Retail Banking Choice of Distribution Strategy in a Competitive Market

Reynold E. Byers  
Graduate School of Management  
University of California, Irvine  
Irvine, CA 92697

Marshall Freimer  
and  
Phillip J. Lederer  
William E. Simon Graduate School of Business Administration  
University of Rochester  
Rochester, NY 14627

February 25, 2002

The service economy is the dominant productive sector of economically advanced nations. Despite this, competition between service-providing firms has been little studied because of the nonconvexity of the production function of service firms. Nonconvexity of production functions arises because of economies of scale inherent in location, inventory and queuing technologies. A major contribution of this paper is demonstrating existence and uniqueness of a competitive equilibrium in markets with economies of scale. These results are shown in the context of modeling the competitive choice of retail distribution strategy by banks. We model many important features of this problem, including: heterogeneous customers, customer choice, cost structures of different distribution channels, entry of banks into the market, and competitive pricing. Banks compete by choice of entry, prices, products, locations and distribution technologies. We show the existence and uniqueness of a competitive entry equilibrium with these choices.

The value of this approach is demonstrated by presenting numerical experiments of competition between retail banks. We show the nature of the competitive equilibrium, and show that for retail banking, the major factor driving evolution of strategic distribution systems is customer preferences, not cost advantages of new technologies. We show that virtual banks will remain unprofitable until the segment that prefers electronic transactions grows much larger than it currently is.
1. Introduction

The service sector has become the dominant sector of the economy of the United States and other advanced nations (Harker 1995). Within this sector, many service businesses are quite competitive. However, to date, competitive analysis of service systems has been difficult to model using traditional mathematical economic approaches because economies of scale exist for many service technologies. In microeconomic models of competition, economies of scale implies that monopoly provision will occur, or the existence of an equilibrium hard to prove.

In this paper, we present a model of competition between banks that choose a distribution strategy to serve customers. An important decision that banks make is the number and location of branches and ATM’s. The benefit of increasing the number of retail locations is decreased customer travel time and cost, and thus higher convenience. The cost of more facilities is higher bank cost. The benefit net of cost exhibits economies of scale effects.

Besides location models with endogenous choice of facilities, many other important service-oriented operations technologies exhibit economies of scale. Some examples are: inventory models with fixed order costs (e.g., EOQ model), inventory models with safety stock (e.g., models with risk pooling) and queuing models.

In this paper we show how to overcome these problems, and present numerical simulations to show the likely direction of competition in the banking industry.

1.1 Retail Banking

Over the past two decades the retail banking industry has experienced significant changes. Chief among the changes are deregulation (increasing competition both among banks and non-banks), technological advances in the distribution system possibilities (Bauer 1995) and associated changes to customer preferences.

This paper examines a retail bank’s choice of distribution strategy in light of the major competitive and consumer changes in the marketplace. The model used in the analysis applies a perfect competition approach. The objective is to show that an equilibrium exists, and then to characterize the equilibrium under a variety of parameters to judge what distribution technology strategies are successful.

The model captures a rich variety of influential market parameters for both banks and consumers. Retail banks choose among various technologies including branches, ATM’s and PC
banking systems to serve their customers. Banks are faced with a tradeoff between the cost of adding another branch or kiosk and the increase in customer convenience from the resulting reduction in average travel distance. This tradeoff creates non-convexities in the banks' cost function. Customers are of several types, each distinguished by preferences over technologies. Further, the model incorporates both fixed and variable costs of providing banking services.

The model is developed and demonstrates sufficient conditions to guarantee the existence of a competitive entry equilibrium. A surprising result is that not only does a competitive entry equilibrium exist despite the banks' nonconvex cost functions, but the equilibrium is unique. Although an equilibrium exists and is unique, a closed-form expression for the equilibrium does not exist; however, an efficient numerical procedure identifies the equilibrium for a given set of market parameters.

Numerical sensitivity analysis shows how demand and cost parameters affect the equilibrium outcome. The analysis demonstrates that the equilibrium mix of banks is most sensitive to the relative size of consumer segments, and is not as sensitive to the absolute size of the market. Further analysis shows that the equilibrium results are relatively insensitive to the fixed and variable costs of technology. The results generally suggest that changing consumer behavior, rather than bank cost structure, drives changes in retail bank competitive strategies at equilibrium.

While there are some papers with related modeling approaches, no other papers address the banking distribution strategy question. The closest paper (Balasubramanian 1998) models competition between retailers and a direct marketer. Our paper addresses several extensions proposed by Balasubramanian, including allowing the firms to own both the physical and direct channels, multiple customer segments and elastic demand. Further, this paper employs a careful model of consumer choice with attention to relevant parameters for the retail banking industry.

Prasad and Harker (2000) address the question of pricing PC banking services considering the existence of network externalities: however, they do not address the issue of strategic choice of distribution technology nor do they address equilibrium. Several articles in the industry literature discuss the nature of branches and locations in a marketing framework. Some articles discuss strategic branch location, focusing on qualitative aspects such as sharing facilities with supermarkets or warehouse stores (Aractingi 1994, Hawk 1990, Stahl 1995, Tandy and Stovel 1989). There is industry discussion of the tradeoffs and strategies involved in choosing a distribution strategy, but none applies an analytical approach (Bauer 1995, Salomon Bros. 1995). Finally, there are some mathematical models of locating bank branches in a market. Most of these models address the optimal location of branches based on
demographic information, but do not address the tradeoffs between alternative service delivery systems (Hopmans 1986, Eliopoulos and Kouzelis 1987, Chelst et al 1988).

The presentation is organized as follows: Section 2 presents the model of the market for retail banking services, including consumers and banks. Section 3 defines a competitive entry equilibrium, and proves the existence and uniqueness of such an equilibrium. Section 4 presents sensitivity analysis, and Section 5 presents the conclusions and ideas for future research.

2. Model of Retail Banking Market
The model considers a competitive geographical market for retail banking services with heterogeneous consumers and banks. Retail banks are profit maximizing and can enter and exit the market. Banks provide retail transactions through a choice of the number and type of different distribution channel technologies, e.g. branches and PC banking, and capacity decisions. Banks also choose pricing strategy, that along with distribution channel choice affects customers' demands. Consumer demand is sensitive to price, distribution technology, and average distance from a retail bank, and customers choose relationships with banks offering the greatest customer value. Customers are heterogeneous and are grouped into segments having homogeneous tastes, but not locations. The overall goal of this research is to define a competitive entry equilibrium, prove its existence and uniqueness, and characterize its properties.

We begin by describing the customers, the banks and the competitive entry equilibrium. We use the notation found in Table 1.

2.1 Customers
We assume a finite set of customers located in a geographical region. Customer's demand for retail bank transactions depends on the total cost of conducting a transaction. For ease, we assume that there are $J$ customer segments, each having homogeneous tastes. We assume that costs are imposed on customers in three ways: price, distance traveled and distribution channel technology preferences. In practice, retail banks charge prices in different ways, many of them indirect. Some examples of the ways that retail banks charge prices are fees for a specific service, minimum account balances, periodic fees and deals for packages of services. The model employed in this paper approximates the multiple pricing methods used by a per-transaction pricing approach.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Remarks</th>
<th>Meaning of Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>Indices to indicate technology</td>
<td>Set of distribution technologies</td>
</tr>
<tr>
<td>( k )</td>
<td>Generic index for technology</td>
<td>Some examples are: branches, ATM’s, PC (at home PC banking systems).</td>
</tr>
<tr>
<td>( J )</td>
<td>Indices used to indicate segment</td>
<td>Set of consumer segments</td>
</tr>
<tr>
<td>( j )</td>
<td>Generic index for segment</td>
<td></td>
</tr>
<tr>
<td>( I )</td>
<td></td>
<td>Set of firm strategies: what subset of distribution technologies a bank has (a subset of ( J x K )).</td>
</tr>
<tr>
<td>( i )</td>
<td>Generic index for strategy</td>
<td></td>
</tr>
</tbody>
</table>

**COST PARAMETERS:**

| \( h_{jk} \) | Disutility imposed on customer segment \( j \) by a transaction through technology \( k \) ($ per transaction) |
| \( \beta_j \) | Coefficient of segment \( j \)'s disutility due to distance ($/mile) |
| \( F_k \) | Yearly fixed cost of placing an outlet using technology \( k \) in the market region |
| \( G_i \) | Yearly fixed cost of maintaining a network of strategy \( i \) ($/year) |
| \( c_k \) | Coefficient of variable cost per transaction’s linear term ($/transaction) |
| \( d_{jk} \) | Coefficient of the variable cost per transaction’s quadratic term ($/transactions^2) |

**BANK’S DECISIONS:**

| \( n_i \) | Number of outlets of bank strategy \( i \) |
| \( p_j; \overline{p} \in \mathbb{R}^J \) | Price per transaction charged to segment \( j \) ($), vector of these prices for all segments ($) |
| \( P_{jk} \) | Full price includes price, travel costs and disutility due to technology type |
| \( p_j; \overline{p} \in \mathbb{R}^J \) | Full price per transaction to segment \( j \) using channel \( k \) ($) |
| \( m_i; \overline{m} \in \mathbb{R}^J \) | Number of firms using strategy \( i \). The vector of firms’ strategy choices. |
| \( \tau_{jk} \in \mathbb{R}^J \) | Number of transactions supplied by a bank using strategy \( i \), serving customer segment \( j \) with distribution technology \( k \). |
| \( \overline{\tau}_i \in \mathbb{R}^{JK} \) | The vector of transactions for strategy \( i \). |
| \( \overline{\tau}_{ij} \in \mathbb{R}^I \) | The vector of total production for segment \( j \) by strategy type. |
| \( \overline{\tau}_{ij} \in \mathbb{R}^K \) | The vector of production for segment \( j \) and strategy \( i \) by distribution channel. |

**DEMAND FUNCTION:**

| \( D_j(\overline{P}); \overline{\Delta}(\overline{P}) \) | Demand by segment \( j \) given full price vector \( \overline{P} \), vector of demand for each customer segment given price vector \( \overline{P} \). |

*Table 1: Notation Used in This Paper*
Consumer segments differ in their attitudes toward distribution channel technology. Some customers are wary of remote technology and prefer the secure feeling of dealing with a teller. Other customer segments are averse to dealing with tellers and prefer the efficiency of an ATM or (at-home) PC banking system. These preferences are modeled as an additional cost, due to disutility, imposed on the customer. We denote the set of distribution channel technologies by $K$, and denote the disutility imposed on customer segment $j$ by a transaction through distribution channel $k$ as $h_{jk}$.

