An optimal pricing theory of transaction fees*

Simon Loertscher† Andras Niedermayer‡

July 22, 2026

Abstract

We derive optimal transaction fees as the solution to an intermediary's pricing problem. Assuming independent private values, we show that no incentive compatible, individually rational mechanism performs better. Decreasing the elasticity of supply or increasing the weight on the intermediary's profits increases the optimal fees, as can increasing the elasticity of demand. Extreme value theory implies asymptotic optimality of linear fees in thin markets arising from vanishing overlap between values and costs. The results generalize to a dynamic setting with short-lived buyers. Under optimality, observed linear fees suffice to identify the seller's type distribution up to truncation.

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

JEL-Classification: C72, C78, L13

*We are grateful to three anonymous referees and the editor, Mark Armstrong, for comments and feedback that have helped us improve the paper substantively. We also thank Arnaud Costinot, Julio González-Díaz, David Genesove, Philipp Kircher, Stephan Lauermann, Volker Nocke, François Ortalo-Magné, Marco Ottaviani, Martin Peitz, Bernard Salanié, Juan Santaló, Mark Satterthwaite, Philipp Schmidt-Dengler, Art Shneyerov, Dan Spulber, Adam Szeidl, Ernst-Ludwig von Thadden, Rakesh Vohra, Cédric Wasser, Asher Wolinsky, Abdullah Yavas, participants at the IIOC 2012 in Washington D.C., the SFB TR15 2012 meeting in Mannheim, the Swiss IO Day 2012 in Bern, EARIE 2012 in Rome, and seminar audiences at Bates White, Bristol, Columbia, Deakin, Georgetown, Hamburg, LSE, Mannheim, Melbourne, Michigan, Munich, UNSW, Northwestern, Sydney, WHU, Zurich, and the University of Pennsylvania for their comments and suggestions. We want to thank Mark Satterthwaite for stating the conjecture of convergence to linear fees as the opportunity costs of trading increase and for having encouraged us to prove this result. Financial support through a research grant by the Faculty of Business and Economics at the University of Melbourne is also gratefully acknowledged. Financial support by the German Science Foundation through SFB TR-15 and project PE 813/2-2 and the ARC Discovery Project DP250100407 is also gratefully acknowledged.

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

‡CY Cergy Paris Université. Email: andras.niedermayer@cyu.fr.

1 Introduction

Internet platforms, brokers, and governments around the globe charge commission fees and taxes that are paid when a transaction occurs. Examples include the indirect taxes imposed by governments, the percentage fees charged by real-estate brokers, headhunters, and stock brokers, and the commission fees charged by auction houses and trading platforms such as Sotheby's, Christie's, eBay, Apple Store, and Amazon. Transaction fees also figure prominently in public policy debates on issues as varied as credit-card fees, allegations of collusive commission fee-setting by auction houses and real-estate agents, the fees Apple and Google charge to app developers, the – at times, drastic – increases of value-added taxes in the wake of the global financial crisis, or proposals for financial-transaction taxes, such as the European Commission's 2013 proposal for enhanced cooperation among participating EU Member States.[1] Little is known to date about the effects and determinants of the structure of such fees and as to when their use dominates alternative trading mechanisms.

In this paper, we analyze the fee structure from the perspective of an intermediary's optimal-pricing problem, using an independent private values model. Our focus is on markets in which every seller owns a unique object and faces only a small number of buyers in every period. With private information on both the seller's and the buyers' sides, we first show that there is no loss of generality by focusing on fee setting — no incentive compatible, individually rational mechanism fares better.

However, in general, the optimal fee can be a complicated function of the transaction price. This contrasts with the linear and proportional fees that are widely used in practice. With that in mind, we then show that the fee that implements the optimal mechanism is linear if the seller's cost is drawn from a reflected generalized Pareto distribution. This fee is proportional, or a percentage fee, if the lower bound of support

[1]The European Commission published its proposal to implement enhanced cooperation in the area of financial transaction tax on February 14, 2013 (see *The Economist* on February 23, 2013: "Bin it - Europe's financial-transactions tax"). 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.

of this distribution is zero. The optimal fee is independent of the demand side if and only if it is linear. Intuitively, when a linear fee is optimal, the seller's profit-maximization problem is perfectly aligned with the intermediary's. Thus, the seller internalizes any demand-side properties in the same way the intermediary would if he set the price. Importantly, this means that if linear fees are optimal, then the optimal fee will not vary across product categories that differ on the demand side, provided the supply side does not vary. Consequently, observing that an intermediary uses the same fee across different product categories for, say, third-party sellers does not imply that the fee is suboptimal in any given product category.

Moreover, using extreme value theory and a model with transaction costs that shrink the overlap of the supports of the buyer's and the seller's distribution in such a way that only the types who are keenest to trade participate, we then show that in the limit, as the overlap goes to zero, the optimal fees converge to linear ones. This convergence arises because, as markets become increasingly thin, the distributions of participants' valuations converge to generalized Pareto distributions, which in turn implies convergence to linear fees.

Interestingly, and counterintuitively, more elastic demand can lead to higher optimal fees. The reason for the counterintuitive demand-elasticity effect is that the seller's price responds endogenously to changes in demand. When this reaction is excessive, the optimal fee needs to increase to partly offset this price endogeneity effect. For the limiting linear fees, the seller's and the intermediary's pricing incentives are perfectly aligned, so that the optimal fee is independent of the elasticity of demand.

Motivated by the relevance of dynamics in many real-world settings of interest, we then extend the model to an infinite-horizon setting in which the mechanism employed by the intermediary maximizes a weighted sum of the intermediary's and the seller's profit, with short-lived buyers arriving randomly in every period. The weight on the seller's profit can be interpreted as a conduct parameter that captures — unmodeled – competition between intermediaries for sellers. This generalization exhibits an invariance result: the asymptotically optimal fee in thin markets is the same in a dynamic setting with an arbitrary number of buyers as in a static setting with one buyer.

Our paper contributes, first and foremost, to the growing literature that applies insights and methods from mechanism design to pertinent questions in industrial organization. Recent and complementary contributions, such as Board (2008), Gomes (2014), Tirole (2016) and Garrett (2016), have applied multi-period mechanism design to intertemporal pricing and to optimal incentive schemes for platform participation. Our paper, which builds on and supersedes the predecessor Loertscher and Niedermayer (2007), is the first paper to connect fee setting to optimal pricing with two-sided private information of the form first studied by Myerson and Satterthwaite (1983). The companion paper, Loertscher and Niedermayer (2026), takes our model to the data and shows that, empirically, percentage fees perform well relative to the optimal mechanism. Partly in response to antitrust cases and regulatory concerns, there has been a recent upsurge of interest in fee setting, which is called the "agency model" in the language used by the courts; see, for example, De los Santos et al. (2026) and the references therein.[2] Baye et al. (2011) derive the click-through fees for advertisements that are optimal for an online platform. However, to the best of our knowledge, none of the existing papers relate fee setting to optimal mechanisms in an independent private values setting.

This combination of a small number of traders and two-sided private information is also what sets our theory apart from the existing theoretical literature on the transaction fees of profit-maximizing intermediaries (Yavas (1992), Caillaud and Jullien (2003), Hagiu (2007), Matros and Zapechelnyuk (2008), Shy and Wang (2011), Niedermayer and Shneyerov (2014), Johnson (2017), and Wang and Wright (2017)) and on the indirect taxes charged by governments (Salanié (2003, Chapter 3), Delipalla and Keen (1992), and Anderson et al. (2001a,b)).[3] Without this combination, the theory would be silent

[2]Loertscher and Niedermayer (2020) analyze fee setting (or "agency") in markets with a continuum of buyers and sellers and show that agency can profitably be used by an intermediary to deter entry of a competing exchange.

[3]Yavas (1992), Caillaud and Jullien (2003), Shy and Wang (2011), Johnson (2017) and Wang and Wright (2017) assume that the seller's cost is public information (or, equivalently, that either there is no uncertainty about the seller's cost or that there is perfect competition between sellers). In Matros and Zapechelnyuk (2008) the seller's cost is sunk after he chooses to go to the intermediary, which can be best seen by considering the one-period version of their model. Therefore, the seller's private cost only matters for his participation decision, but not for anything that happens after he chooses to participate (in particular for the reserve and transaction price). In Niedermayer and Shneyerov (2014) there is a continuum of sellers and buyers, so that by the law of large numbers there is no uncertainty about the realized distribution of sellers' costs.

about the functional form of the fee, that is, as to whether it is fixed, a percentage fee or a non-linear fee. While in thin markets linear fees are needed for optimality, in thick markets they do no harm, being equivalent to alternative ways of raising revenues. Moreover, our model predicts equilibrium price dispersion, which is consistent with the data but absent in most of the aforementioned models.

Optimal mechanisms for profit-maximizing intermediaries in markets with small numbers of traders are studied, for example, by Myerson and Satterthwaite (1983), Baliga and Vohra (2003), Jullien and Mariotti (2006), Lu and Robert (2001) and Loertscher and Wasser (2019). To the best of our knowledge, none of the existing papers contains an analysis of transaction fees, their structure, or their convergence to linearity.[4] The relevance of the long tail, which motivates and underpins our modeling of thin markets, is well documented in the digital economy; see, for example, Brynjolfsson et al. (2003), Anderson (2006) and Waldfogel (2017).[5] Formalizing long tails using extreme value theory, our paper derives the implications of long tails for optimal fees in thin markets.

The remainder of this paper is organized as follows. Section 2 sets up the baseline model, which we analyze in Section 3. In Section 4, we then introduce and analyze the infinite-horizon model. Section 5 discusses extensions and empirical implications, while Section 6 concludes the paper. Proofs and additional background material are in the Appendix.

2 Model

Our baseline model is a static, one-shot setup with independent private values, one intermediary, one seller, and one buyer. The buyer and the seller can only trade via the

[4]Myerson and Satterthwaite (1983) are mostly cited for their impossibility result, but the paper actually also contains a section about a profit-maximizing intermediary in a static bilateral trade market. Baliga and Vohra (2003) focus on market research rather than fees in a static model. Jullien and Mariotti (2006) assume two-sided private information in a static model with one broker and two buyers, but focus on fees that are a function of the reserve price rather than the transaction price without specifying the functional form of these fees. Moreover, as mentioned, we have multiple buyers and multiple periods. In the companion paper, we also structurally estimate the model.

[5]Theoretically, Loertscher and Muir (2025) show that with long tail products, called niche products, the matching benefits of larger markets dwarf the incentive benefits that have been at the center of attention in the double-auctions literature.

intermediary. For example, the costs of finding a trading partner without the intermediary could be prohibitively high.[6] The value of the outside option of not participating is zero for both the seller and the buyer. The buyer, seller, and the intermediary are risk neutral. In the baseline model, the intermediary, the seller and the buyer maximize their profits.

2.1 Fee setting

Motivated by the prevalence of this trading scheme, we assume that the intermediary uses a *fee-setting mechanism*. That is, the intermediary has the seller set a transaction price p to be paid by the buyer and charges the transaction fee $\omega(p)$. The seller has to pay the fee $\omega(p)$ to the intermediary if and only if a transaction occurs at price p. The fee function ω is a given for the seller, but the seller is free to choose the price.

2.2 Payoffs, types and distributions

At first, we simply stipulate that the buyer's value v is drawn from a distribution F with support $[\underline{v}, \overline{v}]$ and density $f > 0$ and the seller's cost c is drawn independently from a distribution G with support $[\underline{c}, \overline{c}]$ and density $g > 0$. We assume $\underline{v} \leq \underline{c} < \overline{v} \leq \overline{c}$, i.e. there are gains from trade for some but not all realizations of valuations and costs. We further assume that these distributions satisfy what, following Myerson (1981), has become known as *regularity*; that is, the *virtual value* and *virtual cost* functions, denoted $\Phi(v)$ and $\Gamma(c)$ and defined as

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

are increasing in their arguments. A buyer with value v who purchases the good at price p nets a payoff of $v - p$ and the seller with cost c who sells at the net price $p - \omega(p)$ nets a payoff of $p - \omega(p) - c$.

For this setting, we derive in Section 3.1 the optimal fee $\omega(p)$ for an intermediary who maximizes his profit. In general, this is a complicated problem because of the two-sided

[6] Alternatively, the buyers' values and the seller's cost may incorporate an exogenous opportunity cost of searching for a trading partner without the intermediary.

private information held by the agents; the seller with cost c will set an optimal price that depends on the fee function $\omega(p)$. Hence, viewed from this angle, the intermediary's problem is to choose $\omega(.)$ anticipating the seller's optimal behavior. To solve this problem, we will use the mechanism design methodology developed by Myerson (1981) and, in particular, Myerson and Satterthwaite (1983) and by doing so, we will show something stronger: Not only will we derive the fee that is optimal for the intermediary among all fee functions, but in fact we will show that restricting attention to fee setting is without loss of generality in the sense that no other incentive compatible and individually rational mechanism generates a larger expected profit for the intermediary than does the one with the optimal fee.

2.3 Thin market setting

To account for what we call *thin markets*, we will introduce sequences of buyers' and seller's distributions with decreasing overlap in the supports and distinguish between the agents' *primitive* and *effective* valuations and costs, which are defined as follows. We denote the agents' primitive values and costs by v_0 and c_0 and their distributions by F_0 and G_0 whose densities are denoted $f_0(v_0)$ and $g_0(c_0)$, respectively. The supports for both distributions are $[0, 1]$ and the densities are strictly positive for all $v_0, c_0 \in (0, 1)$. We continue to maintain the assumption of regularity for these distributions, that is, the *virtual valuation* and *virtual cost* functions $\Phi_0(v) := v - \frac{1-F_0(v)}{f_0(v)}$ and $\Gamma_0(c) := c + \frac{G_0(c)}{g_0(c)}$, both with domain $[0, 1]$, are increasing and continuously differentiable.

We additionally impose the following endpoint conditions.[7] There exist finite constants $\beta, \sigma > 0$ such that

$$\lim_{v\uparrow 1} \Phi_0'(v) = 1 + \frac{1}{\beta} \quad \text{and} \quad \lim_{c\downarrow 0} \Gamma_0'(c) = 1 + \frac{1}{\sigma}.$$

In Section 3.2, we then analyze sequences of economies indexed by j in which the *effective* valuations v_j and costs c_j are linear transformations of the primitive valuations v_0 and costs c_0, such that the supports are $[\underline{v}_j, \overline{v}_j]$ and $[\underline{c}_j, \overline{c}_j]$, respectively. Formally,

[7] These are the finite-endpoint Weibull-domain von Mises conditions for the upper tail of F_0 and the lower tail of G_0, respectively.

