Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration

Neil R. Ericsson, *Federal Reserve Board*

Parameter constancy and a model's mean square forecast error are two commonly used measures of forecast performance. By explicit consideration of the information sets involved, this paper clarifies the roles that each plays in analyzing a model's forecast accuracy. Both criteria are necessary for "good" forecast performance, but neither (nor both) is sufficient. Further, these criteria fit into a general taxonomy of model evaluation statistics, and the information set corresponding to a model's mean square forecast error leads to a new test statistic, forecast-model encompassing. Two models of U.K. money demand illustrate the various measures of forecast accuracy.

1. INTRODUCTION

Parameter constancy and the mean square forecast error (MSFE) are two commonly used measures of the forecast performance of empirical macromodels. Parameter constancy has long been viewed as a desirable economic and statistical property, and it is closely linked to the issue of predictive failure (Chow, 1960; Hendry, 1979). Further, parameter constancy can imply super exogeneity, which is necessary to sustain

---

Address correspondence to Neil R. Ericsson, Stop 24, Federal Reserve Board, 2000 C Street, N.W., Washington, DC 20551.

The author is a staff economist in the Division of International Finance, Federal Reserve Board. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting those of the Board of Governors of the Federal Reserve System or other members of its staff.

I am grateful to Julia Campos, Frank Diebold, Hali Edison, David Hendry, David Howard, Ross Levine, Jaime Marquez, Doug McManus, and Garry Schinasi for useful comments and discussions, and to Hong-Anh Tran for invaluable research assistance. I am particularly indebted to Edison, Schinasi, and P.A.V.B. Swamy, whose empirical papers on modeling and forecasting exchange rates movitated this paper; see Edison (1985), Edison and Klovland (1987), Edison (1991), Schinasi and Swamy (1989), and Swamy and Schinasi (1989). All numerical results were obtained using PC-GIVE Version 6.01 (Hendry, 1989).

Received March 1991; final draft accepted October 1991.

Journal of Policy Modeling 14(4):465–495 (1992)
 0161-8938/92/$5.00

counterfactual policy simulations of an econometric model (Hendry, 1988). Lack of parameter constancy can induce apparent unit roots, posing potential difficulties when testing for cointegration (Hendry and Neale, 1991). The MSFE is a common criterion for evaluating the performance of alternative macromodels; see Fair (1986) for a general discussion and Meese and Rogoff (1983) for a classic example with models of the exchange rate.

Sometimes, the literature has viewed these two forecast criteria as competing rather than complementary. Thus, this paper aims to clarify the roles of parameter constancy and the MSFE in evaluating the forecast accuracy of a model.

Section 2 works through some simple examples to show that (i) parameter constancy is neither necessary nor sufficient for minimizing MSFE across a given set of models, and (ii) both criteria together are necessary but not sufficient to obtain the best forecasting model, even on only the data available from the given set of models. Section 3 explains why, showing that parameter constancy and minimizing MSFE are criteria that evaluate a given model (respectively) against that model's own data and against other models' data. Both are reasonable criteria, but other criteria are also important for determining the forecast adequacy of an empirical model. Section 4 introduces a new model evaluation criterion, "forecast-model" encompassing, and the corresponding test statistic. Further, Section 4 shows that minimum MSFE, forecast encompassing, and forecast-model encompassing parallel variance dominance, variance encompassing, and parameter encompassing, respectively. Section 5 discusses several implications for forecasting integrated and cointegrated variables. Section 6 comments briefly on the role of time-varying coefficient models in forecasting. Section 7 illustrates the various forecast-based criteria with an application to two models of narrow money demand in the United Kingdom.

Before turning to the heart of the paper, three remarks may be helpful. First, the results in this paper are quite general. To illustrate the central concepts, however, simple, static, Gaussian models are used as examples throughout. See Hendry and Richard (1982) for a framework in which more general results may be obtained. Second, to abstract from sampling issues, results are often presented as "asymptotic." This in no way invalidates the results, but simply permits a clearer exposition. Third, the concept of an "adequate forecasting model" is intentionally left vague. Roughly, such a model efficiently uses the information available for creating forecasts. It is defined in part by its negation. For instance, a model is *in*adequate for forecasting if its forecast errors are predictable, a situation including both parameter nonconstancy and lack of forecast encompassing (as will be seen below). For Gaussian processes, minimum MSFE is a condition for forecast adequacy, but Section 2 shows that it is not sufficient because the corresponding errors may still be predictable.

2. PARAMETER CONSTANCY AND MINIMIZING MSFE

This section shows that (i) parameter constancy is neither necessary nor sufficient for minimizing MSFE across a given set of models, and (ii) both criteria together are necessary (but not sufficient) to obtain the best forecasting model on the data available. Four "propositions" establish (i) and (ii), which are illustrated by some simple examples. The analytical form of the MSFE and of Chow's (1960) "prediction interval" statistic clarifies the absence of relationship between minimizing MSFE over a set of models and obtaining parameter constancy for an individual model.

To show the *lack* of connection between parameter constancy and minimizing MSFE, consider the following simple process in which the dependent variable y_t is linearly dependent upon three regressors, each of which is normally and independently distributed.

Models and the data generation process. Suppose that the data (y_t, $t = 1, \ldots, T + n$) are generated by

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t \qquad \epsilon_t \sim \text{NID}(0,\sigma^2), \tag{1a}$$

and the x_{it}'s are normally and independently distributed (NID):

$$x_{it} \sim \text{NID}(0,\omega_{ii}) \qquad i = 1, 2, 3, \tag{1b}$$

where ω_{ii}, the variance of x_{it}, may change over time. The double index on ω_{ii} denotes that the variance is the ith diagonal element from the (diagonal) contemporaneous covariance matrix (Ω) of the $\{x_{it}\}$. Together equations 1a and 1b are referred to as the data generation process (DGP). To exclude trivial cases, σ^2 and all ω_{ii}'s are positive and all β_i's are nonzero.

The econometrician does not know the DGP, and estimates the following (misspecified) models by OLS over a subsample $[1,T]$ and evaluates their forecast performance over n periods $[T + 1, T + n]$.

$$M_1\colon\ y_t = \alpha_1 x_{1t} + u_{1t} \qquad u_{1t} \sim \text{NID}(0,\sigma_1^2) \tag{2a}$$

$$M_2\colon\ y_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + u_{2t} \qquad u_{2t} \sim \text{NID}(0,\sigma_2^2) \tag{2b}$$

$$M_3\colon\ y_t = \delta_1 x_{1t} + \delta_3 x_{3t} + u_{3t} \qquad u_{3t} \sim \text{NID}(0,\sigma_3^2) \tag{2c}$$

The sets of coefficients $\{\alpha_1\}$, $\{\gamma_1,\gamma_2\}$, and $\{\delta_1,\delta_3\}$ are used to distinguish the models' coefficients from the underlying coefficients of the DGP, i.e., $\{\beta_1,\beta_2,\beta_3\}$. For convenience, $\hat{y}_{ij}$ denotes the prediction of y in period j, using the parameter estimates from model M_i estimated over $[1,T]$. For example, $\hat{y}_{2j}$ is

$$\hat{y}_{2j} = \hat{\gamma}_{1T}x_{1j} + \hat{\gamma}_{2T}x_{2j} \qquad j = T+1,\ldots, T+n, \tag{3}$$

where $\hat{\gamma}_{1T}$ and $\hat{\gamma}_{2T}$ are the coefficients γ_1 and γ_2, estimated over $[1,T]$.

The MSFE for the ith model is

$$MSFE_i = \mathscr{E}\left[\sum_{j=T+1}^{T+n} (y_j - \hat{y}_{ij})^2/n\right] \qquad i = 1, 2, 3, \tag{4a}$$

where the expectation $\mathscr{E}[\cdot]$ is over $\{\epsilon_j\}$. For the models discussed in this paper, each term in the summation in Equation 4a has the same expectation, so $MSFE_i = \mathscr{E}[(y_j - \hat{y}_{ij})^2]$, independent of j, for $j = T+1, \ldots, T+n$.

In practice, the MSFE is estimated by the sample average of the squared forecast errors:

$$\widehat{MSFE}_i = \sum_{j=T+1}^{T+n} (y_j - \hat{y}_{ij})^2/n \qquad i = 1, 2, 3, \tag{4b}$$

for the ith model. Most of the discussion in this paper is in terms of the underlying population moments, i.e., the MSFE, thereby abstracting from the additional complication of the sampling distribution of Equation 4b.[1]

Parameter constancy may be evaluated by any of a number of statistics, with Chow's (1960, pp. 594–595) "prediction interval" statistic being one of the more common.[2] The Chow statistic can be written as:

---

[1]Note that the estimation and forecast periods do not overlap. By contrast, e.g., dynamic simulation uses overlapping (usually identical) estimation and "forecast" periods. Mean square forecast errors from such simulations may have quite different properties from those discussed herein. See Hendry and Richard (1982), Chong and Hendry (1986), and Pagan (1989) on the role of dynamic simulation in model comparison.

