Late fees and price discrimination

Sudipta Sarangi*, Randal Verbrugge

Virginia Polytechnic Institute and State University, Department of Economics, Blacksburg, VA 24061-0316, USA

Received 17 May 1999; received in revised form 13 December 1999; accepted 22 December 1999

Abstract

Late penalties for items (like videos) sometimes exceed the rental fee, which itself exceeds the re-renting opportunity cost. This puzzling feature of rental pricing can be justified as price discrimination when consumers have stochastic costs of returning items on time. © 2000 Elsevier Science S.A. All rights reserved.

Keywords: Video rentals; Price discrimination; Cost of time

JEL classification: D4

Late fees are an integral part of everyday life. If you return your videocassette or your rented car late, or are late in mailing your credit card payment, you pay an extra fee. One might argue that this penalty captures the opportunity cost to the seller. However, this could not explain the structure of late fees which existed in early 1999 at Video Update rental stores in the Blacksburg, Virginia, region. In particular, for videos not in the ‘new releases’ category, the rental fee was 47¢ per night rented, and the late fine was 75¢ per additional night — representing a surcharge (over an ordinary rental) of approximately 60%.¹ This presents a challenge to the opportunity cost hypothesis: failure to return a video on time is equivalent to checking the same video out for an additional unit of time. However, re-checking out a video to customer \( j \) in this fashion is less costly for the video store than having the video returned and rented to a different customer \( k \), in that the store avoids labor costs associated with re-stocking, checking out the video a second time, etc.²

While this pricing profile is difficult to justify on the basis of costs to the firm, it can be explained

¹Such pricing structures are not uncommon; other video chains feature similar patterns. (See the following web sites: http://www.netkitchen.com/inv/purdue.html, http://www.seacoast-sales.com/west.htm, and http://capitovideo.com/capitolvideo/html.)

²Furthermore, one might expect a discount for a re-rental, since for the majority of potential customer–video pairings (and any given level of marketing), a customer will rent a particular video only once. (We would expect that especially treasured videos would be purchased outright by the customer.)
straightforwardly in a model in which customers face random shocks to their cost of time (i.e. face random costs to returning rented goods on time). Costs to time are indeed stochastic — at the time of renting a video or a car, consumers do not plan on returning the product late . . . yet some invariably end up doing so! Put differently, late returns typically result from the arrival of unforeseen contingencies. (Interestingly enough, the VCR was initially viewed as a device for ‘time shifting’ television shows to more convenient hours — see Varian and Roehl, 1998). Late fees account for a substantial fraction of the revenue accruing to video rental establishments. According to one official at the local Blockbuster store, on a given day revenues from late penalties can vary between 10 and 50 percent. Industry analysts have suggested that video rental firms have been reluctant to adopt the Divx technology because it eliminates revenues from late fees (see King, 1999). In this note, we show that a rental fee/late fee pricing structure like the one mentioned above could be used to price discriminate between consumers who are heterogeneous in their distribution of shocks to the cost of time.\footnote{We hardly need point out that heterogeneity in costs of time is surely the rule rather than the exception; one need look no further than dispersion of wages to be convinced. Further, it is likely that the sorts of random shocks to the cost of time vary tremendously between, say, families with small children and college students.} We show that profit maximizing video rental firms might well set late penalties larger than ordinary rental fees, under particular conditions on the distribution of cost shocks faced by the customers.

1. A model of price discrimination in video rental markets

We consider two types of risk-neutral consumers, each of measure one. Each type $j$ ($j = 1, 2$) obtains income $y_j$ per period and possesses a utility function of the form

$$\begin{cases} \alpha x_j - \frac{1}{2}x_j^2 + m_j - f x_j & \text{if } f \leq C_j \\ \alpha x_j - \frac{1}{2}x_j^2 + m_j - C_j x_j & \text{if } f > C_j \end{cases}$$

where $x_j$ is the quantity of videos rented during the time period, $f$ is a late fine, $C_j$ is the (random) utility cost of returning the video on time, $\alpha$ is a fixed coefficient, and $m$ refers to money.\footnote{The structure imposed on the utility function allows us to derive a simple demand function.} Individuals 1 and 2 differ in their distribution of shocks to the value of time. Each individual $j$’s shock ($j = 1, 2$) is drawn (independently) from the set $\{c_j, B_j\}$, where $B_j > a$, with respective probabilities $(1 - s_j, s_j)$, where $s_1 < s_2 < \frac{1}{2}$. Thus, agent 2 has a higher probability of receiving a cost shock sufficiently large that paying a late fee is optimal. Since $s_2 < \frac{1}{2}$, there is a sense in which no customer expects to return his video late, and yet sometimes it is unavoidable.

