Online Appendix: Ad-valorem platform fees, indirect taxes and efficient price discrimination

Zhu Wang* and Julian Wright†

August 2016

Abstract

In this online appendix, we provide the detailed working for results in Section 4 of our main paper, which considers the robustness of affine fee schedules to some of our key assumptions.

1 Non-GPD demand functions

In this section, we conduct the following exercise. We assume that the true underlying demand is not from a GPD, but instead comes from some alternative distribution. As an example, we consider that $b \geq 0$ follows a truncated normal distribution. We assume the platform does not observe the true distribution, but rather continues to assume that demand takes the GPD form.

Figures 1 and 2 present our findings based on two informative cases. In each case, we generate 10,000 random draws of $b \geq 0$ from a truncated normal distribution. We then fit the data with a GPD using maximum likelihood estimation to recover the GPD parameters λ and σ that best fit the data. Denote $\Phi(x)$ as the cdf of an underlying mean-zero normal

*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.

distribution. In case 1 and case 2, the draws $b \geq 0$ are from a truncated normal distribution which is a relocation of the 75% upper tail (defined as $x \geq x^*$ for $\Phi(x^*) = 0.75$) and 25% upper tail (defined as $x \geq x^*$ for $\Phi(x^*) = 0.25$) of the underlying normal distribution. In each case, we choose the variance of the underlying normal distribution and the value of d such that the fitted GPD would imply an optimal affine fee schedule matching the observed affine fee schedule used by Amazon (i.e., \$1.35+15% as shown in Table 1 of the main paper).

The results are quite striking. In case 1, the GPD is shown to proxy the 75% upper tail of the underlying normal distribution very well. As a result, the optimal fee schedule based on the fitted GPD is very close to the optimal one based on the true underlying truncated normal distribution, and so are the output and platform profit for each individual good. In fact, the platform achieves more than 99.7% of the maximal profit for each individual good.

In case 2, in which we fit the 25% upper tail of the underlying normal distribution, the GPD does not proxy the true truncated normal distribution as well. Figure 2 shows that the optimal fee schedule based on the fitted GPD is somewhat lower than the optimal one based on the true distribution, and the outputs are somewhat higher. However, in terms of profit, the GPD approximation still performs well, which allows the platform to achieve 90-95% maximal profits for goods below \$10, and 95-99% for goods between \$10 and \$60, and more than 99% for goods above \$60.

We also conducted the exercise by using other alternative underlying distributions and the results are very similar. Figures 3 and 4 report the results for the case of the truncated logistic distribution.

2 Random variation in demand

We provide the full details here for several cases that were noted in the main paper.

2.1 Random multiplicative and additive shocks to demand

The first case noted in the main paper is straightforward. We show our results are not affected at all by random multiplicative and additive shocks to demand.