$v_j := (\overline{v}_j - \underline{v}_j)v_0 + \underline{v}_j$ and $c_j := (\overline{c}_j - \underline{c}_j)c_0 + \underline{c}_j$. Denote the corresponding distributions and densities by F_j, G_j, f_j, and g_j. The additive terms in the linear transformations $\underline{v}_j$ and $\underline{c}_j$ can be interpreted as additive, type-independent transaction costs while multiplicative terms $\overline{v}_j - \underline{v}_j$ and $\overline{c}_j - \underline{c}_j$ can be thought of as type-dependent opportunity costs. Suppose, for this illustration, that the outside-market payoff is the relevant outside option for the types under consideration. As a case in point, the transaction costs can stem from an outside market that offers a payoff of $\theta_j v_0 - p_j$ to a buyer of type v_0, where $\theta_j \in [0,1]$ is the quality offered by the outside market and p_j the price it asks. Consequently, the buyer's willingness to pay for trade via the intermediary becomes $v_0 - (\theta_j v_0 - p_j) = (1-\theta_j)v_0 + p_j$, implying $\overline{v}_j - \underline{v}_j = (1-\theta_j)$ and $\underline{v}_j = p_j$. Likewise, if an outside market offers a seller with cost c_0 a profit of $p_j - \theta_j c_0$, then $\overline{c}_j - \underline{c}_j = 1 - \theta_j$ and $\underline{c}_j = p_j$.

The virtual type functions Φ_j and Γ_j are defined analogously to Φ_0 and Γ_0. It is straightforward to see that the regularity of F_0 and G_0 implies regularity of F_j and G_j for all j. The sequences of economies we study are such that the overlaps of support ratios $u_j^B := (\overline{v}_j - \underline{c}_j)/(\overline{v}_j - \underline{v}_j)$ and $u_j^S := (\overline{v}_j - \underline{c}_j)/(\overline{c}_j - \underline{c}_j)$ go to 0; see Figure 1.

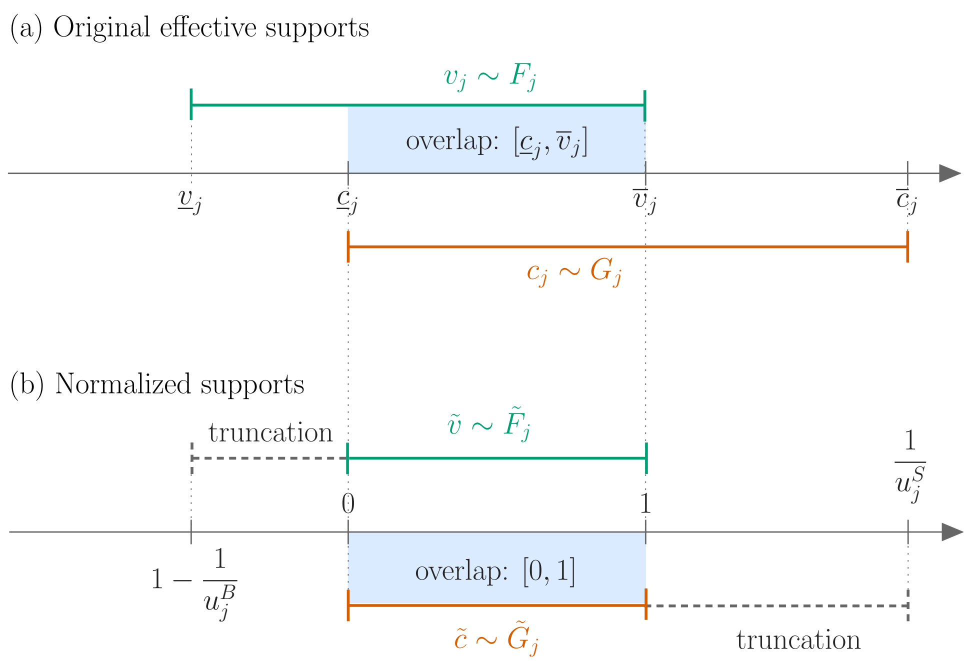


Figure 1: Original effective supports of v_j and c_j and normalized supports of $\tilde{v} = (v_j - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$ and $\tilde{c} = (c_j - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$. The overlap-of-support ratios are $u_j^B = (\overline{v}_j - \underline{c}_j)/(\overline{v}_j - \underline{v}_j)$ and $u_j^S = (\overline{v}_j - \underline{c}_j)/(\overline{c}_j - \underline{c}_j)$. The thin markets assumption is $u_j^B, u_j^S \to 0$.

As mentioned, this way of modeling market thinness is motivated by the fact that revenue from what is known as the "long tail" of the distributions appears to be quantitatively important in the digital economy; see, for example, Brynjolfsson et al. (2003) and Anderson (2006). In our setting, the long tail of the buyer distribution consists of buyers with high values, whereas on the seller side, it consists of sellers with low costs.

3 Optimal transaction fees

We begin by deriving the optimal mechanism for a profit-maximizing intermediary. By the revelation principle, we can focus, without loss of generality as far as optimality is concerned, on direct mechanisms that ask the buyer and the seller to report their types. As a function of these reported types $\hat{v}$ and $\hat{c}$, the mechanism chooses an allocation $Q(\hat{v}, \hat{c}) \in [0, 1]$, which in our setting corresponds to the probability that the buyer and the seller trade, and payments $M_B(\hat{v}, \hat{c})$ and $M_S(\hat{v}, \hat{c})$, where $M_B(\hat{v}, \hat{c})$ is a payment from the buyer to the intermediary and $M_S(\hat{v}, \hat{c})$ is a payment from the intermediary to the seller. The resulting utilities are $vQ(\hat{v}, \hat{c}) - M_B(\hat{v}, \hat{c})$ for the buyer and $M_S(\hat{v}, \hat{c}) - cQ(\hat{v}, \hat{c})$ for the seller. Letting $q_B(v) := \mathbb{E}_c[Q(v, c)]$ and $q_S(c) := \mathbb{E}_v[Q(v, c)]$ be the interim expected probability of trade and $m_B(v) := \mathbb{E}_c[M_B(v, c)]$ and $m_S(c) := \mathbb{E}_v[M_S(v, c)]$ the interim expected payments, anticipating that the other trader reports truthfully for all possible types, the mechanism is Bayesian incentive compatible if and only if, for all $v, \hat{v} \in [\underline{v}, \overline{v}]$ and $c, \hat{c} \in [\underline{c}, \overline{c}]$,

$$q_B(v)v - m_B(v) \geq q_B(\hat{v})v - m_B(\hat{v}) \quad \text{and} \quad m_S(c) - q_S(c)c \geq m_S(\hat{c}) - q_S(\hat{c})c \tag{1}$$

and interim individually rational if for all $v \in [\underline{v}, \overline{v}]$ and all $c \in [\underline{c}, \overline{c}]$

$$q_B(v)v - m_B(v) \geq 0 \quad \text{and} \quad m_S(c) - q_S(c)c \geq 0. \tag{2}$$

That is, a direct mechanism is (Bayesian) incentive compatible if truthful reporting constitutes a Bayes Nash equilibrium and individual rational if each trader is, for each possible type, weakly better off participating than walking away. Given the by now well-known equivalence between Bayesian and dominant strategy incentive compatibility and interim and ex post individual rationality (see e.g. Gershkov et al., 2013), we will simply speak of incentive compatibility and individual rationality. As mentioned, by the payoff equivalence theorem (see e.g. Börgers, 2015), the allocation rule $Q(v,c)$ and incentive compatibility pin down the interim expected payments up to a constant. Under profit maximization, the constants are pinned down by setting $q_B(\underline{v})\underline{v} - m_B(\underline{v}) = 0$ and $m_S(\overline{c}) - q_S(\overline{c})\overline{c} = 0$, that is, by making the least efficient types of each trader indifferent between participating and not. As shown by Myerson and Satterthwaite (1983, Theorem 3), the optimal allocation rule is such that $Q(v,c) = 1$ if $\Phi(v) \geq \Gamma(c)$ holds and $Q(v,c) = 0$ otherwise. Because $\Phi(\overline{v}) = \overline{v}$, it follows that a seller with cost $c > \Gamma^{-1}(\overline{v})$ never trades.

3.1 Implementation via fee-setting mechanisms

Consider now fee-setting mechanisms. As a buyer with value v accepts the price p if and only if $v \geq p$, the problem of a seller with cost c who sets a price p and has to pay the fee $\omega(p)$ if a transaction occurs is to

$$\max_p (p - \omega(p) - c)(1 - F(p)). \tag{3}$$

Therefore, the allocation that induces trade if and only if $\Phi(v) \geq \Gamma(c)$ can be implemented with a fee $\omega(p)$ if and only if a seller with cost $c \leq \Gamma^{-1}(\overline{v})$ sets a price p such that $\Phi(p) = \Gamma(c)$, that is, sets $p = \Phi^{-1}(\Gamma(c))$.

Note that this reasoning only holds for a seller with $c \leq \Gamma^{-1}(\overline{v})$. As a seller with $c > \Gamma^{-1}(\overline{v})$ never trades under the optimal mechanism, sellers with costs above $\Gamma^{-1}(\overline{v})$ need to be incentivized to set prices that are never accepted. Because prices $p > \overline{v}$ never lead to trade in equilibrium, which means that the fee for prices above $\overline{v}$ does not matter, the only remaining concern is that such seller-types might benefit from setting prices below $\overline{v}$. As we will discuss after the following proposition, the optimal fee automatically takes care of this by making it strictly unprofitable for types $c > \Gamma^{-1}(\overline{v})$ to set prices below $\overline{v}$.

The following proposition describes the fee $\omega_I(p)$ that is optimal for the intermediary because it allows the intermediary to obtain the maximum profit derived by Myerson and Satterthwaite (1983, Theorem 3). Throughout the paper, whenever a displayed fee formula is stated only on its equilibrium price range, it is understood to be completed off equilibrium by an extension that leaves the stated pricing rule optimal. The construction after Proposition 1 provides this extension.

Proposition 1. *The optimal fee is* $\omega_I(p) := p - \mathbb{E}_v[\Gamma^{-1}(\Phi(v))|v \geq p]$*, defined on the domain* $[\Phi^{-1}(\underline{c}), \overline{v}]$*. This fee implements the profit-maximizing mechanism.*

The domain $[\Phi^{-1}(\underline{c}), \overline{v}]$ in Proposition 1 is restricted to prices that are set on the equilibrium path of play. These are the prices for which the fee ω_I is pinned down by the optimal mechanism. The displayed formula can be extended to all prices without changing sellers' optimal pricing decisions or the implemented allocation.[8]

Note that the net price $p - \omega_I(p)$ is increasing in p and, by L'Hôpital's Rule, equal to $\Gamma^{-1}(\overline{v})$ at $p = \overline{v}$. Therefore, a seller with cost $c > \Gamma^{-1}(\overline{v})$ makes an expected loss by setting $p < \overline{v}$. This is the sense in which ω_I automatically deters this kind of deviation.

The proof of Proposition 1 is —perhaps surprisingly— simple: plugging the optimal fee into the seller's maximization problem and solving this maximization problem shows that the seller with cost $c \leq \Gamma^{-1}(\overline{v})$ optimally sets the price $\Phi^{-1}(\Gamma(c))$. As noted before the proposition, this pricing rule implements the optimal allocation rule. Further, the least efficient types of the buyer and the seller are easily seen to obtain a utility of zero. For the buyer this follows because the lowest price set in equilibrium is $\Phi^{-1}(\Gamma(\underline{c})) = \Phi^{-1}(\underline{c})$, which is larger than $\underline{c}$ because $\Phi(v) < v$ for all $v < \overline{v}$, and $\underline{c} \geq \underline{v}$. Thus, the buyer with value $\underline{v}$ never trades.

As an illustration, assume that F and G are uniform on $[0, 1]$, in which case $\omega_I(p) = p/2$. It is easy to check that the intermediary's expected profit is $1/24$, the same as the profit Myerson and Satterthwaite (1983) get for the uniform-uniform example they provide. Notice, however, the remarkable generality of the implementation result in Proposition 1. It applies for *any* distributions F and G satisfying regularity.

Moreover, the implementation result via fee setting does not depend on the mechanism being optimal for the intermediary. Indeed, *any* monotone allocation rule can be

[8] For example, fix $\varepsilon > 0$ and set $\omega_I(p) = p - \underline{c} + \varepsilon$ for $p < \Phi^{-1}(\underline{c})$. Then $p - \omega_I(p) = \underline{c} - \varepsilon < c$ for every seller type, so trade at such a price gives every seller a negative payoff. For $p < \underline{c}$, setting $\omega_I(p) = 0$ would also suffice. For $p > \overline{v}$, no buyer accepts, so the fee may be defined arbitrarily.

implemented via fee setting. To see this, assume that there is an allocation rule $H(c)$ such that there is a trade between the seller and the buyer if and only if $v \geq H(c)$ and that H is strictly increasing for all $c \in (\underline{c}, \overline{c})$ such that $H(c) \in (\underline{v}, \overline{v})$. Then the fee function $\omega_H(p) := p - \mathbb{E}_v[H^{-1}(v)|v \geq p]$ implements this allocation rule, which is readily seen by replacing $\omega_I(p)$ in (3) by $\omega_H(p)$ and solving the seller's optimization problem.[9] Further, the implementation via fee setting is also robust insofar as the buyer has a dominant strategy, and anticipating that the buyer plays the dominant strategy, the seller's problem becomes a single-person decision problem. This contrasts, for example, with the multiplicity of equilibria that arises in a k-double auction.[10]

Optimality of linear fees As Proposition 1 shows, in general, the intermediary-optimal fee $\omega_I(p)$ is potentially a complicated, non-linear function that may not even permit closed-form solutions. In the real world, fees are typically simple and often take the linear form $\omega(p) = a + bp$ or even the proportional form $\omega(p) = bp$, where a and b are constant real numbers. If $\omega(p) = bp$, b has the interpretation of a percentage fee.[11] A case in point is real estate brokerage, where percentage fees are almost universally used. Amazon also makes wide use of percentage fees, which it applies in about two thirds of its product categories for third-party sellers.[12] Without calling it a transaction fee, Alibaba charges third-party sellers a "handling" fee of two percent for sales for up

[9] The first-order condition is $0 = -f(p)(H^{-1}(p) - c)$, which is equivalent to $p = H(c)$, and the second-order condition is satisfied whenever the first-order condition is because H and hence H^{-1} is increasing. As an illustration, consider ex post efficiency, in which case $H(c) = c$. If F and G are both uniform on $[0, 1]$, then $\omega_H(p) = (p-1)/2$ and the intermediary's expected profit is $-1/6$. More generally, under ex post efficiency, $\omega_H(p) = p - \frac{\int_p^{\overline{v}} v\,dF(v)}{1-F(p)}$. If F and G have the same support $[0, 1]$, the intermediary's expected profit is $-(\int_0^1 vG(v)dF(v) - \int_0^1 c(1-F(c))dG(c))$, that is, the deficit is equal to the ex ante expected gains from trade.

[10] As observed by Myerson and Satterthwaite (1983), when F and G are uniform, the k-double auction of Chatterjee and Samuelson (1983) has an equilibrium that implements the second-best mechanism for $k = 1/2$. While we do not further elaborate on this here, it can be shown that, with an appropriately chosen fee, the second-best mechanism can be implemented with basically the same robustness as the profit-maximizing mechanism. As an example, when F and G are uniform, the fee $\omega^{SB}(p) = p/2 - 1/4$ induces the second-best allocation rule.

[11] We use the terms proportional and percentage fees interchangeably. Fees of the form $\omega(p) = a + bp$ are linear in the same way as, say, a demand function would be. What we call proportional would be called a linear tariff in much of IO — no terminology is perfect.