Consumers must make their purchasing decisions before the realization of the cost shock. Given sufficiently high income (and $f > c_j$), the solution to consumer $j$’s problem is given by

$$x_j = a - s_j f - (1 - s_j) c_j - p \quad j = 1, 2$$

where $p$ is the rental price. If a consumer receives the high cost shock, she will not return the videos
on time (and will thus pay late fees), whilst if she receives the low cost shock, she will return the videos on time and avoid the late fees (but pay the utility cost of returning the video on time).

We assume that videos are supplied by a monopolist, an extreme assumption made for reasons of simplicity. The monopolist faces the demand of both types of consumers, and chooses \( p \) and \( f \) to maximize profits. The rental period decision is left unmodeled, and the cost structure is deliberately fairly stylized, in order to keep the analysis simple. The marginal cost per unit supplied is \( k \). Tapes returned on time reduce the marginal cost by \( t(s_1, k) \) per unit. Applying the law of large numbers, and provided that \( f \geq \max\{c_1, c_2\} \), the total number of late returns equals \( s_1x_1 + s_2x_2 \). The monopolist’s profit function is then given by:

\[
\Pi(p, f) = (p - k)(x_1 + x_2) + f(s_1x_1 + s_2x_2) + t[(1 - s_1)x_1 + (1 - s_2)x_2]
\]  

(3)

In the following proposition, assuming \( t > \max\{c_1, c_2\} \), we identify conditions under which the late fee can exceed the rental price.

Proposition 1. (Late fee exceeding rental price)

(a) If \( s_1 = s_2 \), optimal behavior by the monopolist only restricts the pair \( (p, f) \) to lie on a particular line segment, so that there exist \( (p, f) \) pairs such that \( f > p \).
(b) If \( s_1 \neq s_2 \), an optimizing monopolist will set the late fee above the price under either of the following sets of conditions:

(i) \( c_1 = c_2 = c, \text{ and } t + c > \frac{a + k}{2} \)

(ii) \( c_1 \neq c_2, t + \frac{c_1 + c_2}{2} - \frac{(c_2 - c_1)(1 - s_1)}{2(s_2 - s_1)} > \frac{a + k}{2} \),

along with the following condition on \( t \):

\[
\begin{align*}
 & t > c_2 + \frac{(c_2 - c_1)(1 - s_1)}{(s_2 - s_1)} & \text{if } c_2 > c_1 \\
 & t > c_1 + \frac{(c_1 - c_2)(1 - s_2)}{(s_2 - s_1)} & \text{if } c_1 > c_2
\end{align*}
\]

6We assume this reduction in costs by noting that tapes returned on time lead to lower overhead costs, fewer total videos needed in inventory, etc. Note that as \( t \) approaches \( k \), the more compelling is the simple opportunity cost explanation of late fees.

5We assume \( t > \max\{c_1, c_2\} \) for technical reasons, to guarantee that the optimal fine will induce each customer to return videos on time whenever he receives his lower cost shock. For example, in the case where \( s_1 = s_2 \) and \( c_2 > c_1 \), if \( t < c_1 \), optimal behavior implies that \( f < c_1 \), thus consumers are always late in returning tapes, and the monopoly collects \( (p + f) \) per video as an overall usage fee. This equilibrium outcome leaves the term ‘late fee’ rather devoid of meaning. Furthermore and unsurprisingly, in this case optimal behavior only restricts the pair \( (p, f) \) to lie on a particular line segment, yielding an infinity of price-late fee pairs which would satisfy \( f > p \). (We are grateful to an anonymous referee for urging us to consider all possible cases, thereby clarifying the conditions under which a ‘late fee’ makes sense, and for pointing out social optimality conditions discussed below).
Proof. See Appendix A.

As the proposition indicates, an optimizing monopolist may set the late fee larger than the rental price in a number of circumstances. Whenever the probability of receiving the high cost shock is identical across consumers, an optimizing monopolist receives identical profits for an infinity of \((p, f)\) pairs, some of which satisfy \(f > p\); thus, it might well choose to set the late fee larger than the rental price. When costs are identical across consumers, but the probabilities of receiving those shocks differ across consumers, an optimizing monopolist will set the late fee above the rental price if condition \(b\) \((i)\) holds. Finally, when both costs and probabilities differ across consumers, an optimizing monopolist will set the late fee above the rental price if the related condition \(b\) \((ii)\) holds.

