The Predictive Failure of the Baba, Hendry and Starr Model of M1

Gregory D. Hess, Christopher S. Jones and Richard D. Porter

Baba et al. (1992) restored a stable specification for M1 demand using an error-correction model which allowed for learning about new assets and incorporated a volatility term in long-term interest rates. Our study replicated their in-sample results, but found that their model completely breaks down over longer sample periods. We argue that this predictive failure could have been anticipated by sensitivity analysis. Their specification appears to have underestimated the interest rate elasticity of money demand because of the learning-adjustment mechanism. Our results also call into question their basic use of volatility in narrow money demand models. © 1998 Elsevier Science Inc.

Keywords: Money demand; Volatility; LSE econometric method

JEL classification: C52, E41, E47

I. Introduction

Numerous authors have attempted to identify a stable demand function for the M1 monetary aggregate [e.g., Goldfeld and Sichel (1992); Hamburger (1977); Lucas (1988); Meltzer (1993)]. Approaches which apply more modern time-series techniques to this problem have been in the majority recently.

... testing of a model requires that the model be tried against data that occurred after the model was built, i.e., after the specifications were decided upon concerning which variables appear in each equation, which variables are endogenous and which exogenous, what is the functional form of each equation, what is the lag structure of each, what restrictions the parameters are to obey.

Carl F. Christ 1975, p. 54)

Consequently, until a model has been rigorously tested against new evidence, it seems hazardous to place much weight on its implications, no matter how 'pleasing' these seem.

(David F. Hendry 1993, p. 426)

Department of Economics, Oberlin College, Oberlin, OH (G.D.H.); The Wharton School, University of Pennsylvania, Philadelphia, PA (C.S.J.); Federal Reserve Board, Division of Monetary Affairs, Washington, DC (R.D.P.)

Address correspondence to: Dr. R. D. Porter, Federal Reserve Board, Division of Monetary Affairs, 20th and Constitution, N.W. Washington, DC 20551, U.S.A.

1 See Anderson and Kavajecz (1993) for a detailed list of the changes made to the definitions and measurement of the monetary aggregates since 1959.

Baba et al. (1992) have made an heroic attempt to resurrect a stable demand function for M1 in the United States. Their aims have been fourfold: First, to introduce both theoretical and empirical evidence that interest rate volatility at the long end of the maturity spectrum is an important explanatory variable in the demand function for M1; second, to provide empirical evidence on the effects of a learning-adjustment mechanism on M1 demand as a result of the deregulation of various accounts and the introduction of new M1 and non-M1 M2 assets; third, to apply a rigorous statistical methodology to the specification of M1 demand; finally, to explain critical aspects of the demand for M1 such as the period of ‘missing money’ and the ‘great velocity slowdown’.²

Our paper analyzes whether the M1 demand model identified by Baba et al. (1992) truly captures the underlying economic and statistical dynamics of this relationship. First, we present their [Baba et al. (1992)] model of M1 demand and discuss their methodology. In particular, we evaluate the theoretical and empirical contribution of three variables which they added to the usual list of explanatory factors: the interest rate spread between the long and short end of the term structure, the volatility of long-term interest rates, and learning-adjusted interest rates on M1 and non-M1 M2 assets. Second, we have found that extending the data sample beyond that used by Baba et al. (1992) leads to a clear, unambiguous rejection of their specification. This rejection is shown using Hendry’s own methodology (Hendry, 1993).

Third, we provide evidence that suggests that the rejection of their model is associated with over-parameterization. To this end, we performed a forecast encompassing test of the Baba et al. (1992) model, the Baba et al. (1992) specification stripped of the specially constructed volatility and learning-adjusted variables, the Boughton (1993) model, and the Mehra (1992) model, which was used as a benchmark for our results. Second, we applied sensitivity analysis used by Leamer (1985) to the original Baba et al. (1992) data set to identify marginal variables in their specification. We found that, in fact, sensitivity analysis would have been helpful in foreshadowing the model’s subsequent predictive failure.³

If the Baba et al. (1992) model has too many parameters, it may not be so surprising that it broke down when confronted with additional data. On this question, we think it is important to assess their use of constructed variables which embodied a number of free parameters. As a rough measure of the importance of constructing explanatory variables to their results, Baba et al. (1992) estimated 14 coefficients for a data sample of 113. However, they selectively used at least 11 other additional free parameters in their variable construction, giving the specification an overall ratio of observations to parameters of nine to two (113/25).⁴

The key term is selective. The extent to which the model was over-parameterized depends not only on these 25 terms, but what is more important, on the highly selective

---

² See Boughton (1993) and the reply by Hendry and Starr (1993) for a discussion of the method used to select the particular M1 specification obtained in the Baba et al. (1992) model.
³ There is a legacy of evaluating money demand models using Leamer-type Extreme Bounds Analysis. For example, see Cooley and LeRoy (1981), McAleer et al. (1985), and the response by Leamer (1985) and Cooley and LeRoy (1986). Although we appreciate the controversy surrounding this methodology, we present this sensitivity analysis to answer the following question: Would an Extreme Bounds Analysis of the Baba et al. (1992) specification, using their own data, have predicted that the model would subsequently break down?
⁴ See footnote 6 for a listing of these variables.
way in which these particular regressors were chosen out of a myriad of possibilities: the
length of the lags on the moving average on volatility and on the learning terms; the use
of the maximum to combine interest rates and an average for other interest rates; the use
of daily average figures on some interest rates and end-of-month measurements on others;
the particular long-term interest rates which were used but which had not been issued for
some time and, thus, had to be derived from other rates using a smoothed version of the
yield curve. The point is that the process of discovering a model that worked had to use
up some degrees of freedom. Moreover, in addition to the degrees of freedom consumed
in discovering a plausible economic specification by this trial-and-error process, one
should also take into account the degrees of freedom lost in fine-tuning the model to fit
the many diagnostic tests the authors considered. All in all, the true degrees of freedom
must inevitably be smaller, perhaps considerably smaller, than the number reported in the
paper.

Our paper is organized as follows: Section II presents an overview of the Baba et al.
(1992) model. Section III presents updated estimates of this model and additional evidence
of model failure discovered in attempting to update their specification. Section IV
discusses potential reasons for the model’s breakdown and provides evidence of potential
problems with the Baba et al. (1992) model of M1. We conclude in Section V.

II. An Overview of the BHS Model

Statistical Overview

The estimation results for the Baba et al. (1992) model are reported in Table 1. For
convenience, we report the original results as published in Baba et al. (1992), equation
(18) in column (I) of Table 1, and in column (II), we present the results over their sample
period, 1960:Q3–1988:Q3, using more recent official updates of the time series, and
making a few minor corrections to their data.\footnote{These include two corrections to their
monthly series for the interest rate on the 20-year government bond obtained from Solomon
Brothers, and a small change in their use of deposit rate series which originated from the
Federal Reserve Board [see the Data Appendix in Baba et al. (1992)].} We also report in Table 2 the results from
a battery of Hendry style diagnostic tests, for possible misspecification. The Baba et al.
(1992) model for real M1 balances is as follows:

\[
\Delta(m - p)_t = \alpha_1 + \alpha_2 AS_t + \alpha_3 \cdot AR11t + \alpha_4 \cdot \Delta Rma_t + \alpha_5 \cdot \Delta AY_t + \alpha_6 \cdot \Delta \hat{\rho}_5 \\
+ \alpha_7 \cdot \Delta_4 p_{t-1} + \alpha_8 \cdot Rnsa_t + \alpha_9 \cdot V_t + \alpha_{10} \cdot \Delta SV_{t-1} + \alpha_{11} \cdot (m - p - \frac{1}{2}y)_{t-2} \\
+ \alpha_{12} \cdot \Delta_4 (m - p)_{t-1} + \alpha_{13} \cdot \Delta^2 (m - p)_{t-4} + \alpha_{14} \cdot D_t, (1)
\]