[12] Third-party sellers on Amazon in the United States are charged percentage fees in 24 out of 35 categories while in Australia, percentage fees apply in 20 out of 31 categories (accessed on April 23, 2026).

to \$5,000 and a flat fee of \$100 for more expensive products. This raises the question under which conditions the optimal fees are linear. We are now going to shed light on this question by first deriving necessary and sufficient conditions on the distributions F and G for this to be the case. In a second step, we provide a thin market based rationale for when one should expect these conditions to hold.

We have already seen that $\omega_I(p) = p/2$ when both F and G are uniform. Our first set of results imply that, up to the no-trade truncation, the optimal fee is $\omega_I(p) = p/2$ if and only if seller costs are uniformly distributed. That is, costs are uniform over the range of seller types that trade with positive probability; the distribution above the no-trade cutoff is not identified and does not affect the fee. More generally, whether the optimal fee is linear (or a percentage fee) only depends on G. To see this, consider the cumulative distribution function

$$G(c) = \left(\frac{c - \underline{c}}{\overline{c} - \underline{c}}\right)^{\sigma}, \tag{4}$$

with support $[\underline{c}, \overline{c}]$ and shape parameter $\sigma > 0$. Note that the virtual cost function associated with the distribution in (4) is $\Gamma(c) = \frac{\sigma+1}{\sigma}c - \frac{1}{\sigma}\underline{c}$, which is linear in c. For example, for $\sigma = 1, \underline{c} = 0$ and $\overline{c} = 1$, $G(c)$ is the uniform on $[0, 1]$ and $\Gamma(c) = 2c$, implying $\Gamma^{-1}(p) = p/2$ as noted. Moreover, $\Gamma(c)$ is independent of $\overline{c}$, which reflects a general truncation-invariance property of virtual costs that we will come back to below.[13] We refer to this distribution as *reflected generalized Pareto distribution* since it is the standard generalized Pareto distribution mirrored around the vertical axis. Because $\Gamma(c)$ is linear, so is its inverse. Hence,

$$\omega_I(p) = p - \mathbb{E}_v[\Gamma^{-1}(\Phi(v)|v \geq p)] = p - \Gamma^{-1}(\underbrace{\mathbb{E}_v[\Phi(v)|v \geq p]}_{=p}) = p - \Gamma^{-1}(p),$$

where the first equality holds by the definition of the optimal fee, the second because Γ^{-1} is linear and the third because of the equality $\mathbb{E}_v[\Phi(v)|v \geq p] = p$, which is well known in mechanism design.[14] In other words, if $\Gamma(c)$ is linear, so is $\omega_I(p)$.

[13]To see this truncation-invariance, denote for an arbitrary $u \in (\underline{c}, \overline{c}]$ by $G(c|u) := \frac{G(c)}{G(u)}$ the distribution of costs conditional on $c \leq u$, whose density we denote by $g(c|u)$. Because $g(c|u) = \frac{g(c)}{G(u)}$, the virtual cost for $c \leq u$ is $c + \frac{G(c)}{g(c)}$.

[14]Intuitively, as observed by Myerson (1981), the buyer's informational rent is given by the expected

Identification with linear fees As we have just seen, if G is a reflected generalized Pareto distribution, then the optimal fee is linear. The converse is also true up to the no-trade truncation: If the optimal fee is linear, then $\Gamma(c)$ has to be linear over the range of costs that trade with positive probability, and the corresponding truncated distribution has to be of the form in (4). A priori, one might think that, through some fluke, F and G might "conspire" to make $\omega_I(p)$ linear even if Γ is not. After all, the optimal fee as shown in Proposition 1 reveals that in general, the fee depends on both the buyer's and the seller's type distribution through Φ and Γ^{-1}, respectively. We show in the following that this is not possible, that is, an optimal linear fee implies a reflected generalized Pareto distribution. Further, this distribution is uniquely identified up to a truncation by the fee.

We provide formal identification results for G and F in Appendix A and a brief summary here. The identification of the distribution G is based on (i) optimal behavior by the seller and (ii) the intermediary optimality of the allocation rule. Take a linear fee $\omega_L(p) := a + bp$. Seller optimality (i) implies that the price p maximizes $(p - \omega_L(p))(1 - F(p))$. The first-order condition of the seller's maximization problem can be rearranged to $\Phi(p) = (c + a)/(1 - b)$. Intermediary optimality (ii) implies $\Phi(p) = \Gamma(c)$. Combining (i) and (ii) implies $\Gamma(c) = (c + a)/(1 - b)$, i.e. Γ is linear and uniquely pinned down by ω_L. Γ linear implies a reflected generalized Pareto distribution G (up to the usual truncation) with parameters identified by the parameters of Γ.

A further interesting property of optimal linear fees is that the distribution of prices posted by the seller alone identifies the distribution F of buyers' valuations (up to a truncation). This is surprising insofar as the price distribution depends in general on both the buyer's and the seller's price distribution. That is, the model not only predicts price dispersion in equilibrium, which is consistent with empirical observations, because sellers' costs and hence their optimal prices are heterogeneous. When linear fees are optimal, the distribution of prices also identifies the distribution of buyers' valuations.

virtual type, and the optimal mechanism for one buyer is a take-it-or-leave-it price. More formally, it can be verified that multiplying both sides of the equation $\mathbb{E}_v[\Phi(v)|v \geq p] = p$ by $1 - F(p)$ and taking derivatives yields the same expression. Further, the two sides of the equation are equal for $p = \overline{v}$.

3.2 Thin markets and the optimality of linear fees

As we have just seen, $\omega_I(p)$ is linear if $\Gamma(c)$ is linear, which is equivalent to saying that $G(c)$ is of the form (4). We are now going to show that in thin markets, defined as markets in which only the types of agents participate who are keenest to trade, linear fees are optimal. Our asymptotic approach relies on Extreme Value Theory, which states that under mild regularity conditions the lower tail of any distribution converges to a reflected generalized Pareto distribution as one moves the truncation point closer to the lower bound of support (see Appendix D). These regularity conditions can be shown to be satisfied in our setup.

We apply these results to sellers and look at truncation points that go to the lower bound of support. This kind of truncation is a natural way of capturing the fact that in many markets, at any given point in time only a small fraction of all potential traders are active. For example, in real estate, typically less than 5 percent of all properties are being offered for sale. Our notion of thin markets thus consists simultaneously of a small number of traders, in particular, one seller, and selection of the keenest trader types.

Formally, we now consider sequences of economies as described in Section 2 indexed by j for which the length of the overlap of support of v_j and c_j goes to zero. The analysis simplifies if we normalize by the length of the overlap to get the normalized variables $\tilde{c} := (c_j - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$, $\tilde{v} := (v_j - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$, and $\tilde{p} := (p - \underline{c}_j)/(\overline{v}_j - \underline{c}_j)$.The distributions of the normalized effective cost $\tilde{c}$ and the normalized effective valuation $\tilde{v}$, truncated to $[\underline{c}_j, \overline{v}_j]$ and denoted, respectively, $\tilde{G}_j$ and $\tilde{F}_j$, are then given as

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

where u_j^B and u_j^S are overlap of support ratios defined in Section 2.3, see Figure 1. Accordingly, the normalized fee is defined as

$$\tilde{\omega}_j(\tilde{p}) := \frac{\omega((\overline{v}_j - \underline{c}_j)\tilde{p} + \underline{c}_j)}{\overline{v}_j - \underline{c}_j}.$$

For parameters $\beta > 0$ and $\sigma > 0$, we refer to the distributions

$$\tilde{F}^*(\tilde{v}) := 1 - (1 - \tilde{v})^\beta \quad \text{and} \quad \tilde{G}^*(\tilde{c}) := \tilde{c}^\sigma$$

with (normalized) support $[0,1]$ as generalized Pareto respectively reflected generalized Pareto distributions. In the statistics literature (see Appendix D for some of the references), the constants β and σ are referred to as the Pareto tail indices. These have been observed to be remarkably invariant over time for a number of applications.

Since there is a one-to-one mapping between $(\underline{c}_j, \overline{c}_j, \underline{v}_j, \overline{v}_j)$ and $(\underline{c}_j, u_j^S, \overline{v}_j, u_j^B)$ we can state our convergence results in terms of the latter.

Proposition 2. *Let $(\underline{c}_j, u_j^S, \overline{v}_j, u_j^B)$ be an arbitrary sequence with u_j^S and u_j^B going to zero as j goes to infinity, satisfying $\underline{c}_j < \overline{v}_j$ for all j. Then:*

(i) the buyer's and the seller's normalized distributions converge to generalized Pareto and reflected generalized Pareto distributions, respectively, that is,

$$\lim_{j\to\infty} \tilde{F}_j(\tilde{v}) = 1 - (1-\tilde{v})^\beta \quad \text{and} \quad \lim_{j\to\infty} \tilde{G}_j(\tilde{c}) = \tilde{c}^\sigma,$$

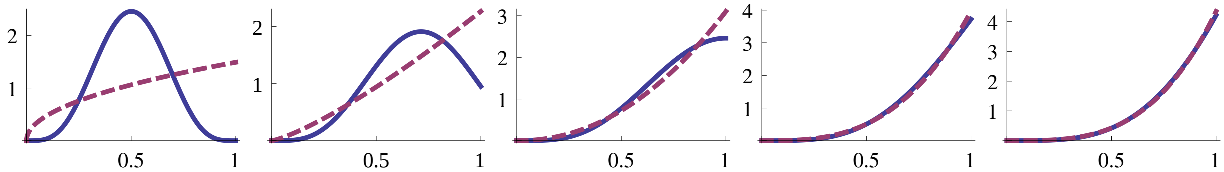
for all $\tilde{v} \in [0,1]$ and $\tilde{c} \in [0,1]$.

(ii) the normalized fee $\tilde{\omega}_j(\tilde{p})$ converges to a linear fee, that is:

$$\lim_{j\to\infty} \tilde{\omega}_j(\tilde{p}) = \frac{1}{1+\sigma}\tilde{p}.$$

Proposition 2 implies that if $\underline{c}_j$ converges to a well-defined limit $\underline{c}$, then the fee ω_j converges to $\omega(p) = p/(1+\sigma) - \underline{c}/(1+\sigma)$, which can be obtained by inverting the relationship between $\tilde{\omega}$, $\tilde{p}$ and ω, p, respectively.

We first provide an intuition for part (i) of Proposition 2. Convergence of the distribution is immediate when the primitive distributions G_0 and F_0 are (reflected) generalized Pareto distributions on $[0,1]$, that is if $G_0(c_0) = c_0^\sigma$ for $\sigma > 0$ and $F_0(v_0) = 1-(1-v_0)^\beta$ with $\beta > 0$, because this implies $\tilde{G}_j(\tilde{c}) = G_0(u_j^S\tilde{c})/G_0(u_j^S) = \tilde{c}^\sigma$ and $1-\tilde{F}_j(\tilde{v}) = (1-F_0(1-u_j^B(1-\tilde{v}))/(1-F_0(1-u_j^B)) = (1-\tilde{v})^\beta$ for all j. In other words, $\tilde{G}_j$ and $\tilde{F}_j$ do not change with j. This is, of course, the well-known property of Pareto distributions that they are invariant to truncation. The intuition for part (ii) of Proposition 2 is that for the limiting reflected generalized Pareto distribution to which G_j converges, the optimal fee is linear.



(a) $u = 1$ (b) $u = 0.7$ (c) $u = 0.5$ (d) $u = 0.3$ (e) $u = 0.2$

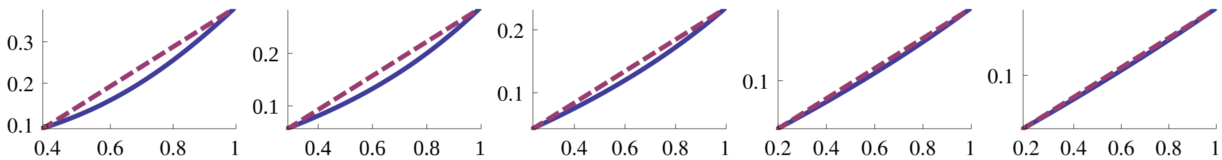
Figure 2: Density of truncated, rescaled distribution for a Beta distribution with support $[0, 1]$ and density $g_0(c) \propto c^4(1-c)^4$ (solid line) compared to an approximating reflected generalized Pareto density with support $[0, 1]$ (dashed).

The reasoning in the above paragraph and part (i) of Proposition 2 do not yet constitute a proof of part (ii) of the proposition since one still has to establish that the transformation from distributions to fees is continuous. We relegate the proof to Appendix B and instead use Figures 2 and 3 here to illustrate the convergence behavior of the underlying distribution to a Pareto distribution and of the optimal fee to a linear fee. The underlying seller's distribution for both figures is a Beta distribution with support $[0, 1]$ and a symmetric density proportional to $c^4(1-c)^4$, and the truncated supports in panels (b) to (e) in both figures are $[0, u]$ with $u \in \{0.7, 0.5, 0.3, 0.2\}$ with associated masses of active traders of $\{0.9, 0.5, 0.1, 0.02\}$.[15] Figure 2 displays the (truncated, rescaled) densities of the Beta distribution and the density of the approximating generalized Pareto distribution (dashed) while Figure 3 shows the optimal fee for the truncated density (solid) and an approximating linear fee (dashed). A truncation that only retains the lower 10 percent of the distribution — that is, $u = 0.3$ — already results in a distribution that is visually hard to distinguish from a reflected generalized Pareto distribution (Figure 2(d)) and an optimal fee function that is barely distinguishable from a linear fee (Figure 3(d)).[16]

To analyze and interpret the true – that is, denormalized – fee $\omega(p)$ that is approxi-

[15] Formally, the displayed density is that of the truncated and rescaled distribution $G_u(c) := G_0(\underline{c} + u(c - \underline{c}))/G_0(\underline{c} + u(\overline{c} - \underline{c}))$ for $u \in \{1, 0.7, 0.5, 0.3, 0.2\}$ for a Beta distribution with support $[0, 1]$ and density $g_0(c) \propto c^4(1-c)^4$.

[16] While for the Pareto distribution, the optimal fee does not depend on the buyer's distribution, the optimal fee in general does. This requires us to make assumptions about the buyer's distribution in Figure 3, which we assume to be the same as the seller's and to be truncated to $[u, 1]$ for the same values of u as for the seller's distribution.



(a) $u = 1$ (b) $u = 0.7$ (c) $u = 0.5$ (d) $u = 0.3$ (e) $u = 0.2$

Figure 3: Optimal fee $\omega(\cdot)$ for truncated, rescaled distributions for the same setup as in Figure 2 (solid line) compared to an approximating linear fee (dashed).

mately optimal away from but sufficiently close to the limit, it is also insightful to consider a j such that u_j^S and u_j^B are "sufficiently small" and to work with the denormalized limiting generalized Pareto distributions

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

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

$$\omega_I(p) = p - \Gamma^{*-1}(p) \tag{5}$$

is linear in p.

