Empirical Economics (1998) 23:401–415

EMPIRICAL ECONOMICS

Friedman and Schwartz (1982) revisited: Assessing annual and phase-average models of money demand in the United Kingdom*

Neil R. Ericsson[1], **David F. Hendry**[2], **Kevin M. Prestwich**[3]

[1] Stop 24, Division of International Finance, Federal Reserve Board, 2000 C Street, N.W., Washington, DC 20551, USA (e-mail: ericsson@frb.gov)
[2] Nuffield College, Oxford OX1 1NF, UK (e-mail: david.hendry@nuffield.oxford.ac.uk)
[3] AMS—American Management Systems, 12601 Fair Lakes Circle, Fairfax, Virginia 22033, USA (e-mail: kevin_prestwich@mail.amsinc.com)

Abstract: Several studies have developed empirical models of U.K. money demand using the century of annual and phase-average data in Friedman and Schwartz (1982). The current paper evaluates key models from those studies, employing tests of constancy and encompassing. The evidence strongly favors an annual model from Ericsson, Hendry, and Prestwich (1998a), whereas models based on the phase-average data fare poorly.

Key words: constancy, encompassing, Friedman and Schwartz, money demand, phase averaging

JEL classification: E41, C52

* The first author is a staff economist in the Division of International Finance, Board of Governors of the Federal Reserve System, Washington, DC 20551 USA. The second author is Leverhulme Personal Research Professor of Economics at Nuffield College, Oxford OX1 1NF. The third author was a research assistant in the Division of International Finance at the Federal Reserve Board when this paper was initially drafted. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. The first author gratefully acknowledges the generous hospitality of Norges Bank, where he revised some of the material herein. The second author gratefully acknowledges financial support from the U.K. Economic and Social Research Council under grant R000234954. We wish to thank Clifford Attfield, David Demery, and Nigel Duck for compiling the data in Attfield, Demery, and Duck (1995); and Neva Kerbeshian, Helmut Lütkepohl, Jaime Marquez, and two anonymous referees for helpful comments. All numerical results were obtained using PcGive Professional Version 9.0: see Doornik and Hendry (1996). The data may be obtained from the Internet at http://wotan.wiwi.hu-berlin.de/oekonometrie/engl/data.html

1. Introduction

Empirical modeling of the demand for money has been a central focus in the economics profession for many years. This interest is reflected in the sequence of studies by Friedman and Schwartz (1982), Hendry and Ericsson (1983), Longbottom and Holly (1985), Escribano (1985), Hendry and Ericsson (1991), Attfield, Demery, and Duck (1995), and Ericsson, Hendry, and Prestwich (1998a), which have developed various empirical models of U.K. money demand using the century of annual and phase-average data in Friedman and Schwartz (1982). Those models differ substantially in their empirical properties and in their purported implications for policy, so some model comparison is of value. The current paper evaluates the key models from those studies, employing tests of constancy and encompassing. The evidence strongly favors an annual model from Ericsson, Hendry, and Prestwich (1998a).

These various studies involve different models estimated over different sample periods and on different measures of the data (annual and phase-average). Two types of model comparison are common in the existing literature. First, a given model can be evaluated over different data periods with the same data measure, generating tests of constancy. Second, different models estimated over the same period and with the same data measure can be compared, as with standard encompassing tests. The current paper assesses both constancy and encompassing for four key models that use Friedman and Schwartz's data. It also evaluates these models by a new encompassing approach, comparing different models estimated over the same sample period but with different data measures. Tests of constancy and encompassing are natural to consider for evaluating these models. As Judd and Scadding (1982) document, constancy is a critical and often elusive feature of empirical money-demand equations. Likewise, encompassing is a key feature in a progressive research strategy and so is sensible to consider for extensions of datasets; see Mizon and Richard (1986) and Hendry (1995, Chapters 9 and 14) *inter alia*.

This paper is organized as follows. Section 2 describes the annual data, and Section 3 discusses the annual models. An annual model in Ericsson, Hendry, and Prestwich (1998a) encompasses previous annual models but not conversely, and that model has empirically constant coefficients. Section 4 defines the relationship between annual and phase-average data and reproduces Friedman and Schwartz's central empirical (phase-average) results on U.K. money demand. Section 5 derives some implications of phase averaging in order to compare the annual and phase-average results. Section 6 then compares annual models from Hendry and Ericsson (1991) and Ericsson, Hendry, and Prestwich (1998a) with phase-average models from Friedman and Schwartz (1982) and finds clear evidence supporting the annual models. Throughout, the analysis emphasizes the progressive nature of empirical research and the practical importance of coherent rather than mechanistic extensions of empirical models. Section 7 concludes.

