Abstract

The generalized second-price (GSP) mechanism is the most widely-used auction format in sponsored search markets. However, figuring out how to bid on GSP auctions presents major theoretical and computational challenges due to the complex nature of the auction format and the infinite number of equilibria. Our study characterizes various equilibrium bidding behaviors in GSP auctions. We develop an algorithm to identify all pure-strategy Nash equilibria and discuss their distribution in the pure-strategy space. This equilibrium distribution can help advertisers formulate bidding strategies, and help search engines calculate their expected revenues.

Keywords: Keyword auctions, Nash equilibrium, generalized second-price, equivalence relation

1. Introduction

Keyword advertising is the fastest-growing sector in Internet advertising market. In the US market, the total revenue of Internet advertising market in 2008 is $23.4 billion, 45% of which came from keyword. The increase rate of keyword advertising is 20%, while the entire Internet advertising market increased only 10.6%, according to IAB. In the Chinese market, the revenue of keyword advertising is about $430 million in the first half year of 2009, with a 29.9% increase over the same period of the previous year, while the revenue increase rate of the entire Internet advertising market is only 2.9%, according to DCCI. Meanwhile, keyword advertising is also the most important and fast growing revenue source for auctioneers. According to the financial report provided by Google, its revenue for the second quarter of 2009 is $5.52 billion, about 95% of which came from keyword advertising. While Yahoo! reported its revenue for the same quarter to be $1.14 billion, of which $359 million came from keyword auctions.

The Generalized Second-Price (GSP) mechanism is the mainstream auction format used by search engine companies to sell keywords to advertisers. In a GSP auction, advertisers compete for keyword-specific sponsored search slots offered by the search engine. When a user performs a Web search, the winning advertisements will be displayed in the sponsored search region of the Web search engine interface and the winning advertiser pays the search engine an amount equaling the next-highest bid below this advertiser’s bid if his or her link is clicked by the user (Jansen et al. 2008; Muthukrishnan 2008).

Despite its tremendous commercial success, figuring out how to formulate a bidding strategy on a GSP auction poses serious academic and practical challenges from the point of view of advertisers. At the same time, auctioneers or search engine companies also find it difficult to evaluate the performance of the GSP mechanism. These challenges stem from the complicated nature of the GSP auction format and the existence of an infinite number of pure-strategy equilibria. Although several auctioneers and third-party companies (e.g., Keyword Country, adSage, etc.) provide various services and software tools to help advertisers make bidding decisions, most of these services or tools are heavily based on ad-hoc heuristics and human intelligence without a proper theoretical or computational foundation (Kitts et al. 2004).

In this paper, we will analyze the Nash equilibria of the GSP auction. Several papers have been published on this topic but they all focus on developing solutions based on Nash equilibrium refinements. (Edelman et al. 2007) shows that GSP has no dominate strategy and “truth-telling” may not always be a Nash equilibrium. Locally Envy-Free (LEF) (Edelman et al. 2007) and Symmetric Nash Equilibrium (SNE) (Varian 2007; Varian 2009) are proposed as the refinements of Nash equilibrium. LEF and SNE are found
to be able to explain certain bidding behavior observed on Google markets. (Borgers et al. 2007) provides the existence conditions for SNE in a non-separated model and uses an example to show that inefficient Nash equilibria exist in GSP auctions.

We present an algorithm to characterize and identify all pure-strategy Nash equilibria of a GSP auction and discuss their distribution in the pure-strategy space. We find that all these equilibria can be divided into several distinct classes and each class forms a convex polyhedron. With an understanding of this distributional property, an advertiser can determine which equilibrium polyhedron serves his best interest, and calculate the best bid adjustments according to the distance between his best-response equilibrium polyhedron and other equilibria. Auctioneers, on the other hand, can compute the minimum and maximum revenue based on this equilibrium structure and can compare these values with revenues from other kinds of auction mechanisms, such as the VCG mechanism, in order to evaluate the relative performance of the GSP mechanism. To the best of our knowledge, our work is the first to study the entirety of the Nash equilibria of GSP auctions and examine their distribution structure.

The remainder of this paper is organized as follows: we introduce the basic model of GSP auctions in Section 2. Section 3 discusses the problem of classifying and characterizing all Nash equilibria and analyzes their distribution. We provide two examples in Section 4 and conclude the paper in Section 5 with a summary of our findings and our future research directions.

2. A Model of GSP Auction

There are \( N \) bidders competing for a single keyword. Let \( N = \{1, 2, \cdots, N\} \) denote this bidder set. Each bidder \( n \) evaluates and determines the average value of a single click on his advertisement link to be \( v_n \). We assume that \( v_1 > v_2 > \cdots > v_N \). The search engine provides \( K \) slots. Let \( K = \{1, 2, \cdots, K\} \) denote this slot set. If the bidder \( n \)'s advertisement is allocated in slot \( k \), this advertisement will be visited with a click-through rate \( c_n^k \). We assume that \( \forall n \in \mathbb{N}, \sum_{k=1}^{K} c_n^k \leq 1 \), \( \sum_{n=1}^{N} c_n^k \leq 1 \), \( c_n^k \geq 0 \), \( c_n^k \geq c_n^{k+1} \), \( c_n^k \geq c_n^{k-1} \), \( c_n^k \geq c_n^l \). We assume that \( \forall n \in \mathbb{N}, c_n^k \leq c_n^l \).

We consider the problem of allocating \( K \) slots to \( N \) bidders. Let \( b_n \) be the bid that bidder \( n \) submitted and \( b=(b_1, b_2, \cdots, b_N) \) be the bidding vector of all \( N \) bidders. Denote the GSP auction mechanism as \( \pi: K \rightarrow N \) is the allocation rule, such that \( \forall k \in K, p: K \rightarrow R_+ \) is the payment rule, such that \( p_k = p(k) \) is the price which the bidder allocated in slot \( k \) must pay when his advertisement is clicked. Under the GSP mechanism the price a bidder pays is equal to the bid of the bidder who allocated just below him, \( p_k = \min_{k' \in \pi^{-1}(k)} b_{k'} \). Thus, if bidder \( n \) is allocated in slot \( k \) and his advertisement is clicked, he will receive an expected profit of \( u_n = c_n^k (v_n - p_{\pi^{-1}(k)}) \).

To be convenient, let \( \pi_{K+1} \) be the bidder who loses in this auction with the highest bid, with \( b_{\pi_{K+1}} \) being the highest lost bid; \( \psi(\cdot) \) be the inverse of \( \pi(\cdot) \); \( \pi(K) \) be the set of bidders who get a slot (thus all bidders in set \( N \setminus \pi(K) \) lose in the auction). We assume \( p_k = 0 \), \( \forall k > K \) and that the reserve price is zero. Furthermore, we assume that all bidders are rational, risk-neutral, and do not have any budget restrictions. Using the notation developed above, we describe a GSP keyword auction as a 5-tuple:

\[
(N, K, M, C, v)
\]

where \( C = \{c_n^k \mid \forall n \in \mathbb{N}, \forall k \in K\} \) is the set of all click-throughs; \( v = (v_1, v_2, \cdots, v_N) \) is the vector of all bidder’s value. As typical with the existing GSP models, we assume complete information.