Figure 1: Demand comes from truncated normal distribution

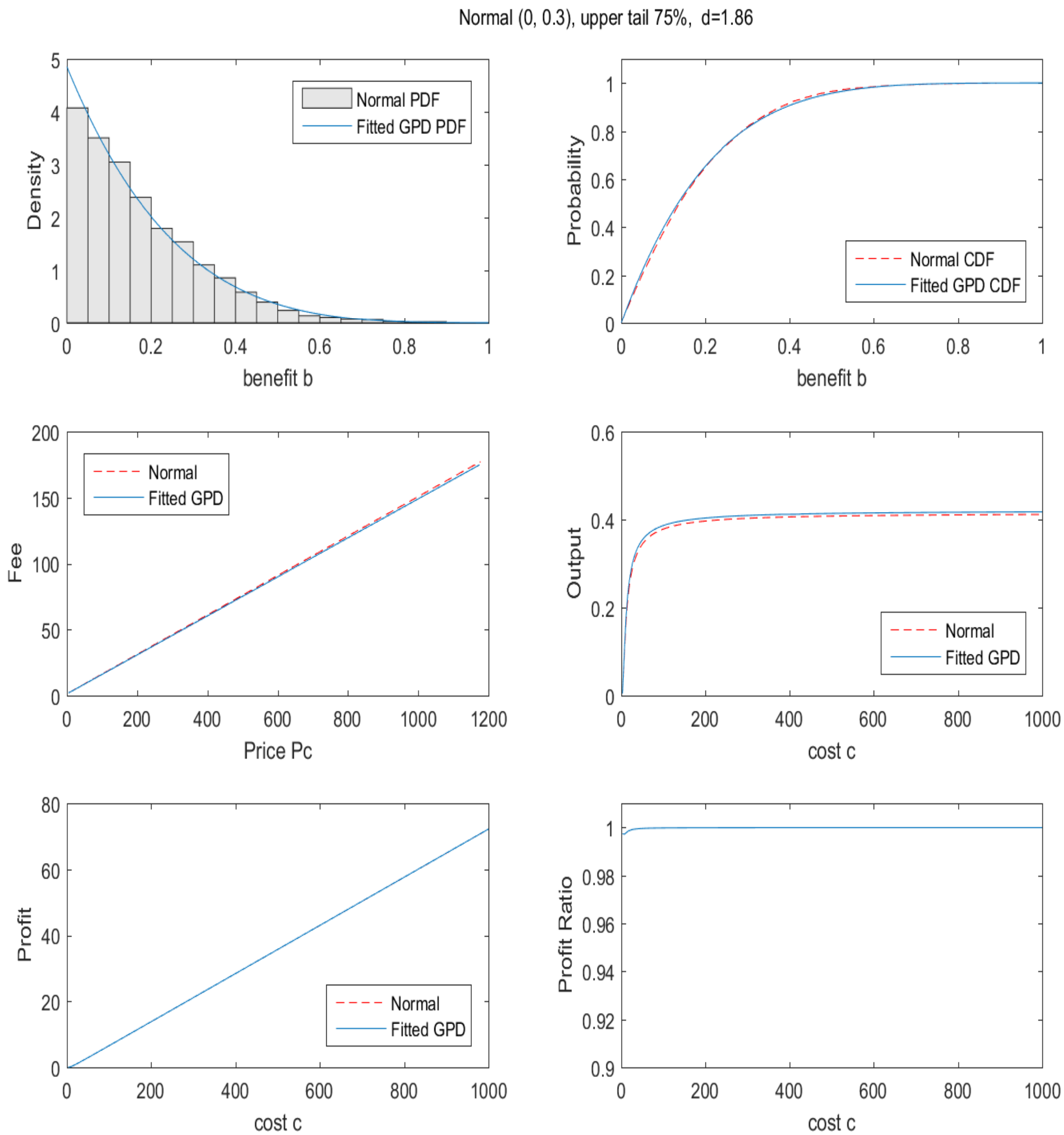


Figure 2: Demand comes from truncated normal distribution

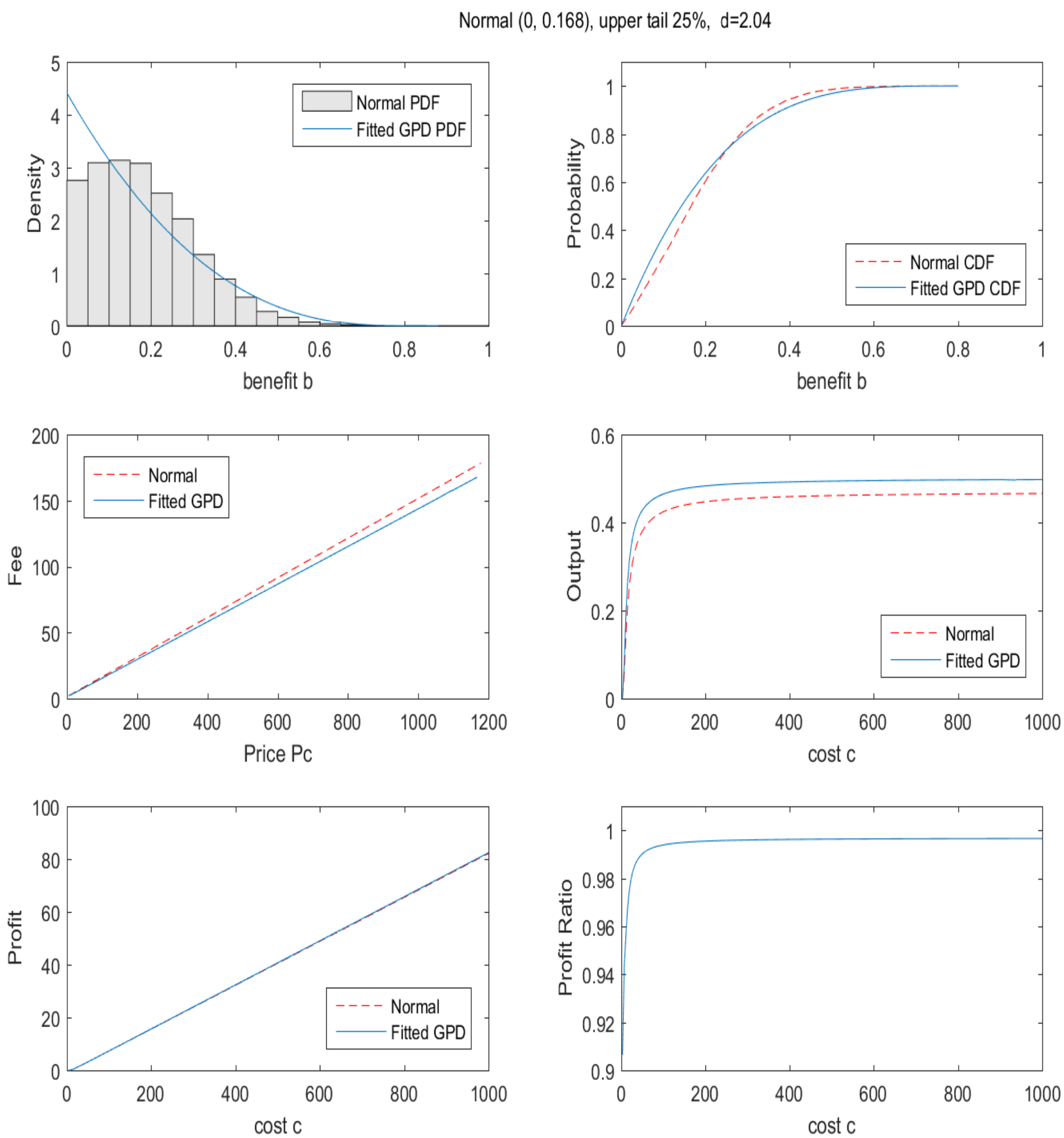


Figure 3: Demand comes from truncated logistic distribution

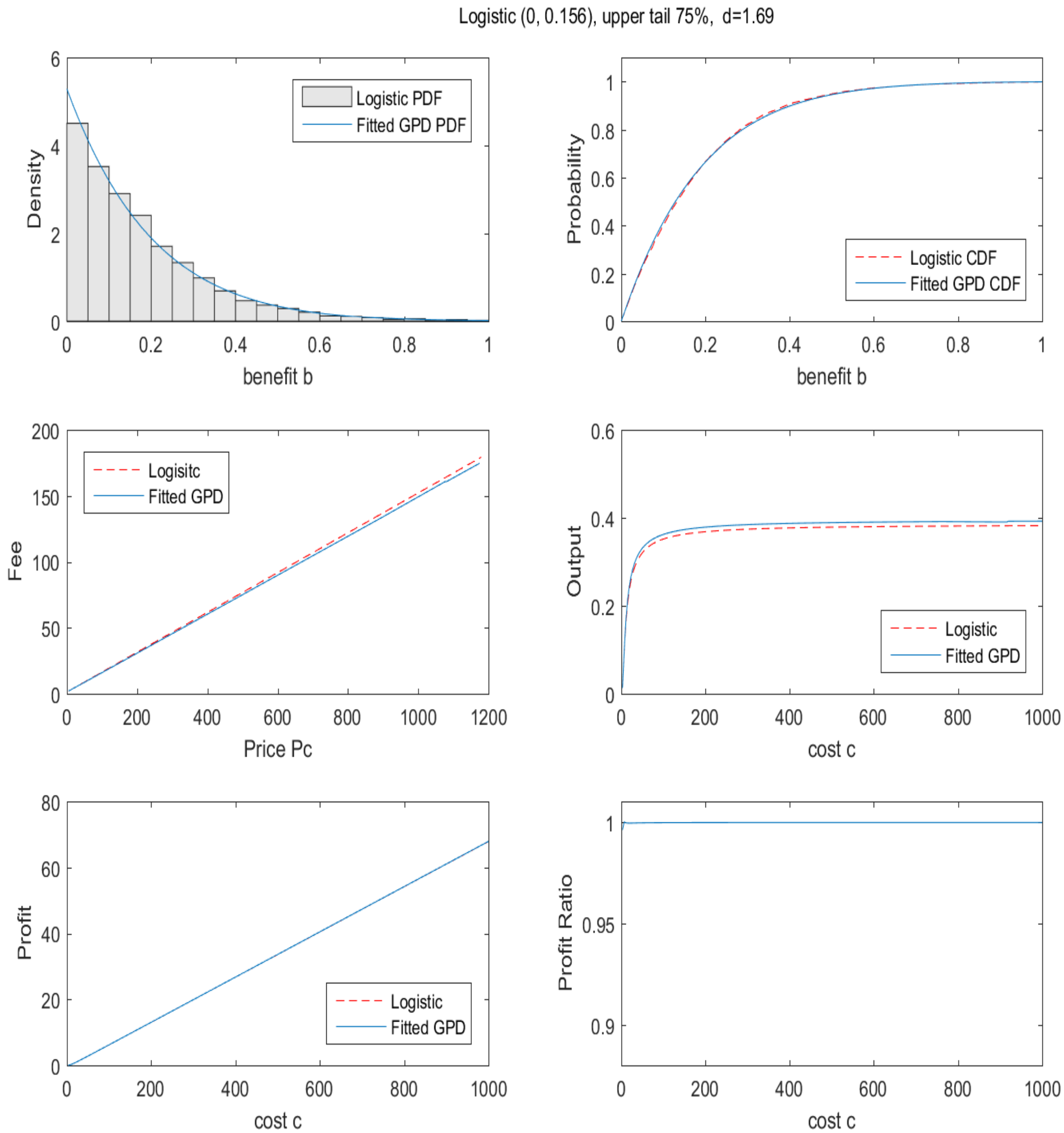
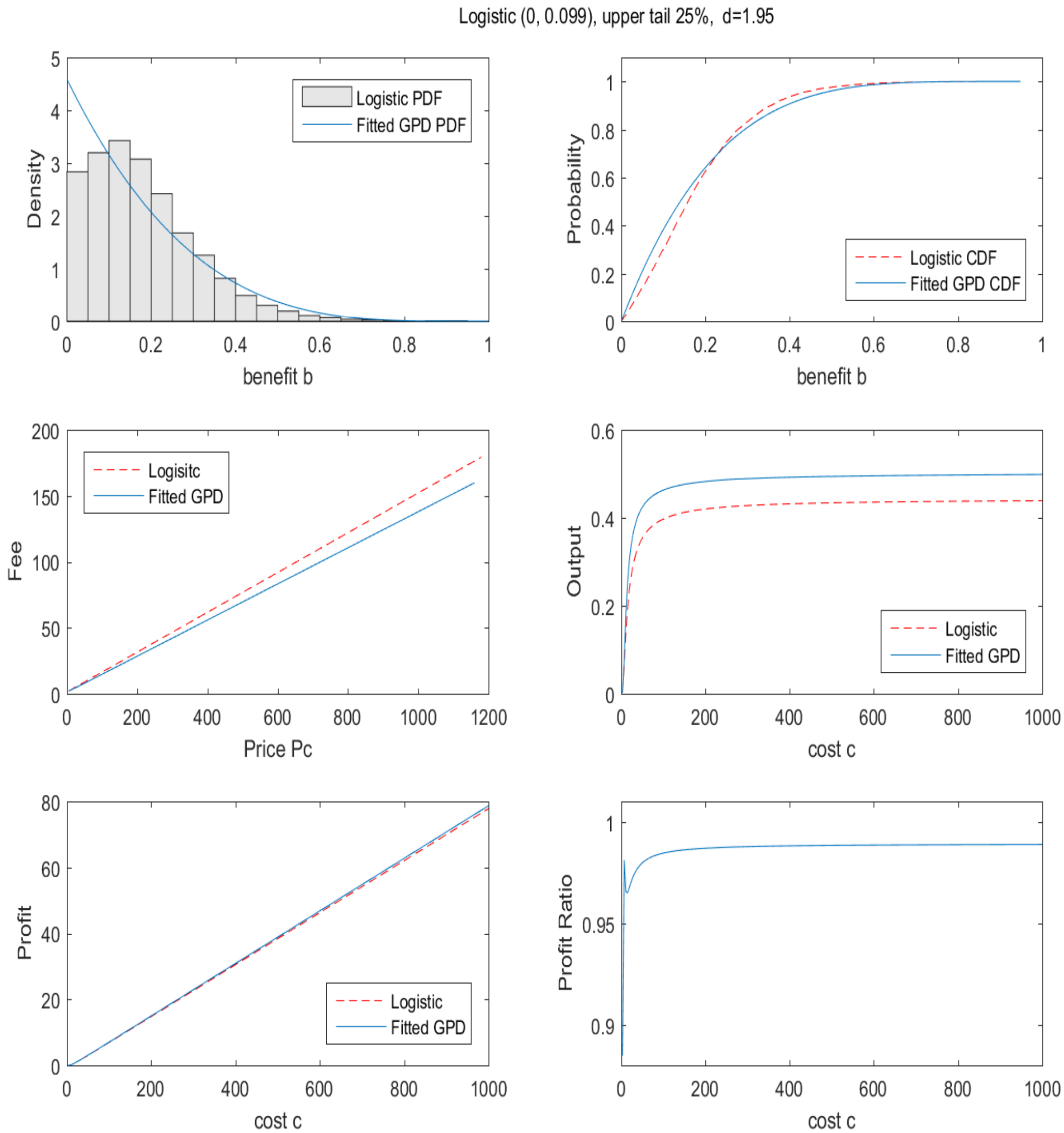


Figure 4: Demand comes from truncated logistic distribution



Specifically, suppose demand can be written as $\xi_c Q\left(1+\frac{T_c}{c}\right)+\varepsilon_c$, where ξ_c is randomly drawn from the distribution Ξ with $E(\xi_c)=1$ and ε_c is drawn (independently) from the distribution ϵ with $E(\varepsilon_c)=0$. Consider the setup used in Section 3 of the main paper on "proportional taxes and fees". In order to maximize revenue for good c, we choose T_c to maximize

$$\int_{\xi_c}\int_{\varepsilon_c} T_c\left(\xi_c Q\left(1+\frac{T_c}{c}\right)+\varepsilon_c\right) d\epsilon\left(\varepsilon_c\right) d\Xi\left(\xi_c\right).$$