Transactions are provided by competing banks. In the next subsection we describe the banks strategies which we denote by an index $i$. We assume customers are sensitive to a measure of distance related to their average distance to a bank outlet, specifically $\beta_i (1/n_i)$, where $n_i$ is the number of bank outlets placed in the market by bank using strategy $i$ and where $\beta_i$ is a cost parameter (in terms of $$/mile) for customer segment $j$.\(^1\) We employ an average distance measure because banks place multiple locations throughout a region, and consumers travel somewhat randomly within the region. It is reasonable to assume that customers do not measure their distance to a single retail bank outlet in evaluating their banking relationship. If customers travel and engage in banking transactions at random times, average distance to a branch or ATM is an appropriate convenience measure.

We assume that some transactions can only be conducted at a branch. It is not possible to obtain cash with PC banking (it could be with mail delivery, but this is relatively slow). Further, some functions, such as personalized investment advice, are available only at a branch or other staffed facility in order to provide greater service and a stronger bank brand. As a result, we assume that there are "required branch transactions" that must be conducted at a branch. For simplicity, the number of required branch transactions will be some fixed fraction $f$ of the total number of regular transactions. For example, if in strategy $i$, the total number of regular transactions for segment $j$ is $\tau$ then the number of required transactions by segment $j$ is just $ft$. Note that branch transactions also generate "required transactions"\(^2\). Customers incur travel costs and disutility for required transactions, as for elective transactions, but the model assumes that the bank cannot charge for required transactions due to the loss of choice of distribution channel.

---

\(^1\)Note that the functional form for average distance, $(1/n_i)$, holds exactly for location using the $\ell^1$ norm, but generalizations are possible for powers of $(1/n_i)^{1/2}$ for $k = 0, 1, 2$, etc. leading to general results.

\(^2\)Alternatively, it can be assumed that regular bank transactions conducted in a branch do not generate required transactions, but the notation becomes slightly more complicated.
Customer demand depends on the total cost of a banking relationship which we call its *full price*. The total cost to the consumer of a banking relationship is a function of price, disutility and convenience costs. Let $p_{jk}$ denote the price per transaction the retail bank using strategy $i$ charges customer segment $j$ using distribution channel $k$. Assume that a bank using strategy $i$ places $nb$ branches, $n_{ATM}$ ATM’s, and provides required transactions at $n_r$ locations in the market. Under strategy $i$, customer segment $j$ using channel $k$ incurs a *full price per regular transaction*:

$$P_j = \begin{cases} 
  p_{jk} + h_{jk} + \frac{\beta_j}{n_k} + f(h_{jr} + \frac{\beta_j}{n_r}) & \text{if } k \in \{b, ATM\} \\
  p_{jk} + h_{jk} + f(h_{jr} + \frac{\beta_j}{n_r}) & \text{if } k \notin \{b, ATM\}
\end{cases}$$

Note that the full price considers the convenience of the fractional required transaction that is generated by a regular transaction. Full price is the per unit total cost to the customer of engaging in a banking transaction. We will be interested in the firm and the distribution channel that minimizes full price for segment $j$, as this sets the *market prices* (prices at which transactions actually take place).

We assume that each customer segment will purchase banking transactions from the bank offering the lowest full price-channel combination. We allow banks to set different prices for each customer segment and each channel. Thus, we allow price discrimination across segments and distribution channels. This requires that banks know the identity of customers so that different prices can be set. We discuss how banks can do this in Section 3.3.

Finally, assume that the number of regular transactions demanded by each customer segment is given by a separable differentiable, strictly decreasing function of the least full prices offered it.

If all banks announce their segment-distribution channel prices, $p_{jk}$, and the type and number of their facilities, then the least full price channel can be found for each segment $j$, say, $P_j$. In this paper we will write $\vec{P}$ as the vector of these least segment full costs. We assume that the vector of demand for each segment is given by the continuous function $\vec{D}(\vec{P}) \in \mathbb{R}^J_+$ that is differentiable and monotone decreasing.

### 2.2 Banks

Banks are profit maximizers and choose prices and strategies to maximize profit. The defining characteristic of a retail bank in our model is its distribution “strategy.” A strategy is defined as a function that maps each customer segment in $J$ into a technology chosen from the set of available
distribution technologies $K$ that will be used to serve that segment. Consider the set of all ordered pairs of elements in $J$ and $K$: $I = \{ (j,k) \mid j \in J, k \in K \cup \emptyset \}$, where $\emptyset$ is the null set. A strategy, $i$, is a subset of the set of all pairs of segments-technologies $I$, $(i \subset I)$. For a strategy $i$, we allow each segment to be mapped into several distribution channels.

In defining set $I$, we allow a customer segment to be mapped into the null set, meaning that this segment is not served. By definition, a retail bank that offers a branch network with ATM’s attached serving segments $k$ and $k'$ employs a different strategy than one offering branches serving $k$ and ATM’s serving $k'$. We will slightly abuse our notation by writing $k \in i$ which means that strategy $i$ employs distribution channel technology $k$.

Each bank using strategy $i$ provides a transaction volume for each technology-customer pair, $(j,k)$. Denote the volume decision variable by $\tau_{j,k}$.

Banks can enter this market, so that the number and strategy choices of banks in the market are endogenously determined. We refer to $\vec{m}$ as the vector of “strategy types in the market.” Understand that $\vec{m}$ is a vector in $R^{I}$ and the $i$th entry of $\vec{m}$, $m_i$ is the number of banks using strategy $i$ in the market$^3$. Further, assume that all banks using strategy $i$ are identical$^4$. Each strategy implies a specific cost structure defined by the component technologies.

Each distribution channel technology involves fixed and variable costs and exhibits diseconomies of scale. In practice, the primary technologies available to a retail bank are branches, ATM’s, and PC banking, but the notation is general enough to include other channel technologies. Each distribution channel technology has a fixed cost $G_k$, a variable technology cost of serving $\tau$ regular and required transactions through technology $k$, $c_i \tau$, and a variable customer cost of serving $\tau_{j,k}$ regular transactions of type $j$ through technology $k$, $d_{j,k} \tau_{j,k}^2$. The linear variable cost reflects the purely operational cost of processing a transaction through technology $k$. The quadratic term represents the diseconomies associated with processing transactions of each type. This diseconomy is important to the model: Without diseconomies of scale there would be no competition in the market;

\textsuperscript{3}Note that this assumption allows for fractional banks; in practice, banks are inherently integral, but this assumption is a technical requirement for finding an equilibrium below; further, if there is a sufficient number of banks in the market, then the assumption of perfect competition is reasonable and a fractional number of firms will make little difference.

\textsuperscript{4}It is possible to generalize this assumption to allow cost asymmetries between banks but we do not do so for simplicity.
the only equilibrium result would be monopoly. The model assumes a quadratic function for simplicity and tractability.\footnote{We assume that required transactions do not add to the diseconomy of scale for technical reasons. The relatively simple results of Proposition 3.1 disappear when required transactions generate diseconomies.}

The form of the diseconomy cost also has implications on information and choice. We note that the cost of processing a transaction of "type" $(j,k)$ is different from the cost of processing a transaction of type $(j',k)$. Thus we are implicitly assuming that transactions by different segments using the same channel are distinguishable from each other. That is, a bank knows what segment demanded a transaction for a specific channel. We discuss this assumption further in section 3.3.

We adopt several assumptions to ease the location analysis. If a bank does not choose to offer branches in its strategy, it still needs some physical presence in the market to offer services associated with required transactions. These required transactions are served by a single central facility which we call a required transactions center. Thus there is always one branch available for required transactions. The fixed cost of this facility is included in fixed cost $G_j$. If a bank offers branches as part of its distribution strategy, then required transactions can be served by any branch. If a strategy adopts both branches and ATM's, these are chosen in the same number and the same locations\footnote{The model could be changed to allow the number of ATMs to be chosen endogenously This change would be accomplished by defining a variable $n_{ATM}$ that represents the number of ATMs to be placed in the market, with associated cost $F_{ATM}$ per ATM. The model does not incorporate this decision variable because combined with the assumption that there are "required transactions," the model becomes intractable. Without the required transactions, the existence and uniqueness proofs would still hold even with the additional decision variable of the number of ATMs. We are thus forced to choose between including endogenous determination of the number of ATMs or including required transactions.}. We leave the assumption that ATM's are attached to branches as an approximation in order to focus on the effects of technological limitation (or strategic choice) that required transactions represent. Further, assume that a retail bank's customers use only ATM's belonging to that bank. Allowing the customer to use other ATM's and the surcharges that are associated is beyond the scope of this paper. As an example: if only ATM's are adopted as part of strategy $i$, then only one required transactions processing center is located in the entire region along with $n_i$ ATM's; if branches and ATM's are adopted by strategy $i$, the number of branches and ATM's are the same. The variable cost of processing a required transaction is the same as the cost of processing a branch transaction and is denoted $c_a$. Further, customer disutility of a required transaction is the same as the disutility of a branch transaction.

Each bank employing branches or ATM's chooses a number of locations, denoted $n_i$, at a cost of $F_i$ dollars per outlet; where $F_i = 0$ if strategy $i$ does not incorporate physical outlets.
We assume that firms are full-price takers, i.e., any one bank is not of significant size to affect the market full-price, \( \bar{P} \). Given a market full prices and its strategy, each bank \( i \) maximizes profit by choosing a regular transaction volume for each customer segment and technology, \( \tau_{ijk} \), a number of branches, \( n_i \), and real prices to charge to each customer segment through each channel, \( P_{ijk} \), to equate a bank’s full price to the market full price. Profit for bank \( i \) is denoted \( \pi_i \) and is given below in (1):

\[
\pi_i = \sum_{j \in J} \sum_{k \in K} P_{ijk} \tau_{ijk} - \sum_{k \in K} c_k \sum_{j \in J} \tau_{ijk} - \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{ijk}^2 - c_b f \sum_{j \in J} \sum_{k \in K} \tau_{ijk} - n_i F_i - \sum_{k \in K} G_k
\]

As banks are full-price takers, it follows that given a market full price \( P_j \) the price a bank can charge for a transaction is given in (2)

\[
P_{ijk} = \begin{cases} 
  P_j - h_{jk} - fh_{jb} - (1 + f) \frac{\beta_j}{n_i} & \text{if } k \in \{b, ATM\} \text{ and } b \in i \\
  P_j - h_{jk} - fh_{jb} - f \frac{\beta_j}{n_i} & \text{if } k \not\in \{b, ATM\} \text{ and } b \in i \\
  P_j - h_{jk} - f(h_{jb} + \beta_j) & \text{if } k \neq ATM \text{ and } b \not\in i \\
  P_j - h_{jATM} - f(h_{jb} + \beta_j) & \text{if } k = ATM \text{ and } b \not\in i 
\end{cases}
\]

To explain the first line, suppose the strategy chosen includes using branches to serve some segment \( (b \in i) \). Then, the required transactions will be processed in a branch. So, if the channel chosen is a branch (or ATM), the customer will pay up to the full price less the convenience cost of using a branch (or ATM) and the associated travel cost, plus the cost of conducting the required transaction. The other expressions are similarly explained.

