ENCOMPASSING THE FORECASTS OF U.S. TRADE BALANCE MODELS

Neil R. Ericsson and Jaime Marquez*

Abstract—Chong and Hendry (1986) propose the concept of *forecast encompassing*—the lack of additional information in another model's forecasts. The corresponding test statistic is based on the regression of one model's forecast errors on the other model's forecasts. This paper generalizes Chong and Hendry's statistic to include *sets* of *dynamic nonlinear* models with *uncertain* estimated coefficients generating *multi-step* forecasts with possibly *systematic* biases. Using stochastic simulation, the generalized statistic is applied to forecasts over 1985Q1–1987Q4 from six models of the U.S. merchandise trade balance, revealing mis-specification in all models.

I. Introduction

ONE of the most noteworthy features of the U.S. external imbalance of the 1980s has been the proliferation of associated trade balance forecasts. For example, Bryant, Holtham, and Hooper (1988, Annex) report forecasts from five well-known models often differing by $150 billion or more. This panorama of choices handicaps the development of policy responses and raises into question the empirical validity of the models themselves.

Dissimilar forecasts are common in economics; and two classes of procedures have been proposed for evaluating models via forecasts, with each class defined by the information set involved. For one, the information set is a *given* model's data across a *different* sample, with corresponding statistics including those of Chow (1960, pp. 594–595) and Hendry (1979) for testing predictive failure. For the other, the information set is *another* model's data across the *same* sample, with forecast encompassing and the comparison of root mean square forecast errors (RMSFEs) falling into this category.

Specifically, this second class compares forecasts from one model over a given period with another model's forecasts over the same period. Different models may capture different aspects of the dependent variable's behavior, in which case forecasts from one model may be informative in explaining the forecast *errors* from another. Conversely, *forecast encompassing*, or the *lack* of additional information in another model's forecasts, is evidence in favor of one's own model. Chong and Hendry (1986) propose this concept and clarify its relationship to the pooling (or combination) of forecasts. Ericsson (1992) shows that having the smallest RMSFE across a set of models is a necessary (but not sufficient) condition for forecast-encompassing those models. That is, forecast encompassing is a more stringent criterion than having the smallest RMSFE.

Thus, in an effort to evaluate trade balance forecasts more thoroughly, we test six econometric models of the U.S. trade balance for forecast encompassing. To emphasize the implications of different modeling strategies, our analysis considers four "structural" models and two time-series models. Moreover, the structural models differ in their dynamic specification, structural behavior, and estimation method.

Applying Chong and Hendry's forecast-encompassing test to the trade models poses technical problems. Their test was designed for static linear models, implying i.i.d. forecast errors from a correctly specified model. However, five of our trade models are nonlinear and dynamic, with (e.g.) multi-step forecasts from even a well-specified model being heteroscedastic and serially correlated. We generalize Chong and Hendry's statistic to address these complications, and also to account for systematic (constant) forecast biases, simultaneous comparison against several models, and coefficient uncertainty.

We find that all our models have highly significant forecast-encompassing test statistics in spite

Received for publication November 28, 1990. Revision accepted for publication December 17, 1991.

* Federal Reserve Board.

The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting those of the Board of Governors of the Federal Reserve System or other members of its staff. We are grateful to William Helkie for encouraging this project and providing the data for the Helkie-Hooper model, and to Ned Prescott and Lucia Foster for excellent research assistance. We also wish to thank Lisa Barrow, Carlo Bianchi, Giorgio Calzolari, Julia Campos, Hali Edison, Ray Fair, William Helkie, David Hendry, Peter Hooper, Fred Joutz, James MacKinnon, Andy Rose, Charlie Thomas, Ralph Tryon, and two anonymous referees for valuable comments and suggestions. All numerical results in this paper were obtained using TROLL Version 13; cf. Intex Solutions (1989).

of large confidence intervals for the forecasts themselves: each model's forecast errors are in part predictable by the forecasts (and so the data) of the other models. Alternatively viewed, the forecast performance of each model indicates room for improved model specification, and the statistics may suggest ways in which that respecification could occur.

Section II describes the class of models considered, the nature of Chong and Hendry's forecast-encompassing test statistic and our generalizations thereon, and the practical implementation of this statistic via stochastic simulation. Section III summarizes the structure of the trade balance models, their estimated coefficients, and their forecasts. Section IV analyzes the empirical results for the forecast-encompassing test statistic. Section V concludes.

II. The Theory of Forecast Encompassing

The question of interest is whether or not the forecast *errors* of a given model may be explained (at least in part) by the *forecasts* of another model. Under the null hypothesis of the first model (model h, say) being correctly specified, the forecast errors from that model are innovations, unpredictable by any information available at the time the forecasts were made, including the data of another model (model l, say) and hence the forecasts of that other model. If, in fact, model h's forecast errors are not predictable by model l's forecasts, then model h "forecast-encompasses" model l. That is, *given model h's forecasts*, model l's forecasts are redundant for predicting the variable of interest. However, if model l's forecasts *are* informative, model h is inadequate for forecasting because its forecast errors are predictable. Model l has useful information that model h does not have, or equivalently, relying on model h alone entails a loss of information.

Forecast encompassing establishes a partial ranking on models: if model h forecast encompasses model l, then the converse cannot be true, at least asymptotically. However, it could be that neither model forecast-encompasses the other, indicating room for improvement in both models.

This section considers how stochastically generated forecasts can help evaluate econometric models via forecast-encompassing tests. Section A describes the class of nonlinear dynamic models under consideration and their forecasts. Section B discusses the mechanics of forecast encompassing, with generalizations. Section C focuses on practical issues which arise in calculating forecasts for the test. For expositions on forecast encompassing, see Chong and Hendry (1986) and Ericsson (1992). On encompassing in general, see Mizon and Richard (1986).

A. *Forecasts from Nonlinear Dynamic Models*

For a given model, we assume the following possibly nonlinear reduced form:

$$y_{T+s} = g(y_{T+s-1}, z_{T+s}, \theta, u_{T+s}),$$
$$s = 1 - T, \ldots, -1, 0, 1, \ldots, S. \quad (1)$$

The function $g(\cdot)$ is an $n \times 1$ vector of equations, y_t is an $n \times 1$ vector of endogenous variables in period t, z_t is an $r \times 1$ vector of (assumedly) weakly exogenous variables, θ is a $c \times 1$ vector of unknown parameters, and u_t is an $m \times 1$ vector of disturbances. If the endogenous variables enter at additional lags, then y_t and its lags can be "stacked" to make (1) the companion form. The first T observations are used to estimate θ, the remaining S observations are forecast. The time index s is such that s is zero or negative for estimation and positive for forecasts. Often, our ultimate interest is in a subset of y_{T+s} or in a linear combination of y_{T+s}, so we introduce a selection matrix A that extracts the variable(s) of interest from $g(\cdot)$.

To forecast $A \cdot y_{T+s}$ s-steps ahead from time T, we use its conditional expectation, denoted μ_{T+s}:

$$\mu_{T+s} \equiv E\left[A \cdot g\left(y_{T+s-1}, z_{T+s}, \hat{\theta}, u_{T+s}\right) \mid (z_{T+s}, y_T)\right]$$
$$s = 1, \ldots, S. \quad (2)$$

$E[\cdot]$ is the expectation operator; and the expectation is taken conditional on (z_{T+s}, y_T), the information available at the time of forecasting. The future value of the exogenous variables z_{T+s} is explicitly assumed known. Uncertainty arises from the future shocks u_{T+s}, from y_{T+s-1} (because y_{T+s-1} is not included in the information set for $s > 1$), and from estimating rather than knowing θ. The last source of uncertainty is $O_p(T^{-1/2})$ and so is asymptotically negligible, but it can be substantial in finite samples. Equation (2) is the

s-step ahead forecast of $A \cdot y_{T+s}$. We also examine the one-step ahead forecast, which is obtained by replacing y_T in (2) with y_{T+s-1}.

