Working Papers

Research Department

WORKING PAPER NO. 00-5

RECOVERING RISKY TECHNOLOGIES

USING THE ALMOST IDEAL DEMAND SYSTEM:

AN APPLICATION TO U.S. BANKING

Joseph P. Hughes
Department of Economics, Rutgers University

William Lang
Office of the Comptroller of the Currency

Loretta J. Mester
Federal Reserve Bank of Philadelphia
and
Finance Department, The Wharton School, University of Philadelphia

Choon-Geol Moon
Department of Economics, School of Business and Economics, Hanyang University

This draft: June 2000
First draft: July 1997

WORKING PAPER NO. 00-05

RECOVERING RISKY TECHNOLOGIES
USING THE ALMOST IDEAL DEMAND SYSTEM:
AN APPLICATION TO U.S. BANKING*

Joseph P. Hughes
Rutgers University

William Lang
Office of the Comptroller of the Currency

Loretta J. Mester
Federal Reserve Bank of Philadelphia
and
The Wharton School, University of Pennsylvania

Choon-Geol Moon
College of Business and Economics, Hanyang University

This Draft: June 2000
First Draft: July 1997

Correspondence to Hughes at Department of Economics, Rutgers University, New Brunswick, NJ 08901-1248; 732-932-7517; jphughes@rci.rutgers.edu. To Lang at Office of the Comptroller of the Currency, 250 E Street SW, Washington, DC 20219; 202-874-5386; William.Lang@occ.treas.gov. To Mester at Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106-1574; 215-574-3807; Loretta.Mester@PHIL.frb.org. To Moon at Department of Economics, College of Business and Economics, Hanyang University, 17 Haengdang-Dong, Seongdong-Gu, Seoul 133-791, Korea; 82-2-2290-1035; mooncg@unitel.co.kr.

Recovering Risky Technologies Using the Almost Ideal Demand System:
An Application to U.S. Banking

Joseph P. Hughes
Rutgers University

William Lang
Office of the Comptroller of the Currency

Loretta J. Mester
Federal Reserve Bank of Philadelphia
and
The Wharton School, University of Pennsylvania

Choon-Geol Moon
College of Business and Economics, Hanyang University

This Draft: June 2000
First Draft: July 1997

Abstract

We argue for a shift in the focus of modeling production from the traditional assumptions of profit maximization and cost minimization to a more general assumption of managerial utility maximization that can incorporate risk incentives into the analysis of production and recover value-maximizing technologies. We show how this shift can be implemented using the Almost Ideal Demand System. In addition, we suggest a more general way of measuring efficiency that can incorporate a concern for the market value of firms' assets and equity and identify value-maximizing firms. This shift in focus bridges the gap between the risk-incentives literature in banking that ignores the microeconomics of production and the production literature that ignores the relationship between production decisions and risk.

Recovering Risky Technologies Using the Almost Ideal Demand System:
An Application to U.S. Banking

Introduction

Risk is a fundamental calculation in production decisions, and as Modigliani and Miller (1958) note, when risk matters, the objective of profit maximization fails to account for the riskiness of production and the interest rate at which profit is discounted. In contrast, the objective of value maximization ranks production plans not just by their expected profitability but also by the discount rate on profit that is required by their riskiness. Consider two production plans, one more profitable and more risky than the other. If we rank these plans by their expected profit, we ignore the effect of their difference in risk on the discounted value of their expected profit. If the more risky plan requires a higher discount rate, the less profitable but less risky plan may imply the higher market value.

Value-maximizing production plans can also differ from static profit-maximizing plans when production involves the potential for costly episodes of financial distress. In such cases trading current expected profit for reduced risk may be a value-enhancing strategy.[1] Thus, the less profitable but less risky production plan may also imply a higher market value because it reduces the expected cost of financial distress.

Risk is a fundamental calculation in commercial banking. A large literature explores banks' contrasting incentives for risk-taking. On the one hand, safety-net subsidies such as mispriced deposit insurance give banks the incentive to take extra risk while, on the other hand, the potential for costly episodes of financial distress involving liquidity crises, regulatory interventions into bank operations, and the possible forfeiture of a valuable charter gives banks the incentive to reduce risk.[2] While this literature explores the effect of these contrasting incentives on banks' risk-taking, it generally ignores the microeconomics of bank production. An equally large literature uses dual cost and profit functions to

---

[1]Tufano (1996) reviews the literature on when trading expected profit for reduced risk is a value-maximizing investment strategy.

[2]See, for example, Merton (1977), Marcus (1984), Grossman (1992), Demsetz, Saidenberg, and Strahan (1996), and Keeley (1990).

investigate the microeconomics of bank production—to measure scale and scope economies and to gauge managerial efficiency.[3] While this literature models production decisions, its reliance on the assumptions of cost minimization and profit maximization leaves it without an analytical framework to incorporate risk-taking and such risk-related phenomena as diversification and moral hazard.

The standard production literature based on profit maximization and cost minimization also lacks a logical apparatus to model *inefficient* production decisions. Although it uses various techniques to measure inefficiency—the failure to achieve maximum profit or minimum cost—its logical framework cannot explain inefficiency. This failure is related to its neglect of risk.

Risk is also a fundamental calculation in many agency relationships and agency problems that result in managerial inefficiency. For example, managers who act in the interests of shareholders may make production decisions that entail extra risk not priced by debtholders and, in doing so, transfer value from debtholders to shareholders.[4] Entrenched managers, whose firm-specific human capital leaves their wealth poorly diversified, may choose less risky production plans to avoid endangering their jobs—enhancing their personal wealth at the expense of their employers' wealth.[5] Less skilled managers who are not entrenched—whose jobs depend on a good realization of profit but whose lower skill reduces the expected profit attained at any level of risk— may take extra risk to attain an acceptable expected value of profit. As long as the risk they take cannot be observed perfectly by the owners, their strategy can maximize their own wealth at the expense of their employers.[6] In all these cases, inefficient production decisions involve risk calculations. The standard analysis of production based on profit maximization and cost minimization is logically unable to explain them. Nor is it able to explain inefficiency due to managerial shirking and expense preferences.

---

[3]See Berger and Mester (1997) for a review of the production literature.

[4]See Calomiris and Kahn (1991) and Flannery (1994) for an analysis of risk shifting in bank production.

[5]See, for example, Saunders, Strock, and Travlos (1990) for an analysis of ownership agency problems in banking.

[6]See Gorton and Rosen (1995).

A. Bridging the Gap Between the Risk-Incentives Literature and the Microeconomics-of-Production Literature

The gap between the risk-incentives literature and the production literature can be bridged by incorporating risk into models of bank production so that the fundamental explanatory framework is logically able to analyze risk-based phenomena such as diversification and moral hazard. Building this bridge requires **a generalization of the managerial objective function** so that it can explain value-maximizing production decisions as well as profit-maximizing decisions—and even decisions that fail to maximize value or profit. And if the generalization of the managerial objective function can explain inefficient decisions, the analysis must also define **an efficiency standard that identifies value-maximizing production plans** and gauges the opportunity cost of firms' inefficiency.

B. Explaining Value Maximization and Inefficiency

In this paper we describe a research program that seeks to shift the focus in modeling bank production from profit maximization to value maximization. We review and synthesize the ideas from a number of papers and present results from an application of some of these ideas to a sample of commercial banks. The research program investigates how to incorporate risk into the analysis of production so that it can explain risk-related phenomena, which include phenomena that result in managerial inefficiency. The program consists of two basic components.

The first component builds a generalized managerial objective function so that managers can rank production plans by higher moments of their implied probability distributions of profit to allow for risk and agency issues to influence their decisions. The goal of this generalization is to build a theoretically consistent model that allows us to explain firms' observed production decisions—even inefficient ones. In contrast, the commonly used maximum profit function explains decisions when risk and inefficiency play no role.

