

Loertscher, Simon; Niedermayer, Andras

Working Paper

Fee-Setting Mechanisms: On Optimal Pricing by Intermediaries and Indirect Taxation

SFB/TR 15 Discussion Paper, No. 434

Provided in Cooperation with:

Free University of Berlin, Humboldt University of Berlin, University of Bonn, University of Mannheim, University of Munich, Collaborative Research Center Transregio 15: Governance and the Efficiency of Economic Systems

Suggested Citation: Loertscher, Simon; Niedermayer, Andras (2012) : Fee-Setting Mechanisms: On Optimal Pricing by Intermediaries and Indirect Taxation, SFB/TR 15 Discussion Paper, No. 434, Sonderforschungsbereich/Transregio 15 - Governance and the Efficiency of Economic Systems (GESY), München,
<https://doi.org/10.5282/ubm/epub.17391>

This Version is available at:

<https://hdl.handle.net/10419/94073>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 434

**Fee-Setting Mechanisms: On
Optimal Pricing by
Intermediaries and Indirect
Taxation**

Simon Loertscher*
Andras Niedermayer**

* University of Melbourne

** University of Mannheim

October 27, 2012

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

Fee-Setting Mechanisms: On Optimal Pricing by Intermediaries and Indirect Taxation *

Simon Loertscher[†] Andras Niedermayer[‡]

October 27, 2012

Abstract

Mechanisms according to which private intermediaries or governments charge transaction fees or indirect taxes are prevalent in practice. We consider a setup with multiple buyers and sellers and two-sided independent private information about valuations. We show that any weighted average of revenue and social welfare can be maximized through appropriately chosen transaction fees and that in increasingly thin markets such optimal fees converge to linear fees. Moreover, fees decrease with competition (or the weight on welfare) and the elasticity of supply but *decrease* with the elasticity of demand. Our theoretical predictions fit empirical observations in several industries with intermediaries.

Keywords: brokers, applied mechanism design, linear commission fees, optimal indirect mechanisms, auction houses.

JEL-Classification: C72, C78, L13

*We want to thank Alp Atakan, Pierre Boyer, Yuelan Chen, Daniele Condorelli, Eddie Dekel, Winand Emons, Péter Eső, Ekaterina Goldfayn, Julio González-Díaz, Philipp Kircher, Stephan Lauermann, Ferenc Niedermayer, François Ortalo-Magné, Marco Ottaviani, Martin Peitz, Bernard Salanié, Mark Satterthwaite, Armin Schmutzler, Art Shneyerov, Dan Spulber, Adam Szeidl, Ernst-Ludwig von Thadden, Cédric Wasser, Asher Wolinsky, Abdullah Yavas, the participants of SED 2008 in Ann Arbor, the Midwest Economic Theory Meeting 2008 in Columbus, IIOC 2009 in Boston, SFB TR15 Meeting 2010 in Caputh and seminars at the Universities of Basel, Bern, Bristol, Mannheim, Melbourne, Michigan, Munich, New South Wales, Pennsylvania, Queensland, Sydney, Zurich, and Columbia, Concordia, Deakin, Georgetown and Northwestern University for their helpful comments. The second author gratefully acknowledges the hospitality of the Managerial Economics and Decision Sciences and Management and Strategy Departments at Northwestern University's Kellogg School of Management. Financial support through a research grant by the Faculty of Business and Economics at the University of Melbourne is also gratefully acknowledged. A previous version of this paper circulated under the title "Fee-Setting Intermediaries: On Real Estate Agents, Stockbrokers, and Auction Houses".

[†]Department of Economics, Economics & Commerce Building, University of Melbourne, Victoria 3010, Australia. Email: simonl@unimelb.edu.au.

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

1 Introduction

Mechanisms according to which sellers are charged a commission fee or a tax levied on the transaction price are prevalent in practice. Examples include the indirect taxes imposed by governments, percentage fees charged by real-estate and by stock brokers and commission fees charged by auction houses and sites such as Sotheby's, Christie's, and eBay. Currently, little is known about the effects and determinants of such mechanisms, which we call *fee-setting mechanisms*. Fee-setting mechanisms have recently fared prominently in public debates on issues as various as credit-card fees, allegations of collusive commission fee setting by auction houses and real estate agents, the – at times drastic – increases of value-added taxes in financially troubled countries, and the introduction of financial transaction taxes in the European Union.

We study fee-setting mechanisms within an independent private values setup. Private information is two-sided insofar as every buyer and every seller is privately informed about his or her type. Buyers have unit demand and sellers have unit capacities. All buyers are *ex ante* symmetric and all sellers are *ex ante* symmetric in the sense that their types are independently and identically distributed. The mechanism designer can be thought of as a private market maker or broker. Alternatively, one can think of a government providing a trading platform via the legal system and law enforcement. The primary focus of our analysis is on the market making problem a mechanism designing intermediary faces whose objective it is to maximize a weighted average of social surplus and revenue accruing to the market maker. We call mechanisms that achieve this objective constrained efficient, or optimal. Such a weighted average is of interest both as a model of imperfect competition between private brokers and as the solution to a Ramsey pricing problem by a government.¹

As to why the services of private intermediaries and governments are useful or necessary, there are a variety of mutually non-exclusive explanations such as for reducing transaction and search costs, certifying quality, improving matching, building reputa-

¹In the world of brokerage, this problem is the analog to the problem of optimal nonlinear pricing for a regulated monopoly in the tradition of Ramsey; see, for example, Wilson (1993).

tion, providing infrastructure that facilitates trade, and enforcing contracts.² Rather than choosing a specific combination of these explanations, we focus on the problem of optimal pricing by intermediaries and optimal indirect taxation, taking as given that governments impose indirect taxes and that buyers and sellers trade via intermediaries.

Our main interest is on thin markets, that is, on environments in which the number of buyers and sellers is small. The results are easiest to summarize for a bilateral trade problem with one buyer and one seller, for which the fee-setting mechanism is such that the intermediary announces a fee that is a function of the transaction price, the seller observes the fee and sets a take-it-or-leave-it price and the buyer accepts or rejects the price.³ Our first main result is that fee-setting mechanisms are without loss of generality. That is, an allocation rule can be implemented through an incentive compatible and interim individually rational Bayesian mechanism if and only if it can be implemented via a fee-setting mechanism with an appropriately chosen fee.

Fee-setting mechanisms are conceptually simple. However, the optimal fee may be a complicated, non-linear function of the transaction price in general. The fee functions that are used in practice are often linear. To address the question under what conditions linear fees are good approximations to optimal fees, we analyze a setup in which only the buyers with the highest valuations and only the sellers with the lowest costs enter. Such a setup can arise, for example, if high transaction costs prevent traders with low expected gains from trade from entering. We show that in increasingly tight markets – that is, as the threshold to enter the market increases for buyers and decreases for sellers – the conditional distributions of entrants converge to (mirrored) Generalized Pareto distributions, or Pareto distributions for short, and that the optimal fees converge to the linear fees implied by Pareto distributions.

Pareto distributions allow for clear-cut results for the effects of elasticities and competition. As the elasticity of supply increases, both the overall fee and the marginal fee decrease. The overall fee also decreases as competition between private intermediaries

²See Spulber (1999) and Salanié (2003) for an overview of the role of private intermediaries and governments, respectively.

³For the sake of brevity we will refer to the entity designing the mechanism as the intermediary, which can be either a private intermediary or a government.

aries becomes more intensive or, equivalently, as the government's need to raise revenue decreases. Somewhat surprisingly, the marginal fee increases as competition between intermediaries becomes more intensive if the elasticity of demand exceeds the elasticity of supply. Even more surprisingly, the overall fee decreases as the elasticity of demand *decreases*.⁴ This seems to contradict the standard notion that inelastically demanded goods are taxed more heavily. Roughly, the intuition is that both the seller and the intermediary want to have higher prices when the elasticity of demand decreases; but absent a reduced fee the seller wants to increase the price by more than the intermediary wishes him to, which makes the intermediary lower his fees. These comparative statics generalize beyond the class of Pareto distributions, at times with the qualification that they hold for the convex lower and the concave upper bound of the fee function (which coincide with the fee for Pareto distribution).

The results summarized so far also generalize naturally to a multilateral trade problem with multiple buyers and one seller: the seller sets a reserve price rather than a take-it-or-leave-it offer to buyers, and buyers participate in a second-price auction. Moreover, auction results can be embedded in a random matching model, in which a seller is matched with a random number of buyers and the bargaining process is modeled as an auction.

Our results on fees and taxes for bilateral and multilateral trade shed light on empirical observations in a variety of industries, such as on commissions of auction houses and websites, bargaining over slotting allowances, markups of retailers, and posted prices of currency exchange offices and used car dealers. Even real estate brokerage fees, which are often considered a prime example of a solution of a principal-agent problem, can be explained naturally as solutions to an optimal pricing setup. Indeed, many of the stylized facts in real estate brokerage that appear puzzling, if viewed from a principal-agent perspective, are predictions of our optimal pricing problem.^{5,6}

⁴This result holds for arbitrarily small positive weights on social welfare. For a zero weight on social welfare and Pareto distributions, the fee does not depend on the elasticity of demand.

⁵Though our analysis is confined to the question of optimal pricing, it is worth noting that fee-setting mechanisms will provide some incentives for the broker to exert effort if effort increases the probability of a sale because the fee is paid if and only if a transaction occurs.

⁶These stylized facts include brokers' use of fee-setting mechanisms, the linearity of these fees, their

Moreover, assuming intermediaries employ constrained efficient mechanisms, our analysis suggests a method for the empirical assessment of how competitive the fee structure in a given industry is, provided one has estimates of the elasticities of demand and supply for that industry. This question is relevant for anti-trust cases such as the case against Sotheby's and Christie's, and for the recent investigation of the real estate brokerage industry (see Department of Justice, 2007).

The present paper relates to the literature on Bayesian mechanism design with two-sided private information, which was pioneered by Myerson and Satterthwaite (1983) and Gresik and Satterthwaite (1989). Our analysis considers the whole frontier that characterizes the efficient tradeoffs between revenue and welfare, whereas previous work has focused on certain points along that frontier, such as on the welfare-maximizing (or *ex post* efficient) point, the revenue-maximizing point or the point that maximizes welfare subject to budget balance.⁷ Moreover, fee-setting mechanisms provide an empirical counterpart to the dominant strategy implementation, which is much more prevalent in theory than in practice. The facts that equilibrium under a fee-setting mechanism is essentially unique, and that fee-setting mechanisms can be used to implement any incentive compatible allocation rule contrast with properties of the double-auction of Chatterjee and Samuelson (1983).⁸ Further, fee-setting mechanisms are *ex post* individually rational, which generalizes a result obtained by Gresik (1991).

Jullien and Mariotti (2006) studied fee setting by intermediaries in a common value setup, while Shy and Wang (2011) compared the performance of percentage fees charged by credit-card companies to fixed fees as a function of the vertical market structure. Matros and Zapechelnyuk (2008) also addressed fee setting by intermediaries in their study of an intermediary that sets a constant fee in an auction with the sequential arrival of an infinite number of buyers. In contrast to these papers, our paper analyzes

invariance to large variations in real estate prices, and the observation that direct sellers on for-sale-by-owner platforms achieve the same price as the gross price achieved by owners selling through a broker. See Hsieh and Moretti (2003) and Hendel, Nevo, and Ortalo-Magné (2009).

⁷See, for example, Myerson and Satterthwaite (1983), Gresik and Satterthwaite (1989), Makowski and Mezzetti (1993), or Baliga and Vohra (2003).

⁸As noted by Myerson and Satterthwaite (1983), for uniform distributions the double-auction has a linear equilibrium that implements the surplus maximizing allocation rule subject to budget balance. However, there are many other equilibria (see, for example, Leininger, Linhart, and Radner (1989)).

thin markets with a small number of traders, two-sided private information, and different levels of competitiveness. This difference in assumptions drives our main findings, such as the shape (and, under some conditions, the linearity) of the optimal fee function, and the effects of the elasticities and of competitiveness.

In public finance, the optimal structure of indirect taxes has received little attention compared to the optimal structure of possibly non-linear income taxes.⁹ To the extent that imperfect competition has been accounted for, this strand of literature has typically stayed within a setup with complete information; see, for example, Anderson, De Palma, and Kreider (2001a,b). We analyze the optimal structure of indirect taxes when both sides of the market are thin and have private information, which is a plausible set of assumptions whenever matching frictions are substantive. We also derive conditions under which, respectively, a specific tax and an *ad valorem* tax are optimal. This contrasts with models of complete information, where *ad valorem* taxes are typically found to be welfare superior (Anderson, De Palma, and Kreider, 2001b).

Lastly, the efficacy of simple mechanisms relative to elaborate optimal mechanism has recently been investigated under a variety of setups; see, for example, McAfee (2002), Rogerson (2003), Chu and Sappington (2007), Chu, Leslie, and Sorensen (2011), and Loertscher and Niedermayer (2012a). The present paper contributes to this strand of literature by providing conditions under which simple (that is, linear) fee-setting mechanisms are optimal, or converge to optimality.

The remainder of this paper is structured as follows. Section 2 introduces the model. Section 3 derives the constrained efficient, or optimal, allocation rules. Sections 4 and 5 analyze, respectively, the case with bilateral and multilateral trade. Section 6 concludes. All proofs are in the appendix.

2 Model

We study different forms of the following trading problem with agents who are privately informed about their types. There are B buyers indexed by $b \in \mathbb{B}$ whose valuations v_b for

⁹See, for example, Salanié (2003, Chapter 3).

one unit of a homogeneous good are independently and identically distributed according to the twice continuously differentiable distribution function F with support $[\underline{v}, \bar{v}]$ and density $f(v_b) > 0$ for all $v_b \in (\underline{v}, \bar{v})$. Similarly, there are S sellers indexed by $s \in \mathbb{S}$ whose costs c_s for producing one unit of a homogeneous good are independently and identically distributed according to the twice continuously differentiable distribution function G with support $[\underline{c}, \bar{c}]$ and density $g(c_s) > 0$ for all $c_s \in (\underline{c}, \bar{c})$.¹⁰ Each buyer demands and each seller supplies at most one unit. All agents are risk-neutral and have quasilinear preferences, so that the expected payoff of a buyer b with valuation v_b who receives a unit of the good with probability q and pays a transfer payment t is $v_b q - t$. Analogously, the expected payoff to a seller s with cost c_s who receives a transfer t and produces the good with probability q is $t - c_s q$. We assume that the value of the outside option of not participating in the mechanism is 0 for all agents. For expositional simplicity, in most of the paper we will assume $\underline{v} = \underline{c}$ and $\bar{v} = \bar{c}$. In the appendix, we deal with the case when this does not hold.

There is also a mechanism designer in this model, which is either a government or a private intermediary, and which we simply refer to as the intermediary. We assume that buyers and sellers can only trade via the intermediary. This intermediary maximizes the weighted average of expected revenue and expected welfare as described in detail below.

The model can be summarized as $\langle B, S \rangle$. Within the independent private values environment, this model is fairly general. It subsumes, for example, the model analyzed by Myerson and Satterthwaite (1983), which focuses on the bilateral trade case $\langle 1, 1 \rangle$ without assuming $\underline{v} = \underline{c}$ and $\bar{v} = \bar{c}$ and the $\langle B, S \rangle$ -model of Gresik and Satterthwaite (1989), which assumes $\underline{v} = \underline{c}$ and $\bar{v} = \bar{c}$.

Define the *virtual valuation function* Φ and the *virtual cost function* Γ , which are well known in mechanism design, as

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

¹⁰These assumptions are slightly weaker than the usually made assumptions $f(v_b) > 0$ for $v_b \in [\underline{v}, \bar{v}]$ and $g(c_s) > 0$ for $c_s \in [\underline{c}, \bar{c}]$. For some of our asymptotic results, more insights can be gained by using our weaker assumption.

Further, define the weighted virtual valuation and weighted virtual cost functions as

$$\Phi_\alpha(v) = (1 - \alpha)v + \alpha\Phi(v) \quad \text{and} \quad \Gamma_\alpha(c) = (1 - \alpha)c + \alpha\Gamma(c),$$

that is, as the weighted average of valuation and virtual valuation and the weighted average of cost and virtual cost. In the main text, we focus our attention on the regular case by assuming that $\Phi(v)$ and $\Gamma(c)$ are strictly increasing functions (that is, that Myerson's regularity condition holds), and we deal with the irregular case in the appendix. Observe that strict monotonicity of $\Phi(v)$ and $\Gamma(c)$ implies that $\Phi_\alpha(v)$ and $\Gamma_\alpha(c)$ are strictly monotone for any $\alpha \in [0, 1]$.¹¹ We further assume that Φ and Γ are continuously differentiable.

As observed by Bulow and Roberts (1989), $\Phi(v)$ can be interpreted as the intermediary's marginal revenue and $\Gamma(c)$ as the marginal cost when increasing the buyer's and seller's probability of trade.¹² v and c are the marginal utilities of the buyer and the seller when the probability of trade increases marginally. Consequently, $\Phi_\alpha(v)$ can be seen as the weighted average of marginal revenue and marginal utility and $\Gamma_\alpha(c)$ as the weighted average of marginal cost and marginal utility.

Denoting by E_x the expectation taken with respect to the distribution of the random variable x , we observe that for any p

$$E_c[\Gamma(c)|c \leq p] = \frac{\int_c^p \Gamma(c)g(c)dc}{G(p)} = p \quad \text{and} \quad E_v[\Phi(v)|v \geq p] = \frac{\int_p^{\bar{v}} \Phi(v)f(v)dv}{1 - F(p)} = p. \quad (1)$$

This can be established, for example, using integration by parts.

3 Optimal Allocation Rules

We first consider the problem of the intermediary as a general mechanism design problem.

When deriving the optimal fee-setting mechanism, we then only need to ensure that it

¹¹Note that for $\alpha > 1$, monotone virtual valuations and virtual costs are not sufficient to guarantee monotonicity of Φ_α and Γ_α . A sufficient condition for this is that the hazard rate $f(v)/(1 - F(v))$ is increasing and that the reverse hazard rate $g(c)/G(c)$ is decreasing.

¹²That is, interpret the probability that $v \geq p$ and $c \leq p$ as quantity demanded and supplied q at price p , that is, $q := 1 - F(p)$ and $q := G(p)$. Thus, the inverse demand and supply functions are $p^d(q) := F^{-1}(1 - q)$ and $p^s(q) := G^{-1}(q)$, yielding $R(q) := qF^{-1}(1 - q)$ and $C(q) := qG^{-1}(q)$ as revenue and cost functions, respectively. Taking derivative w.r.t. q and substituting back in yields $R'(q) = \Phi(p)$ and $C'(q) = \Gamma(c)$.

implements the allocation rule that is the solution to the general mechanism design problem.