The forecast errors associated with (2) are

$$\nu = (\nu_{T+1}, \ldots, \nu_{T+S})' = ([A \cdot y_{T+1} - \mu_{T+1}], \ldots, [A \cdot y_{T+S} - \mu_{T+S}])'. \quad (3)$$

These forecast errors are random because (1) is random, so it is of interest to calculate the errors' dispersion, e.g., as measured by their variance Φ, which is

$$\Phi = E[\nu\nu']. \quad (4)$$

The forecasts $\{\mu_{T+s}\}$, the corresponding forecast errors $\{\nu_{T+s}\}$, and their covariance matrix Φ all play central roles in the tests of forecast encompassing below.

B. *The Mechanics of Forecast Encompassing*

Tests of forecast encompassing are easy to calculate. Following Chong and Hendry (1986, pp. 676–679), regress the forecast errors of one model on the forecasts of another:

$$\nu^{(h)}_{T+s} = \kappa_l \mu^{(l)}_{T+s} + e^{(h)}_{T+s} \qquad s = 1, \ldots, S, \quad (5)$$

where $\nu^{(h)}_{T+s}$ $(\equiv A \cdot y_{T+s} - \mu^{(h)}_{T+s})$ is the (actual) forecast error of model h, $\mu^{(l)}_{T+s}$ is the forecast of model l $(l \neq h)$, and $e^{(h)}_{T+s}$ is the error of the regression. The forecast-encompassing statistic is the t-ratio on κ_l, and is $N(0,1)$ for large T and S when model h is correctly specified.

Chong and Hendry's derivation of the forecast-encompassing statistic assumes forecasts and forecast errors from static linear models with normal independent disturbances, and so models from which all forecasts are effectively one step ahead. We generalize Chong and Hendry's statistic, allowing for

(a) a constant term in (5);
(b) comparison against several models at once, rather than just one;
(c) model nonlinearity;
(d) multi-step ahead forecasts from dynamic models; and
(e) the uncertainty from estimating, rather than knowing, model coefficients.[1]

The regression accounting for (a) and (b) is

$$\nu^{(h)}_{T+s} = \kappa_0 + \Sigma_{l \neq h} \kappa_l \mu^{(l)}_{T+s} + e^{(h)}_{T+s}, \qquad s = 1, \ldots, S. \quad (6)$$

By allowing for a nonzero constant term (κ_0 not necessarily identically zero) and testing $\kappa_l = 0$, a more powerful forecast-encompassing test may result if the forecast errors of model h are systematically biased. Because such a systematic bias may be predictable (i.e., by a constant term), it is also of interest to test two additional hypotheses: $\kappa_0 = 0$ and $\kappa_l = 0$ jointly, and $\kappa_0 = 0$ given $\kappa_l \equiv 0$. Put slightly differently, exclusion of a constant term from (5) forces the regression through the origin. That restriction may be invalid, and its invalidity is of interest in itself.

Equation (6) generalizes (5) by including forecasts from several models [(b)]. Below, we calculate "pairwise" and "multiple-model" forecast-encompassing test statistics, with and without a constant term.

Model nonlinearity [(c)] implies heteroscedastic forecast errors, and s-step (rather than one-step) ahead forecasts [(d)] imply autocorrelation in the forecast errors. The most natural response is GLS estimation of (5) and (6), using $[\Phi^{(h)}]^{-1/2}$ as the correction factor. The matrix $\Phi^{(h)}$ is approximately diagonal for one-step ahead forecasts, in which case the standard heteroscedasticity transformation using the diagonal elements of $\Phi^{(h)}$ is suitable. However, here, and even with linear models, GLS has the advantage of correcting for the (asymptotically negligible) autocorrelation in the forecast errors arising from coefficient uncertainty [(e)].[2] Model nonlinearity also may induce

[1] After presenting an earlier version of this paper at the Brookings conference "Empirical Evaluation of Alternative Policy Regimes," March 8–9, 1990, we learned that independently Fisher and Wallis (1990) had proposed generalizations (a) and (b). However, the "forecasts" in their application are in-sample rather than out-of-sample, so their calculated statistics are *not* forecast-encompassing, but are akin to Davidson and MacKinnon's (1981) C statistic; cf. Fisher and Wallis (1990, pp. 193–194).

[2] Less efficient but valid procedures are available. For instance, Fair and Shiller (1989) estimate equations like (5) and (6) by OLS, and for inference calculate standard errors that suitably adjust for heteroscedasticity and autocorrelation. For (5) and (6), OLS is generally less efficient than GLS, thereby affecting the power of OLS-based tests. Given the high autocorrelation of multi-step ahead forecast errors, the loss in power is probably substantial. Furthermore, *feasible* GLS is simple to implement computationally: see below.

non-normality in forecast errors, but that should not affect the asymptotic (large S) distribution of the forecast-encompassing statistic except for "extreme" non-normality; see Gallant and White (1988). In any event, the observed distribution of simulated forecasts can be scrutinized.

Forecast-encompassing tests with one-step-ahead forecasts and those with s-step ahead forecasts may have different power because the alternative hypotheses being tested are different. That is most easily seen in testing $\kappa_0 = 0$ given $\kappa_l \equiv 0$ with forecasts from a linear model, ignoring coefficient uncertainty for expositional convenience. Whether the regression uses one-step or s-step ahead forecasts, the dependent variable is the same, once rescaled by the appropriate $[\Phi^{(h)}]^{-1/2}$. For one-step ahead forecasts, the regressor is a constant term (aside from a scale factor) because the one-step ahead $\Phi^{(h)}$ matrix is diagonal with the same values on the diagonal. However, for s-step ahead forecasts, the regressor is $[\Phi^{(h)}]^{-1/2} \cdot \iota$ where ι is a vector with unit elements, and the *s-step* ahead $\Phi^{(h)}$ matrix is inherently non-diagonal if the model is dynamic. Thus, the dependent variables are the same but the regressors differ, so the alternative hypotheses are in general different. Likewise, in the more general regressions, the transformed forecasts of another model $[\Phi^{(h)}]^{-1/2} \cdot \mu^{(l)}$ are not invariant to being one-step or s-step ahead.

C. *Stochastically Generated Forecasts*

In practice, Φ and $\{\mu_{T+s}\}$ are themselves unknown, and obtaining them involves multiple integration of the entire model. For instance,

$$\begin{aligned}\mu_{T+s} &\equiv A \cdot E(y_{T+s}) \\ &= A \cdot \int\int g\left(\hat{y}_{T+s-1}, z_{T+s}, \hat{\theta}, u_{T+s}\right) \\ &\quad \cdot f(\hat{\theta}, u) \cdot d\hat{\theta} \cdot du, \end{aligned} \tag{7}$$

where $f(\cdot, \cdot)$ is the joint density of $\hat{\theta}$ and u $[\equiv (u_{T+1}, \ldots, u_{T+s})']$, $\hat{y}_{T+s-1}$ is implicitly a function of $\hat{\theta}$ and u, and we omit the model superscript h or l because the formula is generic. Ideally, we would solve (7) and its counterpart for Φ analytically and evaluate the resulting formulae at $\hat{\theta}$. Chow (1960) solves these formulae for the linear, static, single-equation model. Schmidt (1974) and Baillie (1979) derive easily calculated (approximate) formulae for linear dynamic systems. However, (1) is nonlinear, so neither solution is directly applicable.