The first-order condition is

$$\int_{\xi_c}\int_{\varepsilon_c}\left(\xi_c Q\left(1+\frac{T_c}{c}\right)+\varepsilon_c+\xi_c\frac{T_c}{c}Q'\left(1+\frac{T_c}{c}\right)\right) d\epsilon\left(\varepsilon_c\right) d\Xi\left(\xi_c\right)=0,$$

which given $E(\varepsilon_c)=0$ can be rewritten as

$$\int_{\xi_c}\xi_c d\Xi\left(\xi_c\right)\left(Q\left(1+\frac{T_c}{c}\right)+\frac{T_c}{c}Q'\left(1+\frac{T_c}{c}\right)\right)=0.$$

Given $E(\xi_c)=1$, the first-order condition for determining the optimal per-transaction fee for good c becomes the same as in equation (2) in the main paper. Moreover, equilibrium prices are not affected by the demand heterogeneity since prices are determined entirely by costs so the optimal fee schedule still satisfies equation (4) in the main paper. The same logic applies to profit maximization and the Ramsey regulation problem. Thus, we can easily allow for additive and multiplicative random shocks in the demand function.

The more interesting case arises when different goods can have different shape of demand (or different price elasticity of demand), which we turn to next.

2.2 Random variation in the GPD mean

We first consider random variation in the GPD mean. Suppose the platform faces the GPD demand as previously defined, so that

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

Let the mean of the GPD, $E[b] = \frac{1}{(2-\sigma)\lambda}$, be 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

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

The platform ignores the random variation but rather assumes that the GPD has a fixed mean equal to $\frac{1}{\widehat{\lambda}(2-\sigma)}$. Also, we assume that $d = 0$ for simplicity, given that d is typically small. Therefore, the platform solves for the profit-maximizing per-transaction fee T_c for goods of cost c as follows:

$$\max_{T_c} T_c \left(1 + \widehat{\lambda}(\sigma-1)\frac{T_c}{c}\right)^{\frac{1}{1-\sigma}}.$$

The solution is

$$T_c = \frac{c}{\widehat{\lambda}\,(2-\sigma)}.$$

Given Bertrand competition between sellers we have $p_c = T_c + c$ and this implies the proportional fee schedule

$$T_c = \frac{1}{1+\widehat{\lambda}\,(2-\sigma)} p_c. \tag{2}$$

Because each good c indeed has a different realized value λ due to random variation, the platform earns the average profit π_c from all goods of cost c as

$$\begin{aligned} \pi_c^0 &= E\left[\frac{c}{\widehat{\lambda}\,(2-\sigma)}\left(1+\frac{\lambda(\sigma-1)}{\widehat{\lambda}\,(2-\sigma)}\right)^{\frac{1}{1-\sigma}}\right] \\ &= \frac{c}{\widehat{\lambda}\,(2-\sigma)} E\left[\left(1+\frac{\sigma-1}{\widehat{\lambda}\,(2-\sigma)\left(\frac{1}{\widehat{\lambda}}-(\sigma-2)\,\varepsilon\right)}\right)^{\frac{1}{1-\sigma}}\right], \end{aligned} \tag{3}$$

where $E[\cdot]$ is the expected value of the expression.

In contrast, if the platform knows the realization of λ for each good c, it would set the per-transaction fee

$$T_c = \frac{c}{\lambda(2-\sigma)},$$

and the average platform profit earned from all goods c is

$$\pi_c^1 = E\left[\frac{c}{\lambda(2-\sigma)}\left(1+\frac{\sigma-1}{2-\sigma}\right)^{\frac{1}{1-\sigma}}\right] = \frac{c}{\widehat{\lambda}(2-\sigma)}\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}}.$$

Now the question is how much platform profit is lost if the platform ignores the fact that there is random variation in demand.

To simplify the comparison, we rewrite (3) by taking the second-order Taylor expansion around $E[\varepsilon]=0$ so that

$$\pi_c^0 \simeq \frac{c}{(2-\sigma)\widehat{\lambda}}\left(\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}} - \frac{1}{2}\widehat{\lambda}^2\left(\frac{1}{2-\sigma}\right)^{\frac{1}{1-\sigma}-3} E\left[\varepsilon^2\right]\right)$$

Therefore, we can compute the profit ratio as

$$\frac{\pi_c^0}{\pi_c^1} = 1 - \frac{1}{2}\widehat{\lambda}^2(2-\sigma)^3 E[\varepsilon^2]. \tag{4}$$

We can verify that $\frac{\pi_c^0}{\pi_c^1} < 1$ given that $\sigma < 2$ and $\frac{\pi_c^0}{\pi_c^1} = 1$ when $E[\varepsilon^2]=0$. It is also shown that $\frac{\pi_c^0}{\pi_c^1}$ decreases in $E[\varepsilon^2]$, 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.

Based on some commonly used demand functions, we can further assess the magnitude of the profit loss shown in (4), which tends to be small under a reasonable dispersion of the random variation. Assuming that the observed percentage fee used by Amazon (i.e., 15%) is determined by (2), (1) then pins down the mean of $E[b]$; that is, $\frac{1}{\widehat{\lambda}(2-\sigma)} = 0.176$. Now if we consider the linear demand where $\sigma = 0$, (4) 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. Moreover, for the exponential demand where $\sigma = 1$, the platform earns more than 90% of the optimal profit if the standard deviation of $E[b]$ is smaller than 0.079. In fact, as suggested by (4), such a threshold standard deviation of $E[b]$ increases monotonically in σ, the curvature of the inverse demand function.

Furthermore, note that the profit ratio calculated above in (4) is a conservative estimate in the sense that we assume the platform treats the GPD as having a fixed mean equal to $\frac{1}{\widehat{\lambda}(2-\sigma)}$. Another scenario could be that the platform fully incorporates the random variation in $E[b]$ and maximizes the expected profit by figuring out the best affine fee schedule accordingly. In that case, the platform could potentially earn an even higher ratio of the optimal profit.

2.3 Random variation in the demand elasticity

Aside from different GPD means, it is also interesting to consider random variation in the demand elasticity. To do this we consider the GPD demand with $\sigma = 1 + \frac{1}{\lambda}$ and $\lambda > 1$. In this case, sellers on the platform face a constant-elasticity demand so that

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

The demand elasticity λ is a random variable,

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

where $\widehat{\lambda}$ is the mean of λ and ε is i.i.d. noise with mean zero.