where \(m, p\) and \(y\) are the natural logarithms of M1, the implicit GNP price deflator, and
real GNP, respectively; \(AS, AR1\) and \(AY\) are two-period averages, \(AS_t = \frac{1}{2}(z_t + z_{t-1}); S_t\)
is the spread between a 20-year government bond rate and a one-month bill rate; \(z_t = R1_t\)
is a one-month bill rate; \(Rma\) is a learning-adjusted maximum yield on assets in non-M1
M2; \(Rnsa\) is a learning-adjusted yield on other checkable deposits; \(\Delta \hat{\rho}_5 = \Delta \rho_t + \Delta^2 \rho_t\) is
an inflation indicator; \(V\) is a measure of volatility in long-term government bonds; \(SV =
\max (0, S) \cdot V; D\) is a dummy variable for credit controls equal to \(-1\) in 1980:Q2 and \(+1\)
in 1980:Q3; \(\Delta^2 = \Delta z_t - \Delta z_{t-1}\) and \(\Delta = (z_t - z_{t-1})/\Delta t. M\)
This model specifies the demand for M1 by an error-correction model with an imposed long-run income elasticity of one half, the theoretical value in the Baumol-Tobin model. The long-run portion of the model relates the level of real balances to the level of real income, the volatility of long-term interest rates, the interest rate spread between the 20-year and the one-month interest rates on U.S. government bills, the interest rate on a one-month government bill, the rate of expected inflation, and the learning-adjusted own rate on that portion of M1 which pays explicit interest. Other factors affect only the

\[ m_t \]  
Natural log of M1.

\[ y_t \]  
Natural log of Real GNP.

\[ p_t \]  
Natural log of the Implicit GNP price deflator.

\[ R1_t \]  
1-month Treasury bill rate.

\[ R20_t \]  
20-year Treasury bond yield to maturity.

\[ S_t \]  
\( R20_t - R1_t \).

\[ Rna_t \]  
Maximum learning-adjusted M2 rate, \( \max \{ R_p, Rcda, Rmfa \} \).

\[ \beta_t \]  
Inflation predictor = \( \Delta \log (P_t) + \Delta^2 \log (P_t) \).

\[ Rnsa_t \]  
Average of learning-adjusted NOW and Super NOW rates; \( 1/2 \cdot (Rna + Rnsa) \).

\[ V_t \]  
Nine-quarter moving average of the 12-month moving standard deviation of the holding period yield on 20-year Treasury bonds.

\[ \Delta SV_t \]  
\( \Delta V_t \cdot \max (AS_t, 0) \).

\[ D_t \]  

\[ Az_t \]  
An MA(2) operation such that \( Az = \frac{1}{2}(z_t + z_{t-1}) \).

\[ \Delta z_t \]  
\( z_t - z_{t-1} \).

\[ Diz_t \]  
\( D \cdot z_t \cdot 2 \cdot D \cdot z_t \).

\[ w_A(t) \]  
Weight function on newly-introduced assets, \( A \), at time which is defined over three segments: 0 if \( t \leq t_0 \); 1 if \( t \geq 20 + t_0 \), and \( (1 + \exp (7.2(t - t_0 + 1))^{-1} \) if \( t_0 \leq t \leq 20 + t_0 \).

\[ Rcda \]  
Adjusted Certificate of Deposit rate, \( Rcda = R_p + w_{cd} \cdot Rcd - R_p \).

\[ Rmfa \]  
Adjusted Money Fund rate, \( Rmfa = Rcda + w_{cd} \cdot (Rmf - Rcda) \).

\[ Rna \]  
Adjusted rate on NOW accounts, \( Rna = w_c \cdot Rn \).

\[ Rnsa \]  
Adjusted rate on SuperNOW accounts, \( Rnsa = w_{na} \cdot Rsn \).

Note: See the Data Description Table for additional data definitions and the data sources.

This model specifies the demand for M1 by an error-correction model with an imposed long-run income elasticity of one half, the theoretical value in the Baumol-Tobin model. The long-run portion of the model relates the level of real balances to the level of real income, the volatility of long-term interest rates, the interest rate spread between the 20-year and the one-month interest rates on U.S. government bills, the interest rate on a one-month government bill, the rate of expected inflation, and the learning-adjusted own rate on that portion of M1 which pays explicit interest. Other factors affect only the

6 As indicated above, the explanatory variables embody at least 11 additional free parameters: 1) an income elasticity of \( 1/2 \) (error correction); 2) and 3) the 20-quarter learning curve for deposit rates \( (Rna, Rnsa) \) which is obtained by specifying two parameters; 4) a 12-month standard deviation in calculating volatility \( (V, \Delta S \cdot V) \); 5) a nine-quarter moving average of volatility \( (V, \Delta S \cdot V) \); 6) the construction of \( \beta \) as the second difference plus the first difference of \( p_t \); 7) a two-quarter moving average of real income \( (\Delta Ay) \); 8) a two-quarter moving average of 20-year spread \( (AS) \); 9) a two-quarter moving average of the one-month T-bill \( (AR1) \); 10) the calculation of the rate on M2 as the maximum rate on passbook savings, certificates of deposit and money market mutual funds, and 11) the construction of the interaction term of volatility and the spread as the maximum of zero or \( S \), i.e., \( S \cdot V = \max (0, S) \cdot V \). Other anomalies in their model selection include the recognition of CDS and MMMFs but not MMDAs, in the learning-curve weighted maximum M2 rate \( (Rma) \); the use of a maximum rate for M2 deposits, but an average rate for M1 deposits \( (\Delta Rna, Rnsa) \); the use of levels for M1 deposit rates and differences for non-M1 rates \( (\Delta Rma, Rnsa) \); the use of levels for volatility and differences for spread-adjusted volatility \( (V, \Delta S \cdot V) \), and allowing the volatility of interest rates to enter with a nine-quarter moving average, but directly incorporating adaptive learning with respect to interest rates on new assets.
short-run dynamics of the demand for real M1 balances, e.g., the learning-adjusted rate on non-M1 M2 assets, lagged income and price growth, and an interaction term between the interest rate spread between the 20-year and the one-month bill rate, and the volatility of long-term interest rates.

Baba et al. (1992) vigorously employed a statistical testing methodology that featured residual-based diagnostics to test for heteroscedasticity, serial correlation, non-normality, functional misspecification, and parameter constancy. The objective of their specification search was to find a demand specification for real M1 balances that was parsimonious with constant parameters, passed a battery of diagnostic tests, and made economic sense. Finally, the dynamic structure of the model is based on an error-correction model, which ultimately provides a long-run money demand function which is linear in log levels.

As demonstrated by the coefficient estimates and diagnostic test results in columns (I) and (II) of Tables 1 and 2, the Baba et al. (1992) specification for M1 demand was quite robust over the time period 1960:Q3–1988:Q3. The model’s parameters all have the correct sign and are all significantly different from zero at below the 1% level. Moreover, the model passed all diagnostic tests except for a predictive failure test for the period of the ‘Great Velocity Decline’. Moreover, these results hold for the time period 1960:Q3–1988:Q3, using both the original Baba et al. (1992) data set [column (I)] and the latest data available to the Federal Reserve Board as of April 1994 [column (II)].
The BHS Specification

The three new variables which Baba et al. (1992) contributed to the M1 demand literature are: a learning-adjustment mechanism in which money holders slowly learn about the returns on new M1 and non-M1 M2 assets; the slope of the yield curve; the volatility of long-term interest rates. We discuss each contribution in turn.