2. The annual data

This section describes the annual data. 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), all for the United Kingdom. Data for 1871 through 1975 are from Friedman and Schwartz (1982). Attfield, Demery, and Duck (1995) extended those series over 1976–1993, constructing them from a variety of sources and splicing together several alternative definitions of money. The variables M and H are in £ million; I is in £ million for 1929; N is in millions; $P = 1.00$ in 1929; and RS and $R\ell$ are fractions. Ericsson, Hendry, and Prestwich (1998b, Appendix) give further details on the data, and Hendry and Ericsson (1991) and Ericsson, Hendry, and Prestwich (1998a) graph the data and provide some descriptive statistics.

Some constructed variables are of interest. First, under the quantity theory, the income elasticity is unity, so a key derived variable is velocity $V\ [= (I \cdot P)/M]$.

Second, there are dummy variables. Retaining the notation in Hendry and Ericsson (1991) and Ericsson, Hendry, and Prestwich (1998a), the variables D_1 and D_3 are zero-one dummies for World Wars I and II, and D_4 is a zero-one dummy for 1971–1975. The latter 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. The dummy D_c proxies for both episodes of deregulation, being unity for 1971–1975 and 1986–1989, and zero otherwise.

Third, there are several measures for the opportunity cost of holding money M. Ideally, the measured opportunity cost would incorporate the own rate on money in addition to the outside rate, but a consistent and complete series on the own rate is not currently available. Three feasible alternatives have been advanced, as follows. Friedman and Schwartz (1982) advocated using a fraction of RS, denoted RN and calculated as $(H/M) \cdot RS$. This measure assumes that all components of M except for high-powered money H earn interest at the (outside) short-term rate RS. Hendry and Ericsson (1991) proposed using the short-term interest rate RS itself because, over Friedman and Schwartz's sample, very little of M2 earned interest, and the interest rate on the interest-bearing component of M2 was itself very low. Finally, Ericsson, Hendry, and Prestwich (1998a) suggested modifying Friedman and Schwartz's measure RN so as to account for the changes in the measurement of high-powered money and broad money that occurred over the extended sample. Proper measurement of RN requires the unspliced data, even while the modeled money series is spliced. 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 1975 (for H) and 1987 (for M). Correspondingly, RN denotes $(H/M) \cdot RS$ (as above) with spliced series, and RN^a denotes $(H^a/M^a) \cdot RS$. Over 1871–1975, RS is virtually indistinguishable from RN and RN^a (aside from a scale factor) because H/M and H^a/M^a are essentially constant over that sample period. However, over 1975–1993, H/M and H^a/M^a plummet from 0.22 to 0.09 (for H/M) or 0.04 (for H^a/M^a), implying very different behavior of RS, RN, and RN^a.

Throughout, capital letters denote both the generic name and the level of a variable; logs are in lowercase; and "levels" often means the logarithm of the levels, with the context clarifying the usage. Uppercase delta Δ is the difference operator, 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, $\Delta_j^i x_t = (1 - L^j)^i x_t$ for positive integers i and j. If i or j is not explicit, it is taken to be unity. Single and double asterisks (* and **) adjacent to values of statistics denote significance at the 5% and 1% levels respectively.

3. Constancy, encompassing, and the annual models

This section describes models constructed on the initial annual dataset (1871–1975) and then on the extended dataset (1871–1993), paralleling these models' historical development. This progression traces and summarizes the encompassing nature of the model sequence, where encompassing occurs both across samples for a given model (constancy) and across models for a given sample ("standard" encompassing).

Hendry and Ericsson (1983) developed a linear equilibrium correction model (EqCM) over 1878–1970.[1] 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 EqCM. Each of the two new models encompassed the model in Hendry and Ericsson (1983), but neither could encompass the other. Hendry and Ericsson (1991) developed a model that improved on all three of the previous models, encompassing each one; and they extended the sample through 1975. Because that improved model encompasses the earlier models, it serves like a sufficient statistic for those earlier models. So, for our present purposes, we ignore the earlier models and focus on the improved model, which is an empirically constant EqCM of broad money M, in which the quantity theory of money defines the long-run equilibrium and the data determine the dynamics:

$$\begin{aligned}\Delta(m-p)_t = {} & \underset{(0.06)}{0.47}\,\Delta(m-p)_{t-1} - \underset{(0.04)}{0.11}\,\Delta^2(m-p)_{t-2} - \underset{(0.04)}{0.59}\,\Delta p_t \\ & + \underset{(0.05)}{0.41}\,\Delta p_{t-1} - \underset{(0.005)}{0.017}\,\Delta rs_t - \underset{(0.017)}{0.078}\,\Delta_2 r\ell_t \\ & - \underset{(0.20)}{1.15}(\hat{u}_{t-1} - 0.2)\hat{u}_{t-1}^2 + \underset{(0.002)}{0.007} + \underset{(0.6)}{3.4}(D_1 + D_3)_t \\ & + \underset{(0.9)}{7.1}\,D_{4t} + \underset{(0.027)}{0.090}\,D_{4t}\Delta rs_t \end{aligned} \tag{1}$$

$$T = 98\ [1878\text{–}1975] \quad R^2 = 0.88 \quad \hat{\sigma} = 1.478\% \quad Jt = 1.22 \quad Var = 0.05.$$

For each observation, the equilibrium correction residual $\hat{u}$ is calculated as:

$$\hat{u}_t = (m - p - i)_t - (-0.310 - 7.00 RS_t) \tag{2}$$

$$T = 98[1873\text{–}1970] \quad R^2 = 0.56 \quad \hat{\sigma} = 10.86\%,$$

---

[1] Although Hendry and Ericsson (1983), Longbottom and Holly (1985), Escribano (1985), and Hendry and Ericsson (1991) referred to their models as error correction models, technically speaking they are equilibrium correction models. See Hendry (1995, p. 213) for a discussion of the distinction between the two types of models.

where, for historical reasons, the coefficients in (2) are derived from a static Engle-Granger regression over 1873–1970.[2] Here and below, t is the annual time subscript; T is the number of annual observations; R^2 is the squared multiple correlation coefficient; $\hat{\sigma}$ is the standard deviation of the residuals, expressed as a percentage of real money and adjusted for degrees of freedom; Jt and Var are Hansen's (1992) statistics for testing joint parameter nonconstancy and variance nonconstancy; and OLS standard errors are in parentheses $(\cdot)$. The coefficients on the war dummies $(D_1 + D_3)$ and the deregulation dummy D_4 (and on D_c, below) have been scaled up 100-fold so that they are interpretable as percentages. The coefficient on $D_{4t}\Delta rs_t$ is *not* rescaled, so as to maintain units comparable to those of the coefficient on Δrs_t. While diagnostic tests are important in model evaluation, the papers cited above give a battery of such tests for (1)–(2) and the models examined below, so diagnostic tests (beyond those for constancy and encompassing) are not included herein.

Equations (1)–(2) use the annual data as compiled by Friedman and Schwartz (1982), which end in 1975. Attfield, Demery, and Duck (1995) and Ericsson, Hendry, and Prestwich (1998a, Section III) evaluate mechanistic extensions of this model over 1976–1993 and find 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. To wit, a coherent economic extension of (1)–(2) over the same sample obtains constant coefficients, as shown in Ericsson, Hendry, and Prestwich (1998a, Section V). This economic extension is:

$$\begin{aligned}\Delta(m-p)_t = {} & \underset{(0.05)}{0.48}\,\Delta(m-p)_{t-1} - \underset{(0.04)}{0.10}\,\Delta^2(m-p)_{t-2} - \underset{(0.04)}{0.62}\,\Delta p_t \\ & + \underset{(0.05)}{0.40}\,\Delta p_{t-1} - \underset{(0.006)}{0.020}\,\Delta rn^a_t - \underset{(0.016)}{0.041}\,\Delta_2 r\ell_t \\ & - \underset{(0.33)}{2.26}(\tilde{u}_{t-1} - 0.2)\tilde{u}^2_{t-1} + \underset{(0.002)}{0.004} + \underset{(0.5)}{3.9}(D_1 + D_3)_t \\ & + \underset{(0.7)}{5.2}\,D_{ct} + \underset{(0.026)}{0.100}\,D_{4t}\Delta rs_t \end{aligned} \qquad (3)$$

$$T = 116[1878\text{–}1993] \quad R^2 = 0.87 \quad \hat{\sigma} = 1.622\% \quad Jt = 1.93 \quad Var = 0.82^{**}$$

$$Cov: F(10,95) = 1.31 \quad Chow: F(20,85) = 2.73^{**},$$

where $\tilde{u}$ is the Engle-Granger residual from:

$$\tilde{u}_t = (m-p-i)_t - (-0.318 - 6.67RN^a_t) \qquad (4)$$

$$T = 98[1873-1970] \quad R^2 = 0.59 \quad \hat{\sigma} = 10.57\%.$$

Equations (1) and (2) require two economic extensions to obtain (3) and (4). First, the deregulation dummy D_4 is extended as D_c when entering by itself in

