Ad-valorem platform fees, indirect taxes and efficient price discrimination*

Zhu Wang[†] and Julian Wright[‡]

August 28, 2016

Abstract

This paper explains why platforms such as Amazon and Visa rely predominantly on ad-valorem fees, fees which increase proportionally with transaction prices. It also provides a new explanation for why ad-valorem sales taxes are more desirable than specific taxes. The theory rests on the ability of ad-valorem fees and taxes to achieve efficient price discrimination given that the value of a transaction to buyers tends to vary proportionally with the cost of the good traded.

JEL classification: D4, H2, L5

Keywords: Platforms, Taxation, Ramsey pricing

*We would like to thank the editor and four anonymous referees for their valuable comments and suggestions. We also thank Boyan Jovanovic, Bin Grace Li, llya Segal, Thomas Tregouet, Glen Weyl, and participants at various conferences and seminars for helpful comments, and Andy Lee for excellent research assistance.

†Research Department, Federal Reserve Bank of Richmond. Email: zhu.wang@rich.frb.org. The views expressed are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

‡Department of Economics, National University of Singapore. Email: jwright@nus.edu.sg.

1 Introduction

Ad-valorem fees and taxes which depend positively on transaction prices are widely used in practice. Platforms that facilitate transactions between buyers and sellers typically charge sellers ad-valorem fees, with fees for a transaction being a percentage of the transaction price set by sellers. Well-known examples include online marketplaces (such as Amazon and eBay) and payment card networks (such as Visa and MasterCard). Given these platforms do not incur significant costs that vary with transaction prices, their use of ad-valorem fees raises the question of why ad-valorem fees are so prevalent. Ad-valorem sales taxes are also widely used, with their desirability vis-a-vis per-unit (or specific) taxes the subject of a substantial literature.

In this paper, we provide a new rationale for the use of ad-valorem fees and taxes. The theory explains why profit-maximizing platforms favor ad-valorem fees even if they do not incur any variable costs of enabling trades. The same theory also implies that a regulatory authority that maximizes social welfare subject to covering a platform's fixed costs should similarly make use of ad-valorem fees when regulating fees for such platforms, and that a tax authority should prefer ad-valorem sales taxes rather than per-unit sales taxes.

The key idea behind our theory is that when a market involves many different goods that vary widely in their costs and values that may not be directly observable, then ad-valorem fees and taxes represent an efficient form of price discrimination because the value of a transaction is plausibly proportional to the cost of the good traded. The situation captures the usual scenario facing platforms, regulators or tax authorities. For example, within any specific market category (e.g., electronics), Amazon and eBay have goods traded that are worth a few dollars and others that are worth hundreds or even thousands of dollars. In such a setting, per-unit fees and taxes have the problem that they distort the price elasticity of demand across goods, since they add proportionally more to the price of a low-cost low-value item compared to a high-cost high-value item, thus reducing the efficiency of revenue extraction. Ad-valorem fees and taxes that are proportional to the transaction price do not lead to such a distortion and so can ensure the optimal Ramsey pricing. In fact, we show that in our setting, charging ad-valorem fees and taxes allows the platform, regulator or tax authority to achieve the same level of profit or welfare that could be obtained under third-degree price discrimination as if the relevant authority could perfectly observe the cost and valuation for each good traded and set a different optimal fee for each.

We then extend the theory to accommodate the fact that many platforms charge sellers a small fixed fee for each transaction in addition to the main ad-valorem component. Table 1 illustrates two such examples.[1] To do so, we allow for the fact that platforms typically incur a small marginal cost per transaction. We show that an affine fee schedule (a fixed fee per transaction plus a fee proportional to the transaction price) is optimal if and only if demand for the sellers belongs to the class with constant curvature of inverse demand (which includes linear demand, constant-elasticity, and exponential demand).

Table 1. Platform fee schedules

Amazon				Visa			
DVD	15%	+	\$1.35	Gas Station	0.80%	+	\$0.15
Book	15%	+	\$1.35	Retail Store	0.80%	+	\$0.15
Video Game	15%	+	\$1.35	Restaurant	1.19%	+	\$0.10
Game Console	8%	+	\$1.35	Small Ticket	1.50%	+	\$0.04

This result allows us to explore policy questions surrounding the use of ad-valorem fees by platforms, for example, whether platforms that do not incur proportional costs should be allowed to charge a percentage-based fee structure. The issue is particularly relevant for the payment card industry and is currently under debate in many countries, including Australia, Canada and the United States.[2] We address this policy question by considering a regulated setting in which the regulator seeks to maximize social welfare but allows the platform to recover its costs, including fixed

[1]The fee schedules reported in Table 1 are fees that Amazon and Visa charge to sellers for each transaction on the platform. Note that Visa fees shown in Table 1 are interchange fees for debit card services in the U.S. market. These fees, set by Visa, are paid by merchants to card issuers through merchant acquirers.

[2]Payment card fees are controversial in many countries. One particular issue is why payment card networks set interchange fees (in the case of Visa and MasterCard) or merchant fees (in the case of American Express) that depend on transaction prices rather than being a fixed amount per transaction. Merchants and policymakers point out that debit and prepaid card transactions do not provide credit or float and bear very small fraud risk; therefore, they do not warrant a percentage-based fee structure. See "Transparency, Balance and Choice: Canada's Credit Card and Debit Card Systems," Chapter 5, the Standing Senate Committee on Banking, Trade, and Commerce, Canada, June 2009.

costs, and has to decide whether to use ad-valorem fees to do so. In contrast to the policymakers' concerns regarding the use of ad-valorem fees, we show that for the class of demand functions that rationalizes a platform's use of affine fee schedules, welfare in this constrained case is always higher when a proportional fee is used to recover costs, in addition to a fixed per-transaction fee. Thus, our theory also suggests a policymaker that wishes to regulate platform fees to cover costs should also make use of ad-valorem fees.

In showing the optimality of ad-valorem fees we rely on two important assumptions in order to get strong analytical results: (i) inverse demands are proportional across goods, so that a scaling parameter that captures the cost and value of the good is the only difference across the goods traded, and (ii) sellers compete as identical Bertrand competitors for each good. In Section 4 of the paper, we consider what happens when these assumptions are relaxed. Our robustness exercises clarify that while assumption (i) is the key to pining down the price-discrimination role of ad-valorem fees, it is not a knife-edge result. Rather, ad-valorem fees still allow a platform to do well when there is some unobserved random variation of demand across goods (such as in elasticities). On the other hand, assumption (ii) is primarily for analytical tractability; it allows us to isolate the price discrimination role for ad-valorem fee schedules from their role in mitigating double marginalization, which only arises when sellers have market power. We show that, even when sellers have market power (including the extreme case that they are local monopolists), a platform loses very little by using the same affine fee schedule derived in the Bertrand competition setting.

Our paper is distinct from some recent studies that consider the use of proportional fees by platforms. Shy and Wang (2011) focus on the market power of the platform and sellers, and argue that a proportional fee is preferred over a fixed per-transaction fee since it helps mitigate double marginalization. Several papers (e.g., Foros et. al. 2014, Gaudin and White 2014a and Johnson 2015) have explored the optimality of a retailer (in our context, a platform) that lets suppliers (in our context, sellers) set final prices and receive a share of the revenue. Revenue sharing is equivalent to proportional fees, and this arrangement is known as the agency model. Like Shy and Wang (2011), they also assume suppliers are imperfect competitors but compare the agency model to a different alternative—the traditional wholesale model of retailing in which the retailer sets prices to consumers. The desirability of proportional fees in these papers again relates back to the ability of proportional fees to mitigate, although not eliminate, double marginalization.[3] As noted above, our paper shuts down double marginalization as an explanation for ad-valorem fees by assuming that sellers compete as identical Bertrand competitors for each good, thereby isolating the role ad-valorem fees play as a price discrimination device.

Loertscher and Niedermayer (2012) consider an independent private values setup with privately informed buyers and sellers that trade a good. They show that fee-setting is an optimal mechanism for the intermediary and provide a limit argument for why an intermediary's optimal fees (or a government's optimal taxes) converge to fees (or taxes) that are linear in the traded price as markets become increasingly thin. In their setup, market thinness is captured as the truncation of the buyers' and seller's distribution, with truncation arising because of a fixed cost of a transaction. Their linearity result rests on the property that the conditional distributions of buyers' value and the sellers' cost converge to generalized Pareto distributions as the truncation points approach the upper and lower bounds of the respective supports. They then use the fact that linear fees are optimal under generalized Pareto distributions to show that linear fees are optimal in the limit. Thus, their linearity result relies on the same property for why the optimal per-transaction fee for a particular good is linear in its price in our paper, which comes from the properties of the generalized Pareto distribution, which they independently discovered.