Substituting equations (2) into (1) yields the profit function for strategy \( i \):

\[
\pi_i(\bar{P}) = \begin{cases} 
  \sum_{j \in J} \sum_{k \in K} P_j \tau_{ijk} - \sum_{j \in J} \sum_{k \in K} \frac{\beta_j}{n_i} \left( \tau_{ijk} + \tau_{jATM} + f \left( \sum_{k \in K} \tau_{ijk} \right) \right) - \sum_{j \in J} \sum_{k \in K} h_{jk} \tau_{ijk} - f \sum_{j \in J} \sum_{k \in K} h_{jb} \tau_{ijk} \\
  - \sum_{j \in J} \sum_{k \in K} c_k \tau_{ijk} - fc_b \sum_{j \in J} \sum_{k \in K} \tau_{ijk} - \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{ijk}^2 - \sum_{k \in K} G_k - n_i F_i & \text{if } b \in i \\
  \sum_{j \in J} \sum_{k \in K} P_j \tau_{ijk} - \sum_{j \in J} \sum_{k \in K} \frac{\tau_{jATM}}{n_i} - f \left( \sum_{k \in K} \tau_{ijk} \right) - \sum_{j \in J} \sum_{k \in K} h_{jk} \tau_{ijk} \\
  - f \sum_{j \in J} h_{jb} \sum_{k \in K} \tau_{ijk} - fc_b \sum_{j \in J} \sum_{k \in K} \tau_{ijk} - \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{ijk}^2 - \sum_{k \in K} G_k - n_i F_i & \text{if } b \not\in i 
\end{cases}
\]

(If strategy \( i \) does not include ATM’s, the number of transactions directed to ATM’s is zero, that is, \( \tau_{jATM} = 0 \)).
A profit maximizing bank's choice of number of branches is readily found. Given $\bar{\tau}_i$, the profit function is convex in $n_i$; hence the optimal $n$ is found from the first order condition, as in equation (4):

$$
n_i^* = \begin{cases} 
\frac{\sum_{j \in J} \beta_j \left( \tau_{jb} + \tau_{jATM} + f \left( \sum_{k \in K} \tau_{jk} \right) \right)}{F_i} & \text{if } b \in i \\
\sqrt{\frac{\sum_{j \in J} \beta_j \tau_{jATM}}{F_i}} & \text{if } b \notin i 
\end{cases}
$$

(4)

Notice that the optimal $n_i$ is increasing in transaction volume and consumer sensitivity to distance, $\beta_j$, and is decreasing in the fixed cost of a physical outlet, $F_i$. Further note that when there are no transactions at branches and ATM’s, then the optimal number of outlets is zero. Substituting for $n_i^*$ from (4) into (3):

$$\pi_i^*(\bar{\tau}) = \max_{\tau_i} \pi_i(\bar{\tau_i})$$

$$ \begin{align*}
&= \max_{\tau_i} \sum_{j \in J} \sum_{k \in K} P_j \tau_{jk} - 2\sqrt{F_i} \sum_{j \in J} \sum_{k \in K} \beta_j \left( \tau_{jb} + \tau_{jATM} + f \left( \sum_{k \in K} \tau_{jk} \right) \right) - \sum_{j \in J} \sum_{k \in K} h_{jk} \tau_{jk} \\
&\quad - f \sum_{j \in J} \sum_{k \in K} h_{jb} \tau_{jk} - \sum_{j \in J} \sum_{k \in K} c_k \tau_{jk} - f c_b \sum_{j \in J} \sum_{k \in K} \tau_{jk} - \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{jk}^2 - \sum_{k \in K} G_k & \text{if } b \in i \\
&\quad - \sum_{j \in J} \sum_{k \in K} c_k \tau_{jk} - c_b \sum_{j \in J} \sum_{k \in K} \tau_{jk} - \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{jk}^2 - \sum_{k \in K} G_k & \text{if } b \notin i
\end{align*}
$$

(5)

$$s.t. \quad \tau_{jk} \geq 0 \quad \forall j \in J, \ k \in K
$$

(6)

For notational convenience, define the following functions:

$$R_i(\bar{\tau}_i) = \sum_{j \in J} \sum_{k \in K} P_j \tau_{jk}$$

$$C_i(\bar{\tau}_i) = \begin{cases} 
2\sqrt{F_i} \sum_{j \in J} \sum_{k \in K} \beta_j \left( \tau_{jb} + \tau_{jATM} + f \left( \sum_{k \in K} \tau_{jk} \right) \right) + \sum_{j \in J} \sum_{k \in K} h_{jk} \tau_{jk} + f \sum_{j \in J} \sum_{k \in K} h_{jb} \tau_{jk} \\
+ \sum_{j \in J} \sum_{k \in K} c_k \tau_{jk} + f c_b \sum_{j \in J} \sum_{k \in K} \tau_{jk} + \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{jk}^2 + \sum_{k \in K} G_k & \text{if } b \in i \\
2\sqrt{F_i} \sum_{j \in J} \beta_j \tau_{jATM} + \sum_{j \in J} \sum_{k \in K} \beta_j \tau_{jk} + \sum_{j \in J} \sum_{k \in K} h_{jk} \tau_{jk} + f \sum_{j \in J} \sum_{k \in K} h_{jb} \tau_{jk} \\
+ \sum_{j \in J} \sum_{k \in K} c_k \tau_{jk} + f c_b \sum_{j \in J} \sum_{k \in K} \tau_{jk} + \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{jk}^2 + \sum_{k \in K} G_k & \text{if } b \notin i
\end{cases}$$

Refer to $R_i$ as the “Full Revenue for strategy $i$” function and to $C_i$ as the “Full Cost for bank strategy $i$” function. The objective function in the bank’s optimization problem (5) is not concave everywhere.
due to the nonconvexity of the full cost function. For \( \bar{\tau}_j \) sufficiently large, the quadratic terms dominate the square root term and it is strictly convex, but for smaller values of \( \bar{\tau}_j \), the function exhibits increasing returns to scale. This is formalized in Lemma 3.1.

Define a bank \( i \)'s supply function, \( s_i(\tilde{P}) \), as the production decision that maximizes its profit given the full price vector \( \tilde{P} \). Specifically, for bank \( i \),

\[
\tilde{s}_i(\tilde{P}) = \arg \max_{\pi_i(\tilde{P}, \bar{\tau}_j)} \sum_{j \in J} \pi_i(\tilde{P}, \bar{\tau}_j) \tag{7}
\]

so that \( \tilde{s}_i(\tilde{P}) \in \mathbb{R}_{+}^J \) is strategy \( i \)'s profit maximizing production decision. If all banks maximize their profit by choosing supply, the aggregate supply delivered to the market is given by the function

\[
\tilde{S}(\tilde{P}) = \left( \tilde{s}_i(\tilde{P}) \right)_{j \in J} = \left( \sum_{i \in I} m_i s_i(\tilde{P}) \right)_{j \in J} \tag{8}
\]

Now that the consumers and the banks have been defined, it remains to discuss how those two interact in the market setting. The next section discusses the definition of a competitive entry equilibrium.

3. Equilibrium in the Retail Banking Market
This section defines and proves the existence of an equilibrium solution for the retail banking market. We define a \textit{competitive entry equilibrium}. Then we show its existence and properties.

3.1 Competitive Equilibrium and the Competitive Entry Equilibrium
Given a vector of the number of banks of each type in the market, \( \vec{m} \), a competitive equilibrium is a vector of full prices and a vector of transaction-volume decisions such that each bank maximizes its profit and aggregate supply equals demand.

\textit{Definition:} Let \( \tilde{D}(\tilde{P}) \) represent the vector of customer demands at full prices \( \tilde{P} \). A competitive equilibrium for a given vector of firm types \( \vec{m} \) consists of a full price vector \( \tilde{P}^* \) and a vector of firms' transaction volume decisions, \( \tilde{\tau}^* \), such that \( \tilde{\tau}^*_j = \tilde{s}_j(\tilde{P}^*) \) for all \( i \in I \) and

\[
\tilde{S}(\tilde{P}) = \left( \sum_{i \in I} m_i \tilde{\tau}_j \right)_{j \in J} = \tilde{D}(\tilde{P}) \, .
\]
Given the nonconvexities of the banks' profit function in (3), a competitive equilibrium does not necessarily exist.

Next, we extend the definition of equilibrium to allow entry to and exit from the market. A competitive equilibrium with a fixed number of firms ignores the possibility that if positive profits exist in a market, then new firms can enter. A competitive entry equilibrium will be defined as a competitive equilibrium in which all firms in the market earn zero profits.

**Definition:** A competitive entry equilibrium, \((\bar{m}^*, \bar{P}^*, \bar{\tau}^*)\), is defined to be a vector of full prices \(\bar{P}^*\), a vector of firms' transaction volume decisions, \(\bar{\tau}^*\), and a vector \(\bar{m}^*\) of the number of banks of each type such that \(\bar{P}^*\) and \(\bar{\tau}^*\) form a competitive equilibrium and \(\pi_i(\bar{\tau}_i^*) = 0\) for all \(i \in I\).

Our method of proof of the existence of a competitive entry equilibrium is to prove that \((\bar{m}^*, \bar{P}^*, \bar{\tau}^*)\) is a competitive entry equilibrium if this vector also maximizes social welfare. This method of proof allows demonstration of the existence of an unique entry equilibrium, even with the original, nonconcave profit function in equation (5).