3. All Nash Equilibria of the GSP Auction

In this section, we characterize all pure-strategy Nash equilibria of the GSP auction as defined by Model (1) and analyze the distribution of these equilibria in the pure-strategy space. A bidding vector \( \hat{b} \) is a Nash equilibrium if it satisfies the following fixed-point conditions (Fudenberg et al. 1992).
\[ b^*_n \in \arg \max_{b_n} u_n = c^k_n(v_n - b_{n^*_k}), \quad \forall n \in \mathbb{N} \] 

(2)

At a Nash equilibrium point, no bidder in set \( \pi(K) \) wants to raise or lower slot, and no bidder in set \( \mathbb{N} / \pi(K) \) has incentive to change his bid to get a slot. These properties can be formulated as the following inequalities.

\[
\begin{align*}
 v_n - b_{\pi_{n^*_k}} & \geq 0, \quad \forall n \in \pi(K) \\
c^k_n(v_n - b_{\pi_{n^*_k}}) & \geq c^k_n(v_n - b_{n^*_k}), \quad \forall 1 \leq k < \psi_n, \forall n \in \pi(K) \\
c^k_n(v_n - b_{\pi_{n^*_k}}) & \geq c^k_n(v_n - b_{n^*_k}), \quad \forall \psi_n < k \leq K, \forall n \in \pi(K) \\
v_n - b_{n^*_k} & \leq 0, \quad \forall k \in K, \forall n \in \mathbb{N} / \pi(K)
\end{align*}
\]

(3)

Past analyses of SNE and LEF have shown that there are an infinite number of Nash equilibria (Edelman et al. 2007; Varian 2007; Varian 2009). Let \( E \) denote the set of all Nash equilibria in a GSP auction.

### 3.1 Classification of Pure-Strategy Nash Equilibrium

All inequalities that a Nash equilibrium needs to satisfy are in linear form. However, the problem of finding all Nash equilibria cannot be solved by employing solvers of linear inequalities, such as those methods introduced in (Solodovnikov 1979), because the implicit allocation rule is nonlinear. In practice, the GSP allocation mechanism is realized by the following two steps:

- Sorting the bidders according to their bids in the decreasing order.
- Allocating the first slot to the first bidder, the second slot to the second bidder, and so on.

Tie-breaking is needed in cases which more than one bidder submit bids in the same amount, this situation is typically handled through assigning these bidders slots with equal probability. Fortunately, tie-breaking does not need to be considered in equilibrium bidding analysis if the bidders’ values are all distinct, based on the following theorem.

**Theorem 1.** Suppose \( v_1 > v_2 > \cdots > v_N, \forall b \), a necessary condition for \( b \) to be an Nash equilibrium of GSP auction (1) is

- If \( N \leq K \), \( b \) satisfies \( b_{\pi_1} > b_{\pi_2} > \cdots > b_{\pi_N} \)
- If \( N > K \), \( b \) satisfies \( b_{\pi_1} > b_{\pi_2} > \cdots > b_{\pi_N} > b_{\pi_K} \)

The proof of Theorem 1 is omitted due to the page limit. It indicates that a tie in a GSP auction is not a Nash equilibrium. With ties safely ignored, we can find all Nash equilibria in two steps. First, we define an equivalence relation on \( E \) and partition \( E \) into a finite number of equivalence classes according to this relation. Second, since each equivalence class determines a unique allocation, we can substitute this allocation into inequalities (3) and transform them into linear inequalities on bids (given that the allocation rule \( \pi \) is fixed). The inequality methods introduced in (Solodovnikov 1979) can then be used to compute all solutions. (Note that linear programming techniques can be used to quickly determine whether feasible solutions exist but in many cases they cannot fully characterize these solutions.) We now define one such equivalence relation we have identified in our research, which is able to provide a simple yet surprisingly powerful framework for both theoretical analysis and computation with respect to GSP.

**Definition 1.** “\( \equiv \)” is a relation (also referred to as the “same-slot” relation) on \( E \), such that \( \forall b^1, b^2 \in E, b^1 \equiv b^2 \iff \pi^1(k) = \pi^2(k), \forall k \in K \), where \( \pi^1 \) and \( \pi^2 \) denote the allocation corresponding to \( b^1 \) and \( b^2 \), respectively. Intuitively, this relation groups all bidding vectors together so long as they deliver the same allocation. Proof of the following result concerning the same-slot relation is straightforward, here it has been omitted.

**Theorem 2.** “\( \equiv \)” is an equivalence relation on set \( E \).
Now, we can partition $E$ into distinct equivalence classes based on relation \( \equiv \). For $\forall \vec{b} \in E$, the equivalence class generated by $\vec{b}$ is

$$E_b = \{ \vec{b} \mid \vec{b} \equiv \vec{b}, \forall \vec{b} \in E \}$$  \hspace{1cm} (4)

All different equivalence classes of set $E$ form the factor (quotient) set

$$E/ \equiv \{ E_b \mid \forall \vec{b} \in E \} = \{ E_0, E_1, \ldots, E_{M-1} \}$$  \hspace{1cm} (5)

where $M$ is the number of different equivalence classes. $E_0$ is a special and the most important equivalence class, in which the allocation is the identity mapping, i.e., $\pi^N(i) = i, \forall i \in K$. It is obvious that only Nash equilibria in $E_0$ are efficient and all others are inefficient.

Consider an equivalence class $E_m$. If a bidder, say $i$, revises his bid (given no changes from other bidders) but the revised bid vector is still in the same equivalence class $E_m$, then his bid revision has no impact on the allocation and his utility. As such, all equilibria in the same equivalence class are indifferent from this bidder’s point of view (unless other factors such as risk are considered). However, bidder $i$’s action affects (and can only affect) the utility of the bidder assigned just above him, say $j$. A reduction of $i$’s bid will increase $j$’s utility and an increase will reduce $j$’s utility.

### 3.2 Computing all Nash Equilibrium Classes

First, we designate an upper bound for the possible number of equivalence classes (proof omitted).

**Theorem 3.** *In any GSP keyword auction, the number of different equivalence classes is at most $N!$, if $N \leq K$; or $K! + (N-K)(K-1)!$, if $N > K$.*

In order to find all Nash equilibrium classes, there is a direct, naive algorithm as follows:

- **Begin**
  1) Set $m = 0$
  2) Choose an equivalence class $E_m$ (determine an allocation)
  3) Find all Nash equilibria of $E_m$ (solve inequalities)
  4) If $m < M$, $m \leftarrow m + 1$, goto step 2)
- **End**

In the worst case, $M = K! + (N-K)(K-1)!$ ($N > K$) or $M = N!$ ($N \leq K$), the above procedure can be exponential. Our future work will focus on improving this algorithm.

