The Economic and Social Review, Vol. 27, No. 5, October, 1996, pp. 401-422

On the Constancy of Time-Series Econometric Equations

DAVID F. HENDRY*
Nuffield College, Oxford

Abstract: Parameter constancy is a fundamental requirement for empirical models to be useful for forecasting, analysing economic policy, or testing economic theories. However, there are surprises in defining a constant-parameter model, such that models with time-varying coefficients, and expansion of the parameterisation over time are both compatible with constancy, yet unbiased forecasts may not entail a sensible model choice. In-sample tests cannot determine likely post-sample predictive failure. A comparison of two models of UK money demand illustrates the analysis empirically, as one suffers considerable predictive failure yet the other does not, despite being identical in-sample.

I INTRODUCTION

Parameter constancy is a fundamental requirement for empirical models to be useful for forecasting, analysing economic policy, or testing economic theories. Nevertheless, it remains unclear precisely what constancy entails, what aspects of models should be constant, and what features of models in-sample might help diagnose likely post-sample predictive failure. Consequently, this paper addresses the concept of parameter constancy in econometric models, its characterisation, and the implications of predictive failure.

Dublin Economic Workshop Guest Lecture, Tenth Annual Conference of the Irish Economic Association, Dromoland Castle, 12-14 April, 1996.

*This research was financed in part by grants B00220012 and L11625015 from the UK Economic and Social Research Council. Helpful comments from Gunnar Bårdsen; Mike Clements; Jurgen Doornik; Neil Ericsson; Grayham Mizon; John Muellbauer; and Neil Shephard are gratefully acknowledged.

We begin with the definition offered in Hendry (1995a): a parameter is constant if it does not change over time. Further, a model is constant if none of its parameters changes. There are four surprises hidden in this definition of a constant-parameter model, several of which have apparently not been discussed previously. First, whether or not a given model is constant depends on how its parameter space is viewed: for example, models with time-varying coefficients may be constant, providing the underlying parameters are constant. Section 3.4 illustrates this issue.

Second, the above formulation of constancy allows for what might be deemed considerable non-constancy, including the expansion of the parameterisation — providing existing parameters stay the same. For example, the impact of changes in data definitions, or in the measurement system (e.g., the size of the banking sector; or a measure of opportunity cost) may necessitate adding variables and ostensibly new parameters to a model to capture the relevant changes, yet leave the model constant. Section III discusses this issue, and Section VI illustrates the analysis by an application to the demand for narrow money in the United Kingdom.

Third, although predictive failure will manifest itself as apparent parameter change prior to an extension of the model under scrutiny, and constancy as defined above requires that the pre-existing parameters are recovered after the model extension, nevertheless, these together do not need orthogonality of the new variables to those already present. Rather, that state would ensure no change in existing parameters. If adding new variables to re-establish full-sample constancy necessitates changing pre-existing parameters, then genuine parameter change occurs in the space under analysis.

Fourth, in-sample tests cannot detect whether or not an empirical model will manifest predictive failure out of sample. The illustration in Section VI apparently fails badly when extended out of sample, yet the same in-sample model extended in an alternative way does not fail. Any test that anticipated failure (success) would be correct (wrong) for the first extension, and the reverse for the second. Yet, in general, one cannot know which is the valid extension in an empirical setting. This outcome also helps explain why models may fit better than they predict in practice, since the relevant extension may be contingent on events that are not easily predictable a priori.

Section II reviews historical concerns about constancy and some earlier characterisations. Section III offers more formal definitions and develops some of their implications. The role of economic theory in determining empirical parameter constancy is discussed in Section IV. Then we turn to potential causes of predictive failure in Section V. The empirical example of UK M1 demand in Section VI is a major focus of the paper as it illustrates the various concepts in a well-known setting. Section VII concludes.

II HISTORICAL DEBATE ABOUT CONSTANCY

The historical study of econometrics has produced a wealth of information about earlier views on constancy and invariance (see, *inter alia*, Epstein, 1987; Morgan, 1990; Qin, 1993; and Hendry and Morgan, 1995) confirming that such concepts have long concerned econometricians. Since a detailed analysis of the foundations of econometrics is provided in Hendry and Morgan (1995), the following comments merely serve to set the scene.

The long-standing historical debate on the role of econometrics in economics alluded regularly to the potential — or otherwise — for permanence or constancy in empirical economic relationships (see e.g., Hendry, 1995b). An early volley was fired by Henry Moore (1914) who believed that the dominant economic theory of his day (the comparative statics method of economic analysis) failed as a substitute for the experimental method of the exact sciences. Moore therefore tried to give a "concrete reality" to economic relationships, using multiple regression analyses to accord their proper roles to dynamics and the multivariate structure of economic behaviour. Such empirical relationships implicitly require at least within-sample constancy.

Lionel Robbins (1932) strongly disagreed with Moore's approach, although he actually directed his criticisms at Henry Schultz (1928). In particular, Robbins claimed that the formal categories of economic theory could not be given numerical representations, since neither individual values nor technical causes were uniform over time or space. Thus, statistical laws in economics failed Robbins' basic test: when the world changed, empirical laws changed. Robbins' parody of Dr Blank seems to be the first of several attempts by economic theorists to deny a substantive role for econometrics in economics based on arguments about the non-constancy of the underlying economic reality. However, all forms of empirical evidence would be transient in Robbins' formulation, which may well be true in the long run, but is an extreme view over short periods.

While there were no direct replies to Robbins, Jan Tinbergen (see Tinbergen, 1940) gave many practical demonstrations of his concern with parameter constancy in empirical models. For example, he tested his chosen equations on several sub-periods to check parameter constancy; he used forecasting tests to evaluate each equation's performance; and he tested the robustness of regression coefficients when adding other variables, as well as on the up-swing versus the down-swing of the cycle (see Morgan, 1990; and Hendry and Morgan, 1995).

Nevertheless, Tinbergen's approach was criticised by both John Maynard Keynes and Ragnar Frisch (see Keynes, 1939, and Frisch, 1938 — printed in Hendry and Morgan, 1995). Keynes claimed a number of "pre-conditions" for the validity of inferences from data, including both "time homogeneity" (or parameter constancy) and a complete prior theoretical analysis, so he held to an extreme form of the "axiom of correct specification" (see Leamer, 1978): statistical work in economics was deemed impossible without prior theoretical knowledge. In effect, Keynes' criticisms simply assumed the precedence of "theory" over "evidence". However, as argued in Hendry (1995b), if partial explanations are devoid of use (i.e., we cannot discover empirically anything that is not already known theoretically), Keynes must have believed no science ever progressed.

The critique in Frisch (1938) was more directly focused on autonomy, and hence constancy (see Aldrich, 1989; Epstein, 1987; and Qin, 1989). Frisch defined autonomous (structural) relations as those which were invariant to changes elsewhere in the economic system, whereas confluent relations were those which arose from similar time-series patterns producing correlations which might happen to be regular at a phenomenological level but were not based on underlying "substantive" relations: he cited the Harvard A-B-C curves as an example (see Persons, 1924). He then claimed that only descriptive equations could be discovered from passive observations, but such equations need not be autonomous and may not continue to hold if other equations in the system altered. Frisch saw the need for a "super-structure" which should be given by economic theory, since invariance to hypothetical variations could not be learned from the available data. Consequently, he doubted whether Tinbergen could have found genuinely structural relations. His analysis is a clear statement of an argument later often attributed to Lucas (1976): see Favero and Hendry (1992) for an extensive discussion.

