http://clevelandfed.org/research/review/
Best available copy

ECONOMIC REVIEW

1987 QUARTER 4

2 **Learning, Rationality, the Stability of Equilibrium and Macroeconomics.** The issue of how agents learn to form rational expectations has received increasing attention lately. The approach taken in many papers treats model stability as a problem in learning. In reviewing this literature, the author examines carefully the assumptions about individual behavior required for learning to form rational expectations. The meaning of rationality in a macroeconomy characterized by highly decentralized markets is also discussed.

13 **Airline Hubs: A Study of Determining Factors and Effects.** One of the most widely noted and least studied changes in the airline industry has been the switch to hub-and-spoke networks. While there have been some analyses that explain the advantages and effects of hub-and-spoke networks, there have been no attempts to study empirically the determinants that influence where airlines choose to establish their hubs. This paper provides insights into the future evolution of the airline industry by identifying the quantitative effects of these determinants and examining the effect of hub status on airport traffic.

20 **A Comparison of Risk-Based Capital and Risk-Based Deposit Insurance.** The author develops a bank-risk model, based on six FDIC variables for predicting bank failure or loss, and uses it to compare alternative proposals for controlling the level of bank risk. He finds that both risk-based capital and risk-based insurance systems would affect banks' behavior. The impact of the two systems, however, would most likely not be identical, but implementation of either system would probably lead to significant progress in the effort to control bank risk.

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. Copies of the issues listed here are available through our Public Information Department, 216/579-2047.

Editor: William G. Murmann
Assistant Editor: Robin Ratliff
Design: Michael Galka
Typesetting: Liz Hanna

Opinions stated in Economic Review are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or of Ihe Board of Governors of the Federal Reserve System.

Material may be reprinted provided that the source is credited. Please send copies of reprinted materials to the editor.

ISSN 0013-0281

Learning, Rationality, the Stability of Equilibrium and

by John B. Carlson

John B. Carlson is an economist at the Federal Reserve Bank of Cleveland. The author would like to acknowledge helpful criticisms from Richard Kopcke, Mark Sniderman, E.J. Stevens, Alan Stockman, Owen Humpage, Steve Strongin, William T. Gavin, Randall W. Eberts, and Charles Carlstrom, on earlier drafts.

Introduction

It is sometimes argued that the strength in models that assume rational expectations is the weakness of their competitors. For example, McCallum (1980) says: "Each alternative expectational hypothesis, that is, explicitly or implicity posits the existence of some particular pattern of systematic expectational error. This implication is unattractive, however, because expectational errors are costly. Thus, purposeful agents have incentives to weed out all systematic components."

This alluring intuition, however, glosses over a very difficult problem that remains unsolved in general: How do agents acquire the information and understanding sufficient to enable them to "weed out" systematic error? The acquisition of information is costly and no one actually believes anyone knows the true underlying model of the economy. Discovering systematic error is one thing; knowing what to do about it is another. The central issue is one of learning.

The problem of learning in models that assume rational expectations has received increasing attention lately.[1] The approach taken in many papers treats stability of equilibrium as a problem in learning. That is, the issue of convergence to rational expectations equilibrium (REE) is presumed tantamount to the question of how agents acquire sufficient information to weed out systematic expectational error. While several modeling approaches have found such "stability" under different and reasonably plausible assumptions, there are no general theorems. More importantly, however, even the limited results found in these models presume continuous market clearing. Thus, the meaning of stability is quite restricted. The fundamental issue—how individual behavior will lead to the necessary price adjustment—is never explicitly modeled. Neglect of this issue is not new; it has long hindered progress in general equilibrium theory.

The purpose of this paper is to examine carefully the assumptions about individual behavior required for stability in models where agents learn to form rational expectations. Section one provides a restatement of the importance of stability analysis for deriving meaningful results from equilibrium models, and introduces the idea of developing learning models to describe the transition process to systemic equilibrium.

To illustrate the correspondence between learning processes and stability of REE, two examples are presented. The first, presented in section two, presumes rational agents know the structure (that is, the functional form) of the true economic model, but not the parameters. The example presented in section three presumes agents don't even know the model structure while they are learning. The precise meaning of stability in both models is discussed in section four. A distinction is made between expectational equilibrium and equilibrium of the aggregative economy. In section five, we discuss the difficulties facing the researcher who seeks to model

1 For a concise review of these models see Blume, Bray, and Easley (1982).

learning in an aggregative economy. The issues are developed in a general model employing a notion of equilibrium proposed by Frank Hahn (1973). Section six offers concluding remarks.

I. Importance of Stability

Analysis of positions and characteristicsof equilibrium is by far the most widely accepted mode of economic analysis. Typically, such equilibria are derived from (or presumed to be) the solution of individual optimization problems. A key hypothesis that begets coordination of individual plans (aggregative consistency) is that certain variables—usually prices—take on values that make all individual plans mutually consistent. Under these circumstances, no individual has any incentive for further change. Economists rarely specify a behavioral process that could account for how variables, like prices, adjust to recoordinate individual plans when conditions change. Rather "changes" in equilibrium outcomes are generally developed in comparative static analysis, which compares equilibria corresponding to different values of underlying parameters.

The use of comparative statics in economics was first explained in rigorous detail by Samuelson (1947). He recognized, however, that to obtain definite *operationally meaningful* theorems in comparative statics, one has to specify a hypothesis about the dynamical properties that will lead to equilibrium values. The 'duality' between the problem of stability and the problem of deriving fruitful theorems in comparative statics is what Samuelson called the Correspondence Principle.

The importance of dynamical foundations has recently been restated by Fisher (1983). He argues that if general equilibrium models are to be of any use then we must have some confidence that the system is *stable,* that is, that it must converge to an equilibrium, *and* that such convergence to equilibrium must take place relatively quickly:

> If the predictions of comparative statics are to be interesting in a world in which conditions change, convergence to equilibrium must be sufficiently rapid that the system, reacting to a given parameter shift, gets close to the predicted new equilibrium before parameters shift once more. If this is not the case, and *a fortiori,* if the system is unstable so that convergence never takes place, then what will matter will be the 'transient' behavior of the system as it reacts to disequilibrium. Of course, it will then be a misnomer to call such behavior 'transient' for it will never disappear. (p. 3)

Fisher goes on to emphasize his point in the context of models assuming rational expectations:

> In such models, analysis generally proceeds by finding positions of rational expectations equilibrium if they exist. At all other points, agents in the model will have arbitrage opportunities; one or another group will be able systematically to improve its position; ....The fact that arbitrage will drive the system away from points that are *not* rational expectations equilibria does not mean that arbitrage will force the system to converge to points that are rational expectations equilibria. The latter proposition is one of stability and it requires a separate proof. Without such a proof—and, indeed without a proof that such convergence is rapid—there is no foundation for the practice of analyzing only equilibrium points of a system which may spend most or all of its time far from such points and which has little or no tendency to approach them. (pp. 3-4)

Fisher argues that analysis of this problem requires a full-dress model of disequilibrium — one that is based on explicit behavior of optimizing agents.[2] A general model would accommodate trading, consumption and production while the model is out of equilibrium. That is, such an approach would provide a theoretically based alternative to the Walrasian auctioneer. Arbitrage would follow from *individual* rationality. Unfortunately, practitioners of this approach have not advanced the subject enough to address the stability of model-consistent (that is, of rational) expectations.

The stability of REE has been addressed, extensively, however, on a less fundamental level. This approach presumes that markets clear and that REE is the true underlying *long-run* equilibrium. It examines different processes by which agents might acquire (learn) the information necessary for an expectations equilibrium consistent with REE. An important paper by Cyert and DeGroot (1974) defends the use of models of the learning process:

> The attempt to develop process models immediately opens us to the criticism of developing *ad hoc* models. We acknowledge that there may be a large number of models that could potentially describe the process to equilibrium. Our position is

2 Fisher (1983) does make a contribution in this direction but only under the assumption of perfect foresight His monograph illustrates the burden that lies ahead of any serious theoretician in this matter.

> that, while the models have a certain amount of face validity, our major contribution is the introduction of an explicit learning process described in Bayesian terms. The notion of developing models to describe the transition process toward equilibrium of a system disturbed by some random shocks may be questioned by some economists. The development of comparative statics and the neglect of dynamic analysis is in part a reflection of such attitudes in the profession. Yet without well-developed process models, the concept of rational expectations is essentially a black box. (p. 522)

Thus, models of the learning process are essentially provisional tools that enable us to interpret REE in a more realistic way. We may think of the development of such models as an attempt to justify the use of the rational expectations hypothesis.

These models, at the very least, allow us to ask if it is conceivable that agents could "learn their way" to equilibrium in the model at hand. This problem is not simple. Because agents are presumed to base their decisions on their own estimates of a model's parameters, their actions cannot be considered exogenous to parameter estimation. If estimates of parameters change, agents adjust their behavior accordingly. Moreover, agent actions generate the data on which the estimates of parameters are made, making learning an endogenous process. To correctly specify the model, agents would need to take the endogeneity into account. Conventional econometric techniques are typically not well-suited for this task.

The question of convergence to REE has been examined in two frameworks. The first assumes that agents know the functional form of the model or, at least, the appropriate specification of the likelihood function underlying the generation of the data. In this framework, agents are presumed to learn about the value of parameters either through classical statistical methods, repeated use of Baye's Theorem, or some other statistical method. The second framework does not require that agents know the model, although some of this work assumes that agents base their expectations on the basis of one model chosen from a set that includes the true model.

II. Learning When Agents Know the Model

To illustrate a process of learning and its connection to the stability of REE, we first examine one approach taken by Cyert and DeGroot (1974). They proposed to design models that describe the process by which rational expectations may develop within a market. They build on a version of the cobweb model used by Muth (1961) to propose the concept of rational expectations. Muth posited a partial equilibrium model for a homogeneous good with a production lag. Using the notation of Bray and Savin (1986), the market equations have the following form in any period t:

$$(1)\quad d_t = m_1 - m_2 p_t \qquad \text{(demand)}$$
$$(2)\quad s_t = m_0 + m_5 p_t^e + v_{2t} \qquad \text{(supply)}$$
$$(3)\quad d_t = s_t \qquad \text{(equilibrium),}$$

where m_0, m_1, m_2, and m_5 are fixed parameter values; p, is the market-clearing price of the good; p_t^e is the market-anticipated price before trade takes place; and v_2, is an exogenous shock to supply. It is assumed that all units demanded are consumed in period t and that firms make production decisions before trade takes place. Thus, the deterministic component of supply is fixed in period t.

The assumption of market clearing yields:

$$(4)\quad p_t = M - a p_t^e + u_t,$$

where $M = (m_1 - m_0)m_2^{-1}$, $a = m_5 m_2^{-1}$ and $u = -m_2^{-1} v_{21}$.

Under the usual assumption of rational expectations, the market-anticipated price equals the objective mathematical expectation for price given the model and as conditioned on the data available when the expectation was formed.[3] That is, $p_t^e = E_{t-1}(p_t)$. Cyert and Degroot propose a similar basis for determining p_t^e. They assume that expectations are consistent, meaning that the firms' expectations are based on the mechanism implied by the model. The essence of this distinction is that while agents are presumed to know the correct likelihood functions, they are not required to know the parameter values. Cyert and DeGroot derive an explicit expression for market-anticipated price by taking expections of both sides of (4), substituting p_t^e for $E_{t-1}(p_t)$ and solving for p_t^e:

$$(5)\quad p_t^e = \frac{E_{t-1}M - E_{t-1}[u_t]}{1 + E_{t-1}(a)}$$

Note that since the parameter values are unknown, the market-anticipated price is expressed in terms of *expected* values of the parameters, not true values. Agents (firms) learn to form rational expectations if, with additional data, the expected values of the parameters converge to their true values. Note also that market-anticipated price will differ from actual market

3 It is perhaps more accurate to call such expectations model-consistent instead of "rational." (See Simon 1978).

price both because of expectional error and the supply shock.

The economic process evolves as follows: In each period, the firms form consistent expectations of the price in the next period h-om (5) based on expected parameter values (priors). The actual price is then generated according to the model incorporating the consistent expectations, that is, price is given by (4). The observed values of actual price contains new information that leads firms to change their expectations of the values of the parameters and, hence, to change their expectations of the price in the following period. The actual price in the next period is again generated by the model and the process continues in this manner.

Cyert and DeGroot verify that such a process can, in fact, converge to REE when slope coefficients m_2 and m_5 are known, even if intercepts m_0 and m, are not. In this example, the authors assume that the random (supply intercept) error has a normal distribution with mean 0 and known precision (inverse of variance). Moreover, they posit a posterior distribution for M at the end of period t-1 that is normal with finite mean and precision. Finally, they show that a Bayesian updating of parameter values does converge to the true value of M.

The convergence result was encouraging. It showed that one need not assume all knowledge is innate, but that, h-om a Bayesian point of view, the relationship between expectations and other variables in the model arises naturally when economic agents form expectations in a manner internally consistent with the mechanism generating the data. In simple terms, this means that agents can learn parameter values even though their expectations affect outcomes of the model. An essential assumption is that all agents can correctly specify likelihood functions of unknown parameters, that is, that they "know" the structure of the model.