Direct Mechanisms A direct mechanism asks all buyers and sellers to report their types $v_b \in [\underline{v}, \bar{v}]$ and $c_s \in [\underline{c}, \bar{c}]$. It consists of an allocation rule $\mathbf{Q}(\mathbf{v}, \mathbf{c})$ and a payment (or transfer) rule $\mathbf{T}(\mathbf{v}, \mathbf{c})$, where $\mathbf{v} = (v_1, \dots, v_B)$ and $\mathbf{c} = (c_1, \dots, c_S)$. \mathbf{Q} maps into $[0, 1]^{B+S}$ and \mathbf{T} maps into \mathbb{R}^{B+S} . The elements of the vector $\mathbf{Q}(\mathbf{v}, \mathbf{c})$ represent the probabilities that individual agents will trade, that is $Q_b(\mathbf{v}, \mathbf{c})$ for $b \in \mathbb{B}$ represents buyer b 's probability of buying and $Q_s(\mathbf{v}, \mathbf{c})$ for $s \in \mathbb{S}$ represents seller s 's probability of selling if agents report (\mathbf{v}, \mathbf{c}) . Similarly, $\mathbf{T}(\mathbf{v}, \mathbf{c})$ is the vector of transfers, that is $T_b(\mathbf{v}, \mathbf{c})$ is buyer b 's payment to the intermediary and $T_s(\mathbf{v}, \mathbf{c})$ is the payment seller s receives from the intermediary.

A direct mechanism is feasible if $\sum_{b \in \mathbb{B}} Q_b(\mathbf{v}, \mathbf{c}) \leq \sum_{s \in \mathbb{S}} Q_s(\mathbf{v}, \mathbf{c})$ for all $(\mathbf{v}, \mathbf{c}) \in [\underline{v}, \bar{v}]^B \times [\underline{c}, \bar{c}]^S$, and in what follows we implicitly restrict our attention to feasible mechanisms. Denote by \mathbf{v}_{-b} the vector \mathbf{v} without v_b , that is $\mathbf{v}_{-b} := (v_1, v_2, \dots, v_{b-1}, v_{b+1}, \dots, v_B)$. The interim probabilities of trade for buyer b of type v_b and seller s of type c_s are, respectively, $q_b(v_b) := E_{\mathbf{v}_{-b}, \mathbf{c}}[Q_b(\mathbf{v}, \mathbf{c})]$ and $q_s(c_s) := E_{\mathbf{v}, \mathbf{c}_{-s}}[Q_s(\mathbf{v}, \mathbf{c})]$. Analogously, the interim expected transfer from buyer b and to seller s are, respectively, $t_b(v_b) := E_{\mathbf{v}_{-b}, \mathbf{c}}[T_b(\mathbf{v}, \mathbf{c})]$ and $t_s(c_s) := E_{\mathbf{v}, \mathbf{c}_{-s}}[T_s(\mathbf{v}, \mathbf{c})]$.

A direct mechanism (\mathbf{Q}, \mathbf{T}) is said to be *incentive compatible* if agents have an incentive to report their types truthfully and *interim individually rational* if agents are willing to participate.^{13,14} By the revelation principle, the focus on direct, incentive-compatible, and individually rational mechanisms is without loss of generality. An allocation rule \mathbf{Q} can be implemented via an incentive-compatible direct mechanism if and only if the

¹³Formally, incentive compatibility means $v_b q_b(v_b) - t_b(v_b) \geq v_b q_b(\hat{v}_b) - t_b(\hat{v}_b)$ and $t_s(c_s) - c_s q_s(c_s) \geq t_s(\hat{c}_s) - c_s q_s(\hat{c}_s)$ for all $b \in \mathbb{B}$, all $s \in \mathbb{S}$, all $v_b, \hat{v}_b \in [\underline{v}, \bar{v}]$, and all $c_s, \hat{c}_s \in [\underline{c}, \bar{c}]$. Interim individual rationality means $v_b q_b(v_b) - t_b(v_b) \geq 0$ and $t_s(c_s) - c_s q_s(c_s) \geq 0$ for all $b \in \mathbb{B}$, all $s \in \mathbb{S}$, and all $(\mathbf{v}, \mathbf{c}) \in [\underline{v}, \bar{v}]^B \times [\underline{c}, \bar{c}]^S$.

¹⁴Interim individual rationality constraints are very natural if the mechanism designer is a private intermediary. In the case that government is the mechanism designer, the requirement that the mechanism satisfy individual rationality is less obvious as the power of coercing individuals is part of what distinguishes the state from private agents. Nonetheless, for reasons that are admittedly outside the model, it may still be desirable that government's mechanisms satisfy individual rationality; see Hellwig (2003, Section 7) for a discussion of these questions in the context of providing public goods when agents' valuations are their private information. This is also related to the view widely held in the public finance literature that non-distortionary lump-sum transfers are not feasible (see Salanié (2003)).

allocation rule is such that, for all s and all c_s , $q_s(c_s)$ is weakly decreasing in c_s and, for all b and all v_b , $q_b(v_b)$ is weakly increasing in v_b . It follows that an allocation rule can be implemented as the outcome of a (Bayes Nash) equilibrium of any mechanism if and only if the allocation rule can be implemented in an incentive-compatible direct mechanism. For terminological simplicity, we say that an allocation rule is implementable if it can be implemented through an incentive-compatible direct mechanism.

Optimal Allocation Rules The *ex ante* expected social surplus that is generated under the mechanism (\mathbf{Q}, \mathbf{T}) is $W = E_{\mathbf{v}, \mathbf{c}} [\sum_{b \in \mathbb{B}} v_b Q_b(\mathbf{v}, \mathbf{c}) - \sum_{s \in \mathbb{S}} c_s Q_s(\mathbf{v}, \mathbf{c})]$, which only depends on the allocation rule because lump-sum transfers do not affect welfare. The expected revenue of the intermediary is

$$R = E_{\mathbf{v}, \mathbf{c}} \left[\sum_{b \in \mathbb{B}} T_b(\mathbf{v}, \mathbf{c}) Q_b(\mathbf{v}, \mathbf{c}) - \sum_{s \in \mathbb{S}} T_s(\mathbf{v}, \mathbf{c}) Q_s(\mathbf{v}, \mathbf{c}) \right].$$

The goal of the intermediary is to maximize the weighted average $(1 - \alpha)W + \alpha R$ for some $\alpha \in [0, 1]$. The following proposition provides a uniform statement that encompasses as special cases various problems that have been studied in the literature and that are discussed at the end of this section.

Proposition 1. (a) *For any incentive-compatible, individually rational mechanism*

$$(1 - \alpha)W + \alpha R = E \left[\sum_{b \in \mathbb{B}} \Phi_\alpha(v_b) Q_b(\mathbf{v}, \mathbf{c}) - \sum_{s \in \mathbb{S}} \Gamma_\alpha(c_s) Q_s(\mathbf{v}, \mathbf{c}) \right] - \alpha \left[\sum_{b \in \mathbb{B}} \underline{U}_b + \sum_{s \in \mathbb{S}} \bar{U}_s \right],$$

where $\underline{U}_b = \underline{v} q_b(\underline{v}) - t_b(\underline{v})$ and $\bar{U}_s = t_s(\bar{c}) - \bar{c} q_s(\bar{c})$ are the utilities of the least efficient agents.

(b) *Let the α -allocation rule \mathbf{Q}^α be such that it maximizes $\sum_{b \in \mathbb{B}} \Phi_\alpha(v_b) Q_b(\mathbf{v}, \mathbf{c}) - \sum_{s \in \mathbb{S}} \Gamma_\alpha(c_s) Q_s(\mathbf{v}, \mathbf{c})$ for every realization of (\mathbf{v}, \mathbf{c}) . A mechanism maximizes $(1 - \alpha)W + \alpha R$ if and only if it gives zero utility to the least efficient agents and implements the α -allocation rule \mathbf{Q}^α .*

The following simple interpretation can be given for the α -allocation rule \mathbf{Q}^α . For any realization of reported types (\mathbf{v}, \mathbf{c}) define the α -Walrasian sets of buyers and sellers as those buyers and sellers who would trade in a Walrasian equilibrium if the true types

were $\Phi_\alpha(v_b)$ and $\Gamma_\alpha(c_s)$ for all $b \in \mathbb{B}$ and for all $s \in \mathbb{S}$. Then set $Q_b^\alpha(\mathbf{v}, \mathbf{c}) = Q_s^\alpha(\mathbf{v}, \mathbf{c}) = 1$ for all buyers and sellers who belong to the α -Walrasian sets and $Q_b^\alpha(\mathbf{v}, \mathbf{c}) = Q_s^\alpha(\mathbf{v}, \mathbf{c}) = 0$ for all others. Note that for $\alpha = 0$ the α -allocation rule coincides with the Walrasian equilibrium because at $\alpha = 0$ the objective function is $(1 - \alpha)W + \alpha R = W$ and the α -allocation rule induces the Walrasian sets of agents, defined with respect to their true types v_b and c_s , to trade. For $\alpha = 1$, the objective function is $(1 - \alpha)W + \alpha R = R$. Therefore, the α -allocation rule with $\alpha = 1$ maximizes revenue and induces trade by the Walrasian sets defined with respect to $\Phi(v_b)$ and $\Gamma(c_s)$.

An alternative interpretation of α is as a measure of competition in a standard oligopoly model. As an analogy, consider, for example, a Cournot model with N symmetric firms who sell homogenous goods to consumers with demand $1 - F(p)$ and let R measure total profits. Using arguments analogous to those in footnote 12, one can show that the symmetric equilibrium of this model predicts the same prices, welfare, and revenues as the solution of maximizing $(1 - \alpha)W + \alpha R$ with $\alpha = 1/N$. Further, $\alpha = 1$ corresponds to monopoly and $\alpha = 0$ represents perfect competition. It is also worth noting that the interpretation of α is the same across different oligopoly models.¹⁵

The set of all feasible mechanisms can be represented in revenue-welfare-space, as for the example in Figure 1 (shaded area). Let R_α and W_α denote, respectively, expected revenue and welfare under an α -allocation rule. The set of points (R_α, W_α) for $\alpha \geq 0$ can be viewed as the *maximum revenue frontier*, since it represents all mechanisms that maximize revenue subject to achieving a certain welfare level (dashed and solid lines in Figure 1). The subset of points (R_α, W_α) for $\alpha \in [0, 1]$ can be viewed as the *efficient frontier* in analogy to the term in finance, since it represents all mechanisms that maximize welfare subject to a revenue constraint (solid line in Figure 1).¹⁶ It is not hard to show that the slope and continuity properties of these frontiers are as depicted

¹⁵For example, if competition is Bertrand rather than Cournot, $\alpha = 1$ still represents monopoly and $\alpha = 0$ perfect competition, but the latter occurs already for $N = 2$ because under Bertrand competition two is enough for perfect competition.

¹⁶In finance, the set of all points that minimize variance for a given expected return is the minimum variance frontier. The subset of this frontier that puts positive weight on expected return is the efficient frontier.

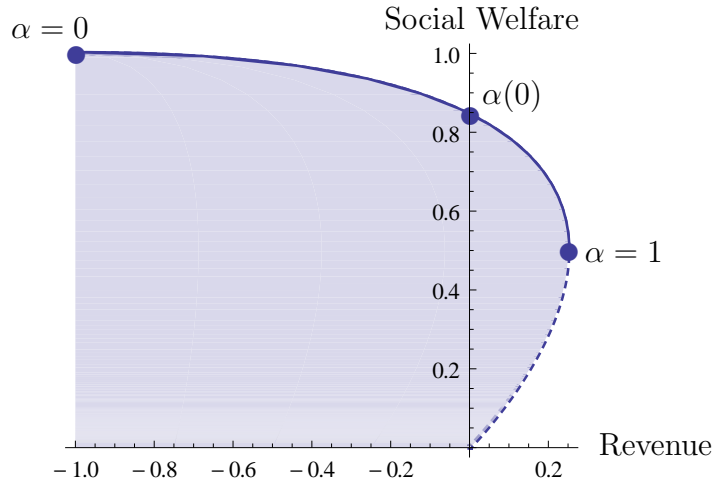


Figure 1: The efficient frontier (solid) and the maximum revenue frontier (solid and dashed) for the bilateral trade problem with uniform distributions F and G on support $[0, 1]$. Social welfare and revenues are normalized by the first-best social welfare $W_0 = 1/6$. The highest achievable revenue is $R_1 = 1/24$. The maximum social welfare achievable under a balanced budget is $W_{\alpha(0)} = 9/64$ with $\alpha(0) = 1/3$.

in the figure in general.¹⁷

Of course, one could also consider the problem of maximizing W subject to $R \geq \underline{R}$ instead of maximizing $(1 - \alpha)W + \alpha R$. Let λ^* be the optimal value of the Lagrange-multiplier for the constraint $R \geq \underline{R}$; that is, the maximizer of the Lagrangian $W + \lambda(R - \underline{R})$. The solution to this constrained optimization problem is the same as the solution to the problem of maximizing $(1 - \alpha)W + \alpha R$ with $\alpha = \lambda^*/(1 + \lambda^*)$,¹⁸ where α has the interpretation of the Ramsey number widely used in public finance and the theory of regulation and nonlinear pricing.¹⁹ Notice that λ^* will depend on \underline{R} , in which case we write $\alpha(\underline{R}) = \lambda^*(\underline{R})/(1 + \lambda^*(\underline{R}))$. However, if one is maximizing welfare in multiple markets and the market under consideration is small relative to the whole economy, one can also treat λ^* as an approximately exogenous parameter that measures the social

¹⁷That the revenue R_0 is negative for the first best allocation $\alpha = 0$ holds for the bilateral trade problem and for any $\langle B, S \rangle$ if $\underline{v} = \underline{c}$ and $\bar{v} = \bar{c}$ holds. It may be positive for $B > 1 = S$ and $\bar{v} > \bar{c}$. See Myerson and Satterthwaite (1983), Gresik and Satterthwaite (1989), and Makowski and Mezzetti (1993), respectively.

¹⁸For the case of bilateral trade, Bulow and Roberts (1989, p.1084) observed the equivalence of maximizing $(1 - \alpha)W + \alpha R$ and the problem of maximizing W subject to $R \geq \underline{R}$, which can be seen as a Ramsey pricing problem. For analogous reasoning in the context of public goods provision, see Hellwig (2003).

¹⁹See, for example, Wilson (Chapter 5 1993).

opportunity costs of public funds, which determines α . Empirical studies assess the social cost of tax revenue raised to be in the range between \$0.2 and \$0.6 per dollar raised.²⁰ This corresponds to an α between 0.17 and 0.38.

Three important special cases have been studied in the literature: $\alpha = \alpha(0)$, which maximizes social surplus subject to budget balance; $\alpha = 1$, which maximizes the revenue of the intermediary; and $\alpha = 0$, which is *ex post* efficient trade.²¹ Moreover, for $B = S = 1$ Gresik (1991) has constructed a mechanism that implements the $\alpha(0)$ -allocation rule and respects individual rationality *ex post*, that is, where agents are willing to participate after the realization of the types of other agents is revealed to them.²²

4 Bilateral Trade

4.1 General Implementation Result

In this section, we study the bilateral trade problem $\langle 1, 1 \rangle$. We begin with the definition of a fee-setting mechanism.

Definition 1. *In a fee-setting mechanism for bilateral trade, the intermediary first announces a fee function $\omega(\cdot)$ and the seller sets a price p . If the buyer is willing to trade, he pays the price p , the intermediary receives the fee $\omega(p)$, and the seller receives the net price $p - \omega(p)$. If the buyer is not willing to trade, there are no payments.*

Now consider an intermediary that wants to ensure that trade occurs if and only if $v \geq P(c)$ with a fee-setting mechanism, where P is some arbitrary increasing function mapping from $[\underline{c}, \bar{c}]$ to $[\underline{v}, \bar{v}]$. The easiest way to derive the optimal fee-setting mechanism is to first derive the dominant strategy implementation: The buyer pays the highest

²⁰See, for example, Ballard, Shoven, and Whalley (1985) for an estimate of the social cost of taxation in the U.S. See also Perrigne and Vuong (2011) for a recent econometric application of this idea and a discussion of the literature.

²¹Myerson and Satterthwaite (1983) derive results for $B = S = 1$ and $\alpha \in \{0, \alpha(0), 1\}$. See Gresik and Satterthwaite (1989) for arbitrary $\langle B, S \rangle$, $\underline{c} = \underline{v}$, $\bar{c} = \bar{v}$, and $\alpha = 0$; Baliga and Vohra (2003) for arbitrary $\langle B, S \rangle$ and $\alpha = 1$; and Makowski and Mezzetti (1993) for $B > 1 = S$, $\bar{v} > \bar{c}$, and $\alpha = \alpha(0)$.

²²Formally, a direct mechanism is said to satisfy *ex post individual rationality* if, for all $b \in \mathbb{B}$, all $s \in \mathbb{S}$, and all $(\mathbf{v}, \mathbf{c}) \in [\underline{v}, \bar{v}]^B \times [\underline{c}, \bar{c}]^S$, $v_b - T_b(\mathbf{v}, \mathbf{c}) \geq 0$ for all buyers b who receive a unit (and $T_b(\mathbf{v}, \mathbf{c}) \leq 0$ for all buyer who do not) and $T_s(\mathbf{v}, \mathbf{c}) - c_s \geq 0$ for all sellers s who produce a unit (and $T_s(\mathbf{v}, \mathbf{c}) \leq 0$ for all sellers who do not).

valuation that he could have reported and still obtained the good, $P(c)$, in case of trade. The seller receives the highest cost he could have reported and still sold the good, $P^{-1}(v)$. If there is no trade, there are no transfers. That reporting one's true type is a dominant strategy for both buyer and seller is easy to verify.

