

# BANKING MARKETS: PRODUCTIVITY, RISK, AND CUSTOMER SATISFACTION

Gerald R. Faulhaber \*

## Abstract

Recent empirical research in banking has found substantial but unexplained differences among bank's costs and profits, indicating that the industry is not in long-run equilibrium. In this paper, three factors are hypothesized as the source of this out-of-equilibrium behavior: (i) variations in banks' abilities to match capacity and demand; (ii) variations in risk management; and (iii) variations in providing customer satisfaction. A structural model is combined with new data sources to estimate these effects using data from 1984-1992. Each factor is found to be significant, and taken together they account for virtually all the observed differences in bank performance.

**Keywords:** banking, structural estimation, quality, productivity, efficiency, risk

---

\* Professor, Department of Public Policy and Management, Wharton School, University of Pennsylvania. This research was supported by a grant from the Wharton Financial Institutions Center funded by the Sloan Foundation and Deloitte & Touche. The author wishes to thank Larry Brown, Kermit Daniel, and Dan Raff as well as participants in the Financial Institutions Center seminars for their comments and suggestions. An anonymous referee provided insightful and very useful comments. Jalal Akhavein and Tony Cai provided invaluable research assistance.

# BANKING MARKETS: PRODUCTIVITY, RISK, AND CUSTOMER SATISFACTION

## 1. Introduction

Much of the recent empirical literature in banking has focused on the rather large differences in X-efficiencies (in the sense of Leibenstein (1966)) among U.S. banks. Traditionally, empirical research in banking examined the presence or absence of economies of scale and scope; this more recent work finds that differences in cost unrelated to scale and scope are of much greater magnitude. This research suggests that “X-inefficiencies account for ... 20% or more of costs in banking, while scale and product mix inefficiencies...account for less than 5% of costs” (Berger, Hunter, and Timme (1993)). These X-inefficiency results have been obtained using various frontier estimation methods on both bank cost functions (see Berger and Humphrey (1991) and profit functions (see Berger, Hancock, and Humphrey (1993), for more powerful results), using cross-section/time series data on U.S. banks. For a review of these frontier methods as applied to banking, see Berger (1993) and more recently Berger and Humphrey (1997).

These somewhat surprising results have refocused the literature away from scale and scope toward explaining both the existence and the magnitude of these X-inefficiencies. Kaparakis, Miller, and Noulas (1994) use a stochastic frontier approach and find that bank size, in particular the number of branches, has a negative impact on short-run efficiency. Mester (1994) uses cost frontier methods to assess the efficiency of banks in the Federal Reserve’s third district, finding significant X-inefficiencies. Grabowski, Rangan, and Rezvanian (1994) use a nonparametric frontier approach to assess the effect of deregulation on bank efficiency, finding a decline over the years 1979-1987, principally focused on technical rather than allocative inefficiency. Overall, these papers have identified potential X-inefficiencies as substantially greater than either scale or scope effects, and have thus turned scholarly attention toward identifying possible sources for these large efficiency differences among banks. For example, Mester (1993) has examined efficiency differences between stock and mutual savings and loan institutions, finding that the mutual form of ownership is associated with greater efficiency. More recently, Berger and Mester (1997) have attempted to examine the sources of these inefficiencies, which is the spirit of this paper as well.

This paper rests on three fundamental predicates: (i) the large and persistent differences in costs/profits constitute strong evidence that the banking industry is not in long-run competitive equilibrium<sup>1</sup>; the

question I address is, What choices are banking managers making that lead to these inefficiencies? (ii) The econometric evidence has been largely in the form of fixed effects models or frontier methods, which identify *that* there are large and persistent differences but not *what explains* these differences; I introduce specific hypotheses to explain these differences<sup>2</sup>. (iii) I use a *structural* model to test these hypotheses and derive results, based on a model of bank and market behavior. The nonlinear estimating equations are derived from the market equilibrium conditions, and are not a reduced form approximation. The unique contribution of this paper is to use structural models and new datasets to test hypotheses that *explain* the source of the X-inefficiencies previously observed<sup>3</sup>, and to determine if the hypothesized explanations explain the magnitude of the previous observed X-inefficiencies.

Three hypotheses are tested regarding the source of the cost/profit inefficiencies:

1. Bank managers differ in their abilities to plan their capacity to match realized demand.
2. Bank managers differ in their abilities to manage risk.
3. Bank managers differ in their abilities to provide customer satisfaction.

A fourth hypothesis tested relates to the power of the structural model to explain the X-inefficiencies previously observed.

4. The structural model reflecting the above bank and market behavior captures virtually all the variation previously observed as X-inefficiencies.

There is ample evidence from micro institutional research to suggest that in the recent past, banks differ markedly in each of these three areas. Frei (1996) reports, as part of a larger study involving interviews and field work at a substantial number of US banks, that institutions differ quite strongly in their ability to effectively use computing capacity, a relatively large component of their fixed cost. Santomero (1995) surveyed bank practices in risk management, reporting that not only do banks differ strongly in their ability to manage risk, but most have not availed themselves of readily available “best practice” technologies in this area. Lastly, Greenwich Associates conducts annual surveys of the commercial customers of banks (which we use below) and consistently finds significant differences in customer satisfaction among US banks.

The key to understanding the results of this paper is that, while the product markets are explicitly assumed to be in equilibrium, the market for managerial control is explicitly assumed not to be in equilibrium, so that systematic differences among banks in the three factors listed above are not assumed to be equilibrated away.<sup>4</sup> I do not seek to “explain” the differences in these factors as an equilibrium outcome, but rather to identify which of these factors account for the observed out-of-equilibrium performance

differences among banks. The estimation of the model parameters enables us to examine each of the three hypotheses above. Our conclusion is that each of these factors is an important contributor to differences among the performance of banks.

However, a stronger conclusion can be drawn from the analysis. It can be shown that by estimating a structural model that reflects these three factors, virtually all of the profit/cost differences among banks identified in previous research are accounted for. To show this, I estimate the structural model both with and without fixed effects, and test the restriction that all fixed effects are zero, which I find cannot be rejected, indicating that the structural model accounts for much of the variation that fixed effects previously accounted for. In order to ensure that this result was truly due to the structural model and not this particular dataset, I estimated a reduced-form quadratic equation both with and without fixed effects, and tested the restriction that all fixed effects are zero, which was strongly rejected. Thus, I conclude that the structural model statistically explains the X-inefficiencies<sup>5</sup> that were previously identified but not explained by reduced form frontier estimation.

The paper is organized as follows. Section 2 lays out the model. Section 3 uses these results to derive the estimating equations. In Section 4, the data is described, including how various anomalies in the several datasets were treated. Section 5 contains the results of the estimation, and Section 6 lists our conclusions as well as possible future research along the lines developed in this paper.

## 2. The Model

Banks sell up to six products to customers in a geographically limited area, which I assume to be a Metropolitan Statistical Area (MSO).<sup>6</sup> The six products are: (i) demand deposits; (ii) time deposits; (iii) commercial and industrial (C&I) loans; (iv) consumer loans (including credit card accounts); real estate loans (including both commercial and residential mortgages); and (vi) off-balance-sheet items (including all counterparty guarantees, such as letters of credit, but not futures contracts, such as swaps).

At the beginning of each period, bank managers estimate the demand for the bank's products in the coming period, and decide how much capacity to provide for the coming period. I take a broad view of "capacity" as any cost which the bank is unwilling to adjust after demand is realized in the subsequent period. This would certainly include bank branches and computer hardware, but it also may include employees such as credit officers, executives, and others who do not face layoffs or a short schedule in the event that expected demand is not realized. In this view, a cost is fixed if the bank treats it as fixed, and is not necessarily associated with physical assets on the balance sheet. This capacity is assumed to be bank-specific but not product-specific. Demand forecasts are assumed to depend upon private information as

well as the ability of managers; thus, banks enter the product markets with different capacities and therefore different variable costs, with higher capacities implying lower variable operating costs. In choosing capacity, the bank is choosing an operating (short-run) cost function for the coming period with imperfect knowledge of demand.

After choosing capacity levels, each bank then faces its actual demand, which in general will differ from their “planned-for” demand. However, they cannot change their operating capacity or variable costs in the short run. Each bank then optimally chooses demand in each product market, based upon its variable operating costs, its risk cost, and the level of customer service characteristic of that bank. In each MSO, there are six product markets, with each bank within the MSO operating in all six markets. The equilibrium concept is Cournot-Nash, in which each bank can earn positive profits. At the end of the period, the game is repeated, with banks choosing capacity for the next period. In the model, there is no linkage between periods.

The timing of the game is as follows:

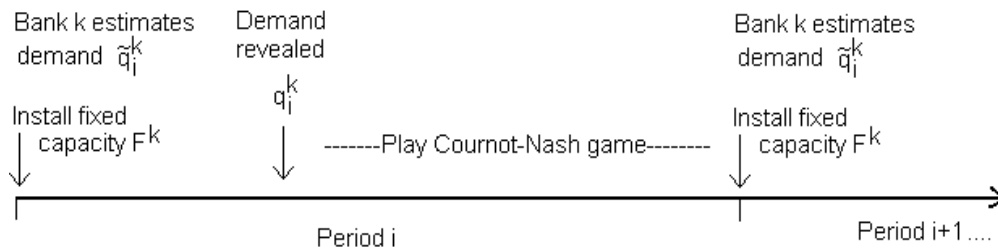


Figure 1

The technology is defined by a class of (short-run) cost functions  $C^k(\mathbf{q}, \mathbf{r}, \mathbf{X}; \xi)$  indexed by the (vector) parameter  $\xi$ , and which depend upon the bank  $k$ , the vector of quantities  $\mathbf{q}$ , a vector of risk characteristics in each market  $\mathbf{r}$ , and a vector of the bank’s customer satisfaction level  $\mathbf{X}$ . Note that (i) costs are bank-specific, so that differences in bank costs are modeled explicitly; (ii) each bank makes a choice of the cost function it wishes to face by selecting from the class defined by  $\{\xi\}$ , the index parameter; and finally, (iii) input prices are not include in the cost function. At this point, this is assumed for expositional convenience; it is shown below that empirically, real input prices are constant across banks and time periods, so they need not enter into the analysis. The cost function is assumed to be separable into (i) operating cost  $C_o$ , which depends upon quantities in each market  $\mathbf{q}$  and the choice of index parameter  $\xi$ , (ii) a bank-specific risk cost  $C_r$ , which depends upon risks in each market  $\mathbf{r}$ , and (iii) a single bank-specific

but not market-specific customer satisfaction measure  $X$ :  $C^k(\mathbf{q}, \mathbf{r}, \mathbf{X}; \xi) = C(C_o(\mathbf{q}; \xi), C_r^k(\mathbf{r}), X^k)$ . For ease of exposition, each component of this cost function is discussed in turn. I start with the family of operating cost functions, then add the risk cost and customer satisfaction cost.

