William A. Barnett and Ge Zhou

William A. Barnett is a professor of economics at Washington University, St. Louis. Ge Zhou recently received a doctorate in economics from Washington University, St. Louis. *Research on this project was partially supported by NSF grant SES 9223557. We wish to thank William Brainard for his comments, which substantially influenced the final revision of this paper.*

Financial Firms' Production and Supply-Side Monetary Aggregation Under Dynamic Uncertainty

THIS PAPER IS FOCUSED ON the production theory of the financial firm and supply-side monetary aggregation in the framework of dynamics and risk. On the demand side, there has been much progress in applying consumer demand theory to the generation of exact monetary aggregates and integrating them into consumer demand system modeling.[1] However, on the supply-side, monetary services are produced by financial firms through financial intermediation, and, hence, exact supply-side monetary aggregation must be based upon financial firm output aggregation. Most of the literature on exact aggregation theory is based upon perfect certainty, which often is a reasonable assumption regarding contemporaneous consumer goods allocation decisions. Risk, however, is an important consideration in modeling the decisions of financial intermediaries. Furthermore, that risk not only applies to future prices and interest rates, but also to contemporaneous interest rates and thereby to the contemporaneous user costs of produced monetary services. In this paper we derive a model of financial firm behavior under dynamic risk, and we find the exact monetary services output aggregate. We estimate the Euler equations that comprise the first-order conditions for optimal behavior by financial firms.

Barnett (1978,1980) introduced economic aggregation and index number theory to demand-side monetary aggregation by applying Diewert's (1976) results on superlative index numbers. The proposed Divisia index in Barnett's work is an element of Diewert's superlative index number class. Analogous to demand-side monetary aggregation, Hancock (1985, 1987), Barnett (1987), and Barnett, Hinich and Weber (1986) have provided results on supply-side monetary aggregation.[2] They use neoclassical economic theory to model

[1]See Barnett, Fisher and Serletis (1992).

[2]"Demand-side" and "supply-side" imply respectively the demand for monetary services by consumers and manufacturing firms, and the production of monetary services by financial intermediaries. Barnett (1987) has shown that consumer's demand for money and manufacturing firm's demand for money result in the identical aggregation problem, at least in the perfect certainty case. However, supply-side aggregation of produced monetary services creates uniquely different aggregation problems resulting from the existence of required reserves, which alter the user cost of produced monetary services. For further results regarding demand for monetary services by manufacturing firms, see Robles (1993) and Barnett and Yue (1991).

financial firms' production, so the existing economic aggregation and index number theory are directly applicable. In fact, throughout the literature on applying economic aggregation and index number theory to monetary aggregation, researchers usually assume perfect certainty. Exceptions are Barnett and Yue (1991) and Poterba and Rotemberg (1987), who generalize to demand-side exact monetary aggregation under risk. Supply-side monetary aggregation under risk has not previously been the subject of research.

Introduction of dynamics and uncertainty into supply-side monetary aggregation requires extensions of earlier research in this area. A financial firm's portfolio is generally diversified across different investment instruments, and the portfolio's rate of return is unknown at the time that the investment decision is made. Hence, the assumption of perfect-certainty and single-period modeling is not appropriate. Furthermore, superlative index numbers, such as the discrete time Divisia index, have known tracking ability only under the assumption of perfect certainty. In this paper, we develop a dynamic approach to supply-side monetary aggregation under uncertainty.

Historically, the literature on financial intermediation has produced many diverse models, often linked only weakly with neoclassical economic theory and having various objectives. The early view of the creation of money by financial firms, primarily viewed to be banks, was the deposit multiplier approach. By this theory in its original form, the process of creating money is simply determined by the reserve requirement ratio. Another approach is based upon the Miller-Modigliani theorem, which asserts the irrelevance of financial firms to the real economy in a setting of a perfect capital market. In recent years, many economists have questioned the appropriateness of either of those two very different propositions and attempts have been made to extend those theories by weakening the underlying assumptions.

Another approach is based upon the capital-asset pricing model (CAPM). Under the assumptions of that model, either the financial firm's portfolio rate of return is normally distributed or investors have a quadratic utility function defined over end-of-period wealth. Under either of those assumptions, the financial firm's optimal portfolio behavior can be represented by maximizing utility over the portfolio's expected rate of return and variance. This approach has been useful in modeling the optimal portfolio allocation decision conditionally upon the real resource inputs, which are not explained endogenously. Another important approach is represented by Diamond and Dybvig (1983). They apply traditional consumption-production theory and use an intertemporal model subject to privately observed preference shocks to examine the equilibrium between banks and depositors. The studies in this tradition have been successful in explaining bank runs. However, banks, serving solely as a production technology to depositors, play only a passive role in that approach.

Another approach is represented by Hancock (1985,1987), Barnett (1987), and Barnett, Hinich and Weber (1986). They treat the financial intermediary in the same manner as a conventional production unit and use neoclassical firm theory to model a financial intermediary's production of output services and employment of inputs subject to the firm's technological feasibility constraint.[3] This approach fully models the role played by financial firms as producers of monetary services. Moreover, it provides the needed tools to apply existing economic aggregation theory to aggregation over financial firms' output monetary services, which comprise the economy's inside money. However, those studies have not developed a dynamic model of financial firms' production under uncertainty. This paper provides that difficult extension of financial firm modeling and output aggregation under neoclassical assumptions with dynamic risk.

With the theoretical model of a financial firm's monetary services production and the derived exact theoretical output aggregate, we estimate the model's parameters and test for weak separability of output services from factor inputs. We then substitute the parameter estimates into the weakly separable output aggregator function to generate the estimated exact supply-side monetary aggregate.[4] To this end, we develop a procedure for testing weak separability and for estimating the parameters of a flexible functional form specification of bank technology. The estimation is accomplished

[3]The papers of Tobin (1961) and Brainard and Tobin (1963, 1968) were the first to argue forcefully for the use of microeconomics and equilibrium theory in modeling the financial firm.

[4]Diewert and Wales (1987) and Blackorby, Schworm and Fisher (1986) have illustrated the difficulty of maintaining flexibility, regularity and weak separability simultaneously.

through Hansen and Singleton's (1982) generalized method of moments approach to estimating Euler equations.

Our empirical results are based upon commercial banking data. Our evidence indicates that banks' outputs are weakly separable from factor inputs in the transformation function. Moreover, even under uncertainty, the Divisia index provides a better approximation to the estimated theoretical aggregate than does the simple-sum or CE index.[5] These findings support the existence of a supply-side monetary aggregate and the potential usefulness of the Divisia index to aggregate over the weakly separable monetary assets on the supply side of money markets. The result is a measure of inside money, in the sense of monetary services produced by private financial firms.

The paper proceeds as follows. In the next section, we construct our theoretical model of monetary service production by financial firms under dynamic uncertainty. The model reduces to a dynamic stochastic choice problem, for which we derive the Euler equations. In the third section, we present our approach to flexible parametric specification, weak separability testing and parameter estimation using Hansen and Singleton's (1982) generalized method of moments estimation. The fourth section formulates the empirical application using banking industry data. The fifth section contains the empirical results, including parameter estimates, weak separability test results, the estimated theoretical aggregate, and the comparison among index number approximations to the estimated exact aggregate, where the index numbers considered include the Divisia, simple-sum and CE indexes. Section 6 brings together the demand side with the supply side to investigate the implications of our model in general equilibrium. Section 7 provides a graphical illustration of the errors-in-the-variables problem produced by the use of the simple-sum index as a measure of the monetary service flow. The final section presents a few concluding remarks.

THEORETICAL MODEL

In this section, we derive our theoretical model of monetary services production by financial firms under dynamic uncertainty. Consider a financial firm which issues its own liabilities and reinvests the borrowed funds in primary financial markets. In this process, real resources such as labor, materials and capital are used as factors of production in creating the services of the produced liabilities. Those produced liabilities are deposit accounts providing monetary service combinations that would not have existed in the economy without the financial firm. The liabilities of the financial firms include, for example, demand deposits and passbook accounts, and are assets to the depositors. The value added through the creation of those assets by a financial intermediary is that firm's contribution to the economy's inside money services. Without the existence of financial firms and the accounts that they create, investors in money markets would be limited to the use of primary money-market securities as the short maturity assets in their portfolios. While the produced liabilities of financial firms may not appear to be "outputs" to an accountant looking at the firm's balance sheet, the produced liabilities of financial firms are the outputs of the firms' production technologies.[6]

The financial firm's profits are made from the interest rate spread between the financial firm's financial assets (loans) and the firm's produced liabilities. That spread must exceed the real resource costs, in order for the firm to profit from its operation. Let Y_t be the real balances of the financial firm's asset (loan) portfolio during period t.[7] Let R_t be the portfolio rate of return, which is unknown at the beginning of each period. Financial firms also hold excess reserves in the form of cash, which has a nominal return of zero. The real balance of cash holding is C_t. Let y_{it} be real balances in the firm's ith produced account type and h_{it} be holding cost per dollar for that liability, where $i=1,...,I$.[8] The amount of the jth real resource used is z_{jt}, and

[5] The formula for computing the Divisia index is in Barnett (1980). Further details regarding the data sources used with the index are in Thornton and Yue (1992), who also provide instructions on downloading the data from the Federal Reserve Bank of St. Louis' public electronic bulletin board, called FRED. The formula for computing the CE ("currency equivalent") index is in Rotemberg, Driscoll and Poterba (1991).

[6] See Barnett (1987).

[7] As used in this paper, portfolio is the sum of all investments.

[8] The holding cost h_{it} is defined as $h_{it}=r_{it}+R_tk_{it}$. In this formula, r_{it} is the account's net interest rate, which is defined such that all the benefits (for example, service charges) and costs (for example, deposit insurance) generated by the borrowed funds have been factored into the interest rate, and R_tk_{it} is the implicit tax rate on the financial firm from the existence of a reserve requirement on that account type. Required reserves are assumed to yield no interest and hence, produce an opportunity cost to the financial firm, since the firm otherwise could have invested the required reserves at a positive rate of return.

its price is w_{jt}, *where* $j=1,...,J$. Let P_t be the general price index, which is used to deflate nominal to real units. All financial transactions are contracted at the beginning of each period, but interest is paid or received at the end of the period. The cost of employing resource z_{jt} is paid at the start of the period.

The firm's variable profit at the beginning of period t in accordance with Hancock's (1991, equation 3.1) formula, is

$$(1)\ \pi_t = (1+R_{t-1})\ Y_{t-1}P_{t-1} - Y_tP_t + C_{t-1}P_{t-1} - C_tP_t + \sum_{i=1}^{I} [y_{it}P_t - (1+h_{i,\,t-1})y_{i,\,t-1}P_{t-1}] - \sum_{j=1}^{J} w_{jt}z_{jt}.$$

The first two terms in equation 1 represent the net cash flow generated from rolling over the loan portfolio during period t. The third and fourth terms represent the change in the nominal value of excess reserves. The fifth term is the net cash flow from issuing produced financial liabilities. The last term is total payments for real resource inputs.

Portfolio Y_t investment, however, is constrained by total available funds, under the assumption that all earnings are paid out as dividends. The relationship is

$$(2)\ Y_tP_t = \sum_{i=1}^{I} (1-k_{it})\ y_{it}P_t - C_tP_t - \sum_{j=1}^{J} w_{jt}z_{jt},$$

where k_{it} is the reserve requirement ratio for the ith produced account type, with $0 \leq k_{it} \leq 1$. Rearranging, equation 2 can be seen to state that total deposits $\sum_{i=1}^{I} y_{it}P_{it}$ are allocated to required reserves, excess reserves, investment in loans, and payments for all real resource inputs. Substituting 2 into 1 to eliminate Y_t, we obtain the firm's profit function subject to its balance sheet constraint:

$$(3)\ \pi_t = \sum_{i=1}^{I} \{[(1+R_{t-1})\ (1-k_{i,\,t-1}) - (1+h_{i,\,t-1})]y_{i,\,t-1}P_{t-1} + k_{it}y_{it}P_t\} - R_{t-1}C_{t-1}P_{t-1} - \sum_{j=1}^{J} (1+R_{t-1})\ w_{j,\,t-1}z_{j,\,t-1}.$$

We assume the financial firm chooses the level of borrowed funds, excess reserves, and real resource inputs to maximize its expected discounted intertemporal utility of variable profits, subject to the firm's technology. We further assume the financial firm's intertemporal utility function is additively separable. Then, the firm's maximization problem can be expressed by the following dynamic choice problem:

$$(4)\ Max\ E_t\Big[\sum_{s=t}^{\infty} \Big(\frac{1}{1+\mu}\Big)^{s-t} U(\pi_s)\Big]$$

$$s.t.\ \ \Omega(y_{1s},...,y_{Is}, C_s, z_{1s},..., z_{Js}) = 0 \quad \forall\ s \geq t,$$

where E_t denotes expectation conditional on the information known at time t, μ is the subjective rate of time preference and is assumed to be constant, U is the utility function, π_s is the variable profit at period s given by equation 3, and Ω is the firm's transformation function, defining the firm's efficient production technology from

$$(5)\ \Omega(y_{1s},...,y_{Is}, C_s, z_{1s},..., z_{Js}) = 0\ \forall\ s \geq t.$$

In accordance with the usual properties of a neoclassical transformation function, Ω is convex in its arguments. In addition, the inputs are distinguished from the outputs by the inequality constraints:[9]

$$(6)\ \frac{\partial\Omega}{\partial C_t} \leq 0,\ \frac{\partial\Omega}{\partial z_{jt}} \leq 0\ \forall\ j=1,...,J$$

and

$$(7)\ \frac{\partial\Omega}{\partial y_{it}} \geq 0\ \forall\ i=1,...,I.$$

We also assume that Ω is continuous and second-order differentiable.

Substituting equation 3 into 4, we have

$$(8)\ Max\ E_t\ \{\sum_{s=t}^{\infty} \Big(\frac{1}{1+\mu}\Big)^{s-t} U(\sum_{i=1}^{I} \{[(1+R_{t-1})\ (1-k_{i,\,t-1}) - (1+h_{i,\,t-1})]y_{i,\,t-1}P_{t-1} + k_{it}y_{it}P_t\} - R_{t-1}C_{t-1}P_{t-1} - \sum_{j=1}^{J} (1+R_{t-1})\ w_{j,\,t-1}z_{j,\,t-1})\}$$

$$s.t.\ \ \Omega(y_{1s},...,y_{Is}, C_s, z_{1s},..., z_{Js}) = 0\ \ \forall\ s \geq t.$$

We now proceed to derive the Euler equations, comprising the first-order conditions, for this stochastic optimal control problem. We use Bellman's method. To do so, we must put the decision into Bellman's form, which requires identifying the state and control variables and determining that the decision, stated in terms of those variables, is in the form providing known Euler equation structure.

[9]See Barnett (1987), Hall (1973) and Diewert (1973).

We assume that the financial firm behaves competitively, so that the prices $h_{i,s-1}$ and $w_{j,s-1}$ are taken as given by the firm. In addition, $h_{i,s-1}$ and $w_{j,s-1}$ are nonstochastic, since they are lagged one period. From the same perfect competition assumption, it follows that R_s, k_{is}, and P_s are random processes that are not controllable by the firm. We select as state variables during period s: $y_{i,s-1}$ $\forall$ i, $z_{j,s-1}$ $\forall$ j, C_{s-1}, R_{s-1}, R_s, k_{is}, $h_{i,s-1}$ $\forall$ i, $w_{j,s-1}$ $\forall$ j, P_{s-1}, and P_s. We choose y_{is} $\forall$ i and z_{js} $\forall$ j to be the control variables during period s.

Define $\boldsymbol{w}_s$ to be the vector of all of the state variables, and define $\boldsymbol{u}_s$ to be the vector of all control variables. Let Λ_s be the subset of state variables defined by $\Lambda_s = (R_s, k_{is}, h_{i,s-1}\ \forall\ i,\ w_{j,s-1}\ \forall\ j, P_s)$. We assume that Λ_s follows a first-order Markov process, with transitions governed by the conditional distribution function $F(\Lambda_{s+1}|\Lambda_s)$. Hence, the transition equation for state variables $(R_{s-1}, R_s, k_{is}, h_{i,s-1}\ \forall\ i,\ w_{j,s-1}\ \forall\ j, P_{s-1}, P_s)$ is implicitly defined by $F(\Lambda_{s+1}|\Lambda_s)$. The transition equations for $y_{i,s-1}$ $\forall$ i and $z_{j,s-1}$ $\forall$ j are the trivial identities

(9) $y_{is} = y_{is}, \forall\ s$

and

(10) $z_{js} = z_{js}, \forall\ s.$

The role played by these two equations in our application of Bellman's method follows from the fact that each of the variables in equations 9 and 10 are included both among the control and state variables, although with a time shift distinguishing them in each of their roles.[10] Hence, with the appropriate time shift in the subscript, equations 9 and 10 can be viewed as connecting together some of the control and state variables. This connection accounts for the function of those equations as transition equations. In particular, the left-hand sides can be identified as next-period state variables, while the right-hand sides can be identified as current-period control variables. Hence, each of those equations can be interpreted as defining the evolution of a state variable conditionally on a control variable. The transition equation for C_{s-1} is implicitly determined by the transformation function 5.