Even so, (7) may be approximated by either numerical integration or Monte Carlo methods. Numerical integration is generally not feasible for model-based forecasts because of the large number of random variables. So, we solve a stochastic analogue to the analytical formula implicit in (7), simulating by Monte Carlo the effects of inherent and coefficient uncertainty. Thus, the remainder of this section discusses how the Monte Carlo estimates of Φ and $\{\mu_{T+s}\}$ are calculated. See Bianchi, Calzolari, and Brillet (1987) and Brown and Mariano (1989) for general expositions on stochastically simulated forecasts and Hendry (1984) on Monte Carlo methods.

The principle and implementation of the Monte Carlo procedures are simple and intuitive. Taking (7) again, for each random variable in that equation, we draw a random number from the distribution which that random variable is supposed to follow. The corresponding set of random numbers is substituted into $g(\cdot)$, and a simulated forecast is obtained. This procedure is repeated many times (i.e., many replications), and the resulting simulated forecasts are averaged to obtain an estimated mean forecast. That averaging precisely parallels the integration of $g(\cdot)$ in (7). A similar procedure follows for estimating Φ.

In detail, the random variables in (7) are u and $\hat{\theta}$. By assumption, we have

$$u_t \sim NI(0, \Omega) \qquad t = 1, \ldots, T+S, \tag{8}$$

where Ω is an $m \times m$ covariance matrix independent of the time period t. Likewise, our estimation procedures imply

$$\hat{\theta} \sim N(\theta, \Sigma) \tag{9}$$

asymptotically, where Σ is the $c \times c$ asymptotic covariance matrix of $\hat{\theta}$.[3] Thus, we draw random numbers u_k $[\equiv (u_{k,T+1}, \ldots, u_{k,T+s})']$ and θ_k from the distributions $NI(0, I_s \otimes \hat{\Omega})$ and $N(\hat{\theta}, \hat{\Sigma})$,

[3] For regularity conditions, see Gallant and White (1988, chapter 5).

respectively. The subscript k indicates the replication number of K possible replications, and circumflexes denote empirical estimates, which replace the (unknown) population values in the distribution functions. The s-step ahead simulated forecast for the k^{th} replication is

$$y_{k,T+s} = g(y_{k,T+s-1}, z_{T+s}, \theta_k, u_{k,T+s}),\quad s = 1,\ldots,S, \tag{10}$$

where, for $s > 1$, (10) is solved forward sequentially via (1) with the initial condition that $y_{k,T} \equiv y_T$.

The mean forecasts $\{\mu_{T+s}\}$, mean forecast errors $\{\nu_{T+s}\}$, and variance-covariance matrix of the latter (Φ) are estimated by their corresponding Monte Carlo sample moments:

$$\tilde{\mu}_{T+s} = \Sigma_k A \cdot y_{k,T+s}/K, \qquad s = 1,\ldots,S, \tag{11}$$

$$\tilde{\nu}_{T+s} = A \cdot y_{T+s} - \tilde{\mu}_{T+s}, \qquad s = 1,\ldots,S, \tag{12}$$

and

$$\tilde{\Phi} = \Sigma_k \phi_k \phi_k'/K, \tag{13}$$

where ϕ_k is the vector of deviations between the k^{th} stochastic simulation of the S forecasts and the estimated mean forecasts:

$$\phi_k = ([A \cdot y_{k,T+1} - \tilde{\mu}_{T+1}],\ldots,[A \cdot y_{k,T+S} - \tilde{\mu}_{T+S}])'. \tag{14}$$

Monte Carlo estimators are denoted by tilde superscripts in order to distinguish them from *empirical* estimators such as $\hat{\theta}$. Thus, in practice, the mean forecast errors used in (5) and (6) are $\{\tilde{\nu}_{T+s}^{(h)}\}$, which are $\{A \cdot y_{T+s} - \tilde{\mu}_{T+s}^{(h)}\}$. The averaging over replications in (11), (12), and (13) parallels the averaging over the population implicit in the expectations operator in (2), (3), and (4).

III. Forecasts of the U.S. Trade Balance

This section summarizes the form of the four structural and two time-series models (section A) and their estimated coefficients' and forecasts' properties (section B). The data are quarterly for the last two decades. Marquez and Ericsson (1990, appendix E; 1993) describe the data, the models, and their forecasts in detail.

A. *Model Structure*

Equations (15)–(19) characterize the four structural models.

$$\ln(P_{xt}) = P_x(P_t, E_t, P_t^*) + upx_t \tag{15}$$
$$\ln(X_t) = X(P_{xt}/(E_t \cdot P_t^*), Y_t^*) + ux_t \tag{16}$$
$$\ln(P_{mt}) = P_m(E_t, P_t^*, P_{xt}) + upm_t \tag{17}$$
$$\ln(M_t) = M(P_{mt}/P_t, Y_t) + um_t \tag{18}$$
$$NX_t = P_{xt} \cdot X_t - P_{mt} \cdot M_t. \tag{19}$$

The variables P_m, P_x, and P are the prices of imports and exports and the general price level; E is the nominal exchange rate (domestic/foreign); M, X, NX, and Y are the volume of imports, the volume of exports, the (nominal) trade balance, and real income, respectively; and the absence or presence of an asterisk denotes a variable measured for the domestic (i.e., U.S.) or foreign country. Equations (15)–(19) are the explicit representation of (1) above, with the selection matrix A extracting NX_t from the vector of left-hand side variables. Each model assumes that the vector of errors $(upx_t, ux_t, upm_t, um_t)'$, which is u_t, is jointly normally and independently distributed as in (8), conditional upon a given dynamic structure for (15)–(18).

Empirical implementation of (15)–(19) requires specification of exogeneity, dynamics, estimation technique, and level of trade disaggregation. Different choices result in the following models. Model M1 is the econometric model of multilateral, commodity-disaggregated U.S. trade in Helkie and Hooper (1988), which is regarded as a standard in the literature. The remaining three structural models are variations on M1. Model M2 allows for Ω to be non-diagonal (M1 does not). Model M3 generalizes M1 further by allowing for richer dynamics. Model M4 is Marquez's (1989) bilateral trade model, which allows for non-diagonal Ω, richer dynamics, and disaggregation across countries (but not across commodities).

The time-series models M5 and M6 serve as benchmarks for the dynamic specification and forecasting performance of models M1–M4. Model M5 is a fourth-order, four-variable vector autoregression (VAR) for the *logarithms* of export and import volumes and price indices. The trade balance is constructed via (19). Model M6 is a first-order univariate autoregression [AR(1)] in the level of the trade balance.

B. Model Estimation and Forecasts

This subsection summarizes estimation and forecast results of the six models.

The estimated elasticities for the structural models are consistent with standard trade theory; see Marquez and Ericsson (1993). Thus, the "economic plausibility" of these estimated models does not help discriminate between them, so we turn to forecast performance.

Figure 1 graphs the *explanatory* variables of the forecast-encompassing regressions, which are the estimated mean forecasts $\{\tilde{\mu}_{T+s}\}$ of the six models for one-step and s-step ahead forecasts. The forecast period is 1985Q1–1987Q4 ($S = 12$) with $K = 1000$ replications. The dispersion of forecasts across models is considerable: up to \$60 billion for one-step ahead forecasts and as much as \$140 billion for s-step ahead forecasts. Such differences in forecasts imply the potential for one model's forecasts explaining a substantial quantity of another model's forecast errors.