**Operating Cost** Each bank has a choice of which short-run operating cost function  $C_o(\mathbf{q}; \xi)$  to adopt in each period. The (vector) index parameter  $\xi$  reflects the usual trade-off between fixed and variable costs. As usual, the envelope of short-run cost functions is the long-run cost function  $c_o(\mathbf{q})$  with the property that for every output vector  $\tilde{\mathbf{q}}$  there exists a short-run cost function which is both equal to and tangent to the long-run cost function. That is, there exists a unique  $\xi$  for which  $C_o(\tilde{\mathbf{q}}; \xi) = c_o(\tilde{\mathbf{q}})$ , and  $\nabla C_o(\tilde{\mathbf{q}}; \xi) = \nabla c_o(\tilde{\mathbf{q}})$ . Since there is a one-to-one relationship between  $\xi$  and  $\tilde{\mathbf{q}}$ , the latter becomes a natural index parameter for the short-run cost function. That is,  $C_o(\mathbf{q}; \tilde{\mathbf{q}})$  is that short-run cost function which is equal to and tangent to the long-run cost function at the output vector  $\tilde{\mathbf{q}}$ . The actual family of cost functions used for estimation purposes is developed in Appendix A.

This model enables the characterization of bank manager behavior regarding the first hypothesized factor: the ability to match capacity and demand. If there is little or no error in estimating demand, then the cost-minimizing bank manager would equate actual demand with planned-for demand:  $q = \tilde{q}$ . If there is significant error in estimating demand, then we might expect that the condition of cost-minimization is  $E[q] = \tilde{q}$ . However, this is not the case; the efficient manager will minimize the expected cost of error, not the expected error. In the case where this cost is symmetric, these are equivalent. However, if the cost of errors is asymmetric, then efficient managers would either provide either more capacity than expected demand or less, depending on the cost asymmetry.

It is assumed that each bank bases its choice of planned-for demand  $\tilde{\mathbf{q}}^k$ , and therefore its short-run cost function, on private (and possibly some public) information and that this choice is not observable to other banks. Thus, each bank may choose a different short-run cost function, and that choice cannot be used to signal to its competitors. Of course, some bank managements may be better at this planning function which could be a source of competitive advantage for successful banks and X-inefficiencies for less successful ones.

**Cost of Risk** In this analysis, the cost of risk is measured rather than modeled. What is unique in this paper is the use of a market-based risk measure: the stock market  $\mathbf{b}$  for each bank in each time period. Other studies have focused on credit risk, using either statistical models or internal accounting data. For example, bank risk was examined in McAllister and McManus (1993), where an inventory model is used

to assess scale economies associated with risk pooling. Hughes and Mester (1994) model bank managers' risk preferences, using the leverage ratio as their risk measure. Additionally, Hughes and Mester (1993) assess the impact of the "too-big-to-fail" doctrine on bank risk-taking, a matter we consider in this paper.

Using a bank's stock market  $\beta$  as the measure of risk has several compelling positive features: (i) it is a market measure, and so reflects the best judgment of those whose money is actually at risk; (ii) it reflects all the risks taken by the bank, including business risk (such as the risk associated with opening new branches, changing lines of business, etc.) as well as the more usual credit risk; (iii) it is prospective in nature, rather than historical; it measures how investors perceive the variation of future returns<sup>7</sup>, not how past returns have varied. It also has some negative aspects: (i) it does not reflect total risk (in the sense of variance of earnings) but rather covariation of returns with the market; this is the appropriate measure of risk to the bank's owners, but perhaps not for other purposes; (ii) the presence of deposit insurance socializes some of the risk to shareowners, which would therefore not be picked up in  $\beta$ . Although shareowners are presumably not protected by deposit insurance, bank bailouts and forced mergers are beneficial to shareowners and thus reduce at least some of the risk they face; (iii) banks are not covered by GAAP accounting, and the information available to investors about banks is arguably less than for other firms. Thus, market judgments regarding bank riskiness may be less reliable than for other firms; (iv) market  $\beta$ 's are only available at the banking holding company level, and not at the state bank level, which is the unit of analysis of this study; (v) the market  $\beta$  is a dimensionless number representing risk and cannot be used directly in a cost analysis.

On balance, the positive aspects of using  $\beta$  as the measure of risk appear to outweigh the negative aspects. Focusing on risk that the shareowners bear is in fact exactly what bank managers ought to be doing; to the extent they are not, they are inefficient. Shareowner risk, therefore, is exactly the correct measure of bank risk management. The fact that banks produce less publicly available accounting data is unlikely to bias the estimate of  $\beta$ , although it may affect how noisy the estimate is, which would not alter the results of this paper. The fact that  $\beta$ 's are only available at the bank holding company level represents a relatively serious estimation problem. The approach taken in this paper is to impute the holding company  $\beta$  to each of its bank subsidiaries. In defense of this approach, I note that a holding company is unlikely to pursue radically different risk management policies in different subsidiaries. However, it is also the case that a bank holding company may own non-bank businesses, so that imputing its  $\beta$  to its state banks may introduce error into our estimation. Lastly, a transformation is developed below that defines a cost (\$/yr) of risk based on the observed stock market  $\beta$ , consistent with the CAPM model.

Costs are normally denominated in dollars per unit time, while the stock market beta is a dimensionless quantity. To incorporate risk into the cost function framework, a transform of the stock market  $\beta$  is

derived which measures the cost of risk commensurate with all other costs. I define a bank's risk cost as the difference between its current earnings stream and what the earnings of the bank must be to achieve the same value of the firm if the bank carried no (nondiversifiable) risk.<sup>8</sup> The risk cost is derived in Appendix A.

**Customer Satisfaction** Banks may differ in their ability and willingness to satisfy customers with their service. Recent reports in the trade press suggest that customer satisfaction may be a highly profitable strategy for banks. "Financial marketers appear to have overlooked the fundamental truth that the longer an institution keeps a customer, the more profitable the customer becomes. [Banks need ]...to maximize their satisfaction with [the] institution," according to Vavra (1995). Recent work has attempted to measure the impact of service quality on bank profits, such as Roth and Jackson (1995) and Soteriou and Zenios (1997), who develop a strategic framework for service quality. In this paper, I employ the related concept of customer satisfaction, which may be thought of as the end objective of service quality.

As with risk, customer satisfaction is measured rather than modeled. The customer satisfaction measure I use is an index derived from an extensive and long-standing survey of commercial bank customers conducted by Greenwich Associates, Inc., a market research firm specializing in the commercial banking sector. The index merges survey responses into a single number  $X^k$  for bank  $k$ , normalized between 1 and 100, with a higher index number corresponding to greater customer satisfaction.

The details of the survey and the construction of the index from the survey instrument are discussed below in Section 4. We note here the key features of the index that validate its use in this study as a measure of customer satisfaction: (i) actual bank customers are surveyed to assess their opinions regarding their degree of satisfaction with their banking relationships. (ii) The survey has been conducted annually for over twenty years; the index is therefore based on a set of questions and responses that have been consistent over the period of our study. (iii) The survey results are considered informative by the banks themselves, as attested to by the fact that Greenwich Associates' principal source of income over this period has been the sale of the survey results to individual banks. The validity of the results has thus passed a "market test."

The customer satisfaction measure may be modeled as either a choice variable of the bank, or a characteristic of the bank over which it has little or no control in the short term. In this paper, we treat customer satisfaction as a characteristic of the firm rather than a choice variable. The principal justification for this assumption is empirical: if customer satisfaction were a choice variable, then in equilibrium profits would be identical for all levels of customer satisfaction. Some banks would offer higher service levels and greater satisfaction, and some would offer less, but all would earn approximately

the same profits. In fact, this is not the case; as we shall see below, profits vary considerable with customer satisfaction, likely to account for the great practitioner interest in this issue.

Customer satisfaction  $X$  is assumed to affect operating costs for each product by the factor  $X^d$ . For  $d > 0$ , operating costs are increasing in customer satisfaction, and the converse for  $d < 0$ .

**The Total Cost Function** Constructing an explicit cost function which captures all of the above elements of the model into a form suitable for equilibrium analysis and empirical estimation is a tedious task which I relegate to Appendix A. At this point we define the relationship of total cost to the three components: operating cost, risk cost, and customer satisfaction cost, without explicit exposition of the functional form of operating cost. The short-run cost function is

$$C^k = (X^k)^d C_o(\mathbf{q}; \tilde{\mathbf{q}}^k) + R^k, \quad (1)$$

where  $\tilde{\mathbf{q}}^k$  is bank  $k$ 's planned-for demand and  $R^k$  is the risk cost for bank  $k$ . The long-run cost function is

$$c = (X^k)^d c_o(\mathbf{q}) + \sum_{i=1}^6 \bar{r}_i^k q_i,$$

where the last terms are the linear decomposition of risk to products (see Appendix A) Each bank  $k$  chooses its planned-for demand  $\tilde{\mathbf{q}}^k$  which determines its fixed cost and variable cost coefficients  $\mathbf{c}$ . This mapping from planned-for demand to the fixed and variable costs solves the tangency and equality conditions of the short-run and long-run cost functions, as explained above.

**Demand** Demand analysis is not the focus of this paper; however, estimates of demand elasticities are essential to deriving equilibrium conditions in imperfectly competitive markets. Further, customer satisfaction levels are also likely to affect demand and therefore need to be explicitly modeled. A simple (inverse) demand system is assumed that allows for interdependent demands:

$$p_i = (X^k)^v A_i \prod_{j=1}^6 Q_j^{h_{ij}}, \quad i = 1, \dots, 6. \quad (2)$$

where the matrix  $\mathbf{h}$  is symmetric and the parameter  $Z$  is the price elasticity with respect to customer satisfaction. This demand system has the property that the self- and cross-flexibilities  $\frac{Q_j}{p_i} \frac{\partial p_i}{\partial Q_j} = h_{ji}$  are

constant. The more familiar demand self- and cross-elasticities can be obtained from the flexibilities by inverting the flexibility matrix, and they are (of course) constant as well:  $h^{-1} = e$ , the elasticity matrix.

Each firm  $k = 1, \dots, m$  supplies a portion  $\mathbf{q}^k = (q_1^k, \dots, q_n^k)$  of this total demand, with  $\mathbf{Q} = \sum_{k=1}^m \mathbf{q}^k$ .

**Product Market Equilibrium** I assume that the number of banks in the market is fixed. In the case at hand, metropolitan statistical areas (MSA) are assumed to be the relevant banking markets. At the beginning of the period before demand is revealed, banks plan for demand  $\tilde{\mathbf{q}}^k$  by selecting a short-run cost function which they cannot change after demand is revealed. After the market opens and demand is revealed all banks in the market then play a Cournot-Nash quantity game, now with full knowledge of demand, in which they offer the quantity vector  $\mathbf{q}^k$  (not necessarily equal to  $\tilde{\mathbf{q}}^k$ ) in the market.

