The UK Demand for Broad Money over the Long Run

Neil R. Ericsson, David F. Hendry, and Kevin M. Prestwich*

May 1, 1997

Abstract

Using data from Friedman and Schwartz (1982), Hendry and Ericsson (1991a) developed an annual model of the demand for broad money in the UK over the sample $1878-1975$. We update that model through 1993 on data extended by Attfield, Demery and Duck (1995), taking account of changed data definitions. The concept of parameter constancy is discussed. Using appropriate measures of opportunity cost and credit deregulation, the model's parameters are empirically constant over the extended sample, which was economically rather turbulent. Models based on Friedman and Schwartz's phase-average data fare poorly, both on the original sample and on the extended dataset. Some policy implications are drawn.

1 Introduction

We are delighted to contribute to a volume celebrating the centenary of the *Scandinavian Journal of Economics*, and we offer our birthday greetings to a journal that has played a distinguished role in the development of our science over the last century. As befits such an occasion, this paper reconsiders a long-standing controversy with important policy implications — the demand for money. Specifically, we examine the demand for broad money using a long historical data record for the United Kingdom.

Evaluation of empirical models over extended samples involves both statistical and conceptual considerations. As a substantive application illustrating both aspects, this paper updates the model in Hendry and Ericsson (1991a) (estimated over 1878–1975) to include data through 1993. Extending the sample creates the opportunity for alternative measures of the opportunity cost of money and of credit derestrictions. Thus, we test for parameter constancy in models using different measures. Mechanistic updates of the measures result in models with predictive failure. However, for coherent measures, the model's short-run and long-run properties remain virtually unchanged over the unusually long forecast period of 18 years, despite major changes in monetary control rules and substantial financial innovation in the economy.

Section 2 briefly reviews the economic theory of the demand for money and defines and describes the data series. Section 3 records the estimated model in Hendry and Ericsson (1991a) and notes its nonconstancy if the model is extended mechanistically over the new data. Section 4 considers how to evaluate and update empirical models over samples involving major changes to the economy. Specifically, for the

*The first and third authors are respectively a staff economist and a research assistant in the Division of International Finance, Federal Reserve Board, Washington, DC. The second author is Leverhulme Personal Research Professor of Economics at Nuffield College, Oxford, England. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting those of the Board of Governors of the Federal Reserve System or other members of its staff. All numerical results were obtained using PcGive Professional Version 8.10 and PcGive Professional Version 9.00: see Doornik and Hendry (1994). Financial support from the UK Economic and Social Research Council under grant R000234954 is gratefully acknowledged.

forecast period 1975–1993, economic extensions of the empirical models are required for changing measurement of money, for the associated changes in the opportunity costs of holding money, and for financial innovation. Data measurement, the opportunity cost, and financial innovation each have implications for parameter constancy. Section 5 extends the economics of the model in Hendry and Ericsson (1991a) to incorporate these issues, and tests for the constancy of the model's parameters. Additionally, this section considers identification and the policy implications. Section 6 concludes. The Appendix compares the performance of the annual model with other models, including with models based on Friedman and Schwartz's phase-average data.

2 Economic Theory and Data Description

This section provides a backdrop for the empirical modeling in Sections 3.1–5. Section 2.1 sketches the standard theory underlying empirical models of money demand. Section 2.2 describes the data modeled and characterizes the data's basic properties.

2.1 Economic Theories of Money Demand

The theoretical and empirical study of the demand for money in the United Kingdom has an impressive history, almost matching the extensive time series now available on money and its main determinants. See Jevons (1884) (reprinted in part as Hendry and Morgan (1995, Chapter 6)), Marshall (1926), Keynes (1930), and Hawtrey (1938) *inter alia* for earlier literature. More recently, the literature has seen an explosion in the modeling of numerous monetary aggregates, both for the United Kingdom and for other countries. Goldfeld and Sichel (1990) extensively review that more recent theoretical and empirical work.

As discussed in these papers and elsewhere, money is demanded for at least two reasons: as an inventory to smooth differences between income and expenditure streams, and as one among several assets in a portfolio. Both demands lead to a long-run specification in which nominal money demanded (M^d) depends on the price level (P), a scale variable (I), and a vector ($\mathbf{R}$, in bold) of rates of returns on various assets:

$$M^d/P \;=\; q(I, \mathbf{R})\,. \tag{1}$$

The function $q(\cdot,\cdot)$ is assumed increasing in I, decreasing in those elements of $\mathbf{R}$ associated with assets excluded from money (M), and increasing in those elements of $\mathbf{R}$ for assets included in M. The opportunity cost is determined through $\mathbf{R}$ and is the focus of Section 4.2.

Commonly, (1) is specified in log-linear form, albeit with interest rates entering in either logs or levels:

$$m^d - p \;=\; \gamma_0 + \gamma_1 i + \gamma_2 R^{own} + \gamma_3 R^{out}\,. \tag{2}$$

Capital letters denote both the generic name and the level, logs are in lower case, and (somewhat symbolically) R^{own} and R^{out} denote the own and outside rates of interest. Anticipated signs of the coefficients are $\gamma_1 > 0$, $\gamma_2 > 0$, and $\gamma_3 < 0$.

2.2 Data Description

This subsection defines the data and presents descriptive properties. Graphs and simple time-series regressions help characterize the data's central properties, which must be considered in empirical modeling.

The basic data series are annual values of the broad money stock (M), real net national income (I), the corresponding deflator (P), short-term and long-term nominal interest rates (RS and $R\ell$), population

(N), and high-powered money (H) for the United Kingdom. Data from 1871 to 1975 are from Friedman and Schwartz (1982). Attfield *et al.* (1995) extend those series over 1976–1993, constructing them from a variety of sources and using several alternative definitions of money. Section 4.1 below discusses these measures of money as such. The variables M, I, H are in £ million; P is such that $1929 = 1.00$; N is in millions; and RS and $R\ell$ are fractions.

Some constructed variables are also of interest. First, under the quantity theory of money, the income elasticity is unity ($\gamma_1 = 1$ in (1)), so a key derived variable is velocity V, constructed as $(I \cdot P)/M$. Second, there are dummy variables. The variables D_1 and D_3 are dummies for World Wars I and II. The dummy D_4 is unity over 1971–75 and zero otherwise: it aims to capture the deregulation of the banking sector with the introduction of Competition and Credit Control in 1971. A similar period of deregulation occurs in 1986–1989 (see Section 4.3 below). As a proxy for both episodes, the dummy D_c is unity for 1971–75 and 1986–1989, and zero otherwise. See Friedman and Schwartz (1982) and Hendry and Ericsson (1991a) for details on the data.

Finally, before examining the data, we introduce a few more conventions. Δ is the difference operator.[1] "Levels" often means the logarithm of the levels, and the context clarifies this. Figures appear as 2×2 panels of graphs, which are labeled with a suffix a, b, c, or d, and which appear in the panels according to their suffix, as follows: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Single and double asterisks (* and **) denote significance at the 5% and 1% levels.

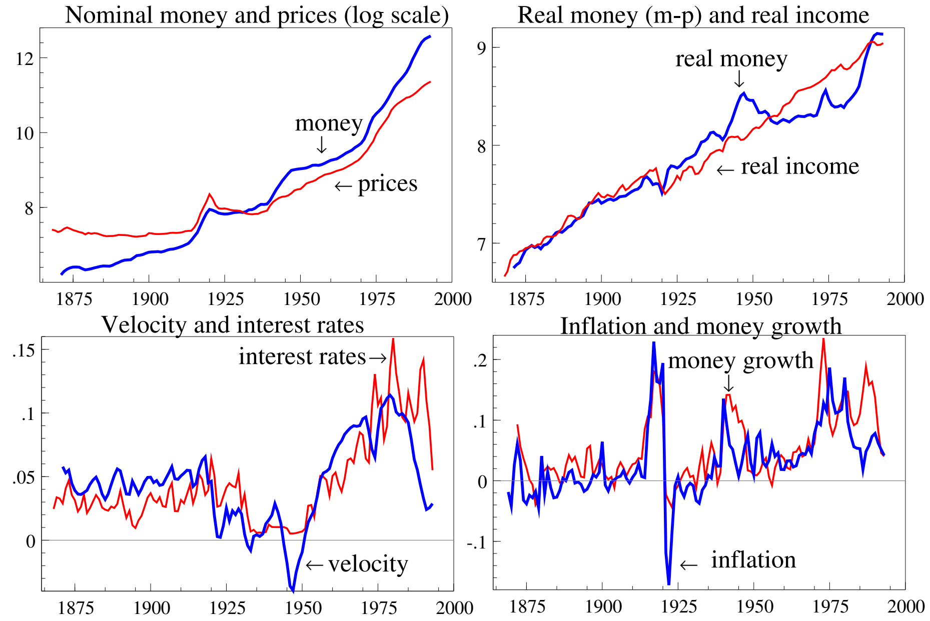

Figure 1 UK time series of (m, p), $(m - p, i)$, v, RS and $\Delta m, \Delta p$.

Figure 1 shows the full-sample time series of (m, p), $(m - p, i)$, (v, RS) and $(\Delta m, \Delta p)$ as a 2×2 panel. Figure 1a emphasizes the huge changes in money and prices over the century. Noting that the figure is in logs (for ease of display), the actual level of money moves from £*492.4* million in 1871 to £*291,173* million in 1993, an increase of almost 600 *fold*. Over the same period, prices rise 55-fold, from 0.579 to 31.38. Thus, a pound sterling in 1993 is equivalent in purchasing power to approximately 2 pence at the beginning of the sample (actually just over 4d). From Figure 1b, real money has increased 11-fold, almost exactly matching the 11.5-fold increase in real national income. To characterize adequately such massive growth and change in money is a serious challenge, particularly in a constant

[1] The difference operator Δ is defined as $(1 - L)$, where the lag operator L shifts a variable one period into the past. Hence, for x_t (a variable x at time t), $Lx_t = x_{t-1}$ and so $\Delta x_t = x_t - x_{t-1}$. More generally, $\Delta_j^i x_t = (1 - L^j)^i x_t$. If i (or j) is undefined, it is taken to be unity.

Table 1 Descriptive unit-root statistics.

	t_{adf}	β	σ	lag	$t_{\Delta Y_i}$	t-prob	F-prob
m	-0.44	0.9975	0.0263	2	-4.58	0.0000	–
p	-1.29	0.9901	0.0396	1	9.47	0.0000	0.167
i	-2.67	0.9133	0.0313	1	3.54	0.0006	0.888
$m-p$	-2.78	0.9345	0.0356	1	7.23	0.0000	0.063
RS	-0.74	0.9849	0.0125	2	-4.13	0.0001	–
$R\ell$	-0.07	0.9993	0.00584	2	-2.22	0.0281	–
v	-2.35	0.9489	0.0428	1	6.05	0.0000	0.5274
Δm	-5.18^{**}	0.7249	0.0262	1	5.01	0.0000	0.4187
Δp	-5.12^{**}	0.6070	0.0394	1	1.59	0.1147	–
Δi	-7.25^{**}	0.0466	0.0315	2	2.35	0.0205	–
$\Delta(m-p)$	-6.72^{**}	0.4105	0.0356	1	2.62	0.0101	0.1514
ΔRS	-8.37^{**}	-0.4998	0.0123	2	1.88	0.063	–
$\Delta R\ell$	-7.23^{**}	-0.0061	0.00576	2	1.80	0.0746	–
Δv	-6.60^{**}	0.4630	0.0434	0	–	–	0.2226

parsimonious model.

As Figure 1b also shows, real money grows most rapidly over the last two decades, notwithstanding the two world wars elsewhere in the sample and despite the U.K. government's attempt at monetary control over much of the 1980s. Correspondingly, velocity falls sharply to levels prevalent in the last century; see Figure 1c. Even relative to the previous century, large changes have occurred over the two decades.

Figure 1d plots the growth in nominal money and the inflation rate. These series move relatively closely overall. There are four notable departures: 1921–22, when prices fell by almost 30% but nominal money by only 5%; World War II and the early 1970s, when money growth greatly exceeded the rate of inflation; and 1985–90, when money growth was about 15% with inflation of 5%.

Univariate autoregressive models can help characterize an important aspect of a time series: its order of integration. Specifically, the augmented Dickey-Fuller (ADF) statistic t_{ADF} can be used to test whether a series is integrated of order k (or $I(k)$), where an $I(k)$ series requires differencing k times to remove all unit roots; see Dickey and Fuller (1981). Table 1 records the ADF statistics over the entire sample 1875–1993 for the key variables. The hypothesis of a unit root cannot be rejected for any variable in levels, whereas it is strongly rejected for all variables in differences, consistent with the visual evidence that differences repeatedly return to their means over the sample. Despite their smooth and trending appearance, m and p appear to be $I(1)$ rather than $I(2)$. The smoothness in m and p may be associated with the main regime shifts over the sample, such as wars and financial innovations. If so, a non-negligible part of their observed variation may be due to non-stationarities arising from structural change. Each ADF statistic is reported for the shortest lag length obtainable (commencing from 2) without dropping a lagged difference significant at the 5% level. However, given their univariate nature and the absence of tests for congruency, such statistics should not be regarded as definitive.