Now we can transform the dominant strategy implementation into a fee-setting mechanism, in which the buyer pays the gross price $P(c)$ and the seller receives the net price $P(c) - \omega(P(c))$. It is easy to check that the buyer will accept if and only if $v \geq P(c)$ and receives the same payoff as with the dominant strategy implementation. Therefore, it is only left to confirm that the seller receives the same expected payoff as in the dominant strategy mechanism, $P^{-1}(v)$. Note that in the dominant strategy implementation, the seller receives $E_v[P^{-1}(v)|v \geq P(c)]$ in expectations over v conditional on trade occurring. Since the seller is risk neutral, the intermediary could alternatively pay the seller the expected value as the net price $P(c) - \omega(P(c))$. Equating the net price with the seller's expected payoff and replacing $P(c)$ with p gives the optimal fee stated in Proposition 2 (b). Part (a) of the proposition is proved in Appendix C.²³

Proposition 2. (a) *The focus on fee-setting mechanisms is without loss of generality. That is, a deterministic allocation rule \mathbf{Q} is incentive-compatible if and only if there exists a fee-setting mechanism that implements it.*

(b) *Let $P(c)$ be an increasing function. An allocation rule in which trade occurs if and only if $v \geq P(c)$ can be implemented by the following fee-setting mechanism. The intermediary charges the fee function*

$$\omega(p) = p - E_v[P^{-1}(v)|v \geq p]$$

for $p \leq \bar{v}$ and an arbitrary $\omega(p) > \bar{v} - p$ for $p > \bar{v}$. A seller with cost $c \leq P^{-1}(\bar{v})$ sets the price $P(c)$ and a buyer with valuation v accepts if and only if $p \leq v$. Moreover, the equilibrium is essentially unique in the sense that equilibria only differ with respect to actions of types that never trade in any equilibrium.

²³Given the fee in Proposition 2, it is easy to verify that the seller's payoff is maximized at $p = P(c)$ because his expected payoff is $(E_v[P^{-1}(v)|v \geq p] - c)(1 - F(p)) = \int_p^{\bar{v}} (P^{-1}(v) - c)dF(v)$, so that the first-order condition with respect to p is satisfied at $P^{-1}(p) = c$.

Proposition 1 implies that for a given α the optimal allocation rule is to allow trade if and only if $\Phi_\alpha(v) \geq \Gamma_\alpha(c)$, or equivalently, $v \geq \Phi_\alpha^{-1}(\Gamma_\alpha(c))$. Hence, setting $P(c) = \Phi_\alpha^{-1}(\Gamma_\alpha(c))$, Proposition 2 gives the following corollary.

Corollary 1. *An optimal allocation rule can be implemented with the following fee setting mechanism. The intermediary sets the fee function*

$$\omega_\alpha(p) = p - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v)) | v \geq p] \quad (2)$$

for $p \leq \bar{v}$ and an arbitrary $\omega_\alpha(p) > \bar{v} - p$ for $p > \bar{v}$. A seller with cost $c \leq P_\alpha^{-1}(\bar{v})$ sets the price $P_\alpha(c) := \Phi_\alpha^{-1}(\Gamma_\alpha(c))$ and a buyer with valuation v accepts if and only if $p \leq v$.

As an illustration, consider the example underlying Figure 1 for which F and G are uniform on $[0, 1]$. In this example, which is also used in Myerson and Satterthwaite (1983) and Chatterjee and Samuelson (1983), a profit-maximizing intermediary will set the fee $\omega_1(p) = p/2$, a welfare-maximizing intermediary will set the fee $\omega_0(p) = p/2 - 1/2$, and an intermediary maximizing welfare subject to budget balance will set the fee $\omega_{\alpha(0)} = p/2 - 1/4$, where $\alpha(0) = 1/3$. Implied social welfare and revenues are of course the same as in Myerson and Satterthwaite (1983) and, for $\alpha(0)$, the same as in Chatterjee and Samuelson (1983).

4.2 Participation by the Most Motivated

We now consider markets in which only the most motivated traders participate, that is, the lowest cost sellers and the highest valuation buyers. We will assume that the fraction participating is very small. The first reason that such an analysis is of interest is that, for many markets, this is a good approximation. For example, in real estate markets typically less than 5 percent of homeowners offer their property for sale at any given time. The second reason is that this analysis provides insights that are valid even if the above assumption is not met. In particular, it allows us to describe fee structures in terms of the fundamental economic concepts of the elasticities of demand and supply and the degree of competitiveness. It also allows for a comparison of optimal fee setting and taxation and optimal one-sided pricing.

Assume now that the reason that only a small fraction of agents – those who are most efficient or most motivated – enters the market is that there are additional opportunity costs of trading for buyers and sellers. It is simplest to think of these opportunity costs as transaction costs. As an example, if a potential buyer's valuation for a house is only slightly higher than the current owner's when ignoring transaction costs, they may be unwilling to trade if costs of moving are taken into account. Opportunity costs may also arise for other reasons, such as the option value of future trade.²⁴ Below are two examples of what these opportunity costs may look like.

Example 1 Consider a buyer and a seller whose net valuation v and net cost c have the distributions F and G with supports $[0, 1]$. Assume that the seller's gross cost is $\hat{c} = c + \rho$ and the buyer's gross valuation is $\hat{v} = v - \rho$, with the additive opportunity cost $\rho \geq 0$. The support of the seller's gross cost distribution is $[\rho, 1 + \rho]$. For the buyer, the support is $[-\rho, 1 - \rho]$. Buyers and sellers with valuation \hat{v} and cost \hat{c} outside the relevant range $[\rho, 1 - \rho]$ trade with probability zero. As ρ goes to $1/2$, the length of the relevant range goes to zero, and so do the masses of buyers $1 - F(2\rho)$ and sellers $G(1 - 2\rho)$ in the relevant range.

Example 2 Consider transaction costs that stretch rather than shift valuations and costs. Take again F and G with supports $[0, 1]$. Assume the seller's gross cost is $\hat{c} = \rho c$ with $\rho \geq 1$. This stretches the seller's support by a factor ρ to $[0, \rho]$. Similarly, let $\hat{v} = 1 - \rho(1 - v)$, which stretches the buyer's support by a factor ρ to $[-(\rho - 1), 1]$. Traders outside the relevant range $[0, 1]$ trade with probability zero. The masses of sellers $G(1/\rho)$ and buyers $1 - F(1 - 1/\rho)$ in the relevant range go to zero as ρ goes to infinity.

The commonality between the two examples is that only the most motivated buyers and sellers enter, and the mass of buyers and sellers in the relevant range goes to zero. Such transformations of distributions are dealt with in *extreme value theory* in statistics.

²⁴It may be worthwhile to delay trade and search for a more efficient trading partner. For this interpretation of opportunity costs, one would need to model the option value of future trade in a search market, which adds a further level of complexity.

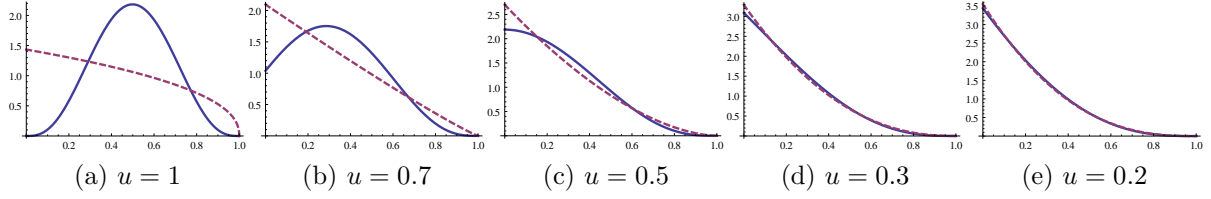


Figure 2: Density of truncated, rescaled distribution $F_u(v) = 1 - [1 - F(\bar{v} - u(\bar{v} - v))]/[1 - F(\bar{v} - u(\bar{v} - \underline{v}))]$, for $u \in \{1, 0.7, 0.5, 0.3, 0.2\}$ for a Beta distribution with support $[0, 1]$ and density $f(v) \propto v^3(1 - v)^3$ (solid line) compared to an approximating Generalized Pareto density with support $[0, 1]$ (dashed). Masses in the relevant range are (a) $1 - F(1 - 1) = 1$, (b) $1 - F(1 - 0.7) = 0.87$, (c) $1 - F(1 - 0.5) = 0.5$, (d) $1 - F(1 - 0.3) = 0.13$, (e) $1 - F(1 - 0.2) = 0.033$. As the mass decreases, the distribution converges to the approximating Pareto distribution and the approximating Pareto distribution converges to the limiting Pareto distribution.

Appendix B contains a brief summary of this theory.

Extreme Value Theory For our purposes, the main result of the theory of exceedences in extreme value theory is summarized in Theorem 1 in Appendix B. The theorem says that for any F that satisfies some weak regularity condition,

$$\lim_{u \rightarrow 0} 1 - \frac{1 - F(\bar{v} - u(\bar{v} - v))}{1 - F(\bar{v} - u(\bar{v} - \underline{v}))} = 1 - \left(\frac{\bar{v} - v}{\bar{v} - \underline{v}} \right)^\beta =: F^*(v), \quad (3)$$

where convergence is uniform and β is some constant. The left-hand side of (3) is the rescaled distribution conditional on being above the threshold $\bar{v} - u(\bar{v} - \underline{v})$. According to Theorem 1, this truncated and rescaled distribution converges to a Generalized Pareto distribution F^* as the threshold $\bar{v} - u(\bar{v} - \underline{v})$ goes to the finite upper bound \bar{v} . See Appendix B for a discussion of the regularity condition (which we show to be satisfied in our setup) and the infinite upper bound version of the theorem.

The motivation for this theory was an empirical regularity found in many situations: that the upper tail of a distribution is well approximated by a (Generalized) Pareto distribution. A prominent example is the distribution of the highest 20 percent of income and wealth in many countries, which was first observed by Vilfredo Pareto.²⁵

²⁵Other 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.

Figure 2 shows the density of a distribution of which only the upper tail is taken. In particular, it shows only the distribution conditional on being on $[1-u, 1]$ rather than the full support $[0, 1]$. As depicted in the figure, moving the truncation point $1-u$ upwards brings the density of the conditional distribution closer to the density of a Generalized Pareto distribution.

We note that symmetric results to Theorem 1 hold for distributions that are truncated from above. In particular, for a distribution G satisfying a weak regularity condition with a finite lower bound \underline{c} , truncation from above leads asymptotically to a mirrored Generalized Pareto distribution, $\lim_{u \rightarrow 0} G(\underline{c} + u(c - \underline{c}))/G(\underline{c} + u(\bar{c} - \underline{c})) = [(c - \underline{c})/(\bar{c} - \underline{c})]^\sigma =: G^*(c)$.

In applied statistics the asymptotic results of extreme value theory are seen as approximations when “close enough” to the limit. In particular, Theorem 1 is understood to mean that for u “close to 0”, a Generalized Pareto distribution is a “good approximation” of the truncated distribution above the threshold $\bar{v} - u(\bar{v} - \underline{v})$. This resembles the application of other asymptotic results, such as the central limit theorem.²⁶ Here, as there, it is then an empirical question of what “close to 0” means. For the distribution of income and wealth the upper 20 percent of the distribution is sufficiently close to 0 to allow for a good approximation by a Pareto distribution. For the numerical example considered in Figure 2, upper tails consisting of the highest 13 percent (Figure 2 (d)) and the highest 3.3 percent (Figure 2 (e)) of the distributions are already very close to a Pareto distribution.

Convergence to Linear Fees We now show how adding opportunity costs to the buyer’s and seller’s initial valuation and cost affects the fee structure. Assume the initial distributions F and G have support $[0, 1]$. This assumption is without loss of generality, as will become evident shortly. We consider a sequence of linear transformations of c and v , indexed by k . Let the seller’s gross cost c_k and the buyer’s gross valuation v_k be

$$c_k = \frac{\bar{v}_k - \underline{c}_k}{u_k^S} c + \underline{c}_k \quad \text{and} \quad v_k = \bar{v}_k - (1 - v) \frac{\bar{v}_k - \underline{c}_k}{u_k^B},$$

²⁶Rather than having an infinite n , a large finite n is seen as a justification for assuming that the distribution of the average of n random draws is approximately normal.

with respective supports $[\underline{c}_k, \underline{c}_k + (\bar{v}_k - \underline{c}_k)/u_k^S]$ and $[\bar{v}_k - (\bar{v}_k - \underline{c}_k)/u_k^B, \bar{v}_k]$. The parameters \underline{c}_k and \bar{v}_k shift the supports. u_k^S can be interpreted as follows. Sellers with costs outside the relevant range $[\underline{c}_k, \bar{v}_k]$ trade with probability zero. The ratio of the length of the relevant range $\bar{v}_k - \underline{c}_k$ to the length of c_k 's support $[(\bar{v}_k - \underline{c}_k)/u_k^S + \underline{c}_k] - \underline{c}_k$ is u_k^S . Therefore, the mass of sellers in the relevant range is $G(u_k^S)$. Similarly, for buyers, the ratio of the relevant range is u_k^B and the mass of buyers in the relevant range is $1 - F(1 - u_k^B)$.

Define the normalized (gross) cost $\tilde{c} = (c - \underline{c}_k)/(\bar{v}_k - \underline{c}_k)$, the normalized (gross) valuation $\tilde{v} = (v - \underline{c}_k)/(\bar{v}_k - \underline{c}_k)$, and the normalized price $\tilde{p} = (p - \underline{c}_k)/(\bar{v}_k - \underline{c}_k)$. The distributions of the normalized gross cost \tilde{c} and the normalized gross valuation \tilde{v} , truncated to $[\underline{c}_k, \bar{v}_k]$ and denoted, respectively, \tilde{G}_k and \tilde{F}_k , are then given as

$$\tilde{G}_k(\tilde{c}) = \frac{G(u_k^S \tilde{c})}{G(u_k^S)} \quad \text{and} \quad \tilde{F}_k(\tilde{v}) = 1 - \frac{1 - F(1 - u_k^B(1 - \tilde{v}))}{1 - F(1 - u_k^B)}. \quad (4)$$

The normalized fee is defined as $\tilde{\omega}(\tilde{p}) = \omega(p)/(\bar{v}_k - \underline{c}_k)$. Now we can return to the examples provided before.

Example 1 revisited Let $\underline{c}_k = 1/2 - 1/(2k)$, $\bar{v}_k = 1/2 + 1/(2k)$, and $u_k^S = u_k^B = 1/k$. For $k = 1$ the supports of c and v are $[0, 1]$. As k goes to infinity, the supports converge to $[1/2, 3/2]$ and $[-1/2, 1/2]$, respectively. For example, let $\alpha = 1$ and F and G be uniform. Plugging this into (2) gives a fee $\omega(p) = (p - \underline{c}_k)/2 = (p - 1/2 - 1/(2k))/2$, which converges to $(p - 1/2)/2$ as k goes to infinity. The normalized fee $\tilde{\omega}(\tilde{p}) = \tilde{p}/2$ is invariant to changes in k .

Example 2 revisited Let $\underline{c}_k = 0$, $\bar{v}_k = 1$, and $u_k^S = u_k^B = 1/k$. For $k = 1$, the supports of c and v are $[0, 1]$. As k goes to infinity, the supports of c and v go to $[0, \infty)$ and $(-\infty, 0]$, respectively. The relevant range remains $[0, 1]$ for all k . For the example of $\alpha = 1$ and F and G uniform, the optimal fee is $\omega(p) = p/2$, and the optimal normalized fee $\tilde{\omega}(\tilde{p}) = \tilde{p}/2$, which is also invariant to changes in k .

As illustrated by the examples, \underline{c}_k and \bar{v}_k are secondary to the analysis. The driving force is that u_k^S and u_k^B go to zero as k goes to infinity. Thus the mass of entering traders vanishes. The general principle behind the convergence of the normalized fee to a linear function observed in the two examples above is as follows.

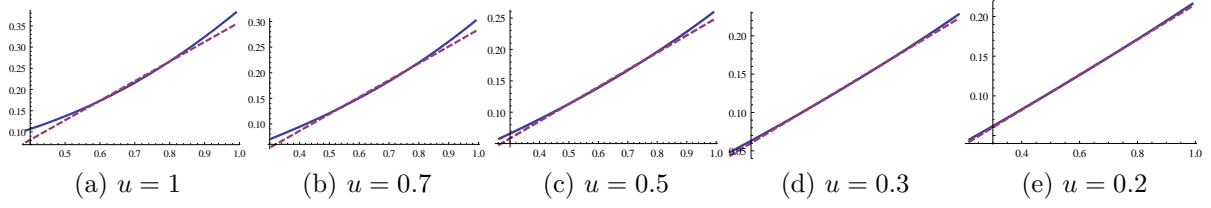


Figure 3: Optimal fee $\omega_\alpha(\cdot)$ for $\alpha = 1$ and truncated, rescaled distributions $F_u(v) = 1 - [1 - F(\bar{v} - u(\bar{v} - v))]/[1 - F(\bar{v} - u(\bar{v} - \underline{v}))]$, and $G_u(c) = G(\underline{c} + u(c - \underline{c}))/G(\underline{c} + u(\bar{c} - \underline{c}))$ for the same setup as in Figure 2; that is, for $u \in \{1, 0.7, 0.5, 0.3, 0.2\}$ for Beta distributions with support $[0, 1]$ and density $f(x) = g(x) \propto x^3(1 - x)^3$ (solid line) compared to an approximating linear fee (dashed).

Proposition 3. Assume F and G have bounded support and Φ and Γ are continuously differentiable. Let the shifting constants $\underline{c}_k, \bar{v}_k$ be arbitrary sequences satisfying $\underline{c}_k < \bar{v}_k$ for all k . Let the ratios of the relevant ranges u_k^S and u_k^B be sequences that go to 0 as k goes to infinity. Then, as $k \rightarrow \infty$,

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

(ii) the normalized fee converges to a linear fee:

$$\tilde{\omega}_\alpha(\tilde{p}) = \alpha \tilde{p} \left[\frac{1}{\alpha + \sigma} \right] - (1 - \alpha)(1 - \tilde{p}) \left[\frac{\sigma}{(\alpha + \sigma)(\beta + 1)} \right] \quad (5)$$

While the proof of convergence of arbitrary distributions with bounded support and continuously differentiable virtual valuation functions relies on results from extreme value theory, it is relatively easy to check that convergence is immediate if one begins with (mirrored) Generalized Pareto distributions $G^*(c) = c^\sigma$ and $F^*(v) = 1 - (1 - v)^\beta$: one only to verify that G^* and F^* are invariant to truncation and rescaling as defined in (4). Further, plugging G^* and F^* into (2) yields a linear fee. Convergence of the fee to linearity is illustrated in Figure 3, which is based on the numerical example from Figure 2.