How might one interpret these conditions? Note that an increase in \(f\) reduces demand, but has two benefits: it increases revenue per late return, and (provided \(f > \max\)) prompts customers who receive their lower cost shock to return videos on time, which reduces costs by \(t\) per video. The larger are the \(c’s\) and \(t\), ceteris paribus, the bigger the benefit to increasing \(f\), and the more likely is \(f\) to be greater than \(p\). When both probabilities and cost shocks differ, the condition on \(c_1, c_2\), and \(t\) is modified. When \(c_2 > c_1\), meaning the types are as different as possible, the monopolist’s ability to price-discriminate is enhanced, making it easier to satisfy the necessary condition on the \(c’s\) and \(t\); when \(c_2 < c_1\), the monopolist’s ability to price-discriminate is weakened, making it harder to satisfy the necessary condition on the \(c’s\) and \(t\).

When both probabilities and cost shocks differ, the ability to price-discriminate increases, and optimizing behavior takes into account the difference in the disutilities as well the difference in the probability of occurrence of the two types.

Given the values of \(f^*_s\) (optimal fee when \(c_1 = c_2\) and \(s_1 = s_2\)) and \(f^*_c\) (optimal fee when \(s_1 \neq s_2\), \(c_1 \neq c_2\)) in Appendix A, one can also compare the size of the late penalty with the monopolist’s cost savings. This has some optimality implications:

**Corollary.** (Behavior is socially optimal) Since \(\max\{c_1, c_2\} < f^* < t\), the consumers’ decisions to return tapes on time is always socially optimal.

**Proof.** Given the values of \(f^*_s\) and \(f^*_c\) in Appendix A, \(\max\{c_1, c_2\} < f^* < t\) always holds. When consumers return tapes on time, we can save the cost by \(t\) socially. However, consumers pay a time–cost for returning tapes on time; when this time cost exceeds \(t\), it is socially optimal for consumers not to return tapes on time. Since \(c_j < f^* < B_j\) always holds, consumers always return tapes on time when doing so is socially optimal.

In this model, consumer heterogeneity has two dimensions: consumers can differ in their disutility parameters and in their shock probabilities. The following proposition demonstrates that, provided consumers differ in the stochastic dimension, the price and late fee structure may be used as a price discrimination tool.

**Corollary.** (Profits from price discrimination) When \(s_1 \neq s_2\), monopoly profits are higher than under \(s_1 = s_2\), since in the former case, the monopolist can use the price and late fee to price discriminate between these groups.
Proof. See Appendix A.

In order to price discriminate, monopolists must be able to exploit heterogeneity amongst their customers in some fashion. As is evident from the demand functions (2), heterogeneity between the two customers in the probabilities of the two shocks implies that any given change in a late fee-rental price combination will provoke different responses by the two customers; this differential response allows the monopolist to use late fee-rental price combinations to price discriminate. In contrast, heterogeneity only in the disutility of returning the videos on time (i.e., differing $c_j$’s) is not exploitable — despite the difference in $c_j$’s, which causes the two types of consumers to differ in their quantities of videos rented, this type of heterogeneity has no impact on the response of each consumer to small changes in $f$ and $p$. Changes in $f$ affect consumers only when they receive their higher cost shock (which both receive with the same probability), and changes in $p$ affect consumers under all states of the world. Hence, both types of consumers respond to changes in late fee-rental price combinations in exactly the same manner.

2. Conclusion

It is commonly thought that late fees merely capture the opportunity cost to the seller. However, in this note we show that late fees may be used to price discriminate between consumers who receive random time cost shocks that have different distributions.

Acknowledgements

Thanks are due to Hans Haller, Rob Gilles and Nancy Lutz, and to an anonymous referee, whose comments were extremely useful. All errors, misinterpretations, and omissions are ours.