Bank  $k$  chooses the quantity vector  $\mathbf{q}^k$  to maximize *economic* profit. Using equations (A11) and (A10), this can be written as:

$$\max_{q_i^k} p^k = (X^k)^V \sum_{i=1}^n p_i(\mathbf{Q}) q_i^k - (X^k)^d C_o(\mathbf{q}^k; \tilde{\mathbf{q}}^k) - \sum_{i=1}^n \bar{r}_i^k q_i^k \quad (3)$$

The first order conditions are:

$$\frac{\partial p^k}{\partial q_i^k} = MR_i^k - MC_i^k = (X^k)^V p_i \left[ 1 + \left( \sum_{j=1}^n s_{ji} f_{ji} q_j^k \right)^{-1} \right] - (X^k)^d \left( \frac{\partial C_o(\mathbf{q}^k; \tilde{\mathbf{q}}^k)}{\partial q_i^k} \right) - \bar{r}_i^k = 0,$$

where

$$s_{ji} = \frac{p_j Q_j}{p_i Q_i}, \quad f_{ji} = \frac{Q_i}{p_j} \frac{\partial p_j}{\partial Q_i} = h_{ji}, \quad q_j^k = \frac{q_j^k}{Q_j},$$

which are, respectively, the share of product  $j$  total market revenues relative to product  $i$  total market revenues, the flexibility of price  $j$  with respect to quantity  $i$  (constant for the assumed functional form of the demand system), and the market share of firm  $k$  for the  $j^{\text{th}}$  product.<sup>9</sup>

This first-order condition can be put into this more familiar form:

$$\frac{p_i - (X^k)^{d-V} \left( \frac{\eta_{C_o}(\mathbf{q}^k; \tilde{\mathbf{q}}^k)}{\eta_{q_i^k}} \right) - \bar{r}_i^k}{p_i} = \sum_{j=1}^n s_{ji} h_{ji} q_j^k, \text{ for } i = 1, 2, \dots, 6, \text{ and all banks } k. \quad (4)$$

For given prices  $\mathbf{p}$ , planned-for demand  $\tilde{\mathbf{q}}^k$ , customer satisfaction index  $X^k$ , and marginal risk cost  $\bar{r}_i^k$ , bank  $k$  determines its optimal quantities  $\mathbf{q}^k$  as the solution to the 6×6 equation system (4). Assuming  $C_o$  is quasiconcave, the equilibrium, if it exists, is unique.

However, equation (4) has a different purpose for this research; for given price  $\mathbf{p}$  and observed quantities  $\mathbf{q}^k$ , the planned-for demand  $\tilde{\mathbf{q}}^k$  can be imputed. Thus, equation (4) is the basis for estimating each bank's planned-for demand, and therefore the differences between actual and planned-for demand. This is an essential step in assessing whether or not mismatches between demand and capacity are a significant source of the bank inefficiency found by earlier researchers.

### Limitations of the Model

The form of the game suggests three questions: (i) is the Cournot-Nash assumption reasonable for these markets? (ii) Are banks the only players in these six product markets? (iii) Is the MSA the relevant market? In fact, the Cournot quantity game with a fixed number of players is rather well-suited to banking markets. Entry into these markets is restricted (though not prohibited) by government regulation, so that the “no entry “ assumption is a reasonable though not perfect approximation to reality, especially for larger banks. Prices tend to be market-determined, and are often strongly affected by capital market activity over which banks have little effect. By choosing capacity, such as size of branch network, extent of trading activities, number of credit officers hired and trained, banks are choosing quantities to offer the market, coincident with the Cournot quantity game.

More troubling is the assumption that the banks in the sample are the only players in the game. In fact, there are many more players in these markets, including smaller banks, non-bank financial institutions, and commercial paper markets. In virtually all of the products in this model, banks in our sample compete with institutions not in our sample. Lack of available data on these non-bank institutions constrains all researchers from addressing this issue. The results of this paper, as well as the results of virtually the entire empirical banking literature, must be understood in light of this deficiency.

Another concern is our assumption that the relevant market is the MSA. For retail products and many middle market products this is a good assumption. However, for some services such as corporate loans the market may well be national. This suggests that the market share of banks is less than that used here, so that price-cost margins would also be less.

### 3. Estimation

The model of the previous section has the following parameters to be estimated: (i) planned-for demands  $\tilde{q}^k$ ; (ii) the customer satisfaction measure elasticities of cost and demand,  $d$  and  $Z$ ; (iii) any remaining fixed effects; (iv) the product-specific risk coefficients  $\bar{r}^k$ ; and (v) the parameters of the operating cost function. I discuss each in turn.

**Planned-for Demand** Estimating planned-for demand for each bank and each time period clearly is unrealistic, since this would consume almost half the degrees of freedom available from the data. In fact, the hypothesis to be tested concerns the *mean* deviation of planned-for from actual demand for each bank. I assume, then, that the ratio of planned-for to actual demands is fixed for each bank, equal to the bank's *bias* parameter:

$$m^k = \frac{\tilde{q}_i^k}{q_i^k}, \text{ for } i=1,\dots,6, \text{ and all time periods.} \quad (5)$$

The bias  $m^k$  is a bank-specific parameter to be estimated. Note that the optimal bias parameter need not be unity. Systematic differences of planned-for and actual demand may be due to nonlinear costs. On the other hand, bank managers may exhibit an “optimism” or “pessimism” bias, consistently over- or under-estimating demand,<sup>10</sup> in which case these mismatches represent allocative inefficiency. An objective of this analysis is to determine the extent of allocative inefficiency due to this source.

**Customer Satisfaction Elasticities** Unfortunately, the Greenwich Associates customer satisfaction measure is based only upon commercial bank customers. This suggests that the customer satisfaction measure is unlikely to apply to all products of the bank, which is confirmed in the empirical analysis. Therefore, we restrict the customer satisfaction coefficients to the terms in the cost and revenue functions specific to C&I loans, the product which most closely represents the Greenwich survey's target market.

**Fixed Effects** It will be recalled that much of the motivation for this research comes from the finding of previous work that there are quite large unexplained differences among banks' cost functions and profit functions. This work is designed to model and measure three specific hypotheses regarding bank inefficiencies, in order to rigorously explain what others have found. To the extent this effort is successful, the importance of fixed effects should diminish or disappear from the empirical analysis.

Fixed effects are therefore explicitly included into the structural equation estimation to determine if this occurs.

**Risk coefficients** Estimation of the risk coefficients is straightforward, using equation (A11).

**Operating Costs** The results of Appendix A are used to derive the operating cost estimating equation. Substituting those results into equation (1), the bank's profit function is

$$\rho^k = \mathbf{p} \cdot \mathbf{q}^k - \sum_{i=1}^n c_i^k (q_i^k)^{2a_{ii}} - \sum_{i=1}^n \sum_{j=i}^n n_{ij} (q_i^k q_j^k)^{a_{ij}} - F(m^k \mathbf{q}^k; \mathbf{c}^k, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{v}) - \sum_{i=1}^k \bar{r}_i^k q_i^k \quad (6)$$

where the functions  $c_i^k$  and  $F^k$  are given in Appendix A.

**Estimating Equations** Economic profit cannot be observed from the books of the bank, as it includes risk costs, among other possible costs not accounted for. We assume in this work that risk costs are the only costs that do not appear on banks' books, so that economic profit is equal to accounting earnings (which is observable) less risk cost:  $\pi^k = \hat{\rho}^k - R^k$ .  $R^k$  in turn depends upon market observables  $r_f$ ,  $r_m$ , and  $b^k$ . In addition, we also include a dummy variable  $B^k$  for each bank. These are the fixed effect coefficients, designed to assess any bank-specific effects not picked up by the parameters of the model. In addition, the customer satisfaction index is suppressed in the estimating equations; this is discussed below in the section on customer satisfaction estimation.

There are thus two estimating equations, one for risk and the other for operational activities. We use actual bank earnings  $\hat{\rho}^k$  as a proxy for uniform expected earnings:

$$R^k = \hat{\rho}^k \left( \frac{b^k (r_m - r_f)}{r_f + b^k (r_m - r_f)} \right) = \sum_{i=1}^6 \bar{r}_i^k q_i^k \quad (7)$$

$$\hat{\rho}^k = \mathbf{p} \cdot \mathbf{q}^k - \sum_{i=1}^n c_i^k (q_i^k)^{2a_{ii}} - \sum_{i=1}^6 \sum_{j=i}^6 n_{ij} (q_i^k q_j^k)^{a_{ij}} - F(m^k \mathbf{q}^k; \mathbf{c}^k, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{v}) + B^k \quad (8)$$

Everything in these two equation is either observable ( $\hat{\rho}^k$ ,  $r_f$ ,  $r_m$ ,  $b^k$ ,  $\mathbf{p}$ ,  $\mathbf{q}^k$ ), is a parameter ( $\boldsymbol{\alpha}$ ,  $\boldsymbol{\gamma}$ ,  $\mathbf{v}$ ,  $\boldsymbol{\lambda}$ ,  $\mathbf{m}$ ,  $\mathbf{B}$ ), or is derived from observables and parameters ( $\mathbf{c}^k$ ,  $F^k$ ). Therefore, it can be estimated using nonlinear methods. Note that these two equations are not simultaneous.

## 4. The Data

Three different datasets were employed in this study: *operating* data, *capital market* data, and *customer satisfaction* data.

***Operating Data*** This data is taken from the Report of Condition and Income (“Call Report”), which all insured banks operating in the US are required to file with Federal regulators. We use quarterly data, from 1984 to 1992 inclusive, for all banks with over \$1 billion in assets in 1984. The variables in this dataset include:

*quantities*; end-of-period balance sheet entry for each bank for each of the six products listed;

*revenues*; total period net revenue for each bank for each of the six products listed;

*earnings*; income before extraordinary items and after taxes for each bank.

All quantities are expressed in thousands of 1982 dollars.

This data is collected from the banks themselves by the FDIC, which maintains the dataset. It is *accounting* data.

***Capital Market Data*** This data is taken from the CRSP (Center for Research in Stock Prices) dataset, which contains end-of-day prices of stocks traded on major exchanges and other securities, such as bonds. The variables in this dataset include:

*market rate of return*; this is the total return (dividends plus capital gain) for the NYSE.

*risk-free rate of return*; this is the interest rate on 90-day Treasury bills.

*beta*; the  $\beta$  of each bank for each quarter was computed using the actual price data together with the market rate and the risk-free rate.

This data is collected from US stock markets by the Center for Research in Stock Prices at the University of Chicago. It is *market* data.

Not all banks had data which covered the entire period. Indeed, many banks were merged into new banks during this period. Since there is nothing in the model that is inherently dynamic, these mergers did not constitute a problem; as of the date of the merger, the old banks dropped from the dataset and the new one was inserted. The total number of data points (banks  $\times$  quarters) is 6190.

**Customer Satisfaction Data** This data is taken from an ongoing survey conducted by Greenwich Associates, Inc., a marketing research firm that has been conducting surveys of commercial customers of US banks since 1972. The survey is designed to elicit from customers (i) their degree of satisfaction with the banks they do business with; and (ii) the specific factors and attributes important to them, and how their banks fared on these items. From this extensive survey data, Greenwich Associates constructed a single index, expressly for this study, to measure overall customer satisfaction for each bank. The details of the survey methods and the construction of the index is contained in Appendix C. The key points about this index: (i) it is scaled to range from 0 to 100, with a mean of 50; higher scores correspond to greater customer satisfaction; (ii) only commercial customers are surveyed; thus, the customer satisfaction index only applies to commercial products, in particular C&I loans; (iii) due to data limitations, only 112 banks (from our larger sample) are included, during the period 1985 to 1992. The surveys are conducted annually (at most), so quarterly data is not available. The total number of data points (banks  $\times$  years) is 476.