An implicit assumption underlying this and all other models obtaining convergence when agents know the model is that the solution concept being employed is Nash equilibria. This means that each agent has no reason to alter his specification of the likelihood function, given his own specification and those of all other agents. Thus, the approach assumes not only that agents know the model, but also that agents know that other agents know the model. The implications of this are discussed by Blume, Bray, and Easley (1982):

> The concept of a Nash equilibrium in learning strategies has much to commend it. Any other learning process is to some degree ad hoc; if some or all of the agents are learning by using mis-specified models, at some stage they should realize this and change the specification. Nash equilibria in learning strategies are rational expectations equilibria in which agents take into account their uncertainty about features of the world which they are assumed to know in standard models of rational expectations equilibria. However, Nash equilibria in learning strategies are liable to be considerably more informationally demanding than conventional rational expectations equilibria, as agents require extensive knowledge about the structure and dynamics of the model that prevails while they learn. There may also be problems with the existence of equilibrium. Thus, while this approach yields convergence to a conventional rational expectations equilibrium, its extreme informational demands make it an unsatisfactory answer to the initial question of how agents learn how to form rational expectations. (p.315)

In sum, employing the Nash solution concept begs the question as to how agents learn the structural form of the underlying model. Moreover, it provides no economic justification for why any agent should believe that all other agents will know what forecast methods other agents use. What incentives are there for such behavior?

III. Learning When Agents Don't Know the Model

When agents know the structural form of the economy, it is a relatively straightforward task to identify informational requirements sufficient to obtain convergence to REE. As we have seen, however, these requirements are quite demanding. They presume that agents have extensive knowledge about what other agents believe as they all learn about the parameters. It is somewhat interesting, however, that in situations where agents don't know the model, convergence can occur under somewhat weaker assumptions about the learning process. These results, however, are model specific. Other, equally reasonable, approaches lead to instability of REE. Achieving convergence depends not only on the nature of learning but on the structural and stochastic parameters of the underlying model.

When agents don't know the model, the problem of learning has been addressed in two distinct ways. The first approach provides an explicit model that allows agents to modify their forecasting rules in light of observable outcomes (see Blume and Easley [1982]). Typically, they choose among a set of models that includes the true one. Convergence occurs when

all agents eventually adopt the true model. In this approach, we find that the results are mixed. In some models, rational expectations equilibria are locally stable but not unique.

The second approach examines the possibility of convergence when agents never switch models, despite the fact that they may have misspecified the model while they are learning. Essentially, this approach considers whether "irrational" learning can lead to rational expectations equilibrium.

An interesting model by Bray and Savin (1986) examines the second kind of learning. An appealing feature of this model is that agents learn using conventional techniques—such as by estimating the parameters of a standard linear-regression model. While this is the correct econometric specification for their postulated model in *equilibrium,* the econometric model is misspecified while people are learning. Moreover, Bray and Savin use simulations to examine the rate at which convergence takes place and to assess the possibility that agents discover that their estimated model is misspecified.

Following Townsend (1978), they extend the cobweb model to include stochastic demand, to allow for exogenous shocks to aggregate supply, and to accommodate diversity of firm expectations and decisions. All firms are assumed to face the same technology as defined by a quadratic cost function

$$c_{it} = q_{it}^2/2m_5,$$

where $m_5 > 0$ and q_{it} is the output of firm i at date t. Under the profit-maximizing postulate, firm i chooses an output level equal to $m_5 p_{it}^e$ where p_{it}^e is the mean of its prior on market-anticipated price.4

The aggregate of these expectations over all firms is denoted as $\bar{p}_t^e$. Their model is thus given by:

(6) $d_t = m_1 - m_2 p_t + v_1,$ (demand)
(7) $s_t = m_5 \bar{p}_t^e + x_t' m_4 v_{2t},$ (supply)
(8) $d_t = s_t$ (equilibrium),

where $x_t' m_4 + v_{2t}$ is an exogenous supply shock and x_t' is observable. Market clearing implies that:

(9) $p_t = x_t' m + a\bar{p}_t^e + u_t,$

where x_t' is redefined to include 1 as the first component and $m = [m_1 : m_4] m_2^{-1}$ and as in (4) $a = m_5 m_2^{-1}$, but $u_t = (v_{1t} - v_{2t}) m_2^{-1}$.

If agents knew both the model structure and the values of the parameters, the REE price forecast would be:

(10) $p_{it}^e = x_t' m (1-a)^{-1}$

for all i, assuming $a \neq 1$. Together (9) and (10) imply that the REE price, for each t, is:

(11) $p_t = x_t' m (1-a)^{-1} + u_t.$

The linear relationship between actual price and exogenous-supply influences applies only in equilibrium when agents all share the same expectations. This simple relationship does not hold when agents are learning the values of the parameters. To illustrate this, Bray and Savin assume agents maintain the hypothesis that:

(12) $p_t = x_t' b + u_t$

satisfies the assumptions of the standard linear model, and estimate b accordingly. They consider the consequences that agents may be classical or Bayesian statisticians. If all agents (firms) are Bayesian statisticians who assume u_t is *i.i.d.* as $N(0,\varphi^2)$, and if firm i's initial prior on b is b_{io} and prior on precision is S_0/σ^2, firm i may obtain revised priors on b after observing $(x_1, p_1), \ldots, (x_{t-1}, p_{t-1})$, which will have mean $b_{i,t-1}$, and precision S_{t-1}/σ^2 where,

(13) $$b_{i,t-1} = \left(S_0 + \sum_{j=1}^{t-1} x_j x_j'\right)^{-1} \left(S_0 b_{i,0} + \sum_{j=1}^{t-1} x_j p_j\right)$$

and

$$S_{t-1} = S_0 + \sum_{j=1}^{t-1} x_j x_j'.$$

Note that the classical statistician is essentially a Bayesian Statistician whose initial prior on b is diffuse ($S_o = 0$).

With this revised prior, agent i's forecast of p_t is $p_{it}^e = x_t' b_{it}$. The aggregate of market-anticipated price is $p_t^e = x_t' b$, where b, is an aggregate of b_{it} over all firms. Substituting this in (9) gives:

(14) $p_t = x_t'(m + ab_{t-1}) + u_t.$

Equation (14) generates the actual observed price given both the market mechanism and the way agents form expectations. Note that the coefficient of x_{t-1}, $(m + ab_{t-1})$, varies with time. Thus, agents are incorrectly assuming that price is generated by a standard linear model with a constant coefficient. The model is incorrect because it fails to take account of the effects of learning on the parameter values. If agents

4 Bray and Savin consider a continuum of firms producing a homogenous good. The set of firms is the unit interval [0,1] indexed by i. Thus, market-anticipated price is a Lebesque intergral. It is in that sense an average expected price.

knew what we know, they would not use linear regressions to form expectations.

Despite the fact that agents may misspecify the model, Bray and Savin are able to show that: (1) the difference between the individual estimates b_{it} and the average estimate b_t tends to zero with probability one as t tends to infinity; and (2) the average estimate b_t cannot converge to any value other than the REE value $m(1-a)^{-1}$. The intuition they offer is that if b, tends to b for large t, the actual price is $p_t = x_t'(m + ba) + u_t$. Since the data generation process closely approximates the standard linear model with coefficient $m + ba$, the estimate b_t tends to $m + ba$, which is impossible unless $b = m(1-a)^{-1}$.

These results enable Bray and Savin to obtain the restrictions on parameters a and b that are necessary and sufficient for existence, uniqueness, and 'stability' of the REE. The conditions are precisely the same conditions for the existence, uniqueness, and tantonnement stability of a market in which supply and demand are simultaneous, that is, a Walrasian model in which supply at time t is based on actual price at t as opposed to market-anticipated price.

The intuition behind the convergence process of the Bray-Savin model is straightforward. Suppose suppliers' beliefs are such that, in the aggregate, they underestimate price corresponding to a given set of exogenous influences. This would lead them to supply less than they otherwise would have done. Consequently, the auction would assure that the market-clearing price would be above the market-anticipated price. Taking account of the newly observed price, suppliers would, on average, raise their estimate of price corresponding to the same set of exogenous influences. Provided they don't overreact, learning would bring them closer to REE in each successive period.

An important feature of the Bray-Savin approach is that the specified learning process is reasonably simple and plausible despite the fact that the underlying mechanism is much more complicated. A potential problem, however, is that agents might discover that they have incorrectly specified the model. Since the estimated model is not the true one while they are learning, the data may confirm the misspecification. On the other hand, if convergence is sufficiently fast, their test may fail to spot the misspecification.

To examine this possibility, Bray and Savin use computer simulations. The simulations suggest that the rate of convergence can be slow if the ratio of the slopes of demand and supply are near the boundary of the stability region, especially if the initial prior mean is incorrect for REE and the prior precision is high. Thus, the fact that equilibrium may be stable may not mean much. Equilibrium behavior may not provide a reasonable enough approximation of the actual behavior to be meaningful.

Bray and Savin also use the simulations to examine the likelihood that agents will discover that their estimated model is misspecified. Agents are presumed to examine the Durbin-Watson statistic as a diagnostic check for model misspecification. The results suggest that if REE is stable, and if the estimates converge rapidly, agents are unlikely to identify the misspecification. Thus, it is reasonable to expect that agents could persist using simple linear (misspecified!) methods and eventually learn all they need to know to form expectations in a manner consistent with REE.

IV. The Meaning of Stability

The major contribution of the learning models discussed above is that they provide an explicit framework for describing a transition process toward equilibrium of a system disturbed by some random shocks.[5] While they successfully demonstrate how rational expectations may develop in a perfectly competitive market, learning models do not provide the kind of underpinnings sought by general equilibrium theorists in stability analysis. They focus only on the development of *expectational* equilibrium. No attempt is made to specify the dynamics of price formation. Rather, the framework implicitly assumes an auction process not substantively different from that required to achieve standard competitive (Walrasian) equilibrium.

Thus, these models beg the central question that continues to plague general equilibrium theorists: how to derive behavioral foundations for price adjustment. This is not a criticism specific to the models at hand, but is a fundamental problem with all equilibrium models, including fixed-price models. To appreciate the problem, it is useful to review briefly the theoretical foundations of the stability of competitive equilibrium.

Stability analysis of competitive equilibrium builds on the earliest notions about price adjustment, which were imbedded in the "law of supply and demand." It essentially holds that in competitive markets, prices will rise when there is excess demand and fall when there is

5 It is the view of Cyert and DeGroot that such a process has to be developed if the rational hypothesis is to be a scientific truth rather than a religious belief.

excess supply. This argument has the familiar dynamic formulation first proposed by Samuelson in 1941 (see 1947):

$$(15)\quad \frac{dp}{dt} = h(D-S),\ h(0) = 0, \text{and } h' > 0 \text{ and}$$

$$(16)\quad D = D(p,a) \quad S = S(p),$$

where D and S are quantities demanded and supplied for a homogeneous good; p is the market price of that good, and a is an exogenous shift parameter. The properties of the *static* demand and supply functions are derived under the standard hypothesis that households and firms maximize familiar objective functions. Formal proofs for the stability of competitive markets essentially derive sufficient conditions for the dynamic relations expressed by (15) to yield time paths of prices that approach their equilibrium values from arbitrary points.[6] Unfortunately, global stability is obtained only under very severe restrictions on excess demand functions, the most notable being the assumption that all goods be gross substitutes.

While the assumption implicit in (15) seems plausible, it is beset by some important conceptual difficulties. The first problem is that (15) has never been deduced as the maximizing response of economic agents to changing data. Sonnenschein (1973) has shown that the standard assumptions about individual behavior do not imply any restrictions on excess demand functions beyond homogeneity of degree zero and Walras' Law—conditions not sufficient for stability. Thus, adjustment to Walrasian equilibrium lacks the rigorous basis that is accorded to the properties of *static* supply and demand functions. Moreover, it is not clear who changes prices when the system is not in equilibrium. In competitive equilibrium, sellers and buyers are typically treated as price takers. Therefore, it is presumed that there is some implicit market manager who sets price.

The idea of a market manager whose behavioral rule for price adjustment is given by (15) was, of course, the ingenious answer given by Walras. This approach is tantamount to an assumption that all consumers and suppliers gather in one place. The market manager quotes a set of prices for each commodity. Then each trader writes on a piece of paper (a ticket) the amounts of each of the commodities he wishes to buy or sell at the given set of prices. If there is excess demand for the commodity i, the manager raises the price of i, if there is an excess supply for commodity j, he lowers the price of j. Each time a new set of prices is quoted, each trader submits a revised ticket. The process continues until excess demand is zero, that is, equilibrium price is determined. *Until then no trade or production takes place.*[7] Essentially, this is a description of a *timeless* process by which market clearing can be achieved and thus fails to help in understanding the *dynamics* of price.

The only difference between this Walrasian situation and the one implied by the Bray-Savin model is that, under the latter, suppliers commit to production levels prior to trade. Suppliers therefore must base their decisions for output levels on the anticipated price for their good. While these anticipated prices may initially differ when suppliers use Bayesian learning models, the observed market-clearing price at any point in time must be the same for all suppliers. Because the model used by suppliers to determine anticipated price specifies the *single* market-clearing price as the dependent variable, a tantonnement process is necessary to generate data that is essential for the process to be operational. Clearly, the auction process plays an essential role in consolidating information that is necessary for convergence.

