pi \Phi_\omega(v)dF_\infty(1)(v)}{1 - F_\infty(1)(p)} \]

where

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

That the two expressions for \( R_\omega \) are equal can be checked by observing that \( R_\omega(\pi) = \pi \) for both expressions and that the derivatives \([R_\omega(p)(1 - F_\infty(1)(p))]'\) can be shown to be equal for both expression for \( R_\omega \) with some algebra and by using (1).

One can also show with some algebra that

\[ R'_\omega(p) = \frac{f_\infty(1)(p)}{1 - F_\infty(1)(p)}(R_\omega(p) - \Phi_\omega(p)) \] (29)

and that

\[ \frac{f_\infty(1)(p)}{1 - F_\infty(1)(p)} \frac{1 - F_\infty(p)}{f_\infty(p)} = \frac{1 - \delta F_\infty(1)(p)}{1 - \delta} \] (30)

Plugging (29) and (30) into (28) yields

\[ \tilde{\Phi}_\omega(p) = R_\omega(p) - (R_\omega(p) - \Phi_\omega(p)) \frac{1 - \delta F_\infty(1)(p)}{1 - \delta} \]
the derivative of which can be rearranged to
\[
\tilde{\Phi}'(p) = \frac{1 - \delta F(\delta)(p)}{1 - \delta} \Phi'(p).
\]

Since this expression for \(\tilde{\Phi}'\) is equal to the expression for \(\Phi'\) in the proposition and since the two expressions for \(\tilde{\Phi}(p)\) are equal to \(v\) for \(p = \bar{v}\), \(\Phi\) is the same in the proposition as in (28).

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

Proof of Lemma 1. Suppose to the contrary that the optimal mechanism, denoted \(\langle \hat{Q}, \hat{M} \rangle\), does not auction off the good at some reserve in period \(t\) and in some states \((v_t, c)\). 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 \(\langle \hat{Q}, \hat{M} \rangle\) except for the states in period \(t\) in which \(\langle \hat{Q}, \hat{M} \rangle\) does not auction off the good. Let the alternative mechanism sell the good to the highest value buyer in all those instances for which \(\langle \hat{Q}, \hat{M} \rangle\) sells it to some other buyer. This alternative mechanism will increase the broker’s payoff \(W_I(c)\) by increasing the revenue it raises while leaving the seller’s payoff \(W_S(c)\) unaffected. Therefore, the mechanism \(\langle \hat{Q}, \hat{M} \rangle\) cannot be optimal.

Proof of Lemma 2. We first derive (7). For \(k \in [v, \bar{v}]\) let
\[
1 - F_T(k) := \max_{(k_t)_{t=0}^T} \left\{ \sum_{t=0}^T q_t(k) \right\} \quad \text{s.t.} \quad \frac{\sum_{t=0}^T q_t(k) k_t}{\sum_{t=0}^T q_t(k)} = k,
\]
and define
\[
1 - F(k) := \lim_{T \to \infty} 1 - F_T(k).
\]
Let \((k^*_t(k))_{t=0}^T\) be a maximizer of \(\max_{(k_t)_{t=0}^T} \left\{ \sum_{t=0}^T q_t(k) \right\} \quad \text{s.t.} \quad \frac{\sum_{t=0}^T q_t(k) k_t}{\sum_{t=0}^T q_t(k)} = k\). Under
stationarity, we have \( k_t^*(k) = k \) for all \( t \). Therefore,

\[
1 - \bar{F}(k) = \lim_{T \to \infty} \sum_{t=0}^{T} \left( \prod_{\tau=0}^{t-1} \delta F(1)(R^{-1}(k)) \right) \left( 1 - F(1)(R^{-1}(k)) \right)
\]

\[
= \lim_{T \to \infty} \frac{1 - \delta^{T+1} F(1)(R^{-1}(k))^{T+1}}{1 - \delta F(1)(R^{-1}(k))} \left( 1 - F(1)(R^{-1}(k)) \right)
\]

\[
= \frac{1 - F(1)(R^{-1}(k))}{1 - \delta F(1)(R^{-1}(k))} = 1 - F(\infty)(R^{-1}(k)),
\]

which is (7).