The second component builds upon the infrastructure of the generalized managerial objective function to address issues of firm inefficiency. This approach shifts the focus of efficiency measurement

from fitting maximum profit and minimum cost functions as frontiers to fitting frontiers that identify value-maximizing firms. This can be done directly with the market values of firms that are publicly traded, or it can be done indirectly with the profit predicted by the generalized managerial objective function. By fitting a frontier of *predicted* profit, rather than fitting the profit function itself as a frontier, each firm's *expected profitability* can be compared to the expected profitability of firms in a varied set of peer groups. These peer groups could be defined by a measure of risk that is a proxy for the discount rate or by a measure of size or even by the arguments of the maximum profit function. To the extent that the predicted profit and a proxy for the discount rate can explain market value, this method incorporates market value in its measure of efficiency and can identify firms whose production decisions approximate value-maximizing decisions.

Thus, the **explanatory goal** of this research program involves not only incorporating risk and inefficiency into firms' decision-making but also developing empirical methods of measuring the efficient frontier and inefficiency. Since the standard profit function assumes that profit is maximized, its logic potentially applies only to the behavior of the firms "close" to the frontier—the "efficient" firms. In contrast, the generalized managerial objective function is potentially able to explain the behavior of all firms because it does not impose efficiency as its managers' objective. And by generalizing efficiency measurement to incorporate market value, firms whose production decisions approximate the goal of value maximization can be identified.

The theoretical framework and the empirical tools that constitute this program have been used to investigate a variety of questions in bank production. We review this literature and its evolution, and we turn to an estimation, some of whose results appeared in an unpublished working paper (Hughes, Lang, Mester, and Moon, 1995),[7] which launched these applications. This empirical application differs from any of those that followed. Here, we update the interpretation of its findings and present new empirical results to show how our framework bridges the gap between the risk-incentives literature and the microeconomics-of-production literature.

---

[7]Hereafter, we shall use HLMM to designate these coauthors.

I. Generalizing the Managerial Objective Function

A. Generalizing Managers' Rankings of Production Plans

Our strategy to incorporate risk into the firm's choice of production plans links the production plan to a subjective, conditional distribution of profit. The technology defines all feasible production plans. Each of these plans is linked to a subjective, conditional probability distribution of profit by managers' beliefs about the probability distribution of future economic states and about how these states interact with production plans to generate profit. Given these beliefs, a firm's choice of production plan is equivalent to a choice of a conditional probability distribution of profit.

We attribute to the firm's managers a ranking of production plans that reflects their objectives. Managers who maximize profit rank plans by their expected profitability—the first moment of their conditional probability distribution of profit. Managers who maximize the firm's value, or who trade the firm's value for personal objectives, also consider higher moments of the plans' conditional probability distributions.

To identify the firm's highest ranked production plan, we represent its ranking of plans by a managerial utility function defined over the production plan and profit. This allows managers to rank plans based on expected profit and the variance of profit, the first two moments of the conditional probability distribution of profit. But it also allows for more general utility functions that include even higher moments. Firms maximize this utility function to obtain their *most preferred production plan* (for inputs and outputs) and their *most preferred profit function*.[8] These choice functions are sufficiently general to subsume profit maximization as well as value maximization and to allow managers to trade value or profit for personal objectives. Hence, they can capture the effects of distress costs as well as safety-net subsidies on banks' production decisions, and they can explain inefficiency, too. Since the logic of the model permits managers to choose plans that do not maximize the value of the firm, a technique of

---

[8]This framework for modeling generalized managerial preferences draws on earlier work by Hughes (1989, 1990) on hospitals and education that allows managers to choose production plans that trade profit for other objectives.

efficiency measurement that identifies value-maximizing firms is required.

B. Specifying Banking Technology

We define a bank's production plan by its output vector, y, which consists of asset categories, such as commercial and industrial loans, consumer loans, real estate loans, and government securities. The production plan also includes the input vector, x, and equity capital, k. Inputs comprise sources of loanable funds, insured and uninsured deposits and other borrowed money, and the labor and physical capital used in intermediating the loanable funds. Thus, a production plan, (y, x, k), consists of the portfolio of loans and securities and the inputs and financial capital used to produce the portfolio. The output price vector, p, represents the interest rates charged on the different components of the asset portfolio, y. The higher the interest rate charged, relative to the risk-free rate, r, the higher the asset's risk premium. We use the average contractual return on assets, p, as a measure of *ex ante* asset quality and the amount of nonperforming loans, n, as a measure of *ex post* asset quality.

Other components of the bank's price environment include w, the input price vector; w_k, the required return on equity capital; and m, income from sources other than the output vector, y, which is noninterest and fee income. Letting t be the tax rate on profit and $p_\pi = 1/(1-t)$, the price of a dollar of after-tax profit in terms of before-tax dollars, the before-tax accounting profit is defined as $p_\pi \pi = p\,y + m - w\,x$.

We represent banking technology by a transformation function, $T(y, x, n, p, r, k) \leq 0$, which defines the feasible production plans, augmented by the three variables (n, p, r) characterizing asset quality.

C. Maximizing a Generalized Managerial Utility Function

Managers of banks rank and choose production plans. Profit-maximizing managers rank plans only by the first moment of their implied probability distributions of profit. Value-maximizing managers and managers pursuing personal goals are influenced by higher moments as well. To allow for the

possibility that higher moments may influence rankings, we represent these **generalized managerial preferences** by a utility function, $U(\pi, \boldsymbol{y}, n, p, r, \boldsymbol{x}, k)$. Managers' beliefs about how states of the world, s, interact with the production plan to determine a realization of profit, $\pi = g(\boldsymbol{y}, n, p, \boldsymbol{x}, k, s)$, plus their beliefs about the probability distribution of states of the world, determine a subjective distribution of profit, $f(\pi; \boldsymbol{y}, n, p, r, \boldsymbol{x}, k)$, that is conditional on the production plan.

Hughes and Moon (1995) demonstrate that this representation of the managerial utility function is a more general way of representing preferences than utility defined over expected profit and profit risk, $U(E(\pi), S(\pi))$. Only under certain restrictive conditions can the distributions $f(\pi; \boldsymbol{y}, n, p, r, \boldsymbol{x}, k)$ be represented by their first two moments, $E(\pi; \boldsymbol{y}, n, p, r, \boldsymbol{x}, k)$ and $S(\pi; \boldsymbol{y}, n, p, r, \boldsymbol{x}, k)$, and so managers' rankings of production plans can be represented by a utility function defined over these moments, $U(E(\cdot), S(\cdot))$. A more general representation allows higher moments to influence preferences and is given by utility defined over the distribution, $f(\cdot)$: $U(\pi; \boldsymbol{y}, n, p, r, \boldsymbol{x}, k)$.

This generalized managerial utility function that ranks subjective conditional distributions of profit is sufficiently general to allow managers to maximize profit, to maximize value, or to maximize some other objective that results from agency problems. When they maximize profit, only the first moment influences the ranking or, equivalently, only π has marginal significance in the function. Hence, the production plan does not directly influence utility except through π. When risk or, say, managerial expense preferences influence rankings, higher moments matter, that is, the production plan influences utility directly and also indirectly through its effect on π.

The **generalized managerial objective function** assumes that managers choose their most preferred production plan or, equivalently, their most preferred subjective conditional probability distribution of profit. Hence, managers maximize the generalized utility function subject to the profit identity and the transformation function:

(1a)
$$\max_{\pi, \boldsymbol{x}} U(\pi, \boldsymbol{x}; \boldsymbol{y}, n, p, r, k)$$

(1b)
$$\text{s.t. } p_\pi \pi = \boldsymbol{p}\,\boldsymbol{y} + m - \boldsymbol{w}\,\boldsymbol{x}$$

(1c) $\qquad\qquad\qquad\qquad T(\boldsymbol{y}, n, p, \boldsymbol{x}, k) \leq 0.$

Note this problem is conditioned on the output vector, $\boldsymbol{y}$, to facilitate the computation of scale economies and on equity capital, k, to allow the profit demand function to be normalized by equity capital to obtain the rate of return on equity. The solution consists of the most preferred profit function, $\pi^* = \pi(\boldsymbol{y}, n, \boldsymbol{v}, m, k)$, and the most preferred input demand functions, $\boldsymbol{x}^* = \boldsymbol{x}(\boldsymbol{y}, n, \boldsymbol{v}, m, k)$, where $\boldsymbol{v} = (\boldsymbol{w}, p, r, p_\pi)$.