3 Previous Estimates and a Mechanistic Extension

Section 3.1 presents the initial annual model in Hendry and Ericsson (1991a) and their alternative model on a slightly different sample. Section 3.2 evaluates a simple mechanistic extension of their model over the sample 1976–1993, finding strong evidence of parameter nonconstancy. However, the way in which a model is extended over a new sample bears directly on its statistical performance on that sample. That leads to economic extensions of empirical models (Section 4) and a re-evaluation of the model in Hendry and Ericsson (1991a) on the new data (Section 5).

3.1 The Annual Models in Hendry and Ericsson (1991a)

Economic theory offers little guidance in modeling the behavior of money out of equilibrium, beyond saying that adjustments to "desired" levels of money holdings are not likely to be instantaneous due to adjustment costs. In that light, Hendry and Ericsson (1991a) developed a dynamic error correction model (ECM) of broad money M, allowing the economic theory above to define the long-run equilibrium while determining short-run dynamics from the data. That ECM is:

$$\begin{aligned}\Delta(m-p)_t \;=\;& \underset{[0.06]}{0.45}\;\Delta(m-p)_{t-1} - \underset{[0.04]}{0.10}\;\Delta^2(m-p)_{t-2} - \underset{[0.04]}{0.60}\;\Delta p_t \\ &+ \underset{[0.05]}{0.39}\;\Delta p_{t-1} - \underset{[0.006]}{0.021}\;\Delta rs_t - \underset{[0.021]}{0.062}\;\Delta_2 r\ell_t \\ &- \underset{[0.59]}{2.55}\;(\hat{u}_{t-1}-0.2)\hat{u}_{t-1}^2 + \underset{[0.002]}{0.005} + \underset{[0.6]}{3.7}\;(D_1+D_3)_t \end{aligned} \tag{3}$$

$T = 93$ [1878–1970] $\mathsf{R}^2 = 0.87$ $\hat{\sigma} = 1.424\%$ $\mathsf{dw} = 1.82$
$AR : \mathsf{F}(2,82) = 1.39$ $\mathsf{Normality} : \chi^2(2) = 1.9$ $\mathsf{ARCH} : \mathsf{F}(1,82) = 1.51$
$RESET : \mathsf{F}(1,82) = 0.41$ $\mathsf{Hetero} : \mathsf{F}(15,68) = 0.87$ $\mathsf{Form} : \mathsf{F}(43,40) = 0.81$

Here and below, T is the number of observations; R^2 is the squared multiple correlation coefficient; and $\hat{\sigma}$ is the standard deviation of the residuals, expressed as a percentage of real money and adjusted for degrees of freedom. OLS standard errors are in parentheses $(\cdot)$, whereas heteroscedasticity-consistent standard errors are in square brackets $[\cdot]$; see White (1980), Nicholls and Pagan (1983), and MacKinnon and White (1985) on the latter. Equation (3) also includes diagnostic statistics for testing against various alternative hypotheses: residual autocorrelation (dw and AR), skewness and excess kurtosis ($Normality$), autoregressive conditional heteroscedasticity ($ARCH$), RESET ($RESET$), heteroscedasticity ($Hetero$), and heteroscedasticity quadratic in the regressors (alternatively, functional form mis-specification) ($Form$).[2] The null distribution is designated by $\chi^2(\cdot)$ or $F(\cdot,\cdot)$, the degrees of freedom fill the parentheses, and (for AR and $ARCH$) the lag order is the first degree of freedom. Also, $\hat{u}$ in (3) is the equilibrium-correction residual from the static Engle–Granger regression over 1873–1970:

$$(m-i-p)_t \;=\; -0.309 - 7.00RS_t \tag{4}$$

$T = 98$ [1873–1970] $\mathsf{R}^2 = 0.56$ $\hat{\sigma} = 10.86\%$ $\mathsf{dw} = 0.33$ $\mathsf{t_{ADF}} = -2.77$;

see Engle and Granger (1987). Based on research in Escribano (1985), $\hat{u}$ enters (3) nonlinearly.

[2]For references on the test statistics, see Godfrey (1978), Doornik and Hansen (1994), Engle (1982), Ramsey (1969), and White (1980), the latter on both $Hetero$ and $Form$.

Equation (3) appears reasonably well specified in sample, given the diagnostic statistics above and additional results in Hendry and Ericsson (1991a). In particular, Hendry and Ericsson (1991a) provide graphical evidence that (3) is empirically constant, using recursive least squares.

However, (3) as specified is nonconstant over the period 1971–75, following the introduction of Competition and Credit Control (CCC). Hendry and Ericsson (1991a) expand (3) to account for CCC by including the dummy D_4, both directly and interactively with Δrs_t. Estimated over 1878–1975, the resulting model is close to (3) for comparable variables:

$$\begin{aligned}\Delta(m-p)_t \;=\;& \underset{[0.06]}{0.47}\;\Delta(m-p)_{t-1} - \underset{[0.04]}{0.11}\;\Delta^2(m-p)_{t-2} - \underset{[0.04]}{0.59}\;\Delta p_t \\ &+ \underset{[0.05]}{0.41}\;\Delta p_{t-1} - \underset{[0.006]}{0.017}\;\Delta rs_t - \underset{[0.019]}{0.078}\;\Delta_2 r\ell_t \\ &- \underset{[0.19]}{1.16}\;(\hat{u}_{t-1}-0.2)\,\hat{u}^2_{t-1} + \underset{[0.002]}{0.007} + \underset{[0.6]}{3.4}\;(D_1+D_3)_t \\ &+ \underset{[0.010]}{0.071}\;D_{4t} + \underset{[0.020]}{0.090}\;D_{4t}\cdot\Delta rs_t \qquad (5)\end{aligned}$$

$$T = 98\;[1878\text{–}1975] \quad \mathsf{R}^2 = 0.88 \quad \hat{\sigma} = 1.48\% \quad \mathsf{dw} = 1.89.$$

Figure 2 shows the fitted and actual values, scaled residuals, residual correlogram, and residual density for (5). The model appears congruent within this sample against the available information.

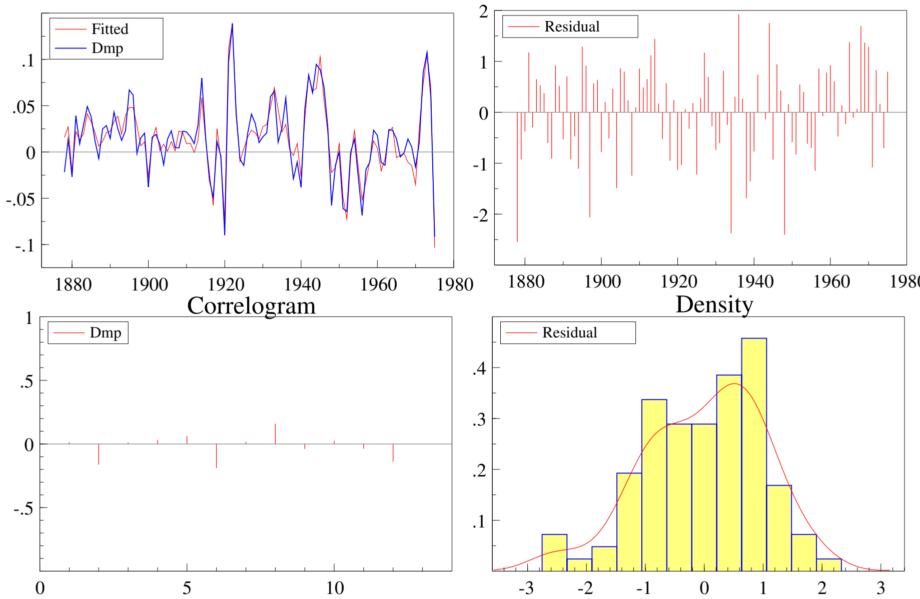

Figure 2 Fitted and actual values, scaled residuals, correlogram and density over 1878-1970.

3.2 Predictive Failure of a Mechanistic Extension of the Model

In evaluating a model over a new sample, a common approach updates the model mechanistically, simply "plugging in" the new data into the existing equation. For (5) and the data described in Section 2.2, the results are as follows:

$$\Delta(m-p)_t \;=\; \underset{[0.07]}{0.54}\;\Delta(m-p)_{t-1} - \underset{[0.05]}{0.11}\;\Delta^2(m-p)_{t-2} - \underset{[0.05]}{0.63}\;\Delta p_t$$

$$
\begin{aligned}
&+ \underset{[0.05]}{0.47}\ \Delta p_{t-1} - \underset{[0.007]}{0.007}\ \Delta rs_t - \underset{[0.0171]}{0.089}\ \Delta_2 r\ell_t \\
&- \underset{[0.045]}{0.056}\ (\hat{u}_{t-1} - 0.2)\hat{u}_{t-1}^2 + \underset{[0.003]}{0.009} + \underset{[0.7]}{3.1}\ (D_1 + D_3)_t \\
&+ \underset{[0.009]}{0.050}\ D_{c,t} + \underset{[0.03]}{0.13}\ D_{4,t} \cdot \Delta rs_t
\end{aligned} \tag{6}
$$

$T = 116$ [1878–1993] $\mathsf{R}^2 = 0.82$ $\hat{\sigma} = 1.9094\%$

$AR : \mathsf{F}(2, 106) = 4.63^{**}$ Normality : $\chi^2(2) = 4.5$ ARCH : $\mathsf{F}(1, 106) = 0.003$

$RESET : \mathsf{F}(1, 107) = 0.87$ Hetero : $\mathsf{F}(20, 87) = 3.69^{**}$

$Chow : \mathsf{F}(18, 90) = 5.265^{**}$ $\mathsf{Jt} = 2.58$ $\mathsf{V} = 0.86^{**}$

where Jt and V are the statistics from Hansen (1992) for testing joint parameter constancy and variance constancy. Despite the benefit of using D_c (the extended dummy for credit loosening), the Chow (1960) predictive-failure statistic ($Chow$) strongly rejects constancy. The statistic V likewise rejects constancy; and Jt nearly rejects at the 90% critical value, even though it has 12 degrees of freedom. Several residual diagnostic tests also reject.

Predictive failure is extensive and not isolated to one or two observations. Figure 3 demonstrates this failure through the fitted, forecast, and actual values for $\Delta(m-p)_t$, their cross-plots, the equation's residuals, and the equation's forecasts (with ± 2 standard error bars) compared with actual outcomes. Both numerically and statistically, the model's forecast performance is far worse than its in-sample behavior.

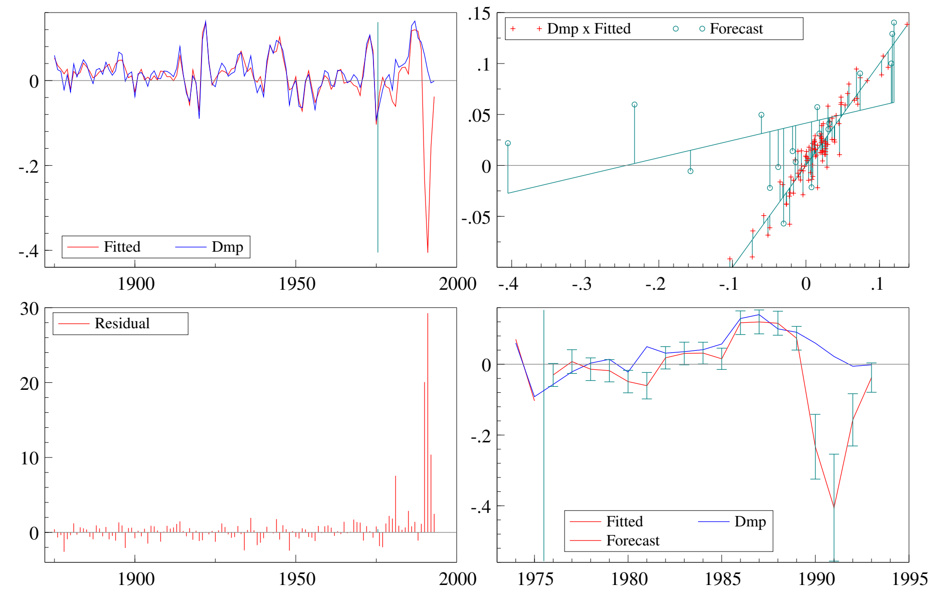

Figure 3 Fitted and actual values, their cross plot, scaled residuals, and forecasts from the mechanistic update.

4 Economic Extensions of Empirical Models

Mechanical extensions of an empirical model, such as (6), can mislead because of their simplistic approach to updating. Rather, changing environments may require economic adaptation of a model. To that end, this section considers four issues. First, Section 4.1 examines how the adopted measure of money changes when extending the data series. Second, Section 4.2 derives the consequent effects on the opportunity cost of holding the new measure of money. Third, Section 4.3 discusses the effect of financial innovation and financial deregulation on both the original measure of money and the new measure. Fourth, Section 4.4 analyzes the statistical implications for a model's constancy on an extended sample with financial innovation and changed data definitions and opportunity costs. Section 4.5 summarizes, and considers alternative approaches.