Finally, Muthers and Wismer (2013) propose that a commitment by a platform to use proportional fees can reduce a hold-up problem that arises from the platform wanting to compete with sellers after they have incurred costs to enter the platform.

All these studies propose different (and complementary) rationales for why "platforms" may choose ad-valorem fees, but they all assume the platform only deals with one type of good. Our setting is one with many goods that come in different scales, with potentially very different costs and valuations. Charging ad-valorem fees and taxes allows the platform, regulator or tax authority to achieve the same level of profit or welfare that could be obtained under third-degree price discrimination, which we believe is a primary reason for the widespread use of ad-valorem fees. Our article

[3]Gaudin and White (2014b) show related results in the context of a revenue-maximizing government in comparing ad-valorem vs. per-unit sales taxes. Llobet and Padilla (2014) show related results in the context of an innovator that has licensed its technology to a downstream firm and has to decide on whether to use ad-valorem or per-unit license fees.

is therefore related to the substantial literature on monopolistic third-degree price discrimination (e.g., see Aguirre et al. 2010, for a general analysis).

Our paper also offers a new perspective on the use of ad-valorem sales taxes. One of the oldest literatures in the formal study of public finance is the choice between a per-unit (or specific) tax and an ad-valorem tax (a tax proportional to transaction prices). The main focus has been on the role of different forms of imperfect competition among sellers (see Keen, 1998 for a survey).[4] In contrast, our paper shuts down such channels by assuming sellers are identical price competitors. Rather, we focus on a previously overlooked aspect that ad-valorem taxes allow the tax authority to efficiently price discriminate across goods with different costs and valuations based on Ramsey pricing principles.[5] Thus, our paper contributes to the voluminous literature on indirect taxation, providing a new theory for why ad-valorem sales taxes should be preferred.

The remainder of this paper is structured as follows. Section 2 introduces the model. Section 3 provides results on the optimality and efficiency of ad-valorem taxes and fees. Section 4 explores the robustness of our findings to alternative specifications, including demand functions outside the class with constant curvature of inverse demand (Section 4.1.1), unobserved random variation in demand (Section 4.1.2), and sellers' market power (Section 4.2). Finally, Section 5 offers some concluding remarks.

2 The model

We consider an environment where multiple goods are traded in a market. For each good traded, there is a unit mass of buyers each of whom wants to purchase one unit of the good. We consider a regulator of the market (it could be a tax authority, a platform operator that facilitates the trades or a regulator of such a platform) that wishes to collect some revenue from traders in the market by levying taxes (in the case of a platform, we will refer to these as fees). The market could consist of a narrow category of goods (e.g., external hard drives) or a wide variety of different

[4]Important contributions include Suits and Musgrave (1953), Delipalla and Keen (1992), Skeath and Trandel (1994), Anderson *et al.* (2001) and Hamilton (2009).

[5]In reality, we often observe per-unit taxes on alcohol, cigarettes and gasoline. One explanation is that ad-valorem taxes on these goods may promote the consumption of lower quality products that could result in worse side effects. Our paper abstracts from this issue.

goods that the regulator cannot easily distinguish (e.g., electronics).

Within such a market, there are multiple identical sellers of each good who engage in Bertrand competition.[6] Different goods within the market are indexed by c, which can be thought of as a scale parameter, so that different goods can be thought of as having similar demands except that they come in different scales. In particular, the per-unit cost of good c to sellers (which is known to all buyers and sellers of the good) is normalized to c and the value of the good to a buyer drawing the benefit parameter $b \geq 0$ is $c(1+b)$, so the scale parameter increases the cost and the buyer's valuation proportionally.[7] We denote the lowest and highest values of c as c_L and c_H respectively, with $c_H > c_L > 0$. We assume $1+b$ is distributed according to some smooth (i.e., twice continuously differentiable) and strictly increasing distribution function F on $\left[1, 1+\overline{b}\right]$, where $\overline{b} > 0$. (We do not require that $\overline{b}$ is finite.) Only buyers know their own b, while F is public information. Let $f = F'$ be the density and $h = f/(1-F)$ the hazard rate of F.

This setup captures the idea that for a given market that can be identified by the regulator, the main difference across the goods traded is their scale (i.e., some goods are worth a little and some a lot). In comparison to the wide range of scales of goods traded, potential differences in the shapes of demand functions across the different goods traded are not likely to be of first-order importance. The assumption that buyers' values for a good can be scaled by c is broadly consistent with a key empirical finding of Einav *et al.* (2013). They examined quasi-experimental observations from a large number of auctions of different goods on eBay and found that the distribution of buyer valuations for a good is proportional to the normal transaction price of the good (taken as the average price across posted-price transactions of the particular good). Since we will show that the transaction price predicted by our model is proportional to c, this finding implies an inverse demand function that is proportional to c, which is the essential property of our model as shown below.[8]

Our demand specification can also be justified on alternative grounds. In Appendix A we provide a setting which generates the same demand specification but there need not be a positive correlation between costs and consumer valuations of

[6] We discuss what happens when sellers have market power in Section 4.2.

[7] The assumption $b \geq 0$ is an innocuous normalization. In fact, we do not have to consider buyers with $b < 0$ given their valuation for a product is less than its cost.

[8] We will show the robustness of our results to this assumption in Section 4.1.2.

the goods traded. Rather, we consider a platform that reduces trading frictions, and assume the loss to buyers of using the less efficient trading environment (i.e., trading without using the platform) is proportional to the cost or price of the goods traded. This would apply whenever the alternative trading environment exposes the buyer to some risk or inconvenience that is proportional to the amount she pays for the good. With this alternative demand specification, we can apply the rest of our model setup to obtain identical results.

We assume the regulator cannot directly observe which good is being traded within a given market (i.e., the scale of the good sold), but only the transaction price of each trade. As a result, taxes can only condition on the price of a transaction (i.e., not on the actual cost of sellers or valuation of buyers). Since sellers of a particular good are identical and compete in price, competition between sellers for a good c will result in an equilibrium in which sellers charge buyers a common price p_c. In general, the tax schedule (equivalently, fee schedule) can be written as $T(p_c)$ for a given price of the transaction p_c, which can be decomposed into a buyer tax schedule $T^b(p_c)$ and seller tax schedule $T^s(p_c)$. Each tax schedule can consist of two components, a fixed per-transaction component (i.e., a specific tax) which is independent of p_c and a variable per-transaction component which varies with p_c (i.e., the ad-valorem component).[9]

Given that identical sellers compete for buyers, any tax charged to sellers will be passed through to buyers. The final price faced by buyers will reflect any taxes, and the buyer treats these the same whether she faces them directly or through sellers. Due to this standard result on the irrelevance of the incidence of taxes across the two sides, we can assume without loss of generality that only the seller side is charged. With this normalization, we write the sellers' tax schedule as simply $T(p_c)$. In case of a platform, we call this a fee schedule.

The number of transactions Q_c for a good c is the measure of buyers who obtain non-negative surplus from buying the good, $\Pr(c(1+b) - p_c \geq 0)$. Therefore, we can

[9]We assume all buyers and sellers have access to the market, so in the case of a platform, we abstract from joining fees. Seller joining fees would not help in our model since the platform already extracts all surplus through sellers due to the Bertrand competition assumption. If buyers have to make participation decisions, buyer joining fees would also be redundant provided we assume buyers know their draw of b before deciding whether to join the platform and given that each buyer wants to purchase exactly one unit of a good.

write the demand function for good c as

$$Q_c(p_c) \equiv Q\left(\frac{p_c}{c}\right) \equiv 1 - F\left(\frac{p_c}{c}\right). \tag{1}$$

The corresponding inverse demand function for good c is $p_c(Q_c) = cF^{-1}(1 - Q_c)$, which note is proportional to c. This form of demand function, hinged on a scaled price, is reminiscent of the one used in Weyl and Tirole (2012), which they refer to as the *stretch parametrization* of a general demand function.

Finally, assume c takes on a finite number of distinct values in $[c_L, c_H]$, with the set of all such values being denoted C. The cumulative distribution of c on C is denoted G and the associated probability for each distinct value of c is denoted g_c, with $\sum_{c \in C} g_c = 1$.

