Working Paper Series

# Should Platforms be Allowed to Charge Ad Valorem Fees?

WP 17-05

Zhu Wang
Federal Reserve Bank of Richmond

Julian Wright
National University of Singapore



FEDERAL RESERVE BANK OF RICHMOND®
Richmond ▪ Baltimore ▪ Charlotte

# Should platforms be allowed to charge ad valorem fees?

Zhu Wang* and Julian Wright†

March 2017

Working Paper No. 17-05

**Abstract**

Many platforms that facilitate transactions between buyers and sellers charge ad valorem fees in which fees depend on the transaction price set by sellers. Given these platforms do not incur significant costs that vary with transaction prices, their use of ad valorem fees has raised controversies about the efficiency of this practice. In this paper, using a model that connects platforms' use of ad valorem fees to third-degree price discrimination, we evaluate the welfare consequences of banning such fees. We find the use of ad valorem fees generally increases welfare, including for calibrated versions of the model based on data from Amazon's marketplace and Visa's signature debit cards.

*JEL classification:* D4, H2, L5

Keywords: platforms, taxation, third-degree price discrimination

*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 platform fees which depend positively on transaction prices are widely used in practice. Well-known examples include online marketplaces (such as Amazon and eBay), payment card platforms (such as Visa, MasterCard and American Express), and hotel booking platforms (such as Booking.com and Expedia). In these cases platforms typically charge sellers percentage fees, as well as sometimes small fixed per-transaction fees. Platform costs, which are largely fixed or dependent on the number (rather than the value) of transactions cannot explain the levels of ad valorem fees set by these platforms. This has led to criticisms of the ad valorem fee structure, given it is not cost reflective. In this paper, we explore whether such ad valorem fees harm welfare, and so whether there may be a case for banning them.

Concerns over the use of ad valorem fees have been raised in the context of payment card platforms. 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. For instance, Summers (2012) criticizes that "*Payment schemes' owners and infrastructure operators also have monopoly power that can be used to set prices far above their production cost. There is abundant evidence of clearing and settlement pricing that is based not on production cost but on methods designed to extract very high returns for use of the infrastructure. Perhaps the most prominent example is ad valorem pricing for payment methods that essentially involve giving bank account holders direct access to their deposits and that do not entail bank credit, as in the case of debit cards*". The Canadian Senate Committee on Banking, Trade and Commerce (June 2009) made the following ruling: "*The Committee believes that there is little rationale for percentage-based interchange, merchant discount and switch fees on debit cards, since this payment method involves a relatively simple and nearly instantaneous transfer of funds. There is no obvious credit risk and no interest-free period to fund in these transfers....The Committee believes that debit card transactions are inherently less risky and costly than credit card transactions; consequently, they do not warrant a percentage-based fee structure, whether at the level of interchange fees or switch fees.*"

To address this issue we make use of the model we developed in Wang and Wright (forthcoming) in which a profit maximizing platform designs its fee structure to take into account heterogeneity in demand across the many products sold over its platform. The key idea captured by the model 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 model implies that the profit maximizing fee structure is affine (consisting of a percentage fee plus a fixed per-transaction component) if and only if the demand faced by sellers belongs to the generalized Pareto class that features constant curvature of inverse demand (which includes linear demand, constant-elasticity, and exponential demand as special cases). This matches the fee structure used by many platforms, as shown in Table 1 below.[1] According to the model, the fixed per-transaction component is present only because the platform incurs a marginal cost for processing each transaction; otherwise a simple percentage fee would be profit maximizing. In Wang and Wright (forthcoming) we used the model to show that ad valorem fees should also be used by an authority that wants to regulate a platform's fees while allowing for the recovery of a certain amount of revenue (e.g., to cover the platform's fixed costs). In contrast, in this paper we investigate a different issue: What would happen if a policymaker banned a platform's use of any fee that depended on the value of transactions (i.e., ad valorem fees) but left the level of the platform's fees unregulated. For policymakers who want platform fees to be determined on the basis of costs but are concerned about directly regulating fee levels, this seems to be a natural approach to consider. However, we show that the welfare results turn out not at all obvious, and are related to the long-standing debate on the welfare effects of third-degree price discrimination.

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 |

To address this question, we first show that the conditions for welfare to increase as a result of banning the use of ad valorem fees turn out to be the same as those which determine whether banning a monopolist from using third-degree price discrimination improves welfare, in which each good traded is treated as a separate observable market over which the monopolist can price. This allows us to draw on the substantial literature on monopolistic third-degree price discrimination (e.g., see Aguirre *et al.*, 2010, for a recent analysis). In the setting usually adopted for price discrimination studies, in which there are only two markets or a continuum of markets, we find welfare is generally higher when platforms are

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

allowed to use ad valorem fees. Specifically, we are able to show in the special case in which the platform trades only two goods, welfare is higher whenever demand is exponential or log-convex within the generalized Pareto class, and also when demand is log-concave within this class provided the two goods are sufficiently dispersed. We then extend these results to show similar qualitative findings hold allowing for a continuum of uniformly distributed goods.

The intuition for the above results relates to the standard intuition for why third-degree price discrimination is desirable when uniform pricing shuts down the low-demand market (or in our setting, trade of the low-demand good). For log-concave demand, there is a choke price at which demand becomes zero. As a result, provided the demand across the two goods is sufficiently different, a monopoly platform that can only set a uniform fee will want to set a fee that leads to the low-demand good not being traded because setting the fee low enough so that both goods are traded will sacrifice too much of the platform's profits. In this case, allowing the platform to price discriminate (i.e., by using ad valorem fees) will not only increase its profit but also consumer surplus and social welfare since the low-demand good is then offered for sale. We show this logic extends to the case with a continuum of uniformly distributed goods provided they are sufficiently dispersed. A similar logic also applies to exponential and log-convex demands in which there is no choke price, provided the two goods are sufficiently different. This is because a platform that can only set a single fee will set a fee very close to the monopoly level for the high-demand good, thereby almost ruling out all sales of the low-demand good. On the other hand, when the two goods are not very different, the existing results of Aguirre *et al.* (2010) can be applied to our exponential and log-convex demands to explain why welfare will be higher with price discrimination.

More generally, the welfare effects of allowing price discrimination when there are many goods can be sensitive to the distribution of goods. In practice, the distribution of prices and sales of goods traded are highly skewed. We therefore use information on the actual distribution of prices and sales measures for two different platforms (Amazon's and Visa's), as well as their fee structures, to calibrate our model. We find, in most cases, welfare would be harmed if ad valorem fees were banned. Our results therefore imply policymakers should be cautious about banning the use of ad valorem fees.

Shy and Wang (2011) also look at the welfare effects of a platform shifting from a fixed per-transaction fee to an ad valorem fee. They consider a setting with a monopoly platform and imperfectly competitive sellers that sell a homogenous good, and demand takes a constant-elasticity form. They find welfare is higher when the platform can use an ad valorem fee, but their result relies on the property that ad valorem fees help mitigate the double marginalization problem. Thus, their work complements ours, which assumes away any double marginalization problem by focusing on the case in which sellers are identical price competitors. In a similar vein, several studies (e.g., Foros *et al.*, 2014; Gaudin and White, 2014a; and Johnson, forthcoming) have explored the advantages of so-called agency model used by mass retailers such as Amazon, where the retailer lets suppliers (i.e., sellers) set final prices and receive a share of the revenue, which is equivalent to using a percentage fee. Like Shy and Wang (2011), they also show that the revenue-sharing used in the agency model has the advantage of mitigating double marginalization, but they differ by focusing on how the agency model affects retail prices compared with the traditional wholesale pricing model.

Our theory can also be used to justify the adoption of ad valorem taxes rather than unit (or specific) taxes in settings in which governments seek to maximize tax revenue (the so-called Leviathan hypothesis; see Brennan and Buchanan, 1977, 1978). In this case, the tax-revenue maximizing government is identical to a revenue-maximizing platform. Our results imply that an ad valorem tax regime welfare dominates a unit tax regime. In this regard, our finding complements that of Gaudin and White (2014b). They also show ad valorem taxes welfare dominate when governments maximize tax revenue, although they obtain their results in a very different setting, in which double marginalization rather than price discrimination is the key driving force.

Finally, our paper also contributes to the extensive literature on the welfare effects of monopolistic third-degree price discrimination (e.g., Schmalensee, 1981; Varian, 1985; Schwartz, 1990; Layson, 1994; Aguirre, 2006, 2008; Cowan, 2007, 2016; and Aguirre et al., 2010). Aguirre *et al.* (2010) is particularly relevant. They focus on the case there are two markets which are always served and consider a general demand specification. Among the cases for which they obtain stronger results is the case in which demand in each market has constant curvature of inverse demand. We study a special case of their specification in which all markets have the same constant curvature of inverse demand and in which inverse demand only varies by a multiplicative term across each market. At the same time, we generalize their setting by allowing for a continuum of markets and that not all markets are served. Our results also extend the findings of Malueg and Schwartz (1994), who consider a continuum of markets each having a linear demand that varies with a multiplicative term, which turns out to be a special case of ours. Moreover, we go beyond purely theoretical discussions by calibrating our model to the data from Amazon's marketplace and Visa's signature debit cards. This allows us to provide quantitative evaluations of the welfare effects of price discrimination at those platforms based on the actual, highly skewed, distributions of goods traded on these platforms.

The rest of the paper proceeds as follows. Section 2 sets out the model. Section 3 provides some analytical results while Section 4 provides results based on a calibrated model of the Visa platform and the Amazon platform. Finally, Section 5 provides some brief concluding remarks.

# 2 The model

We consider an environment where multiple goods are traded over a platform. For each good traded, a unit mass of buyers want to purchase one unit of the good. There are multiple identical sellers of each good who engage in Bertrand competition. Different goods sold on the platform 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. 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.

This setup captures the idea that for a given market that can be identified by the platform, 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 consistent with a key empirical finding of Einav *et al.* (2015) who study quasi-experimental observations from a large number of auctions of different goods on eBay.

In Wang and Wright (forthcoming), we show that this demand specification can also be justified on alternative grounds. Instead of directly assuming positive correlation between costs and consumer valuations of the goods traded, 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. This alternative demand specification delivers exactly the same results and helps further clarify why the difference in demand across the goods traded on the platform is mainly determined by their "scale". Wang and Wolman (2016) provide empirical evidence consistent with this interpretation. Analyzing payment patterns for two billion retail transactions, they find that the value of transaction is the key to explain consumers' choice between using payment cards and cash.

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

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

The corresponding inverse demand function for good $c$ is $p_c\left(Q_c\right)=cF^{-1}\left(1-Q_c\right)$, 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.

Given this setup, in Wang and Wright (forthcoming) we show that the profit-maximizing platform fee is affine if and only if $F$ takes on the generalized Pareto distribution

$$F\left(x\right)=1-\left(1+\lambda\left(\sigma-1\right)\left(x-1\right)\right)^{\frac{1}{1-\sigma}}. \tag{2}$$

Here $\lambda>0$ is the scale parameter and $\sigma<2$ is the shape parameter. Note that the generalized Pareto distribution implies the corresponding demand functions for sellers on the platform are defined by the class of demands with constant curvature of inverse demand

$$Q_c(p_c)=1-F\left(\frac{p_c}{c}\right)=\left(1+\frac{\lambda\left(\sigma-1\right)\left(p_c-c\right)}{c}\right)^{\frac{1}{1-\sigma}}, \tag{3}$$

where $p_c$ is the price of good $c$ on the platform and $Q_c(p_c)$ is the measure of units of good $c$ sold by sellers on the platform at that price.[2] The constant $\sigma$ is the curvature of inverse demand, defined as the elasticity of the slope of the inverse demand with respect to quantity. 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 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 $F\left(x\right)=1-e^{-\lambda(x-1)}$ on the support $[1,\infty)$, with a constant hazard rate $\lambda$. This implies the exponential (or log-linear) demand $Q_c(p_c)=e^{-\frac{\lambda(p_c-c)}{c}}$.