### 3.2 Social Welfare and the Proof of the Existence of a Competitive Entry Equilibrium

Social welfare is defined as the sum of consumers' surplus and bank profits.

\[
SW(\bar{m}, \bar{\tau}) = \sum_{j=1}^{J} \left( \int_0^{\bar{\tau}_j} P_j(x)dx - \bar{P}_j(\bar{m} \cdot \bar{\tau}_j) \bar{m} \cdot \bar{\tau}_j \right) + \sum_{i \in I} m_i \left( R_i(\bar{P}, \bar{\tau}_i) - C_i(\bar{\tau}_i) \right)
\]

where \(P_j(x)\) denotes the separable inverse demand function associated with the demand function \(D_j(P)\). The domain of the variables are \((\bar{m}, \bar{\tau}) \in R^*_+ x R^*_+\). The second term of consumer surplus represents the total full revenue in the market, so the revenue terms cancel:

\[
SW(\bar{m}, \bar{\tau}) = \sum_{j=1}^{J} \int_0^{\bar{\tau}_j} P_j(x)dx - \sum_{i \in I} m_i C_i(\bar{\tau}_i) \tag{9}
\]

If social welfare is maximized on the domain, the necessary conditions are that

\[
\frac{\partial SW}{\partial \tau_{ik}} = m_i \left( P_j(\bar{m} \cdot \bar{\tau}_j) - v_{yk}(\bar{\tau}_i) \right) \leq 0 \quad \text{if} \quad \tau_{yk} > 0
\]

\[
\frac{\partial SW}{\partial m_i} = \sum_{j \in J} P_j(\bar{m} \cdot \bar{\tau}_j) \bar{\tau}_j - C_i(\bar{\tau}_i) \leq 0 \quad \text{if} \quad m_i > 0 \tag{10}
\]
\[
\begin{array}{c}
\bar{m} \geq 0, \bar{v} \geq 0
\end{array}
\] (12)

where \( v_{i,k}(\bar{\tau}_i) = \frac{\partial C_i(\bar{\tau}_i)}{\partial \tau_{i,k}} \) in equation (10). These conditions may not be sufficient because the social welfare function is not concave, which is easily noted since the cost functions are not convex. The next propositions demonstrate important properties of the social welfare function, and each firm’s cost function.

3.3 Properties of the Social Welfare Optimum and Proof of the Equilibrium’s Existence

This section finds important properties of the social welfare optimum and uses them to prove existence and uniqueness of a competitive entry equilibrium. The first result tells us that for any bank strategy, along any straight line in the positive quadrant in \( R_+^{JK} \), a bank’s cost function is first concave and then strictly convex.

**Lemma 3.1** For any \( i \in I \), any point \( \bar{\tau}_i \in R_+^{JK} \) and any vector \( \bar{\nu} \in R_+^{JK} \), define the function \( C_i(\lambda) = C_i(\bar{\tau}_i + \lambda \bar{\nu}) \). Suppose \( \lambda = \arg\min_{\lambda} \{ \lambda \mid \bar{\tau}_i + \lambda \bar{\nu} \in R_+^{JK} \} \), then there exists a \( \lambda_o \) such that \( C_i(\lambda) \) is strictly convex for all \( \lambda > \lambda_o \), and \( C_i(\lambda) \) is concave for all \( \lambda \leq \lambda_o \).

**Proof:** Suppose strategy \( i \) uses branches but not ATM’s. Then, by direct computation:

\[
\begin{align*}
C_i(\lambda) &= \sum_{j \in J} \beta_j \left( \tau_{ij} + \lambda \nu_{ij} + \sum_{k \in K} c_{ik} \left( \tau_{ik} + \lambda \nu_{ik} \right) \right) + \sum_{j \in J} \sum_{k \in K} h_{jk} \left( \tau_{jk} + \lambda \nu_{jk} \right) \\
&\quad + \sum_{j \in J} \sum_{k \in K} d_{jk} \left( \tau_{jk} + \lambda \nu_{jk} \right)^2 + \sum_{k \in K} G_k
\end{align*}
\]

This function has the property of the conclusion. Similar analysis can be done for the case where strategy \( i \) uses ATM’s only, and strategy \( i \) uses ATM and branches. When \( i \) does not use branches or ATMs, the cost function is strictly convex for all \( \lambda \). \( \Box \)

It is possible that \( \lambda_o = \lambda \) so that \( C_i(\lambda) \) is strictly convex for all lambda in the positive quadrant. The next result shows that there is a hyperplane in \( R_+^{JK} \) that separates the convex and non-convex regions of each firm’s cost function. This is a key technical result of this paper.
Proposition 3.1 Define vector $\vec{v} = \left( \frac{\beta_i}{d_{jk}} \right)_{j \in J, k \in K}$. For any $i \in I$, the region where $C_i(\cdot)$ is strictly convex in $R^N_i$, is the points strictly above the hyperplane defined by vector $\vec{v}$ and constant $\xi_i$:

$$H_i = \{ \vec{\tau}_i \in R^N_i \mid \vec{\tau}_i \cdot \vec{v} > \xi_i \}.$$  (13)

The proof is found in the Appendix. The conclusion of the proposition is that the vector $\vec{v}$ provides a direction in which the full cost function is not convex below a certain hyperplane and strictly convex above that hyperplane. (Generally the full cost function is not even concave below the hyperplane.)

This fact is very useful. For some strategy $i$, if cost function $i$ is concave at the point $\vec{\tau}_i$, then one can find two banks using strategy $i$, one producing at the hyperplane boundary and the other producing at one of the axes that has the same total production as $\vec{\tau}$, but lower total cost than the original bank $i$'s output. Thus at the social welfare optimum, all firms produce in the region given by hyperplane (13).

Proposition 3.2: If $(\vec{m}^*, \vec{\tau}^*)$ is a social welfare optimum and bank strategy $i$ is active ($m_i^* > 0$), then bank $i$'s production $\vec{\tau}_i^*$ lies above the hyperplane (13).

The proof is found in the Appendix, and is by contradiction, using a construction. That is, we assume that the conclusion is false and show by construction that social welfare cannot be maximized at the supposed $(\vec{m}^*, \vec{\tau}^*)$.

We next show that a social welfare optimum is also a competitive entry equilibrium.

Proposition 3.3: If the vectors $\vec{\tau}^*$ and $\vec{m}^*$ are a maximizing point for the social welfare function, and $\vec{P}^*$ is the vector of full prices at the social welfare optimum, then $(\vec{m}^*, \vec{P}^*, \vec{\tau}^*)$ is a competitive entry equilibrium.

Proof: The proof is by contradiction: We suppose that a firm is not maximizing its profit, and show that this implies that social welfare cannot be maximized. Suppose that at a social welfare optimizing point given by $\vec{\tau}^*$ and $\vec{m}^*$, there exists some bank strategy $i''$ that is not optimizing profits. Let $\vec{\tau}_{i''}$ be
the vector of production of bank \(i\) at the social welfare optimum. Specifically, there exists some vector of production decisions for bank \(i\), \(\bar{z}_i \neq \bar{z}_i^*\) such that

\[
\bar{P} \cdot \bar{z}_i - C_i(\bar{z}_i) > 0.
\]  
(14)

Let \(\Delta m \in R_+\) be sufficiently small so that \(\|\Delta m \bar{z}_i\|_2 = \epsilon > 0\), where \(\epsilon\) is small. Note that the sum of banks' total cost of producing the aggregate quantity \((\bar{z}_i^* - \Delta m \bar{z}_i^*)\) is equal to

\[
\sum_{i \in I} m_i C_i(\bar{z}_i^*) - \nabla_{\bar{z}_i} C_i(\bar{z}_i^*) \cdot (\Delta m \bar{z}_i^*) = \sum_{i \in I} m_i C_i(\bar{z}_i^*) - \Delta m \bar{P} \cdot \bar{z}_i^*
\]  
(15)

since price is equal to marginal full cost at a social welfare optimum. Now assign \(\Delta m\) additional banks of type \(i\) to each produce the vector \(\bar{z}_i\). The total cost from adding these new banks is \(\Delta m C_i(\bar{z}_i^*)\). Summing (15) and \(\Delta m C_i(\bar{z}_i^*)\), total production is held constant with total cost as

\[
\sum_{i \in I} m_i C_i(\bar{z}_i^*) - \Delta m \bar{P} \cdot \bar{z}_i^* + \Delta m C_i(\bar{z}_i^*) = \sum_{i \in I} m_i C_i(\bar{z}_i^*) - \Delta m (P \cdot \bar{z}_i - C_i(\bar{z}_i)) < \sum_{i \in I} m_i C_i(\bar{z}_i^*) > 0
\]

The quantity in parentheses is >0 as assumed in equation (14). Thus, total production is constant but the total cost of that production is reduced, contradicting the assumption of social welfare optimality. Thus, (14) cannot hold and all active banks are profit maximizers. \(\square\)

We can show that the competitive entry equilibrium is unique because of our next result: The social welfare optimum is unique.

**Proposition 3.4:** The social welfare function, \(SW(\bar{m}, \bar{z})\), given in (9) has a unique optimal solution.

**Proof:** See the Appendix.

The method of proof is to show that the social welfare function is sequentially strictly convex. This means for any fixed \(\bar{m}\), SW is strictly convex in \(\bar{z}\) and defining the SW maximizing production as a function of \(\bar{m}\), \(\bar{z}(\bar{m})\), the function \(SW(\bar{m}, \bar{z}(\bar{m}))\) is strictly convex in \(\bar{m}\). The next proposition shows that a competitive entry equilibrium corresponds to a social welfare optimum.

**Proposition 3.5:** A competitive entry equilibrium is also a social welfare optimum.

**Proof:** By the proof of Proposition 3.4, the social welfare function is sequentially strictly convex. Thus the Kuhn-Tucker conditions given in Equations (10), (11), (12), and (16) are necessary and sufficient for a social welfare optimum. In competitive entry equilibrium full price equals marginal cost, thus Equation (10) holds for any firm that is in the market. Similarly, Equation (11) holds for all
firms in the market. Thus, a competitive equilibrium satisfies the necessary and sufficient first order conditions for social welfare optimization so a competitive equilibrium corresponds to a social welfare optimum. □

The next results characterize the competitive entry equilibrium. We show that in equilibrium full prices are equal to marginal full cost. We also show that all customers pay a real price equal to the marginal production cost of the providing firm.