3 The optimality of ad-valorem taxes and fees

In this section, we rationalize ad-valorem taxes and fees. We first show that when there is no variable cost associated with the regulator, a simple form of ad-valorem taxes and fees, which are proportional to price, can be used to maximize revenue and welfare (subject to a minimum revenue constraint). We then extend the theory to allow for the fact that platforms typically incur a small marginal cost per transaction. We show that an affine fee schedule, as shown in Table 1, is optimal if and only if demand for goods belongs to a particular yet broad family. We show under the same family of demand functions, that an affine fee schedule remains optimal even if the platform is regulated to just cover its costs, including fixed costs.

3.1 Proportional taxes and fees

We first rationalize the use of proportional taxes and fees when there is no variable cost associated with taxing (or handling) trades. Consider the problem facing a regulator in a hypothetical full-information setting in which the regulator can observe c for each good and wants to maximize its revenue by setting a different tax rate T_c for each good. In case the regulator is a profit-maximizing platform operator, this captures the situation in which the variable cost of handling transactions on the platform is zero. If the regulator sets the tax T_c to sellers of good c, the price for good c will be $p_c = c + T_c$ due to Bertrand competition, and the corresponding revenue from trades is $R_c = T_c Q\left(1 + \frac{T_c}{c}\right)$ for good c. The regulator's problem is to choose T_c for each good c to maximize the total revenue across all goods in the market, denoted as $R = \sum_{c \in C} g_c R_c$.

Similarly, we could consider a social planner choosing a tax schedule to maximize welfare subject to raising a minimum level K of revenue. This covers the case of a tax authority trying to optimally raise a certain amount of tax revenue from the market. It also covers the case of a regulatory authority that regulates the fee schedule charged by a platform so as to maximize welfare subject to the platform's break-even condition in case the platform faces a fixed cost of operation of K. In general, if the regulator sets the tax schedule T_c, the welfare generated by the trade of good c is $W_c = \int_{T_c}^{\infty} Q\left(1 + \frac{t}{c}\right) dt + T_c Q\left(1 + \frac{T_c}{c}\right)$. Assuming it is possible to raise at least K in revenue, the regulator's problem is to choose T_c to maximize the total welfare across all goods in the market, denoted as $W = \sum_{c \in C} g_c W_c - K$, subject to the constraint that $R \geq K$.

Denote e_{T_c} as the demand elasticity with respect to T_c. Then, the revenue-maximizing level of T_c, denoted T_c^*, needs to solve the first-order condition

$$e_{T_c^*} = -\frac{T_c^*}{c} \frac{Q'\left(1 + \frac{T_c^*}{c}\right)}{Q\left(1 + \frac{T_c^*}{c}\right)} = 1, \tag{2}$$

while the welfare-maximizing level of T_c, denoted T_c^r, needs to solve the first-order condition

$$e_{T_c^r} = -\frac{T_c^r}{c} \frac{Q'\left(1 + \frac{T_c^r}{c}\right)}{Q\left(1 + \frac{T_c^r}{c}\right)} = \frac{\eta}{1 + \eta}, \tag{3}$$

where $\eta > 0$ is the Lagrangian multiplier associated with the revenue constraint $R \geq K$. Note that the welfare-maximizing tax is characterized by the same first-order condition as the revenue maximization problem except that the tax is set so that the demand elasticity with regard to T_c will be lower in magnitude, reflecting that less than 100% of the weight in the social planner's objective is put on revenue recovery.

It is straightforward to see that in both cases, the optimal tax is pinned down by the ratio $\omega = \frac{T_c}{c}$ that solves (2) or (3). This implies $T_c = \omega c$, which combined with the fact $p_c = c + T_c$, implies the regulator's full-information optimal solution can be achieved using the linear ad-valorem (or proportional) tax schedule

$$T(p_c) = \frac{\omega}{1+\omega} p_c. \tag{4}$$

Implementation is possible even though the regulator cannot identify each good, given the slope of the tax schedule is strictly between 0 and 1, so it never pays for a seller to set a lower price in order to pay the lower tax meant for a lower-cost seller.

The findings are summarized in the following proposition. The proposition, as with other propositions in this section, is formally established in Appendix B.

Proposition 1 ***(Revenue and welfare maximization)***
Suppose the revenue function R_c is strictly concave in either tax T_c or quantity Q_c. Provided there is no variable cost associated with taxing (or handling) trades, a proportional tax schedule is optimal. Specifically,
(i) the proportional tax schedule (4) maximizes the revenue that a regulator can extract from a market, where the value of $\omega = \frac{T_c}{c}$ solves (2);
(ii) the proportional tax schedule (4) maximizes welfare subject to raising the revenue K, where the value of $\omega = \frac{T_c}{c}$ solves (3);
(iii) the tax rate $\frac{\omega}{1+\omega}$ chosen to maximize (constrained) welfare is lower than that chosen to maximize revenue.

The proposition provides a new argument in the long-standing debate between specific versus ad-valorem sales taxes. We show that ad-valorem taxes enable the tax authority to charge different specific taxes to different goods based on the value of transactions, which minimizes distortions in our setting. To better understand why proportional taxes and fees are the most efficient way to raise revenue for a platform, tax authority or a regulator, consider what would happen under the alternative in which all goods in the market face a fixed per-transaction tax T. In our model setup, (1) implies that the demand function for any given good c can be written as

$$Q_c(p_c) = Q\left(1 + \frac{T}{c}\right).$$

Charging a fixed per-transaction tax T for all goods would imply the effective rate of taxation, T/c, is higher for low-cost goods than for high-cost goods. Given the usual condition that the price elasticity of demand increases in price (known as Marshall's second law of demand), this implies that the demand elasticity with respect to T would be higher for low-cost goods than for high-cost goods. Compared to this situation, by shifting taxes from low-cost goods to high-cost goods, more revenue would be extracted from goods with low price elasticity of demand, thereby improving the efficiency of revenue extraction. This is how a proportional tax schedule achieves this result, which is in line with classic Ramsey pricing principles. This logic can also be illustrated by noting that a fixed per-transaction tax would likely result in some very low-value goods not being traded at all due to the high burden a fixed per-transaction tax would imply for such goods. A proportional tax schedule can instead prevent this.

3.2 Affine fee schedules

Proposition 1 helps explain the widespread use of proportional taxes and fees by tax authorities, regulatory authorities and platforms alike. However, in the case of platforms, we often see the use of affine fee schedules, which contain a fixed per-transaction fee in addition to the proportional component. Table 1 gave some examples of such fees. In this section, we show how such fee schedules are optimal for a platform or its regulator once we allow the platform to incur a positive marginal cost of handling each trade. We will show the optimality only holds for a restricted class of demand specifications, albeit a reasonably general one.

In the following analysis, we assume that the platform incurs a cost $d > 0$ for handling each trade, and that $\bar{b} > d/c_L$ so it is efficient for some buyers to trade with sellers even for the good with the lowest scale. Per-transaction costs likely arise in the examples provided in Table 1 given there is a probability any given transaction will require an intervention by the platform, such as in cases where disputes arise. In the case of Visa, card issuers also incur incremental costs for card processing, clearance and settlement.[10]

[10] According to Federal Reserve Board estimates in 2011, the average incremental cost can be as high as 21 cents for a typical debit card transaction in the U.S. market.

3.2.1 A profit-maximizing platform

Consider again the full-information setting in which the platform can observe c for each good and set a fee accordingly. The first-order condition for optimality can be written as

$$e_{T_c^*} = -\frac{T_c^*}{c}\frac{Q'\left(1+\frac{T_c^*}{c}\right)}{Q\left(1+\frac{T_c^*}{c}\right)} = \frac{T_c^*}{T_c^*-d},$$

which requires that

$$\frac{T_c^*-d}{c} = -\frac{Q\left(1+\frac{T_c^*}{c}\right)}{Q'\left(1+\frac{T_c^*}{c}\right)}. \tag{5}$$

Since the right-hand-side of (5) is the inverse hazard rate for F evaluated at $1+T_c^*/c$, we can infer a unique family of distributions and its corresponding demand functions that justify a platform's optimal fee schedule being an increasing affine function of the price of each good. The findings are established formally in the following proposition.