4.1 Measures of Money

This subsection reviews some difficulties in measuring money, and then focuses on the specific statistical measures used herein. Difficulties include vagueness in the concept of money itself, the effects of financial innovation on measurement of a consistent series, and changing definitions of observed series. The measurement of money leads directly to the calculation of its opportunity cost, in Section 4.2.

The concept of money is imprecise and ambiguous, and has been so for decades, if not centuries. In their evidence to the U.K. Parliamentary Committee on Bimetalism, Marshall and Edgeworth cite important changes contributing to these difficulties: the newly introduced cheque book, more rapid transfers between bank accounts in different locations due to the telegraph, and bimetalism itself through the respective quantities and prices of gold and silver. The first and third factors affected the definition of money itself, and all three factors allowed economization in the transactions medium. Later work has emphasized the changing importance of portfolio considerations versus liquidity.

Also, the very form of the transactions medium has changed greatly, becoming interest bearing (for the most part) towards the end of the sample. While cash may not have changed much, its substitutes have. Checks have fallen from favor relative to credit cards, and recently debit card transactions have shown rapid growth. The recorded measures of money also have changed over the period, especially the coverage of the most cited measures — M1, M2, M3 and M4. In particular, building societies (akin to savings and loans associations) have grown rapidly, offering close substitutes for commercial-bank liabilities, and with some converting their legal status to commercial banks from 1989 onwards. Hendry and Ericsson (1991a, Appendix, p. 34) highlighted such reservations on the measurement and interpretation of M. The conclusions in Hendry and Ericsson (1991a) reiterated these concerns.

> Undoubtedly, the model proposed above is not the end of the story since, for example, parameter nonconstancy is evident during 1971–1975. That it [the model] is not perfect is less than surprising, as the data span a century during which financial institutions altered dramatically: witness the growth of building societies and, after 1970, the introduction of Competition and Credit Control regulations and of floating exchange rates. (p. 33)

These difficulties and their consequences are apparent in the measure of money used herein. To start, both Friedman and Schwartz (1982) and Attfield *et al.* (1995) were faced with incomplete series on any measure of broad money over their samples. The measure M2 ("old definition") exists through 1968; M3 from 1963 though 1987; and M4 (adjusted for slight definitional breaks) from 1982 through 1993. To create their annual money series, Friedman and Schwartz (1982) spliced M3 onto M2, multiplying M3 by approximately 1.02 to match the value of M2 in 1968 and using M3 from 1969 through 1975. The scale factor is greater than unity because these measures of M2 and M3 are not nested series. To extend the data beyond 1975, Attfield *et al.* (1995) similarly spliced M4 onto M3 in 1987, with M4 multiplied by approximately 0.545.

The statistical magnitudes of the splicings are apparent from Figure 4, which plots the rescaled logs of the three measures (Figure 4a), their unrescaled logs (Figure 4b), the growth rates of the unrescaled

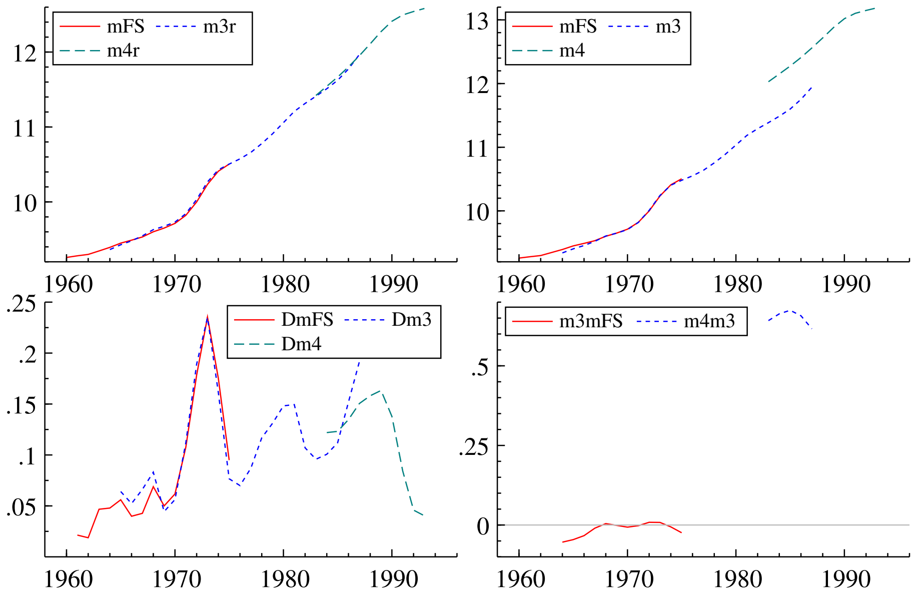

Figure 4 Rescaled money series, unrescaled money series, their growth rates, and logs of the unrescaled differentials.

measures (Figure 4c), and logs of the ratios of the series for periods with overlapping observations (Figure 4d). The first splice appears minor in nature. The second, by contrast, markedly broadens the definition of money, with M4 being 1.84 times M3 in 1987, the date of the second splicing. The non-M3 component of M4 (dominated by building society deposits) is almost entirely interest-bearing, so the implicit opportunity cost on the spliced money series is nearly halved, leading to the discussion in the next subsection.

4.2 The Opportunity Cost of Holding Money

Because older series such as M2 and M3 and some of their components are not published in the latter part of the sample, construction of a "consistent" series is infeasible, so modelers are left with little option other than splicing series together. However, splicing may alter the implied opportunity cost of holding the (measured) money, with the second splice being of primary concern because of the numerical magnitudes involved. The choice of measured opportunity cost in turn can dramatically affect the performance of empirical models of money demand. Thus, this subsection re-examines existing measures of the opportunity cost. They prove inadequate, so we develop a measure of the opportunity cost, which is key to the empirical constancy of (5) when the data are extended through 1993.

Friedman and Schwartz (1982) and Hendry and Ericsson (1991a) propose two different measures of opportunity cost for M. Hendry and Ericsson (1991a) use the short-term interest rate RS as the primary opportunity cost of holding money, as in (3), (4), and (5) above. Friedman and Schwartz (1982) advocate using a fraction of RS, denoted RN and calculated as $(H/M) \cdot RS$. This alternative measure assumes that all components of M except for high-powered money H earn interest at the (outside) short-term rate RS. Three issues bear on the choice between RS and RN and on the suitability of either as an opportunity cost: the ratio H/M, the relationship between the actual own rate and RS, and the measures of H and M for calculating RN.

First, if H/M is nearly constant, empirical results should be relatively unaffected by the choice between RS and RN, aside from a scale factor on the estimated coefficient. For Friedman and Schwartz's sample period, H/M varies almost exclusively within the narrow range $[0.21, 0.28]$. As Hendry and Ericsson (1991a, p. 32, footnote 26) note, their annual models correspondingly are little affected by the choice of interest rate.

Second, if the relationship between RS and the own rate of interest changes, neither RS nor RN may be good proxies for the actual opportunity cost. Hendry and Ericsson (1991a) expressed this concern, particularly because the behavior of the own interest rate did change at the end of Friedman and Schwartz's sample.

> A potential explanation for this finding [the predictive failure over 1971–1975], consistent with the earlier evidence and economic analysis, is presaged by Klovland's (1987) result for pre-1914 data that the *own* interest rate on broad money is an important omitted variable from the present information set.[26] Over much of the sample, U.K. commercial banks acted like a cartel with administered (and generally low) deposit interest rates; this situation changed after 1970 due to the competition regulations. Thus, own interest rates rose rapidly, altering the historical differentials and inducing predictive failure in models that excluded that variable. (p. 32)

Third, constructed RN may use either spliced or unspliced high-powered money and broad money. In Friedman and Schwartz's framework, proper measurement of RN requires the unspliced data, even while the modeled money series is spliced. By way of explanation, note that the second splice itself sharply increases the percentage of (spliced) money bearing interest. That percentage is measured by the ratio of the unspliced high-powered money to broad money. To distinguish between measures using spliced and unspliced series, H and M denote spliced series whereas H^a and M^a denote actual values (superscript ^{a} for actual). Specifically, H^a and M^a are not rescaled for the definitional changes in 1976 (for H) and 1987 (for M). Correspondingly, RN denotes $(H/M) \cdot RS$ (as above) with spliced series, and Rn denotes $(H^a/M^a) \cdot RS$.

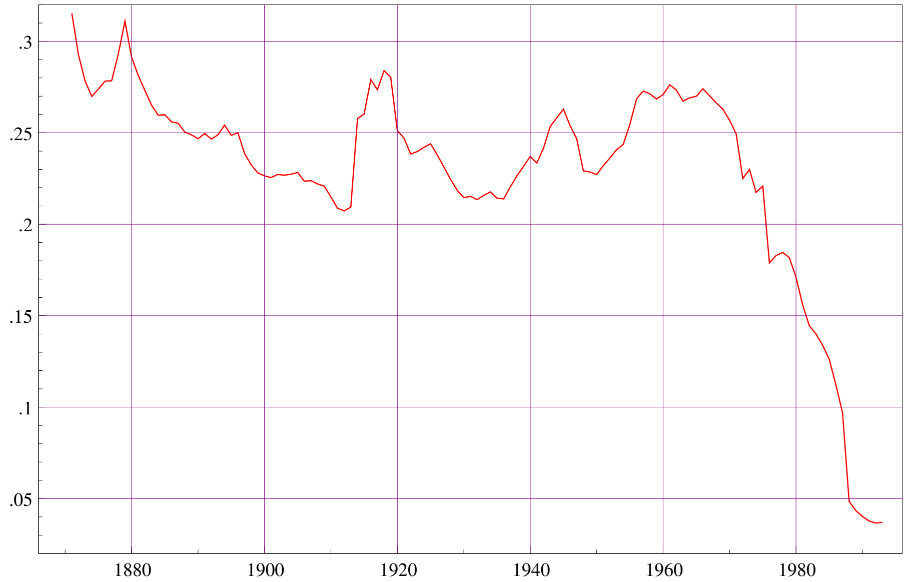

Figure 5 Time series of H^a/M^a used to adjust RS to Rn.

Figure 5 shows the time series of H^a/M^a. The ratio H^a/M^a changes little over the period prior to 1970, but falls sharply and rapidly to near zero from 1971 to 1993. Two of the largest drops in this measure occur directly after 1987, reflecting the switch in M from M3 to M4, and 1976, when the redefinition of H necessitates a rescaling by 0.7865. These effects play an important role in explaining the mis-predictions of the mechanically updated annual model (6), as Section 4.1 shows.

4.3 Financial Innovation and Financial Deregulation

The last century has witnessed considerable financial innovation. Examples early in the sample include the telegraph, the cheque book, and building societies. More recent examples include interest-bearing sight deposits, credit cards, and cash machines. All have combined to alter the financial scene radically. In particular, they have changed the role of money in portfolios, as a source of liquidity, and as the main component of the transactions medium.

Financial deregulation has played a pivotal role as well, both by allowing some forms of financial innovation to occur and by allowing rates of return on existing assets to be determined in a more competitive atmosphere. This subsection considers three key deregulations affecting the latter part of the sample: Competition and Credit Control, the allowance of interest-bearing sight deposits, and a 1986 Act of Parliament. Sections 4.2 and 5 discuss how to account for innovation and deregulation in the model.

As the quote above notes, Competition and Credit Control regulations (CCC) in 1971 deregulated the commercial banking sector, abolishing the earlier cartel arrangements. CCC reduced credit rationing and led to an otherwise unexplained increase in money holdings, which Hendry and Ericsson (1991a) modeled using the dummy variable D_4. For earlier attempts at accounting for this change, see Hacche (1974), Hendry and Mizon (1978), and Lubrano, Pierse and Richard (1986) *inter alia*.

Financial deregulation in 1984 permitted retail sight deposits (checking accounts) to bear interest. The opportunity cost of holding narrow money fell substantially, and the demand for (e.g.) M1 increased correspondingly, as shown in Hendry and Ericsson (1991b). The increase in M1 is large, e.g., 130% in nominal M1 (86% in real M1) between 1984Q3 and 1989Q2.

During 1986–1989, additional financial deregulation loosened credit rationing, particularly for Building Societies following the 1986 Act of Parliament. See Muellbauer (1994) for a discussion of the Act's large effects on consumers' expenditure. Given the similarities between the deregulations of the 1970s and the 1980s, the dummy D_c includes both episodes.

4.4 Constancy

As Judd and Scadding (1982) document, constancy is a critical and often elusive feature of empirical money-demand equations. Indeed, the historical review in Hendry (1996) shows that constancy has long been regarded as a fundamental requirement for empirical modeling, since models with no constancies cannot be used for forecasting, analyzing economic policy, or testing economic theories. Equally, model evaluation is central to any progressive research strategy: retrospective evaluations are particularly valuable for learning about the evolution of the economy and how to improve the performance of existing empirical models.