Appendix A

Proof of Proposition 1. Define $s = \frac{1}{2}(s_1 + s_2)$, $c = \frac{1}{2}(c_1 + c_2)$, and $\hat{c} = \max\{c_1, c_2\}$. Below is $\Pi_a^*(p,f)$, which is the optimal profit obtained if the firm faced the average consumer, i.e., when $s_1 = s_2$ and $c_1 = c_2$:

$$\Pi_a^*(p,f) = 2\left\{ \left( \frac{a-c}{2} \right)^2 + \left( \frac{k-t}{2} \right)^2 + \left( \frac{cs-ts}{2} \right)^2 + \frac{c}{2}(k-t) + \left( \frac{a-k}{2} \right)(c-t) + cst \right\} - \frac{s}{2}(c^2 + t^2)$$

In this case, it is easy to verify that there is an infinity of potential profit-maximizing choices; optimal behavior only restricts the optimal choice of $p^*$ and $f^*$ to a particular line segment, one which lies on the line

$$p^*_a + sf^*_a = \frac{1}{2}(a + k - t - c + s(c + t))$$

and satisfies $f > \hat{c}$ and $p < a$. Further, it is easy to check that even when $c_1 \neq c_2$, as long as $s_1 = s_2$, optimal behavior is identical to the previous case, as are optimal profits.
When \( c_1 = c_2 \) and \( s_1 \neq s_2 \), optimal profits \( \Pi^*_s(p,f) \) are given by

\[
\Pi^*_s(p,f) = \frac{1}{2}(a^2 + k^2 + c^2 + t^2) + t(a - k) + c(k - t) - a(k + c) + \left( \frac{c - t}{2} \right)(s_1^2 + s_2^2) + \left( \frac{s_1 + s_2}{2} \right)(ac + kt + ct - kc - at - c^2 - t^2)
\]

In this case, the optimal price and late fee are

\[
p^*_s = \frac{1}{2}(a - c + k - t) \quad \text{and} \quad f^*_s = \frac{1}{2}(c + t)
\]

As \( t > c \), \( f^*_s > c \) and \( f^*_s < t \). Further, the optimal fee exceeds the optimal price, or \( f^*_s > p^*_s \), whenever \( t + c > \frac{1}{2}(a + k) \). Next, we display \( \Pi^*_s(p,f) \), which represents optimal monopoly profits when \( s_1 \neq s_2, c_1 \neq c_2 \):

\[
\Pi^*_s(p,f) = \frac{1}{2}(a^2 + k^2 + t^2) + \frac{1}{4}(c_1^2 + c_2^2) + (a - k)t - ak + \left( \frac{a - k + 2t}{2} \right)(c_1s_1 + c_2s_2)
\]

\[
- \frac{c_1s_1}{2}(c_1 + ts_1) - \frac{c_2s_2}{2}(c_2 + ts_2) + \left( \frac{c_1 + c_2}{2} \right)(k - t - a) + \left( \frac{ts_1 + ts_2}{2} \right)(k + t - a)
\]

\[
+ \frac{s_1^2}{4}(t^2 + c_1^2) + \frac{s_2^2}{4}(t^2 + c_2^2)
\]

The optimal price and late fee from which this profit function is obtained are

\[
p^*_s = \frac{a + k - t}{2} - \frac{c_1s_1(1 - s_1) - c_2s_2(1 - s_2)}{2(s_2 - s_1)} \quad \text{and} \quad f^*_s = \frac{t}{2} - \frac{c_2(1 - s_2) - c_1(1 - s_1)}{2(s_2 - s_1)}
\]

When \( c_2 > c_1 \), \( t > c_2 + (c_2 - c_1)(1 - s_1)/(s_2 - s_1) \) is required to ensure that \( f^*_s > c_2 \); when \( c_1 > c_2 \), \( t > c_1 + (c_1 - c_2)(1 - s_2)/(s_2 - s_1) \) is required to ensure that \( f^*_s > c_1 \). It is straightforward (but somewhat tedious) to verify that \( f^*_s > p^*_s \), i.e., the optimal late fee is in excess of the optimal rental price, if the following condition is satisfied:

\[
t + \frac{c_1 + c_2}{2} - \frac{(c_2 - c_1)(1 - s_2) - (c_2 - c_1)(1 - s_1)}{2(s_2 - s_1)} > \frac{a + k}{2}
\]

Note that when \( s_1 \neq s_2 \), \( t > f^*_s > \hat{c} \) regardless of whether \( c_1 = c_2 \) or not.

It is straightforward to check that \( \Pi^*_s(p^*_s,f^*_s) > \Pi^*_h(p^*_h,f^*_h) \) and \( \Pi^*_s(p^*_s,f^*_s) > \Pi^*_h(p^*_h,f^*_h) \). Furthermore, it can be shown that \( \Pi^*_s(p^*_s,f^*_s) > \Pi^*_h(p^*_h,f^*_h) \) under most conditions. The simplest and most stringent sufficient condition for this is that \( c_2 > 3c_1 \); the weakest condition is too messy to show here. Finally, it is easily verified that the requirement does not violate any of the other conditions necessary for the propositions.

References