3.3 Comparative statics

The expression in (5) offers a neat interpretation and simple comparative statics: A monopsonist with valuation v who faces the supply $G^*(\hat{p})$ would set the price $\Gamma^{*-1}(v)$. A more elastic supply means a higher monopsony price and hence a lower fee. As an illustration, $\tilde{G}(\tilde{c}) = \tilde{c}^\sigma$ can be seen as isoelastic supply with elasticity σ. As the elasticity σ increases, the fee $\tilde{\omega}(\tilde{p}) = \tilde{p}/(\sigma + 1)$ decreases.[17] Surprisingly, and interestingly, the fee is independent of the distribution of valuations, which is a point we will expand on below.

[17]For the non-normalized price and a reflected generalized Pareto G, the fee $\omega(p) = (p - \underline{c})/(1 + \sigma)$ decreases as σ increases for all prices p that may be set in equilibrium. For out-of-equilibrium prices $p < \underline{c}$, $\omega(p)$ increases in σ. In what follows, we say that the fee decreases if $\omega(p)$ decreases for all prices p set in equilibrium.

Recall the well-known industrial organization interpretation of the virtual type functions Φ and Γ going back to Mussa and Rosen (1978) and Bulow and Roberts (1989): if one views $1-F(p)$ and $G(p)$ as expected quantities demanded and supplied, $\Phi(p)$ and $\Gamma(p)$ have the interpretation of marginal revenue and marginal cost functions.[18] Interpreting $1-F(p)$ and $G(p)$ in this way, one can then define the price elasticity of demand at v as $\eta_d(v) := vf(v)/(1-F(v))$ and the price elasticity of supply at c as $\eta_s(c) := cg(c)/G(c)$.

The comparative statics with respect to η_s are as expected: a global increase in $\eta_s(c)$ leads to lower fees because an increase of $\eta_s(c)$ for all c implies a decrease of $\Gamma(c) = c(1+1/\eta_s(c))$ and hence an increase of Γ^{-1} and a decrease of $\omega_I(p) = p - \mathbb{E}_v[\Gamma^{-1}(\Phi(v))|v \geq p]$, which corresponds to the results on isoelastic supply and monopsony above. The effect of the elasticity of demand η_d (and equivalently of $\Phi(v) = v(1-1/\eta_d(v))$) is clearly more complicated. It is useful to first consider the limit, in which G^* is reflected generalized Pareto, so that both Γ^* and the optimal fee ω_I are linear. Equation (5) shows that in this limit, the elasticity of demand does not play any role for the fee whatsoever. This suggests that the way Γ differs from a linear function determines how η_d affects the optimal fee. As shown below, this is indeed the case, and we also provide an intuitive economic interpretation of the effects of the elasticity of demand.

Proposition 3. *Assume Γ^{-1} can be represented on its domain by its Taylor series around $\overline{v}$. Using the Taylor expansion of $\Gamma^{-1}(x)$ around $\overline{v}$, the net price received by the seller is*

$$\begin{aligned} p - \omega_I(p) = \; & \overbrace{\Gamma^{-1}(p)}^{\textit{monopsony price}} + \overbrace{\frac{\overline{\gamma}_2}{2} Var_{v\sim F}[\Phi(v) - \overline{v}|v \geq p]}^{\textit{second-derivative price endogeneity effect}} \\ & \underbrace{+ \sum_{n=3}^{\infty} \frac{\overline{\gamma}_n}{n!} \left\{ \mathbb{E}_{v\sim F}[(\Phi(v) - \overline{v})^n | v \geq p] - \mathbb{E}_{v\sim F}[\Phi(v) - \overline{v}|v \geq p]^n \right\}}_{\textit{higher-derivative price endogeneity effects}}, \end{aligned} \tag{6}$$

where $\overline{\gamma}_n := (\Gamma^{-1})^{(n)}(\overline{v})$ denotes the n-th derivative of Γ^{-1} at $\overline{v}$.

Recall that a monopsonist with value x would set the monopsony price $\Gamma^{-1}(x)$. Naturally, the intermediary's valuation for the good is p because p is what the buyer will

[18] Setting $q = 1-F(p)$, the revenue when selling q is $F^{-1}(1-q)q$. Consequently, marginal revenue is $[F^{-1}(1-q)q]'_{p=F^{-1}(1-q)} = \Phi(p)$. Analogous reasoning applies to $\Gamma(p)$.

pay. Therefore, absent any other effects, the seller's net price should be $\Gamma^{-1}(p)$, which is exactly what the seller receives when Γ is linear. In other words, for linear Γ and hence linear fees, the price setting incentives of the seller are aligned with the intermediary's objective, so that the intermediary does not have to distort the monopsony pricing rule. However, if Γ is not linear, the intermediary has to take into account that the price p is determined endogenously, which requires the optimal net price the seller receives to be adjusted in order to incentivize the seller, which is reflected in the second- and higher-derivative effects.

According to Proposition 3, η_d has no first-derivative effect on the optimal fee. Indeed, as seen above when G is a reflected generalized Pareto distribution, $\omega_I(p)$ is independent of F. However, the second- and higher-derivative effects can go either way. To see this, assume that $\Gamma^{-1}(x)$ is quadratic, that is, it is of the form $\Gamma^{-1}(x) = \overline{\gamma}_0 + \overline{\gamma}_1 x + \overline{\gamma}_2 x^2/2$.The quadratic form shuts down the higher-derivative effects and allows us to focus on the second-derivative price endogeneity effect. The effect of a change of the elasticity of demand depends on the curvature, as shown in the following proposition.

Proposition 4. *Assume Γ^{-1} is quadratic. Then an overall increase of the demand elasticity leads to (i) an overall increase of the fee if Γ^{-1} is convex, (ii) an overall decrease of the fee if Γ^{-1} is concave, and (iii) no effect if Γ^{-1} is linear.*

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

Important contributions by Bulow and Pfleiderer (1983), Aguirre et al. (2010), Bulow and Klemperer (2012), and Weyl and Fabinger (2013) have identified a number of properties of the demand function such as its curvature, the curvature of the inverse demand function, the pass-through rate, and the markup- or quantity-weighted average pass-through, which prove useful in a variety of contexts in industrial organization. Naturally, one may then wonder whether the counterintuitive result that optimal fees are sometimes higher for a higher elasticity of demand may be explained by alternative properties of the demand function. The example with Γ^{-1} quadratic shows that the answer is no. Any change of any property of the demand function F that leads to higher fees for Γ^{-1} convex will lead to lower fees for Γ^{-1} concave and will have no effect on the fees for a linear Γ^{-1}. This can be seen directly by plugging a quadratic Γ^{-1} into (6), i.e., dropping the higher-derivative price endogeneity effects.

These results may also be of relevance for the practice of competition policy. When there is suspicion that fee-setting intermediaries, such as auction houses and platforms or real-estate brokers, collude, the standard approach would be to estimate the demand function and to then evaluate how closely prices are to the monopoly price implied by the estimates. Our analysis suggests that in thin markets with intermediaries, the more relevant object of interest is the elasticity of *supply* rather than the elasticity of demand. As a first-order approximation, the elasticity of demand does not matter for the fees of a profit-maximizing intermediary. Instead, colluding intermediaries should be expected to leave a net price to the seller which corresponds to the price set by a monopsonist whose valuation, as a first-order approximation, is the gross price.

4 Generalization

Motivated by the widespread use of fee setting in intermediated markets in which buyers arrive over time, we now set up and analyze a more general infinite-horizon model with commitment. By extending the model to allow for multiple and randomly arriving buyers in each period and the intermediary to maximize a weighted average of its and the seller's expected surplus, we will show that the key insights of the previous section—optimality of fees, and optimality of linear fees in thin markets—generalize considerably. Along the way, we introduce the concepts of dynamic virtual valuation functions, weighted virtual valuation functions, and discounted probabilities of trade that allow for a natural generalization of the results in the previous section. In particular, we will see that the optimal fee increases as the intermediary extracts more rents and that if G is a reflected generalized Pareto distribution, the optimal fee is independent of the discount factor and thus the same as in a static setting.

4.1 Preliminaries

We first introduce the setup for the infinite-horizon model and then define direct mechanisms for this setting.

Infinite-horizon setup Time is discrete and indexed by $t = 0, 1, ...$ In every period t, there is a fixed number of potential buyers $\overline{B}$, each of whom enters with the independent probability $\tilde{\pi}$, so that the probability π_B of having exactly $B \leq \overline{B}$ buyers is given by the probability mass function for the binomial $\pi_B := \begin{pmatrix} \overline{B} \\ B \end{pmatrix} \tilde{\pi}^B (1 - \tilde{\pi})^{\overline{B}-B}$. Buyers who participate are sometimes also called bidders. Each bidder draws her valuation v independently from the distribution F with support $[\underline{v}, \overline{v}]$ and density $f(v) > 0$ for all $v \in (\underline{v}, \overline{v})$. The distributions of the highest and second-highest draws from F with support $[\underline{v}, \overline{v}]$ are denoted by $F_{(1)}(v)$ and $F_{(2)}(v)$ and given by $F_{(1)}(v) := \sum_{B=0}^{\overline{B}} \pi_B F(v)^B$ and $F_{(2)}(v) := F_{(1)}(v) + (1 - F(v)) \sum_{B=1}^{\overline{B}} \pi_B B F(v)^{B-1}$, respectively, with densities $f_{(1)}(v)$ and $f_{(2)}(v)$.

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. That said, the interpretation of many concepts used in the mechanism design framework is most straightforward if one interprets δ as a pure survival probability.

The seller and the intermediary contract at date 0. For a given incentive compatible and individual rational dynamic mechanism, denoted $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ and spelled out in more detail below, we denote the seller's and the intermediary's expected discounted profit when the seller's cost is c by $\mathcal{W}_S(c)$ and $\mathcal{W}_I(c)$, respectively. We assume that, subject to incentive compatibility and individual rationality constraints for the buyers and the seller, the contract between the intermediary and the seller maximizes over $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$

$$\mathbb{E}_c[\alpha \mathcal{W}_I(c) + (1 - \alpha)(\mathcal{W}_I(c) + \mathcal{W}_S(c))] \tag{7}$$

where $\alpha \in [0, 1]$ can be thought of as a reduced form parameter that measures competition between intermediaries for sellers with $\alpha = 1$ corresponding to monopoly (or perfect collusion) and $\alpha = 0$ capturing perfect competition. Setting $\delta = 0$, $\alpha = 1$ and $\overline{B} = 1$ with $\pi_1 = 1$ yields the model analyzed in Section 3 as a special case.

The contract between the intermediary and the seller ends when the seller exits, which can happen because of the exogenous probability of exit or because the seller's object is sold. In addition, only buyers who have arrived can be contracted with, and because each buyer is short-lived, buyers only participate in the period in which they arrive.

Dynamic mechanisms with commitment We say that a mechanism is *active* in period t if the seller has not exited prior to t. A mechanism is *direct* if it asks all agents who participate in the mechanism to report their types. For the seller, this simply means that he reports his cost c at date 0. A direct mechanism then asks all buyers b who enter in period $\tau \geq 0$ to report their valuations $v_b \in [\underline{v}, \overline{v}]$ to the mechanism designer. The valuations of buyers who do not enter are set equal to $-\infty$. Thus, $v_b = -\infty$ with probability $1-\tilde{\pi}$, while for $v \in [\underline{v}, \overline{v}]$ its unconditional cumulative distribution function is $\Pr(v_b \leq v) = 1 - \tilde{\pi} + \tilde{\pi}F(v)$. Whether they enter or not, buyers are indexed in increasing order, so that in period 0, there are $\overline{B}$ buyers and by period 1 the set of buyers has $2\overline{B}$ elements. Accordingly, in any given period t the vector of valuations $\boldsymbol{v}^t$ is $\overline{B}$-dimensional and an element of $\mathcal{V}^{\overline{B}}$ where the support is denoted by $\mathcal{V} := (\{-\infty\} \cup [\underline{v}, \overline{v}])$. For each $t \geq 0$, let $\boldsymbol{v}_t := (\boldsymbol{v}^0, \ldots, \boldsymbol{v}^t)$ be the vector that contains all valuations from period 0 to period t, which is $(t+1)\overline{B}$-dimensional. To ease notation, we denote the set of period t buyers by $\mathcal{B}_t := \{t\overline{B}+1, \ldots, (t+1)\overline{B}\}$ and the set of all buyers by $\mathcal{B} := \cup_{t=0}^{\infty} \mathcal{B}_t$.

A direct mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ consists of an allocation rule $\boldsymbol{Q} := (\boldsymbol{Q}_S, \boldsymbol{Q}_B)$ and a payment rule $\boldsymbol{M} := (\boldsymbol{M}_S, \boldsymbol{M}_B)$ as follows. The probability that the seller of type c sells the good in period t given $\boldsymbol{v}_t$ is denoted $Q_S^t(\boldsymbol{v}_t, c)$. The probability that buyer $b \in \mathcal{B}_t$ is allocated the good in period t upon the collection of reports $(\boldsymbol{v}_t, c)$ is denoted $Q_b^t(\boldsymbol{v}_t, c)$. The probability of being allocated the good is also 0 for all buyers who do not enter. Analogously, the payment made by buyer $b \in \mathcal{B}_t$ to the mechanism designer given reports $(\boldsymbol{v}_t, c)$ is denoted $M_b^t(\boldsymbol{v}_t, c)$ and the payment from the mechanism designer to the seller is denoted $M_S^t(\boldsymbol{v}_t, c)$. If the mechanism is no longer active in period t, then all payments and probabilities are 0. Feasibility further requires

$$\sum_{b\in\mathcal{B}_t} Q_b^t(\boldsymbol{v}_t, c) \leq Q_S^t(\boldsymbol{v}_t, c) \tag{8}$$

for all t, all $\mathcal{B}_t$ and all $(\boldsymbol{v}_t, c)$.[19] Conditional on being active at the beginning of period t, the good is sold with probability $Q_S^t(\boldsymbol{v}_t, c)$. If δ is interpreted as a survival probability, the mechanism remains active in period $t+1$ with probability $\delta(1-Q_S^t(\boldsymbol{v}_t, c))$. If δ instead represents pure time discounting, the physical continuation probability is $1 - Q_S^t(\boldsymbol{v}_t, c)$, while $\delta(1 - Q_S^t(\boldsymbol{v}_t, c))$ is the corresponding discounted continuation weight.