**Proposition 3.6:** If \((\tilde{m}^*, \tilde{P}^*, \tilde{v}^*)\) is a competitive equilibrium, then, production rates minimize the total cost of firms’ production and customers’ convenience costs subject to meeting demand. Further in equilibrium, the full price for segment \(j\) customers is equal to the marginal full costs of all firms serving these customers and is no higher than the marginal full costs of all firms not serving the segment. That is, for all \(j \in J\)

\[
\tilde{P}_j^* = \frac{\partial C_i(\tilde{r}_i^*)}{\partial r_{jk}} \quad \text{for all } \tilde{r}_{jk}^* > 0.
\]

\[
\tilde{P}_j^* \leq \frac{\partial C_i(\tilde{r}_i^*)}{\partial r_{jk}} \quad \text{for all } \tilde{r}_{jk}^* = 0.
\]

**Proof:** The first assertion is immediate since the competitive entry equilibrium is also a social welfare optimum where equation (10) holds. The second follows as full price is equal to marginal full cost at a social welfare optimum by Equation (10). The third holds as if not, the Social Welfare optimum does not minimize total production cost, which is a contradiction. □

**Corollary 3.1** For bank strategy \(i\), all segments that purchase distribution channel \(k\) from \(i\) are charged the marginal cost of transaction \(k\) for segment \(j\).

**Proof:** The per transaction price by a firm using strategy \(i\) to serve segment \(j\) using channel \(k\) is given by (2). In competitive entry equilibrium full price is equal to marginal full cost. This implies that prices given by (2) are equal to the marginal costs for the channel \(k\).

Although all customers are charged a real price equal to full marginal cost, channel prices may differ between segments. This is because banks’ marginal costs may not be the same to these segments. In section 2.3, we assumed a cost function that implied that transactions are distinguishable from each other.
Lederer and Li (1997) show that when transactions are distinguishable, incentive compatibility is not an issue because a firm can determine (possibly ex-post) the customer type requesting the transaction. There are many ways that customer types can be distinguished independent of observing single transactions, particular through segmentation strategies (Kotler 2000, Chapter 9). Demographic segmentation is often used by banks through age-related (e.g.: youth and senior) restrictions on sale. Benefits-related segmentation strategies are often observed in retail banking. When usage pattern of a segment can be identified, then usage limits, or equivalently, nonlinear use prices would separate groups. Similarly, similar services can be redesigned so that they are differentiated. As an example, ATM systems can be designed to offer advanced features and functions that only some segments desire to use. Thus banks find ways to price the same type of transaction differently for segments.

4. Sensitivity Analysis of Retail Banking Market Equilibrium

This section presents sensitivity analysis that examines the effect of various parameters on the market equilibrium. The complexity of the model forces the sensitivity analysis to be done numerically. The analysis will examine the effect on equilibrium of changing customer demands, variable costs, and fixed costs of PC banking systems. The analysis suggests that the relative size of customer segments, as opposed to cost structure or total demand, induces strategy shifts of competing banks.

For this numerical study, we assume three distribution channels available to competing banks. The three technologies are teller-staffed branches (denoted $B$), ATM’s (denoted $ATM$) and PC banking systems, which are generally assumed to be home-based, Internet-delivered transactions (denoted $PC$). Distribution channel space is thus $K = \{B, ATM, PC\}$. Of course, any bank in the market must have at least one branch to serve required transactions.

We assume two customer segments, one that prefers branches and one that prefers PC banking. The customer segments are denoted $b$ for the segment that prefers branches and $e$ for the segment that prefers PC banking. Thus, $J = \{b, e\}$.

Enumerating the full cost function requires data for customer utility. The analysis assumes that the customer technology disutility is ordered as follows: for branch preferring customers $0 = h_{bb} \leq h_{bATM} \leq h_{bPC}$, and for electronic distribution preferring customers $0 = h_{ePC} \leq h_{eATM} \leq h_{eB}$. The electronic channel preferring segment incurs its greatest disutility at branches.
We consider all possible bank strategies generated by our set \( J \times K \), which number \( 2^{31} - 1 = 63 \) strategies. There are many interesting strategies in the relevant space: \( i = \{(b,B),(e,B)\} \): all customers are served by branches, \( i = \{(b,B),(e,PC)\} \): the electronic distribution channel preferring segment is served by PC banking and the branch preferring segment is served by branches, etc. Some strategies are uninteresting, of course: \( i = \{(b,PC),(e,B)\} \), \( i = \{(e,B)\} \), \( i = \{(b,ATM)\} \), etc. There are three classes of strategies that arise in equilibrium and it is convenient to name them. A branch strategy serves \( b \) customers with branches and ATMs, with \( e \) customers possibly served by ATMs; an example is \( i = \{(b,B),(b,ATM),(e,ATM)\} \). A full strategy is any strategy that uses all three distribution channels to serve both segments. For example, such a strategy is \( i = \{(b,B),(b,ATM),(b,PC),(e,ATM)\} \). Virtual strategies are those where all customers are served via ATM’s and PC (that is, there are no branches): Such a strategy is \( i = \{(e,PC),(b,ATM)\} \).

Values for the parameters in the analysis are estimates derived from industry publications. References include American Banker Magazine and the book Distribution 2000: Developing and Implementing Strategies for Retail Financial Institutions. Numbers gathered from those sources are representative of a wide range of banks and allow for a reasonable range for approximation. Note that the coefficients on the quadratic cost terms are not readily available from industry literature. The coefficients used in the analysis below are chosen empirically to represent the current number of banks in a market. Table 2 gives the point estimates and ranges for each of the parameters used in the analysis. We assume the diseconomy of scale for branch transactions is greater than that for ATM transactions, which in turn, is larger than that for PC banking.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Point Estimate (Base Value)</th>
<th>Range used in Sensitivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable Costs: linear</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c_1 ): Branch</td>
<td>$0.55 \text{ per transaction}</td>
<td>$0.10 -- $1.00</td>
</tr>
<tr>
<td>( c_2 ): ATM</td>
<td>$0.30 \text{ per transaction}</td>
<td>$0.1 -- $0.775</td>
</tr>
<tr>
<td>( c_3 ): PC</td>
<td>$0.11 \text{ per transaction}</td>
<td>$0.01 -- $0.55</td>
</tr>
<tr>
<td><strong>Variable Costs: quadratic coefficients</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d_{11} ): ( b ) customer at Branch</td>
<td>( 6 \times 10^{-7} )</td>
<td></td>
</tr>
<tr>
<td>( d_{12} ): ( b ) customer at ATM</td>
<td>( 9 \times 10^{-7} )</td>
<td></td>
</tr>
<tr>
<td>( d_{21} ): ( e ) customer at Branch</td>
<td>( 8 \times 10^{-7} )</td>
<td></td>
</tr>
<tr>
<td>( d_{22} ): ( e ) customer at ATM</td>
<td>( 7 \times 10^{-7} )</td>
<td></td>
</tr>
<tr>
<td>( d_{23} ): ( e ) customer at PC</td>
<td>( 6 \times 10^{-7} )</td>
<td></td>
</tr>
</tbody>
</table>
**Customer travel costs**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>customer sensitivity to traveling to a branch/ATM</td>
<td>$1.5 \text{ per mile}$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>customer sensitivity to traveling to a branch/ATM</td>
<td>$2.25 \text{ per mile}$</td>
</tr>
</tbody>
</table>

**Required Transactions**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Customer Disutility**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{bb}$</td>
<td>customer at Branch</td>
</tr>
<tr>
<td>$h_{bATM}$</td>
<td>customer at ATM</td>
</tr>
<tr>
<td>$h_{bPC}$</td>
<td>customer at PC</td>
</tr>
<tr>
<td>$h_{eb}$</td>
<td>customer at Branch</td>
</tr>
<tr>
<td>$h_{eATM}$</td>
<td>customer at ATM</td>
</tr>
<tr>
<td>$h_{ePC}$</td>
<td>customer at PC</td>
</tr>
</tbody>
</table>

**Fixed Costs** *(Annual fixed costs. One-time charges are amortized over 10 years.)*

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Branch Placement (with ATMs): Cost Per Branch</td>
<td>$500,000 \text{ per branch}$</td>
<td></td>
</tr>
<tr>
<td>Branch Network (with ATMs): Fixed Cost</td>
<td>$400,000</td>
<td></td>
</tr>
<tr>
<td>ATM/location</td>
<td>$100,000 \text{ per location}$</td>
<td></td>
</tr>
<tr>
<td>ATM network</td>
<td>$200,000</td>
<td></td>
</tr>
<tr>
<td>PC banking System</td>
<td>$50,000</td>
<td></td>
</tr>
<tr>
<td>Overhead</td>
<td>$1,000,000</td>
<td></td>
</tr>
<tr>
<td>Start-up</td>
<td>$500,000</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Parameter values used in sensitivity analysis

We assume a linear inverse demand function with parameters as shown in Table 3. These parameters represent our estimate of current market sizes and elasticities.

**Customer Segment $b$**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>Intercept</td>
</tr>
<tr>
<td>$b_1$</td>
<td>Slope</td>
</tr>
</tbody>
</table>

**Customer Segment $e$**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_2$</td>
<td>Intercept</td>
</tr>
<tr>
<td>$b_2$</td>
<td>Slope</td>
</tr>
</tbody>
</table>

**Table 3:** Base inverse demand function parameters for numerical experiments
The sensitivity analysis uses the above parameters to examine the effects on the equilibrium mix of banks and transaction volumes of varying demand, variable costs and fixed cost. The social welfare maximization procedure is written in *Mathematica*. The procedure finds the optimal quantities for a given vector of firms and then updates the vector of firms until KKT conditions are satisfied.
4.1 Competitive Entry Equilibrium Without PC Banking Distribution Channel Technology

The first test examines the equilibrium outcome without the possibility of PC banking systems. This is intended to represent the “historical” banking systems that relied solely on branches and ATM’s under current demand and cost conditions. Parameters used for estimation are those found in Tables 1 and 2. The number of competing banks and the transaction volume decisions of each bank in the Pre-PC case are presented in Table 4. We find that one bank strategy occurs in equilibrium. Customer type $b$ is served by branches and ATMs; and $e$ customers are served only by ATMS.

<table>
<thead>
<tr>
<th>“Historical” Case Equilibrium Outcome: Number of regular transactions/year per bank.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy $i^*=(b,B),(b,ATM),(e,ATM)$, $m_{(b,B,K,e,ATM)}=\text{10.46}$</td>
<td></td>
</tr>
<tr>
<td>Segment</td>
<td>Branch</td>
</tr>
<tr>
<td>e Segment</td>
<td>0</td>
</tr>
<tr>
<td>b Segment</td>
<td>1,990,000</td>
</tr>
</tbody>
</table>

Table 4: Equilibrium outcome in market before advent of PC banking. Table entries are annual transaction volumes. Only one strategy exists in equilibrium. In it, the $e$ segment is served by ATM’s, while the $b$ segment is served by branches and ATM’s. Parameters are set using base values in Tables 2 and 3. This represents our prediction of bank strategies, number of banks in the market and transaction shares if PC banking is not available.

