Banks' Demand for Excess Reserves: A Partial Equilibrium Analysis--Rationality and the Inertia Effect Hypothesis

Thomas H. Turner

February 1978

Working Paper #: 107

PACS File #: 3040

NOT FOR DISTRIBUTION

The views expressed herein are solely those of the author and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The material contained is of a preliminary nature, is circulated to stimulate discussion, and is not to be quoted without permission of the author.

Several hypotheses have been advanced to explain the unusually large accumulation of excess reserves in the 1930s. One of these hypotheses is the theory of the inertia effect advanced by Morrison; this model has as its central tenet the notion that a bank's demand for cash reserves varies inversely with its expected or forecast cash inflow.[1/] This paper explores for this type of static partial equilibrium model the implications of the notion that expectations of commercial banks are formed rationally in the sense of Muth.[2/] The imposition of rationality both reduces the stringency of the assumptions necessary for application of the basic model to aggregate data and delivers several empirically testable restrictions on the model.

The model we will consider differs slightly from the one Morrison proposed, although it is one which captures the essential ingredients of his model. Basically, the model is a simple one-period inventory model in which an individual bank acts to minimize its expected losses (or to maximize its expected profits) in the face of both uncertain cash flows and legal reserve requirements. A single bank is presumed to hold only two types of assets--noninterest-bearing cash and fixed interest-bearing "loans" (securities)--and to issue three types of liabilities--noninterest-bearing demand deposits, fixed interest-bearing short-term debt, and equities. In this model, expected loss each period is a function of the interest cost of holding cash plus the expected transactions costs associated with meeting its end of period reserve requirements. These expected transactions (or penalty) costs arise from the possibility of cash drains occuring during the period which leave the bank's cash position short of that necessary to meet its legal reserve requirements.[3/] Minimization of the expected loss function results in a demand for

excess reserves which is a function of the short-term rate of interest, the transactions or penalty costs, and the parameters of the distribution of changes in the bank's excess reserve position resulting from the random actions of depositors. Thus, the resulting demand for excess reserves is given by:

$$x = f(r, g, \mu, \sigma^2)$$

where x represents excess reserves, g is the proxy for the relevant transactions costs, μ is the mean of the distribution of the cash flows, and σ^2 represents some measure of the variance of the distribution of cash flows.

In Morrison's framework, μ actually represents the banks expected or forecast cash inflow over the period. The central hypothesis in his model is that a bank's demand for cash varies inversely with this expected inflow; e.g., if the expected inflow is small or negative, the amount of cash held to protect against the expected cash drain will increase. Morrison gives empirical content to this notion of expected cash inflow by considering a variable he terms transitory deposit potential. Deposit potential is defined as the ratio of total reserves to the average required reserve ratio, and thus it supposedly takes into account the interdependence between total and required reserves. Transitory potential deposits are then the difference between current actual deposit potential and the bank's expected long-run or permanent deposit potential. When transitory potential deposits are high, expected cash inflow is low, and vice-versa. Thus Morrison hypothesizes that excess reserves vary directly with transitory potential deposits. In his model we then have a bank's demand for excess reserves ρ_e (expressed as a percentage

of net deposits subject to reserve requirements) as a function of the short-term interest rate r, the federal reserve discount rate r_d (used as the penalty cost variable), the yield spread P between corporate bonds of different grade (used as a proxy for variations in default risk on earning assets), and transitory potential deposits q (expressed as a percentage of expected potential deposits). Thus,

$$\rho_e = f(r, r_d, P, q). \text{ 4/}$$

Morrison invokes the standard mechanism of expressing the demand function for cash in terms of ratios independent of nominal cash levels. This homogeneity assumption puts the demand function for cash in the usual form associated with the modern quantity theory. The model we are considering, therefore, differs from Morrison's expression in that our formulation expresses a relationship between the level of excess reserves and the level of net deposits subject to reserve requirements. Most of the empirical tests will be conducted twice, however--once using the levels of deposits and once using the created deposit potential series. In addition, Morrison found that for the periods in which we are interested, neither the discount rate nor the yield spread on corporate bonds were significant in explaining reserves; indeed, he dropped the discount rate entirely from the final form of his estimated equations. We will drop both variables from the equation itself; they will, however, enter the empirical work in a manner to be indicated below.

To facilitate the analysis, we define the following variables:

x_{it} = excess reserves

d_{it} = net deposits subject to reserve requirements

c_{it} = total reserves

k = average required reserve ratio at time t

$dp_{it} = \frac{c_{it}}{k}$ = deposit potential

${}_t\hat{dp}_{it+1}$ = expected (permanent) deposit potential at time t

${}_t\hat{d}_{it+1}$ = expected (permanent) deposits at time t

$tdp_{it} = dp_{it} - {}_t\hat{dp}_{it+1}$ = transitory deposit potential

$td_{it} = d_{it} - {}_t\hat{d}_{it+1}$ = transitory deposits

where for each variable without a caret the subscript i refers to the ith bank and the subscript t refers to the tth period. The variables with carets are forecast variables, the preceding subscript indicating the period during which the forecast is made and the second following subscript indicating the period for which the forecast is made.