In his classic work, Haavelmo (1944) did not comment directly on these debates, but instead constructively formalised many of the essential concepts of econometrics, including autonomy. His treatment sets the scene for many later analyses and developments, but to appraise that literature, we now need some formal definitions and concepts, and will use those proposed in Hendry (1995a).

III CHARACTERISING CONSTANCY

3.1 *Parameter*

A parameter is a numerical entity which indexes a stochastic process (or a distribution function), and so does not alter across realisations of the relevant random variables. By definition, therefore, a parameter must remain constant across realisations of the stochastic process, but it may or may not be constant over time, or dependent on other stochastic processes. As such, a parameter vector θ of k elements is a point in a parameter space Θ, which delineates its admissible values, usually $\theta \in \Theta \subseteq \mathbb{R}^k$. Distribution functions are invariant to 1 – 1 reparameterisations $\phi = f(\theta)$ (e.g., variances to standard deviations, noting both are non-negative by definition), so θ and ϕ are equivalent parameterisations. A parameter θ is identifiable if different distributions always result when θ takes different values.

3.2 *Constancy*

The parameter θ is constant over the time period $T = \{\ldots,-2,-1,0,1,2,\ldots\}$ if θ has the same value for all $t \in T$. A model is constant over the time period T if all of its parameters are constant. As the historical debate discussed in Section II showed, constancy has long been regarded as a fundamental requirement for empirical modelling: in practice, models with no constancies cannot be used for forecasting, analysing economic policy, or testing economic theories.

Despite the attempt to be precise and unambiguous, there remain non-negligible problems with these definitions when the "correct" parameterisation of the data generation process is not known a priori. We now consider three inter-related issues: the composition of the parameter vector; the parameterisation itself; and the model formulation.

3.3 *Composition of the Parameter Vector*

First, precisely what are the parameters in any given setting? Depending on how a model is formulated, "over-parameterisation" may occur inadvertently simply by not imposing constraints that are valid in the population. For example, consider a linear model which involves (say) $\theta(x_{1,t} + x_{2,t})$ but is written as $\theta_1 x_{1,t} + \theta_2 x_{2,t}$. Since the (non)-constancy of θ uniquely entails the (non)-constancy of (θ_1,θ_2), the redundancy seems unproblematic. However, since 1 – 1 transformations are allowed, zero may appear as a "parameter", as in the equivalent representation:

$$\theta_1(x_{1,t} + x_{2,t}) + (\theta_2 - \theta_1)x_{2,t} = \phi_1(x_{1,t} + x_{2,t}) + \phi_2 x_{2,t}.$$

In the present example, $\phi_2 \equiv 0$, so is certainly constant. Unfortunately, as will be seen shortly, we have just opened Pandora's box.

3.4 *Time Variation*

Second, constant models can have time-varying coefficients (i.e., initially-conjectured parameters), providing such models have an underlying set of constant parameters which characterise the probability mechanism. For example, consider the model:

$$y_t = \alpha_t z_t + \varepsilon_t \text{ where } \varepsilon_t \sim \mathrm{IN}\left[0, \sigma_\varepsilon^2\right] \tag{1}$$

when, for simplicity:

$$z_t = \lambda z_{t-1} + \omega_t \text{ where } \omega_t \sim \mathrm{IN}\left[0, \sigma_\omega^2\right]$$

with $|\lambda| < 1$, so z_t is stationary and strongly exogenous for α_t. Then the parameterisation $\left(\alpha, \sigma_\varepsilon^2\right)$ in (1) is not constant when any $\alpha_t \neq \alpha$. However, let α_0 be a fixed parameter in the auxiliary process:

$$\alpha_t = \rho\alpha_{t-1} + \nu_t \text{ where } \nu_t \sim \mathrm{IN}\left[0, \sigma_\nu^2\right] \tag{2}$$

where $\rho = 1$, and $\{\nu_t\}$ is independent of $\{\varepsilon_t\}$. When y_t and z_t are defined over a sample space Ω, and $\{\nu_t\}$ does not depend on Ω, then (1) is a random-coefficients model. Nevertheless, it has an underlying constant parameterisation, defined by the constant unit root ρ in (2), the constant linear functional form in (1), the constant means of zero and fixed variances of the constant normal distributions of $\{\varepsilon_t\}$ and $\{\nu_t\}$ (as well as the constant absence of serial dependence in $\{\varepsilon_t\}$ and $\{\nu_t\}$).

To see that the parameterisation of the conditional relation is constant, reformulate the model in (1) as:

$$y_t = \alpha_t z_t + \varepsilon_t = (\alpha_{t-1} + \nu_t) z_t + \varepsilon_t = \alpha_0 z_t + \left(\sum_{i=0}^{t-1} \nu_{t-i}\right) z_t + \varepsilon_t = \alpha_0 z_t + u_t \tag{3}$$

where α_0 is constant. Also, u_t is a mean-zero, normal process conditional on z_t, although it is autocorrelated and highly heteroscedastic, so all parameters in (3) are not constant:

$$V[u_t] = E\left[\sum_{i=0}^{t-1} \nu_{t-i}^2\right] E\left[z_t^2\right] + E\left[\varepsilon_t^2\right] = \frac{\sigma_\nu^2 \sigma_\omega^2}{1-\lambda^2} t + \sigma_\varepsilon^2.$$

In effect, the formulation in terms of $\{\alpha_t\}$ in (1) involved latent variables rather than parameters.

This issue is perhaps clearer in the formulation of structured time-series models as in Harvey (1981) and Harvey and Shephard (1992), such as:

$$x_t = \mu_t + w_t$$
$$\mu_t = \mu_{t-1} + \lambda + \nu_t$$

where $\{w_t\}$ and $\{v_t\}$ are mutually independent, identically distributed processes. At first sight, these again seem not to be "constant-parameter" models since $\mu_t \neq \mu \; \forall t$, but as Harvey shows, differencing x_t yields:

$$\Delta x_t = \lambda + \Delta w_t + v_t,$$

so the generated series has a constant unit root, a constant growth rate λ and a constant negative moving-average error, being a special case of a constant-parameter ARIMA process.

3.5 *Model Expansion*

As noted in the introduction, somewhat surprisingly, the definition of model constancy also does not even preclude model expansion, or adaptation, providing existing parameters stay the same. For example, over a sample up to a time T_1, consider the model:

$$y_t = \theta x_t + \varepsilon_t \text{ where } \varepsilon_t \sim \mathrm{IN}\left[0, \sigma_\varepsilon^2\right] \tag{4}$$

which describes the available data congruently, with constant, and highly significant, $\theta > 0$. The same equation is then fitted to a sample $T_1+1,\ldots,T$, but now describes the data extremely poorly, and delivers the very different parameter $\delta < 0$ (say) for the coefficient of x_t. Apparently, the model is non-constant. In fact, there was a change in the measurement process for x_t such that the "correct measure" after T_1 becomes $x_t^* = x_t + D_t z_t$ where D_t is an indicator variable which is unity after T_1 and zero otherwise, when z_t is the correction for mis-measuring x_t. Thus the whole-period model is

$$y_t = \theta x_t^* + \varepsilon_t \text{ where } \varepsilon_t \sim \mathrm{IN}\left[0, \sigma_\varepsilon^2\right]. \tag{5}$$

Written as in (5), the model is constant; however, written as:

$$y_t = \theta(x_t + D_t z_t) + \varepsilon_t = \theta_1 x_t + \theta_2 D_t z_t + \varepsilon_t \tag{6}$$

we have model expansion, and return to the first issue of redundant parameters. Whether or not the model is constant appears to depend on how it is written, which potentially depends on the inclusion of irrelevant zeroes, despite the invariance of the distribution function to $1-1$ reparameterisations. There is no requirement of orthogonality between the original and extended variables: rather, if excluded variables are, and remain, orthogonal to included, the initial parameters will be unaffected by their exclusion.