Proposition 2 ***(Profit maximization)***
Suppose $d>0$. The platform's optimal fee schedule is an increasing affine function of the price of each good with slope less than unity, that is,

$$T\left(p_c\right) = \frac{\lambda d}{1+\lambda\left(2-\sigma\right)} + \frac{p_c}{1+\lambda\left(2-\sigma\right)}, \tag{6}$$

where $\lambda>0$ and $\sigma<2$ if and only if one of the following equivalent conditions hold:
(i) the distribution of buyers' benefits F has an affine inverse hazard rate (including the special case the inverse hazard rate is constant) with the condition $h'+h^2>0$ (i.e., the hazard rate h never decreases too fast);
(ii) the distribution of buyers' benefits F is the generalized Pareto distribution

$$F\left(x\right) = 1-\left(1+\lambda\left(\sigma-1\right)\left(x-1\right)\right)^{\frac{1}{1-\sigma}}, \tag{7}$$

where $\lambda>0$ is the scale parameter and $\sigma<2$ is the shape parameter;
(iii) the corresponding demand functions for sellers on the platform are defined by the

class of demands with constant curvature of inverse demand

$$Q_c(p_c) = 1 - F\left(\frac{p_c}{c}\right) = \left(1 + \lambda\left(\sigma - 1\right)\left(\frac{p_c}{c} - 1\right)\right)^{\frac{1}{1-\sigma}}. \tag{8}$$

Proposition 2 focuses on the platform's profit-maximizing fee schedule. It characterizes the class of demand functions that rationalizes a platform's optimal fee schedule being an increasing affine function of the price of the good in our model setting. For the demand faced by sellers, this turns out to be the broad class of demand functions that has constant curvature of inverse demand defined by (8), where the curvature of inverse demand σ is the elasticity of the slope of the inverse demand with respect to quantity. As shown in Proposition 2, this class of demands can be derived from the condition that buyers' benefits follow a generalized Pareto distribution (GPD), so we will refer to it as the GPD demand in our following analysis.[11] Note the platform's optimal fee schedule, which is given by (6), has a fixed per-transaction component only if there is a positive cost to the platform of handling each transaction.

Proposition 2 suggests that the proportional component of the platform's optimal fee schedule hinges on the constant pass-through of c, which is a demand shifter. Also, consistent with Bulow and Pfleiderer (1983) and Weyl and Fabinger (2013), we can show GPD demand functions feature constant pass-through of cost, which is d in our setting. Using that $p_c = c + T_c\left(p_c\right)$, both results can be verified by rewriting (6) as

$$T_c^* = \frac{\lambda d + c}{\lambda\left(2 - \sigma\right)}, \tag{9}$$

which implies a pass-through rate of $\frac{1}{\lambda(2-\sigma)}$ for c and a pass-through rate of $\frac{1}{2-\sigma}$ for d. Note it is easily confirmed that (9) is the solution to (5) given the demand specification (8). This result shows that not only does an increasing affine fee schedule allow a platform to achieve its optimal outcome when there is heterogeneity in the goods that are traded (i.e., in c), but as can be seen from (6) it does so without requiring that the platform knows the distribution G of goods that are traded.

For the class of GPDs and implied demand functions, σ is the key parameter. When $\sigma < 1$, the support of F is $[1, 1 + 1/\lambda\left(1 - \sigma\right)]$ and it has increasing hazard. Accordingly, the implied demand functions $Q_c(p_c)$ are log-concave and include the

[11] This class of demands is considered by Bulow and Pfleiderer (1983), Aguirre et al. (2010), Bulow and Klemperer (2012), Weyl and Fabinger (2013) among others for various applications.

linear demand function ($\sigma = 0$) as a special case. Alternatively, when $1 < \sigma < 2$, the support of F is $[1, \infty)$ and it has decreasing hazard. The implied demand functions are log-convex and include the constant elasticity demand function ($\sigma = 1 + 1/\lambda$) as a special case. When $\sigma = 1$, F captures the left-truncated exponential distribution $F(x) = 1 - e^{-\lambda(x-1)}$ on the support $[1, \infty)$, with a constant hazard rate λ. This implies the exponential (or log-linear) demand $Q_c(p_c) = e^{-\lambda\left(\frac{p_c}{c}-1\right)}$.

3.2.2 Ramsey regulation

We now consider the case where a social planner wants to maximize total welfare conditional on allowing the platform to just recover its costs, including a fixed cost K plus the per-transaction cost $d > 0$. We are interested in knowing whether such a planner will also want to use an affine fee schedule to do so.

As was the case without a per-transaction cost, the platform's full-information problem and the social planner's full-information problem are aligned other than that the optimal demand elasticity with regard to T_c is equal to a lower level under the planner's problem. The first-order condition for social planner's problem is

$$e_{T_c^r} = -\frac{T_c^r}{c}\frac{Q'\left(1+\frac{T_c^r}{c}\right)}{Q\left(1+\frac{T_c^r}{c}\right)} = \frac{\eta}{1+\eta}\frac{T_c^r}{T_c^r - d}, \tag{10}$$

where $\eta > 0$ again is the Lagrangian multiplier associated with the revenue constraint. We can therefore obtain the following proposition.

Proposition 3 ***(Ramsey regulation)***
Consider a social planner who maximizes welfare subject to recovering the platform's costs, including a fixed cost K plus the per-transaction cost $d > 0$. The planner will
(i) choose an affine fee schedule given that sellers on the platform face GPD demand with $\lambda > 0$ and $\sigma < 2$;
(ii) set the proportional component lower than that chosen by the platform, and also set the fixed per-transaction component lower than that chosen by the platform when demand is more convex than constant elasticity demand.

We note from (10) that when $d > 0$, the GPD demand requires the optimal solution

$$e_{T_c^r} = \frac{\lambda d + \frac{\eta}{1+\eta} c}{\lambda d(\sigma - 1) + c}, \tag{11}$$

where $\frac{\eta}{1+\eta}$ equals one in the monopoly platform's solution. Given that $\sigma - 1 < 1$ and $\frac{\eta}{1+\eta} \leq 1$, (11) implies $\partial e_{T_c^r}/\partial c < 0$. Therefore, for both the platform and the social planner, fees or taxes should be set so that the demand elasticity with regard to T_c is higher for goods with low costs and values, but lower for goods with high costs and values. This is achieved by setting a fixed per-transaction fee as part of the optimal fee schedule.

Proposition 3 implies that for the class of GPD demand functions that rationalizes a platform's use of affine fee schedules, welfare is always highest when the social planner uses an affine fee schedule to maximize welfare subject to a break-even constraint. This justifies using ad-valorem fees in this fully regulated setting.[12] Because the platform breaks even at the welfare optimum, the social planner's choice of affine fee schedule also maximizes consumer surplus subject to the same break-even constraint.

4 Robustness of affine fee schedules

In this section, we check the robustness of our findings and show that a platform continues to do well with an affine fee schedule even if some of our major assumptions or conditions do not hold exactly. The detailed analysis for the results discussed in this section is given in an online appendix.

4.1 Alternative demand structures

We first consider the robustness of our results to the demand structure we have assumed. There are two main elements to consider.

The first is that we have shown that when $d > 0$, the platform's optimal fee schedule is an affine function of the price of each good if and only if the distribution of buyers' benefits $b \geq 0$ follows a GPD. While the GPD covers a broad family of commonly used distributions, it is still interesting to explore how well an affine fee

[12] Note that provided the platform's profit function is strictly concave in Q_c, the planner's optimal fee schedule is an increasing affine function with slope less than unity if and only if sellers on the platform face GPD demand with $\lambda > 0$ and $\sigma < 2$. In this case, the result in part (i) of Proposition 3 would parallel that in Proposition 2.

schedule approximates the optimal fee schedule under alternative distributions. We do this in Section 4.1.1.

A more fundamental demand assumption in our framework is the multiplicative scaling parametrization of demand. The requires that the willingness to pay of all types of buyers is assumed to grow proportionally with the cost of the good c. In Section 4.1.2, we consider how well an affine fee schedule approximates the optimal fee schedule when demand is subject to random variation across goods.

4.1.1 Non-GPD demand functions

We consider alternative demand functions such that buyers' benefits (i.e., $b \geq 0$) follow an alternative distribution rather than a GPD. The platform does not know the true distribution, but rather continues to assume that demand takes the GPD form. In this case, we show in the online appendix that while the affine fee schedule is no longer optimal, it is nevertheless very close to being optimal for the full range of goods. This result is driven by a well-known empirical regularity that the GPD is a good proxy for the upper tail of other distributions. In the statistics literature, the extreme value theory shows that under some mild regularity conditions, the distribution of a random variable conditional on it being above a certain threshold converges to a GPD asymptotically. Therefore, the GPD approximation is able to offer an affine fee schedule very close to the optimal fee schedule. The better the GPD proxies for the true distribution, the closer the affine fee schedule is to the first best.