[2]Chow (1960) also discusses a parameter constancy test statistic based on the analysis of covariance, in which estimates of the coefficients over the two subsamples are compared for equality; see Fisher (1922) for its original development. This statistic is distributed as $F(k_i,T+n-2k_i)$ under the null hypothesis, with classical assumptions about the regressors and disturbances. This covariance test statistic is sometimes (and confusingly) referred to as the "Chow statistic" although Chow (1960, p. 592) was well aware of its presence in the literature. In the

$$CHOW_i(n,T-k_i) = \{(Y_{T+1}^{T+n} - \hat{Y}_{T+1}^{T+n})'[\mathrm{Var}(Y_{T+1}^{T+n} - \hat{Y}_{T+1}^{T+n})]^{-1}(Y_{T+1}^{T+n} - \hat{Y}_{T+1}^{T+n})/n\}/\{\hat{\sigma}_{iT}^2/\sigma_{iT}^2\}$$
$$= \left\{\sum_{j=T+1}^{T+n} (y_j - \hat{y}_{ij})^2/n\right\}/\hat{\sigma}_{iT}^2 + O_p(T^{-1})$$
$$= \widehat{MSFE}_i/\hat{\sigma}_{iT}^2 + O_p(T^{-1}), \tag{5}$$

where $Y_{T+1}^{T+n} = (y_{T+1} \ldots y_{T+n})'$, $\hat{Y}_{T+1}^{T+n}$ is the forecast of Y_{T+1}^{T+n} by model M_i, $\hat{\sigma}_{iT}^2$ is the estimated equation error variance for model M_i over $[1,T]$, k_i is the number of regressors in model M_i, and $O_p(T^{-1})$ denotes a term of order T^{-1} in probability. As Equation 5 clarifies, the Chow statistic in effect tests whether each of the forecast errors of a given model has a zero mean, i.e., $\mathscr{E}(y_j - \hat{y}_{ij}) = 0$ for $j = T + 1, \ldots, T + n$. It does so by comparing the mean square forecast error against the estimated error variance over the estimation subsample. Under the null hypothesis, and with fixed regressors and normal disturbances, the Chow statistic is distributed as $F(n,T - k_i)$. ''Significant'' Chow statistics are often referred to as ''predictive failure'' (Hendry, 1979).[3]

To simplify the analysis even further, suppose that T is large enough so the uncertainty in estimating the model parameters can be ignored when considering the characteristics of the MSFE and the Chow statistic. This assumption and virtually all assumptions in Equations 1 and 2 are for expositional purposes only, and most of the results below obtain under more general conditions (e.g., nonlinearity; autocorrelated, multicollinear, endogenous regressors; more or fewer regressors relative to those in the models here; non-normality of the errors and/or regressors).

Examples 1 and 2 consider two situations, one in which all population data moments are constant and the other in which some of them change over time.

Example 1: constant population data moments. This is equivalent to having $(\omega_{ii}, \beta_i, i = 1, 2, 3)$ and σ^2 constant in Equation 1.

---

current paper, the phrase ''Chow statistic'' refers exclusively to Chow's prediction interval statistic (Equation 5). Wilson (1978) discusses conditions under which each of the Chow (prediction interval) statistic and the covariance statistic is uniformly most powerful. Fisher (1970) and Dufour (1980) present intuitive derivations of the two statistics.

[3]The first term on the last line of Equation 5 is $1/n$ times Hendry's (1979) χ^2 statistic for testing the *numerical* accuracy of the forecasts. Chow's statistic tests their *statistical* accuracy by accounting for the uncertainty arising from estimating (rather than knowing) the regression coefficients. This affects only the finite sample distribution of the statistic: Hendry's and Chow's statistics *are* asymptotically equivalent. Because coefficient uncertainty is ignored for the most part in this paper, the equivalency proves useful, given the simpler form of Hendry's statistic.

All models (i.e., M_1, M_2, M_3) will have constant parameters because the corresponding (OLS) estimators are functions of the sample data moments, with the sample data moments being constant in expectation (by assumption). For the DGP in Equation 1, OLS for each model in Equation 2 is unbiased for the relevant subset of $\{\beta_1, \beta_2, \beta_3\}$, and is so only because the regressors are uncorrelated with each other and are static. That property does *not* generalize; however, even with correlated regressors, for example, constant population data moments are sufficient for parameter constancy.

From Equations 1 and 2, it follows directly that the mean square forecast errors for M_1, M_2, and M_3 are

$$MSFE_1 = \sigma^2 + \beta_2^2\omega_{22} + \beta_3^2\omega_{33} \tag{6a}$$
$$MSFE_2 = \sigma^2 + \beta_3^2\omega_{33} \tag{6b}$$
$$MSFE_3 = \sigma^2 + \beta_2^2\omega_{22} \ . \tag{6c}$$

Clearly, M_1 has the largest MSFE; the ranking of M_2 and M_3 depends upon the relative magnitudes of $\beta_2^2\omega_{22}$ and $\beta_3^2\omega_{33}$. This indeterminacy leads to the first proposition.

Proposition 1. *If a model has (empirically) constant parameters, it can have either a smaller or a larger MSFE than some other model.*

That is, parameter constancy is not sufficient for obtaining the smallest MSFE among a set of models.

Nonconstant population data moments help demonstrate the lack of necessity.

Example 2: nonconstant population data moments. Suppose that the variance of x_{2t} increases from ω_{22} to ω_{22}^* at time $T + 1$ and remains at ω_{22}^* thereafter.

For models M_1 and M_3, the increase from ω_{22} to ω_{22}^* implies a forecast error variance larger than the estimation subsample error variance, so the Chow statistic will indicate parameter nonconstancy. If regressors are correlated, either or both models may have coefficient nonconstancy, apparent, for example, through graphs of the recursively estimated coefficients.

The mean square forecast errors for the models are

$$MSFE_1 = \sigma^2 + \beta_2^2\omega_{22}^* + \beta_3^2\omega_{33} \tag{7a}$$
$$MSFE_2 = \sigma^2 + \beta_3^2\omega_{33} \tag{7b}$$
$$MSFE_3 = \sigma^2 + \beta_2^2\omega_{22}^* . \tag{7c}$$

Again, M_1 has the largest MSFE, but the ranking of those for M_2 and M_3 could be the same as (or different from) the ranking of Equation 6, depending upon ω_{22}^*. Further, whether or not a model exhibits parameter nonconstancy has little to do with its ranking by MSFE. For instance, M_3 can have a smaller *or* larger MSFE than M_2, depending upon the values of $\beta_2^2\omega_{22}^*$ and $\beta_3^2\omega_{33}$, but M_3 exhibits parameter nonconstancy whereas M_2 does not. This indeterminacy implies another proposition.

Proposition 2. *If a model has (empirically)* non*constant parameters, it can have either a smaller or larger MSFE than some other model.*

That is, parameter constancy is not necessary for obtaining the smallest MSFE among a set of models.

Whether the parameters of the "other" model are constant or not makes no difference to either Proposition 1 or Proposition 2, and this provides a different view on the lack of necessity and sufficiency.

Proposition 3. *For both Propositions 1 and 2, the constancy or otherwise of the "other" model is immaterial.*

For instance, consider a fourth model:

$$M_4\colon y_t = \tau_2 x_{2t} + u_{4t} \qquad u_{4t} \sim \text{NID}(0,\sigma_4^2), \tag{8a}$$

which has MSFE:

$$MSFE_4 = \sigma^2 + \beta_1^2\omega_{11} + \beta_3^2\omega_{33}. \tag{8b}$$

Model M_4 has constant parameters, but its MSFE may be smaller or larger than that for M_1 (which has nonconstant parameters), depending upon the relative variances of the regressors. Hence, parameter constancy is neither necessary nor sufficient for minimizing MSFE across a given set of models. Table 1 summarizes the properties of these models; Figure 1 provides a schematic of their relationships in terms of MSFE.

The ranking of models by MSFE can change across subsamples as well. For instance, the ranking of the models above depends upon the variances of the regressors, and those need not remain constant over subsamples. Unless a model is well specified in a very general sense, i.e., "congruent" with the evidence (see, for example, Hendry, 1987, Campos and Ericsson, 1988, and White, 1989), there is no guarantee whatsoever that an observed ranking in mean square forecast error will obtain over different sample periods.

Finally, consider the relationship between the properties of param-

Table 1: Models for the Discussion of Mean Square Forecast Error (MSFE) and Parameter Constancy

Model	Equation	$MSFE_i$	Constancy[a]
M_1	$y_t = \alpha_1 x_{1t} + u_{1t}$	$\sigma^2 + \beta_2^2\omega_{22}^* + \beta_3^2\omega_{33}$	No
M_2	$y_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + u_{2t}$	$\sigma^2 + \beta_3^2\omega_{33}$	Yes
M_3	$y_t = \delta_1 x_{1t} + \delta_3 x_{3t} + u_{3t}$	$\sigma^2 + \beta_2^2\omega_{22}^*$	No
M_4	$y_t = \tau_2 x_{2t} + u_{4t}$	$\sigma^2 + \beta_1^2\omega_{11} + \beta_3^2\omega_{33}$	Yes
M_5	$y_t = \pi_1 x_{1t} + \pi_2 x_{2t} + \pi_3 x_{3t} + u_{5t}$	σ^2	Yes
DGP	$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t$	σ^2	Yes

[a]Model constancy is evaluated under the condition that the variance of x_{2t} changes at time $T + 1$, i.e., that $\omega_{22}^* \neq \omega_{22}$. Hence, models excluding x_{2t} are nonconstant.