[2] Due to increased numerical accuracy in recent versions of PcGive, the estimated intercept in (2) differs slightly from the one reported in Hendry and Ericsson (1991, equation (9)). See Ericsson, Hendry, and Prestwich (1998a, footnote 4) for details.

order to capture the deregulation in 1986–1989. Second, the measure of the opportunity cost (RS) is replaced by RN^a in order to capture the decline in the non-interest-bearing fraction of measured money during the last two decades.

The coefficients in (1) and (3) are virtually identical, confirming the constancy of these equations' coefficients. Forecasts and recursive estimates of (3) in Ericsson, Hendry, and Prestwich (1998a, Figures 7 and 8) and the insignificance of Jt and the Chow (1960) covariance statistic (Cov) underscore the constancy of (3)'s coefficients. That said, $\hat{\sigma}$ has increased by about 10%, and Var and the Chow (1960) predictive-failure statistic ($Chow$) reject, indicating that the nonconstancy present is due to omitted variables nearly orthogonal to the included variables. Equally, rejection by the predictive-failure and Var tests reflects these tests' high power to detect numerically modest changes in $\hat{\sigma}$. The long sample and the high variance of the data relative to that of the equation error are the proximate reasons for that high power. Overall, the model (3) performs well for this recent turbulent period in the U.K. economy, while its increased error variance reveals that further improvements are possible.

Formal encompassing tests of (1) and (3) strongly favor equation (3). For the samples considered (1878–1973, 1878–1975, and 1878–1993), the parameter-encompassing statistics are: $F(2,83) = 1.22$ [0.30], $F(2,85) = 5.28$ [0.0069], and $F(2,103) = 22.15$ [0.0000] for whether (1) encompasses (3); and $F(2,83) = 0.13$ [0.88], $F(2,85) = 0.17$ [0.85], and $F(2,103) = 0.45$ [0.64] for whether (3) encompasses (1). $F(\cdot,\cdot)$ denotes the asymptotic null distribution, and p-values are in brackets. This analysis and the more detailed results in Ericsson, Hendry, and Prestwich (1998a) establish that (3) encompasses key aspects of (1) on the extended sample through suitable measurement of the opportunity cost and proper adjustment for financial deregulation. Thus, the following sections turn to assessing these annual models in light of phase-average models.

4. Phase-average data and a model thereof

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

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

$$\bar{x}_j = \frac{\frac{1}{2}x_{t_j} + x_{t_j+1} + \cdots + x_{t_j+c_j-1} + \frac{1}{2}x_{t_j+c_j}}{c_j} \quad j = 1, \ldots, J, \tag{5}$$

where an overbar denotes phase averaging, j is the index for phase averaging, c_j is the phase length of the jth phase (in years), t_j is the first year of the jth phase, 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 (1982, Chapters 3 and 4) phase-average 108 annual observations (1868–1975) on the logarithms of money, prices, incomes, and population, and the levels of the interest rates to obtain 38 phase averages for each series.

Friedman and Schwartz (1982, p. 282) present their central regression for U.K. money demand in their Table 6.14. We could closely replicate that equation:

$$(\bar{m}-\bar{p}-\bar{n})_j = \underset{(0.19)}{0.02} + \underset{(0.049)}{0.883}(\bar{\imath}-\bar{n})_j - \underset{(3.29)}{11.22}\,\overline{RN}_j - \underset{(0.29)}{0.21}\,G(\bar{p}+\bar{\imath})_j$$

$$+ \underset{(0.58)}{1.38}\,\overline{W}_j + \underset{(2.7)}{20.6}\,\bar{S}_j \qquad (6)$$

$J = 36$ [1874–1973] $\quad R^2 = 0.97 \quad \hat{\sigma} = 10.36\%$.

The variable $G(\bar{p}+\bar{\imath})$ is the two-sided growth rate of nominal income, $\overline{W}$ is a dummy for "postwar adjustment", and $\bar{S}$ is a data-based dummy for "[a]n upward demand shift, produced by economic depression and war ... " during 1921–1955, where quotes are from Friedman and Schwartz (1982, pp. 228, 281).[3] Friedman and Schwartz also consider a similar equation in rates of change, discussed below.