For the discussion below, it helps to formalize the concept of constancy for econometric equations fitted to time series. Briefly, suppose a parameter θ indexes a stochastic process: by definition, θ is constant across realizations of that process. Even so, θ may vary over time, perhaps depending on other stochastic processes. If θ has the same value throughout the period $\mathcal{T}$, then θ is constant over that period. A model is constant over $\mathcal{T}$ if all of its parameters are constant over $\mathcal{T}$: see Hendry (1995). Hendry (1996) highlights important ambiguities concerning parameterization and model formulation. This subsection (and Section 4.5) draws on that analysis, illustrating some of the subtleties involved through a simple conditional linear model.

Constancy of a model depends upon how the model is formulated and upon how variables are updated for extended samples. In particular, constancy may depend upon inclusion in the model of data that are zero over the original sample. To discuss this and related aspects of constancy, five forms of a conditional linear model are of interest.

$$\begin{array}{lrcll} \textit{Initial Model} & y_t & = & \theta' z_t + u_t & T \in [1, T_1] \quad (7) \\ \textit{Isolikelihood Model} & y_t & = & \theta' z_t + \delta' w_t + u_t & T \in [1, T] \quad (8) \\ \textit{Expanded Model} & y_t & = & \theta' z_t + \gamma' Q w_t + u_t & \quad (9) \\ \textit{Reparameterized Model} & y_t & = & \theta'(z_t + AQw_t) + (\gamma' - \theta' A) Q w_t + u_t & \quad (10) \\ \textit{Translated Model} & y_t & = & \theta'(z_t + A^* Q w_t) + u_t & \\ & & = & \theta' z_t^* + u_t & \quad (11) \end{array}$$

In the *initial model* (7), the variable y_t linearly depends on k variables z_t with coefficient vector θ and error u_t over the sample $[1, T_1]$. In the remaining four models, the data period is extended to include K more observations $[T_1 + 1, T]$. In the second or *isolikelihood model* (8), the initial model is extended to include K linearly independent variables w_t and so has the same likelihood as (7). The variables w_t might enter the isolikelihood model as some subset of linear combinations of w_t — Qw_t, say — with coefficient γ, generating the *expanded model* (9). If at most k combinations of w_t enter the expanded model, then it can be written as the *reparameterized model* (10), which depends upon a linear combination $(I : A)$ of $(z_t : Qw_t)$ with coefficient θ and upon Qw_t itself. Finally, some linear combination of $(z_t : Qw_t)$ may exist [$(I : A^*)$, say] such that the reparameterized model depends on that linear combination alone and not Qz_t in addition. If such a linear combination exists, the resulting *translated model* is (11), so-called because z_t has been translated into z_t^* $(= z_t + A^* Q w_t)$ over the extended sample. Five examples help clarify the interpretation of these models.

Example 1: Chow statistic for predictive failure. The initial model (7) and the isolikelihood model (8) are the basis of Chow's statistic for testing predictive failure. Without loss of generality, suppose that w_t is simply a set of zero-one impulse dummies: specifically, K dummies for the K observations in $[T_1 + 1, T]$. By construction, the coefficient vector θ on z_t remains unchanged as the sample and model expand from (7) to (8). Models (7) and (8) are identical over the initial subsample. The F-statistic testing that $\delta = 0$ is simply the Chow statistic for predictive failure. That is, the statistic tests whether or not extending the sample requires extending the model relative to the initial measurements $\{y_t, z_t\}$.

If $\delta = 0$, lengthening the sample does not require any extension of the model. This is the "purest" form of model constancy. Conversely, for any initial model (7), θ is always unchanged over the extended sample $[1, T]$, provided the model is expanded by K dummies w_t with arbitrary weights δ. This constancy of θ is trivial in that constancy always holds, and holds by construction.

Example 2: A reduced set of additional variables. The expanded model (9) provides a testably simpler representation of the isolikelihood model (8). Suitable choice of Q in (9) can generate any time series over $[T_1 + 1, T]$. If Q has rank less than K, then only some linear combinations are relevant over the extended sample, thus simplifying or reducing (8) to (9). For instance, (5) above extends (3) by 5 observations but appears to require only 2 additional variables to do so: D_{4t} and $D_{4t} \cdot \Delta rs_t$.

Example 3: Measurement of money. Section 4.1 discusses how the measurement of money alters over the sample. That is, y_t is "translated" over time into some new variable y_t^*, much as z_t is translated into z_t^* above. Equation (9), with known γ and Q, includes extensions of y as a special case. In general, A (and A^*) and Q are known, whereas γ is estimated.

Example 4: Financial innovation and interest rate spreads. Achieving a translated model requires both a reduction from w_t to Qw_t to achieve (9), and the restriction that $\gamma' = -\theta'$ to obtain (11). Drawing on Hendry and Ericsson (1991b) and Hendry (1996), suppose $T_1 \approx 1984$, y is nominal M1 in the United Kingdom, z is the Local Authority 3-month deposit rate R_{LA} (the outside rate for M1), Qw is the interest rate on sight deposits R_{SD} (the own rate of M1), and other variables in the relationship are ignored for ease of exposition.

Hendry and Ericsson (1991b) and Hendry (1996) demonstrate the constancy through 1984 of a pre-existing model for U.K. M1, where that model includes R_{LA} but not R_{SD}. Over 1985–1989, that model fails miserably on Chow's statistic. By construction, and trivially, the original parameters of that model are "constant" if a dummy is added for each observation in the forecast period, giving (8). More interestingly, the original model remains constant if just R_{SD} is added to it. That is, the expanded model (9) is such that $Qw_t = R_{SDt}$. Finally, $A^* = -1$ is a statistically acceptable restriction, implying that the interest rate differential $R_{LA} - R_{SD}$ is a suitable translation of the original variable R_{LA}, resulting in (11). Importantly, the translated model has the same parameters as the initial model: only the measurement of the data changes, and that only for the extended portion of the sample. Constancy is an operational concept for both the expanded and translated models. Both are testably not constant over the whole sample period, i.e., relative to the isolikelihood model.

Economically, the introduction of interest-bearing sight deposits requires redefining the opportunity cost of holding transactions money. Initially, the opportunity cost is the outside interest rate R_{LA}. Once sight deposits begin earning interest, the opportunity cost becomes the differential with the own rate, i.e., $R_{LA} - R_{SD}$. Equally, the opportunity cost is the differential $R_{LA} - R_{SD}$ for the full sample, and it is observationally equivalent to the outside rate R_{LA} over the initial subsample.

Example 5: Measurement of the opportunity cost. Sections 3 and 5 empirically analyze a similar issue in the measurement of the opportunity cost for the annual data on broad money demand. This example sets up the algebra of constancy for that analysis. Suppose $T_1 = 1975$; y is nominal broad money m; z is the short-term interest rate RS, which is the interest rate in (4); Qw is $(1 - \frac{H^a}{M^a}) \cdot RS$, which is the own rate on M proposed in Section 4.2 and based on Friedman and Schwartz (1982); and $A^* = -1$. For ease of presentation, other variables in the relationship are ignored (as in Example 4), and B is defined as the term $(1 - \frac{H^a}{M^a})$ in the proposed own rate. Because this example is central to Sections 3 and 5, and because it is somewhat more complicated than Example 4, the model representations are written explicitly.

$$
\begin{array}{rrcll}
\textit{Initial Model} & y_t & = & \theta RS_t + u_t & T \in [1, T_1] \quad (12)\\
\textit{Isolikelihood Model} & y_t & = & \theta RS_t + \delta' w_t + u_t & T \in [1, T] \quad (13)\\
\textit{Expanded Model} & y_t & = & \theta RS_t + \gamma B_t RS_t + u_t & (14)\\
\textit{Reparameterized Model} & y_t & = & \theta\,(RS_t + A \cdot B_t RS_t) + (\gamma - \theta A) B_t RS_t + u_t & (15)\\
\textit{Translated Model} & y_t & = & \theta\,(RS_t - B_t RS_t) + u_t & \\
 & & = & \theta Rn_t + u_t & (16)
\end{array}
$$

Hendry and Ericsson (1991a) demonstrate the constancy through 1975 of the initial model (12), which is (5) in conjunction with (4). Using the Chow statistic implied by (13), $\S$ 3.2 shows that this initial model is nonconstant on the extended sample through 1993. However, model translation recovers constancy. Section 5 below shows that the coefficients in the initial model remain constant if two variables are translated: RS into Rn, as from (12) to (16); and D_4 into D_c, which follows from a similar model path.

Paralleling Example 4, the translated model has parameters equivalent to those in the initial model: only the measurement of the data changes. Aside from a scale factor, RS and Rn are virtually indistinguishable through 1975 because H^a/M^a is nearly constant (Figure 5). Thus, θ in the initial model (12) is the coefficient on the opportunity cost $(H_a/M_a)RS$, multiplied by the (near constant) value of H_a/M_a over that subsample. As H_a/M_a falls in the 1970s and 1980s, the measure of opportunity cost must account for that fall through Rn, as in (11). Otherwise, the model will break down. Mis-measurement of

Rn by RN also can induce predictive failure because H_a/M_a and H/M behave differently in the last two decades of data.

4.5 General Comments

This subsection comments on several related issues: constancy over realizations versus over time, the consequences of nonsingular reparameterizations and data transformations, tests of constancy, data redefinitions, and Divisia indexes.

First, constant models can have time-varying coefficients, provided a deeper set of constant parameters characterizes the probability mechanism. Examples include the structured time-series models in Harvey (1981) and Harvey and Shephard (1992), random-coefficients models, and the smooth transition dynamic models in Teräsvirta and Anderson (1992) and Granger and Teräsvirta (1993). Example 5 above also falls into this category, in that θB_t varies over time even though θ does not. Thus, the existence of constancy may depend on whether raw coefficients or underlying parameters are evaluated.

Second, since one-to-one transformations of parameters are also valid indexes, zero can be the population value for some parameters in an equivalent representation. Consequently, the definition of model constancy above allows for an expanded model, provided the *existing* parameters stay constant. Relatedly, there is often a choice between nearly-equivalent variables at various stages of empirical modeling, with later data clarifying which variables actually determine a sustainable relationship. Example 4 illustrates this in the present context with RS and Rn.

Third, no tests exist in-sample (i.e., using only information up to T_1) for whether an empirical model will manifest predictive failure out of sample. An in-sample test correctly indicating the failure of the initial model would incorrectly indicate the failure of a (constant) translated model, as the initial model is identical to the translated model in-sample. Consequently, predictive failure is uniquely a post-sample problem, requiring change somewhere to induce change elsewhere. Hence, models should not be selected on the basis of their forecast performance unless their sole purpose is forecasting. Clements and Hendry (1996) show that forecasting devices can be robustified against important forms of predictive failure by intercept corrections or differencing; see also Hendry and Mizon (1996).

Fourth, previous experience with updating models of money demand and of consumers' expenditure has revealed many pitfalls, including redefinitions of variables, large changes in measurements between data revision vintages, and important structural changes; see Hendry and Ericsson (1991b) and Hendry (1994) respectively for examples. In the present context, many measures of money exist, their definitions and coverages have altered over time, and the principal measure in Friedman and Schwartz (1982) (namely M2) ceased to be the appropriate one for the United Kingdom and was replaced by M3 and then by M4. The measure M4 (but not M2 or M3) includes the liabilities of Building Societies and seems the most appealing as a measure of broad money; see Hendry and Ericsson (1991a, Appendix). These changes in measured money require adaptations elsewhere in the model: this section has focused on the most obvious adaptation, that for the measured opportunity cost of money itself. Coherent measurement of the opportunity cost (and of data generally) is contextual rather than absolute, in that the measurement may differ depending upon the economic relationship being modeled.

Fifth, to address issues such as deregulation and financial innovation, some researchers have sought to develop better indices of money, weighting money's components by their "liquidity". In Divisia indices, for example, the liquidity of a component of money is inversely related to the corresponding rate of return; see Diewert (1976), Barnett (1980), and la Cour (1996). We do not follow this route because Divisia indexes *per se* seem unlikely to resolve the changes that occurred. For instance, the relevant rates of return were controlled over some subsamples and deregulated over others. A Divisia approach implies (implausibly) that liquidity suddenly changes when interest rates move upon deregulation. Further, mortgage rationing was prevalent over most of the 20th century in the United Kingdom, with Building Societies offering high interest rates on their (highly liquid) deposit and share accounts; see Anderson and Hendry (1984) and Muellbauer (1996).

5 An Economic Extension of the Annual Model

This section re-analyzes the annual model (5) over the extended sample, with the statistical measures of opportunity cost and deregulation adjusted to reflect the economic concepts that they attempt to capture. Section 5.1 examines long-run properties first, then the dynamic equilibrium-correction model, as in Hendry and Ericsson (1991a). Sections 5.2 and Section 5.3 focus on the identification of the model as a money demand equation and on the model's policy implications.

