# Avoiding Ballot Stuffing in eBay-like Reputation Systems

Rajat Bhattacharjee [*]
Stanford University

Ashish Goel [†]
Stanford University

June 16, 2005

## Abstract

We present a preliminary study on the robustness of binary feedback reputation systems (e.g. eBay) to ballot stuffing and bad mouthing. In a feedback based reputation system, a seller can collude with other buyers to undertake fake transactions in order to enhance her reputation. This problem is referred to as ballot stuffing. A seller can also be targeted by a group of buyers to deliberately lower her reputation. This problem is referred to as bad mouthing. For the reputations to be meaningful, any practical reputation system needs to be resistant to these problems. We use a simplified model to give an explicit relation between the reputation premium and the transaction cost that needs to hold in order to avoid ballot stuffing. Thus we draw attention to the necessity of transaction costs for a well functioning reputation system. Our conclusions are confirmed by empirical experiments on eBay.

## 1 Introduction

Reputation systems form an important component of websites/services like eBay, Epinions and Amazon.com. In transactions and interactions on the Internet and peer-to-peer systems, where there is a lack of other traditional indicators of trustworthiness, feedback from previous customers plays a pivotal role. We refer the reader to [6, 13] for a more detailed introduction to reputation systems. Since reputations are essential for trust in electronic commerce, it is important that the purported reputation of an agent is an actual indicator of its trustworthiness. There are numerous fraudulent behaviors possible in a reputation system. Recent surveys (see [8, 11]) have shown high levels of fraud in online markets.

In this paper we concentrate on two strategies that can lead to reputations which are not reflective of the trustworthiness of a seller. A seller can collude with other buyers to undertake fake transactions resulting in positive ratings. These fake transactions have the effect of inflating a seller's reputation. This is known as ballot stuffing [3, 4]. On the other hand, a group of buyers can collude (may be at the behest of a rival seller) to deliberately give negative feedback to a particular seller and hence lower her reputation. This is known as bad mouthing. For reputations to be meaningful, any practical reputation system needs to be resistant to these behaviors. In this paper, we present restrictions on reputation premiums such that there is no incentive for agents to indulge in either ballot stuffing or bad mouthing. We give explicit relations between the reputation premium and the transaction costs that need to hold in order to avoid ballot stuffing. Thus we draw attention to the importance of transaction costs for a well functioning reputation system. We also point out how similar ideas can be extended to bad mouthing in a restricted setting.

We review three reputation systems used on the Internet: eBay, Amazon.com, Epinions. In eBay, for each transaction, the buyer and seller are allowed to rate each other by leaving feedback. Each feedback consists of a rating (positive, negative, or neutral), and a short comment. These ratings that members leave for each other are used to determine the reputation of individual members. The feedback score is the difference between the number of members who left a positive rating and the number of members who left a negative rating. As a measure to check ballot stuffing, eBay only counts the number of unique members who have given a positive rating. In Amazon.com, transaction partners can rate each other using the star rating system, from 5 stars (best) to 1 star (worst). Both buyers and sellers may leave feedback ratings. Although feedback is collected in a scale of 5, it is again changed to a positive/neutral/negative feedback system by assigning positive feedback for 5 or 4 stars, neutral for 3 stars and negative for 2 or 1 stars. Reputation systems are also used in other contexts. For example, Epinions.com uses it for rating reviews left by users. Each reviewer makes his/her own web of trust, consisting

of a network of reviewers whose reviews are useful. Epinions.com uses the Web of Trust system to predict how helpful a review will be to a user.

**Our Results.** As seen in the examples, the reputation systems currently in use tend to present information to the users rather than giving concrete suggestions as to how this information should be used. The issue is important because it is not clear to what extent a reputation system can be trusted when it is exposed to ballot stuffing and bad mouthing. We use the term reputation premium to denote the incremental cost a buyer is willing to pay for dealing with a seller of higher reputation. Our main result is to present constraints on reputation premium such that the reputation is inflation resistant. An inflation resistant reputation premium ensures that there is no incentive for sellers to fake transactions to enhance their reputations. We show that transaction costs ensure that a family of reputation premiums are inflation resistant.

In Section 2, we present related work in the field. Section 3 presents the model with which we work. Section 4 contains the main result and in Section 5 we present a caveat.