The profit function, $\pi(\boldsymbol{y}, n, \boldsymbol{v}, m, k)$, is not necessarily the standard maximum profit function. To the extent that risk matters in production or that managers have expense preferences, it captures these considerations. In this sense, it is a **generalized profit function**. Note, as well, that the tax rate on profit can influence the solution, which is not consistent with profit maximization but can result from value maximization.

II. Using the Almost Ideal Demand System to Recover Managers' Rankings

To implement the model, we adapt the Almost Ideal Demand System developed by Deaton and Muellbauer (1980) to estimate consumers' demands and, hence, to recover their rankings of goods and services from budget data. By applying the AID System to estimate the most preferred profit and input demand functions, we recover managers' rankings of production plans. As noted above, these rankings may be influenced by risk.

A. Estimating the Share Equations and First-Order Condition for Equity Capital

We adapt the expenditure function of the AID System to represent generalized managerial preferences and use it to derive the functional forms for the utility-maximizing demands for profit and the production plan. These demand functions take the form of share equations and are derived by applying Shephard's Lemma to the managerial expenditure function, which is dual to the utility maximization problem (1a-1c). The details of their derivation can be found elsewhere (e.g., HLMM 1996) so we do not repeat them here. Appendix 2 lays out the steps taken in the estimation of the model. The theoretical

derivation of the demands results in shares expressed in terms of total revenue, $\mathbf{p \cdot y} + m$:

$$(2a) \qquad \frac{p_\pi \pi}{\mathbf{p \cdot y} + m} = \frac{\partial \ln \mathbf{P}}{\partial \ln p_\pi} + \mu \left[\ln (\mathbf{p \cdot y} + m) - \ln \mathbf{P} \right].$$

$$(2b) \qquad \frac{w_i x_i}{\mathbf{p \cdot y} + m} = \frac{\partial \ln \mathbf{P}}{\partial \ln w_i} + v_i \left[\ln (\mathbf{p \cdot y} + m) - \ln \mathbf{P} \right] \quad \forall \, i$$

where $\ln \mathbf{P} = \alpha_0 + \sum_i \alpha_i \, \ln z_i + (\tfrac{1}{2}) \sum_i \sum_j \alpha_{ij} \, \ln z_i \, \ln z_j$, and $z = (y, n, v, k)$, and $v = (w, p, r, p_\pi)$. In addition to the share equations, a first-order condition defining the optimal level of equity capital, k, is added to the equations to be estimated. The first-order condition constitutes a restriction on the parameters. The details of its derivation as well as the adding-up, homogeneity, and symmetry restrictions are found in HLMM (1996).

In addition to these regularity conditions, HLMM (1996) derive conditions implied by the comparative-statics of profit maximization that permit a **test of whether the production data are consistent with the assumption of profit maximization**. When these conditions are imposed or are satisfied by the data, equation (2a) becomes identical to the translog profit function and equation (2b) is transformed into the associated translog share equations.[9]

Just as consumers' preferences change over time, managers' beliefs that constitute their rankings of production plans can be expected to change, too. To minimize this problem, cross-sectional data are used, and the production system is estimated using nonlinear two-stage least squares, a generalized method of moments.

B. Applying the AID System to Estimate Scale Economies

The AID System constituted by equations (2a) and (2b) and by the first-order condition defining the optimal demand for equity capital has been used several times to estimate scale economies in U.S. commercial banking. Again, we refer the reader to HLMM (1996) for the details of how scale economies

---

[9]Although the translog functional form is nested in the $\ln \mathbf{P}$ expression of the AID System, any other flexible functional form could be substituted for it.

are computed from the profit share equation (2b). These applications all find large scale economies that increase with the size of the bank—evidence consistent with the view that historical restrictions on branching in the U.S. created a disequilibrium that the recent wave of mergers is resolving. HLMM (1996) estimated scale economies for highest-level U.S. bank holding companies (BHCs) in 1994. These BHCs ranged in asset size from $33 million to $250 billion. They found statistically significant mean scale economies of 1.14 that ranged from 1.12 for the smallest size group to 1.25 for BHCs that exceed $50 billion.[10] They tested the restrictions implied by profit maximization and rejected them. DeYoung, Hughes, and Moon (forthcoming) estimated the AID System with 1994 data on U.S. national banks to study the relationship of banks' CAMEL ratings to their production decisions. This data set included smaller banks than those in HLMM (1996). Although they did not report their estimates of scale economies, they found statistically significant mean scale economies of 1.12 that ranged from 1.08 for the smallest banks to 1.21 for banks larger than $50 billion in assets.

III. Assessing the Effect of Risk on Production Decisions

We now discuss a particular estimation of the AID System.[11] This empirical application differs from any of those that followed in its focus on how risk affects production decisions—in particular, levels of capitalization and volatile funding. We update its findings and their interpretation to show how our framework bridges the gap between the risk-incentives literature and the microeconomics-of-production literature. Our application uses 1990 data on U.S. banks with assets exceeding $1 billion. Compared to samples used by the studies described above, this sample has no small banks, and because of the earlier year and the use of individual banks rather than holding companies, it has no banks larger than $70 billion. Appendix 1 gives the definitions of the variables used; Appendix 2 lays out the empirical model that was estimated; Tables 1 and 2 summarize the data; Table 3 reports the estimates of scale economies and

---

[10]We measure scale economies by the inverse elasticity of cost with respect to the outputs. Hence, values greater than one imply scale economies while values less than one, diseconomies of scale.

[11]Some of the empirical results discussed below are also found in our unpublished working paper HLMM (1995).

elasticities of cost with respect to output and equity capital, elasticities of equity capital with respect to

output price and nonperforming loans, and elasticities of volatile funds (i.e., fed funds, repos, and other

borrowed money) with respect to equity capital and nonperforming loans. We consider how these input

demand elasticities are influenced by risk-taking.

Figure 1 plots each bank's estimated scale economies versus its asset size. As shown in Table 3,

the estimates of scale economies are statistically significant: the sample mean is 1.15 and the estimates

range from 1.10 in the smallest asset-size quartile to 1.21 in the largest. These results are consistent with

those of other studies using the AID System.

A unique feature of the AID System is that it allows derivation of the utility-maximizing level of

equity capital. Including equity capital in the modeling of bank technology is becoming more common,

but it is usually treated as a conditioning argument.[12] When the maximization problem is conditioned on

the level of equity capital, the sensitivity of input demands and output supplies to variations in equity

capital can be calculated. In the formulation of the AID System in (2a) and (2b), the levels of assets

(outputs) are also conditioning arguments. Hence, a variation in equity capital, holding the levels of assets

constant, constitutes a variation in the bank's capital-to-asset ratio, an important component of the bank's

riskiness. Thus, conditioning the demands on equity capital allows the researcher to investigate the

marginal effects of this component of risk. And when equity capital is made endogenous, an important

component of risk-taking can be explained, which is the purpose of including the first-order condition

defining the utility-maximizing capitalization in the model to be estimated.

Table 3 reports the sensitivity of the utility-maximizing demands of equity capital to changes in the

ex ante and *ex post* measures of asset quality. Remember that the asset levels are conditioning arguments

so that the response of equity capital to variations in asset quality is equivalent to the response of the

capital-to-asset ratio. Table 3 also reports the sensitivity of the demand for volatile funding, x_4, to asset

quality, which can be interpreted as the sensitivity of an important component of liquidity risk. The

sensitivity of the demand for volatile funds to a variation in equity capital ($\partial ln\ x_4/\partial ln\ k$) is negative and

---

[12]See Hughes, Mester, and Moon (2000) for a detailed analysis of this issue and a review of studies that have included equity capital in bank technology.

statistically significant for banks in all four size groups. Thus, a decrease in capital or, equivalently, in the capital-to-asset ratio is associated with an increase in the demand for volatile funds. Apparently, banks that accept more insolvency risk from a reduced capital ratio are also likely to accept more liquidity risk. However, in all size groups except the smallest, banks with a higher level of nonperforming loans, n, or, equivalently, a higher ratio of nonperforming loans to total assets rely less on volatile funds. Apparently, the difference in insolvency risk posed by a decrease in equity capital and by an increase in nonperforming loans is not symmetric. Perhaps the realization of nonperformance is more threatening to short-term creditors who supply volatile funds.