Similar to the proof above, we assume that $d = 0$ for simplicity. If the platform ignores the random variation but rather assumes that the demand has a fixed elasticity λ equal to its mean value $\widehat{\lambda}$, the platform would set the profit-maximizing per-transaction fee

$$T_c = \frac{c}{\widehat{\lambda} - 1}$$

for each good. Given Bertrand competition between sellers we have $p_c = T_c + c$ and this implies the proportional fee schedule

$$T_c = \frac{1}{\widehat{\lambda}} p_c. \tag{5}$$

The corresponding platform profit is

$$\pi_c^0 = E\left[\frac{c}{\widehat{\lambda}-1}\left(1+\frac{1}{\widehat{\lambda}-1}\right)^{-\lambda}\right]. \tag{6}$$

In contrast, if the platform knows the realized λ for each good c, it could set the per-transaction fee

$$T_c = \frac{c}{\lambda - 1}$$

for each good, and the average platform profit earned from good c would be

$$\pi_c^1 = E\left[\frac{c}{\lambda-1}\left(1+\frac{1}{\lambda-1}\right)^{-\lambda}\right]. \tag{7}$$

To simplify the comparison, we can rewrite (6) by taking the second-order Taylor expansion around the mean $\widehat{\lambda}$ so that

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

We can do the same for (7), so that

$$\pi_c^1 \simeq c\widehat{\lambda}^{-\widehat{\lambda}}(\widehat{\lambda}-1)^{\widehat{\lambda}-1}\left(1+\frac{1}{2}\left(\left(\ln\frac{\widehat{\lambda}}{\widehat{\lambda}-1}\right)^2+\frac{1}{\widehat{\lambda}-1}-\frac{1}{\widehat{\lambda}}\right)E\left[\varepsilon^2\right]\right).$$

Therefore, the profit ratio is

$$\frac{\pi_c^0}{\pi_c^1} = \frac{1+\frac{1}{2}\left(\ln\frac{\widehat{\lambda}}{\widehat{\lambda}-1}\right)^2 E\left[\varepsilon^2\right]}{1+\frac{1}{2}\left(\left(\ln\frac{\widehat{\lambda}}{\widehat{\lambda}-1}\right)^2+\frac{1}{\widehat{\lambda}-1}-\frac{1}{\widehat{\lambda}}\right)E\left[\varepsilon^2\right]}. \tag{8}$$

We can verify that $\frac{\pi_c^0}{\pi_c^1} < 1$ and $\frac{\pi_c^0}{\pi_c^1} = 1$ when $E\left[\varepsilon^2\right] = 0$. It is also shown that $\frac{\pi_c^0}{\pi_c^1}$ decreases in $E\left[\varepsilon^2\right]$, the variance of the noise term.

Moreover, we can verify the profit loss shown in (8) tends to be small under a reasonable dispersion of the random variation. Assuming that the observed percentage fee used by Amazon (i.e., 15%) is determined by (5), this implies $\widehat{\lambda} = 6.667$. Equation (8) then 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.

3 Sellers' market power

We suppose that for each good there are $n_c \geq 1$ identical quantity-setting sellers on the platform (i.e., Cournot competitors). This includes the special case $n_c = 1$, for which each good is sold by a local monopolist. Each seller obtains the goods at a unit cost c and sells them at a retail price p_c. In Section 3.1 below we show that 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. In Section 3.2 we derive analytical results for the case $d = 0$ and show that while the Bertrand fee schedule is not necessarily the optimal affine fee schedule for all the GPD demands, it can be very close. Finally, in Section 3.3, we consider the general case in which $d > 0$ and the platform is allowed to use a non-linear fee schedule. Calibrating the model to Amazon's DVD sales ranks and fees, we show that the platform does not lose much by using the Bertrand fee schedule rather than the optimal fee schedule.

3.1 Bertrand fee schedule versus optimal per-transaction fees

We first consider the problem of a platform with full information on c and n_c, setting an optimal per-transaction fee for each good.

□ **Optimal per-transaction fees**

Suppose the platform charges a per-transaction fee T_c for good c. Let $q_{c,i}$ denote the output sold by seller i for good c. Each seller i sets $q_{c,i}$ taking the output by competing sellers $q_{c,-i} = Q_c - q_{c,i}$ as given, and maximizes its profit $(p_c - c - T_c)\, q_{c,i}$. Assuming F follows the GPD

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

where $\lambda > 0$ and $\sigma < 2$, the total demand for good c is

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

which implies that the inverse demand is

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

Therefore, an individual seller's profit maximization problem is

$$\max_{q_{c,i}} c\left(1 + \frac{(q_{c,-i} + q_{c,i})^{1-\sigma} - 1}{\lambda(\sigma - 1)}\right) q_{c,i} - (c + T_c)q_{c,i}.$$

The first-order condition for good c is

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

In a symmetric Cournot equilibrium, $q_{c,i} = q_c$ for every seller, so the total sellers' output $Q_c = n_c q_c$. We can then rewrite the first-order condition as

$$\frac{c(n_c q_c)^{1-\sigma} - c}{\lambda(\sigma - 1)} = \frac{c(n_c q_c)^{1-\sigma}}{n_c \lambda} + T_c,$$

and derive

$$Q_c = n_c q_c = \left(\frac{c n_c + \lambda(\sigma - 1)T_c n_c}{c n_c - (\sigma - 1)c}\right)^{\frac{1}{1-\sigma}}. \tag{9}$$

Accordingly, the price of good c is

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

or

$$\frac{p_c}{c} = \frac{n_c}{(n_c + 1 - \sigma)} \frac{T_c}{c} + \frac{1 + (n_c + 1 - \sigma)\lambda}{(n_c + 1 - \sigma)\lambda}.$$

The platform takes (9) as given and maximizes its profit by setting a per-transaction fee for good c as follows

$$\max_{T_c} \ (T_c - d) \left(\frac{cn_c + \lambda(\sigma - 1)T_c n_c}{cn_c - (\sigma - 1)c} \right)^{\frac{1}{1-\sigma}}.$$

The first-order condition implies the optimal per-transaction fee T_c^f:

$$T_c^f = \frac{\lambda d + c}{\lambda (2 - \sigma)}, \tag{11}$$

which yields the same profit-maximizing platform fees as implied by the Bertrand fee schedule (see equation (9) in the main paper). The optimal per-transaction fee does not depend on the number of sellers, and so also holds for a monopoly seller. Note that to ensure a meaningful solution (i.e., $T_c^f > d$) , it is required that

$$d (\sigma - 1) + \frac{c}{\lambda} > 0. \tag{12}$$

This is satisfied for the GPD demand specification as stated in our main paper. Note that when demand is log-linear or log-convex, the GPD specification requires that $\sigma \geq 1$ so the condition in (12) holds. When demand is log-concave, the GPD specification requires that $\sigma < 1$ and $d < \frac{c}{\lambda(1-\sigma)}$, so the condition in (12) again holds.