## 2    Related Work

Resnick et al. [14] report a controlled experiment conducted on eBay. The experiment involved a high-reputation established eBay dealer selling matched pairs of lots (batches of vintage postcards in this case). They observed that, in consonance with the expected, the established identity fared better than the one with no reputation. In this experiment, the difference in buyer's willingness-to-pay was $8.1\%$ of the selling price. In the case of relatively new sellers, one or two negative feedbacks did not affect buyer's willingness-to-pay. Another feature of eBay is the rarity of negative feedback. Resnick and Zeckhauser [12] observe that sellers received negative feedback only $1\%$ of the time. Eliciting feedback from entities is another problem. Many other empirical studies have been done on eBay; we refer the reader to [6, 14, 10] for pointers to these studies. Ba and Pavlou [1] studied eBay data to conclude that trust can be induced through market feedback mechanisms and enhancing trust can lead to more profitable outcomes for sellers. Lucking-Reiley et al. [9] show that higher reputation leads to higher reputation premium. Bolton, Katok and Ockenfels [2] study and compare electronic reputation mechanisms with and without online feedback. They conclude that mechanisms with feedback are more efficient than those without feedback.

The reputation systems for Internet based transactions are different from conventional business transactions as the sellers who have got negative feedback can leave the system and join under a new name. Friedman and Resnick [7] study this scenario and conclude that buyers need to impose some disadvantage on sellers with no feedback at all. Among the suggestions they make, they discuss the possibility of charging for name changes, that is, a newcomer has to make some payment to the system before joining the system.

Dellarocas [5] introduced a formal model for binary reputation mechanisms like the one used by eBay. Here a buyer gives a rating of $+1$ or $-1$ (i.e. "good" or "bad") after a transaction is over. It is shown that leniency of buyer while rating sellers is important for the system to settle down to steady-state levels avoiding oscillations. He argues that the ratio of negative feedback to total feedback is a good indicator of reputation. Also, Dellarocoas [3, 4] considers the problems of ballot stuffing and bad mouthing. The methods suggested in these papers are based on anonymity and cluster filtering. Anonymising buyers and sellers could be potentially useful to curtain bad mouthing. However anonymising will not work for ballot stuffing as the seller can give hidden indications of its identity to its colluders. For example, the seller might signal its colluders by pricing its products at a price having a specific decimal point. The method of cluster filtering involves profiling buyers based on their other transactions and feedback. Then these profiles are used to give weights to specific feedback. Our work contrasts with these methods as we study the inherent capability of binary feedback based reputation system to deter dishonest behavior without using additional heuristics.

## 3    Model

In a binary reputation system, a buyer provides feedback for a seller after a transaction. This feedback is either "positive" or "negative". This is very similar to eBay where the buyer provides feedback in the form of $\{-1, 0, 1\}$ and also any additional comments. However, there is an important difference. In eBay, a bad feedback by a buyer can be countered by a bad feedback from the seller for the buyer. Hence, a buyer while writing a feedback, is mindful of the fact that the seller would write a feedback for the buyer. In eBay, there is also a provision for a seller and buyer to mutually withdraw feedbacks. In other words, the mechanism has reciprocity and retaliation. This is not captured in our simplified model. The reputation of a seller then constitutes the net feedback, i.e., the number of positive feedbacks less the number of negative feedbacks accrued by the seller. Any

additional comment made by the system is also displayed. Dellarocas [5] has studied these binary reputation mechanisms. It is worth noting here that binary reputation systems are not the only kind of reputation systems being used. For example, Amazon asks its users to rate sellers in a scale of 5. There are other schemes which use a more complex feedback. We confine ourselves to the important subclass of binary reputation systems. We make the following assumptions in our model:

**Price of goods:** It is assumed for simplicity that there is only one good that the seller wishes to sell and the manufacturing cost of the good by the seller is 1. We assume that the cost of manufacturing for all the sellers is the same and hence the manufacturing cost of each seller is 1.

**Feedback:** Each transaction results in a positive feedback or a negative feedback. It is assumed that the sums of positive and negative feedbacks displayed by the system spans the entire history of the seller. However, it might make sense to only display results for a particular time window. For example, eBay only spans a history of six months. We note here that our results hold for such systems too.