Two quantitative exercises are used to highlight the results. In each exercise, we fit a GPD to data generated by a truncated normal distribution, for which the best fitted GPD implies an optimal affine fee that matches the fee used by Amazon (i.e., \$1.35+15% shown in Table 1). In the first exercise, we show the GPD proxies very well the 75% tail of the truncated normal distribution, and as a result, the platform achieves more than 99.7% of the maximal profit it would obtain for each good if it knew the true truncated normal distribution. In the second exercise, we find the GPD does not proxy quite as well the 25% tail of the truncated normal distribution. However, in terms of profit, the GPD approximation still performs well, which allows the platform to achieve 90-95% of the maximal profits for goods below \$10, 95-99% for goods between \$10 and \$60, and more than 99% for goods above \$60. These results are robust to the use of other standard distributions.

The analysis discussed so far assumes the platform does not know the true demand but approximates it with GPD demand, showing this approximation does not lead to much loss in the platform's profit. Alternatively, we can also consider a scenario in which the platform knows the true non-GPD demand. In this case, the same quantitative exercises above reveal that the optimal fee schedule remains approximately linear. Aside from the fact non-GPD demand can be approximated by GPD demand, this also reflects that d is relatively small. We know from Proposition 1, that a proportional fee is optimal for general (non-GPD) demand functions when $d = 0$, and this should remain approximately true when $d > 0$ is small.

4.1.2 Random variation in demand

It is straightforward to show that all our existing results in Section 3 continue to hold even if there are random multiplicative and additive shocks to demand. Given our multiplicative scaling of demand, such shocks do not change the elasticity or shape of demand, and so are irrelevant to the determination of the optimal fee schedule. The more interesting case arises when different goods' demands can come from distributions that have different means or different price elasticities. We consider both cases in this section.

□ **Random variation in the GPD mean**

We first study what happens when each good has GPD demand but is generated from a distribution that has a different mean value centered around some common mean. This allows goods with the same cost to have their values independently determined—a good that draws a high mean of b has higher value than another that has a low mean of b.

Suppose the platform faces GPD demand $Q_c = \left(1 + \lambda(\sigma - 1)\frac{T_c}{c}\right)^{\frac{1}{1-\sigma}}$. However, the mean of the GPD, $E[b] = \frac{1}{\lambda(2-\sigma)}$, is now a random variable. To be specific, we assume that

$$\lambda = \frac{1}{\frac{1}{\widehat{\lambda}} + (2 - \sigma)\,\varepsilon},$$

where $\widehat{\lambda}$ and σ are fixed numbers but ε is an i.i.d. random variable with mean zero drawn for each good. Then for any particular good with realization λ, we have

$$E[b] = \frac{1}{\lambda(2-\sigma)} = \frac{1}{\widehat{\lambda}(2-\sigma)} + \varepsilon.$$

Also, given that d is typically small, we focus on the simpler case in which $d = 0$.

We then compare the profits between two scenarios: (i) the platform ignores the random noise but rather assumes that the GPD has a fixed mean equal to $\frac{1}{\widehat{\lambda}(2-\sigma)}$ for each good c, (ii) the platform has perfect information and knows the realization of λ for each good c. Denoting the optimal profit for the two scenarios, π_c^0 and π_c^1 respectively, we can compute that

$$\frac{\pi_c^0}{\pi_c^1} \approx 1 - \frac{1}{2}\widehat{\lambda}^2(2-\sigma)^3 E\left[\varepsilon^2\right]. \tag{12}$$

Note that $\frac{\pi_c^0}{\pi_c^1} < 1$ given that $\sigma < 2$ and $\frac{\pi_c^0}{\pi_c^1} = 1$ when $E\left[\varepsilon^2\right] = 0$. Also, $\frac{\pi_c^0}{\pi_c^1}$ decreases in $E\left[\varepsilon^2\right]$, the variance of the noise term, and increases in σ, the curvature of the inverse demand function. Because the ratio of $\frac{\pi_c^0}{\pi_c^1}$ is independent of c, the same ratio holds for comparing the platform's total profits.

With some commonly used demand functions within the GPD class, we can show the profit loss tends to be small under a reasonable dispersion of the noise. Take the Amazon example. The fee used by Amazon (i.e., 15%) implies the mean of $E[b]$, that is $\frac{1}{\widehat{\lambda}(2-\sigma)} = 0.176$. If we consider linear demand where $\sigma = 0$, (12) implies that as long as the standard deviation of $E[b]$ is smaller than 0.056, the platform earns more than 90% of the optimal profit by sticking to the simple 15% proportional fee. Moreover, for the exponential demand where $\sigma = 1$, as long as the standard deviation of $E[b]$ is smaller than 0.079, the platform earns more than 90% of the optimal profit with the 15% proportional fee. In fact, as suggested by (12), such a threshold standard deviation of $E[b]$ increases monotonically in σ.

Note that the profit ratio calculated in (12) is a conservative estimate in the sense that we assumed the platform continues to use the 15% fee that is optimal when $E[b]$ is constant rather than choosing the best affine fee schedule taking the random variation of $E[b]$ into account.

☐ **Random variation in the demand elasticity**

In this case, we allow that there is random variation in the price elasticity of seller demand which is independent of the costs of goods. Specifically, we consider the GPD demand with $\sigma = 1 + \frac{1}{\lambda}$ and $\lambda > 1$, so that sellers on the platform face a constant-elasticity demand $Q_c = (1 + \frac{T_c}{c})^{-\lambda}$. We assume the demand elasticity λ is a random variable,

$$\lambda = \widehat{\lambda} + \varepsilon,$$

where $\widehat{\lambda}$ is the mean of λ and ε is defined as before.

Using the same approach as above, we can show that the profit ratio is

$$\frac{\pi_c^0}{\pi_c^1} \approx \frac{1 + \frac{1}{2}\left(\ln \frac{\widehat{\lambda}}{\widehat{\lambda}-1}\right)^2 E\left[\varepsilon^2\right]}{1 + \frac{1}{2}\left(\left(\ln \frac{\widehat{\lambda}}{\widehat{\lambda}-1}\right)^2 + \frac{1}{\widehat{\lambda}-1} - \frac{1}{\widehat{\lambda}}\right) E\left[\varepsilon^2\right]}. \tag{13}$$

Again, the profit loss is small under a reasonable dispersion of the noise. For example, in this case, the fee used by Amazon (i.e., 15%) implies $\widehat{\lambda} = 6.667$. Then (13) implies that as long as the standard deviation of λ is smaller than 3.073, the platform earns more than 90% of the optimal profit.

4.2 Sellers' market power

Returning to our GPD demand formulation without random variation, we now relax another important assumption in our analysis, that sellers compete à la Bertrand. In a setting where sellers have market power, we explore two questions. First, how well does our affine fee schedule in (6), referred to hereafter as the "Bertrand fee schedule", perform compared to the optimal affine fee schedule? Second, more generally, how well does this Bertrand fee schedule perform compared to the optimal (non-linear) fee schedule?

To address these questions, we suppose sellers on the platform are identical Cournot competitors with $n_c \geq 1$ such sellers for each good c. Note our formulation includes the most extreme alternative to Bertrand competition, the case in which each seller is a local monopolist. In contrast to the situation with Bertrand competition among sellers, the platform now faces a double marginalization problem. Nevertheless, we find that under GPD demand, the Bertrand fee schedule continues to perform well.

First, we show without information on each good's cost, the platform can continue to use the Bertrand fee schedule and earn a higher profit than if it knew the cost of each good and set the optimal per-transaction fee for each good. This is because the Bertrand fee schedule not only achieves the same benefits of price discrimination but also mitigates the double marginalization problem associated with sellers' market power, a result aligned with the findings of Shy and Wang (2011), Foros et. al. (2014), Gaudin and White (2014a) and Johnson (2015).

Second, we show that while the Bertrand fee schedule is not necessarily the optimal affine fee for all GPD demands, it is nonetheless very close to being optimal. For this exercise we assume $d = 0$ so we can derive general results that do not depend on the distribution of c. We consider three types of demand faced by sellers. For constant elasticity demand, the Bertrand fee schedule is indeed the optimal affine fee schedule. For exponential demand, provided the percentage fee in the Bertrand fee schedule is less than 50% (which requires $\lambda > 1$), then the Bertrand fee schedule achieves 98.5% of the profit that is achievable under the optimal affine fee schedule when $n_c = 1$ (the monopoly seller case), and this percentage converges to 100% as n_c becomes large. For linear demand, the same criteria requires $\lambda > 0.5$, in which case the Bertrand fee schedules achieves more than 97.5% of the profit under the optimal affine fee schedule when $n_c = 1$, and again, an even higher percentage as n_c increases. We summarize our main findings in the following proposition.