For a given $(\boldsymbol{v}_t, c)$, letting $\boldsymbol{Q}^t_{\mathcal{B}_t}(\boldsymbol{v}_t, c) := (Q_b^t(\boldsymbol{v}_t, c))_{b\in\mathcal{B}_t}$ and $\boldsymbol{M}^t_{\mathcal{B}_t}(\boldsymbol{v}_t, c) := (M_b^t(\boldsymbol{v}_t, c))_{b\in\mathcal{B}_t}$, we have $\boldsymbol{Q}_S(\boldsymbol{v}, c) := \left(\boldsymbol{Q}_S^t(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty}$, $\boldsymbol{Q}_B(\boldsymbol{v}, c) := \left(\boldsymbol{Q}^t_{\mathcal{B}_t}(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty}$, $\boldsymbol{M}_S(\boldsymbol{v}, c) := (M_S^t(\boldsymbol{v}_t, c))_{t=0}^{\infty}$ and $\boldsymbol{M}_B(\boldsymbol{v}, c) := \left(\boldsymbol{M}^t_{\mathcal{B}_t}(\boldsymbol{v}_t, c)\right)_{t=0}^{\infty}$, where $\boldsymbol{v} = \boldsymbol{v}_\infty$. Consequently, the allocation and payment rules $\boldsymbol{Q}$ and $\boldsymbol{M}$ are, respectively, collections $\{\boldsymbol{Q}_S(\boldsymbol{v}, c), \boldsymbol{Q}_B(\boldsymbol{v}, c)\}$ and $\{\boldsymbol{M}_S(\boldsymbol{v}, c), \boldsymbol{M}_B(\boldsymbol{v}, c)\}$ for all possible $(\boldsymbol{v}, c)$. That is, $\boldsymbol{Q}_S : \mathcal{V}^{\mathcal{B}} \times [\underline{c}, \overline{c}] \to [0,1]^{\mathbb{N}_0}$, $\boldsymbol{Q}_B : \mathcal{V}^{\mathcal{B}} \times [\underline{c}, \overline{c}] \to [0,1]^{\mathcal{B}}$, $\boldsymbol{M}_S : \mathcal{V}^{\mathcal{B}} \times [\underline{c}, \overline{c}] \to \mathbb{R}^{\mathbb{N}_0}$, and $\boldsymbol{M}_B : \mathcal{V}^{\mathcal{B}} \times [\underline{c}, \overline{c}] \to \mathbb{R}^{\mathcal{B}}$, where, $\boldsymbol{Q} = (\boldsymbol{Q}_S, \boldsymbol{Q}_B)$ satisfies in addition the constraints imposed by feasibility.

A direct mechanism $\langle \boldsymbol{Q}, \boldsymbol{M}\rangle$ is said to satisfy interim individual rationality and (Bayesian) incentive compatibility if it satisfies these constraints for every possible type of every buyer who participates and for the seller in expectations at date 0. Observe that in principle this allows for the seller's individual rationality constraint to be violated in some periods $t > 0$. However, as we will see, with the fee-setting implementation of the optimal mechanism, such violations never occur.

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.

[19] Of course, this is a weak physical feasibility constraint expressed in terms of probabilities. However, as is well known, by the Birkhoff-von Neumann theorem, (8) is also sufficient to guarantee that feasibility can be satisfied ex post.

4.2 Optimal dynamic mechanisms

We first derive the optimal direct mechanism. We say that the good is auctioned off in period t with reserve p_t if $Q_b^t(\boldsymbol{v}^t, c) = 1$ if $v_b = \max_{b' \in \mathcal{B}_t} v_{b'}$ and $v_b \geq p_t$ and $Q_b^t(\boldsymbol{v}^t, c) = 0$ for all $b \in \mathcal{B}_t$ otherwise.[20]

Lemma 1. *A mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ is optimal only if the good is auctioned off in every period t at some reserve p_t.*

Lemma 1 implies that the choice set for allocation rules $\boldsymbol{Q}$ can be narrowed down to sequences of reserves $\boldsymbol{p}(c) := (p_t(c))_{t=0}^{\infty}$, one sequence for every seller type c, with the understanding that, provided the mechanism is still active in period t, the good will be sold to the buyer with the highest valuation present in that period, provided this valuation is no less than $p_t(c)$. The mechanism design problem thus boils down to determining these optimal reserves.

To make headway in this direction, the concept of the *ultimate probability of selling* for a seller who reports type c, which was introduced by Satterthwaite and Shneyerov (2008) and is defined as

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

proves useful. A further concept that simplifies the analysis is the *ultimate conditional expected revenue*. Since this concept is more technical, we defer its definition and discussion to the proof of Lemma 2. The seller's expected discounted payment is

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

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

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

[20] This definition neglects the possibility of ties at the highest value, which have probability 0. To account for such ties explicitly, one can arbitrarily set $Q_b^t(\boldsymbol{v}^t, c) = 1$ for a buyer b with the highest valuation, where, say, b has the highest index among all buyers with the highest value.

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

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

Standard arguments then imply that a direct mechanism satisfies the seller's incentive compatibility constraint if and only if $q_S(c)$ is monotone in c and

$$m_S(c) = q_S(c)c + \int_c^{\overline{c}} q_S(x)dx + \mathcal{W}_S(\overline{c}) \tag{9}$$

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 $\overline{c}$, that is, if $\mathcal{W}_S(\overline{c}) \geq 0$. Because $\mathcal{W}_S(\overline{c})$ enters the objective function as the constant $-\alpha\mathcal{W}_S(\overline{c})$, optimality requires $\mathcal{W}_S(\overline{c}) = 0$ for $\alpha > 0$. For $\alpha = 0$, we set $\mathcal{W}_S(\overline{c}) = 0$ without loss of generality.

As usual, reversing the order of integration in the double integral that results from taking expectations shows that the ex ante expected payment to the seller given the ultimate probability of selling q_S is $\mathbb{E}_c[m_S(c)] = \mathbb{E}_c[\Gamma(c)q_S(c)]$, while the seller's ex ante expected payoff is $\mathbb{E}_c[\mathcal{W}_S(c)] = \mathbb{E}_c[(\Gamma(c) - c)q_S(c)]$. Consequently, the objective in (7) can be rewritten as

$$\mathbb{E}_{\boldsymbol{v},c}\left[\sum_{t=0}^{\infty}\left(\sum_{b\in\mathcal{B}_t} M_b^t(\boldsymbol{v}_t, c)\right)\prod_{\tau=0}^{t-1}\delta(1-Q_S^{\tau}(\boldsymbol{v}_\tau, c))\right] - \mathbb{E}_c[(\Gamma(c) - (1-\alpha)(\Gamma(c)-c))q_S(c)].$$

Notice that $\mathbb{E}_c[(\Gamma(c) - (1-\alpha)(\Gamma(c)-c))q_S(c)] = \mathbb{E}_c[(\alpha\Gamma(c) + (1-\alpha)c)q_S(c)]$. Thus, the relevant object from the intermediary's perspective vis-à-vis the seller is the weighted virtual cost

$$\Gamma_\alpha(c) := \alpha\Gamma(c) + (1-\alpha)c,$$

which is the weighted average of the virtual cost $\Gamma(c)$ and the seller's cost c. Observe that monotonicity of $\Gamma(c)$ implies monotonicity of $\Gamma_\alpha(c)$. Thus, the mechanism design problem with regard to the seller is effectively the same as in the static setup, except that $\Gamma(c)$ is replaced by $\Gamma_\alpha(c)$. Note also that $\Gamma_\alpha(c)$ increases in α and satisfies $\Gamma(c) = \Gamma_1(c)$. Moreover, $\Gamma_\alpha(c)$ increasing in α implies that Γ_α^{-1} decreases in α.

With this in hand, solving the intermediary's mechanism design problem is effectively the same as solving for the optimal selling mechanism for a seller whose cost is $\Gamma_\alpha(c)$ rather than c. As it turns out, by appropriate transformations of distributions and the revenue function, the optimization problem in the dynamic model with random arrival of buyers can be represented and solved as the problem faced by a seller in a static setting with a single buyer.

Specifically, denote by

$$R(p) := \frac{\int_p^{\overline{v}} \Phi(v) dF_{(1)}(v)}{1 - F_{(1)}(p)}$$

the expected revenue conditional on selling when setting a reserve price p. Note that in the static problem with one buyer analyzed in Section 3, we have $R(p) = p$ because $F_{(1)} = F$ and $\frac{\int_p^{\overline{v}} \Phi(v) dF(v)}{1-F(p)} = p$. The optimal selling problem is then to

$$\max_p (R(p) - \Gamma_\alpha(c))(1 - F_\infty(p)), \tag{10}$$

where

$$1 - F_\infty(p) := (1 - F_{(1)}(p)) \left(\sum_{t=0}^{\infty} \delta^t F_{(1)}(p)^t \right) = \frac{1 - F_{(1)}(p)}{1 - \delta F_{(1)}(p)} \tag{11}$$

is the *ultimate probability of selling* when setting the price p in every period.[21] Thus, the optimal reserve in the dynamic setting is the same as the optimal reserve set in a static setting in which a single, albeit fictitious, buyer draws her value from the distribution F_∞ and generates an expected revenue conditional on trade given reserve p of $R(p)$.

The first-order condition for a maximum is

$$0 = f_\infty(p) \left[\Gamma_\alpha(c) - \left(R(p) - \frac{1 - \delta F_{(1)}(p)}{1 - \delta} (R(p) - \Phi(p)) \right) \right],$$

which uses the readily established facts that $R'(p) = \frac{f_{(1)}(p)}{1-F_{(1)}(p)} (R(p) - \Phi(p))$ and $\frac{f_{(1)}(p)}{1-F_{(1)}(p)} \frac{1-F_\infty(p)}{f_\infty(p)} = \frac{1-\delta F_{(1)}(p)}{1-\delta}$. We denote by

$$\tilde{\Phi}(p) := R(p) - \frac{1 - \delta F_{(1)}(p)}{1 - \delta} (R(p) - \Phi(p))$$

[21] This concept is inspired by Satterthwaite and Shneyerov (2007, 2008)'s notion of the "ultimate discounted probability of trade". 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.

the *dynamic virtual valuation*, whose derivative can be shown to satisfy $\tilde{\Phi}'(p) = \frac{1-\delta F_{(1)}(p)}{1-\delta}\Phi'(p) > 0$. Hence, $\tilde{\Phi}$ is invertible and the optimal reserve $p^*(c)$ thus satisfies

$$p^*(c) := \tilde{\Phi}^{-1}(\Gamma_\alpha(c)). \tag{12}$$

In other words, the good is sold in the earliest period in which $\tilde{\Phi}(v) \geq \Gamma_\alpha(c)$, where v is the highest valuation among buyers in that period. When the good is sold, it is sold to the buyer with the highest valuation.

Observe that $\tilde{\Phi}(\overline{v}) = \overline{v}$ and that $\tilde{\Phi}(v)$ decreases in δ for any $v \in [\underline{v}, \overline{v})$. At $\delta = 0$, we have $\tilde{\Phi}(v) = \Phi(v)$. Because $\Gamma_\alpha(c)$ increases in α, it follows that $p^*(c)$ increases in α and in δ. That is, the optimal reserve price increases as the intermediary's market power increases and when the future matters more (which is equivalent to less friction when trying to find a buyer).

We summarize the optimal allocation rule in the following lemma:

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*

$$\max_{b \in \mathcal{B}_t} \tilde{\Phi}(v_b) \geq \Gamma_\alpha(c),$$

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

Lemma 2 generalizes Theorem 3 of Myerson and Satterthwaite (1983) to our dynamic setting with multiple buyers using the concept of the dynamic virtual valuation. It is based on the insight that in any optimal mechanism, the good goes to the buyer with the highest value in any given period if this value is above some threshold, and stays with the seller otherwise. Lemma 2 adds to this the insight that the good goes to the buyer with the highest dynamic virtual valuation, appropriately defined, provided it exceeds $\Gamma_\alpha(c)$. Intuitively, Myerson and Satterthwaite (1983) find in a one-shot setup with one buyer and one seller and $\alpha = 1$ that the good is transferred whenever $\Phi(v)$ exceeds $\Gamma(c)$. In the language of optimal monopoly pricing, this means that the good changes hands whenever marginal revenue is larger than marginal cost (see footnote 18 for why virtual types can be interpreted as marginal revenue and marginal cost). In our setup, the dynamic virtual valuation $\tilde{\Phi}(v)$ has to be used to adjust for the option value of future trade and the weighted virtual cost $\Gamma_\alpha(c)$ to account for the weight on the seller's utility.

4.3 Implementation via optimal fees

We now show that the optimal mechanism can be implemented via an appropriately structured fee that is levied on the transaction price. The seller's choice variable in every period is to set a reserve price. If there are multiple buyers present in that period, they participate in a second-price auction with that reserve.

Lemma 2 derived the optimal allocation rule. This allocation rule can be implemented via fee setting if a seller of type c can be induced to set the reserve price $p^*(c)$ defined in (12) in every period in which he is active. If this is possible, then bidding in the second-price auction will ensure that the object goes to the buyer with the highest virtual valuation while the reserve price $p^*(c)$ ensures that trade only takes place if this virtual value, that is $\tilde{\Phi}$, exceeds $\Gamma_\alpha(c)$. The discounted probability that a seller of type c who always sets the price $p^*(c)$ ever sells is $1 - F_\infty(p^*(c))$. A seller with cost $\Gamma_\alpha^{-1}(\overline{v})$ should optimally set the price $\overline{v}$ and never trade. By a standard revenue equivalence argument (see e.g. Myerson, 1981), the expected discounted payoff $V(c)$ of a seller of type $c \in [\underline{c}, \Gamma_\alpha^{-1}(\overline{v})]$ who always sets the price $p^*(c)$ is pinned down by the allocation rule and given by

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

With this in hand, we can now describe the optimal transaction fees for the dynamic problem, which we denote $\omega_{\alpha,\delta}(p)$.

Proposition 5. *A fee-setting mechanism that in every period $t = 0, 1 \ldots$ in which it is active uses the fee*

$$\omega_{\alpha,\delta}(p) := p - \frac{\int_p^{\overline{v}} \left[\Gamma_\alpha^{-1}(\tilde{\Phi}(v)) + \delta V(\Gamma_\alpha^{-1}(\tilde{\Phi}(v)))\right] f(v)dv}{1 - F(p)} \tag{13}$$

defined on the equilibrium reserve-price range $[\tilde{\Phi}^{-1}(\underline{c}), \overline{v}]$, implements the optimal mechanism, whose allocation rule is described in Lemma 2.

Note that for $\alpha = 1$ and $\delta = 0$, the fee $\omega_{\alpha,\delta}(p)$ is the same as $\omega_I(p)$ derived in Section 3, that is, $\omega_{1,0}(p) = \omega_I(p)$ holds. This also means that the number of buyers does not matter for the optimal static fee $\omega_I(p)$, which reflects the well-known fact that the optimal reserve in a standard first- or second-price auction is independent of the number of buyers and equal to the optimal take-or-leave-it offer to a single buyer. However, in the dynamic setting, the arrival process of buyers affects, in general, the optimal fee because $\tilde{\Phi}$ depends on $F_{(1)}$. Notice also that a lower α means an overall lower fee, because less weight is put on the intermediary's revenue; that is, as α decreases, the integrand increases because Γ_α^{-1} increases and thus the fee decreases. In contrast, the effects of increasing δ cannot be signed in general because $\tilde{\Phi}$ decreases while δV increases in δ. As noted, $p^*(c)$ increases in α and in δ, but reflecting the properties of the static setting with respect to the elasticity of demand and the price-endogeneity effect, the effects of δ on the fee can go either way.

Further, the optimal fee is linear and independent of δ if G is a reflected generalized Pareto distribution. In the dynamic setting, however, the linearity of this fee is not immediately apparent from the general expression for the optimal fee. The reason is that a non-linear fee implies payoffs for a seller *with a fee* that are equivalent to those of a non-risk-neutral seller *without a fee* and utility function $p - \omega_{\alpha,\delta}(p)$, so that the seller takes into account uncertainty and intertemporal trade-offs. (This equivalence is purely mathematical; the seller, of course, remains risk neutral.) However, if the fee is linear, a seller with a fee is equivalent to a risk-neutral seller without a fee. Formally, a seller with cost c facing a linear fee $\omega(p) = a + bp$ solves the maximization problem

$$\max_p((1-b)R(p) - a - c)(1 - F_\infty(p)),$$

mirroring that faced by the intermediary in (10). The optimal price set by the seller satisfies

$$\tilde{\Phi}(p) = \frac{c+a}{1-b}.$$

Thus, choosing a and b such that $\frac{c+a}{1-b} = \Gamma_\alpha(c)$ makes this linear fee optimal.[22]

[22] The seller's problem cannot be collapsed into an equivalent static optimization when the fee is not linear because then $\omega(R(p))$ is not equal to the expected value of transaction fees paid, conditional on trade given the reserve price p.