Hendry and Ericsson (1991) document evidence on the mis-specification of (6), including rejection of residual normality, price homogeneity, and constancy, the last by both Chow and covariance statistics. Hansen's statistics provide further evidence of nonconstancy: $Jt = 2.15^{*}$ and $Var = 0.23$. That said, comparison of (1) and (3) with (6) is still of interest because an otherwise apparently well-specified model may yet fail to encompass a model with known mis-specification. Sections 5 and 6 thus interpret and assess the phase-average results in light of their annual counterparts.

5. An interpretation of phase-average error variances

Encompassing statistics are relatively easy to calculate for annual models 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 (5), 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 required annual model 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 a derivation of the formal encompassing statistic.

---

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

Section 6 below offers two alternatives: it examines the error variances of annual and phase-average models, and it compares phase-average outcomes with annual outcomes transformed to phase averages. Examination of error variances is of interest because a necessary condition for encompassing is variance dominance, where one equation dominates another equation in variance if the former has a smaller error variance; see Hendry and Richard (1982). To assess variance dominance, the error variances from the two types of models must be in the same units. By re-examining the results in Friedman and Schwartz (1982), the current section shows how to ensure that the variances are in the same units.

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. 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 $IN(0, \sigma^2)$: that is, x_t is serially independent and normally distributed with mean zero and variance σ^2. From (5), the phase-average variable $\bar{x}_j$ is distributed as $N(0, \sigma^2/k_j^2)$, where $k_j^2 = 2c_j^2/(2c_j - 1)$. Even though the annual variable x_t is homoscedastic, the phase average $\bar{x}_j$ is not, as its variance depends upon the phase length c_j. To correct this heteroscedasticity, Friedman and Schwartz (1982) used weighted least squares regressions, where the weights depended 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.) $N(0, \sigma^2)$ for $\bar{x}_j$ regressed on a constant. However, Friedman and Schwartz (1982) [and Attfield, Demery, and Duck (1995) as well] used *normalized* weights κ_j (say), where $\kappa_j = k_j \backslash \overline{k}$ and the normalization factor $\overline{k}$ is the approximate sample mean of k_j. Because $\overline{k}$ simply rescales the entire weighted least squares regression, the choice between k_j and κ_j does not affect coefficient estimates, their standard errors, or their t-ratios. The equation standard error *is* affected, though, noting that $\kappa_j \bar{x}_j$ is distributed as $N(0, \sigma^2/\overline{k}^2)$. Thus, for a phase-average regression estimated by weighted least squares with normalized weights, the equation standard error must be multiplied by $\overline{k}$ 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 (6)], 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 $\overline{k} = 1.826$, the $\hat{\sigma}$ comparable to annual results is $5.54\% \cdot 1.826 = 10.12\%$. That is virtually 10.36%, the value for $\hat{\sigma}$ obtained in (6), 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 7.19, 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 actual $\hat{\sigma}$ comparable to annual results is $1.34\% \cdot 8.00 = 10.72\%$.

6. Constancy and encompassing with different measures of data

Using the adjustments from Section 5, Section 6.1 compares estimated values of $\hat{\sigma}$ from annual and phase-average models for U.K. money demand to determine the direction of variance dominance.[4] Variance dominance is a property based on in-sample calculations, and variance dominance is necessary for encompassing. Section 6.1 also performs similar comparisons for out-of-sample calculations, where dominance in root mean squared error (RMSE) is necessary for forecast encompassing; see Ericsson (1992). Once comparable units and time periods are obtained, the annual models clearly dominate the phase-average models, both in variance and in RMSE. Then, Section 6.2 evaluates the models' statistical and numerical constancy. These combined results allow re-assessment of claims in Friedman and Schwartz (1991) and Attfield, Demery, and Duck (1995).

6.1 Encompassing and goodness of fit

Table 1 lists comparably scaled 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 (1), and the translated annual model (3). This subsection compares these two measures of fit, both across models and across samples.

When estimated over the initial sample of 1878–1973 (equivalent to phases 2–36), the annual models (1) and (3) have virtually identical equation standard errors (1.424% and 1.406%) and substantially variance-dominate both phase-average equations, which have equation standard errors of 10.36% and 10.75% respectively. The annual models also variance-dominate the phase-average models when estimated over the new sample (1974–1993) and over the extended sample (1878–1993). Typically, a phase-average model's equation standard error is 7 or more times that of a corresponding annual model, reflecting just how much better the annual models explain the data relative to the phase-average models. Put slightly differently, roughly 98% of the residual variation in a phase-average model is explained by either annual model.