Proposition 4 ***(The Bertrand fee schedule and seller market power)***
Assume that the demand functions for sellers on the platform belong to the generalized Pareto class with $\lambda > 0$ and $\sigma < 2$, and that for each good c there are $n_c \geq 1$ identical sellers that set quantities. Then we have the following results:
(i) the platform obtains a higher profit using the Bertrand fee schedule than if it sets the optimal per-transaction fee for each good;
(ii) if sellers face constant elasticity demand ($\sigma = 1 + \frac{1}{\lambda}$ and $\lambda > 1$) and $d = 0$, the Bertrand fee schedule is the optimal affine fee schedule;
(iii) if sellers face exponential demand ($\sigma = 1$), $\lambda > 1$ and $d = 0$, the Bertrand fee schedule can recover more than 98.5% of the profit under the optimal affine fee schedule;
(iv) if sellers face linear demand ($\sigma = 0$), $\lambda > 0.5$ and $d = 0$, the Bertrand fee schedule can recover more than 97.5% of the profit under the optimal affine fee schedule.

Proposition 4 is a powerful result since in reality a platform typically deals with many categories of goods. Each of the categories could have hundreds or even thousands of different goods, so it seems unrealistic to assume the platform knows the cost c or the exact nature of competition for each good. Using the simple Bertrand fee schedule in (6) allows the platform to earn the highest possible profits in case of competitive sellers, while achieving almost the same profit as the optimal affine fee schedule in the presence of double marginalization. Overall, from this result, we expect that the loss from using the Bertrand fee schedule compared to the optimal affine fee schedule is small.

More generally, when $d > 0$ or if we want to consider a non-linear fee schedule rather than just an affine fee schedule, the analysis becomes more complicated and depends on the distribution of c, the value of d, and the GPD demand parameters. To proceed, we assume sellers face constant elasticity demand. We obtain the values of d and λ by matching the implied Bertrand fee schedule to Amazon's fee schedule for DVD sales from Table 1. We obtain an estimated distribution of c by fitting a log-normal distribution to the actual distribution of sales obtained from sales ranks of DVDs sold on Amazon. We assume each seller is a monopolist (i.e., $n_c = 1$), the most extreme alternative to Bertrand competition, which provides us with the most conservative results.

With these assumptions, we find the platform obtains a profit of 0.383 with a fixed per-transaction fee (i.e., without any price discrimination). If the platform could observe each different good sold by the sellers, it could do better setting the per-transaction fee that is optimal for each good c. This increases its profit by 17.7% to 0.457, which represents the gain due to price discrimination. Consistent with our results in Proposition 4 above, we find that the benefits of price discrimination can be obtained by using the Bertrand fee schedule, which does not require any information on the values of c and has the added benefit of mitigating double marginalization. Indeed, the platform can increase its profit to 0.537, or a further 16.3%, by using the Bertrand fee schedule. The Bertrand fee schedule represents a 33.8% increase in profit compared to relying on a fixed per-transaction fee. Taking account of the fact sellers are monopolists and the particular distribution of c, the platform can increase its profit by a further 1.5% moving to the optimal affine fee schedule.

Finally, we obtain the platform's profit for the optimal non-linear fee schedule, which we obtain by solving for the optimal polynomial fee schedule of degree k, starting with $k = 1$ (the affine fee schedule) and considering higher and higher k until the platform's profit no longer increases.[13] Compared to the optimal affine fee

[13] A more detailed description of the procedure and the resulting optimal fee schedule is given in

schedule, moving to the optimal non-linear fee schedule only increases the platform's profit by a further 1.3%. The results are summarized in Table 2. The table shows similar (indeed, even stronger) results are obtained from repeating the exercise with linear demand.

In summary, very little is lost from restricting fee schedules to affine fee schedules or indeed the Bertrand fee schedule. In case of constant elasticity demand, the platform can get 97.2% and 98.7% of the maximal profit from fee schedules using the Bertrand fee schedule and the optimal affine fee schedule respectively. For linear demand, the same percentages are 99.6% and 99.9% respectively. This suggests that even if sellers have market power, there is little to be gained from using a more complicated optimal non-linear fee schedule compared to the Bertrand fee schedule.

Table 2. Profitability of different types of fees

	Const-elasticity demand		Linear demand	
Fee schedule	Profit	Profit gain (%)	Profit	Profit gain (%)
Fixed per-transaction fee	0.383		0.632	
Per-transaction fee varying by good	0.457	17.7%	0.966	42.4%
Bertrand fee schedule	0.537	16.3%	1.039	7.2%
Optimal affine fee schedule	0.545	1.5%	1.041	0.3%
Optimal fee schedule	0.553	1.3%	1.043	0.1%
Total profit gain (%)		36.7%		50.0%

5 Concluding remarks

In this paper, we investigate a puzzle and possible policy concern. Anecdotal evidence suggests that usually there is no obvious relationship between the value of an exchange and the cost of the intermediation. So why do platforms often charge a fee proportional to the value of the exchange?

We argue that price discrimination provides a natural explanation for the widespread use of ad-valorem fees. A platform has to deal with trade in multiple goods that vary widely in costs and valuations. If only a fixed per-transaction fee is used, a disproportionate amount will be charged to the low-cost and low-value goods, resulting in

the online appendix.

their price elasticity of demand being too high, while high-cost and high-value goods will end up with their price elasticity of demand being too low. Ad-valorem fees and taxes ensure efficient revenue extraction for platforms and planners, and by the same logic, for tax authorities.

We show when platforms face costs of handling each transaction, the use of an affine fee schedule (a proportional fee plus a fixed per-transaction fee) is optimal or nearly optimal for the platform, allowing it to achieve efficient price discrimination. We also show such an affine fee schedule would be chosen by a planner who wants to maximize social welfare (or consumer surplus) subject to the platform covering its costs including fixed costs.

The idea that ad-valorem fees provide a simple and efficient way for price discrimination is likely to apply more generally. Besides the two examples given in the paper, online marketplaces and payment cards, there are many other potential applications of our theory including to auction houses, booking platforms, employment agencies, insurance brokers, investment banks, real estate agencies, etc. It would be interesting to explore the extent to which our theory applies to these different markets.

Our theory can usefully be extended in several directions. First, one could use our setting to evaluate the welfare effects of banning ad-valorem fees in an unregulated environment in which the platform remains free to set the level of its fixed per-transaction fee. We explore this in Wang and Wright (2016). Second, it would be interesting to consider whether ad-valorem fees are justified from a Mirrlees-Atkinson-Stiglitz perspective, rather than the Ramsey perspective we took. A benefit of ad-valorem fees from this perspective is that they track the income of the individual purchasing (in case there is tax evasion that needs to be compensated for with sales taxes, for example, or tax rates would otherwise be too low). Third, one could consider settings in which there is a role for more complicated fee mechanisms. For example, we sometimes observe platforms that set caps (e.g. eBay) or lower percentage rates for transactions that go beyond some high values (e.g. Amazon for major appliances). A possible explanation that could be explored using our alternative theory of demand in Appendix A is that the benefit of using the platform service is proportional to the price of the goods traded only up to some point, but after that the benefit is bounded or does not increase as fast with the price of goods reflecting that other alternatives become viable. Finally, it would be useful to extend our analysis to cover the case of competing platforms, to understand whether inter-platform competition changes the incentives to use ad-valorem fees in any way.

References

[1] Aguirre, Iñaki, Simon Cowan and John Vickers (2010). "Monopoly price discrimination and demand curvature." *American Economic Review*, 100, 1601–1615.

[2] Anderson, Simon, André de Palma and Brent Kreider (2001). "The efficiency of indirect taxes under imperfect competition." *Journal of Public Economics*, 81, 231–251.

[3] Bulow, Jeremy and Paul Pfleiderer (1983). "A note on the effect of cost changes on prices." *Journal of Political Economy*, 91, 182–185.

[4] Bulow, Jeremy and Paul Klemperer (2012). "Regulated prices, rent seeking, and consumer surplus." *Journal of Political Economy*, 120, 160–186.

[5] Delipalla, Sofia and Michael Keen (1992). "The comparison between ad valorem and specific taxation under imperfect competition." *Journal of Public Economics*, 49, 351–367.

[6] Einav, Liran, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan (2013). "Learning from seller experiments in online markets." Working Paper.

[7] Foros, Øystein, Hans Jarle Kind and Greg Shaffer (2014). "Turning the page on business formats for digital platforms: Does Apple's agency model soften competition?" Working Paper.

[8] Gaudin, Germain and Alexander White (2014a). "On the antitrust economics of the electronic books industry." Working Paper.

[9] Gaudin, Germain and Alexander White (2014b). "Unit vs. ad-valorem taxes under revenue maximization." Working Paper.

