The Economic Journal, **103** (*September*), 1158–1169. © Royal Economic Society 1993. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 238 Main Street, Cambridge, MA 02142, USA.

THE DEMAND FOR M1 IN THE USA: A REPLY TO JAMES M. BOUGHTON*

David F. Hendry and Ross M. Starr

We are pleased to see James Boughton's contribution to resolving the problem of the constancy of the US narrow-money demand relationship. The issue of empirical model selection that he raises is an important one and merits careful analysis. Specifically, we believe that Boughton is making four distinct claims:

1. An econometric model other than that in Baba, Hendry and Starr (1992 – denoted BHS) can be found by general-to-simple methods, with the property that it fits almost as well as BHS. However, the models differ greatly in their economic implications and dynamic adjustments. This is an issue of uniqueness.
2. The new model is not encompassed by BHS, but does not encompass BHS either. This is an issue of dominance.
3. The new model is constant although it does not use all of the new variables in BHS. This is an issue of constancy.
4. Therefore, the new variables BHS do not matter for fit or constancy: the improvement over previous studies is due to more flexible dynamics. This is an issue of the robustness of implications of empirical models.

We summarise these four points as arguing that it is possible to find another congruent, undominated, constant model using only the conventional variables in money demand analysis.

Boughton's paper is mainly a critique of previous studies which failed to find constant relations using alternative methods (see Judd and Scadding (1982) for a survey of earlier studies). In most respects, it demonstrates the practical utility of reduction methods applied to well established problems. Nevertheless, it is important to clarify the precise implications of Boughton's findings and check the validity of the claims in 1. to 4. above. Alternatively, our note can be seen as a further contribution to the literature on how to judge between contending, apparently equivalent, empirical models. We address the four issues in turn, and believe all four can be resolved. The outcome confirms the importance of the novel variables in BHS, shows that the encompassing relationship is one-way only here, and highlights an old paradox of 'significance' in a multiple selection exercise.

Section I considers uniqueness, Section II addresses dominance, Section III analyses constancy and Section IV comments on robustness of implications. Section V concludes.

* We are grateful to Mike Clements and Neil Ericsson for helpful comments on a previous draft. Financial support from the United Kingdom Economic and Social Research Council under grant R000231184 is gratefully acknowledged.

I. MODEL UNIQUENESS: GOODNESS-OF-FIT AND ECONOMIC INTERPRETATION

We attempted to replicate equation (2) in Boughton (1992) and although our final effort is close to the equation Boughton reports, it is not identical. Our estimates are recorded as equation (1) below. The data were slightly revised between the data set that Neil Ericsson provided to Boughton (and used in Hendry and Ericsson, 1991) and that underlying the reported results in the final version of BHS. We doubt if any substantive changes would result from using the revised data. Our note is mainly concerned with the principles of comparison between empirical models, so we proceed on the assumption that the two sets of results are close enough not to prejudice the outcome either way. The variables and data sources are fully described in BHS but in summary are quarterly for the period 1960(3)–1988(3), where:

$m_t = \log \mathrm{M1}$, seasonally adjusted; $p_t = \log \mathrm{GNP}$ deflator, seasonally adjusted (base 1982); $y_t = \log$ real GNP, seasonally adjusted; R_t = 20-year Treasury bond yield to maturity; R_{1t} = 1-month Treasury bill coupon equivalent yield; $S_t = R_t - R_{1t}$; AS_t^* = risk-adjusted average spread: $AS_t - 0{\cdot}57 V_t - 6{\cdot}4 \Delta SV_{t-1}$; R_{mat} = learning-adjusted maximum yield on instruments in M2; R_{nsat} = learning-adjusted other checkables rate in M1; V_t = volatility measure based on long-bond holding-period yields; $SV_t = \max(0, S_t) V_t$; $\Delta_i x_t = (x_t - x_{t-1})/i$ for any variable x_t; $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1}$; $Ax_t = \frac{1}{2}(x_t + x_{t-1})$; $\Delta\hat{p}_t = \Delta p_t + \Delta^2 p_t$; D_t = Dummy for 1980(2) = −1; 1980(3) = +1.

Our replication of Boughton's equation (2) is shown in (1):

$$\begin{aligned}\Delta\widehat{(m-p)}_t = {} & \underset{(0{\cdot}004)}{0{\cdot}037} - \underset{(0{\cdot}003)}{0{\cdot}030}\,(m_{t-4} - p_{t-4} - y_{t-1}) - \underset{(0{\cdot}035)}{0{\cdot}386}\,(\Delta R_{t-3} + 4\Delta_4 R_t) \\ & + \underset{(0{\cdot}050)}{0{\cdot}220}\,\Delta y_t - \underset{(0{\cdot}102)}{0{\cdot}753}\,\Delta p_t - \underset{(0{\cdot}122)}{0{\cdot}274}\,\Delta_4 p_{t-2} + \underset{(0{\cdot}058)}{0{\cdot}275}\,\Delta^2 m_{t-1} - \underset{(0{\cdot}062)}{0{\cdot}247}\,\Delta R_{1t-1} \\ & + \underset{(0{\cdot}046)}{0{\cdot}158}\,\Delta(R_{1t-2} + R_{1t-4}) + \underset{(0{\cdot}046)}{0{\cdot}184}\,(R_{1t-1} - R_{t-1} - 2\Delta_2 R_{nsat-2}) + \underset{(0{\cdot}003)}{0{\cdot}024}\,D_t \end{aligned}$$

$$\mathrm{R}^2 = 0{\cdot}859,\ \hat{\sigma} = 0{\cdot}438\%,\ \mathrm{F}(10, 102) = 62{\cdot}39,\ \mathrm{DW} = 1{\cdot}99,\ \mathrm{SC} = -10{\cdot}51. \qquad (1)$$