4.2 Competitive Entry Equilibrium With PC Banking

We now allow banks to adopt PC banking. Again all parameters are set to their base values as presented in the tables above. The equilibrium result is .77 branch banks and 9.71 full banks. The transaction volume that each bank serves is presented in Table 5 below.

Note that the branch banks specialize in serving only branch customers while the full banks serve both customer segments through the customers’ preferred technology. Next we perform sensitivity analysis on these results.
Mix of Bank Strategies at Equilibrium

<table>
<thead>
<tr>
<th>Strategy: # banks</th>
<th>Segment</th>
<th>Branch</th>
<th>ATM</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Branch: .77</td>
<td></td>
<td>1,970,000</td>
<td>44,794</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full: 9.71</td>
<td></td>
<td>2,000,000</td>
<td>86,531</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0</td>
<td>0</td>
<td>287,000</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 5:* Equilibrium results in baseline case. When PC banking is permitted as a distribution channel, almost all banks adopt it along with branch and ATM technologies. There is a small market for banks serving the b segment using branches and ATM's.

### 4.3 Sensitivity Analysis and Relative Market Size

The first sensitivity test varies the relative size of the customer segments and will provide insight into the equilibrium as the e customer segment grows. It is generally assumed that the number of those who are comfortable using PC technology will grow tremendously during the next several years. To conduct the test, the demand curve slopes are held constant while the intercepts are adjusted. The intercepts are varied as to maintain a constant total number of potential transactions (at zero price) of 28 million. The test assumes a constant market size but increasing e segment size. The results are shown in Figures 1 and Figure 2.

Note that as the e customer segment grows and approaches the size of the b customer segment the branch strategy banks exit the market and only full strategy banks compete. As the e customer segment continues to grow and the b segment dwindles, virtual banks enter the market. Three points of interest follow this analysis. First, some full banks are left in the market so long as the b customer segment has some demand. Second, while the total number of banks initially drops, there is net entry into the market once e customers are the dominant segment. The full strategy has the highest fixed cost of all the strategies, hence there are fewer competitors when full banks enter. On the other hand, the virtual strategy has the lowest fixed cost of any strategy so net entry into the market is expected when virtual banks are entering. This result appears to mirror the recent history of the U.S. banking industry. In fact, banks are courting e customers and attempting to educate other customers.
**Figure 1:** Number of banks resulting in market when relative size of the customer segments is varied.

**Figure 2:** Total transaction volume through each technology as relative size of the customer segments is varied. Required transactions are reported as are the regular transactions that are served by branch banks.
by installing PC banking systems, but that segment remains small. In recent years, there has been consolidation in the banking market and the number of banks using only branches and ATM’s has fallen as even the smaller regional banks and credit unions deploy PC banking systems. Further, the model would suggest that as the e customer segment grows over time and the b segment diminishes, those banks that don’t convert to a virtual strategy will be forced out of the market and there eventually will be room for new competitors using a virtual strategy. Third, and last, the virtual strategy will not be viable until about half of the potential market are e customers. Current estimates of PC banking use are about 7-10% of households. Thus, the virtual strategy will not be a viable strategy for some time to come, until the e segment grows much larger than it currently is.

4.4 Sensitivity Analysis on the Fixed Cost of PC Banking

This sensitivity test examines the effect of a range of fixed costs for PC banking systems. All other parameters are set to their base levels and the up front fixed cost of a PC banking system is varied. The results are presented in Figure 3.

As the fixed cost of PC banking falls, the full strategy is substituted for the branch strategy while the total number of banks remains roughly constant. Further, it is surprising that even though the mix of banks is changing, the total number of transactions served in the market remains constant for both customer segments. Thus, as PC banking systems become more affordable, the analysis shows that branch banks are forced out of the market in favor of more full banks, while roughly the same total number of transactions is served. Some have suggested that as PC banking systems become ever cheaper, the e customer segment will be served by virtual banks, regardless of the relative size of the e segment. This analysis contradicts that supposition.
4.5 Sensitivity Analysis on the Density of Demand

This test varies the density of demand, i.e., the ratio of market sizes of the two customer segments is kept constant but the total market size is varied. This enables testing of the model in denser market regions. The results are presented in Figure 4.

Figure 4: Equilibrium mix of banks as total market size increases and relative size of customer segments remains constant. For low demand density, branch bank strategy is not competitive, and is not observed. As the density of demand increases and branch banks exist, the number of full banks increase at a faster rate than those adopting branch strategy.
This figure shows that when total demand is very small, the branch strategy is not viable, only the full strategy can compete. Further, as demand grows, the number of full banks grows faster than the number of branch banks. Note also that the kiosk/PC strategies and the virtual strategies never enter the market. These results are due to the full strategy’s ability to serve both customer segments through their preferred transaction methods, thus allowing for profitable entry despite the low customer demand and high fixed costs associated with the full strategy. Notice that unlike the results from the analysis on the relative size of the customer segments (Figures 1 and 2), there is specialization in serving the $b$ customer segment the branch strategy doesn’t serve any $e$ customers but there is no specialization in the $e$ segment. Some have suggested that one should expect to see virtual banks leveraging their low fixed costs to "steal" profitable $e$ customers from the full banks. The sensitivity analysis supports the opposite. The full strategy is able to serve enough $e$ customers through the lower-cost PC banking channel to prevent entry by virtual banks but leaves volume to be served by branch banks. This test suggests that unless the $e$ customer segment grows relative to the $b$ segment, the full strategy is the only one to serve $e$ customers.

4.6 Sensitivity Analysis on Variable Cost Parameters

This sensitivity test considers the variable cost parameters for branch and PC transactions. The general industry wisdom is that the significant cost savings of PC banking systems will drive banks to migrate their customers to these systems and, thus, there will be a rise in the number of full and virtual banks in the market. In the analysis, both variable cost parameters are tested at low, medium and high values. The variable cost of an ATM transaction is taken as the average of the branch and PC variable cost. The results follow in Table 6.

<table>
<thead>
<tr>
<th>Branch Variable Cost</th>
<th>$0.10</th>
<th>$0.55</th>
<th>$1.00</th>
<th>$3.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01</td>
<td>1.51, 8.9, .64</td>
<td>0, 10.34, 0</td>
<td>0, 9.44, 0</td>
<td>0, 6.91, 0</td>
</tr>
<tr>
<td>$0.10</td>
<td>1.77, 9.03, 0.3</td>
<td>0.86, 9.64, 0</td>
<td>0, 9.7, 0</td>
<td>0, 6.93, 0</td>
</tr>
<tr>
<td>PC $0.25</td>
<td>0, 9.3, 1.34</td>
<td>0, 9.86, 0</td>
<td>0, 6.87, 0</td>
<td></td>
</tr>
<tr>
<td>Variable Cost</td>
<td>$0.55</td>
<td>1.64, 8.71, 0.52</td>
<td>1.2, 9.06, 0</td>
<td>6.88</td>
</tr>
<tr>
<td>$1.00</td>
<td>1.5, 8.2, 0.93</td>
<td>0, 7.17, 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: The number of banks adopting branch, full and virtual strategies as a function of the variable cost of a branch transaction and a PC banking transaction.
The table shows that varying the variable cost parameter only significantly affects the equilibrium mix and number of banks for the highest values of variable cost of branch transactions. Even at the highest cost for branch transactions, while branch banks do exit the market, virtual strategies do not enter. This result suggests that without a shift in customer preferences, variable cost alone does not imply widespread changes in how banks compete in the market.

5. Conclusion and Future Research for Retail Banking Model

Retail banking distribution strategy is an important issue in the banking industry and this paper shows the effect of key banking parameters on the choice of that strategy. This paper presents an equilibrium model of distribution strategy choice, including heterogeneous consumers and banks and allowing for a rich variety of market parameters. The paper further addresses the existence and uniqueness of the equilibrium. Finally, numerical sensitivity analysis shows how several parameters affect the equilibrium outcome and illustrates that the relative size of the customer segments drives the choice of distribution strategy.

A unique market equilibrium is guaranteed despite nonconvexities in the banks’ cost functions. The model allows banks to optimally choose the number of branches or kiosks they place in the market. As customer convenience is an inverse function of the number of banks and bank cost is a linear function of the same, the result is a nonconvexity in the total cost function. It is surprising, then, that despite the nonconvexity; a market equilibrium exists and is unique. The uniqueness of the equilibrium make possible a sensitivity analysis that compares the equilibrium outcome for a wide variety of parameter values.

Most importantly, these results can be used to analyze competitive equilibrium in other situations of operations systems with economies of scale.

Assuming two customer segments and four strategies, the analyses demonstrate that the equilibrium mix of banks is highly sensitive to the relative size of the customer segments. Specifically, when the $e$ customer segment is dominated by the $b$ customer segment, branch and full banks serve the market. Then, as the number of consumers that prefer PC transactions increases and the number of customers that prefer branches decreases, there is first an increase in full banks (as branch banks exit), followed by an increase in the number of virtual banks until only virtual banks are left. This result contrasts with the analysis in which the relative size of the customer segments was
kept constant while the overall size of the market was varied. That test demonstrated that even though the magnitude of the e customer segment increased, virtual banks did not enter the market because the full banks prevented their entry.

Further, the analysis demonstrates that the equilibrium is relatively insensitive to technology cost differences. When the difference between PC and branch transactions is at its greatest, branch banks exit the market leaving only full banks; however, when the difference is not as high, both bank types serve customers.

These results suggest that changing consumer behavior and attitudes, instead of banks’ cost structure, effects significant changes in distribution strategy. If the segment of consumers that prefer PC banking remains small relative to the segment that prefers branches then there will still be a market for specialized branch banks and the full banks can prohibit successful entry by virtual banks. However, if the e segment grows at the expense of the b segment, as some have predicted, the model predicts that branch banks and, eventually, full banks will exit the market in favor of virtual banks.