### 3.3 Characteristics of Equilibrium Classes

In GSP keyword auctions, any equilibrium class is defined by a series of linear inequalities. From the viewpoint of geometry, it determines a convex polyhedron in $R^N$. This result is summarized in the following proposition.

**Proposition 1.** $\forall E_m \in E/ \equiv$, $E_m$ is a convex polyhedron.

Because a non-empty convex polyhedron may be a point, a line segment, or a rectangle, etc, in a GSP auction, an infinite number of inefficient Nash equilibria might exist. Next, we consider two distinct polyhedra $E_m$ and $E_n$, where $m \neq n$. We make the following observation.

**Proposition 2.** $\forall E_m, E_n \in E/ \equiv$, $m \neq n$, their intersection is the null set, that is $E_m \cap E_n = \phi$.

By combining proposition 1 and 2, the separating hyper-plane theorem guarantees that any two distinct polyhedra can be separated by a hyper-plane (Boyd et al. 2004). Note that, the polyhedron determined by inequalities (3) is closed, but the allocation rule as governed by a series of strict inequalities (see Theorem
1), may define an open polyhedron. Thus, the intersection between these sets may be a half open polyhedron such that some boundaries are open whereas others are closed. Proposition 2 guarantees that two distinct equivalence classes do not share any point but they may share a common open boundary. If this is possible, a Nash equilibrium could change from one equivalence class to another through infinitesimal bid modification. We show that this is impossible (proof omitted).

**Proposition 3.** \( \forall \mathbf{E}_m, \mathbf{E}_n \in E/ \cong, m \neq n, \) the intersection of the closures of these two polyhedra is null set, that is \( \mathbf{E}_m \cap \mathbf{E}_n = \emptyset, \) where \( \mathbf{E}_m \) is the closure of set \( \mathbf{E}_m. \)

**4. Illustrating Examples**

We provide two examples on the distribution of equilibria in GSP auctions and discuss how advertisers and auctioneers can make use of this distribution structure.

**Case: \( N = 2, K = 2. \)**

Here, there are at most two distinct allocations: \( \pi^0 \) and \( \pi^1, \) where \( \pi^0 \) is the identity mapping: \( \pi^0(1) = 1, \pi^0(2) = 2, \) and \( \pi^1(1) = 2, \pi^1(2) = 1. \) Define \( s = 1 - 1/\gamma_1. \) If \( sv_1 > v_2, \) then the distribution of the two polyhedra is shown as in Figure 1(a). Otherwise, the distribution is illustrated as in Figure 1(b). In both figures, the shadowed regions \( E_0 \) and \( E_1 \) are equilibrium polyhedra that correspond to allocation \( \pi^0 \) and \( \pi^1, \) respectively. In Figure 1(a) truth-telling \( (v_1, v_2), \) point \( D, \) is a Nash equilibrium but not a dominant one. However, in Figure 1(b), truth-telling \( (v_1, v_2), \) point \( D', \) is not a Nash equilibrium. This illustrates the fact that truth-telling may not always be a Nash equilibrium as proposed in (Edelman et al. 2007).

In \( E_0, \) the bid of bidder 1 can be arbitrarily large, whereas in \( E_1, \) the bid of bidder 2 has the same characteristic (indicated by the arrows). In practice, no bidder will bid a very large value. Although doing so may guarantee a higher slot, the risk of his opponent raising their bids is high as well.

In \( E_0, \) the profits of bidder 1 and 2 are \( c_1^1(v_1 - b_2) \geq c_1^1(v_1 - sv_1) = c_1^2v_1 \) and \( c_2^1v_2, \) respectively. In \( E_1, \) bidder 1 has a profit of \( c_1^2v_1 \) and bidder 2 has a profit of \( c_2^2(v_2 - b_1) \geq c_2^2(v_2 - sv_2) = c_2^2v_2. \) So, bidder 1 prefers \( E_0 \) but bidder 2 prefers \( E_1. \) As such, bidder 1 will bid more than \( sv_2 \) and bidder 2 will bid more than \( sv_1. \) This resulting bid vector is no longer a Nash equilibrium. After a period of bids adjustments, the final outcome will be in either \( E_0 \) or \( E_1. \)

As to the auctioneer, his revenue varies from 0 to \( sv_1 \) and 0 to \( sv_2 \) in \( E_0 \) and \( E_1, \) respectively.

**Case: \( N > 2. \)**

In case \( N = 3, K = 3, \) the number of polyhedra is at most 6, see Figure 2(a). In case \( N \geq 3, K = 2, \) the number of polyhedra is at most \( N \) (Theorem 3). The possible allocations are: \( \pi^0: \pi^0(1) = 1, \pi^0(2) = 2; \pi^1: \pi^1(1) = 2, \pi^1(2) = 1; \pi': \pi'(1) = 1, \pi'(2) = i + 1, i = 2, 4, \ldots, N-1. \) Figure 2(b) illustrates an
example where $N = 3$, $K = 2$. In both figures, arrows attached to each polyhedron indicate that the corresponding bidder can announce an arbitrarily large bid in the sense of Nash.

![Equilibrium classes/polyhedra of a GSP auction with three bidders.](image)

**Figure 2.** Equilibrium classes/polyhedra of a GSP auction with three bidders.

5. Conclusions and Future Researches

In this paper, we characterized all pure-strategy Nash equilibria of a GSP auction. By using the same-slot equivalence relation, the nonlinear allocation rule implicated in the definition of Nash equilibrium can be eliminated from consideration. Well-developed inequality methods can then be employed to identify all Nash equilibria. Generally, each equivalence class of Nash equilibria is a convex polyhedron and all of these polyhedra are distributed in the pure-strategy space separately. In our ongoing work, we are focused on improving the current equilibrium-finding algorithm in terms of efficiency. We are analyzing bidding strategies in repeated GSP auctions and we aim to characterize the corresponding equilibria. On an empirical front, we are collecting data and plan to validate the predictive value of our work.

Acknowledgements

The authors wish to acknowledge research support from the NNSFC (60621001, 60875049, 70890084), MOST (2006AA010106), and CAS (2F07C01, 2F08N03).

References


The Impact of Ranking Mechanism on Quality Score and Revenue in Keyword Auctions

Jing Hao, Syam Menon, Srinivasan Raghunathan, Sumit Sarkar
School of Management
The University of Texas at Dallas
jhao@utdallas.edu, syam@utdallas.edu, sraghu@utdallas.edu, sumit@utdallas.edu

Abstract

We examine two of the most widely used ranking mechanisms in sponsored search. The first, where bidders are ranked based on their bids (rank-by-bid, or RBB) was used by Yahoo! until recently. The other mechanism ranks bidders based on the product of their bids and quality scores of their advertisements (rank-by-revenue, or RBR), and was initiated by Google. Existing literature has implicitly assumed the quality score to be exogenous. Advertisers can usually improve the quality scores of their advertisement at some cost, and the quality score should ideally be endogenous to the decision being made. We identify conditions under which one ranking mechanism is more profitable to the search engine than the other when an advertiser optimizes the quality score of his advertisement and when it does not.