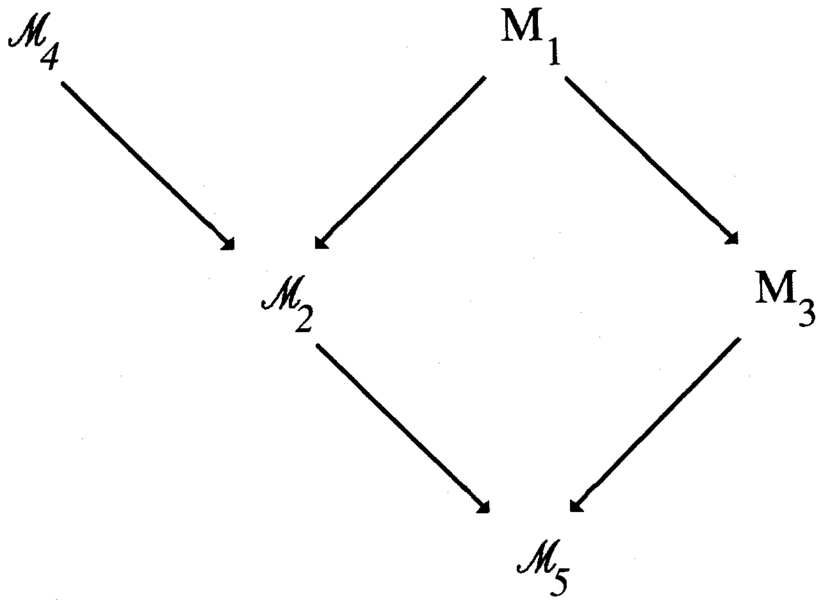

Figure 1. The ranking of $MSFE_t$ across models. Arrows denote direction of *decreasing* MSFE. The ranking of MSFE is indeterminate for each of the following pairs of models: ($\mathcal{M}_2$,M_3), (M_3,$\mathcal{M}_4$), and (M_1,$\mathcal{M}_4$). Models in $\mathcal{script}$ are always constant. Models in roman are nonconstant if $\omega_{22}^* \neq \omega_{22}$, and are constant otherwise.

eter constancy and minimizing MSFE, and an adequate forecasting model.

Proposition 4. *Individually and jointly, parameter constancy and minimizing MSFE are necessary but not sufficient to ensure an adequate forecasting model.*

Necessity is shown by considering the implications of lacking either property. Specifically, forecast errors from a nonconstant model contain a predictable element: for instance, by imposing an incorrect coefficient on the variable with nonconstant moments. For example, M_3 imposes a zero coefficient on x_{2t}. Also, a model that does not minimize MSFE, does not do so because it makes inefficient use of the information available for forecasting. A model that has constant parameters and does minimize MSFE across a set of models meets a necessary condition for being an adequate forecasting model. That condition is only necessary, however, and is not sufficient.

Lack of sufficiency can be shown, as follows. For the DGP and models above, model M_2 has constant parameters and, if $\beta_2^2\omega_{22}^*$ is larger than $\beta_3^2\omega_{33}$, M_2 minimizes MSFE among the models M_1, M_2, M_3, and M_4. However, the forecast errors for M_2 are:

$$\hat{u}_{2j} = y_j - \hat{y}_{2j} = \beta_3 x_{3j} + \epsilon_j \qquad j = T+1, \ldots, T+n, \tag{9}$$

and so are predictable on the data sets available to the models M_1, M_2, M_3, and M_4. The regressor x_{3j} is valuable in forecasting y_j, and M_2 ignores that information, but M_3 does not. Technically speaking, the forecast error $\hat{u}_{2j}$ is not an innovation with respect to the information set generated by models M_1, M_2, M_3, and M_4.

This analysis clarifies why minimizing MSFE is *not* enough for obtaining a good forecasting model. Although model M_2 minimizes MSFE over the set of models M_1, M_2, M_4, and (if $\beta_3^2\omega_{33} < \beta_2^2\omega_{22}^*$) M_3, the forecast errors of each model may be (in part) predictable from some other model's data.

Conversely, it is possible to create a model from the data of those four models that uniformly dominates M_1, M_3, M_4, *and* M_2 in MSFE, and that has constant parameters. One such model, denoted M_5, is

$$M_5\colon y_t = \pi_1 x_{1t} + \pi_2 x_{2t} + \pi_3 x_{3t} + u_{5t} \qquad u_{5t} \sim \mathrm{NID}(0,\sigma_5^2), \tag{10a}$$

with MSFE:

$$MSFE_5 = \sigma^2. \tag{10b}$$

Model M_5 has a smaller MSFE than even M_2 and has constant coefficients. Model M_5 happens to be the DGP and happens to nest M_1, M_2, M_3, and M_4, but neither property is necessary for it to "dominate" the other models in terms of MSFE. Rather, model M_5 dominates M_1, M_2, M_3, and M_4 because the forecast errors of any one of those four models are in part predictable from the regressors used in M_5, but not conversely. Note also that ϵ_j itself may be in part predictable on a larger information set, in which case the corresponding model's MSFE would be smaller than σ^2 in Equation 10b. Hendry (1986) discusses some of these issues in the related context of n-step ahead *ex ante* forecasts from macromodels.

If a model M_x minimizes the MSFE over a set of models $\{M_i\}$, that shows that the other models are worse in a specific sense. It does *not* show that M_x is a good forecasting model, *even on only the data available in the models* $\{M_i\}$. Even jointly, parameter constancy and minimum MSFE do not ensure efficient forecasting from the information available. Hence, there exists a need for more powerful tools in evaluating the forecast performance of models. The key to designing those tools is the information set against which models are being evaluated when MSFEs are compared. In Section 3, information sets resolve the logical status of parameter constancy vis-à-vis minimum MSFE. In Section 4, information sets define a taxonomy for test criteria, with parameter constancy and minimum MSFE being members of that taxonomy.

3. INFORMATION SETS

Information sets help clarify both why the MSFE and parameter constancy are sensible criteria for evaluating how "well" a model forecasts, and why having constant parameters and minimizing MSFE over a set of models are not in general sufficient conditions for obtaining an adequate forecasting model. The MSFE and tests of parameter constancy evaluate a given model against different sources of information, being either other models' MSFEs or the given model's fit over the estimation subsample. The former is obvious; the latter follows from Equation 5, the equation for the Chow statistic. Expressed somewhat differently, the Chow statistic evaluates a given model over different subsamples of that model's data, whereas minimizing MSFE evaluates several models over a given subsample but across the models' different datasets. The informational content of an alternative model's data and of the data of one's own model need not be (and generally are not) equivalent, so tests based on those information sets need not give similar results.

4. THE ROLES OF PARAMETER CONSTANCY AND MSFE IN EMPIRICAL MODELING

This section discusses how parameter constancy and minimizing MSFE fit into a general framework for evaluating (and designing) empirical models. That framework is based on the information sets against which models are evaluated and designed. It clarifies the relationship of parameter constancy and minimizing MSFE to other test statistics. It also results in a new test statistic, forecast-model encompassing, which is a more general and more stringent criteria for evaluating forecast performance than minimizing MSFE.

How well- or poorly designed an empirical economic model is depends upon its ability (or lack thereof) to capture salient features of the data and to deliver reliable inference on economic issues (e.g., coefficient estimates, predictions, and policy effects). Many statistics exist for evaluating such properties of a model; they relate to goodness-of-fit, absence of residual autocorrelation and heteroscedasticity, valid exogeneity, predictive ability, parameter constancy, the statistical and economic interpretation of estimated coefficients, the validity of a priori restrictions, and the ability of a model to account for properties of alternative models. These test statistics can serve as criteria both for evaluating existing specifications and for designing new ones. Table 2 summarizes the statistics, which are arranged by the type of information generating testable null hypotheses:

(A) the data of one's own model,
(B) the measurement system of the data,
(C) economic theory, and
(D) the data of alternative models.

For details, see Spanos (1986), Hendry and Richard (1982), Ericsson and Hendry (1985), Hendry (1987), and Ericsson, Campos, and Tran (1991).

Parameter constancy belongs to category A3 in Table 2 (the relative future of the data of one's own model) and is at the heart of model design, both statistically and economically. Most estimation techniques require parameter constancy for valid inference, and those that seem not to do so still posit "meta-parameters" assumed constant over time. Because economic systems appear far from constant empirically, and the coefficients of derived ("non-structural" or "reduced form") equations may alter when any of the underlying parameters or data correlations change, it is important to identify empirical models that have reasonably constant parameters and that remain interpretable when some change occurs.[4] That puts a premium on good theory. Conversely, empirical models with constant parameterizations in spite of "structural change" elsewhere in the economy may provide the seeds of fruitful research in economic theory. Parameter constancy typically is evaluated by comparing parameter estimates of a given model obtained from different subsamples of data. Recursive estimation of an equation provides an incisive tool for investigating parameter constancy, both through the sequence of estimated coefficient values and via the associated Chow statistics; see, for example, Dufour (1982).

Minimizing MSFE, like parameter constancy, focuses on the "relative future," but on that of alternative models' data rather than on the data of one's own model. Thus, MSFE dominance belongs to category D3. Because the structure of D3 parallels that of D1 (the relative past of alternative models' data), D1 is briefly discussed to

[4]See Goldfeld and Sichel (1990) for a discussion of the nonconstancy of many estimated money-demand equations. That nonconstancy implies nonconstancy in one or more of the equations of the underlying data generation process.

elucidate the connections between the two. Also, for reasons which will be apparent shortly, the criterion of minimizing MSFE will be referred to as *MSFE dominance*.

