Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 626

October 1998

A FRAMEWORK FOR ECONOMIC FORECASTING

Neil R. Ericsson and Jaime Marquez

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.bog.frb.fed.us.

A FRAMEWORK FOR ECONOMIC FORECASTING

Neil R. Ericsson and Jaime Marquez*

Abstract: This paper proposes a tripartite framework of design, evaluation, and post-evaluation analysis for generating and interpreting economic forecasts. This framework's value is illustrated by re-examining mean square forecast errors from dynamic models and nonlinearity biases from empirical forecasts of U.S. external trade. Previous studies have examined properties such as nonlinearity bias and the possible nonmonotonicity and nonexistence of mean square forecast errors in isolation from other aspects of the forecasting process, resulting in inefficient forecasting techniques and seemingly puzzling phenomena. The framework developed reveals how each such property follows from systematically integrating all aspects of the forecasting process.

Keywords: forecasts, mean square forecast error, Monte Carlo, nonlinearity bias, trade balance.

JEL classifications: C53, C15.

*Forthcoming in the *Econometrics Journal*, an electronic journal published by the Royal Economic Society on the WorldWide Web at www.blackwellpublishers.co.uk/ectj/. The authors are staff economists in the Division of International Finance, Board of Governors of the Federal Reserve System, Washington, D.C. 20551 U.S.A., and may be reached on the Internet at ericsson@frb.gov and marquezj@frb.gov respectively. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. We are grateful to William Helkie for providing the data for the Helkie-Hooper model; and to Richard Baillie, Carlo Bianchi, Giorgio Calzolari, Julia Campos, Mike Clements, Russell Davidson, Ray Fair, Giampiero Gallo, William Helkie, Dale Henderson, David Hendry, Peter Hooper, David Howard, Neva Kerbeshian, Mico Loretan, James MacKinnon, Jan Magnus, Doug McManus, Grayham Mizon, Garry Phillips, Jean-François Richard, George Tauchen, Ted Truman, Ken Wallis, and two anonymous referees for valuable comments and suggestions. We are indebted to Lucia Foster, Ned Prescott, and Molly Wetzel for excellent research assistance. The second author gratefully acknowledges the generous hospitality of the Department of Economics at the University of Canterbury, Christchurch, New Zealand, where he revised some of the material herein, and the financial support from a Visiting Erskine Fellowship, which facilitated visiting the University of Canterbury. This paper was presented at the 1996 EC^2 conference and was previously circulated under the title "A Framework for Simulated and Analytical Properties of Economic Forecasts". It draws upon and re-interprets research in Ericsson and Marquez (1989) and Marquez and Ericsson (1990). All numerical results were obtained using TROLL Version 13, Quattro Pro for Windows Version 5.0, and PcGive Professional Version 9: see Intex Solutions (1989), Borland (1993), and Doornik and Hendry (1996). This paper is being simultaneously circulated as International Finance Discussion Paper No. 626 by the Board of Governors of the Federal Reserve System, and as Discussion Paper No. 9812 by the Department of Economics at the University of Canterbury. These discussion papers can be obtained on the WorldWide Web at www.bog.frb.fed.us/pubs/ifdp/1998/626/default.htm and www.econ.canterbury.ac.nz/dp1998.htm respectively.

1 Introduction

Despite the voluminous literature on the theory and practice of economic forecasting, little effort has been devoted to organizing systematically the various aspects of generating and interpreting such forecasts. Even so, the different aspects involved are intimately related, and exploiting such interrelationships elucidates previously neglected properties of forecasts. Accordingly, this paper advances a tripartite framework of design, evaluation, and post-evaluation analysis. These concepts are not new in themselves, and they have been previously discussed in a narrower context for Monte Carlo analysis. The current paper generalizes these concepts and applies them to forecasting, re-interpreting the various aspects of forecasting as parts of a whole instead of as a collection of loosely related tasks. In so doing, this paper draws together theoretical, computational, empirical, and policy aspects of forecasting in a unified approach.

Four examples highlight the value of the proposed framework: the possible nonmonotonicity of mean square forecast errors in the forecast horizon, the possible nonexistence of mean square forecast errors for certain estimation sample sizes and forecast horizons, the uncertainty arising from Monte Carlo simulation of mean square forecast errors when exact analytical results are unavailable, and a parallel uncertainty in estimating the bias from using deterministic forecasts from nonlinear models. Each case is of substantive economic and methodological interest, and each shows how the proposed framework clarifies and improves upon existing results and techniques.

This paper is organized as follows. Section 2 develops the general framework, which partitions forecast activities into design, evaluation, and post-evaluation analysis. Design includes specification of the model for forecasting and selection of the forecast's characteristics of interest. Thus, design includes choice of the variables being forecast, the forecast horizon, the model's specification, and estimation method. Evaluation specifies how the forecasts are actually generated, and includes the choice of analytical or numerical techniques and the type of approximation used. Post-evaluation analysis includes presentation and summarization of the forecasts.

Sections 3 and 4 illustrate the principles of design, evaluation, and post-evaluation analysis for some time-series models that have been studied previously. That choice emphasizes how existing results can be beneficially re-interpreted in the proposed framework. Section 3 presents analytical properties of the mean square forecast error (MSFE) for one-step and multi-step ahead forecasts from vector autoregressions, relying on approximations due to Schmidt (1974) and Baillie (1979b). While the general formula for the MSFE is useful for empirical applications within this class of models, many of its properties can be most easily understood for a special case, the univariate first-order autoregressive [AR(1)] process. As detailed below, Section 4 discusses three potential properties of the MSFE for the AR(1) model: nonmonotonicity, nonexistence, and uncertainty in its estimation by Monte Carlo simulation.

First, Section 4.1 re-interprets Hoque, Magnus, and Pesaran's (1988) exact numerical results on the MSFE for an AR(1) model without an intercept. Mean square forecast errors are commonly viewed as being monotonic in the forecast horizon. However, they need not be so, and whether they are depends on dynamics and on the uncertainty in estimating the forecast model's coefficients. Nonmonotonicity has immediate consequences for policy, as the MSFE may achieve a maximum at forecast horizons typically of interest, e.g., one to two years out. Nonmonotonicity in the MSFE can be difficult to interpret on its own, as noted by Hoque, Magnus, and Pesaran (1988). That said, nonmonotonicity is easily understood through post-simulation analysis in conjunction with the (approximate) analytical formula for the MSFE from Section 3.

Second, Section 4.2 analyzes the MSFE of the AR(1) model using maximum likelihood estimation. As Hoque, Magnus, and Pesaran (1988) show, the MSFE does not exist beyond a certain horizon when the AR(1) model is estimated by least squares. This problem disappears for truncated estimators, such as maximum likelihood. Exact numerical solutions for the MSFE are not known for these truncated estimators, so forecast evaluation is by Monte Carlo. The approximate analytical solution for the MSFE from Section 3 aids interpreting the Monte Carlo results, paralleling its comparison in Section 4.1 to the exact numerical results from Hoque, Magnus, and Pesaran (1988). The framework in Section 2 identifies the need for accurate Monte Carlo outcomes, so control variates are employed for variance reduction.

Third, Section 4.3 re-examines Orcutt and Winokur's (1969) Monte Carlo study of the AR(1) model with an intercept. When exact analytical results are lacking, theoretical properties of forecasts are often ascertained through Monte Carlo simulation. That simulation, *qua* simulation, introduces uncertainty into the results. This calculation (or evaluation) may either clarify or obscure the forecasts' underlying properties, depending upon the particular simulation procedures employed. As in Section 4.2, the analytical properties from Section 3 help interpret the Monte Carlo results. In all three subsections of Section 4, the asymptotic MSFE captures much of the variation across experiments, and the approximate MSFE does even better in doing so. Their degree of inaccuracy relative to the exact MSFE appears related to how close conditions for the existence of the exact MSFE are to being violated when estimating by least squares.

Section 5 considers six models of U.S. imports and exports, some of which have been used in practical forecasting exercises at the Federal Reserve Board. These models' forecasts differ from those of the earlier sections in three respects: the models are nonlinear rather than linear; the data are empirical rather than simulated; and the object of analysis is the nonlinearity bias arising from deterministic forecasts, rather than the MSFE. Even asymptotic analytical solutions are not usually available for nonlinear models, so the analysis in Section 5 is entirely by Monte Carlo. While nonlinearity biases are well-known, existing analyses typically use inefficient techniques

in quantifying them. As in Section 4.2, more efficient techniques are readily apparent, once evaluation is viewed in a general framework for interpreting the properties of economic forecasts. Contrasting with Section 4.2, the variance-reduction technique is antithetic variates rather than control variates, in part because the asymptotic solution for the nonlinearity bias is not known. Efficiency gains from the antithetic variates range from 4-fold to over 2000-fold, illustrating how important good numerical techniques can be, even when the underlying problem is inherently analytical. Section 6 concludes.

2 Forecast Methodology

Forecasting is often viewed by economists as a simple extension of model estimation, in which fitted values are constructed from the estimated model over a sample not used in estimation. However, in order to understand better the properties of forecasts and forecast errors, it is helpful to divide the mechanics of forecasting into three parts:

1. *design*, in which the characteristics of the forecasts and the relationship of interest are specified;

2. *evaluation*, wherein the forecasts are generated; and

3. *post-evaluation analysis*, in which the forecasts are presented.

Here, "evaluation" takes its common meaning "working out the numerical value of", rather than (e.g.) "judging the value of". Table 1 lists details for each of these three categories. While this framework is applied to forecasting here, it could be used to interpret any properties of economic models.

This three-fold partition generalizes discussion in Hendry (1984) and Ericsson (1986) on experimental design, simulation, and post-simulation analysis for Monte Carlo studies. While forecasts may involve Monte Carlo simulation as an evaluation technique, they may also (or instead) rely on analytical calculations or numerical integration. In any case, this framework helps identify and clarify properties of forecasts that are otherwise missed or that generate puzzles. Sections 2.1, 2.2, and 2.3 discuss the three aspects of forecasting at a general level, using Table 1 as a guide. Table 2 (below) gives particulars for the models and forecasts examined in the remaining sections and indicates how this framework aids in the construction and interpretation of the associated results.

2.1 Design

Forecast design specifies the relationship of interest for generating the forecasts and the characteristics of the forecasts being examined. This subsection considers these two aspects of design.

Table 1. A Framework for Generating and Interpreting Forecasts

Design		Evaluation	Post-evaluation Analysis
Relationship of Interest	Characteristics of the Forecasts		

Dimension
  single equation
  multiple equations
	Dimension
	  scalar
	  vector
		Technique
		  analytical
		  numerical
		  Monte Carlo
		  naive
		  antithetic variates
		  control variates
		  bootstrap
			Presentation
			  algebraic
			  graphical
			  tabular
			  response surface

Completeness
  full system
  subsystem			

Temporal form
  static
  dynamic
  I(0)
  integrated
	Forecast horizon
	  one-step ahead
	  multi-step ahead
		Nature
		  [deterministic]
		  asymptotic
		  approximate
		  $O(n^{-1})$
		  higher order
		  exact
			Summarization
			  none
			  response surfaces
			  test statistics

Functional form
  linear
  nonlinear
	Transformations
	  linear
	  nonlinear		

Distributions
  error
  coefficient estimates
  initial conditions
	Distributional property
	  mean (and bias)
	  variance (and MSFE)
	  existence of moments
	  other		

Estimation
  OLS
  ML
  other			

To accommodate a wide range of possibilities, the specification of the estimated model (the relationship of interest) is general at the outset:

$$f(y_t, y_{t-1}, z_t, \theta, u_t) \;=\; 0\,, \qquad t = 1, \ldots, n+S\,. \tag{1}$$

The function $f(\cdot)$ is an $m \times 1$ vector of equations included in the model, y_t is an $m \times 1$ vector of endogenous variables at period t, z_t is a $q \times 1$ vector of (assumedly) weakly exogenous variables, θ is a $c \times 1$ vector of parameters, and u_t is an $m \times 1$ vector of disturbances, whose assumed distribution is specified as part of the relationship of interest (Table 1). The first n observations are used in estimation, and the remaining S are forecast. In describing forecast procedures, it is convenient to solve (1) explicitly for y_t and change the time subscript t to $n+s$, so that forecasts are calculated for positive s. This transformed representation is written as:

$$y_{n+s} \;=\; g(y_{n+s-1}, z_{n+s}, \theta, u_{n+s})\,, \qquad s = 1-n, \ldots, -1, 0, 1, \ldots, S\,, \tag{2}$$

where $g(\cdot)$ is a suitable redefinition of $f(\cdot)$ above. In practice, (2) may be a single equation or multiple equations, with or without exogenous variables and/or dynamics, and either linear or nonlinear. For multi-step ahead forecasts ($s > 1$), z_t is assumed strongly as well as weakly exogenous, or a model for z_t is used in the forecasting process, in effect embedding z in y; see Engle, Hendry, and Richard (1983).