In the main body we have shown that the optimal allocation rule is such that trade takes place as soon as

\[
\Phi(k) \geq \Gamma_{\alpha}(c).
\]

(31)

Let \( k(c) := \Phi^{-1}(\Gamma_{\alpha}(c)) \). Since \( \Phi(k) = F(\infty)(R^{-1}(k)) \), \( \Phi(k) = \Phi(R^{-1}(k)) \). According to (31) trade should take place as soon as \( v \geq R^{-1}(k(c)) \), which because of the afore-noted equalities and the monotonicity of \( \Phi \), is equivalent to \( \Phi(v) \geq \Gamma_{\alpha}(c) \) as claimed in the proposition.

Proof of Proposition 2. The expected profit of a seller with cost who faces a fee \( \bar{w} \) is

\[
(1 - \bar{F}(k))(k - \bar{w}(k) - c).
\]

Substituting \( \bar{w}(k) \) by the expression in the Proposition 2, the maximization problem becomes

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

The first-order condition is

\[
0 = -\bar{F}(k(c)) \left[ \Gamma_{\alpha}^{-1}(\bar{F}(k)) - c \right],
\]

which is equivalent to \( \Phi(k) = \Gamma_{\alpha}(c) \), which is equivalent to the allocation rule in Lemma 2 (see its proof for details). The second-order condition is satisfied whenever the first-order condition is satisfied if \( \Phi(v) \) is monotone.

Proof of Proposition 3. Facing a fee \( \omega(p) \) in the current period and a sequence of optimal fees thereafter, the seller of type \( c \) faces the problem in (9). The first-order condition for
this is
\[ 0 = f_{(1)}(p) \left[ (p - \omega(p)) \frac{1 - F(p)}{f(p)} - (p - \omega(p)) + c + \delta V(c) \right] \]
because, as noted in the proof of Proposition 1, \( F_{(2)}(p) - F_{(1)}(p) = f_{(1)}(p)(1 - F(p))/f(p) \). Substituting the expression for \( \omega(p) \) and simplifying reveals that the first-order condition is equivalent to
\[ 0 = f_{(1)}(p) \left[ -\left( \Gamma^{-1}_\alpha(\Phi(p)) + \delta V(\Gamma^{-1}_\alpha(\Phi(p))) \right) + c + \delta V(c) \right]. \]
Because \( c + \delta V(c) \) is an increasing function, the first-order condition is satisfied if and only if
\[ \Gamma^{-1}_\alpha(\Phi(p)) = c, \]
which is equivalent to implementing the allocation rule in Lemma 2. Lastly, observe that the second-order condition is satisfied strictly if the first-order condition is satisfied because \( \Gamma^{-1}_\alpha(\Phi(p)) \) increases in \( p \).

Proof of Proposition 4. Proof of part (i): \( \Phi \) continuously differentiable implies
\[ \lim_{v \to \overline{v}} \frac{d}{dv} \left[ \frac{1 - F(v)}{f(v)} \right] = \lim_{v \to \overline{v}} \frac{d}{dv} [v - \Phi(v)] = \overline{\beta}, \tag{32} \]
for some constant \( \overline{\beta} \). Equation (32) is the von Mises condition as stated in Theorem 2 in Appendix B. It implies that \( F \) is in the domain of attraction of an extreme value distribution (as defined in Definition 1) by Theorem 2. 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 (43). 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 \( \tilde{v} \). An analogous reasoning holds for the convergence of \( \tilde{G}_j \).

Proof of part (ii): First, define
\[ \bar{\beta} := \lim_{v \to \overline{v}} 1 - \Phi'(v) \quad \text{and} \quad \bar{\sigma} := \lim_{c \to \overline{c}} \Gamma'(c) - 1, \]
\[ \beta := -1/\bar{\beta}, \quad \text{and} \quad \sigma := 1/\bar{\sigma}. \]
Observe that by l’Hôpital’s rule
\[ \lim_{v \to \overline{v}} \frac{(\overline{v} - v) f(v)}{1 - F(v)} = \lim_{v \to \overline{v}} \frac{\overline{v} - v}{v - \Phi(v)} = \lim_{v \to \overline{v}} \frac{-1}{1 - \Phi'(v)} = \beta. \tag{33} \]
The proof proceeds in four steps: we show that (a) for the limiting distributions $\tilde{F}^*$ and $\tilde{G}^*$ the expectational fee $\tilde{\omega}$ is equal to the limiting fee $\tilde{\omega}^*$, (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, (d) the transaction fee converges to $\tilde{\omega}^*$.