Equation (5) lends itself well to an economic interpretation and comparative statistics. It is useful to introduce the concept of shifted elasticities of supply $\bar{\eta}_s(c) = (c - \underline{c})g(c)/G(c) = \sigma$ and demand $\bar{\eta}_d(v) = (\bar{v} - v)f(v)/(1 - F(v)) = \beta$, which deals with

percentage changes of $c - \underline{c}$ rather than c and percentage changes of $\bar{v} - v$ rather than v . Observe that the elasticities of supply $\eta_s(c) = cg(c)/G(c) = \sigma c/(c - \underline{c})$ and demand $\eta_d(v) = |-vf(v)/(1 - F(v))| = \beta v/(\bar{v} - v)$ are proportional to the shifted elasticities σ and β , respectively.²⁷ For the sake of brevity, we will do comparative statics with respect to changes of shifted elasticities, which corresponds to overall changes of the standard elasticities. For $\alpha = 1$, the second term in brackets disappears and the first term in brackets represents the fee charged by a profit-maximizing intermediary. The fee is independent of the shifted elasticity of demand β . For $\alpha = 0$, the first term disappears and the second term represents the subsidy paid by a welfare-maximizing intermediary. For $\alpha = 0$, this term is independent of the shifted elasticity of supply σ . Note that for the highest price $\tilde{p} = 1$, the second term (subsidy) vanishes, whereas for the (hypothetical) lowest price $\tilde{p} = 0$, the first term (profit-maximizing fee) vanishes.

The overall fee $\tilde{\omega}$ and the slope of the fee $\tilde{\omega}'$ both decrease as the shifted elasticity of supply σ increases. This corresponds to the intuition that a higher elasticity leads to a lower markup for a standard monopolist. Further, the overall fee decreases as competition increases; that is, as α decreases. Interestingly, the slope $\tilde{\omega}'$ increases as α decreases if and only if $\beta > \sigma$. Thus, putting more weight on social welfare (or increasing competition) will increase the marginal fee $\tilde{\omega}'$ (and also lower the fixed component of the fee $\tilde{\omega}(0)$), provided demand is more elastic than supply.

An even more surprising result is that the overall fee increases as the shifted elasticity of demand β increases. This seems to contradict the standard notion that more rent will be extracted for less elastically demanded goods. The intuition is that marginal revenue, which is what the seller cares about, is more responsive to a change in the elasticity of demand than is a weighted average of marginal revenue and marginal social welfare, which is what the intermediary cares about. The intermediary lowers the fee to counteract the

²⁷The interpretation of elasticities follows the same logic as in footnote 12: probabilities have to be interpreted as expected quantities. The expected quantity demanded is $1 - F(p)$, which implies the elasticity of demand

$$\eta_d(p) = \left| \frac{d(1 - F(p))/(1 - F(p))}{dp/p} \right| = \left| -\frac{pf(p)}{1 - F(p)} \right|$$

and similarly for supply. For the shifted elasticities, replace dp/p by $d(\bar{v} - p)/(\bar{v} - p)$ and $d(p - \underline{c})/(p - \underline{c})$, respectively.

seller's otherwise excessive price increase if the elasticity of demand decreases.²⁸

We now consider the fee ω that is not normalized, denoting the limits of the boundaries as $\underline{c} := \lim_{k \rightarrow \infty} \underline{c}_k$ and $\bar{v} := \lim_{k \rightarrow \infty} \bar{v}_k$. When doing so, one has to distinguish between the case (i) $\underline{c} = \bar{v}$, for which the length of the relevant range goes to zero as in example 1, and the case (ii) $\underline{c} < \bar{v}$, for which the length does not go to zero as in example 2.

We begin with case (ii), $\underline{c} < \bar{v}$. It is convenient to define an upper bound of the seller's support \bar{c} and a lower bound of the buyer's support \underline{v} , satisfying $\bar{c} \geq \bar{v}$ and $\underline{v} \leq \underline{c}$. The non-normalized truncated distributions are $G^*(\hat{c}) = [(\hat{c} - \underline{c})/(\bar{c} - \underline{c})]^\sigma$ and $F^*(\hat{v}) = 1 - [(\bar{v} - \hat{v})/(\bar{v} - \underline{v})]^\beta$ in the limit. The corresponding weighted virtual type functions are $\Gamma_\alpha^*(\hat{c}) = \hat{c} + (\hat{c} - \underline{c})\alpha/\sigma$ and $\Phi_\alpha^*(\hat{v}) = \hat{v} - (\bar{v} - \hat{v})\alpha/\beta$. The non-normalized fee can be written as

$$\omega_\alpha(p) = \left[\alpha(p - \underline{c}) - (1 - \alpha)\sigma \frac{(\bar{v} - p)}{\beta + 1} \right] \frac{1}{\alpha + \sigma} = p - \Gamma_\alpha^{*-1}(\alpha p + (1 - \alpha)\Phi_\alpha^{*-1}(p)), \quad (6)$$

where the first equality stems from (5) and the definitions of the normalized variables and the second equality can be verified by plugging in the functional forms for Γ_α^* and Φ_α^* .

There is an intuitive interpretation of the right-hand side of (6) for $\alpha = 1$ and $\alpha = 0$. For a purely revenue maximizing intermediary, the fee becomes $\omega_1(p) = p - \Gamma^{*-1}(p)$. Note that $\Gamma^{*-1}(x)$ is the price that would be set by a monopsonist with valuation x (which equals the reserve price in a procurement auction). That is, the intermediary acts as if he ignored that the fee would affect the seller's price setting behavior: if the gross price p charged to the buyer were exogenously given, a profit maximizing intermediary would offer the net price $\Gamma^{*-1}(p)$ to a seller whose distribution is G^* . For a purely welfare maximizing intermediary, the fee is $\omega_0(p) = p - \Phi^{*-1}(p)$. Observe that $\Phi^{*-1}(x)$ is the price that would be set by a monopolist with cost x (which equals the reserve price in a standard auction). This means that, in the case of a welfare-maximizing intermediary, the

²⁸More formally, consider the second term in brackets in (5). This term represents the intermediary's subsidy that counteracts the seller's markup. A higher elasticity of demand means a higher markup by the seller which means a higher subsidy (that is, a lower fee) by the intermediary. However, the slope $\tilde{\omega}'$ increases with β .

seller will set a gross price equal to his cost $\Phi_\alpha^{*-1}(\Gamma_\alpha^*(c))|_{\alpha=0} = c$ and the intermediary will pay the seller the markup $\Phi^{*-1}(c) - c = -\omega(c)$ that a profit maximizing seller (without an intermediary) would have charged as a subsidy. This sheds light on why the fee increases with β : the intermediary compensates the seller for the markup he could earn; this markup decreases as the elasticity of demand increases.

For case (i) (that is, $\underline{c} = \bar{v}$), there are two ways of interpreting non-normalized values, such as p , $\omega(\cdot)$, \underline{v}_k , and \bar{c}_k . First, in the spirit of how asymptotic results are interpreted in applied statistics, linear fees are seen as a good (enough) approximation of optimal fees, even outside the limit, provided that k is sufficiently large (which means $\bar{v}_k - \underline{c}_k$ sufficiently small). With this interpretation in mind, the same reasoning holds as for $\underline{c} < \bar{v}$, except that \underline{c} and \bar{v} are replaced with \underline{c}_k and \bar{v}_k , respectively. Second, set $\bar{v} = \underline{c} =: p^*$, so that the fee becomes

$$\omega(p) = (p - p^*) \left(\alpha \left[\frac{1}{\alpha + \sigma} \right] + (1 - \alpha) \left[\frac{\sigma}{(\alpha + \sigma)(\beta + 1)} \right] \right). \quad (7)$$

The fee is zero at $p = p^*$ and the slope is determined by the expression after $(p - p^*)$, whose interpretation is similar to the one above.

It is easy to extend these results to a setup in which the seller's opportunity cost is due to the maximum of m price offers from outside the intermediated market, each offer c_i drawn from G_i . Similarly, the buyer's valuation is the minimum of n price offers from outside, each v_j drawn from F_j . The seller's cost $c = \max_{i=1}^m c_i$ and the buyer's valuation $v = \min_{j=1}^n v_j$ are then drawn approximately from the Pareto distributions $G(c) = c^{m\tilde{\sigma}}$ and $F(v) = 1 - (1 - v)^{n\tilde{\beta}}$, where $\tilde{\sigma} = (\sum_i \sigma_i)/m$ and $\tilde{\beta} = (\sum_j \beta_j)/n$ are the average elasticities of the G_i s and F_j s.²⁹ Little variation of m , $\tilde{\sigma}$, n , $\tilde{\beta}$ over time would imply little variation of the elasticities (or Pareto indices) $\sigma = m\tilde{\sigma}$ and $\beta = n\tilde{\beta}$ and hence little variation in the fees. Note that for a variety of topics such as income distributions and city sizes, the empirical literature finds Pareto indices to have remarkably little variation.

To summarize, the driving force behind the fees is the level of elasticity of demand and supply. For (mirrored) Generalized Pareto distributions G^* and F^* , the shifted elasticities $\underline{\eta}_s(c) = \sigma$ and $\bar{\eta}_d(v) = \beta$ are constant. For non-Pareto distributions, an

²⁹ c has the exact distribution $G(c) = \prod_i G_i(c)$ and v the exact distribution $F(v) = 1 - \prod_j (1 - F_j(v))$.

additional effect arises because the shifted elasticities $\underline{\eta}_s(c)$ and $\overline{\eta}_d(v)$ change with v and c over the relevant support $[\underline{c}_k, \overline{v}_k]$. As $u_k^S, u_k^B \rightarrow 0$, this additional effect vanishes and the level of the shifted elasticities becomes the only driving force.

The above results also have an interesting implication for empirical analysis. If F and G are approximately Pareto, then the pricing function $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ will be approximately linear, which means that the empirically observed prices p will approximately follow a mirrored Generalized Pareto distribution $G_p(p)$ that is a linear transformation of G : $G_p(p) = [(p - \underline{p})/(\overline{p} - \underline{p})]^\sigma$, where \underline{p} and \overline{p} can be computed from the pricing function. The applied literature on extreme value theory describes several ways of estimating the shape parameter σ of a Generalized Pareto distribution (see Coles (2001)). Given an estimate of the shape parameter σ , one can determine the fee that a monopolist (or colluding oligopolists) would charge, $\omega_1(p) = (p - \underline{c})/(\sigma + 1)$, which simplifies to $\omega_1(p) = p/(\sigma + 1)$ if one assumes $\underline{c} = 0$. Alternatively, if more data is available, \underline{c} and the parameters of F can be estimated.³⁰ With an estimate of F , one can also compute $\omega_\alpha(p)$ for arbitrary α and compare it with the fee observed in the data.

4.3 Results for General Distributions

We now show that, with some additional qualifications, the results for (mirrored) Generalized Pareto distributions extend to distributions outside this class. We use the following definitions to derive convex lower bounds and concave upper bounds for the optimal fee $\omega_\alpha(p)$ for general distributions. Define the convex hull of a function $h(x)$, denoted as $\underline{h}(x)$, as the highest convex function that is weakly below the function h . Analogously, define the concave hull of the function $h(x)$, denoted as $\overline{h}(x)$, as the lowest concave function that is weakly above the function h .³¹ Note that $\underline{\Phi}_\alpha(v) = \alpha \underline{\Phi}(v) + (1 - \alpha)v$ and $\underline{\Gamma}_\alpha(c) = \alpha \underline{\Gamma}(c) + (1 - \alpha)c$, and that this is similar for $\overline{\Phi}_\alpha$ and $\overline{\Gamma}_\alpha(c)$.

³⁰If one has data on many transactions occurring through an intermediary, the probability that the good was sold for a certain price p identifies $F(p)$. In our extension of this setup to an auction, which we will discuss later, p is the reserve price set by the seller. The distribution of bids also provides information on F . Note that in practice, the price distribution may not be exactly Pareto, for example because quality adjusted prices are observed with noise due to imperfect measures of quality. See Loertscher and Niedermayer (2012a) for an empirical analysis that does not assume Pareto distributions.

³¹See Myerson (1981) for a more detailed description of convex hulls.

Lemma 1. $\omega_\alpha(p)$ is bounded from below and above by the following convex and concave functions:

$$p - \underline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)\underline{\Phi}^{-1}(p)) \leq \omega_\alpha(p) \leq p - \overline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)\overline{\Phi}^{-1}(p)). \quad (8)$$

Inequality (8) is the analog of (6) for Pareto distributions. The following three propositions provide extensions of the results for Pareto distributions reported in Proposition 3 to general distributions. We begin with the effect of the elasticity of supply.

Proposition 4. (i) ω_0 does not depend on G (and hence on the elasticity of supply), and $\omega_0(p) = p - E_v[v|v \geq p]$. The fee $\omega_0(p)$ has the following lower bound and upper bound:

$$p - \underline{\Phi}^{-1}(p) \leq \omega_0(p) \leq p - \overline{\Phi}^{-1}(p). \quad (9)$$

(ii) For $\alpha > 0$, ω_α decreases with the elasticity of supply, formally, for two distributions \hat{G} and G with $\hat{\eta}_s(c) > \eta_s(c)$ for all c , $\hat{\omega}_\alpha(p) \leq \omega_\alpha(p)$ for all p , keeping F and α the same.

We will provide an intuitive explanation for the boundaries in (9) after deriving analogous results for the elasticity of demand.

Proposition 5. (i) For $\alpha = 1$, the lower and upper bound

$$p - \underline{\Gamma}^{-1}(p) \leq \omega_1(p) \leq p - \overline{\Gamma}^{-1}(p), \quad (10)$$

do not depend on F (and hence do not depend on the elasticity of demand). For G mirrored Generalized Pareto, the bounds are tight and the fee $\omega_1(p) = p - \Gamma^{-1}(p)$ is independent of the distribution F , which is not restricted to be Pareto.

(ii) For $\alpha < 1$, the lower and upper bounds

$$p - \underline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)E_v[v|v \geq p]) \leq \omega_\alpha(p) \leq p - \overline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)E_v[v|v \geq p])$$

increase as the elasticity of demand increases. Formally, let $\hat{v} \sim \hat{F}$ and $v \sim F$ with $\hat{\eta}_d(p) > \eta_d(p)$ for all p ,

$$p - \underline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)E_{\hat{v}}[\hat{v}|\hat{v} \geq p]) > p - \underline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)E_v[v|v \geq p]),$$

and likewise for the upper bound, keeping G and α constant. For G mirrored Generalized Pareto, the fee increases with the elasticity of demand. That is, $\hat{\omega}_\alpha(p) > \omega_\alpha(p)$ for all p .

Recall that $p - \Phi^{-1}(p)$ is the negative markup of a monopolist with cost p . Hence (9) can be interpreted the following way. For Φ convex – which corresponds to a decreasing monopoly pass-through rate³² – the subsidy paid by the intermediary $\omega_0(p)$ is larger than the negative markup of a monopolist $p - \Phi^{-1}(p)$ with cost p . The opposite holds for Φ concave. The interpretation of (10) is that for convex Γ , the fee charged by the intermediary $\omega_1(p)$ is larger than the markup of a monopsonist with valuation p .³³ The opposite holds for Γ concave.

We next consider the effect of an increase in competitiveness (that is, a decrease of α) on the optimal fee ω_α .

Proposition 6. (i) *The overall fee decreases as the market becomes more competitive. That is, for $\hat{\alpha} < \alpha$, $\omega_{\hat{\alpha}}(p) \leq \omega_\alpha(p)$ for all p .*

(ii) *The marginal fee increases as the market becomes more competitive if the slope of the inverse pricing function increases with competition. The marginal fee decreases with competition if the slope decreases. The marginal fee does not change if competition does not change the slope. Formally, for $\hat{\alpha} < \alpha$, $\omega'_{\hat{\alpha}}(p) \geq \omega'_\alpha(p)$ for all p if $\partial^2 P_\alpha^{-1}(v)/\partial\alpha\partial v < 0$ for all v . $\omega'_{\hat{\alpha}}(p) \leq \omega'_\alpha(p)$ for all p if $\partial^2 P_\alpha^{-1}(v)/\partial\alpha\partial v > 0$ and $\omega'_{\hat{\alpha}}(p) = \omega'_\alpha(p)$ for all p if $\partial^2 P_\alpha^{-1}(v)/\partial\alpha\partial v = 0$.*

Note that $\partial^2 \Gamma_\alpha^{-1}(\Phi_\alpha(v))/\partial v \partial \alpha < 0$ for all v iff $\partial^2 \Phi_\alpha^{-1}(\Gamma_\alpha(c))/\partial c \partial \alpha > 0$ for all c . $\partial \Phi_\alpha^{-1}(\Gamma_\alpha(c))/\partial c$ can be interpreted as the gross pass-through rate, that is how much the gross price (including fees) changes as the seller's costs increase. Hence, the statement in the proposition can be restated as: the marginal fee increases as the market becomes more competitive if the gross pass-through rate increases with competition. For Pareto distributions, the condition simplifies to $\beta > \sigma$.

The comparative statics of the gross price $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ are as follows.

Proposition 7. *The gross price $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ decreases (i) as demand becomes more elastic, (ii) as supply become more elastic, and (iii) as the market becomes more competitive.*

³² Φ convex means $\underline{\Phi}^{-1} = \Phi^{-1}$. Therefore, the monopoly price is a concave function of costs, or equivalently, the pass-through rate of a monopolist $d\Phi^{-1}(c)/dc$ is decreasing in c .

³³ Γ convex means that the monopsony price is a concave function of valuation, or the pass-through rate of a monopsonist is decreasing.

4.4 Applications

Value-Added Tax Consider a seller whose costs consist of the publicly known input price c_0 and a privately observed component c_1 drawn from the distribution G with support $[0, \bar{c}]$ and a government that maximizes revenue; that is, $\alpha = 1$. The total cost $c = c_0 + c_1$ has support $[c_0, c_0 + \bar{c}]$. If the distribution G can be expected to be well approximated by a mirrored Generalized Pareto distribution, the optimal fee is close to $\omega(p) = (p - c_0)/(1 + \sigma)$, which corresponds to a value-added tax levied on the difference of input and output prices $p - c_0$.

Unit Tax We have so far only considered (mirrored) Generalized Pareto distributions with a bounded support. However, results extend (with some modification) to the infinite boundary case. An infinite upper bound special case of the Generalized Pareto distribution is the exponential distribution $F(v) = 1 - e^{-\beta(v-\underline{v})}$ with support $[\underline{v}, \infty)$ and for mirrored Generalized Pareto distributions $G(c) = 1 - e^{-\sigma(\bar{c}-c)}$ with support $(-\infty, \bar{c}]$.³⁴ See Appendix B for more details on Generalized Pareto distributions. Plugging the corresponding virtual valuation and cost functions, $\Phi_\alpha(v) = v - \alpha/\beta$ and $\Gamma_\alpha(c) = c + \alpha/\sigma$, into (2) yields the (constant) unit tax $\omega(p) = \alpha/\sigma - (1 - \alpha)/\beta$.