Standard errors are in parentheses, estimation was by OLS, and the full sample was used. R^2 is the squared multiple correlation coefficient, $\hat{\sigma}$ is the equation standard error, F(10, 102) tests the null of no relationship, DW is the Durbin–Watson statistic, and SC is the Schwarz criterion (see Doornik and Hendry, 1992).

First, the goodness-of-fit of (1) as measured by $\hat{\sigma}$ is somewhat poorer than (2) in Boughton (1993) but remains good. Such a good fit is a natural part of the design process in a general-to-simple modelling exercise since $\hat{\sigma}$ is the standard deviation of the innovation error. For evaluation, it is irrelevant how (2) was obtained, but the method of selection may influence the importance attached to the existence of a well-fitting rival explanation.

Second, the economic interpretation of (1) is far from clear. Many of the parameter constraints are less than transparent, especially the lagged long-short spread being added to twice the change in the own rate lagged two periods, and the term $\Delta R_{t-3}+4\Delta_4 R_t (= R_t + R_{t-3} - 2R_{t-4}$, noting the definition of Δ_j). Also, several coefficients have signs which are difficult to interpret, especially the negative impact of the own rate R_{nsa}, and the mix of positive and negative signs on the bill rate changes. Finally, the error correction coefficient is small so, interpreting (1) as a transactions demand equation, the long lag in responding to income changes is peculiar.

There is always an important role for economic theory in model selection. Although theory is neither definitive nor necessarily the same for all investigators, *ceteris paribus* equations should be interpretable. When theory and evidence conflict in their support of any given equation, an awkward choice may have to be made. However, here, on both these counts, (1) is dominated by BHS.

II. MODEL DOMINANCE: ENCOMPASSING COMPARISONS

Boughton presents an encompassing test between (1) and the model in BHS using the Davidson–MacKinnon J-test. The test outcomes that he reports suggest that neither model encompasses the other. Rejection of (1) is indeed decisive, but the t-test value for (1) against BHS is only 3·27, which is smaller than the value of 3·46 required by BHS to retain variables in their model. The reasons for the apparent failure of encompassing can be determined by examining what happens in the joint model. Further, the F-encompassing test is the only test which is invariant to specifying the alternative model: the latter could be the rival model as formulated by its proprietor; the orthogonal complement of the model under test (relative to the rival model); the joint (nesting) model; or any linear combination of the two models.

For convenience, we first report the estimates of the preferred model (22) from BHS (the small changes from BHS are due to improved accuracy by storing transformed data variables in double precision in PcGive7: see Doornik and Hendry, 1992):

$$\Delta\widehat{(m-p)}_t = \underset{(0\cdot021)}{0\cdot358} + \underset{(0\cdot003)}{0\cdot013}\, D_t + \underset{(0\cdot070)}{0\cdot370}\, \Delta Ay_t - \underset{(0\cdot129)}{1\cdot066}\, \Delta_4 p_{t-1}$$
$$- \underset{(0\cdot046)}{0\cdot341}\, \Delta\hat{p}_t - \underset{(0\cdot049)}{0\cdot260}\, \Delta R_{mat} - \underset{(0\cdot105)}{1\cdot428}\, AS_t^* - \underset{(0\cdot063)}{0\cdot985}\, AR_{1t} + \underset{(0\cdot051)}{0\cdot465}\, R_{nsat}$$
$$- \underset{(0\cdot015)}{0\cdot253}\, (m-p-\tfrac{1}{2}y)_{t-2} - \underset{(0\cdot098)}{0\cdot348}\, \Delta_4(m-p)_{t-1} - \underset{(0\cdot040)}{0\cdot148}\, \Delta^2(m-p)_{t-4}.$$

$$R^2 = 0\cdot889,\ \hat{\sigma} = 0\cdot391\,\%,\ F(11, 01) = 73\cdot32,\ DW = 1\cdot79,\ \dot{S}C = -10\cdot70. \quad (2)$$