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 $[\overline{v}, \overline{v}]$ and $[\underline{c}, \underline{c}]$. 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: $\Gamma^*_{\alpha}(c) = c + (c - \underline{c})\alpha/\sigma$.

Fees given in (??) can be rearranged the following way.

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

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

$$= p - \Gamma^*_{\alpha}^{-1}(p),$$

where the second equality comes from the linearity of $\Gamma^*_{\alpha}$ and the third from observing that for any $p$ and any distribution $\overline{F}$,

$$E_v[F^*(v)|v \geq p] = \frac{\int_p^\overline{v} \overline{F}(v)f(v)dv}{1 - F(p)} = p,$$

which can be established, for example, using integration by parts. Plugging in the functional form for $\Gamma^*_{\alpha}$ yields

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

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

Step (b): Next, we want to 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 $[\underline{c}_j, (\overline{v}_j - \underline{c}_j)/u_j^S + \underline{c}_j]$ is transformed to $[0, 1]$ and the support of the buyer’s distribution becomes $[\overline{v}_j - (\overline{v}_j - \underline{c}_j)/u_j^B, \overline{v}_j]$ to $[u_j - 1, u_j]$ with some $u_j > 0$. Note that as $k \to \infty$, $u_j \to 0$. In part of the following analysis, we will drop the subscript $k$ 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

\[
\hat{\omega}(\hat{p}) = u\hat{p} + \frac{\int_1^\infty \hat{\Gamma}_\alpha^{-1}(\hat{\Phi}(u\hat{v}))d\hat{\Phi}(u\hat{v})}{1 - \hat{\Phi}(u\hat{p})} \tag{35}
\]

where the expression comes from plugging in \( u\hat{p} \) for \( p \) in (??).

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 \( \hat{\omega} \).

By the definition of \( \beta \) we have

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

This implies that

\[
\frac{1}{u} \left[ \frac{1 - \hat{F}(u\hat{v})}{f(u\hat{v})} \right] \overset{u \to 0}{\Rightarrow} \frac{\hat{\Phi}(u\hat{v})}{\hat{\Gamma}_\alpha(1 + \alpha/\sigma)}
\]

and hence

\[
\frac{1}{u} \hat{\Phi}(u\hat{v}) \overset{u \to 0}{\Rightarrow} \hat{\Phi}(1 + \alpha/\sigma)
\]

where the double arrow \( \Rightarrow \) stands for uniform convergence. By a similar logic

\[
\frac{1}{u} \hat{\Gamma}_\alpha^{-1}(u\hat{c}) \overset{u \to 0}{\Rightarrow} \hat{\Gamma}_\alpha^{-1}(1 + \alpha/\sigma)
\]

and hence

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

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

Observe that

\[
\hat{\Phi}(k) = \hat{\Phi}_\infty(\hat{R}^{-1}(k)), \quad \hat{\Phi}_\infty(p) = \frac{1 - \hat{\Phi}(p)}{1 - \delta \hat{\Phi}(p)}, \quad \hat{\Phi}(p) = \sum_{B=0}^\infty \hat{\Phi}(p)^B. \tag{36}
\]

and

\[
\hat{R}(p) = \frac{\int_0^1 \hat{\Phi}(v)d\hat{\Phi}(v)}{1 - \hat{\Phi}(p)}. \tag{37}
\]
By Theorem 1, the expressions in (35) uniformly converge to their respective limits if $\hat{R}^{-1}$ uniformly converges. $\hat{R}^{-1}$ converges uniformly if $\hat{R}$ converges uniformly.

So we are left to show that $\hat{R}$ converges uniformly, in order to show uniform convergence of the integrand in (35) and hence convergence of $\hat{\omega}$.