Table 3 also reports that, for the smallest banks, the utility-maximizing level of capital responds positively to the level of nonperforming loans. These banks may be acting to protect their solvency from the *ex post* realization of poor loan quality by increasing their capitalization, or their regulators may be requiring these banks to increase their capital levels.

IV. Assessing the Effect of Risk on Scale Economies

Most studies of banks' scale economies find either constant returns or slightly decreasing returns for large banks—a result that contradicts the usual rationale given for the megamergers of banks. Why does the AID System so consistently uncover the evidence of scale economies that has eluded so many studies that use the standard cost function? Hughes (1999) and Hughes, Mester, and Moon (2000) (hereafter denoted HMM) contend that banks' elusive scale economies are found on the utility-maximizing expansion path. The key difference between the **utility-maximizing expansion path** of the AID System and the **cost-minimizing expansion path** of the standard cost function is that the AID System allows risk to influence production decisions. If risk influences banks' production decisions, as the risk-incentives literature emphasizes, then the cost-minimizing expansion path misspecifies the bank's equilibrium expansion path. And this point is confirmed by the AID System's consistent rejection of the conditions for profit maximization.

We return to the empirical application of the previous section—the estimation of the AID System

with 1990 data on large U.S. banks—and we apply a Wald test to the restrictions implied by the
assumption of profit maximization. If banks maximize profit, a variation in the tax rate, t, and therefore,
p_π, will not affect the bank's choice of before-tax profit. In addition, the revenue and risk characteristics of
production, represented by output prices, p, will not influence the bank's cost-minimizing production plan,
nor will income from sources other than the output vector, m. In all, there are 31 restrictions. (However,
two restrictions are redundant because of the adding-up conditions these parameters satisfy.) The
restrictions are strongly rejected by the data: the value of the test statistic, which is distributed χ^2, is
294.01 with 29 degrees of freedom. Thus, the behavior of banks in the sample is not consistent with profit
maximization.

Next, we ask what measures of scale economies we would obtain if we imposed these restrictions.
We answer this question by comparing the results of the AID System with those obtained from the
standard cost function estimated with the same data. For this standard cost function model, we drop equity
capital and the measures of asset quality, and we use the standard translog cost formulation, including
input share equations. We find that the estimated scale economies over asset-size categories of banks are
all significantly different from one but are considerably smaller and remarkably familiar: ranging from
1.022 for the smallest asset-size quartile, 1.029 for the second quartile, 1.035 for the third quartile, and
1.050 for the largest quartile.

HMM, using 1994 BHC data from HLMM (1996), also present evidence showing that the first-
order conditions defining the cost-minimizing demands for insured deposits, uninsured deposits, and
equity capital are violated. Notably, most smaller banks overemploy equity capital while many larger
banks underemploy capital—a result consistent with the contrasting incentives for risk-taking emphasized
in the risk-incentives literature. Larger banks may, on balance, be exploiting safety-net subsidies by
reducing their capital ratios while smaller banks may be protecting their charter values with higher capital
ratios.

The evidence from all of these studies strongly suggests that risk-taking incentives influence
banks' optimal expansion path and that the strategy of measuring scale economies along the cost-

minimizing expansion path fails to capture the **empirically relevant equilibrium expansion path**. What is the empirically relevant equilibrium path? This path should be able to explain firms' decisions that respond to incentives for risk-taking. These incentives reflect the influence of risk on the market value of firms' equity and debt and on the personal wealth and utility of its managers. HMM contend that the AID System's utility-maximizing expansion path captures these effects when it measures scale economies, because it allows risk and expense preference to influence managers' rankings of production plans. They provide empirical evidence showing how risk, diversification, and inefficiency affect scale economies.

Thus, we contend that measuring bank technology requires a shift in the focus from the assumption of profit maximization and cost minimization to a focus on a more powerful explanatory framework based on managerial utility maximization. Rather than impose profit-maximizing efficiency on the data, we want to allow the data to speak more generally about the objectives of managers.

Allowing the data to speak generally about the objectives of managers does not mean that the concept of efficiency is discarded—only that standards of efficiency are used to evaluate the *predicted* behavior of the utility-maximizing equilibrium rather than imposed on the equilibrium itself. This principle stands in contrast to the practice of estimating a maximum profit function as a frontier to identify efficient practice. The behavioral function itself cannot explain inefficient production decisions and may not even be consistent with the production decisions of the firms it labels as relatively efficient.

V. Gauging Efficiency and Identifying Value-Maximizing Production Plans

HMM note that the empirically relevant expansion path, then, is not necessarily the efficient expansion path. How do we identify the efficient expansion path? What standard of efficiency do we use? We may be interested in measuring scale economies along the expansion path that maximizes the value of banks' equity. But if the managers, acting in the interest of the owners of their banks, shift risk to debtholders, they reduce the value of the debt and increase the value of the equity. Hence, we may instead be interested in measuring scale economies along the expansion path that maximizes the value of the

banks' assets. And, of course, we may be interested in measuring scale economies along the cost-minimizing expansion path. Any of these standards could be designated as X-efficiency. Once an efficiency standard is chosen, the question is, how can we identify the **X-efficient expansion path**?

A traditional answer has been to estimate the managers' objective function as a frontier, but this solution, we have argued, may not identify an empirically relevant expansion path—certainly not for the firms that are relatively inefficient by the standard of its objective function. On the other hand, the utility-maximizing expansion path is more likely to be empirically relevant but not necessarily X-efficient (unless efficiency is defined by the managers' own utility). Since we are not interested in defining efficiency by managers' personal welfare, it makes no sense to estimate the **generalized profit function** of the AID System as a frontier.

We describe the strategy proposed by HMM. HMM identify X-efficient firms by fitting a frontier to the X-measure of performance. They use three measures of efficiency related to the market value of banks. For publicly trade BHCs, one efficiency measure is derived from a stochastic frontier relating the market value of assets to the book value of assets; another efficiency measure is derived from a stochastic frontier relating the market value of equity to the book value of equity.[13] A third efficiency measure can be computed for the full sample of BHCs, regardless of whether they are publicly traded. This measure is derived from a stochastic frontier relating expected profit based on the generalized profit function to market-priced risk. Thus, this efficiency measure indicates the BHC's lost potential profit given its level of profit risk. HLMM (1999) show that the profit predicted by the generalized profit function and a production-based measure of risk (the econometric prediction risk of this profit) are systematically related to the market value of equity for the subsample of publicly traded banks. Hence, this third measure of efficiency is an empirically good measure of market-value efficiency. HMM identify the most efficient quarter of their BHCs using each of these efficiency measures. The utility-maximizing expansion paths of these BHCs are used to approximate the value-maximizing expansion path, and mean scale economies are estimated for these efficient BHCs by asset-size group.

---

[13]These two techniques were developed by Hughes, Lang, Moon, and Pagano (1997) and applied by HLMM (1999) to study the efficiency effects of bank consolidation.

A. An Application

We return now to our empirical application—the estimation of the AID System with 1990 data on large U.S. banks—and apply the methodology of HMM to measure mean scale economies on the value-maximizing expansion path. First, we compute the third efficiency measure described above for our sample of banks. For each bank we compute its expected return by dividing the profit predicted for the bank from the estimated generalized profit function (2a) by its capital level, k. So $ER = E(p_\pi \pi)/k$. The bank's risk is proxied by the standard error of predicted profit divided by k, i.e., $RK = S(E(p_\pi \pi))/k$. Then a stochastic frontier is estimated regressing ER on RK and RK^2:

(3) $\qquad ER_i = \gamma_0 + \gamma_1 RK_i + \gamma_2 RK_i^2 + \epsilon_i$

where $\epsilon_i \equiv v_i - \mu_i$ is a composite error term, with v_i normally distributed with a zero mean, and μ_i positive and half-normally distributed. Bank i's inefficiency is measured by the conditional mean of μ_i given ϵ_i, $E(\mu_i/\epsilon_i)$.

Once bank efficiency is measured, we identify the quarter of banks that are most efficient in the sample and in each of the four asset-size quartiles of banks. As in HMM, these banks' utility-maximizing expansion path is considered to approximate the value-maximizing expansion path, and the mean of the estimated scale economies of these banks is considered to approximate scale economies along the value-maximizing expansion path. As shown in Table 4, the results for the value-maximizing expansion path remain qualitatively similar to the results for the utility-maximizing expansion path, with banks in all size quartiles exhibiting strong scale economies.