5.1 Constancy of an Economic Extension of the Model

In the annual model (5) with (4), the measure of opportunity cost is RS and so does not incorporate the definitional and institutional changes described in Section 4, which occur primarily after 1975. To address these changes, RS is replaced by Rn, which leaves the estimates in (5) and (4) virtually unaffected. To show the value of model translation, (4) and (5) are each estimated with Rn rather than RS, first over their initial sample and then over the extended sample. For ease of comparison between coefficients on RS and Rn, Rn is divided by 0.25, the approximate mean of H^a/M^a for the sample 1878–1975.[3]

Using Rn rather than RS to measure the opportunity cost, the Engle-Granger regression (4) becomes:

$$(m-i-p)_t \;=\; -0.318 - 6.67 Rn_t \tag{17}$$
$$T = 98\ [1873\text{–}1970] \quad \mathsf{R}^2 = 0.59 \quad \hat{\sigma} = 10.57\% \quad \mathsf{dw} = 0.31 \quad \mathsf{t}_{\mathsf{ADF}} = -2.82,$$

which is essentially (4).[4] The residuals from (17) provide the equilibrium correction term $\tilde{u}_t$, which proxies for long-run excess demand and replaces the residual $\hat{u}_t$ from (4). Figure 6a graphs the two residuals ($\hat{u}_t$ and $\tilde{u}_t$) through 1970. Figure 6b graphs the corresponding measures of opportunity cost (RS and Rn) over the same period. In both figures, the differences between the old and new measures are minor. Figures 6c and 6d plot the same series on the new sample: both figures show marked differences between the old and new measures, as implied by H^a/M^a in Figure 5.

The translated model is (17) estimated over the full sample:

$$(m-i-p)_t \;=\; -0.344 - 6.30 Rn_t \tag{18}$$
$$T = 121\ [1873\text{–}1993] \quad \mathsf{R}^2 = 0.63 \quad \hat{\sigma} = 11.74\% \quad \mathsf{dw} = 0.42 \quad \mathsf{t}_{\mathsf{ADF}} = -4.80.$$

The coefficients in (18) are hardly altered from (17) and (4).

Following Hendry and Ericsson (1991a), regression residuals from either (17) or (18) enter as the ECM in the dynamic equation. Equation (17) has the advantage that its sample pre-dates the forecast period. Conversely, assuming that the relation does cointegrate, (18) should be more precise than (17). In practice, the choice makes little difference, and (17) is used below.

[3] Because the data here are annual rather than (e.g.) quarterly, we have not attempted to adjust Rn for agents' learning of financial innovations; cf. Hendry and Ericsson (1991b) and Baba, Hendry and Starr (1992).

[4] To mirror the approach in Hendry and Ericsson (1991a), this section presents Engle-Granger regressions rather than (say) the cointegration analysis in Johansen (1995) based on vector autoregressions. We intend to report such an analysis at a later date for the complete system of money, prices, income, and interest rates. A subsystem analysis appears in Ericsson and Irons (1995).

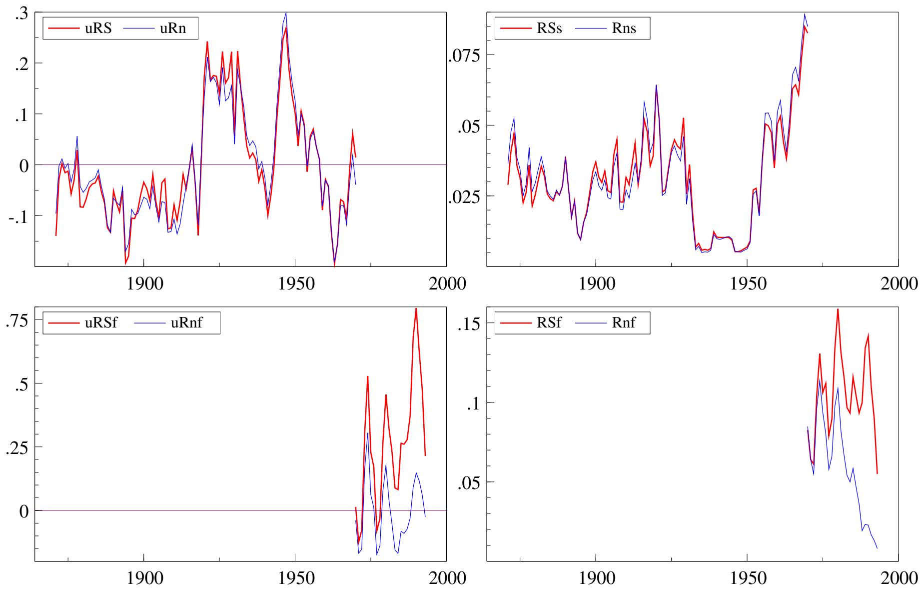

Figure 6 Cointegrating vectors based on RS and Rn.

With Rn as the opportunity cost measure in the cointegrating relation, the sample to 1973 delivers the following re-estimated dynamic model.

$$\begin{aligned}\Delta(m-p)_t \;=\;& \underset{[0.06]}{0.46}\;\Delta(m-p)_{t-1} - \underset{[0.04]}{0.11}\;\Delta^2(m-p)_{t-2} - \underset{[0.04]}{0.59}\;\Delta p_t \\ &+ \underset{[0.05]}{0.41}\;\Delta p_{t-1} - \underset{[0.005]}{0.020}\;\Delta rn_t - \underset{[0.018]}{0.066}\;\Delta_2 r\ell_t \\ &- \underset{[0.56]}{2.78}\;(\tilde{u}_{t-1}-0.2)\tilde{u}^2_{t-1} + \underset{[0.002]}{0.004} + \underset{[0.5]}{3.6}\;(D_1+D_3)_t \\ &+ \underset{[0.009]}{0.065}\;D_{4t} + \underset{[0.026]}{0.09}\;D_{4t}\cdot\Delta rs_t \end{aligned} \tag{19}$$

$T = 96$ [1878–1973] $\mathrm{R}^2 = 0.88$ $\hat{\sigma} = 1.41\%$
AR : $\mathrm{F}(2,83) = 1.11$ ARCH : $\mathrm{F}(1,83) = 1.5$ Normality : $\chi^2(2) = 0.85$
$Hetero$: $\mathrm{F}(18,66) = 0.65$ $\mathrm{Jt} = 1.25$ $\mathrm{V} = 0.10$.

The switch from RS to Rn leaves coefficients and $\hat{\sigma}$ in (19) virtually unaltered relative to those in (3).[5]

Estimation of (19) over the sample to 1993 obtains:

$$\begin{aligned}\Delta(m-p)_t \;=\;& \underset{[0.05]}{0.48}\;\Delta(m-p)_{t-1} - \underset{[0.04]}{0.10}\;\Delta^2(m-p)_{t-2} - \underset{[0.04]}{0.62}\;\Delta p_t \\ &+ \underset{[0.05]}{0.40}\;\Delta p_{t-1} - \underset{[0.006]}{0.020}\;\Delta rn_t - \underset{[0.016]}{0.039}\;\Delta_2 r\ell_t \\ &- \underset{[0.33]}{2.25}\;(\tilde{u}_{t-1}-0.2)\tilde{u}^2_{t-1} + \underset{[0.002]}{0.004} + \underset{[0.5]}{3.9}\;(D_1+D_3)_t \\ &+ \underset{[0.007]}{0.052}\;D_{ct} + \underset{[0.026]}{0.10}\;D_{4t}\cdot\Delta rs_t \end{aligned} \tag{20}$$

[5]Closely similar estimates result for the sample period extended through 1975 or truncated at 1970, the latter for the model without the two terms involving D_4. We chose 1973 as the end point because it allows the minimum sample consistent with estimating the impact of CCC.

$$T = 116\ [1878\text{–}1993] \quad \mathsf{R}^2 = 0.87 \quad \hat{\sigma} = 1.62\%$$
$$AR: \mathsf{F}(2, 103) = 3.5^{*} \quad \text{Normality}: \chi^2(2) = 0.1 \quad \text{ARCH}: \mathsf{F}(1, 103) = 0.00$$
$$Hetero: \mathsf{F}(18, 86) = 0.87 \quad \text{Jt} = 1.77 \quad \text{V} = 0.73^{*},$$

where the deregulation dummy D_4 has been extended as D_c when entering by itself. In this translated model, the coefficients are virtually unaltered from the initial model (19), although $\hat{\sigma}$ has increased by about 15%. Correspondingly, the joint parameter-constancy test (Jt) is insignificant, whereas the variance-change statistic (V) rejects constancy. Consistent with this evidence, the covariance statistic for testing parameter constancy over 1974–1993 yields $\mathsf{F}(10, 95) = 1.44$, whereas the Chow predictive-failure statistic is $\mathsf{F}(20, 85) = 2.619^{**}$. The latter has power to detect changes in equation error variances as well as in regression coefficients. A single outlier in 1981 by itself produces this rejection.

Figures 7, 9, and 8 summarize additional information on the performance of (19) and (20). Figure 7 graphs descriptive statistics for model (19): fitted, forecast, and actual values, plotted as time series and cross-plotted; the corresponding residuals; and the forecast and actual values, with ± 2 standard error bands for the forecasts. Figure 7 for (19) parallels Figure 3 for (5). Comparison of these two figures highlights the different consequences of mechanistic and economic extensions of a model. Figures 9a, 9b, and 9c plot the same model descriptions as those in Figures 7a, 7b, and 7d, but with $(m - p)_t$ as the derived dependent variable. Figure 9d is a variant of Figure 9c, but without the unit root from the dependent variable $\Delta(m - p)_t$ imposed. Figure 8 graphs the recursively estimated coefficients (first nine panels), the 1-step residuals and $0 \pm 2\hat{\sigma}_t$, the 1-step Chow statistics normalized by their one-off 1% critical values, and the break-point Chow statistics likewise normalized, all for (20). From all of these graphs, the translated model (20) performs well over this turbulent period for the U.K. economy.

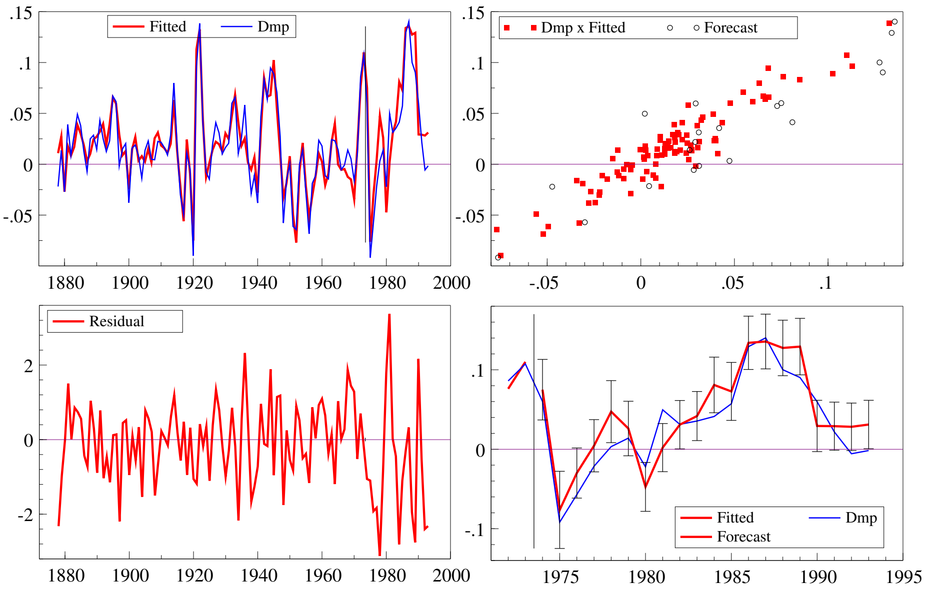

Figure 7 Annual model fit, forecasts and residuals in differences.