The results from estimating the joint model over the full sample are revealing and are recorded in (3). Common regressors are shown first, then those from (1), and finally those from (2), with the three sets separated by ‖:

$$\widehat{\Delta(m-p)}_t = \underset{(0.041)}{0.250} + \underset{(0.003)}{0.015}\,D_t$$
$$\Vert -\underset{(0.014)}{0.016}\,(m_{t-4}-p_{t-4}-y_{t-1}) + \underset{(0.065)}{0.053}\,(R_{1t-1}-R_{t-1}-2\Delta_2 R_{nsat-2})$$
$$+\underset{(0.073)}{0.052}\,\Delta y_t - \underset{(0.066)}{0.190}\,\Delta p_t - \underset{(0.442)}{0.208}\,\Delta_4 p_{t-2} + \underset{(0.066)}{0.134}\,\Delta^2 m_{t-1}$$
$$-\underset{(0.066)}{0.142}\,\Delta R_{1t-1} - \underset{(0.052)}{0.006}\,(R_{1t-2}+R_{1t-4}) - \underset{(0.058)}{0.103}\,(\Delta R_{t-3}+4\Delta_4 R_t)$$
$$\Vert -\underset{(0.177)}{1.033}\,AS_t^* - \underset{(0.120)}{0.671}\,AR_{1t} - \underset{(0.061)}{0.256}\,\Delta R_{mat} + \underset{(0.108)}{0.337}\,\Delta Ay_t$$
$$-\underset{(0.177)}{0.244}\,\Delta\hat{p}_t - \underset{(0.551)}{0.691}\,\Delta_4 p_{t-1} + \underset{(0.096)}{0.290}\,R_{nsat}$$
$$-\underset{(0.036)}{0.164}\,(m-p-\tfrac{1}{2}y)_{t-2} - \underset{(0.135)}{0.257}\,\Delta_4(m-p)_{t-1} - \underset{(0.055)}{0.135}\,\Delta^2(m-p)_{t-4}$$
$$R^2 = 0.905,\ \hat{\sigma} = 0.378\%,\ F(20,92) = 43.97,\ DW = 2.14,\ SC = -10.48$$
$$(1)\ v\ (3):\ F(9,92) = 1.79;\ (2)\ v\ (3): F(10,92) = 4.45. \qquad (3)$$

Almost no variables from Boughton's model (1) actually matter when added to BHSs model (2): only two coefficients exceed twice their standard error and none exceeds 2·5. By way of contrast, six variables from BHS retain significance even on the criterion of $|t| > 3$. The alternative proxies for inflation are highly collinear (Δp_t and $\Delta_4 p_{t-2}$ versus $\Delta\hat{p}_t$ and $\Delta_4 p_{t-1}$) as are the lagged endogenous dynamics, but apart from those effects every other coefficient from BHS remains highly significant. Further, most coefficients in (3) from BHS are recognisably similar to those in (2), whereas most from (1) have been halved or even reduced close to zero. We believe that it is important to report the joint model whenever possible, albeit that such a model may have no economic significance. The reported simplification encompassing F-test from (3) to (1) of 1·8 has a probability P of 0·08 under the null, so we do not concur that equation (2) fails to encompass (1) above. The converse F-test of the reduction from (3) to (2) has a probability of essentially zero under the null of a valid reduction.

The results are even more clear-cut against the baseline model (18) in BHS which Boughton uses (this did not restrict the role of the volatility variables V_t and ΔSV_{t-1} through AS_t^*): eight BHS regressors maintain values of $|t| > 3$

whereas none does from Boughton's model and the largest |t| is just over 2 (for $\Delta^2 m_{t-1}$). The simplification F-test of BHSs (18) against the union of itself and (1) above is F (9, 90) = 1·34 (P = 0·23), with the $\hat{\sigma}$ of the joint model being 0·379% as against BHSs value of 0·385%.

It might be argued that collinearity between regressors in (3) acting as substitutes for a given economic effect could camouflage the significance of a subset of the variables from (1). To check such a possibility, variables with |t|-statistics less than 2 were sequentially eliminated from the smallest upwards to give the largest chance to retaining variables. After such a sequential simplification, the only variable from (1) with $|t| > 2$ after others with $|t| < 2$ were eliminated was $\Delta^2 m_{t-1}$. At only one stage would there have been any doubt about eliminating a variable from (1) first, namely when $\Delta^2 m_{t-1}$, ΔR_{1t-1} and $(\Delta R_{t-3} + 4\Delta_4 R_t)$ remained with t-statistics of 2·22, −2·16 and −1·75, whereas $\Delta_4(m-p)_{t-1}$ had a t of −1·60 (the next smallest t-statistic from the original BHS variables was 4·2). If the odd combination of interest rate changes is eliminated, ΔR_{1t-1} then ceases to be significant also, leaving just $\Delta^2 m_{t-1}$ with $|t| = 2{\cdot}77$. However, this is a variable which in fact was explicitly deleted by the BHS reduction rule since we required $|t| \geqslant 3{\cdot}46$ unless deletion induced a residual diagnostic problem (see BHS, p. 33), which omitting $\Delta^2 m_{t-1}$ did not.

As before, the baseline model in BHS yields clearer results: there is never any doubt about the order of deletion of variables, and no elimination of a variable from (1) other than $\Delta^2 m_{t-1}$ is anywhere near significant.

Thus, we do not accept the claim that (2) does not encompass (1), since it parsimoniously encompasses (3) at the significance levels stated above. Of course, if a lower critical |t|-value is allowed for individual coefficients, then BHS may require extension by $\Delta^2 m_{t-1}$, but we do not deem that to be a major alteration. The issue is similar to that discussed in Hendry and Ungern-Sternberg (1981) as a critique of the decision to eliminate seasonals by Davidson *et al.* (1978): an (in)significant outcome on a joint test is compatible with either significant or insignificant individual tests when these are viewed as one-off tests.