Sections 3–5 consider numerous (but not exhaustive) combinations of these possibilities. Here and below, analytical expositions are typically for s-step ahead forecasts only. However, the formal structure developed applies to both s-step and one-step ahead forecasts, and both types of forecasts are examined in Sections 3–5.

Throughout, (2) is assumed to be the data generation process as well as the econometric model. The framework in Table 1 could allow for a distinction between the data generation process and the model. Doing so would result in an additional column under "design" in Table 1 with entries paralleling those under "relationship of interest". While this distinction would serve no immediate purpose, it is crucial for forecasting in general. See Clements and Hendry (1996, 1998a, 1998b) and Hendry (1997) on the generic effects of model mis-specification on forecasting, and Gallo (1991, 1996) on the specific effects of data mis-measurement through provisional data.

The characteristics of the forecasts include the forecast horizon, transformations of y_{n+s} to obtain the actual forecasts (and hence the dimension of the forecasts), and the distributional property of interest. By assumption, (2) is the process generating y_{n+s}, so (2) naturally serves for generating $\widehat{y}_{n+s}$ (the forecast of y_{n+s}), given some choice of assumptions about y_{n+s-1}, z_{n+s}, θ, and u_{n+s}. The particular choice adopted determines the type of forecast. If the actual value of y_{n+s-1} is used in $g(\cdot)$, then the S forecasts $\{\widehat{y}_{n+s};\ s = 1, \ldots, S\}$ are one-step ahead, i.e., forecasts of the endogenous variables are conditional on the observed endogenous variables lagged one period. Alternatively, the previous forecast $\widehat{y}_{n+s-1}$ could replace y_{n+s-1}, but with observed y_n

starting the forecast process. This sequential solution for $\widehat{y}_{n+s}$ in (2) generates a set of s-step ahead forecasts ($s = 1, \ldots, S$).

The variable z_{n+s} may be weakly or strongly exogenous in fact, and the modeler may (correctly or incorrectly) assume some exogeneity status for z. Even if z is strongly exogenous, either actual or forecast values of z might appear in $g(\cdot)$, where the forecast values would contribute additional uncertainty, both because z would be forecast rather than known and because the forecasting model for z would likely be estimated.

For a given forecast horizon s and a given treatment of y_{n+s-1} and z_{n+s}, the choice of θ and u_{n+s} in $g(\cdot)$ fully specifies the forecast procedure. Numerous choices of θ and u_{n+s} exist, the most common being "deterministic simulation" (u_{n+s} and θ fixed, e.g., at $u_{n+s} = 0$ and $\theta = \widehat{\theta}$) and "stochastic simulation" (either u_{n+s} random and $\theta = \widehat{\theta}$, or u_{n+s} and θ both random). Sections 3 and 4 focus on just deterministic simulations, whereas Section 5 compares the numerical solutions for both types of simulations.

While the econometric model generates the forecast $\widehat{y}_{n+s}$, economic interest may be in some transformation of $\widehat{y}_{n+s}$, and in some property of that transformation. Specifically, consider a linear transformation P of $\widehat{y}_{n+s}$ and a general property $h(\cdot)$. (The assumption of linearity is without loss of generality, noting that $g(\cdot)$ is arbitrarily nonlinear.) In this context, the object of the forecast analysis is the expectation μ_{n+s}:

$$\mu_{n+s} \equiv \mathcal{E}[h(P'\widehat{y}_{n+s})] = \iint h(P'\widehat{y}_{n+s}) \cdot \upsilon(u, \widehat{\theta}) \cdot du \cdot d\widehat{\theta} \,, \tag{3}$$

where $\mathcal{E}(\cdot)$ is the expectations operator, $\upsilon(\cdot, \cdot)$ is the joint density of u [defined as $(u_{n+1}, \ldots, u_{n+s})'$] and $\widehat{\theta}$, $\widehat{y}_{n+s-1}$ is implicitly a function of $\widehat{\theta}$ and u, and the selection of P and $h(\cdot)$ may affect whether or not the expectation μ_{n+s} exists. The sections below focus on aspects of the first two moments of the forecasts: the mean square forecast error, and the bias of deterministic forecasts. In Sections 4.1 and 4.2, y is a single variable, $P = 1$, and $h(P'\widehat{y}_{n+s})$ is $\widehat{y}^2_{n+s}$, making μ_{n+s} the MSFE of the variable in the model. Section 4.3 also examines the MSFE, but for one of two variables in a bivariate vector y by having $P = (1\ 0)'$. In Section 5, $h(P'\widehat{y}_{n+s}) = P'\widehat{y}_{n+s}$, and P selects one variable from a multivariate vector of forecasts $\widehat{y}_{n+s}$, so μ_{n+s} is the unbiased forecast of the variable selected. Equation (3) generalizes to allow linear combinations of forecasts from different forecast horizons, as in Campos (1992), who considers *inter alia* quarterly forecasts from monthly models. Below, for notational simplicity, P is assumed to be the identity matrix unless explicitly required otherwise.

2.2 Evaluation

Exact analytical solutions to (3) are known for only very simple models. Alternatives to such solutions include numerical integration (Section 4.1), analytical approximations (Section 3), Monte Carlo simulation (Sections 4–5), and bootstrapping (see

Efron (1982), Peters and Freedman (1985), and Hall (1994) *inter alia*). Each approach has its own strengths and weaknesses. Using multiple approaches for a given problem, as is done below, extracts each approach's advantages while ameliorating the associated shortcomings. Additionally, algorithms may be available for improving the computational efficiency with which (3) is solved. For Monte Carlo simulation in particular, Sections 4.2 and 5 employ two generic variance-reduction techniques, control variates and antithetic variates.

For each alternative technique, an application might ignore or account for the uncertainty arising from u and/or $\widehat{\theta}$. These two sources of uncertainty are denoted "inherent" and "coefficient". A technique ignoring both sources of uncertainty is "deterministic". A technique accounting for inherent uncertainty alone is "asymptotic", so-called because the uncertainty from u affects the distribution of the forecast error $\widehat{y}_{n+s} - y_{n+s}$, even when the estimation sample size n is large. If a technique incorporates at least the limiting distribution of $\widehat{\theta}$ as well as inherent uncertainty, then the technique is "approximate".

Even at this stage, the framework in Table 1 provides direct insight into the nature of stochastic simulation of forecasts. That stochastic simulation is aimed at solving a nonstochastic problem, namely (3), which itself may be analytically intractable. In that light, it is highly desirable to adopt computationally efficient algorithms for stochastic simulation, e.g., antithetic variates and control variates. Such variance-reduction techniques aim to reduce the Monte Carlo sampling variability introduced by changing the nonstochastic problem (3) to its stochastic analogue. See Hammersley and Handscomb (1964) and Hendry (1984) for further discussion, and Hendry and Trivedi (1972) and Hendry and Harrison (1974) for early examples using antithetic variates and control variates in the econometrics literature.

2.3 Post-evaluation Analysis

Once the forecasts or their characteristics have been calculated, some form of presentation is adopted, with tables and graphs being common modes. Other possibilities exist and, even for graphs and tables, both "good" and "bad" presentations are feasible. Tufte (1983, 1990, 1997) shows how small multiples of tables and graphs can improve presentation; cf. the figures below. Sometimes, properties of forecasts are summarized, as in Chow (1960) statistics and forecast-encompassing statistics; see Chong and Hendry (1986), Fair and Shiller (1989), Lu and Mizon (1991), Ericsson (1992), and Ericsson and Marquez (1993) on the latter.[1]

[1]In-sample encompassing statistics are also feasible, as in Fisher and Wallis (1990), being of the form suggested by Davidson and MacKinnon (1981) and Mizon and Richard (1986), noting Pagan (1989).

3 MSFEs for Vector Autoregressive Models

Mean square forecast errors have had a longstanding presence in the economics profession, both for calculating forecast confidence intervals and predictive failure tests through corresponding forecast variances, and for directly comparing models' forecast performance. Clements and Hendry (1993) identify important limitations for the latter use of MSFEs. Still, MSFEs occupy a central role in forecasting, particularly in light of their now-established role in forecast encompassing; cf. Ericsson (1992). Vector autoregressions (VARs) are a common and general class of models for forecasting, as they include univariate autoregressive models, reduced form econometric models, and simultaneous equations models. This section thus sketches some analytical properties of MSFEs for VARs, setting the backdrop for Section 4 on AR(1) models.

Schmidt (1974) and Baillie (1979b) build on Goldberger, Nagar, and Odeh (1961) to derive the approximate MSFE for two variants of the general vector autoregressive model. Schmidt considers the linear dynamic simultaneous equations model where the non-endogenous variables are known for the forecast period. That is, Schmidt's model is a subsystem conditional VAR with known, strongly exogenous variables. Baillie considers a complete VAR. His framework allows for the strongly exogenous variables in Schmidt's model but, if they are present, requires that they be forecast as well (i.e., true *ex ante* forecasting). Baillie (1979a, 1981) derives the approximate MSFE in Schmidt's framework with autoregressive errors. West (1996) provides further generalizations. To help understand the properties of the exact MSFE in Section 4, the current section sketches the derivation of the Schmidt-Baillie approximation and discusses its analytical properties. Equally, the Schmidt-Baillie approximation permits comparison of different evaluation techniques across different orders of approximation and estimators (Section 4).

The first horizontal block in Table 2 characterizes the framework for this section. Section 3.1 designs the forecasting structure, describing the relationship of interest (the VAR) and the characteristics of the forecasts (MSFEs for linear combinations of vector one-step and s-step ahead forecasts). Section 3.2 analytically derives the approximate MSFE for forecasts from that model (evaluation), and Section 3.3 considers properties of the approximate MSFE (post-evaluation analysis). The exposition in Sections 3.1–3.2 follows Chong and Hendry (1986) because of the latter's accessibility.

3.1 Design

Let y_t be a vector of variables generated by the first-order autoregressive process:

$$y_t = Ay_{t-1} + u_t \qquad u_t \sim IN(0, \Omega) \qquad t = 2, \ldots, n+S\,, \tag{4}$$

where A is an $m \times m$ matrix of feedback coefficients, u_t is distributed independently

Table 2. A Guide to the Forecasts Examined

Section	Model	Design		Evaluation	Post-evaluation Analysis
		Relationship of Interest	Characteristics of the Forecasts		
3	VAR	multiple equations full system dynamic, I(0) linear OLS	vector 1- and s-step linear MSFE	analytical, approximate	algebraic
4.1	AR(1) without an intercept	single equation full system dynamic linear OLS	scalar 1- and s-step linear MSFE	analytical, including exact	algebraic Figures 1–3
4.2	AR(1) without an intercept	same as above, with estimation by ML	same as above	analytical Monte Carlo naive, CV	algebraic Figures 1, 4
4.3	AR(1) with an intercept	same as above, with estimaton by OLS	same as above	analytical Monte Carlo naive	algebraic Figure 5
5	Trade models M1–M6	multiple equations (M1–M5) single equation (M6) subsystem (M1–M4) full system (M5–M6) dynamic nonlinear (M1–M5) linear (M6) various estimation procedures	scalar 1- and s-step nonlinear (M1–M5) linear (M6) mean and bias	Monte Carlo naive, AV	Tables 3–4 various statistics

and normally $[IN(\cdot,\cdot)]$ with mean zero and covariance matrix Ω, and y_1 is given. Although (4) appears limited to first-order processes, it is not. If the underlying process is of a higher order, it always can be stacked to give a first-order process. Because of that stacking, or for other reasons, the variables of interest for forecasting may be a subset (or some linear combination) of y_{n+s}, with the selection matrix P extracting the vector of interest as $P'y_{n+s}$. Further, the matrix A may be restricted (e.g., have zeros), so it is useful to recognize explicitly how A is a function of its unconstrained elements θ:

$$\alpha \equiv A^{\nu} = R\theta + r \,, \tag{5}$$

where $(\cdot)^{\nu}$ denotes the column vectorizing operator, α is the vectorization of A, and all the elements of R and r are known. For convenience, A is assumed to be such that y_t is integrated of order zero, although this condition is more restrictive than necessary.