First, Baba et al. (1992) specified learning-adjusted interest rates on both new M1 assets and new non-M1 M2 assets by incorporating an ad hoc weighting function in the difference between the return on the newly-introduced asset and those on the existing asset. A problem with their approach is that the learning process is only with respect to the rate of return on the asset and is therefore independent of how much of the asset is held. Indeed, the same set of learning parameters is applied to each newly-introduced asset, a restriction which Baba et al. (1992) did not test. This criticism is particularly important as the learning effects take as much as two years to be fully incorporated in the model, whether or not consumers have revealed learning about the asset via changes in

Table 1. Estimates of the BHS Specification of Real M1 Demand

<table>
<thead>
<tr>
<th>Variable</th>
<th>(I)</th>
<th>(II)</th>
<th>(III)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.352***</td>
<td>1.431***</td>
<td>0.790***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.095)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>$A_{S_t}$</td>
<td>-1.409**</td>
<td>-1.319**</td>
<td>-0.382***</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.112)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>$AR_{1_{t}}$</td>
<td>-0.973***</td>
<td>-0.910***</td>
<td>-0.487***</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.068)</td>
<td>(0.082)</td>
</tr>
<tr>
<td>$\Delta R_{ma_{t}}$</td>
<td>-0.255***</td>
<td>-0.203***</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.054)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>$\Delta A_{Y_{t}}$</td>
<td>0.395***</td>
<td>0.387***</td>
<td>-0.102</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.081)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>$\Delta p_{t}$</td>
<td>-0.330***</td>
<td>-0.354***</td>
<td>-0.400***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.054)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>$\Delta q_{t,1}$</td>
<td>-1.097***</td>
<td>-1.190***</td>
<td>-0.343*</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.156)</td>
<td>(0.203)</td>
</tr>
<tr>
<td>$R_{nsa_{t}}$</td>
<td>0.435**</td>
<td>0.335***</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.059)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>$V_{t}$</td>
<td>0.859***</td>
<td>0.897***</td>
<td>0.275***</td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
<td>(0.092)</td>
<td>(0.109)</td>
</tr>
<tr>
<td>$\Delta SV_{t,1}$</td>
<td>11.680***</td>
<td>10.593***</td>
<td>5.107***</td>
</tr>
<tr>
<td></td>
<td>(1.490)</td>
<td>(1.629)</td>
<td>(2.296)</td>
</tr>
<tr>
<td>$(m - p - \frac{1}{2}y)_{t-2}$</td>
<td>-0.249***</td>
<td>-0.235***</td>
<td>-0.129***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>$\Delta (m - p)_{t-1}$</td>
<td>-0.334***</td>
<td>-0.372***</td>
<td>0.571***</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.108)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>$\Delta^2 (m - p)_{t-4}$</td>
<td>-0.156</td>
<td>-0.090**</td>
<td>-0.209***</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.043)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>$D_{t}$</td>
<td>0.013***</td>
<td>0.016***</td>
<td>0.010***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

Notes: The dependent variable is $D(\hat{m} - \hat{p})_{t}$.

* *, ** *, *** Denote statistical significance at below the .10, .05 and .01 levels, respectively.

The BHS Specification

The three new variables which Baba et al. (1992) contributed to the M1 demand literature are: a learning-adjustment mechanism in which money holders slowly learn about the returns on new M1 and non-M1 M2 assets; the slope of the yield curve; the volatility of long-term interest rates. We discuss each contribution in turn.

First, Baba et al. (1992) specified learning-adjusted interest rates on both new M1 assets and new non-M1 M2 assets by incorporating an ad hoc weighting function in the difference between the return on the newly-introduced asset and those on the existing asset. A problem with their approach is that the learning process is only with respect to the rate of return on the asset and is therefore independent of how much of the asset is held. Indeed, the same set of learning parameters is applied to each newly-introduced asset, a restriction which Baba et al. (1992) did not test. This criticism is particularly important as the learning effects take as much as two years to be fully incorporated in the model, whether or not consumers have revealed learning about the asset via changes in
Table 2. Diagnostic Tests for the BHS Specification of Real M1 Demand

<table>
<thead>
<tr>
<th>Episode Predictive Failure Tests</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing money</td>
<td>1.77</td>
</tr>
<tr>
<td>New operating procedures</td>
<td>7.09</td>
</tr>
<tr>
<td>Great velocity decline</td>
<td>191.89</td>
</tr>
<tr>
<td>M1 explosion</td>
<td>1.52</td>
</tr>
<tr>
<td>Post-BHS sample</td>
<td>37.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residual Normality ( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.452</td>
</tr>
<tr>
<td>0.202</td>
</tr>
<tr>
<td>0.133</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order 1 RESET Test ( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.073</td>
</tr>
<tr>
<td>0.216</td>
</tr>
<tr>
<td>0.073</td>
</tr>
</tbody>
</table>
their actual holdings. Also, one can question the way in which incentives enter into the learning formulation. If opportunity costs are high, presumably deposit holders will have a greater incentive to learn about the new accounts than if opportunity costs are low.\footnote{A comparison of the introduction of nationwide ATS accounts in the fall of 1978 with the introduction of nationwide NOW accounts in January of 1981 supports this contention. Because the accounts provide functionally about the same services, the observation that the NOW account introduction had a much bigger impact is most likely due to the much higher level opportunity costs prevailing in 1981.} In the absence of micro-foundations, why should the learning effects be with respect to own rates, rather than opportunity costs, and shouldn’t evidence of learning also be reflected in the amount of the new asset held?

With regard to the learning of newly introduced M1 and non-M1 M2 assets, the Baba et al. (1992) methodology introduced several, although not all, innovations in the treatment of new M2 assets on the demand for M1. They modeled the introduction of: 1) certificates of deposit in 1965:Q4; 2) money market mutual funds in 1974:Q3; 3) NOW accounts in 1981:Q1, and 4) SuperNOW accounts in 1983:Q1. However, there are a few problems with their treatment of various institutional details. First, since approximately 1986:Q1, NOW and Super NOW accounts have paid almost identical rates, in contrast to the series used by Baba et al. (1992). Second, the Baba et al. (1992) data set had fluctuating rates for NOW accounts and Super NOWs prior to 1983, a time in which virtually all accounts were at fixed interest rate ceilings and Super NOWs did not exist.\footnote{Rates on regular NOW accounts were under fixed regulatory ceilings until 1986.} Finally, Baba et al. (1992) excluded the introduction of MMDAs in the learning-curve weighted maximum M2 rate ($R_{ma}$), yet in many ways MMDAs had the most dramatic effect on M2, increasing this aggregate more than 200 billion dollars (at over a 20\% annual rate) in their first full quarter of existence (1983:Q1).\footnote{MMDA accounts were actually introduced on December 15, 1982.} Note, however, there is no discernable change in their model if these institutional factors are correctly implemented and, hence, we did not feel these criticisms fundamentally altered the contribution of Baba et al. (1992).

Baba et al. (1992) also argued that a spread between the long-term and short-term rates is an essential explanatory variable in the model. However, as M1 balances can be re-optimized each period, only the short-term rate (known with certainty) and the expected holding period yields on long-term assets should affect the demand for real balances. In fact, by allowing the slope of the yield curve to affect the demand for real M1 balances, Baba et al. (1992) were arguing against a pure rational expectations view of the term...
structure, because the holding period yields should be equal for short- and long-term assets if there is no risk premia for holding a longer-term asset. Of course, the yield curve effect may have come from de-linking long rates and short rates in their partial equilibrium model for money. However, if agents face adjustment costs for changing M1 balances, then the yield-curve spread may provide an individual with the expected path of future short rates, which could affect the present demand for M1.10

Finally, Baba et al. (1992) argued that the volatility of long interest rates has substantial effects on the demand for M1. They presented a model adapted from Ando and Shell (1975). In the original model, consumers chose their holdings of money, savings (short-term assets) and equities (longer-term assets) in order to maximize utility. Money and short-term assets each received a fixed nominal interest rate, and the long-term interest rate was subject to risk. In this simple model, the money holdings/savings decision is determined by transactions cost considerations, whereas the portfolio balance between short-term and long-term assets is determined by financial risk and return.11 Therefore, because savings and money have relative returns which are both exposed to the same risk (inflation), the spreads between savings and money are not affected by risk and, hence, the relationship between savings and money can be established with certainty at time t.12 There is no need, therefore, for the riskiness of equities to affect the relation between money and short-term savings.