Substituting (11) into (9) and (10), we get

$$p_c = \frac{n_c d}{(2 - \sigma)(n_c + 1 - \sigma)} + \frac{n_c + (2 - \sigma) + (2 - \sigma)(n_c + 1 - \sigma)\lambda}{(2 - \sigma)(n_c + 1 - \sigma)\lambda} c, \tag{13}$$

and

$$Q_c = \left(\frac{\lambda(\sigma - 1)n_c d + cn_c}{(2 - \sigma)(cn_c - (\sigma - 1)c)} \right)^{\frac{1}{1-\sigma}}. \tag{14}$$

As a result, the platform profit from good c is

$$\pi_c = \left(\frac{(\sigma - 1)d}{2 - \sigma} + \frac{c}{(2 - \sigma)\lambda} \right) \left(\frac{\lambda(\sigma - 1)n_c d + cn_c}{(2 - \sigma)(cn_c - (\sigma - 1)c)} \right)^{\frac{1}{1-\sigma}}.$$

□ **Affine fee schedule**

Now consider Cournot sellers facing an affine fee schedule $T(p_c) = t_0 + t_1 p_c$ for each transaction. With GPD demand, the sellers' problem is to choose $q_{c,i}$ to maximize

$$((1-t_1)p_c - c - t_0)q_{c,i},$$

where

$$p_c = c\left(1 + \frac{(q_{c,-i}+q_{c,i})^{1-\sigma}-1}{\lambda(\sigma-1)}\right).$$

In a symmetric Cournot equilibrium, $q_{c,i} = q_c$ for every seller, so the total sellers' output $Q_c = n_c q_c$. The first-order condition then requires

$$(1-t_1)c\left(\frac{1}{\lambda(\sigma-1)}-1\right)+c+t_0 = \frac{(1-t_1)cQ_c^{1-\sigma}}{\lambda}\left(\frac{1}{\sigma-1}-\frac{1}{n_c}\right). \tag{15}$$

□ **Comparing Bertrand fee schedule and optimal per-transaction fees**

Substituting the Bertrand fee schedule from equation (6) in the main paper into (15) gives the same price and output for a given c as we found above in (13) and (14) for the full information case. That is, the price and output for each good are identical to that implied by the optimal per-transaction fee (11). However, the per-transaction fee for good c implied by the Bertrand fee schedule is now

$$\begin{aligned} T^*(p_c) &= t_0 + t_1 p_c = \left(\frac{\lambda}{1+(2-\sigma)\lambda} + \frac{n_c}{(1+(2-\sigma)\lambda)(2-\sigma)(n_c+1-\sigma)}\right)d \\ &\quad + \left(\frac{1}{1+(2-\sigma)\lambda}\right)\left(\frac{n_c+(2-\sigma)+(2-\sigma)(n_c+1-\sigma)\lambda}{(2-\sigma)(n_c+1-\sigma)\lambda}\right)c, \end{aligned}$$

which is strictly higher than the fee in (11) if and only if the condition (12) holds. This implies the platform earns a higher profit using the Bertrand fee schedule than if it used the optimal per-transaction fee for each different good assuming full information. This result holds for any $n_c \geq 1$, and so also holds for monopoly sellers.

This result shows that the Bertrand fee schedule can be used in this setting to solve the price discrimination problem. It delivers the same price and output for each good without using any information on each good's cost. At the same time, the Bertrand fee schedule generates a higher profit for the platform because it mitigates the double marginalization problem associated with using the optimal per-transaction fee for each good, allowing the platform to set a higher fee for each good while achieving the same level of final price and output.

3.2 Bertrand fee schedule versus the optimal affine fee schedule

In this section, assuming $d = 0$, we show that the Bertrand fee schedule (6) from the main paper is very close to the optimal affine fee schedule in the presence of double marginalization.[1] In this case, the Bertrand fee schedule (6) in the main paper implies the following proportional fee schedule

$$T^* (p_c) = \left(\frac{1}{1 + (2 - \sigma)\lambda} \right) p_c. \tag{16}$$

We can then check whether this is the optimal affine fee schedule in general.

Consider a platform maximizing its profit by using an affine fee schedule $t_0 + t_1 p_c$. We assume that the platform cannot subsidize sellers to operate by setting $t_0 < 0$. Doing so is likely to create an adverse selection problem, in which some sellers sign up just to collect t_0 and then do not sell anything. This imposes the requirement that $t_0 \geq 0$.

Subject to this requirement, the platform then solves the following problem:

$$\pi = \max_{t_0, t_1} \sum_c g_c \left(t_0 + t_1 p_c \right) \left(1 - F\left(\frac{p_c}{c} \right) \right)$$

subject to two additional conditions

$$p_c = c \left(1 + \frac{Q_c^{1-\sigma} - 1}{\lambda(\sigma - 1)} \right) \tag{17}$$

and

$$(1 - t_1)c \left(\frac{1}{\lambda(\sigma - 1)} - 1 \right) + c + t_0 = \frac{(1 - t_1)c Q_c^{1-\sigma}}{\lambda} \left(\frac{1}{\sigma - 1} - \frac{1}{n_c} \right), \tag{18}$$

where (17) is given by the GPD demand and (18) is from the first-order condition derived above. We can verify that the constraint $t_0 \geq 0$ is binding at the maximum (that is, $\frac{\partial \pi}{\partial t_0} < 0$ when evaluated at the optimal t_1 and $t_0 = 0$), so the optimal affine fee schedule is also just a proportional fee schedule. Moreover, given that $t_0 = 0$, p_c/c does not depend on c, so the platform can solve for the optimal t_1 without knowing the distribution of c. The first-order

[1] If $d > 0$, the results will depend on the distribution of c. We explore this case in Section 3.3 below.

condition on t_1 requires

$$\begin{aligned}
& (1+\lambda(1-\sigma))(1-t_1-t_1\lambda(1-\sigma))(1-t_1)-t_1\lambda(1+\lambda(1-\sigma)) \\
= \; & n_c\left(\frac{t_1}{1-t_1}\lambda^2(2-\sigma)-\lambda\right). \qquad (19)
\end{aligned}$$

The optimal proportional fee implied by (19) is in general not equal to the proportional fee implied by (16), but based on an examination of some common demand functions, it is very close and so are the profits.