Equations (17) and (20) are together a single model, even though they are estimated sequentially. If (17) and (20) were estimated jointly, their coefficients would differ from those reported above, both because the dynamics in (20) would affect the estimates in (17) and because $\tilde{u}_t$ enters (20) nonlinearly. Joint estimation by nonlinear least squares obtains:

$$\Delta(m-p)_t = \underset{(0.06)}{0.49}\ \Delta(m-p)_{t-1} - \underset{(0.04)}{0.10}\ \Delta^2(m-p)_{t-2} - \underset{(0.04)}{0.62}\ \Delta p_t$$

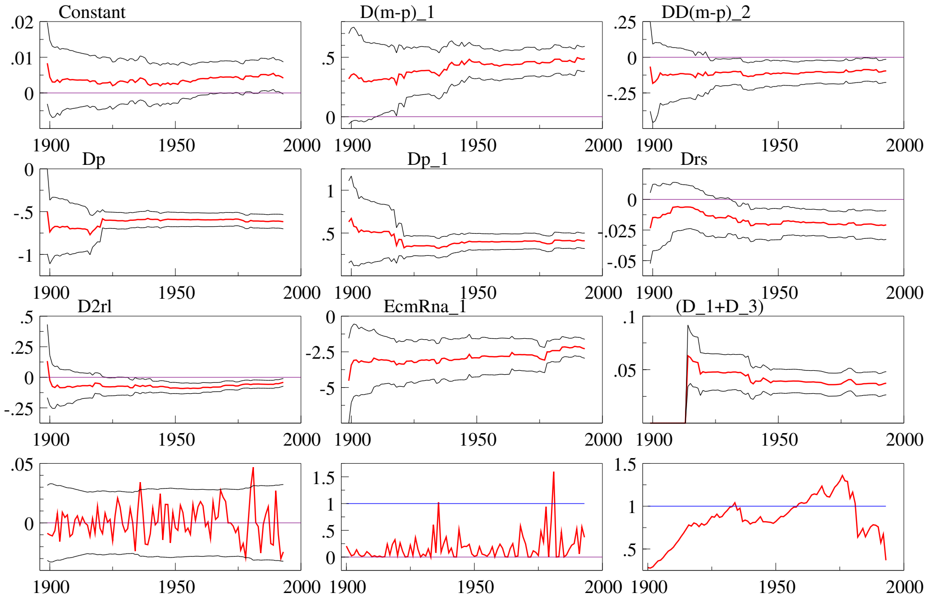

Figure 8 Annual model recursive estimates.

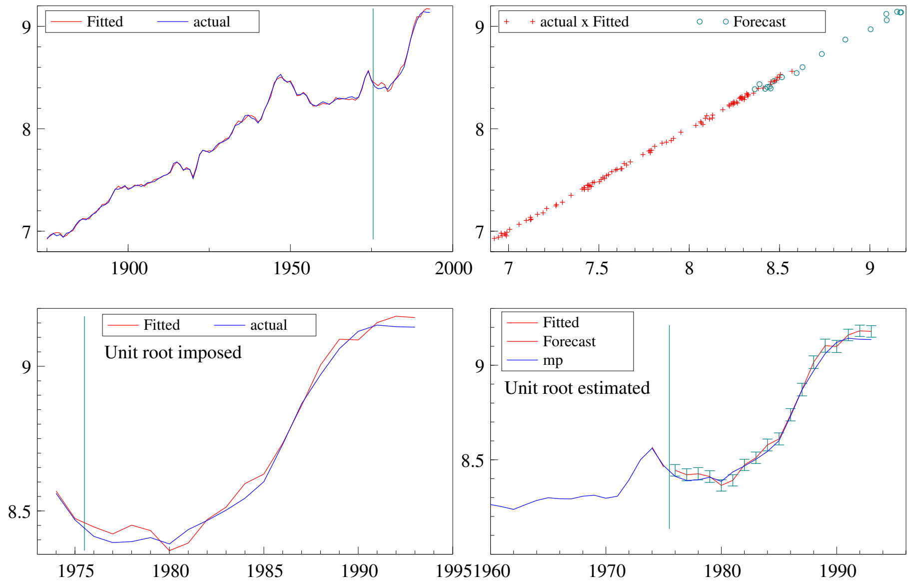

Figure 9 Annual model fit, cross plot and forecasts in levels.

$$
\begin{aligned}
& + \underset{(0.05)}{0.41}\ \Delta p_{t-1} - \underset{(0.006)}{0.020}\ \Delta rn_t - \underset{(0.016)}{0.040}\ \Delta_2 r\ell_t \\
& - \underset{(0.48)}{2.13}\ (\bar{u}_{t-1} - \underset{(0.11)}{0.13}\ )\bar{u}^2_{t-1} + \underset{(0.002)}{0.005} + \underset{(0.6)}{3.7}\ (D_1 + D_3)_t \\
& + \underset{(0.008)}{0.054}\ D_{ct} + \underset{(0.028)}{0.090}\ D_{4t} \cdot \Delta rs_t \qquad (21)
\end{aligned}
$$

$T = 116$ [1878–1993] $\mathsf{R}^2 = 0.87$ $\hat{\sigma} = 1.6206\%$ $\mathsf{dw} = 1.77$

$AR : \mathsf{F}(2, 100) = 2.8$ $\mathsf{Normality} : \chi^2(2) = 0.1$ $\mathsf{ARCH} : \mathsf{F}(1, 100) = 0.06$

$Hetero : \mathsf{F}(22, 79) = 1.06,$

where

$$\bar{u}_t = (m - p - i)_t - (-\underset{(0.043)}{0.302} - \underset{(0.39)}{6.89}\ Rn_t) \qquad (22)$$

The coefficients in (21) and (22) are little changed from those in (20) and (17), although the "interwar equilibrium departure" in the nonlinear ECM is somewhat lower than previously, now being estimated at $+13\%$ relative to the remainder of the period. No residual diagnostics are significant. Overall, (21) and (22) confirm the nonlinear feedback reaction; see Escribano (1996).

Hendry and Ericsson (1991a) emphasized the role of constancy in their modeling approach.

> Parameter constancy is at the heart of model design from both statistical and economic perspectives. Since economic systems are far from being constant and the coefficients of derived ("nonstructural" or "reduced form") equations may change when any of the underlying parameters or data correlations change, it is important to identify empirical models that have reasonably constant parameters, which remain interpretable when change occurs. (p. 21)

From Section 4, and in particular from Section 4.4, parameter constancy is compatible with (and may even require) economic extensions of the data and model. With the economic extensions for the opportunity cost of money and the proxy for deregulation, the ECM in Hendry and Ericsson (1991a) remains constant over the additional two decades of data. Coherent rather than mechanical extensions are critical for obtaining that constancy.

5.2 Identification of Money Demand

The estimated relation, (17) and (20), is interpreted as a demand for money function for three reasons. First, being a conditional model, its parameterization is unique. Second, from institutional knowledge, the supply equation in the United Kingdom is an interest rate policy function, and it shifted substantially as economic policy regimes changed over the sample. Specific regimes include the gold standard and the policy of low interest rates during the interwar period. Consequently, any combination of the shifting supply function with the demand equation would be nonconstant, yet the estimated demand function is constant. In effect, these shifts in the supply function (over-)identify the demand function in the sense of the Cowles Commission. Third, the estimated coefficients have sensible interpretations as demand responses, but they are problematic in a policy reaction function. In particular, if the estimated equation is viewed as an inverted policy reaction function, it is difficult to understand why — or how — the Bank of England might seek to control the differential between the outside interest rate and the own interest rate.

The dummies for wars and credit deregulation complicate the issue of identification, as they might represent supply perturbations or changed (unmodeled) conditions in money demand. The risks of wartime might induce additional money demand relative to usual opportunity costs, with the model implying a 3.9% increase in demand. Conversely, exigencies of financing war may have led to excessive printing of money. The equality of the coefficient across the two wars is more consistent with the former interpretation, given the changes induced by Keynes (1940), but is hardly definitive proof. Similarly, the proxy for deregulation could capture a shift in either demand or supply. Interpretation as a demand shift again seems preferable. Supporting evidence includes 'round-tripping' (borrowing from a commercial bank to relend the money in the market at a higher rate) and withdrawal of housing equity (borrowing extra on a mortgage and redepositing some when the after-tax liquidity cost is low); see Green (1987) and Patterson (1993).

5.3 Policy Implications

Policy implications fall into (at least) three distinct categories, each involving a pair of related concepts: constancy and prediction, causation and endogeneity, and expectations and the Lucas critique. See Banerjee, Hendry and Mizon (1996) for a general discussion.

First, parameter change and predictive failure need not be germane to the policy under analysis if, for instance, the predictive failure arises from data mis-measurement. As shown above, mechanistic extension of a model can result in predictive failure whereas economic extension of the same model can obtain constant parameters. So, predictive failure in itself is not a sufficient cause for discounting a model; see Hendry and Mizon (1996). Equally, econometric models may fit better than they forecast, simply because the proper economic extension is unknown *ex ante*.

That said, existing econometric models can provide information on the effects of future structural change. For example, in (20), the effect of the two World Wars on money demand are partialed out by using the dummy variables D_1 and D_3. Their coefficients are statistically insignificantly different from each other, so the effect on money demand from the First World War could predict in large measure the effect from the Second World War. A parallel result holds for the two major episodes of credit derationing, where both episodes have similar effects numerically on money demand.

Second, for the empirical models above, specific policy issues center on the direction of causation between money and prices. Does money growth cause inflation; or is it the converse, with inflation being responsible for increased money holdings? While causation cannot be wholly resolved from a single-equation study, Hendry and Ericsson (1991a) demonstrate super exogeneity of prices and interest rates in (5) and the non-invertibility of that conditional model; see Engle and Hendry (1993). Those results carry over to (20). The evidence on super exogeneity and non-invertibility is consistent with a constant demand equation in which money is endogenously determined by private sector decisions, with policy determining the outside interest rate to which agents react.

Targeting the stock of an endogenous money stock could be problematic, especially if the induced higher volatility in short-term interest rates increased the risk premium on long-term interest rates; see Baba *et al.* (1992). Both the Bank of England and the Federal Reserve Board targeted various monetary aggregates during the late 1970s and early 1980s, and both experienced difficulties in achieving (or chose not to achieve) the set targets.

Relatedly, deregulation and financial innovation can affect monetary growth rates, so interpretation of the latter must account for the former. In particular, from 1984 on, the large increases in real money might simply reflect portfolio adjustments due to changed interest-rate differentials, or they might induce higher inflation in the future. These alternatives would suggest very different responses on the part of the central bank. The evidence supports the interpretation of a portfolio shift, including the model (20) itself, the historical track record of the Bank of England, and the actual low inflation of the late 1980s and early 1990s.

Third, super exogeneity rules out the role of model-based expectations in (20), in which case the Lucas critique is not empirically germane. *Data-based* predictors are allowable, so error correction models such as (20) can have a forward-looking interpretation; see Hendry and Ericsson (1991b). Specifically, current and lagged inflation enter (20) as $-0.3(\Delta p_t + \Delta^2 p_t)$ (approximately), where $\Delta p_t + \Delta^2 p_t$ is a natural predictor of next period's inflation. Flemming (1976) proposes similar models for forming expectations about inflation.

Even when super exogeneity holds, policy can and (in general) does affect agent behavior. It does so through the variables entering the conditional model, albeit not through the parameters of that model. Government policy might well affect inflation and interest rates, and so the demand for money. However, under super exogeneity, the precise mechanism that the government adopts for such a policy does not affect agent behavior, except insofar as the mechanism affects actual outcomes.

6 Conclusions

This paper develops a framework for economic extensions of empirical models, and it applies that framework to the dynamic model of broad money demand in Hendry and Ericsson (1991a). The analysis clarifies the importance of coherently measured time series when building and evaluating empirical models. It also reflects the progressive nature of empirical research. Econometric equations are not 'one-off laws' cast in stone, but are flexible tools for understanding the economy and helping guide both policy and forecasting; see Pagan (1987).

Empirically, the broadening of measured money and the concurrent financial innovations required a measure of the opportunity cost that reflects those changes. Consequently, a measure was developed that depends on the proportion of non-interest-bearing money. With that measure, the cointegrating vector remains constant over the extended sample. The coefficients in the dynamic model itself are also virtually unchanged for the extended sample, once the new measure of opportunity cost is incorporated and the dummy for deregulation is extended to account for the 1986 Building Societies Act. Even so, the dynamic equation's error variance does increase over the extended sample. The increase may reflect effects from interest rate volatility similar to the effects on U.S. M1 from the risk of capital loss induced by volatile long-term interest rates; see Baba *et al.* (1992).

The model's degree of empirical constancy seems reasonable, especially in light of the very large changes over the last 120 years in the variables' magnitudes, in the underlying economic structure, and in the financial system. Nevertheless, several developments could improve the model's fit in the most recent period and thereby improve its constancy. First, a more consistent series for money might be constructed, allowing for the liabilities of Building Societies throughout. Second, the measure of opportunity cost also could be improved. Specifically, the effects of interest-rate volatility and learning adjustment could be modeled with higher frequency data and then mapped to annual values. Third, indicator (dummy) variables require more study. They are endogenously chosen to remove nonconstancies and may reflect non-modeled factors, such as data measurement errors, structural changes, and omitted transitions. Equally, war, financial innovation, and deregulation have complex effects on the economy, typically generating the most perturbed and informative data of the sample. Yet, adding dummies for these events wastes information by essentially excluding such data.

Finally, econometrics has progressed during the model's forecast period. The technology of multivariate cointegration analysis is bound to yield additional insights if the marginal models can be developed as congruent relations. Methods for estimating nonlinear equilibrium-correction models are now available, as are techniques for estimating transition effects; see Escribano (1996) and Granger and Teräsvirta (1993) respectively.