Despite these apparent drawbacks, our earlier definition of constancy has operational content. First, (4) is not constant over the whole sample period in the space of (y_t, x_t); that issue is unambiguous. Using (4) to forecast over $T_1 + 1,\ldots,T$ would be inadvisable. Second, in terms of (6), constancy requires $\theta_1 = \theta$, namely the original parameter of (4). Thus, if in (6), $\theta_1 \neq \theta$, the expanded model would remain non-constant. In turn, that would not preclude a parameterisation in which the model was constant.

3.6 *Predictive Failure Tests*

Finally, it follows that there are no possible in-sample tests to detect whether or not an empirical model will manifest predictive failure out of sample. Any test that would have predicted the failure of (4) when extended out of sample using just x_t as a regressor would do so correctly when the existence of $D_t z_t$ was unknown. Unfortunately, the same test would deliver the wrong answer when the in-sample model was extended in the alternative way as in (6) which does not fail. Any test that anticipated failure (success) would be correct (wrong) for the first extension, and the reverse for the second. Yet, in general in economics, it cannot be known for certain even after the event which is the valid extension.

Of course, tests of the *ex post* non-constancy when using just x_t as a regressor do deliver the correct result for that model, and so remain a useful tool for rejecting inappropriate specifications. The preceding paragraph refers to using a test based on information up to T_1 only, in order to anticipate how a given model will fare thereafter.

Consequently, predictive failure is uniquely a post-sample problem, requiring change somewhere to "cause" change elsewhere. As argued in Hendry (1979), in large samples from weakly stationary processes, models based on first and second moments of the data will fit as well to later samples as they did to the one they were selected from (subject to caveats from rare events, and perhaps "overfitting"). Thus, when they fail to do so, some form of non-stationarity must be operating.

In many respects, the issue is analogous to discriminating a long cyclical upswing from secular growth. In a sufficiently long sample, the former will eventually turn down and reveal itself — but for most practical purposes, a trend would suffice. Similarly, an apparent structural break may be just another drawing from a rare-event distribution, such that the composite distribution is stationary: for example, the collapse of Bretton Woods might just be one more instance in the historical sequence of joining and leaving the gold standard, the ERM etc. As each occurs, temporary predictive failure results; after enough exemplars, their effects can be modelled in a constant unconditional distribution.

3.7 *Forecast Accuracy*

A final issue is that of selecting or rejecting models on the basis of their forecast performance. When the purpose is forecasting, then clearly forecast performance is all that matters (although that may not be easy to judge: see Clements and Hendry, 1993, and the related discussion). However, econometric models often have many purposes including testing theories and advising on alternative economic policies. Both because of the reasons enunciated in the previous subsection, and because forecasting devices can be "robustified" against some forms of predictive failure as we will now show, it is inadvisable to select models for policy using forecast performance alone.

Consider a closed, linear dynamic system after transformation to I(0) space in n variables x_t, written as a first-order vector autoregression:

$$x_t = \gamma + \Gamma x_{t-1} + \nu_t \quad \text{where} \quad \nu_t \sim \mathrm{IN}_n[0, \Omega_\nu]. \tag{7}$$

By assumption, x_t has the unconditional mean:

$$\mathrm{E}[x_t] = (\mathrm{I}_n - \Gamma)^{-1}\gamma = \psi. \tag{8}$$

The 1-step ahead forecasts at time T from (7) are:

$$\hat{x}_{T+1} = \hat{\gamma} + \hat{\Gamma} x_T \tag{9}$$

where '^'s on parameters denote estimates, and on random variables, forecasts. The forecast errors are:

$$\hat{\nu}_{T+1} = x_{T+1} - \hat{x}_{T+1}.$$

Between the estimation and forecast periods, $(\gamma : \Gamma)$ changes to $(\gamma^* : \Gamma^*)$. The data are now generated by:

$$x_{T+1} = \gamma^* + \Gamma^* x_T + \nu_{T+1}. \tag{10}$$

From (9) and (10), assuming correctly measured initial conditions x_T, the 1-step ahead forecast error is in fact:

$$\hat{\nu}_{T+1} = \gamma^* - \hat{\gamma} + \Gamma^* x_T - \hat{\Gamma} x_T + \nu_{T+1}, \tag{11}$$

or, for consistently-estimated parameter values:

$$\begin{aligned}\hat{v}_{T+1} = &(\gamma^* - \gamma) - (\hat{\gamma} - \gamma)\\ &+ (\Gamma^* - \Gamma)\,(x_T - \psi) - (\hat{\Gamma} - \Gamma)\,(x_T - \psi).\\ &+ (\Gamma^* - \Gamma)\psi - (\hat{\Gamma} - \Gamma)\psi + v_{T+1}\end{aligned} \tag{12}$$

Then, from (8), letting $\gamma^* = (I_n - \Gamma^*)\psi^*$, and assuming approximately unbiased estimates (which is surprisingly reasonable here, given the symmetric error distribution assumption, using the antithetic-variate argument in Hendry and Trivedi, 1972):

$$\begin{aligned}E\left[\hat{v}_{T+1}\right] &= (\gamma^* - \gamma) + (\Gamma^* - \Gamma)\psi\\ &= (I_n - \Gamma^*)\,(\psi^* - \psi).\end{aligned} \tag{13}$$

Thus, forecasts are biased only to the extent that the long-run mean shifts. Importantly, the bias is zero for mean-zero processes $(\gamma = \gamma^* = \psi = \psi^* = 0)$, or when shifts in γ^* offset those in Γ^* to leave ψ unaffected $(\psi^* = \psi)$. Alternatively expressed, shifts in the deterministic factors, either directly or as a consequence of changes in other parameters, are the main determinants of serious forecast errors. There are variance effects as well, and the *ex ante* forecast-error variance estimate will mis-estimate that ruling *ex post*, but these seem to be dominated by the mean shift at times of structural breaks (see Clements and Hendry, 1996).

When the parameter change occurs, forecasts will be incorrect from almost any statistical procedure. However, consider the following period. Persisting with forecasts from (9) will generate the same bias at T + 2 as in (13). However, robustness to regime shifts which bias forecasts can be obtained either by intercept corrections that carry forward the shift at time T + 1; or by suitable differencing to eliminate the changed intercept in later periods. Such devices can greatly improve forecast accuracy on bias measures, yet entail nothing about the usefulness for other purposes of the forecasting model. For example, at time T + 1 to forecast time T + 2, adding in the previous forecast error in (12) will on average produce unbiased forecasts (at a cost in forecast-error variance). Alternatively, simply first differencing produces the naîve forecast $\Delta\tilde{x}_{T+2} = 0$, or $\tilde{x}_{T+2} = x_{T+1}$ with mean forecast error:

$$\begin{aligned}E\left[\tilde{v}_{T+2}\right] &= E\left[x_{T+2} - \tilde{x}_{T+2}\right]\\ &= \gamma^* + \left(\Gamma^* - I_n\right)E\left[x_{T+1}\right]\\ &= \left(I_n - \Gamma^*\right)\psi^* + \left(\Gamma^* - I_n\right)\psi^* = 0.\end{aligned}$$

Although the resulting forecast is less biased than that from (9), such an outcome hardly sustains the choice of that model for policy, or as a baseline for future modelling exercises. This issue is discussed in more detail in Clements and Hendry (1996) and Hendry and Mizon (1996), and is illustrated below.