While the *dependent* variables for the forecast-encompassing regressions are implicit in figure 1, it is fruitful to examine them explicitly. Figure 2 presents the estimated one-step ahead mean forecast errors $\{\tilde{\nu}_{T+s}\}$, model by model, with approximate (one-off) 95% confidence intervals given by $0 \pm 2\sqrt{\tilde{\Phi}_{SS}}$, where $\tilde{\Phi}_{SS}$ is the s^{th} diagonal element of $\tilde{\Phi}$. Figure 3 contains the corresponding information for s-step ahead forecasts.

Three general features are evident. First, the forecast errors typically lie within the estimated confidence regions. Unsurprisingly, tests of forecast accuracy are not rejected; see Marquez and Ericsson (1993). Still, the errors often appear systematic, even for the one-step ahead errors, which should be innovations under the hypothesis of correct specification.

Second, the trade balance forecasts are subject to a wide margin of error, even in the short run. The smallest one-step ahead confidence interval is $\pm$\$20 billion, and that is for the AR(1) model. The s-step ahead confidence intervals often are larger than the realized trade balance itself. The s-step ahead confidence intervals for models M1–M4 are large in spite of taking "future" production costs, degree of rationing, trend factors, real incomes, GDP deflators, and *nominal exchange rates* as known. Standard errors for comparable ex ante forecasts of the trade balance would almost invariably be larger, perhaps similar to those for the (ex ante) time-series forecasts from models M5 and M6.

Third, forecast uncertainty is time-dependent, with both dynamics and model nonlinearity as contributing factors. This implies the importance of accounting for autocorrelation and heteroscedasticity in the forecast-encompassing regressions, and large potential gains in efficiency from using GLS rather than OLS. Although not monotonic, the forecast variance tends to increase with the forecast horizon, particularly so for s-step ahead forecasts.

IV. Forecast Encompassing and Models of the U.S. Trade Balance

Using the forecasts for 1985Q1–1987Q4, we apply feasible GLS to (5) and (6) and test for the significance of the associated coefficient estimates. In spite of large confidence intervals in figures 2 and 3, the forecast-encompassing tests indicate that forecast errors from each model are predictable by other models' forecasts. So, these tests caution against the use of any of these models for forecasting and policy simulation (i.e., counter-factual analyses). This section describes these results and their implications; additional results appear in Marquez and Ericsson (1990).

A. Empirical Results

Tables 1 and 2 report the test statistics for one-step and s-step ahead forecasts, respectively. Looking at the pairwise tests on table 1 first, the mean predictions from the univariate time-series model M6 help explain the mean forecast errors of the structural model M4 when a constant term is included [$F(1,10) = 14.3$]. That suggests dynamic mis-specification of M4 because dynamics are the only information in M6. However, M6 cannot forecast-encompass M5 [$F(1,10) = 4.9$]. Further, models M1–M3 cannot encompass M4, perhaps due to aggregation over countries [$F(1,10) = 5.7, 8.5, 6.5$]. The multiple-model tests reject models M1–M4, and nearly reject M5 [$F(5,6) = 5.3, 5.3, 5.5, 5.0, 3.4$].[4]

[4] Generally, the simulated forecasts appear normally distributed, justifying critical values from the F distribution; see Marquez and Ericsson (1993, appendix C).