Taking as given that demand belongs to the generalized Pareto class, we allow $c$ to take

[2] This class of demands has been considered by Bulow and Pfleiderer (1983), Aguirre *et al.* (2010), Bulow and Klemperer (2012), and Weyl and Fabinger (2013), among others.

on potentially many different values in $[c_L, c_H]$, with the set of all such values being denoted $C$. The distribution of $c$ on $C$ is denoted $G$. We allow for the possibility $c$ takes only a finite number of values in $C$, or that it is continuously distributed. We let $g_c$ capture the probability (or density in case $G$ is continuous) corresponding to the realization $c$.

The platform incurs a cost of $d \geq 0$ per transaction.[3] If it charges sellers the fee schedule $T(p_c)$, the platform's profit is $\Pi_c = (T(p_c) - d)\, Q_c(p_c)$ for good $c$. Note given Bertrand competition between sellers, the price $p_c$ for good $c$ solves

$$p_c = c + T(p_c). \tag{4}$$

The platform's problem is to choose $T(p_c)$ to maximize

$$\Pi = \sum_{c \in C} g_c \Pi_c. \tag{5}$$

In Wang and Wright (forthcoming), we show that under these assumptions, the optimal fee schedule is affine, given by

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

which maximizes (5). Note the platform's optimal fee schedule has a fixed per-transaction component only if there is a positive cost to the platform of handling each transaction (i.e., $d > 0$). Given $\lambda > 0$ and $\sigma < 2$, the fee schedule is increasing (higher prices imply higher fees are paid) but with a slope less than unity (this implies (4) has a unique solution for any given $c > 0$). The result in (6) also implies the platform can maximize its profit without knowing the distribution $G$ of goods that are traded on its platform. Finally, note in Wang and Wright (forthcoming) we show the solution in (6) is equivalent to charging a different fixed per-transaction fee

$$T_c^* = \frac{\lambda d + c}{\lambda(2 - \sigma)} \tag{7}$$

for each different good $c$, which would be possible if the platform could identify each good $c$ and set a different fee accordingly.

[3] Note that if the platform is a tax authority, then based on the conventional approach of ignoring collection costs, we can set $d = 0$.

# 3 Banning ad valorem fees: analytical results

In this section, we consider whether banning the use of ad valorem fees (i.e., any fees that depend on transaction prices) would harm welfare in an otherwise unregulated environment. Without any restrictions on its pricing, the platform chooses a fee schedule to maximize (5), which results in the affine fee schedule (6). If instead the platform cannot condition on transaction prices in any way, it must choose a single fixed per-transaction fee $T$ across all goods. In this case, it will choose $T$ to maximize

$$\Pi = \sum_{c \in C} g_c \left(T - d\right) Q_c \left(p_c\right), \tag{8}$$

where from (3) and (4), $Q_c \left(p_c\right) = 1 - F\left(\frac{p_c}{c}\right) = 1 - F\left(1 + \frac{T}{c}\right)$. Our problem of interest is thus what happens to total welfare in going from the platform's optimal fee schedule (6), which maximizes (5), to the single fee $\widehat{T}$, which maximizes (8). In other words, is banning ad valorem fees desirable?

The solution to this problem can be found by solving a dual problem, which amounts to the welfare effects of banning third-degree price discrimination. The dual problem involves considering a standard monopolist firm that sells in distinct and identifiable markets, each indexed by $c$. It sets $T_c$ for each market $c$ to maximize profit

$$\Pi_c = \left(T_c - d\right) Q_c \left(T_c\right). \tag{9}$$

In our context, $Q_c \left(T_c\right)$ can be interpreted as the demand function that the platform faces, $T_c$ is the relevant price in each market, and $c$ is a parameter which shifts demand across different markets. The expression of $Q_c \left(T_c\right) = 1 - F\left(1 + \frac{T_c}{c}\right)$ is given by

$$Q_c \left(T_c\right) = \left(1 + \frac{\lambda \left(\sigma - 1\right) T_c}{c}\right)^{\frac{1}{1-\sigma}}, \tag{10}$$

which also belongs to the generalized Pareto class. Note that while the platform deals with different goods, one can equivalently think of the platform providing a homogenous service to different markets with demand in each market varying by $c$, and the platform's output in terms of transaction numbers is addable across these markets.

If third-degree price discrimination is banned, the monopolist will instead choose a uniform price $T$ across all markets to maximize (8). Given our demand specification, the resulting $T$ (denoted $\widehat{T}$) is between $T^*_{c_L}$ and $T^*_{c_H}$, the lowest and highest optimal discriminatory prices.[4] The following duality result follows from the equivalence between solving for the optimal $T_c$ for each observed good (or market) $c$ and solving for the optimal fee schedule in case the latter is an increasing affine function of the transaction price with slope less than one.

**Proposition 1** ***(Duality):*** *The welfare effect of banning a platform from using an ad valorem fee is identical to the welfare effect of banning third-degree price discrimination in the dual problem in which a monopolist can observe the demand of each different market $c$ as given by (10) and charge different (optimal) prices accordingly.*

This duality result allows us to draw on the existing literature on the welfare effects of monopolistic third-degree price discrimination. Consider then a monopolist firm (i.e., the platform) facing demand in market $c$ given by (10) where $T_c$ is the price set by the monopolist in market $c$ and $c$ is a demand shift parameter. When price discrimination is allowed, the monopolist chooses prices $T_c$ for each market $c$ to maximize its profit

$$\sum_{c \in C} g_c(T_c - d)Q_c(T_c).$$

If price discrimination is banned, it must set the same $T_c$ in each market.

In this section, to derive analytical results, we first consider a setting with just two goods $c_L$ and $c_H$ being sold on the platform, with $c_H > c_L$, and $g_c = 1$ for each good. We then extend this analysis to the case with a continuum of goods drawn from the uniform distribution on $[c_L, c_H]$ so that the platform's profit becomes

$$\left(\frac{1}{c_H - c_L}\right) \int_{c_L}^{c_H} [(T_c - d)Q_c(T_c)]dc.$$

In each case, the results we obtain on price discrimination directly apply for the welfare effects of allowing a platform to use ad valorem fees given the duality result above.

## 3.1 Two goods

Aguirre *et al.* (2010) focus on the case in which the monopolist sells in two distinct markets, and will continue to sell in both markets even with uniform pricing. They provide general

[4] Given $\sigma < 2$, it can be shown that $d\Pi_c/dT < 0$ if $T > T_c^*$ and $d\Pi_c/dT > 0$ if $T < T_c^*$, so $\Pi_c$ is single-peaked. In this case, as shown in Nahata, Ostaszewski and Sahoo (1990), the optimal uniform price is bounded above and below by the highest and lowest optimal discriminatory prices. The intuition is straightforward: Since $T_c^*$ is increasing in $c$, $\widehat{T}$ would never be below $T_{c_L}^*$ (above $T_{c_H}^*$) since a higher $T$ (lower $T$) can increase profit in each market in which there is trade on the platform.

conditions to sign the output and welfare effects of price discrimination under non-linear demand. When the curvature of inverse demand function $\sigma$ is common across markets, as it is in our case, a sufficient condition for total output to increase is that $\sigma$ is positive and constant for each good. This implies that in our setting with generalized Pareto demand, price discrimination will always expand output (i.e., the number of transactions on the platform) provided $\sigma > 0$ (so demand is more convex than linear demand). Aguirre *et al.* (2010) also show that when $\sigma > 0$, price discrimination raises (reduces) welfare when demand is log-convex (log-concave) if the discriminatory prices are not far apart. This applies to our setting, but given platforms often deal with transactions of widely different values, we also need to consider cases where discriminatory prices could be far apart.

Define $k \equiv c_H/c_L > 1$ as the measure of dispersion of the two demand levels. We first establish a new result on the welfare effects of third-degree price discrimination. (The proof, which is lengthy, is given in Appendix A.)

**Proposition 2** ***(Welfare effects with two markets)****: Assume that demand is given by (10) and the monopolist has zero marginal costs (i.e., $d = 0$). If there are two markets with demand characterized by $c_L$ and $c_H$, then banning price discrimination across the two markets lowers welfare if demand is exponential ($\sigma = 1$) or log-convex ($1 < \sigma < 2$), and will also lower welfare if demand is log-concave ($\sigma < 1$) provided $k \equiv c_H/c_L$ is sufficiently large.*

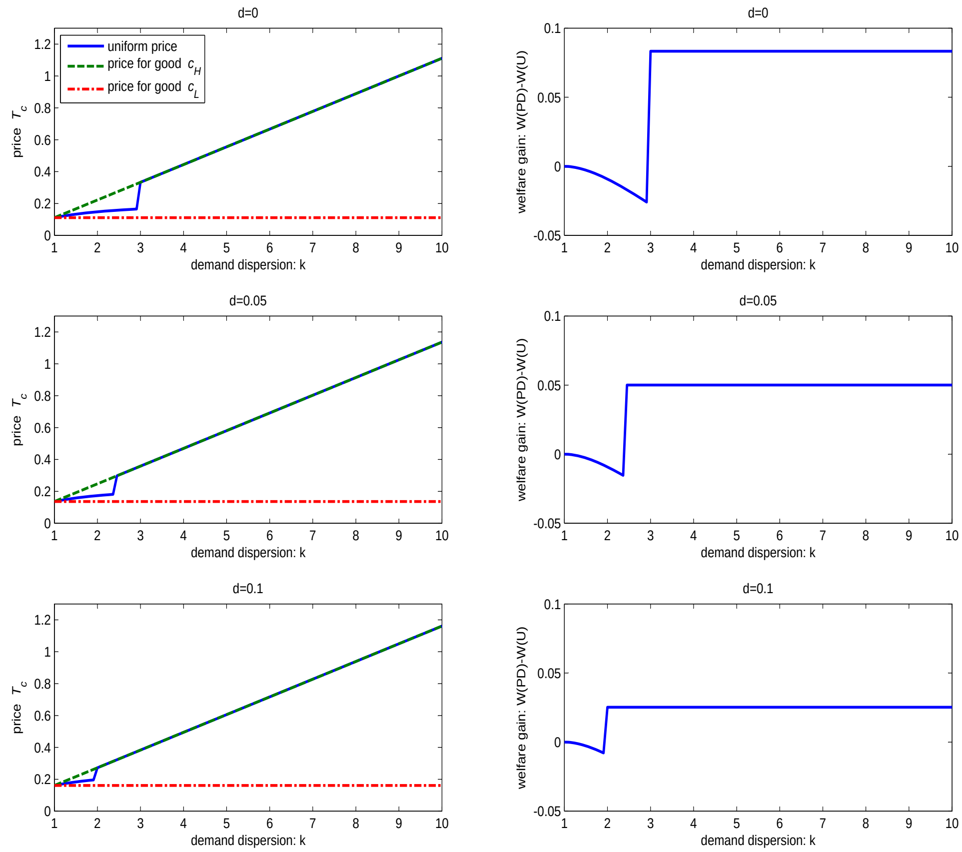
From the duality result in Proposition 1, Proposition 2 implies that for exponential and log-convex demand from the generalized Pareto class given by (10) and for log-concave demand given by (10) but with high enough $k$, banning a monopolist platform from using ad valorem fees will lower welfare.