IV ECONOMIC THEORY AND PARAMETER CONSTANCY

There are major difficulties in determining a priori what aspects of a model will manifest constancy: i.e., what parameterisation will be constant. Potential contenders include: "tastes and technology", but there are important caveats about fads and fashions on the one hand, and innovation, discovery, R&D, technical progress and learning on the other (compare the attitude of, say, Robbins, 1932); behavioural patterns such as propensities (as Keynes, 1936), but here again there are similar caveats; and we may even consider a biological basis, but this too is subject to changes in the relative delivery costs of alternative sources of satisfaction from the same caveats as before.

At a more practical level, Koopmans (1937) argued against correlations being constants in economics, and in favour of partial derivatives, thereby setting the agenda for regression analysis. I believe this remains good advice, but in what transformations of the variables does one seek these constant derivatives? Levels; log levels; differences of logs (growth rates); or other non-linear functions; should one seek for propensities or elasticities, etc.? Are short-run or long-run effects more likely to be constant? Can clever (orthogonal) parameterisations help isolate the constancies and allow marginalisation with respect to non-constancies? Or are indirect constancies, as in (1)-(3), more likely than direct as in (4)?

Further, it can be nearly impossible to deduce an appropriate functional form that delivers all the theoretical and empirical requirements. For example, long-term nominal interest rates appear to be an integrated process, yet are non-negative with a standard deviation that seems roughly proportional to the level, and inversely related to the price of long-term bonds: this suggests working with logs, except that the cost of borrowing (or gain from lending) manifestly depends on the (non-logged) level.

Since non-linearities do not appear to be extreme in economics, I suspect that the precise transformation of the variables to deliver the correct functional form of derivatives is less important in practice than working with the correct partial: non-constancies induced by changing omitted variables seem more pernicious empirically than locally approximating a curve by a straight line. At first sight, countering the "omitted-variables" problem appears to require omniscience (to get the complete set of determining variables), but does so only if one is unwilling to contemplate a progressive research strategy (see the Keynes debate in Section II above). In any case, on this issue economic theory is at best on a par with empirical econometrics: if either theory or evidence suggests the required extension of the information set, then it is useful to follow it up; if either omits key variables, then its results will be tainted thereby.

V POSSIBLE CAUSES OF PREDICTIVE FAILURE

In-sample parameter constancy can offer no guarantee of out-of-sample constancy, but it is not ruled out either. Conversely, one cannot prove that the former is necessary for the latter. Possible causes of predictive failure of previously-constant models include the so-called Lucas (1976) critique: shifts in underlying equilibria (e.g., taste changes); major financial innovations; non-modelled variables changing; alterations to measurement systems; policy regime changes; and major catastrophes; as well as technical change and learning.

There have been many regime shifts historically: international examples include the formation and breakdown of the Bretton Woods agreement, and the two "oil crises" of the 1970s; in Europe, the creation of major trading blocks and the various monetary systems; in the USA, the experiment with monetary control based on the New Operating Procedures; and in the UK, nationalisation then privatisation, Competition and Credit Control regulations, the introduction of interest-bearing retail sight deposits, and substantive switches in fiscal policy. Some specific markets, such as housing, have been subject to almost interminable changes affecting taxation (the extent of mortgage interest deductibility and the rate of deduction, the introduction and later abolition of Schedule A tax; effects from changes in capital gains taxes on other assets; alterations, sometimes retrospective, to tenants' rights; leasehold reform; the abolition of the mortgage cartel in favour of competition; financial deregulation; from rates, to the poll tax, to the council tax; and so on). Thus, it cannot be a surprise that predictive failure is common, and that models grow over time to adapt to such regime changes.

This analysis also helps explain why models may fit better than they predict, since the relevant extension may be contingent on events that are not easy to predict a priori but are easy to incorporate later. While unhelpful for forecasting, an increasing complexity of empirical explanations over time is less problematic for modelling. Perhaps an attitude change is needed to predictive failure that is later resolved to re-create constant parameters: instead of condemning it as yet another example of the emptiness of empirical work, it should be carefully evaluated to see if it is a positive development in progressive research — where we have learned from our mistakes.

VI EMPIRICAL CONSTANCY IN UK MONEY DEMAND

To illustrate the preceding analysis empirically, we consider UK M1 quarterly data (seasonally adjusted), and study the impact of a major financial innovation. The model postulated by Hendry and Mizon (1993) was estimated over the sample 1963(3) to 1983(2). The forecasts therefrom failed badly when the data period was extended to 1989(2). Rebuilding their model over the whole sample confirms a failure of cointegration and does not yield any improvement in forecasting. However, the own interest rate (learning adjusted) was added by Hendry and Ericsson (1991) who thereby recovered the earlier model's parameter estimates, found cointegration again, and avoided predictive failure. Alternatively, adding a step-shift dummy to allow a separate intercept (autonomous growth) over the forecast period also essentially rescues the forecasts, similar to those from the "correct" model: this is a form of intercept correction (see Hendry and Clements, 1994).

The M1 demand model is from Hendry and Mizon (1993), fitted over T = 1963(4)-1983(2). Let m denote nominal M1, i total final expenditure, p its deflator, and R the Local-Authority three-month bill interest rate: lower case denotes logs, and Δ the first difference. They find that the regressors (i, Δp, R) are weakly exogenous for the parameters of a conditional equation for m – p, which can therefore be modelled in isolation. Their initial unrestricted linear dynamic equation had the following long-run solution (cointegrating vector):

$$\begin{aligned} \mathrm{Ecm} &= m - i - p - 0.26 + 6.7\Delta p + 7.1R \\ t_{ur} &= -7.0^{**} \end{aligned} \tag{14}$$

where t_{ur} strongly rejects a unit root in the dependent-variable lag polynomial (see Doornik and Hendry, 1994). Using (14) as the equilibrium-correction mechanism (Ecm), reduction of the unrestricted dynamic equation led to the more parsimonious model:

$$\begin{aligned} \Delta(m-p)_t &= -\underset{(0.06)}{0.29}\,\Delta(m-i-p)_{t-1} - \underset{(0.16)}{0.76}\,\Delta_2\Delta p_t \\ &\quad - \underset{(0.08)}{0.62}\,(\Delta R_t + \Delta R_{t-1}) - \underset{(0.009)}{0.094}\,\mathrm{Ecm}_{t-2} \\ R^2 &= 0.70 \quad \hat{\sigma} = 1.30\% \quad V = 0.25 \quad J = 0.77 \\ F_{ar}(5, 70) &= 1.3 \quad F_{arch}(4, 67) = 0.9 \quad \chi^2_{nd}(2) = 2.1 \quad F_{het}(8, 66) = 0.66 \end{aligned} \tag{15}$$

In (15), R^2 is the squared multiple correlation coefficient; $\hat{\sigma}$ is the standard deviation of the residuals (as a percentage of (m – p)), adjusted for degrees of freedom; and OLS standard errors are shown in parentheses. The diagnostic tests are of the form $F_j(k, T - N)$ which denotes an F-test against the alternative hypothesis j for: 5^{th}-order serial correlation (F_{ar}: see Godfrey, 1978), 4^{th}-order autoregressive conditional heteroscedasticity (F_{arch}: see Engle, 1982), heteroscedasticity (F_{het}: see White, 1980); and a chi-square test for normality (χ^2_{nd} (2): see Doornik and Hansen, 1994): * and ** denote significance at the 5 per cent and 1 per cent levels respectively. V and J are the variance-change and joint parameter-constancy tests from Hansen (1992). $\Delta x_t = x_t - x_{t-1}$ and $\Delta_2 x_t = x_t - x_{t-2}$.