The objective function in equation 8 is an infinite summation of discounted utilities of variable profits, starting at period t. Recalling the time shifts appearing in our definition of the state and control variables during period s, we see that the discounted utility of variable profit at period s depends only on that period's state variables and control variables. By examining the transition equations, it is evident that each state variable is a function of only previous controls and not of previous values of the states. In particular, if we let $\boldsymbol{g}$ represent the vector of all transition functions, we can rewrite the dynamic decision problem as

$$Max\ E_t\ \{\sum_{s=t}^{\infty} (\frac{1}{1+\mu})^{s-t} U[\pi_s(\boldsymbol{w}_s, \boldsymbol{u}_s)]\}$$

$$s.t.\ \boldsymbol{w}_{s+1} = \boldsymbol{g}(\boldsymbol{u}_s),\ s \geq t.$$

This dynamic problem meets all of the conditions to be a recursive problem in the Bellman form. Using Bellman's principle, we can derive the first-order conditions for solving the dynamic problem 8. The Bellman recursive equation is

$$v(\boldsymbol{w}_t) = \max_{\boldsymbol{u}_t} E_t\{U[\pi_t(\boldsymbol{w}_t, \boldsymbol{u}_t)] + \frac{1}{1+\mu} v(\boldsymbol{w}_{t+1}) \mid \boldsymbol{w}_t,\ s.t.\ \boldsymbol{w}_{t+1} = \boldsymbol{g}(\boldsymbol{u}_t)\},$$

where $v(\boldsymbol{w}_t)$ is the optimized value of the objective function.

The first-order conditions for the Bellman equation are

$$(11)\ E_t\Big[\frac{\partial U}{\partial \pi_t}(\pi_t)\frac{\partial \pi_t}{\partial \boldsymbol{u}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t) + \frac{1}{1+\mu}\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{u}_t}(\boldsymbol{u}_t)\frac{\partial v}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_{t+1}) \mid \boldsymbol{w}_t\Big] = \mathbf{0}.$$

The functional form of v is unknown. However, since $\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{w}_t} = \mathbf{0}$ we can use the Benveniste and

[10]The use of such trivial identities as transition equations (laws of motion) in optimal control and dynamic programming is not unusual. For example, it is common in optimal growth models to define current capital stock to be a state variable, while next period's capital stock is defined to be a control, with those state and control variables tied together by a trivial identity. The nontrivial dynamics is found in the objective function of such models. See, for example, Sargent (1987, p. 24).

Scheinkman equations to eliminate $\frac{\partial v}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_{t+1})$.[11]

The general form of the Benveniste and Scheinkman equations is

$$\frac{\partial v}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_t) = \frac{\partial U}{\partial \pi_t}(\pi_t)\frac{\partial \pi_t}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t) + \frac{1}{1+\mu}E_t\left[\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t)\frac{\partial v}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_{t+1})\right].$$

Since $\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{w}_t} = \boldsymbol{0}$, the above equation implies

$$(12)\quad \frac{\partial v}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_t) = \frac{\partial U}{\partial \pi_t}(\pi_t)\frac{\partial \pi_t}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t).$$

Substituting 12 into 11, we get

$$(13)\quad E_t\left[\frac{\partial U}{\partial \pi_t}(\pi_t)\frac{\partial \pi_t}{\partial \boldsymbol{u}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t) + \frac{1}{1+\mu}\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{u}_t}(\boldsymbol{u}_t)\frac{\partial U}{\partial \pi_t}(\pi_{t+1})\frac{\partial \pi_t}{\partial \boldsymbol{w}_t}(\boldsymbol{w}_{t+1}, \boldsymbol{u}_{t+1}) \mid \boldsymbol{w}_t\right] = \boldsymbol{0}.$$

A very general specification of utility to represent risk is the hyperbolic absolute risk aversion (HARA) class, defined by

$$(14)\quad U(\pi_t) = \frac{1-\rho}{\rho}\left(\frac{h}{1-\rho}\pi_t + d\right)^{\rho},$$

where ρ, h and d are three parameters to be estimated. The following useful utility functions are fully nested special cases of the HARA class:[12]

a. risk neutrality: $\rho=1$, $U(\pi_t)=h\pi_t$,

b. quadratic: $\rho=2$, $U(\pi_t) = -(1/2)(-h\pi_t + d)^2$,

c. negative exponential: $\rho=-\infty$ *and* $d=1$, $U(\pi_t)=-e^{-h\pi_t}$,

d. power: $d=0$ *and* $\rho<1$, $U(\pi_t) = (\pi_t^{\rho}/\rho)$,

e. logarithmic: $d=\rho=0$, $U(\pi_t) = \log \pi_t$.

The general HARA specification for $U(\pi_t)$ satisfies the relevent theoretical regularity conditions when the domain of $U(\pi_t)$ is constrained to $\{\pi_t : \frac{h}{1-\rho}\pi_t + d > 0\}$ with h constrained to satisfy $h>0$. When $\rho>1$, absolute risk aversion (Arrow-Pratt) is decreasing, and when $\rho>1$, absolute risk aversion is increasing. The power utility function special case is very widely used. Since that functional form exhibits constant relative risk aversion (CRRA), the power utility function often is called the CRRA or isoelastic case.[13]

Differentiating (14) with π_t, we get

$$(15)\quad \frac{\partial U}{\partial \pi_t} = h\left(\frac{h}{1-\rho}\pi_t + d\right)^{\rho-1}.$$

Using equations 13 and 15 along with the defined state variables, control variables and transition equations, we obtain

$$(16)\quad E_t\Big\{P_t k_{it}\left(\frac{h}{1-\rho}\pi_t + d\right)^{\rho-1} + P_t\frac{1}{1+\mu}\big[(1+R_t)(1-k_{it}) - (1+h_{it})\big] + R_t\frac{\partial\Omega/\partial y_{it}}{\partial\Omega/\partial C_t}\Big]\left(\frac{h}{1-\rho}\pi_{t+1} + d\right)^{\rho-1}\Big\} = 0 \quad \forall\, y_{it},\ i=1,\dots,I$$

and

$$(17)\quad E_t\left[P_t R_t\frac{\partial\Omega/\partial z_{jt}}{\partial\Omega/\partial C_t}\left(\frac{h}{1-\rho}\pi_{t+1} + d\right)^{\rho-1} - (1+R_t)w_{jt}\left(\frac{h}{1-\rho}\pi_{t+1} + d\right)^{\rho-1}\right] = 0 \quad \forall\, z_{jt},\ j=1,\dots,J.$$

Equations 16 and 17 are a system of $I+J$ nonlinear equations. Theoretically, from 16 and 17 plus the transformation function 5, we could solve for $(Y_{1t},\dots, Y_{It}, C_t, Z_{1t},\dots,z_{Jt})$. However, in practice the solution could be produced only numerically, since a closed form algebraic solution rarely exists for such Euler equations.

---

[11] See Sargent (1987) for an excellent presentation of dynamic programming.

[12] See Ingersoll (1987, pp. 37-40). In case (d) below, imposing the restriction $d=0$ alone on equation 14 will not produce the exact form provided for the power function. However, the form acquired subject to that sole restriction is a positive affine transformation of the power function. Hence both forms represent the same risk behavior.

[13] See, for example, Barnett and Yue (1991).

In the following discussion, we extend the dynamic decision 8 into the more general case incorporating learning by doing technological change. In the econometric literature on estimating returns to scale in manufacturing, increasing returns to scale usually are found, despite the fact that increasing returns to scale violates the second-order conditions for profit maximization. We believe that a likely source of this paradox is the potential to confound technological change with returns to scale, when learning by doing technological change exists but is not incorporated within one's model.

Let $\boldsymbol{y}_t$ be the vector of $\boldsymbol{y}_{it}$ for all i and $\mathbf{z}_t$ be the vector of $\mathbf{z}_{jt}$ for all j. We then write the maximization problem as

$$(18)\ Max\ E_t\Big[\sum_{s=t}^{\infty}\Big(\frac{1}{1+\mu}\Big)^{s-t}U(\pi_s)\Big]$$

$$s.t.\ \Omega(\boldsymbol{y}_s, C_s, \mathbf{z}_s, \boldsymbol{y}_{s-1}) = 0 \qquad \forall\ s \geq t.$$

The appearance of $\boldsymbol{y}_{s-1}$ in the transformation function represents learning by doing. Firm technology improves through experience.

At the present stage of this research, we are not using the learning by doing extension of our model in our empirical work, so we only provide the Euler equations below, without supplying the details of the derivation. Those Euler equations under learning by doing are

$$(19)\ E_t\Big\{\frac{\partial U}{\partial \pi_t}(\pi_t)\frac{\partial \pi_t}{\partial \boldsymbol{y}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t) + \frac{1}{1+\mu}\Big[\frac{\partial U}{\partial \pi_t}(\pi_{t+1})\frac{\partial \pi_t}{\partial \boldsymbol{y}_{t-1}}(\boldsymbol{w}_{t+1}, \boldsymbol{u}_{t+1}) - \frac{1}{1+\mu}\frac{\partial\Omega/\partial \boldsymbol{y}_{t-1}}{\partial\Omega/\partial C_t}(\boldsymbol{w}_{t+1}, \boldsymbol{u}_{t+1})\frac{\partial U}{\partial \pi_t}(\pi_{t+2})\frac{\partial \pi_t}{\partial C_{t-1}}(\boldsymbol{w}_{t+2}, \boldsymbol{u}_{t+2}) - \frac{\partial\Omega/\partial \boldsymbol{y}_t}{\partial\Omega/\partial C_t}(\boldsymbol{w}_t, \boldsymbol{u}_t)\frac{\partial U}{\partial \pi_t}(\pi_{t+1})\frac{\partial \pi_t}{\partial C_{t-1}}(\boldsymbol{w}_{t+1}, \boldsymbol{u}_{t+1})\Big]\Big\} = \mathbf{0} \qquad \forall \boldsymbol{y}_t$$

and

$$(20)\ E_t\Big\{\frac{\partial U}{\partial \pi_t}(\pi_t)\frac{\partial \pi_t}{\partial \mathbf{z}_t}(\boldsymbol{w}_t, \boldsymbol{u}_t) + \frac{1}{1+\mu}\Big[\frac{\partial U}{\partial \pi_t}(\pi_{t+1})\frac{\partial \pi_t}{\partial \mathbf{z}_{t-1}}(\boldsymbol{w}_{t+1}, \boldsymbol{u}_{t+1}) - \frac{\partial\Omega/\partial \mathbf{z}_t}{\partial\Omega/\partial C_t}(\boldsymbol{w}_t, \boldsymbol{u}_t)\frac{\partial U}{\partial \pi_t}(\pi_{t+1})\frac{\partial \pi_t}{\partial C_{t-1}}(\boldsymbol{w}_{t+1}, \boldsymbol{u}_{t+1})\Big]\Big\} = \mathbf{0} \qquad \forall \mathbf{z}_t$$

Equations 19 and 20 are generalizations of (16) and (17). If learning by doing is excluded by imposing $\partial\Omega/\partial \boldsymbol{y}_{t-1} = \mathbf{0}$, then (19) and (20) reduce to (16) and (17), respectively. In the rest of the current paper, we return to the special case of no technological change.

A further nested special case is also interesting. We acquire risk neutrality by setting $\rho=1$. As is conventional under risk neutrality, discounting is acquired objectively by replacing the subjective rate of time discount, μ, by R_t.[14] One reason for interest in that special case is that, in general equilibrium theory, the assumption of complete contingent claims markets combined with perfect competition can be shown, under certain additional assumptions, to produce the conclusion that firms will be risk neutral, even if their owners are risk-averse. The risk aversion of the owners then is captured within the contingent claims prices, which are taken as given by the firms' managers under perfect competition.[15]

While this theoretical issue is interesting, we do not consider it alone to be a convincing reason to impose risk neutrality on the management of an industry that behaves in a manner exhibiting clear risk aversion. However, we are interested in that fact that the Divisia index, along with virtually all of the literature on index number theory, is produced under the assumption of perfect certainty. This fact would suggest that the tracking ability of such index numbers may degrade as the level of risk aversion in-

[14]While the risk-neutral case is acquired directly by making those substitutions in the original decision problem, the resulting Euler equations are not acquired simply by making those substitutions in the risk-averse Euler equations, 16 and 17. The reason is that a cancellation within the Euler equations that is produced when the rate of discount is the constant, μ, does not apply when the rate of discount becomes the variable, R_t. In particular, after replacing ρ with 1.0, and μ with R_t, it also is necessary to multiply the two terms within equation 17 *by* $1/(1+R_t)$ to get the risk neutral case Euler equations. No such adjustment is needed within equation 16, since no relevant factors cancelled out in the derivation of equation 16. This observation also is relevant to the risk-neutral Euler equations 80 and 81 below.

[15]See, for example, Debreu (1959, ch. 7) and Duffie (1991, section 6.3). Regarding the complications produced by incomplete markets, see Magill and Shafer (1991, section 4).

creases. Hence, we produce results both with and without risk neutrality imposed, as a means of exploring the extent to which the tracking ability of index numbers is degraded in the risk averse case relative to the risk-neutral case.

Under risk neturality, our Euler equations reduce to[16]

$$(19') \quad E_t\left[P_t \frac{R_t(1-k_{it}) - r_{it}}{1+R_t} + P_t \frac{R_t}{1+R_t} \frac{\partial\Omega/\partial y_{it}}{\partial\Omega/\partial C_t}\right] = 0 \qquad \forall\, y_{it},\ i=1,\ldots,I$$

and

$$(20') \quad E_t\left[P_t \frac{R_t}{1+R_t} \frac{\partial\Omega/\partial z_{jt}}{\partial\Omega/\partial C_t} - w_{jt}\right] = 0 \qquad \forall\, z_{jt}, j=1,\ldots,J.$$

The assumption of perfect competition is itself sufficient for the existence of a representative firm. See Debreau 1959, p. 45, result 1. Hence, the theory acquired from our model can be applied with data aggregated over banks.[17]

SUPPLY-SIDE MONETARY AGGREGATION AND A WEAK SEPARABILITY TEST

Having formulated our dynamic model of financial firm production under uncertainty and having derived the Euler equations, we can proceed to investigate the exact supply-side monetary aggregates that are generated, if the firm's output monetary services are weakly separable from inputs.

Supply-Side Aggregation

Most money in modern economies is inside money, which is simultaneously an asset and a liability of the private sector. Inside money provides net positive services to the economy, as a result of the value added that is created by the financial intermediation that produces the inside money. In our model, the borrowed funds that are outputs produced by financial intermediaries are inside money. Inside money may take various forms such as demand deposits, interest-bearing checking accounts, small time deposits, and checkable money market deposit accounts. The sum of the dollar value in such accounts does not measure the services of inside money, any more than the sum of subway trains and roller skates measures transportation services, since the components of the aggregate are not perfect substitutes. The aggregation-theoretic exact quantity aggregate does, however, measure the service flow.[18]

The procedures involved in identifying and generating the exact quantity aggregates of microeconomic theory are described in detail by Barnett (1980). The approach necessarily involves two steps: identification of the components over which exact aggregation is admissible and determination of the aggregator function defined over those components. The first step determines whether or not an exact aggregate exists, and the second step creates the exact aggregate that is consistent with microeconomic theory. The second step cannot be applied unless the first step succeeds in identifying a component cluster that satisfies the existence condition. That existence condition, which is the basis for the first stage clustering of components, is blockwise weak separability. In accordance with the definition of weak separability, a blocking of components is admissible if and only if the goods in the block can be factored out of the structure of an economy through a subfunction. In other words, it must be possible to formulate the economic structure in the form of a composite function, with the goods in the cluster being the sole variables entering into the inner function of the structure. If that condition

[16]Observe that only one time subscript exists in the risk-neutral Euler equations, so that the solution becomes static. Once the nonlinear utility function has been removed from the objective function, the terms with common time subscripts can be grouped together. However, under risk aversion, even under our assumption of intertemporal strong separability, more than one time subscript exists within the utility function for each time period, since both current and lagged t appear as subscripts in equation 3 for each value of profit, π_t. Hence, the dynamics found within the objective function of equation 4 cannot be removed by regrouping terms.

[17]In fact, Debreu's theorem can be used to aggregate over all firms of all types in the economy to produce the aggregated technology of the country. The representative firm maximizes profits subject to that aggregated technology. However, we use the theorem only to aggregate over the firms in one industry. It should be observed that the ease of aggregation over firms under perfect competition is in marked contrast with the complexity of the theorems on aggregating over consumers.

[18]See, for example, Blackorby, Schworm and Fisher (1986) regarding the importance of using appropriately aggregated output data from firms.

is satisfied, an exact quantity aggregate exists over the goods in the cluster and the aggregator function that produces the exact aggregate over those goods is the inner function within the composite function.

Let $\boldsymbol{y}=(y_{1t},\ldots,y_{It})'$ and $\boldsymbol{x}=(C_t, z_{1t},\ldots,z_{Jt})'$ where $\boldsymbol{y}$ is the vector of the firm's outputs and $\boldsymbol{x}$ is the vector of the firm's inputs. The transformation function becomes

$$\Omega(\boldsymbol{y},\boldsymbol{x})=0.$$

An exact supply-side aggregator exists over all of the elements of $\boldsymbol{y}$ if and only if $\boldsymbol{y}$ is weakly separable from $\boldsymbol{x}$ within the structure of Ω. Mathematically, that statement is equivalent to the existence of two functions H and y_0 such that

$$\Omega(\boldsymbol{y},\boldsymbol{x})=H(y_0(\boldsymbol{y}),\boldsymbol{x}),$$

where $y_0(\boldsymbol{y})$ is a convex function of $\boldsymbol{y}$.[19] In aggregation theory, $y_0(\boldsymbol{y})$ is called the output aggregator function. Furthermore, suppose that $y_0(\boldsymbol{y})$ is linearly homogeneous in $\boldsymbol{y}$. Under this assumption, if each y_i grows at the same common rate, the theoretical aggregate $y_0(\boldsymbol{y})$ will grow at that rate. Clearly, without that condition, $y_0(\boldsymbol{y})$ could not serve as a reasonable aggregate.[20]

As shown by Leontief (1947a, 1947b), the weak separability condition is equivalent to

$$(21)\quad \frac{\partial}{\partial x_k}\left(\frac{\partial\,\Omega(\boldsymbol{y},\boldsymbol{x})/\partial y_i}{\partial\,\Omega(\boldsymbol{y},\boldsymbol{x})/\partial y_j}\right) = 0 \quad \textit{for all } k.$$

If a subset of the components of $\boldsymbol{y}$ were weakly separable from all of the other variables in Ω, then an exact output aggregate would exist only over the services of that subset of components and not over the services of all outputs. If we can test for the separability structure of the transformation function and acquire the functional form of $y_0(\boldsymbol{y})$, when $\boldsymbol{y}$ is weakly separable from $\boldsymbol{x}$, then we could estimate the parameters of $y_0(\boldsymbol{y})$ to acquire an econometric estimate of the exact output aggregate.