Parameter encompassing, variance encompassing, and variance dominance. Consider the following two alternative non-nested linear models, both claiming to explain y_t.

$$M_1:\quad y_t = \delta_1' z_{1t} + \nu_{1t} \qquad \nu_{1t} \sim \mathrm{NID}(0,\sigma_1^2) \tag{11a}$$

$$M_2:\quad y_t = \delta_2' z_{2t} + \nu_{2t} \qquad \nu_{2t} \sim \mathrm{NID}(0,\sigma_2^2) \tag{11b}$$

The notation is distinct from that in Section 2 above. In Equations 11a and 11b, δ_1 and δ_2 are $k_1 \times 1$ and $k_2 \times 1$ vectors of unknown parameters. The vectors z_{1t} and z_{2t} are of k_1 and k_2 regressors respectively, with each vector having at least some variables that are not in common with those in the other vector. For simplicity, assume that none are in common. To ensure the feasibility of the parameter-encompassing and forecast-model encompassing statistics, assume that $T > k_1 + k_2$ and $n > \max(k_1,k_2)$.

As alternative models, Equation 11a entails the irrelevance of z_{2t} in explaining y_t, given z_{1t}; and vice versa for Equation 11b. In any event, the variables y_t, z_{1t}, and z_{2t} are generated by *some* process, and, under the simplifying (but inessential) assumption of joint normality, z_{1t} and z_{2t} can be linked using

$$z_{1t} = \Pi z_{2t} + \zeta_{1t}, \tag{12}$$

where Π is defined by $\mathscr{E}(z_{2t}\zeta_{1t}') = 0$, and (again for expositional simplicity) $\mathscr{E}(\zeta_{1t}\,\zeta_{1t}') = \Omega$. Substituting Equation 12 into Equation 11a obtains

$$\begin{aligned} y_t &= \delta_1' z_{1t} && + \quad \nu_{1t} \\ &= (\delta_1'\Pi) z_{2t} && + \quad (\nu_{1t} + \delta_1'\zeta_{1t}) \\ &= (\delta_2') z_{2t} && + \quad \nu_{2t}. \end{aligned} \tag{13}$$

In Equation 13, the parameter δ_2 and the error ν_{2t} are derived from Equations 11a and 12, being $\Pi'\delta_1$ and $\nu_{1t} + \delta_1'\zeta_{1t}$ respectively. Consequently, Equation 13 is what model Equation 11a predicts model Equation 11b should find, and it implies several hypotheses, including

$$H_a:\quad \delta_2 = \Pi'\delta_1 \tag{14a}$$

and

$$H_b:\quad \sigma_2^2 = \sigma_1^2 + \delta_1'\Omega\delta_1. \tag{14b}$$

Table 2: Evaluation/Design Criteria (*continued*)

Information set	Null hypothesis	Alternative hypothesis	References
(A) own model's data			
(A1) relative past	innovation errors	first-order residual autocorrelation	Durbin and Watson (1950, 1951)
"	"	qth-order residual autocorrelation	Box and Pierce (1970); Godfrey (1978), Harvey (1981, p. 173)
"	"	invalid parameter restrictions	Johnston (1963, p. 126)
"	"	qth-order ARCH	Engle (1982)
"	"	heteroscedasticity quadratic in regressors	White (1980, p. 825), Nicholls and Pagan (1983)
"	"	qth-order RESET	Ramsey (1969)
"	normality of the errors	skewness and excess kurtosis	Jarque and Bera (1980)
(A2) relative present	weakly exogenous regressors	invalid conditioning	Sargan (1958, 1980), Engle, Hendry, and Richard (1983)
(A3) relative future	constant parameters, adequate forecasts	parameter nonconstancy, predictive failure	Fisher (1922), Chow (1960), Brown, Durbin, and Evans (1975), Hendry (1979)
(B) measurement system	data admissibility	"impossible" predictions of observables	
(C) economic theory	theory consistency	"implausible" coefficients, predictions; no conintegration	Engle and Granger (1987)

Table 2: Evaluation/Design Criteria (*continued*)

Information set	Null hypothesis	Alternative hypothesis	References
(D) alternative models' data			
(D1) relative past	variance dominance	poor fit relative to an alternative model	Hendry and Richard (1982)
"	variance encompassing	inexplicable observed error variance	Cox (1961, 1962), Pesaran (1974), Hendry (1983)
"	parameter encompassing	significant additional variables	Johnston (1963, p. 126), Mizon and Richard (1986)
(D2) relative present	exogeneity encompassing	inexplicable valid conditioning	Hendry (1988)
(D3) relative future	MSFE dominance	poor forecasts relative to those of alternative models	Granger (1989), Granger and Deutsch (1992)
"	forecast encompassing	informative forecasts from alternative models	Chong and Hendry (1986)
"	forecast-model encompassing	regressors from alternative models valuable for forecasting	(this paper)

These hypotheses are called *parameter encompassing* and *variance encompassing*, respectively, and the positive definiteness of Ω in the latter implies *variance dominance*:

$$H_c:\quad \sigma_1^2 < \sigma_2^2. \tag{14c}$$

H_a, H_b, and H_c are implications of omitted variable bias in Equation 11b, assuming that Equations 11a plus 12 are the DGP. These three hypotheses, albeit in reverse order, generate the evaluation criteria for D1 on Table 2 (see Hendry, 1983, and Mizon and Richard, 1986).

Parameter encompassing by Equation 11a of Equation 11b may be tested using the formula in Equation 14a or by testing whether z_{2t} is irrelevant if added to Equation 11a. To see the latter, let δ_2 be unconstrained, and define the $k_2 \times 1$ vector γ as $\delta_2 - \Pi'\delta_1$. H_a is equivalent to $\gamma = 0$. By substitution of $\delta_2 = \Pi'\delta_1 + \gamma$ in (13),

$$y_t = \delta_1' z_{1t} + \gamma' z_{2t} + \nu_{1t}. \tag{15}$$

Thus, H_a is equivalent to claiming that z_{2t} has no power in explaining y_t, given z_{1t} (or in explaining the residuals from Equation 11a). In practice, it is simplest to estimate δ_1 and γ jointly in Equation 15 and test $\gamma = 0$ with the standard F statistic.

Variance encompassing may be tested either by using Equation 14b or by testing the insignificance in Equation 11a of the *fitted* values from Equation 11b. That is, hypothesis H_b can be tested by jointly estimating δ_1 and α in:

$$y_t = \delta_1' z_{1t} + \alpha \hat{y}_{2t} + \nu_{1t}, \tag{16}$$

and testing that $\alpha = 0$ (where $\hat{y}_{2t} = \hat{\delta}_{2T}' z_{2t}$). Testing $\alpha = 0$ is equivalent to testing $\sigma_2^2 = \sigma_1^2 + \delta_1'\Omega\delta_1$; see Davidson and MacKinnon (1981, p. 789), Hendry (1983), and Mizon and Richard (1986). Equally, H_b is equivalent to claiming that a certain linear combination of z_{2t}, namely $\hat{y}_{2t}$, has no power in explaining the residuals from Equation 11a. Because Equation 16 is testing for the insignificance of that certain linear combination, rather than of any linear combination (as in Equation 15), the test of $\alpha = 0$ is a narrower test than that of $\gamma = 0$. The t statistic on α in Equation 16 is Davidson and MacKinnon's (1981) P statistic: it is asymptotically $N(0,1)$ when M_1 is true and is asymptotically equivalent to Cox's (1961) statistic for testing non-nested hypotheses, as applied to linear regression models by Pesaran (1974).

The logic of the hypotheses in Equation 14 is as follows: variance dominance is necessary, but not sufficient, for variance encompassing, which in turn is necessary, but not sufficient, for parameter encompassing. Conversely, if $\gamma = 0$ in Equation 15, then α also must be zero in Equation 16, from which it follows that σ_1^2 is less than σ_2^2 because z_{1t} is not an exact linear transformation of z_{2t}.

In light of the preceding analysis, it readily follows that MSFE dominance parallels variance dominance, and that two other forecast criteria (forecast encompassing and the new forecast-model encompassing) parallel variance encompassing and parameter encompassing. The remainder of this section explores those connections between D1 and D3 in Table 2.

Forecast-model encompassing, forecast encompassing, and MSFE dominance. Assume that the two alternative models, Equations 11a and 11b, have been estimated over the sample period $[1,T]$ and are being used to forecast over $[T + 1, T + n]$. The forecasts from the two models are

$$\hat{y}_{1j} = \delta_1' z_{1j} \tag{17a}$$

$$\hat{y}_{2j} = \delta_2' z_{2j} \qquad j = T + 1, \ldots, T+n, \tag{17b}$$

ignoring (again) the uncertainty arising from estimating coefficients over $[1,T]$. As with Equation 13 above, under M_1,

$$\begin{aligned} y_j &= \delta_1' z_{1j} + \nu_{1j} \\ &= (\delta_1'\Pi) z_{2j} + (\nu_{1j} + \delta_1'\zeta_{1j}) \\ &= (\delta_2') z_{2j} + [(y_j - \hat{y}_{1j}) + \delta_1'\zeta_{1j}] \\ &= \hat{y}_{2j} + (y_j - \hat{y}_{2j}), \end{aligned} \tag{18}$$

where the last line follows from Equation 17b and the equality $y_j = y_j$. Equation 18 implies two testable hypotheses:

$$H_a^*: \delta_2 = \Pi'\delta_1 \tag{19a}$$

and

$$H_b^*: \mathscr{E}(y_j - \hat{y}_{2j})^2 = \mathscr{E}(y_j - \hat{y}_{1j})^2 + \delta_1'\Omega^*\delta_1, \tag{19b}$$

where an asterisk * denotes the corresponding hypothesis or matrix over the *forecast* period. The second hypothesis may be written as

$$H_b^*: MSFE_2 = MSFE_1 + \delta_1'\Omega^*\delta_1, \tag{19c}$$

and implies

$$H_c^*\colon MSFE_1 < MSFE_2. \tag{19d}$$

These three hypotheses H_a^*, H_b^*, and H_c^* are called *forecast-model encompassing, forecast encompassing*, and *MSFE dominance*. Forecast encompassing could be called MSFE encompassing as well; see Chong and Hendry (1986). From Equation 19d, it follows that an adequate forecasting model must minimize MSFE (asymptotically), but doing so is a necessary (and not sufficient) condition for obtaining an adequate forecasting model, as discussed in Section 2. The design of tests of H_a^* and H_b^* parallels those of H_a and H_b.