Equation (15) satisfies all the reported diagnostic tests, with interpretable parameters in a parsimonious model, and as Figure 1 shows, the recursive estimates are constant. Reading from left to right, then top to bottom, the first four graphs show the parameter estimates recursively, with ±2 standard errors on either side; then the innovations and the recursive residuals with $0 \pm 2\hat{\sigma}_t$; and finally, 1-step, break-point, and forecast Chow (1960) statistics, scaled by their 1 per cent significance levels (shown as a straight line at unity). Despite using one-off significance levels, none of the tests exceeds its 1 per cent critical value anywhere in the sample.

Nevertheless, updating (15) to 1989(2) yields a Chow statistic of $F_c(24, 75) = 7.98^{**}$. This is massive predictive failure, as seen graphically in Figure 2 which shows the forecast statistics (fitted or forecast values with the actual outcomes; a cross plot of those; the residuals; and the forecasts with 95 per

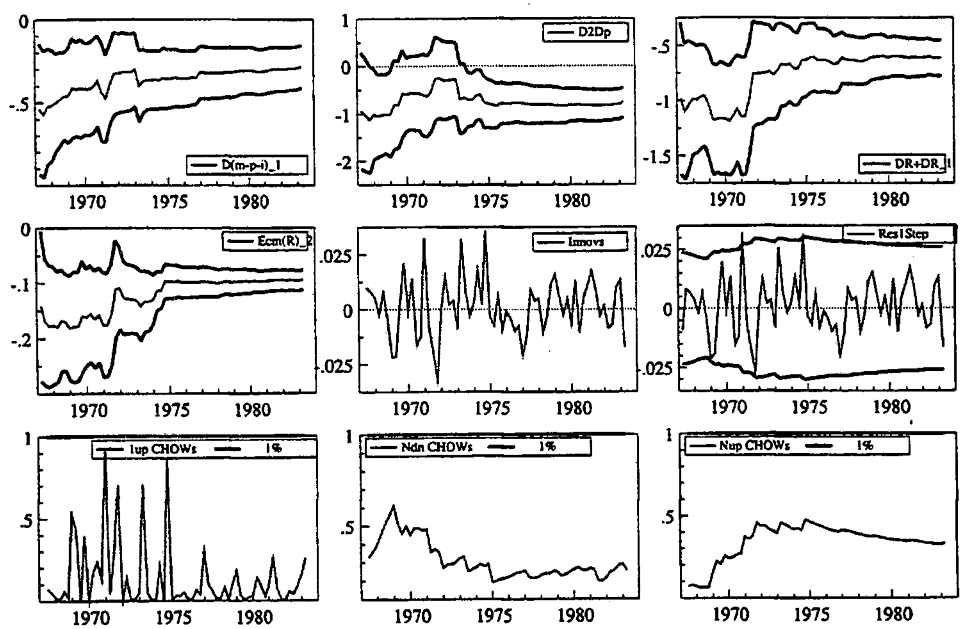

Figure 1: *Short-sample Recursive Estimates*

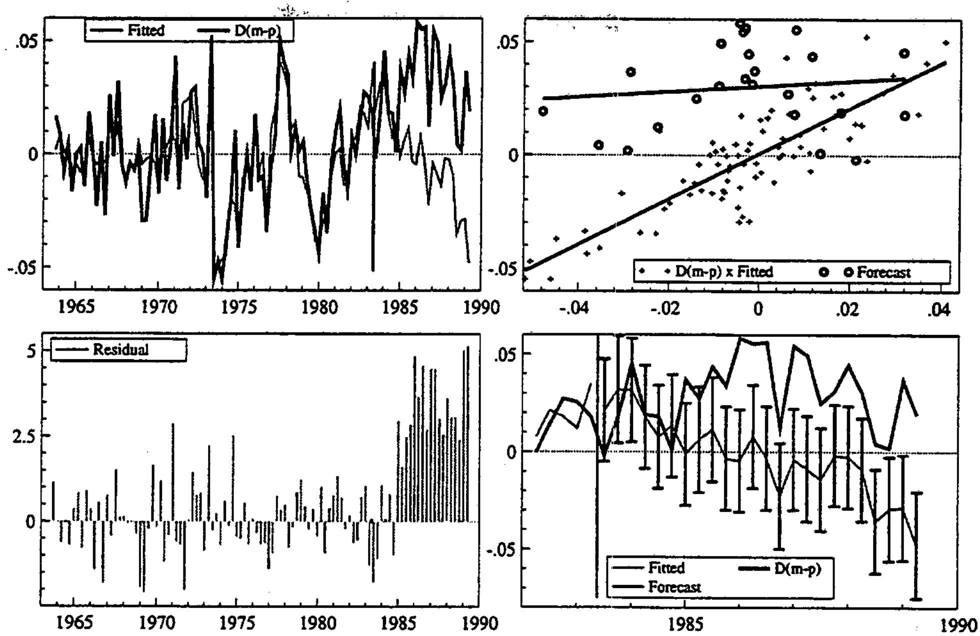

Figure 2: *Short-sample Estimation with Forecast Statistics*

cent confidence bars centred on the forecasts). The forecasts depart ever further from the outcomes as the horizon increases, the residuals are dramatically larger out of sample, and the 95 per cent confidence bars fail to include most of the realised values.

Such predictive failure is not simply earlier overfitting, since re-estimation over the extended sample T = 1963(4) – 1989(2) does not help:

$$
\begin{aligned}
\Delta(m-p)_t &= -\underset{(0.09)}{0.08}\,\Delta(m-i-p)_{t-1} - \underset{(0.23)}{0.73}\,\Delta_2\Delta p_t \\
&\quad - \underset{(0.12)}{0.41}\,(\Delta R_t + \Delta R_{t-1}) - \underset{(0.012)}{0.060}\,\mathrm{Ecm}_{t-2} \qquad (16)\\
R^2 &= 0.37 \quad \hat{\sigma} = 2.13\% \quad V = 1.73^{**} \quad J = 4.00^{**} \\
F_{ar}(5, 94) &= 14.20^{**} \quad F_{arch}(4, 91) = 11.48^{**} \quad \chi^2_{nd}(2) = 4.0 \quad F_{het}(8, 90) = 1.89
\end{aligned}
$$

The full-sample recursive estimates starkly confirm the predictive failure: the residual standard deviation has almost doubled, and all the instability tests reveal non-constancy. Figure 3 shows the changes over the full sample in the estimates, as well as the large increase in the standard deviation. As will be shown below, this predictive failure could not be predicted by any in-sample statistical test, and does not reveal a failure of methodology, nor a failure of rigorous testing. Rather, it highlights that equilibrium-correction mechanisms are precisely that, namely they correct to a specific equilibrium, here defined by (14). Should that equilibrium shift for any reason, the model will still correct to (14), and hence will make increasingly large forecast errors after a permanent shift to try and "drag" the model back to the old equilibrium. Thus, between equilibria, they do not error correct, so the original name in Davidson, Hendry, Srba and Yeo (1978) was a misnomer.