Proposition 2 covers two types of cases. In the first case, demand is log-concave, so $\sigma < 1$. As an example, consider the case $\sigma = 0$ so that the monopolist faces linear demand for good $c$ given by

$$Q_c(T_c) = \left(1 - \frac{\lambda T_c}{c}\right). \tag{11}$$

Then there is a choke price at which demand becomes zero. As a result, provided $k$ is sufficiently high, a monopolist that can only set a single price will want to stop selling the good $c_L$ (i.e., the low-demand good) by setting its single profit maximizing price at the monopoly level for good $c_H$ (i.e., the high-demand good). This is because continuing to sell the low-demand good will sacrifice too much of the monopolist's profit from the high-demand good. Allowing it to price discriminate will not only increase the monopolist's profit but also consumer surplus and so welfare since sales of low-demand goods are enabled. The condition for this to arise in the linear demand example given in (11) is $k > 3$ when $d = 0$, so that the dispersion of demand across goods does not have to be very high for the result to hold. The proposition shows the same logic holds for any log-concave demand. Figure 1 illustrates this for the linear demand case.[5] It also shows as $d$ increases, the critical value of $k$ declines, so our welfare finding continues to hold for $d > 0$.

Figure 1: Monopoly prices with linear demand (two-good case)

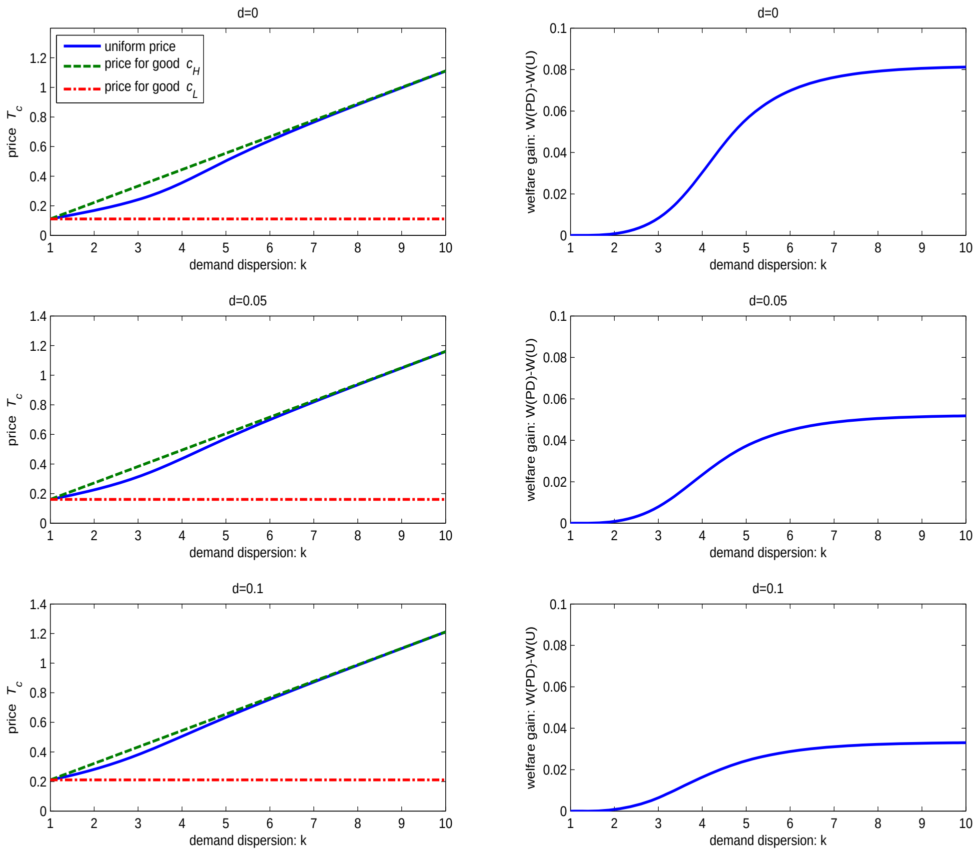



For the exponential and log-convex case, the logic is actually similar even though there is no choke price at which demand becomes zero. When demand for the two goods is sufficiently dispersed, a monopolist that can only set a single price will set a price very close to the monopoly price for the high-demand good, thereby almost ruling out all sales of the low-demand good, and implying a welfare gain of allowing price discrimination. Figure 2 illustrates this property, showing the monopolist's optimal prices with and without price discrimination as $k$ varies in the particular case of exponential demand. In the proof of Proposition 2 we note this property formally by showing that as $k \to \infty$, the uniform price in the absence of price discrimination converges to the price that the monopolist would set for good $c_H$ under price discrimination, both when demand is exponential and when it is

[5] To plot the figure, we normalize $\lambda = 4.5$ and $c_L = 1$. Note that the values of $\lambda$ and $c_L$ just scale the results, but do not affect welfare findings in any of our exercises in this section.

log-convex. This explains the result when the two goods are sufficiently dispersed. On the other hand, when the two goods are not dispersed very much, we know already from Aguirre *et al.* (2010) that price discrimination also raises welfare, which is consistent with our finding in Proposition 2.

Figure 2: Monopoly prices with exponential demand (two-good case)




The welfare effects of allowing price discrimination (i.e., an ad valorem fee schedule) for the different cases captured by Proposition 2 are summarized in Figure 3. The figure considers three different values of $d$ and for a range of values of $k$ and $\sigma$. The dark blue area in the figure indicates a welfare loss due to price discrimination, while the light orange area indicates a welfare gain. In the log-concave demand case ($\sigma < 1$), there is a discrete jump between these two areas when $k$ becomes sufficiently large, reflecting that the low-demand good gets shut down if price discrimination is not allowed. In the log-concave case with $d = 0$, the critical level of $k$ for which welfare is higher under price discrimination than under uniform pricing is $k > 3.5$, so quantitatively we do not require unreasonably high dispersion in the demand for goods to get the welfare result.[6] In the exponential or log-convex case,

[6]While not shown in the figure, the critical level of $k$ for which welfare becomes higher under price discrimination declines as $\sigma$ decreases below $-1$, so the sufficient condition $k > 3.5$ continues to hold.

Figure 3: Welfare comparison for two goods

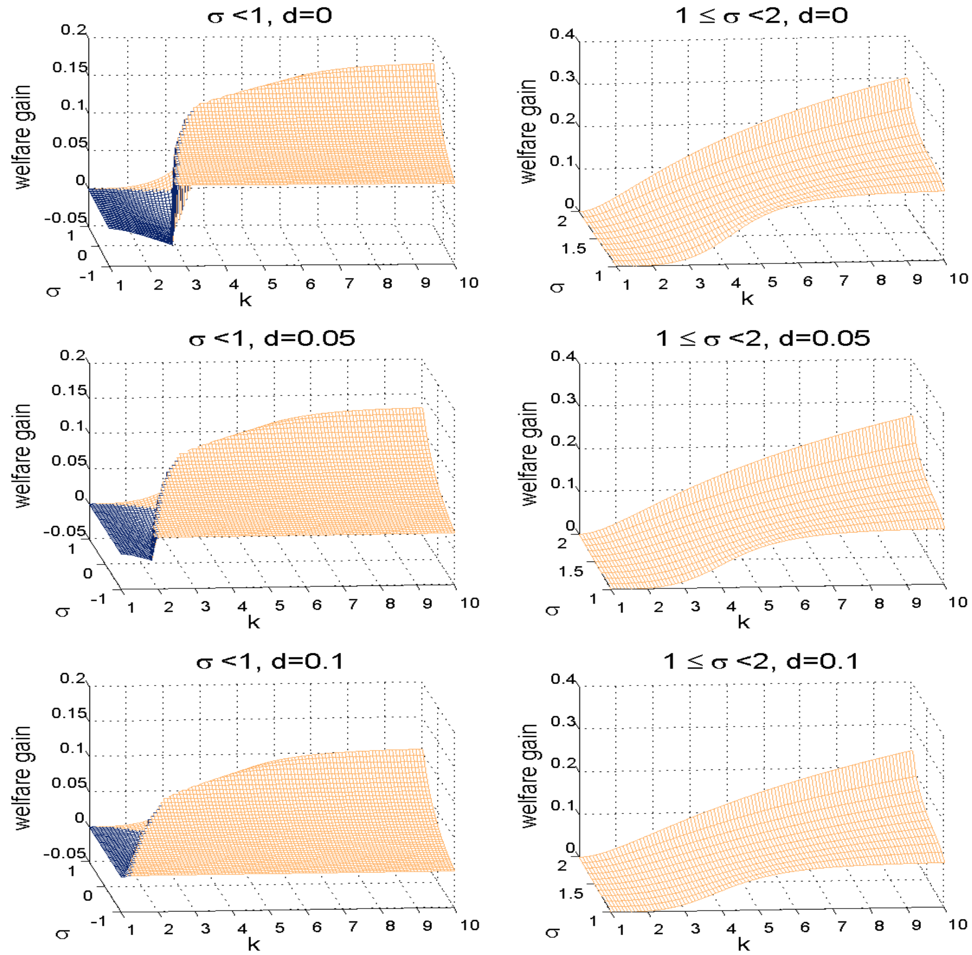



Figure 3 shows that welfare is always higher under price discrimination regardless of the level of dispersion $k$. For both cases, Figure 3 shows that the welfare finding extends to $d > 0$.

## 3.2 Continuum of uniformly distributed goods

The qualitative conclusions on the welfare-gains of price discrimination (or equivalently, allowing ad valorem fees) can hold when there are many markets (or equivalently, goods) rather than just two. In this section we will assume that $c$ is uniformly distributed between $c_L > 0$ and $c_H = kc_L$, with $k \equiv c_H/c_L$ and $k > 1$. We first derive the welfare results for the special case in which $\sigma = 0$ (so demand is linear) and $d = 0$. We then establish that welfare is always higher under price discrimination whenever demand is log-concave provided there is enough dispersion in $c$ when $d = 0$. In particular, we show there is a cutoff level of $c$ equal to $xc_L$ (where $1 < x < k$) such that all markets below the cutoff will be shut down by the monopolist, provided that the dispersion in $c$ is large enough (i.e., $k > k_0$). We show that the threshold $k_0$ depends only on $\sigma$, and the cutoff value $x$ is a constant fraction of $k$ provided $k > k_0$. Finally, we explore graphically the welfare effects of price discrimination for the full range of $\sigma$, allowing $d > 0$.

### 3.2.1 Linear demand

We first consider the special case with $d = 0$ and $\sigma = 0$, so demand is linear. Then the inverse demand faced by the monopolist for good $c$ is

$$T_c(Q_c) = \frac{c(1 - Q_c)}{\lambda}.$$

Then the problem is stated in exactly the same form as the third-degree price discrimination problem analyzed by Malueg and Schwartz (1994), except that we allow inverse demand to be multiplied by a constant positive parameter and we allow that the uniform distribution on $c$ does not have to be centered at unity.[7] It turns out what matters for Malueg and Schwartz's results is the ratio of the highest to lowest value of $c$ in the support of the distribution, i.e. $k$. Therefore reinterpreting the relevant part of their Proposition 1 to our setting, it implies that for large enough dispersion $k > k_0$, some markets are shut down under uniform pricing; in this range, the ratio of welfare under price discrimination to welfare under uniform pricing increases monotonically with dispersion and exceeds 1 when dispersion is sufficiently large.