Baba et al. (1992) modified the Ando and Shell (1975) model by introducing a ‘capital market imperfection’ whereby individuals face a different rate at which they can lend or borrow in the short term. This modification had the following result: first, in the cases where an agent either borrowed or lent short-term assets, the original result held. However, if a capital market imperfection created a wedge between borrowing and lending short that was sufficient to make an agent not participate in the short-term asset market, then the Ando and Shell (1975) separation result did not hold. In particular, the agent must have equated the savings in transactions costs from holding money to the risk adjusted marginal utility from holding long-term instruments. In this way, a rise in the volatility of long rates leads to an increase in money demand.

Baba et al. (1992) also assumed that as the return on longer-term assets rises relative to that on shorter-term instruments, the capital market imperfection becomes more binding and, hence, an interaction term between volatility and the equity-savings spread should have an enhanced effect on the demand for money. Their justification of consumer avoidance of the short-term asset market, however, may be misleading. That the borrowing rate is higher than the lending rate is a necessary but not a sufficient condition for their model result to hold. It also must be true that the rate on equities must be much larger than the rate on savings, so as to make short-term assets relatively unattractive, a condition which will not certainly be satisfied at all times. It is unfortunate that their empirical work did not include a borrowing/lending spread to affect the demand for money directly, or at least present evidence that the capital market imperfection they focused on is, in fact, the driving force between the introduction of volatility and the level of long rates in their model.

10 For example, Cuthbertson (1988) presented a theoretical and empirical treatment of a buffer stock model for M1 with adjustment costs. However, as Goodfriend (1985) has noted, the transaction costs of adjusting M1 are fixed and small so that there is scant justification for including lags in the specification. Goodfriend (1985) attributed the observed lags in empirical money demand to the presence of measurement error in the regressors.
11 This result is due to the assumption that consumption in period t is independent of the portfolio choice at time t.
12 In other words, as both are exposed to inflation risk, the effect cancels out.
Finally, volatility is inherently a short-run, high-frequency concept. The notion that volatility is related to uncertainty and that an increase in volatility might lead, in turn, to a temporary run up in M1 demand has some appeal. However, the reverse should hold as well. This flight to liquidity should reverse itself once the level of volatility has returned to more normal levels.

To see how large an effect their volatility measure has on the low-frequency movements in the M1 demand, Figure 1 plots the long-run elasticity of real balances with respect to volatility (bottom panel), and its impact on real long-run equilibrium M1 holdings (top panel). From Table 1 and equation (1), the long-run elasticity with respect to volatility is approximately $3.4 \cdot V$. Looking at the two panels together, Baba et al. (1992) helped to explain the M1 explosion during the early 1980s due to the permanent increase in volatility. However, the long swings in the elasticity measure seem excessive and unnaturally dominated by low-frequency movements. For example, the average volatility elasticity of real M1 was 4.4% during the 1960s, 8.3% during the 1970s, but 13.2% in the 1980s. We question whether the average individual increased his holdings of real M1 balances from the 1960s to the late 1980s due to the increased volatility of long rates.

We believe that the extremely backward-looking fashion in which volatility is constructed imbues it with an unrealistic, permanent component. The Baba et al. (1992) volatility proxy is a nine-quarter moving average of the quarterly average of the 12-month moving standard deviation of the 20-year Treasury bond. This measure is a highly backward-looking, smoothed, *ex post* measure of interest rate changes which may have little to do with the *ex ante* perceptions of market agents about subsequent interest rate changes. Figure 2 plots the raw volatility series, contrasted against its four-quarter moving average and the nine-quarter moving average used in their paper. Uncertainty about future returns is potentially uncorrelated with the much delayed response of the Baba et al. (1992) measure, the nine-quarter moving average of volatility. For example, during the period when the Federal Reserve targeted non-borrowed reserves (shaded), a period of increased movements in interest rates, the smoothed volatility measure is relatively dormant and takes two years to feed into the nine-quarter smoothed volatility measure used in their model. In contrast, the unsmoothed volatility measure (which is itself a 12-month moving standard deviation) peaks during the non-borrowed reserves period. Moreover, a cross plot between the Baba et al. (1992) volatility measure and an implied *ex ante* volatility measure—based on the Black-Scholes option pricing formula as implemented by a large Wall Street brokerage house for data since 1983—is presented in Figure 3. The figure presents a cross plot between the Baba et al. (1992) quarterly series (unsmoothed) and the implied volatility series, and shows a correlation coefficient of only 0.60. Given the model’s reliance on the volatility series and the theory’s emphasis on the *ex ante* volatility of long-term interest rates, the Baba et al. (1992) *ex post* measure may be a poor proxy.13

---

13 In fact, we show in Figure 6 that the correlation coefficient between the volatility of the long rate and the long rate itself is equal to 0.80. We also found that a daily version of the Baba et al. (1992) volatility measure has a correlation of only 0.33, with the daily measure obtained from a large brokerage house. In addition, one can question why agents do not learn about volatility in the same way that they learn about the returns on new assets, rather than using a simple nine-quarter moving average.
Equilibrium Real M1 and the Contribution of Volatility
(log scale)

- Equilibrium M1
- Equilibrium M1 with zero volatility

Elasticity of Volatility
(3.4 * Volatility)

The vertical lines mark the end of the BHS sample period.

Figure 1.
Figure 2.

The shaded area represents the non-borrowed reserves targeting period. The vertical line marks the end of the BHS sample period.

Figure 3.

Two Measures of Volatility
(quarterly)

correlation: 0.6

Goldman Sachs Measure

BHS Measure
III. Updating the BHS Model

Updated estimates for the Baba et al. (1992) model are presented in column (III) of Table 1, and the results from the diagnostic tests are reported in column (III) of Table 2. The results are striking! Three of the coefficient estimates are no longer statistically different from zero at conventional significance levels (ΔRma, ΔAy and Rnsa), and although one coefficient is still significantly different from zero, it has switched sign (Δ4(m – p)1t). The diagnostic tests also demonstrate a statistical breakdown of the model. The diagnostic tests now reveal significant levels of serial correlation, parameter instability as indicated by the rejection of the Chow test over the post-Baba et al. (1992) sample, predictive failure over the same sample, and non-normal errors. It is important to note that the predictive failure is not due to the updating of the time-series data, which may involve slight data re-definitions and the inclusion of newly-revised seasonal factors, as over the initial sample used by Baba et al. (1992), their model still fit well and passed all the diagnostic statistics [compare columns (I) and (II) of Tables 1 and 2].

Figure 4 presents a multi-panel figure to help evaluate the source of the model’s breakdown with respect to parameter constancy. This figure plots the updated coefficients plus their two standard deviation confidence intervals for 12 parameters in the model. We summarize the results from these plots as follows: First, the error-correction coefficient (α11) rapidly approaches zero, suggesting a misspecification in the levels relationship for M1 demand. Second, volatility (α9) and the interest rate coefficients (α2, α3, α4 and α8) show a strong tendency to approach zero as well, suggesting that these variables were either miscalculated or spuriously included, or both. As these variables were the key innovation to the determinants of money demand in the Baba et al. (1992) specification, this finding places their basic contribution in question.