**Intrinsic error rate:** The rate at which a seller commits errors (i.e., unsatisfactory transactions) is denoted by $\epsilon$. The error rate, $\epsilon$, is assumed to be intrinsic and hence fixed to a seller. All the errors result in negative feedback. Also, any genuine transaction which is not an error results in positive feedback. An error can be a result of manufacturing errors or things not in the hands of the seller like postal delay. It should be noted that these assumptions are not simple ones as empirical studies have shown that buyers tend to negotiate before giving negative feedback. Moreover, it is difficult to elicit feedback from buyers [12, 13].

**Ballot stuffing rate:** The rate at which fake transactions leading to a positive feedback occur (for a particular seller) is denoted by $\rho$, i.e. for each genuine transaction undertaken by the seller, there are $\rho$ fake transactions on the average which all lead to positive feedback.

**Perceived error rate:** The rate at which the system observes a seller making error is denoted by $\epsilon_p$:

$$\epsilon_p = \frac{\epsilon}{1 + \rho}$$

Thus a seller's perceived error rate can decrease with an increase in $\rho$.

**Central System:** It is assumed that there is a central system that keeps record of the feedback and collects transaction costs.

**Reputation premium:** Buyers use the reputation of a seller in order to make decision on a purchase. We model this effect by means of a function we call reputation premium, $\xi$. It is a function of the perceived error rate $\epsilon_p$. Reputation premium implies that the seller can sell the product at a price $1 + \xi$, i.e. it represents the buyers' willingness-to-pay for higher reputation. We make the assumption that $\xi$ is continuous and differentiable. Based on the work of Dellarocas [5], we assume that $\xi$ is a function of the perceived error rate, $\epsilon_p$. We also assume that the reputation premium decreases with an increase in perceived error rate, $\epsilon_p$, i.e.,

$$\xi'(\epsilon_p) \leq 0. \tag{1}$$

**Transaction Costs:** Each transaction has a fixed seller transaction cost, $\tau$, paid by the seller to the central system. We note that this is a simplistic assumption and ordinarily the transaction cost would be a percentage of the selling cost. We analyse this scenario in the extended version of this paper and show that this scenario is not different from that of fixed transaction cost in the regime of small (relative to selling price) reputation premium and transaction costs.

# 4 Reputation Premiums resistant to Ballot Stuffing

**Definition.** An $(\epsilon, \rho)$-Bernoulli seller is a seller for which the intrinsic error rate is $\epsilon$ and the ballot stuffing rate is $\rho$. More precisely, the seller errs in a genuine transaction with probability $\epsilon$ independent of all other transactions. Moreover, the probability that a transaction is fake is $\frac{\rho}{1+\rho}$ independent of all other transactions.

**Definition.** The transaction cost per genuine transaction is denoted by $\psi(\rho)$ and is equal to $\tau(1 + \rho)$.

**Definition.** A reputation premium, $\xi$, is inflation resistant if the following relation holds:

$$\left.\frac{\partial}{\partial \rho}(\xi - \psi)\right|_{\rho=0} \leq 0. \tag{2}$$

In other words, $\xi$ should be such that the increase in reputation premium due to increase in positive fake transactions is offset by the increase in transaction costs. An inflation resistant reputation premium gives no

3

incentive for sellers to fake transactions to enhance their reputations. Note that $\xi$ is a function of the perceived error rate, which is a function of the actual error rate and the number of fake transactions. Although it is clear that a higher perceived error rate should result in a lower reputation premium and a lower perceived error rate should result in a higher reputation premium, it is not clear what the exact dependence should be. The next theorem gives a restriction on the reputation premium such that any reputation premium satisfying the constraint would be inflation resistant.