To calculate these points precisely, define $k_0 > 1$ which solves $1 + 2\ln k_0 = k_0$, so $k_0 \simeq 3.513$. Then the point at which dispersion is sufficiently large for welfare to increase under price discrimination arises when[8]

$$k > \frac{3k_0 - \sqrt{3k_0(4 - k_0)}}{k_0 - 4 + \sqrt{3k_0(4 - k_0)}} \simeq 4.651.$$

Thus, provided there is sufficient dispersion in $c$, welfare is unambiguously higher with price discrimination (or equivalently, with ad valorem fees).

The result is illustrated in Figure 4, which replicates Figure 1 for this continuum case.

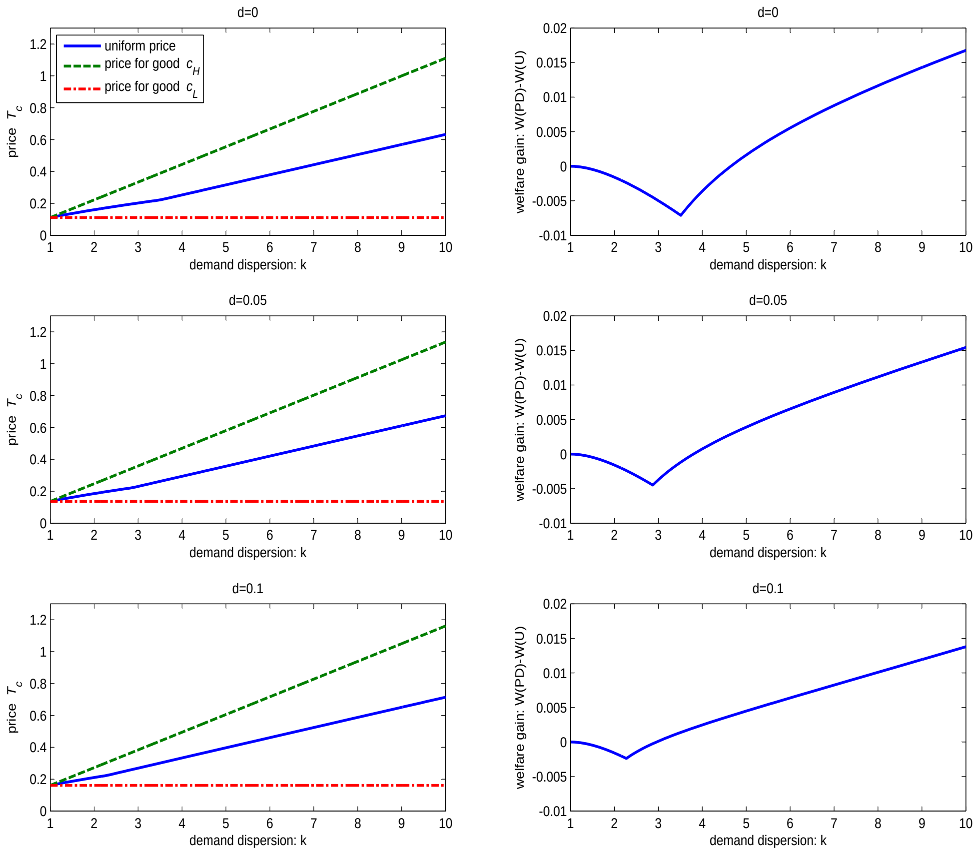
### 3.2.2 Log-concave demand

We can generalize Malueg and Schwartz's result on the positive welfare effects of price discrimination when goods are sufficiently dispersed to log-concave generalized Pareto demands.

[7]Their specification can be obtained by setting $\lambda = 1$, $c = a$, $c_L = 1 - x$ and $c_H = 1 + x$.

[8]There is a typo in Malueg and Schwartz's stated formula for this threshold (in their footnote 17) which does not generate the approximate numerical value they state in the footnote. However, their stated numerical value corresponds to ours, which we derived directly with our specification. I.e. if their threshold is denoted $x_e$ and ours is denoted $k_e$, then it can be checked that $k_e = (1 + x_e) / (1 - x_e)$.

Figure 4: Monopoly prices with linear demand (continuum case)




The demand for each good $c$ is given by (10) and inverse demand is

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

where $\sigma < 1$ given demand is log-concave. With this specification, we obtain the following result on the welfare effects of price discrimination. (The proof is given in Appendix B).

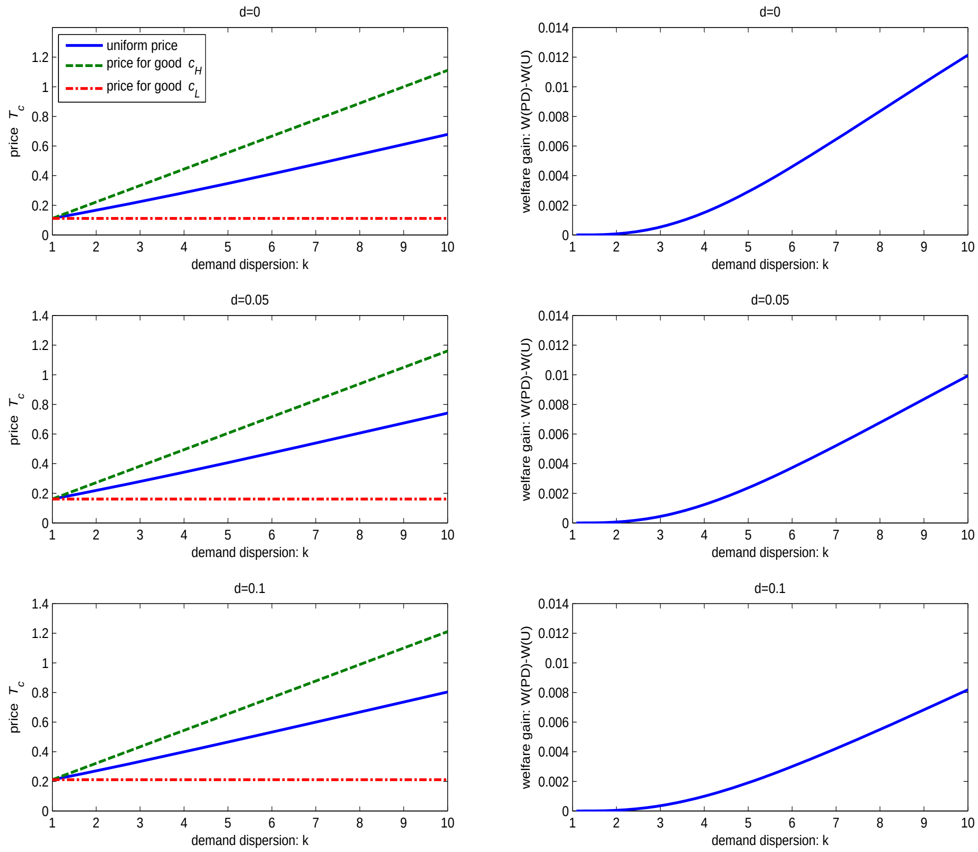
**Proposition 3** ***(Welfare effects with a continuum of markets)****: Assume demand is given by (10) and the monopolist has zero marginal costs (i.e., $d = 0$). If there are a continuum of markets, uniformly distributed between $c_L$ and $c_H$, then banning price discrimination across the markets lowers welfare if demand is log-concave ($\sigma < 1$) provided $k \equiv c_H/c_L$ is sufficiently large.*

In the proof of the proposition we show that the monopolist will set the price such that goods with $c$ lower than some cutoff level $xc_L$ will be dropped, provided that the dispersion of demand across goods is large enough (i.e., $k > k_0$). The threshold $k_0$ depends only on $\sigma$, and the cutoff value $x$ is a constant fraction of the dispersion $k$.

### 3.2.3 Exponential and log-convex demand

For cases with log-convex demand, we present the results graphically. The case corresponding to Figure 2, with exponential demand, is given in Figure 5.

Figure 5: Monopoly prices with exponential demand (continuum case)




More generally, Figure 6 shows that provided $k$ is large enough when demand is log-concave, and for all $k$ when demand is exponential or log-convex, welfare is higher under price discrimination (and so when a platform can use ad valorem fees). In the log-concave case with $d = 0$, the critical level of $k$ for which welfare is higher under price discrimination than under uniform pricing is $k > 5$, and the critical value of $k$ declines as $d$ increases, so the dispersion of demand across goods does not have to be very high for price discrimination to generate higher welfare than uniform pricing. In the exponential or log-convex case, Figure 6 suggests welfare is always higher under price discrimination regardless of the level of $k$ and $d$.

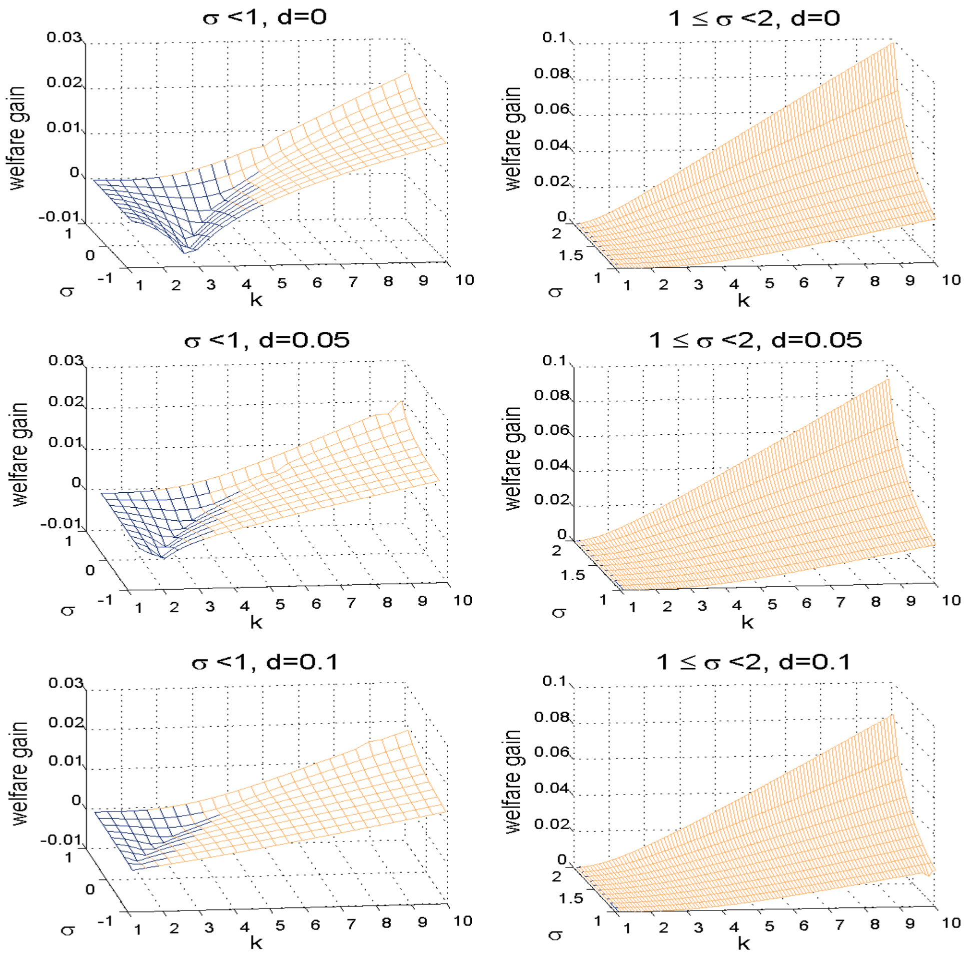



Figure 6: Welfare comparison for a continuum of goods

# 4 Banning ad valorem fees: calibrated model

In practice, platforms deal with thousands of different goods, and $c$ will not be uniformly distributed across these goods. For example, there are typically many more transactions using Visa debit cards in the \$10-\$50 range than in the \$60-\$100 range. The same is true for goods traded on Amazon's marketplace, where the sales distribution is highly skewed. The welfare effects of price discrimination in such settings are largely unexplored in the literature. In this section, we wish to work out the welfare effects of banning ad valorem fees (i.e., banning price discrimination in the equivalent setting where different $c$ captures distinct markets with observably different demand) using realistic sales distributions. The data we use are from Visa signature debit card transactions and DVD listings on Amazon's marketplace. In both cases, the platforms have adopted affine fee schedules, as shown in Table 1.