For an individual bank at time t, excess reserves are a linear function of r_t and td_{it}, i.e.,

$$(1) \qquad x_{it} = \alpha_i r_t + \beta_i (d_{it} - {}_t\hat{d}_{it+1}) + \varepsilon_{it} = \alpha_i r_t + \beta_i dt_{it} + \varepsilon_{it}$$

where the sequence ε_{it}, i=1, ..., N represents serially independent mutually uncorrelated random terms with mean 0 and finite variance.

Tests of the hypothesis hinge critically on the specification of an expectations mechanism by which banks form their forecasts of expected or permanent deposits; Morrison employed the adaptive expectations framework of Cagan and Friedman.[5/] According to this notion, the expected or permanent level of deposits is revised at a rate which is proportional to the difference between actual deposits at time t and expected deposits, i.e.,

$$\Delta_t \hat{d}_{it+1} = b_i(d_{it} - {}_t\hat{d}_{it+1}).$$

Aside from a trend component, the resulting forecasting scheme for the ith bank is of the form

$$(2) \qquad {}_t\hat{d}_{it+1} = (1-\lambda_i) \sum_{k=0}^{\infty} \lambda_i^k d_{it-k} \quad \underline{6/}$$

We seek now to explore the implications for the behavior of an aggregated class of banks of inclusion in a model like (1) of the notion that forecasts of the expected or permanent levels of deposits for the ith bank are optimal economic and statistical forecasts, based on the information available at time t.

1. Aggregation Over a Class of Banks

The model posits behavior for an individual bank, but empirically the goal is to examine the performance of the model for the aggregate data we possess for aggregated classes of banks. Morrison employed the not unusual technique of positing that an equation of exactly the same form as (1) also describes the behavorial relationship for an aggregated class of N banks, i.e.,

$$(3) \qquad x_t = \alpha r_t + \beta(D_t - {}_t\hat{D}_{t+1}) + \varepsilon_t$$

$$= \alpha r_t + \beta TD_t + \varepsilon_t$$

where x_t represents aggregate excess reserves, D_t represents aggregate deposits, ε_t is an aggregated disturbance term with mean 0 and finite variance, ${}_t\hat{D}_{t+1}$ represents the forecast of the permanent or expected level of aggregate deposits at time t+1 based on information available at time t, and TD_t is aggregate transitory deposits. In addition, this forecast is formed using aggregate data in a manner identical to that used by an individual bank to forecast its own permanent level of deposits:

$$(4) \qquad {}_t\hat{D}_{t+1} = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k D_{t-k}.$$

Morrison correctly acknowledges the resulting aggregation bias introduced in the estimation of (3) and (4) via this formulation,[7/] but it is interesting here to consider briefly the types of restrictions necessary to yield (3) and (4) as appropriate aggregations of (1) and (2).

Rewriting (1) and summing over N banks, we have

$$(5a) \qquad \sum_{i=1}^{N} x_{it} = \sum_{i=1}^{N} \alpha_i r_t + \sum_{i=1}^{N} \beta_i d_{it} - \sum_{i=1}^{N} \beta_i \, {}_t\hat{d}_{it+1} + \sum_{i=1}^{N} \varepsilon_{it}$$

which yields

$$(5b) \qquad x_t = \alpha r_t + \sum_{i=1}^{N} \beta_i d_{it} - \sum_{i=1}^{N} \beta_{i\ t} d_{it+1} + \varepsilon_t$$

where $x_t = \sum_{i=1}^{N} x_{it}$, $\alpha = \sum_{i=1}^{N} \alpha_i$, and $\varepsilon_t = \sum_{i=1}^{N} \varepsilon_{it}$. Now consider that from (5b) and (3) we have

$$(7a) \qquad \beta D_t = \sum_{i=1}^{N} \beta_i d_{it}$$

and

$$(7b) \qquad \beta_t \hat{D}_{t+1} = \sum_{i=1}^{N} \beta_{i\ t} \hat{d}_{it+1}.$$

From (7a) we have

$$D_t = \sum_{i=1}^{N} \frac{\beta_i}{\beta} d_{it} = \sum_{i=1}^{N} w_i d_{it},$$

which says that the aggregate deposit level is a weighted sum of individual bank deposits. If deposits are simply nominal figures, it is difficult to suggest a meaningful interpretation for the w_i's. We may get around this problem by using the perhaps somewhat restrictive assumption of identical banks. This notion may be given content in one of two ways: (a) by assuming $\beta_i = \beta$, i=1, ..., N, in which case $D_t = \sum_{i=1}^{N} d_{it}$, or (b) by assuming that all banks face the identical distribution of deposit flows and that each bank maintains a constant fraction δ_i of aggregate deposits, in which case

$$\beta D_t = \sum_{i=1}^{N} \beta_i \delta_i D_t$$

so that

$$\beta = \sum_{i=1}^{N} \beta_i \delta_i .$$

A less restrictive assumption, and one which may be closer to what people have in mind when aggregating, is that the coefficient β_i is statistically uncorrelated with d_{it} so that

$$\beta D_t = \sum_{i=1}^{N} \beta_i d_{it} = N^{-1} \sum_{i=1}^{N} \beta_i \sum_{i=1}^{N} d_{it}$$