Keywords: Sponsored search, keyword auctions, ranking mechanism, Symmetric Nash Equilibrium

1. Introduction

Sponsored search is a form of online advertisement where advertisers (bidders) bid for certain keywords and the search engine (auctioneer) allocates available slots to the advertisers following a pre-determined ranking mechanism. Since the launch of this business model in the late 90’s, different search engines have applied different ranking mechanisms for slot allocation. For example, Yahoo! used to rank advertisements (ads) by the Cost-Per-Click (CPC) bid submitted by advertisers, where the ad with a higher bid was displayed at a higher position (Rank-By-Bid, or RBB). Google ranks ads by the product of the CPC bid and a “quality score” (called ad rank), where ads with higher ranks appear in higher positions (Rank-By-Revenue, or RBR). In February 2007, Yahoo! switched to its own version of RBR rankings. Although several studies have analyzed the performances of both the RBB and the RBR mechanisms under different settings (e.g., Feng et al. 2007, Lahaie 2006), they have implicitly assumed that the quality score used in the RBR ranking is exogenous to the advertiser - implying that the quality of ads are fixed. However, in reality, search engines encourage advertisers to provide high quality ads, with higher quality ads being rewarded with a higher position for the same bid amount. Higher quality ads enable advertisers to place relatively lower bids, which in turn increase their ROI. Google determines the quality score for an ad rank using several factors, including the historical click-through rate (CTR) of the ad and the matched keyword, the relevance of the keyword and the ad to the search query, the account history of all the ads and keywords in an advertiser’s account as measured by the CTR, and other relevant factors. Note that the quality score is different from the quality of the products or services of the advertisers.

In order to improve the quality score (referred to as "quality" where appropriate), advertisers can invest some effort to better target each ad in their ad campaign. This could be through the use of auctioneer-provided resources, or by hiring professional help from a host of firms that have emerged to cater to such needs. By incurring these additional costs, an advertiser can affect the quality scores assigned to his ads and thereby impact their rankings. We examine how the two ranking mechanisms – RBB and RBR – affect the quality choices and bidding behaviors of advertisers, and impact the revenues to the auctioneer as a result. We use a stylized model that has one slot and two advertisers. We allow the two advertisers to have the same valuations (Scenario 1) or different valuations (Scenario 2) for a potential click. Under each scenario, we first analyze outcomes where the quality score is exogenous, (i.e., advertisers cannot
improve the quality of their ads even by incurring a cost), and then revisit the analyses where advertisers optimally choose their quality by considering the improvement cost. Our research shows the following. Under Scenario 1, the RBB mechanism performs at least as well as the RBR mechanism when quality is exogenous. When advertisers choose their quality, one or both of the advertisers may have an incentive to improve the quality of their ads under RBR; further, RBR generates at least as much revenue as RBB, and does better for a wide range of cost and other parameter values. Under Scenario 2, we find that the revenue comparison depends on the correlation between valuation and quality when the advertisers cannot improve quality. If the advertisers have the ability to improve quality, either mechanism may be better depending on the relative magnitude of cost and other parameters. Our analysis demonstrates that RBR is superior to RBB only when RBR induces a sufficient level of quality competition to offset the less intense bid competition among advertisers.

Many researchers have studied various aspects of these mechanisms (Edelman et al. 2007, Feng et al. 2007). Ghose and Yang (2009) find that keyword attributes (which impact the quality of an advertisement) affect click-through rates and conversion rates. Kempe and Mahdian (2008) examine the externality effect in search ads where the value of an ad to the user is affected by the set of other ads on the page. Two papers closely related to our work are Varian (2007) and Lahaie (2006). Varian (2007) considers a simultaneous game with complete information and analyzes equilibria under RBB and RBR. He also introduces the concept of Symmetric Nash Equilibria. Lahaie (2006) analyzes RBR and RBB under both complete and incomplete information for first and second price payment rules. He shows via example that no general revenue ranking of RBB and RBR is possible, and surmises that multiple Pure Strategy Nash Equilibria may exist under the complete information setting in the second price payment rule. By allowing advertisers to incur a cost and set their quality, we extend the scope of the bidding game. As search engines provide feedback and incentives to improve ad quality, the quality-setting decision should be endogenous to the game being played, and this work represents the first step in that direction.

We provide an overview of our model in Section 2. The analyses for the equal and different valuation scenarios are discussed in Sections 3 and 4, respectively. Section 5 concludes the paper.

2. Model Description

We consider a search engine that auctions a single slot to two advertisers A and B. The advertisers make their CPC bids simultaneously. Advertiser i has a value of $V_i$ per click and bids $b_i$ for a slot; we assume $b_i \leq V_i$. We use $q_i$, the probability that advertiser i’s ad is clicked given it is displayed at the slot as the proxy for his quality score. The search engine uses the generalized second price auction to determine $p$, the price paid by the winning bidder if his ad is clicked. Under RBB, $p$ is equal to the bid of the next highest bidder, and under RBR, $p$ is equal to the minimum bid he needs to retain his position.

We focus on two scenarios: in Scenario 1, the bidders have the same valuations, i.e., $V_A=V_B=V_i$; in Scenario 2, we assume $V_A> V_B$ WLOG. For each scenario, we examine the cases when quality is exogenous, and when it is endogenous. We consider two levels of quality, High (H) and Low (L). Advertisers are initially at the Low level, and do not incur a cost to stay there. They can, however, incur a cost of $c$ and move to the High level. We assume that all exogenous parameters are common knowledge (this assumption is widely used in prior research, e.g., Edelman et al. (2007), and Varian (2007). We consider a planning horizon with $n$ expected auctions, and assume that the bid and ad quality do not change within this planning horizon.

3. Scenario 1: Equal Valuations

3.1 Exogenous Quality Score

Here, the advertisers cannot alter the quality score. The sequence of events is as follows.
- Stage 1: Bidders place their bids;
- Stage 2: A search query is entered by a user;
Stage 3: The search engine allocates the slot using its ranking mechanism (RBB or RBR), and the advertiser whose ad is clicked pays the search engine. RBB ranks advertisers by bids. Therefore, the higher bidder gets the slot irrespective of quality, and pays an amount equal to the bid of the other bidder. We identify the equilibrium bids using the Symmetric Nash Equilibrium (SNE) concept (V...
We follow a similar procedure as before to derive the equilibrium qualities. For the choice \( (q_A = q_B = q_H) \), either A or B will be assigned the slot with a probability \( 0.5 \). The winning bidder pays \( V \) to the search engine if his ad is clicked. If A is randomly assigned the slot, his payoff is \( u_A = n(V - V)q_A - c = -c \) and the payoff to B is \( u_B = -c \). Similarly, if B is randomly assigned to the slot, the payoffs are \( u_A = u_B = -c \). The complete payoff matrix for RBR is presented in Figure 2.