The current analysis is illuminating but leaves room for further work. One promising addition to the model would be to create classes of transactions. For example, it is currently not possible withdraw cash from a PC banking system and, despite technological advances, it may never be a viable possibility. It would make sense then to create a class of transactions that must be conducted through an ATM or branch. This change may increase the persistence of strategies that include branches and may provide additional importance to the kiosk/PC strategy. A finer segmentation of the consumers in the market may also provide additional insights into the equilibrium. Finally, other work has established the existence of network effects in consumer acceptance of retail banking technology (Prasad and Harker 2000); incorporating those effects into an equilibrium model may provide additional insight.

References


Appendix

Proof of Proposition 3.1 The bank i’s cost function can be written in the form:

\[
C_i(\tau_i) = 2\sqrt{F_i} \left( \sum_{j \in J} \sum_{k \in K} \bar{\beta}_{jk} \tau_{jk} + \sum_{j \in J} \sum_{k \in K} \bar{h}_{jk} \tau_{jk} + \right.
\]

\[
+ \sum_{j \in J} \sum_{k \in K} \bar{c}_k \tau_{jk} + \sum_{j \in J} \sum_{k \in K} d_{jk} \tau_{jk}^2 + \sum_{j \in J} d_{jk} \tau_{jk} + \sum_{k \in K} G_k
\]

for some coefficients \( \bar{\beta}_{jk} \) and \( \bar{c}_k \). The Hessian matrix of \( C_i \) is the matrix of second partials of \( C_i \) with respect to \( \tau_i \) and is denoted \( H(\tau_i) \). Let \( \vec{c} \in \mathbb{R}^K \) be a vector in \( JxK \) with \( \vec{c}^T \) its transpose. \( C_i(\vec{c}) \) is positive definite at \( \tau_i \) iff \( \vec{c}^T H(\vec{c}) \vec{c} > 0 \) for all \( \vec{c} \neq 0 \). If \( C_i(\vec{c}) \) is positive definite in a compact set, it is also convex in that set. Define the vector \( \vec{V}_i = \left( \frac{\bar{\beta}_{jk}}{d_{jk}} \right)_{j \in J, k \in K} \).
Write $\text{Det}(\tilde{\tau}_i)$ for the determinant of the matrix $H(\tilde{\tau}_i)$. It is possible to show by induction that

$$\frac{\tilde{v}^T H(\tilde{\tau}_i) \tilde{v}}{\text{Det}(\tilde{\tau}_i)} = \frac{\sum_{j \in J} \prod_{k \in K} v^2_{j,k}}{2^{2j-1} \prod_{j \in J} \prod_{k \in K} v^2_{j,k}} > 0.$$ 

Thus, $\tilde{v}^T H(\tilde{\tau}_i) \tilde{v} > 0$ iff $\text{Det}(\tilde{\tau}_i) > 0$. Also, by induction one shows that $\text{Det}(\tilde{\tau}_i) = 2^{2jK-1} \prod_{j \in J} \sum_{k \in K} v^2_{j,k} - 2^{2JK-1} \prod_{j \in J} \prod_{k \in K} v^2_{j,k}.$ The determinant of $H$ at $\tilde{\tau}_i$ is positive iff

$$\sum_{j \in J} \sum_{k \in K} \tilde{v}_{j,k} \tilde{v}_{j,k} > 2^{2K-1} \prod_{j \in J} \prod_{k \in K} v^2_{j,k} \sum_{j \in J} \sum_{k \in K} v^2_{j,k}.$$ 

Thus the hyperplane defined by

$$\left\{ \tilde{\tau}_i \in R^K \mid \tilde{\tau}_i \cdot \tilde{\beta} > \frac{2^{2K-1} \prod_{j \in J} \prod_{k \in K} v^2_{j,k} \sum_{j \in J} \sum_{k \in K} v^2_{j,k}}{2^{2K-1} \prod_{j \in J} \prod_{k \in K} v^2_{j,k}} \right\}$$

identifies the region where $\text{Det}(\tilde{\tau}_i)$ is strictly positive with the points above the hyperplane.

We next show that if the determinant of the Hessian of $C_i(\tilde{\tau}_i)$ evaluated at $\tilde{\tau}_i$ is positive, then the Hessian of $C_i(\tilde{\tau}_i)$ evaluated at $\tilde{\tau}_i$ is positive definite. Our method of proof is to show that if determinant of the Hessian of $C_i(\tilde{\tau}_i)$ is positive, then all of the principal minors of the Hessian of $C_i(\tilde{\tau}_i)$ are too. We know that if the principal minors of the Hessian are all positive, then the function is strictly positive definite and thus is strictly convex. Order the elements in strategy $i$, labeling them $q=1, 2, ..., Q$. Denote the principal minor of order $k$ of the Hessian as the matrix containing the first $q$ rows and columns of the Hessian. For all $q = 1, ..., Q$, the principal minor of order $q$ of the Hessian has the property that

$$\text{Det}D_q = 2d_q \text{Det}D_{q-1} - \frac{2d_q^2 c_q^2}{4(c_1^q + c_2^2 + ... + c_q^q)^2}.$$ 

Thus if $\text{Det}D_q > 0$, then $\text{Det}D_{q-1}, \text{Det}D_{q-2}, ..., \text{Det}D_1 > 0$. This tells us that if $\text{Det}(\tilde{\tau}_i) > 0$, then the determinant of all the principal minors are also positive we can conclude that the Hessian is positive definite, and the function $C_i(\tilde{\tau}_i)$ is strictly convex in an open neighborhood of $\tilde{\tau}_i$.

We conclude that if
\[
\sum_{j \in J} \sum_{k \in K} \beta_{jk} \tau_{jk} \geq \frac{2^{JK-1}}{2^{JK-1}} \sum_{j \in J} \prod_{k \in K} v_{jk} \sum_{j \in J} \prod_{k \in K} v_{jk}^{-1} \beta_{jk}^2,
\]  

(19)

the function \( C_i(\bar{\tau}_i) \) is strictly convex in an open neighborhood of \( \bar{\tau}_i \). If the sign is reversed, the cost function is not convex in an open neighborhood of \( \bar{\tau}_i \), because the Hessian is not even positive semi definite at that point, and thus the cost function is not convex.

**Proof of Proposition 3.2** We prove by contradiction, using a construction. That is, we assume that the conclusion is false and show by construction that social welfare cannot be maximized at the supposed \((\bar{m}', \bar{\tau}')\). The construction uses the following algorithm. The whole algorithm converges in a finite number of iterations. It is useful to define an ordering over all strategies in \( I \) that is consistent with the partial ordering given by \( \subset \).

**Definition A.1:** Define a complete strict ordering, \( \preceq \) over all elements of \( I \) that is consistent with the partial ordering given by \( \subset \) on \( I \). That is, if \( s', s'' \in Z, s' \neq s'' \) and \( s' \subset s'' \), then \( s' \preceq s'' \).

**Algorithm**

*Operation 0:* Order the strategies \( Z \) using \( \preceq \). Let \( i_0 \) be the maximum of \( Z \) under the strict ordering \( \preceq \). Set \( i = i_0 \).

*Iteration i*

Consider strategy \( i \).

Perform *Operation 1* on all production decisions \( \bar{\tau}_i \) using strategy \( i \).

Perform *Operation 2* on all production decisions \( \bar{\tau}_i \) using strategy \( i \) that lie in the strictly convex region of the cost function.

Go to the next strategy in the ordering (if there is one), Set \( i \) to this strategy and return to (*), If this was the last strategy go to End.

End

*Operation 0* applies a strict ordering to the strategies. *Operation 1* takes any production decisions within the nonconvex portion of the cost function and creates two new production decisions whose total production is the same as the original production decision. One decision is in the strictly convex portion of the cost function and the other is now a production strategy of lower dimension (where one segment-channel production decision that was nonnegative now equals zero). This operation leaves total production the same and reduces cost. *Operation 2* combines production
decisions that correspond to the same strategy and creates a new production decision in the strictly convex region with identical total production but strictly lower cost.

There are only a finite number of strategies. For strategy $i$, each application of Operation 1 eliminates a production decision outside the strictly convex region of the cost function $C_i()$. When Operation 1 is complete, all production decisions corresponding to strategy $i$ must lie in the strictly convex region of $C_i$. Operation 1 creates new production decisions for strategies of lower dimension that will be operated on by Operations 1 and 2 later when those strategies are processed. Operation 2 reduces the number of remaining production decisions for strategy $i$ to a single one. When the algorithm ends, total production is the same, total cost is strictly less and all production decisions are within the strictly convex region of $C$. Thus social welfare has strictly increased. Next we prove each of these claims about the results of each step.

Operation 1: Eliminating points in the nonstrictly convex region for each strategy.

Suppose that $\bar{\tau}_i^*$ is not in the interior of the strictly convex region of $C_i$ on $R^X_i$. Consider bank $i$ and suppose that $\bar{\tau}_i$ corresponds to a nonstrictly convex point in the cost function for bank $i$.

Now create a convex mixture by projecting $\bar{\tau}_i$ along the direction $\nu_i$ in one direction to the boundary of the hyperplane, and in the other direction onto an axis. See Figure 5. Denote the points of intersection with the hyperplane's boundary and with the axis as $\bar{\tau}_i'$ and $\bar{\tau}_i''$, respectively.

We create a new point above the hyperplane. Select a sufficiently small $\varepsilon>0$, and define $\bar{\tau}_i = \bar{\tau}_i' + \varepsilon \bar{\nu}_i$.

Choose $\lambda \in (0,1)$ such that $\lambda \bar{\tau}_i^* + (1-\lambda)\bar{\tau}_i'' = \bar{\tau}_i', \text{in other words, the production of the convex mixture is equal to the original production. The total production in the market from type } i \text{ banks is}$

\[ m_i \tau_i = m_i (\lambda \bar{\tau}_i' + (1-\lambda)\bar{\tau}_i'') = \lambda m_i \bar{\tau}_i' + (1-\lambda)m_i \bar{\tau}_i''. \]

so that the number of banks producing $\bar{\tau}_i'$ is $\lambda m_i$ and the number of banks producing $\bar{\tau}_i''$ is $(1-\lambda)m_i$. The banks producing $\bar{\tau}_i'$ are interior to the hyperplane's boundary, thus insuring a strictly convex point on the cost function. The banks producing $\bar{\tau}_i''$ are no longer active in a dimension, thus are equivalent to a strategy $i'$ contained in $i$, specifically, $\bar{\tau}_i' \subset \bar{\tau}_i$.