Since the integrand in the integral in $\hat{R}$ converges uniformly, $\hat{R}$ converges pointwise to its limit. Further, observe that the sequence $\hat{R}_j$ monotonically increases in $k$: a higher truncation point implies a higher expected revenue $\hat{R}_j$. Pointwise convergence and monotonicity of the sequence imply uniform convergence of $\hat{R}_j$ by Dini’s theorem.

Putting this together implies that $\hat{\omega}$ converges to $\omega^\ast$.

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 $\omega(p) = \bar{\omega}(p)$ is also linear in the limit.

Step (d): Next, we turn to convergence of the normalized transaction fee $\hat{\omega}$. For the sake of notational simplicity we will drop the hat and write $\omega$. We obtained previously that the transaction fee is

$$\omega(p) = p - \frac{\int_{\bar{\Phi}^{-1}(c)}^{\bar{\Phi}^{-1}(\hat{\Phi}(v)))dF(v)}{1 - F(p)}$$

where $S(c) = c + \delta V(c)$ and $V(c)$ is given by (10).

Observe that plugging in $\bar{\omega}$ leads to a simplified version of $V(c)$:

$$V(c) = \int_{k(c)}^{\bar{\Phi}} \Gamma^{-1}_\alpha(\Phi(v))d\bar{F}(v)$$

which simplifies further to

$$V(c) = \int_{\Phi^{-1}(\Gamma\alpha(c))}^{\bar{\Phi}} \Gamma^{-1}_\alpha(\Phi(v))d\bar{F}(v)$$

Plugging the expressions for $S$ and $V$ into the expression for $\omega$ yields

$$\omega(p) = p - \frac{\int_{p}^{\Gamma^{-1}(\Phi(v)))dF(v) + \delta \int_{p}^{\bar{\Phi}} \Gamma^{-1}_\alpha(\Phi(v)))\Gamma^{-1}_\alpha(\Phi(y))d\bar{F}(y)dF(v)}{1 - F(p)}$$

$$= p - \frac{B + \delta A}{1 - F(p)}$$
Using \( \tilde{\Phi}(p) = \Phi(R(p)) \) it is clear that the integrand in \( B \) converges uniformly and hence \( B \) converges. Using this expression for \( \tilde{\Phi} \) yields for \( A \)

\[
A = \int_p^\infty \int_{R(v)}^\infty \Gamma_\alpha^{-1}(\tilde{\Phi}(R^{-1}(y)))dF(y)dF(v)
\]

\[
= \int_p^\infty \left[ \int_p^{R^{-1}(y)} dF(v) \right] \Gamma_\alpha^{-1}(\tilde{\Phi}(R^{-1}(y)))dF(v)
\]

\[
= \int_p^\infty \left[ F(R^{-1}(y)) - F(p) \right] \Gamma_\alpha^{-1}(\tilde{\Phi}(R^{-1}(y)))dF(v),
\]

where the second equality follows from reversing the order of integration. 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 5.** The Taylor expansion of \( \Gamma_\alpha^{-1}(x) \) around \( v \) is

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

Denote the \( n \)th derivative at \( v \) as \( \gamma_n := [\Gamma_\alpha^{-1}(v)]^{(n)} \). We further use the shorthand \( \varphi := \Phi(v) \). The net price received by the seller can be rearranged as

\[
p - \omega_\alpha(p) = E[\Gamma_\alpha^{-1}(\varphi)|v \geq p]
\]

\[
= \sum_{n=0}^{\infty} \frac{\gamma_n}{n!} E[(\varphi - v)^n|v \geq p] \tag{38}
\]

\[
= \sum_{n=0}^{\infty} \frac{\gamma_n}{n!} \left\{ E[(\varphi|v \geq p) - v]^n - E[\varphi|v \geq p] - v]^n + E[(\varphi - v)^n|v \geq p] \right\} \tag{39}
\]

\[
= \Gamma_\alpha^{-1}(E[\varphi|v \geq p]) + \sum_{n=0}^{\infty} \frac{\gamma_n}{n!} \left\{ E[(\varphi - v)^n|v \geq p] - (E[\varphi|v \geq p] - v)^n \right\} \tag{40}
\]

\[
= \Gamma_\alpha^{-1}(p) + \sum_{n=0}^{\infty} \frac{\gamma_n}{n!} \left\{ E[(\varphi - v)^n|v \geq p] - (E[\varphi|v \geq p] - v)^n \right\}, \tag{41}
\]

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 6.** If the distribution of reserve prices is the same, the distribution and thus the mean of the transaction prices are the same. Thus, the second part of the statement follows once the first is shown.