Next, assume that θ is estimated by $\widehat{\theta}$, which is asymptotically distributed as:

$$\sqrt{n}\cdot(\widehat{\theta}-\theta) \xrightarrow{D} N(0,\Psi) \,, \tag{6}$$

where $\xrightarrow{D}$ denotes convergence in distribution as $n \to \infty$, $N(0,\Psi)$ denotes a normal distribution with mean zero and covariance matrix Ψ, and Ψ is the asymptotic covariance matrix for $\widehat{\theta}$. Combining (5) and (6), it follows that:

$$\sqrt{n}\cdot(\widehat{\alpha}-\alpha) \xrightarrow{D} N(0,\Gamma) \,, \tag{7}$$

where $\Gamma = R\Psi R'$. In finite samples, the approximate distribution of $\widehat{\theta}$ is:

$$\widehat{\theta} \underset{app}{\sim} N(\theta, \frac{\Psi}{n}) \,, \tag{8}$$

where $\underset{app}{\sim}$ denotes "is approximately distributed as". For instance, least-squares estimation of the VAR obtains (6)–(8) for $R = I$. For the remainder of the derivation, (8) is treated as if it were the exact distribution of $\widehat{\theta}$, i.e., terms smaller than $O_p(n^{-1/2})$ in the distribution of $\widehat{\theta}$ are ignored.

Using the data $[y_1 \ldots y_n]$ to forecast y_{n+s} gives:

$$\widehat{y}_{n+s} = \widehat{A}^s y_n \,, \tag{9}$$

the *ex ante* s-step ahead forecast. By repeated substitution of (4) into itself at successive lags, the actual outcome y_{n+s} is:

$$y_{n+s} = A^s y_n + \sum_{i=0}^{s-1} A^i u_{n+s-i} \,, \tag{10}$$

where $A^0 \equiv I$ if $A = 0$. Thus, the discrepancy between actual and forecast y_{n+s} is:

$$(y_{n+s} - \widehat{y}_{n+s}) \;=\; \left(\sum_{i=0}^{s-1} A^i u_{n+s-i}\right) + (A^s - \widehat{A}^s) y_n \; . \tag{11}$$

Selecting the variable of interest gives the corresponding forecast error $\widehat{e}_{n+s}$:

$$\begin{aligned} \widehat{e}_{n+s} &\equiv P'(y_{n+s} - \widehat{y}_{n+s}) \\ &= P' \left(\sum_{i=0}^{s-1} A^i u_{n+s-i}\right) + P'(A^s - \widehat{A}^s) y_n \; . \end{aligned} \tag{12}$$

The mean square forecast error for $\widehat{e}_{n+s}$ is the distributional property of interest, i.e., $\mathcal{E}(\widehat{e}_{n+s}\widehat{e}'_{n+s}|y_n)$.

3.2 Evaluation

The two terms on the right-hand side of (12) correspond directly to the two sources of uncertainty being investigated. The first, $P' \left(\sum_{i=0}^{s-1} A^i u_{n+s-i}\right)$, is the cumulation of the shocks to which y_t is subject over the interval $[n+1, n+s]$, where each shock is weighted by the degree to which it influences $P'y_{n+s}$, the variable being predicted. The second, $P'(A^s - \widehat{A}^s)y_n$, reflects the uncertainty present from using an estimated value of A rather than its true value in forecasting $P'y_{n+s}$. For convenience, these two terms are denoted $a_{n,s}$ and $b_{n,s}$. The approximate MSFE for $\widehat{y}_{n+s}$ can be calculated by evaluating the variances of these two terms.

Straightforwardly, the variance of the first term is:

$$\mathrm{var}(a_{n,s}|y_n) \;=\; P' \left[\sum_{i=0}^{s-1} A^i \Omega (A^i)'\right] P \;\equiv\; \mathrm{AsyMSFE} \; , \tag{13}$$

which is the asymptotic (i.e., large n) MSFE. The approximate variance of the second term is:

$$\mathrm{var}(b_{n,s}|y_n) \;\cong\; n^{-1} \cdot (I \otimes y'_n)[D(s)'\Gamma D(s)](I \otimes y_n) \; , \tag{14}$$

where

$$D(s)' \;=\; \frac{\partial (P'A^s)^\nu}{\partial \alpha'} \;=\; (P' \otimes I) \left[\sum_{i=0}^{s-1} A^i \otimes (A^{s-i-1})'\right] \; ; \tag{15}$$

the symbol $\otimes$ is the Kronecker product; and, in Mann and Wald's (1943) order notation, a term of $O_p(n^{-1})$ is ignored in $b_{n,s}$ when calculating $\mathrm{var}(b_{n,s}|y_n)$. The derivation of (14) is more complicated than that of (13) and appears in Ericsson and Marquez (1989, Appendix A).[2] Because $\widehat{A}^s$ and $\{u_{n+s-i}, i = 0, \ldots, s-1\}$ are independent (by assumption), then $a_{n,s}$ and $b_{n,s}$ are as well, so the variance of $a_{n,s}$

[2] A higher-order approximation could be obtained by employing a higher-order Taylor-series expansion for the distribution of $\widehat{\theta}$ in (6) and using distributional results in Shenton and Johnson (1965); cf. (23).

and the approximate variance of $b_{n,s}$ can be added together to obtain the approximate MSFE (AppMSFE):

$$\begin{aligned} \text{AppMSFE}(\widehat{e}_{n+s}|y_n) &= P' \left[\sum_{i=0}^{s-1} A^i \Omega (A^i)' \right] P \\ &\quad + n^{-1} \cdot (I \otimes y_n')[D(s)'\Gamma D(s)](I \otimes y_n) \; . \end{aligned} \tag{16}$$

Equation (16) is relatively easy to implement in a computer program because it involves only sums of products of matrices. P and y_n are known, and the unknown elements of A, Ω, and Γ may be replaced by consistent estimates of them. Calzolari (1981, 1987) provides additional computational simplifications for models with restrictions on A.

3.3 Post-evaluation Analysis

The two terms on the right-hand side of (16) behave differently in terms of the forecast horizon s: $\text{var}(a_{n,s}|y_n)$ increases monotonically in s, whereas $\text{var}(b_{n,s}|y_n)$ may increase before eventually decreasing to zero. Hence, as s increases, the approximate MSFE may decrease as well as increase. Section 4.1 examines this nonmonotonicity in greater detail. Because $\text{var}(b_{n,s}|y_n)$ is positive for finite n and s, the approximate (and so the exact) MSFEs are always larger than the asymptotic MSFE.

4 MSFEs for AR(1) Models

The second, third, and fourth horizontal blocks in Table 2 characterize the framework for Sections 4.1, 4.2, and 4.3, which analyze the possible nonmonotonicity of MSFEs, the existence of MSFEs, and the simulation uncertainty from estimating MSFEs by Monte Carlo. While each facet of forecasting enters these three subsections, Sections 4.1, 4.2, and 4.3 focus on post-evaluation analysis, design, and evaluation respectively, with the ordering of the subsections reflecting the natural presentation of the models and estimators examined.

4.1 Nonmonotonicity of MSFEs

This section summarizes the design, evaluation, and post-evaluation analysis of the AR(1) model by Hoque, Magnus, and Pesaran (1988), using the asymptotic and approximate MSFEs derived in Section 3 to interpret their exact results. Post-evaluation analysis is a particular issue here because of the observed and apparently puzzling nonmonotonicity in the exact MSFE.

Design. One special case of (4) is the AR(1) model, which is:

$$y_t = \beta y_{t-1} + u_t \qquad u_t \sim IN(0, \sigma^2) \qquad t = 2, \ldots, n+S \tag{17}$$

in Hoque, Magnus, and Pesaran's notation, where y_t and u_t are scalar variables, β is the autoregressive coefficient, and σ^2 is the variance of u_t. In the notation of Section 3, $A = \alpha = \theta = \beta$, $\Omega = \sigma^2$, $R = 1$, and $r = 0$. For $\mid \beta \mid < 1$, the ordinary least squares (OLS) estimator of β is asymptotically distributed as:

$$\sqrt{n} \cdot (\widehat{\beta} - \beta) \quad \stackrel{D}{\rightarrow} \quad N(0, [1 - \beta^2]) \ , \tag{18}$$

in which case various matrices in Section 3 simplify: $\Psi = \Gamma = (1 - \beta^2)$, $P = 1$, and $D(s) = s\beta^{s-1}$. Thus, the forecast error (12) is:

$$\widehat{e}_{n+s} \quad = \quad \sum_{i=0}^{s-1} (\beta^i) \cdot u_{n+s-i} + (\beta^s - \widehat{\beta}^s) \cdot y_n \ . \tag{19}$$

For β estimated by OLS, the unbiasedness of the s-step ahead forecast $\widehat{\beta}^s y_n$ for y_{n+s} follows immediately from (19), provided that the expectation of (19) actually exists; see Fuller and Hasza (1980).

Taking advantage of the explicit relationship between the OLS estimator and the disturbances $\{u_t\}$ in calculating $\widehat{e}_{n+s}$, Hoque, Magnus, and Pesaran (1988) derive an analytical expression for the exact MSFE for the AR(1) model (17). To do so, they must specify the distribution of the initial condition y_1, which they posit is:

$$y_1 \quad \sim \quad IN(0, \delta^2 \sigma^2) \ , \tag{20}$$

where δ is an arbitrary constant. Hoque, Magnus, and Pesaran examine two cases: $\delta^2 = (1 - \beta^2)^{-1}$ (the "stationary case") and $\delta^2 = 1$ (the "non-stationary case"). Additionally, the triplet (β, n, s) must be specified, and they choose a full factorial design of:

$$\begin{aligned} \beta &= (0.0, 0.1, 0.2, \ldots, 0.8, 0.9, 0.95, 0.99), \\ n &= (10, 15, 20, 25), \text{ and} \\ s &= (1, 2, 3, 4) \ \ [S = 4]. \end{aligned}$$

The error variance σ^2 can be set to unity without loss of generality, as it simply scales y.

Evaluation. For each set of values of (β, n, s, δ), Hoque, Magnus, and Pesaran calculate the exact MSFE (ExMSFE) by numerical integration of an explicit form for (3), as given in their Theorem 2. Details appear in Hoque, Magnus, and Pesaran (1988, Appendices A–C).

The approximate MSFE also can be calculated, noting that (16) simplifies to:

$$\text{AppMSFE}(\widehat{e}_{n+s} | y_n) \quad = \quad \sigma^2 \left(\frac{1 - (\beta^2)^s}{1 - \beta^2} \right) + (n^{-1} \cdot y_n^2) \cdot (s\beta^{s-1})^2 \cdot (1 - \beta^2) \ . \tag{21}$$

Equation (21) identifies the separate contributions of the different sources of uncertainty. The first term on the right-hand side is the asymptotic term, and the second

is the part arising from coefficient uncertainty. Although the original derivation of (21) is difficult to ascertain, it appears as early as 1970 in Box and Jenkins (1970, p. 269). For the one-step ahead forecast, (21) simplifies to the more familiar formula:

$$\text{AppMSFE}(\widehat{e}_{n+1}|y_n) \;=\; \sigma^2 + \frac{y_n^2(1-\beta^2)}{n} \;\cong\; \sigma^2\left(1 + \frac{y_n^2}{\sum_{t=2}^{n} y_{t-1}^2}\right); \qquad (22)$$

cf. Chow (1960) and Hendry (1979).

Post-evaluation Analysis. Hoque, Magnus, and Pesaran (1988) tabulate their calculations; and they also plot a subset of those exact MSFEs, both as pure MSFEs (for $s = 1$ and for $n = 15$) and as MSFEs relative to the asymptotic MSFEs (for $n = 15$). Hoque, Magnus, and Pesaran (1988, pp. 333, 335) note with some surprise that the exact MSFE can *decrease* as the forecast horizon increases: the formula for the approximate MSFE in (21) above provides an intuitive explanation of why that occurs. The remainder of this subsection first uses the approximate MSFE to explain that observed nonmonotonicity, and then examines more generally the relationship between the asymptotic, approximate, and exact MSFEs.

The approximate MSFE (21) differs from the exact MSFE by ignoring a term of $O_p(n^{-1})$ in $\widehat{e}_{n+s}$. The analytical properties of the approximate MSFE can explain the nonmonotonicity in the exact MSFE. The properties of the first term on the right-hand side of (21) are relatively simple and well-known. This term is the asymptotic MSFE and starts at the conditional variance of y_t (σ^2) for $s = 1$. The asymptotic MSFE increases monotonically in s, tending to the unconditional variance of y_t, which is $\sigma^2/(1-\beta^2)$.