Forecast-model encompassing by Equation 11a of Equation 11b may be tested using the formula in Equation 19a or by testing whether z_{2j} is irrelevant in explaining the forecast errors from Equation 17a. As with parameter encompassing, let $\gamma = \delta_2 - \Pi'\delta_1$. By substitution in Equation 18,

$$y_j = \delta_1' z_{1j} + \gamma' z_{2j} + \nu_{1j} \qquad j = T + 1, \ldots, T + n, \tag{20a}$$

or

$$y_j - \hat{y}_{1j} = \gamma' z_{2j} + \nu_{1j}. \tag{20b}$$

H_a^* is equivalent to $\gamma = 0$ and so to claiming that z_{2j} has no power in explaining the forecast errors from Equation 17a. For large T, fixed z_{ij}'s, and normal ν_{1j}, it is straightforward to show that the standard F statistic testing $\gamma = 0$ in Equation 20b is distributed as $F(k_2, n - k_2)$ under M_1; for stochastic (weakly exogenous) z_{ij}'s, it is distributed as $F(k_2, n - k_2)$ for large n. See the Appendix for details.

An exact test of forecast-model encompassing also exists. It can be motivated most easily by recognizing why the F statistic for $\gamma = 0$ in Equation 20b does not have a simple exact distribution: the statistic conditions upon the *estimated* value of δ_1, thereby ignoring the uncertainty inherent in the corresponding estimator of δ_1. The solution is simple: estimate δ_1 and γ jointly. To do so, consider the following model:

$$y_t = \delta_1' z_{1t} + \gamma' z_{2t}^* + \nu_{1t} \qquad t = 1, \ldots, T + n, \tag{21}$$

where the *entire* data sample $[1, T + n]$ is used, and z_{2t}^* is zero over $[1, T]$ and equal to z_{2t} over $[T + 1, T + n]$. The F statistic for $\gamma = 0$ in Equation 21 is exactly distributed as $F(k_2, T + n - k_1 - k_2)$ under M_1 for fixed z_{ij}'s and normal ν_{1j}, asymptotically so for stochastic (weakly exogenous) z_{ij}'s. Instrumental variable and recursive generalizations of the test statistics for $\gamma = 0$ in Equations 20b and 21 follow naturally.[5]

Forecast encompassing by Equation 11a of Equation 11b may be tested either by using Equation 19c directly or by testing for the insignificance of the *forecast* values given by Equation 17b in explaining the forecast errors from Equation 17a. Thus, H_b^* can be tested by estimating α in

$$y_j - \hat{y}_{1j} = \alpha\hat{y}_{2j} + \nu_{1j} \qquad j = T + 1, \ldots, T + n, \tag{22}$$

and testing that $\alpha = 0$. That is equivalent to testing that $MSFE_2 = MSFE_1 + \delta_1'\Omega^*\delta_1$, following the logic used for variance encompassing. Noting that $\hat{y}_{1j} = \delta_1' z_{1j}$, Equation 22 is similar to Equation 16, the principal difference being that the time period is $[T + 1, T + n]$ rather than $[1,T]$. Chong and Hendry (1986) have shown that, for large T *and* n, the t statistic on α is $N(0,1)$.[6]

The logic of the hypotheses in Equation 19 is as follows: MSFE dominance is necessary, but not sufficient, for forecast encompassing, which in turn is necessary, but not sufficient, for forecast-model encompassing. Conversely, if $\gamma = 0$ in Equation 20b, then α (in Equation 22) must be zero because $\hat{y}_{2j} = \delta_2' z_{2j}$. If $\alpha = 0$, then $MSFE_1$ is less than $MSFE_2$ because z_{1j} is not an exact linear transformation of z_{2j}.

Forecast-type encompassing and parameter constancy. As illustrated in Section 2, even if the "structural" relationship has constant parameters, e.g., (π_i, σ_5^2) in Equation 10a or (δ_1, σ_1^2) in Equation 11a, nonconstant population data moments have implications for the empirical constancy (or lack thereof) of mis-specified models. Nonconstant population data moments also have implications for forecast-type encompassing tests. For instance, if the (reduced-form) variance matrix Ω changes, $MSFE_2$ in H_b^* will alter as that new matrix, i.e., Ω^* in

---

[5]The structure of Equation 21 also leads to two classes of forecast-based encompassing tests, one which assumes constancy between the estimation and forecast samples, and one which does not.

[6]An extensive literature has developed on the combination or "pooling" of forecasts, i.e., where some (usually linear) combination of forecasts from different models is taken to obtain a new forecast. In comparison with any of the individual model forecasts, that new forecast may have better properties, usually being a smaller MSFE. Given the discussion in the text above, finding such a pooled forecast is prima facie evidence of all individual models being mis-specified, and may well indicate that a single model can be constructed that has a smaller MSFE than even the pooled forecast. See Clemen (1989) for a review and bibliography on combining forecasts, and Granger (1989, pp. 187–191) and Diebold (1989) for recent analyses.

Equation 19c, does. Likewise, if the Π matrix changes, because M_2 (falsely) assumes δ_2 is constant, M_2 will have systematic forecast errors that are a function of the changing Π matrix. These "predictions" about model behavior suggest a more general encompassing strategy, including predicting problems in alternative models of which their proponents are unaware. Corroborating such phenomena adds credibility to the claim that the successful model reasonably represents the data generation process, whereas disconfirmation clarifies that it does not.

5. FORECASTING WITH INTEGRATED AND COINTEGRATED VARIABLES

This section describes a certain lack of invariance present in the forecast-encompassing statistics, and illustrates that lack of invariance with a cointegrated process. The forecast-encompassing statistic is then modified to produce an "invariant" test statistic, which tests "forecast-differential encompassing". See Lu and Mizon (1991) for a related discussion.

The forecast errors $y_j - \hat{y}_{1j}$ and $y_j - \hat{y}_{2j}$ are invariant to nonsingular linear transformations of the corresponding models' data, $(y_j, z'_{1j})'$ and $(y_j, z'_{2j})'$. The forecasts themselves, however, are *not* invariant to such transformations, and so neither is the forecast-encompassing test statistic from Equation 22. Specifically, suppose that both z_{1j} and z_{2j} include the lagged dependent variable y_{j-1}, in which case models M_1 and M_2 may be written without loss of generality with either y_j or Δy_j as the dependent variable, where Δ is the first-difference operator. In the first case, with y_j, the auxiliary equation for the forecast-encompassing statistic is Equation 22 as written. In the second case, $\Delta\hat{y}_{2j}$ replaces $\hat{y}_{2j}$ as the right-hand side variable in Equation 22. In both cases, the t statistic on α is asymptotically N(0,1) under the null hypothesis of M_1 being correctly specified. However, the two t statistics are not necessarily equivalent under mis-specification of M_1. This is most apparent when y_j is an integrated process.

To illustrate, suppose that y_j, z_{1j}, and z_{2j} are each I(1) processes, and that each model (M_1 and M_2) represents a cointegrating relationship. This could arise if, e.g., z_{1j} and z_{2j} involved different lag structures of the same underlying variables. From Granger (1986), the forecast errors $y_j - \hat{y}_{1j}$ and $y_j - \hat{y}_{2j}$ are each $I(0)$, whereas the forecasts $\hat{y}_{1j}$ and $\hat{y}_{2j}$ are $I(1)$. Thus, in order for Equation 22 to be "balanced" in terms of orders of integration, α must be zero. Surprisingly, α must be zero even if M_1 is mis–specified and M_2 is the correct model. That is, the forecast-encompassing test may have no power when the dependent variable is $I(1)$.[7,8]

For y_j, z_{1j}, and z_{2j} with the properties specified, both M_1 and M_2 have error-correction representations. In the error-correction representation, the dependent variable is Δy_j, rather than y_j. The corresponding forecast errors remain unchanged numerically, but the right-hand side variable in Equation 22 becomes $\Delta \hat{y}_{2j}$, an I(0) variable, in contrast to $\hat{y}_{2j}$, which is $I(1)$. Balance is unaffected by the value of α in the regression with $\Delta \hat{y}_{2j}$, so the corresponding forecast-encompassing test appears more promising for good power properties than the test based on Equation 22 with $\hat{y}_{2j}$ on the right-hand side. This feature supports formulating forecast models "in I(0) space" as error-correction models, rather than "in I(1) space" in terms of the original [I(1)] levels variables.

It may be desirable to have a forecast-encompassing test that is invariant to nonsingular linear transformations of the data. Such a test may be constructed as follows. As Equation 22 stands, the coefficient on $\hat{y}_{1j}$ is constrained to be unity, while the coefficient on $\hat{y}_{2j}$ is estimated unrestrictedly. Instead, both coefficients could be estimated, with their sum constrained to be unity. The resulting equation can be written as

$$y_j - \hat{y}_{1j} = \alpha^* (\hat{y}_{2j} - \hat{y}_{1j}) + \nu_{1j} \qquad j = T + 1, \ldots, T + n, \tag{23}$$

where α^* is estimated unrestrictedly, and $\alpha^* = 0$ is tested. This equation would parallel Davidson and MacKinnon's (1981) J statistic if the coefficients in $\hat{y}_{1j}$ were estimated jointly with α^*.