A key distinction between the Bray-Savin process and a pure Walrasian process involves a restriction on what suppliers can learn about the *aggregate* supply function. In a standard Walrasian auction, suppliers are free to adjust the quantities they would produce for all the prices quoted. In this way, the auction process also synthesizes for all agents all the relevant properties about both aggregate supply and demand. In the Bray-Savin model, on the other hand, suppliers offer the same quantity for all prices quoted. The auction essentially determines the point on the demand curve that corresponds to the predetermined level of output. That is, the auction synthesizes only responses of consumers to the array of price quotes. Suppliers learn from the (temporaty) equilibrium price about whether they under or overestimated prices, but they do not know how well other suppliers estimated prices and, consequently, how aggregate supply might adjust to different prices. This information is revealed only through a succession of auction outcomes.

Notwithstanding information lags, the situation in the Bray-Savin model may not be very plausible for markets where prices are not

6 See Arrow and Humicz, (1958) and Arrow, Humicz, and Block (1959).

7 The requirement that no trade take place before equilibrium is determined is essential if such a process is to converge to a unique equilibrium. Fisher (1983) shows how trading at "false" prices affects endowments of agents and, hence, the ultimate outcome of the process. Thus equilibrium would depend not only on initial endowments, but also on the process that achieves equilibrium. Such a property is sometimes called hystersis.

determined by auction processes, even though the markets may appear competitive. Arrow (1959) noted that there is an inconsistency between the assumptions required of individuals in a state of equilibrium and those necessary to explain behavior in disequilibrium. He argued that, in situations of excess demand, firms do not behave as price takers but, in fact, use price-setting tactics similar to the profit-maximizing tactics of a monopolist.

The problem is somewhat more complex in that a firm's competitors will also be raising prices. Moreover, on an individual basis, no seller would have the incentive to agree to an auctioneer, since the market-clearing price would be less than what he could obtain in disequilibrium. In situations of excess supply, Arrow shows that firms are still monopolists, but buyers are monopsonists; thus, it is a joint decision that establishes price. The lesson is that disequilibrium price adjustment may need to recognize elements of imperfect competition.

Theories of imperfect competition require elements of strategic behavior, that is, situations in which two or more agents choose strategies that interdependently affect each other. Such problems involve game theory. Arrow (1986) recently concluded that analysis of games with structures that are extended over time leads to very weak implications—in the sense that there are a continua of equilibria. The fact is that we know very little about how economic man interacts with other economic men in situations of excess demand or supply. Unfortunately, the learning models considered above provide no shortcuts around this problem.

V. Learning in the Macroeconomy

While Bray-Savin learning shows that agents using "plausible" models can "learn their way" to REE in *auction* markets, it is doubtful that such a result could obtain for a highly decentralized market economy. This section identifies some difficulties, apart from the problems of modeling strategic behavior, that confront a modeler seeking to extend the Bray-Savin result to the macroeconomy. The issues are sketched using a notion of equilibrium proposed by Frank Hahn (1973).

It is the essence of a decentralized economy that individuals have different information.[8] Furthermore, each individual is specialized in certain activities and has, in general, specialized knowledge about those activities. There is no reason to believe that individuals base their expectations on the rather general kind of information that econometricians use. Instead, different individuals base their decisions on different sets of information. In short, a "plausible" model of learning in macroeconomics would need to incorporate the existence of heterogenous information.

The problem of learning when agents have incomplete and different information has recently been studied by Marcet and Sargent (1986b).[9] In their approach, agents use least-squares estimation to formulate expectations that they think are relevant to understanding the underlying law of motion as it affects them. Marcet and Sargent assume that agents do not respecify their regressions over time, but maintain the same "theory" about the world they observe. As with Bray-Savin, their model accommodates feedback from agent expectations to the actual law of motion of the system. Marcet and Sargent show that the existence of informational asymmetries does not preclude convergence to REE when the law of motion is a linear stochastic process.

While the class of learning models studied by Marcet and Sargent imposes some restrictions on the economic environment, the mechanism can accommodate a wide class of economic theories. Nothing inherent in the least-squares learning schemes precludes convergence to a non-Walrasian equilibrium.

The idea that an economic system might converge to a non-Walrasian equilibrium is, no doubt, difficult to accept for some economists. For example, won't arbitrage opportunities arise? Although there would be such opportunities vis-a-vis a Walrasian ideal, it is not evident that agents can perceive the ideal to identify the opportunities. Because agents don't observe continuous market-clearing equilibrium outcomes in a non-Walrasian environment, there is no reason that their expectations will ever become consistent with Walrasian equilibrium in the long run.

The point here is that agents' expectations *could* become consistent with the conventions (including price-setting mechanisms) that determine the laws of motion of the system. While equilibrium expectations would not be systematically inconsistent with observed outcomes of the model, agent choices would not necessarily be Pareto-optimal. Nevertheless, to the extent that market forces operate, it is conceivable that price-setting conventions could develop

8 This point and the following were made by Arrow (1978) as a criticism of the use of Muthian expectations to the aggregate economy.

9 See Marcent and Sargent (1986a).

that would lead to an equilibrium that is "approximately competitive."[10]

To understand what "approximately competitive:' might mean, it is useful to introduce a notion of equilibrium proposed by Hahn (1973). In Hahnian equilibrium, each agent holds his own theory about the way the economy will develop *and* about the consequences of his own actions.'' The agent abandons his theory when it produces systematic and persistent errors. To the extent the agent maintains a theory, his actions are conditioned on his perceptions about the laws of motion of such a system. The agent is said to be in equilibrium when he maintains his theory. The economy is said to be in equilibrium if it doesn't produce outcomes systematically and persistently inconsistent with agents' perceptions.

In the context of Marcet-Sargent learning, the theories agents hold are embodied in the regressors they choose. Under the assumption that the true law of motion is linear, agents will *ultimately* not be able to falsify their theories.[12] Thus, they would have no reason to abandon the theory. In the context of Hahn's notion, each agent would be considered in equilibrium. Moreover, since the actual outcomes would not be inconsistent with predictions of agents' theories, the economy would be in equilibrium.

Although Hahn was not completely precise about his notion of equilibrium, he clearly intended it to be more general than the equilibrium obtained in Marcet-Sargent learning. For Hahn, the *structure* of true "laws of motion" need not be independent of the theories agents choose. The theories could determine the structure of the laws of motion—a structure that could have nonlinearities that agents could never comprehend. In the model of Sargent and Marcet, the underlying structure is constrained to obey a linear (stochastic) law of motion.

Another important difference is that Hahnian equilibrium would accommodate agent behavior that could be inconsistent at any point in time, but not persistently so. In the Marcet-Sargent limit point, agents ultimately learn enough so that their expectational error is white noise, that is, agent actions lead to a steady-state equilibrium. This means that agent expectations would ultimately become mutually consistent in every period, given what they can know. Because Hahn only imposes that actions (expectations) of agents not be systematically *and persistently* inconsistent, his equilibrium would not be unique. Hence, at any point in time, equilibrium would be distinct from a steady state. Local stability would mean that, for short enough periods and for small enough disturbances, the set of equilibria is large but that it shrinks.

It is useful to stress here that the agents in the Hahnian concept of equilibrium are rational in the spirit of McCallum's intuition. That is, agents do not maintain their "theory" when systematic errors are sufficiently persistent for falsification of the theory. However, the meaning of rationality is much less restrictive (hence more plausible) than is presumed in conventional formulations of rational expectations. Agents in Hahnian equilibrium are rational only in a subjective sense. Nothing inherent in the Hahnian approach would assure that aggregate economic outcomes would converge to a stationary stochastic process with a unique *objective* probability distribution. Without such convergence, agents' subjective expectations could not coincide with an objective expectation of aggregate outcomes. Imposing the restriction that agents' *subjective* expectations be mutually consistent with each other and with a particular *objective* probability distribution underlying a given model seems too restrictive to be very useful in practice. This point has been developed in an alternative model proposed by Swamy, Barth, and Tinsley (1982).[13]

An attractive feature of Hahnian equilibrium concept is that it can accommodate more plausible market structures such as the "approximately competitive" economy suggested above. Agents may adopt stable reaction rules that allow them to cope in a competitive environment without requiring unreasonable computational abilities necessary for analyzing the aggregative

10 The meaning of "approximately competitive" equilibrium developed below is different from the sense that allocations in the core are said to be approximately competitive. The latter refers to outcomes of a bargaining process, while the former refers to outcomes derived from habitual behavior that allows agents to "survive" in a competitive economy.

11 Clearly, this notion abstracts from many difficult problems posed by strategic behavior For a more complete description of Hahn's notion of equilibrium and a comparison to the Austrian view, see Littlechild (1982).

12 It is not evident that agents would maintain their theories in the early stages of learning For any given mode one might want to provide sensitivity analysis a-la Bray-Savin.

13 Swamy et. al., show how confounding 'objective' and 'subjective' notions of probability may violate the axiomatic basis of statistical theory. They propose an alternative model for aggregation of subjective expectations. The problem with conventional formulations of the rational expectations hypothesis in macroeconomic models lies not with the concept of individual rationality but with the context in which it is developed—namely in the representative agent model. Once one allows agents to differ both in the information they have and in the theories they hold, a model can accommodate arbitrage opportunities that are deemed essential for a process leading to a rational expectations equilibrium. How agents learn to recognize arbitrage opportunities, however, remains an open, but difficult, issue.

impacts of strategic behavior. Moreover, the equilibrium of such a model would accommodate a wide variety of nonstationarities in the variables. Nevertheless, Hahnian equilibrium too has some severe limitations.

A key difficulty for a researcher modeling approximately competitive environments is that an infinite set of plausible conventions could be developed that would lead to "model consistent" (rational, in the sense of Hahn?) expectations. This may not be relevant for the individual agent in Hahnian equilibrium. The agent could be satisfied with his own conventions for dealing in his specialized corner of the world. A macromodeler, on the other hand, may not have access to all relevant information. His estimates of underlying relationships would be inconsistent because of omission of relevant explanatory variables bias. Thus, it may be impossible for a modeler of *aggregate* economic activity to discover adequately the law of motion for the economy as a whole, even when the economy is in Hahnian equilibrium. This, of course, is the essence of the Austrian criticism of macroeconomics, both Keynesian and New Classical.[14]

The most difficult problem for modeling learning in an approximately competitive model, however, is the situation in which agents change theories.[15] In the context of Hahnian equilibrium, this is the problem of global stability. That is, when a shock to equilibrium is so big, it causes agents to change their theories. Hahn argued that it is impossible to make any claims about global stability. He concluded that this limitation was imposed by the current state of economic knowledge. Economists know very little about how agents adapt to a changing economic environment.

When confronted with the limits of equilibrium analysis, economists are often more willing to invoke a convenient fiction than to modify their fundamental tools. The urge to close the model typically prevails over a venture into a methodological frontier. As is often noted, some people searching for a lost wallet at night prefer to look under a street lamp even though it may be more likely that they lost the wallet in the dark alley. Hahn's proposed reformulation of equilibrium was useful in illuminating the problems of learning in a large, decentralized economy. In this sense, it demonstrates the potential value of building new streetlamps.

VI. Concluding Remarks

This paper opened with the idea that rational, purposeful individuals have incentives to weed out systematic errors in their own expectations. Thus, it is argued that economic models should not allow expectational errors to persist. Conventional formulations of rational expectations, which assume Walrasian market-clearing, do not violate this restriction. The implicit auction process works to assure that all decisions are mutually consistent both with what agents can know about the model and with the underlying model.

This paper presented the Bray-Savin result that shows that agents may use "plausible" learning mechanisms to "learn their way" to rational expectational equilibrium in auction markets. Thus, learning models extend the results of tatonnement stability analysis to situations where agents form model-consistent expectations about the environment they are in. The restriction that economic models not permit systematic expectational errors to persist, however, does not require that agents behave in a mutually consistent manner in each period of time as in Walrasian equilibrium. The restriction is weaker than that and hence allows for a broader scope in the meaning of rationality than is generally considered in conventional formulations of the rational expectations hypothesis. That is, the restriction allows a broader class of economic models than the Walrasian economy.

The model of "approximately competitive" equilibrium sketched in this paper illustrates one potential subclass of such models. The sketch provides a plausible example of how rational, self-seeking agents might "learn their way" to non-Walrasian equilibria. Without an auctioneer in each and every market, a modeler cannot rule out such equilibria *a priori* simply by assuming agents have incentives to weed out systematic expectational errors.

14 Another way of looking at the same problem is that the specification of "approximately competitive" behavior in this paper is too general to have empirical content. Nevertheless, the researcher is free to specify his own set of conventions — provided, of course, that they are logically consistent. Because of the difficulties in falsifying economic theories, one might choose among alternative specifications on the basis of out-of-sample forecasts. The foundations of such a method are found in Swamy, Conway, and von zur Muehlen (1985).

15 This is what Hahn calls learning. It is also the sense of learning examined by Blume and Easley.

REFERENCES

Arrow, Kenneth,J., L Hurwicz, and H.D. Block, "On the Stability of Competitive Equilibrium II," *Econometrica,* vol. 27, (January, 1959) pp. 82-109.