Summarizing all of this, we have the following corollary to Proposition 5:

Corollary 1. *The optimal fee $\omega_{\alpha,\delta}(p)$ increases in α. It is linear if the seller's distribution is a reflected generalized Pareto distribution, in which case it is pinned down by α and the parameters of the seller's distribution and thus independent of δ, F and the arrival process.*

As an illustration, reconsider the static setup by setting $\delta = 0$, which has been studied extensively in the literature. For any fixed number of buyers $\overline{B}$ (i.e. $\pi_{\overline{B}} = 1$ for some positive integer $\overline{B}$) and any distribution F satisfying regularity, our fee-setting mechanism specializes to the optimal auction with reserve $p^*(c) = \Phi^{-1}(c)$ of Myerson (1981) if all the weight is on the seller's welfare ($\alpha = 0$).

Thin dynamic markets We now return to the thin market setting for types and distributions introduced in Subsection 2.3 and show that, just like in the static setting, the optimal fees become linear as the overlap in supports between the buyers' and the sellers' distributions vanishes. That simple, linear fees are optimal in the dynamic setting is remarkable given that, in general, the structure of the optimal fee $\omega_{\alpha,\delta}$ is substantially more involved than that of the optimal fee ω_I in the static setup. The following proposition shows that as the market becomes thin, asymptotically the optimal fee becomes linear.

Proposition 6. *Let $(\underline{c}_j, u_j^S, \overline{v}_j, u_j^B)$ be an arbitrary sequence, with u_j^S and u_j^B going to zero as $j \to \infty$, satisfying $\underline{c}_j < \overline{v}_j$ for all j. Then:*

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

(ii) the normalized fee $\tilde{\omega}_j(\tilde{p})$ converges to a linear fee, that is:

$$\lim_{j\to\infty} \tilde{\omega}_j(\tilde{p}) = \frac{\alpha}{\alpha+\sigma}\tilde{p}. \tag{14}$$

As before, if $\underline{c}_j$ converges to a well-defined limit $\underline{c}$, then the fee converges to $\omega(p) = [\alpha/(\alpha+\sigma)]p - [\alpha/(\alpha+\sigma)]\underline{c}$. For $\alpha = 1$, the limit fee is the same as in the static one-buyer setting in Section 3. The fee decreases as α decreases and becomes 0 for $\alpha = 0$. This shows that the results on the convergence to linear fees are fairly robust and generalize beyond the one-buyer one-period setting.

5 Discussion

In this section, we provide a brief discussion of extensions, robustness, and empirical implications.

Non-stationarity Let us first briefly explain how our analysis can be extended to account for non-stationary environments at what is essentially a cost in notation. Assuming that the sequences of time varying discount factors δ_t, distributions F_t and random arrival processes π_B^t are known, one can proceed in analogy to the way we proceeded under the assumption of stationarity. Although the optimal transaction fee ω_t in period t will in general vary over time because of non-stationarity, a simple argument based on what we call "expectational fees" (which are defined and derived in Lemma 3 in Appendix B) shows that, in the limit as markets become thinner, the optimal (normalized) transaction fees will be linear and stationary in the limit, too. The limit results also hold in richer dynamic environments, including stochastic changes in buyers' valuation distributions and deadlines after which the seller exits the market for sure. Appendix C contains more details. The stationarity of the optimal limiting fees is particularly remarkable because the optimal reserve price path chosen by the seller will in general be non-stationary.

Non-durable and durable goods The general, dynamic model can be seen both as a model of non-durable goods (which provide a one-off consumption utility when consumed) and of durable goods (which provide a consumption utility in each period they are consumed). A durable goods setting translates in a straightforward way into our setting with valuations (or opportunity costs of selling) being the net present value of consumption utilities.[23]

[23] To see this, consider the non-durable good setup in our model, in which a seller of type c has a payoff of 0 if he never trades and an expected discounted payoff of $\delta^T(p-c)$ if he trades at price p in period T. Alternatively, we could assume that the good is durable, i.e., the good gives the seller a utility of $\tilde{c}$ in

Intermediary's opportunity costs per transaction In many markets, transactions are not necessarily costless for the intermediary, either because they involve a direct resource cost or because they compete with goods that the designer could sell directly to buyers. We denote this opportunity cost per transaction by $\kappa \geq 0$ and revisit, for simplicity, the static bilateral-trade setting analyzed in Section 3, in which the intermediary maximizes its expected profit. If $\kappa \geq \overline{v} - \underline{c}$, the optimal mechanism never induces trade; hence, suppose that $\kappa < \overline{v} - \underline{c}$. A straightforward extension of Myerson and Satterthwaite (1983) implies that the optimal mechanism trades if and only if $\Phi(v) \geq \Gamma(c) + \kappa$. Similarly, a straightforward generalization of Proposition 1 shows that this mechanism is implemented by $\omega_I(p,\kappa) := p - \mathbb{E}_v[\Gamma^{-1}(\Phi(v) - \kappa) \mid v \geq p]$, defined on $[\Phi^{-1}(\underline{c} + \kappa), \overline{v}]$. This fee is increasing in κ. If $G(c)$ is of the form in (4), then $\omega_I(p,\kappa) = p - [\underline{c} + \sigma(p-\kappa)]/(1+\sigma)$. Thus, the optimal fee is linear in p and κ, independent of F, and satisfies $\partial\omega_I(p,\kappa)/\partial\kappa = \sigma/(1+\sigma)$.[24]

Linear fees, first-price auctions, and the informed principal problem Given linear transaction fees ω, the payoff of the seller upon selling at some price $\breve{p}$ is linear in the transaction price. Consequently, linear fees correspond to the case where the seller is a risk-neutral agent with a linear von Neumann-Morgenstern utility function. As is well known, with risk-neutral agents the revenue equivalence theorem applies.[25] This implies that, given linear fees, using a first-price auction in which the seller sets the reserve is equivalent to the second-price or the strategically equivalent English auction we have worked with. Moreover, due to the results for the informed principal problem in linear environments with independent private values of Mylovanov and Tröger (2014), keeping fixed the linear fee the expected payoffs conditional on type would be unchanged if the seller could choose the trading mechanism after having learned his type (in any strongly

every period in which he consumes it. Under this assumption, if he never sells, his expected discounted utility is $\tilde{c}/(1-\delta)$. In contrast, if he sells at price p in period T, his expected discounted payoff is $\delta^T p + \tilde{c}(1-\delta^T)/(1-\delta)$. Taking the difference, one obtains $\delta^T(p - \tilde{c}/(1-\delta))$. Thus, for $c = \tilde{c}/(1-\delta)$ the two models are isomorphic.

[24]To see this, substitute the proposed fee into the seller's pricing problem. Every seller type $c \leq \Gamma^{-1}(\overline{v} - \kappa)$ optimally sets $p = \Phi^{-1}(\Gamma(c) + \kappa)$, while seller types above this cutoff do not trade.

[25]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).

neologism-proof perfect Bayesian equilibrium). Thus, under linear fees, the broker could even delegate the choice of sale mechanism to the seller. This has the important implication that the intermediary could generate his revenues in a decentralized way: he simply sets the linear fees and can leave the choice of the details of the bargaining protocol to the seller.

Empirical implications Our theoretical results have empirical implications that are consistent with observations in several markets. A first implication is that in a static model with an English auction (which is a reasonable approximation of auction houses, such as Sotheby's and Christie's, and online auction sites, such as eBay), the fee is independent of the number of bidders in the auction. This result is similar in spirit to the aforementioned result that the profit-maximizing reserve price in a first- or second-price auction with private values is independent of the number of bidders. Moreover, it is consistent with the observation that the fees of auction houses and platforms do not depend on the number of bidders.

Second, if the optimal fees are linear, then the fee should be independent of the number of buyers and the distribution of their valuations even in a non-stationary setting. Empirically, one observes that indeed the fees of real-estate agents and of auction houses and platforms exhibit little variation over time and across regions (see Hsieh and Moretti (2003) who report the surprising invariance of real estate brokerage fees in the US).

For the specific example of real estate brokerage fees, a number of additional observations are consistent with the predictions of our theory. First, sellers with a higher loan-to-value ratio ask higher prices (Genesove and Mayer, 1997). Second, sellers who had bought their houses when average real estate prices were high, ask for higher prices than those who had bought when prices were low (Genesove and Mayer, 2001). Third, quality-adjusted prices and time on market of houses are positively correlated in cross-sectional data and negatively correlated in longitudinal data. Fourth, direct sellers – that is, private owners who sell their houses without a broker or brokers who sell their own houses – sell at higher prices than private owners who sell through a broker (see Rutherford et al., 2005; Levitt and Syverson, 2008; Hendel et al., 2009). All these observations are either predicted by or at least consistent with our theory.[26]

6 Conclusions

We provide a parsimonious theory of optimal transaction fees and show that the optimal fees converge to linear fees as markets become increasingly thin. While the main purpose of this article is to develop a general model of transaction fees as optimal pricing, a positive side effect of having such a theory is that it resolves many of the puzzling observations documented in the empirical literature on real-estate brokerage fees by providing an alternative to the principal-agent view. Moreover, if linear fees are optimal, then up to the usual truncation the fee pins down the seller's distribution and the posted prices pin down the buyers' distribution.

The mechanism design approach is a powerful tool for monopoly pricing problems. By leveraging the revelation principle and focusing on direct mechanisms, it permits deriving optimal mechanisms without any contractual restrictions. The downside to this approach is that direct mechanisms are typically not used in practice, which makes it difficult to relate optimal mechanisms to observables. The present paper takes a step toward bridging the gap by connecting practically used mechanisms — transaction fees and their functional form — to theoretically optimal mechanisms. This suggests a possible productive path forward of first formulating and solving the mechanism design problem and then looking for practical mechanisms that implement that solution. Rather than imposing contractual restrictions, researchers embarking on that path might then consider restricting the family of type distributions to those that are consistent with optimality and the observed mechanisms. As a case in point, if linear fees are optimal, then that restricts the seller's distribution to be a reflected generalized Pareto distribution.

[26]The first and the second observation can be explained by the seller's opportunity cost of selling being correlated with the loan-to-value ratio and the seller's valuation at the previous transaction, respectively. For the third observation, the cross-sectional prediction is simply due to higher prices leading to a longer time on market. For the longitudinal prediction, observe that in a static model, the seller's reserve price is independent of the number of bidders arriving, which naturally generates high prices and short times on market when there are many buyers. It generates the opposite if there are few buyers. We show in Proposition 4 of our companion paper (Loertscher and Niedermayer (2026)) that the fourth stylized fact is consistent with our theory.

References

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

ANDERSON, C. (2006): *Long Tail: Why the Future of Business Is Selling Less of More*, New York: Hyperion.

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

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

BALIGA, S. AND R. VOHRA (2003): "Market Research and Market Design," *Advances in Theoretical Economics*, 3, Article 5.

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

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

BAYE, M. R., X. GAO, AND J. MORGAN (2011): "On the Optimality of Clickthrough Fees in Online Markets," *Economic Journal*, 121, F340–F367.

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

BÖRGERS, T. (2015): *An Introduction to the Theory of Mechanism Design*, New York: Oxford University Press.

BRYNJOLFSSON, E., Y. HU, AND M. D. SMITH (2003): "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49, 1580–1596.

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

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

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

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

Chatterjee, K. and W. Samuelson (1983): "Bargaining under Asymmetric Information," *Operations Research*, 31, 835–851.

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

De los Santos, B., D. P. O'Brien, and M. R. Wildenbeest (2026): "Agency Pricing and Bargaining: Evidence from the E-Book Market," *American Economic Journal: Microeconomics*, 18, 146–91.

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

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

Garrett, D. F. (2016): "Intertemporal price discrimination: Dynamic arrivals and changing values," *American Economic Review*, 106, 3275–3299.

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

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

Gershkov, A., J. K. Goeree, A. Kushnir, B. Moldovanu, and X. Shi (2013): "On the Equivalence of Bayesian and Dominant Strategy Implementation," *Econometrica*, 81, 197–220.

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

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

Hendel, I., A. Nevo, and F. Ortalo-Magné (2009): "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com," *American Economic Review*, 99, 1878 – 98.

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

Johnson, J. P. (2017): "The agency model and MFN clauses," *Review of Economic Studies*, 84, 1151–1185.

Jullien, B. and T. Mariotti (2006): "Auction and the Informed Seller Problem," *Games and Economic Behavior*, 225–258.

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

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

Loertscher, S. and E. Muir (2025): "The matching benefits of market thickness," *Games and Economic Behavior*, 153, 42–66.

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

——— (2020): "Entry-Deterring Agency," *Games and Economic Behavior*, 119, 172–188.

——— (2026): "Percentage Fees in Thin Markets: An Empirical Perspective," Working Paper.

Loertscher, S. and C. Wasser (2019): "Optimal Structure and Dissolution of Partnerships," *Theoretical Economics*, 14, 1063–1114.

Lu, H. and J. Robert (2001): "Optimal Trading Mechanisms with Ex Ante Unidentified Traders," *Journal of Economic Theory*, 97, 50–80.

Luo, Y., I. Perrigne, and Q. Vuong (2018): "Structural analysis of nonlinear pricing," *Journal of Political Economy*, 126, 2523–2568.

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

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

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

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

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

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

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

Rutherford, R., T. Springer, and A. Yavas (2005): "Conflicts between principals and agents: evidence from residential brokerage," *Journal of Financial Economics*, 76, 627–665.

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

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

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

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

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

Waldfogel, J. (2017): "How Digitization Has Created a Golden Age of Music, Movies, Books, and Television," *Journal of Economic Perspectives*, 31, 195–214.

Wang, Z. and J. Wright (2017): "Ad-valorem platform fees and efficient price discrimination," *RAND Journal of Economics*, 48, 467–484.

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

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

Appendix

A Details of identification with linear fees

First, observe that the cost distribution of sellers whose costs are so high that they never sell does not affect the fees. So we can only make statements about sellers who sell with positive probability. Denote the truncation threshold above which sellers do not participate (because they sell with probability zero) as c_{trunc}. Given that the intermediary's optimal allocation rule is $\Phi(v) \geq \Gamma(c)$, sellers with $\Gamma(c) > \Phi(\overline{v})$ sell with probability 0, so that $c_{\text{trunc}} = \Gamma^{-1}(\Phi(\overline{v})) = \Gamma^{-1}(\overline{v})$. Denote the truncated distribution of sellers' costs as $G_{\text{trunc}}(c) := G(c)/G(c_{\text{trunc}})$. In the following, we show that observing a linear fee function $\omega_L(p)$ uniquely identifies G_{trunc} up to the truncation threshold c_{trunc}.