The test from Equation 23 has two important features. First, because the right-hand side variable is the differential between the two forecasts $[(\hat{y}_{2j} - \hat{y}_{1j})]$ rather than either forecast alone, the right-hand side variable is unaffected by nonsingular linear transformations of the models' data. Thus, the test of $\alpha^* = 0$ is invariant to such transformations. Second, for integrated y_j, z_{1j}, and z_{2j} with both models cointegrated, the right-hand side variable $(\hat{y}_{2j} - \hat{y}_{1j})$ is I(0), preserving balance. This follows because $\hat{y}_{1j}$ and $\hat{y}_{2j}$ must each cointegrate with y_j (with unit coefficients), and so $\hat{y}_{1j}$ cointegrates with $\hat{y}_{2j}$ (also with unit coefficients). The t statistic on α^* will be called the forecast-differential encompassing statistic, noting the form of the right-hand side variable.

[7]I am grateful to Stephen Hall for bringing to my attention (via David Hendry) the apparently low power empirically of Chong and Hendry's forecast-encompassing test with I(1) forecasted variables. Also, see Hendry (1989, pp. 95–97) on implications of nonsingular linear transformations of a linear model's data.

[8]Either or both of models M_1 and M_2 might lack cointegration, in which case the distributions of the forecast-based test statistics may change. We do not consider such cases here.

As a practical matter, any of a model's forecast errors, its forecasts, or the forecast differential $(\hat{y}_{2j} - \hat{y}_{1j})$ may have a nonzero mean. Thus, the power of the tests from Equations 22 and 23 can be affected by the inclusion of a constant term in the auxiliary regression. Under the null of correct specification, the constant term should have a zero coefficient, so it is appropriate to test that α (or α^*) and the constant term's coefficient are jointly zero.

Before applying several of the above tests to a pair of empirical models, we briefly consider the class of models with time-varying coefficients.

6. FORECASTING AND MODELS WITH TIME-VARYING COEFFICIENTS

Time-varying coefficient (TVC) models have been proposed as a means of improving forecast performance. The results above can clarify when that may (or may not) be so.

First, if the data are generated with time-varying coefficients and a correctly specified TVC model is estimated and used for forecasting, then the TVC forecasts will minimize MSFE, and (in general) fixed-coefficient models will have a higher MSFE. Evidence that the TVC model satisfied the evaluation criteria listed on Table 2 would be necessary for the model to be credible as representing the data process.

Second, sometimes an estimated TVC model is recognized as being mis-specified, but it is claimed that the TVC model will forecast better than fixed-coefficient models because the former accounts (in part, at least) for observed parameter nonconstancy in the latter (see, for example, Chow, 1984). For general parameter nonconstancy, however, a TVC model need not minimize MSFE relative to a fixed-coefficient model, even asymptotically and even if the TVC model nests the fixed-coefficient model. An example suffices.

Suppose that the data are generated as

$$y_t = \delta_{1t}' z_{1t} + \nu_{1t} \qquad \nu_{1t} \sim \mathrm{NID}(0, \sigma_1^2), \tag{24}$$

where z_{1t} is stationary, distributed as $\mathrm{N}(0, \Psi_{11})$; and $\delta_{1t} = \delta - \theta$ $(\theta \neq 0)$ for the first half of the estimation period, $\delta_{1t} = \delta + \theta$ for the second half of the estimation period, and $\delta_{1t} = \delta$ for the forecast period, which is the single observation $T + 1$ (for expository convenience). The fixed-coefficient model is Equation 24 but assumes that δ_{1t} is constant. The TVC model specifies, e.g., that $(\delta_{1t} - \bar{\delta}) = \phi(\delta_{1t-1} - \bar{\delta}) + \xi_t$ where ξ_t is assumed to be, e.g., white noise, $|\phi|$

< 1, and $\bar{\delta}$ is the unconditional mean of δ_{1t} (see, for example, Swamy and Tinsley, 1980, and Chow, 1984).

The fixed-coefficient model, although manifesting parameter nonconstancy in sample, has $\mathscr{E}(\hat{\delta}_{1T}) \approx \delta$, and so has a MSFE of approximately σ_1^2. The TVC estimate $\tilde{\delta}_{1T}$ is approximately $\delta + \theta$ because the TVC estimator places more weight on recent data than on older data. Here, the TVC model does so by obtaining estimates of $\bar{\delta}$ and ϕ which are approximately δ and unity respectively. For the forecast observation $T + 1$, the TVC model uses the prediction $\tilde{\delta}'_{1,T+1}z_{1,T+1}$, which is approximately $\tilde{\delta}'_{1,T}z_{1,T+1}$ or $(\delta + \theta)z_{1,T+1}$. Thus, the TVC model has MSFE of approximately $\sigma_1^2 + \theta'\Psi_{11}\theta$, which is greater than σ_1^2.

7. AN EMPIRICAL ILLUSTRATION: MODELS OF NARROW MONEY DEMAND IN THE UNITED KINGDOM

This section calculates the MSFE and forecast-type encompassing statistics for two models of U.K. money demand from Hendry and Ericsson (1991). The first model is an error correction model, the second is a partial adjustment model, and the forecast period is the 1980s.

Hendry and Ericsson (1991) develop a constant, parsimonious, error correction model of narrow money in the United Kingdom for the period 1964:3–1989:2. Estimating their equation through 1979 and forecasting over the 1980s obtains the following:

$$\widehat{\Delta(m-p)}_t = \underset{(0.20)}{-\ 0.80}\,\Delta p_t \underset{(0.07)}{-\ 0.20}\,\Delta(m-p-y)_{t-1} \underset{(0.10)}{-\ 0.63}\,R_t^* \underset{(0.013)}{-\ 0.102}\,(m-p-y)_{t-1} \underset{(0.007)}{+\ 0.021} \qquad (25)$$

$T = 62$ [1964:3–1979:4] + 38 forecasts $R^2 = 0.69$ $\hat{\sigma} = 1.389$ percent.

The data are nominal M_1 (M), 1985 price total final expenditure (Y), the corresponding deflator (P), and a (learning-adjusted) net interest rate (R^*). Lower case denotes logarithms, and standard errors are in parentheses. Hendry and Ericsson (1991) describe the data in their appendix and discuss the statistical and economic merits of Equation 25 in their Section 4.

Hendry and Ericsson (1991) also estimate a partial adjustment model for real narrow money with an autoregressive error, in the spirit of Goldfeld (1973). Contrasting with results on U.S. data, the partial adjustment model appears reasonably constant during the missing money period: the Chow statistic is $F[12,29] = 1.89$ [p value $= 0.079$] for forecasts over 1973:1–1975:4. When estimated through 1979 and forecast over the 1980s, the estimates for the partial adjustment model are as follows:

$$\begin{aligned}\widehat{(m-p)_t} = {} & \underset{(0.024)}{0.955}\ (m-p)_{t-1} + \underset{(0.020)}{0.087}\ y_t \\ & - \underset{(0.07)}{0.78}\ R_t^* - \underset{(0.42)}{0.43} - \underset{(0.13)}{0.31}\ \hat{u}_{t-1} \end{aligned} \tag{26}$$

$$T = 62\ [1964{:}3\text{–}1979{:}4] + 38 \text{ forecasts} \quad \hat{\sigma} = 1.572 \text{ percent.}$$

The coefficient on $\hat{u}_{t-1}$ is the estimated parameter of the (modeled) first-order autoregressive disturbance.

For money-demand equations, the 1980s are of particular interest to forecast, as Goldfeld and Sichel (1990, p. 300) note: "... in the 1980s, U.S. money demand functions, whether or not fixed up to explain the 1970s, generally exhibited extended periods of underprediction as observed velocity fell markedly." In the United Kingdom, velocity fell by twice the percentage drop in the United States. Somewhat surprisingly, neither model fails the Chow test, as seen in Table 3.

Figure 2 graphs the forecast errors from Equations 25 and 26. Visually, the errors from Equation 25 appear uncorrelated with near zero mean, whereas those from Equation 26 are highly autocorrelated, trending from large and positive in 1980 to large and negative in 1989. These series are the dependent variables in the auxiliary regressions for calculating the forecast-type encompassing test statistics. The particular type of test determines the "independent" variables in those auxiliary regressions. Figures 3 and 4 present those variables for the forecast-encompassing tests. Figure 3 graphs the forecasts of $\Delta(m - p)_t$ from Equations 25 and 26, whereas Figure 4 graphs the forecasts of $(m - p)_t$ from Equations 25 and 26. For reference, Figures 3 and 4 also include the series being forecast. The initial underprediction and subsequent overprediction by Equation 26 is clear in both graphs although, from the Chow statistic, these deviations are not statistically detectable *as a structural break*. Even so, the forecast-type encompassing test statistics detect the systematic (and hence predictable) nature of Equation 26's forecast errors.