from which

$$\beta = N^{-1} \sum_{i=1}^{N} \beta_i = \bar{\beta}.$$

Now consider that we are requiring (2), (4), and (7b) to hold simultaneously. From (7b)

$$_t\hat{D}_{t+1} = \sum_{i=1}^{N} \frac{\beta_i}{\beta} \, _i\hat{d}_{it+1} = \sum_{i=1}^{N} w_i \, _t\hat{d}_{it+1}$$

so that the forecast of aggregate deposits next period is a weighted sum of individual bank forecasts of their deposit levels. Using (4) and (2), we have

$$(8) \qquad _t\hat{D}_{t+1} = \sum_{k=0}^{\infty} (1-\lambda)\lambda^k D_{t-k} = \sum_{k=0}^{\infty} \sum_{i=1}^{N} (1-\lambda_i)\lambda_i^k w_i d_{it-k}.$$

Now, if we impose the notion of identical banks by assuming $\beta_i = \beta$, i-1, ..., N so that w_i is identically unity, then

$$_t\hat{D}_{t+1} = \sum_{k=0}^{\infty} \sum_{i=1}^{N} (1-\lambda_i)\lambda_i^k d_{it-k}.$$

For (4) to obtain we can assume that the weights $(1-\lambda_i)\lambda_i^k$ are statistically uncorrelated with d_{it-k} so that

$$_t\hat{D}_{t+1} = \sum_{k=0}^{\infty} (1-\lambda)\lambda^k D_{t-k} = \sum_{k=0}^{\infty} W_k D_{t-k}$$

where

$$W_k = N^{-1} \sum_{i=1}^{N} (1-\lambda_i)\lambda_i^k.$$

In this case it appears that W_k will decline geometrically as (4) requires. If the notion of identical banks is imposed by assuming $d_{it} = \delta_i D_t$ for all t, then

$$_t\hat{D}_{t+1} = \sum_{k=0}^{\infty} V_k D_{t-k} \text{ where } V_k = \sum_{i=1}^{N} (1-\lambda_i)\lambda_i^k w_i \delta_i.$$

In this case, $\lim_{k\to\infty} V_k = 0$ but V_k does not decline geometrically unless we make the additional restrictive assumption either that $\lambda_i = \lambda$, i=1, ..., N, or that $w_i\delta_i$ = constant. On the other hand, if we assume in (8) that the weights $(1-\lambda_i)\lambda_i^k w_i$ are statistically uncorrelated with d_{it-k}, then

$$_t\hat{D}_{t+1} = \sum_{k=0}^{\infty} U_k D_{t-k} \text{ where } U_k = N^{-1} \sum_{i=1}^{N} (1-\lambda_i)\lambda_i^k w_i.$$

Again, $\lim_{k\to\infty} U_k = D$ but U_k does not decline geometrically unless either w_i is constant (e.g., unity) or $\lambda_i = \lambda$, i=1, ..., N.[8/]

In summary, it appears that the least restrictive set of assumptions which yields the desired aggregation is that all N banks respond identically to changes in their own transitory deposits, i.e., $\beta_i = \beta$ for $i \varepsilon [i, N]$, and that the geometric weights employed by each individual bank in forming its forecast of its own level of permanent deposits are uncorrelated with the past levels of own bank deposits. The important point in all of this is not so much the particular set of assumptions we choose to justify the aggregation but the observation that such assumptions are necessary because each individual bank forms its forecast based only on its own series of deposits. Each bank is

totally ignorant of, or at the very least, doesn't care what is happening to the other banks in its class and therefore takes no cognizance of the pattern of aggregate deposits. The objective of the model, however, is to explain the forecast of permanent aggregate deposits and, ultimately, the way in which the aggregate level of excess reserves responds to changes in aggregate deposit flows. Thus, it would seem that a desirable feature to incorporate into the model would be the formulation of an expectations mechanism for an individual bank which takes into account the information contained in the aggregate data on which we have observations. Positing that banks' expectations of their permanent level of deposits are rational provides such a vehicle which, in turn, yields a somewhat less restrictive way of deriving the aggregate formulation.

Invoking rationality here means that the ith bank's subjective forecast of its expected or permanent level of deposits is equal to the objective mathematical expectation of future deposit levels conditional on the information appropriate to this forecast. For a single period horizon, this amounts in (1) above to imposing the condition that

$$(9) \qquad {}_t\hat{d}_{it+1} = E(d_{it+1}|\theta_t) = E_t d_{it+1}$$

where E is the mathematical expectation operator and θ_t is the set of information available at time t which economic theory implies is relevant for forecasting future deposits. If the error in predicting d_{t+1} is given by

$$\xi_{it+1} = d_{it+1} - E(d_{it+1}|\theta_t),$$

then rationality implies that

$$E(\xi_{it+1}|\theta_t) = E[d_{it+1}-E(d_{it+1}|\theta_t)|\theta_t] = 0,$$

i.e., that it is impossible to predict the prediction error.

