



FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall
Philadelphia, Pennsylvania 19106-1574
(215) 574-6428, www.phil.frb.org

Working Papers

Research Department

WORKING PAPER NO. 97-8

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
Hanyang University

Revised July 1997

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
Hanyang University

Revised July 1997

*The views expressed in this paper do not necessarily represent those of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, the Comptroller of the Currency, or the Department of the Treasury. Authors Hughes and Moon gratefully acknowledge the financial assistance of the Rutgers University Research Council on a preliminary project.

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

Abstract

Using modern duality theory to recover technologies from data can be complicated by the risk characteristics of production. In many industries, risk influences cost and revenue and can create the potential for costly episodes of financial distress. When risk is an important consideration in production, the standard cost and profit functions may not adequately describe the firm's technology and choice of production plan. In general, standard models fail to account for risk and its endogeneity. We distinguish between exogenous risk, which varies over the firm's choice sets, and endogenous risk, which is chosen by the firm in conjunction with its production decision. We show that, when risk matters in production decisions, it is important to account for risk's endogeneity.

For example, better risk diversification that results, e.g., from an increase in scale, improves the reward to risk-taking and may under certain conditions induce the firm to take on more risk to increase the firm's value. A choice of higher risk at a larger scale could add to costs and mask scale economies that may result from better diversification.

This paper introduces risk into the dual model of production by constructing a utility-maximizing model in which managers choose their *most preferred production plan*. We show that the utility function that ranks production plans is equivalent to a ranking of subjective probability distributions of profit that are conditional on the production plan. The most preferred production plan results from the firm's choice of an optimal profit distribution. The model is sufficiently general to incorporate risk aversion as well as risk neutrality. Hence, it can account for the case where the potential for costly financial distress makes trading profit for reduced risk a value-maximizing strategy.

We implement the model using the Almost Ideal Demand System to derive utility-maximizing share equations for profit and inputs, given the output vector and given sources of risk to control for choices that would affect endogenous risk. The *most preferred cost function* is obtained from the profit share equation and we show that, if risk neutrality is imposed, this system is identical to the standard translog cost system except that it controls for sources of risk.

We apply the model to the U.S. banking industry using 1989-90 data on banks with over \$1 billion in assets. Consistent with the significant regulatory and financial costs of bank distress, we find evidence that managers trade return for reduced risk. In addition, we find evidence of significant scale economies that help explain the recent wave of large bank mergers. Using these same data, we also estimate the standard cost function, which does not explicitly account for risk, and we obtain the usual results of essentially constant returns to scale, which contradicts the often-stated rationale for bank mergers.

1. Introduction

Using modern duality theory to recover technologies from data can be complicated by the risk characteristics of production. The standard cost and profit functions usually ignore risk. However, in many industries, risk can be an important consideration in formulating the production plan. Risk influences cost and revenue and can create the potential for costly episodes of financial distress; responding to these incentives, firms adjust their exposure to risk. The failure of the standard cost and profit functions to account for firms' exogenous risk environment and for their endogenous responses to risk can lead to misleading results.

For example, a firm's technology may link the diversification of risk to its scale of operations or to the spread of its operations across different geographic regions. In such cases, the better diversification that follows from an increase in scale reduces the cost of managing risk and can lead to scale economies and enhanced profitability. However, the standard cost and profit functions may not be able to detect these scale economies because they do not account for the firm's choice of risk. Better diversification reduces not only the total cost of managing risk but also the marginal cost and, hence, increases the marginal return to risk-taking. If the improved return on risk-taking induces the firm to take more risk, the firm could incur additional costs of risk management that may "use up" the cost saved by better diversification. Hence, when the standard cost function measures the response of cost to a scaled increase in output, it might detect constant or decreasing returns to scale rather than increasing returns, because it has not controlled for the endogenous response of risk to better diversification. Risk in this context plays a role similar to that of product quality. When product quality is endogenous, any measure of scale economies that fails to control for product quality is likely to be misleading.

Consider a second example where risk creates the potential for costly episodes of financial distress. In such a case, the standard cost and profit functions may provide misleading results because the production plan that maximizes the firm's value may involve trading short-term profitability for reduced risk.¹ A firm that incurs extra costs and diminished revenues to reduce risk will not be minimizing the objective function that defines the standard cost function nor will its optimal production plan maximize the objective function that generates the standard profit function. By ignoring risk, the standard cost and profit functions implicitly assume that managers do not trade profit for reduced risk and, hence, that there are no important costs attached to financial distress. However, in many industries, and especially in commercial banking, controlling risk to reduce the potential for financial distress is an essential consideration in production.²

¹Tufano (1996) reviews the literature on why managers might exhibit risk-averse behavior.

²Even if managers choose production plans that trade expected profit for less risk, they presumably

To summarize, when risk is an important consideration in production, the standard cost and profit functions may not adequately describe the firm's technology and choice of production plan. In general, they fail to account for risk and, in particular, they ignore its endogeneity. In the sections that follow, we amend the standard cost and profit functions to account for these aspects of risk; we use the amended functions to recover banking technology from data on a sample of U.S. banks in 1990; and we show that the technology recovered by these amended functions is consistent with the often-stated rationale for the current wave of large bank mergers, i.e., scale economies. In contrast, most studies of U.S. banking technology that use the standard cost function find no evidence of these significant scale economies.

1.1. Incorporating Risk into a Model of Production

Our strategy to incorporate risk into the firm's choice of production plans links the production plan to a subjective, conditional distribution of profit (see Hughes and Moon, 1995). 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 the expected costs of financial distress, their resulting attitude toward risk, and their assessment of the probability distribution of profit conditional on alternative production plans. If the firm's choice of plans follows from a ranking that considers only the plan's expected profit or, equivalently, the first moment of the subjective, conditional distribution of profit, then the firm is risk neutral. The standard cost and profit functions are consistent with this case. On the other hand, when managers also consider risk in choosing the production plan, their rankings of production plans must include higher moments of the distribution.

To identify the firm's highest ranked production plan, we represent its ranking by a managerial utility function defined over the production plan and profit. Firms maximize

maximize profits, given the amount of risk they assume. Stating the problem this way suggests that the standard profit and cost functions can be amended to account for risk simply by conditioning them on appropriate measures of risk. But this would not account adequately for the behavior of a firm that is not risk-neutral. For example, the equilibrium of the non-risk-neutral firm can be influenced by the tax rate on profit as well as by fixed costs, while that of a risk-neutral, profit-maximizing firm is not affected by these variables. The strategy of conditioning the standard profit function on measures of risk fails to account for these critical differences in equilibria. To capture these differences, the firm's objective function must be amended to include a broader objective than profit maximization. Santomero (1984) considers the important factors in specifying an objective function for banks.

this function, subject to the technology and the profit identity, to obtain their *most preferred production plan* (for inputs and outputs) and *most preferred profit function*.³ These choice functions are sufficiently general to subsume the equilibria of the non-risk-neutral firm as well as the risk-neutral firm, since they allow for the possibility that the firm's equilibrium is influenced by the tax rate on profit and by fixed charges and revenues. Hence, the most preferred profit function generalizes the standard profit function to account for the endogeneity of risk and for the possibility that trading profitability for reduced risk is a value-maximizing strategy.

1.2. Recovering Risky Technology with the Almost Ideal Demand System

To implement the model, we borrow the Almost Ideal (AI) Demand System from consumer theory (see Deaton and Muellbauer, 1980). We adapt the AI expenditure function to represent generalized managerial preferences and use it to derive the functional forms for the utility-maximizing demand functions for profit and the production plan. We show that when the parameter values implied by risk neutrality are imposed on the demand functions, they become identically equal to the translog profit (cost) function and share equations. Hence, a standard translog representation of the maximum profit function is subsumed by this more general specification. Moreover, these parameter values provide a test of risk neutrality.

To estimate our model, we turn to the commercial banking industry, whose business includes risk-taking and risk-diversification. Commercial banks rely on demandable debt, which is part of the payments system, to fund their portfolios of loans. In addition, banks are better informed about the quality of these nonmarket assets than are their depositors so that the credit risk they assume in their asset portfolios entails significant liquidity risk, risk that can spill over to other banks and threaten the safety of the payments system. To control this critical systemic externality, regulatory authorities charter banks and supervise them to ensure their safety and soundness. Banks that encounter financial distress can suffer severe regulatory penalties, liquidity crises, and even loss of their valuable charters. Consequently, bank managers may well trade profit for reduced risk to protect their banks from costly episodes of financial distress. In fact, our empirical tests reject the hypothesis of risk neutrality.⁴

We estimate a model that allows us to compute scale economies. We focus on scale economies because most studies based on the standard cost function fail to find evidence of the scale economies that bankers cite as a rationale for the current merger

³This framework for modeling generalized managerial preferences draws on earlier work by Hughes (1989, 1990) on hospitals and education that allowed managers to choose production plans that trade profit or net income for other objectives.

⁴Keeley (1990) shows that declining charter values, eroded by increasing competition in banking over the last several decades, have contributed to increased risk-taking by U.S. commercial banks.

wave (see Berger and Humphrey, 1992). These studies typically find that banking is characterized by slightly increasing returns to scale at relatively small banks and by slightly decreasing returns to scale at relatively large banks. This empirical result seems at odds with statements from bank management that have justified the recent wave of large bank mergers on the basis of cost savings. We consider whether this scale economy puzzle might be due to the standard cost function's failure to account for risk and its endogeneity.