The converse encompassing direction is definitely not acceptable, so we conclude that model (2) captures money demand behaviour better than (1) and hence that a unique empirical model does present itself in this instance. The high value of the J-test cited by Boughton is primarily due to the role of $\Delta^2 m_{t-1}$: for example, if $\Delta^2 m_{t-1}$ is included in the BHS baseline model, the J-test statistic becomes F(1, 97) = 1·79. Of independent interest, the correlation between the residuals on (1) and (2) is only 0·66.

III. MODEL CONSTANCY: A PLETHORA OF EMPIRICALLY CONSTANT MODELS

Even if the analysis in Section II is accepted, so that (2) is deemed to successfully encompass (1), a puzzle remains. Until BHS, there were no empirically constant models and now more than one has been found. We concur with Boughton that (1) satisfies recursive testing; it also passes the new tests of constancy proposed in Hansen (1992) (assuming that all the regressors in (1) are essentially I(0)). Since (1) does not encompass (2) on any simplification test, and the regressors in (2) have non-constant marginal representations as (e.g.) autoregressions, how can (1) be constant?

First, the discovery of an empirically constant relation using the conventional money demand variables demonstrates the power of general-to-simple methods for constructing statistical models. Such equations may suffer from sample dependence, but provide a useful baseline against which to assess econometric models. Second, two different yet empirically constant econometric models of a non-constant data generation process are not impossible (see, for example, the analysis in Hendry (1979) of a situation where a time-series model is constant when an econometric model is not, and the case noted by Smith, 1991). In particular, dynamics can convert structural shifts in levels (which would induce serious predictive failure) to blips in differences (which generate single outliers at the points of change). The Monte Carlo study in Hendry and Neale (1991) shows this result in operation. Thus, constancy tests may have low power in dynamic models against certain non-constant alternatives. Third, the relatively low power of diagnostic tests for constancy is well known, notwithstanding the regularity with which they reject models in practice.

Nevertheless, we do not wish to hide the apparent paradox: given that (1) is empirically constant, then the variables it omits in (3) apparently form a constant linear combination; but if so, then how can they be the explanation for the failure of previous models? Such an outcome would support Boughton's fourth claim that the 'common' variables in (1) and (2) account for constancy. A formal analysis casts some light on this issue.

Let y_t denote the dependent variable and $\mathbf{x}_{1t}$, $\mathbf{x}_{2t}$ and $\mathbf{x}_{3t}$ denote the distinct regressor variables in (1) and (2) respectively, and the common regressors. The nesting linear model is given by:

$$y_t = \mathbf{x}_{1t}'\boldsymbol{\beta}_1 + \mathbf{x}_{2t}'\boldsymbol{\beta}_2 + \mathbf{x}_{3t}'\boldsymbol{\beta}_3 + u_t, \tag{4}$$

where $u_t \sim \mathrm{IN}(0, \sigma_u^2)$. Equation (1) corresponds to omitting $\mathbf{x}_{2t}$, which involves marginalising by the model:

$$\mathbf{x}_{2t} = \boldsymbol{\pi}_{21t}\mathbf{x}_{1t} + \boldsymbol{\pi}_{23t}\mathbf{x}_{3t} + \mathbf{v}_t, \tag{5}$$

where $\mathbf{v}_t \sim \mathrm{IN}(\mathbf{0}, \boldsymbol{\Omega}_t)$, allowing for possible parameter and variance non-constancies in the marginal model. Eliminating $\mathbf{x}_{2t}$ from (4) using (5) yields:

$$y_t = \mathbf{x}_{1t}'\boldsymbol{\gamma}_{1t} + \mathbf{x}_{3t}'\boldsymbol{\gamma}_{3t} + w_t, \tag{6}$$

where $\boldsymbol{\gamma}_{1t} = \boldsymbol{\beta}_1 + \boldsymbol{\pi}_{21t}'\boldsymbol{\beta}_2$, with an error $w_t = \mathbf{v}_t'\boldsymbol{\beta}_2 + u_t$. Since $\boldsymbol{\beta}_1 = \mathbf{0}$ when BHS is the correct specification, the coefficient vector $\boldsymbol{\gamma}_{1t}$ of (1) should equal $\boldsymbol{\pi}_{21t}'\boldsymbol{\beta}_2$, with an equation error variance of $\sigma_{wt}^2 = \sigma_u^2 + \boldsymbol{\beta}_2'\boldsymbol{\Omega}_t\boldsymbol{\beta}_2$. The empirical constancy of $\boldsymbol{\gamma}_{1t}$ in (1) suggests that $\boldsymbol{\pi}_{21t}$ is relatively constant, matching the claim that the functions of the conventional variables in (1) pick up effects that registered as non-constant in other models.