Under rationality (1) now becomes

$$x_{it} = \alpha_i r_t + \beta_i[d_{it}-E(d_{it+1}|\theta_t)] + \varepsilon_{it} \tag{10}$$

where ε_{it} is a serially uncorrelated random term with mean 0 and finite variance. Summing (10) over all N banks in the same class, we get

$$\sum_{i=1}^{N} x_{it} = r_t \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \beta_i d_{it} - \sum_{i=1}^{N} \beta_i E(d_{it+1}|\theta_t) + \sum_{i=1}^{N} \varepsilon_{it}, \tag{11}$$

or, assuming that β_i is statistically uncorrelated with d_{it} and with $E(d_{it+1}|\theta_t)$, we have

$$x_t = \alpha r_t + \beta \sum_{i=1}^{N} d_{it} - \beta \sum_{i=1}^{N} E(d_{it+1}|\theta_t) + \varepsilon_t \tag{12}$$

here $x_t = \sum_{i=1}^{N} x_{it}$, $\beta = N^{-1} \sum_{i=1}^{N} \beta_i$, and $\varepsilon_t = \sum_{i=1}^{N} \varepsilon_{it}$. By virtue of the linearity of the expectation operator we see that

$$\sum_{i=1}^{N} E(d_{it+1}|\theta_t) = E(\sum_{i=1}^{N} d_{it+1}|\theta_t) = E(D_{t+1}|\theta_t)^{9/}$$

from which

$$x_t = \alpha r_t + \beta D_t - \beta E(D_{t+1}|\theta_t) + \varepsilon_t \tag{13a}$$

or

$$x_t = \alpha r_t + \beta TD_t + \varepsilon_t \tag{13b}$$

where TD_t represents aggregate transitory deposits.

Consider again equation (10). Under rationality it is conceivable that x_{it} would itself be useful in predicting d_{it+1}, i.e., $x_{it} \varepsilon \theta_t$. In

general then we would not expect $E(\varepsilon_{it}|\theta_t) = 0$. Indeed, if r_t and d_{it} are also likely candidates for forecasting d_{it+1}, then we would have, for $(x_{it}, r_t\ d_{it})\varepsilon\theta_t$

$$E(\varepsilon_{it}|\theta_t) = x_{it} - \alpha_i r_t - \beta_i d_{it} + \beta_i E(d_{it+1}|\theta_t)$$

$$= \varepsilon_{it},$$

i.e., ε_{it} could be predicted exactly. Similarly, for $\{x_{it}, d_{it}, r_t|i=1, \ldots, N\}\varepsilon\theta_t$, $E(\varepsilon_t|\theta_t) = \varepsilon_t$, i.e., the aggregate disturbance is exactly predictable. Under rationality then it does not seem permissable to assume $E(\varepsilon_{it}|\theta_t) = 0$. We might consider the notion, however, that ε_{it} is statistically independent of all components of θ_t except x_{it} so that $E[\varepsilon_{it}|(\theta_t - x_{it})] = 0$ where $(\theta_t - x_{it})$ represents all variables in θ_t except x_{it}. Note that this assumption would imply that ε_{it} is statistically independent of the current and past deposit levels of banks other than the ith bank, as long as those observations are included in θ_t. The corresponding aggregate condition is thus $E[\varepsilon_t|(\theta_t - x_t)] = 0$.

Unfortunately, this assumption does not solve all of our problems. In particular, r_t and d_{it} (hence r_t and D_t) remain likely candidates for use in forecasting d_{it+1} (or D_{t+1}). Equations (10) and (13) represent behavorial relationships posited for commercial banks and, in general, r_t and D_t are endogenous variables determined simultaneously with x_t (or their components d_{it} and x_{it}). Thus we cannot in general expect that $E(\varepsilon_{it}|r_t, d_{it}) = 0$ or, alternatively, that $E(\varepsilon_t|r_t, D_t) = 0$. Without this condition, however, estimates of equations like (13) yield inconsistent estimates. This is a problem not uncommon in empirical work but one which is often ignored. Imposing rationality rather clearly demonstrates the importance of these difficulties.

What we are left with, in order to further pursue the implications of rationality in this model, is one of two approaches. First, we can assume that $E[\varepsilon_t|(\theta_t-x_t)] = 0$ and that $E[\varepsilon_t|r_t, D_t] = 0$. These assumptions amount to assuming to r_t and D_t (r_t and d_{it}) are exogenous with respect to equations (13) (or equation (10)). To justify this assumption it would really be appropriate to construct a complete macroeconomic model in which r_t and D_t are determined independently of the other determinants of banks' excess reserve positions. It may be possible to construct such a model, but that attempt is not made here. Alternatively, we might seek some sort of rule for the monetary authority to follow so that random shocks to the excess reserve position of banks would be followed by an action to offset any effects of that shock on the interest rate or deposit flows. Such a rule is likely to be rather complicated and its realism might be questioned.

A second approach might be to simply assume nonstochastic relations, i.e., to posit that

$$(10') \qquad x_{it} = \alpha_i r_t + \beta_i[d_{it}-E(d_{it+1}|\theta_t)]$$

so that

$$(13a') \qquad x_t = \alpha r_t + \beta D_t - \beta E(D_{t+1}|\theta_t)$$

and

$$(13b') \qquad x_t = \alpha r_t + \beta T D_t.$$

This approach is certainly even more restrictive than the first alternative, and it is perhaps one against which many economists would have strong priors. Nevertheless, it is a version of the model which is amenable to the tests suggested below and, as such, remains a viable alternative.