FIGURE 1.—FORECASTS OF THE TRADE BALANCE

Figure 1a: one-step ahead forecasts

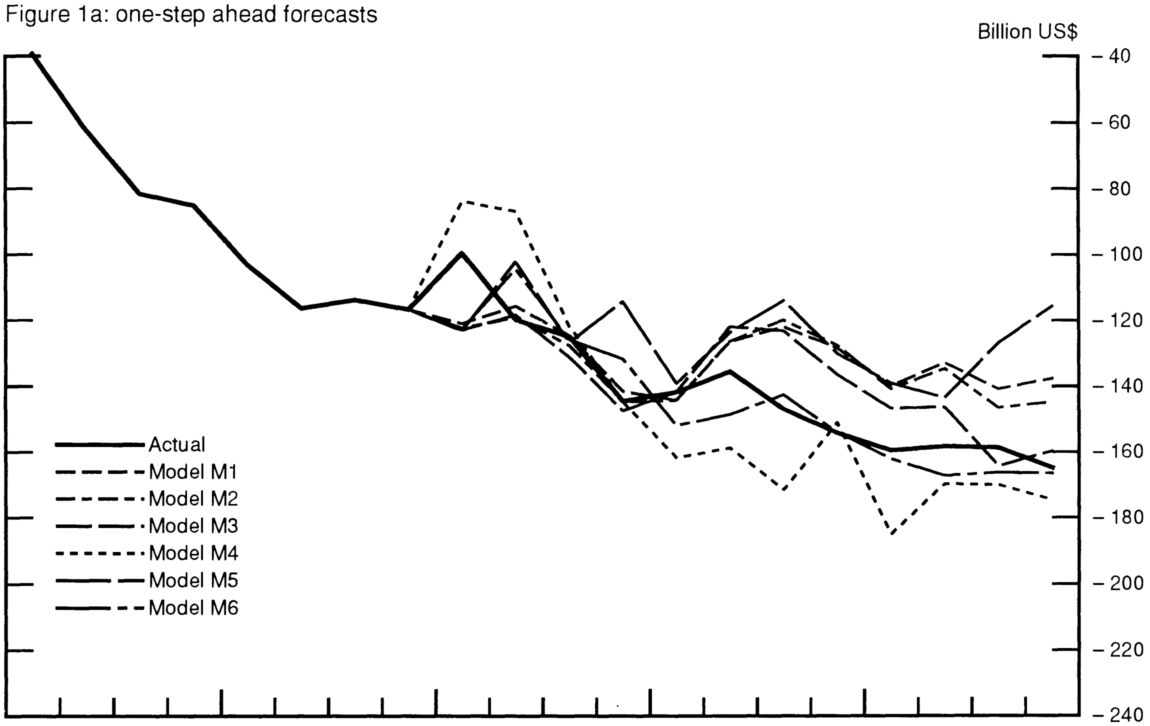


Figure 1b: s-step ahead forecasts

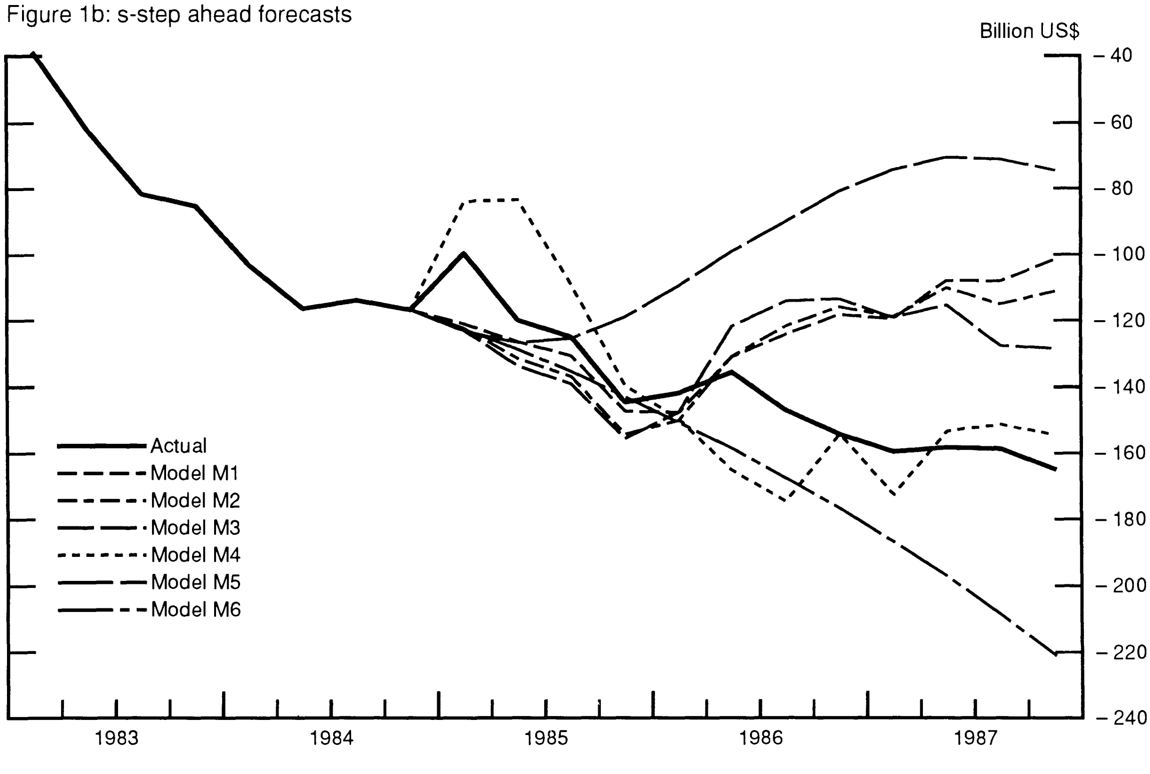


FIGURE 2.—95% CONFIDENCE INTERVALS FOR TRADE-BALANCE FORECAST ERRORS: ONE-STEP AHEAD

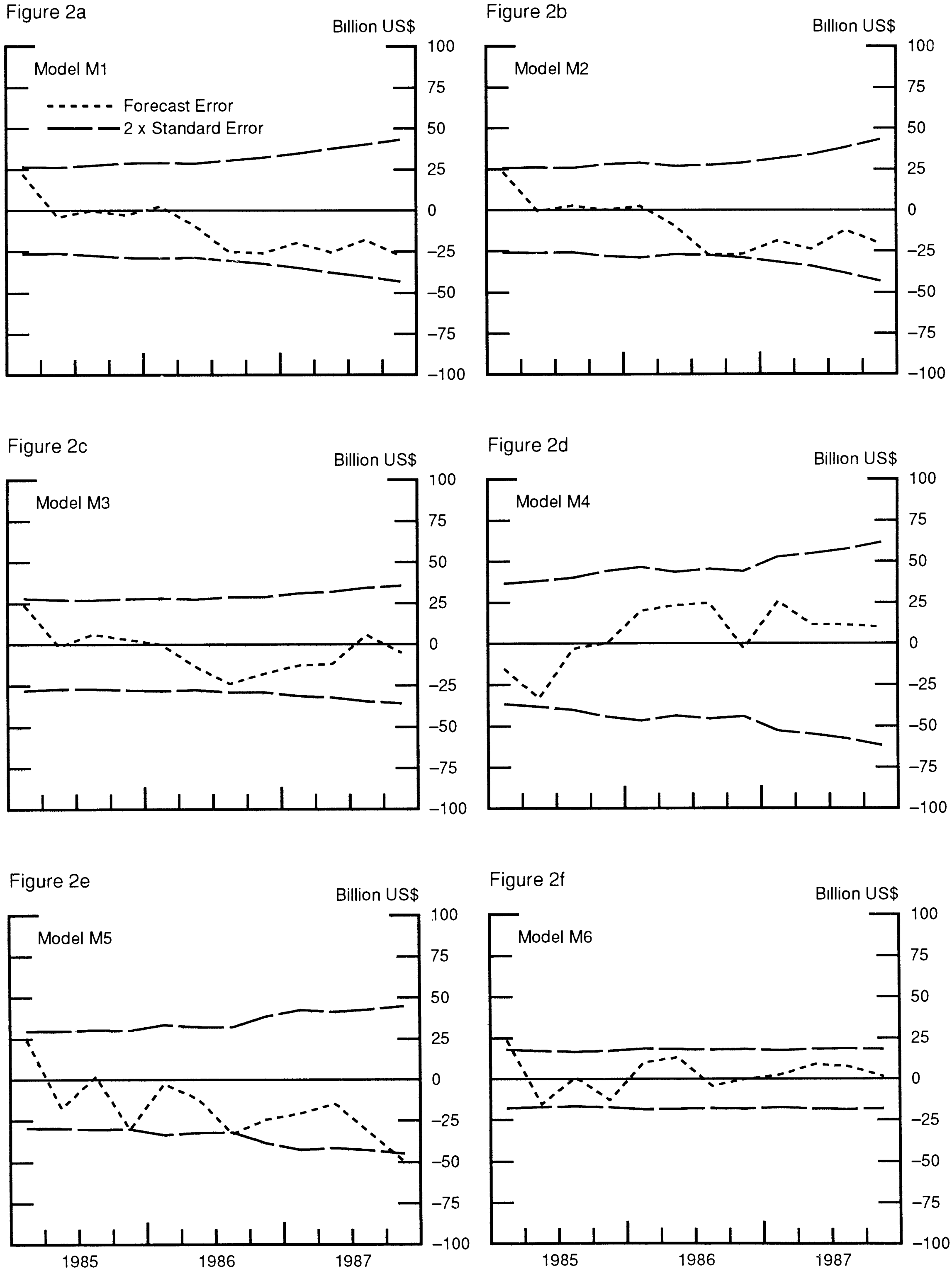
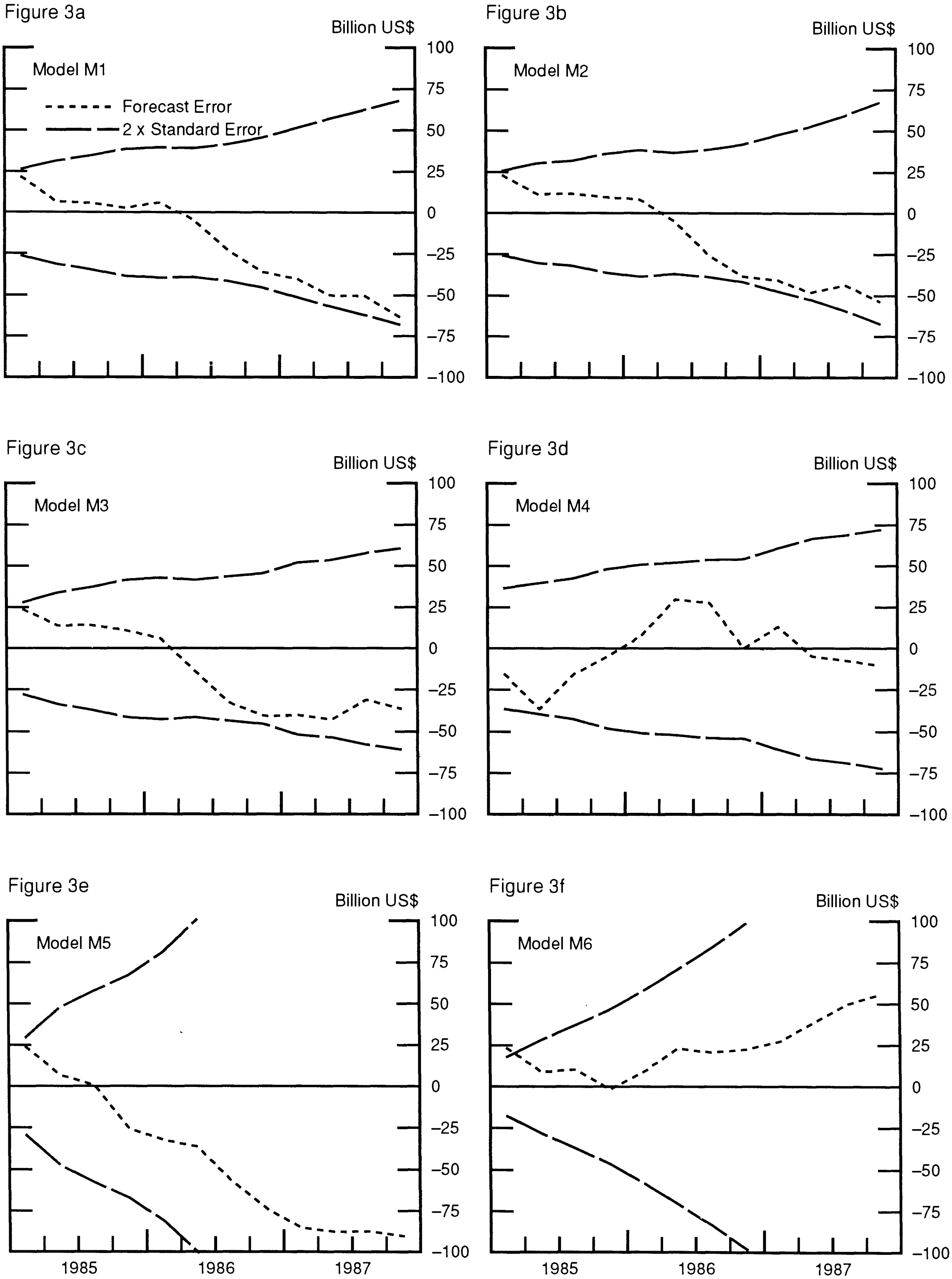