VI. Generalizing Efficiency Measurement

Each of the efficiency techniques described in the previous section has a common general form. Letting X_i denote a measure of the financial performance of the i-th firm, such as the market value of its assets or its profit predicted by the generalized profit function, and G_i denote a measure defining the peer group used to compare the i-th firm's financial performance, such as risk or the replacement cost of assets, we can represent this general form by the definition of the stochastic frontier that gives the highest

potential value of X_i given G_i:

(4) $$X_i = \alpha_0 + \alpha_1 G_i + \alpha_2 (G_i)^2 + \epsilon_i$$

where $\epsilon_i \equiv v_i - \mu_i$ is an error term composed of two terms: v_i, which is normally distributed with a zero mean, and μ_i, which is positive and half-normally distributed when the frontier is fitted as an upper envelope. The deterministic kernel of the stochastic frontier is $[\alpha_0 + \alpha_1 G_i + \alpha_2 (G_i)^2]$. The i-th bank's inefficiency is measured by the conditional mean of μ_i given ϵ_i, $E(\mu_i/\epsilon_i)$.

For example, we can measure banks' financial performance, X, by their profit predicted by the estimated generalized profit function. The frontier gives the upper envelope of X, profit, relative to the peer grouping G. What is the appropriate peer grouping? A measure of risk is one candidate, say, for example, the econometric prediction risk of X, a measure of the bank's market beta and perhaps its interest rate beta, or a historical measure of the risk of X. Since X in this case is defined as accounting profit, another peer group could be the level of equity or the level of assets. In the latter case, the frontier defines the maximum potential profit for any given size of the bank, and in the former case, the maximum given the equity investment. If risk defines the peer group, the choice of risk is excluded from the gauge of efficiency. Similarly, if the investment in assets or equity defines peers, their optimality is not taken into account.

The choice of the appropriate measure of financial performance and of peer groupings depends on the objective of the investigation. To measure scale economies along the value-maximizing expansion path, HMM contend that the market-value measure of asset efficiency is well suited to the task because it accounts both for *discounted* profit as well as size of the bank. It answers the question, given the asset size of the bank, what is its highest potential *discounted* profit? Hence, it measures inefficiency due to suboptimal risk as well as suboptimal profit. Its failure to account for inefficiency due to a suboptimal size is no drawback when the objective is to measure scale economies at any given size.

In summary, the AID System and its generalized profit function afford a high level of explanatory generality in modeling production decisions. Although this generality requires the scrapping of the traditional profit function and its associated efficiency techniques, it does not require abandoning

efficiency analysis. Instead, it offers a far richer set of efficiency concepts that include a concern for market values of assets and equity. In fact, it allows the same definition of peer groups as the standard maximum profit function fitted as a frontier.

Instead of using the profit predicted by the generalized profit function to fit (4), consider fitting (4) as a maximum profit function. The peer grouping, G, is a vector of the usual arguments of the maximum profit function. If we use the arguments of the nonstandard maximum profit function, the peer grouping controls for the output vector, y, and the input price vector, w.[14] In this case, peers are defined by the composition of their output vector as well as their input prices, and profit inefficiency due to suboptimal choices of outputs is excluded from the measure of inefficiency. In contrast, if we fit (4) using the profit predicted by the generalized profit function and if we use the same peer grouping as the nonstandard profit function, we obtain a measure of profit efficiency that differs conceptually from that obtained by the maximum profit function in a critically important way: **both banks' predicted profit and the frontier fitted to it are derived from a model of managerial behavior that can potentially explain all observations–inefficient as well as efficient**.

VII. Conclusions

We have argued for a shift in the focus of modeling production from the traditional assumptions of profit maximization and cost minimization to a more general assumption of managerial utility maximization, and we have suggested more general ways of measuring efficiency that can incorporate a concern for the market value of firms' assets and equity. This shift in focus bridges the gap between the risk-incentives literature in banking that ignores the microeconomics of production and the production literature that ignores the relationship between production decisions and risk.

---

[14]For details of the nonstandard profit function, see Humphrey and Pulley (1997).

Bibliography

Berger, Allen N., and Loretta J. Mester, 1997, "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance*, 21, 895-947.

Calomiris, Charles W., and Charles M. Kahn, 1991, "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *American Economic Review*, 81, 497-513.

Deaton, Angus, and John Muellbauer, 1980, "An Almost Ideal Demand System," *American Economic Review*, 70, 312-326.

Demsetz, Rebecca S., Marc R. Saidenberg, and Philip E. Strahan, 1996, "Banks with Something to Lose: The Disciplinary Role of Franchise Value," Federal Reserve Bank of New York *Economic Policy Review*, 2, 1-14.

DeYoung, Robert, Joseph P. Hughes, and Choon-Geol Moon, "Efficient Risk-Taking and Regulatory Covenant Enforcement in a Deregulated Banking Industry," forthcoming in the *Journal of Economics and Business*.

Flannery, Mark J., 1994, "Debt Maturity and the Deadweight Cost of Leverage: Optimally Financing Banking Firms," *American Economic Review*, 84, 320-331.

Gorton, Gary, and Richard Rosen, 1995, "Corporate Control, Portfolio Choice, and the Decline of Banking," *Journal of Finance*, 50, 1377-1420.

Grossman, Richard S., 1992, "Deposit Insurance, Regulation, and Moral Hazard in the Thrift Industry: Evidence from the 1930's," *American Economic Review*, 82, 800-821.

Hughes, Joseph P., 1989, "Hospital Cost Functions: The Case Where Revenues Affect Production," Working Paper No. 1990-01, Rutgers University, Department of Economics, November.

Hughes, Joseph P., 1990, "The Theory and Estimation of Revenue-Driven Costs: The Case of Higher Education," unpublished manuscript, Rutgers University, Department of Economics, February.

Hughes, Joseph P., 1999, "Incorporating Risk into the Analysis of Production," *Atlantic Economic Journal*, 27:1, 1-23.

Hughes, Joseph P., William Lang, Loretta J. Mester, and Choon-Geol Moon, 1995, "Recovering

Technologies that Account for Generalized Managerial Preferences: An Application to Non-Risk-Neutral Banks," Wharton Financial Institutions Center, Working Paper 95-16 and Federal Reserve Bank of Philadelphia Working Paper No. 95-8/R.

Hughes, Joseph P., William Lang, Loretta J. Mester, and Choon-Geol Moon, 1996, "Efficient Banking Under Interstate Branching," *Journal of Money, Credit, and Banking*, 28, 1045-1071.

Hughes, Joseph P., William Lang, Loretta Mester, and Choon-Geol Moon, 1999, "The Dollars and Sense of Bank Consolidation," *Journal of Banking and Finance*, 23, 291-324.

Hughes, Joseph P., William Lang, Choon-Geol Moon, and Michael Pagano, 1997 (revised 1999), "Measuring the Efficiency of Capital Allocation in Commercial Banking," Federal Reserve Bank of Philadelphia, Working Paper 98-2.

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," *Journal of Productivity Analysis*, 4, 292-315.

Hughes, Joseph P., Loretta J. Mester, and Choon-Geol Moon, 2000, "Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Bank Production," Working Paper, Federal Reserve Bank of Philadelphia.

Hughes, Joseph P., and Choon-Geol Moon, 1995, "Measuring Bank Efficiency When Managers Trade Return for Reduced Risk," Working Paper, Rutgers University, Department of Economics.

Humphrey, D.B., and L.B. Pulley, 1997, "Banks' Responses to Deregulation: Profits, Technology, and Efficiency," *Journal of Money, Credit, and Banking*, 29, 73-93.

Keeley, Michael C., 1990, "Deposit Insurance, Risk, and Market Power in Banking," *American Economic Review*, 80, 1183-1200.

Marcus, Alan J., 1984, "Deregulation and Bank Financial Policy," *Journal of Banking and Finance*, 8, 557-565.

Merton, Robert C., 1977, "An Analytic Derivation of the Cost of Deposit Insurance Loan Guarantees," *Journal of Banking and Finance* 1, 3-11.

Modigliani, F., and M.H. Miller, 1958, "The Cost of Capital, Corporation Finance, and the Theory of

Investment," *American Economic Review* 48, 261-297.