Thus, all firms that are left in the market following strategy $i$ are at a strictly convex point of $C_i$. Note also that the total cost of the banks is now strictly less than the total cost under the old production
Figure 5: Suppose that strategy $i$ uses a production plan $\bar{\tau}_i$ that does not lie in the strictly convex region for cost function $C_i$. Then, we can decrease cost by splitting this production into two new production vectors, $\bar{\tau}_i'$ and $\bar{\tau}_i'''$, where the first vector lies on one of the axes in production space and the other lies in the strictly convex region of $C_i$. We choose the number of firms employing production plans to be $m_i'$ and $m_i'''$ so as to keep total production constant.

level. This reduction comes from two sources: First, as the new production is a convex mixture and the production decisions are respectively, within in a concave region and arbitrarily close to the concave region of the cost function, the convex mix of cost functions is strictly less than the cost of the original production as $\lambda$ was chosen to maintain the original production level. Second, since the banks corresponding to $i'$ are no longer active in some dimension, the fixed cost in the lower dimension is less than the fixed cost of the higher dimension.

Operation 2: Combining several production vectors that all use the same strategy $I$ and lie within the strictly convex portion of $C_i()$.

Now suppose that there are more than one active firms using the same strategy $i$ with all producing within the strictly convex region of the cost function. Choose two production vectors with this property. Any a convex mixture of the production points of each type of bank (adjusting the number of banks to hold total production constant) will produce a new type of bank with lower total
cost and employing the same strategy. It is always possible to make this choice because it is equivalent to solving the following set of equations that choose a \( \lambda \in (0,1) \) and a new number of banks \( m \) such that
\[
m(\lambda \bar{\tau}_i, + (1-\lambda)\bar{\tau}_j) = m_r \bar{\tau}_i + m_l \bar{\tau}_l.
\]
Let \( i' \) represent the banks created from the above procedure and let \( i \) be the banks that were already naturally following that strategy. Choose
\[
\lambda = \frac{m_r}{m_r + m_l}, m = m_r + m_l,
\]
which satisfies the requirements for \( \lambda \in (0,1), m > 0 \), and keeps production constant.

**Proof of Proposition 3.4** Assume \( \bar{m} \) is fixed. Let \( \lambda \in (0,1) \) and \( \bar{\tau}_1, \bar{\tau}_2 \) be any two points such that \( \bar{\tau}_1 \neq \bar{\tau}_2 \) and for each \( i \), lies in the strictly convex portion for function \( C_i() \), i.e., is restricted to be above the hyperplanes defined by (19). As Proposition 1 showed that a social welfare optimizing point has to satisfy equation (19), adding the constraint does not delete any optimal solutions. Consider the function of \( \lambda \), \( SW(\bar{m}, \lambda \bar{\tau}_1 + (1-\lambda)\bar{\tau}_2) \). To show that the optimal \( \bar{\tau} \) is unique in the region defined by region (17), given \( \bar{m} \), it will suffice to show that SW is a strictly concave function of \( \lambda \). Clearly, as the cost functions are convex in this region, the second term of the social welfare function (18) is a sum of concave functions (each is multiplied by \(-1\)), hence, it is concave. It remains to be shown that the first term of \( SW(\bar{m}, \lambda \bar{\tau}_1 + (1-\lambda)\bar{\tau}_2) \) is strictly concave. Consider the second derivative with respect to \( \lambda \) of the first part of the Social Welfare function (which is just):
\[
PSW(\bar{m}, \bar{\tau}_j) = \int_0^{\bar{m} \cdot \bar{\tau}_j} P_j(x) dx.
\]
\[
\frac{d^2 PSW}{d\lambda^2} = \sum_{j \in J} P_j'((\bar{m} \cdot (\lambda \bar{\tau}_j + (1-\lambda)\bar{\tau}_j))((\bar{m} \cdot (\bar{\tau}_j - \bar{\tau}_j))^2.
\]
The first term in the product is negative because the demand function (hence, the inverse demand function) is assumed to be strictly decreasing, and the second term is a squared constant, hence the product is negative. This shows that \( SW(\bar{m}, \lambda \bar{\tau}_1 + (1-\lambda)\bar{\tau}_2) \) is the sum of concave and strictly concave functions hence it is a strictly concave function; thus, for a given vector \( \bar{m} \), the optimal \( \bar{\tau} \) is unique.
From here on, consider \( \tilde{\tau} \) as a function of \( \tilde{m} \), denoted \( \tilde{\tau}(\cdot) \), specifically, the unique optimizing point, in the region defined by (19), given \( \tilde{m} \). Through the rest of this proof, all lower case \( \tau \) are assumed to be functions of the optimal \( \tilde{m} \), but the functional notation is dropped for ease of exposition. Similarly as before, let \( \lambda \in (0, 1) \) and \( \tilde{m}^1, \tilde{m}^2 \) such that \( \tilde{m}^1 \neq \tilde{m}^2 \). Define \( \tilde{m}(\lambda) = \lambda \tilde{m}^1 + (1 - \lambda) \tilde{m}^2 \). As before, it will suffice to show that \( SW(\tilde{m}(\lambda), \tilde{\tau}(\cdot)) \) is a strictly concave function of \( \lambda \), which will imply a unique optimal solution. The first derivative with respect to \( \lambda \) is

\[
\frac{dSW(\tilde{m}(\lambda), \tilde{\tau}(\cdot))}{d\lambda} = \nabla_{\tilde{m}} SW \cdot \nabla_{\lambda} \tilde{m}(\lambda) + \lambda \cdot D_{t_s} \left( SW(\tilde{m}(\lambda), \tilde{\tau}(\cdot)) \right) \cdot D_{\tilde{m}}(\tilde{\tau}(\cdot)) \cdot \nabla_{\lambda} \tilde{m}(\lambda)
\]

(20)

where \( \nabla_{\tilde{m}}(SW) = \left( \frac{\partial SW}{\partial \tilde{m}_1}, \ldots, \frac{\partial SW}{\partial \tilde{m}_i} \right) \) and \( \nabla_{\lambda} \tilde{m}(\lambda) = \tilde{m}^1 - \tilde{m}^2 \). Noting that \( \tilde{\tau}(\cdot) \) is chosen optimally given \( \tilde{m} \) and applying the Envelope Theorem (Varian, 1992) shows that \( D_{t_s} \left( SW(\tilde{m}(\lambda), \tilde{\tau}(\cdot)) \right) = 0 \), so the second term of (20) vanishes. Now consider the first term of (20).

\[
\frac{\partial SW}{\partial \tilde{m}_i} = \sum_{j \in J} P_j \left( \tilde{m} \cdot \tilde{\tau}_j \right) \sum_{k \in K} P_k \left( \tilde{m} \cdot \tilde{\tau}_k \right) \frac{\partial \tau_{jk}}{\partial \tilde{m}_i} - \frac{\partial \tau_{jk}}{\partial \tilde{m}_i} \sum_{j \in J} \left( \frac{1}{m_{ik}} \right) P_j \left( \tilde{m} \cdot \tilde{\tau}_j \right) \frac{\partial \tau_{jk}}{\partial \tilde{m}_i}
\]

Note that the third term of the second line is the familiar full price and marginal cost terms. In choosing the optimal \( \tilde{\tau}(\cdot) \), if bank \( i \) serves segment \( j \) with technology \( k \), then price must equal marginal full cost, so the term is zero. The following is the result.

\[
\frac{dSW(\tilde{m}(\lambda), \tilde{\tau}(\cdot))}{d\lambda} = \sum_{i \in I} \left( \sum_{j \in J} \sum_{k \in K} P_j \left( \tilde{m} \cdot \tilde{\tau}_j \right) \frac{\partial \tau_{jk}}{\partial \tilde{m}_i} \left( m_{ik} \right) - \frac{\partial C_i}{\partial \tau_{jk}} \frac{\partial \tau_{jk}}{\partial \tilde{m}_i} \right)
\]

Note that taking the second derivative with respect to \( \lambda \) will result again in terms involving full price and marginal full cost. As before, those terms vanish leaving the following

\[
\frac{d^2 SW(\tilde{m}(\lambda), \tilde{\tau}(\cdot))}{d\lambda^2} = \sum_{i \in I} \sum_{j \in J} P_j \left( \tilde{m} \cdot \tilde{\tau}_j \right) \frac{d\tilde{m} \cdot \tilde{\tau}_j}{d\lambda} (m_{i}^1 - m_{i}^2)
\]

(21)

Note that

\[
\frac{d(\tilde{m} \cdot \tilde{\tau}_i)}{d\lambda} = \sum_{j \in J} \frac{d(\tilde{m} \cdot \tilde{\tau}_j)}{dm_i} (m_{i}^1 - m_{i}^2)
\]

(22)
Consider the following second partial derivatives,

\[
\frac{\partial^2 PSW}{\partial m_i \partial m_j} = P'_j(m \cdot \tau_j) \frac{d(m \cdot \tau_j)}{dm_j} \tau_j
\]

\[
\frac{\partial^2 PSW}{\partial m_i \partial m_j} = P'_i(m \cdot \tau_i) \frac{d(m \cdot \tau_i)}{dm_i} \tau_i
\]

Equating the cross partials gives the following relationship in equation (23)

\[
\frac{d(m \cdot \tau_j)}{dm_j} = \frac{1}{\tau_i} \frac{d(m \cdot \tau_i)}{dm_i} \tau_j
\]

(23)

Substituting (23) into (22) and then into (21) produces

\[
\frac{d^2 SW(m(\lambda), \bar{\tau}(\lambda))}{d\lambda^2} = \sum_{j \neq i} P'_j(m \cdot \tau_j) \left( \frac{1}{\tau_j} \frac{d(m \cdot \tau_i)}{dm_i} \right) \left( \sum_{i \neq j} \tau_j (m_i - m_i^0) \right)^2 < 0
\]

The first term in the product is strictly negative because the inverse demand function is assumed to be strictly decreasing. The second term in the product is positive by choosing \( i' \) such that \( \tau_{i'} > 0 \) and noting that total supply to customer segment \( j \) must increase as the number of firms of a given type increases because total variable cost declines with more firms. Finally, the third term is squared, hence always positive. Thus, the second derivative with respect to \( \lambda \) is negative, implying that \( SW \) is strictly concave with respect to \( \lambda \), hence \( SW \) is strictly concave everywhere. Thus, by solving for \( \bar{\tau} \) as a function of \( \bar{m} \), and then optimizing with respect to \( \bar{m} \), a unique optimal solution is obtained.