## 5. Empirical Results

### Hypotheses Testing

**Risk** In long run equilibrium, some banks may take on more risk than others if they are rewarded in terms of higher earnings. However, there would be little variation in the risk-cost/earnings ratio among banks. If however this were a source of inefficiency, then rather large variations in this ratio would be observed. To test this, I regress computed bank risk cost against bank-specific dummy variables. This regression is significant ( $F = 1.575$ ,  $df = 219, 5971$ ,  $Prob = 0$ ); the hypothesis of zero cross-bank effects can be rejected. I conclude that *there are significant differences across banks in their ability to manage risk.*

**Demand-Capacity Mismatch** In long run equilibrium, all banks would match demand and capacity to minimize cost on average. The bias parameters of banks would show little variation, and would differ from unity only to the extent cost minimization required it. If however this were a source of inefficiency, then large variations in this parameter would be observed. To test this, I regress the operating cost equation using the slope dummy variables  $m^k$ , and then impose the restriction that all  $m^k$  are identical. The  $F$ -test ( $F = 1.944$ ,  $df = 219, 5857$ ,  $Prob = 0$ ) rejects the restriction of identical bias coefficients. I conclude that *there are significant differences across banks in their ability to plan for capacity.*

**Customer Satisfaction** In long run equilibrium, some banks would opt for higher customer satisfaction and some would choose lower customer satisfaction, but profits would equilibrate across customer satisfaction choices. If however customer satisfaction were a source of performance differences across banks, then

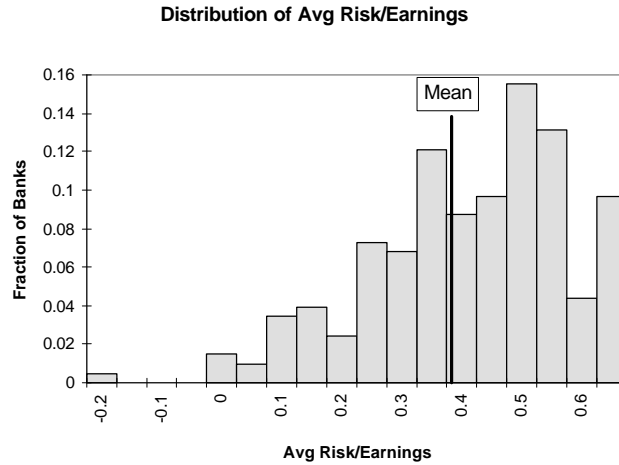
profitability would vary across bank customer satisfaction. I regress customer satisfaction on bank-specific dummies; the regression is highly significant ( $F = 9.144874$ ,  $df = 24, 88$ ,  $Prob = 0$ ), rejecting the restriction that all banks have the same level of customer satisfaction. Therefore, I conclude that *there are significant difference across banks in their ability to deliver customer satisfaction.*

*Fixed Effects* Recall that the motivation for the structural model estimation is to identify the sources of inefficiency identified in previous work that used fixed effects/frontier methods. If the above three hypotheses explain some or all of these observed bank performance differences, then the importance of the fixed effects coefficients should decrease or even disappear. To test this, I estimate the full structural model with bank-specific fixed effects and without, and test the restriction that bank fixed effects are zero. The  $F$ -test for the restriction is not significant ( $F = 1.02177$ ,  $df = 218, 5858$ ,  $Prob = 0.404$ ), so that the restriction cannot be rejected.

This assumes, however, that the previously observed X-inefficiencies using fixed effects/frontier analysis with reduced form models would also appear when tested on this dataset. If not, then the failure to reject the zero fixed effects restriction may well be due to the dataset, not the structural model. To test this, I regressed a full quadratic reduced form model both with and without fixed effects. The  $F$ -test for the restriction is significant ( $F = 2.8927$ ,  $df = 218, 5944$ ), so the restriction of zero fixed effects is rejected for the reduced form model. This demonstrates that the structural model is simply not picking up the X-inefficiencies because it is a flexible functional form (such as a full quadratic). Therefore, I conclude that *the three factors modeled in the structural equation explain most of the X-inefficiencies previously observed.*

### **Additional Results**

*Risk Estimation* The risk cost/earnings ratio (on the left-hand side of (7)) was computed<sup>11</sup> for each bank in each quarter for which data was available. The empirical cumulative distribution of the mean (across all quarters) of this ratio is plotted below:



**Figure 2**

On average, the cost of risk accounts for 38% of earnings, or approximately 3% of booked cost. The rather substantial spread of risk/earnings confirms the hypothesis of differences across banks in risk management abilities; for some banks risk accounts for 2/3 of their booked earnings, indicating that accounting earnings overstates their economic profit by a factor of three. Other banks apparently are able to manage risk more successfully, achieving negative b 's and leading to an economic profit greater than accounting earnings.

Perhaps even more interesting is the relationship between size and risk. For banks under \$1 billion in assets, McAllister and McManus (1993) found that increasing size permitted lower risk costs for banks resulting from inventory economies. Our findings, based on banks over \$1 billion in assets and a stock market-based risk measure, are overall that the risk-earnings ratio is increasing in bank size. A linear regression of annual risk-earnings ratio on bank revenues yields a positive coefficient that is significant at the 99% level. . However, the relationship between size and risk is a complex one; in order to understand the fine structure of the data, the nonparametric curve-fitting Trewess method (Velleman, 1980, 1993; Hutcheson, 1995) was used.<sup>12</sup> The data and results of these analyses:

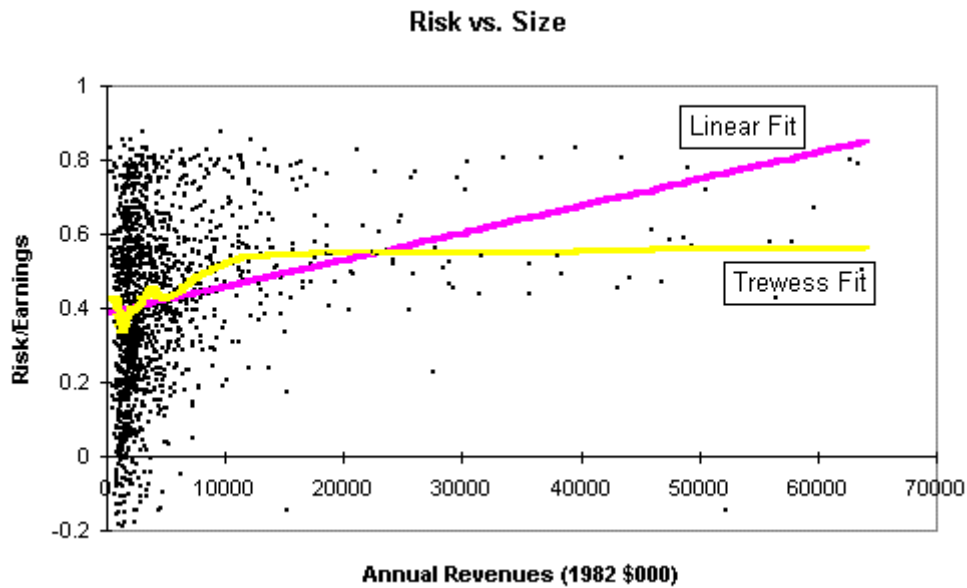


Figure 3

The Trewess fit yields interesting results: risk decreases with size for very small banks (< \$1.5 M in revenues) and then increases sharply with size up to about \$15 M in revenues, and is then relatively constant for larger banks.

The decrease of risk for very small banks is consistent with the results of McAllister and McManus (1993). The increase in risk with size is more puzzling. There are at least three possible explanations of the positive and significant risk-size relationship: (i) it is possible that as banks increase in size and layers of management, effective control of risk in the field is more difficult. (ii) It could also be that large banks become large by management's overly aggressive growth, which drives them to utilize excess capacity with riskier business. (iii) It may also be an optimal response to the so-called "too big to fail" doctrine; if bank managers and depositors of large banks are protected from losses, they are encouraged to undertake riskier actions. Equity holders are not so protected, so these risky actions are reflected in what equity investors are willing to pay for the shares of such banks.

The first two hypotheses would suggest that risk would increase with size without limit, which is not what is observed.<sup>13</sup> The "too big to fail" hypothesis, however, is consistent with the observations; for banks between \$2.5 M and \$18 M there is some probability that in the event of a failure they will be bailed out, and this probability is increasing in size. The higher this probability, the more willing the bank managers are to engage in risky behavior, as they may avoid exposure to this risk via a bailout. Banks above \$18 M are virtually assured of a bailout, so that increasing size does not yield increases in risk-taking; these

banks are already at the maximum risk they wish to take, given the costs of bailouts. Note that this result is consistent with that of Hughes and Mester (1993).

**Operating Costs** The overall regression results are

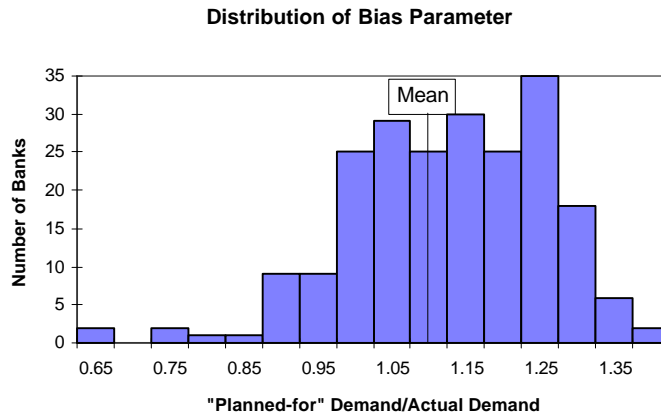
**Table 1 – Overall Regression Results**

$R^2$	0.850842
Adjusted $R^2$	0.849516
F-statistic	641.8394
log likelihood	-49045.23

The estimation of equation (8) had several problems associated with it, which are described in Appendix B, along with the solution and resolution of these problems.

**Scale and Scope Effects** The estimation was not designed to test for scale or scope economies, although it does permit their evaluation, both in the short run and the long run. The full results are reported in Appendix B. In brief, neither scope nor scale economies are present in either the long run or the short run (except the rather weak form of scale economies implied by the assumed short-run fixed costs).

**Bias Parameters** The bank-specific bias parameter represents the ratio of planned-for demand to expected demand for each bank. The theory of the model suggests that *ex post* costs are minimized if this ratio is one, and costs increase as this ratio increases or decreases. A bias parameter is separately estimated for each bank. Of the 219 banks in the sample, 218 had sufficient data to estimate  $m^k$ . The distribution is shown below:



**Figure 4**

The mean ratio of planned-for demand to expected demand is 1.10, indicating substantial overcapacity on average. While a few banks are seriously pessimistic in their forecasting, most banks appear rather aggressive in installing capacity to meet future needs. 76% of the banks had bias parameters significantly greater than 1 at the 90% level.

Before concluding that US banks are deficient in capacity planning, two questions must be answered. First, is this mismatch of demand and capacity an efficient response to forecast uncertainty? In Appendix C, a model of optimal capacity choice is developed in which banks may optimally choose a capacity level that does not correspond to expected demand. A few key results from the Appendix B analysis: (i) this planned-for demand need not equal expected demand; under simplifying assumptions, it is optimal to provide for {more than, less than, the same as} the expected quantity, as  $a_{ii} \{>, <, =\} 0.5$ ; and (ii) under reasonable bounding assumptions about  $a_{ii}$  and forecast error, the theoretical difference between planned-for demand and expected demand is quite small, most likely less than 1/2%. The results of Appendix C suggest that such “efficient” over (or under) capacity is likely to be very small. Therefore, our empirical finding of substantial mismatches between capacity and demand is almost surely the result of inefficient bias, and not an efficient response to forecast uncertainty.

It is also possible that mismatches between demand and capacity come about because the underlying process is non-stationary, and thus particularly costly to forecast well. However, the results of Appendix C suggest that even a simple-minded forecasting method seems to outperform the bank average, suggesting that it is not the difficulty of the forecasting problem that is leading to these mismatches.