## 4.1 Methodology

First, we illustrate how our theoretical model can be calibrated to real world data. Rather than assume there is a unit mass of buyers for each good, we allow for the possibility that for each good there can be a different number of potential buyers so that even goods with the same $c$ can sell different amounts.

The number of transactions for a distinct good $i$ with cost $c$ is $Q_{i,c} = n_{i,c}Q_c$ and the platform makes a profit $\Pi_{i,c} = n_{i,c}\Pi_c$, where $Q_c$ and $\Pi_c$ are the quantity and profit expressions from Section 2 based on a unit mass of potential buyers, and $n_{i,c}$ is the number of potential buyers for good $i$ with cost $c$. We denote the number of distinct goods with cost $c$ as $n_c$. A platform's total profit is therefore

$$\Pi = \sum_{c\in C}\sum_{i=1}^{n_c} n_{i,c}\Pi_c. \tag{12}$$

Given (12), all our previous analysis holds except that we need to change the mass $g_c$ to $\sum_{i=1}^{n_c} n_{i,c}$. This follows because, as shown in Wang and Wright (forthcoming), the optimal platform fee does not depend on $g_c$. Accordingly, an affine fee schedule (such as those in Table 1) can be rationalized by the platform facing generalized Pareto demands, and the profit maximizing platform fee is still given by (6).

Given an observed platform fee schedule $T(p_c) = a_0 + a_1 p_c$, (6) implies that we can uniquely identify the values of $\lambda$ and $d$ for a given value of $\sigma$. Our welfare comparisons will then consider all the possible values of $\sigma$.[9] Note that

$$\frac{\lambda d}{1+\lambda(2-\sigma)} = a_0, \quad \frac{1}{1+\lambda(2-\sigma)} = a_1,$$

so $\lambda = (1/a_1 - 1)/(2-\sigma)$ and $d = a_0/\lambda + a_0(2-\sigma)$. Given the value of $\lambda$, and the observed price $p_c$ and quantity $Q_{i,c}$ for each good traded on the platform, we can then identify the number of potential buyers $n_{i,c}$ for each good. Substituting $T(p_c) = a_0 + a_1 p_c$ into

$$Q_{i,c} = n_{i,c}\left(1+\frac{\lambda(\sigma-1)T_c}{c}\right)^{\frac{1}{1-\sigma}},$$

[9] Alternatively, we could pin down $d$ and $\sigma$ for a given value of $\lambda$, and then conduct welfare comparisons based on different values of $\lambda$. However, given that $\sigma$ determines the shape of demand functions, deriving welfare results based on different values of $\sigma$ is more informative. It allows us to compare our results with those in the literature that typically assume specific demand functions (e.g., linear demand).

we derive

$$n_{i,c} = \frac{Q_{i,c}}{\left[1 + \frac{\lambda(\sigma-1)(a_0+a_1 p_c)}{(1-a_1)p_c - a_0}\right]^{\frac{1}{1-\sigma}}}. \tag{13}$$

With $n_{i,c}$ determined, we can use the weight $\sum_{i=1}^{n_c} n_{i,c}$ in place of $g_c$ when calculating profit and welfare. Our theory also allows us to identify the lower bound of $\sigma$ from the data. Recall when $\sigma < 1$, the generalized Pareto demand has finite support on $[1, 1+\frac{1}{\lambda(1-\sigma)}]$. This means that if we observe any good with positive sales, its price has to satisfy $\frac{p(c)}{c} < 1+\frac{1}{\lambda(1-\sigma)}$. Since $p(c) = c + T^*(p_c)$, this requires that $\frac{T^*(p_c)}{c} < \frac{1}{\lambda(1-\sigma)}$. Substituting in that $T^*(p_c) = a_0 + a_1 p_c$, $c = (1-a_1)p_c - a_0$ and the expression for $\lambda$ from above, the equivalent inequality can be written as $\sigma > 1 + a_1 - \frac{a_1(1-a_1)p_c}{a_0}$. Thus, the minimum price we observe in the data pins down the minimum value of $\sigma$ that our model permits.

## 4.2 Visa debit cards

We use data from the Diary of Consumer Payment Choice (DCPC), conducted in October 2012 by the Boston, Richmond, and San Francisco Federal Reserve Banks to calibrate the model. The DCPC collects consumer payments data on the dollar value of purchases, the payment instrument used, and the category of expense. A national representative sample of 2,468 U.S. respondents were selected, who each recorded all their payments over a three-day period. Since respondents were spread over the entire month of October 2012, this sampling methodology provides reasonable probability estimates of all consumers. For transactions made with payment cards, respondents were asked to report the dollar amount, the exact card type and the card network's brand name.

Based on the DCPC data, we identify 1,048 Visa signature debit card transactions in four distinct market categories, namely retail, restaurant, gas station, and small ticket, to form our empirical transaction distributions. For each market category, we use the interchange fee schedule published by Visa (shown in Table 1) to infer its platform pricing. Given merchant acquirers are highly competitive in the U.S. market, the interchange fee schedules posted by Visa mirror very closely the actual fee schedules passed onto sellers.

Figure 7 plots the raw density distribution of transaction prices in each market category. The distributions are quite skewed. Based on the raw transaction distributions and the fee schedules, we then numerically calculate percentage welfare gains under the observed affine fee schedule and the counterfactual optimal uniform per-transaction fee for each possible value of $\sigma$ assuming the underlying demand takes the generalized Pareto form. The results are presented in Figure 8.

Figure 8 shows that in three out of the four markets, welfare is consistently higher when

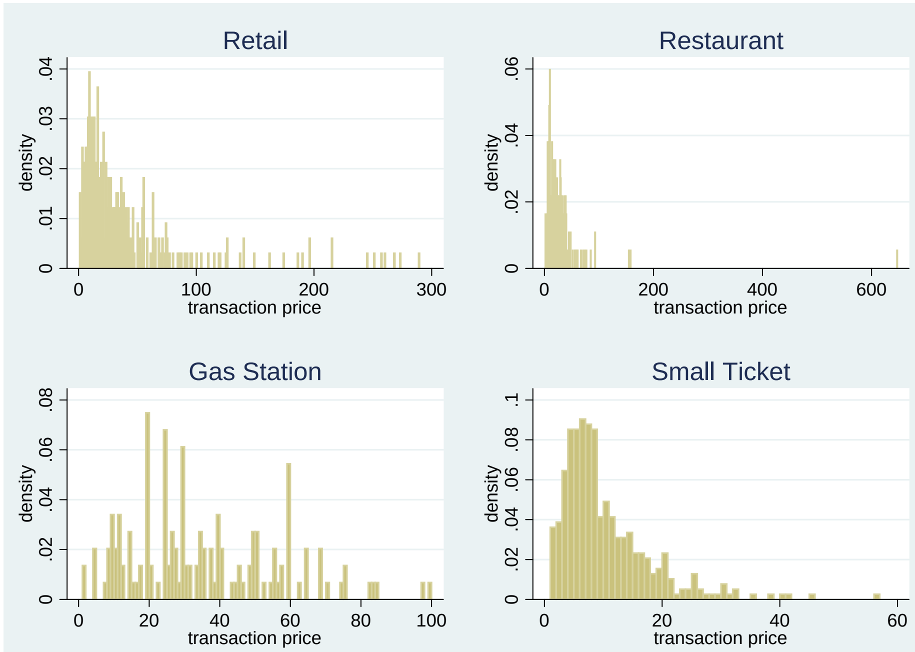



Figure 7: Visa Signature Debit Card Transaction Distributions

ad valorem fees are allowed for any possible value of $\sigma$.[10] The only exception is for the Restaurant market for lower values of $\sigma$. However, this result is driven by a single outlier which has an unusually large transaction price, as can be seen from Figure 7. If that outlier is removed from the sample, welfare would also be consistently higher in the Restaurant market when ad valorem fees are allowed.

Note that the percentage welfare gain (or loss) from allowing ad valorem fees is calculated by assuming that the platform only incurs a marginal cost $d$ but not an overall fixed cost $K$. The percentage change in welfare would be even higher once a positive level of $K$ is taken into account.[11] Moreover, the absolute level of welfare change is likely to be substantial given the size of the payment card industry. In 2011, debit cards were used in 49 billion transactions for a total value of \$1.8 trillion in the U.S. market, in which 60 percent were signature debit card transactions, with Visa's share of these being about 75 percent.

[10]Note that the implied value of $d$ from our calibrated model varies from zero in the limit as $\sigma$ tends to its highest allowed value (i.e., 2) up to 16 cents as $\sigma$ tends to its lowest allowed value as determined by the lowest observed price. This is consistent with the Federal Reserve's study based on comprehensive cost surveys of debit card issuers, as mandated by the Durbin Amendment to the Dodd-Frank Act. According to the study, most issuers incur a marginal cost no more than 21 cents per transaction (See Federal Reserve Board, 2011).

[11]It is easy to see for any $W_1 > W_2$, $(W_1 - K)/(W_2 - K)$ is an increasing function of $K$.

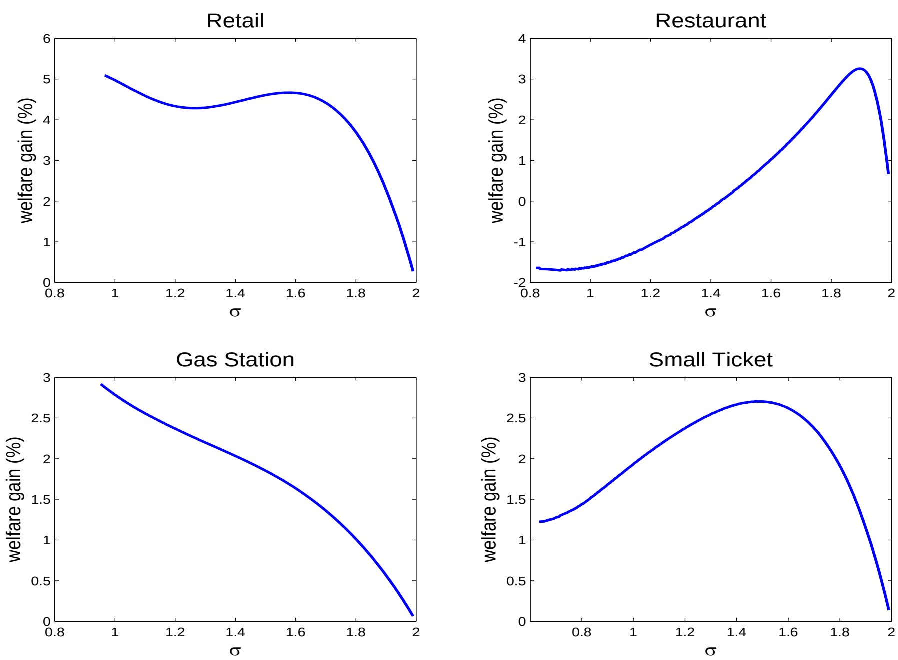



Figure 8: Visa Signature Debit Cards: Welfare Gain from Ad valorem Fees

## 4.3 Amazon marketplace

We also calibrate our model using data from Amazon's marketplace. We focus on DVDs given that it is a well defined market category for which we can be sure all the goods identified are subject to the same fee schedule, and also since we can collect consistent sales ranks for this category. Using a web robot, we collected data on every DVD that was listed under "Movies & TV" on Amazon's marketplace in January 2014. We selected "New" under "Condition" and de-selected the "Out of Stock" option, and ended up with a total of 295,171 distinct items. The data collected include the title, unique ASIN number identifying the DVD, the price, and sales rank of each DVD.[12] Given shipping fees are often not included in the listed price, we also separately collected data on only those items where the listed price included free shipping, resulting in a sample with 191,280 distinct items. Since some DVDs are listed with extreme prices, we restrict our sample to DVDs selling for under $1,000, which includes around 99% of the items collected. For robustness, we also tried alternative price limits, including $500 and $2,000, and the results are very similar.[13]

[12] The price is taken as the price posted at Amazon's marketplace for the DVD. It is the price a buyer will face when they add the item to their cart and go to the checkout – i.e., the "buy-box" price.