Arrow, Kenneth,J., and L Hurwicz. "On the Stability of the Competitive Equilibrium I," *Econometrica,* vol. 26, (October, 1958) pp. 522-552.

Arrow, Kenneth J. "Rationality of Self and Others in An Economic System." TheJournal of Business, vol. 59, no. 4, part 2 (October 1986), pp. 385-399.

Arrow, Kenneth J. "Toward a Theory of Price Adjustment," The *Allocation* of Economic Resources, Stanford University Press, (1959), pp. 41-51.

Arrow, Kenneth J. "The Future and the Present in Economic Life," Economic *Inquiry,* vol. XVI, No. 2, (April 1978), pp. 157-169.

Blume, Lawrence E. and D. Easley. "Learning to be Rational,"Journal of Economic Theory, vol. 26 no. 2, (April 1982), pp. 340-351.

Blume, LE., M.M. Bray, and D. Easley. "Introduction to the Stability of Rational Expectation Equilibrium,"Journal of Economic Theory, vol. 26., no. 2 (April 1982), pp. 317-317.

Bray, M.M., and N.E. Savin. "Rational Expectations Equilibria, Learning and Model Specification," *Econometrica,* vol. 54, no. 5, September 1986. pp. 1129-1160.

Bray, M.M. "Convergence to Rational Expectations Equilibrium."Individual *Forecasters & Aggre*gate Outcomes, Cambridge University Press. 1983, (pp. 123-137).

Cyert, Richard M. and Morris H. DeGroot. "Rational Expectations and Bayesian Analysis." Journal of Political *Economics,* vol. 82 no. 3, (May/June 1974), pp. 521-536.

Fisher, Franklin M. "Disequilibrium Foundations of Equilibrium Theory," Econometric Society Monographs in Pure Theory *#6,* New York: Cambridge University Press, 1983.

Hahn, F.H. "On the Notion of Equilibrium in Economics," Cambridge: Cambridge University Press, (1973).

Littlechild, S. C. "Equilibrium and the Market Process," Method Process, and Austnan Economics: Essays in Honor of *Ludwig* Von *Mises,* ed. Israel M. Kirzner, Lexington Books, 1982.

Marcet, Albert, and Thomas J. Sargent. "Convergences of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," Manuscript, October 1986.

Marcet, Albert, and Thomas J. Sargent. "Convergence of Least Squares Learning In Environments with Hidden State Variables and Private Information," Manuscript, October 1986.

Muth, John F. "Rational Expectations and the Theory of Price Movements," *Econometrica,* vol. 29, no. 3 (July 1961), pp. 315-55.

McCallum, Bennett T. "Rational Expectations and Macroeconomic Stabilization Policy: An Overview,"Journal of Monty, Credit, and Banking - vol. XII, no. 4, part 2, (November 1980), pp. 697-746.

Samuelson, Paul A. Foundations of Economic Analysis, Harvard University Press, Cambridge, Mass., and London, England (1947).

Simon, HA *Models* of Man, Wiley Press, New York, 1957.

Simon, HA "Rationality as Process and as Product of Thought," American Economic Review, vol. 68, no. 2, (May 1978), pp. 1-16.

Sonnenschien, Hugo. "Do Walras' Identity and Continuity Characterize the Class of Community Excess Demand Functions?" *Journal* of Economic Theoy, vol. 6, no. 4, (August 1973), pp. 345-354.

Swamy, P.A.V.B., R. K. Conway, and P. von zur Muehlen. "The Foundations of Econometrics - Are There Any?" *Econometric* Reviews, vol. 4, no. 1 (1985), pp. 1-61.

Swamy, P.A.V.B., J. R. Barth, and PA Tinsley. "The Rational Expectations Approach to Economic Modelling,"Journal of Economic Dynamics and *Control,* Vol. 4 no. 2, (May 1982), pp. 125-147.

Taylor, John B., "Monetary Policy During a Transistion to Rational Expectations," *Journal* of Policital Economy, vol. 83, no. 5, (October 1975), pp. 1009-1021.

Townsend, Robert M. "Market Anticipations, Rational Expectations, and Bayesian Analysis," International Economic Review, vol. 19, no. 2, (June 1978), pp. 481-494.

Airline Hubs: A Study of Determining Factors and Effects

by Paul W. Bauer

Paul W. Bauer is an economist at the Federal Reserve Bank of Cleveland. The author would like to thank James Keeler, Randall W. Eberts, and Thomas J. Zlatoper and others who provided useful comments on earlier drafts of this paper. Paula Loboda provided valuable research assistance.

Introduction

The Airline Deregulation Act (ADA) of 1978 caused many changes in the industry. For the first time in 40 years, new airlines were permitted to enter the industry, and all airlines could choose the routes they would serve and the fares they would charge. Airlines were also free to exit the industry (go bankrupt), if they made poor choices in these matters. Naturally, this has led to many changes in the way airlines operate.

Many aspects of airline behavior, particularly fares, service quality, and safety, have been subjected to intense study and debate. The development of hub-and-spoke networks is one of the most important innovations in the industry since deregulation, and it has affected all of these aspects. Yet comparatively little research has been done on this phenomena.

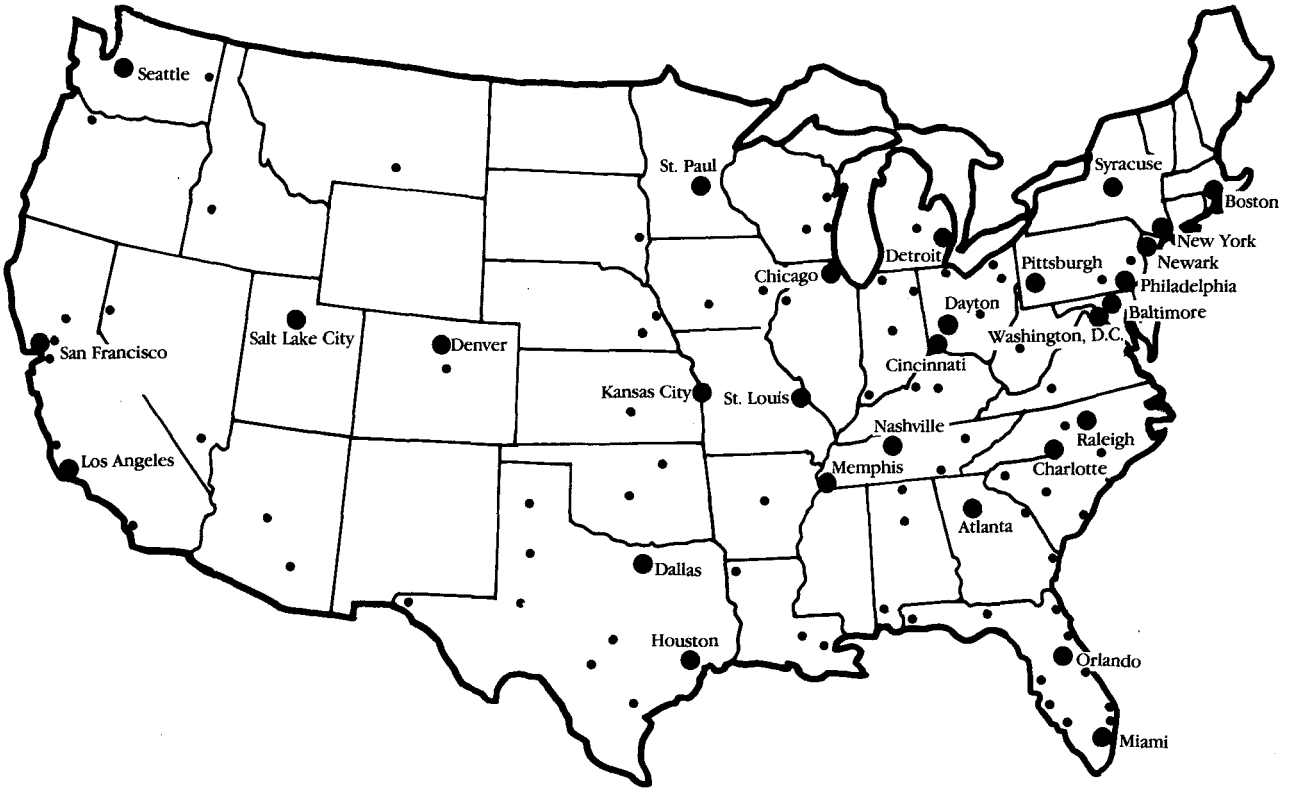
A hub-and-spoke network, as the analogy to a wheel implies, is a route system in which flights from many "spoke" cities fly into a central "hub" city. A key element of this system is that the flights from the spokes all arrive at the hub at about the same time so that passengers can make timely connections to their final destinations. An airline must have access to enough gates and takeoff and landing slots at its hub airports in order to handle the peak level of activity.

An example of a hub-and-spoke network can be seen in figure 1, which shows the location of the hub and spoke cities used in this study. From Pittsburgh, USAir offers service to such cities as Albany, Buffalo, Cleveland, Dallas-Fort Worth, London, New York, Philadelphia, and Syracuse to name just a few. Hub cities tend to have much more traffic than spoke cities. Much of the hub-city traffic centers on making connections. For example, over 60 percent of the passengers who use the Pittsburgh airport hub are making connections, vs. 25 percent at the Cleveland spoke airport.

The advantages of hub-and-spoke networks have been analyzed by several sets of researchers. Bailey, Graham, and Kaplan (1985) discussed the effects of hubbing on airline costs and profitability. Basically, hubbing allows the airlines to fly routes more frequently with larger aircraft at higher load factors, thus reducing costs. Morrison and Winston (1986) looked at the effects of hubbing on passenger welfare, finding that, on average, passengers benefited from the switch to hub-and-spoke networks by receiving more frequent flights with lower fares and slightly shorter travel times.

It is important to note, however, that while passengers benefit on average from hub-and-spoke networks, there are some detrimental effects such as the increased probability of missing connections or losing baggage and having direct service converted into connecting service through a hub (although this is partially offset in many cases by more frequent service). Current public perceptions about the state of airline service have been strongly influenced by the transitory problems many of the carriers have had integrating acquired airlines into their service network.

Hub and Spoke Network



Source: **Author**

FIGURE 1

McShan (1986) and Butler and Huston (1987) have shown another aspect of the switch to hub-and-spoke networks. McShan argues that airlines with access to the limited gate space and takeoff and landing slots at the most desirable hub locations before deregulation have benefited the most from deregulation. Butler and Huston have shown that the airlines are very adept at employing their hub market power, charging lower fares to passengers flying through the hub (who typically have more than one choice as to which hub they pass through) than to passengers flying to the hub (who have fewer options).

Some of these authors have speculated as to why hubs exist in some locations but not in others. Bailey, Graham, and Kaplan (1985) and McShan (1986) have suggested that an ideal hub network would have substantial local traffic at the hub and would be centrally located to allow noncircuitous travel between the airline's hub and spoke cities. However, no empirical exploration of this issue has yet been attempted.

In an attempt to more fully understand the hubbing phenomena, this paper looks for the main factors that airlines consider in evaluating existing and potential hubs, and investigates the impact of the hubbing decision on airport traffic.

The paper is organized as follows. Section I discusses the cost and demand characteristics of the airline industry that lead to hub and spoke networks. From these stylized facts about the airline industry, a two-equation empirical model is constructed in section II. The first equation predicts whether a city is likely to have a hub airline and the second equation estimates the total revenue passenger enplanements the city is likely to generate as a result of the hub activity. Empirical estimates are obtained for this model, using data from a sample of the 115-largest airports in the U.S., and are discussed in section III. The implications of these results on the present and future structure of the U.S. airline industry are discussed in section IV.

I. Characteristics of Airline Demands and Costs

To understand the factors that influence the location of hubs, it is first necessary to look at the demand determinants and costs for providing air service. Basically, people travel for business or pleasure. Travelers usually can pick from several transportation modes. The primary modes of intercity travel in the U.S., are automobiles, airlines, passenger trains, and buses. A traveler's choice of transport is influenced by the distance to be traveled, the relative costs of alternative transportation, and the traveler's income and opportunity cost of time spent traveling.

Aggregating up from individual travelers to the city level, the flow of airline passengers between any two cities is largely explained by the following factors:

1) the air fare between the two cities and the cost of alternative transportation modes,
2) the median income of both cities,
3) the population of both cities,
4) the quality of air service (primarily the number of intermediate stops and the frequency of the flights),
5) the distance between the two cities, and lastly,
6) whether either of the cities is a business or tourist center.

It is important to distinguish between business and tourist travelers. While both generate traffic, business travelers are more time-sensitive and less price-sensitive than tourist travelers. Business travelers would prefer to pay more for a convenient flight, whereas tourists would prefer to pay less, even if it means spending more time en route. These factors influence the demand for air service. The cost of providing that service can now be discussed.

As with any firm, airline costs are determined by how much output is produced and by the price of the inputs required to produce that output. Output in the airline industry is usually measured in revenue passenger miles (rpm), which is defined as one paying passenger flown one mile. Average cost per revenue passenger mile declines as either the average stage length (the average number of miles flown per flight) or the average load factor (the average number of seats sold per flight) increases.