An important issue in evaluating the predictive breakdown of the Baba et al. (1992) model is how long did it take for the model to get off track? Not long at all. Figure 5 plots the one-step-ahead forecast errors, based on a sequence of rolling regressions, and their two standard deviation bands for the Baba et al. (1992) model from 1970:Q1–1993:Q4. In the first quarter beyond the Baba et al. (1992) sample period, the model’s forecast began to exceed its two standard deviation confidence bands. As a measure of the model’s post sample reliability, the model’s one-step-ahead forecast, using the updated Federal Reserve data series, exceeded its two standard deviation bands 19 out of 75 quarters over the period 1970:Q1–1988:Q3 (25.3% of the time), but 18 out of 25 quarters over 1988:Q4–1993:Q4 (72.0% of the time).14 Indeed, a casual analysis of the post-sample period suggests that it may only be a chance coincidence that the Baba et al. (1992) model did this well in that period. Looking at the plot, it appears that the line for actual money growth crossed the predicted line just once, and only in this neighborhood did the forecasts land within the 95% confidence band. By way of contrast, a count of the number of crossings over the in-sample period is considerably larger.

The sense that the model is going off track can also be seen in various diagnostic statistics, which act as hidden or implicit parameters in the Baba et al. (1992) methodology. These terms might be viewed as having the status of a parameter in their approach because in their search procedure, a tentative specification was jiggled until the diagnostic

14 Note that using the actual Baba et al. (1992) data series, the model’s one-step-ahead forecast exceeded its two standard deviation bounds 15 times out of 75 quarters during the period 1970:Q1–1988:Q3 (20% of the time).
Figure 4.

Recursive Coefficient Estimates and 95% Confidence Intervals

Vertical lines mark the end of the BHS sample period.

A

Recursive Coefficient Estimates and 95% Confidence Intervals

Vertical lines mark the end of the BHS sample period.

B
statistics fell into an acceptable region.\footnote{Leamer makes this point in his discussion of Hendry’s methodology \cite{Hendry1990}.} In any event, a number of these diagnostic parameters also veered off after the sample period—see Figure 6a for the Durbin Watson statistic, Figures 6b and 6c for the autocorrelated residuals for lags from 1–4 and lags from 5–8, respectively, Figure 6d for the RESET test, and Figure 6e for the ARCH residuals for orders 1–4.\footnote{The behavior of some of these statistics during the sample period, for example, RESET, strikes us as somewhat odd, too.}

To further investigate whether the predictive failure of the Baba et al. (1992) specification over the time period 1988:Q4 –1993:Q4 is shared by other money demand specifications, we conducted a forecast encompassing test of the Baba et al. (1992) model, the Mehra model (1992), the Boughton model (1993), and the Baba et al. model stripped of its learning-adjusted interest rates and volatility measures. This test evaluates whether the forecasting power of one equation is encompassed by the forecasting power of another equation.\footnote{Chong and Hendry (1986) introduced this method.} Consider what would be required to assert that the forecasting power of equation \( j \) encompasses that of equation \( i \). Let the pairs \((f'_i, e'_i)\) and \((f'_j, e'_j)\) denote the forecasts and errors made by equation \( i \) and \( j \), respectively. If the coefficient, \( \beta_j \), in the following regression is statistically significant, it means that a divergence of \( f'_j \) from \( f'_i \) helps predict the error in the \( i^{th} \) equation:

\[
e'_i = \beta_j \cdot (f'_i - f'_j) + \nu'_i.
\]
Figure 6: Rolling Diagnostic Statistics

Figure 6a -- Durbin-Watson

Figure 6b -- Autocorrelation lags 1-4

Figure 6c -- Autocorrelation lags 5-8

Figure 6d -- RESET, quadratic terms

Figure 6e -- ARCH Residuals, orders 1-4

Vertical line in plots denotes end of BHS sample period
This regression, though, is insufficient to show that one equation forecast encompasses another, because both forecasts may be helpful in explaining each other’s errors. Therefore, we also performed the regression:

$$e_t^j = \beta_i \cdot (f_t^j - f_t^i) + v_t^j.$$  

If $\beta_j$ is significant but $\beta_i$ is not, equation $j$ encompasses equation $i$: Knowledge of $f_t^j$ helps predict the error made by the $f_t^i$ equation, but knowledge of $f_t^i$ does not help us predict the $f_t^j$ equation. Table 3 reports the forecast encompassing tests for the Baba et al. (1992) model, the same model stripped of its volatility and learning-adjusted interest rates, and the Mehra (1992) model. The $t$ statistic, $t(i,j)$, for the coefficient when $e_t^i$ is regressed on $f_t^j$ is shown in row $i$ and column $j$.

The results of the forecast encompassing tests reinforce the view that despite the Baba et al. (1992) model’s good in-sample performance during their estimation time period, 1960:Q3–1988:Q3, their model’s predictive performance deteriorates markedly when the sample is extended. In fact, the Baba et al. (1992) model is encompassed by both the Mehra (1992) model and the Baba et al. (1992) model stripped of its constructed variables at the 5% level of statistical significance. Also, the Boughton (1993) model encompasses the Baba et al. (1992) model at the 10% level of statistical significance. This result is interesting given that neither of the models which encompass the Baba et al. (1992) model passed all the in-sample specification tests. The finding that the Baba et al. (1992) model is the only model which does not encompass another model suggests that it likely over-fits the demand for M1 in-sample, and thus should be likely to fail when confronted with an extended data sample.

Figures 7a–7d present two-panel plots of a dynamic simulation of the Baba et al. (1992) and Mehra (1992) models for two sample periods, the one used by Baba et al. (1992) and the full sample period 1960:Q3–1993:Q4. The upper panel shows the level of M1 as the solid line and its predicted value as the dotted line; the lower panel displays the growth rates of both of these series. Over the Baba et al. (1992) sample period, their model tracked money very closely even in the dynamic simulation. However, as soon as the sample period ended, the equation went astray, as shown by the vertical line in the plot. While the

### Table 3. Forecast Encompassing Tests

<table>
<thead>
<tr>
<th>Model</th>
<th>BHS</th>
<th>Simplified BHS</th>
<th>Mehra</th>
<th>Boughton</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHS</td>
<td></td>
<td>−6.38***</td>
<td>−7.46***</td>
<td>−1.67*</td>
</tr>
<tr>
<td>Simplified BHS</td>
<td>1.88*</td>
<td></td>
<td>−3.00***</td>
<td>−0.05</td>
</tr>
<tr>
<td>Mehra</td>
<td>0.51</td>
<td>0.17</td>
<td></td>
<td>−0.22</td>
</tr>
<tr>
<td>Boughton</td>
<td>−0.61</td>
<td>−5.11***</td>
<td>−6.85***</td>
<td>−0.22</td>
</tr>
</tbody>
</table>

Notes: The $(i,j)$th elements contain the $t$ statistic for the models in row $i$ and column $j$. Diagonal elements are not applicable. ***, ** and * refer to statistical significance at the 1%, 5% and 10% levels, respectively. The forecast range is 1988:Q4 to 1993:Q4.

18 Hendry and Starr (1993) reported that the Boughton specification did not pass a test which endogenously identifies parameter stability.

19 The Mehra (1992) model encompasses all the models at the 1% level of statistical significance. We recognize that the Mehra (1992) model was specified after the Baba et al. (1992) specification, and therein may lie its advantage. We re-performed the forecast encompassing test for the period after the Mehra (1992) model was specified and found a pattern of forecast encompassing which is identical to those reported in Table 3.
BHS model, Fed data, estimated through 1988:Q3

levels

[Graph showing actual and simulated levels over time]

growth rates

[Graph showing actual and simulated growth rates over time]

Figure 7a.
Figure 7b. Predicted Failure of M1 Models

Mehra model, Fed data, estimated through 1988:Q3

levels

growth rates

Figure 7b.
BHS model, Fed data, estimated through 1993:Q4

levels

growth rates

Figure 7c.
Mehra model, Fed data, estimated through 1993:Q4

levels

growth rates

Figure 7d.
Mehra (1992) model did not track money as well over the sample period, it appears to give much more accurate results over the post-sample period. Giving each of the specifications more data (Figures 7c and 7d) helped the Baba et al. (1992) specification over the post-sample, but this improvement comes at the expense of the initial period of fit, which now looks considerably worse. It is obvious that the Baba et al. (1992) specification with constant coefficients cannot explain the whole period.