A related question of interest is whether the distribution F is identifiable given the distribution F_p of prices $\Phi^{-1}(\Gamma(c))$ posted by the seller, i.e. whether we know F if we hypothetically observe an infinite sample size of prices. We show that we can identify c_{trunc} and the distribution F up to a truncation at the bottom, that is, we can identify the distribution of the valuations of buyers who buy with positive probability. Formally, the truncated distribution is $1 - F_{\text{trunc}}(v) := (1 - F(v))/(1 - F(v_{\text{trunc}}))$, where v_{trunc} is the truncation point.

Proposition 7. *Assume the intermediary's fee is optimal and that the seller's pricing is optimal.*

(i) The optimal fee is linear if and only if the seller's truncated cost distribution G_{trunc} is a reflected generalized Pareto distribution as defined in (4). A fee $\omega_L(p) := a+bp$, with $b \in (0,1)$, identifies G_{trunc} as defined in (4) with $\underline{c} = -a/b$, $\overline{c} = c_{trunc}$, and $\sigma = (1-b)/b$.

(ii) If we additionally observe the distribution of prices F_p, then the seller's truncation point c_{trunc} is identified as $c_{trunc} = (1-b)F_p^{-1}(1) - a$ and the buyer's truncated distribution F_{trunc} is identified as

$$F_{trunc}(p) = 1 - \exp\left(-\int_{v_{trunc}}^{p} \frac{dx}{x - \gamma(x)}\right) \tag{15}$$

where

$$\gamma(x) = \frac{(a + bc_{trunc})F_p(x)^{\frac{b}{1-b}} - a(1-b)}{b(1-b)} \tag{16}$$

and $v_{trunc} = F_p^{-1}(0)$.

Proof. Let $\Gamma_{\text{trunc}}(c) := c + G_{\text{trunc}}(c)/g_{\text{trunc}}(c)$ and $\Phi_{\text{trunc}}(v) := v - (1 - F_{\text{trunc}}(v))/f_{\text{trunc}}(v)$. For the truncations considered here, $\Gamma_{\text{trunc}} = \Gamma$ and $\Phi_{\text{trunc}} = \Phi$ on the relevant supports.

Part (i): We have already shown that $G_{\text{trunc}}(c)$ as in (4) implies a linear virtual cost function and hence a linear fee. We are thus left to show that linear fees imply $G_{\text{trunc}}(c)$ as in (4). Plugging in ω_L in the seller's objective function yields $((1-b)p-a-c)(1-F(p)) = (1-b)\left(p-(c+a)/(1-b)\right)(1-F(p))$. The seller's first-order condition can be rearranged to $\Phi(p) = (c+a)/(1-b)$. Implementing the optimal allocation rule requires $\Phi(p) = \Gamma(c)$. Thus, the linear fee is optimal if and only if

$$\frac{c+a}{1-b} = \Gamma(c),$$

that is, if and only if $\Gamma(c)$ is linear.

Writing Γ in terms of G_{trunc} and g_{trunc} in the above equation yields a linear differential equation in G_{trunc}: $g_{\text{trunc}}(c) = G_{\text{trunc}}(c)(1-b)/(a+bc)$. Combining this differential equation with the initial condition $G_{\text{trunc}}(c_{\text{trunc}}) = 1$, we get an initial value problem that has a unique solution by the Picard-Lindelöf Theorem. The solution is the distribution in (4) with $\underline{c} = -a/b$, $\overline{c} = c_{\text{trunc}}$, and $\sigma = (1-b)/b$, which can be easily verified by plugging in the solution into the initial value problem.

Part (ii): Assume that we observe the price distribution F_p. c_{trunc} is such that a seller of this type sets the price $\overline{v}$ and is just indifferent between selling and not selling, i.e. $c_{\text{trunc}} = \overline{v} - \omega(\overline{v})$. Substituting in $\omega(p) = a + bp$ and $\overline{v} = F_p^{-1}(1)$ yields the expression for c_{trunc} in the Proposition. v_{trunc} is the lowest price set in equilibrium, hence $v_{\text{trunc}} = F_p^{-1}(0)$.

F_{trunc} can be obtained in the following way. A seller of type c sets the price $p = \Phi^{-1}(\Gamma(c))$, therefore, $c = \Gamma^{-1}(\Phi(p))$ holds for the seller's cost $c \sim G$ and hence the price distribution F_p satisfies

$$F_p(p) = G(\Gamma^{-1}(\Phi(p)))$$

for all $p \in [v_{\text{trunc}}, \overline{v}]$. Applying inverse functions, the above equation can be rewritten as $\gamma(p) = \Phi(p)$, where $\gamma(p) := \Gamma(G^{-1}(F_p(p)))$. Writing Φ in terms of F_{trunc} and f_{trunc}

and rearranging yields the differential equation $f_{\text{trunc}}(p) = (1 - F_{\text{trunc}}(p))/(p - \gamma(p))$. Together with the initial condition $F_{\text{trunc}}(v_{\text{trunc}}) = 0$, we get an initial value problem with the unique solution provided in (15), which can be verified by plugging in the expression in (15) into the initial value problem. The expression for γ in (16) can be obtained by plugging in the previously obtained expression for Γ and the G_{trunc}^{-1} implied by the previously obtained G_{trunc} and rearranging. □

That F is identified by the price distribution is surprising given that in general, the price distribution $F_p(p) = G(\Gamma^{-1}(\Phi(p)))$ depends on both the buyer's and the seller's type distributions.

Note that the non-identifiability of the distribution of buyer types who trade with probability zero (i.e. identification up to a truncation) is standard in the literature on identification in non-linear pricing; see Luo et al. (2018). In our case, this extends to sellers who trade with probability zero.

B Proofs

Proof of Proposition 1. The proof is in the main text. □

Proof of Proposition 2. See the proof of Proposition 6, which is a generalization of this Proposition. □

Proof of Proposition 3. The Taylor expansion of $\Gamma^{-1}(x)$ around $\overline{v}$ is

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

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

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

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

where the second equality stems from the Taylor expansion, the fourth from reversing a Taylor expansion, and the fifth from the fact that $\mathbb{E}_v[\Phi(v)|v \geq p] = p$. 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 $\mathrm{Var}[\varphi - \overline{v}|v \geq p] = \mathbb{E}_v[(\varphi - \overline{v})^2 | v \geq p] - (\mathbb{E}_v[\varphi|v \geq p] - \overline{v})^2$. This completes the proof. □

Proof of Proposition 4. Let $h(v) := f(v)/(1 - F(v)) = \eta_d(v)/v$, and let $\rho_p(y) := (1 - F(y))/(1 - F(p))$. First note that

$$\mathbb{E}_{v \sim F} \left[(\Phi(v) - \overline{v})^2 \mid v \geq p \right] = (\bar{v} - p)^2 + \int_p^{\bar{v}} \frac{\rho_p(y)}{h(y)} \, dy.$$

This equality can be verified by multiplying both sides by $1 - F(p)$, taking derivatives with respect to p, and checking that both sides are zero at $p = \bar{v}$. Since $\mathbb{E}_{v \sim F}[\Phi(v) \mid v \geq p] = p$, it follows that $\left(E_{v \sim F}[\Phi(v) - \overline{v} \mid v \geq p]\right)^2 = (p - \overline{v})^2$ and

$$\mathrm{Var}_{v \sim F} \left[\Phi(v) - \bar{v} \mid v \geq p \right] = \int_p^{\bar{v}} \frac{\rho_p(y)}{h(y)} \, dy.$$

Now consider two distributions F and $\widehat{F}$, with associated hazard rates h and $\widehat{h}$, such that $\widehat{\eta}_d(v) > \eta_d(v)$ for all relevant v. Equivalently, $\widehat{h}(v) > h(v)$. Since $\rho_p(y) = \exp(-\int_p^y h(x) \, dx)$, we have $\widehat{\rho}_p(y) \leq \rho_p(y)$. Hence $\widehat{\rho}_p(y)/\widehat{h}(y) \leq \rho_p(y)/h(y)$ for all $y \geq p$, and therefore

$$\mathrm{Var}_{v \sim \widehat{F}} \left[\widehat{\Phi}(v) - \bar{v} \mid v \geq p \right] \leq \mathrm{Var}_{v \sim F} \left[\Phi(v) - \bar{v} \mid v \geq p \right].$$

If Γ^{-1} is quadratic with second derivative γ_2, Proposition 3 gives $\omega(p) = p - \Gamma^{-1}(p) - \frac{\gamma_2}{2}\operatorname{Var}_{v\sim F}\left[\Phi(v) - \bar{v} \mid v \geq p\right]$. Thus an overall increase in demand elasticity raises the fee when $\gamma_2 > 0$, lowers the fee when $\gamma_2 < 0$, and has no effect when $\gamma_2 = 0$.

□

Proof of Lemma 1. Suppose to the contrary that the optimal mechanism, denoted $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$, does not auction off the good at some reserve in period t and in some states $(\boldsymbol{v}_t, c)$. This implies that with positive probability the good is sold in period t to a buyer whose valuation is not the highest amongst all the buyers present. Consider then an alternative mechanism that coincides with $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ except for the states in period t in which $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ does not auction off the good. Let the alternative mechanism sell the good to the highest value buyer in all those instances for which $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ sells it to some other buyer. This alternative mechanism will increase the broker's payoff $\mathcal{W}_I(c)$ by increasing the revenue it raises while leaving the seller's payoff $\mathcal{W}_S(c)$ unaffected. Therefore, the mechanism $\langle \hat{\boldsymbol{Q}}, \hat{\boldsymbol{M}} \rangle$ cannot be optimal. □

Proof of Lemma 2. Letting

$$k_t := R(p_t) := \frac{p_t\left(F_{(2)}(p_t) - F_{(1)}(p_t)\right) + \int_{p_t}^{\overline{v}} v dF_{(2)}(v)}{1 - F_{(1)}(p_t)}$$

denote the expected transaction price in period t, conditional on a transaction occurring in period t, choosing a sequence of reserves $\boldsymbol{p}(c)$ is equivalent to choosing a sequence $\boldsymbol{k}(c) := (k_t(c))_{t=0}^{\infty}$ of expected transaction prices (conditional on a transaction occurring). Note that $R(p_t)$ is increasing, because it is a conditional expectation, and therefore invertible.

For any sequence of expected transaction prices $\boldsymbol{k} = (k_t)_{t=0}^{\infty}$, let

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

denote the discounted probability of a sale occurring in period t. To get an intuition for discounted probabilities, assume for the moment that discounting stems from a survival probability, that is, $1 - \delta$ is the probability of the seller dying (or dropping out of the market). Then $\delta F_{(1)}(R^{-1}(k_\tau))$ is the probability that the seller survives and does not sell in period τ, and $q_t(\boldsymbol{k})$ is the probability that the seller survived and did not sell until period t and ends up selling then. If discounting does not purely stem from survival (but from impatience or a combination of survival and impatience), then the intuition is that a sale that happens later is less valuable to the seller and hence discounted. The number $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t$ is then the expected discounted transaction price, or average price for short, given $\boldsymbol{k}$ while $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})$ is the ultimate probability of selling. The number

$$k := \frac{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t}{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})} \tag{17}$$

simplifies the following analysis and has the interpretation of an *ultimate conditional expected revenue*. Again, an intuition for k can be gained by interpreting δ as a survival probability. With this interpretation, k is the expected revenue conditional on trading and not dropping out.[27]

A mechanism can only be optimal if it maximizes the *ultimate conditional expected revenue* for a given ultimate probability of selling, because $\boldsymbol{k}$ enters both $\mathcal{W}_S(c)$ and $\mathcal{W}_I(c)$ only through $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t$ and $\sum_{t=0}^{\infty} q_t(\boldsymbol{k})$. Equivalently, it must maximize the ultimate probability of selling for a given ultimate conditional expected revenue. For $k \in [\underline{v}, \overline{v}]$, define this frontier by

$$1 - \overline{F}(k) := \sup_{\boldsymbol{k}} \left\{ \sum_{t=0}^{\infty} q_t(\boldsymbol{k}) : \frac{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})k_t}{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})} = k \right\}.$$

We next show directly that this frontier is attained by a constant reserve path. For a scalar z, consider the stationary infinite-horizon problem

$$W(z) := \sup_{\boldsymbol{k}} \sum_{t=0}^{\infty} q_t(\boldsymbol{k})(k_t - z).$$

After a failure to sell, the problem returns to the same state. Its Bellman equation is therefore

$$W(z) = \max_{p} \left\{ (1 - F_{(1)}(p))(R(p) - z) + \delta F_{(1)}(p) W(z) \right\}.$$

[27]If impatience stems from a time preference rather than a drop-out probability, k cannot be simply interpreted as a conditional expectation, but should be viewed as an abstract mathematical concept that helps to unite probabilities and discounting and simplifies calculations. One of the advantages of using the ultimate conditional expected revenue is that it avoids the problem of a standard conditional expected revenue with a time preference interpretation of discounting: for any positive constant per period probability of sale, the seller eventually trades with probability 1, so that conditioning on trade occurring would not be a useful concept.

A stationary maximizing reserve exists, and its value can equivalently be found from

$$W(z) = \max_p (R(p) - z) \frac{1 - F_{(1)}(p)}{1 - \delta F_{(1)}(p)} = \max_p (R(p) - z)(1 - F_\infty(p)).$$

The derivative of the last objective has the sign of $z - \tilde{\Phi}(p)$. Since $\tilde{\Phi}$ is increasing, the unique maximizer satisfies $\tilde{\Phi}(p) = z$.