Appendix. Encompassing Other Empirical Models

This appendix demonstrates that the annual translated model (20) encompasses other existing models on this dataset. Section A.1 briefly discusses encompassing of other annual models. The remaining sections focus on models using the phase-average data in Friedman and Schwartz (1982). Section A.2 defines the relationship between annual and phase-average data, and reproduces Friedman and Schwartz's central empirical results on U.K. money demand. Section A.3 offers an interpretation of phase averaging in order to compare annual and phase-average results. Section A.4 does so for Friedman and Schwartz's model. From all the evidence, (20) encompasses the other empirical models, whereas none of the latter can encompass (20).

Encompassing is a key feature in a progressive research strategy and so is natural to consider for extensions of datasets. A necessary condition for encompassing is variance dominance, where one equation variance-dominates another equation if the former has a smaller error variance. That result is critical in the comparison of models below. For the formal structure of encompassing, see Hendry and Richard (1982) and Mizon and Richard (1986).

A.1 Encompassing Other Annual Models

Hendry and Ericsson (1983) developed an annual model over 1878–1970 with a linear error correction term. Using the same data, Longbottom and Holly (1985) and Escribano (1985) obtained improved specifications, the first through the role of interest rates and the second through a nonlinear ECM. Each of the two new models encompassed the model in Hendry and Ericsson (1983), but neither could encompass the other. Equation (3) improved on all three of those models, encompassing each one.

Section 5 establishes additional encompassing results: namely, (20) encompasses (3) on the extended sample through suitable measurement of the opportunity cost and proper adjustment for financial deregulation. The increased error variance in (20) indicates that further improvements are possible, as discussed in Section 6.

A.2 Phase-average Data and Models Thereof

While the original data are annual, Friedman and Schwartz (1982) analyze phase averages of that annual data. This section describes what phase averaging is, and documents Friedman and Schwartz's central phase-average model for U.K. money demand.

In phase averaging, the data are averaged separately over contraction and expansion phases of reference business cycles. That aims to remove the short-term fluctuations from the data. For a typical annual series $\{x_t; t = 1, \ldots, T\}$, the phase-average series is constructed as:

$$\bar{x}_j = \frac{\dfrac{x_t}{2} + x_{t+1} + \cdots + x_{t+n_j-1} + \dfrac{x_{t+n_j}}{2}}{n_j} \qquad j = 1, \ldots, J, \tag{A.1}$$

where an overbar denotes phase averaging, j is the index for phase averaging, n_j is the phase length in years, and J is the number of phase average observations corresponding to T annual observations. Beginning and ending years are weighted by one-half and appear in adjacent phases as well. Friedman and Schwartz phase average the logarithms of money, prices, incomes, and population, and the original values of the interest rates.

Friedman and Schwartz (1982, p. 282) present their central regression for U.K. money demand. We could closely replicate that equation:

$$\begin{aligned}(\bar{m} - \bar{p} - \bar{n})_j \;=\; & \underset{(0.19)}{0.020} + \underset{(0.049)}{0.883}\,(\bar{i} - \bar{n})_j - \underset{(3.29)}{11.22}\,\overline{RN}_j - \underset{(0.29)}{0.21}\,G(\bar{p} + \bar{i})_j \\ & + \underset{(0.58)}{1.38}\,\bar{W}_j + \underset{(2.7)}{20.6}\,\bar{S}_j \end{aligned} \tag{A.2}$$

$J = 36$ [1874–1973] $\mathsf{R}^2 = 0.97$ $\hat{\sigma} = 10.36\%$ $\mathsf{dw} = 1.50$ $\mathsf{Jt} = 2.15^*$ $\mathsf{V} = 0.23$.

The variable $G(\bar{p}+\bar{\imath})$ is the growth rate of nominal income, $\bar{W}$ is a dummy for 'postwar adjustment', $\bar{S}$ is a data-based dummy for '[a]n upward demand shift, produced by economic depression and war ...' during 1921–1955, and quotes are from Friedman and Schwartz (1982, pp. 228, 281).[6] Friedman and Schwartz also consider a similar equation in rates of change, discussed below.

Hendry and Ericsson (1991a) document evidence on (A.2)'s mis-specification, including rejection of residual normality, price homogeneity, and constancy (both Chow and covariance statistics). Hansen's statistics, reported in (A.2), provide further evidence of nonconstancy. That said, the performance of (A.2) *relative to* (5) and (20) is still of interest, so Section A.3 shows how to compare annual and phase-average results.

A.3 An Interpretation of Phase-average Results

Encompassing statistics are relatively easy to calculate for annual models (such as those discussed in Section A.1) and for phase-average models. However, encompassing statistics for comparing annual models with phase-average models are as yet undeveloped, although the approach is clear. Using (A.1), the annual model is reduced to a phase-average representation, whose derived coefficients are compared with the coefficients of the estimated phase-average model. Except under highly restrictive assumptions, the annual model for encompassing is a full system of equations for money, prices, income, and interest rates, and not just a conditional money-demand equation. Construction of a full system is beyond the scope of this paper, as is derivation of the formal encompassing statistic. Section A.4 offers two alternatives. It examines annual and phase-average models for variance dominance, and it compares phase-average outcomes with annual outcomes transformed to phase averages. To do the former, the error variances from the two types of models must be in the same units. This section shows how to ensure that they are by re-examining the results in Friedman and Schwartz (1982).

In estimating their phase-average regressions, Friedman and Schwartz (1982) used weighted least-squares to correct for the heteroscedasticity introduced by variable-length phase-averaging; see also Campos, Ericsson and Hendry (1990). If unnormalized weights are applied in the phase-average regressions, the resulting equation standard errors are directly comparable to equation standard errors from annual regressions. However, to replicate the regressions in Friedman and Schwartz (1982), the weights must be normalized such that the average weight is unity for regressions in levels, and slightly less than unity for regressions in rates of change. Normalization is immaterial to the validity of the heteroscedasticity transform, but comparisons of the equation standard error for phase-average and annual regressions must account for the normalization factor, which rescales the phase-average equations by a non-unit scalar.

An example clarifies the implications of the normalization. Suppose an annual variable x_t is distributed as $IN(0, \sigma^2)$: serially independent and normal with mean zero and variance σ^2. From (A.1), the phase average variable $\bar{x}_j$ is distributed as $IN(0, \sigma^2/k_j^2)$, where $k_j^2 = 2n_j^2/(2n_j - 1)$. Even though the annual x_t is homoscedastic, the phase average $\bar{x}_j$ is not, having a variance depending upon the phase length n_j. To correct this heteroscedasticity, Friedman and Schwartz (1982) use weighted least squares regressions, where the weights depend on k_j. If the weights had been the k_j themselves, then the weighted errors in the regression equation would have been distributed as (e.g.) $IN(0, \sigma^2)$ for $\bar{x}_j$ regressed on a constant. However, Friedman and Schwartz (1982) and Attfield *et al.* (1995) use *normalized* weights κ_j (say), where $\kappa_j = k_j/\bar{k}$ and the normalization factor $\bar{k}$ is the approximate sample mean

[6] Equation (1) in Hendry and Ericsson (1991a) and equation (A.2) above differ slightly in their reported numbers because of rounding errors in the data. The former uses the phase-average data published in Friedman and Schwartz (1982), which are rounded, whereas the latter uses phase-average data calculated directly from the annual data in Friedman and Schwartz (1982).

of k_j. Because $\bar{k}$ simply rescales the entire weighted least squares regression, the choice between the weights k_j and κ_j does not affect coefficient estimates, their standard errors, or t-ratios. The equation standard error *is* affected, though, noting that $\kappa_j \cdot \bar{x}_j$ is distributed as $IN(0, \sigma^2/\bar{k}^2)$. Thus, for a phase-average regression estimated by weighted least squares with normalized weights, the equation standard error must be multiplied by $\bar{k}$ to transform it into units comparable to those of the equation standard error from a regression with annual data.

Normalization has consequences for Friedman and Schwartz's phase-average regressions, both in levels and in rates-of-change. For the phase-average equation in levels (comparable to (A.2)), Friedman and Schwartz (1982, p. 282) report that $\hat{\sigma}$ is 5.54% from the regression with normalized weights. Because the mean of the unnormalized weights is $\bar{k} = 1.825$, the $\hat{\sigma}$ comparable to annual results is $5.54\% \cdot 1.825 = 10.11\%$. That is essentially 10.36%, the value for $\hat{\sigma}$ obtained in (A.2), using unnormalized weighted least squares. Second, for the phase-average equations with rates of change in (e.g.) Friedman and Schwartz (1982, Table 6.8), the normalization factor $\bar{k}$ is 8.00 exactly, which is close to 7.20, the average of the unnormalized weights. Thus, for the $\hat{\sigma}$ of 1.34% in the 'final' rates-of-change equation in Friedman and Schwartz (1982, p. 282), the $\hat{\sigma}$ comparable to annual results is $1.34\% \cdot 8.0 = 10.72\%$.

As mentioned, variance dominance is necessary for encompassing. These results are for in-sample calculations. They also carry over to out-of-sample calculations, where dominance in root mean squared error (RMSE) is necessary for forecast encompassing; see Ericsson (1992) and Clements and Hendry (1993). The following section considers both variance dominance and RMSE dominance.

A.4 Encompassing the Phase-average Models

This section compares the fit of annual and phase-average models for U.K. money demand (Section A.4.1) and their statistical and numerical constancy (Section A.4.2). Once comparable units and time periods are obtained, the annual models clearly dominate the phase-average models, both in variance and in RMSE. These results also allow re-assessment of claims in Friedman and Schwartz (1991) and Attfield *et al.* (1995).

A.4.1 Goodness of Fit

Table A.1 lists values of $\hat{\sigma}$ and RMSE for four equations: the phase-average equations in levels and rates of growth from Friedman and Schwartz (1982, p. 282), the initial annual model (5), and the translated annual model (19). Over the initial sample 1878–1973 (phases 2–36), (5) and (19) have virtually identical equation standard errors (1.424% and 1.406%) and substantially variance-dominate both phase-average equations (equation standard errors of 10.36% and 10.75%). The annual models also variance-dominate the phase-average models over the new sample (1974–1993) and over the extended sample (1878–1993). Almost regardless of sample period or particular model, the ratio of phase-average to annual values is 7 or more, reflecting just how much better the annual model explains the data relative to the phase-average model. Put slightly differently, roughly 98% or more of the residual variation in the phase-average model is explained by the annual model.

The first set of RMSEs in Table A.1 are obtained by estimating each models over its initial sample and forecasting over the remainder of the sample. For the longest matching forecast sample (phases 37–42, or 1974–1988), the annual equations dominate the phase-average equations in RMSE. The ratio of RMSEs for (A.2) and (19) is over 5, similar to the results in-sample. When the forecast sample is extended through 1993, the RMSEs for the annual models increase: marginally for (19) and dramatically for (5). The change in RMSE for (5) arises solely from that equation's poor measure of opportunity cost (RS) and a lack of accounting for financial deregulation. RMSEs for the phase-average models can be computed through only phase 42 (1985–1988). One regressor in the phase-average models is the two-sided growth rate $G(\bar{p}+\bar{i})$, which, for its calculation in phase 43, would require $\bar{p}+\bar{i}$ for the (not yet complete) phase 44 (1991 onward). Still, the RMSE for the annual model (19) is much smaller than the equation standard errors and RMSEs of the phase-average models over any of the samples.

In Table A.1, $\hat{\sigma}$ (and RMSE) for the phase-average and annual equations are in comparable units because the underlying phase-average regressions used unnormalized weighted least squares. By contrast, Friedman and Schwartz (1982, p. 282) used normalized weighted least squares, and Friedman and Schwartz (1991) do not account for that normalization in their comparisons of phase-average and annual models. Because of this technical error, Friedman and Schwartz (1991, pp. 46–47) are incorrect in stating that their regressions variance-dominate those in Hendry and Ericsson (1991a). Rather, the converse is true, as Table A.1 shows.

Equation (A.1) provides an additional method for evaluating the annual and phase-average models: transform the annual models' fitted and forecast values to phase averages and compare those values directly with those from the phase-average models. Figure A.4.1 does precisely this, focusing on the initial annual model (5) and the phase-average model in levels, (A.2). Figure A.4.1a plots the actual, fitted, and forecast values over the extended sample (phases 1–43), and Figure A.4.1c plots the corresponding residuals and forecast errors. Figures A.4.1b and A.4.1d are comparable plots, but over the forecast sample alone. Figure A.4.1 (particularly in Figure A.4.1d) highlights that the phase-average model generally has larger errors than the annual model, with an error in phase 37 of nearly 20%. This outcome is consistent with the comparison of Figures 4 and 7 in Hendry and Ericsson (1991a).