Figs. 1 and 2 respectively show the 1-step forecast errors with $0 \pm 2\hat{\sigma}$ from

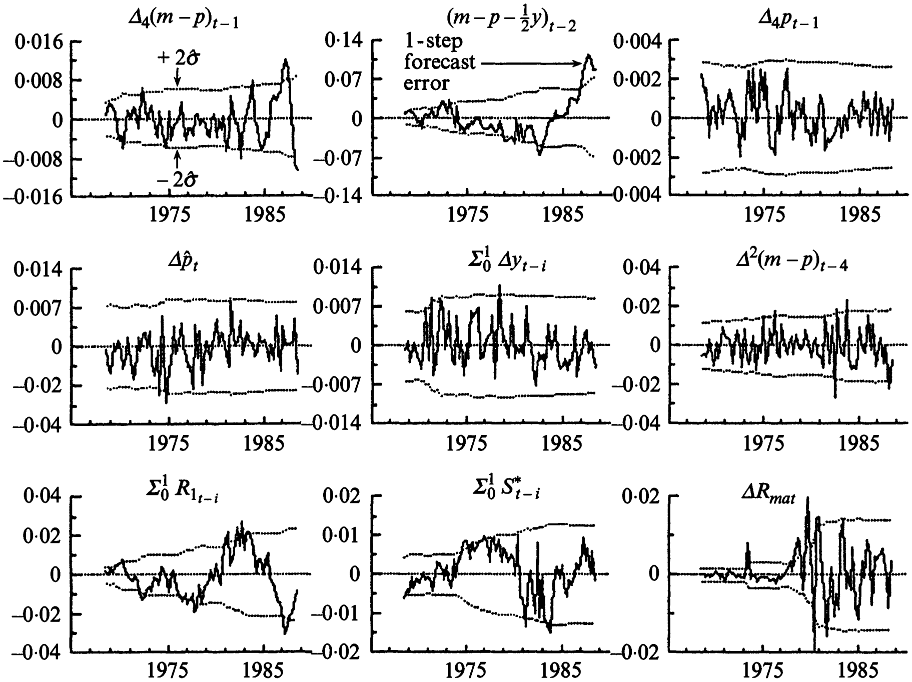

Fig. 1. 1-Step forecast errors with $0 \pm 2\hat{\sigma}$ for $\mathbf{x}_{2t}$ on $\mathbf{x}_{1t}$.

regressing each variable in $\mathbf{x}_{2t}$ on $\mathbf{x}_{1t}$ and $\mathbf{x}_{3t}$, and the sequence of break-point Chow tests scaled by their 0·1 % significance level at each sample size (R_{nsat} is omitted from the graphs as uninformative). The inflation and income relations are relatively constant, but the constancy of all other elements is rejected empirically at a high level of significance. Consequently, the apparent constancy of $\boldsymbol{\gamma}_{1t}$ is not due to the constancy of $\boldsymbol{\pi}_{21t}$.

Even more surprising is that $\mathbf{v}_t' \boldsymbol{\beta}_2$ should have a constant variance although at first sight that too seems consistent with the evidence in (1). By imposing the coefficients in (1) at their full sample estimates $\boldsymbol{\gamma}_{1(T)}$ and $\boldsymbol{\gamma}_{3(T)}$ for $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_3$ from (6), consider the equation:

$$y_t - \mathbf{x}_{1t}' \boldsymbol{\gamma}_{1(T)} - \mathbf{x}_{3t}' \boldsymbol{\gamma}_{3(T)} = w_{t(T)}. \tag{7}$$

Given (7), from (4):

$$w_{t(T)} = \mathbf{x}_{1t}'(\boldsymbol{\beta}_1 - \boldsymbol{\gamma}_{1(T)}) + \mathbf{x}_{2t}' \boldsymbol{\beta}_2 + \mathbf{x}_{3t}'(\boldsymbol{\beta}_3 - \boldsymbol{\gamma}_{3(T)}) + u_t. \tag{8}$$

Estimation of (8) using the full sample residuals for $w_{t(T)}$ from (1), where $\hat{\sigma}_{w(T)} = 0{\cdot}438\,\%$, yields five significant coefficients in $(\boldsymbol{\beta}_1 - \boldsymbol{\gamma}_{1(T)})$. This suggests that $\boldsymbol{\gamma}_{1(T)}$ is not in fact equal to a constant vector $\boldsymbol{\beta}_1$.

Alternatively, it is possible to marginalise with respect to the variables $\mathbf{x}_{1t}$ in (1) using:

$$\mathbf{x}_{1t} = \mathbf{P}_{1(T)}' \mathbf{x}_{2t} + \mathbf{P}_{3(T)}' \mathbf{x}_{3t} + \boldsymbol{\xi}_t, \tag{9}$$

which is the full sample reverse direction of regression to (5). Figs. 3 and 4 show