FIGURE 3.—95% CONFIDENCE INTERVALS FOR TRADE-BALANCE FORECAST ERRORS: *s*-STEP AHEAD



Note: The confidence intervals for Figures 3e and 3f are omitted after 1986Q2 and 1986Q4 respectively because they exceed ±100 billion US$.

TABLE 1.—FORECAST-ENCOMPASSING TEST STATISTICS FOR ALTERNATIVE TRADE-BALANCE MODELS ONE-STEP AHEAD

	F-values of Pairwise Forecast-encompassing Test Statistics (tail probabilities in parentheses)								Multiple-model ($\forall l \neq h$) Forecast-encompassing Test Statistics (tail probabilities in parentheses)		
Encom-passing Model			Model To Be Encompassed (l)								
(h)	Null Hypothesis	d.f.	M1	M2	M3	M4	M5	M6	Null Hypothesis	d.f.	F-value
M1	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.0 (0.96)	0.0 (0.92)	0.0 (0.97)	0.7 (0.41)	0.0 (0.96)	0.1 (0.74)	$\forall\kappa_l = 0; \kappa_0 \equiv 0$	5, 7	2.8 (0.11)
	$\kappa_l = 0$	1, 10		0.1 (0.77)	0.2 (0.69)	5.7 (0.04)	0.0 (0.96)	2.6 (0.14)	$\forall\kappa_l = 0$	5, 6	5.3 (0.03)
	$\kappa_l = \kappa_0 = 0$	2, 10		0.0 (0.96)	0.1 (0.92)	2.8 (0.10)	0.0 (1.00)	1.3 (0.31)	$\forall\kappa_l = \kappa_0 = 0$	6, 6	4.4 (0.05)
M2	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.0 (0.91)	0.0 (0.90)	0.0 (0.85)	1.6 (0.24)	0.0 (0.87)	0.4 (0.56)	$\forall\kappa_l = 0; \kappa_0 \equiv 0$	5, 7	3.1 (0.09)
	$\kappa_l = 0$	1, 10	0.0 (0.91)		0.1 (0.77)	8.5 (0.02)	0.0 (0.86)	3.9 (0.08)	$\forall\kappa_l = 0$	5, 6	5.3 (0.03)
	$\kappa_l = \kappa_0 = 0$	2, 10	0.0 (0.99)		0.1 (0.95)	4.2 (0.05)	0.0 (0.98)	2.0 (0.19)	$\forall\kappa_l = \kappa_0 = 0$	6, 6	4.4 (0.05)
M3	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.0 (0.85)	0.0 (0.90)	0.1 (0.83)	1.1 (0.31)	0.1 (0.81)	0.3 (0.60)	$\forall\kappa_l = 0; \kappa_0 \equiv 0$	5, 7	3.1 (0.08)
	$\kappa_l = 0$	1, 10	0.1 (0.82)	0.4 (0.55)		6.5 (0.03)	0.0 (0.89)	1.9 (0.20)	$\forall\kappa_l = 0$	5, 6	5.5 (0.03)
	$\kappa_l = \kappa_0 = 0$	2, 10	0.1 (0.95)	0.2 (0.81)		3.3 (0.08)	0.0 (0.97)	1.0 (0.41)	$\forall\kappa_l = \kappa_0 = 0$	6, 6	4.6 (0.04)
M4	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.0 (0.85)	0.1 (0.81)	0.1 (0.80)	0.2 (0.67)	0.0 (0.91)	0.0 (0.84)	$\forall\kappa_l = 0; \kappa_0 \equiv 0$	5, 7	2.0 (0.20)
	$\kappa_l = 0$	1, 10	4.6 (0.06)	2.0 (0.18)	1.0 (0.35)		3.5 (0.09)	14.3 (0.00)	$\forall\kappa_l = 0$	5, 6	5.0 (0.04)
	$\kappa_l = \kappa_0 = 0$	2, 10	2.4 (0.14)	1.1 (0.36)	0.6 (0.58)		1.9 (0.20)	7.4 (0.01)	$\forall\kappa_l = \kappa_0 = 0$	6, 6	4.2 (0.05)
M5	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.3 (0.61)	0.3 (0.59)	0.5 (0.50)	1.4 (0.26)	0.2 (0.67)	0.4 (0.54)	$\forall\kappa_l = 0; \kappa_0 \equiv 0$	5, 7	2.7 (0.11)
	$\kappa_l = 0$	1, 10	0.3 (0.57)	0.6 (0.46)	1.6 (0.23)	3.5 (0.09)		0.6 (0.45)	$\forall\kappa_l = 0$	5, 6	3.4 (0.09)
	$\kappa_l = \kappa_0 = 0$	2, 10	0.3 (0.77)	0.4 (0.69)	0.9 (0.43)	1.9 (0.20)		0.4 (0.68)	$\forall\kappa_l = \kappa_0 = 0$	6, 6	2.9 (0.11)
M6	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.0 (0.99)	0.0 (0.99)	0.0 (0.98)	0.0 (0.97)	0.1 (0.75)	0.0 (0.98)	$\forall\kappa_l = 0; \kappa_0 \equiv 0$	5, 7	0.7 (0.64)
	$\kappa_l = 0$	1, 10	0.1 (0.76)	0.0 (0.90)	0.0 (0.96)	0.0 (0.87)	4.9 (0.05)		$\forall\kappa_l = 0$	5, 6	0.9 (0.56)
	$\kappa_l = \kappa_0 = 0$	2, 10	0.0 (0.95)	0.0 (0.99)	0.0 (1.00)	0.0 (0.99)	2.5 (0.14)		$\forall\kappa_l = \kappa_0 = 0$	6, 6	0.7 (0.65)