**Theorem 1** *A reputation premium is inflation resistant if for all $x \in (0,1)$:*

$$-x\xi'(x) \le \tau, \qquad (3)$$

*where $\tau$ is the transaction cost.*

**Proof:** Suppose equation (3) holds for $\xi$. From (2), $\xi$ is inflation resistant if

$$\frac{\partial}{\partial \rho}(\xi - \psi)\Big|_{\rho=0} \le 0$$

$$\Leftrightarrow \quad \frac{\partial \xi(x)}{\partial \rho}\Big|_{\rho=0} \le \frac{\partial(\psi)}{\partial \rho}\Big|_{\rho=0} = \tau.$$

However,

$$\begin{aligned}
\frac{\partial \xi(x)}{\partial \rho}\Big|_{\rho=0} &= \frac{\partial}{\partial \rho}\xi\left(\frac{\epsilon}{1+\rho}\right)\Big|_{\rho=0} \\
&= -\frac{\epsilon}{(1+\rho)^2}\xi'\left(\frac{\epsilon}{1+\rho}\right)\Big|_{\rho=0} \\
&= -x\xi'(x) \le \tau.
\end{aligned}$$

Hence, $-x\xi'(x) \le \tau$, $\forall x \in (0,1)$ is a sufficient condition for $\xi$ to be an inflation resistant reputation premium. $\square$

The above result only gives a restriction on the reputation premium to ensure that the reputation premium is inflation resistant. The actual reputation premium to be used might depend on many other factors. For example, buyers might choose to pay much higher premiums if say they attach a very high importance to the guarantee of quality. However, if the reputation premium meets the above constraint then it would be inflation resistant.

### 4.1 How high can reputation premiums be?

We now consider the following question: assuming a reputation premium is inflation resistant, how high can a reputation premium be? Suppose $\xi$ is an inflation resistant reputation premium satisfying equation (4) by equality at

each point. We assume that $\xi(1) = 0$, i.e., if the perceived error rate of a seller is 1, then she doesn't receive any reputation premium. Therefore,

$$\xi(\epsilon_p) = -\int_{\epsilon_p}^1 \xi'(x)dx = \int_{\epsilon_p}^1 \frac{\tau}{x}dx = (\tau)\log\frac{1}{\epsilon_p}. \quad (4)$$

Note that $\xi \to \infty$ as $\epsilon_p \to 0$.

### 4.2 A Linear Reputation Premium

An obvious choice for reputation premium is $\xi(\epsilon_p) = \tau(1 - \epsilon_p)$. The reputation premium is clearly inflation resistant $\forall x \in [0,1]$,
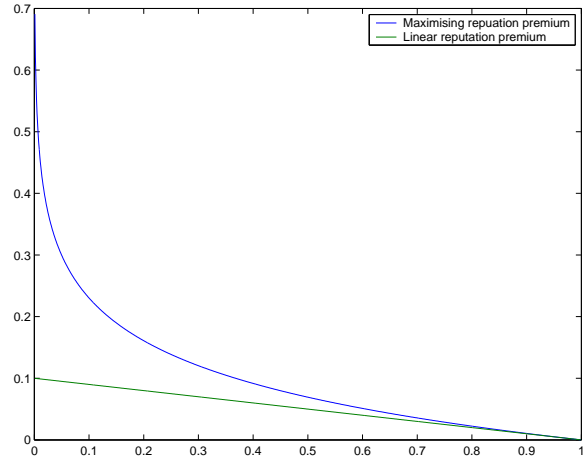
$$-x\xi'(x) = x\tau \le \tau$$

Figure 1: $\tau = 0.1$

### 4.3 Importance of Transaction Costs

The most important contribution of this paper is that it highlights the necessity of transaction costs for a reputation system to work well. Not only is the transaction cost important for a reputation system to function well, the reputation premium accrued to the seller should be related to the transaction cost. This is in consonance with the empirical results of Resnick et al. [14]. It is demonstrated by experiments that the difference in buyer's willingness-to-pay is $8.1\%$ of the selling price, when buyers with established reputation and no reputation are compared. We note that the order of the reputation premium is the same as that of the transaction cost.