It is easy to see why costs behave in this manner. First, every flight must take off and land. These activities incur high fixed costs. In addition to the usually modest takeoff and landing fees, much more fuel is used up when taking off than at other stages of the flight. Taxiing to and from the runways also takes up a significant amount of time. Those costs are unrelated to the distance of the flight or to the number of passengers. By comparison, flying at the cruising altitude is relatively inexpensive. Thus, with each mile flown the high fixed costs per flight are distributed over more and more miles, which lowers the average cost per revenue passenger mile. Second, average cost per revenue passenger mile declines as the average load factor is increased, because it is cheaper to fly one airplane completely full than it is to fly two planes half full.

Studies have shown that the cost of airline operations do not exhibit increasing returns to scale.[1] In other words, large airlines do not enjoy cost advantages over small airlines if load factors and stage lengths are taken into account. This does not mean that large airlines may not have other advantages over their smaller rivals. One advantage that they may have is that they have more flights to more destinations with more connections, so that they may be able to achieve higher load factors, which reduces cost. Frequent-flyer programs also tend to favor larger airlines, since passengers will always try to use one airline to build up their mileage credits faster. The larger airlines, having more flights and more destinations, are more likely to be able to satisfy this preference.

Under these cost and demand conditions, the chief advantage to establishing a successful hub is the increase in the average load factor, which lowers average cost. Hubbing enables an airline to offer more frequent nonstop flights to more cities from the hub because of the traffic increase from spoke cities. Passengers originating from the hub city thus enjoy a higher level of service quality than would have been possible if spoke travelers were not making connections there. Passengers from the spoke cities may also enjoy better service, because they can now make one-stop flights to many cities that they may have only previously reached by multistop flights.

Hubbing has a significant effect on the demand for air travel through its effects on both air fares and the quality of air service. Passengers prefer nonstop flights to flights with intermediate stops, and if there are intermediate stops, passengers prefer making "online" connections (staying with the same air carrier) to making "interline" connections. Nonstop and online flights minimize flying time and are less stressful and exhausting to passengers. The development of a new hub increases the number of nonstop and one-stop flights in a region, while reducing multistop flights, which were common on some routes prior to deregulation. In general, service

1 See Bauer (1987 working paper) and White (1979).

quality increases for both the hub city and the spoke cities when a hub-and-spoke network is created. However, some of the larger spoke cities could end up worse off, because they may lose some nonstop service to other cities that may now have to be reached by flying through the hub.

Now the problem of how to determine whether a particular city might make a successful hub, and the resulting implications for the volume of air traffic at the airport, can be considered.

II. Empirical Model of the Hubbing Phenomenon

The potential for airlines to serve a number of city pairs and the flow of passengers between those city pairs depends upon the demand and cost factors discussed in the last section. Given these factors, airlines trying to maximize profits face the simultaneous problem of choosing which cities to serve and how to serve them, that is, which cities to make hubs, which cities to make spokes, and which pairs to join with nonstop service. This is a complicated problem since the choice of a hub affects fares and service quality and, hence, passenger flows. Decisions by the airline's competitors will also affect the passenger flows within its system.

To investigate how important each of the various demographic factors discussed below is in deciding whether a given city would make a viable hub, a data-set containing information on 115 cities with the largest airports in the U.S. was compiled. These cities range in size from New York City, to Bangor, Maine and are shown in figure 1 with the hub cities in green and the spoke cities in orange. Notice that most of the hubs are located east of the Mississippi in cities surrounded by a large population base.

The data were collected from several sources. Information on whether a city was considered to have a hub airline (if the i-th city had a hub airline, then $h_i = 1$, otherwise $h_i = 0$) and the total revenue passenger miles handled by the city was obtained from 1985 Department of Transportation statistics. Data on the population (pop), and the per capita income (*inc*) of the city were obtained from the State and Area Data Handbook (1984) and from the Survey of Current Business (April 1986 issue).

In addition, a set of variables was collected to identify whether the city was a business or tourist center. The first of these variables (*DBTP*, "Dummy Business-Tourist-Proxy") is a dummy variable that is set equal to one if the total receipts from hotels, motels, and other lodging places for each city is greater than an arbitrary threshold, and is zero if otherwise. This series was also collected from the State and Area Handbook (1984). A value of one for this variable should correspond to cities that are either a business or tourist center. Unfortunately, this variable only measures the joint effect of both activities and does not distinguish between business and tourist travelers.

To construct separate measures of business and tourist activity, three variables are introduced. The number of Standard and Poors 500 companies headquartered in each city (*corp*) was compiled to be used as a proxy for the business traffic that each city is likely to generate. Measures of the likelihood that a city will generate significant tourist activity are obtained from the Places Rated Almanac published by Rand-McNally. The measures are respectively the rank of the city in recreation (*rec*) and the rank of the city in culture (cult). These variables were transformed so that the higher the rank the higher the city's scores were in that catagory.

In this study, a long-run approach is implicitly taken that ignores individual airport characteristics. In the long run, runways, gates, and even whole airports can be constructed.[2] The decision concerning where to locate hubs in the long run is determined by the location of those cities and by demographic variables that determine the demand for travel between cities. Unfortunately, deriving an economically meaningful measure of location is difficult in this context. Hubs can be set up to serve either a national or regional market, or to serve east-west or north-south routes. Thus, while location is an important factor in determining the location of hubs, constructing an index that measures the desirability of a city's location is beyond the scope of the current study.[3]

A more formal model of the hubbing decision can be constructed as follows. Let the viability of a given airport as a potential hub be a log linear function of the demographic variables discussed above where:

$$(1) \quad h_i^* = a_0 + a_1\, ln(pop_i) + a_2\, ln(inc_i) + a_3\, \text{DBTP}, + a_4\, ln(corp_i) + a_5\, ln(rec_i) + \text{a},\ ln(\text{cult},) + v_i.$$

Here, h_i^* measures the viability of a hub in the i-th city. If this index is above a given threshold (at which point the marginal cost of setting up the hub is equal to the marginal revenue that the

2 For short-run analysis, information on individual airport characteristics is required, This approach will be employed in future research.

3 Future research will attempt to look at this question more directly.

Parameter Estimates from Decision to Hub Equation

Parameter	Estimate	*t*-statistic
Constant	-0.347	-0.627
POP	0.869	1.60
inc	-1.57	-0.795
DBTP	0.478	0.920
corp	0.138	1.29
rec	-0.00232	-0.902
cult	0.0110	1.46

Percentage of predictions correct = **87.0.**
Chi-squared statistic = **69.4**
SOURCE: **Author.**

TABLE 1

hub brings in), then an airline will set up a hub there. Thus, h_i is related to h_i as follows:

$$(2) \quad h_i = \begin{cases} 1, \text{ if } h_i^* \geq k \\ 0, \text{ otherwise,} \end{cases}$$

where k is the threshold between hubs and nonhubs and v_i is statistical noise.

The traffic an airport can be expected to handle will depend on the same demographic variables that also influence whether a city is a hub, and by whether or not the city actually is a hub. Thus, traffic, as measured by revenue passenger miles (rpm), can be modeled as a log linear function of the demographic variables and the hub variable:

$$(3) \quad ln(rpe_i) = b_0 + b_1\ ln(pop_i) + b_2\ ln(inc_i) + b_3\ DBTP_i + b_4\ ln(corp_i) + b_5\ ln(rec_i) + b_6\ ln(cult_i) + b_7 h_i + e_i.$$

where e_i is statistical noise.

Since the model is diagonally recursive (only one of the equations includes both endogenous variables and it is assumed that there are no cross equation correlations), each equation of the model can be estimated separately.[4] The equation predicting the viability of the hub was estimated using the Probit maximum likelihood method. The traffic equation was estimated by ordinary least squares.

Estimates from Revenue Passenger Enplanements Equation

Parameter	Estimate	t-statistic
Constant	16.6	118.0
POP	0.545	5.13
inc	1.15	2.73
DBTP	0.914	5.53
corp	-0.0131	-1.46
rec	0.00101	1.71
cult	0.00107	0.922
hub	0.795	4.98

R-squared = **0.850.**
Fstatistic = 86.3.
SOURCE: Author.

TABLE 2

III. Results

Results from estimating the above model are presented in tables 1 and 2. Table 1 presents the parameter estimates from the equation that predicts the viability of a hub in any given city. The overall prediction power of the model is quite good. The point estimates of the parameters all have the expected signs except for the coefficient on per-capita income, though the level of statistical significance is very weak. The high correlation among most of the demographic variables suggests that multicollinearity is a problem and that the standard errors are inflated leading to lower *t*-statistics. Even with this problem, estimates from this equation do correctly predict whether or not a city will be a hub 87 percent of the time.

A city is more likely to become a hub as its population, lodging receipts (DBTP), or number of S&P 500 corporations increase, or as its ranking for recreation or culture improves. Business travelers (being more time-sensitive and less price-sensitive) should be more important to an airline than tourist travelers in the location of hubs, so that the number of S&P 500 corporations should be more important than either recreation or culture. One-tailed tests conducted at the 90 percent confidence level indicate that increasing a city's population and number of S&P 500 corporations, and improving the cultural ranking, all have nonnegative effects on the viability of a hub for a given city, other things being equal. It would have been reasonable to expect that increases in per-capita income would also increase the viability of the hub, but higher per-capita incomes reduce the likelihood of a city being a hub, although this result is not statistically significant.

The results from the estimation of the traffic equation are presented in table 2. Most of the parameter estimates are statistically significant in this equation. All the estimates have the expected sign, except the coefficient on the number of S&P 500 corporations, although it is not statistically significant.

Given the construction of the model, some of these parameters can be interpreted as elasticities. For example, a one percent

4 The results reported here are not sensitive to the assumption of no cross equation correlations.

Outlier Cities

Likely, but do not have a hub	Unlikely, but do have a hub
Cleveland	Raleigh
San Diego	Syracuse
New Orleans	Orlando
Phoenix	Nashville
Tampa	Kansas City

SOURCE: Author.

TABLE 3

increase in a city's population would lead to a 0.55 percent increase in revenue passenger enplanements, while a one percent increase in a city's per capita income would lead to a 1.15 percent increase in revenue passenger enplanements. The coefficient of lodging receipts (DBTP) can be interpreted as follows. From these estimates, it can be calculated that cities classified as business/tourist centers have roughly *2.49* times the traffic that other cities have.

The coefficient for the hub variable has a similar interpretation, given its construction. If two cities are identical, except that one has a hub and the other does not, then the city with the hub can be expected to have over 2.19 times more revenue passenger enplanements than the other city. For example, Cleveland and Pittsburgh have very similar demographic characteristics, yet as a result of USAir's hub, Pittsburgh has about *2.3* times the revenue passenger enplanements that Cleveland has. It was noted earlier that passengers making connections in Pittsburgh account for most of this difference because only 25 percent of the passengers who use Cleveland's airport are there making connections, whereas over 60 percent of the passengers at Pittsburgh's airport are there making connections. Clearly, the creation of a hub greatly increases the activity occurring at an airport.

Table *3* presents two lists of outliers as a by-product of the estimation process. The first list is of cities that the model predicts should be hubs, but are not. The second list is of cities that the model predicts should not be hubs, but are. It is likely that San Diego, Phoenix, and Tampa would not be outliers if a location variable were included in the model, since these cities lie in the southwest and southeast corners of the country (see figure 1). Cleveland and New Orleans, on the other hand, appear to be more likely candidates for future hubs. Other midwest cities to watch are Indianapolis and Columbus.

Two factors can explain why most cities made the second list: location and measurement problems with the hub variable. Although it is hard to develop an index for location, it is easy to get an intuitive feel for it. Both Kansas City and Nashville are situated near the center of the country, giving them an advantage over Phoenix or San Diego in the competition for hubs. The second factor involves the problem of deciding what constitutes hub service at a city. Clearly the activity going on in Chicago by both United Airlines and American Airlines is quantitatively different from what USAir is doing in Syracuse, yet in this study both cities are counted as hubs.

IV. Summary and Implications for the Future

This paper has explored the characteristics that influence hub location and the effect on airport traffic as a result of hub activity. The results indicate that population is the most important factor determining hub location. An increase in per-capita income leads to a larger proportional increase in revenue passenger enplanements, whereas an increase in population leads to a less than proportional increase. One of the most interesting findings was that the creation of a hub at a city leads to a more than doubling of revenue passenger enplanements generated at that city.

The framework developed here is implicitly long run: airlines, passengers, and airports are assumed to have fully adjusted to the new deregulated environment. Given the recent merger wave in the industry, this does not appear to be the case, and many changes are likely in the coming years. More cities will probably become hubs, as traffic cannot increase much further at some large airports that have almost reached their capacity limits using current technology.

The only question is where to hub, not whether to hub. As the airline industry evolves, it will be interesting to track what happens to the air service provided to the communities listed in table *3*. Given the expected growth in future air travel, cities on the first list are more likely to receive hub service than cities on the second list are to lose hub service.

REFERENCES

Bailey, Elizabeth E., David R. Graham, and David P. Kaplan. *Deregulating the Airlines.* Cambridge, *MA:* The MIT Press, 1985.