The structure and properties of the second term on the right-hand side of (21) can be easily explained and interpreted via its derivation. The term $\widehat{\beta}^s$ in (19), viewed as a function of $\widehat{\beta}$, is approximated by a first-order Taylor-series expansion about β to give:

$$\widehat{\beta}^s \;=\; \beta^s + s\beta^{s-1}(\widehat{\beta} - \beta) + O_p(n^{-1}) \,. \qquad (23)$$

Substitution into $(\beta^s - \widehat{\beta}^s) \cdot y_n$ gives $s\beta^{s-1}(\beta - \widehat{\beta}) \cdot y_n + O_p(n^{-1})$, from which the second term on the right-hand side of (21) follows immediately, using (18).[3] That term is always non-negative (and generally positive) for finite n and s, and it vanishes as either s or n becomes large. However, for a given sample size n, final observed value y_n, and nonzero β, the MSFE can either decrease monotonically as the forecast horizon s increases, or increase first and then decrease. Its path depends upon the behavior of the sequence $\{(s\beta^{s-1});\ s = 0, 1, 2, \ldots\}$, and so upon the particular value of β. The contribution of coefficient uncertainty to the approximate MSFE can be large or small relative to the latter, so the functional relationship between the approximate

[3]The effect of s on the distribution of $\widehat{\beta}^s$ also can be seen by analogy to moments of a standardized normal variate x, for which $\mathcal{E}(x^s)^2 = (2s)!/(2^s \cdot s!)$: e.g., $\mathcal{E}(x^s)^2 = 1, 3, 15, 105$ for $s = 1, 2, 3, 4$. Clearly, taking a power of $\widehat{\beta}$ can increase its variance dramatically.

MSFE and the forecast horizon s itself depends upon β and y_n. Thus, the potentially large and varying contribution of coefficient uncertainty to the approximate MSFE can explain the noted nonmonotonicity in s of the exact MSFE; see also Chong and Hendry (1986, p. 685) and Clements and Hendry (1998a, 1998b). The second component of the approximate MSFE provides a simple analytical explanation of this behavior, to the extent that the approximate MSFE offers a good approximation to the exact MSFE. So, this subsection concludes by comparing the asymptotic, approximate, and exact MSFEs for Hoque, Magnus, and Pesaran's design.

The asymptotic, approximate, and exact MSFEs are functions of the parameters (β, s, n).[4] In forecasting, it is of interest to know how and why those MSFEs vary as their determinants vary. Additionally, the exact MSFE is not available in many practical situations whereas asymptotic and approximate MSFEs are, so Hoque, Magnus, and Pesaran's results grant the unusual opportunity of assessing how close the asymptotic and approximate MSFEs are to the exact MSFE.

Figure 1 plots the asymptotic, approximate, and exact MSFE (stationary case) as a function of the forecast horizon s for selected values of β and n: $\beta = (0.2, 0.7, 0.9)$ and $n = (10, 20)$. (Figure 1 also includes estimated exact MSFEs for maximum likelihood estimation of β, examined in Section 4.2 below.) The values of β imply y_t ranging from being nearly white-noise to highly autoregressive. Both here and in following section, the term y_n^2 in (21) is chosen to be a representative value, $\sigma^2/(1-\beta^2)$, i.e., equal to its unconditional expectation.[5]

For all values of β and n, the asymptotic MSFE increases in the forecast horizon s, with the curvature of the MSFE depending on β. The discrepancy between the one-step and infinite-step ahead asymptotic MSFEs increases with β, and the asymptotic MSFE levels out more gradually for larger β. Both results follow directly from (21).

Figure 1 reveals three distinct patterns for the approximate MSFE: a decreasing approximate MSFE, an increasing approximate MSFE, and an approximate MSFE that increases and then decreases. Which pattern prevails depends upon both β and n. For small β, the approximate MSFE is declining almost uniformly as the horizon increases, with the initial (one-step ahead) approximate MSFE being the *largest*. That arises because the uncertainty from estimating β is large (from (18)), but that uncertainty is unimportant in forecasting y_{n+s} except for $s = 1$. Mathematically, $s\beta^{s-1}$ in (21) is approximately zero except for $s = 1$, when it is approximately unity. Because the asymptotic MSFE changes little as s increases, the approximate MSFE is declining from the start (Figures 1a and 1b).

For larger values of β, the variance from coefficient uncertainty increases first and

[4]Without loss of generality, $\sigma^2 = 1$ because σ^2 is a scale factor in (17), and hence in (21).

[5]Another justification for this choice is that the unconditional expectation $\mathcal{E}[(\beta^s - \widehat{\beta}^s)^2 y_n^2]$ is approximately $\mathcal{E}[(\beta^s - \widehat{\beta}^s)^2] \cdot \mathcal{E}[y_n^2]$ because $\widehat{\beta}$ and y_n are approximately independent. See Phillips (1979) for an extensive discussion on the conditional and unconditional finite sample distributions of the forecast error.

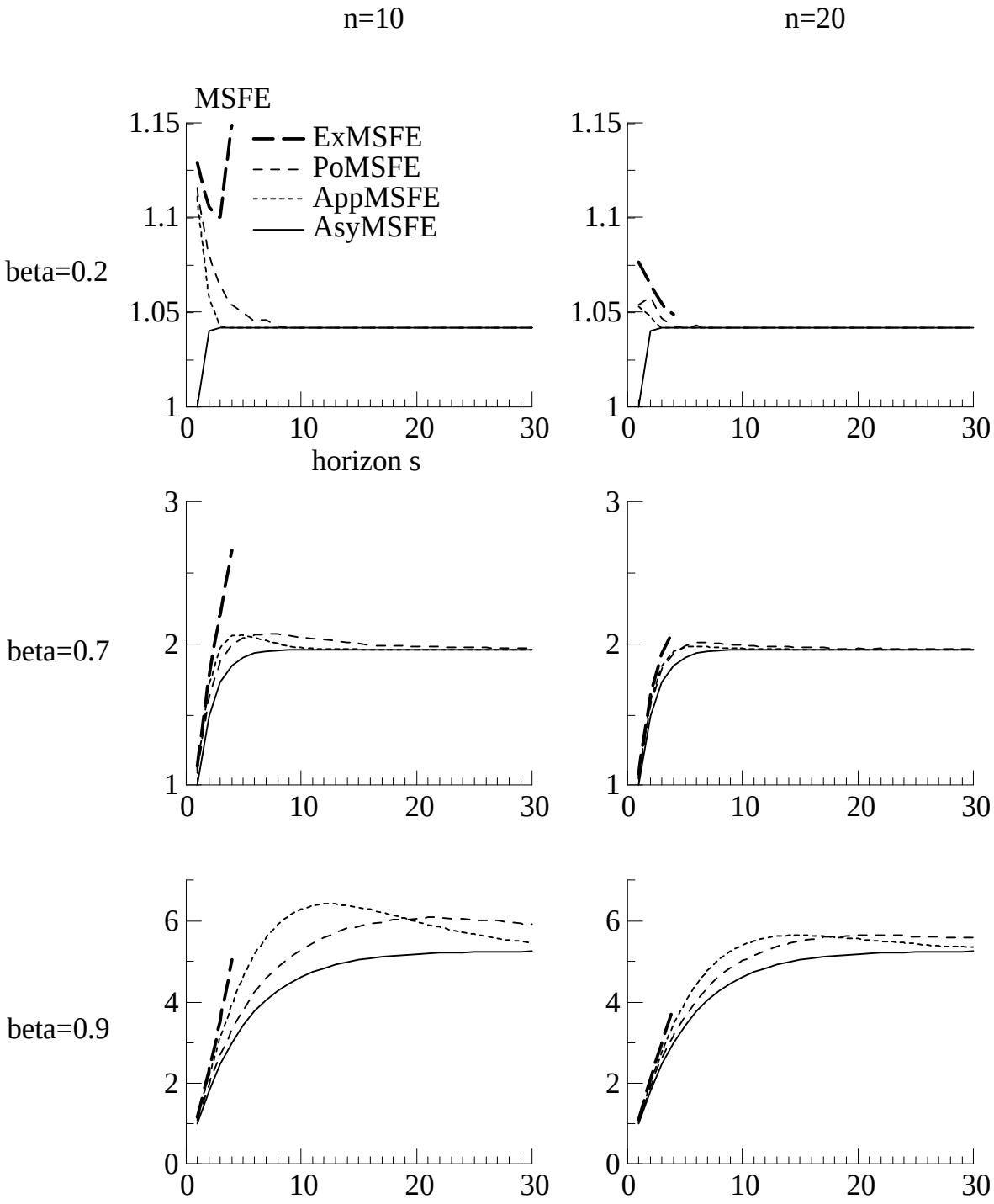

Figure 1: The exact, pooled, approximate, and asymptotic mean square forecast errors for autoregressive coefficient values of $\beta = (0.2, 0.7, 0.9)$ and sample sizes of $n = (10, 20)$ as a function of the forecast horizon s ($s = 1, \ldots, 30$).

then falls because the multiplicative coefficient s in $s\beta^{s-1}$ dominates for small s but the exponential term (β^{s-1}) dominates for large s. As the sum of two components, one monotonically increasing and the other increasing and then tending to zero, the approximate MSFE can either increase first and then fall towards the unconditional variance of y_t (Figures 1c–1f) or increase monotonically, approaching that asymptotic variance (also increasing in s) from above. The latter would be the case for all the figures if the sample size n were large enough.

Although there are some notable discrepancies between the approximate and exact MSFEs in the figures (primarily for $n = 10$ with $s = 4$), the approximate MSFE does remarkably well in approximating the exact MSFE, so well that it is difficult to distinguish them at $n = 20$. However, the dependence on β, n, and s of discrepancies between the approximate (or asymptotic) MSFE and the exact MSFE is more apparent by plotting all of Hoque, Magnus, and Pesaran's calculations. Figures 2 and 3 do so, using small multiples of three-dimensional graphs.

Graphs in the first column of Figure 2 plot the percent discrepancies between the asymptotic MSFE in (13) and the associated exact MSFE from Hoque, Magnus, and Pesaran (1988, Tables 1 and 2) for the stationary case. Each graph in the column is defined by an estimation sample size n and, for that sample size, plots the percent discrepancies as a function of the autoregressive coefficient β and the forecast horizon s. Graphs in the second column of Figures 2 likewise plot the discrepancies between the approximate and exact MSFEs.[6] Figure 3 plots the corresponding results for the nonstationary case ($\delta^2 = 1$).

The asymptotic formula captures the behavior of the exact MSFE well for small to medium values of β and for large n, with deviations of the order of 5%–15%; but it does poorly otherwise. The approximate MSFE fares better: typical departures are 2%–3% or less. As with the asymptotic MSFE, more sizable discrepancies appear for large s paired with small n and/or large β. The effective sample size, as measured here by $(n-1)/(1-\beta^2)$, predicts the possibility of such departures under those conditions; cf. Sims (1974) and Hendry (1984). The approximate MSFE is almost invariably smaller than the exact MSFE, although this inequality need not hold in general; cf. Peters and Freedman (1985, p. 258) and Bianchi (1990, p. 98).

The analytical formulae in (19) and (21) help identify the sources of discrepancies between the three types of MSFEs. Discrepancies between the asymptotic and exact MSFE arise because $\widehat{\beta}$ is not identically β. Discrepancies between the approximate and exact MSFE arise for several reasons. First, the asymptotic and finite sample variances of $\widehat{\beta}$ differ, and hence so do the respective variances of $\widehat{\beta}^s$ in (19). Second, $\widehat{\beta}$ is biased in finite samples but not asymptotically (likewise for $\widehat{\beta}^s$). Third, the

[6]Because only $(n-1)$ observations are actually used in estimating β, the calculation of (21) for the tables uses $(n-1)$ rather than n. Ericsson and Marquez (1989, Appendix B) list the values of the asymptotic and approximate MSFEs for the experiments in Hoque, Magnus, and Pesaran (1988) and for those in Orcutt and Winokur (1969) discussed in Section 4.2 below.