Specifically, as shown in Table 3, Equation 26 fails all forecast-type encompassing tests, except for the forecast-encompassing test with $(m - p)_t$ as the forecast variable and no constant term included in Equation 22. That single lack of failure is due to the approximately

Table 3: Chow, Encompassing, and Related Statistics

Statistic	Null hypothesis (i.e., hypothesized encompassing model)[a,b]			
	Error correction (Equation 25)		Partial adjustment (Equation 26)	
	no constant	constant	no constant	constant
Chow statistic	0.73		0.81	
	[0.84]		[0.75]	
	$F[38,57]$		$F[38,57]$	
$\hat{\sigma}$	1.389%		1.572%	
Root MSFE	1.232%		2.507%	
Forecast encompassing[c]				
Variable forecast is: $\Delta(m - p)_t$	0.00	0.70	4.22	3.78
	[0.96]	[0.50]	[0.047]	[0.032]
	$F[1,37]$	$F[2,36]$	$F[1,37]$	$F[2,36]$
Variable forecast is: $(m - p)_t$	1.12	1.21	0.09	63.14
	[0.30]	[0.31]	[0.77]	[0.000]
	$F[1,37]$	$F[2,36]$	$F[1,37]$	$F[2,36]$
Forecast-differential encompassing	2.27	1.65	125.54	63.35
	[0.14]	[0.21]	[0.000]	[0.000]
	$F[1,37]$	$F[2,36]$	$F[1,37]$	$F[2,36]$
Forecast-model encompassing	0.85		29.38	
	[0.53]		[0.000]	
	$F[5,33]$		$F[5,33]$	
Forecast-model encompassing[d]	0.37		3.27	
("exact")	[0.87]		[0.009]	
	$F[5,90]$		$F[5,90]$	

[a]The three entries for a given statistic and equation are: the value of the statistic, the right-hand tail probability associated with that statistic, and the statistic's distribution under the null hypothesis of that equation being correctly specified. The estimation period is 1964:3–1979:4 [$T = 62$]; the forecast period is 1980:1–1989:2 [$n = 38$].

[b]The phrases "constant" and "no constant" denote whether or not a constant term is included in the auxiliary regression, i.e., in Equations 22 or 23.

[c]Quantitatively similar results obtain for the following pairs of variables forecast: Δm_t and m_t, and $\Delta(m - p - y)_t$ and $(m - p - y)_t$.

[d]The full-sample ($T = 100$) values of $\hat{u}_{t-1}$ in Equation 26 are used to calculate the "exact" forecast-model encompassing test statistic.

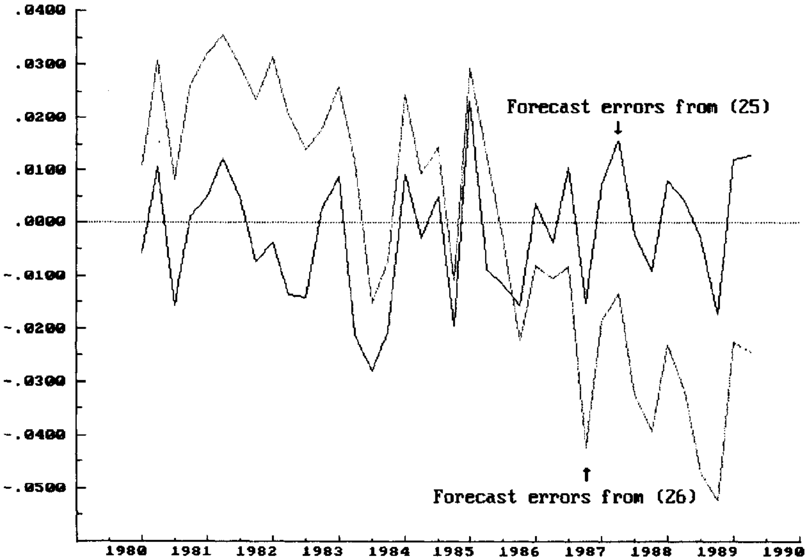

Figure 2. Forecast errors from two models of narrow money demand in the United Kingdom: Equation 25 (the error correction model) and Equation 26 (the partial adjustment model).

zero mean of Equation 26's forecast errors and the large nonzero mean of $(m - p)_t$: see Figures 2 and 4, respectively. Inclusion of a constant term results in rejection at the 0.1 percent level, with the (upwardly) trending $(m - p)_t$ ''explaining'' the (downwardly) trending forecast errors.

Equation 25 dominates Equation 26 substantially in terms of MSFE. Additionally, Equation 25 encompasses Equation 26 according to all forecast-based encompassing tests.

The results for Equations 25 and 26 show how two models may be empirically constant, yet (at least) one may be inadequate for forecasting. This parallels Proposition 1. In evaluating models of the U.S. trade balance, Marquez and Ericsson (1992) find an empirically nonconstant model that obtains the minimum MSFE with respect to all other models considered. That parallels Proposition 2.

8. SUMMARY

Parameter constancy and minimizing the mean square forecast error are sensible criteria that evaluate empirical models against two different

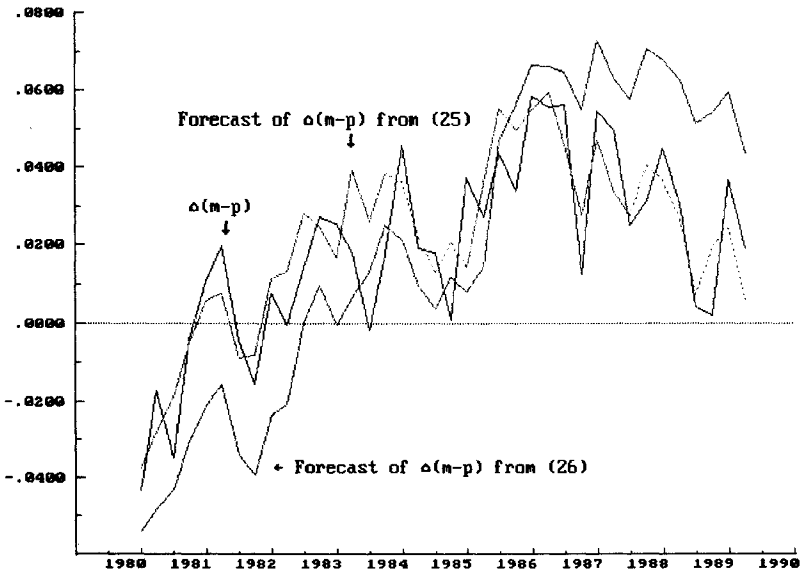

Figure 3. Actual values of the growth rate of real money in the United Kingdom [$\Delta(m - p)_t$] and the forecast values thereof from Equations 25 and 26.

information sets, the data of one's own model and the data of alternative models. MSFE dominance is a necessary condition for two more general criteria for evaluating forecast performance: forecast encompassing and forecast-model encompassing. Parameter constancy, MSFE dominance, and the two types of forecast encompassing fit into a general taxonomy of model evaluation criteria. Satisfying *all* those evaluation criteria (and not just those of parameter constancy and MSFE dominance) are in general necessary for obtaining an adequate forecasting model. Two models for money demand in the United Kingdom help illustrate the concepts developed.

APPENDIX: DISTRIBUTIONS OF STATISTICS FOR TESTING FORECAST-MODEL ENCOMPASSING

The basis for the forecast-model encompassing test statistic is the auxiliary regression in Equation 20b:

$$y_j - \hat{y}_{1j} = \gamma' z_{2j} + \nu_{1j} \qquad j = T + 1, \ldots, T + n. \tag{A1}$$

Under M_1, for large T, fixed z_{ij}'s, and normal ν_{1j}, the dependent variable in (A1) *is* ν_{1j} and γ is zero, so the standard F statistic testing $\gamma = 0$

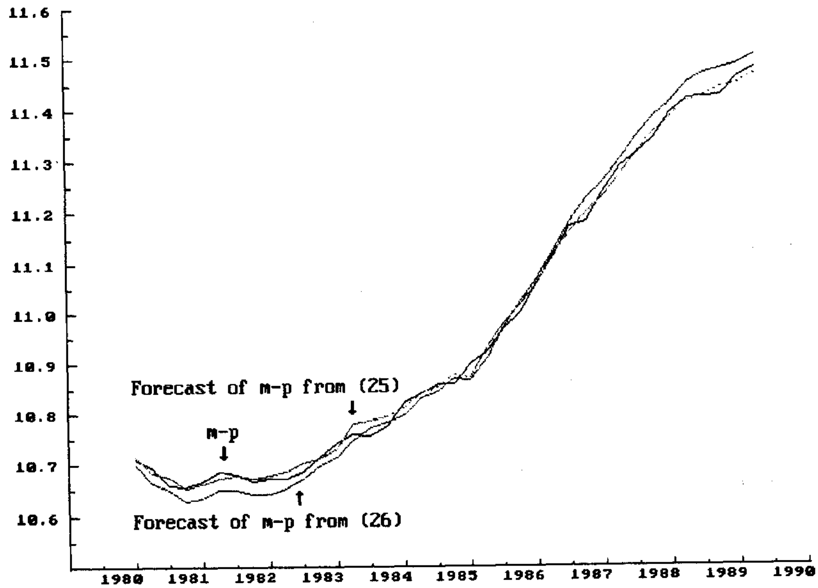

Figure 4. Actual values of the logarithm of real money in the United Kingdom $[(m - p)_t]$ and the forecast values thereof from Equations 25 and 26.

is distributed as $F(k_2,n - k_2)$. For stochastic (weakly exogenous) z_{ij}'s, it is distributed as $F(k_2,n - k_2)$ for large n. See Hendry (1979) and Kiviet (1986).

The distribution of the modified forecast-model encompassing test statistic from the auxiliary regression in Equation 21 follows directly from, e.g., Johnston (1963, Chapter 4).

REFERENCES

Box, G.E.P., and Pierce, D.A. (1970) Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models, *Journal of the American Statistical Association* 65: 1509–1526.