Percentage Fees and Taxes Consider a setup in which the initial distributions of valuations and costs have support $[0, \infty)$. If sellers below a threshold \bar{c} and buyers above a threshold \underline{v} enter, the supports of G and F will be $[0, \bar{c}]$ and $[\underline{v}, \infty)$, respectively. The asymptotic Generalized Pareto distribution for the seller is then $G(c) = (c/\bar{c})^\sigma$, which implies $\eta_s(c) = \sigma$. For support $[\underline{v}, \infty)$, the buyer's asymptotic Generalized Pareto distribution is either an exponential, a Type I Pareto, or a Type II Pareto distribution (see Appendix B). For the classical Type I Pareto distribution $F(v) = 1 - (v/\underline{v})^\beta$ with

³⁴For expositional clarity, we have stated our results for distributions with finite boundaries. However, similar results for infinite boundaries hold both for the optimal fee and for asymptotic results in extreme value theory. The explanation for the lower truncation point \underline{v} for buyers and upper truncation point \bar{c} for sellers has to be slightly different, though. For finite boundary distributions, the simple assumption of transaction costs is sufficient as an explanation. For infinite boundary distributions, one has to assume *participation* costs, which are incurred even if there is no transaction. As an example, a seller with cost c is not willing to enter if his expected profit without fees $\max_p (p - c)(1 - F(p))$ is less than his participation cost. See also Appendix C for $\underline{v} \neq \underline{c}$, $\bar{v} \neq \bar{c}$ and for participation fees.

$\beta < -1$ (which implies $\eta_d(v) = -\beta$), the fee is proportional to the price, given by (7) with $p^* = 0$. For the Type II Pareto distribution $F(v) = 1 - (1 + (v - \underline{v})/\phi)^\beta$, with $\beta < -1$ and $\phi > 0$, the fee is linear and the fixed component positive if and only if $\underline{v} > \phi$.³⁵

Taxation of Multiple Markets Suppose there are two markets $m = 1, 2$, with one buyer and one seller each. The seller's cost is uniformly distributed on $[0, 1]$ in both markets. The buyer's valuation is distributed $F_m(v) = 1 - (1 - v)^{\beta_m}$ and support $[0, 1]$ in $m = 1, 2$. Assume $\beta_1 = 1$ and $\beta_2 = 2$; that is, demand is more elastic in the second market. Consider an intermediary that maximizes the sum of welfare in the two markets $W^1 + W^2$ subject to an overall balanced budget $R^1 + R^2 \geq 0$. Maximizing the Lagrangian $W^1 + W^2 + \lambda(R^1 + R^2)$ yields $\alpha^* = \lambda^*/(1 + \lambda^*) \approx 0.345$, which implies fees $\omega_{\alpha^*}^1(p) \approx 0.5p - 0.24$ and $\omega_{\alpha^*}^2(p) \approx 0.42p - 0.16$ and revenues $R_{\alpha^*}^1 \approx 0.0022$ and $R_{\alpha^*}^2 \approx -0.0022$. Revenues collected in the less elastic first market are used to subsidize the more elastic second market, as conventional wisdom would suggest. However, contrary to conventional wisdom, fees are higher in the second market for all $p \in [0, 1]$.

The interpretation of the asymptotic boundaries of support is somewhat subtle. For instance, the distribution $G(c) = c^{15}$ with support $[0, 1]$ implies $\Gamma(c) = (16/15)c$ and fees $\omega_1(p) = (p - \underline{c})/(\sigma + 1) \approx 0.06p$. The lower bound of the support $\underline{c} = 0$ is necessary to obtain exact optimality of percentage fees without a fixed component. However, *approximate* optimality also holds for positive lower bounds. Take for example a distribution with support $[7/15, 1]$ and the piecewise linear virtual cost function $\hat{\Gamma}(c) = 2c - 7/15$ for $c \in [7/15, 1/2]$ and $\hat{\Gamma}(c) = (16/15)c$ for $c \in [1/2, 1]$.³⁶ While the virtual cost for this distribution differs from the one for $G(c) = c^{15}$ for $c < 1/2$, the mass of sellers below $1/2$ is negligible: $G(1/2) = 2^{-15} < 0.01$ percent. Therefore, a linear approximation $\Gamma(c) = (16/15)c$ (corresponding to $G(c) = c^{15}$ or $\underline{\eta}_s(c) = 15$) and the implied fee $\omega_1(p) \approx 0.06p$ are close to optimal, even if $\underline{c} = 7/15$. A similar reasoning holds for \bar{v} .

³⁵For Type II Pareto with $\underline{v} < \phi$ and for exponential distributions, the fixed component of the fee is negative. Results can be obtained by using the virtual valuation functions $\Phi_\alpha(v) = v - \alpha/\beta$ (exponential) $\Phi_\alpha(v) = v(1 + \alpha/\beta)$ (Type I Pareto), and $\Phi_\alpha(v) = v(1 + \alpha/\beta) - (\underline{v} - \phi)\alpha/\beta$ (Type II Pareto).

³⁶The corresponding distribution is $\hat{G}(c) = (15c - 7)2^{-14}$ for $c \in [7/15, 1/2]$ and $\hat{G}(c) = c^{15}$ for $c \in [1/2, 1]$.

5 Multilateral Trade

5.1 Auctions

One of the most common forms of multilateral trade observed in practice are auctions with 1 seller and B buyers, which are run by an auctioneer who charges a commission fee. This provides a motivation to consider the following mechanism, which adds fees to an otherwise standard second-price auction. The seller sets a reserve price p . Buyers submit sealed bids. If at least two buyers submit a bid above the reserve price, the highest bidder receives the good for a gross transaction price \hat{p} which is equal to the second highest bid. If exactly one buyer bids above the reserve price, he gets the good and pays the reserve price as the gross transaction price \hat{p} . If no one bids above the reserve price, there is no transaction. In the case of a transaction, the intermediary charges a fee $\omega_\alpha(\hat{p})$ and the seller receives the net transaction price $\hat{p} - \omega_\alpha(\hat{p})$. Call this mechanism a second-price auction with fees ω_α .

Proposition 8. *A second-price auction with fees ω_α and an arbitrary number B of buyers, where*

$$\omega_\alpha(\hat{p}) = \hat{p} - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v)) | v \geq \hat{p}], \quad (11)$$

induces the seller to set the reserve price $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ and implements the α -allocation rule.

Note that the fee and the reserve price are the same as for bilateral trade. Since the fee is the same as for the one-buyer case, Propositions 3 and 8 imply that as the fraction of the most motivated traders that have a positive probability of trade goes to zero, fees converge to linear fees as stated in (6). All of the comparative statics of the optimal fee will be the same as for bilateral trade, irrespective of whether the optimal fee is linear.

A natural question to ask is whether revenue equivalence between first- and second-price auctions carries over to auctions with fees, where the rules of a first-price auction with fees are the same as those of a second-price auction with fees, except that the winner pays the highest bid. The following proposition addresses this question.

Proposition 9. (i) Assume there are $B > 1$ buyers and let $\omega(\hat{p})$ be the non-linear fee for both the first-price and the second-price auction with fees. The two auction formats with fees will (a) induce the seller to set different reserve prices and (b) lead to different expected revenues for the intermediary.

(ii) For a linear fee $\omega(\hat{p}) = a\hat{p} + b$, a first-price and a second-price auction with fees will lead to the same reserve price and the same revenues.

An intuition for this result is that the net price $\hat{p} - \omega(\hat{p})$ can be seen as the seller's von Neumann-Morgenstern utility function, so that linear fees correspond to auctions with a risk-neutral seller, while non-linear fees correspond to non-risk-neutrality.

It is worthwhile mentioning that while a first-price and a second-price auction generate different revenues for a fixed (non-linear) fee function, the same revenue can be achieved in both auction formats if the fee functions differ for the two auction formats. For a first-price auction, the fee function has to depend on the number of buyers participating. Interestingly, real world auctioneers that charge transaction fees like Sotheby's, Christie's, and eBay use second-price auctions or the strategically equivalent English auction. Sotheby's and Christie's used a fixed 20 percent fee before being caught for colluding, eBay uses a concave piecewise linear fee function.³⁷

5.2 Slotting Allowances

Real world retailers often require upfront payments by sellers to allocate scarce shelf space and then charge a fee on the revenue generated by the seller. This procedure has become known as *slotting allowances* (see, for example, Marx and Schaffer, 2007, 2010). We now show that slotting allowances can be interpreted as a form of fee-setting mechanism.

We model slotting allowances as a two-stage procedure with $B = 1$ buyers and $S \geq 1$ sellers. In the first stage, the sellers participate in a second-price auction, bidding for the right to serve the buyer. The winner pays the second highest bid, provided this bid

³⁷Note that in terms of our theory, only the sum of the fee (also called commission and premium in the context of auctions) paid by the buyer and the seller matters. The decomposition in the buyer's and the seller's fee is not important.

is positive, and 0 otherwise. If no seller's bid is positive, the game ends with payoffs of 0 for every player, including the intermediary. In the second stage, the winner of the first stage sets a take-it-or-leave-it price offer, facing the fee $\omega(p)$ the intermediary announces at the outset. The fee can also be interpreted as the markup of the retailer.

Proposition 10. *Any α -allocation rule can be implemented using slotting allowances with the fee $\omega_\alpha(p)$.*

For any F and G , $\omega_\alpha(p)$ will be independent of the number of sellers S . Moreover, if F and G are (mirrored) Generalized Pareto distributions, then $\omega_\alpha(p)$ will be linear. However, the expected participation fee, which is determined endogenously via the second-price auction of the first stage, will increase in S .

5.3 Real Estate Brokers

While our baseline model already provides some insights into real estate brokerage, we now modify this setup slightly and consider a random matching extension of our multilateral trade setup, thereby essentially embedding intermediaries in a one-shot version of the Satterthwaite and Shneyerov (2007) model.

Assume that there is mass B of buyers and mass S of sellers and that each real estate broker has a local monopoly or that brokers collude, which corresponds to $\alpha = 1$. Each broker is matched with at most one seller.³⁸ Buyers choose which seller/broker pair to visit at random. The number of buyers that visit a seller can then be shown to follow a Poisson distribution with arrival rate $\zeta = B/S$.³⁹ The transaction price at any given broker is determined via a second-price auction with fees, which can be seen as a plausible approximation of the bargaining between the seller and buyers.

We know from our results on auctions with fees that the reserve price set by the seller and the fee charged by the intermediary are independent of the number of buyers B . Therefore, neither the reserve price nor the fee will depend on whether the number

³⁸Alternatively, one can assume that the goods in questions are so heterogeneous that the multiple sellers a broker may attract are from different markets.

³⁹Alternatively, if there is some heterogeneity and a buyer and a seller are a match only with probability ζ_0 , then it is $\zeta = \zeta_0 B/S$.

of buyers is deterministic or random, nor will they depend on the distributions of the number of buyers. Therefore, all of the results on auctions with fees carry over to real estate brokerage, as do the results on the convergence to linear fees and the analysis of how fees depend on the elasticities of demand and supply.

Proposition 11. *(i) For $\alpha = 1$ and a Poisson arrival rate ζ of buyers, the optimal fee is the same as in a second-price auction, $\omega(\hat{p}) = \hat{p} - E[\Gamma^{-1}(\Phi(v))|v \geq \hat{p}]$, which induces the seller to set the reserve price $\Phi^{-1}(\Gamma(c))$. Both the fee ω and the reserve $\Phi^{-1}(\Gamma(c))$ are independent of ζ .*

(ii) The distribution of the highest valuation is $\hat{F}(v) := e^{-\zeta(1-F(v))}$. For the reserve price p , the probability of selling is $1 - \hat{F}(p)$ and the expected transaction price conditional on trade is

$$\frac{p\zeta(1 - F(p))\hat{F}(p) + \int_p^{\bar{v}} y\zeta(1 - F(y))[\hat{F}(y)]'dy}{1 - \hat{F}(p)}. \quad (12)$$

Both the probability of selling and the expected transaction price conditional on trade are increasing in ζ .

This random matching model provides insights into a number of relevant questions in real estate brokerage that appear puzzling from a principal-agent perspective. First, consider the observation that real estate brokers in the U.S. almost always charge 6 percent of the transaction price as fees and that there is very little variation over time and across regions.⁴⁰ This invariance is especially surprising because during booms real estate prices can be twice as high as during busts (see Hsieh and Moretti, 2003). As the costs of brokerage seem unlikely to double during booms and drop to original levels after a boom, brokerage costs cannot explain why brokers' revenue is proportional to the average price. Conversely, our model can, provided only that one assumes that the cyclical variation of real estate prices is due to variations of the ratio of potential buyers B and potential sellers S , so that during booms the Poisson arrival rate of buyers $\zeta = B/S$ is higher than during busts. A higher arrival rate of buyers leads to a higher expected transaction price \hat{p} in the auction but does not affect the fee function $\omega(\hat{p})$.

⁴⁰This observation has been made by the Department of Justice (2007) and by Hsieh and Moretti (2003).

Moreover, real estate markets are typically thin, so that G is well approximated by a mirrored Generalized Pareto distribution. This implies that for $\alpha = 1$, a change of the buyer's distribution F would have no effect on the fee.

If we maintain the assumption that variations in ζ are the cause of price differences in booms and busts, then Proposition 11 (ii) is also consistent with the widely made observation that during booms prices are high and houses sell easily and during recessions prices are low and the probability of a house selling is low. It is also consistent with the empirical observation that in cross-sectional data, houses offered at higher prices are more difficult to sell.⁴¹

Next consider the following empirical observation, made independently by Rutherford, Springer, and Yavas (2005), Levitt and Syverson (2008), and Hendel, Nevo, and Ortalo-Magné (2009): after controlling for the quality of the property, direct sellers – that is, owners who sell without a broker – receive on average the same or even higher gross prices than do indirect sellers – that is, owners who sell through a broker.⁴² We now show how a simple extension of our model can shed light on this observation. For simplicity, assume that some sellers have prohibitively high costs of selling directly, either because of lack of time or lack of searching and bargaining skills. These sellers become indirect sellers. Other sellers can sell directly without any additional costs, either because they are brokers selling their own homes or because they have the time and ability to sell their property over a for-sale-by-owner platform efficiently. These sellers become direct sellers. To allow for a comparison, assume further that the distribution of costs is the same for potential direct and indirect sellers and that sellers that trade with probability zero do not bother to enter the market and are hence not observable. Due to this endogenous entry by sellers, there are two countervailing effects: fees exclude some of the highest cost indirect sellers who would have entered were they direct sellers, and an indirect seller with the same cost will set a higher price than will a direct seller.

⁴¹See Loertscher and Niedermayer (2012a) for a discussion of these empirical observations.

⁴²In Hendel, Nevo, and Ortalo-Magné (2009) direct sellers are sellers selling through a for-sale-by-owner platform. In Rutherford, Springer, and Yavas (2005) and Levitt and Syverson (2008) direct sellers are brokers selling their own property through the Multiple Listing Service platform. In all papers, indirect seller are sellers selling through a broker through the Multiple Listing Service.

Proposition 12. *Assume F is Generalized Pareto, $\alpha = 1$ and that only sellers with a positive probability of trade enter.*

(a) *For G mirrored Generalized Pareto, the distributions of reserve prices for direct and indirect sellers are the same.*

(b) *Compared to the average reserve price of indirect sellers, the average reserve price of direct sellers is (i) equal if Γ is linear, (ii) higher if Γ is concave, and (iii) lower if Γ is convex.*

Proposition 12 (a) implies that the distribution of *transaction* prices (and not just reserve prices) will also be identical for direct and indirect prices in thin markets; that is, for Pareto distributions. An argument that is sometimes put forth to account for the observations by Rutherford, Springer, and Yavas (2005), Levitt and Syverson (2008), and Hendel, Nevo, and Ortalo-Magné (2009) is that there is an agency problem because the 6 percent fees give the brokers insufficient incentive to exert effort for a good price for the seller. However, this begs the question why such inefficient contracts are used in practice.⁴³ No such questions arise if the mechanisms employed by real estate brokers are viewed from an optimal pricing perspective, with the role of choosing the mechanisms assigned to the brokers, who are the long-run players in the market. From this angle, linear fee-setting mechanisms arise naturally as optimal mechanisms, and so does the prediction that there is no price difference among direct and indirect sellers.

A structural estimation based on an infinite horizon dynamic random matching extension of this setup (Loertscher and Niedermayer, 2012a) shows that linear and percentage fees are a good approximation of the optimal pricing mechanism.⁴⁴

⁴³For example, why do the seller and the broker not choose a contract with a steeper incentive scheme? It is also hard to explain why, when both a buyer's and the seller's broker are involved in a transaction, the commission is split between the two brokers, giving each 3 percent, as a percentage fee paid to the buyer's broker gives the broker an incentive to act *against* the buyer's interest.

⁴⁴In a previous version of the current paper (Loertscher and Niedermayer, 2012b), we also considered an extension of our (somewhat modified) setup to a dynamic random matching model. Dynamics adds too much complexity, so that an analytical solution is not possible. However, numerical calculations suggest that results carry over to the more complex dynamic setup.

5.4 Limit Results and Price Posting

Fee setting with many buyers and sellers For the $\langle B, S \rangle$ -setup, we now construct a generalized fee-setting mechanism with uniform pricing for buyers and sellers, which specializes to the fee-setting mechanism for a second-price auction for $S = 1$. For simplicity, we confine attention to allocation rules that are strictly monotone in the sense that $q_b(v_v)$ and $q_s(c_s)$ are strictly monotone whenever they are neither 0 nor 1. Let $P(c)$ be the monotone pricing function that will implement this allocation rule if, in equilibrium, all sellers set prices according to $P(c)$ and if all buyers bid their values.

This generalized fee-setting mechanism works as follows. The intermediary first announces a fee function $\omega(p)$, which is levied on the uniform transaction price p . Observing their own cost c_s and the fee $\omega(\cdot)$, all sellers $s \in S$ set prices p_s simultaneously, and all buyers submit their valuations simultaneously.⁴⁵ Let w be the Walrasian quantity defined with respect to submitted bids by buyers and prices set by sellers. After relabeling buyers in decreasing order and sellers in increasing order, the uniform transaction price is $p = \max\{v_{w+1}, p_w\}$, where p_w is the w -th lowest price set by a seller and v_{w+1} is the $(w + 1)$ th highest bid by a buyer. Notice that buyers will have a weakly dominant strategy to bid truthfully, which is already reflected in this notation. If $w \geq B$, let $v_{w+1} = \underline{v}$. The sets of buyers and sellers who trade are given by the Walrasian sets. All buyers who trade pay p while all other buyers pay 0 and all sellers who trade receive $p - \omega(p)$ while all other sellers receive 0. All that remains is to derive the fee function $\omega(p)$ that induces sellers of type c to set the price $P(c)$ for $c \in [\underline{c}, P^{-1}(\bar{v})]$ and an arbitrary price no less than \bar{v} for any $c > P^{-1}(\bar{v})$.