In the following we consider the case where cost \( c \) is either higher or lower than \( nV(q_H - q_L) \). When \( c > nV(q_H - q_L) \), Low quality is the dominant strategy for both advertisers. The revenue for the search engine is \( R_{RBR} = nVq_L \). When \( c < nV(q_H - q_L) \), there are two Pure Strategy Nash Equilibria: \( (q_A = q_H, q_B = q_L) \) and \( (q_A = q_L, q_B = q_H) \). The expected revenue under either equilibrium is \( R_{RBR} = nVq_L \). In addition, there is also a mixed strategy with probability \( \theta_A = 1 - \theta_B = 1 - c/[nV(q_H - q_L)] \) for advertiser to choose High quality. Under the mixed strategy, the revenue to the auctioneer is \( R_{RBR} = nVq_L + nV(1 - c/[nV(q_H - q_L)])^2(q_H - q_L) \) and the expected payoffs to the bidders are \( u_A = u_B = 0 \). Table 2 summarizes the above analyses. A comparison of RBB and RBR yields the following.

**Proposition 2:** When advertisers have equal click valuations and the quality score is endogenous, (i) \( R_{RBB} \geq R_{RBR} \); (ii) Neither advertiser is worse off under RBR relative to RBB.

**Table 2: Revenue Comparison When Quality is Endogenous under Equal Valuations**

<table>
<thead>
<tr>
<th>Nash Equil.</th>
<th>Payoff to A</th>
<th>Payoff to B</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>((q_L, q_H))</td>
<td>0</td>
<td>0</td>
<td>( R_{RBB} = nVq_L )</td>
</tr>
<tr>
<td>((q_L, q_L))</td>
<td>0</td>
<td>0</td>
<td>( R_{RBB} = nVq_L )</td>
</tr>
<tr>
<td>((q_H, q_L))</td>
<td>( nV(q_H - q_L) - c )</td>
<td>0</td>
<td>( R_{RBR} = nVq_L )</td>
</tr>
<tr>
<td>((q_H, q_H))</td>
<td>( nV(q_H - q_L) - c )</td>
<td>0</td>
<td>( R_{RBR} = nVq_L )</td>
</tr>
</tbody>
</table>

**4. Scenario II: Unequal Valuations**

Next, we analyze how RBB and RBR perform in an asymmetric valuation scenario. WLOG, we assume \( V_A > V_B \). Since the steps in the analyses are identical to those in Scenario I, we present only the results.

**4.1 Exogenous Quality Score**

Under RBB, since \( V_A > V_B \), A always wins the slot. B bids his valuation \( V_B \). Therefore, the revenue to the search engine is \( nV_{sd_A} \). The payoffs to advertisers A and B are, respectively, \( n(V_A - V_B)^2q_A \) and 0. Under RBR, either A or B can win the slot depending on whose ad has the higher ad rank. The advertiser who does not win the auction bids his true valuation. The revenue to the search engine and the payoffs to the advertisers under various conditions are shown in Table 3. We observe that, depending on the correlation between valuation and quality, either RBB or RBR can yield higher revenue. If the two are sufficiently negatively correlated such that \( V_{sd_A} \leq V_{sd_B} \), we have \( R_{RBR} = nV_{sd_A} > R_{RBB} = nV_{sd_A} \). On the other hand, if valuation and quality score are positively correlated, i.e., \( q_A > q_B \) while \( V_A > V_B \), then \( R_{RBR} = nV_{sd_B} < R_{RBB} = nV_{sd_B} \). These results are consistent with the simulation results of Feng et al. (2007).

**Proposition 3:** When advertisers have unequal valuations and quality is exogenous, (i) \( R_{RBB} \geq R_{RBR} \) if valuation and quality are sufficiently negatively correlated so that \( V_{sd_A} \leq V_{sd_B} \) and (ii) \( R_{RBB} < R_{RBR} \) if value and quality are positively correlated.
4.2 Endogenous Quality Score

Under RBB, A wins the slot because $V_A > V_B$ and B bids $V_B$ in the Symmetric Nash Equilibrium. We obtain the payoff matrix in Figure 3 for the quality-setting game between advertisers. It shows that B always chooses Low quality. If $c < n(V_A - V_B)(q_H - q_L)$, $(q_A = q_H, q_B = q_L)$ is the Nash Equilibrium and the revenue for the search engine is $R_{RBB} = nV_Bq_L$. Otherwise, $(q_A = q_H, q_B = q_L)$ is the equilibrium and the revenue is $R_{RBB} = nV_Bq_L$. Our analysis implies that A has an incentive to improve the quality of his ad as long as $c$ is small, even though quality is not explicitly considered by the search engine for ranking. However, a smaller difference between $V_A$ and $V_B$ makes $c > n(V_A - V_B)(q_H - q_L)$ leading to a Low quality choice for both advertisers.

Under RBR, A is guaranteed to win except when $(q_A = q_L, q_B = q_H)$. When $(q_A = q_L, q_B = q_H)$, B can win the slot if $V_A \leq V_Bq_Hq_L$, and A wins otherwise. The matrix for the game is in Figure 4. When $V_A > V_Bq_Hq_L$, B always chooses Low quality. If $c < nV_A(q_H - q_L)$, $(q_A = q_H, q_B = q_L)$ is the Nash Equilibrium and A is better off by providing a High quality ad. If $c > nV_A(q_H - q_L)$, $(q_A = q_H, q_B = q_L)$ is the Nash Equilibrium. When $V_A \leq V_Bq_Hq_L$, either (or both) may choose High quality in equilibrium. We have the same Nash Equilibrium as in the case of $V_A > V_Bq_Hq_L$ for the same cost threshold of $nV_A(q_H - q_L)$, except for a possible sub-region within the region of $n(V_A - V_B)q_H < c < n(V_Bq_H - V_Bq_L)$, there is an additional Nash Equilibrium $(q_A = q_L, q_B = q_H)$ and a mixed strategy Nash Equilibrium. Under both of these new equilibria, $R_{RBB}$ is strictly better than $R_{RBB}$. The revenue to the search engine and the advertiser payoffs are summarized in Table 4.

**Proposition 4:** When advertisers have unequal valuations and the quality is endogenous, (i) $R_{RBR} > R_{RBB}$ if the difference in the valuation of a click of the advertisers is sufficiently small, i.e., $V_A < \min\{V_B + c/(nq_H), V_Bq_Hq_L - c/(nq_L)\}$, and (ii) $R_{RBB} \leq R_{RBB}$ otherwise.