Bauer, Paul W. "An Analysis of Multiproduct Technology and Efficiency Using the Joint Cost Function and Panel Data: An Application to the U.S. Airline Industry." Ph.D. Dissertation, University of North Carolina at Chapel Hill, 1985.

Butler, Richard V., and John H. Huston. "Actual Competition, Potential Competition, and the Impact of Airline Mergers on Fares." Paper presented at the Western Economic Association meetings, Vancouver, B.C., July 1987.

McShan, William Scott. "An Economic Analysis of the Hub-and-Spoke Routing Strategy in the Airline Industry," Northwestern University, Ph.D., 1986.

Morrison, Steven, and Clifford Winston. *The Economic Effects of Airline Deregulation.* Washington, D.C.: Brookings Institution, 1986.

White, Lawrence, J. "Economies of Scale and the Question of "National Monopoly" in the Airline Industry," *Journal of Air Law and Commerce,* vol. *44,* no. **3,** (1979) pp. 545-573.

A Comparison of Risk-Based Capital and Risk-Based Deposit Insurance

by Robert B. Avery
and Terrence M. Belton

Robert B. Avery is a senior economist in the Division of Research and Statistics at the Board of Governors of the Federal Reserve System. Terrence M. Belton is a senior economist at the Federal Home Loan Mortgage Corporation.

The authors would like to thank Randall W. Eberts. Edward Ettin, Gerald Hanweck, Myron Kwast, James Thomson, and Walker Todd for helpful comments and suggestions.

An earlier version of this paper was presented at the Federal Reserve Bank of Cleveland's fall seminar on the role of regulation in creating1 solving problems of risk in financial markets — November 3, 1986.

Introduction

The perception of increased bank risk-taking has raised concerns as to whether changes and improvements are needed in our system of regulatory supervision and examination. These concerns clearly underlie recent proposals for risk-based capital standards issued by all three bank regulatory agencies—the Federal Reserve Board, the Federal Deposit Insurance Corporation (FDIC), and the Comptroller of the Currency—as well as proposals by the FDIC and Federal Savings and Loan Insurance Corporation (FSLIC) for risk-based deposit insurance premiums. None of these approaches has, as yet, been implemented, and each is still under active consideration by at least one regulatory body.

As part of an ongoing evaluation of the potential effectiveness of various methods of controlling bank risk-taking, this paper presents a comparison of risk-based capital and risk-based deposit insurance premium proposals. Although these proposals may appear to represent quite different methods of controlling bank risk, the results presented below suggest that this need not be the case and that, if implemented properly, the two methods can produce a similar level of bank risk-taking.

The paper also suggests that differences that exist between the two methods lie not in the fact that one controls premiums and the other capital levels, but that one prices risk and the other sets a risk standard. This is discussed informally in section I, while evidence of how both a risk-based insurance and risk-based capital system could be implemented using similar measures of risk is presented in the section that follows.

I. Discussion

In the current regulatory environment, commercial banks are subject to a fixed minimum level of primary capital per-dollar of assets and a fixed deposit insurance premium per-dollar of domestic deposits regardless of the risk that they present to the FDIC. As many critics have pointed out, this presents a potential problem of incentives in that banks may not bear the full social costs of increased risk-taking. Both a risk-based capital and risk-based insurance system are designed to address this problem by inducing banks to internalize the expected costs that their risk-taking imposes on the FDIC and society in general.' The programs appear to differ significantly, however, in how they attempt to achieve this goal.

As proposed, a risk-based deposit insurance system would explicitly price risk-taking behavior on the part of insured banks. Periodically, the FDIC would assess the risk represented by each bank and charge an insurance premium reflecting the expected social

1 Another objective may be to distribute the costs of risk-taking more equitably across banks even if such differences stem from exogenous factors and if issues of moral hazard and allocative efficiency are irrelevant.

Risk Variables

Symbol	Definition
KTA	percent ratio of primary capital to total assets,
PD90MA	percent ratio of loans more than 90 days past due to total assets,
LNNACCA	percent ratio of nonaccruing loans to total assets,
RENEGA	percent ratio of renegotiated loans to total assets,
NCOFSA	percent ratio of net loan charge-offs (annualized) to total assets,
NETINCA	percent ratio of net income (annualized) to total assets.

Source: Board of Governors of the Federal Reserve System.

TABLE 1

costs attributable to it.[2] Because banks would in principle bear the full expected cost of their actions, they would either be deterred from excessive risk-taking or would pay the full expected costs to the FDIC.

A risk-based capital standard works by setting a standard that, by absorbing losses, limits the amount of risk an insured bank can impose on the FDIC, rather than by explicitly pricing risk. If the regulators determine that a bank represents a risk above the allowable standard at its current level of capital, they would require the bank to raise more capital. By adjusting capital "buffers," regulators can control the size of potential losses irrespective of bank behavior.

The regulator uses information on differences in risk-taking behavior across banks to require different amounts of capital or co-insurance, not to charge different premiums. Indeed, since adjustment of the capital buffer is used to reduce the risk represented by each bank to the same level, it is then appropriate that they be charged a flat premium rate.[3] Bank risk-taking behavior may be deterred because banks would recognize that they will incur higher expected capital costs, an implicit price, even though banks do not face explicit prices for risk. In both schemes, overall system risk-taking would be reduced because banks would take full account of the expected consequences of their actions, either through explicit insurance premiums or implicit prices via higher capital costs.

Current Proposals on Risk-Based Deposit Insurance and Risk-Based Capital

In recent years, there have been several specific proposals made by the federal regulatory agencies for basing insurance premiums or capital requirements on the perceived risk of depository institutions. In 1986, for example, the FDIC asked for legislation authorizing the adoption of a risk-based deposit insurance system and has developed a specific proposal for implementing such a system. More recently, the Federal Reserve Board, in conjunction with the Bank of England and with other U.S. banking regulatory authorities has published for public comment a proposal for risk-based capital requirements.

The FDIC proposal for risk-based deposit insurance utilizes two measures for assessing bank risk-taking.* The first measure is based on examiner-determined CAMEL ratings for individual commercial banks. CAMEL ratings, which range from 1 through 5 (with 5 representing the least healthy bank) are intended to measure the bank's capital adequacy (*C*), asset quality (A), management skills (*M*), earnings (*E*), and liquidity (*L*). The FDIC's problem-bank list consists of all banks with CAMEL ratings of 4 and 5.

The second measure of bank risk employed in the FDIC proposal is a risk index developed by the FDIC that is based on publicly available Call Report data. The index is defined as:

$$(1) \quad I = .818 - .151KTA + .211PD90MA + .265LNNACCA + .177RENEGA + .151NCOFSA - .347NETINCA,$$

where all variables are defined in table 1. The weights in the index were estimated from historical data with a probit model that predicts whether or not an individual bank is on the FDIC's problem-bank list. The index can be interpreted as providing a measure of the likelihood that a bank is a problem bank. Banks with higher index values of the index are more likely to be problem institutions and therefore more likely to impose higher expected costs on the FDIC.

Premiums would be assessed, under the FDIC proposal, by defining two premium classes. Banks having a positive value of the risk index and a CAMEL rating of 3, 4, or 5, would be classified as above-normal risk. These

2 If the FDIC cannot fully assess the ex-ante risk represented by each bank, perhaps because monitoring costs would be excessive, then the "optimal" risk premium would also include "penalties" over and above the FDIC's estimate of each bank's expected social cost.

3 Assuming the risk-based capital requirement is binding so that no institution holds capital in excess of its requirement.

4 The proposal is described in "Risk-Related Program," FDIC Discussion Paper, September 20, 1985, and Hirschhorn, E., "Developing a Proposal for Risk-Related Deposit Insurance," *Banking and Economic Review*, FDIC, September/October 1986.

Summary of Risk Weights and Major Risk Categories for State Member Banks and Bank Holding Companies

Category **A1**	(0 percent weight) Cash—domestic and foreign Claims on Federal Reserve Banks
Category **A2**	(10 percent weight) Short-term (one year or less) claims on U.S. Government and its Agencies.
Category **A3**	(25 percent weight) Cash items in process of collection. Short-term claims on domestic depository institutions and foreign banks, including foreign central banks. Claims (including repurchase agreements) collateralized by cash or U.S. Government or Agency debt. Claims guaranteed by the U.S. Government or its Agencies. Local currency claims on foreign central governments to the extent that bank has local currency liabilities. Federal Reserve Bank stock.
Category ***A4***	(50 percent weight) Claims on U.S. Government-sponsored Agencies. Claims (including repurchase agreements) collateralized by U.S. Government-sponsored Agency debt. General obligation claims on states, counties and municipalities. Claims on multinational development institutions in which the U.S. is a shareholder or contributing member.
Category **A5**	(100 percent weight) All other assets not specified above, including: Claims on private entities and individuals. Long-term claims on domestic and foreign banks. All other claims on foreign governments and private obligators.

Source: Board of Governors of the Federal Reserve System.

TABLE 2

institutions would be charged an annual premium equal to one-sixth of one percent of domestic deposits, or twice the current premium level. All other institutions (that is, institutions having either a negative value for the risk index or a CAMEL rating of 1 or 2) would be classified as normal-risk banks and be charged the current premium of one-twelfth of one percent.

The risk-based capital requirement proposed by the Federal Reserve Board, in conjunction with other regulatory authorities, measures bank risk-taking in a somewhat different fashion than the FDIC's deposit insurance proposal. Capital requirements would be assessed, under the Board's proposal, as a fraction of the on- and off-balance-sheet activity of individual commercial banks.[5] Specifically, the proposal defines five asset categories that are shown in table 2. These categories are intended to measure, in broad terms, assets having varying degrees of credit risk. Cash and claims in Federal Reserve Banks (category A1) are deemed to have no credit risk and require no capital support. Commercial loans to customers other than banks, (Category A5) are deemed to have the greatest amount of credit risk. The minimum primary capital level, K, required under the proposal would be defined as:

$$(2) \qquad K = a(0A1 + .10A2 + .25A3 + .5A4 + 1A5),$$

where a denotes the minimum required ratio (not yet specified in the proposal) and A1 to A5 denote the asset categories defined in table **2.**

The requirement shown in equation (2) effectively imposes different minimum capital standards on each of the five asset categories. If a is set at 7 percent, for example, all

5 The proposal is described in two press releases of the Board of Governors of the Federal Reserve System titled "Capital Maintenance: Revision to Capital Adequacy Guidelines," dated February 12, 1987 and March 18, 1987.

commercial loans, except those to other banks (category A5), would effectively have minimum required capital ratios equal to 7 percent; claims on U.S. government-sponsored agencies (category A3) would have required capital ratios equal to 1.75 percent; and short-term treasury securities (category A2) would have required capital ratios of 0.7 percent.[6]

It is clear that a major difference between the risk-based capital and risk-based deposit insurance proposals just described is the type of information that is used to assess bank risk-taking. The risk-based deposit insurance proposal focuses on measures of bank performance, such as earnings and asset quality; the risk-based capital proposal focuses on the types of activities in which banks are involved. The former view is based on statistical evidence that suggests these performance measures provide the best forecast of future bank problems? The latter approach to measuring bank risk-taking is based on the view that certain activities are inherently more risky than other activities and that these more risky activities should be capitalized at higher levels.

In contrasting the two approaches to measuring bank risk, it should be emphasized that the different measures used do not represent an inherent difference between risk-based capital and risk-based insurance. Indeed, both systems could, in principle, use identical information in assessing the risk of individual banks. The difference between the two systems lies not in what information the regulator collects, nor in how it uses that information to assess bank risk; rather, the difference results primarily because one system controls risk by a ***standard*** and the other by *explicit prices*. In the next subsection, we describe how these differences affect both banks and bank regulators.

Differences Between Risked-based Capital and Risk-based Deposit Insurance

Because one system is based on a minimum standard and the other on a price, a number of differences are likely to exist between risk-based capital and risk-based insurance. One difference is that enforcement of a risk-based capital system is likely to offer the ***regulator more*** *flexibility* and potential for discretion than a risk-based premium system. If an annual insurance assessment appeared on a bank's income statement, and therefore was public, it would be difficult to waive or adjust the fee without alerting competing banks, financial market participants, and the public. Moreover, enforcement would likely be very mechanical. Banks would be assessed a fee, and examiners would have to deal individually only with those banks that could not or would not pay.

However, enforcement of a risk-based capital standard is likely to be of a very different nature. Enforcement might focus only on those firms close to or under the standard, and would likely entail more individual examiner input. Moreover, the judgement of whether or not a bank with a continually changing balance sheet meets the standard—and if not, how long it has to comply—is likely to offer considerable potential for discretion. Thus, in a regulatory environment based on judgement and discretionary supervision and regulation, a risk-based capital standard might be more attractive.

Another difference is that because a risk-based premium system prices risk rather than limiting it by forced capital adjustments, it is likely to offer *banks* ***a*** *more flexible,* and therefore potentially more efficient, means of response. Under a risk-based capital system, a risky bank facing abnormally high capital costs does not have the option of paying the FDIC for the right to take excessive portfolio risk even though this may be its most cost-effective response.[8] This feature is likely to favor a risk-based premium approach under virtually all regulatory environments. It might be argued that banks should not be allowed too much freedom as they may not properly respond to prices. However, this could be accommodated in a risk-based premium system by shutting down banks with excessive risk-taking or by altering their behavior by other supervisory means.