Although aggregation theory can provide us with the tools to estimate the exact aggregator function, the resulting aggregate is specification and estimator dependent. Alternatively, the literature on statistical index number theory provides nonparametric approximations to aggregator functions when the existence of the aggregator can be demonstrated through a weak separability test. Statistical index numbers provide only approximations to the theoretical aggregate, however, and when uncertainty exists, little is known about the tracking ability of statistical index numbers as approximations to the exact aggregates of microeconomic theory. In this paper we consider the Divisia, simple-sum and CE indexes to explore their abilities to track the econometrically estimated exact output aggregate.[21] We produce our econometric estimate of the exact theoretical aggregate, for comparison with the index numbers, by using generalized method of moments (GMM) estimation of the parameters of the Euler equations under rational expectations. We do the GMM estimation both under risk aversion and under the imposition of risk neutrality, to investigate sensitivity of our conclusions to risk aversion.

Flexibility, Regularity and Weak Separability

In empirical applications, there are two widely used approaches to testing for the weak separability condition that is necessary for economic aggregation: the nonparametric, nonstochastic approach based upon revealed preference and the statistical, parametric approach.[22] Since we are working from within a parametric specification, the conventional parametric approach to testing the hypothesis is to be preferred. In fact, we shall see that weak separability will be a strictly nested null hypothesis within our parametric specification, and, hence, conventional statistical testing is available immediately. In addition, the nonparametric approach, at its current state of development, is nonstochastic and, hence, has unknown power.

[19]See Barnett (1987).

[20]Without linear homogeneity of y_0, the exact aggregate would become the distance function, rather than y_0, and would reduce to y_0, only under linear homogeneity of y_0. We do not pursue that generalization in this study, but see Barnett (1980) for details.

[21]The Divisia monetary aggregate index was introduced by Barnett (1978, 1980). The simple-sum index is the traditional monetary index acquired by simply adding up the component quantities without weights. The CE index is the currency equivalence aggregate, originated by Rotemberg (1991) and Rotemberg, Driscoll and Poterba (1991). For an alternative interpretation of the CE index as an economic monetary stock index connected with the Divisia service flow, see Barnett (1991).

[22]See Swofford and Whitney (1987).

Restrictive parametric specifications can bias inferences. As a result, flexible functional forms have been developed and are widely used in current studies. A flexible functional form, by definition, has enough free parameters to approximate locally to the second-order any arbitrary function.[23] However, using flexible functional forms creates a new problem. These models, unlike earlier, more restrictive models, may not globally satisfy the regularity conditions of economic theory, including the monotonicity and curvature conditions. It would be desirable to be able to impose global theoretical regularity on these models, but most of the models in the class of flexible functional forms lose their flexibility property, when regularity is imposed.[24] We use a model that permits imposition of regularity, without compromise of flexibility.

While flexibility and regularity are desirable in any neoclassical empirical study, weak separability in some blocking of the goods is also needed to permit aggregation over the goods in that block. We again are presented with the risk of losing flexibility by imposing a restriction, and in fact imposing weak separability on many flexible functional forms greatly damages the specifications' flexibility. For example, imposing weak separability on the translog function does great damage to its flexibility.[25] Because of the difficulties in imposing regularity and separability simultaneously without damage to flexibility, parametric tests of weak separability have been slow to appear and have been applied only to the static, perfect certainty case in which duality theory is available. In our case of dynamic uncertainty, very little duality theory is currently available.

In this section, we develop an approach that permits testing and imposing blockwise weak separability within a globally regular and locally flexible transformation function that is arising from a dynamic, stochastic choice problem. Our approach uses Diewert and Wales' (1991) symmetric generalized McFadden functional form to specify the technology of the firm.[26] In the discussion to follow, we first specify the model's form under the null hypothesis of weak separability in outputs. We then provide the more general form of the model that remains valid without the imposition of weak separability.

Using the notations defined previously, if $\boldsymbol{y}$ is weakly separable from $\boldsymbol{x}$, then

$$\Omega(\boldsymbol{y},\boldsymbol{x}) = H(y_0(\boldsymbol{y}),\boldsymbol{x}).$$

We further assume that the transformation function is linearly homogenous. Instead of specifying the form of the full transformation function Ω directly and thereafter imposing weak separability in $\boldsymbol{y}$, we impose weak separability directly by specifying $H(y_0,\boldsymbol{x})$ and $y_0(\boldsymbol{y})$ separately. We acquire our weakly separable form for Ω by substituting $y_0(\boldsymbol{y})$ into $H(y_0,\boldsymbol{x})$. Since our specifications of $y_0(\boldsymbol{y})$ and $H(y_0,\boldsymbol{x})$ are both flexible, it follows that our specification of Ω is flexible, subject to the separability restriction.

We specify H to be the symmetric generalized McFadden functional form

$$(22)\quad H(y_0,\boldsymbol{x}) = a_0 y_0 + \boldsymbol{a}'\boldsymbol{x} + \frac{1}{2}\,[\,y_0,\ \boldsymbol{x}'\,]\,\overline{\boldsymbol{A}}\begin{bmatrix} y_0 \\ \boldsymbol{x} \end{bmatrix} / \ \alpha'\boldsymbol{x},$$

with $\alpha'\boldsymbol{x}\neq 0$, where a_0, $\boldsymbol{a}'=(a_1,\ldots,a_n)$, and $\overline{\boldsymbol{A}}$ consist of parameters to be estimated. The matrix $\overline{\boldsymbol{A}}$ is $(n+1)\times(n+1)$ and symmetric. The vector $\alpha'=(\alpha_1,\ldots,\alpha_n)$ is a fixed vector of nonnegatvie constants.[27] The division by $\alpha'\boldsymbol{x}$ in 22 makes H linearly homogeneous in y_0 and $\boldsymbol{x}$.

To conform with the partitioning of the vector $(y_0,\boldsymbol{x}')$, we partition the matrix $\overline{\boldsymbol{A}}$ as

$$\overline{\boldsymbol{A}} = \begin{bmatrix} A_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A} \end{bmatrix}$$

where A_{11} is a scalar, $\boldsymbol{A}_{12}$ is a $1\times n$ row vector,

[23]The flexibility here is sometimes called Diewert-flexible or second-order flexible. See Diewert (1971). The flexibility applies only locally. However, Gallant (1981, 1982) introduced the Fourier semi-nonparametric functional form, which can provide global flexibility asymptotically. Barnett, Geweke and Wolfe (1991) have developed the alternative seminonparametric asymptotically ideal model (AIM), which is globally flexible asymptotically and has advantages in terms of regularity.

[24]See Gallant and Golub (1984), Lau (1978) and Diewert and Wales (1987). However, if we can choose a model whose regularity region contains the data, then the regularity will be satisfied without imposing additional restrictions.

[25]See Blackorby, Primont and Russell (1977). Denny and Fuss (1977) propose a partial solution to avoid destroying flexibility. Their approach is to impose weak separability conditions at a point. However, local weak separability is not sufficient for the existence of a global aggregator function.

[26]Diewert and Wales (1987) alternatively also developed the generalized Barnett model. Although we have not used that model in this study, the generalized Barnett model has been applied to the analogous perfect-certainty case by Barnett and Hahm (1994). Regarding the merits of the generalized Barnett model in testing for weak separability, also see Blackorby, Schworm and Fisher (1986).

[27]We use the term "fixed constants" to designate constants that the researchers can select *a priori* and treat as constants during estimation.

A_{21} is an $n\times 1$ column vector, and $\bar{A}$ an $n\times n$ symmetric matrix. Since $\bar{A}$ is symmetric, it follows that $A_{12}=A'_{21}$.

Let $(y_0^*, x^*)\neq 0$ be the point about which the functional form is locally flexible. That point is selected by the researcher in advance, in a manner analogous to the selection of the point about which a Taylor series is expanded. Since the transformation function is assumed to be linearly homogeneous, the specification in the above form is not parsimonious, and hence, we further can restrict the model without losing the local flexibility property.[28] We therefore impose

$$(23)\ \alpha' x^* = 1,$$

$$(24)\ A_{11}y_0^* + A_{12}x^* = 0,$$

and

$$(25)\ A'_{12}y_0^* + Ax^* = 0_n,$$

where 0_n is an n-dimensional vector of zeros. Under 23, 24 and 25, it can be verified that the number of free parameters in equation 22 equals the minimum number of free parameters needed to maintain flexibility.

Solving 24 and 25 for A_{11} and A_{12}, we have

$$(26)\ A'_{12} = -Ax^*/y_0^*$$

and

$$(27)\ A_{11} = x^{*\prime}Ax^*/y_0^{*2}.$$

Substituting 26 and 27 into 22 yields

$$(28)\ H(y_0,x) = a_0 y_0 + a'x + \frac{1}{2}(\alpha' x)^{-1}x'Ax - (\alpha' x)^{-1}x^{*\prime}Ax(y_0/y_0^*) + \frac{1}{2}(\alpha' x)^{-1}x^{*\prime}Ax^*(y_0/y_0^*)^2.$$

Diewert and Wales (1987) have proved that $H(y_0,x)$, defined by equation 28, is flexible at (y_0^*, x^*).

In a similar way, we define $y_0(y)$ to be

$$(29)\ y_0(y) = b'y + \frac{1}{2}y'By/\beta' y,$$

with the parameters satisfying

$$(30)\ \beta' y^* = 1,$$

$$(31)\ y_0^* = b'y^*,$$

and

$$(32)\ By^* = 0_m,$$

where $b'=(b_1,\ldots,b_m)$, and the $m\times m$ symmetric matrix B consists of parameters to be estimated, $\beta'=(\beta_1,\ldots,\beta_m)$ is a fixed vector of nonnegative constants, and $y^*\neq 0$ is the point at which local flexibility of equation 29 is maintained.

Substituting 29 into 28, we get

$$\begin{aligned}(33)\ \Omega(y,x) &= H(y_0(y),x)\\ &= a_0\left(b'y+\frac{1}{2}(\beta' y)^{-1}y'By\right)+a'x\\ &+\frac{1}{2}(\alpha' x)^{-1}x'Ax\\ &-(y_0^*\alpha' x)^{-1}x^{*\prime}Ax\left(b'y+\frac{1}{2}(\beta' y)^{-1}y'By\right)\\ &+\frac{1}{2}(y_0^{*2}\alpha' x)^{-1}x^{*\prime}Ax^*(b'y)\\ &+\frac{1}{2}\left((\beta' y)^{-1}y'By\right)^2,\end{aligned}$$

which is a flexible functional form for $\Omega(y,x)$ and satisfies weak separability in outputs.

Neoclassical curvature conditions require $\Omega(y,x)$ and $y_0(y)$ to be convex functions, and neoclassical monotonicity requires $\partial\Omega/\partial y\geq 0$ and $\partial\Omega/\partial x\leq 0$. Diewert and Wales (1987), theorem (10) have shown that $H(y_0,x)$, defined by 28, and $y_0(y)$, defined by 29, are globally convex if and only if A and B are positive semidefinite.

[28] A flexible functional form is parsimonious if it has the minimum number of parameters needed to maintain flexibility. Diewert and Wales (1988) have acquired the minimum number of parameters needed to provide a second-order approximation to an arbitrary function. If a specification for an arbitrary function with n variables is flexible, it must have at least $1+n+n(n+1)/2$ independent parameters. In our case, the linear homogeneity imposes $1+n$ extra constraints on the first and second derivatives of H, so the minimum number of parameters needed to acquire flexibility is reduced by $1+n$.

For $\Omega(\boldsymbol{y},\boldsymbol{x})$ to be convex, we further need

$$(34)\quad \frac{\partial H(y_0,\boldsymbol{x})}{\partial y_0} \geq 0.$$

If 34 holds, then $\Omega(\boldsymbol{y},\boldsymbol{x})$ is globally convex in $(\boldsymbol{y},\boldsymbol{x})$, when $H(y_0,\boldsymbol{x})$ is convex in $(y_0,\boldsymbol{x})$ and $y_0(\boldsymbol{y})$ is convex in $\boldsymbol{y}$.[29]

If the unconstrained estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ are not positive semidefinite symmetric matrices, positive semidefiniteness can be imposed without destroying flexibility by the substitution

$$(35)\quad \boldsymbol{A} = \boldsymbol{qq'}$$

and

$$(36)\quad \boldsymbol{B} = \boldsymbol{uu'},$$

where $\boldsymbol{q}$ is a lower triangular $n \times n$ matrix and $\boldsymbol{u}$ is a lower triangular $m \times m$ matrix.[30] In estimation, we replace $\boldsymbol{A}$ and $\boldsymbol{B}$ by lower triangular matrixes $\boldsymbol{qq'}$ and $\boldsymbol{uu'}$, so that the function 33 is globally convex if 34 is true.

Monotonicity restrictions are difficult to impose globally. However, we can impose local monotonicity with simple restrictions. Differentiating 33 with respect to $(\boldsymbol{y},\boldsymbol{x})$, we get

$$\begin{aligned}(37)\quad \frac{\partial \Omega}{\partial \boldsymbol{y}} = {} & a_0\,[\boldsymbol{b}+\frac{1}{2}\,(2(\beta'\boldsymbol{y})^{-1}\boldsymbol{By}-(\beta'\boldsymbol{y})^{-2}\beta\boldsymbol{y}'\boldsymbol{By})] \\ & -\,(y_0^*\alpha'\boldsymbol{x})^{-1}\boldsymbol{x}^{*\prime}\boldsymbol{Ax}[\boldsymbol{b}+\frac{1}{2}\,(2(\beta'\boldsymbol{y})^{-1}\boldsymbol{By} \\ & -(\beta'\boldsymbol{y})^{-2}\beta\boldsymbol{y}'\boldsymbol{By})] \,+\, (y_0^{*2}\alpha'\boldsymbol{x})^{-1}\boldsymbol{x}^{*\prime}\boldsymbol{Ax}^* \\ & [\boldsymbol{b}+\frac{1}{2}\,(2(\beta'\boldsymbol{y})^{-1}\boldsymbol{By}-(\beta'\boldsymbol{y})^{-2}\beta\boldsymbol{y}'\boldsymbol{By})] \\ & (\boldsymbol{b}'\boldsymbol{y}+\frac{1}{2}\,(\beta'\boldsymbol{y})^{-1}\boldsymbol{y}'\boldsymbol{By})\end{aligned}$$

and

$$\begin{aligned}(38)\quad \frac{\partial \Omega}{\partial \boldsymbol{x}} = {} & \boldsymbol{a}+\frac{1}{2}\,[2(\alpha'\boldsymbol{x})^{-1}\boldsymbol{Ax}-(\alpha'\boldsymbol{x})^{-2}\alpha\boldsymbol{x}'\boldsymbol{Ax}] \\ & -[(y_0^*\,\alpha'\boldsymbol{x})^{-1}\boldsymbol{Ax}^*-(y_0^*\,\alpha'\boldsymbol{x})^{-2}y_0^*\alpha\boldsymbol{x}^{*\prime}\boldsymbol{Ax}] \\ & (\boldsymbol{b}'\boldsymbol{y}+\frac{1}{2}\,(\beta'\boldsymbol{y})^{-1}\boldsymbol{y}'\boldsymbol{By}) \\ & -\frac{1}{2}\,(y_0^{*2}\alpha'\boldsymbol{x})^{-2}y_0^{*2}\,\alpha\boldsymbol{x}^{*\prime}\boldsymbol{Ax}^*(\boldsymbol{b}'\boldsymbol{y} \\ & +\frac{1}{2}(\beta'\boldsymbol{y})^{-1}\boldsymbol{y}'\boldsymbol{By})^2.\end{aligned}$$

If we evaluate these derivatives at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$, we have

$$(39)\quad \frac{\partial \Omega}{\partial \boldsymbol{y}} = a_0\boldsymbol{b}$$

and

$$(40)\quad \frac{\partial \Omega}{\partial \boldsymbol{x}} = \boldsymbol{a}.$$

Applying the method of squaring technique, we impose on 39 and 40 the monotonicity conditions[31]

$$\begin{aligned}(41)\quad & \frac{\partial \Omega}{\partial \boldsymbol{y}}(\boldsymbol{y}^*,\boldsymbol{x}^*) = a_0\boldsymbol{b} \geq 0 \\ & \textit{and}\ \frac{\partial \Omega}{\partial \boldsymbol{x}}(\boldsymbol{y}^*,\boldsymbol{x}^*) = \boldsymbol{a} \leq 0.\end{aligned}$$

Equation 41 assures that the monotonicity conditions are satisfied locally at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$.

We have shown that the functional form defined by equation 33 and restricted to satisfy equations 23, 30-32, 34-36 and 41 is flexible, locally monotone, and globally convex, provided that the assumed weakly separable structure is true. Although we do not impose global monotonicity, we do check and confirm that monotonicity is satisfied at each observation within our data. In the following discussion, we will define a more general flexible functional form that does not require weak separability.