As this comparison shows, the EqCMs (1) and (3) fit far better than Friedman and Schwartz's phase-average equations in levels and rates of change. One potential explanation is the treatment of dynamics: (1) and (3) incorporate dynamics directly, whereas Friedman and Schwartz's equations

[4] The situation is further complicated by the *data-based* selection of the phases, contrasting with the annual model being fit and forecast to data at fixed 1-year intervals. However, Campos, Ericsson, and Hendry (1990) show that the dominant effects of Friedman and Schwartz's phase averaging appear to be captured by fixed-length phase averaging, so the analysis below ignores the consequences of endogenously selecting the phases.

Table 1. Estimated equation standard errors ($\hat{\sigma}$), RMSEs, and constancy statistics for phase-average and annual models

Statistic [phase observations in brackets]	Phase-average Data		Annual Data	
	Log-level model (6)	Growth-rate model	Initial model (1)	Translated model (3)
$\hat{\sigma}$				
[2–36] 1878–1973	10.36	10.74	1.424	1.406
[37–44] 1974–1993	21.70	28.37	2.915	2.644
[2–44] 1878–1993	10.95	12.67	1.931	1.622
RMSE: original data				
[37–42] 1974–1988	13.93	20.76	6.750	2.648
[37–44] 1974–1993	–	–	28.124	2.755
RMSE: Figure 1				
[1–36]	10.15	–	–	–
[2–36]	10.19	–	1.346	–
[37–42]	14.75	–	6.110	–
[37–43]	–	–	13.533	–
Chow statistic				
value	1.71	3.28^{*}	5.40^{**}	2.73^{**}
p-value	0.1536	0.0139	0.0000	0.0007
degrees of freedom	(6, 30)	(6, 29)	(20, 85)	(20, 85)
Covariance statistic				
value	0.30	1.03	4.95^{**}	1.31
p-value	0.8736	0.4065	0.0000	0.2342
degrees of freedom	(4, 32)	(4, 31)	(10, 95)	(10, 95)
Hansen statistic				
Jt	1.71	1.10	2.58	1.83
Var	0.28	0.58^{*}	0.86^{**}	0.78^{**}

Notes:
1. All values of $\hat{\sigma}$ and RMSE are in percent.
2. Listed dates correspond to estimation periods (for $\hat{\sigma}$) and forecast periods (for the RMSE).
3. Phase observations appear in square brackets. For $\hat{\sigma}$, the listed phase observations are the literal equivalents to the years listed. Samples for the estimation of phase-average models are 1–36, 37–42, and 1–42 for the log-level model, and 2–36, 37–42, and 2–42 for the growth-rate model. For RMSEs, the listed phase observations are the literal equivalents to the years listed and are also the actual samples. Values of $\hat{\sigma}$ and RMSE for phase-average samples are converted to annual units for comparability.
4. Statistics for testing constancy are calculated over phases 1–42 and 2–42 for the phase-average equations in levels and growth rates and over 1878–1993 for the annual equations.

do not. That said, the latter do implicitly adjust for dynamics through phase averaging, which uses 64 degrees of freedom, contrasting with only 11 degrees of freedom for dynamics in (1) and (3). In that light, comparison of the two phase-average equations with the (static) annual Engle-Granger regressions (2) and (4) offers a stark inference. The equation standard errors for those four equations are essentially the same, implying that phase averaging failed entirely in its ostensible purpose—to remove the short-run fluctuations in the data.

The first set of RMSEs in Table 1 are obtained by estimating each model over its initial sample (e.g., phases 2–36, or 1878–1973) 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 (6) and (3) is over 5, similar to the in-sample results. When the forecast sample is extended through 1993, the RMSEs for the annual models increase: marginally for (3) and dramatically for (1). The change in RMSE for (1) arises from that equation's poor measure of opportunity cost (RS) and a lack of accounting for data splicing. RMSEs for the phase-average models can be computed through only phase 42 (1985–1988) because one regressor in the phase-average models is the two-sided growth rate $G(\bar{p}+\bar{\imath})$. Its calculation for phase 43 requires $\bar{p}+\bar{\imath}$ for the (not yet complete) phase 44 (1991 onward). Still, the RMSE over 1974–1993 for the annual model (3) is much smaller than the equation standard errors and RMSEs of the phase-average models over any of the samples.