The two proposals are also likely to have significant differences in the amount of information that they reveal to the public. At

6 In addition to imposing capital requirements on various balance-sheet asset categories, the proposal also addresses the risk from off-balance-sheet activities. Capital requirements for those activities are determined by first converting the face-amount of off-balance-sheet items to a balance-sheet equivalent. This is done by multiplying the face amount of the off-balance-sheet contract by an appropriate credit conversion factor. The resulting balance-sheet equivalent is then assigned to one of the five risk categories depending on the identity of the obligator and, in certain cases, on the maturity of the instrument.

7 In addition to the empirical work on predicting problem banks, the literature also suggests that earnings, capital and asset quality measures are important predictors of future bank failure. See J. Bovenzi, J. Marino, and F. McFadden, "Commercial Bank Failure Prediction Models," in Economic Review, Federal Reserve Bank of Atlanta (November 1983) and Robert B. Avery, Gerald A. Hanweck and Myron L. Kwast, "An Analysis of Risk-Based Deposit Insurance for Commercial Banks," *Preceedings* of a Conference on Bank Structure and Competition *(1985)*, Federal Reserve Bank of Chicago.

8 Technically, raising capital is not the only adjustment available to the bank as it can adjust any factor used in the regulator's assessment of risk. Thus, the relevant price banks face is the price of the minimum-cost method of meeting the standard. If this price is not equal to the regulator's price, there will be an inefficiency.

most, a risk-based capital standard would reveal only whether or not a bank met the standard. One could not even infer that a bank adding capital was doing so because it had become excessively risky; the extra capital might be needed because of anticipated expansion, etc. However, it would be very difficult to keep a bank's insurance premium confidential. Low-risk banks would have an incentive to advertise this fact and investors would have incentives to identify high-risk banks. This might cause particular problems in the use of confidential data to calculate premiums. Knowledge of a bank's premium could be used to draw strong inferences about values of any confidential inputs used. To the extent that this would deter the use of confidential data in a risk-based premium system, it might mean that risk assessment with a risk-based capital system would be more accurate and therefore fairer.

Moreover, even if confidential data were not used, public disclosure of a bank's premium might create the possibility of bank runs. The official declaration of the FDIC that a bank was risky, even if based on a mechanical calculation from publicly available balance sheet data, might be sufficient to induce significant withdrawals.

Yet another difference between the two methods is likely to occur in the regulatory response lag. Because it is based on a standard, a risk-based capital system may have a built-in response lag that is not present with a risk-based premium system. Under a risk-based premium system, a bank could be required to compensate the FDIC immediately for its risk exposure. In contrast, particularly if it entails raising new capital, adherence to a capital standard would likely entail some lag, thereby delaying the ability of the insurer to control its risk exposure.

Finally, even if the FDIC's assessment rate were adjusted so that it bore equivalent actuarial risk, there may be some differences in the number of bank failures under the two systems. Either system should reduce the number of bank failures from current levels because of the reduced risk-taking that should result when banks are required to bear the full costs of their risk-taking? The magnitude of this reduction, however, may differ for the two systems. As noted earlier, risk-based deposit insurance systems allow banks the flexibility of holding capital levels below those required under a comparable risk-based capital system and of offsetting the higher risk by paying larger insurance premiums. For those banks that opt to hold capital levels below those required under a capital standard and pay correspondingly larger insurance premiums, the incidence of failure would be higher under a risk-based insurance system than that observed under a risk-based capital standard.

By the same token, a risk-based insurance system would provide other banks the flexibility of holding capital levels well above those required under a risk-based capital standard and of being compensated for this increased capital by paying lower insurance premiums. For such banks, the incidence of failure will be lower under a risk-based insurance system than under a capital standard. This difference between the two systems stems from the fact that a capital standard does not reward banks for having capital greater than the minimum standard; a risk-based insurance system provides such a reward in the form of a reduced premium.

The foregoing analysis suggests that, in the aggregate, it is unclear which of the two systems would reduce bank failures by the greatest amount. Prediction of whether an individual bank's capital would be greater under a risk-based capital standard than under a risk-based premium system depends on the cost of capital faced by the bank and upon the degree to which the risk-based insurance system penalizes banks for reductions in their capital. When the cost of raising capital in the private market (or other adjustment methods) is high relative to the penalty rate charged by the deposit insurer for reductions in capital, banks will be more likely to choose lower capital levels under a risk-based insurance scheme than that required under a risk-based capital standard. Conversely, when the insurance system assigns a relatively steep penalty rate for reductions in bank capital, individual banks would be more likely to hold larger amounts of capital under a risk-based insurance system, implying a lower incidence of bank failure.

Despite these differences, if based on the same method of assessing bank risk, proposals for risk-based capital and risk-based insurance should have a similar impact on bank risk-taking. To provide a glimpse as to how such proposals might work, a practical system of risk-based deposit insurance and risk-based capital is developed and presented in the next section. Both proposals are based on the same method of

9 Some critics have charged that a risk-based capital or deposit insurance system might actually increase failures and incentives for risk-taking because regulators would measure risk poorly or misprice it. While this may be true, it should be pointed out that the current system assumes all banks represent the same risk. The relevant question, therefore, is not whether regulators would do a perfect job, but whether they could differentiate among banks at all.

Sample Variable Statistics

Variable	Means of Failed Banks	Means of Nonfailed Banks
KTA	6.14	9.26
PD90MA	3.41	0.77
LNNACCA	3.64	0.57
RENEGA	0.28	0.07
NCOFSA	2.89	0.43
NETINCA	-2.94	0.90

Source: Board of Governors of the Federal Reserve System.

TABLE 3

assessing bank risk. As this represents only part of an on-going effort to develop such systems, we only briefly summarize our work.[10]

II. A Model of Bank Risk

Both the risk-based capital and risk-based insurance premium proposals require an accurate method of assessing bank risk. Forming an index or rank ordering of banks by risk entails two steps. First, variables must be selected that are good predictors of risk; and second, weights must be calculated to transform values of the vector of predictor variables into a single-valued index.

Development of a good index is a substantial task and is well beyond the scope of this paper. It was decided somewhat arbitrarily, therefore, to use the same six predictor variables used by the FDIC in its risk-based insurance proposal (see table 1). One good method of forming weights for the index is to use historical data to "fit" values of the predictor variables to an observable ex-post measure of loss. Candidates for ex-post measures of bank performance might be bank failure and FDIC losses when failure occurs, or bank earnings or loan charge-offs. Although we use other measures of bank performance in other work, for the illustrative proposals developed for this paper it was decided to utilize bank failure. The basic strategy followed was to use historical data on bank failure to estimate weights that could be used to transform values of the six variables listed in table 1 into an index of risk. This index forms the basis of both our risk-based capital and risk-based deposit insurance proposals.

In selecting data used in this study for both estimation and model evaluations, the following specific procedures were used. The sample was restricted to insured commercial banks headquartered in the United States. Mutual savings banks were excluded. Microdata were collected for each bank for each of the five semiannual call and income reports filed from December 1982 through December 1984.[11]

Each of the "calls" represented a potential observation with the following adjustments (thus each bank could appear in the sample five times). Because new banks are thought to follow a different behavioral process, all calls were eliminated whenever a bank had not been in continuous existence for three years at that point. Banks without assets, deposits, or loans were also eliminated. The sample was further reduced by eliminating all banks with assets above $1 billion (approximately two percent of all banks) because of the virtual absence of large bank failures.[12] These adjustments reduced the banks available in December 1984, for example, from 14,460 to 13,388. The actual estimation sample was further reduced by only using 10 percent (randomly selected) of the calls reported by banks that did not fail within a year of the call.

This stratification of the nonfailed banks (which was corrected for in the estimation procedure) was done to create an estimation data-set of manageable size. All calls where the bank failed within a year of the call were used (thus a failed bank could contribute two calls to the sample). The final estimation sample consisted of 6,869 observations, 160 of which represented calls for banks that failed within six months of the call and 138 for banks that failed between six months and a year after the call.

The data used for the study were taken directly from the bank's filed call report, with slight adjustment. June values for the two income variables—charge-offs and net income—were recalculated to reflect performance over the previous year rather than the 6-month period reported. Means of the variables for the estimation data are given in table 3. The data were fit using a logistic model to predict bank failure

10 See Robert B. Avery and Gerald A. Hanweck, "A Dynamic Analysis of Bank Failures," Proceedings of a Conference on Bank Structure and Competition *(1984)*, Federal Reserve Bank of Chicago; Robert A. Avery, Gerald A. Hanweck and Myron L. Kwast, "An Analysis of Risk-Based Deposit Insurance for Commercial Banks," Proceedings of a Conference on Bank Structure and Competition *(1985)*, Federal Reserve Bank of Chicago; and Terrence M. Belton, "Risk-Based Capital Standards for Commercial Banks," presented at the Federal Reserve System Conference on Banking and Financial Structure, New Orleans, Louisiana, September 1420, 1985.

11 More time periods could have been used. However, it was decided to limit the length of the estimation period so that an "out of sample" measure of the model's performance could be computed.

12 The elimination of large banks had virtually no effect on the results,

where a bank was deemed to have failed if it failed within a year following the call. The estimated risk index is:

$$(3) \quad R = -2.42 - .501\,KTA + .428\,PD90MA + .314\,LNNACCA + .269\,RENEGA + .223\,NCOFSA - .331\,NETINCA,$$

(3.07) (4.89) (5.16) (4.31) (1.07) (1.60) (2.68)

where the logistic form of the model implies that the probability that a bank will fail within a year is,

$$(3a) \quad \text{PROB} = \frac{1}{1 - exp\,(-R)}$$

T-statistics for the estimated coefficients are given in parenthesis under each weight.[13] All weights are statistically significant except those for NCOFSA (which has a perverse sign) and RENEGA.[14]

Although the overall fit of the model suggests that predicting bank failure is difficult, the failed banks in the sample had an average predicted probability of failure of 0.24, a number 69 times larger than the average predicted failure probability of nonfailed banks in the sample. Hence, the model clearly does have some ability to discriminate between high- and low-risk banks.

III. Risk-Based Deposit Insurance Premiums

Several somewhat arbitrary assumptions were used to convert the estimated risk-assessment model into a risk-based deposit insurance premium system. First, the FDIC's expected cost of insuring each bank (per-dollar of deposits) was computed as the estimated probability of failure (from the formula in [3]) times the average FDIC loss when failure occurs (13.6 cents/ per dollar).[15] Assessment of this premium, which averaged 7.2 basis points per dollar of deposits in December 1985, would be actuarially fair if there were no monitoring or social costs. Since these factors are not known, and to provide comparability with the current system, an intercept (or flat premium) of 1.1 basis points per dollar of deposits was added to the risk-based assessment so that the total assessment would be equivalent to the FDIC's actual revenues as of December 1985 (with the current flat-rate assessment of 8.3 basis points). While certainly not a necessary ingredient of a risk-based system, the FDIC revenue constraint was adopted in order to allow the concentration of effort and discussion on estimating the risk-based component of the premium while not having to address the issue of what the appropriate level of gross revenues should be. Finally premiums were "capped at 100 basis points because of the belief that premiums above this level would be difficult to collect.

Estimates of December 1985 risk-based premiums under this system are presented in table 4. Premiums are computed across seven asset-size classes of banks (rows [1] through [7]) and six premium-size intervals (columns [1] through [6]). It should be emphasized that while premiums for banks with over $1 billion in assets are computed and reported, these are extrapolations as no banks of this size were included in the sample used to estimate the risk index. Rows (8) and (9) show the premium distribution for banks that subsequently failed in 1986 and 1987 (through September 30), giving an idea of the system's capacity to identify and penalize risky banks. Row (10) and column (7) present totals for all banks. The first number in each cell is the average risk-based premium expressed in basis points of total domestic deposits. The second number is the average estimated (percentage) probability of failure by banks in that cell, and the third figure is the number of banks, based on the total of 13,522 banks used to compute the premium, that are predicted to fall into each size and risk-class category.

The primaty conclusion to be drawn from table 4 is that the risk-based system depicted there would divide banks into three major groups. First, even with the FDIC revenue constraint imposed, the vast majority of banks

13 Coefficients for a logistic model have a less straightforward interpretation that those in regression models. When multiplied by PROB (1-PROB) each coefficient represents the expected change in the probability of failure resulting from a one-unit change in the variable. Thus, if a bank with a probability of failure of 0.1 raised its capital ratio one percentage point, the model implies that its probability of failure would fall by .045, that is, (-.501 x .1 x .9). Although they were estimated using the same variables, and with data drawn from similar time periods, the coefficients in (3) differ somewhat from those in (1). This occurs, in part, because the FDIC model was estimated using a probit rather than logistic specification, which effects the scaling of the variables (logistic coefficients should be approximately 1.8 times as large). It also stems from the fact that the FDIC used problem-bank status rather than bank failure as a dependent variable.