[10] Hamilton, Stephen F. (2009). "Excise taxes with multiproduct transactions." *American Economic Review*, 99, 458–71.

[11] Johnson, Justin (2014). "The agency model and MFN clauses." Working Paper.

[12] Keen, Michael (1998). "The balance between specific and ad valorem taxation." Fiscal Studies, 19, 1–37.

[13] Llobet, Gerard and Jorge Padilla (2014). "The optimal scope of the royalty base in patent licensing," *Journal of Law and Economics*, forthcoming.

[14] Loertscher, Simon and Andras Niedermayer (2012). "Fee-setting mechanisms: On optimal pricing by intermediaries and indirect taxation." Working Paper.

[15] Muthers, Johannes and Sebastian Wismer (2013). "Why do platforms charge proportional fees? Commitment and seller participation." Working Paper.

[16] Shy, Oz and Zhu Wang (2011). "Why do payment card networks charge proportional fees?" *American Economic Review*, 101, 1575–1590.

[17] Skeath, Susan and Gregory Trandel (1994). "A Pareto comparison of ad valorem and specific taxation in noncompetitive environments." *Journal of Public Economics*, 53, 53–71.

[18] Suits, D. B. and R. A. Musgrave (1955). "Ad valorem and unit taxes compared." *Quarterly Journal of Economics,* 67, 598–604.

[19] Wang, Zhu and Julian Wright (2016). "Should platforms be allowed to charge ad-valorem fees?" Working paper.

[20] Weyl, Glen and Jean Tirole (2012). "Market power screens willingness-to-pay." *Quarterly Journal of Economics*, 127(4), 1971-2003.

[21] Weyl, Glen and Michal Fabinger (2013). "Pass-through as an economic tool: Principle of incidence under imperfect competition." *Journal of Political Economy*, 121, 528–583.

Appendix A: Alternative trading environment

In this appendix, we consider an alternative setting which gives rise to the same demand specification (1) as in Section 2 and for which the results in the rest of the paper continue to hold. Suppose buyers can buy from sellers through an efficient trading environment which involves a "tax" or through an alternative less efficient trading environment. In case of a tax authority, the alternative trading environment may be to trade in the underground economy, thereby avoiding any tax. In the case of a platform, the alternative trading environment may be trading directly or through a different means (e.g., from the seller's own website rather than through eBay or Amazon, or using cash instead of Visa or MasterCard payment cards), thereby avoiding any platform fees.

Different goods within a market are indexed by c. For good c, sellers all incur a unit cost c to provide the good, and buyers all value the good at $v_c > c$, which is assumed sufficiently large such that all buyers want to buy one unit of the good regardless of the trading environment. Note the buyers' valuation for each different good need not be scaled by c. There are assumed to be sufficient identical sellers so that there is homogenous Bertrand competition in both trading environments. Since there are no fees involved, the price sellers set in the alternative trading environment will exactly equal c.

Buyers purchasing through the efficient trading environment avoid an inconvenience or loss that is expected when purchasing through the alternative means, an expected loss which is equal to bc, where $b \geq 0$. Buyers have heterogenous valuations of this loss, so draw b from a distribution as in the model in Section 2. Thus, the key assumption in this alternative setting is that the loss from using the less efficient trading environment is proportional to the cost or price of the goods traded. In case of a tax collected by authorities, the assumption requires that the expected loss from making a trade at the price c in the underground economy is proportional to c, perhaps reflecting the amount at risk to each party in any such trade. In case of a payment card platform such as that offered by Visa, the assumption requires that the risk and inconvenience of having to use cash for a payment equal to c is proportional to c, reflecting the opportunity cost of the cash involved and the risk of theft.

Suppose taxes (or platform fees) are only on the seller side. Thus, when a transaction takes place, a buyer of good c that draws b and faces a price for good c through the efficient trading environment of p_c receives a surplus of $v_c - p_c$ using the efficient trading environment and $v_c - (1 + b)\, c$ using the inefficient trading environment. The assumptions on $1 + b$ and other parameters then follow as in Section 2. The number of transactions on the platform Q_c for a particular good c is the measure of buyers whose surplus from using the efficient trading environment is greater than using the alternative, i.e., $\Pr\left(1+b \geq \frac{p_c}{c}\right)$, which gives rise to the same demand function (1) as in Section 2. The rest of the analysis follows as before.

Appendix B: Proof of Propositions

Proof of Proposition 1. Equations (2) and (3) provide necessary conditions for maximization. We show here that under our assumption, that the revenue function R_c is strictly concave in either tax T_c or quantity Q_c, (2) and (3) also pin down the unique solution for the maximization problem.

(i) If the revenue function R_c is strictly concave in tax T_c, there is a unique T_c^* that satisfies (2), which maximizes the revenue. If R_c is strictly concave in quantity Q_c, there is a unique Q_c^* that satisfies $\frac{\partial R_c}{\partial Q_c}=0$, which maximizes the revenue. Because $\frac{\partial R_c}{\partial T_c}=\frac{\partial R_c}{\partial Q_c}\frac{\partial Q_c}{\partial T_c}$ and $\frac{\partial Q_c}{\partial T_c}<0$, T_c^* (corresponding to Q_c^*) is the only solution that satisfies (2). Therefore, in either case, (2) pins down the global maximum.

(ii) The welfare maximization problem is

$$\max \sum_{c\in C} g_c W_c - K \qquad s.t. \quad \sum_{c\in C} g_c R_c \geq K.$$

First, we consider the case where the revenue function R_c is strictly concave in T_c. In this case, the platform chooses the vector of taxes T_c, one for each good, to maximize welfare. Given that the objective function $W_c = \int_{T_c}^{\infty} Q\left(1+\frac{t}{c}\right)dt + T_c Q\left(1+\frac{T_c}{c}\right)$ is strictly declining in T_c for each good, the objective function $\sum_{c\in C} g_c W_c$ is quasi-concave in T_c. Also, because the revenue function R_c is strictly concave in T_c, the weighted sum $\sum_{c\in C} g_c R_c$ is strictly concave in T_c. As a result, according to the Kuhn-Tucker maximum theorem, the first-order condition given by (3) pins down the unique T_c^r solving the maximization problem (see Arrow and Enthoven, 1961).

Second, we consider the case where the revenue function R_c is strictly concave in quantity Q_c. In this case, the platform chooses the vector of quantities Q_c, one for each good, to maximize welfare. We can rewrite the objective function as $W_c = \int_0^{Q_c} T_c(x)dx = \int_0^{Q_c} c[F^{-1}(1-x)-1]dx$. Note that W_c is strictly concave in Q_c for each good, so the objective function $\sum_{c\in C} g_c W_c$ is strictly concave in Q_c. Also, given the revenue function R_c is strictly concave in Q_c, the weighted sum $\sum_{c\in C} g_c R_c$ is strictly concave in Q_c. As a result, the Kuhn-Tucker maximum theorem implies that there is a unique Q_c^r solving the maximization problem. Again, because $\frac{\partial Q_c}{\partial T_c} < 0$, T_c^r (corresponding to Q_c^r) is the only solution that satisfies (3). Therefore, (3) pins down the global maximum.

(iii) Finally, note that the value of ω solving (3) will be lower than that solving (2). To see this, we first consider the case where the revenue function R_c is strictly concave in T_c. Denote the solution to (2) as T_c^* and $\omega^* = \frac{T_c^*}{c}$, we have $\frac{\partial R}{\partial T_c}|_{T_c=T_c^*} = Q(1+\omega^*) + \omega^* Q'(1+\omega^*) = 0$. Denote the solution to (3) as T_c^r and $\omega^r = \frac{T_c^r}{c}$. Given that $\eta > 0$, (3) implies that $\frac{\partial R}{\partial T_c}|_{T_c=T_c^r} = Q(1+\omega^r) + \omega^r Q'(1+\omega^r) > 0$. Given that $\frac{\partial^2 R}{\partial T_c^2} < 0$, we have $\omega^r < \omega^*$, and so $\frac{\omega^r}{1+\omega^r} < \frac{\omega^*}{1+\omega^*}$. Similarly, for the case where the revenue function R_c is strictly concave in Q_c, note that the welfare maximizing quantity is greater than the revenue maximizing one (i.e., $Q_c^r > Q_c^*$). Accordingly, we again have $\omega^r < \omega^*$, and so $\frac{\omega^r}{1+\omega^r} < \frac{\omega^*}{1+\omega^*}$. ■

Proof of Proposition 2. We proceed in four steps.