2. Modeling Risky Production: An Application to Banking

Risk-taking and risk diversification are the business of banking. Using their depositors' funds, commercial banks produce diversified, leveraged portfolios of information-intensive assets. They assess credit risk, they write loan contracts, and they monitor borrowers' behavior to control moral hazard. Banks' efforts at managing individual loan risk are enhanced by the diversification achieved by combining loans in a portfolio. The credit risk of any particular loan can be represented by the probability distribution of the present value of the stream of its future payments. Banks that charge higher interest rates on their loans will typically attract borrowers with higher ex ante probability of default. If there is a market trade-off for risk and return in lending, these banks are sacrificing loan quality for higher expected profitability. Diversification can reduce the return risk on the bank's portfolio of loans. As banks increase in size, the diversification of the loan portfolio improves and its return variance diminishes at any given level of loan quality. If larger banks respond to the improved diversification by raising the interest rates on loans, increasing their expected return and risk, loan quality is reduced and costs will be higher than at a higher level of loan quality. Thus, diversification may generate cost-savings at a given level of loan quality and may also create the incentive to reduce the level of loan quality, generating higher costs of managing the reduced quality. The measurement of bank costs must control for loan quality to distinguish the cost-saving effects of diversification from the potentially cost-increasing changes in loan quality.⁵

⁵Several studies of bank costs have modified the standard cost function to introduce risk. McAllister and McManus (1993) conditioned the cost function on financial capital and control for the risk of insolvency by adjusting each bank's level of financial capital so that all banks in their sample have the same probability of insolvency. They find essentially constant returns to scale. If improved diversification reduces return risk, it also reduces the risk of insolvency. Hence, adjusting capital to control for the probability of insolvency may obscure scale economies due to improved diversification. Because managers may "overemploy" financial capital to reduce the probability of financial distress, Hughes and Mester (1993) estimated a banking cost function conditioned on the level of financial capital so that the level need not minimize cost. They also controlled for loan quality. However, in measuring scale economies, they varied financial capital proportionately with the output vector. They, too, obtained constant returns to scale. The equiproportionate variation in financial capital would obscure any economies

We define a bank's production plan by its output vector, \mathbf{y} , which consists of asset categories, such as commercial and industrial (C&I) loans, consumer loans, real estate loans, and government securities. The production plan also includes the input vector, \mathbf{x} , and financial (or equity) capital, k . Inputs comprise sources of loanable funds such as insured and uninsured deposits and other borrowed money, and the labor and physical capital used in intermediating the loanable funds. Thus, a production plan, $(\mathbf{y}, \mathbf{x}, k)$, consists of the portfolio of loans and securities and the inputs and financial capital used to produce the portfolio. The output vector, \mathbf{p} , represents the interest rates charged on the different components of the asset portfolio, \mathbf{y} . The higher the interest rate, relative to the risk-free interest rate, r , the higher the asset's risk-premium. An increase in the loan rate, given the risk-free rate, typically reduces the quality of the loan applicants, since borrowers with better credit ratings will seek lenders offering lower rates. Hence, the higher loan rate increases return risk and higher risk may be accompanied by higher expected returns. Thus, the risk premium provides an *ex ante* proxy of asset quality. Lower quality assets are likely to have a higher loss rate, too. Hence, the amount of nonperforming loans, designated by n , gives some indication of *ex post* asset quality.

Other components of the bank's price environment include \mathbf{w} , the input price vector; w_k , the price of financial capital (rate of return on equity); and m , income from sources other than those accounted for by the output vector, \mathbf{y} . Letting t be the tax rate on profit and $\tilde{p}_\pi (= 1)$ be the nominal "price" of a real dollar, the price of a dollar of real, after-tax profit in terms of nominal, before-tax dollars is $p_\pi = \frac{\tilde{p}_\pi}{1-t}$.⁶ Nominal, before-tax accounting profit is, thus, defined as

$$p_\pi \pi = \mathbf{p} \cdot \mathbf{y} + m - \mathbf{w} \cdot \mathbf{x}. \quad (1)$$

Nominal accounting profit is composed of before-tax economic profit, $p_\pi \hat{\pi}$, and the required payment to equity, $w_k k$, which will depend on the riskiness of the bank. Hence, let $w_k = r \cdot g(\mathbf{s})$, where, as noted above, r is the risk-free rate of return and $g(\mathbf{s}) \geq 1$ is a risk premium. The risk premium, $g(\mathbf{s})$, is assumed to be homogeneous of degree zero in (\mathbf{p}, r) . Thus, a proportional variation in the risk-free rate r and the asset returns \mathbf{p} results in an equiproportional variation in w_k . The other arguments that affect the

that might be obtained when better diversification allows banks to reduce their capital-to-asset ratios. When Hughes and Mester (forthcoming) embedded the conditional cost minimization problem into the problem of maximizing managerial utility, they obtained the utility-maximizing demand for financial capital, which need not be the level that minimizes cost and which allows financial capital to vary optimally when scale economies are computed. Again, they controlled for asset quality. Using the same sample of banks as in their previous paper, they found substantial economies in capitalization and relatively large overall scale economies.

⁶The "price," \tilde{p}_π , facilitates stating the homogeneity conditions: a proportional variation in \tilde{p}_π implies the same variation in p_π so homogeneity will be stated in terms of the latter.

premium are discussed below. Thus,

$$\begin{aligned} p_\pi \pi &= p_\pi \left[\frac{w_k k}{\tilde{p}_\pi} + \hat{\pi} \right] \\ &= p_\pi \left[\frac{r \cdot g(\mathbf{s}) \cdot k}{\tilde{p}_\pi} + \hat{\pi} \right]. \end{aligned} \quad (2)$$

The nominal, before-tax return on equity is then $\frac{p_\pi \pi}{k} = \frac{p_\pi r \cdot g(\mathbf{s})}{\tilde{p}_\pi} + \frac{p_\pi \hat{\pi}}{k}$, which consists of the required return and the economic rent.

Managers of a bank rank and choose production plans. The standard profit function assumes that the bank ranks plans by the value of (2) and that it chooses the plan that maximizes (2). However, when production plans entail risk, profit is more appropriately characterized by a subjective probability distribution that is conditional on the production plan and asset quality.⁷ Thus, when production is risky, the standard profit function implicitly assumes that managers rank and choose plans by their expected profitability.

Not only do the standard profit and cost functions assume risk neutrality, they also assume that there are no costs attached to financial distress. Given the potential for costly episodes of financial distress, the production plan that maximizes the value of the firm will not maximize profit defined by (2). The value-maximizing production plan will involve trading profit for reduced risk—in particular, a reduced probability of large negative realizations of profit. In contrast to the ranking based on the expected value of profit, the ranking that accounts for financial distress costs must include higher moments of the distribution that reflect risk. To allow for the possibility that higher moments may also influence the rankings, we represent these *generalized* managerial preferences by the utility function $U(\pi, \mathbf{s})$, where $\mathbf{s} = (\mathbf{y}, \mathbf{x}, k, \mathbf{p}, r, n)$. When managers do not trade return for reduced risk, only profit has marginal significance in the utility function. In this case, the production plan and asset quality influence utility only through their effect on profit so that utility is essentially a function of a single argument, profit. Hence, when utility is maximized, profit is also maximized, and the cost of producing the optimal output and quality vector is minimized. On the other hand, when the value-maximizing strategy involves trading profit for reduced risk, the production plan and quality arguments will directly influence utility *in addition to* affecting profit, because they influence higher moments of the subjective, conditional distribution of profit. In such cases, production plans that involve, for example, extra labor for more intensive credit evaluation and loan monitoring or that use more expensive, but less volatile funding sources to reduce risk may be ranked ahead of more profitable but riskier plans.

The utility function's ranking of production plans depends on managers' beliefs about how the production plans interact with future economic conditions to generate

⁷Under certain conditions, these subjective, conditional probability distributions of profit can be summarized by their first two moments, the expected value and variance (or risk).

profit and on their beliefs about the probability distribution of future economic states. These beliefs link production plans and asset quality to subjective, conditional distributions of profit. Thus, the utility function's ranking of production plans reflects these beliefs and can be interpreted as an implicit ranking of subjective, conditional distributions of profit. When managers choose their *most preferred production plan*, they can be characterized equivalently as choosing their *most preferred conditional probability distribution of profit*. This choice reflects their beliefs about the revenues and costs attached to risk-taking, about opportunities for risk diversification, and about potential costs of financial distress.

We formalize the definition of the most preferred production plan as the solution to the problem of maximizing the managerial utility function with respect to the production plan and profit, subject to the definition of profit and to the technology that specifies feasible production plans, $T(\mathbf{y}, \mathbf{x}, n, k) \leq 0$. We condition the maximization problem on (i) the output (asset) vector, \mathbf{y} , so that we can readily gauge how cost varies with a proportionate change in the output vector; on (ii) the components of output (asset) quality, (\mathbf{p}, r, n) , so that the measure of scale economies is not biased by an endogenous variation in asset quality; and on (iii) financial capital, k , so that we can investigate the effect of capitalization on the organization of production. The choice of financial capital will be derived in a second-stage optimization. Hence, the most preferred production plan and the most preferred level of profit solves the following problem:

$$\max_{\pi, \mathbf{x}} U(\pi, \mathbf{x}; \mathbf{y}, \mathbf{p}, r, n, k) \quad (3)$$

$$\text{s.t. } \mathbf{p} \cdot \mathbf{y} + m - \mathbf{w} \cdot \mathbf{x} - p_\pi \pi = 0 \quad (4)$$

$$T(\mathbf{x}; \mathbf{y}, n, k) \leq 0 \quad (5)$$

Letting the price vector be represented by $\mathbf{v} = (\mathbf{w}, \mathbf{p}, r, p_\pi)$, the most preferred production plan that produces the asset portfolio, \mathbf{y} , characterized by asset quality, (\mathbf{p}, r, n) , with financial capital, k , is given by $\mathbf{x}(\mathbf{y}, n, \mathbf{v}, m, k)$, and the most preferred level of after-tax profit implied by this production plan is $\pi(\mathbf{y}, n, \mathbf{v}, m, k)$. As functions of input prices and output levels, these utility-maximizing demand functions for inputs and profit resemble the standard cost-minimizing ones. As functions of prices and income (revenue), they resemble consumers' demand functions. As functions of the tax rate on profit, they resemble neither, but afford consistency with the theory of the non-risk-neutral firm. Since this solution is also conditioned on the amount of financial capital, the utility-maximizing level of profit is readily translated into the optimal rate of return on equity. When managers do not trade profit for reduced risk, the utility-maximizing level of profit is the maximum profit or, equivalently, the maximum rate of return on equity.