Further, the results of Appendix C suggest that this 10% mismatch costs US banks about 2.2% of total costs, or more than 25% of the average bank margin of 7.8%. Also, the evidence does not support that

overcapacity increases customer satisfaction; we found no correlation between the bias parameter and customer satisfaction.

*Other Estimation Problems* It will be recalled that the cost function did not include input prices, as theory would suggest. This formulation is valid as long as input prices are constant across time and across banks. In order to test the time-independence assumption, the model was estimated using year dummies; all year coefficients were zero. This suggests that using constant dollars captures the full effect of time in the operating results estimation, in contrast to the risk estimation in which year is important.

However, there may still be variation in input prices across regions of the country. While the cost of capital and other materials is not likely to vary by region, it is possible for both labor costs and land costs to vary considerably. Two regions of the US suggest themselves as high-cost candidates: New York and California. To test for such differences in costs, the model was estimated using dummy variables to represent location in one of these regions. Chicago was also included as a relatively low-cost region. All three regional dummy coefficients were zero. These results are consistent with the hypothesis that input prices can be safely ignored.

Finally, the nonlinear and iterative nature of the estimation is cause for concern that the parameter estimates may depend upon the starting points. Although the size of the problem precluded extensive testing, all results were verified by starting the estimation process at least three different initial values, with no effect on the outcome. It should be recalled, however, that the variance-covariance matrix derived from nonlinear regressions is at best an approximation to the true matrix. Therefore, the quoted standard errors should be viewed with this reservation in mind.

*Customer Satisfaction Estimation* The fact that many banks are willing to pay Greenwich Associates for customer satisfaction survey data year after year suggests that customer satisfaction is important to banks. There are several ways in which the effect of customer satisfaction on a bank is realized: cost, price, and demand.

*Cost* As discussed above, the operating results estimation was completed on the full 6190-observation dataset. The fitted model was used to evaluate two terms: those relating to costs of C&I loans ( $C_3$ ) and all other costs ( $C_{-3}$ ):

$$C(\mathbf{q}) = F + \sum_{i \neq 3} c_i q_i^{2a_{ii}} + \sum_{i \neq 3} \sum_{j \neq i} n_{ij} (q_i q_j)^{a_{ij}} + X^d \left( c_3 q_3^{2a_{33}} + \sum_{j \neq 3} n_{3j} (q_3 q_j)^{a_{3j}} \right)$$

$$C(\mathbf{q}) = q(C_{-3} + X^d C_3)$$

This equation was estimated on the 476-observation dataset, with the following results:

**Table 2 - Cost-Customer satisfaction Estimation**

	Coefficient	Std. Error	T-statistic	$p$ -value
q	1.769108	0.049293	35.88932	0.0000
d	0.011379	0.003201	3.554510	0.0000

$R^2$	0.702536	Adjusted $R^2$	0.701908
log likelihood	-4469.949	F-statistic	1119.470

Clearly, customer satisfaction increases cost but not very much. The elasticity parameter is statistically significant but not economically significant; doubling customer satisfaction only increases costs by 1%. For example, increasing customer satisfaction from the median level of 51 to the 75th percentile level of 59, a 15% change, increases costs by on 0.15%.

*Price* High customer satisfaction firms are often able to command a price premium for their services. Four-star hotels, high-end jewelers, and other retailers seem to earn rents from the higher prices they are able to charge. In some cases, though, higher quality commands only a very small price premium, but does lead to higher volume. Some department stores well-known for customer satisfaction stay price-competitive but handle higher volume with greater efficiency.

The model  $\log p_3 = q + f \log X$  was estimated, with the following results:

**Table 3 - Price-Customer satisfaction Estimation**

	Coefficient	Std. Error	T-statistic	$p$ -value
q	0.039351	0.019669	2.000693	0.0460
f	0.010535	0.005031	2.093846	0.0368

$R^2$	0.009165	Adjusted $R^2$	0.007074
log likelihood	972.0862	F-statistic	4.384192

Clearly, there are many factors other than customer satisfaction which affect price. It is the case that customer satisfaction does affect price positively (as is expected); again, the result is statistically significant but not economically significant; doubling customer satisfaction permits a mere 1% price premium.

However, though both the customer satisfaction elasticity of cost and of price are small, they are statistically significant and it is of interest to note that they are essentially equal. The small increase in cost from higher customer satisfaction is recovered in the equally small increase in the price premium the bank can charge for customer satisfaction; margins do not suffer at all with higher customer satisfaction.

*Quantity* Does customer satisfaction affect the quantity demanded? The estimation of this relationship is substantially more difficult. A simple regression of quantity on customer satisfaction does indeed lead to a strongly positive and significant elasticity; however, this may merely mean that large banks are better at providing good service. In order to sort out whether quantity leads to greater customer satisfaction (economies of scale in customer satisfaction provision) or customer satisfaction leads to greater quantity (a market response to better service), several models were estimated that included proxies for size as well as the customer satisfaction measure.

The literal interpretation of the model suggests that the only reason banks differ in size is because of their capacity choices. Thus, the first model uses the bias parameter as a size proxy:

$\log q_3 = q_0 + q_1 \log X + q_2 \log m^k$ . The second model uses total revenues as a size proxy in linear form:  $q_3 = q'_0 + q'_1 X + q'_2 R$ . The third model (unreported here) uses a combination of

$q_1$ ,  $q_2$ , and  $q_6$  in quadratic form, with equally strong results. The estimated coefficients from the first two models:

**Table 4 - Quantity-Customer satisfaction Estimation**

	Coefficient	Std. Error	T-statistic	p-value
<b>Model 1</b>				
$q_0$	5.954614	0.793794	7.501460	0.0000
$q_1$	0.980932	0.203279	4.825551	0.0000
$q_2$	-3.058770	0.510000	-5.997588	0.0000
<b>Model 2</b>				
$q'_0$	-24122.57	6444.394	-3.743187	0.0002
$q'_1$	607.5021	125.4451	4.842772	0.0000
$q'_2$	8.252161	0.266152	31.00547	0.0000

Using different proxies for size all lead to strongly positive significant elasticities, ranging from 0.16 to 0.98.

This set of results is consistent with the hypothesis that though banks cannot command a price premium for customer satisfaction, it is a means of attracting and keeping customers. Once a bank adapts itself to a high-quality technology, they can realize greater demand at about the same cost.

*Profit* The above results suggest that customer satisfaction should have a positive impact on profit. We test this directly by estimating a simple profit-customer satisfaction relationship. Once again, size must be controlled for. Both revenues and the bias parameter were used as size controls, with similar results. We report on the model  $\hat{p} = \alpha_0 + \alpha_1 X + \alpha_2 R$ :

**Table 5 - Profit-Customer satisfaction Estimation**

	Coefficient	Std. Error	T-statistic	p-value
$\alpha_0$	-542.0245	441.7871	-1.226891	0.2205
$\alpha_1$	16.04065	8.599727	1.865251	0.0628
$\alpha_2$	0.065855	0.018246	3.609369	0.0003

Customer satisfaction has the expected sign and is significant at the 6% level.

## 6. Conclusions

A model of banking markets is developed which explicitly incorporates risk, customer satisfaction, capacity planning, short-run and long-run cost functions, and market equilibrium. The model is estimated in structural form, using data from operating results, capital markets, and customer surveys. The principal findings are:

- (i) Banks differ widely in their ability to manage risk; larger banks take on relatively more risk; on average, risk cost accounts for 38% of bank earnings. I conclude that risk cost is a significant source of bank performance differences.
- (ii) There are substantial inefficiencies due to demand/capacity mismatches. On average, banks are over-optimistic by 10% in the demand they plan for, and this cost them about 2.2% of total costs. This is substantially more than can be justified by “optimal overshooting” in the face of planning uncertainty, and amounts to over 25% of average bank margin. I conclude that capacity planning is a significant source of bank performance differences.

- (iii) Greater customer satisfaction correlates with greater profitability, principally due to greater demand; the effect of customer satisfaction on cost and price is minimal. I conclude that the ability to provide customer satisfaction is a source of bank performance differences.
- (iv) Bank-specific fixed effects are relatively insignificant; the very significant bank-specific effects that previous research discovered appear to have been largely captured and directly estimated in the structural model. I conclude that the three factors account for the previously observed but unexplained performance differences among banks.
- (v) Confirmation of the results of previous research that there are no significant long-run economies of scale or scope.

The paper employs methodological innovations as well:

- (i) The Capital Asset Pricing Model is used to measure bank risk, thereby capturing all risk in a measure based on market behavior.
- (ii) A new dataset is introduced to study the ability of banks to satisfy their customers.
- (iii) Structural model estimation is used to expand the range of questions that can be empirically addressed; it is also used to explain and directly estimate inefficiencies hitherto captured only by inference in fixed effects models.

This paper is but a start on unraveling the problem of bank inefficiencies, focusing on structural modeling and new datasets. Understanding why banks (or other financial institutions) are different requires new models, new variables, and new datasets. The previous research paradigm of reduced-form estimation on generic datasets using frontier methods is most likely near the end of its useful life; the challenge to researchers is to develop the modeling and data tools required to address this new set of issues.

-- References --

- Berger, Allen N. (1993) "'Distribution-Free' Estimates of Efficiency in the US Banking Industry and Tests of the Standard Distributional Assumptions," *J. Productivity Analysis*, **4**, 261-292.
- Berger, Allen N. and David B. Humphrey (1991) "The Dominance of Inefficiencies Over Scale and Product Mix Economies in Banking," *J. Monetary and Financial Econ*, **28**, 117-148.
- Berger, Allen N. and David B Humphrey (1997), "Efficiency of financial institutions: International survey and directions for future research," *European Journal of Operational Research*, **98**, 175-212.
- Berger, Allen N., Diana Hancock, and David B. Humphrey (1993) "Bank Efficiency Derived from a Profit Function," *J. Banking and Finance*, **17**, 317-347.
- Berger, Allen N. and Loretta J Mester (1997), "Inside the black box: What explains differences in the efficiencies of financial institutions?" *Journal of Banking & Finance*, **21**, 895-947.
- Berger, Allen N., William C. Hunter, and Stephen G. Timme (1993) "The Efficiency of Financial Institutions: a Review and Preview of Research Past, Present, and Future," *J. Banking and Finance*, **17**, 221-249.
- Berger, Allen N., John H. Leusner, and John J. Mingo (1994) "The Efficiency of Bank Branches," Wharton Financial Institutions Center Working Paper 94-27.
- Frei, Frances X. (1996) "The Role of Process Designs in Efficiency Analysis: an Empirical Investigation of the Retail Banking Industry," doctoral dissertation, Wharton School, Univ. of Pennsylvania (University Microfilm: Ann Arbor MI).
- Fried, Harold O., C.A. Knox Lovell, and Phillipe Vanden Eeckaut (1993) "Evaluating the Performance of US Credit Unions," *J. Banking and Finance*, **17**, 251-265.
- Grabowski, Richard, Nanda Ragan, and Rasoul Rezvanian (1994) "The Effect of Deregulation on the Efficiency of US Banking Firms," *J. Economics and Business*, **46**, 39-54.
- Hughes, Joseph P. And Loretta J. Mester (1993) "A Quality and Risk-Adjusted Cost Function for Banks: Evidence on the "Too-Big-To-Fail" Doctrine, *J. Productivity Analysis*, **4**, 293-315.
- Hughes, Joseph P. and Loretta J. Mester (1994) "Bank Managers' Objectives" Federal Reserve Bank of Philadelphia Working Paper 94-8.