To look more closely at what might underlie these differences, Table 4 decomposes money growth into its basic determinants, interest rates (including volatility in the Baba et al. (1992) specification) and income and prices, over the post-sample period.20 As, over the post-sample period, short- and long-term rates came down by relatively sizeable amounts, one might presume that the net contribution of interest rates and volatility would tend to stimulate money growth over the post-sample period. It is thus surprising that the Baba et al. (1992) specification with constant coefficients cannot explain the whole period.

Table 4. Contribution of Various Variables to Money Demand Prediction for Alternative Models and Estimation Periods (percentage points)

<table>
<thead>
<tr>
<th>Model</th>
<th>Period of Fit</th>
<th>Actual</th>
<th>Error</th>
<th>Predicted</th>
<th>Rates and Volatility</th>
<th>Income and Prices</th>
<th>Initial Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHS</td>
<td>BHS sample</td>
<td>3.8</td>
<td>2.7</td>
<td>1.1</td>
<td>−0.1</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Full sample</td>
<td>3.8</td>
<td>1.4</td>
<td>2.4</td>
<td>0.2</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Mehra</td>
<td>BHS sample</td>
<td>3.8</td>
<td>1.0</td>
<td>2.7</td>
<td>1.7</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Full sample</td>
<td>3.8</td>
<td>0.4</td>
<td>3.4</td>
<td>2.2</td>
<td>1.4</td>
<td>−0.1</td>
</tr>
</tbody>
</table>


Mehra (1992) model did not track money as well over the sample period, it appears to give much more accurate results over the post-sample period. Giving each of the specifications more data (Figures 7c and 7d) helped the Baba et al. (1992) specification over the post-sample, but this improvement comes at the expense of the initial period of fit, which now looks considerably worse. It is obvious that the Baba et al. (1992) specification with constant coefficients cannot explain the whole period.

To look more closely at what might underlie these differences, Table 4 decomposes money growth into its basic determinants, interest rates (including volatility in the Baba et al. (1992) specification) and income and prices, over the post-sample period.20 As, over the post-sample period, short- and long-term rates came down by relatively sizeable amounts, one might presume that the net contribution of interest rates and volatility would tend to stimulate money growth over the post-sample period. It is thus surprising that the Baba et al. (1992) model assigns hardly any contribution to these effects. Indeed, over the shorter sample period, the combined effect of interest rates and volatility is negative, not positive! Using the parameter values from the whole period, this contribution remains small but turns positive. By contrast, the Mehra (1992) model has the correct sign for the impact throughout and a much stronger interest-rate contribution (over both estimation periods) and, as a result, produces a much smaller error.

If one believed the claims of Baba et al. (1992), one might conclude that there has been a major new shift in the stable demand for money function identified in their study, and try to identify the reasons for this shift. However, the fact that the Mehra (1992) model did not experience anything like the difficulties faced by the Baba et al. (1992) specification in the post-sample period, suggests otherwise. At first blush, one might be inclined to say that the Mehra (1992) model captures the interest rate effects while the Baba et al. (1992)

---

20 The decomposition, which is obtained by linear superposition, also includes a small initial conditions effect. The error term in the equation represents the cumulative effect of the future shocks in the model which were not known at the beginning of the sample period. Given the structure of the model, the implied dynamic error process is autocorrelated. Comparing this plot with Figure 5, we see that the in-sample fit is worse, as would be expected. Nonetheless, the in-sample fit is impressive.
model doesn’t, but the argument is more complicated. Over the initial sample period, the
net coefficient on the short-run interest rate is very roughly the same in the two models. 21

Nonetheless, given the pattern of interest rates over the post-sample period, it must be true
that our basic insight from the Mehra (1992) model is correct. But what also matters is
the behavior of volatility. Over the post-sample period, volatility drifted down somewhat (see
Figure 2). Thus, the net interest rate effects, which generally acted to boost money growth
over this period, tend to be offset in the Baba et al. (1992) specification by the downtrend
of volatility, which acted to retard money growth. 22

We believe that the misspecification Baba et al. (1992) made is with the lag on
volatility, if indeed, volatility has much of a bearing on real M1 demand at all. Essentially,
Baba et al. (1992) built in too much inertia in their volatility measure. Although, given the
rest of their specification, the long moving average they introduced works better than a
shorter one, the rest of the specification is not the right benchmark. Given the incentives
to learn, as measured by high opportunity costs on deposits, depositors appear to have
learned about the new accounts in a few quarters at most, thus the 20-quarter lag on
learning appears spurious. 23 The reason the particular Baba et al. (1992) specification does
not require a shorter learning curve has to do with the rather extreme values that both
volatility and perhaps the long-term rate assumed when nationwide NOW accounts were
introduced. As a result, the particular specification they constructed, the evidence for a
quicker learning response, was masked.

IV. Could Sensitivity Analysis Have Been Helpful?

The evaluation of money demand specifications using Extreme Bounds Analysis (here-
after EBA) has a rich and controversial history in econometrics. Although we do not wish
to argue the merits of EBA versus the alternative approaches to model selection used by
Baba et al. (1992), we do wish to explore one issue: Given that the Baba et al. (1992)
model for M1 failed to predict with any accuracy outside of the sample period, despite
passing numerous in-sample specification tests, we wonder whether an EBA on the
original data would have anticipated the failure which did occur subsequently?

In brief, EBA requires the researcher to classify from an unrestricted model, a base
specification which contains only free variables, and removes all doubtful variables. By
re-estimating the base specification by including different combinations of the doubtful

21. To make this comparison, we collapse the interest rate structure of the model to the short-term rate, on the
assumption that long rates are a random walk and the spread between the long rate and the short rate is mean
reverting [see Gilles (1994)]. We also assume that deposit rates in the long run are an affine transformation of
short-term rates with slope coefficient equal to one minus the reserve requirement against their accounts. For
simplicity, we also used the model in which the spread between the long and the short rate is risk adjusted [Baba
et al. (1992, equation 22)]. Finally, we assume volatility is fixed in the calculation.

22. Indeed, if we split out the volatility and net interest rate contributions in row 1 of Table 4, we find that
the net interest rate contribution is positive, about 1.0 percentage points, while the volatility measure depresses
M1 growth over the period by about 1.3 percentage points. (The two terms only approximately add up to the net
contribution shown in the table because of the small interaction effect in the specification, the cross product term
in the model.)

23. For example, the introduction of nationwide NOW accounts in January of 1981 did draw sizeable balances
into M1 from outside of this narrow aggregate almost immediately. Most M1 specifications need to account for
this rapid growth in some fashion. The reason that the Baba et al. (1992) specification does not require it is
because of the sharp increase in volatility which is underway, as can be seen in Figure 2. But Figure 2 also
reveals that if Baba et al. (1992) had used a model with no delay in measured volatility, a considerably different
outcome would have occurred.
variables, one can obtain, for example, the lowest and highest estimated 95% confidence bound for each coefficient in the unrestricted model. As well, one can obtain the minimum and maximum estimates, $\beta_{min}$ and $\beta_{max}$, for each coefficient in the unrestricted model.