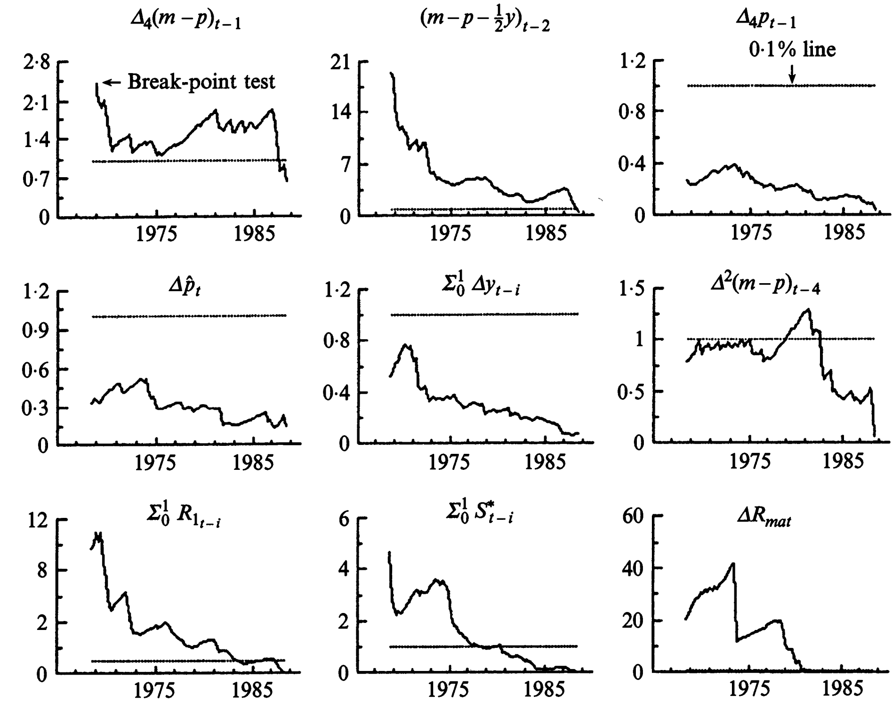

Fig. 2. Break-point Chow tests scaled by 0·1% significance levels for $\mathbf{x}_{2t}$ on $\mathbf{x}_{1t}$.

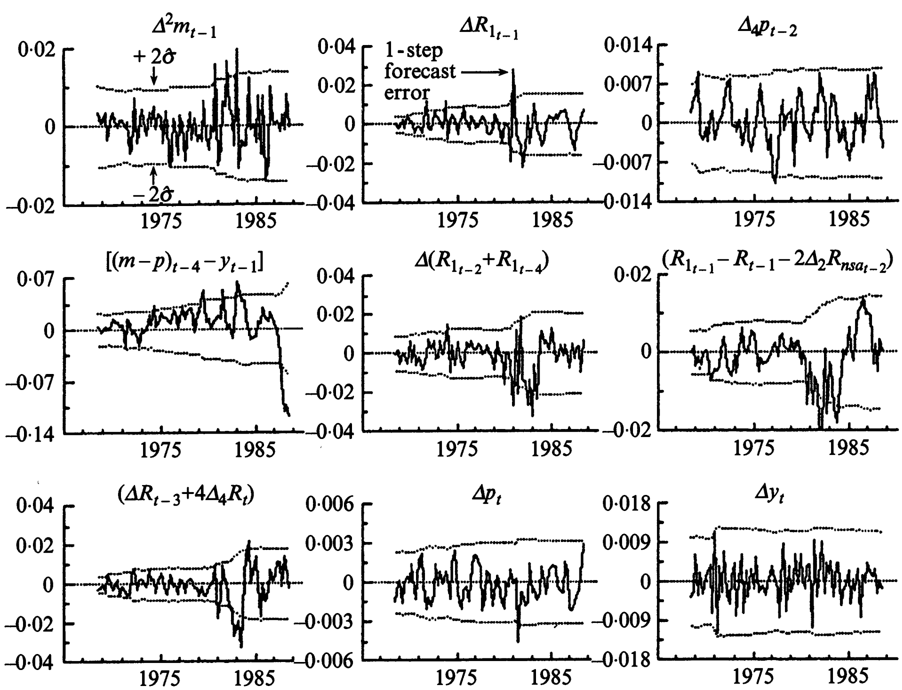

Fig. 3. 1-Step forecast errors with $0 \pm 2\hat{\sigma}$ for $\mathbf{x}_{1t}$ on $\mathbf{x}_{2t}$.