The number of independent parameters in equation 33 is

$$(42)\quad 1+n+\frac{n(n+1)}{2}+m-1+\frac{m(m\text{-}1)}{2}.$$

We know that the minimum number of parameters required to maintain flexibility for a linearly homogeneous function with $n+m$ variables is

$$(43)\quad 1+n+m+\frac{(n+m)\ (n+m+1)}{2} - (1+n+m).$$

Subtracting 42 from 43, we get $n(m-1)$, which is the number of additional parameters that must be introduced into equation 33 to acquire

[29]See Diewert and Wales (1991) for the proof.

[30]See Lau (1978) and Diewert and Wales (1987).

[31]See Lau (1978).

a flexible functional form for a general transformation function. Let

$$(44)\quad \Omega(\boldsymbol{y},\boldsymbol{x}) = H(y_0(\boldsymbol{y}),\boldsymbol{x}) + \boldsymbol{c}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{C}\boldsymbol{x} \,/\, (\gamma'\boldsymbol{y} + \lambda'\boldsymbol{x}),$$

where γ and λ are vectors of nonnegative fixed constants, the vector $\boldsymbol{c}'=(c_1,\ldots,c_m)$ and the $m\times n$ matrix $\boldsymbol{C}$ are new parameters to be estimated, and the division by $\gamma'\boldsymbol{y}+\lambda'\boldsymbol{x}$ makes Ω linearly homogeneous. Because of the linear homogeneity property, we have more free parameters than needed for flexibility and, hence, we can impose the following additional restrictions without losing local flexibility:

$$(45)\quad \gamma'\boldsymbol{y}^*+\lambda'\boldsymbol{x}^*=1,$$

$$(46)\quad \boldsymbol{c}'\boldsymbol{y}^*=0,$$

$$(47)\quad \boldsymbol{y}^{*\prime}\boldsymbol{C}=\boldsymbol{0}'_n,$$

and

$$(48)\quad \boldsymbol{C}\boldsymbol{x}^*=\boldsymbol{0}_m,$$

where $(\boldsymbol{y}^*,\boldsymbol{x}^*)$ is the point at which local flexibility is maintained. Under equations 45-48, the number of new free parameters added into 44 is exactly equal to $n(m-1)$. Diewert and Wales (1991) have proved that the function 44 is a flexible functional form at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$ for a general nonseparable transformation function.

Global convexity is difficult to impose in this case. However, we can derive the restrictions for local convexity at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$. Deriving the Hessian matrix of 44 and evaluating at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$, we have

$$(49)\quad \nabla^2\Omega(\boldsymbol{y}^*,\boldsymbol{x}^*) = \begin{bmatrix} a_0\boldsymbol{B}+\boldsymbol{b}\boldsymbol{b}'\boldsymbol{x}^{*\prime}\boldsymbol{A}\boldsymbol{x}^*/y_0^{*2} & \boldsymbol{C}-\boldsymbol{b}\boldsymbol{x}^{*\prime}\boldsymbol{A}/y_0^* \\ \boldsymbol{C}'-\boldsymbol{A}\boldsymbol{x}^*\boldsymbol{b}'/y_0^* & \boldsymbol{A} \end{bmatrix}.$$

If $\nabla^2\Omega(\boldsymbol{y}^*,\boldsymbol{x}^*)$ is positive semidefinite, then $\Omega(\boldsymbol{y}^*,\boldsymbol{x}^*)$ is convex at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$. Let

$$(50)\quad \boldsymbol{A} = \boldsymbol{q}\boldsymbol{q}',$$

$$(51)\quad \boldsymbol{C} = \boldsymbol{v}\boldsymbol{q}',$$

and

$$(52)\quad \boldsymbol{B} = a_0^{-1}[\boldsymbol{v}\boldsymbol{v}'+\boldsymbol{u}\boldsymbol{u}'],$$

where $\boldsymbol{q}$ and $\boldsymbol{u}$ are lower triangular matrices introduced for reasons described above, and $\boldsymbol{v}$ is an unrestricted $m\times n$ matrix. Then $\nabla^2\Omega(\boldsymbol{y}^*,\boldsymbol{x}^*)$ is a positive semidefinite symmetric matrix.[32]

Using 50-52, we rewrite 47, 48 and 32 as

$$(53)\quad \boldsymbol{y}^{*\prime}\boldsymbol{v}=\boldsymbol{0}'_n,$$

$$(54)\quad \boldsymbol{v}(\boldsymbol{q}'\boldsymbol{x}^*) = \boldsymbol{0}_m,$$

and

$$(55)\quad \boldsymbol{u}'\boldsymbol{y}^* = \boldsymbol{0}_m.$$

The function defined by 44 and satisfying 23, 30-31, 45-46 and 50-55 is a flexible functional form for a general transformation function at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$. In addition, local convexity is satisfied.

We now turn to imposing local monotonicity. Differentiating 44 with respect to $(\boldsymbol{y},\boldsymbol{x})$ and evaluating at $(\boldsymbol{y}^*,\boldsymbol{x}^*)$, we have

$$(56)\quad \frac{\partial\Omega}{\partial\boldsymbol{y}} = a_0\boldsymbol{b} + \boldsymbol{c}$$

and

$$(57)\quad \frac{\partial\Omega}{\partial\boldsymbol{x}} = \boldsymbol{a}.$$

As above, we use the method of squaring to impose nonnegativity on 56 and nonpositivity on 57. The estimated results then satisfy local monotonicity.

Comparing 33 with 44, we see that weak separability of outputs in 44 is equivalent to:

$$(58)\quad H_0{:}\boldsymbol{c} = \boldsymbol{0}_m \textit{ and } \boldsymbol{v}_{m\times n} = \boldsymbol{0}_{m\times n}.$$

Note that under the null hypothesis, H_0, equation 44 reduces to 33. Hence, $\boldsymbol{y}$ is weakly separable from $\boldsymbol{x}$ if and only if H_0 is true.

We have derived two flexible functional forms with appropriate regularity properties. One structure holds in the general case and the other under the null hypothesis of weak separability. We now are prepared to test weak separability and to estimate the parameters of the transformation function. The basic tool is Hansen and Singleton's generalized method of moments (GMM) estimator.

[32] See Lau (1978) and Diewert and Wales (1991).

Substituting the functional form given by either 33 or 44 into the Euler equations 16 and 17, we obtain our structural model, which consists of a system of integral equations. A closed form solution to such Euler equations rarely exists. However, GMM permits estimating nonlinear rational expectations models defined in terms of Euler equations. Hansen (1982) has proved that under very weak conditions, the GMM estimates are consistent and asymptotically normally distributed.[33]

In the GMM framework, there are two methods of testing hypotheses.[34] The first approach applies Hansen's asymptotic χ^2 statistic to test for no overidentifying restrictions. We impose the weak separability restrictions 58 on the flexible functional form 44, estimate the restricted system, and then run Hansen's test for no overidentifying restrictions. Since 44 reduces to 33 after imposing the weak separability restrictions, we can substitute equation 33 itself directly into the Euler equations to impose the null for testing. If the test of no overidentifying restrictions is rejected, then the restrictions imposed under the null hypothesis are rejected, where in our case the null is the weakly separable structure imposed on the transformation function.

The second approach to hypothesis testing with GMM is based on the asymptotically normal distribution of the GMM parameter estimators. Let θ be the vector of parameters to be estimated in equation 44. Then the GMM estimator $\hat{\theta}$ has an asymptotically normal distribution with mean θ and covariance matrix Σ.

Let τ be an $[n(m-1)]\times 1$ vector which contains all $n(m-1)$ independent parameters in the vector $\boldsymbol{c}$ and the matrix $\boldsymbol{v}$. The hypothesis of weak separability can be rewritten now as $\tau=\mathbf{0}$ or equivalently as a set of linear restrictions of the form

$$(59)\quad \boldsymbol{S}\theta=\tau=\mathbf{0},$$

where $\boldsymbol{S}$ is an $[n(m-1)]\times[(n+m+1)/2]$ matrix whose elements are all zeros and ones.

From the known asymptotic distribution of $\hat{\theta}$, we have

$$(60)\quad \sqrt{T}\,(\boldsymbol{S}\hat{\theta}-\boldsymbol{S}\theta) \overset{a}{\sim} N(\mathbf{0},\boldsymbol{S}\Sigma\boldsymbol{S}'),$$

where T is the number of observations. Under the null hypothesis, H_0: $\boldsymbol{S}\theta = 0$, we have

$$\sqrt{T}\,\hat{\tau} \overset{a}{\sim} N(\mathbf{0},\boldsymbol{S}\Sigma\boldsymbol{S}'),$$

where $\hat{\tau} = \boldsymbol{S}\hat{\theta}$. We obtain the following χ^2 statistic

$$(61)\quad \boldsymbol{\Phi} = (\sqrt{T}\,\hat{\tau})'[\boldsymbol{S}\Sigma\boldsymbol{S}']^{-1}(\sqrt{T}\,\hat{\tau})$$

$$= T\hat{\tau}'[\boldsymbol{S}\Sigma\boldsymbol{S}']^{-1}\hat{\tau} \overset{a}{\sim} \chi^2_{n(m-1)}\,.$$

Although Σ is unknown, we can replace it by a consistent estimate without changing the asymptotic results. The test is one sided, with the null of separability rejected if $\boldsymbol{\Phi}$ is large.

EMPIRICAL APPLICATION

Barnett and Hahm (1994), and Hancock (1985, 1987, 1991) have analyzed monetary service production by the banking industry in detail, under the assumptions of perfect certainty and neoclassical joint production. The balance sheet of a bank consists of fund-providing functions and fund-using functions. The fund-providing functions include demand deposits, time deposits and nondeposit funds.[35] The fund-using functions include investment, real estate mortgage loans, installment loans, credit card loans and industrial loans. In our theoretical model, the sources of funds are the firm's borrowed funds, and the uses of funds are the firm's portfolio. The total available funds on the balance sheet are total assets minus premises and other assets.

On the average, demand deposits and time deposits account for over 85 percent of total available funds. The equity capital included in the non-deposit funds can be treated as a fixed factor that does not enter the variable profit

[33] Hansen (1982), Hansen and Singleton (1982), and Newey and West (1987) provide a detailed discussion of GMM estimation.

[34] See Mackinlay and Richardson (1991).

[35] Demand deposits consist of checking accounts, official checks, money orders, treasury tax accounts and loan accounts. Time deposits consist of regular savings, money market deposit accounts, other time accounts, retirement accounts, and certificates of deposit under $100,000. Non-deposit funds consist of equity capital, federal funds purchased, borrowed money, capital notes and debentures, time deposits of $100,000 and over, other money market instruments, and other liabilities.

function.[36] For these reasons, we only choose demand deposits and time deposits as borrowed funds in our model. Turning to inputs, excess reserves are total cash balances minus required reserves. Other real resource inputs are labor, materials and capital.[37] Capital is treated as fixed, and we include only variable factors in the transformation function. An obvious direction for possible future extension of this research would be the incorporation of some or all capital as variable factors to produce inferences applicable to a longer run perspective than that implicit in our definition of variable and fixed factors.

Using equations 16 and 17, the Euler equations are

$$(62)\quad E_t \{P_t k_{1t} (\frac{h}{1-\rho}\pi_t + d)^{\rho-1} + P_t \frac{1}{1+\mu}[(1+R_t)\,(1-k_{1t}) - (1+h_{1t}) + R_t \frac{\partial\Omega/\partial D_t}{\partial\Omega/\partial C_t}]\,(\frac{h}{1-\rho}\pi_{t+1} + d)^{\rho-1}\} = 0,$$

$$(63)\quad E_t \{P_t k_{2t} (\frac{h}{1-\rho}\pi_t + d)^{\rho-1} + P_t \frac{1}{1+\mu}[(1+R_t)\,(1-k_{2t}) - (1+h_{2t}) + R_t \frac{\partial\Omega/\partial T_t}{\partial\Omega/\partial C_t}]\,(\frac{h}{1-\rho}\pi_{t+1} + d)^{\rho-1}\} = 0,$$

$$(64)\quad E_t \{[P_t R_t \frac{\partial\Omega/\partial L_t}{\partial\Omega/\partial C_t} - (1+R_t)w_{1t}]\,(\frac{h}{1-\rho}\pi_{t+1} + d)^{\rho-1}\} = 0,$$

and

$$(65)\quad \mathrm{E}_t \{[P_t R_t \frac{\partial\Omega/\partial M_t}{\partial\Omega/\partial C_t} - (1+R_t)w_{2t}]\,(\frac{h}{1-\rho}\pi_{t+1} + 1) + d)^{\rho-1}\} = 0,$$

where D_t is demand deposits, T_t is time deposits, L_t is labor input, M_t is materials input, and w_{1t} and w_{2t} are the prices of labor and materials respectively.

Using the notations in section three, we can write

$\boldsymbol{y}' = (D_t, T_t)$ and $\boldsymbol{x}' = (C_t, L_t, M_t)$.

If the weakly separable structure of the transformation function is true, then equation 33 is the transformation function. As discussed in section three, the weak separability hypothesis can be tested by applying Hansen's χ^2 statistic.

The derivatives of Ω with respect to its arguments are given by equations 37 and 38. The fixed constants and the center of the local approximation need to be selected before estimation. We choose

$y_0^* = 1$, $\boldsymbol{y}^{*\prime} = (1,1)$, *and* $\boldsymbol{x}^{*\prime} = (1,1,1)$

as the center of approximation. To locate that center within the interior of the observations, we rescale the data about the midpoint observation

$$(66)\quad \tilde{x}_i^t = x_i^t \,/\, x_i^{t^*} \quad \forall\ i = 1,\ 2,\ 3 \ \text{and}$$

$$\tilde{y}_i^t = y_i^t \,/\, y_i^{t^*} \quad \forall\ i = 1,\ 2,$$

where t^* represents the midpoint observation.[38] We correspondingly rescale each price by multiplication by the midpoint observation. That rescaling of prices keeps dollar expenditures on each good unaffected by the rescaling of its quantity.

We select the fixed nonnegative constants α_i and β_i such that

$$(67)\quad \alpha_i = |\bar{\tilde{x}}_i| / \sum_{j=1}^{3} |\bar{\tilde{x}}_j| \qquad \forall\ i = 1,\ 2,\ 3$$

and

$$(68)\quad \beta_i = |\bar{\tilde{y}}_i| / \sum_{j=1}^{2} |\bar{\tilde{y}}_j| \qquad \forall\ i = 1,\ 2,$$

where $\bar{\tilde{x}}$ and $\bar{\tilde{y}}$ are the sample means of $\tilde{x}$ and $\tilde{y}$ respectively. Note that α_i and β_i satisfy equations 23 and 30, as is required. With our data sam-

[36]See Barnett (1987). Equity capital includes preferred and common stocks, surplus, undivided profits and reserves, and valuation reserves.

[37]Labor includes managerial labor and nonmanagerial labor. Materials include stationery, printing and supplies, telephone, telegraph, postage, freight and delivery.

[38]The data point at which all quantities are set to unity can be arbitrary.

ple, we find $\alpha_1=0.33$, $\alpha_2=0.35$, $\alpha_3=0.32$, $\beta_1=0.58$, and $\beta_2=0.42$.

Before estimating the independent parameters, we need only impose the inequality restrictions. Equation 31 implies $b_2=1-b_1$, and the monotonicity condition (41) requires $b_i\geq 0$. Hence, it also follows that $b_i\leq 1$. Combining these conditions, we can replace b_1 and b_2 by

$$(69)\ b_1 = \sin^2(\xi) \text{ and } b_2 = \cos^2(\xi)$$

and estimate ξ. Since $\Omega(\boldsymbol{y},\boldsymbol{x})=0$, we also normalize $a_0=1$.

The monotonicity condition 41 requires $a_i\leq 0$, which we impose by replacing a_i by $-\tilde{a}_i^2\ \forall\ i=1,2,3$, where $\tilde{a}_i\ \forall\ i=1,2,3$, are the new parameters to be estimated. The convexity conditions are imposed by replacing $\boldsymbol{A}$ and $\boldsymbol{B}$ by the lower triangular matrices $\boldsymbol{qq'}$ and $\boldsymbol{uu'}$ respectively, where $\boldsymbol{q}$ and $\boldsymbol{u}$ are

$$\boldsymbol{q} = \begin{bmatrix} q_{11} & 0 & 0 \\ q_{21} & q_{22} & 0 \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

and

$$\boldsymbol{u} = \begin{bmatrix} u_{11} & 0 \\ u_{21} & u_{22} \end{bmatrix}.$$

Equation 32 implies

$$(70)\ \begin{bmatrix} u_{11} & 0 \\ u_{21} & u_{22} \end{bmatrix}\begin{bmatrix} u_{11} & u_{21} \\ 0 & u_{22} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Solving 70, we get $u_{21} = -u_{11}$ and $u_{22} = 0$. Substituting them into equation 36, we have

$$(71)\ \boldsymbol{B} = u_{11}^2\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

The above discussion identifies all the independent parameters to be estimated in the specification of the transformation function. They are ξ, u_{11}, the lower triangular matrix $\boldsymbol{q}$, and the vector $\tilde{\boldsymbol{a}}'=(\tilde{a}_1,\tilde{a}_2,\tilde{a}_3)$.