Let $q(p) := q_s(P^{-1}(p))$ be the probability that a seller who sets the price p trades under the fee-setting mechanism that implements the allocation rule $P(c)$ when setting the price p and when every other seller s sets the price according to $p_s = P(c_s)$ and every buyer plays his weakly dominant strategy. Further, define $a(p) := \int_1^p \frac{q'(y)}{q(y) - (1 - H(y))} dy$, where H is the distribution of the Walrasian price. The formal expressions for $q(p)$ and H are provided in the proof in the appendix.

⁴⁵It does not matter whether buyers observe the prices (p_1, \dots, p_S) before submitting their bids.

Proposition 13. *With B buyers and S sellers, the allocation rule $P(c)$ can be implemented via a generalized fee-setting mechanism with the fee*

$$\omega(p) = e^{a(\bar{v})-a(p)}(\bar{v} - P^{-1}(\bar{v})) - \int_p^{\bar{v}} e^{a(y)-a(p)}[1 + (y - P^{-1}(y))a'(y)]dy. \quad (13)$$

It is not hard to verify that for $B \geq S = 1$, (13) specializes to the by now familiar expression $\omega(p) = p - E_v[P^{-1}(v)|v \geq p]$.

Price Posting A mechanism that is widely used in practice, particularly in thick markets, and that is also useful for our theoretical analysis of fees, is price posting. Formally, the price posting mechanism is defined as follows. Let B and S be the finite numbers of buyers and sellers. The intermediary announces a buyer price p^B and a seller price p^S , which are such that every buyer who trades pays p^B and every seller who sells receives p^S . Let n_B and n_S be the number of buyers and sellers that are willing to trade at those prices. The number of trades that occur at the posted prices p^B and p^S is $n := \min\{n_B, n_S\}$, with agents on the long side being rationed randomly.

Our first result is a negative one:

Proposition 14. *Assume $B = S = 1$ and consider an intermediary who is restricted to set an ask price p^B and a bid price p^S . The weighted sum of welfare and profit under price posting is maximized by prices satisfying the equation system $\Phi_\alpha(p^B) = \alpha p^S + (1 - \alpha)E_c[c|c \leq p^S]$ and $\Gamma_\alpha(p^S) = \alpha p^B + (1 - \alpha)E_v[v|v \geq p^B]$. Price posting is not optimal.*

For two uniform distributions F and G with support $[0, 1]$, prices are $p^B = 1 - 2/(3(\alpha + 1))$ and $p^S = 2/(3(\alpha + 1))$. It is easy to verify, for example, that for $\alpha = 1$, the expected revenue with these posted prices is $1/27$, which is lower than the revenue of $1/24$ generated by a fee-setting mechanism with the fee ω_1 .

The reason price posting is not optimal is that it implements a different allocation than the optimal α -allocation: it does not allow for flexibility of the transaction price for different realizations of v and c .⁴⁶ By the same reasoning as for bilateral trade, price posting is not optimal for any finite number of buyers and sellers.

⁴⁶For example, the threshold for v above which a buyer trades, p^B , is independent of the realization of c , whereas for the optimal allocation rule, $P_\alpha(c)$ does depend on c .

However, as the number of buyers and sellers goes to infinity, price posting converges to optimality. To see this, consider a K -fold replica of the $\langle B, S \rangle$ economy, keeping the ratio $\zeta = B/S$ fixed. We have the following positive result.

Proposition 15. *Let prices be given by the system of equations $\Phi_\alpha(p^B) = \Gamma_\alpha(p^S)$ and $\zeta(1 - F(p^B)) = G(p^S)$. As K goes to infinity, (i) the optimal mechanism converges to a price-posting mechanism with these prices and (ii) the price-posting mechanism with these prices converges to optimality.*

For two uniform distributions F and G on $[0, 1]$ and $\zeta = 1$, prices are $p^B = 1 - 1/(2(1 + \alpha))$ and $p^S = 1/(2(1 + \alpha))$.

The driving force behind this result is that it becomes unnecessary to condition on the realized distribution of (\mathbf{v}, \mathbf{c}) , since uncertainty about the realized distribution vanishes in the limit. Note that this result also extends to a setup in which an intermediary meets one buyer and one seller at a time with an infinite time horizon and where the good can be stored without cost.⁴⁷ It is also interesting to note that intermediaries that use price posting – for example, used car dealers and currency exchange offices – either have large numbers of traders or store the good.

We now return to optimal fees for the $\langle B, S \rangle$ -setup, while letting B and S go to infinity, keeping ζ fixed. Here, the intermediary can charge the bid-ask spread given by price posting in Proposition 15 as a fee, $\omega(p^B) = p^B - p^S$. For prices $p \neq p^B$ the fee has to be such that a seller does not have an interest in setting a price different from p^B . This implies the same allocation rule and the same payments as for price posting, which we know is optimal for infinite B and S .

6 Conclusions

We study fee-setting mechanisms within a fairly general model of market making. Fee-setting mechanisms are used by real-estate brokers around the globe, by auction houses

⁴⁷Storeability of the good means that the intermediary can consider all buyers and sellers as present at the same time and by Proposition 15 price posting is optimal. This proposition is in line with the prevalence of price posting in thick markets and in markets with storable goods. The assumption that there is a large number of traders is particularly relevant for standardized or storable commodities.

and auction sites on the world wide web, and by governments in most developed countries to generate tax revenue via indirect taxation.

Our pricing based model ascribes the central role of choosing the trading mechanism to the intermediaries who make the market. The model's predictions are consistent with a number of observations in various markets including real-estate brokerage, auction houses, retailers with slotting allowances, and used car dealers. We show that linear fees are optimal in increasingly thin markets because thin markets are characterized by (mirrored) Generalized Pareto distributions on the demand and supply side. It is hoped that these results will prove fruitful for future applied work, given the host of results from statistics for this family of distributions.

References

- 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(1), Article 5.
- BALKEMA, A., AND L. DE HAAN (1974): "Residual Life Time at Great Age," *Annals of Probability*, 2(5), 792–804.
- BALLARD, C., J. SHOVEN, AND J. WHALLEY (1985): "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States," *American Economic Review*, 75(1), 128–138.
- BARVINEK, E., I. DALER, AND J. FRANCU (1991): "Convergence of Sequences of Inverse Functions," *Archivum Mathematicum*, 27(3-4), 201–204.
- BULOW, J., AND J. ROBERTS (1989): "The Simple Economics of Optimal Auctions," *Journal of Political Economics*, 97(5), 1060–90.
- CHATTERJEE, K., AND W. SAMUELSON (1983): "Bargaining under Asymmetric Information," *Operations Research*, 31, 835–851.
- CHU, L. Y., AND D. SAPPINGTON (2007): "Simple Cost-Sharing Contracts," *American Economic Review*, 97(1), 419–428.
- CHU, S., P. LESLIE, AND A. SORENSEN (2011): "Bundle-size Pricing as an Approximation to Mixed Bundling," *American Economic Review*, 101(1), 263–303.

- COLES, S. (2001): *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- DE HAAN, L., AND A. FERREIRA (2006): *Extreme Value Theory: An Introduction*. Springer Verlag.
- DEPARTMENT OF JUSTICE (2007): “Competition in the Real Estate Brokerage Industry,” *Report by the Federal Trade Commission and the U.S. Department of Justice*.
- FALK, M., J. HÜSLER, AND R. REISS (2010): *Laws of Small Numbers: Extremes and Rare Events*. Springer Verlag.
- GRESIK, T. (1991): “Ex Ante Efficient, Ex Post Individually Rational Trade,” *Journal of Economic Theory*, 53, 131–145.
- GRESIK, T., AND M. SATTERTHWAITE (1989): “The Rate at Which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms,” *Journal of Economic Theory*, 48, 304–332.
- HELLWIG, M. F. (2003): “Public-Good Provision with Many Participants,” *Review of Economic Studies*, 70(3), 589 – 614.
- HENDEL, I., A. NEVO, AND F. ORTALO-MAGNÉ (2009): “The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com,” *American Economic Review*, 99(5), 1878 – 98.
- HSIEH, C.-T., AND E. MORETTI (2003): “Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry,” *Journal of Political Economy*, 111(5), 1076–1122.
- HU, A., S. MATTHEWS, AND L. ZOU (2010): “Risk Aversion and Optimal Reserve Prices in First-and Second-Price Auctions,” *Journal of Economic Theory*, 145(3), 1188–1202.
- JULLIEN, B., AND T. MARIOTTI (2006): “Auction and the Informed Seller Problem,” *Games and Economic Behavior*, (56), 225–258.
- KRISHNA, V. (2002): *Auction Theory*. Elsevier Science, Academic Press.
- LEININGER, W., P. B. LINHART, AND R. RADNER (1989): “Equilibria of the Sealed-Bid Mechanism for Bargaining with Incomplete Information,” *Journal of Economic Theory*, 48(1), 63 – 106.
- LEVITT, S. D., AND C. SYVERSON (2008): “Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions,” *Review of Economics and Statistics*, 90(4), 599–611.
- LOERTSCHER, S., AND A. NIEDERMAYER (2012a): “Assessing the Performance of Simple Contracts Empirically: The Case of Percentage Fees,” *Working Paper*.

- (2012b): “Fee Setting Intermediaries: On Real Estate Agents, Stock Brokers, and Auction Houses,” *Working Paper*.
- MAKOWSKI, L., AND C. MEZZETTI (1993): “The Possibility of Efficient Mechanisms for Trading and Indivisible Object,” *Journal of Economic Theory*, 59(2), 451–465.
- MARX, L., AND G. SCHAFFER (2007): “Upfront Payments and Resale Price Maintenance: A Comparison of Facilitating Practices,” *RAND Journal of Economics*, 22, 120–135.
- (2010): “Slotting Allowances and Scarce Shelf Space,” *Journal of Economics & Management Strategy*, 19(3), 575–603.
- MATROS, A., AND A. ZAPECHELNYUK (2008): “Optimal Fees in Internet Auctions,” *Review of Economic Design*, 12(3), 155–63.
- MCAFEE, P. (2002): “Coarse Matching,” *Econometrica*, 70(5), 2025–34.
- MYERSON, R. (1981): “Optimal Auction Design,” *Mathematical Operations Research*, 6(1), 58–73.
- MYERSON, R., AND M. SATTERTHWAITE (1983): “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory*, 29(2), 265–281.
- PERRIGNE, I., AND Q. VUONG (2011): “Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard,” *Econometrica*, 79(5), 1499 – 1539.
- PICKANDS, J. (1975): “Statistical Inference Using Extreme Order Statistics,” *Annals of Statistics*, pp. 119–131.
- ROGERSON, W. (2003): “Simple Menus of Contracts in Cost-Based Procurement and Regulation,” *American Economic Review*, 93(3), 919–26.
- RUTHERFORD, R., T. SPRINGER, AND A. YAVAS (2005): “Conflicts between Principals and Agents: Evidence from Residential Brokerage,” *Journal of Financial Economics*, 76, 627–65.
- SALANIÉ, B. (2003): *The Economics of Taxation*. MIT Press.
- SATTERTHWAITE, M., AND A. SHNEYEROV (2007): “Dynamic Matching, Two-Sided Information, and Participation Costs: Existence and Convergence to Perfect Competition,” *Econometrica*, 75(1), 155–200.
- SHY, O., AND Z. WANG (2011): “Why Do Payment Card Networks Charge Proportional Fees?,” *American Economic Review*, 101(4), 1575–1590.
- SPULBER, D. F. (1999): *Market Microstructure: Intermediaries and the Theory of the Firm*. Cambridge University Press, Cambridge.
- WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press.

Online Appendix

A Proofs

Proof of Proposition 1. (a) We first show that for any incentive-compatible mechanism (\mathbf{Q}, \mathbf{T}) that satisfies interim individual rationality and gives zero utility to the most inefficient types, the revenue of the intermediary is

$$R = E_{\mathbf{v}, \mathbf{c}} \left[\sum_{b \in \mathbb{B}} \Phi(v_b) Q_b(\mathbf{v}, \mathbf{c}) - \sum_{s \in \mathbb{S}} \Gamma(c_s) Q_s(\mathbf{v}, \mathbf{c}) \right] - \sum_{b \in \mathbb{B}} U_b(\underline{v}) - \sum_{s \in \mathbb{S}} U_s(\bar{c}), \quad (14)$$

where $U_b(v_b) := v_b q_b(v_b) - t_b(v_b)$ and $U_s(c_s) := t_s(c_s) - c_s q_s(c_s)$, is a buyer's and a seller's utility if they report their types truthfully. By a standard envelope theorem argument, incentive compatibility implies

$$U_b(v_b) = U_b(\underline{v}) + \int_{\underline{v}}^{v_b} q_b(y) dy \quad \text{and} \quad U_s(c_s) = U_s(\bar{c}) + \int_{c_s}^{\bar{c}} q_s(y) dy,$$

see for example Krishna (2002, chapter 5). Thus, every agent's expected payoff at the interim stage (that is, after learning his type and expected payment and the probability conditional on this type but nothing more) is pinned down, up to a constant, which is $U_b(\underline{v})$ if he is a buyer and $U_s(\bar{c})$ if he is a seller, by the allocation rule. Since the integral terms are non-negative, interim individual rationality is satisfied if and only if $U_b(\underline{v}) \geq 0$ and $U_s(\bar{c}) \geq 0$. Using the definitions $U_b(v_b)$ and $U_s(c_s)$ and assuming that the least efficient types \underline{v} and \bar{c} never trade, the expected payment from a buyer of type v_b and the expected payment to a seller of type c_s are

$$\begin{aligned} t_b(v_b) &= t_b(\underline{v}) + v_b q_b(v_b) - \int_{\underline{v}}^{v_b} q_b(y) dy, \\ t_s(c_s) &= t_s(\bar{c}) + c_s q_s(c_s) - \int_{c_s}^{\bar{c}} q_s(y) dy. \end{aligned}$$

Changing the order of integration in the resulting double integral and making use of the assumption that types are independent, the *ex ante* expected transfer from a buyer b , denoted $E_{v_b}[t_b(v_b)]$, is

$$E_{v_b}[t_b(v_b)] = t_b(\underline{v}) + \int_{\underline{v}}^{\bar{v}} \Phi(y) q_b(y) f(y) dy,$$

where v_b inside $E_{v_b}[\cdot]$ denotes a random variable. Similarly, the *ex ante* expected transfer to a seller s , denoted $E[t_s(c_s)]$, can be expressed as

$$E_{c_s}[t_s(c_s)] = t_s(\bar{c}) + \int_{\underline{c}}^{\bar{c}} \Gamma(y) q_s(y) g(y) dy.$$

Therefore, *ex ante* expected revenue R of the mechanism (\mathbf{Q}, \mathbf{T}) is

$$R = E_{\mathbf{v}, \mathbf{c}} \left[\sum_{b \in \mathbb{B}} \Phi(v_b) Q_b(\mathbf{v}, \mathbf{c}) - \sum_{s \in \mathbb{S}} \Gamma(c_s) Q_s(\mathbf{v}, \mathbf{c}) \right] - \sum_{b \in \mathbb{B}} U_b(\underline{v}) - \sum_{s \in \mathbb{S}} U_s(\bar{c}). \quad (15)$$

Since $R = \sum E_{v_b}[t_b(v_b)] - \sum E_{c_s}[t_s(c_s)]$ and $U_b(\underline{v}) = t_b(\underline{v})$ and $U_s(\bar{c}) = t_s(\bar{c})$. The expression in the proposition follows from using (15) in $(1 - \alpha)W + \alpha R$.

(b) Individual rationality requires $U_b(\underline{v}) \geq 0$ and $U_s(\bar{c}) \geq 0$ for all b and s . $U_b(\underline{v}) = U_s(\bar{c}) = 0$ for all b and s ensure that this constraint is just satisfied. Further, since the α -allocation rule maximizes $\sum_b \Phi_\alpha(v_b) Q_b^\alpha(\mathbf{v}, \mathbf{c}) - \sum_s \Gamma_\alpha(c_s) Q_s^\alpha(\mathbf{v}, \mathbf{c})$ for every realization of (\mathbf{v}, \mathbf{c}) , it also maximizes its expected value $(1 - \alpha)W + \alpha R$. \square

Proof of Proposition 3. (i) Φ continuously differentiable implies

$$\lim_{v \rightarrow \bar{v}} \frac{d}{dv} \left[\frac{1 - F(v)}{f(v)} \right] = \lim_{v \rightarrow \bar{v}} \frac{d}{dv} [v - \Phi(v)] = \bar{\beta}, \quad (16)$$

for some constant $\bar{\beta}$. Equation (16) is the von Mises condition as stated in Theorem 2 in Appendix B. It implies that F is in the domain of attraction of an extreme value distribution (as defined in Definition 2) by Theorem 2. By the Pickands-Balkema-de Haan theorem (see Theorem 1), this in turn implies that F has a Generalized Pareto upper tail as defined in (31). This implies convergence of the normalized distribution \tilde{F}_k to $\tilde{F}^*(\tilde{v}) = 1 - (1 - \tilde{v})^\beta$, because of the definition of the normalized variable \tilde{v} . An analogous reasoning holds for the convergence of \tilde{G}_k .