3. The Most Preferred Cost Function

Scale economies are obtained from the *most preferred (MP) cost function*, which is defined by:

$$C(\mathbf{y}, n, \mathbf{v}, m, k) \equiv \mathbf{w} \cdot \mathbf{x}(\mathbf{y}, n, \mathbf{v}, m, k) \equiv \mathbf{p} \cdot \mathbf{y} + m - p_\pi \pi(\mathbf{y}, n, \mathbf{v}, m, k). \quad (6)$$

There are several notable features of the MP cost function. From (6), it is clear that the cost function is embedded in the utility-maximizing demand for profit. As will be discussed later, measures of technology such as scale and scope economies can be derived from the utility-maximizing profit equation. Second, when outputs and inputs as well as profit affect utility marginally, revenue influences cost. Not only will output-based revenue, $\mathbf{p} \cdot \mathbf{y}$, affect the optimum, so will fixed revenue, m (and fixed cost). Additionally, the tax rate the bank pays on its profit will, in general, influence the optimum. Of course, in the special case of a risk-neutral manager, where only profit has marginal significance in the utility function, revenue and tax rates will not influence cost. Finally, notice that input mixes on the interior of the input requirement sets (isoquants) can be utility-maximizing.

Unlike the standard cost function, the homogeneity properties of the MP cost function include output prices and fixed revenues. The input demand functions $\mathbf{x}(\mathbf{y}, n, \mathbf{v}, m, k)$ are homogeneous of degree zero in (\mathbf{v}, m) , while the nominal profit function $p_\pi \pi(\mathbf{y}, n, \mathbf{v}, m, k)$ is homogeneous of degree one in (\mathbf{v}, m) . Hence, the MP cost function $\mathbf{w} \cdot \mathbf{x}(\mathbf{y}, n, \mathbf{v}, m, k)$ is homogeneous of degree one in (\mathbf{v}, m) .

4. Deriving the MP Cost Function from the Almost Ideal Demand System

The functional forms for the utility-maximizing input demands and profit equation can be obtained from the AI Demand System. In the special case where only profit has marginal significance in the utility function, these functional forms are identical to the translog cost function and share equations.

Our strategy is to adapt the expenditure function of the AI Demand System to represent generalized managerial preferences, to apply Shephard's lemma to obtain the expenditure-minimizing demand system for inputs and profit, and then to substitute the indirect utility function into the demand system to convert it into the utility-maximizing system that we estimate.

The expenditure function describes the amount of expenditure required to achieve a given level of utility U^0 . The managerial expenditure function is defined by the following problem:

$$\min_{\pi, \mathbf{x}} \mathbf{w} \cdot \mathbf{x} + p_\pi \pi \quad (7)$$

$$\text{s.t. } U^0 - U(\pi, \mathbf{y}, \mathbf{x}, \mathbf{p}, r, n, k) = 0 \quad (8)$$

$$T(\mathbf{x}; \mathbf{y}, n, k) \leq 0, \quad (9)$$

whose solution yields the constant-utility demand functions $\mathbf{x}^u(\mathbf{y}, n, \mathbf{v}, k, U^0)$ and $\pi^u(\mathbf{y}, n, \mathbf{v}, k, U^0)$. Substituting these demand functions into (7) yields the expenditure function $E(\mathbf{y}, n, \mathbf{v}, k, U^0)$. The expenditure-minimization problem (7) is dual to the utility-maximization problem (3) so that $E(\mathbf{y}, n, \mathbf{v}, k, U^0) = \mathbf{p} \cdot \mathbf{y} + m$. Also, the demand functions obtained from (3) and (7) are identically equal when the indirect utility function, $V(\mathbf{y}, n, \mathbf{v}, m, k)$, derived by inverting the expenditure function, is substituted for the utility index in the expenditure-minimizing demands:

$$\mathbf{x}^u(\mathbf{y}, n, \mathbf{v}, k, V(\mathbf{y}, n, \mathbf{v}, m, k)) \equiv \mathbf{x}(\mathbf{y}, n, \mathbf{v}, m, k) \quad (10)$$

$$\pi^u(\mathbf{y}, n, \mathbf{v}, k, V(\mathbf{y}, n, \mathbf{v}, m, k)) \equiv \pi(\mathbf{y}, n, \mathbf{v}, m, k). \quad (11)$$

Adapting the framework of the AI Demand System to accommodate the generalized managerial preferences yields the expenditure function,

$$\ln E(\cdot) = \ln P + U \cdot \beta_0 \left(\prod_i y_i^{\beta_i} \right) \left(\prod_j w_j^{\nu_j} \right) p_\pi^\mu k^\kappa, \quad (12)$$

where

$$\begin{aligned} \ln P = & \alpha_0 + \sum_i \alpha_i \ln p_i + \sum_i \delta_i \ln y_i + \sum_j \omega_j \ln w_j \\ & + \eta_\pi \ln p_\pi + \tau \ln r + \vartheta \ln n + \rho \ln k \\ & + \frac{1}{2} \sum_i \sum_j \alpha_{ij} \ln p_i \ln p_j + \frac{1}{2} \sum_i \sum_j \delta_{ij} \ln y_i \ln y_j \\ & + \frac{1}{2} \sum_s \sum_t \omega_{st}^* \ln w_s \ln w_t + \frac{1}{2} \eta_{\pi\pi} (\ln p_\pi)^2 \\ & + \frac{1}{2} \tau_{rr} (\ln r)^2 + \frac{1}{2} \vartheta_{nn} (\ln n)^2 + \frac{1}{2} \rho_{kk} (\ln k)^2 \\ & + \sum_i \sum_j \theta_{ij} \ln p_i \ln y_j + \sum_i \sum_s \phi_{is} \ln p_i \ln w_s + \sum_i \psi_{i\pi} \ln p_i \ln p_\pi \\ & + \sum_i \psi_{ir} \ln p_i \ln r + \sum_i \psi_{in} \ln p_i \ln n + \sum_i \psi_{ik} \ln p_i \ln k \\ & + \sum_j \sum_s \gamma_{js} \ln y_j \ln w_s + \sum_j \gamma_{j\pi} \ln y_j \ln p_\pi + \sum_j \gamma_{jr} \ln y_j \ln r \\ & + \sum_j \gamma_{jn} \ln y_j \ln n + \sum_j \gamma_{jk} \ln y_j \ln k \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} \sum_s \omega_{s\pi}^* \ln w_s \ln p_\pi + \frac{1}{2} \sum_s \omega_{\pi s}^* \ln p_\pi \ln w_s \\
& + \sum_s \omega_{sr} \ln w_s \ln r + \sum_s \omega_{sn} \ln w_s \ln n + \sum_s \omega_{sk} \ln w_s \ln k \\
& + \eta_{\pi r} \ln p_\pi \ln r + \eta_{\pi n} \ln p_\pi \ln n + \eta_{\pi k} \ln p_\pi \ln k \\
& + \tau_{rn} \ln r \ln n + \tau_{rk} \ln r \ln k + \vartheta_{nk} \ln n \ln k.
\end{aligned} \tag{13}$$

Hence, from (12) the indirect utility function is

$$V(\cdot) = \frac{\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P}{\beta_0 \left(\prod_i y_i^{\beta_i} \right) \left(\prod_j w_j^{\nu_j} \right) p_\pi^\mu k^\kappa} \tag{14}$$

Applying Shephard's lemma to (12) to obtain the constant-utility input demand equations and profit equation and then substituting the indirect utility function (14) into these equations yields the utility-maximizing choice functions:

$$\begin{aligned}
\frac{\partial \ln E}{\partial \ln w_i} &= \frac{w_i x_i}{\mathbf{p} \cdot \mathbf{y} + m} = \frac{\partial \ln P}{\partial \ln w_i} + \nu_i [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] \\
&= \omega_i + \sum_s \omega_{si} \ln w_s + \sum_j \phi_{ji} \ln p_j + \sum_j \gamma_{ji} \ln y_j \\
&\quad + \omega_{\pi i} \ln p_\pi + \omega_{ir} \ln r + \omega_{in} \ln n + \omega_{ik} \ln k \\
&\quad + \nu_i [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P]
\end{aligned} \tag{15}$$

$$\begin{aligned}
\frac{\partial \ln E}{\partial \ln p_\pi} &= \frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} = \frac{\partial \ln P}{\partial \ln p_\pi} + \mu [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] \\
&= \eta_\pi + \eta_{\pi\pi} \ln p_\pi + \sum_j \psi_{j\pi} \ln p_j + \sum_j \gamma_{j\pi} \ln y_j \\
&\quad + \sum_s \omega_{s\pi} \ln w_s + \eta_{\pi r} \ln r + \eta_{\pi n} \ln n + \eta_{\pi k} \ln k \\
&\quad + \mu [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P]
\end{aligned} \tag{16}$$

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

Symmetry requires $\alpha_{ij} = \alpha_{ji}$ and $\delta_{ij} = \delta_{ji}$ in addition to $\omega_{s\pi} = \omega_{\pi s}$ and $\omega_{si} = \omega_{is}$. The first two symmetry conditions must be imposed in the estimation of the share equations, since the constituent coefficients cannot be separately identified. However, the

latter two symmetry conditions 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 and permit a test of symmetry when the condition is not imposed on the estimation. In summary, symmetry requires that

$$\begin{aligned} \text{(S1)} \quad \alpha_{ij} &= \alpha_{ji} \quad \forall i, j, & \text{(S2)} \quad \delta_{ij} &= \delta_{ji} \quad \forall i, j, \\ \text{(S3)} \quad \omega_{s\pi} &= \omega_{\pi s} \quad \forall s, \text{ and} & \text{(S4)} \quad \omega_{si} &= \omega_{is} \quad \forall s, i. \end{aligned}$$