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

\[
G(\Gamma_{\alpha}^{-1}(x)) = G(x).
\]

To see that this is true, observe first that \( \Gamma_{\alpha}^{-1}(x) = \frac{\sigma}{\sigma + \alpha}x + \frac{\alpha}{\sigma + \alpha}\overline{v} \). Therefore,

\[
G(\Gamma_{\alpha}^{-1}(x)) = \left( \frac{\sigma}{\sigma + \alpha} \right)^{\sigma} \left( \frac{x - \overline{v}}{\overline{v} - \overline{v}} \right)^{\sigma}.
\]

Consequently, \( G(\Gamma_{\alpha}^{-1}(\overline{v})) = \left( \frac{\sigma}{\sigma + \alpha} \right)^{\sigma} \), whence \((43)\) follows.

**B Extreme Value Theory**

**Extreme Value Theory** For our purposes, the main result of the theory of exceedences in extreme value theory is summarized in Theorem \( \text{I} \). The theorem says that for any \( F \) that satisfies some weak regularity condition,

\[
\lim_{u \to 0} 1 - \frac{1 - F(\overline{v} - u(\overline{v} - v))}{1 - F(\overline{v} - u(\overline{v} - v))} = 1 - \left( \frac{\overline{v} - v}{\overline{v} - \overline{v}} \right)^{\beta} =: F^{*}(v),
\]

where convergence is uniform and \( \beta \) is some constant. The left-hand side of \((44)\) is the rescaled distribution conditional on being above the threshold \( \overline{v} - u(\overline{v} - \overline{v}) \). According to Theorem \( \text{II} \) this truncated and rescaled distribution converges to a Generalized Pareto distribution \( F^{*} \) as the threshold \( \overline{v} - u(\overline{v} - \overline{v}) \) goes to the finite upper bound \( \overline{v} \).
The motivation for this theory was an empirical regularity found in many situations: that the upper tail of a distribution is well approximated by a (Generalized) Pareto distribution. A prominent example is the distribution of the highest 20 percent of income and wealth in many countries, which was first observed by Vilfredo Pareto.

This appendix gives a brief review of extreme value theory. The theory of exceedences within extreme value theory deals with the distribution of a random variable conditional on being above a high threshold (for the original articles see Balkema and De Haan (1974), Pickands (1975); for a textbook see Falk, Hüsler, and Reiss (2010)). The theory relates to the fact that one often observes that the upper tail of a distribution is well approximated by a Generalized Pareto distribution.

The general principle behind this observation is described by the Pickands-Balkema-de Haan theorem (also called the second theorem of extreme value theory). For expository simplicity, we provide a simplified version of the theorem, since it is sufficient for our purposes. See Pickands (1975, Theorem 7) and Balkema and De Haan (1974) for the theorem itself. The theorem establishes a connection between the behavior of the maximum of a distribution and its upper tail. The relevant concept for the maximum is the domain of attraction:

**Definition 1.** A distribution $F$ is in the domain of attraction of an extreme value distribution if there exists a sequence of constants $a_n > 0$ and $b_n$ real for $n = 1, 2, \ldots$, such that

$$
\lim_{n \to \infty} F_n(a_n x + b_n) = F_{\text{max}}(x)
$$

for every continuity point $x$ of $F_{\text{max}}$ for some non-degenerate distribution function $F_{\text{max}}$ (see De Haan and Ferreira, 2006, p. 4).

This means that for $n$ independently and identically distributed random variables, $(\max\{X_1, X_2, \ldots, X_n\} - b_n)/a_n$ has a non-degenerate distribution as $n$ goes to infinity.

44 Other examples include the distribution of the strength of earthquakes in historical data (which tend to contain only the most severe earthquakes); and for the discrete type variant of the Pareto distribution – Zipf's law – the distribution of the frequency of the most common words in a larger text and the sizes of the largest cities in most countries.
The following theorem holds.