Saunders, Anthony, E. Strock, and N.G. Travlos, 1990, "Ownership Structure, Deregulation, and Bank-Risk-Taking," *Journal of Finance*, 45, 643-654.

Tufano, Peter, 1996, "Who Manages Risk? An Empirical Examination of Risk Management Practices in the Gold Mining Industry," *Journal of Finance*, 50, 1097-1137.

van Wissen, Leo and Thomas F. Golob (1992), "A Dynamic Model of Car Fuel-Type Choice and Mobility," *Transportation Research*, 26B, 77-96.

Appendix 1

Data and Definitions of Variables

We estimate the AID System using data on U.S. banks that reported at least $1 billion in assets as of the last quarter of 1998. The data are taken from the Consolidated Reports of Condition and Income for the fourth quarters of 1989 and 1990. Banks in unit banking states and special-purpose banks chartered under Delaware's Financial Center Development Act and Consumer Credit Bank Act are excluded from the sample. A total of 286 banks, ranging in size from $1.025 billion to $69.612 billion, are included in the data set. The data are summarized in Tables 1 and 2.

We specify five outputs, each measured as the average dollar amount in the fourth quarters of 1989 and 1990: y_1, real estate loans, including commercial as well as non-commercial; y_2, commercial and industrial loans, lease financing receivables, and agricultural loans; y_3, loans to individuals for household, family, and other personal expenditures; y_4, other loans (such as loans for purchasing and carrying securities, unplanned overdrafts to deposit accounts, loans to nonprofit institutions, and loans to individuals for investment purposes); y_5, securities, assets in trading accounts, federal funds sold, securities purchased under agreement to resell, and total investment securities.

Equity capital, k, is the average amount of equity capital, loan-loss reserves, and subordinated debt in 1990. Hence, it consists of Tier 1 and Tier 2 capital. In addition to equity capital, five other inputs are incorporated into the model: x_1, labor, whose price, w_1, is measured by salaries and benefits paid in 1990 divided by the average number of employees in 1990; x_2, physical capital, whose price, w_2, is proxied by the ratio of occupancy expense in 1990 to the average dollar value of net bank premises in 1990; x_3, insured deposits, whose price, w_3, is computed as the ratio of interest paid in 1990 on deposits under $100,000, net of service charges received by the bank, to the average dollar amount of interest-bearing deposits net of CDs over $100,000; x_4, other borrowed money, whose price, w_4, is the ratio of the total expense of federal funds purchased, securities sold under agreement to repurchase, obligations to the U.S. Treasury, and other borrowed money in 1990 to the average amount of these funds in 1990; and x_5,

uninsured deposits, whose price, w_5, is the ratio of the interest expense in 1990 of deposits over $100,000 to the average amount of those deposits.

Although some formulations have *assumed* that deposits are outputs, Hughes and Mester (1993) derive a test for determining whether deposits are inputs or outputs.[15] In their data set, which is very similar to the one here, they conclude that *insured and uninsured deposits are inputs*. Hughes, Mester, and Moon (2000) reach the same conclusion with later data on highest-level BHCs. Thus, we treat them as inputs here as well.

The amount of nonperforming loans, n, is measured by the sum of the average level of loans past due 90 days or more and still accruing interest and the average level of nonaccruing loans. The price or yield, p_i, on the i-th output is measured by the ratio of total interest income from the i-th output to the average amount of the i-th output that is accruing interest. This price is not just a component of revenue. Its magnitude relative to the risk-free rate indicates the risk premium incurred by the output and, hence, suggests the average quality of the asset. To reduce the number of parameters to be estimated, the vector of output prices is replaced by the weighted average of output prices, $\tilde{p} = \Sigma_i\, p_i\, (y_i / \Sigma_j\, y_j)$.

The variable, m, is measured by the amount of noninterest income received in 1990. Revenue is the sum, $\boldsymbol{p \cdot y} + m$. The variable, t, is the tax rate on profit. For each bank, It is the sum of the federal tax rate, which does not vary across banks and the state tax rate in the state in which the bank is headquartered. The federal tax rate is 34% and state tax rates are obtained from *The Book of the States*, published by the Council of State Governments, and from *Significant Aspects of Fiscal Federalism*, published by the U.S. Advisory Commission on Intergovernmental Relations. Once t is constructed, we construct the price of a dollar of after-tax profit in terms of before-tax dollars as $p_\pi = 1/(1-t)$.

---

[15]Hughes and Mester (1993) show that when deposits are inputs (outputs), variable cost (i.e., the cost of all nondeposit inputs) will be decreasing (increasing) in the level of deposits.

Appendix 2

Empirical Model and Estimation

The model we estimate comprises the profit share equation (2a), the input share equations (2b), and the first-order condition for the optimal level of equity capital, k, which is a conditioning argument in the share equation. The profit and input demand functions are shares expressed as shares of total revenue, $\boldsymbol{p\,y} + m$, and sum to one. They are derived by applying Shephard's Lemma to the managerial expenditure function, which is dual to the utility maximization problem (1a-1c). Thus, the model to be estimated is:

$$(A2.1) \qquad \frac{p_\pi \pi}{\boldsymbol{p\cdot y} + m} = \frac{\partial \ln \boldsymbol{P}}{\partial \ln p_\pi} + \mu[\ln(\boldsymbol{p\cdot y} + m) - \ln \boldsymbol{P}].$$

$$(A2.2) \qquad \frac{w_i x_i}{\boldsymbol{p\cdot y} + m} = \frac{\partial \ln \boldsymbol{P}}{\partial \ln w_i} + v_i[\ln(\boldsymbol{p\cdot y} + m) - \ln \boldsymbol{P}] \quad \forall\, i$$

$$(A2.3) \qquad \frac{\partial V(\cdot)}{\partial k} = \frac{\partial V(\cdot)}{\partial \ln k}\frac{\partial \ln k}{\partial k} = 0,$$

where $\ln \mathbf{P} = \alpha_0 + \sum_i \alpha_i \ln z_i + (\tfrac{1}{2})\sum_i \sum_j \alpha_{ij} \ln z_i \ln z_j$, and $z = (\boldsymbol{y}, n, \boldsymbol{v}, k)$, and $\boldsymbol{v} = (\boldsymbol{w}, p, r, p_\pi)$, and

$$(A2.4) \qquad V(\cdot) = \frac{\ln(\boldsymbol{p\cdot y} + m) - \ln P}{\beta_0 \left(\prod_i y_i^{\beta_i}\right)\left(\prod_j w_i^{v_j}\right) p_\pi^\mu k^\kappa}.$$

As discussed in Appendix 1, to save on degrees of freedom, we replaced the vector of output prices, $\boldsymbol{p}$, with the weighted-average output price, $\tilde{p} = \Sigma_i p_i (y_i/\Sigma_j y_j)$. The risk-free rate, r, is the same for all banks, so it is not included is the estimating equations. Thus, written out, the equations to be estimated are:

(A2.1′)

$$\frac{p_\pi \pi}{\boldsymbol{p} \cdot \boldsymbol{y} + m} = \eta_\pi + \eta_{\pi\pi} \ln p_\pi + \psi_{p\pi} \ln \tilde{p} + \sum_j \gamma_{j\pi} \ln y_j + \sum_s \omega_{s\pi} \ln w_s$$

$$+ \eta_{\pi n} \ln n + \eta_{\pi k} \ln k$$

$$+ \mu [\ln (\boldsymbol{p} \cdot \boldsymbol{y} + m) - \ln P]$$

(A2.2′)

$$\frac{w_i x_i}{\boldsymbol{p} \cdot \boldsymbol{y} + m} = \omega_i + \sum_s \omega_{si} \ln w_s + \phi_{pi} \ln \tilde{p} + \sum_j \gamma_{ji} \ln y_j + \omega_{\pi i} \ln p_\pi$$

$$+ \omega_{in} \ln n + \omega_{ik} \ln k$$

$$+ \nu_i [\ln (\boldsymbol{p} \cdot \boldsymbol{y} + m) - \ln P]$$

(A2.3′)

$$\rho + \rho_{kk} \ln k + \psi_{pk} \ln \tilde{p} + \sum_j \gamma_{jk} \ln y_j + \sum_s \omega_{sk} \ln w_s + \eta_{\pi k} \ln p_\pi$$