### 4.4 Bad Mouthing

Here we show that our analysis carries over to bad mouthing, although in a restricted setting. We now as-

4

sume that the rate at which fake transactions leading to a negative feedback occur is denoted by $\eta$, i.e., for each genuine transaction undertaken by the seller, there are $\eta$ fake transactions on the average which all lead to negative feedback. The perceived error rate, $\epsilon_p$, now becomes $\frac{\epsilon+\eta}{1+\eta+\rho}$. We assume that each transaction now also has a transaction cost for the buyer which we denote by $\beta$ (this may correspond to shipping and handling costs). It is further assumed that $\beta > \tau$. We note that this is a very restrictive condition, as it is saying that the transaction cost of the buyer is more than the transaction cost of the seller. The transaction cost for the buyers intending to deflate the sellers reputation is denoted by $\chi(\eta) = \beta\eta$. A reputation premium, $\xi$, is deflation resistant if the following relation holds:

$$\frac{\partial}{\partial \eta}(-\xi + \psi - \chi)\Big|_{\eta=0} \leq 0. \qquad (5)$$

In other words, $\xi$ should be such that the decrease in reputation premium and increase in transaction costs of the seller should not be more than the transaction costs of the malicious buyers. We claim that a reputation premium is deflation resistant if

$$-(1-x)\xi'(x) \leq \beta - \tau.$$

The proof is similar to the inflation resistant premium along with the observation that

$$
\begin{aligned}
\frac{\partial \xi(x)}{\partial \eta}\Big|_{\eta=0} &= \frac{\partial}{\partial \eta}\xi\left(\frac{\epsilon+\eta}{1+\rho+\eta}\right)\Big|_{\eta=0} \\
&= \left(\frac{1}{1+\rho+\eta} - \frac{\epsilon+\eta}{(1+\rho+\eta)^2}\right)\xi'\left(\frac{\epsilon+\eta}{1+\rho+\eta}\right)\Big|_{\eta=0} \\
&= -\frac{1-x}{1+\rho}\xi'(x) \leq -(1-x)\xi'(x) \leq \beta - \tau.
\end{aligned}
$$

The maximizing $\xi$ would now be the one which has $-\xi'(x) = \frac{\tau}{x}$, $\forall x \in [\frac{\tau}{\beta}, 1]$ and $-\xi'(x) = \frac{\beta-\tau}{1-x}$, $\forall x \in (0, \frac{\tau}{\beta})$. Hence, $\xi(\epsilon_p) = \tau \log \frac{1}{\epsilon_p}$, $\forall x \in [\frac{\tau}{\beta}, 1]$ and $\xi(\epsilon_p) = \tau \log \frac{\beta}{\tau} + (\beta-\tau) \log \frac{(1-\epsilon_p)\beta}{\beta-\tau}$, $\forall x \in (0, \frac{\tau}{\beta})$. It is easy to verify that the linear reputation premium, $\xi(\epsilon_p) = \tau(1-\epsilon_p)$ is also deflation resistant for $\beta \geq 2\tau$. Figure 2 plots the maximizing and the linear reputation premium for $\beta = 0.2, \tau = 0.1$.

## 5 Caveat

In this section, we draw attention to the problem of an initial strategizing seller.

**Definition.** A $\rho$-initial strategizing seller is one who does $\rho k$ positive fake transactions and then undertakes $k$ genuine transaction with a Bernoulli error rate $\epsilon$. An honest Bernoulli seller is nothing but the 0-initial
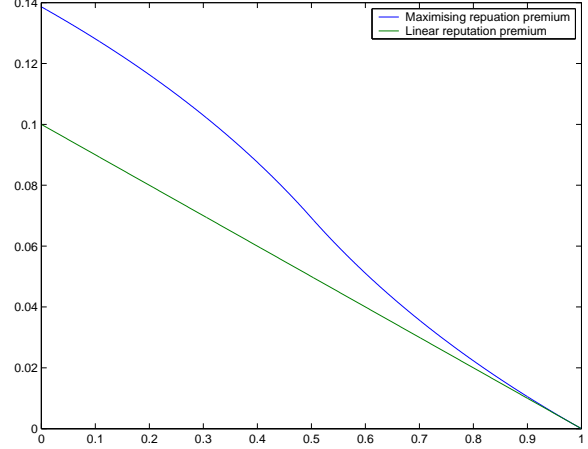


Figure 2: $\tau = 0.1, \beta = 0.2$

strategizing seller.