[13] A concern with extreme DVD prices is that the prices listed are unlikely to reflect the prices at which transactions actually take place. For instance, some sellers post extreme prices as placeholders to avoid a temporary delisting when they are out of stock or away for vacation. Others may be errors in the seller's entry of its prices.

Given we do not directly observe the sales of each DVD, we use a power law to infer it from the sales rank, so $Q_{i,c} = aR_{i,c}^{-\phi}$, where $Q_{i,c}$ is the estimated sales of an item and $R_{i,c}$ is the corresponding sales rank.[14] The parameter $a$ does not affect our results, so we normalize it by setting $a = 1$. We try different values for the parameter $\phi$, including $\phi = 0$ (where sales rank is assumed to be irrelevant), $\phi = 1$ (Zipf's law) and $\phi = 1.7$ (which is the number suggested by Smith and Telang (2009) in an experimental study on DVD sales on Amazon, although it implies very little weight is placed on items with sales ranks below the top ten).

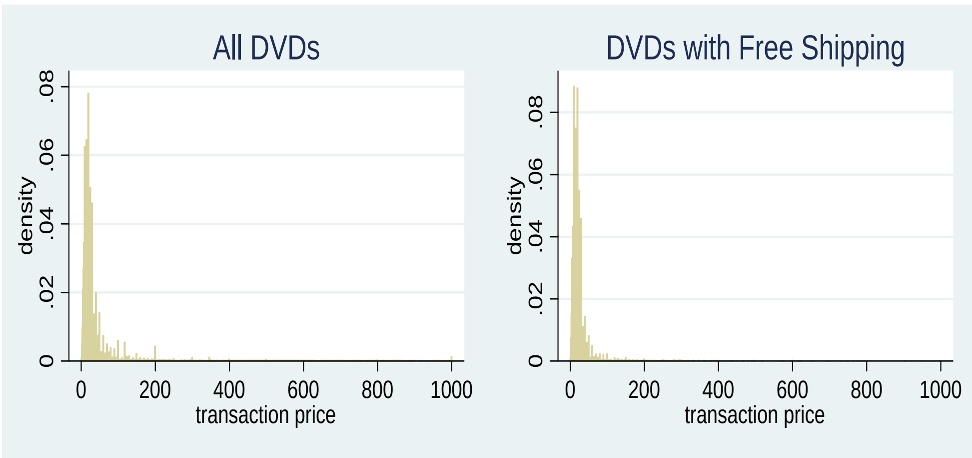



Figure 9: DVD Sales Distribution

Figure 9 plots the density of items listed at each price which corresponds to the sales distribution under the assumption that $\phi = 0$. The distributions are highly skewed with a majority of items listed at prices below \$50. With $\phi = 1$ or $\phi = 1.7$, the distributions become even more skewed.

Based on each of the sales distributions and the fee schedule from Table 1, we numerically calculate percentage welfare gains under the observed affine fee schedule and the counterfactual optimal uniform per-transaction fee for each possible value of $\sigma$ assuming the underlying demand takes the generalized Pareto form. The results are presented in Figure 10, which shows that once sales ranks are taken into account, welfare is consistently higher under an affine fee schedule than under a uniform per-transaction fee.[15]

[14] Power law distributions are widely used to describe rank data, with the well-known "Zipf's law" being a special case. See Chevalier and Goolsbee (2003) for detailed discussions as well as an application to online sales data.

[15] In the Amazon case, the implied value of $d$ from our calibrated model varies from zero up to \$1.65 as $\sigma$ varies from its highest possible value to its lowest.

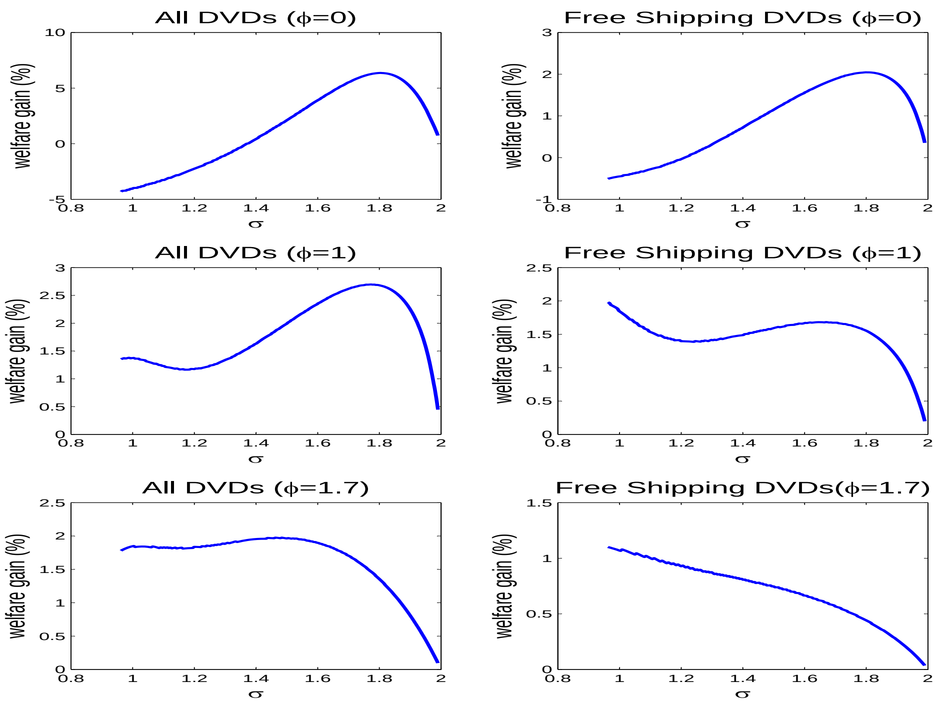



Figure 10: DVDs: Welfare Gain from Ad valorem Fees

# 5 Concluding remarks

Many platforms that facilitate transactions between buyers and sellers charge ad valorem fees in which fees depend on the transaction price set by sellers. Given these platforms do not incur significant costs that vary with transaction prices, their use of ad valorem fees has raised controversies about the efficiency of this practice. For policymakers who would want to align platform fees with costs but are concerned about directly regulating fee levels, it seems natural to consider regulating fee structures, such as banning platforms from using ad valorem fees. However, we have shown that such regulation tends to have negative welfare outcomes, including when we calibrate our model to data on sales of DVDs on Amazon's marketplace and data for Visa signature debit card transactions. Therefore, caution should be taken when policymakers consider this option. A similar result would also apply to a government that wanted to maximize tax revenue—welfare would be higher when it does so using an ad valorem tax. The key feature that drives these results is that when a market involves many different goods that vary widely in their costs and values, ad valorem fees and taxes represent an efficient form of price discrimination. In comparison, uniform fees or taxes could adversely affect low-cost low-value goods so that the total welfare is reduced.

There are several avenues for future research. First, as Shy and Wang (2011) showed, another reason why banning ad valorem fees could lower welfare is that ad valorem fees help to mitigate double marginalization. This suggests that allowing for imperfect competition between sellers would add to the welfare loss of banning ad valorem fees. Using our demand assumptions, one could analyze a ban on ad valorem fees in their environment to evaluate the overall effects. However, in this case, affine fee schedules will not necessarily maximize platform profits, although they may still do so approximately. Moreover, one would no longer be able to rely on the duality result which allowed us to draw on the existing literature on third-degree price discrimination. Thus, combining these two mechanisms in a single model would be a challenging exercise for future research. Second, one could consider demand functions outside the generalized Pareto class. The reason that we focused on the generalized Pareto demand is because it covers a broad family of commonly used demand functions that rationalize platforms' use of ad valorem fees. In reality, however, ad valorem fees may be used as an approximation to more complicated optimal fee schedules. Therefore, it might be useful to consider demand specifications outside the generalized Pareto class and conduct robustness checks for our results. Finally, it might be interesting to consider alternative regulations on platform fees, such as allowing for ad valorem fees but with a cap for high-value transactions. Such a regulation may achieve better welfare outcomes, even though there would be the additional complication of choosing the appropriate level of the cap.

# References


[1] Aguirre, Iñaki (2006). "Monopolistic Price Discrimination and Output Effect under Conditions of Constant Elasticity Demand." *Economics Bulletin*, 4(23), 1–6.

[2] Aguirre, Iñaki (2008). "Output and Misallocation Effects in Monopolistic Third-Degree Price Discrimination." *Economics Bulletin*, 4(11), 1–11.

[3] Aguirre, Iñaki, Simon Cowan and John Vickers (2010). "Monopoly Price Discrimination and Demand Curvature." *American Economic Review*, 100, 1601–1615.

[4] Brennan, G., and J. M. Buchanan (1977). "Towards a Tax Constitution for Leviathan." *Journal of Public Economics*, 8(3), 255–273.

[5] Brennan, G., and J. M. Buchanan (1978). "Tax Instruments as Constraints on the Disposition of Public Revenues." *Journal of Public Economics*, 9(3), 301–318.

[6] Bulow, Jeremy and Paul Pfleiderer (1983). "A Note on the Effect of Cost Changes on Prices." *Journal of Political Economy*, 91, 182–185.

[7] Bulow, Jeremy and Paul Klemperer (2012). "Regulated Prices, Rent Seeking, and Consumer Surplus." *Journal of Political Economy*, 120, 160–186.

[8] Chevalier, Judith and Austan Goolsbee (2003). "Measuring Prices and Price Competition Online: Amazon.Com and Barnesandnoble.Com." *Quantitative Marketing and Economics*, 1(2), 203–222.

[9] Cowan, Simon (2007). "The Welfare Effects of Third-Degree Price Discrimination with Nonlinear Demand Functions." *RAND Journal of Economics*, 38(2), 419–28.

[10] Einav, Liran, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan (2015). "Assessing Sale Strategies in Online Markets using Matched Listings." *American Economic Journal: Microeconomics*, 7, 215-247.

[11] Federal Reserve Board (2011). *Regulation II (Debit Card Interchange Fees and Routing) Final Rule,* June 29, www.federalreserve.gov/newsevents/press/bcreg/20110629a.htm.

[12] 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.

[13] Gaudin, Germain and Alexander White (2014a). "On the Antitrust Economics of the Electronic Books Industry." Working Paper.

[14] Gaudin, Germain and Alexander White (2014b). "Unit vs. Ad valorem Taxes under Revenue Maximization." Working Paper.

[15] Johnson, Justin. "The Agency Model and MFN Clauses." *Review of Economic Studies*, forthcoming.

[16] Layson, Stephen (1994). "Market Opening under Third-Degree Price Discrimination." *Journal of Industrial Economics*, 42(3), 335–40.