$$+ \vartheta_{nk} \ln n + \kappa [\ln (\boldsymbol{p} \cdot \boldsymbol{y} + m) - \ln P] = 0$$

where

$$\ln P = \alpha_0 + \alpha_p \ln \tilde{p} + \sum_i \delta_i \ln y_i + \sum_i \omega_i \ln w_j$$

$$+ \eta_\pi \ln p_\pi + \vartheta \ln n + \rho \ln k + \frac{1}{2}\alpha_{pp}(\ln \tilde{p})^2 + \frac{1}{2}\sum_i \sum_j \delta_{ij} \ln y_i \ln y_j$$

$$+ \frac{1}{2}\sum_s \sum_t \omega_{ij}^* \ln w_s \ln w_t + \frac{1}{2}\eta_{\pi\pi}(\ln p_\pi)^2$$

$$+ \frac{1}{2}\vartheta_{nn}(\ln n)^2 + \frac{1}{2}\rho_{kk}(\ln k)^2$$

$$+ \sum_j \theta_{pj} \ln \tilde{p} \ln y_j + \sum_s \phi_{ps} \ln \tilde{p} \ln w_s + \psi_{p\pi} \ln \tilde{p} \ln p_\pi$$

$$+ \psi_{pn} \ln \tilde{p} \ln n + \psi_{pk} \ln \tilde{p} \ln k$$

$$+ \sum_j \sum_s \gamma_{js} \ln y_j \ln w_s \; + \; \sum_j \gamma_{j\pi} \ln y_j \ln p_\pi$$

$$+ \sum_j \gamma_{jn} \ln y_j \ln n \; + \; \sum_j \gamma_{jk} \ln y_j \ln k$$

$$+ \frac{1}{2} \sum_s \omega_{s\pi}^* \ln w_s \ln p_\pi \; + \; \frac{1}{2} \sum_s \omega_{\pi s}^* \ln p_\pi \ln w_s$$

$$+ \sum_s \omega_{sn} \ln w_s \ln n \; + \; \sum_s \omega_{sk} \ln w_s \ln k$$

$$+ \; \eta_{\pi n} \ln p_\pi \ln n \; + \; \eta_{\pi k} \ln p_\pi \ln k$$

$$+ \; \tau_{rk} \ln r \ln k \; + \; \vartheta_{nk} \ln n \ln k \, ,$$

$$\omega_{si} = \frac{1}{2}(\omega_{si}^* + \omega_{is}^*) = \omega_{is} \text{ and } \omega_{s\pi} = \frac{1}{2}(\omega_{s\pi}^* + \omega_{\pi s}^*) = \omega_{\pi s} \, ,$$

and $\qquad p_\pi = 1/(1-t)$.

We impose several conditions on the parameters of the model. <u>Symmetry</u> requires that

(S1) $\delta_{ij} = \delta_{ji} \; \forall \; i,j,$ $\qquad$ (S2) $\omega_{s\pi} = \omega_{\pi s} \; \forall \; s,$ $\qquad$ and (S3) $\omega_{si} = \omega_{is} \; \forall \; s,i.$

(S1) must be imposed in the estimation of the share equations, since the constituent coefficients cannot be separately identified. However, (S2) and (S3) involve coefficients of prices that are used by Shephard's lemma to obtain the share equations. Consequently, they appear in separate share equations and are, thus, identifiable. It is a judgment call as to whether one imposes these symmetry conditions or not. In the consumer demand literature they are typically not imposed, and we follow that tact here.

The input and profit revenue share equations sum to one, which implies the following <u>adding up</u> conditions:

(A1) $\Sigma_i \, \omega_i + \eta_\pi = 1,$ $\qquad\qquad$ (A2) $\Sigma_i \, \omega_{si} + \omega_{s\pi} = 0, \; \forall \; s,$

(A3) $\Sigma_i \, \phi_{pi} + \psi_{p\pi} = 0,$ $\qquad\qquad$ (A4) $\Sigma_i \, \gamma_{ji} + \gamma_{j\pi} = 0 \; \forall \; j,$

(A5) $\Sigma_i \, \omega_{\pi i} + \eta_{\pi\pi} = 0,$ $\qquad\qquad$ (A6) $\Sigma_i \, \omega_{ir} + \eta_{\pi r} = 0,$

(A7) $\Sigma_i \, \omega_{ik} + \eta_{\pi k} = 0,$ $\qquad\qquad$ (A8) $\Sigma_i \, \omega_{in} + \eta_{\pi n} = 0,$ and

(A9) $\Sigma v_j + \mu = 0.$

Notice that (A6) contains coefficients on variables involving the risk-free rate, r. These coefficients are not estimated, since r does not vary across banks, so condition (A6) is not imposed in the estimation.

The input and profit share equations are <u>homogeneous of degree zero</u> in $(w, \tilde{p}, r, p_\pi)$, which implies the following conditions:

(H1) $\Sigma v_j + \mu = 0,$

(H2) $\alpha_p + \Sigma \omega_j + \eta_\pi + \tau = 1,$

(H3) $\alpha_{pp} + \Sigma_t \phi_{pt} + \psi_{pr} + \psi_{p\pi} = 0,$

(H4) $\phi_{pt} + \Sigma_s \omega_{st} + \omega_{tr} + \omega_{\pi t} = 0 \; \forall \; t,$

(H5) $\tau_{rr} + \psi_{pr} + \Sigma_s \omega_{sr} + \eta_{\pi r} = 0,$

(H6) $\theta_{pj} + \Sigma_t \gamma_{jt} + \gamma_{j\pi} + \gamma_{jr} = 0 \; \forall \; j,$

(H7) $\eta_{\pi\pi} + \psi_{p\pi} + \Sigma_s \omega_{s\pi} + \eta_{\pi r} = 0$

(H8) $\psi_{pn} + \Sigma_s \omega_{sn} + \tau_{rn} + \eta_{\pi n} = 0,$

(H9) $\psi_{pk} + \Sigma_s \omega_{sk} + \tau_{rk} + \eta_{\pi k} = 0,$ and

(H10) $\frac{1}{2}\alpha_{pp} + \frac{1}{2}\Sigma_s \Sigma_t \omega_{st} + \Sigma_t \phi_{pt} + \frac{1}{2}\tau_{rr}$

$$+ \frac{1}{2}\eta_{\pi\pi} + \psi_{p\pi} + \psi_{pr} + \Sigma_s \omega_{sr}$$

$$+ \frac{1}{2}\Sigma_s \omega_{s\pi} + \frac{1}{2}\Sigma_s \omega_{\pi s} + \eta_{\pi r} = 0.$$

Note that the only homogeneity condition imposed in the estimation is (H1), which is equivalent to the adding-up condition (A9). The other homogeneity conditions contain coefficients on variables involving the risk-free rate, r. These coefficients are not estimated, since r does not vary across banks, but the homogeneity conditions can be used to recover these coefficients.

To summarize: in estimating the model, we imposed (S1), (A1)-(A5), (A7)-(A8), and (A9) $\equiv$ (H1). The adding-up constraint (A6) amends the homogeneity constraint (H5), so that the last two terms of (H5) drop out. The homogeneity conditions (H2)-(H10) are then used to recover the parameters on variables involving the risk-free rate, r, which does not vary over the cross-section of banks.

We estimated the model using nonlinear two-stage least squares, a generalized method of moments. Starting values were obtained by setting the constant terms, ω_i, in the input share equations at the average value of the input share across banks in the sample, the constant term, η_π, in the profit share equation at the average value of the profit share across banks in the sample, and all other parameters in the input share, profit share, and equity capital demand equation equal to zero.