**Theorem 1.** *(Simplified version of the Pickands-Balkema-de Haan Theorem)* Assume $F$ has a finite upper bound and $f(v) > 0$ for all $v \in (v, \overline{v})$. Then $F$ has a Generalized Pareto upper tail, formally

\[
\lim_{u \to 0} \frac{1 - F(\overline{v} - u(\overline{v} - v))}{1 - F(\overline{v} - u(\overline{v} - v))} = 1 - \left(\frac{\overline{v} - v}{\overline{v} - v}\right)^\beta, \tag{45}
\]

for some constant $\beta$, where convergence is uniform, if and only if $F$ is in the domain of attraction of an extreme value distribution.

The left-hand side of (45) is the rescaled distribution conditional on being above the threshold $\overline{v} - u(\overline{v} - v)$. The right-hand side is the cumulative distribution function of a finite upper bound Generalized Pareto distribution.

**Proof of Theorem 1.** See Theorem 7 in Pickands (1975). Note that for our setup ($\overline{v}$ finite and $f(v) > 0$ for all $v \in (v, \overline{v})$) the definition of $F$ having a Generalized Pareto upper tail given in Definition 4 in Pickands (1975) simplifies to (45).

The literature on extreme value theory states several sufficient conditions for a distribution to be in the domain of attraction of an extreme value distribution. We state the one most suitable for our purposes.

**Theorem 2.** Assume $F$ has a finite upper bound. $F$ is in the domain of attraction of an extreme value distribution if the von Mises condition

\[
\lim_{v \to \overline{v}} \frac{d}{dv} \left[ \frac{1 - F(v)}{f(v)} \right] = \overline{\beta}, \tag{46}
\]

for some constant $\overline{\beta}$, holds.

**Proof.** See, for example, Theorem 1.1.8 in De Haan and Ferreira (2006, p. 15).

As stated in the literature, even this sufficient condition is weak and is satisfied by all “textbook” continuous distributions, such as uniform, Beta, bounded Generalized Pareto, inverse Weibull and (for the infinite upper bound counterpart of the condition) the normal, exponential, Cauchy, and infinite upper bound Generalized Pareto distribution.
Often, the Generalized Pareto distribution is defined with the parametrization

\[ F^*(v) = 1 - \left(1 + \frac{\xi(v - \mu)}{\sigma}\right)^{-1/\xi}. \]

For \( \xi < 0 \) the distribution has a finite upper bound and corresponds to the parametrization used in this paper with \( \underline{v} = \mu, \overline{v} = \mu - \sigma/\xi \), and \( \beta = -1/\xi \). For \( \xi \geq 0 \), it has an infinite upper bound and lower bound \( \mu \). One obtains the exponential distribution as a special case as \( \lim_{\xi \to 0} F^*(v) = 1 - e^{-(v-\mu)/\sigma} \). For \( \xi > 0 \) and \( \sigma = \mu \xi \) one obtains the classical Type I Pareto distribution \( F(v) = 1 - (v/\mu)^{1/\xi} \). For \( \xi > 0 \) one obtains the Type II Pareto distribution.

For infinite upper bounds, convergence can be stated as

\[ \left(1 - \frac{1 - F(u + x)}{1 - F(u)}\right) - F_u^*(x) \xrightarrow{u \to \infty} 0, \]

for some Generalized Pareto distribution \( F_u^* \). See the abovementioned references for more details.

Note that the characteristic property of Generalized Pareto distributions is that the inverse hazard rate is linear: \( [(1 - F(v))/f(v)]' = \xi \). The special cases can be seen as the inverse hazard rate decreasing (bounded Generalized Pareto distribution), constant (exponential distribution), and increasing ((Non-Generalized) Pareto distribution). \( \xi < 0 \) corresponds to the common monotone hazard rate condition (that is, \( f(v)/(1 - F(v)) \) is increasing). \( \xi < 1 \) corresponds to Myerson’s regularity condition \( \Phi'(v) > 0 \) and is also necessary to ensure that the distribution has a finite mean.

References


---


REFERENCES


REFERENCES