- Hutcheson, Matthew C. (1995) "Trimmed Resistant Weighted Scatterplot Smooth," Department of Statistics, Cornell University, Ithaca, N.Y.
- Kaparakis, Emmanuel I., Stephen M. Miller, and Athanasios G. Noulas (1994) "Short-run Cost Inefficiency of Commercial Banks: A Flexible Stochastic Frontier Approach," *J. Money, Credit, and Banking*, **26**, 875-893.
- Leibenstein, Harvey (1966) "Allocative vs. 'X-Efficiency'," *American Economic Review*, **56**, 392-415.
- Mester, Loretta J. (1993) "Efficiency in the Savings and Loan Industry," *J. Banking and Finance*, **17**, 267-286.
- Mester, Loretta J. (1994) "How Efficient Are Third District Banks?," *Business Review*, (Jan/Feb) 3-18.
- McAllister, Patrick H. And Douglas McManus (1993) "Resolving the Scale Efficiency Puzzle in Banking," *J. Banking and Finance*, **17**, 389-405.
- Pulley, Lawrence B. and David B. Humphrey (1993) "The Role of Fixed Costs and Cost Complementarities in Determining Scope Economies and the Cost of Narrow Banking Proposals," *J. Business* **66**, 437-462.
- Rohlf, Jeffrey (1978) "A Model of the Bell System Cost Function," submitted to U.S. District Court by AT&T.
- Roth, Aleda V. and William E Jackson III (1995), "Strategic determinants of service quality and performance: Evidence from the banking industry," *Management Science*. 41(11): 1720-1733.
- Santomero, Anthony M. (1995) "Commercial Bank Risk Management: an Analysis of the Process," Working Paper 95-11, Wharton Financial Institutions Center
- Souterious, Andreas and Stavros A. Zenios (1997) "Operations, Quality, and Profitability in the Provision of Banking Services," Report 97-13, Dept of Public and Business Administration, University of Cyprus, Nicosia, Cyprus.
- Vavra, Terry G. (1995) "Selling After the Sale," *Bank Marketing*, **27**, 27-30.
- Velleman, Paul F. (1980), "Definition and comparison of robust nonlinear data smoothing algorithms," *J. Amer. Statist. Assoc.*, **75**, 609-615.
- Velleman, Paul F.(1993), "Trewess Proposal," Ithaca, N.Y: Date Description, Inc.

## Appendix A – Cost Functions

### The Cost of Risk

Let the value of the bank be  $V$ , the earnings of the bank in future period  $t$  be the random variable  $\Pi_t$ .

Then the market discount rate of the bank is  $s$ , where  $s$  satisfies

$$V = E \left[ \sum_{t=1}^{\infty} \frac{\Pi_t}{(1+s)^t} \right],$$

and  $E$  is the expectation operator. The discount rate  $s$  incorporates the stock market's assessment of risk  $b$ .<sup>14</sup> Define the *uniform expected earnings* as  $\bar{p}$ , the uniform certain level of earnings which would yield

the same value of the firm if discounted at  $s$ :

$$V = \sum_{t=1}^{\infty} \frac{\bar{p}}{(1+s)^t}.$$

Now define the *risk-free earnings* as  $\tilde{p}$ , the uniform certain level of earnings which would yield the same value of the firm if it were completely free of risk, and the earnings discounted at  $r_f$ :

$$V = \sum_{t=1}^{\infty} \frac{\tilde{p}}{(1+r_f)^t}, \text{ or}$$

$$\tilde{p} = \bar{p} \cdot \left( \frac{r_f}{r_f + b(r_m - r_f)} \right)$$

Now the risk cost is defined to be the additional earnings,  $R = \bar{p} - \tilde{p}$ , that investors demand because of the risk that the bank carries. Normalizing this risk cost by expected earnings, this is simply

$$\frac{R}{\bar{p}} = \left( \frac{b(r_m - r_f)}{r_f + b(r_m - r_f)} \right), \text{ the risk/earnings ratio.} \quad (\text{A9})$$

If all banks were equally efficient at risk management, this risk/earnings ratio would differ little among banks. Those that chose to bear more risk would reap the reward in terms of higher earnings, leading to a slightly lower risk/earnings ratio.

Aggregate risk for a bank is comprised of risks from many individual transactions, each of which is associated with a specific product. For purposes of this paper, an estimate of the cost of risk for each

product is required. The most straightforward reduced form model that decomposes total bank risk into risk by product is assumed:

$$R^k = \sum_{i=1}^6 \bar{r}_i^k q_i^k . \quad (\text{A10})$$

A more general functional form involving higher order terms would no doubt conform more closely to a richer view of product risk. However, *bank-specific* risk coefficients are estimated in this study, and the per-bank sample size is quite limited, thereby restricting the parameters that can usefully be estimated.<sup>15</sup> As a result, we must be content with capturing first-order effects only.

### Operating Cost Function

Each bank's production of its six outputs in a time period is measured by the outstanding stock of each output on the bank's balance sheet at the end of the period (for products 1-5, and analogously for product 6). The criteria I use in selecting a functional form for the short-run cost function are (i) it must incorporate a tradeoff between fixed and variable cost; (ii) it must reflect possible cost complementarities and economies of scope (implying the need for interaction terms); (iii) it must behave reasonably throughout the positive orthant; in particular, if a single output is not produced and others are, its incremental cost must be zero but total costs must not be; and (iv) it is helpful to be able to estimate partial scale and scope effects directly. These criteria are satisfied by the following *family of short-run cost functions* of the form

$$C(\mathbf{q}; F, \mathbf{c}) = F + \sum_{i=1}^6 c_i q_i^{2a_{ii}} + \sum_{i=1}^5 \sum_{j=i+1}^6 \rho_{ij} (q_i q_j)^{a_{ij}} , \quad (\text{A11})$$

where  $\mathbf{v}$  and  $\boldsymbol{\alpha}$  are symmetric matrices common to the entire family of short-run cost functions. The bank chooses a specific member of this family by selecting an  $F$  and  $\mathbf{c} = (c_1, \dots, c_6)$ , which choice variables completely characterize the menu of short-run cost functions. In turn, the lower envelope of this family of short-run cost functions defines the *long-run cost function*; for every output vector in the positive orthant,  $\mathbf{q} \in \mathbb{R}^6$ , there exists a short-run cost function that (i) is equal to the long-run cost function at  $\mathbf{q}$ ; (ii) is tangent to the long-run cost function at  $\mathbf{q}$ ; and (iii) lies everywhere above the long-run cost function. Alternatively, the long-run cost function may be viewed as defining the short-run options  $F$  and  $\mathbf{c} = (c_1, \dots, c_6)$ , which is the view taken here. The long-run cost function should have similar properties to the short-run functions, but reflect the property that in the long run, no costs are fixed. The long-run cost function is assumed to be of the form

$$C(\mathbf{q}) = \sum_{i=1}^6 \sum_{j=i}^6 l_{ij}(q_i, q_j)^{g_{ij}} \quad (\text{A12})$$

where  $\lambda$  and  $\gamma$  are symmetric  $6 \times 6$  matrices. For each vector  $\tilde{\mathbf{q}}$  in the positive orthant, there exists a unique short-run cost function  $(F, \mathbf{c})$  defined by the total condition (i):

$$C(\tilde{\mathbf{q}}; F, \mathbf{c}) = C(\tilde{\mathbf{q}}), \text{ or}$$

$$F(\tilde{\mathbf{q}}; \mathbf{c}, \alpha, \gamma, \lambda, \nu) = \sum_{i=1}^6 \sum_{j=i}^6 l_{ij}(\tilde{q}_i, \tilde{q}_j)^{g_{ij}} - \sum_{i=1}^6 c_i \tilde{q}_i^{2a_{ii}} - \sum_{i=1}^5 \sum_{j=i+1}^6 n_{ij}(\tilde{q}_i, \tilde{q}_j)^{a_{ij}} \quad (\text{A13})$$

and the tangency conditions (ii):

$$\nabla C(\tilde{\mathbf{q}}; F, \mathbf{c}) = \nabla C(\tilde{\mathbf{q}}), \text{ or}$$

$$c_i = \frac{\tilde{q}_i^{-2a_{ii}}}{2a_{ii}} (2b_{ii} l_{ii} \tilde{q}_i^{2g_{ij}} + \sum_{j \neq i} [b_{ij} l_{ij}(\tilde{q}_i, \tilde{q}_j)^{g_{ij}} - a_{ij} n_{ij}(\tilde{q}_i, \tilde{q}_j)^{a_{ij}}]), i = 1, \dots, n. \quad (\text{A14})$$

We assume the second-order conditions are satisfied.

The intuition here is straightforward. Banks choose a short-run cost function from the available menu; if they expect relatively low demand, they choose a low fixed cost, high variable cost technology. If they expect relatively high demand, they choose a high fixed cost, low variable cost technology. The available menu of short-run cost functions is defined by its lower envelope, which is therefore the long-run cost function.

Using equation (4), the variable cost coefficients  $\mathbf{c}^k$  that are consistent with the first-order conditions for this functional form are

$$c_i^k = \frac{p_i (1 - \sum_{j=1}^n s_{ji} h_{ji} q_j^k) - \sum_{j \neq i} a_{ij} n_{ij} (q_i^k)^{a_{ij}-1} (q_j^k)^{a_{ij}} - \bar{r}_i^k}{2a_{ii} (q_i^k)^{2a_{ii}-1}} \quad (\text{A15})$$

The fixed cost  $F^k$  consistent with these first-order conditions is given in equation (A13), substituting for  $c_i^k$  from equation (A15).

## Appendix B – Estimation of the Operating Cost Function

### Estimation Problems

The estimation of equation (8) presents several problems. The first problem is the number of parameters and the estimating equation's nonlinearity. The full model has 99 structural parameters ( $\alpha$ ,  $\gamma$ ,  $\lambda$ ,  $\nu$ ), 219 intercept dummies ( $B^k$ ) and 219 slope dummies ( $m^k$ ). The full model is estimated using an iterative procedure; first, the dummy variables are assumed known and the 99 parameters are estimated using nonlinear least squares. Second, the estimated structural parameters are assumed known, the intercept dummies set equal to the bank average residuals, and the slope dummies are estimated using nonlinear least squares. These dummy values are then used to repeat the first step; the procedure is continued until convergence of all parameters is achieved.

The second problem is that while the operating results are available for 6190 data points, the customer satisfaction results are only available for a subset of size 473, a reduction of over an order of magnitude. The loss of degrees of freedom involved in this reduction of sample size, especially with a nonlinear equation, is substantial. The full model was therefore estimated in four configurations: (i) the full 6190 sample, without customer satisfaction; (ii) an annual sample of 1533 data points,<sup>16</sup> without customer satisfaction; (iii) the reduced sample of 473, without customer satisfaction; and (iv) the reduced sample of 473, *with* customer satisfaction. The structural parameter estimates were nearly identical for all four configurations; however, the standard errors of the estimates differed substantially, with the largest sample size yielding rather good estimates of many parameters. The stability of the parameter estimates suggested a sequential estimation strategy: estimate the operating parameters using the 6190 dataset, then reduce the fitted values to the 473 dataset to estimate only the customer satisfaction parameters. In this section, therefore, the results of the operating estimation is reported.