We proceed with the analysis by adopting the assumptions that $E[\varepsilon_{it}|(\theta_t - x_{it})] = 0$ and $(r_t, d_{it}|i=1, \ldots, N)\varepsilon\theta_t$ so that $E[\varepsilon_{it}|r_t, d_{it}] = 0$. Now suppose we take the mathematicl expectation of x_t conditional on some subset θ_{1t} of the information set $(\theta_t - x_t)$. We then have $E(x_t|\theta_{1t}) = E(\sum_{i=1}^{N} x_{it}|\theta_{1t}) = \sum_{i=1}^{N} E(x_{it}|\theta_{1t})$. Using (12), $E[x_t|\theta_{1t}] = \alpha E[r_t|\theta_{1t}] + \beta E[\sum_{i=1}^{N} d_{it}|\theta_{1t}] - \beta E[\sum_{i=1}^{N} E(d_{it+1}|\theta_t)|\theta_{1t}] + E[\varepsilon_t|\theta_{1t}]$ or

$$E[x_t|\theta_{1t}] = \alpha E[r_t|\theta_{1t}] + \beta E[D_t|\theta_{1t}] - \beta E[D_{t+1}|\theta_{1t}] + E[\varepsilon_t|\theta_{1t}]. \tag{14}$$

The third term on the right side of (14) was derived from the line above it by noting that

$$E[\sum_{i=1}^{N} E(d_{it+1}|\theta_t)|\theta_{1t}] = \sum_{i=1}^{N} E[E(d_{it+1}|\theta_t)|\theta_{1t}].$$

Now $\theta_{1t} \subseteq \theta_t - x_t \subseteq \theta_t$; it can be shown in this case for a random variable u_t that $E[E(u_t|\theta_t)|\theta_{1t}] = E[u_t|\theta_{1t}]$.[10/] Using this result, we then see that the above becomes

$$\sum_{i=1}^{N} E[d_{it+1}|\theta_{1t}] = E[\sum_{i=1}^{N} d_{it+1}|\theta_{1t}] = E[D_{t+1}|\theta_{1t}].$$

Thus we have used the assumptions that β_i is statistically independent of the bank deposit levels and forecasts of bank permanent deposits and that banks' forecasts are rational to derive, conditional only on <u>subsets</u> of the relevant economic data, a testable relationship between the aggregate level of excess reserves and the interest rate, current aggregate deposits, and rational forecasts of next period's aggregate permanent deposit level. In particular, suppose that θ_{1t} consists only of the short-term rate of interest and current and past aggregate deposits: $\theta_{1t} = [r_t, D_t, D_{t-1}, \ldots, D_{t-k}] \subseteq \theta_t - x_t \subseteq \theta_t$.

Then (14) implies that an unconstrained regression of x_t on $[r_t, D_t, D_{t-1}, \ldots, D_{t-k}]$ should be equal to $\alpha r_t + \beta TD_t$, where $TD_t = D_t - E[D_{t+1}|r_t, D_t, \ldots, D_{t-k}]$, aggregate transitory deposits at time t. In deriving this result we have used $E[r_t|r_t, D_t, \ldots, D_{t-k}] = r_t$ and $E[D_t|r_t, D_t, \ldots, D_{t-k}] = D_t$ and our earlier assumptions on ε_{it} to give $E[\varepsilon_t|r_t, D_t, \ldots, D_{t-k}] = E[\sum_{i=1}^{N} \varepsilon_{it}|r_t, D_t, \ldots, D_{t-k}] = \sum_{i=1}^{N} E[\varepsilon_{it}|r_t, D_t, \ldots, D_{t-k}] = 0$. Thus we have a way of imposing at least one test on the notion of rationality itself via comparison of the performance of an unconstrained regression of x_t on θ_{1t} with that of a regression of x_t on r_t and TD_t. Note also the important condition which has led to formulation of these tests: individual banks need to forecast rationally only with respect to subsets of all available data, i.e., the aggregate data on which we have observations.

Now note that (13b) is, apart from the exact form of the expectations mechanism itself, Morrison's aggregate model. Morrison employed the forecasting scheme (4); this method generally incorporates restrictions on the weights in the distributed lag, usually taking the form of a single restriction requiring the sum of the weights to equal unity.[11/] Imposition of such an arbitrary restriction is probably, in general, not the appropriate restriction to impose and hence can be expected to yield suboptimal forecasts. On the other hand, direct substitution of (4) into (13b) would yield least squares estimates of the coefficients α, $\beta\lambda$, $\beta\lambda(1-\lambda)$, $\beta\lambda(1-\lambda)^2$, ...; β is thus not identifiable with this scheme. Morrison employed a two-step procedure in an attempt to determine the most appropriate weighting scheme for forecasting in various periods and for estimating β. First, λ was varied in (4) and the resultant ${}_t\hat{D}_{t+1}$ and TD_t series were calculated and, second, α and β

were estimated via least squares regression in (13b). The λ and β which maximized R^2 were selected as best estimates.