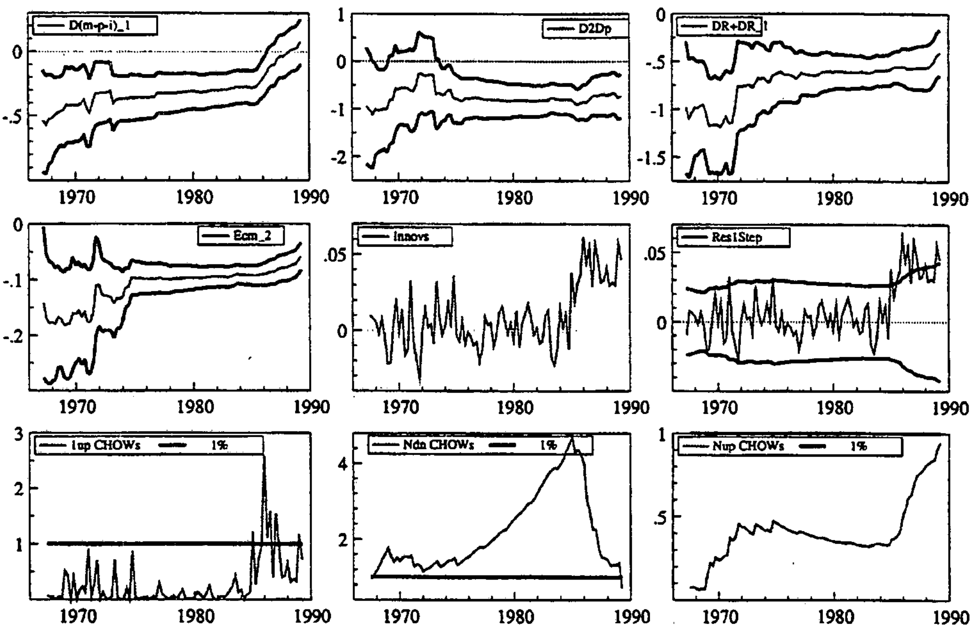

Figure 3: *Long-sample Recursive Estimates*

In fact, returning to the unrestricted model highlights the disintegration of cointegration: the long-run outcome is badly determined with uninterpretable coefficient magnitudes, and the unit-root t-test does not reject the null of no cointegration. In terms of the preceding analysis, there is a clear structural break. Such problems often seem to occur in empirical research, and may appear to cast doubt on the modelling strategy advocated in e.g., Hendry (1995a). The interesting issue is whether there are additional variables that re-create a constant-parameter money-demand equation in which the original parameters are closely reproduced. Given the full-sample estimates, that does not seem possible within the present information set of linear functions of m – p, i, R, and Δp.

The solution proposed by Hendry and Ericsson (1991) is to add a measure of the own interest rate on M1, a variable that was zero until 1984 (other than implicit interest payments used to offset transactions costs, for which commercial banks did not charge), when a change in the law allowed interest on retail sight deposits (interest-bearing checking accounts). This induced a large change in the opportunity cost of holding M1. Let $R_{o,t}$ denote the (learning-adjusted) own rate on M1 used in Hendry and Ericsson (1991), and shown in the first block of Figure 4 with the real money stock $m - p$. The sharp rise in $m - p$ coincides with the jump from zero in R_o, as is necessary for constancy to occur in the five-dimensional system $m - p$, i, R, R_o and Δp. The opportunity cost of holding M1 after 1984 is measured by the net rate, $R_{n,t} = R_t - R_{o,t}$. Figure 4 also shows the two interest rates R_t and $R_{n,t}$, the corresponding disequilibrium measures, and the changes in the interest rates in the "competing" models.

Re-estimation on replacing R_t by $R_{n,t}$ over T = 1963(4) – 1989(2) delivers:

$$\begin{aligned}
\Delta(m-p)_t &= -\underset{(0.06)}{0.27}\ \Delta(m-i-p)_{t-1} - \underset{(0.14)}{0.83}\ \Delta_2\Delta p_t \\
&\quad - \underset{(0.07)}{0.59}\ (\Delta R_{n,t} + \Delta R_{n,t-1}) - \underset{(0.006)}{0.093}\ \mathrm{Ecm}_{t-2} \qquad (17) \\
R^2 &= 0.77 \quad \hat{\sigma} = 1.28\% \quad V = 0.22 \quad J = 0.68 \\
F_{ar}(5, 94) &= 1.81 \quad F_{arch}(4, 91) = 0.81 \quad \chi^2_{nd}(2) = 1.2 \quad F_{het}(8, 90) = 0.87
\end{aligned}$$

The final parameter estimates reported in Equation (17) impose the restriction that the outside and own rates have equal-magnitude, opposite-sign effects, so only their net differential $R_{n,t}$ affects money demand ($R_{n,t-2}$ is used in Ecm_{t-2}). The estimates and fit are very similar to those in (15), and no diagnostic test is significant. Figure 5 shows the forecast statistics for the extended model estimated on the short sample and forecasting the previously difficult period. As can be seen, the forecast failure has been removed, confirming constancy in the extended set of variables. The test of parameter constancy over 1983(3)-1989(2) yields $F_c(24, 75) = 0.89$ revealing no breakdown in the extended model. The close similarity of new and old parameter estimates suggests that the constancy is not spurious.

Several morals follow from this illustration of the earlier analysis. First, it is an example of "extended constancy" in a relation among the variables m, i, p, R_t, $R_{o,t}$: when $R_{o,t}$ is used as an unrestricted regressor, the model is enlarged, but the crucial index of its constancy is that all the previous parameters retain their original values. This is an essential attribute of a constant model. Second, updating models requires the use of sensible measurements which adapt to changing environments: retaining R_t is a bad proxy for opportunity cost post 1984. Third, as noted, there is no possible within-sample test of later behaviour: whether or not predictive failure is