(ii) Define

$$\bar{\beta} := \lim_{v \rightarrow \bar{v}} 1 - \Phi'(v) \quad \text{and} \quad \bar{\sigma} := \lim_{c \rightarrow \underline{c}} \Gamma'(c) - 1,$$

$\beta := -1/\bar{\beta}$, and $\sigma := 1/\bar{\sigma}$. Observe that by l'Hôpital's rule

$$\lim_{v \rightarrow \bar{v}} \frac{(\bar{v} - v)f(v)}{1 - F(v)} = \lim_{v \rightarrow \bar{v}} \frac{\bar{v} - v}{v - \Phi(v)} = \lim_{v \rightarrow \bar{v}} \frac{-1}{1 - \Phi'(v)} = \beta. \quad (17)$$

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 $[\underline{v}, \bar{v}]$ and $[\underline{c}, \bar{c}]$. The distributions are hence $F^*(v) = 1 - [(\bar{v} - v)/(\bar{v} - \underline{v})]^\beta$ and $G^*(c) = [(c - \underline{c})/(\bar{c} - \underline{c})]^\sigma$. Virtual type functions are linear, $\Phi_\alpha^*(v) = v - (\bar{v} - v)\alpha/\beta$ and $\Gamma_\alpha^*(c) = c + (c - \underline{c})\alpha/\sigma$. Fees given in (2) can be rearranged the following way.

$$\begin{aligned} \omega^*(p) &= p - E_v[\Gamma_\alpha^{*-1}(\Phi_\alpha^*(v))|v \geq p] \\ &= p - E_v[\Gamma_\alpha^{*-1}(\alpha\Phi^*(v) + (1 - \alpha)\Phi^{*-1}(\Phi(v)))|v \geq p] \\ &= p - \Gamma_\alpha^{*-1}(\alpha E_v[\Phi^*(v)|v \geq p] + (1 - \alpha)\Phi^{*-1}(E_v[\Phi^*(v)|v \geq p])) \\ &= p - \Gamma_\alpha^{*-1}(\alpha p + (1 - \alpha)\Phi^{*-1}(p)), \end{aligned}$$

where the first equality comes from the definition of $\Phi_\alpha(v) = \alpha\Phi(v) + (1 - \alpha)v = \alpha\Phi(v) + (1 - \alpha)\Phi^{*-1}(\Phi(v))$, the second equality from the linearity of Γ_α^* and Φ^* , and the third from (1). Plugging in the functional forms for Γ_α^* and Φ^* yields

$$\omega(p) = \alpha(p - \underline{c}) \left[\frac{1}{\alpha + \sigma} \right] - (1 - \alpha)(\bar{v} - p) \left[\frac{\sigma}{(\alpha + \sigma)(\beta + 1)} \right],$$

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

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

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}_k and \hat{G}_k , with

$\hat{G}_k(\hat{c}) = G(\hat{c})$ and $\hat{F}_k(\hat{v}) = F(\hat{v} + (1 - u))$. The transformed fee is

$$\begin{aligned}\hat{\omega}(\hat{p}) &= u\hat{p} + \frac{\int_{\hat{p}}^1 \hat{\Gamma}_{\alpha}^{-1}(\hat{\Phi}_{\alpha}(u\hat{v})) u \hat{f}(u\hat{v}) d\hat{v}}{1 - \hat{F}(u\hat{p})} \\ &= u\hat{p} + \left[\frac{1 - \hat{F}(u\hat{p})}{1 - \hat{F}(u)} \right]^{-1} \int_{\hat{p}}^1 \hat{\Gamma}_{\alpha}^{-1}(\hat{\Phi}_{\alpha}(u\hat{v})) \frac{1}{1 - \hat{v}} \frac{u(1 - \hat{v}) \hat{f}(u\hat{v})}{1 - \hat{F}(u\hat{v})} \frac{1 - \hat{F}(u\hat{v})}{1 - \hat{F}(u)} d\hat{v},\end{aligned}$$

where the first expression comes from plugging in $u\hat{p}$ for p in (2) and the second expression is obtained by multiplying by $(1 - \hat{F}(u))/(1 - \hat{F}(u))$, $(1 - \hat{F}(u\hat{v}))/ (1 - \hat{F}(u\hat{v}))$, and $(1 - \hat{v})/(1 - \hat{v})$.

By Theorem 1, we know that

$$\left[\frac{1 - \hat{F}(u\hat{p})}{1 - \hat{F}(u)} \right]^{-1} = \left[\frac{1 - F(1 - u(1 - \hat{p}))}{1 - F(1 - u)} \right]^{-1} \rightarrow (1 - \hat{p})^{-\beta},$$

as $u \rightarrow 0$.

As a next step, we need to show that the integrand uniformly converges to the expression given by the limiting distributions. Uniform convergence of the integrand then implies convergence of the integral.

Again, by Theorem 1

$$\left[\frac{1 - \hat{F}(u\hat{p})}{1 - \hat{F}(u)} \right] \xrightarrow{u \rightarrow 0} (1 - \hat{v})^{\beta},$$

where the double arrow \Rightarrow stands for uniform convergence.

$$\frac{u(1 - \hat{v}) \hat{f}(u\hat{v})}{1 - \hat{F}(u\hat{v})} = \frac{(1 - \hat{v}) u f(1 - u(1 - \hat{v}))}{1 - F(1 - u(1 - \hat{v}))} = \frac{(1 - v') f(v')}{1 - F(v')},$$

where $v' = 1 - u(1 - \hat{v})$. As $u \rightarrow 0$, $v' \rightarrow 1$ for arbitrary \hat{v} , which implies that the above expression goes to β by (17).

By the definition of β we have

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

This implies that

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

and hence

$$\frac{1}{u}\hat{\Phi}_\alpha(u\hat{v}) \xrightarrow{u \rightarrow 0} \hat{v} - \alpha \frac{1 - \hat{v}}{\beta}$$

By a similar logic

$$\frac{1}{u}\hat{\Gamma}_\alpha(u\hat{c}) \xrightarrow{u \rightarrow 0} \hat{c} \left(1 + \frac{\alpha}{\sigma}\right)$$

and hence

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

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

Since both $\hat{\Phi}_\alpha$ and $\hat{\Gamma}_\alpha^{-1}$ uniformly converge to the functions implied by the limiting distributions, so does their combination.

Taken together, one can see that the integrand uniformly converges to the function implied by the limiting distributions. This implies convergence of the integral, which in turn implies convergence of the fees to linear fees. \square

Proof of Lemma 1. First take the convex hulls. Since $\underline{\Gamma} \leq \Gamma$ and convex, we have $\underline{\Gamma}^{-1} \geq \Gamma$ and $\underline{\Gamma}^{-1}$ concave. The same reasoning applies to $\underline{\Gamma}_\alpha^{-1}$ and $\underline{\Phi}^{-1}$.

Therefore, $p - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p] = p - E_v[\Gamma_\alpha^{-1}(\alpha\Phi(v) + (1 - \alpha)\Phi^{-1}(\Phi(v))|v \geq p]$ is bounded from below by $p - E_v[\underline{\Gamma}_\alpha^{-1}(\alpha\Phi(v) + (1 - \alpha)\underline{\Phi}^{-1}(\Phi(v))|v \geq p]$. Now take the chain of equalities that led from $\omega(p)$ to $p - \Gamma_\alpha^{-1}(\alpha p + (1 - \alpha)\Phi^{-1}(p))$ above. Whenever the linearity of Φ and Γ is used to take the expectations into a function, use the concavity of $\underline{\Phi}^{-1}$ and $\underline{\Gamma}^{-1}$ and Jensen's inequality. This results in a chain of inequalities that shows that $\omega(p) \geq p - \underline{\Gamma}_\alpha^{-1}(\alpha p + (1 - \alpha)\underline{\Phi}^{-1}(p))$. The same reasoning applies to the concave hulls and the upper bound. \square

Proof of Proposition 4. For $\alpha = 0$, $\Phi_\alpha(v) = v$ and $\Gamma_\alpha(c) = c$. Hence

$$\omega_\alpha(p) = p - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p] = p - E_v[v|v \geq p].$$

Equation (9) is obtained by plugging $\alpha = 0$ into (8).

(ii) Note that

$$\Gamma_\alpha(c) = c \left(1 + \frac{\alpha}{\eta_s(c)}\right).$$

Therefore, $\hat{\eta}_s > \eta_s$ implies $\hat{\Gamma}_\alpha(c) < \Gamma_\alpha(c)$ for all c , which in turn implies $\hat{\Gamma}_\alpha^{-1}(x) > \Gamma_\alpha^{-1}(x)$ for all x . Thus,

$$\hat{\omega}_\alpha(p) = p - E_v[\hat{\Gamma}_\alpha^{-1}(\Phi_\alpha(v))|v \geq p] \leq p - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p] = \omega_\alpha(p)$$

□

Proof of Proposition 5. (i) Equation (10) is obtained by plugging $\alpha = 1$ into (8). For G mirrored Generalized Pareto, Γ is linear. Hence the lower bound and the upper bound coincide: $\omega_1(p) = p - \Gamma^{-1}(p)$.

(ii) By the reasoning in Lemma 1, ω_α has the lower and upper bounds

$$p - \underline{\Gamma}_\alpha(\alpha p + (1 - \alpha)E_v[v|v \geq p]) \leq \omega_\alpha(p) \leq p - \bar{\Gamma}_\alpha(\alpha p + (1 - \alpha)E_v[v|v \geq p]), \quad (18)$$

and similarly for \hat{v} and \hat{F} . $\hat{\eta}_d(p) > \eta_d(p)$ for all $p \in [\underline{v}, \bar{v}]$ implies

$$\frac{\hat{f}(p)}{1 - \hat{F}(p)} \geq \frac{f(p)}{1 - F(p)},$$

for all p , that is, hazard rate dominance. Hazard rate dominance implies $E_{\hat{v}}[\hat{v}|\hat{v} \geq p] \geq E_v[v|v \geq p]$ for all p . Combining this with (18) gives the first statement in the proposition. For G Pareto, the boundaries in (18) are tight, that is,

$$\omega_\alpha(p) = p - \Gamma_\alpha(\alpha p + (1 - \alpha)E_v[v|v \geq p]),$$

which implies the second statement in the proposition by the above reasoning. □

Proof of Proposition 6. (i) Since

$$\Phi_\alpha(v) = v - \alpha \frac{1 - F(v)}{f(v)}, \quad \Gamma_\alpha(c) = c + \alpha \frac{G(c)}{g(c)},$$

Φ_α decreases and Γ_α increases as α increases. Therefore, an increase of α decreases $\Gamma_\alpha^{-1}(\Phi_\alpha(v))$ for all v , and hence increases

$$\omega_\alpha(p) = p - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p]$$

for all p , which proves the statement.

(ii) Note that

$$\omega'_\alpha(p) = \frac{\partial}{\partial p} [p - E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p]]$$

can be rearranged to

$$\begin{aligned} \omega'_\alpha(p) &= 1 - \frac{f(p)}{1 - F(p)} [E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p] - \Gamma_\alpha^{-1}(\Phi_\alpha(p))] \\ &= 1 + \frac{f(p)}{1 - F(p)} [E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(p)) - \Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p]]. \end{aligned}$$

Therefore, $\partial\omega'_\alpha(p)/\partial\alpha$ has the same sign as

$$\frac{\partial E_v[\Gamma_\alpha^{-1}(\Phi_\alpha(p)) - \Gamma_\alpha^{-1}(\Phi_\alpha(v))|v \geq p]}{\partial\alpha},$$

which is positive if

$$\frac{\partial^2 \Gamma_\alpha^{-1}(\Phi_\alpha(v))}{\partial\alpha\partial v} < 0, \quad (19)$$

for all v . This and the analogous results for positive and zero cross-derivatives in (19) prove the statement. \square

Proof of Proposition 7. As η_d increases, $\Phi_\alpha(v) = v(1 - \alpha/\eta_d(v))$ increases. Therefore, $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ decreases. As η_s increases, $\Gamma_\alpha(c) = c(1 + \alpha/\eta_s(c))$ decreases and so does $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$. As α decreases, Φ_α increases and Γ_α decreases. Therefore, $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ decreases. \square

Proof of Proposition 8. By Proposition 1 we know that for one seller and B buyers the optimal allocation rule is to allow trade if and only if $\max_{b=1}^B \Phi_\alpha(v_b) \geq \Gamma_\alpha(c)$. The second-price auction with ω_α fees implements this allocation rule for the following reason. A buyer has the incentive to bid his own valuation by a standard (Vickrey) dominant strategies argument. The highest bid $\max_b v_b$ will be greater than or equal to the reserve price $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$ if and only if $\max_b \Phi_\alpha(v_b) \geq \Gamma_\alpha(c)$. Therefore, the auction implements the optimal allocation rule. It only remains to be shown that the seller has the incentive to set the reserve price $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$.

The profit of the seller in a second-price auction is

$$(p - \omega_\alpha(p) - c)(1 - F(p))BF(p)^{B-1} + \int_p^{\bar{v}} (y - \omega_\alpha(y) - c)(1 - F(y))B [F(y)^{B-1}]' dy,$$

where the first term is the seller's profit if the reserve price p is binding (that is, exactly one buyer is above the reserve price) and the second term – the integral – is the profit if the reserve price is not binding (that is, at least two buyers are above the reserve price). The profit is the same as in a standard auction (see, for example, Krishna (2002, p. 25)), except for the fee $\omega_\alpha(\cdot)$. The seller's first order condition with respect to p simplifies to

$$[(p - \omega_\alpha(p) - c)(1 - F(p))]' BF(p)^{B-1} = 0.$$

Since $BF(p)^{B-1}$ is always positive, the first-order condition simplifies further to $[(p - \omega_\alpha(p) - c)(1 - F(p))]' = 0$, which is the same as in the case with one buyer. We know that for the one-buyer case, the seller's optimal reserve price is $\Phi_\alpha^{-1}(\Gamma_\alpha(c))$. Further, since the mechanism gives zero utility to the most inefficient agents \bar{c} and \underline{v} , it is an optimal mechanism. \square

Proof of Proposition 9. (i) The net transaction price $\hat{p} - \omega(\hat{p})$ received by the seller can be reinterpreted as a von Neumann-Morgenstern utility function. Non-linear fees imply a non-linear net price, which corresponds to a non-risk-neutral seller. A non-risk-neutral seller will set a reserve price that depends on the number of buyers in a first-price auction (for a proof see Hu, Matthews, and Zou (2010)). Since we know that for a second-price auction the reserve price is independent of the number of buyers, the reserve price has to differ in a first-price auction compared to in a second-price auction. Different reserve prices lead to different revenues.

(ii) A linear net transaction price corresponds to a risk-neutral seller. It is a well-known result in auction theory that for risk-neutral sellers, a first-price and a second-price auction generate the same expected revenues (see Myerson (1981) or Krishna (2002)). The seller also sets the same reserve price in both auction formats. Further, since the fee is linear, the two auction formats generate the same revenue for the intermediary. \square

Proof of Proposition 10. We know that setting $p = P_\alpha(c)$ is the unique optimal price for any seller-type who trades with positive probability in the second stage. In the first stage, every seller has a dominant strategy to place a bid equal to his expected profit of participating in the second stage, given cost c and $\omega_\alpha(p)$. Since these profits are strictly

decreasing in costs, the seller with the lowest cost wins in the first stage. Thus, the mechanism implements the α -allocation rule. \square

Proof of Proposition 11. (i) This is a straightforward implication of Proposition 8: since the fee and the reserve price are independent of the number of buyers B , they are also independent of the random arrival rate of buyers ζ .

(ii) The probability mass function of the Poisson distributed number of buyers B is

$$\frac{e^{-\zeta}\zeta^B}{B!}.$$

Observe that for a fixed number of buyers B , the highest valuation follows the distribution $F(v)^B$. Taking expectations over B yields

$$\sum_{B=0}^{\infty} \frac{e^{-\zeta}\zeta^B}{B!} F(v)^B = e^{-\zeta(1-F(v))}. \quad (20)$$

Therefore, the probability of selling is $1 - e^{-\zeta(1-F(p))}$. Further, observe that

$$\sum_{B=1}^{\infty} \frac{e^{-\zeta}\zeta^B}{B!} BF(p)^{B-1} = \zeta \sum_{B=1}^{\infty} \frac{e^{-\zeta}\zeta^{B-1}}{(B-1)!} F(p)^{B-1} = \zeta e^{-\zeta(1-F(p))}, \quad (21)$$

where the second equality follows from (20). The expected revenue in a second-price auction with a fixed number of buyers B is

$$p(1 - F(p))BF(p)^{B-1} + \int_p^{\bar{v}} y(1 - F(y))[BF(y)^{B-1}]' dy,$$

see, for example, Krishna (2002, p. 25). Taking expectations over B yields

$$\begin{aligned} & p(1 - F(p)) \left[\sum_{B=1}^{\infty} \frac{e^{-\zeta}\zeta^B}{B!} BF(p)^{B-1} \right] + \int_p^{\bar{v}} y(1 - F(y)) \left[\sum_{B=1}^{\infty} \frac{e^{-\zeta}\zeta^B}{B!} BF(y)^{B-1} \right]' dy \\ &= p(1 - F(p))\zeta e^{-\zeta(1-F(p))} + \int_p^{\bar{v}} y(1 - F(y))[\zeta e^{-\zeta(1-F(y))}]' dy, \end{aligned}$$

by (21). The expected transaction price is the expected revenue divided by the probability of selling, which leads to (12).

Further, the cumulative distribution function of the number of buyers $\sum_{i=0}^B e^{-\zeta}\zeta^i/i!$ decreases in ζ . Therefore, a distribution of buyers with a higher arrival rate ζ first-order stochastically dominates a distribution with a lower ζ . Since both the probability of trade and the expected transaction price increase in B , they also increase in ζ . \square

Proof of Proposition 12. (a) For F and G Pareto, Φ and Γ are linear. Since the price of both the direct seller $\Phi^{-1}(c)$ and the indirect seller $\Phi^{-1}(\Gamma(c))$ are linear functions of c , the distributions of reserve prices have to be mirrored Generalized Pareto for both direct and indirect sellers with shape parameter σ . Further, the highest price set is \bar{v} for both direct and indirect sellers. The lowest price set by direct sellers is $\Phi^{-1}(\underline{c})$ and by indirect sellers $\Phi^{-1}(\Gamma(\underline{c})) = \Phi^{-1}(\underline{c})$, which is again the same. Having the same lower and upper bound and shape parameter, the distributions of the reserve prices of direct and indirect sellers are identical.