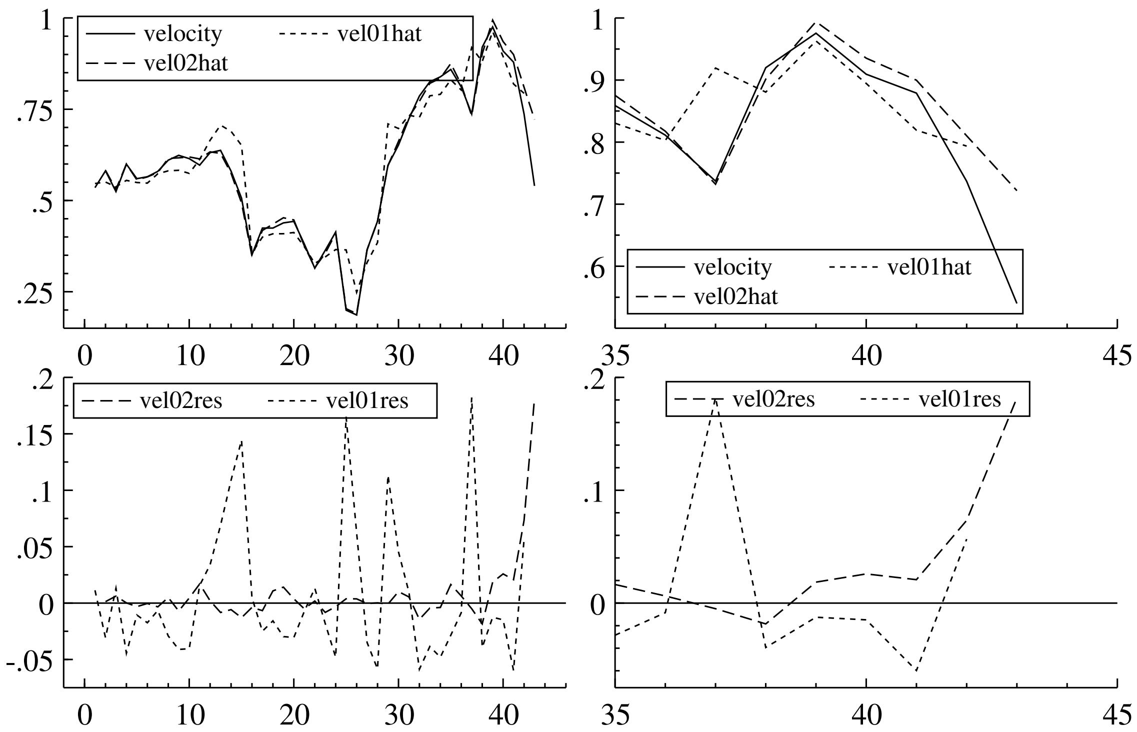

Figure A.4.1: phase-average and annual equation forecasts and residuals in a common phase mapping

The second set of RMSEs in Table A.1 numerically summarize the graphical results in Figure A.4.1.[7] As with the earlier comparisons, the annual model (5) always dominates the phase-average model (A.2) for matched samples, reflecting the superior performance of the annual model. Over phases 37–43, the

[7] Because the annual model used the sample 1878–1970 due to lagged values, its phase-average fitted values begin in phase 2, not phase 1.

RMSE for the annual model is 13.533% because of the large outlier in phase 43 (1988–1991). As noted above, the phase-average model cannot be tested for that phase.

A.4.2 Tests of Constancy

Tests of constancy are yet another metric for assessing the relative fit of models. The annual models (5) and (19) both fail the Chow and Hansen V tests, and (5) fails the covariance test as well; see Table A.1. These failures reflect the models' increased variances over the new and extended samples. By contrast, the phase-average models show little evidence of nonconstancy over the extended sample.

The apparent contradiction between variance dominance and constancy has an immediate resolution. Variance (and RMSE) dominance compare the fit of different models across the same sample, whereas the Chow, covariance, and Hansen statistics compare the fit of a given model across different samples. Furthermore, the power of a constancy test depends not only on the magnitude of the change in the coefficients, but also on the fit of the model and the number of observations available. In sample, the annual models fit the data well and have white noise errors, and their coefficient estimates are precisely estimated. These features, in combination with many observations in the forecast period, ensure that constancy tests have high power to detect even small nonconstancies in the annual model. Conversely, the phase-average models fit rather poorly in sample, and (A.2) in particular is detectably nonconstant in sample even though that sample is small. Thus, further nonconstancies in the phase-average model are hard to detect statistically.

Put somewhat differently, the numerical and statistical forecast properties of a model need not be the same, as the values of $\hat{\sigma}$, RMSE, and the Chow statistic demonstrate. The phase-average model appears constant, in fair part because it fits the data so poorly. The annual model is statistically detectably nonconstant because it fits the data so well.

The results in Table A.1 and Figure A.4.1 allow a re-assessment of Attfield *et al.* (1995), which tests for the constancy of (A.2) over phases 37–42 [1973–88] and of (5) over 1976–93. Using Chow and covariance tests, Attfield *et al.* (1995, Tables 1–7) find that (A.2) appears constant whereas (5) does not. In concluding, they claim:

> . . . the set of estimates based on the *phase-average* data and the approach adopted by [Friedman and Schwartz (1982)] appear to provide a reasonably good explanation of phase-average money holdings since 1975, at least until the very last phase now available. And some formal tests of stability suggest that these estimates are stable. In contrast, the estimates based on annual observations and the approach adopted by [Hendry and Ericsson (1991a)] do not provide a satisfactory explanation for the annual observations after 1975 and appear to be highly unstable. (p. 10)

These inferences are incorrect or misleading, for four reasons. First, while the annual model (5) is statistically less constant than the phase-average model (A.2), the annual model provides a far better $explanation$ of phase-average money than does the phase-average model; see Figure A.4.1. Annual modeling for phase-average data is analogous to modeling seasonally unadjusted data and then seasonally adjusting the outcomes, assuming that seasonally adjusted outcomes were desired. Attfield *et al.* (1995) confuse goodness of fit and constancy, which are not equivalent concepts.

Second, the phase-average models appear statistically constant in fair part because they fit so poorly. Tests of constancy out-of-sample are further contaminated for the phase-average model in levels because that model is known to be nonconstant in-sample. Third, in Attfield *et al.* (1995), comparison of annual and phase-average models is based on mismatched samples, biasing the constancy tests in favor of the phase-average models. Their Figures 1a, 1b, and 2 of the models' forecast errors seem impressive, but the worst performance of the annual model is over data that the phase-average model does not attempt to predict. Fourth, Attfield *et al.* (1995) ignore the economic consequences of redefining the dependent variable. In particular, the annual model is mechanically extended when its constancy is tested over 1975–1993. If that model is economically extended, as in (19), it performs quite respectably.

In summary, the constancy tests in Attfield *et al.* (1995) point to the practical importance of coherent rather than mechanistic extensions of empirical models. The translated annual model (20) performs well over a period with radical changes in economic policy. Its doing so is directly dependent upon sensible economic choices for extending the time series of the regressors to reflect their altered measurement and the altered measurement of the dependent variable. The annual model, whether mechanistically or economically extended, still explains the data better than either of the phase-average models. The remaining nonconstancy in the economically extended annual model indicates the high power of constancy tests in well-fitting models and some remaining room for model improvement.

References

Anderson, G. J., and Hendry, D. F. (1984). An econometric model of United Kingdom building societies. *Oxford Bulletin of Economics and Statistics*, **46**, 185–210.

Attfield, C. L. F., Demery, D., and Duck, N. W. (1995). Estimating the UK demand for money function: A test of two approaches. Mimeo, Economics department, University of Bristol.

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

Banerjee, A., Hendry, D. F., and Mizon, G. E. (1996). The econometric analysis of economic policy. *Oxford Bulletin of Economics and Statistics*, **58**, 573–600.

Barnett, W. A. (1980). Economic monetary aggregates: An application of index number and aggregation theory. *Journal of Econometrics*, **14**, 11–48.

Campos, J., Ericsson, N. R., and Hendry, D. F. (1990). An analogue model of phase-averaging procedures. *Journal of Econometrics*, **43**, 275–292.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.

Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.

Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural breaks. *Journal of Applied Econometrics*, **11**, 475–494.

Dickey, D. A., and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057–1072.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, **4**, 115–145.

Doornik, J. A., and Hansen, H. (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.

Doornik, J. A., and Hendry, D. F. (1994). *PcGive 8: An Interactive Econometric Modelling System*. London: International Thomson Publishing, and Belmont, CA: Duxbury Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.

Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.

Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.

Ericsson, N. R. (1992). Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. *Journal of Policy Modeling*, **14**, 465–495.

Ericsson, N. R., and Irons, J. S. (1995). Book review of applied econometric techniques. *Econometric Theory*, **14**, 121–133.

Escribano, A. (1985). Non-linear error correction: The case of money demand in the UK (1878–1970). Mimeo, University of California at San Diego.

Escribano, A. (1996). Nonlinear error correction: The case of money demand in the UK (1878–1970). Mimeo, Universidad Carlos III, Madrid.

Flemming, J. S. (1976). *Inflation*. Oxford: Oxford University Press.

Friedman, M., and Schwartz, A. J. (1982). *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*. Chicago: University of Chicago Press.

Friedman, M., and Schwartz, A. J. (1991). Alternative approaches to analyzing economic data. *American Economic Review*, **81**, 39–49.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.

Goldfeld, S. M., and Sichel, D. E. (1990). The demand for money. In Friedman, B. M., and Hahn, F. H. (eds.), *Handbook of Monetary Economics*, Vol. 1, pp. 299–356. Amsterdam: North-Holland.

Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Green, C. J. (1987). Money market arbitrage and commercial banks' base rate adjustments in the United Kingdom. *Bulletin of Economic Research*, **39**, 273–296.

Hacche, G. (1974). The demand for money in the United Kingdom: Experience since 1971. *Bank of England Quarterly Bulletin*, **14**, 284–305.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, **14**, 517–533.

Harvey, A. C. (1981). *Time Series Models*. London: Philip Allan.

Harvey, A. C., and Shephard, N. (1992). Structural time series models. In Maddala, G. S., Rao, C. R., and Vinod, H. D. (eds.), *Handbook of Statistics*, Vol. 11. Amsterdam: North-Holland.

Hawtrey, R. G. (1938). *A Century of Bank Rate*. London: Longmans, Green and Co.

Hendry, D. F. (1994). HUS revisited. *Oxford Review of Economic Policy*, **10**, 86–106.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1996). On the constancy of time-series econometric equations. *Economic and Social Review*, **27**, 401–422.

Hendry, D. F., and Ericsson, N. R. (1983). Assertion without empirical basis' an econometric appraisal of 'monetary trends in... the United Kingdom' by Milton Friedman and Anna J. Schwartz. Discussion paper 27, Board of Governors. Monetary Trends in the United Kingdom, Bank of England Panel of Academic Consultants, Paper No. 22; (1987).

Hendry, D. F., and Ericsson, N. R. (1991a). An econometric analysis of UK money demand in monetary trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz. *American Economic Review*, **81**, 8–38.

Hendry, D. F., and Ericsson, N. R. (1991b). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.

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*, **88**, 549–563. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Hendry, D. F., and Mizon, G. E. (1996). Selecting econometric models for policy analysis by forecast accuracy. Mimeo, Nuffield College, University of Oxford.

Hendry, D. F., and Morgan, M. S. (1995). *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Hendry, D. F., and Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, **20**, 3–33. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press and in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Jevons, W. S. (1884). *Investigations in Currency and Finance*. London: Macmillan.

Johansen, S. (1995). *Likelihood based Inference on Cointegration in the Vector Autoregressive Model*. Oxford: Oxford University Press.

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

Keynes, J. M. (1930). *A Treatise on Money*. London: Macmillan. Reprinted in 1971 by the Royal Economic Society.

Keynes, J. M. (1940). *How to Pay for the War*. London: Macmillan.

la Cour, L. F. (1996). The problem of measuring money: Results from an analysis of Divisia monetary aggregates for Denmark. mimeo, Institute of Statistics, University of Copenhagen.

Longbottom, A., and Holly, S. (1985). Econometric methodology and monetarism: Professor Friedman and Professor Hendry on the demand for money. Discussion paper no. 131,, London Business School, London.

Lubrano, M., Pierse, R. G., and Richard, J.-F. (1986). Stability of a UK money demand equation: A Bayesian approach to testing exogeneity. *Review of Economic Studies*, **53**, 603–634.

MacKinnon, J. G., and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, **29**, 305–325.

Marshall, A. (1926). *Official Papers*. London: Macmillan and Company.

Mizon, G. E., and Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.

Muellbauer, J. N. J. (1994). The assessment: Consumer expenditure. *Oxford Review of Economic Policy*, **10**, 1–41.

Muellbauer, J. N. J. (1996). Measuring mortgage market deregulation. Unpublished paper, Nuffield College, Oxford.

Nicholls, D. F., and Pagan, A. R. (1983). Heteroscedasticity in models with lagged dependent variables. *Econometrica*, **51**, 1233–1242.

Pagan, A. R. (1987). Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, **1**, 3–24. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

Patterson, K. D. (1993). The impact of credit constraints, interest rates and housing equity withdrawal on the intertemporal pattern of consumption – a diagrammatic analysis. *Scottish Journal of Political Economy*, **40**, 391–407.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, **31**, 350–371.

Teräsvirta, T., and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, **7**, 119–139.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.