The third potential problem is the use of quarterly *Call Report* data vs. annual data. Most empirical banking research has focused on annual *Call Report* data, citing concerns that end-of-year bank reports are likely to be more accurate than preceding quarters. To examine this question, the model was run using end-of-year data only. The result was that while the parameter estimates changed almost not at all, the *t*-statistics were significantly worse moving from quarterly to annual data. If the quarterly data introduces more noise into the estimation, it certainly did not outweigh the additional degrees of freedom.

Initial estimates of the full model yielded insignificant results for both the long-run and the short-run cost coefficients for the following cross-terms (refer to Section 2 for product numbers): (1,2), (1,3), (1,4), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6). Consequently, these terms were dropped in subsequent estimation.

**Scale and Scope Effects**

*Scale Effects* Ray scale economies are defined in the usual way. For a scalar  $h$  and any output vector  $q$ , the long-run scale elasticity is defined as

$$S_{LR} = \frac{\sum_{i=1}^6 \sum_{j=i}^6 2g_{ij} |_{ij} (h^2 q_i q_j)^{g_{ij}}}{\sum_{i=1}^6 \sum_{j=i}^6 |_{ij} (h^2 q_i q_j)^{g_{ij}}},$$

with {constant returns, increasing returns, decreasing returns} to scale as  $S_{LR}$  is {=,<,>} 1. It is clear from this formula that if all  $g_{ij} \geq 0.5$ , then there cannot be increasing returns to scale. If the inequality is strict for at least one coefficient, then there are decreasing returns to scale. Inspection of the left-hand side of Table 6 shows that this is indeed the case. All coefficients are either not significantly different from 0.5, or they are significantly greater than 0.5 ( $g_{11}$  and possibly  $g_{23}$ ).

The scale analysis of the short-run cost functions is complicated by the fact that the fixed costs lead to a weak form of scale economies: “spreading the fixed cost.” However, if we examine only variable costs, the left-hand side of Table 6 shows highly significant increasing marginal cost, with no coefficient significantly less than 0.5 and all the direct coefficients significantly above this number.

**Table 6 – Coefficient Estimates**

	Estimate	Std. Error	p-Value*		Estimate	Std. Error	p-Value*
$g_{11}$	0.923659	0.114651	0.000110896	$a_{11}$	1.626377	0.548304	0.019996674
$g_{22}$	0.533569	0.126315	0.39521833	$a_{22}$	1.666517	0.489423	4.87509E-06
$g_{33}$	0.383381	0.12275	0.171063035	$a_{33}$	2.115984	0.175854	2.72262E-20
$g_{44}$	0.546865	0.038757	0.113318185	$a_{44}$	2.056322	0.780928	0.023159515
$g_{55}$	0.024372	51.18033	0.495912859	$a_{55}$	1.683526	0.889139	0.091606246
$g_{66}$	0.319622	0.215415	0.201215351	$a_{66}$	3.123559	0.525182	3.02229E-07
$g_{16}$	0.459534	0.07297	0.289609966	$a_{15}$	0.254758	14.18534	0.493103571
$g_{23}$	0.620312	0.090278	0.091344774	$a_{16}$	1.061438	0.143995	4.88542E-05
$g_{56}$	0.473472	0.126587	0.417008016	$a_{23}$	1.190735	0.04016	5.83302E-65
				$a_{56}$	-0.755504	17103787	0.5

\* probability that the true value of the coefficient lies on the opposite side of 0.5 than the point estimate.

The short-run scale parameters are all substantially greater than 0.5, indicating significant increasing marginal costs in the short run. This appears consistent with the results of Kaparakis, Miller, and Noulas (1994).

*Scope Economies* A necessary and sufficient condition for long-run cost complementarities between products  $i$  and  $j$  is that  $\lambda_{ij} < 0$ , and for short-run cost complementarities,  $\eta_{ij} < 0$ . The estimation results are not consistent with the existence of cost complementarities:

**Table 7**

	Coefficient	Std.Error	p-value		Coefficient	Std.Error	p-value
$\lambda_{15}$	0.09333	0.154031	0.5446	$\eta_{14}$	0.460942	136.5392	0.9973
$\lambda_{16}$	-0.01545	0.032797	0.6376	$\eta_{15}$	4.59E-07	1.58E-06	0.7707
$\lambda_{23}$	0.001395	0.003181	0.6609	$\eta_{16}$	1.05E-07	9.50E-08	0.267
$\lambda_{56}$	0.016984	0.05546	0.7594	$\eta_{36}$	0.157856	25732798	1.0000

None of the cross-terms are significantly different from zero; these results are consistent with the hypothesis that there are no cost complementarities, either short-run or long-run, in banking. Note that the existence of short-run fixed costs implies that simply spreading these fixed costs across several product lines does give rise to a weak form of scope economies.

## Appendix C – Optimal Capacity Choice

**A Model of Optimal Capacity Choice Based on Demand Forecasting** Bank  $k$  has available to it certain *public* information  $I$  and certain *private* information  $I^k$  regarding their demand for the next period. Using this information, bank  $k$  estimates the cumulative distribution function of its next-period demand as  $G(\mathbf{x}; I, I^k) = G^k(\mathbf{x})$ . The parameters depend upon both public and private information, so they need not be the same for all banks.

Bank  $k$  chooses the technology  $(\tilde{F}^k, \tilde{\mathbf{c}}^k)$  that minimizes its expected cost (assuming risk neutrality):

$$(\tilde{F}^k, \tilde{\mathbf{c}}^k) = \arg \min_{F, \mathbf{c}} \int_{-\infty}^{\infty} \left( F + \sum_{i=1}^6 c_i x_i^{2a_i} + \sum_{i=1}^5 \sum_{j=i+1}^6 \rho_{ij} (x_i x_j)^{a_{ij}} \right) dG^k(\mathbf{x}) \quad (\text{C16})$$

To simplify the analysis, we consider the case in which all products are independent, both in the economic sense ( $\rho_{ij} = 0$ ) and in the statistical sense ( $dG^k(\mathbf{x}) = g^k(\mathbf{x}) = \prod g_i^k(x_i)$ ). In this case, the short-run and long-run cost functions are (removing redundant double subscripts)

$$C(\mathbf{q}; F, \mathbf{c}) = F + \sum c_i q_i^{2a_i}, \quad C(\mathbf{q}) = \sum l_i q_i^{2g_i}.$$

For each  $\mathbf{q} > 0$ , there is a corresponding  $(F, \mathbf{c})$ , derivable from the conditions

$$C(\mathbf{q}; F, \mathbf{c}) = C(\mathbf{q}), \quad \nabla C(\mathbf{q}; F, \mathbf{c}) = \nabla C(\mathbf{q}):$$

$$c_i = \frac{l_i g_i}{a_i} q_i^{2(b_i - a_i)}, \quad F = \sum (l_i q_i^{2g_i} - c_i q_i^{2a_i}) = \sum \left( l_i q_i^{2g_i} \left[ \frac{a_i - g_i}{a_i} \right] \right). \quad (\text{C17})$$

Evaluating the expected cost in the single-product case, we obtain

$$\int (F + \sum c_i x_i^{2a_i}) dG^k(\mathbf{x}) = F + \sum c_i \int x_i^{2a_i} g_i^k(x_i) dx_i = \sum \left( l_i q_i^{2g_i} \left[ \frac{a_i - g_i}{a_i} \right] + \frac{l_i g_i}{a_i} q_i^{2(b_i - a_i)} \int x_i^{2a_i} g_i^k(x_i) dx_i \right).$$

Differentiating this with respect to  $q_i$  and setting the result equal to zero yields

$$2 l_i g_i q_i^{2g_i - 1} \left( \frac{a_i - g_i}{a_i} \right) = 2 l_i g_i q_i^{2g_i - 2a_i - 1} \left( \frac{a_i - g_i}{a_i} \right) \int x_i^{2a_i} g_i^k(x_i) dx_i$$

so that the optimal “planned-for” demand is

$$\tilde{q}_i^k = \left( \int x_i^{2a_i} g_i^k(x_i) dx_i \right)^{\frac{1}{2a_i}}. \quad (C18)$$

The actual technology ( $F, c$ ) can be determined by substituting (A3) into (A2).

Clearly, for  $a_i = 0.5$  (short-run constant marginal cost), the optimal capacity to install is that which corresponds with the mean demand:  $\tilde{q}_i^k = \bar{q}_i^k$ . It is easy to show that for  $a_i = 1$  (short-run increasing marginal cost), the optimal capacity corresponds to planned-for demand greater than the mean demand:

$\tilde{q}_i^k = \left( \text{Var } q_i^k + (\bar{q}_i^k)^2 \right)^{\frac{1}{2}} > \bar{q}_i^k$ , which is increasing in  $\text{Var } q_i^k$ . In addition, planned-for demand is an

increasing function of  $a_i$ , since  $\frac{\partial \tilde{q}_i^k}{\partial a_i} = \left( \int x_i^{2a_i} g_i^k(x_i) dx_i \right)^{\frac{1-2a_i}{2a_i}} \cdot \int x_i^{2a_i-1} g_i^k(x_i) dx_i > 0$ , for all  $a_i >$

0, assuming the support of  $g_i^k$  is strictly positive. We can therefore conclude that in the independent product case, it is optimal to provide for {more than, less than, the same as} the expected quantity, as  $a_i$  {>,<,<=} 0.5.

If banks are making unbiased demand forecasts (in the sense that on average expected demand equals actual demand), then observed mismatches between actual demand and installed capacity may be efficient in that the mismatch may be an efficient response to forecast uncertainty and the nonlinear cost function. On the other hand, such mismatches could also be due to consistently biased demand forecasts, which would be inefficient. This could occur if, for example, bank managers are consistently over-optimistic about next period demand, consistently installed excess capacity as a result, and did not learn from past mistakes. Is it possible to distinguish between efficient mismatches of demand and capacity and inefficient mismatches?

Some simple numerical examples suggest that efficient mismatching of demand and capacity is unlikely to be significant in practice. In the empirical results, the average  $a_i$  is found to be 2.04. To bound the coefficient of variation of the forecast distribution, we analyzed the simplest possible forecast method: next period's demand will equal last period's demand. Applying this method to the data for all six products yields an overall empirical coefficient of variation of 0.085. Assuming a normal forecast distribution, this leads to planned-for demand 1.1% larger than expected demand. Of course, this simplest possible forecast method should be easy to beat; if bank forecasters can reduce the coefficient of variation by half (a very modest target, it would seem), then planned-for demand is only 0.3% larger than expected demand. In the case at hand, efficient mismatching of demand and capacity are not significant.