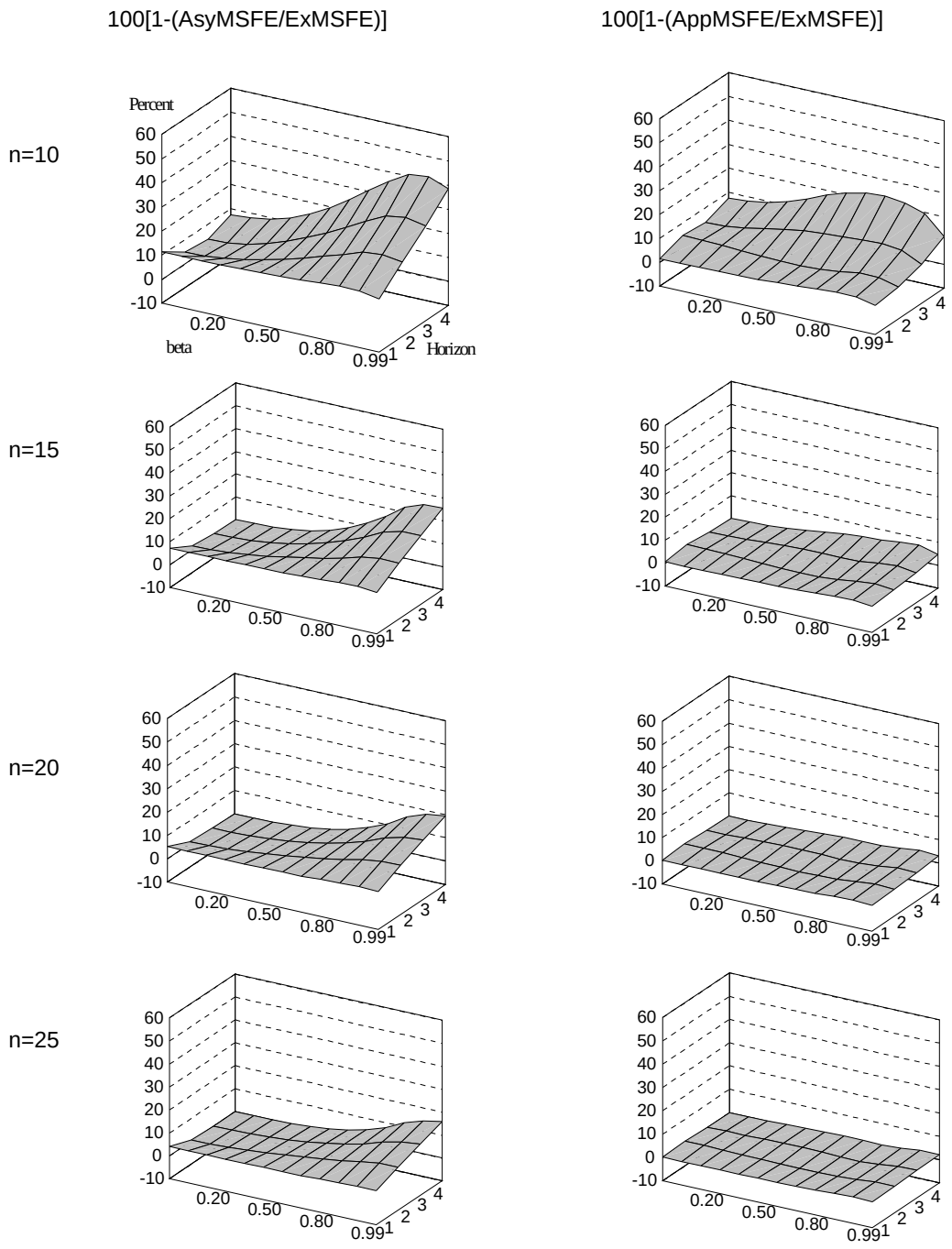

Figure 2: Percent deviations of the asymptotic and approximate MSFEs from the exact MSFE, as a function of sample size n, forecast horizon s, and autoregressive coefficient β (stationary case).

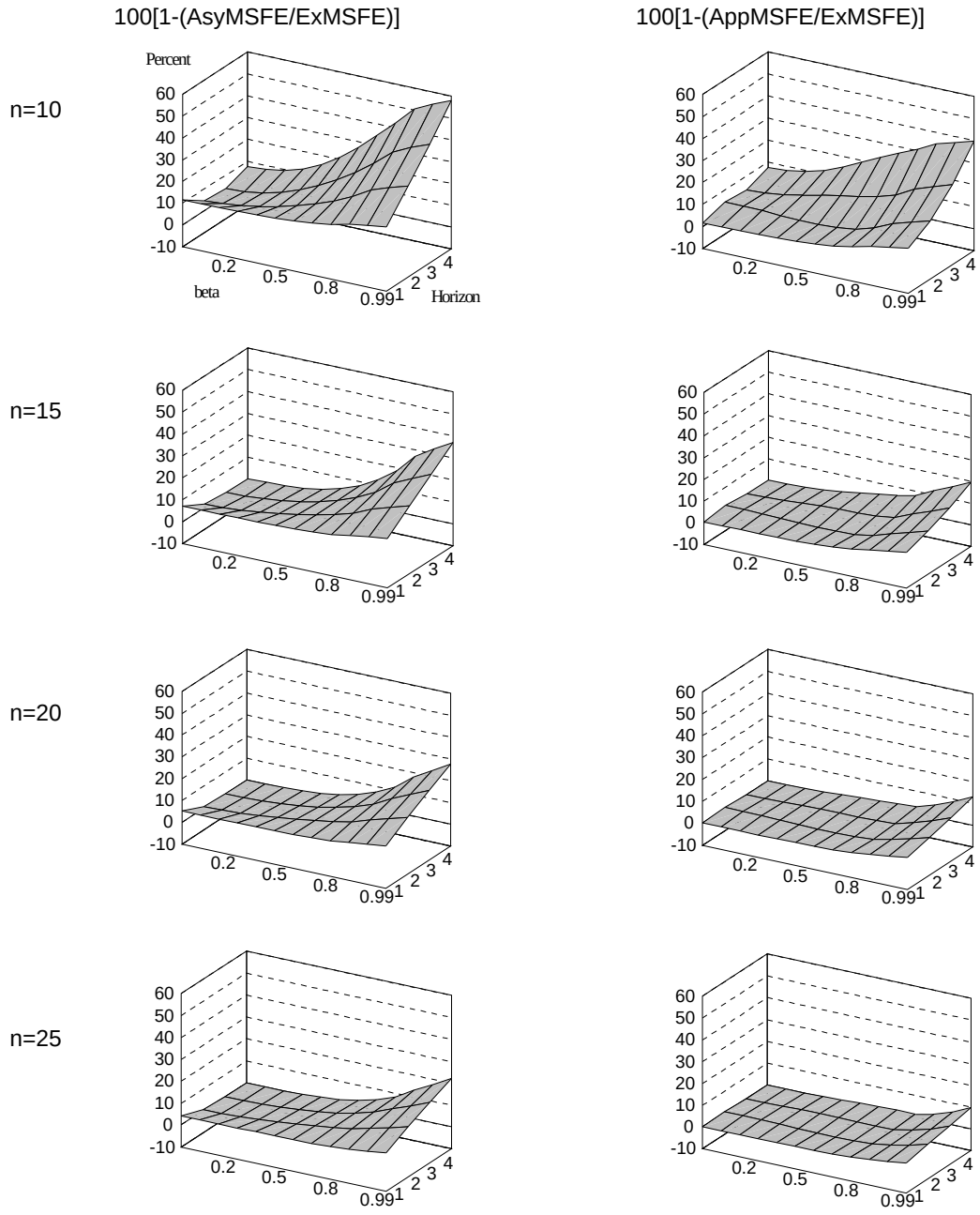

Figure 3: Percent deviations of the asymptotic and approximate MSFEs from the exact MSFE, as a function of sample size n, forecast horizon s, and autoregressive coefficient β (nonstationary case).

first-order Taylor-series approximation of $\widehat{\beta}^s$ about β^s may ignore important terms. The accuracy and the generality of the approximate MSFE do not in any way belittle Hoque, Magnus, and Pesaran's exact results. To the contrary, exact results are highly desirable because they involve no approximation error, because they are so rare, and because they are essential for assessing the accuracy of approximations such as (21).

4.2 The Existence of MSFEs

Hoque, Magnus, and Pesaran (1988) show that the exact MSFE exists for the AR(1) model if and only if the forecast horizon s is not greater than $(n-2)/2$. Non-existence of the exact MSFE for longer forecast horizons raises several issues. Specifically, Hoque, Magnus, and Pesaran's result was derived for the least-squares estimator of β, and it does not generalize to other estimators of β. This subsection shows that the exact MSFEs for truncated estimators of β are finite at all forecast horizons and considers additional properties of MSFEs for truncated estimators. The exact maximum likelihood estimator is a truncated estimator of particular interest, so this subsection analyzes its MSFE in detail. Forecast design is thus central here, both through the choice of n and s (to ensure existence of the MSFE for OLS) and through the choice of estimator. That said, evaluation of the MSFE also plays an important role in this subsection.

A different perspective arises from current forecasting practice, which often treats estimation as a mechanical procedure for replacing unknown coefficients with numbers. When this practice entails estimated dynamic models by OLS, the resulting MSFE may fail to exist at certain forecast horizons. Maximum likelihood estimation solves this problem. To understand why, it helps to view model estimation as an integral part of the forecasting exercise (i.e., of forecast design) and not as technique for merely quantifying unknowns.

The effects of estimator truncation on MSFE follow straightforwardly. Suppose an estimator $\bar{\beta}$ is truncated such that $|\bar{\beta}| \leq \gamma$ for some positive bound γ. The corresponding forecast error is:

$$\overline{e}_{n+s} = \sum_{i=0}^{s-1} \beta^i u_{n+s-i} + (\beta^s - \bar{\beta}^s) \cdot y_n \ , \tag{24}$$

paralleling (19). A bound can be placed on the exact MSFE for $\overline{e}_{n+s}$ by application of the triangle and Schwartz inequalities and by noting that $\mathcal{E}(y_n^2) = \sigma^2/(1-\beta^2)$.

$$\begin{aligned}
\text{ExMSFE}(\overline{e}_{n+s}) &= \sigma^2 \left[\frac{1-(\beta^2)^s}{1-\beta^2}\right] + \mathcal{E}[(\beta^s - \bar{\beta}^s)^2 \cdot y_n^2] \\
&\leq \sigma^2 \left[\frac{1-(\beta^2)^s}{1-\beta^2}\right] + \mathcal{E}[(\beta^s - \bar{\beta}^s)^2] \cdot \mathcal{E}(y_n^2) \\
&\leq \sigma^2 \left[\frac{1-(\beta^2)^s}{1-\beta^2}\right] + \mathcal{E}[(|\beta^s| + |\bar{\beta}^s|)^2] \cdot \mathcal{E}(y_n^2)
\end{aligned}$$

$$\leq \quad \sigma^2 \left[\frac{1 - \beta^{2s} + (1 + \gamma^s)^2}{1 - \beta^2} \right] \tag{25}$$

If the bound is the unit circle ($\gamma = 1$), then a slightly looser bound exists that is independent of s: i.e., $5\sigma^2/(1 - \beta^2)$, five times the unconditional variance of the process. The existence of the exact MSFE does not require that the estimator be consistent for any value whatsoever. Conversely, because neither bound makes use of the asymptotic properties of $\bar{\beta}$, neither converges to the asymptotic MSFE as $n \to \infty$. Even so, the existence of a bound (and so of the exact MSFE) indicates how sensitive the existence results are to minor changes in the distributional assumptions of the estimator being used.[7]

The condition for existence of the exact MSFE also bears on the accuracy of approximation. For instance, for $n = 10$ in Figures 1–3, the exact MSFE exists for $s \leq 4$ only. The worsening of the approximation error as s increases may be due to the declining number of moments of the forecast error. The effects of the existence of moments are also suggested by Figures 1–2, where the convergence of the exact and approximate MSFE appears faster than $O(n^{-2})$ for large s.

Because the Schmidt-Baillie approximation relies on only the asymptotic distribution of the estimator used, the formulae in Section 3 are valid for all estimators asymptotically equivalent to OLS; see (7) and (18) above. One such estimator is exact maximum likelihood (ML), which accounts for the distribution of y_1 through (20) with $\delta^2 = 1/(1 - \beta^2)$.[8] Maximum likelihood estimation has several desirable features in the present context. In particular, its exact MSFE exists at all forecast horizons, independent of the size of the estimation period. That follows because the ML estimator is bounded by the unit circle, i.e., $\gamma = 1$. Furthermore, $\gamma = 1$ implies a bound on the exact MSFE for ML, independent of the forecast horizon. The first of these results permits assessing the generality of the Schmidt-Baillie approximation at forecast horizons longer than those feasible for OLS. However, because the analytical formula for the exact MSFE with ML is unknown and because Hoque, Magnus, and Pesaran's numerical approach is specific to least squares, this subsection compares the approximate MSFE with Monte Carlo estimates of the exact MSFE of ML. Again, the approximate MSFE is remarkably close to the exact MSFE, even for short estimation periods and long forecast horizons.

Design. In order to assess the closeness of the approximate MSFE to the exact MSFE for ML without the advantage of exact analytical formulae, the exact MSFE is estimated by Monte Carlo for a wide range of β, n, and s, and compared with the

[7]For instance, suppose that the estimator is OLS truncated with $\gamma = 10^{3\,10}$, entailing a range often permitted by floating-point double-precision calculations on a computer. That value of γ implies existence of the exact MSFE for all s, yet the truncated estimator will look like OLS for virtually all practical purposes.