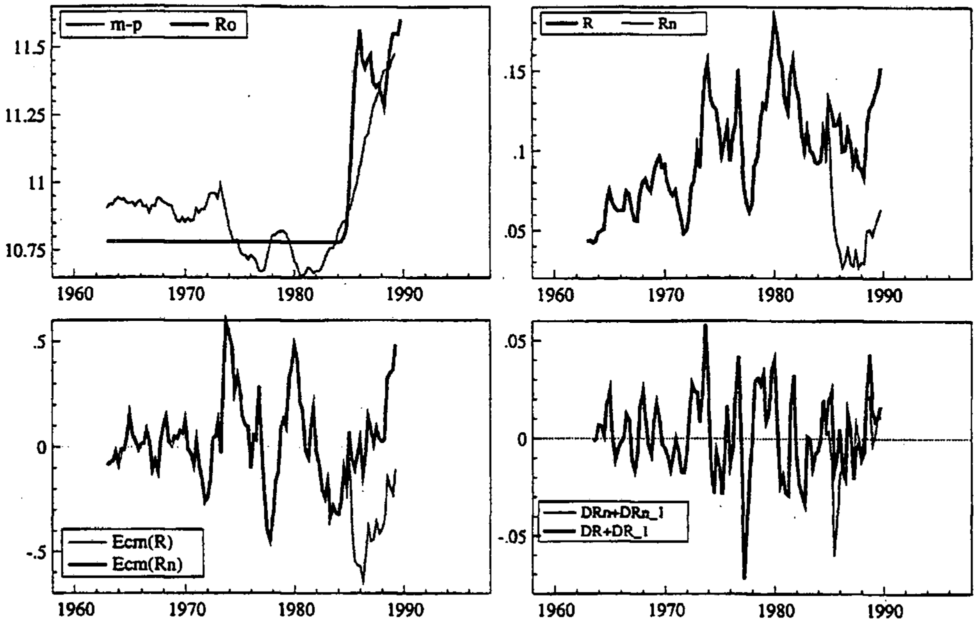

Figure 4: *Time-series of Interest Rates, Real Money, and Disequilibria*

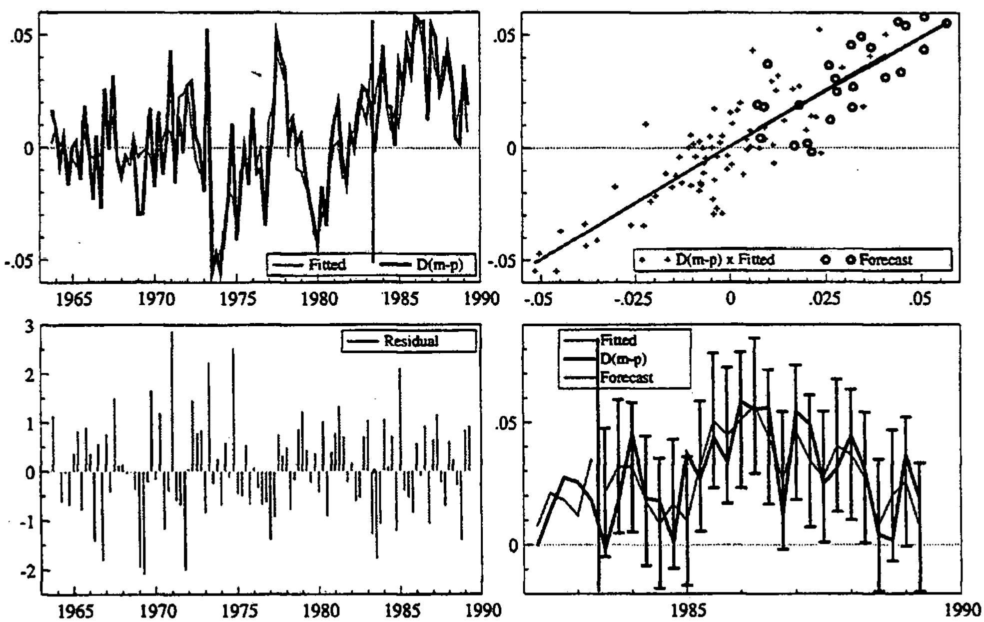

Figure 5: *Estimation of Extended Model with Forecast Statistics*

manifested depends on how the model is updated, not on the in-sample behaviour. Next, both evidence and economic theory played crucial roles in re-establishing constancy, the latter "predicting" the same magnitude, opposite sign effect that allowed the creation of the net interest rate. Further, the decision to alter the law concerning the payment of interest on checking accounts was not readily predictable in advance, so the empirical model sequence fits better than it forecasts.

Finally, to illustrate the analysis of forecast robustness in Section 3.7, if the crucial shift is that in the equilibrium mean of the original model, then a step-shift dummy should remove the predictive failure, akin to the intercept correction (see Hendry and Clements, 1994). There is a problem of how to date the break point at which the dummy commences (i.e., when predictive failure first becomes noticeable), but empirical experimentation suggests 1985(1), leading to an indicator variable that is unity thereafter. The dummy mimics the effect of R_0 remarkably well, inducing almost the same residual variance, and long-run outcome. Indeed, despite using only three non-zero in-sample values for estimation, respectable forecasts result: Figure 6 records the estimation and forecast statistics for the original model with a shift dummy, but without $R_{0,t}$, in terms of the level of m_t.

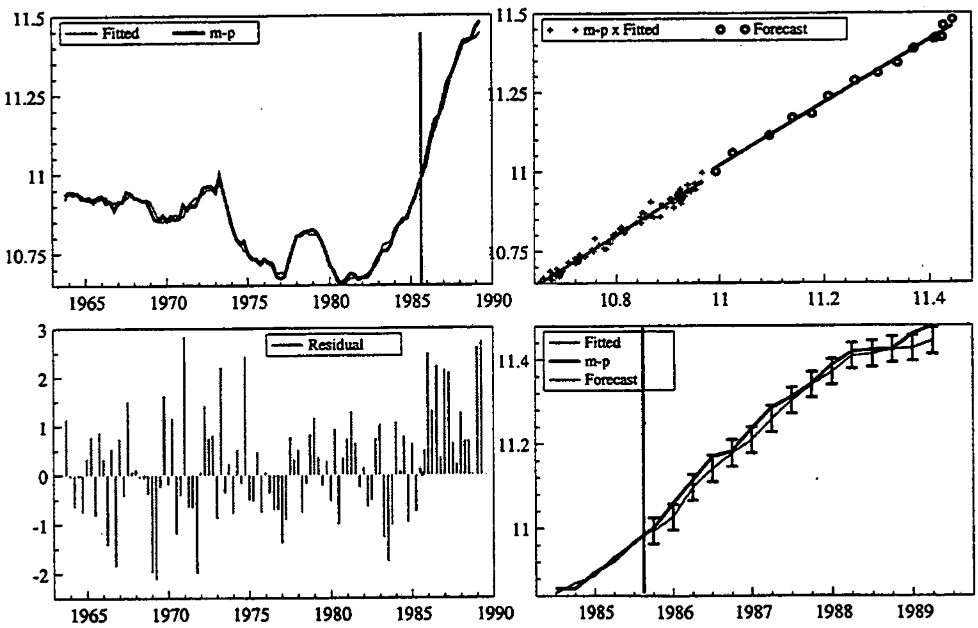

Figure 6: *Forecast Statistics for the Model with a Shift Dummy*

Similarly, the first-difference projection $\Delta(m-p)_{t+1} = \Delta(m-p)_t$ yields a Chow test value of $F_c(24, 79) = 0.66$, so differencing also removes the equilibrium mean shift, but does not thereby justify that model beyond its forecast-error robustness.

VII CONCLUSION

The constancy of a model is a more subtle concept than might have been expected, and was explored above. A number of implications of the definition seem surprising, particularly that model expansion need not be inconsistent with constancy. Constancy therefore depends on precisely how an empirical model is interpreted and updated. *Ex ante* constancy could be construed as the necessary ingredient for adequate forecasts, whereas *ex post* constancy of the same model indicates a progression in understanding. In the empirical example, the interest rate regressor was interpreted as the measure of the opportunity cost of holding idle money for transactions purposes (see e.g., Hendry, 1985, p. 81), so the extension using R_t rather than $R_{n,t}$ was the inappropriate one, albeit the first to suggest that misprediction was occurring.