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

$$\begin{aligned} \text{(H1)} \quad \sum_j \nu_j + \mu &= 0, & \text{(H2)} \quad \Sigma \alpha_i + \Sigma \omega_j + \eta_\pi + \tau &= 1, \\ \text{(H3)} \quad \sum_i \alpha_{ij} + \sum_t \phi_{jt} + \psi_{jr} + \psi_{j\pi} &= 0 \quad \forall j, & \text{(H4)} \quad \sum_i \phi_{it} + \sum_s \omega_{st} + \omega_{tr} + \omega_{t\pi} &= 0 \quad \forall t, \\ \text{(H5)} \quad \tau_{rr} + \sum_i \psi_{ir} + \sum_s \omega_{sr} + \eta_{\pi r} &= 0, & \text{(H6)} \quad \sum_i \theta_{ij} + \sum_t \gamma_{jt} + \gamma_{j\pi} + \gamma_{jr} &= 0 \quad \forall j, \\ \text{(H7)} \quad \eta_{\pi\pi} + \sum_i \psi_{i\pi} + \sum_s \omega_{s\pi} + \eta_{\pi r} &= 0, & \text{(H8)} \quad \sum_i \psi_{in} + \sum_s \omega_{sn} + \tau_{rn} + \eta_{\pi n} &= 0, \\ \text{(H9)} \quad \sum_i \psi_{ik} + \sum_s \omega_{sk} + \tau_{rk} + \eta_{\pi k} &= 0, \text{ and} & \text{(H10)} \quad \frac{1}{2} \sum_i \sum_j \alpha_{ij} + \frac{1}{2} \sum_s \sum_t \omega_{st} & \\ & & + \sum_i \sum_t \phi_{it} + \frac{1}{2} \tau_{rr} + \frac{1}{2} \eta_{\pi\pi} + \sum_i \psi_{i\pi} & \\ & & + \sum_i \psi_{ir} + \sum_s \omega_{sr} + \sum_s \omega_{s\pi} + \eta_{\pi r} &= 0. \end{aligned}$$

The input and profit shares sum to one, which implies the following adding up conditions:

$$\begin{aligned} \text{(A1)} \quad \sum_i \omega_i + \eta_\pi &= 1, & \text{(A2)} \quad \sum_i \omega_{si} + \omega_{s\pi} &= 0 \quad \forall s, \\ \text{(A3)} \quad \sum_i \phi_{ji} + \psi_{j\pi} &= 0 \quad \forall j, & \text{(A4)} \quad \sum_i \gamma_{ji} + \gamma_{j\pi} &= 0 \quad \forall j, \\ \text{(A5)} \quad \sum_i \omega_{\pi i} + \eta_{\pi\pi} &= 0, & \text{(A6)} \quad \sum_i \omega_{ir} + \eta_{\pi r} &= 0, \\ \text{(A7)} \quad \sum_i \omega_{ik} + \eta_{\pi k} &= 0, & \text{(A8)} \quad \sum_i \omega_{in} + \eta_{\pi n} &= 0, \text{ and} \\ \text{(A9)} \quad \sum_j \nu_j + \mu &= 0 \quad (\text{which is also a homogeneity condition}). \end{aligned}$$

5. Managerial Objectives: Profit Maximization?

If banks maximize profit (which is equivalent to maximizing return on equity here, since financial capital is treated as exogenous), a variation in the tax rate and, equivalently, in $p_\pi (= \frac{1}{1-t})$ will not affect the bank's choice of before-tax profit. This implies that

$$\text{(P1)} \quad \eta_\pi = \eta_{\pi\pi} = \psi_{i\pi} = \gamma_{j\pi} = \omega_{s\pi} = \eta_{\pi r} = \eta_{\pi n} = \eta_{\pi k} = 0 \quad \forall i, j, s.$$

Thus, (16) is simplified to

$$\frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} = \mu [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P]. \quad (17)$$

In addition, the revenue and risk characteristics of production represented by the output price vector will not influence the bank's cost-minimizing production plan so that

$$\text{(P2)} \quad \alpha_i = \alpha_{ij} = \theta_{ij} = \phi_{is} = \psi_{i\pi} = \psi_{ir} = \psi_{in} = \psi_{ik} = 0 \quad \forall i, j, s.$$

Similarly, variations in m have no marginal significance for the optimal input demand \mathbf{x} . Hence, the numerators, $w_i x_i$, of the shares (15) are unaffected by a variation in m . Instead, the variation in m solely affects profit so that $\frac{\partial p_\pi \pi}{\partial m} = 1$. Using these results in differentiating equations (15) and (16) with respect to $\ln m$ yields the following parameter values in the case of profit maximization:

$$(P3) \quad \begin{aligned} \nu_i &= \frac{\partial \left(\frac{w_i x_i}{\mathbf{p} \cdot \mathbf{y} + m} \right)}{\partial \ln m} \\ &= - \frac{w_i x_i}{\mathbf{p} \cdot \mathbf{y} + m}, \end{aligned} \quad (18)$$

$$\begin{aligned} \mu &= \frac{\partial \left(\frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} \right)}{\partial \ln m} \\ &= 1 - \frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m}. \end{aligned} \quad (19)$$

Therefore, we can test for profit maximization (and cost minimization) by testing the restrictions (P1), (P2), and (P3).

Substituting (18) into (15) and (19) into (16) yields,

$$w_i x_i \left[\frac{1 + \ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P}{\mathbf{p} \cdot \mathbf{y} + m} \right] = \frac{\partial \ln P}{\partial \ln w_i} \quad (20)$$

$$p_\pi \pi \left[\frac{1 + \ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P}{\mathbf{p} \cdot \mathbf{y} + m} \right] = \ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P, \quad (21)$$

and using (21), an expression for cost can be constructed,

$$C = \mathbf{p} \cdot \mathbf{y} + m - p_\pi \pi = \frac{\mathbf{p} \cdot \mathbf{y} + m}{1 + \ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P}. \quad (22)$$

Substituting (22) into (20) and (21) shows that *in the case of profit maximization*, the share equations are cost shares (and are identical to the translog cost function and corresponding share equations for inputs):

$$\frac{w_i x_i}{C} = \frac{\partial \ln P}{\partial \ln w_i} \quad (23)$$

$$- \frac{p_\pi \pi}{C} = \ln P - \ln(\mathbf{p} \cdot \mathbf{y} + m). \quad (24)$$

6. Deriving the Demand for Financial Capital

Conditioning the MP cost function and input demands on the level of financial capital allows us to investigate how a bank's underlying financial condition affects its production decisions. However, the more basic decision centers on the level of financial capital itself because, as a cushion against insolvency, it signals the bank's tolerance for risk. Thus, the utility-maximization framework of (3) must be expanded to include a second stage where the financial capital level is determined.

The utility-maximizing demands for inputs and profit derived from (3) are conditioned on the level of financial capital, k . It is straightforward to add a second stage to the maximization problem to determine the bank's choice of capital. Writing the Lagrangian function for (3) and evaluating it at the first-stage optimum, conditional on k , the conditional indirect utility function is obtained:

$$\begin{aligned} V(\mathbf{y}, n, \mathbf{v}, m, k) &\equiv U(\pi(\cdot), \mathbf{x}(\cdot); \mathbf{y}, \mathbf{p}, r, n, k) \\ &\quad + \lambda(\cdot)[\mathbf{p} \cdot \mathbf{y} + m - \mathbf{w} \cdot \mathbf{x}(\cdot) - p_\pi \pi(\cdot)] \\ &\quad + \gamma(\cdot)[T(\mathbf{x}(\cdot); \mathbf{y}, \mathbf{n}, \mathbf{k})]. \end{aligned} \quad (25)$$

The demand for financial capital follows from maximizing (25) with respect to k . Using the definition from (2) that $p_\pi \pi = p_\pi \left[\frac{r \cdot g(\mathbf{s}) \cdot k}{\tilde{p}_\pi} + \hat{\pi} \right]$, and differentiating (25) with respect to k yield the first-order condition

$$\frac{\partial V(\cdot)}{\partial k} = \frac{\partial U(\cdot)}{\partial k} - \lambda(\cdot) \frac{p_\pi \left[r \cdot g(\mathbf{s}) + r k \frac{\partial g(\mathbf{s})}{\partial k} \right]}{\tilde{p}_\pi} + \gamma(\cdot) \frac{\partial T(\cdot)}{\partial k} = 0, \quad (26)$$

whose solution is the demand for financial capital, $k(\mathbf{y}, n, \mathbf{v}, m)$.

The AI system's conditional indirect utility function (14) implies this first-order condition is

$$\begin{aligned} \frac{\partial V(\cdot)}{\partial k} &= \frac{\partial V(\cdot)}{\partial \ln k} \frac{\partial \ln k}{\partial k} \\ &= - \frac{1}{k \left[\beta_0 \left(\prod_i y_i^{\beta_i} \right) \left(\prod_j w_j^{\nu_j} \right) p_\pi^\mu k^\kappa \right]} \left[\frac{\partial \ln P}{\partial \ln k} + \kappa [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] \right] = 0 \\ &\implies \rho + \rho_{kk} \ln k + \sum_i \psi_{ik} \ln p_i + \sum_j \gamma_{jk} \ln y_j + \sum_s \omega_{sk} \ln w_s + \eta_{\pi k} \ln p_\pi \\ &\quad + \tau_{rk} \ln r + \vartheta_{nk} \ln n + \kappa [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] = 0. \end{aligned} \quad (27)$$

Not surprisingly, the demand for financial capital follows readily from the parameters of the conditional system of input demands (15) and profit (16) and, thus, constitutes additional parameter restrictions.

The output vector can be made endogenous in an analogous fashion; however, a simpler system can be derived by eliminating the portion of the two-stage problem that results from conditioning the utility maximization on the output vector. This modified procedure is developed in Hughes, Lang, Mester, and Moon (1993).