### Table 3: Revenue Comparison when Quality is Exogenous under Unequal Valuations

<table>
<thead>
<tr>
<th>Rank order</th>
<th>RBB Ranking</th>
<th>RBR Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A \rightarrow B$</td>
<td>$V_Bq_L &gt; V_Bq_H$</td>
<td>$V_Bq_L &lt; V_Bq_H$</td>
</tr>
<tr>
<td>Payoff to</td>
<td>$A \rightarrow B$</td>
<td>$B \rightarrow A$</td>
</tr>
<tr>
<td>A</td>
<td>$n(V_A - V_B)(q_H - q_L)$</td>
<td>$n(V_Aq_H - V_Aq_L)$</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Revenue</td>
<td>$R_{RBB} = nV_Bq_L$</td>
<td>$R_{RBB} = nV_Bq_L$</td>
</tr>
</tbody>
</table>

### Table 4: Revenue Comparison when Quality is Endogenous under Unequal Valuations

<table>
<thead>
<tr>
<th>RBB ranking</th>
<th>RBR ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c &lt; n(V_A - V_B)q_Hq_L$</td>
<td>$c &gt; n(V_Aq_H - V_Aq_L)$</td>
</tr>
<tr>
<td>Nash Equilibrium</td>
<td>$(q_H, q_L)$</td>
</tr>
<tr>
<td>Payoff to</td>
<td>$n(V_A - V_B)q_Hq_L$</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>Revenue</td>
<td>$R_{RBB} = nV_Bq_L$</td>
</tr>
</tbody>
</table>

*: Mixed Strategy Nash Equilibrium ; †: $R_{RBR} = nq_B(nBq_H^2 - nBq_H) + nBq_H^2 + nBq_Hq_L + nBq_H$
Figure 5 shows the regions where one mechanism outperforms the other. We find that in region 1 (where $V_A > V_B + c/(n(q_H - q_L))$), $R_{RBB} = nV_Bq_H > R_{RBR} = nV_Bq_L$. In region 2 (where $V_A < V_B + c/(nq_H)$ and $V_A < V_Bq_H/c(nq_L)$, $R_{RBR} \geq R_{RBB}$. In all other regions $R_{RBB} = R_{RBR}$. We summarize the findings below:

### 5. Conclusions

Our results can be explained through the quality and bid competition between the advertisers induced by the two mechanisms. The mechanism that generates greater competition between advertisers yields more revenue for the search engine. When the quality score is exogenous, the advertisers engage in bid competition only, and the competition is more intense under RBB than under RBR unless the low valuation advertiser has sufficiently higher quality relative to the high valuation one. When quality is endogenous, the advertisers engage in both quality competition and bid competition. The intensity of quality competition is mitigated when the cost to improve quality is either too low or too high, or when the difference in the valuations of the advertisers is large. In such cases, the bid competition determines whether RBR or RBB yields higher revenue for the search engine (just as when quality score is exogenous). If the cost to improve quality is moderate and the advertisers have similar valuations, then the intensity of quality competition is higher under RBR than under RBB, which results in a higher revenue for the search engine under RBR than under RBB.

We have also investigated a multiple slot setting with equal valuations. Preliminary findings show that under exogenous quality, either mechanism can generate greater revenue depending on parameter values, while under endogenous quality, RBR is guaranteed to generate more revenue if there are two or more high quality advertisers. Future work will examine scenarios where different advertisers face different costs to improve quality. Our research questions can also be examined under an incomplete information setting.

### References


ABSTRACT

Online peer to peer (P2P) lending has received great coverage in media but little attention from academic researchers. In this study, we focus on risk and return of investments on prosper (P2P lending website). We find that on average, loans through prosper provide negative return. We then use decision tree analysis to segment loans in terms of different return and risk profiles. We further determine the efficient frontier for investments on prosper and calculate the efficiency of loans in various segments. We found that (1) within each credit grade, there exist subgroups which give positive return and for these subgroups risk is aligned with return and (2) group of loans with lower credit grade are more efficient in terms of risk and return than higher credit grade. Our study provides investment guidelines for lenders and design implications for online peer to peer lending websites.

Keywords: Peer to peer lending, risk and return, decision tree, efficient frontier

1. Introduction

Online peer-to-peer (P2P) lending is a new e-commerce phenomenon where individual lenders provide unsecured loans directly to individual borrowers without the traditional intermediaries such as banks. Since 2006, many internet-based lending services have emerged and online consumer lending has attracted millions of investors and have become increasingly popular alternative channel for consumers to get personal loans. They demonstrate how web 2.0 can facilitate creation of new markets competing with traditional providers such as credit card issuers leading to disintermediation of the financial services industry. According to a Wall Street Journal article (Kim 2007), the amount of new P2P loans issued in these lending websites was around $100 million in 2007 and will increase to as much as $1 billion by 2010.

As P2P lending bypasses the intermediaries and thus the associated cost, it can provide investors an opportunity to earn return that may be higher than that available from traditional investments. It also provides loans to borrowers who are not able to get loan from traditional financial markets or who get loans at higher rate in traditional markets. However, as the individual lenders are short of experience and monitoring capabilities compared to banks, these benefits are accompanied with high uncertainty and risk. Also, since most borrowers in these P2P lending sites are one-time borrowers, there is no reputation mechanism that can be built up over time to convey borrowers’ credibility like the one used for sellers on eBay. It is therefore essential for lenders to draw similarities on loans across borrowers based on available information to judge on expected risk and return of a new loan request.

In this research we focus on risk and return associated with loans. In particular, how can we segment loans in terms of different return and risk profiles based on available loan and borrower characteristics? How is the return aligned with the associated risk? What are the factors that influence the efficiency of different loan segments in terms of return and associated risk? Answers to these questions will contribute to the understanding of this new and rapidly growing lending market. Our findings can direct lenders towards ways to optimize their investment strategy and help these online lending service providers improve consumer welfare.

As a relatively new phenomenon, online P2P lending has received great coverage in media, but little attention from academic researchers. Two recent studies (Freedman and Jin 2008; Lin, Prabhala and
Viswanathan (2008) studied the impact of social connections on the likelihood of loan requests being successfully funded, the resulted interest rates and loan default rates. Herzenstein et al. (2008) further compared the effects of demographic attributes on the likelihood of funding success for online lending to those documented for the practices of traditional financial institutions. While these three studies examined this market from the borrower’s point of view, we set out our analysis from the lender’s point of view and focus on risk and return alignment in P2P lending.

2. Background and Data

In this paper we did our analysis on the Prosper.com (online peer to peer lending website) market place. Prosper was opened to the public on February 13, 2006 and as of August 1, 2008, Prosper has registered 750,000 members and originated 26,273 loans that total over 164 million US dollars. All prosper loan are fixed rate, unsecured, three year, and fully amortized with simple interest. Borrowers create listing on prosper and specify the amount of loan requested (from $1000 to $25,000) and maximum interest rate they are willing to pay. Lenders also have access to information related to borrowers’ credit history provided by prosper and other additional voluntary information (pictures, purpose of the loan etc.) provided by borrowers. The auction process is similar to proxy bidding on the eBay. Lender can bid on any listing by specifying the lowest interest rate he is willing to accept and the amount he wants to contribute (any amount above $50). A listing gets fully funded if the total amount of bids exceeds the amount requested by borrower. Minimum interest rate specified by the first lender excluded from funding the loan becomes the contractual rate for the loan.

Our study utilizes the publicly available data downloaded on August 1, 2008. This data set includes listings that began on or after June 1, 2006 and ends on or before July 2008. We focus on all the loans that were listed from February 2007 to July 2007, which leaves the first few months of the market and ensures that we have at least two year of repayment data for any loan. The final dataset has 4504 loans.