The primary data source is the Federal Reserve's Functional Cost Analysis (FCA).[39] We got our data from the Federal Reserve Bank of St. Louis. The data used are the National Average FCA Report, which contains annual data from 1966 to 1990. Hence, there are a total of 25 observations in our annual data. Monthly data is not available from the FCA. From the FCA, we acquired banks' portfolio rate of return, the net interest rates on demand deposits and time deposits, and the nominal quantity of demand deposits, time deposits and cash balances.[40] The prices and quantities of labor and materials are aggregate producer prices and quantity indexes from the data in the FCA Report and the *Survey of Current Business.*[41] The required reserve ratio is from the *Federal Reserve Bulletin.* The implicit price deflator is the implicit GNP deflator from the Citibank data base. We deflate the nominal dollar balances of all financial goods to convert them into real balances.

EMPIRICAL RESULTS

We use the GMM estimator in the TSP mainframe version (version 4.2) to estimate our model. In the disturbances we allow for conditional heteroskedasticity and second-order moving average serial correlation. Using the spectral density kernels in TSP, our estimated results are robust to heteroskedasticity, autocorrelation and positive semidefinite weighting matrix. To use the GMM method, instrumental variables must be selected. We choose as instruments the constant, the federal funds rate, the discount window rate, the composite bond rate (maturities over 10 years), the holding cost of demand deposits and time deposits, the lagged banks' portfolio rate of return, excess cash reserves, and capital. In estimation, we replace h by h^2 to impose nonnegativity of the resulting h^2. That nonnegativity is needed for regularity in the definition of the HARA class.

The GMM parameter estimates, subject to imposition of weak separability of outputs from inputs, are reported in Table 1. All three parameters in the utility function are statistically

[39] The Functional Cost Analysis program is a cooperative venture between the Federal Reserve Banks and the participating banks. This program is designed to assist a participating bank in increasing overall bank earnings as well as to improve the operational efficiency of each bank function.

[40] The net interest rate equals the interest paid minus service charges earned plus FDIC insurance premiums paid.

[41] See Barnett and Hahm (1994) for a detailed discussion about the aggregation of labor and material.

Table 1

GMM Estimates Using the HARA Utility Function with Weak Separability in Outputs Imposed

Parameter	Estimate	Standard error	t-Statistic
h^2	0.003	0.122	0.024
$\rho-1$	2.330	25.625	0.091
d	0.001	0.044	0.012
$\mu+1$	1.090	0.165	6.602
ξ	58.382	0.201	290.459
u_{11}	0.232	0.418	0.555
q_{11}	0.186	0.078	2.372
q_{21}	0.418	0.106	3.931
q_{31}	0.105	0.048	2.178
q_{22}	0.477	0.101	4.725
q_{32}	0.120	0.162	0.743
q_{33}	0.116	0.505	0.230
$\tilde{a}_1$	0.323	0.035	9.117
$\tilde{a}_2$	0.436	0.058	7.523
$\tilde{a}_3$	0.280	0.038	7.448

insignificant at the 5 percent level. As a result of the very low precision of those three parameter estimates, it is clear that we have introduced risk aversion in a manner incorporating too many parameters for the available sample size. Hence, we need to restrict HARA to one of its less deeply parameterized special cases. As observed in the second section, the HARA class reduces to the popular power (CRRA isoelastic) utility function. We now test whether that popular special case is accepted.

Equation (61) in the third section provides a statistic to test that a set of parameters is jointly equal to zeros. When the set of parameters includes only one element, the χ^2 test statistic Φ, given by equation (61), equals the number of observations multiplied by the square of the t-statistic of that parameter. We calculated that $\Phi=0.0033$, while the critical value is 6.635 at the one percent significance level. Hence, we cannot reject $d=0$, and the power utility function is accepted. We reestimate the model using that specification.

To impose the inequality restriction $0 < \rho < 1$, which is sufficient for regularity of the power utility function special case, we replace ρ by $\sin^2(\tilde{\rho})$ and estimate $\tilde{\rho}$. In addition, to prevent the implausible possibility of a negative subjective rate of time discount, we replace μ by $\tilde{\mu}^2$ and estimate $\tilde{\mu}$.[42] The estimated results, subject to imposition of weak separability of outputs from inputs, are reported in Table 2.[43] All parameters are significantly different from zero at the 5 percent level except for $\tilde{\mu}$, u_{11}, *and* q_{33}. Monotonicity is necessarily satisfied at $(\boldsymbol{y}^*, \boldsymbol{x}^*)$, since local monotonicity was imposed at that point. We use the estimated parameters to determine whether monotonicity is satisfied elsewhere in the sample. Substituting the estimated parameters into equations 37 and 38, we find that $\partial\Omega/\partial\boldsymbol{y}>0$ and $\partial\Omega/\partial\boldsymbol{x}<0$ everywhere in the sample. Hence, no violations of monotonicity occurred within the sample. Regarding curvature, we have imposed global convexity on $H(y_0,\boldsymbol{x})$ and $y_0(\boldsymbol{y})$. To verify global convexity of $\Omega(\boldsymbol{y},\boldsymbol{x})$, we must check equation 34 at each data point.

[42]Actually only the upper bound imposed on ρ is required by theory. Hence, if we had found that the lower bound implied by our substitution was binding, we would have switched to the more sophisticated substitution of 2−cosh $(\tilde{\rho})$ in place of ρ. But in practice our estimate of ρ was strictly positive, so we did not have to resort to the introduction of hyperbolic functions. Furthermore, our imposition of nonnegativity on μ was equally as harmless, since no corner solutions were acquired on that inequality restriction either. In fact, in the HARA case, we did not impose nonnegativity on μ at all, since we got nonnegativity from our estimates without the need to impose it, and in retrospect it is evident that we could have done the same in the power utility case.

[43]The instrumental variables are the constant, the federal funds rate, the discount window rate, the composite bond rate (over 10 years), the three-month T-bill rate, the yields on demand deposits and time deposits, the lagged bank's portfolio rate of return, and capital.

Table 2

GMM Estimates Using the Power Utility Function with Weak Separability in Outputs Imposed

Parameter	Estimate	Standard error	t-Statistic
$\tilde{\rho}$	−524.629	9.410	−55.754
$\tilde{\mu}$	0.351	0.187	1.877
ξ	60.692	0.019	3122.720
u_{11}	0.171	0.283	0.605
q_{11}	0.240	0.050	4.821
q_{21}	0.461	0.077	5.980
q_{31}	0.103	0.018	5.908
q_{22}	0.418	0.047	8.958
q_{32}	0.093	0.029	3.147
q_{33}	−0.025	0.412	−0.062
$\tilde{a}_1$	0.330	0.031	10.762
$\tilde{a}_2$	0.482	0.045	10.607
$\tilde{a}_3$	0.217	0.020	10.836

Differentiating $H(y_0,\boldsymbol{x})$ with respect to y_0, we get

$$(72)\quad \frac{\partial H(y_0,\boldsymbol{x})}{\partial y_0} = a_0-(\alpha'\boldsymbol{x})^{-1}\boldsymbol{x}^{*\prime}\,\boldsymbol{A}\boldsymbol{x}/y_0^{*} + (\alpha'\boldsymbol{x})^{-1}\boldsymbol{x}^{*\prime}\boldsymbol{A}\boldsymbol{x}^{*}y_0/y_0^{*2},$$

where y_0 is given by equation 29. Substituting the estimated parameters into equation 72, we find that $\partial H(y_0,\boldsymbol{x})/\partial y_0>0$ at every data point. Convexity of Ω is satisfied throughout the sample.

The weak separability hypothesis is tested by using Hansen's χ^2 test for no overidentifying restrictions. His test statistic is

$$(73)\quad \Phi = TQ\sim\chi^2_{e-f},$$

where T is the number of observations, Q is the value of the objective function, e is the number of orthogonal conditions, and f is the number of parameters estimated.[44] The calculated statistic is 27.6, while the critical value is 41.64 at the 1 percent significance level. We cannot reject the weak separability hypothesis. Hence, the existence of a theoretical monetary aggregate over the outputs produced by banks is accepted.

Substituting the parameter estimate of ξ from Table 2 into equation 69, we obtain $b_1 = 0.76$ and $b_2 = 0.24$. The estimated theoretical aggregate then is acquired by substituting the estimated parameters and fixed constants into equation 29 to get

$$(74)\quad y_0(D_t,T_t)=0.76D_t+0.24T_t+\frac{1}{2}\left[\frac{.17^2(D_t-T_t)^2}{.58D_t+0.42T_t}\right].$$

It is important to recognize that this aggregator function should not be used for forecasting or simulation outside the region of the data, and hence its usefulness is limited to research within the sample. While we have confirmed monotonicity within the region of the data, this aggregator function is not globally regular for all possible nonnegative values of the variables outside that region.

Having our econometrically estimated theoretical supply-side monetary aggregate, we now proceed to investigate whether any of the well known nonparametric statistical index numbers can track the estimated exact aggregate adequately. By converting from $\tilde{\rho}$ back to ρ and then computing the degree of relative risk aversion, $1-\rho$, we find that the degree of relative risk aversion is $1-.07=.93$. Since risk neutrality occurs only for zero values of relative risk aversion, we do not have risk neutrality. But there is no currently available theory regarding the tracking ability of nonparametric statistical index numbers when risk aversion exists. Hence, our only method of investigating the tracking

[44]The value of the objective function is defined as $Q = g_N(\hat{\theta})'\hat{W}_N g_N(\hat{\theta})$, where $g_N(\hat{\theta})$ is the sample mean of the moment conditions and $\hat{W}_N$ is the weighting matrix that defines the metric in making $g_N(\hat{\theta})$ close to zero in the GMM estimation procedure.

ability of the more easily computed nonparametric statistical indexes is to estimate the exact index econometrically, as we just have done, and compare its behavior with that of the statistical index numbers.

In this paper, we compare the estimated theoretical aggregate with the Divisia, simple-sum and CE indexes. Rotemberg, Driscoll and Poterba (1991) have found that the growth rate of the CE index is very volatile with monthly data. Hence, they have proposed (see their footnote 11) a method of smoothing that index's growth rates by replacing the index's weights by 13-month, centered moving averages. Since we are using annual data, there already is a form of smoothing implicit in the data construction. Nevertheless, in addition to computing the annual contemporaneous CE index, we compute the smoothed index in accordance with the method selected by Rotemberg, Driscoll and Poterba.

To parallel the 13-month centered moving-average smoothing as closely as possible with annual data, we use a three-year centered moving average. In a sense, our results with unsmoothed annual data slightly undersmooths relative to Rotemberg, Driscoll and Poterba's method, while the three-year centered moving average oversmooths relative to Rotemberg, Driscoll and Poterba's method. Nevertheless, as we shall see, the CE index's growth rate remains too volatile. A centered moving average is not defined at the start and end of a sample. Hence, a special method is needed to phase in the centered moving average at the start of the sample and phase it out at the end of the period. For that purpose, we use the procedure advocated by Rotemberg, Driscoll and Poterba. Figure 1 contains plots of the levels of all those aggregates. Figure 2 contains plots of their growth rates. We also separately plot the growth rate of each of the four statistical index numbers (simple sum, Divisia, unsmoothed CE and smoothed CE), with the growth rate of the estimated theoretical path superimposed. These plots are given in Figures 3, 4, 5 and 6.

While no econometric estimation is needed to compute the Divisia index, it is important on the supply side to incorporate the required reserves implicit tax into the user cost formula, when computing the Divisia output index. The user-cost formula is needed to compute the prices of monetary services, since the Divisia quantity index is a function of prices as well as quantities. On that subject, also see Barnett and Hahm (1994), Barnett, Hinich and Weber (1986), Hancock (1985, 1987, 1991) and Barnett (1987), who derive and supply the user cost of supplied monetary services, when required reserves yield no interest. The resulting real user-cost price for account type i is

$$(75)\quad \Phi_{ift} = \frac{(1-k_{it})\,R_t - r_{it}}{1+R_t}$$

$$(76)\quad \phantom{\Phi_{ift}} = \Phi_{ict} - \frac{k_{it}R_t}{1+R_t},$$

where r_{it} is the own rate of return defined in footnote 8, and where

$$(77)\quad \Phi_{ict} = \frac{R_t - r_{it}}{1+R_t}.$$

The nominal user cost is $P_t\,\Phi_{ift}$. The second term on the right-hand side of equation 76 is the discounted implicit tax on banks resulting from the nonpayment of interest on required reserves. Equation 77 is the same form as the user-cost price paid on the demand side by depositors, where R_t is the benchmark yield on a pure investment asset producing no services other than its own yield, so that equation 77 is the discounted foregone interest given up by the depositor in return for the services provided by asset type i.

Clearly the Divisia index tracks the theoretical aggregate more accurately than any of the other two indexes. The smoothed and unsmoothed CE index's level paths are almost identical to each other, as shown in Figure 1, despite the improvement in the performance of the CE index's growth rate plot after smoothing. Before 1972, the Divisia and estimated theoretical index are almost identical. After 1972, a small gap opens between them.

The CE index almost always underestimates the theoretical aggregate throughtout the sample period, with the gap growing to be larger after 1980. The simple-sum index always overestimates the theoretical aggregate, with the gap growing to be large and remaining large after only a few years. In terms of levels, the tracking error of the CE index is smaller than that of the simple-sum index, especially early in the same period. However, the CE index is much more volatile than the theoretical aggregate, especially from 1979 to 1983. Comparing Figures 5 and 6, we see that the CE index with smoothed weights is less volatile than the unsmoothed in-

Figure 1

Levels of Five Monetary Aggregates (parameters of theoretical monetary aggregate estimated with risk aversion permitted)

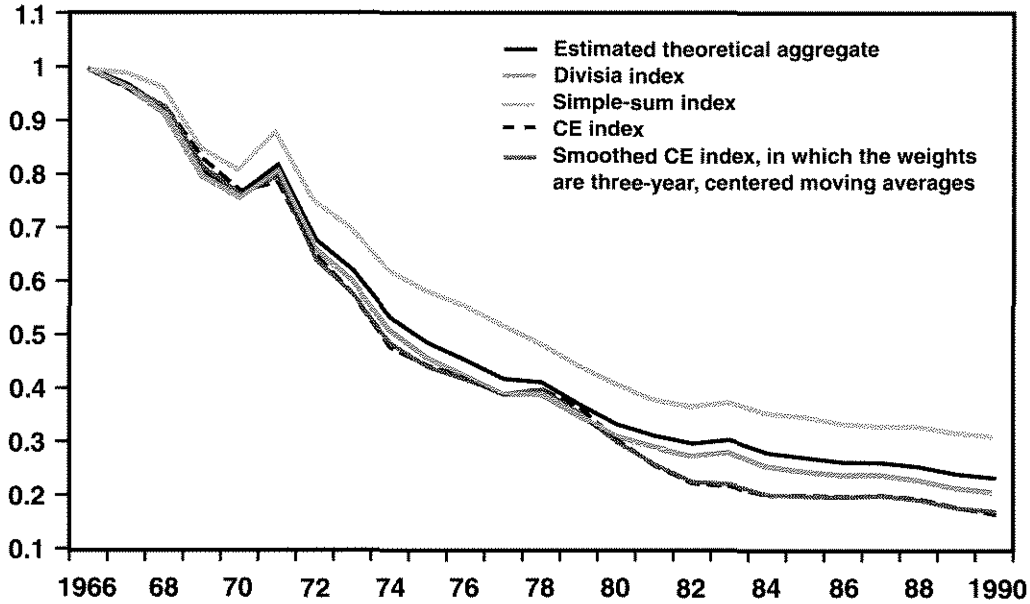


Figure 2

Growth Rates of Five Monetary Aggregates (parameters of theoretical monetary aggregate estimated with risk aversion permitted)

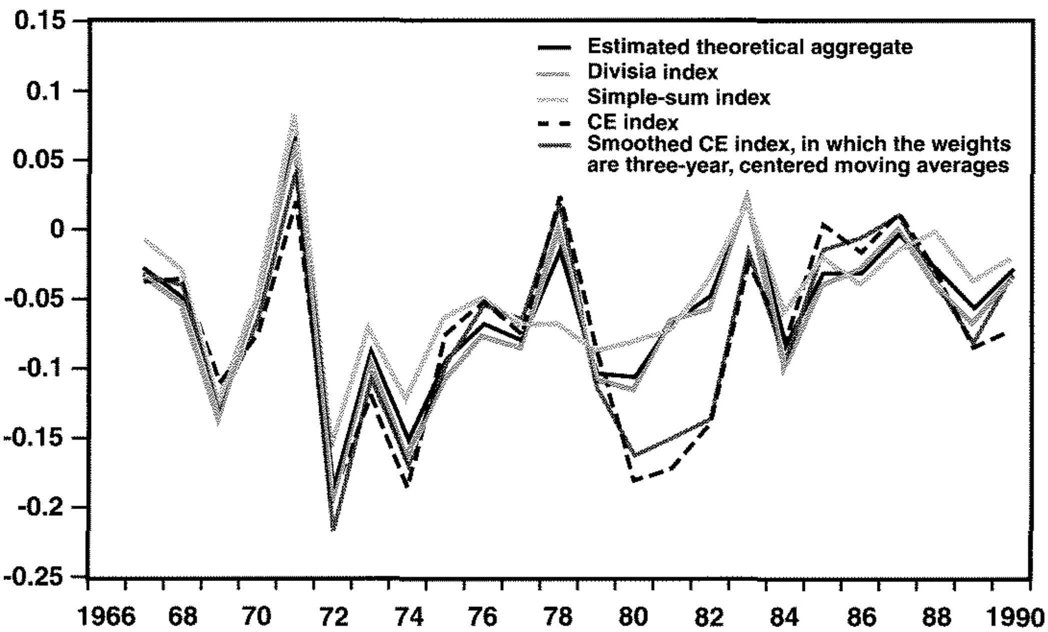


Figure 3

Growth Rates of Theoretical Monetary Aggregate and Divisia Index (with risk aversion permitted)