[17] Malueg, David and Marius Schwartz (1994). "Parallel Imports, Demand Dispersion, and International Price Discrimination." *Journal of International Economics*, 37, 167–195.

[18] Nahata, Babu, Krzysztof Ostaszewski, and Prasanna K. Sahoo (1990). "Direction of Price Changes in Third-Degree Price Discrimination." *American Economic Review*, 80(5), 1254–58.

[19] Schmalensee, Richard (1981). "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination." *American Economic Review*, 71(1), 242–47.

[20] Schwartz, Marius (1990). "Third-Degree Price Discrimination and Output: Generalizing a Welfare Result." *American Economic Review*, 80(5), 1259–62.

[21] Shy, Oz and Zhu Wang (2011). "Why Do Payment Card Networks Charge Proportional Fees?" *American Economic Review*, 101, 1575–1590.

[22] Smith, Michael and Rahul Telang (2009). "Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy." *MIS Quarterly*, 33, 321–338.

[23] Standing Senate Committee on Banking, Trade, and Commerce. *Transparency, Balance and Choice: Canada's Credit Card and Debit Card Systems*, Canada, 2009.

[24] Summers, Bruce (2012). "Facilitating Consumer Payment Innovation through Changes in Clearing and Settlement." *Consumer Payment Innovation in the Connected Age*, Conference Proceedings, Federal Reserve Bank of Kansas City.

[25] Varian, Hal (1985). "Price Discrimination and Social Welfare." *American Economic Review*, 75(4), 870–75.

[26] Wang, Zhu and Alexander Wolman (2016). "Payment Choice and Currency Use: Insights from Two Billion Retail Transactions." *Journal of Monetary Economics*, 84, 94-115.

[27] Wang, Zhu and Julian Wright. "Ad Valorem Platform Fees, Indirect Taxes, and Efficient Price Discrimination." *RAND Journal of Economics*, forthcoming.

[28] Weyl, Glen and Jean Tirole (2012). "Market Power Screens Willingness-to-Pay." *Quarterly Journal of Economics*, 127(4), 1971-2003.

[29] 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: Proof of Proposition 2

We consider three cases.

**(i) Demand is log-concave:** Suppose demand is log-concave so $\sigma < 1$. Then there is a choke price $T_c' = c/\left(\lambda\left(1-\sigma\right)\right)$ at which demand becomes zero for market $c$. Let $c_L$ be fixed and consider increasing $k$ and so $c_H$. Let $z = \left(\frac{1}{2-\sigma}\right)\left(1-\frac{1-\sigma}{2-\sigma}\right)^{\frac{1}{1-\sigma}}$ where $0 < z < e^{-1}$ given $\sigma < 1$. Under price discrimination, the profit from the high-demand market is $c_H z/\lambda \to \infty$ as

$k \to \infty$. The profit from the low-demand market is fixed at $c_L z/\lambda$. Total profit is unbounded as $k$ increases. On the other hand, with a uniform price the profit is bounded if both markets continue to operate since the price cannot exceed the choke price for market $c_L$, which is $c_L / (\lambda (1-\sigma))$. Therefore, there exists a high enough $k$ such that the monopolist will give up on the low-demand market if it is forced to set a single price. The threshold $k_0$ such that the monopolist will no longer keep the low-demand market open whenever $k \geq k_0$ is determined by

$$\left(\frac{1}{2-\sigma}\right)^{\frac{2-\sigma}{1-\sigma}} = \frac{1}{(k_0+1)} \left[ \left(\frac{k_0+\sigma}{k_0+1}\right)^{\frac{1}{1-\sigma}} + \left(\frac{k_0\sigma+1}{k_0+1}\right)^{\frac{1}{1-\sigma}} \right], \tag{14}$$

which is obtained by comparing the monopolist's profit with and without shutting down the low-demand market under uniform pricing. Note $k_0$ only depends on $\sigma$. For example, in the case of linear demand, solving (14) with $\sigma = 0$ implies $k_0 = 3$. With price discrimination, the monopolist will set the same price for the high-demand market as it would under uniform pricing, and set a lower price for the low-demand market to ensure it operates, thereby generating additional profit, consumer surplus and welfare.

**(ii) Demand is exponential:** Suppose demand is exponential so $\sigma = 1$. Then there is no choke price at which demand becomes zero. We compare welfare directly. Welfare from market $c$ is

$$W_c = \int_0^{Q_c} T_c(Q)\, dQ = \int_0^{Q_c} (-\frac{c}{\lambda} \ln Q) dQ,$$

so that $W_c(T_c^*) = 2ce^{-1}/\lambda$ under price discrimination given $Q_c(T_c) = e^{-\frac{\lambda T_c}{c}}$ and (7) with $d = 0$ and $\sigma = 1$. Therefore, welfare from both markets under price discrimination is $W_{PD} = 2c_L(1+k)\, e^{-1}/\lambda$.

Now consider welfare without price discrimination. The monopolist will set the uniform price $T$ to maximize

$$\Pi = T\left(e^{-\frac{\lambda T}{c_L}} + e^{-\frac{\lambda T}{c_H}}\right).$$

The optimal uniform price $\widehat{T}$ solves the first-order condition

$$e^{-\frac{\lambda \widehat{T}}{c_L}} \left(1 - \frac{\lambda \widehat{T}}{c_L}\right) + e^{-\frac{\lambda \widehat{T}}{c_H}} \left(1 - \frac{\lambda \widehat{T}}{c_H}\right) = 0.$$

The solution can be written as $\widehat{T} = \rho c_L/\lambda$, where $\rho$ solves

$$(1-\rho)\, e^{-\rho} + \left(1 - \frac{\rho}{k}\right) e^{-\frac{\rho}{k}} = 0.$$

Note $\rho$ is just a function of $k$. Welfare under uniform pricing is

$$\begin{aligned} W_U &= \widehat{T}\left(e^{-\frac{\lambda \widehat{T}}{c_L}} + e^{-\frac{\lambda \widehat{T}}{c_H}}\right) + c_L\left(\frac{e^{-\frac{\lambda \widehat{T}}{c_L}}}{\lambda}\right) + c_H\left(\frac{e^{-\frac{\lambda \widehat{T}}{c_H}}}{\lambda}\right) \\ &= \frac{c_L}{\lambda}\left((k+\rho)\, e^{-\frac{\rho}{k}} + (1+\rho)\, e^{-\rho}\right). \end{aligned}$$

Therefore,

$$W_{PD} - W_U = \frac{c_L}{\lambda}\left(2\,(1+k)\, e^{-1} - (1+\rho)\, e^{-\rho} - (k+\rho)\, e^{-\frac{\rho}{k}}\right).$$

Since $\rho$ is just a function of $k$, and the term in brackets in $W_{PD} - W_U$ is just a function of $\rho$ and $k$, the sign of $W_{PD} - W_U$ just depends on $k$. We can verify $W_{PD} - W_U > 0$ for all $k > 1$, and so welfare is higher under price discrimination.

The limit case as $k \to \infty$ provides some insight into what happens as demands become more dispersed across markets. In the limit as $k \to \infty$, it can be shown $\rho \to k$. Accordingly, $\widehat{T} \to kc_L/\lambda = c_H/\lambda = T^*_{c_H}$ and $W_{PD} - W_U \to 2c_L e^{-1}/\lambda$. In other words, for large $k$, the uniform price converges to the optimal discriminatory price that the monopolist would set for the high-demand market, and the welfare gain converges to the discriminatory profit that the monopolist would make from the low-demand market.

**(iii) Demand is log-convex:** Suppose demand is log-convex so $1 < \sigma < 2$. Welfare from market $c$ is

$$W_c = \int_0^{Q_c} T_c(Q)\, dQ = \int_0^{Q_c} \frac{c\,(1 - Q_c^{1-\sigma})}{\lambda\,(1-\sigma)} dQ, \tag{15}$$

so that

$$W_c(T^*_c) = \frac{c}{\lambda\,(\sigma-1)}\left[\left(\frac{1}{2-\sigma}\right)^{2+\frac{1}{1-\sigma}} - \left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}}\right]$$

under price discrimination where demand given in (10) and price $T^*_c$ given in (7) have been substituted into (15). Therefore, welfare from both markets under price discrimination is

$$W_{PD} = \frac{c_L\,(1+k)}{\lambda\,(\sigma-1)}\left[\left(\frac{1}{2-\sigma}\right)^{2+\frac{1}{1-\sigma}} - \left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}}\right]. \tag{16}$$

Now consider welfare without price discrimination. The monopolist will set the uniform price $T$ to maximize

$$\underset{T}{Max}\ \Pi = T\left[\left(1 + \frac{\lambda\,(\sigma-1)\,T}{c_L}\right)^{\frac{1}{1-\sigma}} + \left(1 + \frac{\lambda\,(\sigma-1)\,T}{c_H}\right)^{\frac{1}{1-\sigma}}\right].$$

The optimal uniform price $\widehat{T}$ solves the first-order condition

$$\begin{aligned}&\left(1+\frac{\lambda\left(\sigma-1\right)\widehat{T}}{c_L}\right)^{\frac{1}{1-\sigma}}+\left(1+\frac{\lambda\left(\sigma-1\right)\widehat{T}}{c_H}\right)^{\frac{1}{1-\sigma}}\\ =\;&\frac{\lambda\widehat{T}}{c_L}\left(1+\frac{\lambda\left(\sigma-1\right)\widehat{T}}{c_L}\right)^{\frac{\sigma}{1-\sigma}}+\frac{\lambda\widehat{T}}{c_H}\left(1+\frac{\lambda\left(\sigma-1\right)\widehat{T}}{c_H}\right)^{\frac{\sigma}{1-\sigma}}.\end{aligned}$$

The solution can be written as $\widehat{T}=\rho c_L/\left(\lambda\left(\sigma-1\right)\right)$, where for any given $\sigma$, the term $\rho$ is just a function of $k$ which solves

$$(1+\rho)^{\frac{1}{1-\sigma}}-\frac{\rho}{\sigma-1}(1+\rho)^{\frac{\sigma}{1-\sigma}}+\left(1+\frac{\rho}{k}\right)^{\frac{1}{1-\sigma}}-\frac{\rho}{k\left(\sigma-1\right)}\left(1+\frac{\rho}{k}\right)^{\frac{\sigma}{1-\sigma}}=0.$$

Welfare under uniform pricing is

$$W_U=\frac{c_L}{\lambda\left(\sigma-1\right)}\left[\frac{(1+\rho)^{\frac{2-\sigma}{1-\sigma}}}{2-\sigma}-(1+\rho)^{\frac{1}{1-\sigma}}+\frac{k\left(1+\frac{\rho}{k}\right)^{\frac{2-\sigma}{1-\sigma}}}{2-\sigma}-k\left(1+\frac{\rho}{k}\right)^{\frac{1}{1-\sigma}}\right]. \tag{17}$$

Since $\rho$ is just a function of $k$ and the term in brackets in $W_U$ is just a function of $\rho$ and $k$ for any given $\sigma$, the sign of $W_{PD}-W_U$ just depends on $k$ for any particular $\sigma$. Evaluating (16) and (17) confirms $W_{PD}-W_U>0$ for all $k>1$ for any $1<\sigma<2$, so that welfare is higher under price discrimination.