It then followed that in-sample tests could not reveal the likely predictive failure of a model on a later sample. An analogy might be a spacecraft to a distant planet being exactly on course and forecast to land successfully, just before being destroyed by a meteorite. Also note that the resulting predictive failure in such a case hardly refutes the underlying physical theories.

Further, we showed that forecast accuracy is not the ultimate test of a model, so forecast dominance is not necessarily a reflection of usefulness for other purposes. This result was established theoretically for changes in deterministic terms, and empirically for whatever caused the initial specification of UK M1 not to forecast.

Thus, despite the existence of many empirically unsuccessful models, as argued in Hendry (1995b), the development of congruent, theory-consistent and constant models remains a viable route in econometrics, even with evolving data. Further, since progressive research can discover structure in part without prior knowledge of the whole, Keynes's worry about the need for knowledge of the complete specification in advance of empirical research is misplaced: we can learn from our mistakes.

REFERENCES

ALDRICH, J., 1989. "Autonomy", *Oxford Economic Papers*, Vol. 41, pp. 15-34.

CHOW, G.C., 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, Vol. 28, pp. 591-605.

CLEMENTS, M.P., and D.F. HENDRY, 1993. "On the Limitations of Comparing Mean

Squared Forecast Errors", *Journal of Forecasting*, Vol. 12, pp. 617-637 (With discussion).

CLEMENTS, M.P., and D.F. HENDRY, 1996. "Intercept Corrections and Structural Breaks", *Journal of Applied Econometrics*, Vol. II, pp. 475-494.

DAVIDSON, J.E.H., D.F. HENDRY, F. SRBA, and S. YEO, 1978. "Econometric Modelling of the Aggregate Time-series Relationship Between Consumers' Expenditure and Income in the United Kingdom", *Economic Journal*, Vol. 88, pp. 661-692. Reprinted in D.F. HENDRY, 1993, *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

DOORNIK, J.A., and H. HANSEN, 1994. A Practical Test for Univariate and Multivariate Normality, Discussion paper, Nuffield College.

DOORNIK, J.A., and D.F. HENDRY, 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*, Vol. 50, pp. 987-1,007.

EPSTEIN, R., 1987. *A History of Econometrics*, Amsterdam: North-Holland.

FAVERO, C., and D.F. HENDRY, 1992. "Testing the Lucas Critique: A Review", *Econometric Reviews*, Vol. 11, pp. 265-306.

FRISCH, R., 1938. Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum "Autonomy of Economic Relations", 6 November 1948, Oslo, Universitets Økonomiske Institutt. Reprinted in D.F. HENDRY and M.S. MORGAN, 1995, *The Foundations of Econometric Analysis*, Cambridge: Cambridge University Press.

GODFREY, L.G., 1978. "Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables", *Econometrica*, Vol. 46, pp. 1,303-1,313.

HAAVELMO, T., 1944. "The Probability Approach in Econometrics", *Econometrica*, Vol. 12, pp. 1-118. Supplement.

HANSEN, B.E., 1992. "Testing for Parameter Instability in Linear Models", *Journal of Policy Modelling*, Vol. 14, pp. 517-533.

HARVEY, A.C., 1981. *Time Series Models*, London: Philip Allan.

HARVEY, A.C., and N. SHEPHARD, 1992. "Structural Time Series Models", in G.S. Maddala, C.R. Rao, and H.D. Vinod (eds.), *Handbook of Statistics*, Vol. 11, Amsterdam: North-Holland.

HENDRY, D.F., 1979. "Predictive Failure and Econometric Modelling in Macro-Economics: The Transactions Demand for Money", in P. Ormerod (ed.), *Economic Modelling*, pp. 217-242, London: Heinemann. Reprinted in D.F. HENDRY, 1993, *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

HENDRY, D.F., 1985. "Monetary Economic Myth and Econometric Reality", *Oxford Review of Economic Policy*, Vol. 1, pp. 72-84. Reprinted in D.F. HENDRY, 1993, *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

HENDRY, D.F., 1995a. *Dynamic Econometrics*, Oxford: Oxford University Press.

HENDRY, D.F., 1995b. "Econometrics and Business Cycle Empirics", *Economic Journal*, Vol. 105, pp. 1,622-1,636.

HENDRY, D.F., and M.P. CLEMENTS, 1994. "On a Theory of Intercept Corrections in Macro-economic Forecasting", in S. Holly, (ed.), *Money, Inflation and Employment: Essays in Honour of James Ball*, pp. 160-182, Aldershot: Edward Elgar.

HENDRY, D.F., and N.R. ERICSSON, 1991. "Modelling the Demand for Narrow Money in the United Kingdom and the United States", *European Economic Review*, Vol. 35, pp. 833-886.

HENDRY, D.F., and G.E. MIZON, 1993. "Evaluating Dynamic Econometric Models by Encompassing the VAR", in P.C.B. Phillips (ed.), *Models, Methods and Applications of Econometrics*, pp. 272-300, Oxford: Basil Blackwell.

HENDRY, D.F., and G.E. MIZON, 1996. "Selecting Econometric Models for Policy Analysis by Forecast Accuracy", Mimeo, Nuffield College, University of Oxford.

HENDRY, D.F., and M.S. MORGAN, 1995. *The Foundations of Econometric Analysis*, Cambridge.

HENDRY, D.F., and P.K. TRIVEDI, 1972. "Maximum Likelihood Estimation of Difference Equations with Moving-average Errors: A Simulation Study", *Review of Economic Studies*, Vol. 32, pp. 117-145.

KEYNES, J.M., 1936. *The General Theory of Employment, Interest and Money*, London: Macmillan.

KEYNES, J.M., 1939. "Professor Tinbergen's Method", *Economic Journal*, Vol. 44, pp. 558-568.

KOOPMANS, T.C., 1937. *Linear Regression Analysis of Economic Time Series*, Haarlem: Netherlands Economic Institute.

LEAMER, E.E., 1978. *Specification Searches. Ad-Hoc Inference with Non-Experimental Data*, New York: John Wiley.

LUCAS, R.E., 1976. "Econometric Policy Evaluation: A Critique", in K. Brunner and A. Meltzer (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19-46, Amsterdam: North-Holland Publishing Company.

MOORE, H.L., 1914. *Economic Cycles — Their Law and Cause*, New York: Macmillan.

MORGAN, M.S., 1990. *The History of Econometric Ideas*, Cambridge: Cambridge University Press.

PERSONS, W.M., 1924. *The Problem of Business Forecasting*, No. 6 in Pollak Foundation for Economic Research Publications, London: Pitman.

QIN, D., 1989. "Formalisation of Identification Theory", *Oxford Economic Papers*, pp. 73-79.

QIN, D., 1993. *The Formation of Econometrics: A Historical Perspective*, Oxford: Clarendon Press.

ROBBINS, L., 1932. *An Essay on the Nature and Significance of Economic Science*, London: Macmillan.

SCHULTZ, H., 1928. *The Theory and Measurement of Demand*, University of Chicago Press.

TINBERGEN, J., 1940. *Statistical Testing of Business-Cycle Theories*, Geneva: League of Nations, Vol. I: *A Method and Its Application to Investment Activity.*

WHITE, H., 1980. "A Heteroskedastic-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, Vol. 48, pp. 817-838.