Consider first the case of constant elasticity demand, where $\sigma = 1 + \frac{1}{\lambda}$ and $\lambda > 1$. In this case, both (16) and (19) yield $t_1 = 1/\lambda$, and so identical profits. Thus, in this case the Bertrand fee schedule coincides with the optimal affine fee schedule.

Next consider the case of exponential demand where $\sigma = 1$. Then (19) implies the optimal proportional fee satisfies

$$(1-t_1)^3 + \lambda(1-t_1)(n_c - t_1) = n_c t_1 \lambda^2,$$

which has a unique solution. In contrast, (16) implies the proportional fee

$$t_1 = \frac{1}{1+\lambda}.$$

The two fees are not exactly equal, but they are very close. For the empirically meaningful range where the proportional term t_1 of the Bertrand fee schedule satisfies $t_1 \leq 50\%$ (or equivalently, $\lambda \geq 1$), the Bertrand fee schedule can recover more than 98.5% of the profit under the optimal affine fee schedule when all sellers are monopolists (so $n_c = 1$ for all c). Moreover, the profit gap between using the Bertrand fee schedule and using the optimal affine fee schedule decreases monotonically in n_c, and the two converge as the number of Cournot sellers gets large.

Finally, consider the case of linear demand where $\sigma = 0$. Then (19) implies the optimal proportional fee satisfies

$$\begin{aligned} & (1-t_1)^2(1+\lambda)(1-t_1-t_1\lambda)-t_1(1-t_1)\lambda(1+\lambda) \\ = \;& n_c\left(2t_1\lambda^2-\lambda(1-t_1)\right), \end{aligned}$$

which has a unique solution. In contrast, (16) implies the proportional fee

$$t_1=\frac{1}{1+2\lambda}.$$

For the empirically meaningful range where the proportional term t_1 of the Bertrand fee schedule satisfies $t_1 \leq 50\%$ (or equivalently, $\lambda \geq 0.5$), the Bertrand fee schedule can recover more than 97.5% of the profit under the optimal affine fee schedule when all sellers are monopolists (so $n_c = 1$ for all c). Again, the profit gap between using the Bertrand affine fee schedule and using the optimal affine fee decreases monotonically in n_c, and the two converge as the number of Cournot sellers gets large.

3.3 Bertrand fee schedule versus the optimal fee schedule

Finally, we compare the platform's profit from the Bertrand fee schedule with its profit from the optimal (non-linear) fee schedule. Once we allow for a non-linear fee schedule, the optimal fee schedule will depend on the distribution of goods $G(c)$. This is also true for the optimal affine fee schedule once we allow $d > 0$. Therefore, to proceed, we need to assume some realistic distribution for c and calculate the profitability of different fee schedules numerically. We use the distribution based on fitting a log-normal distribution to the actual distribution of sales obtained from sales ranks of DVDs sold on Amazon (as explained below). We assume sellers face constant elasticity demand—recall, when $d = 0$, this is the benchmark for which the Bertrand fee schedule coincides with the optimal affine fee schedule regardless of the number of sellers that are competing. We use the parameter values $d = 1.35$ and $\sigma = 1.15$ so that the Bertrand fee schedule matches the fee schedule used by Amazon for DVDs (i.e., \$1.35+15% shown in Table 1 in the main paper). We assume each seller is a monopolist (i.e., $n_c = 1$).

With these assumptions, we compare the profit the platform obtains for different types of fees and fee schedules. Note we continue to restrict our attention to fee schedules and do not consider other more complicated mechanisms. Thus, our focus is on how much better a platform can do by allowing the ad-valorem component of the transaction fee to be any non-linear function of price.[2]

As noted above, the optimal fee schedule will in general depend on the distribution of c we assume. Rather than using an arbitrary and possibly unrealistic distribution, we proceed by calibrating our model to a realistic distribution of goods being sold. The platform we study is Amazon's marketplace for DVDs. Using a web robot, we collected data on every DVD that was listed under "Movies & TV" on Amazon's marketplace in January 2014. Given shipping fees are often not included in the listed price, we focus on the items where the listed price included free shipping, resulting in a sample with 191,280 distinct items. The data collected include the title, unique ASIN number identifying the DVD, the price, and sales rank of each DVD. Since some DVDs are listed with extreme prices, we restrict our sample to DVDs selling for no more than \$300.[3]

Given we do not directly observe the sales of each DVD, we use a power law to infer it from the sales rank data, 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.[4] The parameter a does not affect our results, so we normalize it by setting $a = 1$. We assume $\phi = 1.7$, which is the number suggested by Smith and Telang in an experimental study on DVD sales on Amazon.[5] With the value of d and λ as noted above (i.e., calibrated to Amazon's fee schedule for DVD sales), we derive the distribution of DVDs being

[2] Note that such fee schedules (indeed affine fee schedules) are optimal mechanisms when sellers are perfectly competitive but will not be in general once sellers have market power. For instance, such fee schedules rule out the possibility that the platform can charge sellers joining fees. In practice, sellers are rarely charged joining fees. A formal analysis of why joining fees are not commonly used is outside the scope of our paper, but presumably there are practical difficulties for platforms to use such fees given that they face tremendous heterogeneity across different goods, for example, in the cost of sellers, the market size for each good, and the competition facing each seller. Without detailed information on each of these, it could be difficult for a platform to set appropriate joining fees in order to earn a higher profit than simply using the Bertrand affine fee schedule.

[3] 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.

[4] Power law distributions are widely used to describe rank data. See Judith Chevalier and Austan Goolsbee (2003). "Measuring prices and price competition online: Amazon.com and Barnesandnoble.com." *Quantitative Marketing and Economics*, 1(2), 203–222 for detailed discussions as well as an application to online sales data.