There are two types of inferences which are considered by EBA advocates to be fragile. First, an inference is Type A fragile if $\beta_{max} - \beta_{min} > k \cdot \sigma(b_0)$, where $k$ is a constant chosen by the researcher and $\sigma(b_0)$ is the estimated standard error of $\beta$ in the unrestricted model. Type A fragility suggests that if the interval $[\beta_{min}, \beta_{max}]$ is large relative to sampling uncertainty, then different models will provide different inferences and, hence, inference will be sensitive to the model being considered. Second, there is evidence of Type B fragility when the 95% extreme bounds switch sign. If Type B fragility occurs, then one is on fragile grounds to decide whether a given variable belongs in the model (i.e., is non-zero).

Cooley and Leroy (1981) used EBA to show that the empirical estimates of a negative interest rate elasticity for the demand for M1 reflects the researchers priors rather than the information content of the data. They [Cooley and Leroy (1981, p. 836)] state that 'the data are such that an energetic specification search will give back almost whatever interest elasticity one wishes to extract, particularly if more than one interest rate is included and if the specification search involves extended tinkering with dynamic effects.'

In a series of articles and responses, McAleer et al. (1985), Leamer (1985) and Cooley and LeRoy (1985) have debated the relative merits of extreme bounds analysis in general, especially as it relates to the demand for money. In summary, McAleer et al. (1985) criticized the original Cooley and LeRoy (1981) money demand EBA by arguing that distinguishing variables into classes of doubtful and free is arbitrary, and that their results about the fragility of estimated interest rate elasticities using EBA were themselves sensitive to this classification. Instead, they argued in favor of an approach to model selection where a general specification is estimated and parameter restrictions are systematically imposed to ensure that the model is parsimonious, and that it passes a battery of rigorous diagnostic tests. In essence, the Baba et al. (1992) approach to modeling the demand for money follows the McAleer et al. (1985) study, while introducing a number of new explanatory factors into the specification and implementing an error-correction type analysis which combines the short-run dynamics and long-run relationships of M1 demand.

In any case, one can ask whether the Leamer (1985) methodology of sensitivity analysis would have been helpful in predicting the model’s breakdown? The results from an EBA of the Baba et al. (1992) model are reported in Table 5. To not overwhelm the reader, we simply report whether the variable passed the Type A and B fragility tests. A more detailed set of results for the 95% extreme bounds, the minimum and maximum estimates, $\beta_{min}$ and $\beta_{max}$, and the standard errors associated with the extreme estimates using a bootstrap procedure [McAleer and Veall (1989)], are reported in the working paper version of this paper [Hess et al. (1994)]. Although we consider alternative base models in columns (II) and (III), we specify the number of doubtful variables in columns (I) to be four: $\Delta Rma, Rnsa, V$, and $\Delta S \cdot V$. Of course, the designation of these variables as doubtful is our call. The basic reason for doing so is the idea that these variables were perhaps over-manufactured to satisfy the in-sample criteria followed by Baba et al. (1992). Again, we are asking whether a researcher who felt that these variables were doubtful, and

---

24 See Breusch (1985) for a useful analytical survey on extreme bounds which clarifies many of the arguments made in McAleer et al. (1985).
implemented an EBA on the model, would have predicted the model to fail. The results suggest that, measured by both types of fragility (A and B), inference about the variables in question is quite fragile.

From the results in column (I) of Table 5, we find that the coefficients on the intercept, $A$, $AR1$, $Dp$, $(m - p - \frac{1}{2})2$, are susceptible to Type A fragility, that the coefficient on $Dma$ is susceptible to Type B fragility, and the coefficients on $Dy$ and $D(m - p)2$ are susceptible to both types of fragility. Therefore, if a researcher were to perform an EBA of the Baba et al. (1992) specification and view their constructed learning-adjusted interest rates and volatility as doubtful, the researcher would have reason to believe that this model would eventually break down, as it subsequently did.

The results presented in columns (II) and (III) demonstrate the EBA when the list of doubtful variables changed. In column (II), the base model was specified such that only the learning-adjusted interest rates were considered doubtful, whereas in column (III), only the volatility variables were deemed doubtful. As noted by McAleer et al. (1985) in their criticism of EBA, as the list of doubtful variables changes, so do the fragility measures. In fact, as the number of doubtful variables decreases, fewer variables are considered fragile. Although we believe that this is an important criticism of EBA, it is of utmost interest to note that a common element of Table 5 is that the variables, $Dy$ and $D(m - p)2$, are always considered both Type A and B fragile. What is even more striking is how this EBA result compares with the evolution of the estimated coefficients

<table>
<thead>
<tr>
<th>Coefficient Parameter</th>
<th>Type A Fragility</th>
<th>Type B Fragility</th>
<th>Type A Fragility</th>
<th>Type B Fragility</th>
<th>Type A Fragility</th>
<th>Type B Fragility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$A$</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$AR1$</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$Dy$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$Dp$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$(m - p - \frac{1}{2})2$</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$D(m - p)2$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$Dma$</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$Rnsa$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$V$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$\Delta S \cdot V_{t-1}$</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Notes: These results are based on the original Baba et al. (1992) data set. The baseline specification is equation (1) in the text.

\[
\Delta(M - p) = \alpha_1 + \alpha_2 A + \alpha_3 \cdot AR1t + \alpha_4 \cdot Dma + \alpha_5 \cdot \Delta Dy + \alpha_6 \cdot \Delta p + \alpha_7 \cdot \Delta(y)_{t-1} + \alpha_8 \cdot Rnsa + \alpha_9 \cdot V, \\
+ \alpha_{10} \cdot \Delta S \cdot V_{t-1} + \alpha_{11} \cdot (m - p - \frac{1}{2})2 + \alpha_{12} \cdot \Delta(m - p)_{t-1} + \alpha_{13} \cdot \Delta(m - p)_{t-2} + \alpha_{14} \cdot D. \quad (1)
\]

The doubtful variables in column (I) are $Dma$, $Rnsa$, $V$ and $\Delta S \cdot V$. The doubtful variables in column (II) are $Dma$ and $Rnsa$, while in column (III), they are $V$ and $\Delta S \cdot V$. For column (I), there are 15 combinations of doubtful variables which can be added to the base specification, and the results reported in this table are derived from these 15 regression models. For columns (II) and (III), there are four combinations. There is evidence of Type A fragility when $\frac{\beta_{max} - \beta_{min}}{\sigma(b_{o})}$ is greater than $k$, where $k$ is set to 4. There is evidence of Type B fragility when the 95% extreme bounds switch sign.
of the Baba et al. (1992) model as presented in Table 1. For the full sample, the coefficient on \( D_{Ay} \) becomes insignificantly different from zero, and the coefficient on \( D_{4(m^2p^2/2y)}t^2 \) switches sign. We believe that the identification of these variables as marginal by EBA, and the subsequent poor performance of these variables, is no coincidence. We thus conclude that sensitivity analysis might have been helpful in foreshadowing the predictive failure of the Baba et al. (1992) model.

Table 6 presents extreme bounds results in which perturbations on the specification of the volatility measure were considered by varying both the width of the standard deviation observation interval and the length of the moving average with which volatility enters the Baba et al. (1992) specification over two periods: The original Baba et al. (1992) sample and the full sample. The results in the table demonstrate that the impact of volatility becomes more ambiguous, ranging from about 0 to 1 in the Baba et al. (1992) sample period for the 95% bounds. Over the entire period, the sign of the volatility term is completely ambiguous, with about as much chance of being positive as negative, at least judging from the 95% results. On balance, we believe that the evidence supporting the use of the volatility measure is quite limited. In addition, apart from the spread terms, the short-term interest rate, and the inflation rate, all of the other coefficients in the specification do not appear to be very precisely estimated in either of the sample periods with the imprecision, of course, greater over the full sample period.