In the special case of profit maximization, variations in m should not affect the optimal demand for capital so that another restriction is added to (P3) above:⁸

$$(P3) \quad \kappa = 0.$$

7. Deriving Scale Economies from the MP Cost Function

Scale economies are defined by the inverse of the elasticity of cost with respect to output. Using the definition of the MP cost function (6) and substituting the utility-maximizing demand for financial capital into (6), the degree of scale economies is given by

$$\begin{aligned} SCALE &= \frac{C}{\sum_i y_i \left(\frac{\partial \hat{C}}{\partial y_i} + \frac{\partial \hat{C}}{\partial k} \frac{\partial k}{\partial y_i} \right)} & (28) \\ &= \frac{\mathbf{p} \cdot \mathbf{y} + m - p_\pi \pi}{\sum_i y_i \left(p_i - \frac{\partial p_\pi \pi}{\partial y_i} - \frac{\partial p_\pi \pi}{\partial k} \frac{\partial k}{\partial y_i} \right)} \\ &= \frac{\mathbf{p} \cdot \mathbf{y} + m - p_\pi \pi}{\sum_i \left[p_i y_i - (\mathbf{p} \cdot \mathbf{y} + m) \frac{\partial \left(\frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} \right)}{\partial \ln y_i} - \left(\frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} \right) p_i y_i - (\mathbf{p} \cdot \mathbf{y} + m) \frac{\partial \left(\frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} \right)}{\partial \ln k} \frac{\partial \ln k}{\partial \ln y_i} \right]} \end{aligned}$$

The final expression in (28) is stated in terms of derivatives of the profit share equation (16).

8. The Data

We estimate the AI Demand System using data on U.S. banks that reported at least \$1 billion in assets as of the last quarter of 1988. 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

⁸If the regulatory capital constraint were binding on banks, changes in m would not affect even a non-risk-neutral bank's demand for capital. But in our sample, and in general, the capital constraint is not binding.

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 (C&I) 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); and y_5 , securities, assets in trading accounts, federal funds sold, securities purchased under agreements to resell, and total investment securities.

Financial capital, k , is the average amount of equity capital, loan-loss reserves, and subordinated debt in 1990. In addition to financial 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 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) derived a test for determining whether deposits are inputs or outputs.⁹ In their data set, which is very similar to the one here, they concluded that *insured and uninsured deposits are inputs*. Thus, we treat them as inputs here as well.

In addition to financial capital, another indicator of a bank's underlying financial condition is its amount of nonperforming loans, n , which 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

⁹Hughes and Mester (1993) show that when deposits are inputs (outputs), variable costs (i.e., the cost of all nondeposit inputs) will be decreasing (increasing) in the level of deposits.

risk-free rate indicates the risk premium incurred by the output and, hence, suggests the average quality of the asset.

The variable, m , is measured by the amount of noninterest income received in 1990. Revenue is the sum, $\mathbf{p} \cdot \mathbf{y} + m$, and before-tax accounting profit is $\mathbf{p} \cdot \mathbf{y} + m - \mathbf{w} \cdot \mathbf{x}$. Banks pay taxes on their income, with the tax rates reflected in p_π . 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.

9. Estimation

The system to be estimated consists of the share equations, (15) and (16), given the definition of $\ln P$ found in (13), and the first-order condition (27), which defines the level of capitalization. Because a cross-section is used, there is no variation in the risk-free interest rate, r , so it is dropped from the estimating equations. However, its parameters can be recovered by using the homogeneity conditions.

To reduce the number of parameters to be estimated, the vector of output returns \mathbf{p} is reduced to its weighted average

$$\tilde{p} = \sum_i p_i \left[\frac{y_i}{\sum_j y_j} \right]. \quad (29)$$

Dropping r and using \tilde{p} simplifies the model to be estimated. The amended equations are given in the appendix and are indicated by primes attached to their equation numbers (thus, equation (13) in the text becomes equation (13') in the appendix).

We use nonlinear two-stage least squares to estimate the following system of nonlinear simultaneous equations (subject to the parameter restrictions (A1)-(A2), (A3'), (A4)-(A5), (A7)-(A9), and (S2)):

$$\mathbf{f}_t(\mathbf{y}_t, n_t, \mathbf{v}_t, m_t \mid \Theta) \equiv \begin{bmatrix} s_{1t} - (\text{r.h.s. of (15')} \text{ with } i = 1) \\ s_{2t} - (\text{r.h.s. of (15')} \text{ with } i = 2) \\ s_{3t} - (\text{r.h.s. of (15')} \text{ with } i = 3) \\ s_{4t} - (\text{r.h.s. of (15')} \text{ with } i = 4) \\ s_{\pi t} - (\text{r.h.s. of (16')}) \\ \text{l.h.s. of (27')} \end{bmatrix} = \mathbf{u}_t \quad (30)$$

where r.h.s. designates the right hand side of the indicated equation and l.h.s., the left hand side; t is the bank index, ranging from 1 to T ; s_{it} is the i -th input's revenue share at the t -th bank, i.e., $s_{it} \equiv \frac{w_{it}x_{it}}{\mathbf{p}_t \cdot \mathbf{y}_t + m_t}$ for $i = 1 \dots 5$; $s_{\pi t}$ is the profit share at the t -th bank, i.e., $s_{\pi t} \equiv \frac{p_{\pi t} \pi_t}{\mathbf{p}_t \cdot \mathbf{y}_t + m_t}$; \mathbf{u}_t are i.i.d. over t with the cross-equation covariance matrix Σ ; and Θ is the set of all identifiable parameters excluding those in Σ . The standard

errors and t -statistics we report are based on the asymptotic covariance matrix of the estimate of Θ , which penalizes for not using cross-equation dependence (see Gallant (1977)). The contemporaneously correlated error terms \mathbf{u}_t reflect optimization errors (i.e., errors in utility maximization).

10. The Empirical Findings

The parameter estimates are reported in Table 3. There are four striking primary findings. First, risk neutrality or, equivalently, cost minimization, is conclusively rejected: managers appear to trade return for reduced risk. Second, the estimate of scale economies is larger than that found by most studies that do not control for endogenous risk. Third, the estimate of scale economies increases with bank size, which is consistent with the wave of mergers among large banks. Last, banks with higher capital levels or higher levels of nonperforming loans rely less on volatile funding sources.

10.1. Risk Neutrality

We used a Wald test to test the 31 restrictions (P1'), (P2'), and (P3') (given in the appendix) implied by risk neutrality or profit maximization. The value of the test statistic was 294.01 with 29 degrees of freedom (two of the restrictions are redundant because of the adding-up conditions that these parameters satisfy). Thus, the restrictions are strongly rejected, indicating that banks in the sample are not behaving in a risk-neutral manner.¹⁰

10.2. Scale Economies Measured by Other Studies

The many bank cost studies based on estimating a flexible functional form differ in the following ways: (1) how inputs and outputs are defined and, hence, how cost is constituted, (2) whether financial capital is ignored, included as a conditioning argument, or included as an element of cost, and (3) whether an average practice cost function or best practice cost frontier is estimated. One might expect these differences to yield a variety of scale estimates, but instead, the estimates are quite similar. Most studies of large banks (whose assets exceed \$1 billion) find either slight scale economies or slight diseconomies, and they usually find that scale economies decrease with bank size.¹¹

¹⁰We also computed the test statistic after removing (P3'), the restriction related to m . The value of this test statistic was 52.83 with 28 degrees of freedom. Again, profit maximization is rejected with a p -value of 0.003.

¹¹To maintain consistency with our measure of scale economies in (28), the discussion below transforms published measures of scale economies so that *values greater than one imply scale economies, while values less than one imply diseconomies*.

Treating noninterest-bearing deposits as a quasi-fixed input while characterizing the other inputs and outputs as we do here, Noulas, Ray, and Miller (1990) examine large banks in 1986 and find that scale economies decrease from 1.02 for the smallest banks (\$1-3 billion) to 0.97 for the largest banks (\$10+ billion). In a similar study that differs primarily in controlling for the number of branches, Hunter and Timme (1991) find that in 1986 scale economies range from 1.123 for the smallest group (\$1-1.5 billion) to 0.950 for the largest group (\$25+ billion). When they omit branches, the measures drop to values very close to those of Noulas, Ray, and Miller: 1.037 for the smallest banks to 0.977 for the largest banks.

Berger and Humphrey (1991) calculate scale economies using a thick frontier and find mild diseconomies, 0.98, for banks with \$1-2 billion in assets, decreasing to 0.97 for banks with over \$10 billion in assets. When they use a conventional approach, the range drops to 0.96 for the smallest banks and 0.94 for the largest banks.

Two other studies find approximately constant returns to scale overall but much wider ranges in scale economies. Hunter, Timme, and Yang (1990) obtain values ranging from 1.09 for the smallest banks (\$1-1.5 billion) to 0.90 for the largest (\$25+ billion) using 1986 data. Excluding interest payments from cost, Evanoff and Israilevich (1991) find measures ranging from 1.11 at \$0.72 billion to 0.76 at \$30 billion.

McAllister and McManus (1993) apply nonparametric estimation to 1984-90 data and find increasing returns to scale up to \$0.5 billion and constant returns from \$0.5-10 billion, the largest bank in their sample. Using 1988 data on all banks with over \$1 billion in assets, Pulley and Braunstein (1992) report an average measure ranging from 1.04 to 1.06 depending on the estimation procedure. Evanoff, Israilevich, and Merris (1990) estimate a shadow cost function to account for regulatory distortions over the period 1972-87 and obtain values of 1.07 for multibank holding companies and 1.10 for one-bank holding companies.