(b) Direct sellers set the price $\Phi^{-1}(c)$ (see Myerson (1981)) and enter if $c \leq \bar{v} = \bar{c}$. Indirect sellers set the price $\Phi^{-1}(\Gamma(c))$ (see Corollary 1) and enter if $c \leq \Gamma^{-1}(\Phi(\bar{v})) = \Gamma^{-1}(\bar{v}) = \Gamma^{-1}(\bar{c})$. Therefore, average prices of direct and indirect sellers are $P_A^D = E_c[\Phi^{-1}(c)]$ and $P_A^I = E_c[\Phi^{-1}(\Gamma(c)) | c \leq \Gamma^{-1}(\bar{c})]$. Using the linearity of Φ and (1), prices can be rewritten as $P_A^D = \Phi^{-1}(E_c[c])$ and $P_A^I = \Phi^{-1}(\Gamma^{-1}(\bar{c}))$. Since Φ and Γ are increasing, $P_A^D \geq P_A^I$ is equivalent to $\Gamma(E_c[c]) \geq \bar{c}$. This implies the statement because of $E_c[\Gamma(c)] = \bar{c}$ by (1) and Jensen's inequality. \square

Proof of Proposition 13. The profit of seller j when of type $c_j \in [\underline{c}, P^{-1}(\bar{v})]$ and when setting price p_j , given $\omega(p)$ and given that every other seller sets the price according to $P(c)$, can be written as

$$\begin{aligned} \Pi_j(p_j, c_j) &= \Pr(p_j = p)[p_j - \omega(p_j)] + \Pr(p_j < p)[E_p[p - \omega(p) | p > p_j] + (1 - \Pr(p_j \leq p))c_j] \\ &= [q(p_j) - (1 - H(p_j))][p_j - \omega(p_j)] + \int_{p_j}^{\bar{v}} [p - \omega(p)]dH(p) + (1 - q(p_j))c_j, \end{aligned}$$

where $q(p_j) := q_s(P^{-1}(p_j))$ is the probability that the seller of type $c_j = P^{-1}(p_j)$ trades and $H(p)$ is the cumulative distribution function of p . The probability $q(p_s)$ is $q(p_s) =$

$$\sum_{j=0}^{S-1} \binom{S-1}{j} G^j(P^{-1}(p_s))(1 - G(P^{-1}(p_s)))^{S-1-j} \sum_{i=j+1}^B \binom{B}{i} F^{B-i}(p_s)(1 - F(p_s))^i.$$

As will be shown below, what matters for the fee function is the probability $q(p_s) - (1 - H(p_s))$ rather than $H(p_s)$. This probability is $q(p_s) - (1 - H(p_s)) =$

$$\sum_{j=0}^{S-1} \binom{S-1}{j} G^j(P^{-1}(p_s))(1 - G(P^{-1}(p_s)))^{S-1-j} \binom{B}{j+1} F^{B-(j+1)}(p_s)(1 - F(p_s))^{j+1}.$$

The first-order condition for seller j 's problem is

$$q'(p_j)[p_j - \omega(p_j) - c_j] + [q(p_j) - (1 - H(p_j))][1 - \omega'(p_j)] = 0. \quad (22)$$

Defining

$$a(p) := \int_1^p \frac{q'(y)}{q(y) - (1 - H(y))} dy \quad (23)$$

and substituting $P^{-1}(p_j) = c_j$, one can re-write (22) as

$$\omega'(p_j) + \omega(p_j)a'(p_j) = 1 + [p_j - P^{-1}(p_j)]a'(p_j). \quad (24)$$

The boundary condition for this differential equation is that a seller of type $P^{-1}(\bar{v})$ makes zero profit even if he sold a unit (which will happen with probability 0); that is,

$$\bar{v} - \omega(\bar{v}) - P^{-1}(\bar{v}) = 0. \quad (25)$$

The unique solution to the differential equation (24), subject to the boundary condition (25), is the expression in (13). \square

Proof of Proposition 14. In the case that trade occurs, the weighted average of the intermediary's revenue and agents' utility is $\alpha(p^B - p^S) + (1 - \alpha)(v - c)$. Integrating over the area in which trade occurs yields

$$\int_{\underline{c}}^{p^S} \int_{p^B}^{\bar{v}} [\alpha(p^B - p^S) + (1 - \alpha)(v - c)] dF(v) dG(c). \quad (26)$$

The first-order condition with respect to p^S , $\partial(\alpha R + (1 - \alpha)W)/\partial p^S = 0$, yields after some algebra $G(p^S)f(p^B)[- \Gamma_\alpha(p^S) + \alpha p^S + (1 - \alpha)E_v[v|v \geq p^S]] = 0$. Similarly, the first-order condition with respect to p^B yields $(1 - F(p^B))g(p^S)[- \Gamma_\alpha(p^S) + \alpha p^B + (1 - \alpha)E[v|v \geq p^B]] = 0$. This proves the first part of the statement. The second part – that price posting is not optimal – follows from the fact that for the optimal α -allocation rule the trading threshold for the buyer, $v \geq \Phi_\alpha^{-1}(\Gamma_\alpha(c))$, depends on c , whereas for price posting, $v \geq p^B$, it does not. \square

Proof of Proposition 15. (i) For a finite number of buyers and sellers a dominant strategy implementation of the α -allocation rule is optimal: everyone reports their valuations and costs. Every buyer who trades pays $\max\{v_{w+1}, \Phi_\alpha^{-1}(\Gamma_\alpha(c_w))\}$, where w is the Walrasian

quantity defined with respect to the α -allocation rule. Every seller who trades receives $\min\{c_{w+1}, \Gamma_\alpha^{-1}(\Phi_\alpha(v_w))\}$.

The valuation of the marginal trading and non-trading buyers and the marginal seller's cost plus the spread charged by the intermediary converge in probability to the same value, which we denote as p^B :

$$\text{plim}_{K \rightarrow \infty} v_w = \text{plim}_{K \rightarrow \infty} v_{w+1} = \text{plim}_{K \rightarrow \infty} \Phi_\alpha^{-1}(\Gamma_\alpha(c_w)) =: p^B. \quad (27)$$

Similarly,

$$\text{plim}_{K \rightarrow \infty} c_w = \text{plim}_{K \rightarrow \infty} c_{w+1} = \text{plim}_{K \rightarrow \infty} \Gamma_\alpha^{-1}(\Phi_\alpha(v_w)) =: p^S. \quad (28)$$

For the fraction of buyers and sellers who trade we have

$$\text{plim}_{K \rightarrow \infty} \frac{w}{KB} = \text{plim}_{K \rightarrow \infty} \frac{\max\{i | v_i \geq p^B\}}{KB} = 1 - F(p^B), \quad (29)$$

$$\text{plim}_{K \rightarrow \infty} \frac{w}{KS} = \text{plim}_{K \rightarrow \infty} \frac{\max\{i | c_i \geq p^S\}}{KS} = G(p^S). \quad (30)$$

(27), (28), (29), and (30) imply that the optimal mechanism converges to price posting with p^B and p^S that satisfy $\Phi_\alpha(p^B) = \Gamma_\alpha(p^S)$ and $\zeta(1 - F(p^B)) = G(p^S)$.

(ii) Define $n_B := \max\{i | v_i \geq p^B\}$ and $n_S := \max\{i | c_i \leq p^S\}$ as the number of buyers and sellers willing to trade under the price posting mechanism. As $\Phi_\alpha(p^B) = \Gamma_\alpha(p^S)$,

$$\text{plim}_{K \rightarrow \infty} \frac{n_B}{KB} = 1 - F(p^B) = G(p^S) = \text{plim}_{K \rightarrow \infty} \frac{n_S}{KS}.$$

Further, since $\zeta(1 - F(p^B)) = G(p^S)$, we have $\Phi_\alpha(v_{n_B}) \geq \Phi_\alpha(p^B) = \Gamma_\alpha(p^S) \geq \Gamma_\alpha(c_{n_S})$ and by analogy $\Phi(v_{n_B+1}) < \Gamma(c_{n_S+1})$. Therefore, the fraction of traded quantity is in the limit

$$\text{plim}_{K \rightarrow \infty} \frac{\min\{n_B, n_S\}}{K \min\{B, S\}} = \frac{\max\{i | \Phi(v_i) \geq \Gamma(c_i)\}}{K \min\{B, S\}} \stackrel{\text{def}}{=} \frac{w}{K \min\{B, S\}},$$

which is the fraction of the Walrasian quantity defined with respect to the α -allocation rule. \square

B Extreme Value Theory

This appendix gives a brief review of extreme value theory. The theory of exceedences within extreme value theory deals with the distribution of a random variable conditional

on being above a high threshold (for the original articles see Balkema and De Haan (1974), Pickands (1975); for a textbook see Falk, Hüsler, and Reiss (2010)). The theory relates to the fact that one often observes that the upper tail of a distribution is well approximated by a Generalized Pareto distribution.

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

Definition 2. *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 $n = 1, 2, \dots$, such that*

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

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

This means that for n independently and identically distributed random variables, $(\max\{X_1, X_2, \dots, X_n\} - b_n)/a_n$ has a nondegenerate 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 $v \in (\underline{v}, \bar{v})$. Then F has a Generalized Pareto upper tail, formally*

$$\lim_{u \rightarrow 0} 1 - \frac{1 - F(\bar{v} - u(\bar{v} - v))}{1 - F(\bar{v} - u(\bar{v} - \underline{v}))} = 1 - \left(\frac{\bar{v} - v}{\bar{v} - \underline{v}} \right)^\beta, \quad (31)$$

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 (31) is the rescaled distribution conditional on being above the threshold $\bar{v} - u(\bar{v} - \underline{v})$. 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 (\bar{v} finite and $f(v) > 0$ for all $v \in (\underline{v}, \bar{v})$) the definition of F having a Generalized Pareto upper tail given in Definition 4 in Pickands (1975) simplifies to (31). \square

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*

$$\lim_{v \rightarrow \bar{v}} \frac{d}{dv} \left[\frac{1 - F(v)}{f(v)} \right] = \bar{\beta}, \quad (32)$$

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

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

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

Often, the Generalized Pareto distribution is defined with the parametrization

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

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

For infinite upper bounds, convergence can be stated as

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

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

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

C The General Mechanism Design Problem

In the main text we have focused on the regular case in which Φ_α and Γ_α are increasing and F and G have the same support, that is $\underline{v} = \underline{c}$ and $\bar{v} = \bar{c}$. We have also focused on allocation rules that are strictly monotone for all types that do not trade with probability 0 or 1. In this appendix we show that results hold with some slight modifications if we drop these assumptions.

First, consider $\underline{v} \neq \underline{c}$ and/or $\bar{v} \neq \bar{c}$. If $\Phi_\alpha(\underline{v}) \leq \Gamma_\alpha(\underline{c})$ and $\Phi_\alpha(\bar{v}) \leq \Gamma_\alpha(\bar{c})$ (which is satisfied for $\underline{v} = \underline{c}$ and $\bar{v} = \bar{c}$ as assumed in the main text), then results hold without modification. This is because the most inefficient traders – a buyer with valuation \underline{v} and a seller with cost \bar{c} – trade with probability zero and hence receive utility zero in a fee-setting mechanism.

Before deriving the optimal fee for the cases in which $\Phi_\alpha(\underline{v}) > \Gamma_\alpha(\underline{c})$, $\Phi_\alpha(\bar{v}) > \Gamma_\alpha(\bar{c})$, $\Phi_\alpha(v)$ decreasing for some v , or $\Gamma_\alpha(c)$ decreasing for some c , we introduce more general definitions of virtual valuation functions and inverses.

Define the ironed virtual valuation and virtual cost, denoted respectively as $\Phi_{\alpha,iron}$ and $\Gamma_{\alpha,iron}$, by following the procedure in Myerson (1981, p.68). That is, for a buyer,

define

$$J_{b,\alpha}(q) := \text{conv} \int_0^q \Phi_\alpha(F^{-1}(r))dr,$$

where conv stands for the convex hull of a function. The ironed virtual valuation function of the buyer can now be defined as $\Phi_{\alpha,iron}(v) = J'_{b,\alpha}(F(v))$. Wherever Φ_α^{iron} is strictly increasing, $\Phi_{\alpha,iron}(v) = \Phi_\alpha(v)$.

For the seller, construct the ironed virtual cost function in a similar way by defining

$$J_{s,\alpha}(q) := \text{conv} \int_0^q \Gamma_\alpha(G^{-1}(r))dr,$$

and the ironed virtual cost function as $\Gamma_{\alpha,iron}(c) = J'_{s,\alpha}(G(c))$.

Lemma 2. *In the general setup (in which Myerson's regularity assumption may not hold), it is optimal for the intermediary to allow trade if and only if $\Phi_{\alpha,iron}(v) \geq \Gamma_{\alpha,iron}(c)$ and to give zero utility to the most inefficient traders.*

Proof. By the same logic as Myerson (1981)'s theorem on p. 69 and Myerson and Satterthwaite (1983). \square

The easiest way to derive the optimal fee is by the following brief detour through the dominant strategy implementation of the optimal allocation rule. Similar to Myerson and Satterthwaite (1983, p.280), we can implement the allocation rule with dominant strategies in the general case by letting the buyer pay $P_{\alpha,gen}(c) := \Phi_{\alpha,gen}^{-1}(\Gamma_{\alpha,iron}(c))$ and the seller receive $P_{\alpha,gen}^{-1}(v) := \Gamma_{\alpha,gen}^{-1}(\Phi_{\alpha,iron}(v))$ in the case of trade. Here, gen stands for general and the generalized inverses are defined as

$$\begin{aligned} \Phi_{\alpha,gen}^{-1}(x) &:= \min\{v | v \geq \underline{v} \text{ and } \Phi_{\alpha,iron}(v) \geq x\}, \\ \Gamma_{\alpha,gen}^{-1}(x) &:= \max\{c | c \leq \bar{c} \text{ and } \Gamma_{\alpha,iron}(c) \leq x\}. \end{aligned}$$

Without trade, both get 0. We can now use the same logic as in the main text to derive the optimal fee-setting mechanism. The dominant strategy implementation already gives the right price for the buyer, $P_{\alpha,gen}(c)$. To ensure that the seller's payment only depends on c , we have to consider the expectation of $P_{\alpha,gen}^{-1}(v)$, which he gets in the dominant

strategy implementation. The net price the seller receives is $E_v[P_{\alpha,gen}^{-1}(v)|v \geq P_{\alpha,gen}(c)]$, since we have to take expectations conditional on trade taking place. The optimal fee is the gross minus the net price, in the indirect mechanism representation. Using p rather than $P_{gen}(c)$, this is

$$\omega_{\alpha,gen}(p) = p - E_v[P_{\alpha,gen}^{-1}(v)|v \geq p]$$

for $p \leq \bar{v}$ and any $\omega_{\alpha,gen}(p) \geq \bar{v} - p$ for $p > \bar{v}$. This is the same as in the main text, except that we are using ironed virtual valuations.

The generalized inverses cover the case $\Phi_{\alpha}(\underline{v}) > \Gamma_{\alpha}(\underline{c})$, since the price $P_{\alpha,gen}(c)$ set by the seller ensures that the price is never below \underline{v} , which ensures that the most inefficient buyer \underline{v} has zero utility. The inverse Φ_{α}^{-1} in the main text represented the solution of the first-order condition, assuming that there is an inner solution, the generalized inverse $\Phi_{\alpha,gen}^{-1}$ handles boundary solutions.

If $\Phi_{\alpha,iron}(\bar{v}) > \Gamma_{\alpha,iron}(\bar{c})$,⁴⁸ the most inefficient seller \bar{c} could obtain the positive profit $(p - \omega_{\alpha,gen}(p) - \bar{c})(1 - F(p))$, where $p = P_{\alpha,gen}(\bar{c})$. To correct for this, the optimal fee has to be chosen as

$$\omega_{\alpha,gen}(p) = p - E_v[P_{\alpha,gen}^{-1}(v)|v \geq p] + \left\{ \frac{1 - F(P_{\alpha,gen}(\bar{c}))}{1 - F(P_{\alpha,gen}(c))} (E[P_{\alpha,gen}^{-1}(v)|v \geq P_{\alpha,gen}(\bar{c})] - \bar{c}) \right\}, \quad (33)$$

where the expression in brackets ensures that sellers of type \bar{c} have no informational rents; that is $t_s(\bar{c}) - \bar{c}q_s(\bar{c}) = 0$. The price paid by the buyer, $P_{\alpha,gen}(c)$ is a left-continuous function (the discontinuities, if any, being determined by the discontinuities of $\Phi_{\alpha,iron}^{-1}$) with horizontal intervals where $\Gamma_{\alpha,iron}$ is horizontal. The inverse $P_{\alpha,gen}^{-1}(v)$ is a right continuous function (the discontinuities, if any, arising due to $\Gamma_{\alpha,iron}^{-1}$) with horizontal intervals where $\Phi_{\alpha,iron}$ is horizontal.

The fee in (33) relies only on charging a transaction fee. Another possible implementation is to charge the simple transaction fee $\omega_{\alpha,gen}(p) = p - E_v[P_{\alpha,gen}^{-1}(v)|v \geq p]$ and charge an additional participation fee $(1 - F(P_{\alpha,gen}(\bar{c}))(E[P_{\alpha,gen}^{-1}(v)|v \geq P_{\alpha,gen}(\bar{c})] - \bar{c})$ to the seller, which ensures that the most inefficient seller receives zero utility.

⁴⁸This can be the case, for example, if $\bar{v} > \bar{c}$ and if α small enough. Recall that $\bar{v} > \bar{c}$ is one of the necessary conditions for the possibility result for *ex post* efficient trade of Makowski and Mezzetti (1993).

Finally, we provide the proof of part (a) of Proposition 2.

Proof of part (a) of Proposition 2. From the revelation principle we know that no mechanism can implement \mathbf{Q} if it is not incentive-compatible. Thus we are left to show that a fee-setting mechanism can implement any incentive-compatible \mathbf{Q} . From the Revenue Equivalence Theorem (see, for example, Krishna (2002)) we know that \mathbf{Q} is incentive-compatible if and only if $q_b(v)$ is nondecreasing and $q_s(c)$ is nonincreasing.

Let the correspondence $a_{\mathbf{Q}}$, which maps from $[\underline{c}, \bar{c}]$ to $[\underline{v}, \bar{v}]$, be the boundary in (v, c) -space of the allocation rule \mathbf{Q} , such that $Q_b(v, c) = Q_s(v, c) = 1$ if and only if $v \geq a_{\mathbf{Q}}(c)$. Whenever $a_{\mathbf{Q}}$ is single-valued, then it is non-decreasing if \mathbf{Q} is incentive-compatible. Let $P : [\underline{c}, \bar{c}] \rightarrow [\underline{v}, \bar{v}]$ be the left-continuous function that has a discontinuity at every point at which $a_{\mathbf{Q}}$ is multi-valued and whose graph coincides with the graph of $a_{\mathbf{Q}}$ wherever $a_{\mathbf{Q}}$ is single-valued. Let $P_{gen}^{-1}(x) := \max\{c \leq \bar{c} \text{ and } P(c) \leq x\}$ be the generalized inverse. Note that a fee-setting mechanism implements \mathbf{Q} if it is such that a seller of type $c = P_{gen}^{-1}(p)$ optimally sets the price p . Thus, we are left to show that such a fee-setting mechanism exists. Let $\omega_{gen}(p) = p - E_v[P_{gen}^{-1}(v)|v \geq p]$ be the fee the intermediary charges. In this case, the seller of type c who sets the price p has an expected profit of $(p - \omega_{gen}(p) - c)(1 - F(p)) = \int_p^{\bar{v}} (P_{gen}^{-1}(v) - c)f(v)dv$. The first derivative with respect to p is $-(P_{gen}^{-1}(p) - c)f(p)$. Hence, the first-order condition is satisfied at $P_{gen}^{-1}(p) = c$, or equivalently at $p = P(c)$. Moreover, because $P_{gen}^{-1}(\cdot)$ is a weakly increasing function, the second-order condition is satisfied weakly whenever the first-order condition is, proving that the seller's profit function is quasiconcave. \square