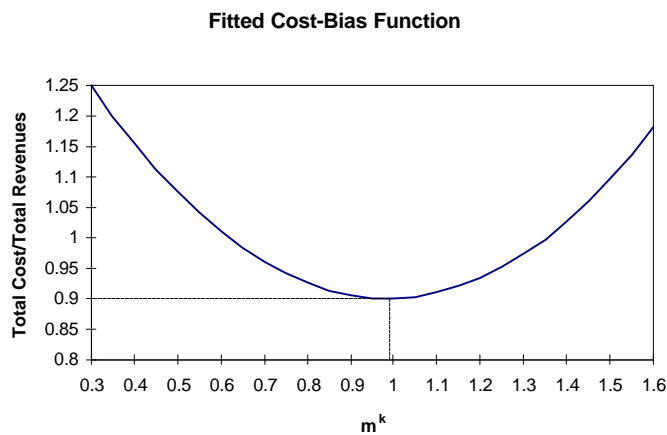
### Cost of Mismatching Capacity and Demand

Does the mismatch between capacity and demand matter? It could be that such mismatches result in only a very small increase in cost, so it would not be optimal to invest much effort into getting the forecast right. A direct test of this would be to examine the ratio of short-run to long-run costs as computed from the model. Unfortunately, the fact that many of the parameters of the nonlinear model are not statistically significant leads to this ratio being more noise than information; in particular, this calculated ratio is less than one for about 35% of the observations, contrary to the theory. In a less direct method of assessing the cost of bias, the ratio of total cost to total revenues was regressed against a quadratic function of the bias parameter (assuming orthogonality with other coefficients), with the following results:

**Table 8 - Cost/Revenue vs. Bias**

	Coefficient	Std. Error	T-Statistic
Constant	1.622114	0.166863	9.721210
$m$	-1.465286	0.314564	-4.658150
$m^2$	0.743318	0.147404	5.042740

While the theory suggests that this function should be minimized at  $m = 1$ , there is nothing in the data that forces this to be the case. Therefore, it is surprising that the minimum of this empirically fitted function actually does occur at 0.986, with an optimal value of the cost/revenue ratio of 0.9, as shown in **Error! Reference source not found..**



**Figure 5 - Fitted Cost/Revenue vs. Bias**

This suggests that biased forecasts can be costly, especially those of large magnitude. On balance, however, the quadratic model applied to the bias parameter estimates yields the result that the average

cost increase associated with biased capacity planning amounts to 2.2% of costs (with standard deviation of 2%). While this appears small relative to the 10% average over-forecasting, it is in fact a rather large number compared to the average bank margin of 7.8% (earnings/costs). On average, over 25% of bank earnings are absorbed by the costs of the allocative inefficiency of demand/capacity mismatches.

However, allocative inefficiency is not the only possible cost of poor forecasting. Since most banks seem to err on the side of overcapacity, banks may increase their risk as a result. Excess capacity and low marginal cost may tempt bank managers to lower their credit standards to find enough customers to “fill the pipeline.” This hypothesis implies that the bias parameter and the risk/earnings ratio will be positively correlated. The data does not confirm this hypothesis, however; the relationship is actually somewhat negative.

It is also possible that excess capacity could translate into higher customer satisfaction; more branches, for example, could lead to greater convenience for customers, who might then be willing to pay a price premium and/or give more business to the bank. For C&I loans, we have a direct measure of customer satisfaction, which is not correlated with the bias parameter. For all other products, the bias parameter shows either no correlation with prices and quantities or else a small negative correlation. Thus, higher capacity is not correlated with customer satisfaction, either measured directly or via prices and quantities. The bias parameter thus represents allocative inefficiency.

## Appendix D – Customer Satisfaction Dataset

### The Greenwich Associates' Dataset

Greenwich Associates is a strategic research and consulting firm for financial service providers, offering over 60 programs in commercial banking, investment management, stockbroking, bond dealing, investment banking, and foreign exchange dealing. The firm has been in business and conducting financial market research since 1972.

Each year Greenwich Associates conducts 32,000 interviews with senior financial officers at corporations in 19 countries. The data collected from these interviews captures qualitative information on the satisfaction of clients with their current providers of financial services.

Research in the US corporate credit market is split into three programs:

**Large Corporate Banking** research (covering companies with over \$500 million in sales) has over 1200 respondents. Nearly 75% of the population is covered and the research is conducted annually.

**Middle Market Banking** (companies with sales of \$50-500 million) interviews officials at over 3000 companies. Roughly half of the population is covered and the research is conducted every odd numbered year.

**Commercial Banking** (companies with sales of \$5-50 million) analyzes the performance of over 300 banks on a state-by-state basis. Approximately 20,000 interviews are conducted, giving a coverage rate of 20% of the population. Regions covered rotate in a two year cycle.

Bank clients are asked to describe their relationship with the specific bank; the types of relationship are:

**Customer:** This means that the corporation has been a customer of the bank over the period covered by the survey.

**Principal Bank:** The bank is one of the corporation's three or four most important banks.

**Lead Bank:** The bank is the corporation's main or first bank.

Data collected for this study from the Large Corporate and Middle Markets was based on relationships of type "Principal Bank," while data from Commercial Banking was based on relationships of type

“Customer.” So the Competitive Situation Report for Large Corporate Banking, for example, would show what percentage of clients from relationship type “Principal Bank” indicated a positive response to a question.

As might be expected, the survey instrument is quite large, and varies from year to year as well as from segment to segment. In order to ensure as much uniformity across the sample as possible, eight specific questions were selected. These questions appear in all survey instruments in all years in all market segments. On the basis of a principal components analysis, Greenwich Associates avers that these questions capture the key elements of customer satisfaction. In summary form, these questions are:

A) How Do You Rate This Bank in:

- i-willingness to lend
- ii- competitive loan pricing
- iii- cash management capabilities
- iv- international service capabilities

B) How Do You Rate This Bank’s Account Officers in:

- i-convincing bank to meet credit needs
- ii- prompt follow-up
- iii- knowledge of cash management services
- iv- advice on corporate finance

For each question, the respondent had five choices:

- 1- Poor            2- Fair            3- Good 4- Very Good            5-Excellent

For each of the items above, the average percent of respondents who rated the bank a “4” or “5”, referred to as “Above Average,” is recorded, and constitutes the core data of this study.

The dataset consists of observations on 112 banks over several years between 1985 and 1992, resulting in a total of 476 usable observations. Each observation consists of data drawn from the three market segments, each of which has:

- 1) the number of customer respondents;
- 2) for each of the eight questions, the percentage of all respondents who rated the bank “Above Average”; and
- 3) the average revenue for respondents in this segment (from Dun and Bradstreet).

The overall index is constructed as follows:

- For each market segment (Large Corporate, Middle Market, and Commercial), the average of (2) above for all eight questions was computed. In each segment, this measure was Z-score normalized.
- The three normalized indices were combined by weighting them by the average size of clients in each market and the number of clients each bank had in each market and then adding them together.
- This produced an index with mean zero and a unit standard deviation. The range of the index was from -4.23 to 4.40.

This index was then re-scaled, so that -4.23 was mapped to “1” and 4.4 was mapped to “100.” This yields a resulting index ranging from 1 to 100, with a mean of 50.

-- Notes --

<sup>1</sup> This is hardly surprising, as it is unlikely that banking is either competitive, in equilibrium, or in the long run. The limited extent of bank markets<sup>1</sup> suggests banking is more oligopolistic than competitive. The continued importance of bank regulation suggests that the market is regulated away from equilibrium. And the significant and continuing changes in computing, telecommunications, regulation, and global competition over the past fifteen years suggest the industry is not in the long run either.

However, the presence of large variations in observed profits and costs does not *prove* that the industry is out of equilibrium. Indeed, it may be that there is an equilibrium model which “explains” these large variations. There are no such explanations in the extant literature, and the results of this paper make it somewhat less likely that there is one.

<sup>2</sup> While the previous research that identified large inter-bank differences in performance was expressed in terms of frontier analysis, it is the view of this paper that all of the econometric methods used in frontier methods are variants of fixed effects models, with an X-efficiency interpretation added. This is by no means the only view in the literature; Berger (1993) identifies the fixed effects model as merely one of several approaches to calculating the frontier, a quite different perspective. I adopt the more conservative approach of focusing only on observed differences, leaving to others possible X-efficiency interpretations. Note that the fixed effects interpretation does not apply to data envelopment analysis, which is not an econometric method.

<sup>3</sup> This is not the first paper to attempt to explain these inefficiencies. In an excellent article, Berger and Mester (1997) attempt to solve this puzzle. Another ingenious approach is in Berger, Leusner, and Mingo (1994), in which a generic cost function approach is combined with a unique dataset on bank branches to estimate a specific type of potential inefficiency.

<sup>4</sup> For example, in a long-run competitive equilibrium, we expect that (i) bank managers would on average match capacity to demand (though more on this below); (ii) bank managers would manage risk relative to reward equally efficiently; and (iii) markets would lead to different banks offering different quality levels but that profitability would be about the same across the quality spectrum. We explicitly do not impose these conditions on our estimation.

<sup>5</sup> A rigorous economic “explanation” of observed phenomena would involve constructing a game in which these phenomena are equilibrium outcomes. This is beyond the scope of this paper.

<sup>6</sup> The geographic basis of competition is a limitation of this model. It clearly applies to much of traditional banking which was locally based. However, it clearly does not apply to the so-called money center banks, and its application to the newer “super-regionals” is questionable. In the empirical results, the unit of analysis is the state bank (the legal reporting unit), not the holding company. In the case of super-regional banks, changing the ownership of a previously independent state bank catering to consumers and middle-market firms to a new holding company is unlikely to change its essentially geographic focus, so that the MSO assumption seems a good approximation. In the case of money center banks, most such banks have their headquarters in New York, so that the New York/New Jersey MSO captures this class of banks which in fact compete with each other. Unfortunately, it also captures non-money center banks whose market really is New York/New Jersey consumers and middle-market firms. While this anomaly involves relatively few firms, they are among the most important, so that the results must be interpreted with some caution.

---

<sup>7</sup> Of course, any market data must be historical by its nature. For example, if the stock's beta is measured today using data from one month ago, the beta reflects investors' perceptions of the stock's future prospects as of one month ago, not today.

<sup>8</sup> Again, the benchmark is not zero risk, but rather a zero beta, meaning a zero correlation with the market portfolio. As shall be seen below, some banks actually achieve a negative beta.

<sup>9</sup> This expression was first derived in Rohlfs (1978) as a condition for Ramsey pricing in the presence of demand interdependencies. Rohlfs coined the term "superelasticity" for the term in brackets, by analogy to the role of the self-elasticity in the simpler case of Ramsey pricing with no demand interdependencies.

<sup>10</sup> In the long run, of course, such sub-optimal behavior would not be sustainable in a competitive market. But the maintained hypothesis of this paper is that banking markets are neither competitive nor in the long run.

<sup>11</sup> In order to smooth out variations in market-wide observables, we estimated the difference  $r_m - r_f$  from the data as 0.088, which we use throughout. In addition, several methods for estimating  $g$  were tried, including a linear trend and an exponential trend of the actual data. In both cases, a substantial number of negative growth rates, as well as growth rates that led to a negative value of the firm, were encountered. As a result, the growth rate was not used in computing the risk cost.

<sup>12</sup> Trewess, for Trimmed Resistant Weighted Scatterplot Smooth, is a kernel scatterplot smoother. It is one of a class of nonparametric curve estimation methods, designed to produce a smooth curve across a scatterplot.

<sup>13</sup> The second hypothesis is disconfirmed below.

<sup>14</sup> Let  $r_f$  be the risk-free rate (say, on short-term US Treasury bills) and  $r_m$  be the rate of return on the market as a whole (say, the S&P 500). The bank's  $b$  satisfies the Capital Asset Pricing Model equation:  $s = r_f + b(r_m - r_f)$ .

<sup>15</sup> If we were willing to forego bank-specific coefficient estimates for total-sample estimates, then it is obvious a richer model could be estimated. However, we suspect that the interesting variation is among banks rather than among products; we prefer an approximate solution to the more interesting problem rather than a more precise solution to the less interesting problem.

<sup>16</sup> The annual sample was included to determine if the estimation was sensitive to the elimination of non-end-of-year data. Some have suggested that banks are less serious about quarterly results than they are about end-of-year results, and may even view such intermediate results as "window dressing," shading their estimates optimistically. Our results do not support such allegations.