[8]Cf. Maekawa (1987), who shows an equivalence to $O(n^{-1})$ between the complete distributions of the forecast error for OLS and approximate ML.

corresponding asymptotic and approximate MSFEs. Specifically, the Monte Carlo experiment is a full factorial design of:

$$\begin{aligned} \beta &= (0.0, 0.1, 0.2, \ldots, 0.8, 0.9, 0.95, 0.99), \\ n &= (10, 15, 20, 25, 40), \text{ and} \\ s &= (1, 2, \ldots, 30) \ [S = 30], \end{aligned}$$

with $\delta^2 = \sigma^2/(1-\beta^2)$ to ensure stationarity. Again, $\sigma^2 = 1$, and without loss of generality. This design embeds the range of values evaluated by Hoque, Magnus and Pesaran for OLS with a stationary AR(1) process. However, because the exact MSFE for ML exists for all forecast horizons, its values can be compared with the approximate MSFE at much longer forecast horizons than available to Hoque, Magnus and Pesaran.

Evaluation. To obtain reasonably accurate Monte Carlo estimates, 10,000 replications per experiment are used. Recognizing that simulation is a tool for solving an unknown analytical problem leads immediately to the construction and use of Monte Carlo procedures that are more efficient than pure simulation. The simulations of the MSFE in this subsection incorporate a control variate proposed and implemented by Ericsson and Marquez (1989) and also applied by Clements and Hendry (1995, p. 136) to forecasts in cointegrated systems. The remainder of this subsection describes the control variate's construction, its properties, and the simulation results obtained with it.

The basic Monte Carlo simulation was conducted as follows. For each experiment, which is characterized by a (β, n) pair, K replications of $n+S$ normal pseudo-random numbers $\{(u_{k,t},\ t = 1, \ldots, n+S),\ k = 1, \ldots, K\}$ were generated from pairs of uniform pseudo-random numbers using Box and Muller's (1958) transformation.[9] For the kth replication, a set $(y_{k,t},\ t = 1, \ldots, n+S)$ was created from (17) and (20) with $\delta^2 = (1-\beta^2)^{-1}$, and the ML estimate was found by solving the cubic in β from setting the score of the likelihood equal to zero; see Koopmans (1942), Anderson (1971, p. 354), and Beach and MacKinnon (1978). Given the ML estimate, the forecast error was calculated for each value of s ($s = 1, \ldots, S$). Explicitly, let $\tilde{\beta} = \tilde{\beta}(\beta_h, n_j, k)$ denote the ML estimate for the kth replication of the experiment with the hth value of β (β_h) and the jth estimation sample size n (n_j). Further, let $\{\widetilde{e}_{k,n_j+s}(\beta_h);\ s = 1, \ldots, S\}$ be

[9]The uniform random number generators are the mixed-congruential generator RNDM from Carrier, Atkins, and Taylor (1969) (but converted from COMPASS to FORTRAN) and the multiplicative-congruential generator G05CAF from the Numerical Algorithms Group (1984). Different random number generators were used for each number in the pair of uniform pseudo-random numbers in order to avoid the potential difficulties with Box and Muller's transformation described in Neave (1973).

the corresponding set of observed forecast errors, where:

$$\begin{aligned}\widetilde{e}_{k,n_j+s}(\beta_h) &\equiv y_{k,n_j+s} - \tilde{y}_{k,n_j+s} \\ &= \sum_{i=0}^{s-1}(\beta_h^i)\cdot u_{k,n_j+s-i} + (\beta_h^s - \tilde{\beta}^s)\cdot y_{k,n_j} \ . \end{aligned} \tag{26}$$

The usual Monte Carlo estimator of the MSFE (McMSFE) is:

$$\text{McMSFE}_{n_j+s}(\beta_h) = \sum_{k=1}^{K}\frac{[\widetilde{e}_{k,n_j+s}(\beta_h)]^2}{K} \ , \tag{27}$$

which, in Hendry's (1984) terminology, is the "naive" Monte Carlo estimator. When normalized by the exact MSFE, this Monte Carlo estimator is approximately distributed as $N(1,[2/K])$. In the current experimental design, K is 10^4, so the standard deviation of McMSFE/ExMSFE is $(2/10^4)^{1/2}$ or about 1.4%. Increasing K tenfold would reduce the standard deviation to only about 0.5%, an indication of the difficulties in obtaining precise estimates by such Monte Carlo techniques. For comparison, see Ansley and Newbold (1980), who compute McMSFE for several estimators (including ML) of various ARMA processes, but use 1000 or fewer replications per experiment.

Control variates provide a powerful method for variance reduction of naive Monte Carlo estimators, as discussed by Hammersley and Handscomb (1964) and Hendry (1984) *inter alia*. To be useful, a control variate (CV) should be highly correlated with the naive estimator and should have a known distribution. Because the purpose of the Monte Carlo study is to estimate a moment which is unknown, those two properties might appear to conflict. However, often it is possible to partition a statistic into an asymptotic component and a finite sample one, with the former having an exact distribution; cf. Hendry and Harrison (1974) and Hendry (1984) on doing so for econometric estimators. The Monte Carlo estimator McMSFE has a natural control variate because the first term on the right-hand side of the equality in (26) (and more generally, of (12)) is exactly normal, and independent of the second term. The implied control variate (denoted CvMSFE) is:

$$\text{CvMSFE} = \frac{1}{K}\sum_{k=1}^{K}\left[\sum_{i=0}^{s-1}(\beta_h^i)\cdot u_{k,n_j+s-i}\right]^2 \ , \tag{28}$$

which is exactly distributed as $\text{AsyMSFE}\cdot\chi^2(K)/K$ and has a mean of AsyMSFE; see also Section 4.3 below.

The CV is used to reduce the simulation uncertainty of the naive estimator by subtracting the CV from the naive estimator (with which it is positively correlated) and adding back the known mean of the CV. The resulting Monte Carlo estimator is called a pooled estimator. Here, the pooled estimator (PoMSFE) is:

$$\begin{aligned}\text{PoMSFE} &\equiv \text{McMSFE} - \text{CvMSFE} + \mathcal{E}(\text{CvMSFE}) \\ &= \text{AsyMSFE} + \frac{1}{K}\sum_{k=1}^{K}\left[(\beta_h^s - \tilde{\beta}^s)\cdot y_{k,n_j}\right]^2 \ . \end{aligned} \tag{29}$$

By construction, the pooled estimator has the same expectation as the naive estimator. Its variance is smaller than that of the naive estimator by the extent to which the CV is correlated with the naive estimator. In the present case, the reduction in variance is obvious because the CV has eliminated the term in the naive estimator which simulates the asymptotic MSFE. That is, the naive estimator McMSFE simulates a term of $O_p(1)$ whereas the pooled estimator PoMSFE simulates a term of only $O_p(n^{-1})$. The efficiency of the CV will vary across experiments, but from a cursory comparison of the fluctuations in the naive and pooled estimates, they are considerable for the design chosen. Further, the CV requires little additional computational expenditure because the asymptotic MSFE is trivial to calculate and is the same for all n. An even more efficient CV is feasible by utilizing the asymptotic distribution of $s\beta^{s-1}(\widehat{\beta} - \beta)$ from (23), in which case the pooled estimator simulates a term of only $o_p(n^{-1})$ in the MSFE. However, this CV was not implemented, given the considerable accuracy of the CV in (28).

Post-evaluation Analysis. Figure 1 plots values of the pooled MSFE for selected β and n, in addition to the exact MSFE for OLS, the asymptotic MSFE, and the approximate MSFE. For $\beta = 0.2$ with $n = 10$, the approximate MSFE is even closer to the exact MSFE for ML than to the exact MSFE for OLS. At medium and large values of β (0.7 and 0.9) with $n = 10$, the approximate MSFE still does well, but it over-estimates the exact MSFE at short to medium horizons and under-estimates it at very long horizons. Relatedly, the hump so evident in the approximate MSFE is less pronounced (but still present) in the exact MSFE for ML. For $n = 20$, the deviations between the exact and approximate MSFE are much smaller than for $n = 10$, as expected.

The full simulation results appear in Figure 4 as deviations of the pooled MSFEs from the asymptotic and approximate MSFEs, paralleling Figure 2. To condense presentation, values for $s > 10$ appear for s at multiples of five. At long horizons, the exact, asymptotic, and approximate MSFE all change slowly as a function of s in any case. All values of the respective MSFEs are tabulated in Ericsson and Marquez (1989, Appendix D). As with OLS, the asymptotic MSFE does reasonably well for small to medium values of β and for large n, with the approximate MSFE doing better, and over a wider range of β and n. As n increases, both deviations generally decline, as would be expected because the exact MSFE is tending to the asymptotic MSFE. Unlike with OLS, the approximate MSFE often over-estimates the exact MSFE for β in the range of $[0.80, 0.95]$. However, for β very close to the unit circle ($\beta = 0.99$), the approximate MSFE again under-estimates the exact MSFE. The boundedness of the ML estimator is affecting the exact MSFE for large β, but little more can be said without considering terms smaller than $O(n^{-1})$ in the MSFE. Even so, the approximate MSFE offers a remarkably simple and accurate summary of the behavior of the exact MSFE for ML.

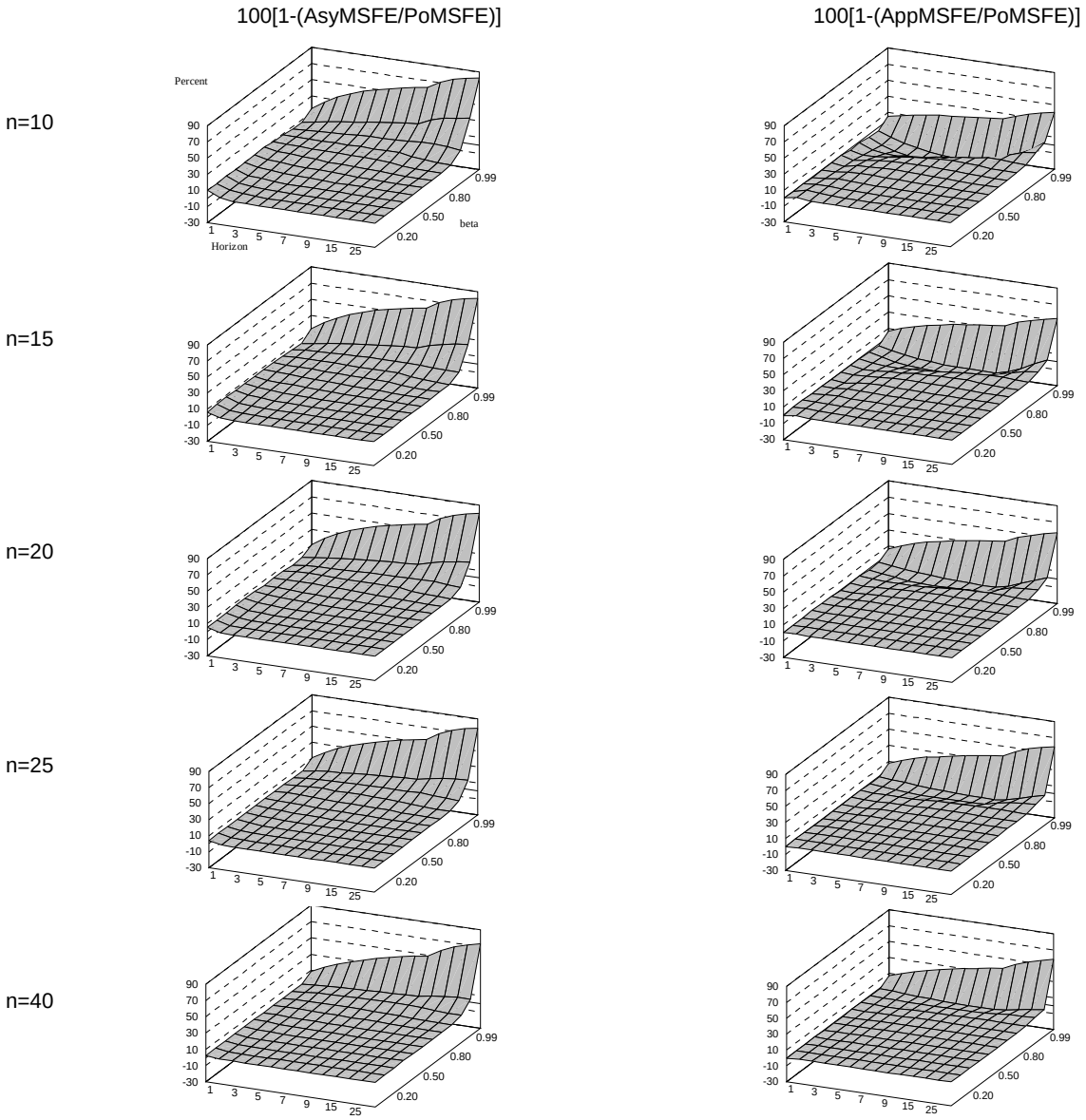

Figure 4: Percent deviations of the asymptotic and approximate MSFEs from the (estimated pooled) exact MSFE for maximum likelihood, as a function of sample size n, forecast horizon s, and autoregressive coefficient β.