Notes:

a. The regressions for pairwise and multiple-model forecast encompassing are

$$\tilde{\nu}^{(h)}_{T+s} = \kappa_0 + \kappa_l \tilde{\mu}^{(l)}_{T+s}$$

and

$$\tilde{\nu}^{(h)}_{T+s} = \kappa_0 + \Sigma_{l \neq h} \kappa_l \tilde{\mu}^{(l)}_{T+s},$$

respectively, where the l^{th} model(s) is (are) being encompassed by the h^{th} model (the encompassing model). Each regression uses GLS, i.e., premultiplying each equation by $[\tilde{\Phi}^{(h)}]^{-1/2}$ in order to account for the autocorrelation and heteroscedasticity of the dependent variable $\tilde{\nu}^{(h)}_{T+s}$.

b. The diagonal elements in the block of pairwise forecast-encompassing test statistics are statistics testing $\kappa_0 = 0$ with $\kappa_l \equiv 0$.

c. "d.f." denotes the degrees of freedom for the corresponding F-statistic.

d. Ericsson (1992) modifies the forecast-encompassing statistic to acount for integrated and cointegrated variables. Statistically, the trade balance appears $I(1)$, so tables 1 and 2 were re-calculated with this new ("forecast-differential encompassing") statistic, and stronger rejections usually resulted.

TABLE 2.—FORECAST-ENCOMPASSING TEST STATISTICS FOR ALTERNATIVE TRADE-BALANCE MODELS
s-STEP AHEAD

	F-values of Pairwise Forecast-encompassing Test Statistics (tail probabilities in parentheses)								Multiple-model ($\forall l \neq h$) Forecast-encompassing Test Statistics (tail probabilities in parentheses)		
			Model To Be Encompassed (l)								
Encom-passing Model (h)	Null Hypothesis	d.f.	M1	M2	M3	M4	M5	M6	Null Hypothesis	d.f.	F-value
M1	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	1.4	2.2	2.0	0.1	3.5	0.1	$\forall \kappa_l = 0; \kappa_0 \equiv 0$	5, 7	5.9
			(0.27)	(0.16)	(0.19)	(0.72)	(0.09)	(0.76)			(0.02)
	$\kappa_l = 0$	1, 10		2.1	1.0	1.7	14.5	33.8	$\forall \kappa_l = 0$	5, 6	5.0
				(0.18)	(0.35)	(0.22)	(0.00)	(0.00)			(0.04)
	$\kappa_l = \kappa_0 = 0$	2, 10		1.8	1.2	1.6	8.7	19.6	$\forall \kappa_l = \kappa_0 = 0$	6, 6	4.8
				(0.22)	(0.35)	(0.25)	(0.01)	(0.00)			(0.04)
M2	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	2.6	1.5	2.6	0.1	4.3	0.1	$\forall \kappa_l = 0; \kappa_0 \equiv 0$	5, 7	7.7
			(0.14)	(0.24)	(0.14)	(0.76)	(0.06)	(0.73)			(0.01)
	$\kappa_l = 0$	1, 10	3.3		2.5	2.6	27.1	32.1	$\forall \kappa_l = 0$	5, 6	5.8
			(0.10)		(0.14)	(0.14)	(0.00)	(0.00)			(0.03)
	$\kappa_l = \kappa_0 = 0$	2, 10	2.6		2.1	2.1	16.1	18.9	$\forall \kappa_l = \kappa_0 = 0$	6, 6	5.6
			(0.12)		(0.17)	(0.17)	(0.00)	(0.00)			(0.03)
M3	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	1.4	1.6	0.8	0.1	3.0	0.0	$\forall \kappa_l = 0; \kappa_0 \equiv 0$	5, 7	6.3
			(0.26)	(0.23)	(0.40)	(0.80)	(0.11)	(0.96)			(0.02)
	$\kappa_l = 0$	1, 10	2.2	3.2		5.4	31.5	14.7	$\forall \kappa_l = 0$	5, 6	5.0
			(0.17)	(0.10)		(0.04)	(0.00)	(0.00)			(0.04)
	$\kappa_l = \kappa_0 = 0$	2, 10	1.5	2.1		3.2	17.2	8.2	$\forall \kappa_l = \kappa_0 = 0$	6, 6	4.5
			(0.26)	(0.18)		(0.08)	(0.00)	(0.01)			(0.04)
M4	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	0.6	0.8	1.2	1.0	1.8	0.6	$\forall \kappa_l = 0; \kappa_0 \equiv 0$	5, 7	3.8
			(0.44)	(0.38)	(0.29)	(0.34)	(0.21)	(0.47)			(0.06)
	$\kappa_l = 0$	1, 10	0.8	0.1	0.4		1.9	0.5	$\forall \kappa_l = 0$	5, 6	3.2
			(0.39)	(0.74)	(0.55)		(0.20)	(0.50)			(0.10)
	$\kappa_l = \kappa_0 = 0$	2, 10	0.9	0.5	0.7		1.5	0.7	$\forall \kappa_l = \kappa_0 = 0$	6, 6	3.0
			(0.44)	(0.61)	(0.54)		(0.27)	(0.51)			(0.11)
M5	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	3.7	3.4	3.1	1.6	5.4	3.2	$\forall \kappa_l = 0; \kappa_0 \equiv 0$	5, 7	2.3
			(0.08)	(0.09)	(0.10)	(0.24)	(0.04)	(0.10)			(0.16)
	$\kappa_l = 0$	1, 10	3.7	5.2	6.0	4.6		19.1	$\forall \kappa_l = 0$	5, 6	6.6
			(0.08)	(0.04)	(0.04)	(0.06)		(0.00)			(0.02)
	$\kappa_l = \kappa_0 = 0$	2, 10	5.2	6.3	6.9	5.8		16.7	$\forall \kappa_l = \kappa_0 = 0$	6, 6	8.7
			(0.03)	(0.02)	(0.01)	(0.02)		(0.00)			(0.01)
M6	$\kappa_l = 0; \kappa_0 \equiv 0$	1, 11	2.4	2.1	1.8	1.5	4.4	5.2	$\forall \kappa_l = 0; \kappa_0 \equiv 0$	5, 7	2.1
			(0.15)	(0.18)	(0.20)	(0.24)	(0.06)	(0.04)			(0.19)
	$\kappa_l = 0$	1, 10	8.7	8.8	6.8	0.6	0.4		$\forall \kappa_l = 0$	5, 6	1.1
			(0.02)	(0.01)	(0.03)	(0.44)	(0.52)				(0.44)
	$\kappa_l = \kappa_0 = 0$	2, 10	8.8	8.8	7.4	2.9	2.7		$\forall \kappa_l = \kappa_0 = 0$	6, 6	1.8
			(0.01)	(0.01)	(0.01)	(0.10)	(0.12)				(0.24)

Notes: See notes a–d in table 1.

The *s-step* ahead tests in table 2 roughly parallel those for one-step ahead forecasts, but often with numerically larger values of the test statistics. All models but M4 fail the pairwise tests, and all but M4 and M6 fail the multiple-model tests. The ability of the mean predictions from the structural models M1–M4 to help explain the mean forecast errors of each of the time-series models suggests that the latter lack important economic determinants.[5]

[5] The failure of the time-series models to forecast-encompass models M1–M4 may also reflect the larger information set of the latter, notably including future values of assumedly strongly exogenous variables.