Brown, R.L., Durbin, J., and Evans, J.M. (1975) Techniques for Testing the Constancy of Regression Relationships over Time, *Journal of the Royal Statistical Society, Series B* 37: 149–192 (with discussion).

Campos, J., and Ericsson, N.R. (1988) Econometric Modeling of Consumers' Expenditure in Venezuela, International Finance Discussion Paper No. 325, Board of Governors of the Federal Reserve System, Washington, D.C.

Chong, Y.Y., and Hendry, D.F. (1986) Econometric Evaluation of Linear Macro-economic Models, *Review of Economic Studies* 53: 671–690.

Chow, G.C. (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica* 28: 591–605.

Chow, G.C. (1984) Random and Changing Coefficient Models, In *Handbook of Econometrics, Volume 2* (Z. Griliches and M.D. Intriligator, Eds.). Amsterdam: North-Holland. 1213–1245.

Clemen, R.T. (1989) Combining Forecasts: A Review and Annotated Bibliography, *International Journal of Forecasting* 5: 559–583.

Cox, D.R. (1961) Tests of Separate Families of Hypotheses, In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1* (J. Neyman, Ed.). Berkeley: University of California Press. 105–123.

Cox, D.R. (1962) Further Results on Tests of Separate Families of Hypotheses, *Journal of the Royal Statistical Society, Series B* 24: 406–424.

Davidson, R., and MacKinnon, J.G. (1981) Several Tests for Model Specification in the Presence of Alternative Hypotheses, *Econometrica* 49: 781–793.

Diebold, F.X. (1989) Forecast Combination and Encompassing: Reconciling Two Divergent Literatures, *International Journal of Forecasting* 5: 589–592.

Dufour, J.-M. (1980) Dummy Variables and Predictive Tests for Structural Change, *Economics Letters* 6: 241–247.

Dufour, J.-M. (1982) Recursive Stability Analysis of Linear Regression Relationships: An Exploratory Methodology, *Journal of Econometrics* 19: 31–76.

Durbin, J., and Watson, G.S. (1950) Testing for Serial Correlation in Least Squares Regression. I, *Biometrika* 37: 409–428.

Durbin, J., and Watson, G.S. (1951) Testing for Serial Correlation in Least Squares Regression. II, *Biometrika* 38: 159–178.

Edison, H.J. (1985) The Rise and Fall of Sterling: Testing Alternative Models of Exchange Rate Determination, *Applied Economics* 17: 1003–1021.

Edison, H.J. (1991) Forecast Performance of Exchange Rate Models Revisited, *Applied Economics* 23: 187–196.

Edison, H.J., and Klovland, J.T. (1987) A Quantitative Reassessment of the Purchasing Power Parity Hypothesis: Evidence from Norway and the United Kingdom, *Journal of Applied Econometrics* 2: 309–333.

Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50: 987–1007.

Engle, R.F., and Granger, C.W.J. (1987) Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica* 55: 251–276.

Engle, R.F., Hendry, D.F., and Richard, J.-F. (1983) Exogeneity, *Econometrica* 51: 277–304.

Ericsson, N.R., Campos, J., and Tran, H.-A. (1991) PC-GIVE and David Hendry's Econometric Methodology, International Finance Discussion Paper No. 406, Board of Governors of the Federal Reserve System, Washington, D.C.; forthcoming in the *Revista de Econometria*.

Ericsson, N.R., and Hendry, D.F. (1985) Conditional Econometric Modeling: An Application to New House Prices in the United Kingdom, In *A Celebration of Statistics: The ISI Centenary Volume* (A.C. Atkinson and S.E. Fienberg, Eds.). New York: Springer-Verlag. 251–285.

Fair, R.C. (1986) Evaluating the Predictive Accuracy of Models, In *Handbook of Econometrics, Volume 3* (Z. Griliches and M.D. Intriligator, Eds.). Amsterdam: North-Holland. 1979–1995.

Fisher, F.M. (1970) Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note, *Econometrica* 38: 361–366.

Fisher, R.A. (1922) The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients, *Journal of the Royal Statistical Society* 85: 597–612.

Goldfeld, S.M. (1973) The Demand for Money Revisited, *Brookings Papers on Economic Activity* 1973: 577–646 (with discussion).

Goldfeld, S.M., and Sichel, D.E. (1990) The Demand for Money, In *Handbook of Monetary Economics, Volume 1* (B.M. Friedman and F.H. Hahn, Eds.). Amsterdam: North-Holland. 299–356.

Godfrey, L.G. (1978) Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables, *Econometrica* 46: 1293–1301.

Granger, C.W.J. (1986) Developments in the Study of Cointegrated Economic Variables, *Oxford Bulletin of Economics and Statistics* 48: 213–228.

Granger, C.W.J. (1989) *Forecasting in Business and Economics, Second Edition*. Boston: Academic Press.

Granger, C.W.J., and Deutsch, M. (1992) Comments on the Evaluation of Policy Models, *Journal of Policy Modeling*, 14: 497–516.

Harvey, A.C. (1981) *The Econometric Analysis of Time Series*. Oxford: Philip Allan.

Hendry, D.F. (1979) Predictive Failure and Econometric Modelling in Macroeconomics: The Transactions Demand for Money, In *Economic Modelling* (P. Ormerod, Ed.). London: Heinemann Education Books. 217–242.

Hendry, D.F. (1983) Comment, *Econometric Reviews* 2: 111–114.

Hendry, D.F. (1986) The Role of Prediction in Evaluating Econometric Models, *Proceedings of the Royal Society of London, Series A* 407: 25–34 (with discussion).

Hendry, D.F. (1987) Econometric Methodology: A Personal Perspective, In *Advances in Econometrics, Volume 2* (T.F. Bewley, Ed.). Cambridge: Cambridge University Press. 29–48.

Hendry, D.F. (1988) The Encompassing Implications of Feedback versus Feedforward Mechanisms in Econometrics, *Oxford Economic Papers* 40: 132–149.

Hendry, D.F. (1989) *PC-GIVE: An Interactive Econometric Modelling System, Version 6.0/6.01*. Oxford: University of Oxford, Institute of Economics and Statistics and Nuffield College.

Hendry, D.F., and Ericsson, N.R. (1991) Modeling the Demand for Narrow Money in the United Kingdom and the United States, *European Economic Review* 35: 833–881.

Hendry, D.F., and Neale, A.J. (1991) A Monte Carlo Study of the Effects of Structural Breaks on Tests for Unit Roots, In *Economic Structural Change: Analysis and Forecasting* (P. Hackl and A.H. Westlund, Eds.). Berlin: Springer-Verlag. 95–119.

Hendry, D.F., and Richard, J.-F. (1982) On the Formulation of Empirical Models in Dynamic Econometrics, *Journal of Econometrics* 20: 3–33.

Jarque, C.M., and Bera, A.K. (1980) Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals, *Economics Letters* 6: 255–259.

Johnston, J. (1963) *Econometric Methods*. New York: McGraw-Hill.

Kiviet, J.F. (1986) On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships, *Review of Economic Studies* 53: 241–261.

Lu, M., and Mizon, G.E. (1991) Forecast Encompassing and Model Evaluation, In *Economic Structural Change: Analysis and Forecasting* (P. Hackl and A.H. Westlund, Eds.). Berlin: Springer-Verlag. 123–138.

Marquez, J., and Ericsson, N.R. (1992) Evaluating Forecasts of the U.S. Trade Balance, In *Empirical Evaluation of Alternative Policy Regimes* (R. Bryant, P. Hooper, C. Mann, and R. Tryon, Eds.). Washington, D.C.: Brookings Institution, in press.

Meese, R.A., and Rogoff, K. (1983) Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?, *Journal of International Economics* 14: 3–24.

Mizon, G.E., and Richard, J.-F. (1986) The Encompassing Principle and Its Application to Testing Non-nested Hypotheses, *Econometrica* 54: 657–678.

Nicholls, D.F., and Pagan, A.R. (1983) Heteroscedasticity in Models with Lagged Dependent Variables, *Econometrica* 51: 1233–1242.

Pagan, A. (1989) On the Role of Simulation in the Statistical Evaluation of Econometric Models, *Journal of Econometrics* 40: 125–139.

Pesaran, M.H. (1974) On the General Problem of Model Selection, *Review of Economic Studies* 41: 153–171.

Ramsey, J.B. (1969) Tests for Specification Errors in Classical Linear Least-squares Regression Analysis, *Journal of the Royal Statistical Society, Series B* 31: 350–371.

Sargan, J.D. (1958) The Estimation of Economic Relationships Using Instrumental Variables, *Econometrica* 26: 393–415.

Sargan, J.D. (1980) Some Approximations to the Distribution of Econometric Criteria Which Are Asymptotically Distributed as Chi-squared, *Econometrica* 48: 1107–1138.

Schinasi, G.J., and Swamy, P.A.V.B. (1989) The Out-of-Sample Forecasting Performance of Exchange Rate Models When Coefficients Are Allowed to Change, *Journal of International Money and Finance* 8: 375–390.

Spanos, A. (1986) *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Swamy, P.A.V.B., and Schinasi, G.J. (1989) Should Fixed Coefficients Be Re-estimated Every Period for Extrapolation?, *Journal of Forecasting* 8: 1–17.

Swamy, P.A.V.B., and Tinsley, P.A. (1980) Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients, *Journal of Econometrics* 12: 103–142.

White, H. (1980) A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48: 817–838.

White, H. (1989) *Specification Analysis in Econometrics*. Cambridge: Cambridge University Press, forthcoming.

Wilson, A.L. (1978) When Is the Chow Test UMP?, *American Statistician* 32: 66–68.