In our case, imposing rationality yields some testable over-identifying restrictions on β. Equation (14) may be rewritten as

$$(15) \qquad E[x_t | r_t, D_t, D_{t-1}, \ldots, D_{t-k}] = \alpha r_t + \beta D_t - \beta E[D_{t+1} | r_t, D_t, D_{t-1}, \ldots, D_{t-k}] + E[\varepsilon_t | r_t, D_t, D_{t-1}, \ldots, D_{t-k}].$$

Interpreting the conditional mathematical expectations here as earlier as linear regressions on the conditioning variables, the left side of (15) is simply

$$(16a) \qquad E[x_t | r_t, D_t\ D_{t-1}, \ldots, D_{t-k}] = a_0 r_t + b_0 D_t + b_1 D_{t-1} + \ldots + b_k D_{t-k}.$$

Imposing rationality here amounts to forecasting D_{t+1} via a regression on the above subset of conditioning variables, i.e.,

$$(16b) \qquad E[D_{t+1} | r_t, D_t, D_{t-1}, \ldots, D_{t-k}] = d_0 r_t + c_0 D_t + c_1 D_{t-1} + \ldots + c_k D_{t-k}$$

Substituting (16b) into (15) gives

$$(16c) \qquad E[x_t | r_t, D_t, D_{t-1}, \ldots, D_{t-k}] = (\alpha - \beta d_0) r_t + \beta(1-c_0) D_t - \beta c_1 D_{t-1} - \ldots - \beta c_k D_{t-k}.$$

From (15) we may equate coefficients in (16a) and (16c) so that

$$
\begin{aligned}
(17) \qquad a_0 &= \alpha - \beta d_0 \\
b_0 &= \beta(1-c_0) \\
b_1 &= -\beta c_1 \\
&\vdots \\
b_k &= -\beta c_k
\end{aligned}
$$

and β is obviously overdetermined. In general, we would not expect all k+2 equalities in (17) to be simultaneously satisfied by the estimated coefficients. Thus k of these equalities represent overidentifying restrictions on β and these may be tested. Let H_0 be that the k equalities $\beta = -b_1/c_1 = -b_2/c_2 = \ldots = -b_k/c_k$ are all valid restrictions. H_1 is then that one or more of these is incorrect, i.e., is in some sense too restrictive. The totally unconstrained regression in (16a) will yield estimates of a_0, b_0, b_1, ..., b_k; the regression (16b) similarly provides estimates of d_0, c_0, ..., c_k. The series TD_t may now be formed and a least squares regression of x on r_t and TD_t gives the estimated coefficients $\hat{\alpha}$ and $\hat{\beta}$. Now with the estimates of d_0, c_0, ..., c_k, and $\hat{\alpha}$ and $\hat{\beta}$, we may from (17) determine the values of a_0, b_0, ..., b_k implied by the restrictions on β; call these restricted coefficients $\hat{\hat{a}}_0$, $\hat{\hat{b}}_0$, ..., $\hat{\hat{b}}_k$. Now the sum of squared residuals from the regression of x_t on r_t and TD_t is that appropriate the imposition of all k+2 restrictions in (17). Using the original estimates of b_1, ..., b_k from (16a) plus the estimates $\hat{\hat{a}}_0$ and $\hat{\hat{b}}_0$ to calculate predictions of x_t will then give us a sum of squared residuals appropriate when only the first two restrictions in (17) are imposed. The test statistic is then the ratio of the difference between these sums of squared residuals to the sum of squared residuals appropriate to the partially constrained regression, the ratio being corrected for

degree of freedom. The test statistic will be distributed as $F(k, T-k-2)$, where T = sample size. Note that, in view of our earlier discussion, in order to derive consistent estimates of α and β, we really ought to use the method of instrumental variables, substituting predicted values of r_t and D_t from a first-stage regression into the regression of x_t on r_t and the constructed TD_t series. Thus, if $\bar{r}_t$ and $\bar{D}_t$ are the predicted values from a first-stage regression, then $\hat{\alpha}$ and $\hat{\beta}$ are the estimates from the regression

$$x_t = \alpha\bar{r}_t + \beta[\bar{D}_t - E(D_{t+1}|\theta_{1t}] + V_t$$

where V_t is the residual term including the original disturbance plus the first-stage residuals.

2. The Implications of Rationality for the Stochastic Process Generating Deposits in the Inertia Effect Model

Note that in Morrison's model, his predictor at time t for deposits one period hence (4) is also optimal for the predictor at time t for n periods forward. Thus at time t

$$(18) \qquad {}_t\hat{D}_{t+n} = \ldots = {}_t\hat{D}_{t+1} = (1-\lambda)\sum_{i=0}^{\infty}\lambda^i D_{t-i} \; .$$

From (3) we see that

$$(19) \qquad x_{t+1} = {}^{\alpha}r_{t+1} + {}^{\beta}(D_{t+1} - {}_{t+1}\hat{D}_{t+2}) + {}^{\varepsilon}{}_{t+1}$$

where ${}_{t+1}\hat{D}_{t+2}$ is the forecast via (4) at time t+1 of deposits at t+2, i.e., the permenent level of deposits as seen at t+1. Now, for an autoregressive forecasting scheme such as (4), it can be shown that

$${}_{t+1}\hat{D}_{t+j+1} = {}_t\hat{D}_{t+j+1} + gu_{t+1}$$

where $u_{t+1} = D_{t+1} - {}_t\hat{D}_{t+1}$ and g is a constant.[12/] But at time t, the forecast of the forecast error ${}_tu_{t+1} = 0$. Thus, at time t the forecast of ${}_{t+1}\hat{D}_{t+2}$ is simply ${}_t\hat{D}_{t+2}$. Therefore, according to Morrison's forecasting scheme (18) and (19) imply that

$${}_t\hat{x}_{t+1} = {}^{\alpha}_t\hat{r}_{t+1},$$

i.e., that forecasts of excess reserves one period in the future should depend only on forecasts of interest rates one period in the future.