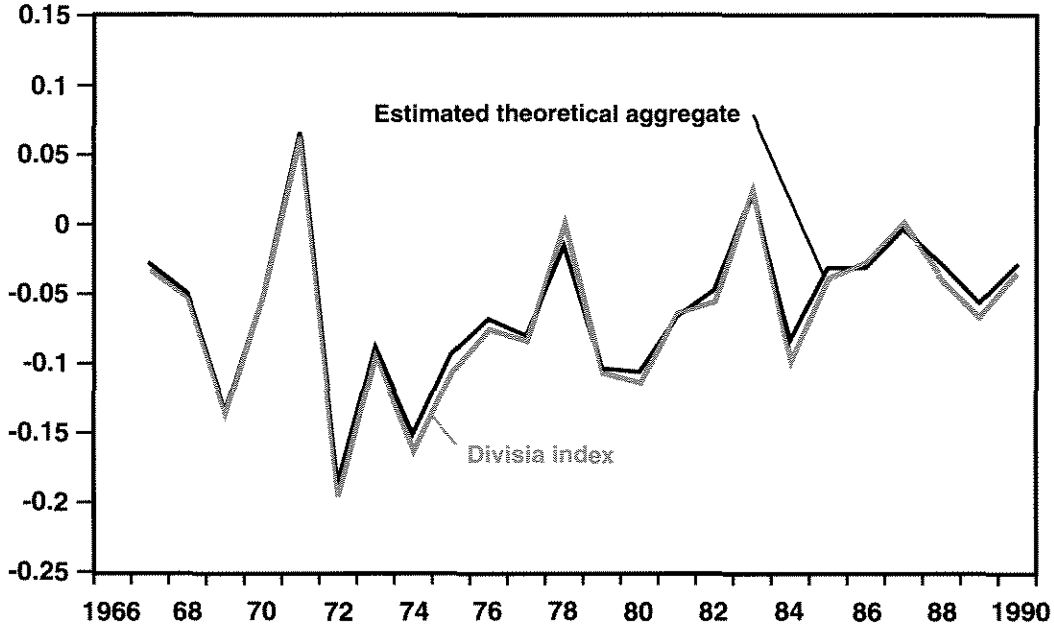


Figure 4

Growth Rates of Theoretical Monetary Aggregate and Simple-Sum Index (with risk aversion permitted)

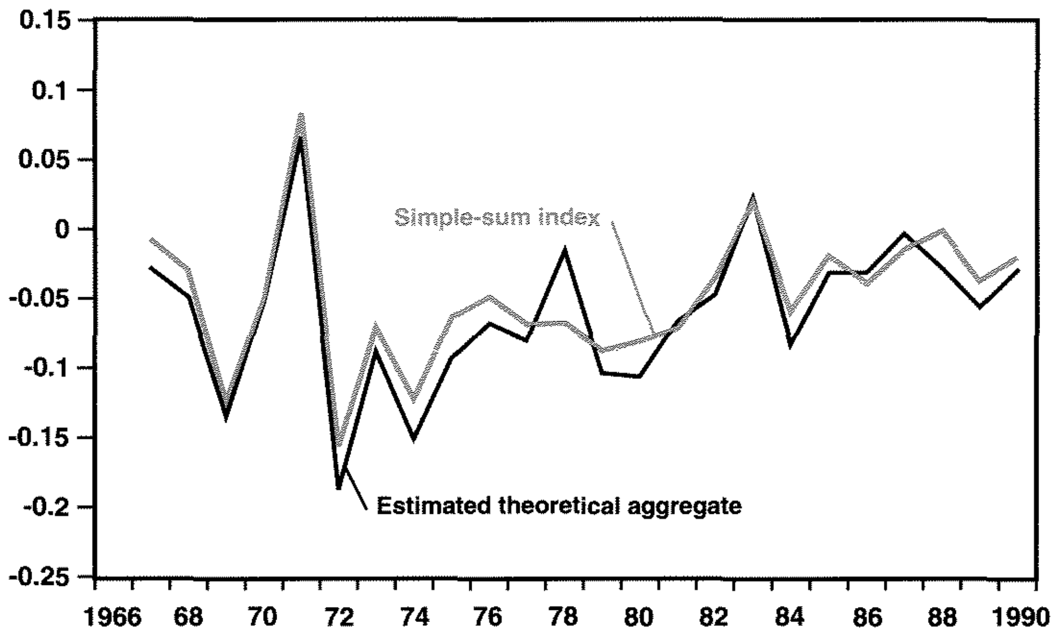


Figure 5

Growth Rates of Theoretical Monetary Aggregate and CE Index (with risk aversion permitted)

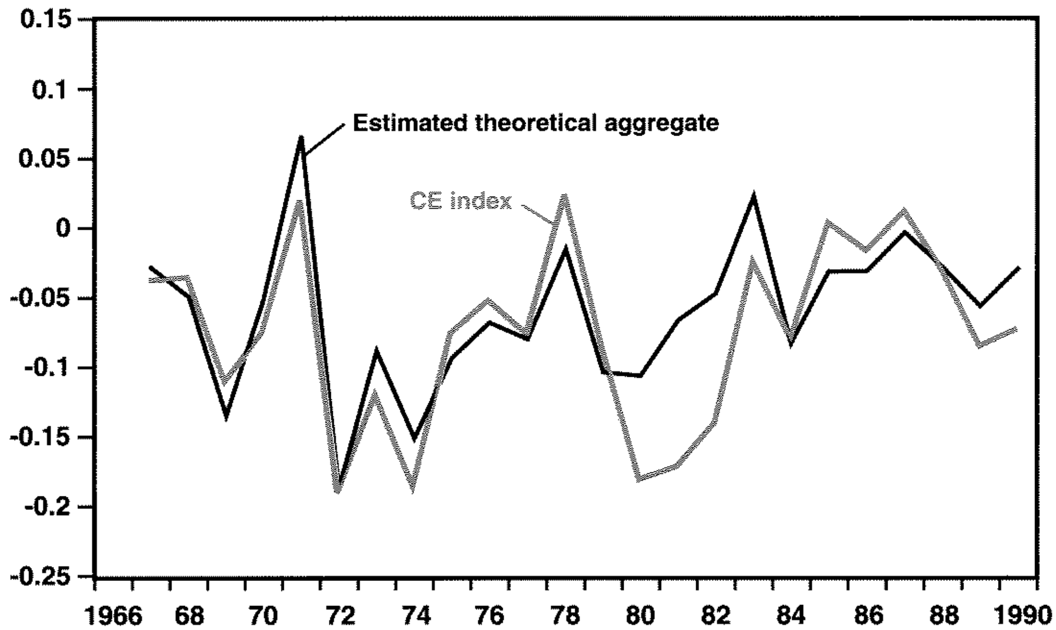


Figure 6

Growth Rates of Theoretical Monetary Aggregate and Smoothed CE Index (with risk aversion permitted)

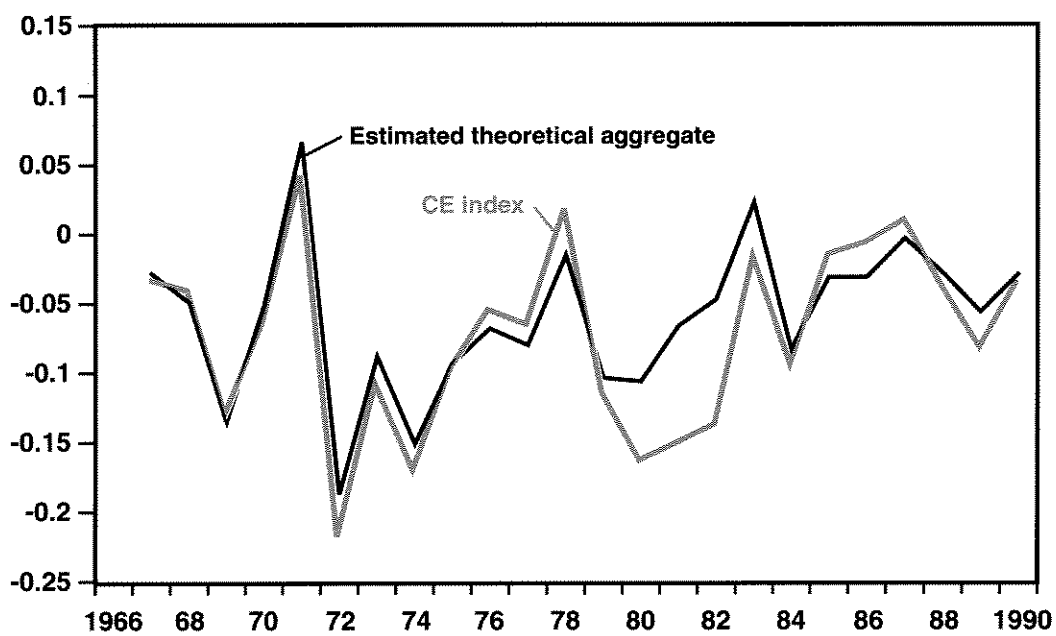


Table 3
GMM Estimates with Weak Separability in Outputs and Risk Neutrality Imposed

Parameter	Estimate	Standard error	*t*-Statistic
ξ	61.82	0.005	11968.80
u_{11}	0.27	0.019	14.31
q_{11}	0.18	0.005	32.78
q_{21}	0.38	0.022	17.01
q_{31}	0.07	0.007	10.14
q_{22}	0.44	0.023	19.09
q_{32}	0.11	0.063	1.68
q_{33}	0.16	0.132	1.25
$\bar{a}_1$	0.33	0.002	162.74
$\bar{a}_2$	0.50	0.003	164.39
$\bar{a}_3$	0.23	0.006	36.50

dex, but the volatility still remains larger than that of the estimated theoretical index. We could experiment with even more smoothing of the CE index than is advocated by Rotemberg, Driscoll, and Poterba, but we feel that further experimentation in that direction would produce an index having dynamics determined more by the ad hoc method of smoothing than by the theory that produces the index. Furthermore, we suspect that smoothing adequate to fix the index between 1979 and 1983 would oversmooth elsewhere. Hence, it seems that there is no way that the CE index can track the growth rates adequately throughout the sample.

In short, as a measure of the level of the money stock, the simple-sum index performs most poorly, while in terms of growth rates, the CE index performs most poorly. In both cases, the Divisia index performs best. These results are in the accordance with index number theory, although most of that theory is available in rigorous form only under the assumption of perfect certainty. Our weak separability test supports the existence of an inside-money output aggregate in banking, and our plots support the use of the Divisia index as the best currently available statistical index for tracking that output aggregate.

For comparison purposes, we repeat the above estimation and testing in the special case of risk neutrality. The Euler equations, 62-65, under risk neutrality become[45]

$$(78)\quad E_t\{P_t\frac{R_t(1-k_{1t})-r_{1t}}{1+R_t}+P_t\frac{R_t}{1+R_t}\frac{\partial\Omega/\partial D_t}{\partial\Omega/\partial C_t}\} = 0,$$

$$(79)\quad E_t\{P_t\frac{R_t(1-k_{2t})-r_{2t}}{1+R_t}+P_t\frac{R_t}{1+R_t}\frac{\partial\Omega/\partial T_t}{\partial\Omega/\partial C_t}\} = 0,$$

$$(80)\quad E_t\{P_t\frac{R_t}{1+R_t}\frac{\partial\Omega/\partial L_t}{\partial\Omega/\partial C_t}-w_{1t}\} = 0,$$

and

$$(81)\quad E_t\{P_t\frac{R_t}{1+R_t}\frac{\partial\Omega/\partial M_t}{\partial\Omega/\partial C_t}-w_{2t}\} = 0.$$

The parameter estimates acquired from GMM estimation under risk neutrality, with weak separability in outputs imposed, are in table 3.[46] Substituting the parameter estimate of ξ in the risk-neutrality case into equations 69, we obtain b_1=0.777 and b_2=0.223. The estimated theoretical aggregate then is acquired by substituting the estimated parameters and fixed constants into equation 29 to get

$$(82)\quad y_0(D_t,T_t) = 0.777D_t+0.223T_t + \frac{1}{2}\left[\frac{.275^2(D_t-T_t)^2}{.58D_t+0.42T_t}\right].$$

The value of the weak separability test statistic, equation 73, is 9.25, while the critical value is 21.666 at the 1 percent significance level. We cannot reject the weak separability hypothesis and, hence, the existence of a theoretical mone-

[45]In producing equations 80 and 81 as special cases of the corresponding risk-averse Euler equations, recall footnote 14.

[46]The instrumental variables are the constant, the discount rate, the lagged banks' portfolio rate of return, excess cash reserves and capital.

tary aggregate over the outputs produced by banks again is accepted. Furthermore, monotonicity and convexity again are accepted throughout the region of the data.

Figures 7-12 provide the risk-neutral plots analogous to those in Figures 1-6 under risk aversion. Imposing risk neutrality produced negligible gain in tracking ability for any of the indexes. Hence, at least with this data, risk aversion does not seriously compromise index number theory.

THE REGULATORY WEDGE

Although the imposition of risk neutrality did not improve the tracking ability of any of our indexes, the risk-neutral special case does permit especially simple graphical illustration of equilibrium phenomena through the use of separating hyperplanes. In particular, with risk neutrality and complete contingent claims markets, each consumer maximizes utility and each firm maximizes profits conditionally upon any fixed, realized contingency (i.e., state). Hence, perfect certainty methods of graphical illustration are available in the risk neutral case, with the understanding that the illustration is conditional upon the realization of all contingencies.

If no regulatory wedge exists between the demand and supply side, a hyperplane separates tastes from technology. But in the case of commercial banks, a regulatory wedge does indeed exist. This conclusion follows from the observation in footnote 8 that an implicit tax is imposed upon banks through the existence of noninterest bearing required reserves. Hence, the user cost price received by banks for the production of monetary services differs from the user cost price paid by depositors for the consumption of those services. The difference is the implicit tax.

The formulas for the user cost prices on each side of the market for produced monetary services was derived by Barnett (1978, 1980, 1987) and computed by Barnett, Hinich and Weber (1986). The result is most easily illustrated in the case of an economy with one consumer, who consumes all of the economy's monetary services, one financial intermediary, which produces all of the economy's monetary services, and two monetary assets. Equilibrium in the monetary sector of the economy at a fixed contingent state is illustrated in Figure 13, when no reserve requirements exist. Money market equilibrium at a fixed contingent state, when one or both of the monetary assets is subject to reserve requirements, is illustrated in Figure 14.

In Figure 13, equilibrium is produced by the familiar separating hyperplane. The separating hyperplane simultaneously supports an indifference curve from below and a production possibility curve from above. The axes represent quantities of each of the two monetary assets demanded and supplied. Equilibrium in the two markets exists at the mutual tangency of the separating hyperplane, the indifference curve, and the production possibility curve at a given optimal level of factor use. In equilibrium, the quantities demanded of each asset are equal to the quantities supplied at the equilibrium point $\mathbf{y}^e=(y_1^e,y_2^e)$. In addition, the gradient vector to the separating hyperplane produces the equilibrium user-cost prices. The vector of user-cost prices paid by the consumer, $\boldsymbol{\phi}_c$, are equal, in equilibrium, to the vector of user-cost prices received by the financial intermediary, $\boldsymbol{\phi}_f$. The user cost price of asset type i is defined by equation 77 above.

With factor employment assumed to be set in advance at its optimum, $\mathbf{x}^*$, the optimum level of aggregate monetary service production, y^*, is defined to be the solution for y^* to the equation $H(y^*,\mathbf{x}) = 0$, where $y^*=y_0(\mathbf{y})$ and where $\Omega(\mathbf{y},\mathbf{x}) = H(y^*,\mathbf{x}) = H(y_0(\mathbf{y}),\mathbf{x})$, as explained in the subsection above. Hence, Figure 13 is drawn conditionally upon that fixed setting of y^*, so that the production possibility surface is the set $\{(y_1,y_2): y_0(y_1,y_2)=y^*\}$.