Studies that find scale economies increasing with bank size are rare. Mester (1992) defines outputs to capture the information-producing and -processing role of banks and, using 1988 data, finds slight scale economies that increase with bank size. Banks in the smallest group (\$1-1.5 billion) exhibit slight economies, 1.0305, and this increases to 1.0426 for banks in the largest group (\$5+ billion). Clark (1996), who examines the period 1988-91, treats core deposits as outputs and estimates a thick frontier. He finds economies of around 1.05 for the smallest banks (up to \$4 billion) and constant returns in all larger size categories (the largest of which is +\$20 billion in assets).

10.3. Scale Economies Measured Without Imposing Risk Neutrality

The studies discussed in the previous section do not explicitly incorporate risk. In contrast, we find that when the structural model of production is amended to account for risk and to control for its endogeneity, measured scale economies are much larger,

as shown in Table 4 and Figure I. These larger measures are more consistent with the recent wave of large bank mergers. The measures range from an average of 1.101 in the smallest asset-size quartile, 1.128 in the second quartile, 1.146 in the third, and 1.208 in the fourth. All are strongly significantly different from one. In addition to being larger than those found in previous studies, the measures also increase with bank asset size, contrary to most previous studies.

To compare the results of our amended model of production, which incorporates risk, with the standard cost function, which does not, we estimate the standard model using our data. We drop financial capital and our measures of asset quality (nonperforming loans and the average return on assets), and we use the standard translog cost formulation, including share equations, which results when we impose risk neutrality on our amended model. We find that the estimated scale measures are all significantly different from one but are considerably smaller and remarkably familiar: ranging from 1.022 for the smallest quartile, 1.029 for the second quartile, 1.035 for the third, and 1.050 for the fourth. These values are surprising only in that they increase in magnitude with asset size.

10.4. Other Results

Table 4 shows that for banks of all sizes, the capital level, k , and volatile funding, x_4 , are inversely related. Thus, banks with higher capital are less likely to rely on volatile sources to fund their assets. It appears that banks choosing lower insolvency risk also choose lower liquidity risk. And all size quartiles except the smallest, banks with a higher level of nonperforming loans, n , which is an *ex post* measure of asset quality, rely less on volatile funding sources. As nonperforming loans are likely to be less liquid than other assets, these banks may prefer more stable funding sources. To the extent that the prices of some of the components of volatile funding include a risk-premium, this may also be a cost-reducing strategy, since the risk-premium is likely to be higher for banks with poor loan quality.

Table 4 also reports that, for the smallest banks, the most preferred 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.

Table 4 also indicates that financial capital, k , and the average return on assets, \tilde{p} , are inversely related (although not significantly so). Lower quality assets have an *ex ante* higher return, but this return is a gamble. Our results suggest that the greater the gamble, that is, the higher \tilde{p} is, the lower the amount of financial capital, k , that is bet.

11. Conclusions

We have proposed amendments to the standard dual models of production to account for risk and, in particular, the endogenous element of risk-taking. While our amendments add obvious complexity to the model, their importance can be evaluated only by comparing the empirical adequacy of explanations that incorporate risk with those that abstract from it. When applied to banking, the most preferred profit and cost functions we develop in this paper explain important production phenomena that elude the approach using the standard cost function. Most notably, our approach reveals large scale economies, while previous studies that did not explicitly account for risk found only slight scale economies or constant returns to scale. Our results help explain the recent wave of mergers among very large banks and are consistent with the rationale frequently cited by participants in these mergers—the quest for cost savings due to a larger scale of operations. Our results also suggest that in industries where financial distress costs are high, managers may well trade profitability for reduced risk. Although such trade-offs are consistent with maximizing the value of the firm, they are not adequately captured by the standard profit and cost functions, which ignore the potential for costly financial distress.

References

- [1] Berger, Allen N. and David Humphrey (1992), "Megamergers in Banking and the Use of Cost Efficiency as an Antitrust Defense," *Antitrust Bulletin*, 37, 541-600.
- [2] Berger, Allen N. and David Humphrey (1991), "The Dominance of Inefficiencies over Scale and Product Mix Economies in Banking," *Journal of Monetary Economics*, 28, 117-148.
- [3] Clark, Jeffrey A. (1996), "Economic Cost, Scale Efficiency and Competitive Viability in Banking," *Journal of Money, Credit, and Banking*, 28, 342-364.
- [4] Deaton, Angus and John Muellbauer (1980), "An Almost Ideal Demand System," *American Economic Review*, 70, 312-326.
- [5] Evanoff, Douglas D. and Phillip R. Israilevich (1991), "Scale Elasticity and Efficiency for U.S. Banks," Working Paper 91-15, Federal Reserve Bank of Chicago.
- [6] Evanoff, Douglas D., Philip R. Israilevich, and Randall C. Merris (1990), "Relative Price Efficiency, Technical Change, and Scale Economies for Large Commercial Banks," *Journal of Regulatory Economics*, 2, 281-298.
- [7] Gallant, A. Ronald (1977), "Three-Stage Least-Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," *Journal of Econometrics*, 5, 71-88.
- [8] Hughes, Joseph P. (1989), "Hospital Cost Functions: The Case Where Revenues Affect Production," Working Paper No. 1990-01, Rutgers University, Department of Economics, November.
- [9] 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.
- [10] Hughes, Joseph P., William Lang, Loretta J. Mester, and Choon-Geol Moon (1993), "Risk-Taking by U.S. Thrifts: Modelling Risk-Indicative, Endogenous Interest Rates," Working Paper (revised January 1994).
- [11] Hughes, Joseph P. and Loretta J. Mester (forthcoming), "Evidence of Scale Economies in Risk Management and Signaling," *Review of Economics and Statistics*.
- [12] 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.

- [13] 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.
- [14] Hunter, William C. and Stephen G. Timme (1986), "Technical Change, Organizational Form, and the Structure of Bank Production," *Journal of Money, Credit, and Banking*, 18:2, 152-166.
- [15] Hunter, William C. and Stephen G. Timme (1991), "Some Evidence on the Impact of Quasi-Fixed Inputs on Bank Scale Estimates," *Economic Review*, Federal Reserve Bank of Atlanta, May/June, 12-20.
- [16] Hunter, William C., Stephen G. Timme, Won Keun Yang (1990), "An Examination of Cost Subadditivity and Multiproduct Production in Large U.S. Banks," *Journal of Money, Credit, and Banking*, 22:4, 504-515.
- [17] Keeley, Michael C. (1990), "Deposit Insurance, Risk, and Market Power in Banking," *American Economic Review*, 80, 1183-1200.
- [18] McAllister, Patrick H. and Douglas McManus (1993), "Resolving the Scale Efficiency Puzzle in Banking," *Journal of Banking and Finance*, 17:2-3, 389-406.
- [19] Mester, Loretta J. (1992), "Traditional and Nontraditional Banking: An Information-Theoretic Approach," *Journal of Banking and Finance*, 16, 545-566.
- [20] Mester, Loretta J. (1994), "How Efficient Are Third District Banks?" *Business Review*, Federal Reserve Bank of Philadelphia, January/February, 3-18.
- [21] Noulas, Athanasios G., Subhash C. Ray, and Stephen M. Miller (1990), "Returns to Scale and Input Substitution for Large U.S. Banks," *Journal of Money, Credit, and Banking*, 22:1, 94-108.
- [22] Pulley, Lawrence B. and Yale M. Braunstein (1992), "A Composite Cost Function for Multiproduct Firms with an Application to Economies of Scope in Banking," *Review of Economics and Statistics*, 74:2, 221-230
- [23] Santomero, Anthony M. (1984), "Modeling the Banking Firm," *Journal of Money, Credit, and Banking*, 16, 576-602.
- [24] Shaffer, Sherrill and Edmond David (1991), "Economies of Superscale in Commercial Banking," *Applied Economics*, 23, 283-293.
- [25] Tufano, Peter (1996), "Who Manages Risk? An Empirical Examination of Risk Management Practices in the Gold Mining Industry," *Journal of Finance*, 50, 1097-1137.
- [26] van Wissen, Leo and Thomas F. Golob (1992), "A Dynamic Model of Car Fuel-Type Choice and Mobility," *Transportation Research*, 26B, 77-96.

Appendix

Substituting \tilde{p} , defined in section 9, into the model, and dropping the risk-free rate, r , since it does not vary across banks, simplifies some of the equations in the model to be estimated.

The amended $\ln P$ is

$$\begin{aligned}
\ln P = & \alpha_0 + \alpha_p \ln \tilde{p} + \sum_i \delta_i \ln y_i + \sum_j \omega_j \ln w_j \\
& + \eta_\pi \ln p_\pi + \vartheta \ln n + \rho \ln k \\
& + \frac{1}{2} \alpha_{pp} (\ln \tilde{p})^2 + \frac{1}{2} \sum_i \sum_j \delta_{ij} \ln y_i \ln y_j \\
& + \frac{1}{2} \sum_s \sum_t \omega_{st} \ln w_s \ln w_t + \frac{1}{2} \eta_{\pi\pi} (\ln p_\pi)^2 \\
& + \frac{1}{2} \vartheta_{nn} (\ln n)^2 + \frac{1}{2} \rho_{kk} (\ln k)^2 \\
& + \sum_j \theta_{pj} \ln \tilde{p} \ln y_j + \sum_s \phi_{ps} \ln \tilde{p} \ln w_s + \psi_{p\pi} \ln \tilde{p} \ln p_\pi \\
& + \psi_{pn} \ln \tilde{p} \ln n + \psi_{pk} \ln \tilde{p} \ln k \\
& + \sum_j \sum_s \gamma_{js} \ln y_j \ln w_s + \sum_j \gamma_{j\pi} \ln y_j \ln p_\pi \\
& + \sum_j \gamma_{jn} \ln y_j \ln n + \sum_j \gamma_{jk} \ln y_j \ln k \\
& + \frac{1}{2} \sum_s \omega_{s\pi} \ln w_s \ln p_\pi + \frac{1}{2} \sum_s \omega_{\pi s} \ln p_\pi \ln w_s \\
& + \sum_s \omega_{sn} \ln w_s \ln n + \sum_s \omega_{sk} \ln w_s \ln k \\
& + \eta_{\pi n} \ln p_\pi \ln n + \eta_{\pi k} \ln p_\pi \ln k + \vartheta_{nk} \ln n \ln k.
\end{aligned} \tag{13'}$$