Under rationality (19) becomes

$$(19') \qquad x_{t+1} = {}^{\alpha}r_{t+1} + {}^{\beta}[D_{t+1} - E(D_{t+2}|\theta_{t+1})] + {}^{\varepsilon}{}_{t+1}.$$

The regression of x_{t+1} on some subset of information $\theta_{2t} \subseteq \theta_t \subseteq \theta_{t+1}$ is thus

$$E[x_{t+1}|\theta_{2t}] = \alpha E[r_{t+1}|\theta_{2t}] + \beta E\{[D_{t+1} - E(D_{t+2}|\theta_{t+1})]|\theta_{2t}\}$$

$$+ E[\varepsilon_{t+1}|\theta_{2t}].$$

While we discussed the problems of assuming $E[\varepsilon_t|\theta_t] = 0$, we have not ruled out assuming $E[\varepsilon_t|\theta_{t-1}] = 0$. Using this assumption and noting that $E[E(D_{t+2}|\theta_{t+1})|\theta_{2t}] = E[D_{t+2}|\theta_{2t}]$, we now have

$$(20) \qquad E[x_{t+1}|\theta_{2t}] = \alpha E[r_{t+1}|\theta_{2t}] + \beta[E(D_{t+1}|\theta_{2t}) - E(D_{t+2}|\theta_{2t}].$$

In Morrision's model, the second term on the right in (20), forecasts at time t of transitory deposits at t+1, ${}_t\hat{TD}_{t+1}$, will be 0, even if $TD_t \neq 0$. This version of the model therefore implies that

$$(21) \qquad E[x_{t+1}|\theta_{2t}] = \alpha E[r_{t+1}|\theta_{2t}] ,$$

i.e., that these two regressions differ by the constant α. This implication may be tested by merely computing the regressions and testing for proportionality.

It is important to note that in general we would not expect the equality

$$(22) \qquad E[D_{t+1}|\theta_{2t}] = E[D_{t+2}|\theta_{2t}]$$

to hold. Consequently, the interesting issue to consider at this point is the set of conditions on the sequence D_t under which the geometrically declining lag forecasting scheme will be a rational scheme with the resulting conditions (22) and (21).

First, we see that in order to have forecasts of D_{t+1} depend only on current and lagged values of the same variable, we are imposing the condition that

$$_t\hat{D}_{t+1} = E[D_{t+1}|D_t, D_{t-1}, \ldots, D_{t-k}, \theta_{3t}] = E[D_{t+1}|D_t, D_{t-1}, \ldots, D_{t-k}]$$

where θ_{3t} consists of all observations on all variables in θ_{2t} except current and lagged values of deposits. This condition amounts to saying that D_{t+1} is exogenous with respect to all of the elements of θ_{3t}. Thus if we knew what variables economic theory implies are reasonable ones to include in θ_{2t}, we could test the above proposition directly by calculating the two regressions and testing for zero coefficients on the components of θ_{3t}.

Next, we note that in view of (4) or (18) we may posit that D_t follows a scheme

$$D_{t+1} = (1-\lambda) \sum_{i=0}^{\infty} \lambda^i D_{t-i} + u_{t+1}$$

where u_{t+1} is assumed to be a normally distributed random term with finite mean and variance and which obeys the relation $E[u_{t+1}|\theta_t] = 0$. Note that in particular we are assuming that u_t is a serially uncorrelated random term. Taking the expectation of D_{t+1} conditional on θ_{2t} then gives

$$E[D_{t+1}|\theta_{2t}] = (1-\lambda) \sum_{i=0}^{\infty} \lambda^i D_{t-i}$$

as desired. We may rewrite the above expression for D_t as

$$D_t = \frac{(1-\lambda)}{(1-\lambda L)} D_{t-1} + u_t$$

where L is the lag operator such that $L^n x_t = x_{t-n}$ and where $E[u_t|\theta_{2t-1}] = 0$. From this

$$D_t = \frac{(1-\lambda)L}{(1-\lambda L)} D_t + u_t$$

which gives

$$(1-L)D_t = (1-\lambda L)u_t, \text{ or}$$

$$\Delta D_t = (1-\lambda L)u_t. \tag{23}$$

Thus, as long as $|\lambda| < 1$, this sequence has an autoregressive representation. On the other hand, the sequence D_t does not have a moving average representation, although the sequence ΔD_t does possess one. Thus, the theory predicts that for (4) to be a rational forecasting scheme for D_t, ΔD_t must be a stationary process as prescribed in (23) and the errors must not be predictable, given other information. To test this hypothesis, we may calculate the sample autovariance function and the sample partial autocorrelation function and then compare these patterns with that predicted by the theory.[13/]