However, the situation is very different, when required reserves exist. In that case, two different supporting hyperplanes exist in equilibrium. One supporting hyperplane exists for the financial intermediary, and another exists for the consumer. In Figure 14, the line with gradient equal to the consumer's monetary-asset user-cost prices, $\boldsymbol{\phi}_c$, is the consumer's supporting hyperplane and it is his budget constraint in equilibrium. That line is tangent to the displayed indifference curve in equilibrium. The financial intermediary's supporting hyperplane has gradient equal to the financial intermediary's user-cost prices, $\boldsymbol{\phi}_f$. That hyperplane is the financial intermediary's iso-revenue line, which is tangent to the firm's production possibility curve at the equilibrium point. While the user-cost price paid by the consumer for the services of asset type i

Figure 7

Levels of Five Monetary Aggregates (parameters of theoretical monetary aggregate estimated with imposed risk neutrality)

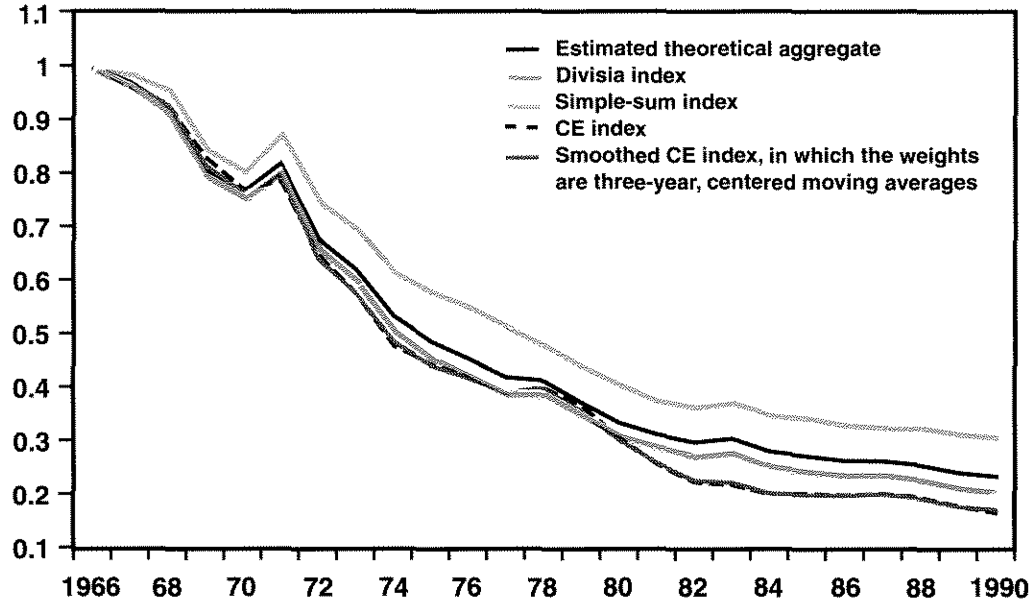


Figure 8

Growth Rates of Five Monetary Aggregates (parameters of theoretical monetary aggregate estimated subject to imposed risk neutrality)

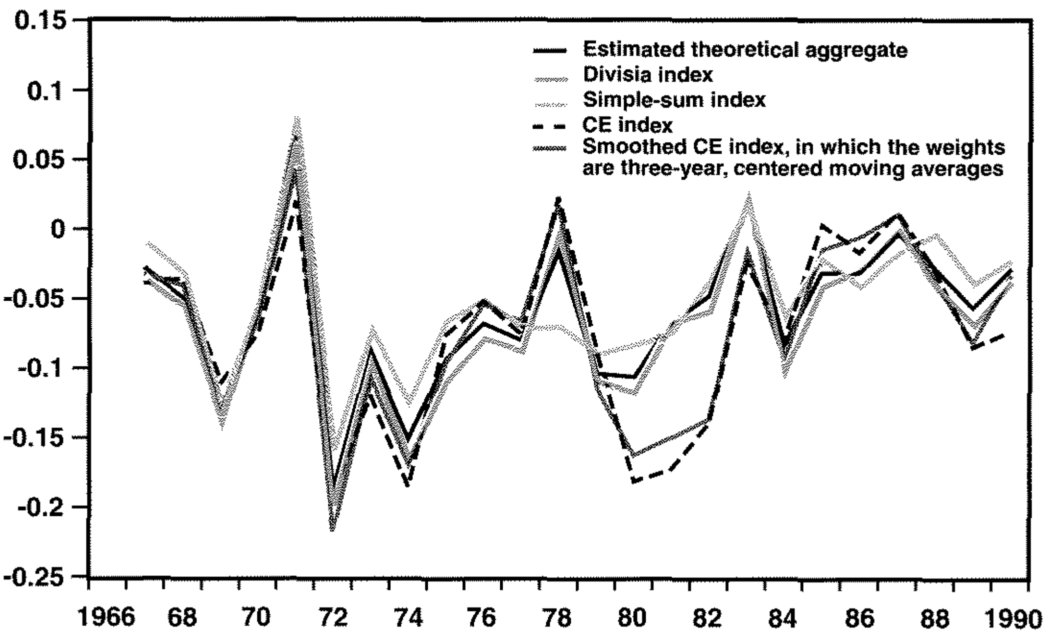


Figure 9

Growth Rates of Theoretical Monetary Aggregate and Divisia Index (with imposed risk neutrality)

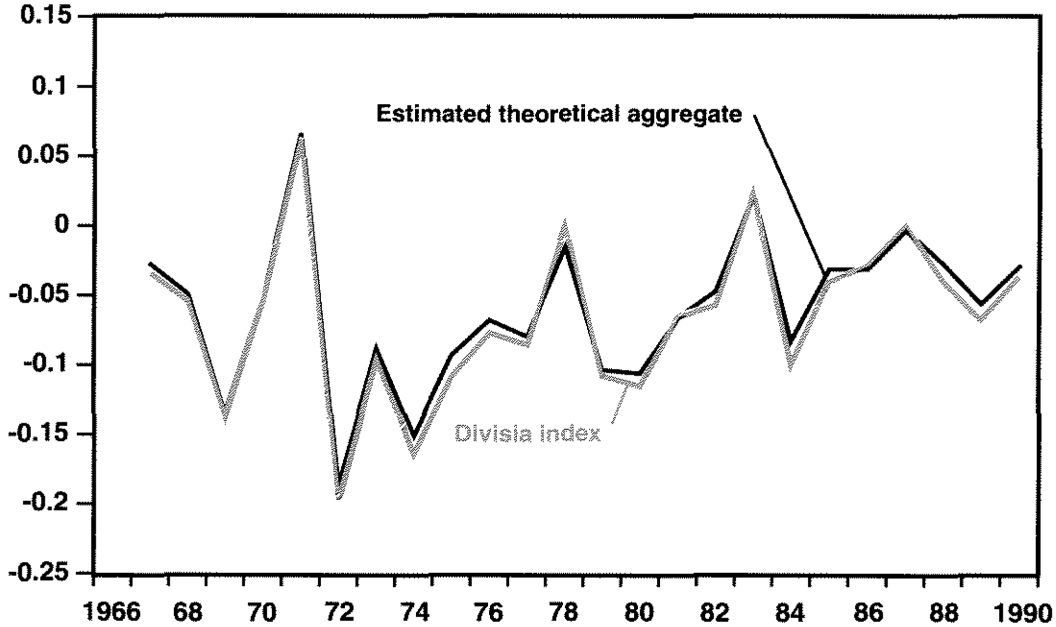


Figure 10

Growth Rates of Theoretical Monetary Aggregate and Simple-Sum Index (with imposed risk neutrality)

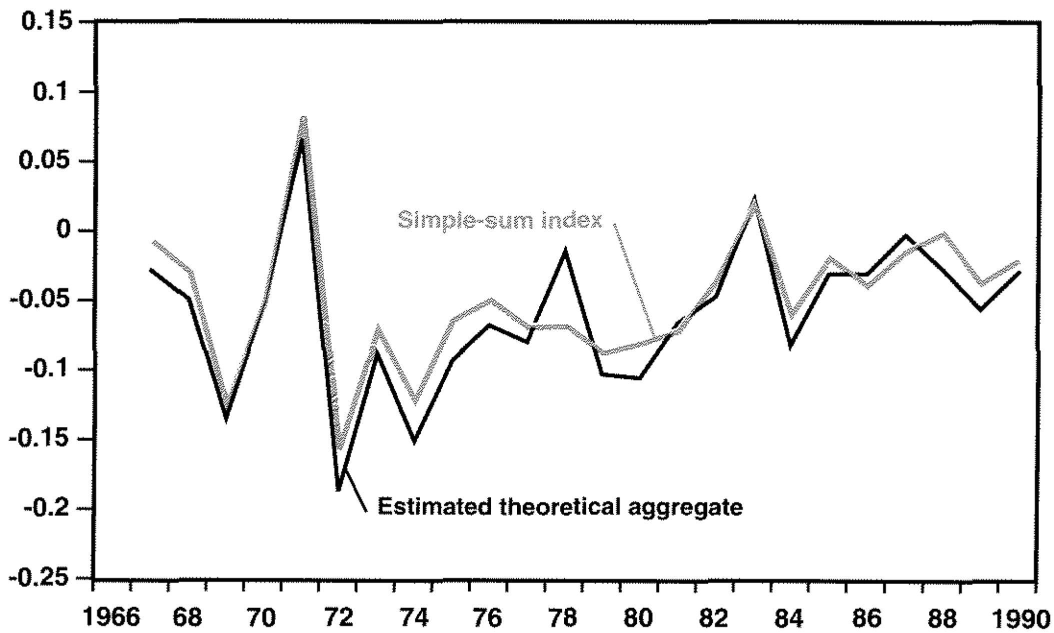


Figure 11

Growth Rates of Theoretical Monetary Aggregate and CE Index (with imposed risk neutrality)

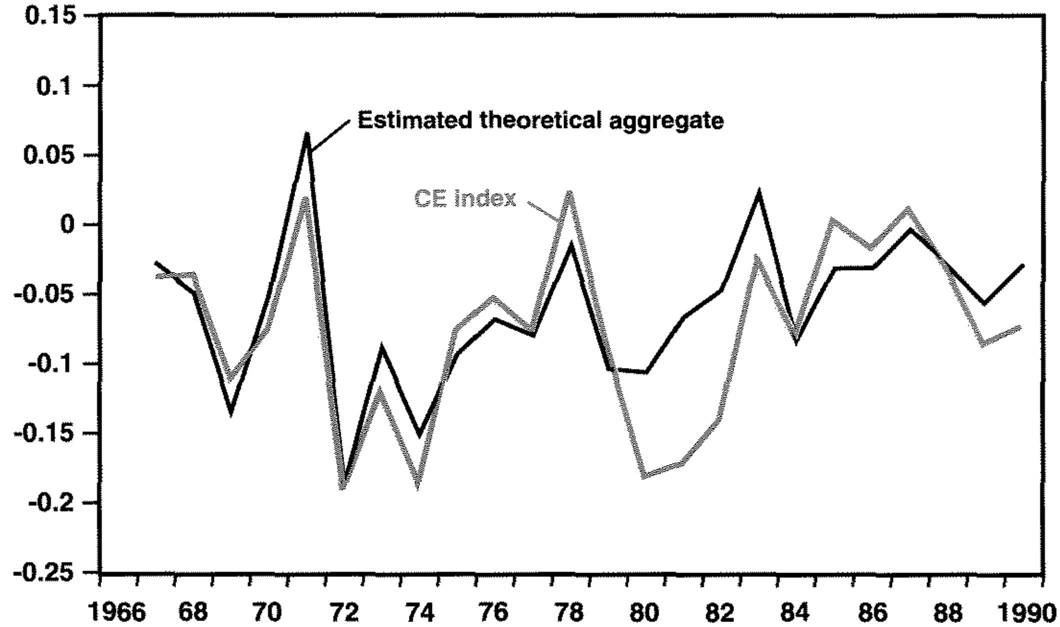


Figure 12

Growth Rates of Theoretical Monetary Aggregate and Smoothed CE Index (with imposed risk neutrality)

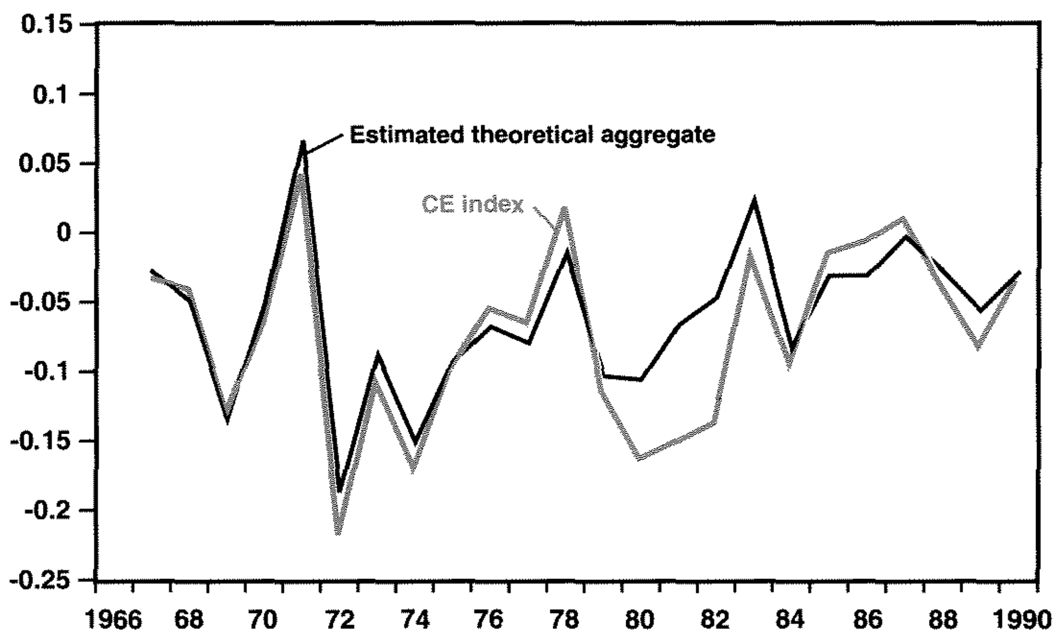


Figure 13
Equilibrium with No Required Reserves

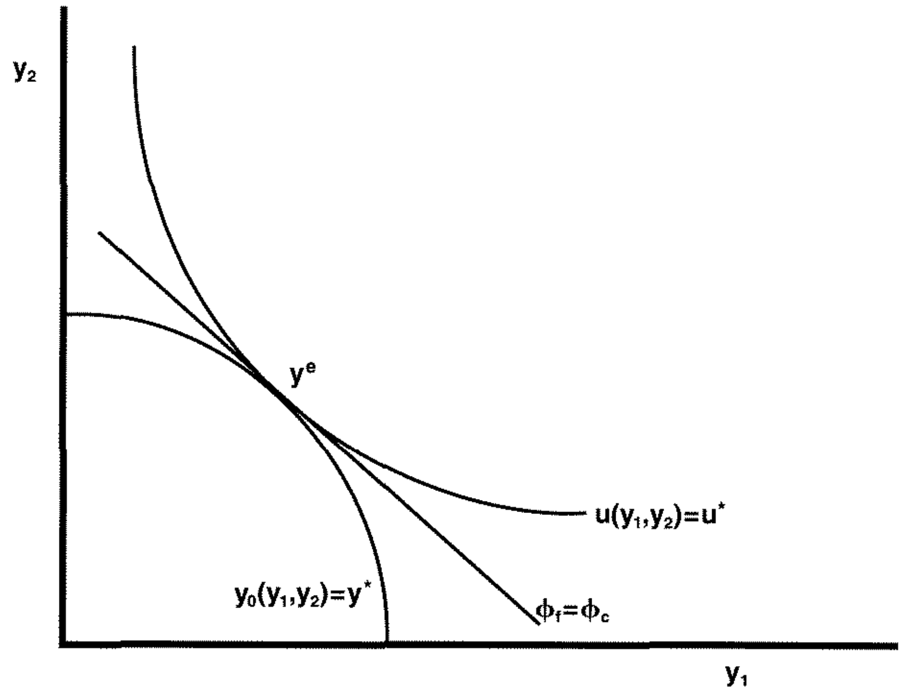
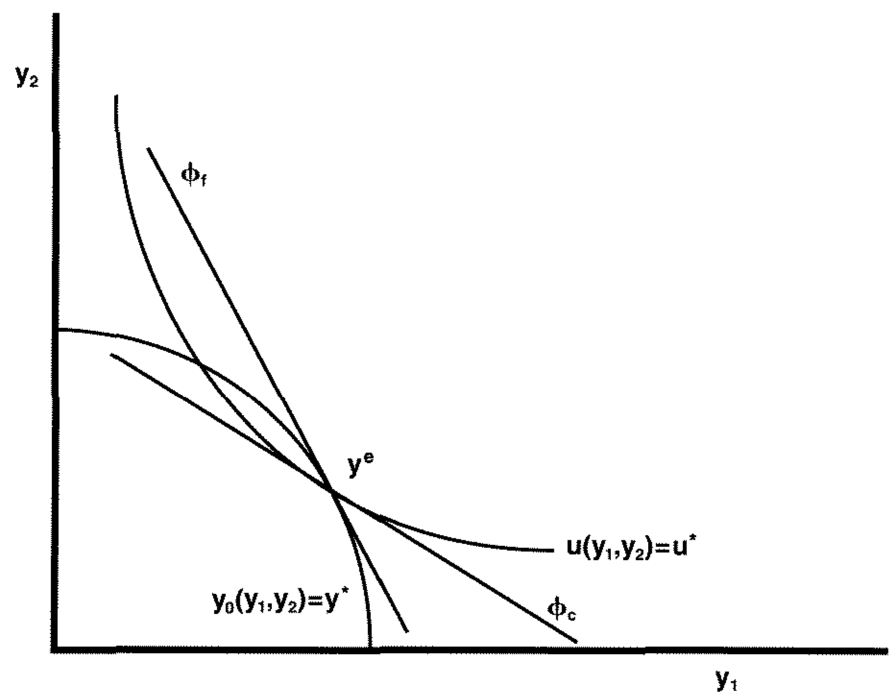


Figure 14
Equilibrium with Required Reserves



is still defined by equation 77, the user-cost price received by the bank for producing those services now is defined by equation 75, which does not equal equation 77 unless no required reserves exist.

The equilibrium point is the point $\mathbf{y}^e$ at which the two supporting hyperplanes intersect, and the angle between them is the regulatory wedge produced by the implicit reserve requirement tax paid by the financial intermediary in the form of foregone interest on required reserves. At the equilibrium point both markets are cleared, and the consumer is maximizing utility subject to the displayed budget constraint, while the financial intermediary is maximizing revenue subject to the displayed production possibility curve.

THE ERRORS-IN-THE-VARIABLES PROBLEM

This same figure also can be used to illustrate the magnitude of the errors-in-the-variables problem produced by the use of the simple-sum index as a measure of the flow of monetary services. Figure 15 illustrates the range of the error on the demand side, while Figure 16 does the same on the supply side. The same illustration could be produced on the supply side by replacing the two indifference curves that are convex to the origin with two production possibility curves, that are concave to the origin. The conclusion would be the same.

In both figures, the hyperplane represents the set

$$A = \{(y_1,y_2): \; y_1+y_2 = M_{ss}\},$$

where M_{ss} is the measured level of the simple-sum index, while A is the set of possible values of the monetary asset component quantities (y_1,y_2) that are consistent with the measured level of the simple-sum index.