(1) We first show the result in the "if" direction. Suppose F has an affine inverse hazard rate $\rho(x) = \rho_0 + \rho_1 x$, where the hazard rate h satisfies $h'(x) + h^2(x) > 0$ for $x \in [1, 1+\bar{b}]$. Given $F'(1) > 0$ we have $\rho_0 + \rho_1 > 0$. The property $h'(x) + h^2(x) > 0$ implies $\rho_1 < 1$. If we can observe c and choose a corresponding T_c^*, the first-order condition for optimality can be written as

$$\frac{T_c^* - d}{c} = -\frac{Q\left(1 + \frac{T_c^*}{c}\right)}{Q'\left(1 + \frac{T_c^*}{c}\right)}. \tag{14}$$

(That T_c^* satisfying this gives the global maximum when F has an affine inverse hazard rate is shown in step 4 below.) Because the right-hand-side of (14) is the inverse hazard rate for F evaluated at $1 + T_c^*/c$, we must have

$$T_c^* = \frac{d}{1-\rho_1} + \frac{\rho_0 + \rho_1}{1-\rho_1} c. \tag{15}$$

This can be implemented by the affine fee schedule

$$T(p_c) = \frac{d}{1+\rho_0} + \frac{\rho_0 + \rho_1}{1+\rho_0} p_c, \tag{16}$$

where $\rho_0+\rho_1>0$ and $\rho_1<1$ implies $\rho_0>-1$, so that the intercept is positive and the slope is between 0 and 1. Note implementation is possible as sellers of type c have to cover their costs $c+T(p_c)$ and so equilibrium prices for good c are given by the lowest p_c such that $p_c \geq c+T(p_c)$. The equilibrium price is characterized by the unique solution to $p_c = c+T(p_c)$ for each good c, which is $p_c = \frac{d+c(1+\rho_0)}{1-\rho_1}$. Given (16), the fee set for good c will indeed be (15) as required.

(2) We next consider the "only if" part of the proof. Without loss of generality, suppose we start with (16) being the optimal fee schedule, with $\rho_0+\rho_1>0$ and $\rho_1<1$. This implies the equilibrium price for good c will be given by $p_c = \frac{d+c(1+\rho_0)}{1-\rho_1}$. Given the slope of the fee schedule is between 0 and 1, a seller cannot make more profit by lowering its price below this level to enjoy a lower fee. Substituting this price into (16) implies the optimal fee for good c is given by (15). Substituting this into (14) implies

$$\frac{\rho_0+\rho_1\left(1+\frac{d}{c}\right)}{1-\rho_1} = -\frac{Q\left(1+\frac{\rho_0+\rho_1+\frac{d}{c}}{1-\rho_1}\right)}{Q'\left(1+\frac{\rho_0+\rho_1+\frac{d}{c}}{1-\rho_1}\right)}, \tag{17}$$

which must hold for all values of c in the support of G.

Given we assume c takes on only a finite number of values, there may be multiple fee schedules that can implement the optimal outcome implied by (17) for the platform. Provided they satisfy optimality around the prices implied by these values of c, there may be some flexibility for the fee schedule for other prices. If this is the case, we assume that the platform will choose the fee schedule from among the multiple fee schedules that remains optimal for as many other values of c between c_L and c_H (i.e., which are not in the support of G). Thus, by optimality we rule out a weakly dominated schedule—one that produces the same profit for the given distribution of c but would be lower in case c takes on some other values outside of G. Given this requirement, optimality can only hold if the inverse hazard rate (the function $-Q(\cdot)/Q'(\cdot)$) is the increasing affine function with intercept ρ_0 and slope ρ_1, where $h'(x)+h^2(x)>0$ follows from $\rho_1<1$.

(3) The equivalence between (i), (ii) and (iii) is shown as follows. Without loss of generality, the requirement that the inverse hazard rate be linear is equivalent to requiring

$$\frac{F'(x)}{1-F(x)} = \frac{\lambda}{1+\lambda(\sigma-1)(x-1)} \tag{18}$$

for some constant parameters λ and σ. Evaluating (18) at $x = 1$, implies that $\lambda > 0$ and $\sigma < 2$ given the requirements that $\rho_0 + \rho_1 > 0$ and $\rho_1 < 1$. Solving the differential equation, the solution requires (7), with the special case $F(x) = 1 - e^{-\lambda(x-1)}$ when $\sigma \to 1$.

Accordingly, the demand faced by sellers is the broad class of demand functions that has a constant curvature of inverse demand σ. It is specified by (8), which includes linear demand ($\sigma = 0$) and constant-elasticity demand ($\sigma = 1 + 1/\lambda$), as well as exponential demand ($\sigma \to 1$). The demand faced by the platform for good c is given by

$$Q_c(T_c) = 1 - F\left(1 + \frac{T_c}{c}\right) = \left(1 + \frac{\lambda(\sigma - 1)T_c}{c}\right)^{\frac{1}{1-\sigma}}, \tag{19}$$

which also belongs to this constant-curvature class of inverse demand functions, though it does not cover the full class (e.g., the platform does not face a constant-elasticity demand).

Given the platform demand function (19), the optimal fee satisfying (14) is

$$T_c^* = \frac{\lambda d + c}{\lambda(2 - \sigma)}. \tag{20}$$

Substituting $p_c = c + T_c^*$, the optimal fee can be written as (6), which is equivalent to the affine fee schedule (16) derived above with $\rho_0 = 1 - \sigma + 1/\lambda$ and $\rho_1 = \sigma - 1$.

(4) We can verify that T_c^* satisfying (20) indeed yields the maximal profit for the class of GPD demand functions identified by (19). Note that the assumption $\bar{b} > d/c_L$ ensures that T_c^* satisfying (20) is always an interior solution. Also, given $\sigma < 2$, the GPD demand implies that the profit function, denoted by Π_c, is strictly concave in Q_c, so there is a unique Q_c^* that satisfies $\frac{\partial \Pi_c}{\partial Q_c} = 0$, which maximizes the profit. Because $\frac{\partial \Pi_c}{\partial T_c} = \frac{\partial \Pi_c}{\partial Q_c}\frac{\partial Q_c}{\partial T_c}$ and $\frac{\partial Q_c}{\partial T_c} < 0$, T_c^* (corresponding to Q_c^*) is the only solution that satisfies (20). Therefore, (20) pins down the unique global maximum of Π_c. ■

Proof of Proposition 3. (i) The first-order condition for the social planner's problem (10) can be written as

$$\frac{1 + \eta}{\eta}\frac{(T_c^r - d)}{c} = -\frac{Q\left(1 + \frac{T_c^r}{c}\right)}{Q'\left(1 + \frac{T_c^r}{c}\right)}. \tag{21}$$

Again, the right-hand-side of (21) is the inverse hazard rate for F evaluated at $1 + T_c^r/c$. Therefore, the first-order condition implies an affine fee schedule given that the demand functions for sellers belong to the GPD class with $\lambda > 0$ and $\sigma < 2$. Accordingly, the optimal regulated platform fee is

$$T_c^r = \frac{(1+\eta)\,d}{1+\eta\,(2-\sigma)} + \frac{\eta c}{\lambda[1+\eta\,(2-\sigma)]}, \tag{22}$$

and the corresponding optimal affine fee schedule is

$$T(p_c) = \frac{(1+\eta)\,\lambda d}{\lambda + [1+(2-\sigma)\,\lambda]\eta} + \frac{\eta p_c}{\lambda + [1+(2-\sigma)\,\lambda]\eta}, \tag{23}$$

with the ad-valorem term $\frac{\eta}{\lambda+(1+(2-\sigma)\lambda)\eta} \in (0,1)$ given $\eta > 0, \lambda > 0$ and $\sigma < 2$. To verify that T_c^r satisfying (22) indeed yields the maximal welfare, note that the assumption $\bar{b} > d/c_L$ ensures that T_c^r satisfying (22) is always an interior solution. We can verify that the welfare function is strictly concave in Q_c. Also, given $\sigma < 2$, the GPD demand implies that the profit function is also strictly concave in Q_c. Hence, there is a unique Q_c^r solving the welfare maximization problem subject to the break even constraint. Again, given $\frac{\partial Q_c}{\partial T_c} < 0$, T_c^r (corresponding to Q_c^r) is the only solution that satisfies (22).

(ii) Comparing the platform's optimal fee schedule (6) with that set by the social planner (23), we find that the proportional component chosen by the planner is always lower than that chosen by the platform given $\eta > 0$, $\lambda > 0$ and $\sigma < 2$. The fixed per-transaction component chosen by the social planner is also lower if and only if $\sigma > 1 + 1/\lambda$. ■