14 The model's log-likelihood R squared, a concept similar to the R squared in a regression model, is 0.22. The sign on the weight of NCOFSA may be not be as perverse as it appears. The coefficient on charge-offs represents the marginal impact on failure holding net income constant. Because charge-offs are also in net income, they are effectively counted twice. The positive sign on charge-offs indicates they have less impact on failure than other contributory factors toward earnings. The total impact of charge-offs (the sum of the coefficients of NCOFSA and NETINC) has the expected negative sign.

15 This number is the average ratio of the FDIC's loss reserve to total domestic deposits calculated for banks that failed between 1981 and 1984. See Avery, Hanweck, and Kwast "An Analysis of Risk-based Deposit Insurance."

Estimated Commercial Bank Risk-based Premiums — December 1985

(Basis Points of Total Domestic Deposits)

First number is the average premium for banks in the cell. Second number is average estimated probability of failure in percent. Third number is number of banks.

Asset Size Class ($ millions)	Premium Size Class						
	(1) < 8.3	(2) 8.3-12.4	(3) 12.5-24	(4) 25-49	(5) 50-99	(6) 100	(7) All Banks
(1) < $10	2.4	10.1	17.2	32.1	61.6	100.0	6.3
	.1	.6	1.2	2.3	4.5	34.5	1.1
	933.0	29.0	23.0	16.0	9.0	25.0	1035.0
(2) $10 — $25	2.6	10.0	17.2	33.3	68.8	100.0	6.9
	.1	.7	1.2	2.4	5.0	42.7	1.2
	3135.0	109.0	131.0	61.0	44.0	78.0	3558.0
(3) $25 — $50	2.9	10.1	17.1	35.0	70.4	100.0	5.9
	.1	.7	1.2	2.5	5.1	33.6	.7
	3258.0	112.0	105.0	47.0	26.0	54.0	3602.0
(4) $50 — $100	3.1	9.9	16.8	33.9	74.3	100.0	5.9
	.2	.7	1.2	2.4	5.4	35.6	.7
	2485.0	116.0	72.0	29.0	19.0	36.0	2757.0
(5) $100 — $500	3.7	9.8	16.4	32.9	71.7	100.0	5.7
	.2	.6	1.1	2.3	5.2	71.1	.5
	1859.0	85.0	65.0	28.0	7.0	16.0	2060.0
(6) $500 — $1000	4.3	9.3	17.3	29.4	69.7	100.0	7.5
	.2	.6	1.2	2.1	5.0	54.8	.9
	171.0	14.0	9.0	3.0	3.0	2.0	202.0
(7) > $1000	5.1	9.8	15.9	37.7	78.8	0.0	7.0
	.3	.6	1.1	2.7	5.7	0.0	.4
	230.0	60.0	15.0	2.0	1.0	0.0	308.0
(8) Banks failing in 1986	4.8	10.8	17.1	38.1	71.5	100.0	68.7
	.3	.7	1.2	2.7	5.2	51.8	30.1.
	17.0	8.0	9.0	12.0	12.0	75.0	133.0
(9) Banks failing in 1987	4.6	10.2	16.9	32.2	69.8	100.0	37.3
	.3	.7	1.2	2.3	5.1	35.6	9.3
	44.0	11.0	20.0	17.0	9.0	31.0	132.0
(10) All Banks	3.0	9.9	16.9	33.6	69.8	100.0	6.2
	.1	.7	1.2	2.4	5.0	37.4	.8
	12071.0	525.0	420.0	186.0	109.0	211.0	13522.0

Source: Board of Governors of the Federal Reserve System.

TABLE 4

would pay a lower insurance premium under the estimated risk-based scheme than the current gross premium of 8.3 basis points. As may be seen from the table, this is true for all size classes, with the proportion paying less ranging from a low of 75 percent to 90 percent. Overall, 89 percent of all institutions are estimated to pay less with an average premium of 3.0 basis points.

The second group of banks is composed of the 9 percent of all banks that would pay an increased premium ranging from a low of 8.3 basis points to 99 basis points (columns 2 through 5). This range of almost 92 basis points is quite large and appears wide enough both to provide a strong incentive to alter current risk-taking behavior by banks and to deter excessive risk-taking in the future. Some perspective on the size of the estimated risk-based premium is given by noting that the average bank's return on total deposits in 1985 was only 82 basis points. The average bank's premium would have been almost 1 percent of its previous year's total capital, and somewhat over 4 percent of its net income. But in the higher risk categories (columns 4-6), the capital percentages range up to 25.5 percent.

The third group of banks is the one percent that would have been asked to pay an insurance premium of over one percent (capped at 100 basis points) of total domestic deposits in 1985 (column 6 of table 4). For these banks it is not unusual for the average expected cost imposed on the FDIC to exceed 500 basis points. Indeed, the total cost that would have been expected to be imposed on the FDIC in 1986 by the 211 banks in column 6 was $477 million, or 25 percent of the total expected cost of $1.9 billion for all 13,522 commercial banks for which premiums were computed. Clearly, because the size of the assessment might be sufficient, by itself, to force these banks into insolvency, special measures might be needed to deal with them.

The ability of the system to identify risky banks in advance is illustrated by the premiums that would have been charged in December 1985 to banks that subsequently have failed. Over 87 percent of the banks that failed in 1986 would have been required to pay higher premiums than they pay currently, a figure in sharp contrast to the overall figure of 11 percent. Over one-half of the 1986 failed banks would have been assessed premiums at the highest rate of 100 basis points. Figures for banks that failed in 1987 are somewhat less dramatic. Still, 67 percent of 1987 failed banks would have been required to pay higher premiums in 1985, and almost one-fourth would have been placed in the highest risk class.

IV. Risk-based Capital

Conversion of the bank failure model estimates into a risk-based capital system was somewhat more complicated than procedures used for the risk-based insurance premium system. To ensure comparability with the current system, it was decided to set a standard so that if all banks held exactly the required capital ratio, the expected losses to the FDIC would be identical to its expected losses under the current system. It was determined that this would occur if each bank in December 1985 were required to hold enough capital so that its probability of failure was 0.7 percent (about 95 expected bank failures per year).

A floor and ceiling were also imposed so that no bank would be required to have a capital ratio of less than 3 percent nor more than 15 percent. This particular standard was chosen in order to make the expected losses to the FDIC of the risk-based capital system as close as possible to the risk-based insurance system outlined in the previous section. Imposition of the 3 percent minimum floor was similar to the addition of an intercept term in the risk-based premium system, and is a tacit admission that any realistic risk-based capital system would have to have a floor. The 15 maximum capital standard is similar to the cap imposed on the risk-based premium.

Solution for the amount of capital each bank would have to hold follows straightforwardly from the estimated risk index. The formula given in equation (3a) implies that a bank with a risk index value of -4.95 would have a probability of failure of precisely 0.7 percent. Equation (3), therefore, implies that the required minimum capital level, KTA^*, must satisfy

$$(4)\quad -4.95 = -2.42 - .501KTA^* + .428PD90MA + .314\text{LNNACCA} + .269\text{RENEGA} - .223NCOFSA - .331\text{NETINCA},$$

or,

$$(5)\quad KTA^* = 5.04 + .854PD90\text{MA} + .627LNNACCA + .537RENEGA - .445NCOFSA - .661NETINCA,$$

which can be solved for each bank.[16]

Table 5 gives an indication as to how a risk-based capital system might work. It shows the December 1985 distribution of required capital by bank-size class and future failure. Rows (1) through (7) represent banks of increasing size, row (8) shows banks that failed in 1986, row (9) shows banks that failed in 1987 (through September 30), and row (10) shows the sum of all banks. The columns show the number and percent of banks in each size class that would have been assigned to various required capital classes. For each cell, the first number given is the average required capital level for banks in the cell, the second number is the percentage of banks that would have to raise capital to meet the new standard, and the third number is the number of banks in the cell.

The numbers in table 5 suggest several interesting conclusions. Eighty-six percent of all banks would have a risk-based capital assessment below 6.5 percent. A middle group would be required to hold capital ratios between 6.5 and 10 percent; and a small group (3.4 percent of the total) would have to hold capital of over 10 percent of assets. There is an indication that banks with higher risk already hold more capital than required. Thus, almost 92 percent of banks would not have to raise more capital under the risk-based standard. However, there is a small

16 The formula implies that a bank would *reduce* its index value by 0.501 for each percentage point rise in its capital ratio. Thus. a bank with a 5.5 percent capital ratio and a risk index of -3.70 would be required to raise its capital ratio 2.5 percentage points to 8 percent, that is 2.5 = [4.95 - 3.71]/.501. Banks with risk indices below -4.95 would be allowed to divest one percentage point of capital for each 0.501 they were below -4.95.

Estimated Commercial Bank Risk-based Required Capital — December 1985

(Percent of **total assets)**

First number is the average capital ratio for banks in the cell. Second number is percent of banks that would have to raise capital. Third number is number of banks.

Asset Size Class ($ millions)	Required Capital Class						
	(1) < 5.5	(2) 5.5-6.4	(3) 6.5-7.4	(4) 7.5-9.9	(5) 10.0-14.9	(6) 15.0	(7) All Banks
(1) < $10	4.6	6.0	7.0	8.5	11.8	15.0	6.1
	0.0	1.0	3.3	27.7	76.1	84.6	8.5
	529.0	198.0	119.0	130.0	46.0	13.0	1035.0
(2) $10 — $25	4.7	5.9	7.0	8.5	11.6	15.0	5.9
	.1	.9	9.0	50.0	92.9	97.1	10.4
	1936.0	775.0	365.0	326.0	141.0	35.0	3558.0
(3) $25 — $50	4.8	5.9	6.9	8.5	11.8	15.0	5.7
	.2	1.1	14.0	54.0	95.7	100.0	8.3
	2158.0	749.0	336.0	252.0	92.0	15.0	3602.0
(4) $50 — $100	4.8	5.9	6.9	8.4	11.7	15.0	5.6
	.4	3.0	16.7	53.8	90.2	91.7	7.8
	1752.0	535.0	239.0	158.0	61.0	12.0	2757.0
(5) $100 — $500	4.9	5.9	6.9	8.3	11.7	15.0	5.5
	.1	4.0	24.1	69.8	100.0	100.0	7.2
	1366.0	448.0	116.0	96.0	31.0	3.0	2060.0
(6) $500 — $1000	4.9	5.9	6.9	8.7	10.9	15.0	5.5
	1.5	10.8	27.8	100.0	100.0	100.0	10.4
	137.0	37.0	18.0	6.0	3.0	1.0	202.0
(7) > $1000	5.0	5.9	6.8	8.6	10.2	0.0	5.4
	3.1	29.0	47.4	100.0	100.0	0.0	15.3
	191.0	93.0	19.0	4.0	1.0	0.0	308.0
(8) Banks failing in 1986	4.6	5.9	7.1	9.0	12.4	15.0	11.5
	0.0	33.3	53.3	86.4	98.1	100.0	86.5
	5.0	3.0	15.0	22.0	54.0	34.0	133.0
(9) Banks failing in 1987	5.0	6.0	6.8	8.8	12.1	15.0	9.2
	9.1	16.7	21.0	75.5	96.7	72.7	61.4
	11.0	12.0	19.0	49.0	30.0	11.0	132.0
(10) All Banks	4.8	5.9	6.9	8.4	11.7	15.0	5.7
	.3	2.9	13.7	51.1	91.7	94.9	8.8
	8069.0	2815.0	1212.0	972.0	375.0	79.0	13522.0

Source: Board of Governors of the Federal Reserve System.

TABLE 5

group that would have to raise a substantial amount of additional capital. The efficiency of a risk-based system is evident from the fact that aggregate bank capital would be reduced by 18 percent from the actual December 1985 total, yet expected FDIC losses would be exactly the same as under the current system. This happens because the risk-based system shifts capital to those banks most likely to fail.

The evidence of the banks that failed in 1986 and 1987 is particularly telling. All but 18 of the 133 banks that failed in 1986 would have been required to raise additional capital in December 1985. *As* a group, these banks would have been required to almost double their aggregate capital. Over 60 percent of the banks that failed in 1987 would have been required to raise additional capital and over 90 percent would have been assigned a capital ratio above the current standard.

V. Final Comments

The systems presented here are meant to be illustrative and would probably require substantial modification before they could be actually implemented. They do show, however, that both risk-based capital and risk-based insurance systems could be constructed that discriminate between banks in a way that would likely affect behavior.

The similarities between the distribution of banks shown in the tables summarizing the proposals is striking. This, however, should not be surprising since both systems are based on the same risk measure. Indeed, if we had arrayed banks by the amount of new capital they would have to raise, instead of by required levels, the rank orderings of banks in the two systems would have been identical. They differ in the arrangements shown only because some banks that would otherwise have higher risk hold more capital than required under the current system, and thus, would reduce their premiums.

This does not mean that the two systems would have identical impacts on bank behavior or on overall system risk As argued earlier, the regulatory environment surrounding each system is likely to differ. If banks face prices for risk in the capital market different from those charged by the FDIC, there will be inefficiencies in a risk-based capital standard that could produce different levels of system risk.

The incentives for banks to alter their risk-taking activities are very likely to differ between the two systems. It is not clear, however, that the impact of such differences would be major. Both systems share a common basis in the principle of differentially regulating banks according to the risk they represent to society. Implementation of either type of system is likely to lead to significant progress in the battle to control bank risk.