For any such measurement on the simple-sum index, the value of the demand-side monetary service flow received by asset holders could be anywhere within the set

$$(83) \quad \{u(y_1,y_2): \; (y_1,y_2)\in A\}.$$

The range of that set is the gap between the utility levels at which the two indifference curves are drawn in Figure 15. Clearly, the upper indifference curve is the one which intersects the hyperplane A at the highest possible utility level, while the lower indifference curve is the one which intersects the hyperplane A at the lowest possible utility level. We see that magnitude of the errors-in-the-variables problem in that illustration, when measured by the range of the set (83), is $M_{max} - M_{min}$. The same conclusion is produced on the supply side from Figure 16, but with set 83 replaced by[47]

$$\{y_0(y_1,y_2): \; (y_1,y_2)\in A\}.$$

The simple-sum monetary aggregates produce a disturbingly large and entirely unnecessary errors-in-the-variables problem. Figures 15 and 16 illustrate the reason. Figures 1-12 illustrate the effect, under circumstances that are most favorable to the simple-sum aggregates: a low level of aggregation over assets having similar yields. With broader aggregation over assets having very different own rates of return, including currency with a zero rate of return, the continued use of simple-sum monetary aggregates by central banks becomes even more difficult to comprehend. The days when all monetary components had zero-own rates of return are long gone.

CONCLUSIONS

In this paper, we develop a theoretical model of monetary service production by financial firms. Earlier models either have permitted risk, but with minimal connection with neoclassical economic theory, or have made full use of neoclassical production theory, but under the assumption of perfect certainty. The latter case has been developed extensively by Barnett (1987), Barnett and Hahm (1994), and Hancock (1985, 1987, 1991). We extend that latter fully neoclassical production approach to the case of risk aversion, subject to Diewert and Wales's symmetric generalized McFadden technology. Our approach permits risk aversion without compromising second-order flexibility or neo-

[47]The magnitude of the gap, $M_{max} - M_{min}$, may differ, when a regulatory wedge is produced by required reserves, but the difference between the conclusions on the demand and supply side is not likely to be large. If the errors-in-the-variables problem is large on one side of the market it is likely to be approximately as large on the other side of the market. See Barnett, Hinich and Weber (1986) for relevant empirical evidence.

Figure 15
Demand Side Errors-in-Variables

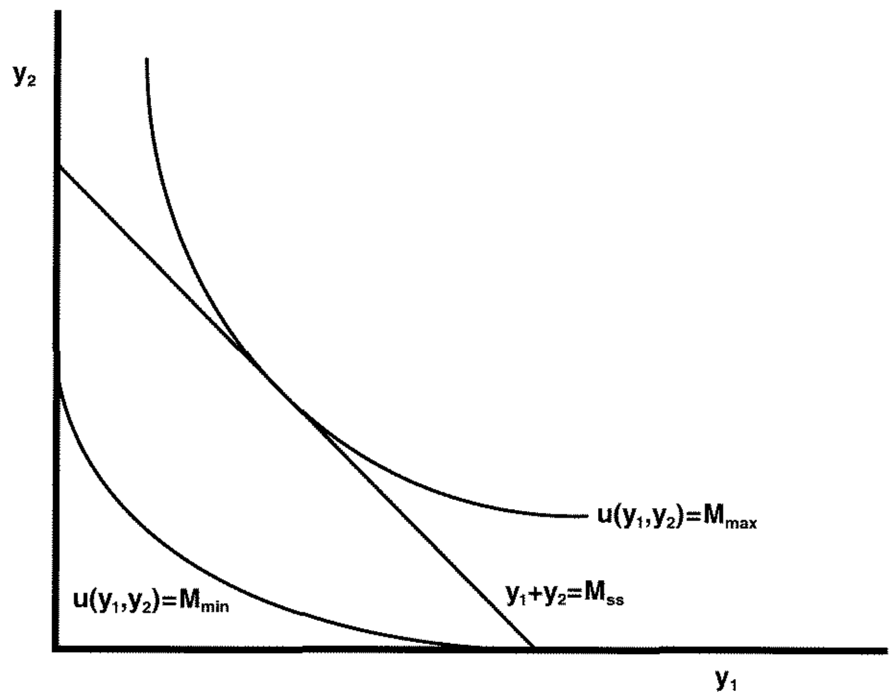
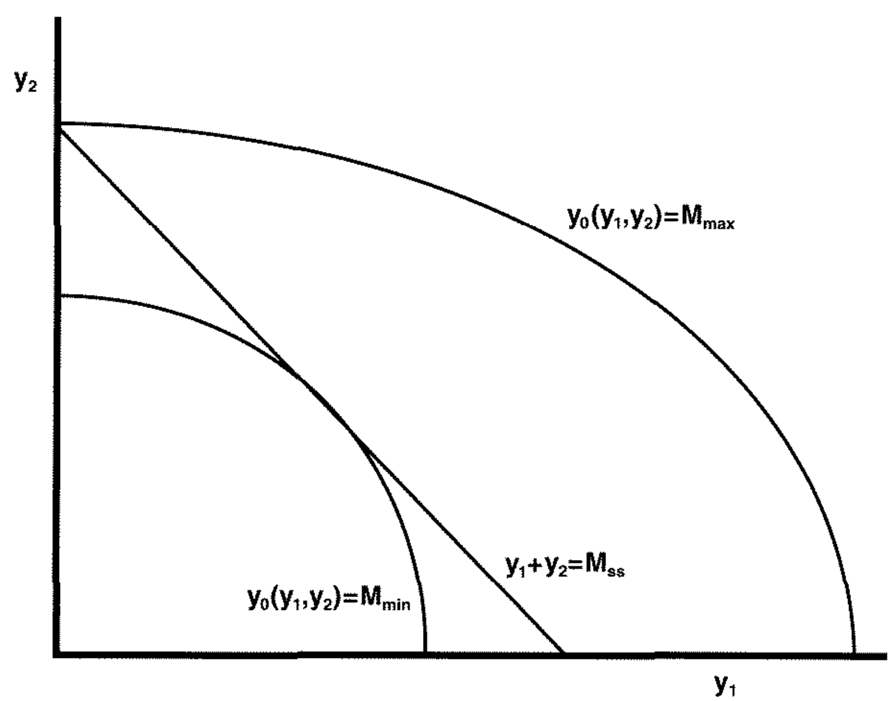


Figure 16
Supply Side Errors-in-Variables



classical regularity of the specification. This is true with or without the imposition of global blockwise weak separability, which we therefore are able to test and to impose, when accepted. Using the resulting Euler equations, we explore exact output aggregation in this paper.

Although applicable to all financial intermediaries, we apply our approach only to the banking industry. While it is possible to impose regularity in curvature conditions upon the generalized McFadden specification, monotonicity can be imposed only locally without damaging the models flexibility. Diewert and Wales's alternative specification, called the generalized Barnett model, is globally regular both in terms of curvature and monotonicity, and hence, that model was used by Barnett and Hahm (1994) in the perfect certainty case. However, in the current paper, using the generalized McFadden model, the estimated parameters satisfy the neoclassical monotonicity and convexity conditions for all observations, even though only convexity was imposed globally. Hence, we doubt that our conclusions would have been much different if we had used the generalized Barnett model in producing our estimated Euler equations. The hypothesis that bank's outputs are weakly separable from inputs is accepted. Hence, the existence of an exact supply-side theoretical monetary aggregate is accepted for banks. The resulting output aggregate is the banking industry's contribution to the economy's inside money services.

While our theory provides a means of econometrically estimating the exact supply-side monetary aggregate, no theory currently is available to support the use of a nonparametric statistical index number as an approximation to the parametrically estimated exact aggregate. Considering the complexities of the GMM estimation involved in producing the estimated exact aggregate, a nonparametric statistical index would, in practice, be much easier to compute and use. We compute the currently most popular of those indexes and find that at least for our sample, the Divisia index tracks the estimated theoretical index more accurately than the others. This conclusion holds regardless of whether or not we impose risk neutrality during estimation of the exact theoretical aggregate. Risk aversion does not appear to produce appreciable degradation of the tracking ability of the Divisia index with our data.

We believe that the approach developed in this paper could be used to investigate technological change in banking, economies of scale and scope in banking, value added in banking and its connection with inside money creation, and the transmission mechanism of monetary policy. We have in fact taken a first step in the direction of producing one of those extensions: We have derived and supplied the Euler equations with learning by doing technological change included in technology. A longer-run framework for the theory also could be productive. In particular, some of the factors excluded from the variable cost function as fixed factors could be incorporated among the variable factors. Bank capital is one such example. Incorporating capital among the variable factors could permit integration of the model with economic growth theory, in which capital evolves endogenously in accordance with a law of motion. In short, this is just a start in a direction that we expect will be very productive for researchers interested in the role of financial institutions in the economy.

REFERENCES

Barnett, William A. "A Reply to Julio Rotemberg," in Michael T. Belongia, ed., *Monetary Policy on the 75th Anniversary of the Federal Reserve System.* Kluwer Academic Publishers, 1991, pp. 232–43.

________. "The Microeconomic Theory of Monetary Aggregation," in William A. Barnett, and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*. Cambridge University Press, 1987, pp. 115–68.

________. "New Indices of Money Supply and the Flexible Laurent Demand System," *Journal of Business and Economic Statistics* (January 1983), pp. 7–23.

________. "The Optimal Level of Monetary Aggregation," *Journal of Money, Credit, and Banking* (November 1982), pp. 687–710.

________. "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory," *Journal of Econometrics* (vol. 14, 1980), pp. 11–48.

________. "The User Cost of Money," *Economic Letters* (vol. 1, no. 2, 1978), pp. 145–49.

________, and Jeong Ho Hahm. "Financial-Firm Production of Monetary Services: A Generalized Symmetric Barnett Variable-Profit-Function Approach," *Journal of Business and Economic Statistics* (January 1994), pp. 36–46.

________, Douglas Fisher, and Apostolos Serletis. "Consumer Theory and the Demand for Money," *Journal of Economic Literature* (December 1992), pp. 2086–119.

________, and Piyu Yue. "Exact Monetary Aggregation Under Risk," Washington University *Working Paper No. 163* (December 1991).

________, John Geweke, and Michael Wolfe. "Seminonparametric Bayesian Estimation of the Asymptotically Ideal Production Model," *Journal of Econometrics* (July/August 1991), pp. 5–50.

________, Melvin J. Hinich and Warren E. Weber. "The Regulatory Wedge Between the Demand-Side and Supply-Side Aggregation-Theoretic Monetary Aggregates," *Journal of Econometrics* (October/November 1986), pp. 165–85.

________, and Yul W. Lee. "The Global Properties of the Minflex Laurent, Generalized Leontief, and Translog Flexible Functional Forms," *Econometrica* (November 1985), pp. 1421–37.

________, Edward K. Offenbacher, and Paul A. Spindt. "The New Divisia Monetary Aggregates," *Journal of Political Economy* (December 1984), pp. 1049–85.

Belongia, Michael T., and James A. Chalfant. "The Changing Empirical Definition of Money: Some Estimates from a Model of the Demand for Money Substitutes," *Journal of Political Economy* (April 1989), pp. 387–97.

Bernanke, Ben, and Mark Gertler. "Banking and Macroeconomic Equilibrium," in William A. Barnett, and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*. Cambridge University Press, 1987, pp. 89–111.

Bhattacharya, Sudipto, and Douglas Gale. "Preference Shocks, Liquidity, and Central Bank Policy," in William A. Barnett, and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*. Cambridge University Press, 1987, pp. 69–88.

Blackorby, Charles, William Schworm, and Timothy Fisher. "Testing for the Existence of Input Aggregates in an Economy Production Function," University of British Columbia Discussion Paper No. 86-26 (1986).

________, Daniel Primont, and R. Robert Russell. "On Testing Separability Restrictions with Flexible Functional Forms," *Journal of Econometrics* (March 1977), pp. 195–209.

Brainard, William, and James Tobin. "Pitfalls in Financial Model Building," *The American Economic Review* (May 1968, Papers and Proceedings 1967), pp. 99–122.

________. "Financial Intermediaries and the Effectiveness of Monetary Controls," *The American Economic Review* (May 1963, Papers and Proceedings 1962), pp. 383–400.

Brown, Randall S., Douglas W. Caves, and Laurits R. Christensen. "Modeling the Structure of Cost and Production for Multiproduct Firms," *Southern Economic Journal* (July 1979), pp. 256–73.

Debreu, Gerard. *Theory of Value, An Axiomatic Analysis of Economic Equilibrium*, Cowles Foundation Monograph 17, Yale University Press, 1959.

Denny, Michael, and Melvyn Fuss. "The Use of Approximation Analysis to Test for Separability and Existence of Consistent Aggregates," *The American Economic Review* (June 1977), pp. 404–18.

Diamond, Douglas W., and Phillip H. Dybvig. "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* (June 1983), pp. 401–19.

Diewert, W. E. "Exact and Superlative Index Numbers," *Journal of Econometrics* (May 1976), pp. 115–45.

________. "Functional Forms for Profit and Transformation Functions," *Journal of Economic Theory* (June 1973), pp. 284–316.

________. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *Journal of Political Economy* (May/June 1971), pp. 481–507.

________, and T.J. Wales. "Flexible Functional Forms and Tests of Homogeneous Separability," University of British Columbia *Discussion Paper No. 91-12* (1991).

________. "A Normalized Quadratic Semiflexible Functional Form," *Journal of Econometrics* (March 1988), pp. 327–42.

________. "Flexible Functional Forms and Global Curvature Conditions," *Econometrica* (January 1987), pp. 43–68.

Duffie, Darrell. "The Theory of Value in Security Markets," in Werner Hildenbrand, and Hugo Sonnenschein, eds., *Handbook of Mathematical Economics*, vol. IV. North-Holland, 1991.

Gallant, A. Ronald. "Unbiased Determination of Production Technologies," *Journal of Econometrics* (vol. 20, 1982), pp. 285–323.

________. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics* (vol. 15, 1981), pp. 211–45.

________, and Gene H. Golub. "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics* (December 1984), pp. 295–321.

Hall, Robert E. "The Specification of Technology with Several Kinds of Output," *Journal of Political Economy* (July/August 1973), pp. 878–92.

Hancock, Diana. *A Theory of Production for the Financial Firm*. Kluwer Academic Publishers, 1991.

________. "Aggregation of Monetary Goods: A Production Model," in William A. Barnett, and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*. Cambridge University Press, 1987, pp. 200-18.

________. "The Financial Firm: Production with Monetary and Nonmonetary Goods," *Journal of Political Economy* (October 1985), pp. 859-80.

Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* (July 1982), pp. 1029–54.

________, and Kenneth J. Singleton. "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models," *Econometrica* (September 1982), pp. 1269–86.

Ingersoll, Jonathan E. *Theory of Financial Decision Making*. Rowman and Littlefield Publishers, Inc., 1987.

Jacklin, Charles J., and Sudipto Bhattacharya. "Distinguishing Panics and Information-Based Bank Runs: Welfare and Policy Implications," *Journal of Political Economy* (June 1988), pp. 568–92.

Lau, L. J. "Testing and Imposing Monotonicity, Convexity and Quasi-Convexity Constraints," in Melvyn Fuss, and Daniel McFadden, eds., *Production Economics: A Dual Approach to Theory and Application*. North-Holland, 1978, pp. 409–53.

Leontief, Wassily. "Introduction to a Theory of the Internal Structure of Functional Relationships," *Econometrica* (October 1947a), pp. 361–73.

________. "A Note On The Interrelation Of Subsets Of Independent Variables Of A Continuous Function With Continuous First Derivatives," *Bulletin of the American Mathematical Society* (April 1947b), pp. 343–50.

Mackinlay, A. Craig, and Matthew P. Richardson. "Using Generalized Method of Moments to Test Mean-Variance Efficiency," *Journal of Finance* (June 1991), pp. 511–27.

Magill, Michael, and Wayne Shafer. "Incomplete Markets," in Werner Hildenbrand and Hugo Sonnenschein, eds., *Handbook of Mathematical Economics*, vol. IV. North-Holland, 1991.

Newey, Whitney K., and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* (May 1987), pp. 703–08.

Poterba, James M., and Julio J. Rotemberg. "Money in the Utility Function: An Empirical Implementation," in William A. Barnett, and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*. Cambridge University Press, 1987, pp. 219–40.

Robles, Bárbara J. "The Optimal Demand for Money in U.S. Manufacturing: A Dynamic Micro Theoretic Approach," University of Colorado at Boulder Discussion Paper No. 93-15 (June 1993).

Rotemberg, Julio J. "Monetary Aggregates and Their Uses," in Michael T. Belongia, ed., *Monetary Policy on the 75th Anniversary of the Federal Reserve System*. Kluwer Academic Publishers, 1991, pp. 223–31.

________, John C. Driscoll, and James Poterba. "Money, Output, and Prices: Evidence from a New Monetary Aggregate," NBER *Working Paper No. 3824* (August 1991).

Sargent, Thomas J. *Dynamic Macroeconomic Theory*. Harvard University Press, 1987.

Swofford, James L., and Gerald A. Whitney. "Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisure, and Money," *Review of Economics and Statistics* (August 1987), pp. 458–64.

Thornton, Daniel L., and Piyu Yue. "An Extended Series of Divisia Monetary Aggregates," this *Review* (November/December 1992), pp. 35–52.

Tobin, James. "Money, Capital, and Other Stores of Value," *The American Economic Review* (May 1961), pp. 26–46.

Woodland, Alan D. "On Testing Weak Separability," *Journal of Econometrics* (December 1978), pp. 383–98.