B. *Econometric Implications*

In addition to its implications for trade modeling, our analysis raises issues concerning the application of forecast-encompassing tests in general. First, pairwise and multiple-model tests could differ even asymptotically, with rejection by multiple-model tests being "correct" if that rejection occurs. That brings home the importance of modeling and testing from general to simple, rather than the reverse; cf. Hendry (1979).

Second, forecast-encompassing tests may suggest the mis-specification present and hence ways in which to improve a model. Specifically, tables 1–2 imply that every model contains information valuable for improving the other models, and that every model could be improved. Reliance on any given model involves a loss of information relative to the other models, implying that better forecasts are obtainable from the data currently used in modeling. For instance, rejection of model M1 using s-step ahead forecasts from the VAR suggests the need for better dynamic specification in model M1. More generally, further disaggregation may be required to account for (e.g.) trade in computers. Still, such inferences are speculative rather than definitive because rejection of the null does not imply the alternative. Also, to assess the *magnitude* of the implied reduction in forecast uncertainty, we must await the development of an improved model.

Third, the numerous rejections from forecast-encompassing tests contrast with little evidence of mis-specification via tests of predictive failure *over the same sample period*; cf. Marquez and Ericsson (1993). This apparent incongruity arises because the two types of tests differ in their information sets (as noted in the introduction) and in their measures of uncertainty. Tests of predictive failure employ the confidence intervals of a given model's forecasts to measure the uncertainty of those forecasts. Forecast-encompassing tests employ a measure of uncertainty derived from regression (6). Because the two measures of uncertainty are obtained under different information sets, those measures need not be equal, and so the tests' outcomes may differ.

Fourth, notwithstanding our computations with both one-step and s-step ahead forecasts, we agree with Pagan (1989) that analyses of one-step ahead residuals, s-step ahead residuals, and fitted residuals are conceptually equivalent for linear models, provided that his implicit assumption of strong exogeneity is valid; see Chong and Hendry (1986, p. 681) and Engle, Hendry, and Richard (1983). Indeed, several caveats to this equivalence justify the multiple calculations that we undertook. From a practical standpoint, only one type of residual may be available, in which case it is useful to have equivalent procedures for analyzing all three types. Also, forecast-encompassing statistics are not invariant for even linear models, as shown in section IIB. Finally, if strong exogeneity is invalid, forecast-based test statistics calculated with the s-step ahead residuals *assuming* strong exogeneity are no longer the correct s-step ahead forecast test statistics, even though the former as (incorrectly) calculated are still numerically equivalent to the one-step ahead statistics. The proper s-step statistics would require modeling the feedback mechanism of the endogenous variables onto the right-hand side *weakly* exogenous variables.

V. Summary and Conclusions

This paper generalizes Chong and Hendry's (1986) test of forecast encompassing to account for systematic (constant) forecast biases, simultaneous comparison against several models, and the statistical effects of multi-step ahead forecasts, model nonlinearity, and coefficient uncertainty. The test statistics are easily implemented with stochastic simulation, which we apply to six econometric models of the U.S. trade balance over 1985Q1–1987Q4. All these models fail the forecast-encompassing tests, so each model's forecast errors are in part explained by the data generating the forecasts of competing models. These failures also indicate the potential for model improvement. Thus, the predictive performance of these models has implications for econometric practice, trade modeling, and the role of these models in policy making.

REFERENCES

Baillie, Richard T., "Asymptotic Prediction Mean Squared Error for Vector Autoregressive Models," *Biometrika* 66 (3) (1979), 675–678.

Bianchi, Carlo, Giorgio Calzolari, and Jean-Louis Brillet, "Measuring Forecast Uncertainty: A Review with Evaluation Based on a Macro Model of the French Economy," *International Journal of Forecasting* 3 (2)

(1987), 211–227.
Brown, Bryan W., and Roberto S. Mariano, "Predictors in Dynamic Nonlinear Models: Large-sample Behavior," *Econometric Theory* 5 (3) (1989), 430–452.
Bryant, Ralph C., Gerald Holtham, and Peter Hooper (eds.), *External Deficits and the Dollar: The Pit and the Pendulum* (Washington, D.C.: Brookings Institution, 1988).
Chong, Yock Y., and David F. Hendry, "Econometric Evaluation of Linear Macro-economic Models," *Review of Economic Studies* 53 (4) (1986), 671–690.
Chow, Gregory C., "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28 (3) (1960), 591–605.
Davidson, Russell, and James G. MacKinnon, "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica* 49 (3) (1981), 781–793.
Engle, Robert F., David F. Hendry, and Jean-François Richard, "Exogeneity," *Econometrica* 51 (2) (1983), 277–304.
Ericsson, Neil R., "Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration," *Journal of Policy Modeling* 14 (4) (1992), 465–495.
Fair, Ray C., and Robert J. Shiller, "The Informational Content of Ex Ante Forecasts," this REVIEW 71 (2) (1989), 325–331.
Fisher, Paul G., and Kenneth F. Wallis, "The Historical Tracking Performance of UK Macroeconometric Models 1978–85," *Economic Modelling* 7 (2) (1990), 179–197.
Gallant, A. Ronald, and Halbert White, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (Oxford: Basil Blackwell, 1988).
Helkie, William L., and Peter Hooper, "An Empirical Analysis of the External Deficit, 1980–86," chapter 2 in R. C. Bryant, G. Holtham, and P. Hooper (eds.), *External Deficits and the Dollar: The Pit and the Pendulum* (Washington, D.C.: Brookings Institution, 1988), 10–56.
Hendry, David F., "Predictive Failure and Econometric Modelling in Macroeconomics: The Transactions Demand for Money," chapter 9 in P. Ormerod (ed.), *Economic Modelling* (London: Heinemann Education Books, 1979), 217–242.
______, "Monte Carlo Experimentation in Econometrics," chapter 16 in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics* Volume II (Amsterdam: North-Holland, 1984), 937–976.
Intex Solutions, *TROLL Version 13*, Reference Manuals, Volumes 1–6 and Updates (Needham, MA: Intex Solutions, 1989).
Marquez, Jaime, "Income and Price Elasticities of Foreign Trade Flows: Econometric Estimation and Analysis of the US Trade Deficit," chapter 6 in L. R. Klein and J. Marquez (eds.), *Economics in Theory and Practice: An Eclectic Approach* (Dordrecht, The Netherlands: Kluwer Academic Publishers, 1989), 129–176.
Marquez, Jaime, and Neil R. Ericsson, "Evaluating the Predictive Performance of Trade-account Models," International Finance Discussion Paper No. 377, Board of Governors of the Federal Reserve System, Washington, D.C., Mar. 1990.
______, "Evaluating Forecasts of the U.S. Trade Balance," in R. C. Bryant, P. Hooper, and C. L. Mann (eds.), *Evaluating Policy Regimes: New Research in Empirical Macroeconomics* (Washington, D.C.: Brookings Institution, 1993).
Mizon, Grayham E., and Jean-François Richard, "The Encompassing Principle and Its Application to Testing Non-nested Hypotheses," *Econometrica* 54 (3) (1986), 657–678.
Pagan, Adrian, "On the Role of Simulation in the Statistical Evaluation of Econometric Models," *Journal of Econometrics* 40 (1) (1989), 125–139.
Schmidt, Peter, "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model," *Econometrica* 42 (2) (1974), 303–309.