Now fix k and set $p = R^{-1}(k)$ and $z = \tilde{\Phi}(p)$. The constant path $p_t = p$ is optimal in the Bellman problem. For every path with ultimate conditional expected revenue k, the Bellman objective equals $(k - z) \sum_{t=0}^{\infty} q_t(\boldsymbol{k})$. Since $k - z > 0$ for $p < \overline{v}$, no such path can have a larger ultimate probability of selling than the constant path; here $k - z = [1 - \delta F_{(1)}(p)](R(p) - \Phi(p))/(1 - \delta) > 0$. The endpoint follows by continuity. Consequently,

$$1 - \overline{F}(k) = \frac{1 - F_{(1)}(R^{-1}(k))}{1 - \delta F_{(1)}(R^{-1}(k))} = 1 - F_\infty(R^{-1}(k)).$$

Denote by $\overline{f}$ the corresponding density and by $\overline{\Phi}$ the corresponding virtual valuation. Thus, for a given c and k, we now have $\mathcal{W}_I(c) = k(c)(1 - \overline{F}(k(c))) - m_S(c)$ and $\mathcal{W}_S(c) = m_S(c) - q_S(c)c$. Using incentive compatibility (9) and $\mathcal{W}_S(\overline{c}) = 0$ by individual rationality, the objective given c becomes

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

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

$$\max_{k(c)} \int_{\underline{c}}^{\overline{c}} [k(c) - \Gamma_\alpha(c)] (1 - \overline{F}(k(c)))g(c)dc. \tag{18}$$

The integral can be maximized pointwise by choosing k such that

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

which is equivalent to $k(c) = \overline{\Phi}^{-1}(\Gamma_\alpha(c))$. This is a monotone function and thus consistent with incentive compatibility. Moreover, the second-order condition for a maximum is satisfied whenever the first-order condition is satisfied if $\overline{\Phi}(v)$ is monotone. This means that under the optimal allocation rule trade takes place as soon as

$$\overline{\Phi}(k) \geq \Gamma_\alpha(c). \tag{19}$$

Let $k^*(c) := \overline{\Phi}^{-1}(\Gamma_\alpha(c))$. Since $\overline{F}(k) = F_\infty(R^{-1}(k))$, $\overline{\Phi}(k) = \tilde{\Phi}(R^{-1}(k))$. According to (19) trade should take place as soon as $v \geq R^{-1}(k^*(c))$, which because of the afore-noted equalities and the monotonicity of $\tilde{\Phi}$, is equivalent to $\tilde{\Phi}(v) \geq \Gamma_\alpha(c)$ as claimed in the lemma. □

Proof of Proposition 5. The proof that this fee induces the seller of type c to set the price $p^*(c)$ in every period essentially follows from taking the first-order condition, assuming that the seller sets $p^*(c)$ in every future period. By the one-period-deviation principle, we can confine attention to a deviation by the seller in the present period to some reserve price p and assume that the seller sets the price $p^*(c)$ in every period after that, whereby he gets $\delta V(c)$. For a deviation within the displayed domain, the expected payoff given the fee $\omega_{\alpha,\delta}(p)$ defined in (13) is $(R_{\omega_{\alpha,\delta}}(p) - c)(1 - F_{(1)}(p))$ in the period of deviation and $F_{(1)}(p)\delta V(c)$ afterwards, where

$$R_{\omega_{\alpha,\delta}}(p) := \frac{(p - \omega_{\alpha,\delta}(p))(F_{(2)}(p) - F_{(1)}(p)) + \int_p^{\overline{v}} (y - \omega_{\alpha,\delta}(y)) dF_{(2)}(y)}{1 - F_{(1)}(p)}.$$

The expected payoff can be rearranged to

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

The first-order condition for a maximum is

$$0 = f_{(1)}(p) \left[-\Gamma_\alpha^{-1}(\tilde{\Phi}(p)) - \delta V(\Gamma_\alpha^{-1}(\tilde{\Phi}(p))) + c + \delta V(c) \right], \tag{20}$$

which follows after substituting $\omega_{\alpha,\delta}$ from (13) and cancelling terms (in particular, using the fact that $F_{(2)}(p) - F_{(1)}(p) = f_{(1)}(p)(1 - F(p))/f(p))$. The first-order condition is satisfied at $p = p^*(c) = \tilde{\Phi}^{-1}(\Gamma_\alpha(c))$. Moreover, the term in brackets in (20) decreases in p; therefore, the objective function is quasi-concave, implying that the first-order condition is sufficient for a maximum. □

Proof of Proposition 6. The proof of part (i) relies on Extreme Value Theory, which we summarize in Appendix D. Φ continuously differentiable implies that, for some constant $\overline{\beta}$,

$$\lim_{v \to \overline{v}} \frac{d}{dv} \left[\frac{1 - F(v)}{f(v)} \right] = \lim_{v \to \overline{v}} \frac{d}{dv} [v - \Phi(v)] = \overline{\beta}. \tag{21}$$

Equation (21) is the von Mises condition as stated in Theorem 2 in Appendix D. By Theorem 2, this implies that F is in the domain of attraction of an extreme value distribution (see Definition 1). By the Pickands-Balkema-de Haan Theorem (see Theorem 1), this in turn implies that F has a generalized Pareto upper tail as defined in (26). This implies uniform convergence of the normalized distribution $\tilde{F}_j$ to $\tilde{F}^*(\tilde{v}) = 1-(1-\tilde{v})^{\beta}$, because of the definition of the normalized variable $\tilde{v}$. Analogous reasoning applies for the convergence of $\tilde{G}_j$.

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

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

The following two constructs are used in the remainder of the proof: First, instead of setting a reserve price p that leads to expected revenue $k = R(p)$ conditional on trade in this period, one can alternatively and hypothetically assume that the seller sets an expected transaction price k, conditional on trade ever occurring, that leads to trade with probability $1-\overline{F}(k) := 1-F_\infty(R^{-1}(k))$. Second, the intermediary can levy an "expectational fee" $\overline{\omega}(k)$ on the expected transaction price k. The following lemma derives the expectational fee $\overline{\omega}(k)$ that implements the allocation rule derived in Lemma 2.

Lemma 3. *The expectational transaction fee that implements the optimal mechanism described in Lemma 2 is*

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

defined on the equilibrium expected-transaction-price range $[\overline{\Phi}^{-1}(\underline{c}), \overline{v}]$.

Proof of Lemma 3. The expected profit of a seller with cost c who faces a fee $\overline{\omega}$ is

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

Substituting $\overline{\omega}(k)$ by the expression in Lemma 3 for k in the displayed domain, the maximization problem becomes

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

The first-order condition is

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

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

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

Step (a): First, we show that linearity of fees holds for the denormalized limiting distributions F^* and G^*. For simplicity, denote the supports of the denormalized limiting distributions as $[\underline{v}, \overline{v}]$ and $[\underline{c}, \overline{c}]$. The distributions are hence $F^*(v) := 1-[(\overline{v}-v)/(\overline{v}-\underline{v})]^{\beta}$ and $G^*(c) := [(c-\underline{c})/(\overline{c}-\underline{c})]^{\sigma}$. The virtual cost function is linear: $\Gamma_{\alpha}^*(c) := c+(c-\underline{c})\alpha/\sigma$. The optimal expectational fees given in Lemma 3 can be rearranged to yield

$$\overline{\omega}^*(p) = p - \mathbb{E}_v[\Gamma_{\alpha}^{*-1}(\overline{\Phi}^*(v))|v \geq p] = p - \Gamma_{\alpha}^{*-1}(\mathbb{E}_v[\overline{\Phi}^*(v)|v \geq p]) = p - \Gamma_{\alpha}^{*-1}(p),$$

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

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

This implies that the equation for $\overline{\omega}^*$ in (14) holds for the limiting distributions F^* and G^*, because of the definitions of $\tilde{\omega}$ and $\tilde{p}$.

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

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

$$\hat{\overline{\omega}}(\hat{p}) := \frac{\overline{\omega}(u\hat{p})}{u} = \hat{p} - \frac{\int_{\hat{p}}^1 u^{-1}\hat{\Gamma}_\alpha^{-1}(\hat{\overline{\Phi}}(u\hat{v}))d\hat{\overline{F}}(u\hat{v})}{1 - \hat{\overline{F}}(u\hat{p})} \tag{22}$$

where the expression comes from plugging in $u\hat{p}$ for p in the expression in Lemma 3 and dividing by u.

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

By the definition of β we have

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

This implies that

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

and hence

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

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

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

and hence

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

because uniform convergence of a function implies uniform convergence of its inverse (see for example Barvinek et al. (1991)).

Observe that

$$\hat{\overline{F}}(k) = \hat{F}_\infty(\hat{R}^{-1}(k)), \qquad 1 - \hat{F}_\infty(p) = \frac{1 - \hat{F}_{(1)}(p)}{1 - \delta \hat{F}_{(1)}(p)}, \qquad \hat{F}_{(1)}(p) = \sum_{B=0}^{\overline{B}} \pi_B \hat{F}(p)^B \tag{23}$$

Reinstating subscripts j, define, for $x \in [0,1]$,

$$H_j(x) := 1 - \frac{1 - \hat{F}_{(1),j}(u_j x)}{1 - \hat{F}_{(1),j}(0)}, \qquad \varphi_j(x) := \frac{\hat{\Phi}_j(u_j x)}{u_j},$$

and define the normalized expected-revenue map by

$$r_j(x) := \frac{\hat{R}_j(u_j x)}{u_j} = \frac{\int_x^1 \varphi_j(t) dH_j(t)}{1 - H_j(x)}, \qquad x < 1, \tag{24}$$

with $r_j(1) := \varphi_j(1)$.

The convergence established above gives uniform convergence of φ_j to $\varphi(x) := x - (1-x)/\beta$. We next show that H_j converges uniformly to $H(x) := 1 - (1-x)^\beta$. Let $s_j := 1 - \hat{F}_j(0)$ and $S_j(x) := (1 - \hat{F}_j(u_j x))/s_j$. Theorem 1 gives uniform convergence of S_j to $S(x) := (1-x)^\beta$. Since the number of buyers is bounded and $\sum_B B\pi_B > 0$, uniformly in x,

$$1 - \hat{F}_{(1),j}(u_j x) = \sum_{B=0}^{\overline{B}} \pi_B \left[1 - (1 - s_j S_j(x))^B\right] = s_j \left(\sum_{B=0}^{\overline{B}} B\pi_B\right) S_j(x) + O(s_j^2).$$

Dividing this expression by its value at $x = 0$ proves that H_j converges uniformly to H.

We now prove uniform convergence of r_j. Fix $\varepsilon > 0$. On $[0, 1-\varepsilon]$, the denominators in (24) are bounded away from zero. The increasing functions φ_j are uniformly bounded, and hence have uniformly bounded total variation. Uniform convergence of φ_j and H_j, together with integration by parts, therefore implies uniform convergence of $N_j(x) := \int_x^1 \varphi_j(t) dH_j(t)$ on this interval. Hence r_j converges uniformly on $[0, 1-\varepsilon]$. For $x \in [1-\varepsilon, 1]$, the fact that $r_j(x)$ is an average of φ_j over $[x, 1]$ implies

$$|r_j(x) - \varphi_j(1)| \le \sup_{t \in [x,1]} |\varphi_j(t) - \varphi_j(1)|.$$

Uniform convergence of φ_j to the continuous function φ makes the right-hand side uniformly small as ε goes to zero. Combining the two intervals yields uniform convergence of r_j on $[0,1]$. The limiting map satisfies

$$r(x) = \frac{\int_x^1 \varphi(t) dH(t)}{1 - H(x)} = x,$$

where the last equality is the standard conditional-expectation identity for virtual values. Consequently, uniform convergence of the inverse functions gives

$$\frac{\hat{R}_j^{-1}(u_j x)}{u_j} = r_j^{-1}(x) \rightrightarrows x.$$

Together with (23), this gives uniform convergence of the normalized versions of $F_{\infty,j}$ and $\overline{F}_j$. The identity $\tilde{\Phi}_j(p) = \overline{\Phi}_j(R_j(p))$ and the definition of the dynamic virtual valuation then give uniform convergence of the normalized $\overline{\Phi}_j$. Combining these conclusions with the uniform convergence of the normalized $\Gamma_{\alpha,j}^{-1}$ established above gives convergence of the conditional-expectation term in (22) for every $\hat{p} < 1$; the endpoint follows by continuous extension. Therefore, $\hat{\overline{\omega}}$ converges to $\overline{\omega}^*$.

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

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

From Proposition 5 and the arguments that precede it, we know that the optimal transaction fee $\omega(p)$ is given by

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

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

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

The limits of integration and the integrand in the last expression are combinations 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. □

C Extension: Non-Stationarity

Assume now that the environment is described by known sequences $\boldsymbol{\delta} := (\delta_t)_{t=0}^{\infty}$ and $\boldsymbol{F} := (F_t)_{t=0}^{\infty}$, where δ_t is the discount factor in period t, F_t is the distribution from which buyers' types are drawn in period t, and $\Phi_t(v) := v - \frac{1-F_t(v)}{f_t(v)}$ is monotone in v for all t. One example is exponential discounting (or a constant dropout 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 π_B^t 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 draws in t. Expected revenue given a reserve p in period t, conditional on trade in period t, is then given as $R_t(p) := \frac{\int_p^{\overline{v}} \Phi_t(v) dF_{(1),t}(v)}{1-F_{(1),t}(p)}$.

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

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

Next define the frontier

$$1 - \overline{F}(k) := \sup_{\boldsymbol{k}} \left\{ \sum_{t=0}^{\infty} q_t(\boldsymbol{k}) : \frac{\sum_{t=0}^{\infty} q_t(\boldsymbol{k}) k_t}{\sum_{t=0}^{\infty} q_t(\boldsymbol{k})} = k \right\}.$$

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 by (18), yielding the allocation rule (19). Consequently, the functional form of the expectational fees $\overline{\omega}(k)$ under non-stationarity will be the same as under stationarity. Hence, it is as given in Lemma 3 in the proof of Proposition 6. This also implies that in the limit, as G converges to a reflected generalized Pareto distribution, the optimal expectational fee will be linear as in the stationary case.

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

D Extreme Value Theory

Extreme Value Theory For the convenience of the reader, this appendix provides a summary of the results of the theory of exceedances in extreme value theory that are the most important ones for the purposes of our paper. This summary is the content of Theorem 1 below. The theorem says that for any F that satisfies some weak regularity condition,

$$\lim_{u\to 0} 1 - \frac{1-F(\overline{v}-u(\overline{v}-v))}{1-F(\overline{v}-u(\overline{v}-\underline{v}))} = 1 - \left(\frac{\overline{v}-v}{\overline{v}-\underline{v}}\right)^{\beta} =: F^*(v), \tag{25}$$

where convergence is uniform and β is some constant. The left-hand side of (25) is the rescaled distribution conditional on being above the threshold $\overline{v}-u(\overline{v}-\underline{v})$. According to Theorem 1, this truncated and rescaled distribution converges to a generalized Pareto distribution F^* as the threshold $\overline{v}-u(\overline{v}-\underline{v})$ goes to the finite upper bound $\overline{v}$.

The motivation for this theory was the 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.[28] The theory of exceedances 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 et al. (2010)).

The general principle is described by the Pickands-Balkema-de Haan theorem (the second theorem of extreme value theory). For expositional simplicity, we provide a simplified version of the theorem, which 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 $n = 1, 2, \ldots$,*

[28]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.

such that

$$\lim_{n\to\infty} [F(a_n x + b_n)]^n = F_{max}(x)$$

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

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

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 $v \in (\underline{v}, \overline{v})$. Then F has a generalized Pareto upper tail, formally*

$$\lim_{u\to 0} 1 - \frac{1 - F(\overline{v} - u(\overline{v} - v))}{1 - F(\overline{v} - u(\overline{v} - \underline{v}))} = 1 - \left(\frac{\overline{v} - v}{\overline{v} - \underline{v}}\right)^{\beta}, \tag{26}$$

for some constant β, 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 (26) is the rescaled distribution conditional on being above the threshold $\overline{v} - u(\overline{v} - \underline{v})$. 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 $v \in (\underline{v}, \overline{v})$) the definition of F having a generalized Pareto upper tail given in Definition 4 in Pickands (1975) simplifies to (26). □

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*

$$\lim_{v\to\overline{v}} \frac{d}{dv}\left[\frac{1 - F(v)}{f(v)}\right] = \overline{\beta}, \tag{27}$$

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 counterpart of the condition with an unbounded support) the normal, exponential, Cauchy, and generalized Pareto distribution without a finite upper bound of support.

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 $\xi \geq 0$, its support is $[\mu, \infty)$. 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 $F(v) := 1 - (\mu/v)^{1/\xi}$. For $\xi > 0$ one obtains the Type II Pareto distribution.

Without a finite upper bound, convergence can be stated as

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

for some generalized Pareto distribution F_u^*. See the above-mentioned 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 (standard 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.