[5] See Michael Smith and Rahul Telang (2009). "Competing with free: The impact of movie broadcasts on DVD sales and internet piracy," *MIS Quarterly*, 33, 321–338.

sold on Amazon by extending our theory to allow for the possibility that for each good there can be a different number of potential buyers. The number of transactions for a distinct good i with cost c is $Q_{i,c} = m_{i,c}Q_c$ and the platform makes a profit $\Pi_{i,c} = m_{i,c}\Pi_c$, where Q_c and Π_c are the quantity and profit expressions from our theory based on a unit mass of potential buyers, and $m_{i,c}$ is the number of potential buyers for good i with cost c. We denote the number of distinct goods with cost c as m_c. A platform's total profit is therefore

$$\Pi = \sum_{c \in C} \sum_{i=1}^{m_c} m_{i,c}\Pi_c. \tag{20}$$

Given (20), all our previous analysis holds except that we need to change the mass g_c to $\sum_{i=1}^{m_c} m_{i,c}$. Given the value of λ, 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 $m_{i,c}$ for each good. Substituting $T(p_c) = a_0 + a_1 p_c$ into

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

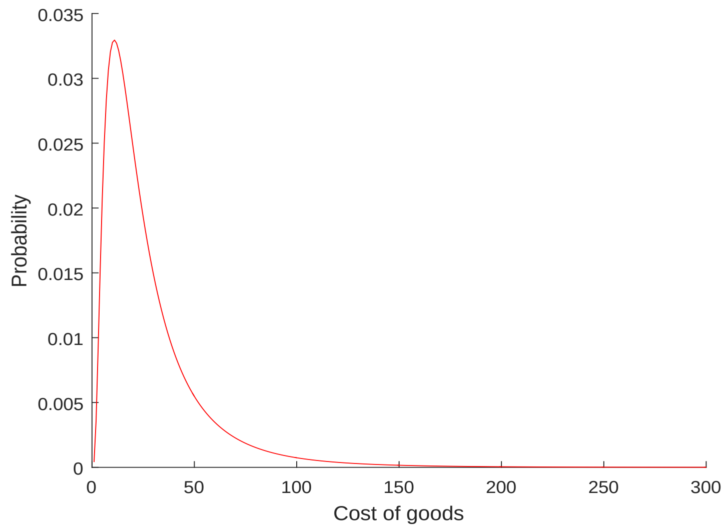
we derive

$$m_{i,c} = \frac{Q_{i,c}}{\left(\frac{p_c}{(1-a_1)p_c - a_0}\right)^{-\lambda}}.$$

The resulting weight $\sum_{i=1}^{m_c} m_{i,c}$ is the empirical measure of mass g_c. Finally, we use a log-normal distribution to fit the empirical distribution of g_c, which is evaluated at each of the 300 one-dollar bins of costs c up to \$300. The fitted log-normal distribution has two parameters (log mean =3.0347 and log standard deviation =0.8048). It is shown in Figure 5.

With this distribution of c given over the integer values of c from 1 to 300, we first work out the platform's profit when it can only choose a fixed per-transaction fee (i.e., without any price discrimination). The optimal fixed per-transaction fee turns out to be \$8.556, which results in the platform obtaining a total profit of 0.383 when integrating over all goods for which there are sales. Obviously, if the platform can observe each different good sold by the sellers, it can do better setting the per-transaction fee that is optimal for each good c. This increases its profit to 0.457. This 17.7% increase in platform profits is the gain attributable to price discrimination.

Figure 5: Fitted distribution of DVD sales



Consistent with our results in Section 3.1 above, we find that the benefits of this price discrimination can be obtained by using the Bertrand fee schedule, which has the added benefit of mitigating double marginalization. Indeed, the platform can increase its profit to 0.537, a further 16.3% increase in profits by using the Bertrand fee schedule. Taking account of the fact sellers are monopolists and the particular distribution of c, the platform can only increase its profit by a further 1.5% moving to the optimal affine fee schedule from the Bertrand fee schedule which is also affine.

The final and most challenging step is to work out the optimal (non-linear) fee schedule $T(p)$. We start from the optimal affine fee schedule, which is a degree one polynomial. We then increase the degree of the polynomial by one (to a quadratic) and find again the optimal polynomial fee schedule. In this case, the platform's profit increases by 1.2%. We repeat this procedure until the platform's profit no longer increases by more than 0.01% with the increase in the degree of the polynomial. Since a polynomial can approximate any continuous fee schedule, this approach is a practical way to search for the optimal function from the infinite possibilities for the optimal function.[6] Even this polynomial approach is computationally demanding, given

[6]It seems reasonable to expect the optimal transaction fee schedule should not only be continuous but smooth given there are 300 different values of c, the empirical distribution of which we have approximated with the smooth log-normal distribution. Any sudden change in fees, which could be desirable for extracting more from one monopolist with a particular value of c will likely reduce the amount that can be extracted from other monopolists with somewhat higher or lower values of c.

that we need to find $k + 1$ parameters for a k-degree polynomial, where for every feasible combination of the $k + 1$ parameters, we need to solve the monopolist's pricing problem for each of the 300 values of c, to obtain the platform's corresponding profit.[7]

The optimal fee schedule based on this approach turns out to be a 5^{th}-degree polynomial, which increases the platform's profit by less than 0.0005% compared to the 4^{th}-degree polynomial. The resulting fee schedule is convex for high prices. The overall effect on profits, however, is small. Compared to the optimal affine fee schedule, moving to the optimal fee schedule only increases the platform's profit by a further 1.3%.

We have repeated this exercise with linear demand. We use the same distribution of c and the same value of d, but assume $\sigma = 0$ and $\lambda = \frac{0.85}{0.3}$, which implies the Bertrand fee schedule has the same 15% proportional fee consistent with Table 1 in the main paper. The results are similar to those with constant elasticity demand; indeed, the Bertrand fee schedule does even better in this case.

The improvements in the platform's profit from moving from one fee schedule to another are summarized in Table 2 in the main paper. These, together with the different fee schedules, are shown graphically in Figure 6 below. The top panel in Figure 6 shows the fee schedules (fees as a function of the seller's monopoly price) for constant-elasticity demand (on the left) and linear demand (on the right). The green horizontal line in these plots represents the fixed per-transaction fee. The red dotted line represents the fee schedule implied by allowing the per-transaction fee to differ for each good in the hypothetical case in which each good's c is observed by the platform. The three remaining fee schedules are the Bertrand fee schedule, optimal affine fee schedule, and optimal non-linear fee schedule respectively. As can be seen, they are not very different, especially in the case of linear demand. The bottom panel of Figure 6 shows the platform's profit from each different good ($c = 1$, $c = 2$, ... $c = 300$). As can be seen the platform obtains higher profit on every good when it switches from a per-transaction fee for each good to the Bertrand fee schedule. However, there is almost no difference in profit between the Bertrand fee schedule, the optimal affine fee schedule, and the optimal non-linear fee schedule.

[7] We used the particle swarm algorithm in Matlab to speed up the optimization.

Figure 6: Fees and associated platform profits