Table 1: Summary Statistics of the Data

variable	sample mean	sample std. dev.	minimum	maximum
y_1†	1490521.75	2739095.05	1038.50	26541000.00
y_2†	1606152.95	3114969.78	15956.00	23962500.00
y_3†	672172.08	1070269.29	8762.50	11781500.00
y_4†	470269.42	1486811.00	713.00	11868570.50
y_5†	1326574.19	2232463.87	13549.50	20454834.50
p_1‡	0.109	0.017	0.026	0.205
p_2‡	0.107	0.020	0.022	0.187
p_3‡	0.123	0.027	0.028	0.279
p_4‡	0.085	0.065	0.001	0.533
p_5‡	0.088	0.014	0.055	0.174
$\tilde{p}$‡	0.104	0.011	0.064	0.161
w_1¶	33.092	9.857	18.140	92.178
w_2‡	0.396	0.178	0.116	1.378
w_3‡	0.060	0.009	0.028	0.108
w_4‡	0.087	0.033	0.039	0.350
w_5‡	0.081	0.017	0.027	0.233
p_π‡	1.664	0.112	1.515	1.871
s_1	0.146	0.041	0.043	0.309
s_2	0.050	0.018	0.011	0.141
s_3	0.277	0.095	0.010	0.542
s_4	0.101	0.076	0.004	0.469
s_5	0.314	0.128	0.109	1.208
s_π	0.337	0.071	0.122	0.749
$p_\pi \pi$†	223763.03	360436.38	17125.04	3116164.55
$p \cdot y + m$†	667979.16	1020863.51	90000.55	8419110.74
n†	157495.47	387914.46	1254.00	3629843.00
k†	561765.95	1069979.99	69516.50	8787000.00
m†	124239.66	277310.33	488.00	2060000.00
total assets†	5865120.40	8037258.81	1025143.00	69611500.00

† in thousands of dollars   ‡ in dollars per dollar   ¶ in thousands of dollars per employee

y_1 = real estate loans; y_2 = C & I loans; y_3 = loans to individuals; y_4 = other loans (to purchase securities, overdrafts, etc.); y_5 = securities, fed funds sold, repos, assets in trading accounts; p_i = price of output i; $\tilde{p}$ = weighted average of output prices; w_1 = price of labor; w_2 = price of physical capital; w_3 = price of insured deposits; w_4 = price of other borrowed money (repos, fed funds purchased, etc.); w_5 = price of uninsured deposits; p_π = price of real after-tax profit; s_i = input i share; s_π = profit share; $p_\pi \pi$ = nominal before-tax accounting profit; $p \cdot y + m$ = expected revenue; n = nonperforming loans; k = equity capital; m = noninterest income.

Table 2: Means of the Variables by Asset-Size Quartiles

variable	1st quartile	2nd quartile	3rd quartile	4th quartile
y_1†	357760.68	657490.28	1281000.08	3680519.81
y_2†	249271.21	467700.81	952710.22	4780167.50
y_3†	188318.96	326874.65	639945.94	1538866.02
y_4†	32843.49	80126.14	157397.63	1620612.07
y_5†	325200.96	552054.45	964296.82	3480755.74
p_1‡	0.108	0.111	0.112	0.105
p_2‡	0.115	0.106	0.109	0.098
p_3‡	0.124	0.119	0.125	0.125
p_4‡	0.083	0.095	0.082	0.082
p_5‡	0.086	0.085	0.086	0.093
$\tilde{p}$‡	0.105	0.103	0.105	0.101
w_1¶	29.372	30.886	33.052	39.090
w_2‡	0.382	0.387	0.393	0.421
w_3‡	0.060	0.060	0.059	0.060
w_4‡	0.080	0.080	0.086	0.102
w_5‡	0.080	0.081	0.084	0.081
p_π‡	1.651	1.662	1.669	1.675
s_1	0.141	0.144	0.144	0.157
s_2	0.047	0.050	0.048	0.053
s_3	0.316	0.309	0.271	0.211
s_4	0.072	0.072	0.107	0.155
s_5	0.289	0.293	0.297	0.376
s_π	0.333	0.337	0.342	0.337
$p_\pi \pi$†	47708.37	84558.38	169048.99	596467.64
$p \cdot y + m$†	140029.89	249238.77	489462.14	1801597.91
n†	18162.45	49266.10	84958.04	480141.30
k†	106021.51	187606.21	359011.36	1602550.27
m†	18240.58	33365.32	68254.24	379166.96
total assets†	1335138.53	2393107.19	4607230.76	15191623.49

† in thousands of dollar   ‡ in dollars per dollar   ¶ in thousands of dollars per employee

y_1 = real estate loans; y_2 = C & I loans; y_3 = loans to individuals; y_4 = other loans (to purchase securities, overdrafts, etc.); y_5 = securities, fed funds sold, repos, assets in trading accounts; p_i = price of output i; $\tilde{p}$ = weighted average of output prices; w_1 = price of labor; w_2 = price of physical capital; w_3 = price of insured deposits; w_4 = price of other borrowed money (repos, fed funds purchased, etc.); w_5 = price of uninsured deposits; p_π = price of real after-tax profit; s_i = input i share; s_π = profit share; $p_\pi \pi$ = nominal before-tax accounting profit; $p \cdot y + m$ = expected revenue; n = nonperforming loans; k = financial capital; m = noninterest income.

Table 3: Scale Economies and Other Elasticities

measure	whole sample	1st quartile	2nd quartile	3rd quartile	4th quartile
SCALE	1.146***	1.101***	1.128***	1.146***	1.208***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
$\dfrac{\partial ln\, C}{\partial ln\, y_1}$	0.317***	0.343***	0.343***	0.355***	0.261***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\dfrac{\partial ln\, C}{\partial ln\, y_2}$	0.197***	0.192***	0.177***	0.192***	0.227***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\dfrac{\partial ln\, C}{\partial ln\, y_3}$	0.120***	0.122***	0.122***	0.131***	0.107***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\dfrac{\partial ln\, C}{\partial ln\, y_4}$	0.031***	0.025***	0.029***	0.028***	0.042***
	(0.01)	(0.01)	(0.005)	(0.01)	(0.01)
$\dfrac{\partial ln\, C}{\partial ln\, y_5}$	0.225***	0.252***	0.231***	0.205***	0.213***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\dfrac{\partial ln\, C}{\partial ln\, k}$	−0.032**	−0.033**	−0.033**	−0.032**	−0.032**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\dfrac{\partial ln\, k}{\partial ln\, n}$	0.054	0.262*	0.089	−0.010	−0.124
	(0.11)	(0.14)	(0.11)	(0.11)	(0.16)
$\dfrac{\partial ln\, k}{\partial ln\, \tilde{p}}$	−1.624	−2.681	−1.605	−1.072	−1.148
	(1.44)	(1.95)	(1.55)	(1.30)	(1.31)
$\dfrac{\partial ln\, x_4}{\partial ln\, \tilde{p}}$	−1.612	−6.125*	−0.823	0.102	0.363
	(1.56)	(3.22)	(1.87)	(1.00)	(0.64)
$\dfrac{\partial ln\, x_4}{\partial ln\, k}$	−0.328***	−0.513***	−0.382***	−0.246***	−0.173***
	(0.09)	(0.15)	(0.11)	(0.07)	(0.05)
$\dfrac{\partial ln\, x_4}{\partial ln\, n}$	−0.111	0.212	−0.250**	−0.199***	−0.204***
	(0.10)	(0.24)	(0.13)	(0.07)	(0.07)

Standard errors in parentheses; *significant at the 0.10 level, **significant at the 0.05 level, ***significant at the 0.01 level, asymptotic two-tailed t-tests.

Note: SCALE's significance is measured against 1; the other statistics' significance levels are measured against 0. Each entry is an average of individual elasticities of the whole sample or of the quartile subsamples. (See van Wissen and Golob (1992) for the rationale of using the average of the elasticities rather than the elasticity evaluated at the means of the data.)

Table 4: Scale Economies Along the Value-Maximizing Expansion Path

	whole sample	1st quartile	2nd quartile	3rd quartile	4th quartile
Scale economies on the utility-maximizing expansion path	1.146	1.101	1.128	1.146	1.208
Scale economies on the value-maximizing expansion path	1.138	1.111	1.110	1.174	1.185

Scale economies on the utility-maximizing expansion path are those based on the estimated generalized managerial production model. Each entry is the average scale estimate for all banks in the whole sample or in each asset-size quartile subsample.

Scale economies on the value-maximizing expansion path are also based on the estimated generalized managerial production model for efficient banks in the sample. Each entry is the average scale estimate for the quarter most efficient banks in the whole sample or in each asset-size quartile subsample.

Figure 1: Scale Economies Against Total Assets

scale economies

ln(total assets)