4.3 Monte Carlo Simulation Uncertainty

Even while omitting an intercept, the AR(1) model in (17) has been useful for discussing generic properties of the MSFE from a VAR (Section 3), exact results from Hoque, Magnus, and Pesaran (1988) (Section 4.1), and the simulation for ML (Section 4.2). However, the role of an included unknown intercept is also important, both because intercepts usually appear in empirical models and because Orcutt and Winokur (1969), in a pivotal Monte Carlo study of the AR(1) model, include an intercept. Thus, this subsection analyzes Orcutt and Winokur's Monte Carlo simulations of MFSEs for an AR(1) model with an intercept. Inclusion of an intercept is easily incorporated into the Schmidt-Baillie approximation, albeit with a minor twist. Details appear in Ericsson and Marquez (1989, Appendix B). Estimating the intercept introduces additional uncertainty, so the resulting approximate MSFE is always larger than that for equations with a known intercept, even if (as in Orcutt and Winokur's experiments) the intercept is zero. The framework in Table 2 identifies two critical issues: imprecision, and the lack of existence of the MSFE for some of Orcutt and Winokur's experiments.

Design. Orcutt and Winokur's Monte Carlo study evaluates numerous facets of the AR(1) model estimated by OLS. As part of their study, they estimate by Monte Carlo the exact MSFE for the AR(1) model with an intercept:

$$y_t \;=\; \tau + \beta y_{t-1} + u_t \qquad\qquad u_t \;\sim\; IN(0,\sigma^2) \qquad\qquad t = 2,\ldots,n+S\;, \tag{30}$$

where τ is the intercept. Their design is a full factorial of:

$$\begin{aligned} \tau &= 0, \\ \beta &= (0.0, 0.3, 0.6, 0.9, 1.0), \\ n &= (10, 20, 40), \text{ and} \\ s &= (1, 2, 3, 4) \;\; [S = 4]. \end{aligned}$$

Equation (30) can be written as a two-equation version of (4), with the first equation being the AR(1) process and the second equation defining the intercept τ:

$$\begin{aligned} y_t &= \left[\begin{array}{c} y_{1t} \\ y_{2t} \end{array}\right] \\ &= A y_{t-1} + u_t \\ &= \left[\begin{array}{cc} \beta & \tau \\ 0 & 1 \end{array}\right] \cdot y_{t-1} + \left[\begin{array}{c} u_{1t} \\ 0 \end{array}\right] , \end{aligned} \tag{31}$$

where $y_{21} = 1$, $\theta = (\tau \;\; \beta)'$, and $P = (1 \;\; 0)'$. The vectors and matrices necessary for solving (16) follow directly, with the resulting forecast error being:

$$\widehat{e}_{n+s} \;=\; \sum_{i=0}^{s-1} \beta^i u_{n+s-i} \;+\; (\beta^s - \widehat{\beta}^s) \cdot y_n \;+\; \tau \sum_{i=0}^{s-1} (\beta^i - \widehat{\beta}^i) \;+\; (\tau - \widehat{\tau}) \sum_{i=0}^{s-1} \widehat{\beta}^i \;. \tag{32}$$

The first two terms on the right-hand side of (32) are the same as in (19), the third term arises from having a nonzero intercept τ, and the fourth term is the contribution from estimating rather than knowing τ.

Evaluation. Orcutt and Winokur use 1000 replications per experiment and estimate the MSFEs as in (27), that is, without any variance-reduction techniques. Given the imprecision of naive Monte Carlo for even $10,000$ replications in Section 4.2, this subsection derives an approximate lower bound on the uncertainty of Monte Carlo estimates of the exact MSFE. This bound is valid for the AR(1) model both with and without estimating an intercept, and it generalizes straightforwardly to the MSFE from a general linear dynamic system by using a Wishart rather than a χ^2 distribution.

In (12), and so in (19) and (32), the source of inherent uncertainty is independent from that of coefficient uncertainty. For instance, in a Monte Carlo analysis such as Orcutt and Winokur's, all the u_t are simulated; thus, both the future errors u_{n+s-i} and the $\widehat{\beta}$ are simulated. The following derivation ignores the latter effect because the first dominates, at least for large n. From independence, that results in a lower bound on the associated variability from Monte Carlo simulation.

Ignoring coefficient uncertainty, the forecast errors $\widehat{e}_{n+s}$ are a linear combination of the future shocks $u_{k,n+s-i}$:

$$\widehat{e}_{n+s} \;=\; (y_{k,n+s} - \widehat{y}_{k,n+s}) \;\cong\; \sum_{i=0}^{s-1} (\beta^i) \cdot u_{k,n+s-i} \; . \tag{33}$$

The $u_{k,n+s-i}$ are jointly normal, so the linear combination of them on the right-hand side is normal:

$$\sum_{i=0}^{s-1} (\beta^i) \cdot u_{k,n+s-i} \;\sim\; N(0, \text{AsyMSFE}) \; . \tag{34}$$

The Monte Carlo estimator of the MSFE is the average of the squared forecast errors:

$$\text{McMSFE} \;=\; \sum_{k=1}^{K} \frac{(y_{k,n+s} - \widehat{y}_{k,n+s})^2}{K} \;\underset{app}{\sim}\; \text{AsyMSFE} \cdot \frac{\chi^2(K)}{K} \; , \tag{35}$$

where the approximation in distribution reflects the approximation in (33). Because the first two moments of a $\chi^2(K)$ are K and $2K$, and because K typically is quite large, the following approximation is useful:

$$\frac{\text{McMSFE}}{\text{ExMSFE}} \;\cong\; \frac{\text{McMSFE}}{\text{AsyMSFE}} \;\underset{app}{\sim}\; \frac{\chi^2(K)}{K} \;\underset{app}{\sim}\; N\left(1, \left[\frac{2}{K}\right]\right) \; . \tag{36}$$

That is, the estimated MSFE is unbiased for the exact MSFE, with a percent standard deviation approximately equal to $100 \cdot (2/K)^{1/2}$. Unbiasedness is an exact result, following directly from the estimated MSFE being a sample mean of the exact MSFE. Monte Carlo simulation indicates that the approximation errors in (36) are small for

the values of (β, n, s) in (e.g.) Hoque, Magnus, and Pesaran's (1988) design and with $K \geq 100$. Orcutt and Winokur (1969) use $K = 1000$, so the 95% confidence interval on a typical estimate in their study is approximately $\pm 9.0\%$. Purely numerical (non-analytical) calculation of the Monte Carlo uncertainty is also feasible, as in Calzolari and Sterbenz (1986).

As in Section 4.2, the approximate MSFE provides a useful benchmark for the Monte Carlo estimates of the exact MSFE. Fuller and Hasza (1980) derive a formula for the approximate MSFE directly from (32). However, in Orcutt and Winokur's experiments, $\tau = 0$ in fact, simplifying (32), Ψ, and hence Γ and $D(s)$. In Orcutt and Winokur's case, the approximate MSFE is:

$$\text{AppMSFE} \;=\; \sigma^2 \left(\frac{1 - (\beta^2)^s}{1 - \beta^2} \right) \;+\; (n^{-1} \cdot \sigma^2) \cdot \left[(s\beta^{s-1})^2 + \left(\frac{1 - \beta^s}{1 - \beta} \right)^2 \right] , \qquad (37)$$

where y_{1n}^2 is set to $\sigma^2/(1 - \beta^2)$, its unconditional expectation. The first term on the right-hand side of (37) is the asymptotic MSFE and is the same as when the intercept is known. The first term in the braces is the effect from estimating β (as in (21)); the second term is from estimating τ and is additional to what appears in (21). These three terms arise from the first, second, and fourth terms on the right-hand side of (32).

Equation (37) provides both the asymptotic and approximate MSFEs for their Monte Carlo, with one exception. Equation (37) is not valid for $\beta = 1$, so $\beta = 0.9999$ is used instead. That should (and does) offer a good approximation, given the difficulty in finite samples of distinguishing between a unit root and a root close to (but less than) unity.

Post-evaluation Analysis. Two issues are of concern: the accuracy of the estimated exact MSFEs and the existence of the exact MSFEs themselves. The graphs in Figure 5 plot the percent discrepancies of the asymptotic and approximate MSFEs from Orcutt and Winokur's estimated exact MSFEs. The overall pattern parallels that in Figures 1–4: the approximate MSFE generally fares better than the asymptotic MSFE, with the latter almost invariably underestimating the estimated exact MSFE. Both the asymptotic and approximate MSFE fare better for larger n and smaller β. From (36), the standard error for their estimated exact MSFE is about 4.5%, so the discrepancies between their estimates and the approximate MSFEs appear to be due almost exclusively to simulation uncertainty. The accuracy of the approximate MSFE here and for the AR(1) model in Sections 4.1–4.2 adds to Chong and Hendry's (1986) Monte Carlo evidence on the accuracy of the Baillie-Schmidt approximation for a two-equation model.

Magnus and Pesaran (1989) show that the exact MSFE exists for the AR(1) model with an estimated intercept if and only if the forecast horizon s is not greater than $(n - 3)/2$. Interestingly, the exact MSFE does not exist for Orcutt and Winokur's

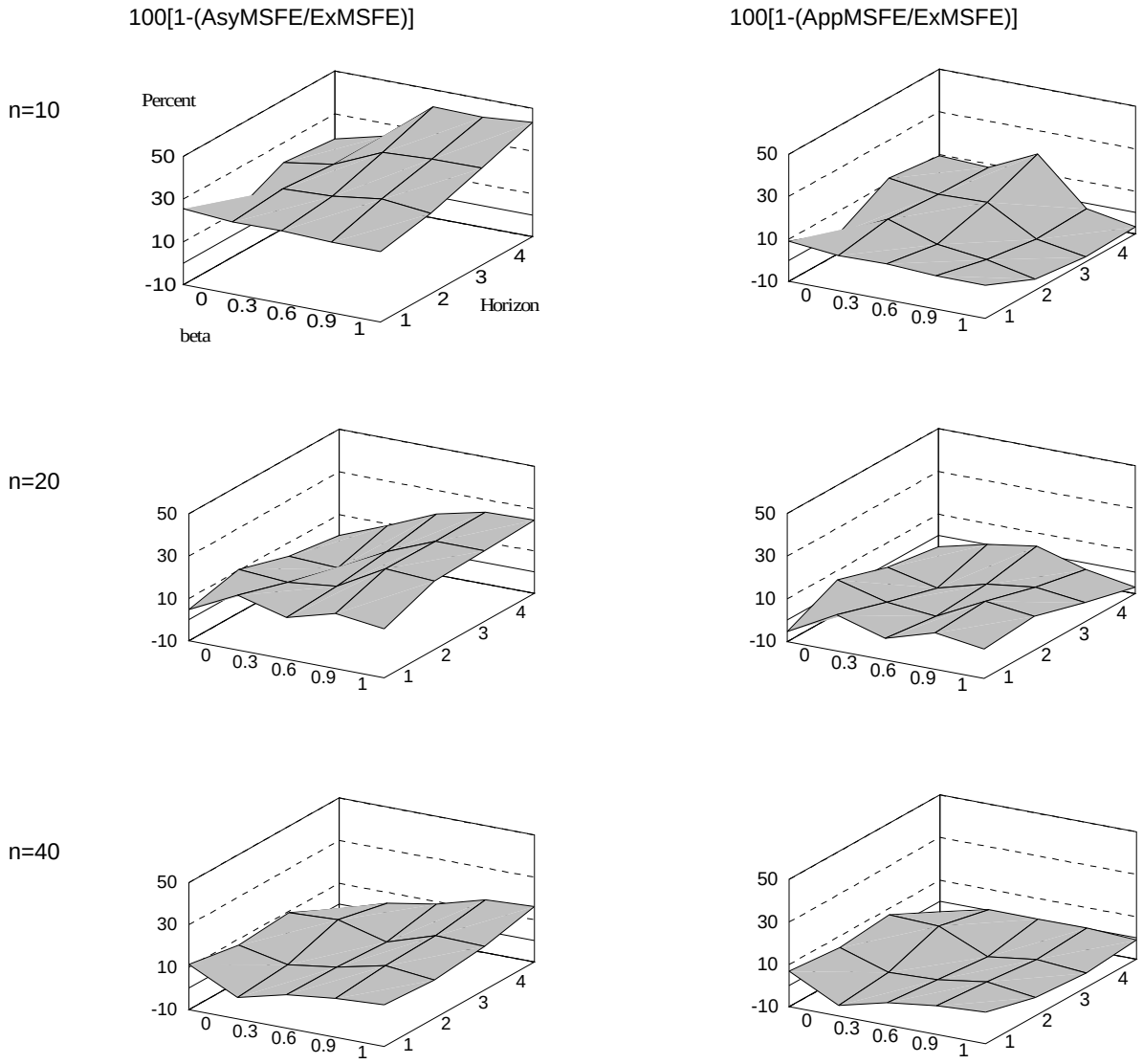

Figure 5: Percent deviations of the asymptotic and approximate MSFEs from Monte Carlo estimates of the exact MSFE in Orcutt and Winokur (1969).

experiments with ($n = 10,\ s = 4$), yet the approximate MSFE still does quite well at approximating their estimates. This surprising result can be interpreted as follows. For values of s for which the exact MSFE does not exist, the approximate MSFE still can be calculated, and it may provide accurate confidence intervals for the forecasts. However, because there is a significant probability of $\widehat{\beta}$ being greater than unity and thus causing the forecast error to explode for large s, the tails of the exact density of the forecast error are too thick for its variance to exist. Sargan (1982) examines a similar situation in which an estimator is well-behaved asymptotically but has no moments in finite samples. In light of his paper, the approximation in (16) may be interpreted as analogous to the Nagar approximation for the moments of an estimator. Conversely, the lack of existence of the exact MSFE (when that occurs) *must* be due to terms smaller than $O_p(n^{-1})$ in the squared forecast error (i.e., $o_p(n^{-1})$ and probably $O_p(n^{-3/2})$) because the approximate MSFE accounts for all terms $O_p(n^{-1})$ and larger, and the approximate MSFE exists for all s.