3. Methodology

3.1 Return on Investment (ROI) calculations

To compare the performance of different loans, we first calculate return on investment (ROI) based on monthly discounted payments over 2 year period for each loan in our dataset. Since we do not have data on actual payments and can only observe principal balances, we use the following information from prosper and some assumptions for the payment calculations:

1. As such, interest accrues on each loan's principal balance on a daily basis.
2. For simplicity, we assume that all payments for loans with status being “current” are paid at the scheduled day.
3. If there is no change in principal balance in a particular period (the loan status will change from “current” to “late”), we assume no payment has been made in that period.
4. On prosper any time borrowers make a loan payment, funds are applied to their loan balances in this order: (1) late fees (2) accrued interest (3) loan principal balance. If change in principal balance is positive (the status changes from “n-month late” to “current”), we assume late fee and interest accrued are paid in full. Late fee charged by prosper is calculated as the maximum of $15 or 5% of the outstanding unpaid payments and is passed on to lenders.
5. The service fee is charged to the lenders based on the current outstanding principal balance and its rate varies from 0.5% to 1% depending on the lenders’ credit grades. This policy has been changing over time. For simplicity we use 1% as the service fee rate for all the loans. Service fee is charged only in those periods when lenders receive payments.

The return on investment is given as:

$$ROI = \frac{CDP - L}{L}$$

where $L$ is the loan amount and $CDP$ is the cumulative discounted payment. The CDP is calculated as follows:
\[
C_{DP} = \frac{P_1}{(1 + \frac{d}{12})^1} + \frac{P_2}{(1 + \frac{d}{12})^2} + \cdots + \frac{P_{24}}{(1 + \frac{d}{12})^n} + F \times P_Bn
\]

\[
P_n = \Delta P_Bn + PB_{n-1} \times \frac{l_n}{365} \times r + L_n - S_n
\]

where \( P_n \) is the net payment lenders receive at the end of \( n^{th} \) month, \( PB_n \) is the remaining principal balance at the end of \( n^{th} \) month and \( d \) is the discount factor. If a loan does not get defaulted or paid earlier, then \( n \) is 24; otherwise, \( n \) is the month in which the loan gets defaulted or paid in full. \( r \) is the interest rate for the loan, \( l_n \) is the length of time period in days (e.g. \( l \) is 31 for January), \( L_n \) is the late payment and \( S_n \) is the service fee charged in period \( n \).

### 3.2 Decision Trees to Group Loans Based on ROI

Our calculations in previous section suggest that prosper loans vary significantly in their return. While this may have attracted attention from both academic researchers and practitioners, what has been missing is whether the significantly varied return is aligned with the associated risk. To study this, we need to group loans with similar characteristics together (to create the distribution of ROI) and calculate both the return (the mean ROI) and risk (the variance of ROI) within each group to see if they are aligned when compared across groups. In this study, we first rely on the decision tree methodology to divide loans into groups such that loans within the same group have similar characteristics and distribution of ROI. We then calculate risk and return for each group and examine their positions in the composition of optimal investment portfolios.

When applying the decision tree methodology to divide loans into different groups, we use variables related to borrower and loan characteristics which lenders can observe before making their investment decisions as explanatory variables and use ROI as the dependent variable. We use F-test as the splitting criteria when dividing groups such that loans that are not significantly different in ROI will be pooled in the same group. We partition the data set into training and validation data sets. The training data set is used to generate the rules to segment the data set. The validation data set is then used to compare the mean predicted value and mean target value of ROI to assess the performance of these rules. Figure 1 depicts the comparison of target and mean value of ROI for the validation data set.

![Validation Data Set](image-url)

**Figure 1.** Comparison of target and predicted value of ROI for validation data set

---

1 For brevity the list of variables used and their definition are not provided. This list is available upon request from the authors.
3.3 Efficient Frontier

The groups identified using a decision tree have different distributions of ROI and represent different investment opportunities. Each group will give a different return but will also have a different risk associated with it. The mean of ROI within each group will give the expected return from the investment in that group and the variance of ROI will give the associated risk. For an efficient market, the risk and return should be aligned i.e. the loans with high risk should give high return.

Lenders can combine loans from different groups as portfolios to achieve optimal return based on the amount of risk they are willing to take. To assess the efficiency of each group, we first calculate the efficient set which gives maximum return for a given risk. The expected return and variance for an investment portfolio containing loans from different groups are given as:

\[ E(ROI_p) = w_1 E(ROI_1) + w_2 E(ROI_2) + \cdots + w_n E(ROI_n) + \cdots w_{28} E(ROI_{28}) \]

\[ V(ROI_p) = (w_1)^2 V(ROI_1) + (w_2)^2 V(ROI_2) + \cdots + (w_n)^2 V(ROI_n) + \cdots (w_{28})^2 V(ROI_{28}) \]

where \( E(ROI_n) \) and \( V(ROI_n) \) are the return (mean value of the ROI) and risk (variance of ROI) obtained by investing \( w_n \% \) fraction of total investment in \( n^{th} \) group. \( E(ROI_n) \) and \( V(ROI_n) \) are the mean and variance of ROI for \( n^{th} \) group. Since the groups are disjoint (in terms of decision rules), we assume that the covariance between ROI of different groups is zero. To get the maximum return for a given level of risk, we solve the following optimization problem:

\[
\max_{w_1, w_2, \ldots, w_{28}} E(ROI_p) \\
\text{subject to} \\
V(ROI_p) \leq K \quad (1) \\
\sum_{n=1}^{28} w_n \leq 1 \quad (2)
\]

The first constraint ensures that the risk of the investment portfolio is less than the given level of risk (\( K \)) and the second constraint is a budget constraint (it ensures that the sum of the proportion of total money invested in different groups is not over 1). After solving this optimization problem we are able to calculate the maximum return for a given risk. Hence for each group we calculate the efficiency measure as follows:

\[ \text{Eff}_n = \frac{E(ROI_n)}{E^*(ROI_p)} \]

where \( E(ROI_n) \) is the return (mean value of the ROI) for the \( n^{th} \) group and \( E^*(ROI_p) \) is the maximum return for the level of risk (variance of ROI) associated with \( n^{th} \) group.

4. Results and Conclusions

We find that on average, loans through Prosper provide negative return. The result is different from previous studies (Freedman and Jin 2008) and Prosper website which found positive estimated rate of return on Prosper. In our study we use only those loans which have at least 24 months of loan history to calculate return whereas other studies have used loans of all ages including loans which have only few months of history. Since on Prosper a loan becomes defaulted only if it is more than 4 months late, including the loans with only few months of history to calculate the overall return will give biased results.