Again, the limit case as $k\to\infty$ provides some insight into what happens as demands become more dispersed across markets. In the limit as $k\to\infty$, it can be shown that $\rho\to k\left(\sigma-1\right)/\left(2-\sigma\right)$. Accordingly, $\widehat{T}\to kc_L/\left(\lambda\left(2-\sigma\right)\right)=c_H/\left(\lambda\left(2-\sigma\right)\right)=T^*_{c_H}$ and $W_{PD}-W_U\to\frac{c_L}{\lambda(\sigma-1)}\left[\left(\frac{1}{2-\sigma}\right)^{2+\frac{1}{1-\sigma}}-\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}}\right]$. In other words, for large $k$, the uniform price converges to the optimal discriminatory price that the monopolist would set for the high-demand market, and the welfare gain converges to the discriminatory profit that the monopolist would make from the low-demand market.

# Appendix B: Proof of Proposition 3

We break the proof up into three steps.

**(i) Price discrimination is allowed:** If price discrimination is allowed, the monopolist solves the following problem for each market $c$:

$$\underset{T_c}{Max}\ \Pi_c = T_c \left(1 - \frac{\lambda (1-\sigma) T_c}{c}\right)^{\frac{1}{1-\sigma}}.$$

The first-order condition yields the optimal price

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

The corresponding demand in market $c$ is

$$Q_c(T_c^*) = \left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}},$$

and the monopolist's profit is

$$\Pi_c = \frac{c}{\lambda}\left(\frac{1}{2-\sigma}\right)^{\frac{2-\sigma}{1-\sigma}}.$$

The resulting welfare from market $c$ is

$$\begin{aligned} W_c &= \int_0^{Q_c} T_c(Q)\, dQ = \int_0^{Q_c} \frac{c\left(1-Q^{1-\sigma}\right)}{\lambda(1-\sigma)} dQ \\ &= \frac{c}{\lambda(1-\sigma)}\left[\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}} - \left(\frac{1}{2-\sigma}\right)^{2+\frac{1}{1-\sigma}}\right]. \end{aligned}$$

Therefore, the monopolist's profit from all markets is

$$\Pi^{PD} = \left(\frac{1}{c_H - c_L}\right)\int_{c_L}^{c_H} \Pi_c dc = \left(\frac{1}{2-\sigma}\right)^{\frac{2-\sigma}{1-\sigma}} \frac{(c_H + c_L)}{2\lambda},$$

and the overall social welfare is

$$W^{PD} = \left(\frac{1}{c_H - c_L}\right)\int_{c_L}^{c_H} W_c dc = \frac{(c_H + c_L)}{2\lambda(1-\sigma)}\left[\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}} - \left(\frac{1}{2-\sigma}\right)^{2+\frac{1}{1-\sigma}}\right].$$

**(ii) Price discrimination is not allowed:** If price discrimination is not allowed, the monopolist solves for the following problem:

$$\underset{x,T}{Max}\ \Pi^U = \left(\frac{T}{c_H - c_L}\right)\int_{xc_L}^{c_H}\left(1 - \frac{\lambda(1-\sigma)T}{c}\right)^{\frac{1}{1-\sigma}} dc$$

$$s.t. \;\; x \geq 1.$$

The Lagrangian is

$$L = \left(\frac{T}{c_H - c_L}\right) \int_{xc_L}^{c_H} \left(1 - \frac{\lambda (1-\sigma) T}{c}\right)^{\frac{1}{1-\sigma}} dc + \gamma(x-1),$$

where $\gamma$ is the Lagrangian multiplier.

The first-order condition for $x$ when $x \geq 1$ is not binding is

$$\frac{\partial L}{\partial x} = 0 \Longrightarrow xc_L = (1-\sigma)\lambda T.$$

The first-order condition for $T$ is

$$\frac{\partial L}{\partial T} = 0 \Longrightarrow \int_{(1-\sigma)\lambda T}^{c_H} \left(1 - \frac{(1-\sigma)\lambda T}{c}\right)^{\frac{1}{1-\sigma}} \left(\frac{c-(2-\sigma)\lambda T}{c-(1-\sigma)\lambda T}\right) dc = 0. \tag{18}$$

Define $c/(\lambda T) = t$. We can rewrite (18) as follows:

$$\lambda T \int_{1-\sigma}^{\frac{c_H}{\lambda T}} \left(1 - \frac{(1-\sigma)}{t}\right)^{\frac{1}{1-\sigma}} \left(\frac{t-(2-\sigma)}{t-(1-\sigma)}\right) dt = 0.$$

Let the optimal fee be denoted $\widehat{T}$. Accordingly, the optimal solution requires $c_H$ and $\widehat{T}$ always being proportional, i.e. $\widehat{T} = c_H/(z\lambda)$, where $z$ is a constant satisfying

$$\int_{1-\sigma}^{z} \left(1 - \frac{(1-\sigma)}{t}\right)^{\frac{1}{1-\sigma}} \left(\frac{t-(2-\sigma)}{t-(1-\sigma)}\right) dt = 0.$$

Therefore, the larger the $c_H$, the larger the $\widehat{T}$ and $x$. Define the threshold $k_0 = \frac{z}{(1-\sigma)}$ for a given $\sigma$. When $k = \frac{c_H}{c_L} > k_0$, some low-$c$ markets are shut down because

$$xc_L = (1-\sigma)\lambda\widehat{T} > c_L. \tag{19}$$

Given $\widehat{T} = c_H/(z\lambda)$, (19) implies that the cutoff value $x$ is a constant fraction of $k$, i.e.

$$x = \frac{(1-\sigma)}{z}k, \tag{20}$$

which implies that $x$ is uniquely determined by $\sigma$ but not $\lambda$ (i.e., $\lambda$ is a scale parameter which does not affect $x$).

In the following discussion, we assume $k > k_0$, so some low-$c$ markets are shut down.

The corresponding welfare from market $c$ is

$$W_c^U = \int_0^{Q_c(\widehat{T})} \frac{c\left(1-Q^{1-\sigma}\right)}{\lambda(1-\sigma)} dQ = \frac{c}{\lambda(1-\sigma)}\left[Q_c\left(\widehat{T}\right) - \frac{Q_c\left(\widehat{T}\right)^{2-\sigma}}{2-\sigma}\right],$$

and the total welfare is

$$\begin{aligned} W^U &= \frac{1}{c_H-c_L}\int_{(1-\sigma)\lambda\widehat{T}}^{c_H} \frac{c}{\lambda(1-\sigma)}\left[Q_c\left(\widehat{T}\right) - \frac{Q_c\left(\widehat{T}\right)^{2-\sigma}}{2-\sigma}\right] dc \\ &= \frac{1}{c_H-c_L}\int_{(1-\sigma)\lambda\widehat{T}}^{c_H} \frac{c}{\lambda(1-\sigma)}\left[\left(\frac{c-(1-\sigma)\lambda\widehat{T}}{c}\right)^{\frac{1}{1-\sigma}} - \frac{\left(\frac{c-(1-\sigma)\lambda\widehat{T}}{c}\right)^{\frac{2-\sigma}{1-\sigma}}}{2-\sigma}\right] dc \\ &= \frac{1}{c_H-c_L}\int_{\frac{(1-\sigma)c_H}{z}}^{c_H} \frac{c}{\lambda(1-\sigma)}\left[\left(1-\frac{(1-\sigma)c_H}{cz}\right)^{\frac{1}{1-\sigma}} - \frac{\left(1-\frac{(1-\sigma)c_H}{cz}\right)^{\frac{2-\sigma}{1-\sigma}}}{2-\sigma}\right] dc. \quad (21) \end{aligned}$$

Define $c/c_H = s$. We can rewrite (21) as

$$W^U = \frac{Rc_H^2}{c_H-c_L},$$

where $R$ is a constant satisfying

$$R = \int_{\frac{1-\sigma}{z}}^{1} \frac{s}{\lambda(1-\sigma)}\left[\left(1-\frac{1-\sigma}{sz}\right)^{\frac{1}{1-\sigma}} - \frac{\left(1-\frac{1-\sigma}{sz}\right)^{\frac{2-\sigma}{1-\sigma}}}{2-\sigma}\right] ds.$$

**(iii) Welfare Comparison:** As shown above, the welfare under price discrimination is

$$W^{PD} = ac_H + ac_L,$$

where

$$a = \frac{1}{2\lambda(1-\sigma)}\left[\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}} - \left(\frac{1}{2-\sigma}\right)^{2+\frac{1}{1-\sigma}}\right]. \quad (22)$$

In contrast, the welfare under uniform price is

$$W^U = \frac{Rc_H^2}{c_H-c_L},$$

where

$$R = \int_{\frac{1-\sigma}{z}}^{1} \frac{s}{\lambda(1-\sigma)} \left[ \left(1 - \frac{1-\sigma}{sz}\right)^{\frac{1}{1-\sigma}} - \frac{\left(1 - \frac{1-\sigma}{sz}\right)^{\frac{2-\sigma}{1-\sigma}}}{2-\sigma} \right] ds, \tag{23}$$

and $z$ is a constant satisfying

$$\int_{1-\sigma}^{z} \left(1 - \frac{(1-\sigma)}{t}\right)^{\frac{1}{1-\sigma}} \left(\frac{t-(2-\sigma)}{t-(1-\sigma)}\right) dt = 0. \tag{24}$$

Normalize $c_L = 1$, so $c_H = k$ and the welfare difference is

$$W^{PD} - W^{U} = ak + a - \frac{Rk^2}{k-1}.$$

Given that $a > R$ for $\sigma < 1$, we have

$$W^{PD} - W^{U} > 0 \Leftrightarrow k > \sqrt{\frac{a}{a-R}}.$$

Hence, welfare is always higher under price discrimination when there is enough demand dispersion across markets; i.e. $k > \sqrt{\frac{a}{a-R}}$.

Note from above, when there is a continuum of uniformly distributed markets and demand is log-concave, we find the monopolist that is not allowed to price discriminate will set the price such that markets below the cutoff level $xc_L$ are shut down, provided that the dispersion of demand across markets is large enough (i.e., $k > k_0$). As (20) suggests, the cutoff value $x$ is a constant fraction of the dispersion $k$ and is unique for each given $\sigma$, i.e.

$$x = \frac{(1-\sigma)}{z}k.$$

Accordingly, the fraction of markets shut down is

$$\frac{\frac{k(1-\sigma)}{z} - 1}{k-1},$$

which increases in $k$ given $(1-\sigma)/z$ is a fraction less than one.

Again, take the linear demand $\sigma = 0$ as an example. Equation (22) can be written as

$$a = \frac{1}{2\lambda}\left[\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)^3\right] = \frac{3}{16\lambda}. \tag{25}$$

Equation (23) can be rewritten as

$$R = \frac{1}{2\lambda} \int_{\frac{1}{z}}^{1} \left[ s - \frac{1}{sz^2} \right] ds = \frac{1}{2\lambda} \left[ \frac{1}{2} - \frac{1}{2z^2} + \frac{1}{z^2} \ln \left( \frac{1}{z} \right) \right]. \tag{26}$$

Note that $z$ is a constant satisfying (24):

$$\int_1^z \left( 1 - \frac{2}{t} \right) dt = 0,$$

which implies

$$z - 1 - 2 \ln z = 0,$$

so $z \simeq 3.513$, which corresponds to $k_0$ in the analysis of Section 2. For any $k > z$, there is a cutoff level $k/z$ such that any markets $c < k/z$ will be shut down.

Given $z \simeq 3.513$, we can also compare (25) and (26),

$$W^{PD} - W^U > 0 \Leftrightarrow k > \sqrt{\frac{a}{a - R}} = \sqrt{\frac{3}{4/3.513 - 1}} \simeq 4.651,$$

as we found for the linear demand case.

In conclusion, we have shown for the continuum case, that when demand given by (10) is log-concave, welfare is higher under price discrimination provided there is sufficient dispersion of demand across markets.