For the purpose of discussion in this section, we assume that the reputation premium used by the system is the linear reputation premium discussed above. The total reputation premium accrued by a $\rho$-initial strategizing seller is approximately equal to:

$$
\begin{aligned}
\sum_{i=1}^{k} \tau\left(1 - \frac{\epsilon i}{\rho k + \epsilon i}\right) &\approx k \int_0^1 \tau\left(1 - \frac{\epsilon x}{\rho + x}\right) dx \\
&= k\tau\left(1 - \epsilon\left(1 - \rho \log\left(\frac{1+\rho}{\rho}\right)\right)\right)
\end{aligned}
$$

The total reputation premium accrued by a 0-initial strategizing seller (honest) is: $\tau(1-\epsilon)$. Hence, the net reputation premium that the seller makes by being a $\rho$-initial strategizing seller is:

$$k\tau\epsilon\rho\log\left(\frac{1+\rho}{\rho}\right)$$

On the other hand, the excess transaction cost paid by a $\rho$-initial strategizing seller is: $\tau\rho$. Hence in the scenario

$$k\tau\epsilon\rho\log\left(\frac{1+\rho}{\rho}\right) > k\tau\rho$$

$$\rho < \frac{1}{exp\left(\frac{1}{\epsilon}\right) - 1}$$

the initial strategizing seller is going to benefit.

The discussion in this section reinforces the need for a special treatment to new sellers. Please see [7] for a detailed treatment of this subject. In our setting, the extra profit that the initial strategizing seller can make is unbounded (see appendix A), making the problem even more challenging. Providing disincentives against initial-strategizing remains an important open problem.

# References

[1] S. Ba and P.A. Pavlou. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quart.*, 26:243–268, 2002.

[2] G.E. Bolton, E. Katok, and A. Ockenfels. How effective are electronic reputation mechanisms? an experimental investigation. *Managament Science*, 50(11):1587–1602, November 2004.

[3] C. Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discrimatory behavior. *Proceedings of the second ACM Conference on Electronic Commerce*, October 2000.

[4] C. Dellarocas. Mechanisms for coping with unfair ratings and discriminatory behaviour in online reputation reporting systems. *International conference on Information Systems*, December 2000.

[5] C. Dellarocas. Analysing the economic efficiency of ebay-like online reputation reporting mechanisms. *Proceedings of the third ACM Conference on Electronic Commerce*, October 2001.

[6] C. Dellarocas. The digitization of word-of-mouth: promise and challenges of online reputation systems. *Management Science*, October 2003.

[7] E. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(1):173–199, 2001.

[8] GartnerG2. Online transaction fraud and prevention get more sophisticated. http://www.gartnerg2.com/rpt/rpt-0102-0013.asp. 2002.

[9] D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves. Pennies from ebay: The determinants of price in online auctions. *Working paper, Vanderbilt University, Nashville, TN*, 1999.

[10] M.I. Melnik and J. Alm. Does a seller's e-commerce reputation matter? *J. Indust. Econom.*, 50(3):337–349, 2002.

[11] U.S. Department of Justice. Internet fraud. http://www.internetfraud.usdoj.gov. 2002.

[12] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: empirical analysis of ebay's reputation system. *The economics of the Internet and e-commerce*, 11:127–157, 2002.

[13] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. *Communications of the ACM*, 43(12):45–48, December 2000.

[14] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment. *Working paper originally presented at the ESA conference*, June 2002.

# A  Unbounded returns for the initial-strategizing seller

Let $\pi(k, \rho)$ denote the extra profit (i.e. the extra reputation premium minus the extra reputation cost) that a $\rho$-initial-strategizing seller can make after $k$ genuine transactions. In section 5 we showed that

$$
\begin{aligned}
\pi(k, \rho) &\approx k\tau\epsilon\rho \log\left(\frac{1+\rho}{\rho}\right) - k\tau\rho \\
&\geq k\tau\epsilon\rho \log\left(\frac{1}{\rho}\right) - k\tau\rho.
\end{aligned}
$$

The second quantity above is maximized when $\rho = \exp(-1 - 1/\epsilon)$, and with this value of $\rho$, we obtain $\pi(k, \rho) = k\rho\tau\epsilon$. Here $\tau$ and $\epsilon$ are intrinsic parameters of the system and the merchant respectively, and $\rho$ depends only on $\epsilon$. Hence, by letting $k$ grow arbitrarily large, the extra profit $\pi(k, \rho)$ becomes unbounded. Interestingly, the rate of return (i.e. the ratio of the extra profit to extra investment) for this choice of parameters is $\approx \epsilon$.