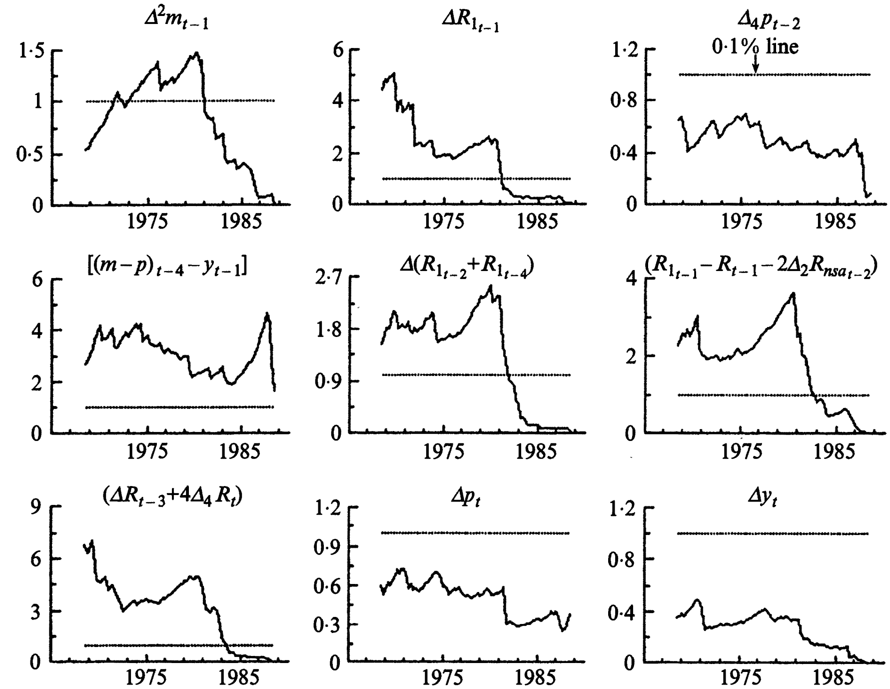

Fig. 4. Break-point Chow tests scaled by 0·1 % significance levels for $\mathbf{x}_{1t}$ on $\mathbf{x}_{2t}$.

the 1-step forecast errors with $0 \pm 2\hat{\sigma}$ from regressing each variable in $\mathbf{x}_{1t}$ on $\mathbf{x}_{2t}$ and $\mathbf{x}_{3t}$, and the associated sequence of break-point Chow tests scaled by their 0·1 % significance levels as in Fig. 2. Again the inflation and income relations are constant, but all other relationships are empirically non-constant at a high level of significance. Using (9) in (8) yields:

$$w_{t(T)} = \mathbf{x}'_{2t}\boldsymbol{\delta}_{2(T)} + \mathbf{x}'_{3t}\boldsymbol{\delta}_{3(T)} + \eta_t, \tag{10}$$

where $\eta_t = u_t + \boldsymbol{\xi}'_t(\boldsymbol{\beta}_1 - \boldsymbol{\gamma}_{1(T)}) = u_t + \boldsymbol{\xi}'_t\boldsymbol{\delta}_{1(T)}$. Hence, $w_{t(T)}$ can be decomposed using (10) and yields $\boldsymbol{\delta}$ and σ_η (= 0·415 %), both empirically constant. However, calculating the series $\mathbf{x}'_{2t}\hat{\boldsymbol{\delta}}_{2(T)}$ revealed its full sample standard deviation to be 0·141 % but highly non-constant (on both recursive Chow tests and the tests in Hansen, 1992). This term raises σ_η over σ_u by only a small percentage, so is probably undetectable against the sampling variability in the multivariate analysis. Indeed, if $(w_{t(T)} - u_{t(T)})$ is regressed on $\mathbf{x}_{2t}$ and $\mathbf{x}_{3t}$, non-constancy is barely detectable, despite the evidence in Figs. 3 and 4 for the non-constancy of the variance of $\boldsymbol{\xi}_t$.

Such an analysis suggests that σ_w is actually non-constant, but hard to detect, consistent with the substitute nature of some of the regressors in the competing models (1) and (2). To summarise this result, the error variance of (1) contains a non-constant component, but the constancy tests fail to detect this.

An important additional result now follows: empirical constancy within sample can be achieved by design just like, for example, innovation errors or homoscedastic residuals. Using an explicit notation for time, write a regression model as:

$$\mathbf{y}_t^1 = \mathbf{X}_t^1 \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t^1 \quad \text{for} \quad t = 1, \ldots, T, \tag{11}$$

where there are K regressors and at sample point t:

$$\hat{\boldsymbol{\beta}}_t = (\mathbf{X}_t^{1\prime}\mathbf{X}_t^1)^{-1}\mathbf{X}_t^{1\prime}\mathbf{y}_t^1 \quad \text{for} \quad t = M > K, \ldots, T. \tag{12}$$

From the sequence of coefficient estimates, generate the 1-step residuals:

$$\nu_t = y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}}_{t-1} \quad \text{with} \quad \text{RSS}_t = \text{RSS}_{t-1} + \nu_t^2/\omega_t, \tag{13}$$

where

$$\boldsymbol{\lambda}_{t+1} = (\mathbf{X}_t^{1\prime}\mathbf{X}_t^1)^{-1}\mathbf{x}_{t+1} \quad \text{and} \quad \omega_t = (1 + \mathbf{x}_t' \boldsymbol{\lambda}_t). \tag{14}$$

The $\{\nu_t\}$ in (13) are the basis for Chow tests, and changes in their variances are crucial since:

$$\text{RSS}_j = \sum_{t=M}^{j} \nu_t^2/\omega_t. \tag{15}$$

Thus, the key issue is whether it is possible to 'stabilise' the variances of the $\{\nu_t\}$ over the sample. Here, we consider doing so by parameter restrictions which remove the power of constancy tests just as COMFAC transforms remove the power of low-order tests for residual autocorrelation (see Sargan (1980) and Hendry and Mizon, 1978). Let the full-sample residuals be:

$$\hat{\epsilon}_t = y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}}_T = \nu_t - (\hat{\boldsymbol{\beta}}_T - \hat{\boldsymbol{\beta}}_{t-1})' \mathbf{x}_t, \tag{16}$$

then we also have:

$$\text{RSS}_\text{T} = \sum_{t=1}^{T} \hat{\epsilon}_t^2. \tag{17}$$

Consider an extreme example where $\hat{\boldsymbol{\beta}}_t$ changes greatly if estimated in each subsample and the $\{\nu_t\}$ reflect this, so that non-constancy is detected on many subsamples. Fix $\boldsymbol{\beta}$ at $\hat{\boldsymbol{\beta}}_T$ so no free parameters remain to estimate. By definition, RSS_T cannot alter by imposing such a restriction, but from (16), the 1-step residuals essentially become the full sample residuals. This formulation will increase the RSS in the earlier part of the period, bringing it closer to constancy, especially on recursive constancy tests. The resulting $\hat{\epsilon}_t$ may be heteroscedastic, but we already know how to transform such residuals to homoscedasticity. Thus, both parameter and variance non-constancy can be camouflaged in part, either inadvertently, or deliberately if that is a design aim: such an outcome may or may not be 'spurious' (see Smith, 1991). The result will certainly look spurious if every parameter has to be fixed, but most models have many parametric restrictions imposed (as do both (2) and (1)), which may serve to induce greater constancy as well as interpretability. To summarise this result, within-sample empirical constancy of a model may be a feature which can be achieved by design, even when the underlying process is not constant. In practice, it may be hard to determine the precise restrictions required, but in principle it seems to be possible.

IV. MODEL ROBUSTNESS: DO THE NEW BHS VARIABLES MATTER?

It is not clear that any issue of robustness of implications remains once (2) encompasses (1) but not conversely. The constancy of (1) by itself does not prove that the new variables do not matter: section (2) above demonstrated that they do matter in several senses. Since pure autoregressions in nominal money have standard errors of less than 0·7%, relatively small differences in the residual standard error can correspond to significant omissions.

The encompassing failure of (1) for (2) is important from a policy perspective. Suppose a policy intends to alter a given variable, say, z_t, where z_t is not included in a specific empirical model, but is in a rival theoretical or empirical model and the behaviour of z_t has previously altered. Then the empirical constancy of the first model is only necessary, and is not sufficient, to justify ignoring the policy variable: a direct test of irrelevance is required. The joint nesting test is one possibility which was shown above to yield fruitful insights.

V. CONCLUSION

We conclude that the BHS model survives the critique presented by Boughton's model, and leaves the former as the only constant, congruent, encompassing model of M1 demand in the United States. The BHS model transpires to be highly robust to an attempt to respecify it, and none of its implications are impugned by the existence of an encompassed rival.

Nuffield College

University of California at San Diego

Date of receipt of final typescript: February 1993

REFERENCES

Baba, Y., Hendry D. F. and Starr, R. M. (1992). 'The demand for M1 in the USA, 1960–1988.' *Review of Economic Studies*, vol. 59, pp. 25–61.

Boughton, J. M. (1993). 'The demand for M1 in the United States: a comment on Baba, Hendry and Starr.' ECONOMIC JOURNAL, this issue, pp. 1154–7.

Davidson, J. E. H., Hendry, D. F., Srba, F. and Yeo, S. (1978). 'Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom.' ECONOMIC JOURNAL, vol. 88, pp. 661–92.

Doornik, J. A. and Hendry, D. F. (1992). *Pc-Give 7: An Interactive Econometric Modelling System.* Oxford: Institute of Economics and Statistics, University of Oxford.

Hansen, B. E. (1992). 'Testing for parameter instability in linear models.' *Journal of Policy Modeling*, vol. 14, no. 4, pp. 517–33.

Hendry, D. F. (1979). 'Predictive failure and econometric modelling in macroeconomics: the transactions demand for money.' In *Economic Modelling* (P. Ormerod, ed), pp. 217–42. London: Heinemann Education Books.

Hendry, D. F. and Ericsson, N. R. (1991). 'Modeling the demand for narrow money in the United Kingdom and the United States.' *European Economic Review*, vol. 35, pp. 833–81.

Hendry, D. F. and Mizon, G. E. (1978). 'Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England.' ECONOMIC JOURNAL, vol. 88, no. 351, pp. 549–63.

Hendry, D. F. and Neale, A. J. (1991). 'A Monte Carlo study of the effects of structural breaks on tests for unit roots'. In (P. Hackl and A. H. Westlund, eds), pp. 95–119. *Economic Structural Change, Analysis and Forecasting*. Springer-Verlag.

Hendry, D. F. and von Ungern-Sternberg, T. (1981). 'Liquidity and inflation effects on consumers' expenditure.' In (A. S. Deaton, ed), *Essays in the Theory and Measurement of Consumer Behaviour*, pp. 237–61. Cambridge: Cambridge University Press.

Judd, J. P. and Scadding, J. L. (1982). 'The search for a stable money demand function: a survey of the post-1973 literature'. *Journal of Economic Literature*, vol. 20, pp. 993–1023.

Sargan, J. D. (1980). 'Some tests of dynamic specification for a single equation.' *Econometrica*, vol. 48, no. 4, pp. 879–97.

Smith, R. P. (1991). 'Spurious structural stability.' *Manchester School*, vol. 59, pp. 419–23.