Credit grade is the most used variable to analyze the loans performance on Prosper both by lenders and by Prosper. For example, initially borrower with any credit score can list loans on Prosper. However from February 2007, borrowers with credit score less than 520 were not allowed and from October 2008 onwards
borrowers with credit score less than 560 (credit grade HR) are not allowed to list loans on Prosper. However, we found that the average return for loans within each credit grade (including credit grade AA (credit score greater than 760) is also negative. We then use the decision tree analysis to determine groups within each credit grade which were different from each other in term of ROI. Except for credit grade HR (which has now been discontinued by prosper), within each credit grade, there exist subgroups which give positive return. Table 1 provides the borrower and loan characteristics for which a loan within each credit grade will provide positive results. The observed mean and variance of ROI provides the risk and return associated with these segments. Hence, lenders on prosper should make their decisions based not only on credit grades but should also take other variables in to account.

We further found that for groups which give positive return; risk is aligned with return i.e. as return increases risk also increases. We then derive the efficient frontier for investments on prosper and determine the optimal portfolio for investments for a given level of risk. The maximum possible return for given risk is summarized in Table 1. Table 2 summarizes the optimal % investment in each subgroup to maximize return at various levels of risk. The key finding of this analysis is that sub group of loans with lower credit score are more efficient in terms of risk and return i.e. they give more return for a given risk as compared to loans with higher credit score. We also found that borrowers, who are not willing to take high risk, should diversify their portfolio (i.e. invest in subgroups of loans of all credit grades) to maximize return, whereas borrowers who are willing to take more risk can maximize their return by investing in only subgroups of loans with lower credit grade. Thus prosper market will be more efficient for lenders who are willing to take more risk.

<table>
<thead>
<tr>
<th>Credit Grade (Credit Score)</th>
<th>Mean ROI</th>
<th>Total Loans</th>
<th>Total Loans</th>
<th>Mean ROI (Return)</th>
<th>Standard Deviation ROI (RISK)</th>
<th>Total Min ROI</th>
<th>Max ROI</th>
<th>Splitting Criterion</th>
<th>Max ROI for given RISK</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA (760+)</td>
<td>-0.01</td>
<td>532</td>
<td>341</td>
<td>0.02</td>
<td>0.10</td>
<td>-0.88</td>
<td>0.34</td>
<td>Inq6m (Inquiries in last 6 Months) &lt;= 1</td>
<td>0.059</td>
<td>0.377</td>
</tr>
<tr>
<td>A (720-759)</td>
<td>-0.05</td>
<td>475</td>
<td>178</td>
<td>0.03</td>
<td>0.11</td>
<td>-0.55</td>
<td>0.14</td>
<td>BMR (Borrower Maximum Rate) &lt; 0.155 and RCB (Revolving Credit Balance) between 1251 and 16886</td>
<td>0.062</td>
<td>0.485</td>
</tr>
<tr>
<td>B (680-720)</td>
<td>-0.09</td>
<td>594</td>
<td>39</td>
<td>0.03</td>
<td>0.11</td>
<td>-0.55</td>
<td>0.12</td>
<td>Listing open for duration and AB (Amount Borrowed) &lt; 2500</td>
<td>0.062</td>
<td>0.485</td>
</tr>
<tr>
<td>C (640-679)</td>
<td>-0.12</td>
<td>857</td>
<td>94</td>
<td>0.05</td>
<td>0.21</td>
<td>-0.72</td>
<td>0.28</td>
<td>Owns House and RCB between 2007 and 5579.5</td>
<td>0.0835</td>
<td>0.647</td>
</tr>
<tr>
<td>D (600-639)</td>
<td>-0.12</td>
<td>924</td>
<td>94</td>
<td>0.06</td>
<td>0.19</td>
<td>-0.76</td>
<td>0.28</td>
<td>Inq6m &lt;= 1 and BMR &lt; 0.195</td>
<td>0.081</td>
<td>0.737</td>
</tr>
<tr>
<td>E (560-599)</td>
<td>-0.16</td>
<td>483</td>
<td>56</td>
<td>0.09</td>
<td>0.28</td>
<td>-0.96</td>
<td>.37</td>
<td>AB &lt; 3000 and BCU (Bank Card Utilization) between 0.465 and 0.955</td>
<td>0.092</td>
<td>1</td>
</tr>
<tr>
<td>HR (520-559)</td>
<td>-0.25</td>
<td>639</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>-0.12</td>
<td>4504</td>
<td>772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. Optimal investment for given risk

<table>
<thead>
<tr>
<th>Return</th>
<th>Risk</th>
<th>Group1 (AA credit grade subgroup)</th>
<th>Group2 (A credit grade subgroup)</th>
<th>Group3 (B credit grade subgroup)</th>
<th>Group4 (D credit grade subgroup)</th>
<th>Group5 (C credit grade subgroup)</th>
<th>Group6 (E credit grade subgroup)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0476</td>
<td>0.0050</td>
<td>0.1180</td>
<td>0.2198</td>
<td>0.1995</td>
<td>0.1807</td>
<td>0.1328</td>
<td>0.1490</td>
</tr>
<tr>
<td>0.0590</td>
<td>0.0100</td>
<td>0.0000</td>
<td>0.1531</td>
<td>0.1304</td>
<td>0.2711</td>
<td>0.1896</td>
<td>0.2557</td>
</tr>
<tr>
<td>0.0660</td>
<td>0.0150</td>
<td>0.0000</td>
<td>0.0693</td>
<td>0.0465</td>
<td>0.3290</td>
<td>0.2236</td>
<td>0.3315</td>
</tr>
<tr>
<td>0.0713</td>
<td>0.0200</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.3632</td>
<td>0.2367</td>
<td>0.4001</td>
</tr>
<tr>
<td>0.0750</td>
<td>0.0250</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.3276</td>
<td>0.1712</td>
<td>0.5012</td>
</tr>
<tr>
<td>0.0777</td>
<td>0.0300</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.3013</td>
<td>0.1232</td>
<td>0.5755</td>
</tr>
<tr>
<td>0.0799</td>
<td>0.0350</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2796</td>
<td>0.0833</td>
<td>0.6371</td>
</tr>
<tr>
<td>0.0818</td>
<td>0.0400</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2606</td>
<td>0.0484</td>
<td>0.6910</td>
</tr>
<tr>
<td>0.0836</td>
<td>0.0450</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2435</td>
<td>0.0171</td>
<td>0.7394</td>
</tr>
<tr>
<td>0.0852</td>
<td>0.0500</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2143</td>
<td>0.0000</td>
<td>0.7857</td>
</tr>
<tr>
<td>0.0866</td>
<td>0.0550</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1697</td>
<td>0.0000</td>
<td>0.8303</td>
</tr>
<tr>
<td>0.0880</td>
<td>0.0600</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1286</td>
<td>0.0000</td>
<td>0.8714</td>
</tr>
<tr>
<td>0.0892</td>
<td>0.0650</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0905</td>
<td>0.0000</td>
<td>0.9095</td>
</tr>
<tr>
<td>0.0904</td>
<td>0.0700</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0547</td>
<td>0.0000</td>
<td>0.9453</td>
</tr>
<tr>
<td>0.0915</td>
<td>0.0750</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0208</td>
<td>0.0000</td>
<td>0.9792</td>
</tr>
<tr>
<td>0.0921</td>
<td>0.0782</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.0921</td>
<td>0.0782</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

### References