The crucial aspect of Morrison's theory is that banks' notions of the stochastic process generating the sequence D_t differ significantly in different time periods. In particular, the theory says that following a bank "crisis" period, the value of λ increases significantly and the effective length of the lag in (4) increases, placing much less weight on the current observation of D_t than in "normal" or precrisis periods. Morrison interprets this increase in λ as a crisis induced drop in the speed of adjustment coefficient in the adaptive expectations formulation, hence the descriptive label "inertia effect." He tests this notion in his model by comparing the estimates of λ for "normal periods" only with the estimates obtained over the entire period between 1921 and 1952, including both crisis and noncrisis periods. The imposition of rationality

suggests that an appropriate testing procedure would be to examine, in view of the relationship (20), the actual sequences D_t over various subperiods. Use of the covariagram and sample partial autocorrelation function over subperiods should provide us with some insight into not only the validity of (23) but also the validity of the notion that the stochastic process generating D_t actually changed following crises.[14/]

As a final note, the inertia effect theory says that[15/]

> "the normal state of affairs is that from the point of view of bank expectations, the bank's current level of deposit-creating potential...and their long run, expected, or 'permanent' level of deposit potential do not diverge widely, because the expected level is rapidly revised to recognize changes in the current level."

As a limiting case, this statement would seem to imply that in normal periods the relation between D_t and $E[D_{t+1}|\theta_t]$ holds approximately as an equality. This amounts to saying that in normal periods the sequence D_t follows a martingale or that in (23) $\lambda = 0$. Morrison never really implies that he believes $\lambda = 0$ exactly in any normal periods, and his regression tests suggest that indeed λ is not zero for the periods he examined. Nevertheless, examination of the spectral density function of ΔD_t in various subperiods will provide a test of this limiting implication of the theory. In particular, if Morrison's theory is correct, the series ΔD_t should be considerably "whiter" in normal periods than in post-crisis periods.

Footnotes

1/ Morrison, George R., Liquidity Preferences of Commercial Banks, 1966, The University of Chicago Press, Chicago, Illinois.

2/ Muth, John F., "Rational Expectations and the Theory of Price Movements," Econometrica, Vol. 29, No. 3, July 1961, pp. 315-335.

3/ Morrison's model differs slightly from this formulation. In his model, he works with a cash/deposits ratio and no legal reserve requirements. His penalty cost derives from restoring his cash balance to zero in face of cash drains in excess of the initial cash ratio. In our model, we are assuming that a bank's reserve requirements are met if end-of-period excess reserves are nonnegative. In addition, Morrison includes a term reflecting anticipated capital gains on initial holdings of securities.

4/ Morrison, p. 19.

5/ Cagan, P. "The Monetary Dynamics of Hyperinflation," in Studies in the Quantity Theory of Money, Milton Friedman , ed., University of Chicago, 1956, pp. 25-120. See also Friedman, Milton, A Theory of the Consumption Function, NBER, Princeton University Press, 1957.

6/ The introduction of a trend component α_i in the solution for ${}_t\hat{d}_{it+1}$ would result in a forecasting scheme of the form

$$ {}_t\hat{d}_{it+1} = \frac{b_i}{b_i-\alpha_i}\,(1-e^{\alpha_i-b_i}) \sum_{k=0}^{\infty} e^{(\alpha_i-b_i)k} d_{it-k}. $$

In practice, Morrison found that was never larger than about one-sixth the value of b and, in noncrisis periods, was only about 1/250th as large as b. For our purposes, no generality is lost by setting $\alpha=0$.

7/ Morrison, p. 20.

8/ This point has been made succinctly in Bierwag, G. O., and M. A. Grove, "Aggregate Koyck Functions," Econometrica, Vol, 34, No. 4, October 1966, pp. 828-832.

9/ This rather standard but quite important result is proven in most basic probability tests, e.g., see Feller, An Introduction to Probability Theory and Its Applications, Vol. I, 3rd ed., John Wiley, p. 222.

10/ For a proof of this proposition see Shiller, R., "Extracts from 'Rational Expectations and the Term Structure of Interest Rates'," manuscript, Massachusetts Institute of Technology, 1972, pp. 8-15. Also, T. J. Sargent has illustrated the use of this theorem in "Interest Rates and Expected Inflation: A Review Article," manuscript, University of Minnesota, November 1974, pp. 24-25.

11/ See, e.g., Cagan, pp. 37-41.

12/ Box, Jenkins, and Bacon, "Models for Forecasting Seasonal and Non-Seasonal Time Series" in Advanced Seminar on Spectral Analysis of Time Series, ed. by Bernard Harris, John Wiley, 1967, pp. 217-311.

13/ Ibid.

14/ Morrison did make some subperiod tests by determining the λ value which maximized the simple correlation between the excess reserve ratio and either transitory potential deposits or expected deposit potential. He was primarily concerned with biases in the estimates of λ due to measurement errors in the common variables appearing in the correlated ratios. He interpreted his results as tending to confirm the notion that λ increased in post-crisis periods. While these results may be indicative of and consistent with the notion that the process governing D_t and D_{pt} changed over these subperiods, the tests still did not constitute a direct test of the nature of the sequences themselves.

15/ Morrison, p. 59.