The amended share equations are

$$\begin{aligned}
\frac{\partial \ln E}{\partial \ln w_i} &= \frac{w_i x_i}{\mathbf{p} \cdot \mathbf{y} + m} = \frac{\partial \ln P}{\partial \ln w_i} + \nu_i [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] \\
&= \omega_i + \sum_s \omega_{si} \ln w_s + \phi_{pi} \ln \tilde{p} + \sum_j \gamma_{ji} \ln y_j \\
&\quad + \omega_{\pi i} \ln p_\pi + \omega_{in} \ln n + \omega_{ik} \ln k \\
&\quad + \nu_i [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P]
\end{aligned} \tag{15'}$$

$$\begin{aligned}
\frac{\partial \ln E}{\partial \ln p_\pi} &= \frac{p_\pi \pi}{\mathbf{p} \cdot \mathbf{y} + m} = \frac{\partial \ln P}{\partial \ln p_\pi} + \mu [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] \\
&= \eta_\pi + \eta_{\pi\pi} \ln p_\pi + \psi_{p\pi} \ln \tilde{p} + \sum \gamma_{j\pi} \ln y_j \\
&\quad + \sum_s \omega_{s\pi} \ln w_s + \eta_{\pi n} \ln n + \eta_{\pi k} \ln k \\
&\quad + \mu [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P]. \tag{16'}
\end{aligned}$$

The first-order condition defining the bank's optimal capital level is estimated as

$$\begin{aligned}
\rho + \rho_{kk} \ln k + \psi_{pk} \ln \tilde{p} + \sum_j \gamma_{jk} \ln y_j + \sum_s \omega_{sk} \ln w_s + \eta_{\pi k} \ln p_\pi \\
+ \vartheta_{nk} \ln n + \kappa [\ln(\mathbf{p} \cdot \mathbf{y} + m) - \ln P] = 0. \tag{27'}
\end{aligned}$$

The symmetry conditions are given in section 4 above. Condition (S1) becomes moot once \tilde{p} , the weighted average of the p_i s, is used. We impose (S2) in the estimation, but do not impose (S3) and (S4), so that these conditions can be tested. The homogeneity conditions (amended, since we are using \tilde{p}) are used to recover the coefficients on variables involving the risk-free rate in expenditure function, equation (12) in section 4. These homogeneity conditions are (H1), as given in section 4, plus the following:

$$\begin{aligned}
(\text{H2}') \quad \alpha_p + \sum \omega_j + \eta_\pi + \tau = 1, & \quad (\text{H3}') \quad \alpha_{pp} + \sum_t \phi_{pt} + \psi_{pr} + \psi_{p\pi} = 0, \\
(\text{H4}') \quad \phi_{pt} + \sum_s \omega_{st} + \omega_{tr} + \omega_{\pi t} = 0 \quad \forall t, & \quad (\text{H5}') \quad \tau_{rr} + \psi_{pr} + \sum_s \omega_{sr} + \eta_{\pi r} = 0, \\
(\text{H6}') \quad \theta_{pj} + \sum_t \gamma_{jt} + \gamma_{j\pi} + \gamma_{jr} = 0 \quad \forall j, & \quad (\text{H7}') \quad \eta_{\pi\pi} + \psi_{p\pi} + \sum_s \omega_{s\pi} + \eta_{\pi r} = 0, \\
(\text{H8}') \quad \psi_{pn} + \sum_s \omega_{sn} + \tau_{rn} + \eta_{\pi n} = 0, & \quad (\text{H9}') \quad \psi_{pk} + \sum_s \omega_{sk} + \tau_{rk} + \eta_{\pi k} = 0, \text{ and} \\
(\text{H10}') \quad \frac{1}{2} \alpha_{pp} + \frac{1}{2} \sum_s \sum_t \omega_{st} + \sum_t \phi_{pt} + \frac{1}{2} \tau_{rr} + \frac{1}{2} \eta_{\pi\pi} \\
+ \psi_{p\pi} + \psi_{pr} + \sum_s \omega_{sr} + \sum_s \omega_{s\pi} + \eta_{\pi r} = 0.
\end{aligned}$$

The adding up conditions are (A1), (A2), (A4), (A5), (A7), (A8), and (A9) as given in section 4, plus the amended condition:

$$(\text{A3}') \quad \sum_i \phi_{pi} + \psi_{p\pi} = 0.$$

(Note that (A6) is dropped.)

The amended conditions for risk neutrality are then:

$$(\text{P1}') \quad \eta_\pi = \eta_{\pi\pi} = \psi_{p\pi} = \gamma_{j\pi} = \omega_{s\pi} = \eta_{\pi n} = \eta_{\pi k} = 0 \quad \forall j, s,$$

$$(\text{P2}') \quad \alpha_p = \alpha_{pp} = \theta_{pj} = \phi_{ps} = \psi_{p\pi} = \psi_{pn} = \psi_{pk} = 0 \quad \forall j, s,$$

and, omitting the restrictions on ν_j and μ , since testing whether they hold is not feasible,

$$(\text{P3}') \quad \kappa = 0.$$

Table 1: Summary Statistics of the Data

variable	sample mean	sample std. dev.	minimum	maximum
y_1^\dagger	1490521.75	2739095.05	1038.50	26541000.00
y_2^\dagger	1606152.95	3114969.78	15956.00	23962500.00
y_3^\dagger	672172.08	1070269.29	8762.50	11781500.00
y_4^\dagger	470269.42	1486811.00	713.00	11868570.50
y_5^\dagger	1326574.19	2232463.87	13549.50	20454834.50
p_1^\ddagger	0.109	0.017	0.026	0.205
p_2^\ddagger	0.107	0.020	0.022	0.187
p_3^\ddagger	0.123	0.027	0.028	0.279
p_4^\ddagger	0.085	0.065	0.001	0.533
p_5^\ddagger	0.088	0.014	0.055	0.174
\tilde{p}^\ddagger	0.104	0.011	0.064	0.161
w_1^\blacklozenge	33.092	9.857	18.140	92.178
w_2^\ddagger	0.396	0.178	0.116	1.378
w_3^\ddagger	0.060	0.009	0.028	0.108
w_4^\ddagger	0.087	0.033	0.039	0.350
w_5^\ddagger	0.081	0.017	0.027	0.233
p_π^\ddagger	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
$p_\pi \pi^\dagger$	223763.03	360436.38	17125.04	3116164.55
$\mathbf{p} \cdot \mathbf{y} + m^\dagger$	667979.16	1020863.51	90000.55	8419110.74
n^\dagger	157495.47	387914.46	1254.00	3629843.00
k^\dagger	561765.95	1069979.99	69516.50	8787000.00
m^\dagger	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 $^\blacklozenge$ 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, nonprofits, 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 = share of input i ; s_π = profit share; $p_\pi \pi$ = nominal, before-tax accounting profit; $\mathbf{p} \cdot \mathbf{y} + m$ = expected revenue; n = nonperforming loans; k = financial 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^\dagger	357760.68	657490.28	1281000.08	3680519.81
y_2^\dagger	249271.21	467700.81	952710.22	4780167.50
y_3^\dagger	188318.96	326874.65	639945.94	1538866.02
y_4^\dagger	32843.49	80126.14	157397.63	1620612.07
y_5^\dagger	325200.96	552054.45	964296.82	3480755.74
p_1^\ddagger	0.108	0.111	0.112	0.105
p_2^\ddagger	0.115	0.106	0.109	0.098
p_3^\ddagger	0.124	0.119	0.125	0.125
p_4^\ddagger	0.083	0.095	0.082	0.082
p_5^\ddagger	0.086	0.085	0.086	0.093
\tilde{p}^\ddagger	0.105	0.103	0.105	0.101
w_1^\blacklozenge	29.372	30.886	33.052	39.090
w_2^\ddagger	0.382	0.387	0.393	0.421
w_3^\ddagger	0.060	0.060	0.059	0.060
w_4^\ddagger	0.080	0.080	0.086	0.102
w_5^\ddagger	0.080	0.081	0.084	0.081
p_π^\ddagger	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
$p_\pi \pi^\dagger$	47708.37	84558.38	169048.99	596467.64
$\mathbf{p} \cdot \mathbf{y} + m^\dagger$	140029.89	249238.77	489462.14	1801597.91
n^\dagger	18162.45	49266.10	84958.04	480141.30
k^\dagger	106021.51	187606.21	359011.36	1602550.27
m^\dagger	18240.58	33365.32	68254.24	379166.96
total assets †	1335138.53	2393107.19	4607230.76	15191623.49

† in thousands of dollars ‡ in dollars per dollar $^\blacklozenge$ 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, nonprofits, 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 = share of input i ; s_π = profit share; $p_\pi \pi$ = nominal, before-tax accounting profit; $\mathbf{p} \cdot \mathbf{y} + m$ = expected revenue; n = nonperforming loans; k = financial capital; m = noninterest income.