### V. Conclusion

This paper has cataloged the predictive failure of the Baba et al. (1992) money demand model, and suggested that in their search for a stable money demand function which would fit both long-run and short-run behavior, Baba et al. (1992) over-fit their model in-sample through their specification methodology and data construction. Consequently, their model

<table>
<thead>
<tr>
<th>BHS Data and Sample</th>
<th>95% Extreme Bounds</th>
<th>Extreme Estimates</th>
<th>95% Extreme Bounds</th>
<th>Extreme Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>( \beta_{\text{min}} )</td>
<td>( \beta_{\text{max}} )</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.140</td>
<td>0.395</td>
<td>0.185</td>
<td>0.352</td>
</tr>
<tr>
<td>( AS_t )</td>
<td>-1.651</td>
<td>-0.240</td>
<td>-1.425</td>
<td>-0.427</td>
</tr>
<tr>
<td>( AR_t^2 )</td>
<td>-1.113</td>
<td>-0.309</td>
<td>-0.981</td>
<td>-0.434</td>
</tr>
<tr>
<td>( \Delta R_t^2 )</td>
<td>-0.418</td>
<td>-0.040</td>
<td>-0.289</td>
<td>-0.186</td>
</tr>
<tr>
<td>( \Delta Ay_t )</td>
<td>-0.089</td>
<td>0.541</td>
<td>0.104</td>
<td>0.400</td>
</tr>
<tr>
<td>( \Delta \rho_{t-1} )</td>
<td>-0.537</td>
<td>-0.214</td>
<td>-0.401</td>
<td>-0.309</td>
</tr>
<tr>
<td>( V_t )</td>
<td>0.047</td>
<td>0.571</td>
<td>0.208</td>
<td>0.459</td>
</tr>
<tr>
<td>( \Delta SV_{t-1} )</td>
<td>0.012</td>
<td>1.033</td>
<td>0.109</td>
<td>0.869</td>
</tr>
<tr>
<td>( (m - p - \frac{1}{2}y)_{t-2} )</td>
<td>0.117</td>
<td>18.79</td>
<td>3.043</td>
<td>14.018</td>
</tr>
<tr>
<td>( \Delta (m - p)_{t-1} )</td>
<td>-0.279</td>
<td>-0.098</td>
<td>-0.250</td>
<td>-0.130</td>
</tr>
<tr>
<td>( \Delta (m - p)^{-1}_{t-4} )</td>
<td>-0.528</td>
<td>0.447</td>
<td>-0.334</td>
<td>0.191</td>
</tr>
<tr>
<td>( D_t )</td>
<td>0.003</td>
<td>0.024</td>
<td>0.011</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Note: See Table 5.
might have been expected to fail on an out-of-sample basis. Furthermore, we have shown that sensitivity analysis might have been helpful in foreshadowing the breakdown of their model.

In particular, we think that their estimated interest rate semi-elasticity was too low, and that models such as that of Mehra (1992), which have done better in out-of-sample forecasting, do so because they have a higher interest rate elasticity. Moreover, expansive use of volatility in Baba et al. (1992) is not convincing. In particular, this variable entered peculiarly, as a nine-quarter moving average of a moving standard deviation of ex post rate movements, and was not robust to changes in the way it appeared. However one views the in-sample evidence, though, the out-of-sample results speak for themselves. For a specification as complicated as that of Baba et al. (1992), confirming evidence is clearly necessary. Thus, the overwhelming predictive failure in the post-sample period is sufficient to throw out the Baba et al. (1992) specification, in line with the dicta put forth in the Christ-Hendry epigraphs of this paper.25

As emphasized by Goldfeld and Sichel (1990, p. 349), the demand for M1 may be faced with ‘recurrent bouts of instability’, which would make it particularly difficult to establish robust properties of the short-run demand for money. Despite the failure of the Baba et al. (1992) model, we believe that understanding the long-run fundamentals of the demand for M1 is still an important econometric endeavor, due to its aid in implementing monetary policy. Interestingly, Lucas (1988), Poole (1988) and Meltzer (1993) all have argued that a long rate of interest is more important than a short rate of interest when explaining the long-run demand for M1. This dependence may arise because, when assessing their M1 balances, it matters most to agents what will happen to the path of short rates and expected inflation over the future, and the long-term interest rate is better at capturing this information. Although Baba et al. (1992) did not intentionally model the long rate separately in their model, we believe they unknowingly confirmed the conclusion reached by Lucas (1988), Poole (1988) and Meltzer (1993).

To demonstrate this point, Figure 8 plots the relationships between the Baba et al. (1992) smoothed volatility series and the 20-year Treasury bond. As shown in the top panel, both series have the dominant feature of rising dramatically during the early 1980s. However, as mentioned above, the Baba et al. (1992) series requires more than two years of smoothing to obtain this hump, whereas the long-term interest rate is contemporaneously revised. Moreover, in the bottom panel we present a cross-plot of the data for the two series for the full sample, and note that the correlation coefficient between the two series is 0.80. The Baba et al. (1992) volatility result, therefore, could be replicating the long-run fundamentals of money demand as discussed by Lucas (1988), Poole (1988), and Meltzer (1993).

Appendix

Data Availability Issues

- The passbook savings rate, the commercial bank savings rate, and the money market mutual fund rate are used only to construct a learning curve-adjusted maximum M2 rate. Hence, the only rate required is the maximum for a given

---

25 Allan Meltzer stressed this point in private correspondence with us.
20-Year Bond Rate Versus MA(9) Volatility

Figure 8.
period. After the introduction of small time certificates of deposits (CDs) in 1965, they quickly dominated passbook savings. A property of the Baba et al. (1992) data is that starting in the mid 1980s, rates on money funds dominated those paid on CDs. However, this conclusion is not supported by the Federal Reserve data. According to the Federal Reserve data, the two rates have been extremely close since 1985, with CDs more likely than money funds to have a higher rate. These two rates are close enough that either one might be a good proxy for the other. This is still a messy issue which should be resolved.

- In the beginning of 1986, the rate ceiling on NOW accounts was eliminated and there ceased to be any distinction between NOWs and Super NOWs, and hence no data were collected on separate accounts. Baba et al. (1992) averaged the two rates over their entire sample period. This led to the averaging of a positive rate with zero in the period prior to the introduction to Super NOWs. Without taking learning into account, this procedure means that the average rate is therefore about 2.5% prior to Super NOWs, and about 6% a year after their introduction. In actuality, Super NOWs were rarely more than 2% above regular NOWs. And considering their minimum balance requirements, which were as high as $2,500 over this period, putting so much weight on them seems unreasonable. In any case, our data set ends this averaging in 1985. From 1986 on, we used the learning-adjusted rate on Super NOWs, which should be the same thing as NOWs anyway. It is true that the Baba et al. (1992) NOW and Super NOW rates were very different in the post-1986:Q1 period. This should not be the case.

- The Baba et al. (1992) NOW rate fluctuated during the early 1980s, when in fact NOW rates were fixed by ceilings.

- The 20-year Treasury bond was discontinued at the end of 1986. The Baba et al. (1992) source for this series, AHYYS, reports that the 20-year rate is interpolated from data on the 10- and 30-year rates.

- Although MMDAs were excluded by Baba et al. (1992) from the calculation of ΔRMA, it appears that after learning adjustments they were never the maximum rate paid within M2. Thus, their inclusion would not modify ΔRMA.

- We have been unable, as of yet, to fully replicate the last two years of the M1 series used in Baba et al. (1992). They reported that the M1 series was obtained from the Citibase October 1989 data tape. We have examined numerous M1 data series from the Federal Reserve’s archives and have pulled data from the October 1989 Citibase data bank available on floppy disks and these do not correspond to their data. As a final note, Citibase reports that they receive their M1 data from the Federal Reserve Board.
References