Choice of units for $\hat{\sigma}$ can affect inference. In Table 1, values of $\hat{\sigma}$ (and of the 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) did not account for that normalization in their comparisons of phase-average and annual models. Because of this technical mistake, Friedman and Schwartz (1991, pp. 46–47) stated that their regressions variance-dominate those in Hendry and Ericsson (1991). Rather, the converse is true, as Table 1 shows. On a related issue, because (1) and (3) are EqCMs, both are equations in log-levels, even though their dependent variables are growth rates. Thus, these equations' values of $\hat{\sigma}$ require no additional rescaling to be comparable to $\hat{\sigma}$ for (6), contrary to Friedman and Schwartz's (1991, footnote 9) claim.

The relationship between annual and phase-average data 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 1 does precisely this for the initial annual model (1) and the phase-average model in levels, (6). Figure 1*a* plots the actual, fitted, and forecast values for velocity v over the extended sample (phases 1–43), and Figure 1*c* plots the corresponding residuals and forecast errors. Figures 1*b* and 1*d* are comparable to Figures 1*a* and 1*c* but plot values over the forecast sample alone. The phase-average model generally has larger errors than the annual model, an outcome consistent with the comparison of Figures 4 and 7 in Hendry and Ericsson (1991). The phase-average model's forecast error in phase 37 is particularly large, being nearly 20%. Using the transformed annual model (3) rather than the initial annual model (1) obtains an even more pronounced contrast between the annual and phase-average results.

The second set of RMSEs in Table 1 numerically summarize the graphical results in Figure 1.[5] As with the earlier comparisons, the annual model (1) always dominates the phase-average model (6) for matched samples, reflecting the superior performance of the annual model. Over phases 37–43, the 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.

[5] 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.

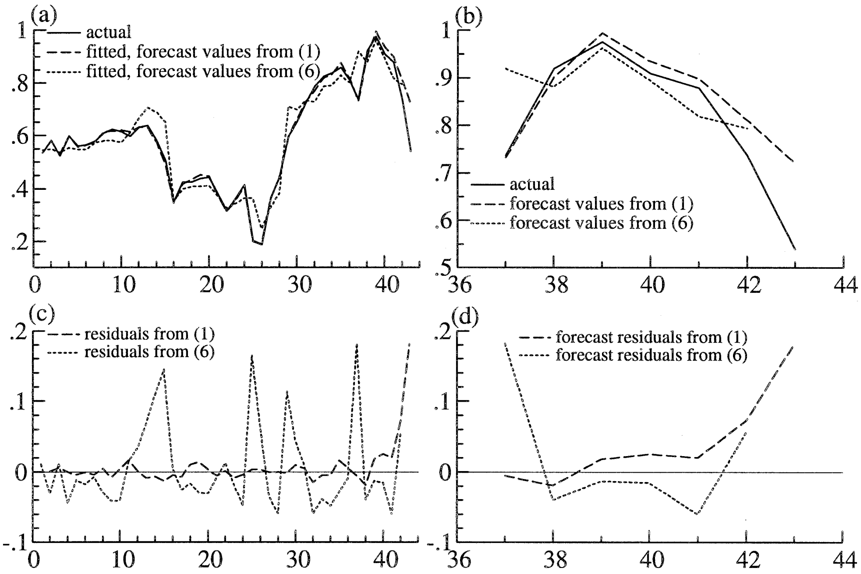

Fig. 1. Actual, fitted, and forecast values for velocity v from the annual equation (1) (––) and the phase-average equation (6) (---), and the corresponding residuals, all in a common phase mapping

6.2 Tests of constancy

Tests of constancy are an alternative metric for assessing the relative fit of models, so this subsection analyzes the constancy of the phase-average and annual models. The annual models (1) and (3) both fail the Chow predictive-failure and Hansen *Var* tests, and (1) fails the covariance test as well: see Table 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 dominance 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 on the number of observations available. The annual models fit the data well in sample, their coefficients are precisely estimated, and the forecast period has a large number of observations: 20 for 1974–1993, or about 20% the number in the initial dataset. Combined, these features ensure that the constancy tests have high power to detect even small nonconstancies in the annual models. Conversely, the phase-average models fit poorly in sample, equation (6) in particular is detectably nonconstant in sample even though that sample is small, and the number of observations forecast is small (only 6). 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: the values of $\hat{\sigma}$, RMSE, and the Chow statistic demonstrate this distinction; see Ericsson (1992). 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 distinction between statistical and numerical properties has immediate implications for inferences, as did rescaling $\hat{\sigma}$ in Section 6.1. Specifically, the results in Table 1 and Figure 1 allow a re-assessment of Attfield, Demery, and Duck (1995), who tested for the constancy of (6) over phases 37–42 [1973–1988] and the constancy of (1) over 1976–1993. Using Chow and covariance tests, Attfield, Demery, and Duck (1995, Tables 1–8) found that (6) appears constant whereas (1) does not. In concluding, they claimed:

... 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 (1991)] 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 (1) is statistically less constant than the phase-average model (6), the annual model provides a far better *explanation* of phase-average money than does the phase-average model; see Figure 1. Attfield, Demery, and Duck (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 the data so poorly. For the phase-average model in levels, tests of constancy out-of-sample are further contaminated because that model is known to be nonconstant in-sample; see Hendry and Doornik (1997). Third, in Attfield, Demery, and Duck (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 1*a*, 1*b*, 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, Demery, and Duck (1995) ignore the economic consequences of redefining the dependent variable. In particular, they mechanically extended the annual model (1) when testing its constancy over 1976–1993. If that model is economically extended, as in (3), it performs quite respectably.

7. Conclusions

The central role of money demand relationships in economic policy has stimulated many empirical studies, including those cited above, which developed various empirical models of U.K. money demand using Friedman and Schwartz's annual and phase-average data. This paper evaluates key models from those studies, employing tests of encompassing and constancy. A historical sequence of annual models has led to an equilibrium correction model on data through 1993, where that model encompasses the earlier annual models. That model also captures salient features in the data from 1975 (the end of Friedman and Schwartz's sample) through 1993—a period with radical changes in economic policy. The model's performance on the new data depends directly upon sensible economic choices for extending the time series of those data to reflect their altered measurement. That said, the annual model, whether mechanistically or economically extended, explains the data far better than the phase-average models. This result holds over the fitted sample and forecast periods; and it holds regardless of whether outcomes are compared across the different datasets on a uniform metric, or whether the outcomes from the annual model are transformed into phase averages and then compared with the outcomes from 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 it points to the possibility of further progress in model specification.

References

Attfield CLF, Demery D, Duck NW (1995) Estimating the UK demand for money function: A test of two approaches. Mimeo, Department of Economics, University of Bristol, Bristol, England

Campos J, Ericsson NR, Hendry DF (1990) An analogue model of phase-averaging procedures. Journal of Econometrics 43, 3:275–292

Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 3:591–605

Doornik JA, Hendry DF (1996) PcGive Professional 9.0 for Windows. International Thomson Business Press, London

Ericsson NR (1992) Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. Journal of Policy Modeling 14, 4:465–495

Ericsson NR, Hendry DF, Prestwich KM (1998a) The demand for broad money in the United Kingdom, 1878–1993. Scandinavian Journal of Economics 100, 1:289–324

Ericsson NR, Hendry DF, Prestwich KM (1998b) Friedman and Schwartz (1982) revisited: Assessing annual and phase-average models of money demand in the United Kingdom. International Finance Discussion Paper, Board of Governors of the Federal Reserve System, Washington, D.C., available on the WorldWide Web at www.bog.frb.fed.us/pubs/ifdp/1998/

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

Friedman M, Schwartz AJ (1982) Monetary trends in the United States and the United Kingdom: Their relation to income, prices, and interest rates, 1867–1975. University of Chicago Press, Chicago

Friedman M, Schwartz AJ (1991) Alternative approaches to analyzing economic data. American Economic Review 81, 1:39–49

Hansen BE (1992) Testing for parameter instability in linear models. Journal of Policy Modeling 14, 4:517–533

Hendry DF (1995) Dynamic econometrics. Oxford University Press, Oxford

Hendry DF, Doornik JA (1997) The implications for econometric modelling of forecast failure. Scottish Journal of Political Economy 44, 4:437–461

Hendry DF, Ericsson NR (1983) Assertion without empirical basis: An econometric appraisal of 'Monetary trends in … the United Kingdom' by Milton Friedman and Anna Schwartz. In: Monetary trends in the United Kingdom, Bank of England Panel of Academic Consultants, Panel Paper No. 22, Bank of England, London, October, pp. 45–101

Hendry DF, Ericsson NR (1991) An econometric analysis of U.K. money demand in 'Monetary trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz. American Economic Review 81, 1:8–38

Hendry DF, Richard J-F (1982) On the formulation of empirical models in dynamic econometrics. Journal of Econometrics 20, 1:3–33

Judd JP, Scadding JL (1982) The search for a stable money demand function: A survey of the post-1973 literature. Journal of Economic Literature 20, 3:993–1023

Longbottom A, 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, England

Mizon GE, Richard J-F (1986) The encompassing principle and its application to testing non-nested hypotheses. Econometrica 54, 3:657–678