Table 3: Coefficient Estimates

parameter	estimate	parameter	estimate	parameter	estimate
α_0	101.846 (69.98)	α_p	24.576* (14.92)	δ_1	0.199 (2.00)
δ_2	-0.683 (1.99)	δ_3	-2.199 (1.74)	δ_4	3.045* (1.68)
δ_5	-0.705 (2.00)	ω_1	-2.857* (1.56)	ω_2	-1.235 (0.76)
ω_3	-7.345 (5.23)	ω_4	8.667 (5.32)	ω_5	-1.187 (1.23)
η_π	4.958* (3.00)	ϑ	-0.342 (1.33)	ρ	-0.979 (0.62)
δ_{11}	0.080 (0.08)	δ_{12}	-0.147* (0.08)	δ_{13}	-0.065 (0.04)
δ_{14}	0.006 (0.04)	δ_{15}	-0.024 (0.07)	δ_{22}	0.088 (0.07)
δ_{23}	0.163** (0.08)	δ_{24}	0.001 (0.04)	δ_{25}	-0.117 (0.08)
δ_{33}	0.025 (0.06)	δ_{34}	-0.022 (0.03)	δ_{35}	0.015 (0.05)
δ_{44}	0.045** (0.02)	δ_{45}	0.008 (0.04)	δ_{55}	0.00007 (0.11)
α_{pp}	2.128 (2.93)	ω_{11}	0.131*** (0.04)	ω_{12}	0.040** (0.02)
ω_{13}	0.088 (0.10)	ω_{14}	-0.155 (0.10)	ω_{15}	0.025 (0.03)
ω_{21}	0.022 (0.02)	ω_{22}	0.019** (0.01)	ω_{23}	0.061 (0.05)
ω_{24}	-0.066 (0.05)	ω_{25}	0.019 (0.01)	ω_{31}	0.104 (0.09)
ω_{32}	0.050 (0.05)	ω_{33}	0.786** (0.33)	ω_{34}	-0.523 (0.33)
ω_{35}	-0.021 (0.07)	ω_{41}	-0.179* (0.09)	ω_{42}	-0.077* (0.05)
ω_{43}	-0.595* (0.31)	ω_{44}	0.667** (0.32)	ω_{45}	-0.117 (0.08)
ω_{51}	0.018 (0.02)	ω_{52}	0.004 (0.01)	ω_{53}	0.082 (0.08)
ω_{54}	-0.091 (0.08)	ω_{55}	0.074*** (0.03)	ϑ_{nn}	0.112* (0.07)

continued on next page

continued from the previous page

ρ_{kk}	0.012** (0.01)	θ_{p1}	-0.524 (0.50)	θ_{p2}	0.019 (0.61)
θ_{p3}	-0.192 (0.36)	θ_{p4}	0.816* (0.46)	θ_{p5}	-0.901 (0.56)
$\psi_{p\pi}$	0.625 (0.51)	ϕ_{p1}	-0.268 (0.26)	ϕ_{p2}	-0.127 (0.12)
ϕ_{p3}	-0.993 (0.93)	ϕ_{p4}	0.740 (0.95)	ϕ_{p5}	-0.037 (0.17)
ψ_{pn}	-0.036 (0.44)	ψ_{pk}	-0.065 (0.10)	$\gamma_{1\pi}$	-0.036 (0.08)
$\gamma_{2\pi}$	-0.060 (0.07)	$\gamma_{3\pi}$	-0.015 (0.06)	$\gamma_{4\pi}$	0.051 (0.04)
$\gamma_{5\pi}$	-0.048 (0.08)	γ_{11}	0.009 (0.04)	γ_{12}	0.005 (0.02)
γ_{13}	0.042 (0.14)	γ_{14}	-0.030 (0.15)	γ_{15}	0.009 (0.02)
γ_{21}	0.024 (0.04)	γ_{22}	0.009 (0.02)	γ_{23}	0.114 (0.13)
γ_{24}	-0.119 (0.14)	γ_{25}	0.032 (0.03)	γ_{31}	0.022 (0.03)
γ_{32}	0.009 (0.01)	γ_{33}	0.073 (0.11)	γ_{34}	-0.068 (0.11)
γ_{35}	-0.020 (0.02)	γ_{41}	-0.023 (0.02)	γ_{42}	-0.011 (0.01)
γ_{43}	-0.104 (0.08)	γ_{44}	0.103 (0.08)	γ_{45}	-0.016 (0.02)
γ_{51}	0.018 (0.04)	γ_{52}	0.007 (0.02)	γ_{53}	0.047 (0.15)
γ_{54}	-0.030 (0.15)	γ_{55}	0.006 (0.02)	γ_{1n}	0.077 (0.06)
γ_{2n}	0.003 (0.06)	γ_{3n}	-0.031 (0.04)	γ_{4n}	-0.066* (0.04)
γ_{5n}	0.004 (0.06)	γ_{1k}	-0.00004 (0.02)	γ_{2k}	0.011 (0.01)
γ_{3k}	0.006 (0.01)	γ_{4k}	-0.009 (0.01)	γ_{5k}	0.007 (0.02)
ω_{1n}	-0.030 (0.03)	ω_{2n}	-0.009 (0.01)	ω_{3n}	-0.083 (0.10)
ω_{4n}	0.090 (0.10)	ω_{5n}	-0.012 (0.02)	$\omega_{1\pi}$	-0.130** (0.06)

continued on next page

continued from the previous page

$\omega_{2\pi}$	-0.055* (0.03)	$\omega_{3\pi}$	-0.395** (0.19)	$\omega_{4\pi}$	0.302* (0.18)
$\omega_{5\pi}$	-0.087* (0.05)	$\omega_{\pi 1}$	-0.064 (0.07)	$\omega_{\pi 2}$	-0.036 (0.03)
$\omega_{\pi 3}$	-0.474** (0.19)	$\omega_{\pi 4}$	0.368** (0.18)	$\omega_{\pi 5}$	-0.099 (0.08)
$\eta_{\pi n}$	0.045 (0.06)	$\eta_{\pi k}$	-0.009 (0.02)	$\eta_{\pi\pi}$	0.306** (0.14)
ω_{1k}	0.023 (0.01)	ω_{2k}	0.009 (0.01)	ω_{3k}	0.033 (0.04)
ω_{4k}	-0.075** (0.04)	ω_{5k}	0.019 (0.01)	ϑ_{nk}	-0.010 (0.01)
ν_1	-0.034** (0.01)	ν_2	-0.015** (0.01)	ν_3	-0.123*** (0.04)
ν_4	0.126*** (0.04)	ν_5	-0.019 (0.01)	μ	0.065*** (0.02)
κ	-0.013*** (0.005)				

Note: Estimates reported are for the model without the symmetry restrictions (S3) and (S4) imposed. A Wald test of the 15 symmetry conditions yielded a test statistic of 39.7, implying a p-value of 0.0005. Although these symmetry conditions were rejected, the results reported in section 10 are qualitatively similar whether symmetry is imposed or not.

The parameters are defined in the main text. The standard errors are provided in the parentheses. The symbols *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively, for an asymptotic two-tailed t -test.

Table 4: Scale Economies and Other Elasticities

measure	whole sample	1st quartile	2nd quartile	3rd quartile	4th quartile
SCALE	1.146*** (0.02)	1.101*** (0.03)	1.128*** (0.02)	1.146*** (0.02)	1.208*** (0.03)
$\frac{\partial \ln C}{\partial \ln y_1}$	0.317*** (0.01)	0.328*** (0.01)	0.343*** (0.01)	0.335*** (0.01)	0.261*** (0.01)
$\frac{\partial \ln C}{\partial \ln y_2}$	0.197*** (0.01)	0.192*** (0.01)	0.177*** (0.01)	0.192*** (0.01)	0.227*** (0.01)
$\frac{\partial \ln C}{\partial \ln y_3}$	0.120*** (0.01)	0.122*** (0.01)	0.122*** (0.01)	0.131*** (0.01)	0.107*** (0.01)
$\frac{\partial \ln C}{\partial \ln y_4}$	0.031*** (0.01)	0.025*** (0.01)	0.029*** (0.005)	0.028*** (0.01)	0.042*** (0.01)
$\frac{\partial \ln C}{\partial \ln y_5}$	0.225*** (0.01)	0.252*** (0.01)	0.231*** (0.01)	0.205*** (0.01)	0.213*** (0.01)
$\frac{\partial \ln C}{\partial \ln k}$	-0.032** (0.01)	-0.033** (0.01)	-0.033** (0.01)	-0.032** (0.01)	-0.032** (0.01)
$\frac{\partial \ln k}{\partial \ln y_1}$	0.182 (0.14)	0.066 (0.17)	0.123 (0.15)	0.261 (0.16)	0.279 (0.21)
$\frac{\partial \ln k}{\partial \ln y_2}$	0.549** (0.21)	0.575** (0.23)	0.554** (0.21)	0.529** (0.21)	0.538** (0.27)
$\frac{\partial \ln k}{\partial \ln y_3}$	0.046 (0.14)	0.287** (0.13)	0.168 (0.12)	0.052 (0.14)	-0.327 (0.26)
$\frac{\partial \ln k}{\partial \ln y_4}$	-0.117 (0.11)	-0.247* (0.14)	-0.115 (0.10)	-0.121 (0.11)	0.014 (0.13)
$\frac{\partial \ln k}{\partial \ln y_5}$	-0.253 (0.24)	-0.444 (0.31)	-0.374 (0.28)	-0.214 (0.22)	0.023 (0.20)
$\frac{\partial \ln k}{\partial \ln n}$	0.054 (0.11)	0.262* (0.14)	0.089 (0.11)	-0.010 (0.11)	-0.124 (0.16)
$\frac{\partial \ln k}{\partial \ln \bar{p}}$	-1.624 (1.44)	-2.681 (1.95)	-1.605 (1.55)	-1.072 (1.30)	-1.148 (1.31)
$\frac{\partial \ln x_4}{\partial \ln \bar{p}}$	-1.612 (1.56)	-6.125* (3.22)	-0.823 (1.87)	0.102 (1.00)	0.363 (0.64)
$\frac{\partial \ln x_4}{\partial \ln k}$	-0.328*** (0.09)	-0.513*** (0.15)	-0.382*** (0.11)	-0.246*** (0.07)	-0.173*** (0.05)
$\frac{\partial \ln x_4}{\partial \ln n}$	-0.111 (0.10)	0.212 (0.24)	-0.250** (0.13)	-0.199*** (0.07)	-0.204*** (0.07)

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.) The Note in Table 3 also applies here.

Figure I: Scale Economies Against Total Assets