Remarks. Before turning to the empirical forecasts in Section 5, five issues are worth brief mention: sample size, exogenous variables, model linearity, forecast horizons, and the treatment of y_n.

First, the smallest sample size in the simulations above ($n = 10$) is very small in the context of empirical work. However, it may be a reasonable number to use for comparison with empirical work, given that only one or two coefficients are being estimated. The sample size relative to the number of coefficients is a plausible measure in this context; cf. Sargan (1975). Since much empirical work involves fewer than ten observations per coefficient estimated, $n = 10$ may be large rather than small for practical purposes.

Second, forecasts of endogenous variables often are based on forecasts (rather than known values) of exogenous variables, adding another source of uncertainty. The algebra of Section 3 readily addresses this because it analyzes a complete system: exogenous variables can be included in the system in the same manner as the endogenous variables, but the former are not simultaneously determined with the endogenous variables nor are they Granger-caused by the endogenous variables. However, for even relatively small systems, the approximate MSFE in (16) can become awkward to compute because of the large matrices arising from vectorizing and from Kronecker products. Calzolari (1987) provides an ingenious analytical technique that can dramatically reduce the computational burden, making calculation of the approximate MSFE feasible for medium- to large-scale models. One alternative is to simulate stochastic exogenous variables in the calculation of forecasts properties, as in Haitovsky and Wallace (1972) and Fair (1980).

Third, the system in Section 3 is linear. Analytic approximations to confidence intervals could be constructed for nonlinear equations (or systems) as well, but Mariano and Brown (1983) show that simulation may be preferable, not only for the MSFE but for the forecast itself. In the context of (12), both the u_{n+s-i} and the $\widehat{A}$ would

be replicated a number of times by Monte Carlo simulation according to their estimated distributions, and Monte Carlo estimates of the forecast mean and the MSFE would be constructed from the resulting pseudo-forecasts. Following this approach, Bianchi, Calzolari, and Brillet (1987) estimate MSFEs from a multi-sectoral model of the French economy, and Section 5 (below) and Marquez and Ericsson (1993) estimate mean forecasts and MSFEs from various models of the U.S. trade balance. Marquez (1991) applies this simulation approach to estimate confidence intervals for the response of the U.S. trade balance to alternative exchange rate realizations. That analysis examines the sensitivity of the confidence intervals to the two types of uncertainty addressed here. His application also demonstrates that the uncertainty of coefficient estimates can have implications for economic questions other than just those dealing with forecasts, e.g., paths of dynamic multipliers.

Fourth, the maximum of the MSFE over all forecast horizons may occur at a relatively short forecast horizon. Estimated autoregressive coefficients in dynamic econometric models range from the very small (e.g., for equations in first differences) to those close to unity (e.g., for equations in levels). Even for the corresponding (and wide) range of values for β, the maximum of the approximate MSFE is often between one and twelve periods, precisely the range over which economists are often interested in forecasting most accurately. That is also the range for which the asymptotic MSFE appears the poorest approximation to the approximate and exact MSFE. In fact, the asymptotic MSFE generally underestimates the approximate (and exact) MSFE for finite horizons, and the former need not even be the main component of the latter.

Fifth, in actual forecasting, y_n is given, in which case the conditional MSFE seems more appropriate than the unconditional MSFE. At another level, y_n is often subject to data revisions, so it may be invalid to treat its (latent) value as known; see Trivellato and Rettore (1986), Gallo (1991, 1996), and Clements and Hendry (1998a). Uncertainty from data revision goes beyond the current analysis, albeit while falling within the framework of Section 2.

5 Biases in Forecasts from Nonlinear Models

This section analyzes the forecasts of the U.S. trade balance from four structural and two time-series models, denoted Models M1–M6. All but one of these models is nonlinear, so the biases associated with deterministic forecasts from nonlinear models are examined. All six models have statistically significant biases for most forecast periods, but only three models have economically significant biases, which are primarily for s-step ahead forecasts at longer horizons. Imprecision in the Monte Carlo simulation of the biases is a concern, as it was in simulating the MSFE for ML in Section 4.2. In both instances, variance-reduction techniques reduce that imprecision. Here, antithetic variates are used because the asymptotic properties for constructing control

variates are not readily available, although Sterbenz and Calzolari (1990) give one potential approach. The lack of analytical results places an even greater premium on efficient Monte Carlo analysis. The last horizontal block in Table 2 summarizes the forecasts' design, evaluation, and post-evaluation analysis. Throughout this section, the forecast bias is assumed to exist, possibly requiring conditions on the estimation sample size relative to the forecast horizon *inter alia*, as in the linear autoregressive model; see Magnus and Pesaran (1991).

5.1 Design

This subsection describes the models to be analyzed and the corresponding forecasts of interest. The first structural model is that developed in Helkie and Hooper (1988); the remaining three structural models are treated as variations on it. The econometric model of U.S. trade in Helkie and Hooper (1988) (Model M1) is a set of four equations explaining the logarithms of import and export prices and volumes. The Helkie-Hooper model makes a variety of assumptions in those equations regarding exogeneity, dynamics, estimation technique, and level of trade disaggregation, and these assumptions are modified in Models M2–M4. Model M2 allows for a non-diagonal error covariance matrix and so uses Full Information Maximum Likelihood (FIML) to estimate the coefficients. Model M3 modifies assumptions on rationing and dynamics. Model M4, from Marquez (1989), disaggregates to bilateral trade flows. Model M5 is a fourth-order, four-variable VAR for the logarithms of export and import volumes and price indices; and Model M6 is an AR(1) model of the level of the nominal trade balance. Details of the models appear in Marquez and Ericsson (1993). Although the choice of models is not exhaustive, the Helkie-Hooper model is widely used in the literature, and VARs and univariate time-series models are often taken as non-structural alternatives.

For the current purposes, the nominal trade balance is the variable of interest. It is calculated as the difference between nominal export and import values, implying that models M1–M5 are nonlinear for the trade balance, both through the loglinear specification of the behavioral equations and from multiplying prices and volumes to obtain nominal trade flows.

5.2 Evaluation

This subsection describes deterministic forecasts, the nature of the bias from these forecasts, naive Monte Carlo procedures for estimating that bias, and the variance-reduction technique "antithetic variates" (AVs), which can reduce the imprecision of the estimated bias.

Deterministic forecasts are a common method in forecasting and were (implicitly) the forecasts analyzed in Sections 3 and 4. Formally, deterministic forecasts solve (2)

numerically, setting θ equal to its in-sample estimated value $\widehat{\theta}$ and u_{n+s} equal to its expected value (usually zero). The associated one-step ahead forecasts δ_{n+s} are:

$$\delta_{n+s} \;=\; P'g(y_{n+s-1}, z_{n+s}, \widehat{\theta}, 0) \;, \qquad s = 1, \ldots, S \;. \tag{38}$$

The s-step ahead forecasts use the same formula, but with δ_{n+s-1} replacing y_{n+s-1} ($s = 2, \ldots, S$). Equation (38) is computationally simple to implement for both one-step and s-step ahead forecasts, but, by ignoring $\widehat{\theta}$ and u_{n+s} as sources of randomness, it generates biased forecasts for nonlinear $g(\cdot)$. Bias arises because the expectation of a nonlinear function is not usually equal to the function of the expectations. That is, the conditional expectation of $P'y_{n+s}$ is:

$$\begin{aligned} \mu_{n+s} \;&\equiv\; \mathcal{E}(P'g(\widehat{y}_{n+s-1}, z_{n+s}, \widehat{\theta}, u_{n+s})) \\ &\neq\; \delta_{n+s} \;\equiv\; P'g(\delta_{n+s-1}, z_{n+s}, \widehat{\mathcal{E}}(\widehat{\theta}), \mathcal{E}(u_{n+s})) \;, \qquad s = 1, \ldots, S \;, \end{aligned} \tag{39}$$

where $\widehat{y}_n = \delta_n \equiv y_n$, $\mathcal{E}(\cdot)$ is the expectations operator, and the expectations in (39) are assumed to exist. For Models M1–M5, $g(\cdot)$ is the vector of equations for import and export prices and volumes, and additionally $g(\cdot)$ includes the identity defining the nominal trade balance. The selection matrix P extracts the nominal trade balance from $g(\cdot)$. Even if the coefficients are known, bias exists because the disturbances enter $g(\cdot)$ nonlinearly.[10] See Nagar (1969), Howrey and Kelejian (1969), and Calzolari (1979) *inter alia* for early studies that recognize systematic biases in deterministic forecasts arising from model nonlinearities. More recent studies include Mariano and Brown (1983), Wallis (1984), Fair (1984, 1988), Fisher and Salmon (1986), and Brown and Mariano (1989). Wallis (1995) provides a lucid overview of deterministic forecast bias and other issues arising with large-scale macroeconometric models.

The bias of the deterministic forecast is $\mu_{n+s} - \delta_{n+s}$. Almost by assumption, μ_{n+s} is not known analytically; and obtaining its value numerically is a complicated process involving multiple integration, i.e., the integral (3), with $g(\cdot)$ in (2) being the whole trade model. The difficulties in numerically evaluating (3) stem from both the number of random variables and the absence of a closed-form solution.[11] As an alternative to numerical integration, μ_{n+s} is estimated by Monte Carlo, using the variance-reduction technique known as antithetic variates. To discuss antithetic variates, a brief description of naive Monte Carlo simulation for the trade models is helpful.

[10] To illustrate, suppose y is normally distributed $N(\phi, \sigma^2)$ and consider the bias in the deterministic forecast of Y, where Y is the exponential of y [$Y \equiv \exp(y)$]. The deterministic forecast is $\exp(\mathcal{E}(y)) = \exp(\phi)$, but the actual expectation of Y is $\mathcal{E}(Y) = \mathcal{E}(\exp(y)) = \exp(\phi + \frac{1}{2}\sigma^2)$. Thus, the nonlinearity bias is $\exp(\phi) \cdot [\exp(\frac{1}{2}\sigma^2) - 1]$ and depends on both ϕ and σ.

[11] Note that $\widehat{\theta}$ is independent of u, a feature that would simplify solving (3) directly by numerical integration. Even so, this task remains a difficult one computationally, given the number of integrals involved. For example, model M4 has 127 coefficients and 21 (times 12 forecast periods) disturbances for a total of 379 random variables to be integrated out.

Monte Carlo analysis solves a stochastic analogue to the unknown analytical formula for μ_{n+s} in (3), simulating the effects of inherent and coefficient uncertainty, which are analytically derivable for linear models. If the bias were derived analytically, assumptions would be made about the distributional properties of the future disturbances $\{u_{n+i},\ i = 1, \ldots, S\}$ and of the coefficient estimates $\widehat{\theta}$, as in (4) and (6). Similar assumptions are made when solving for the distributional properties of forecasts by stochastic simulation. Values of $\{\widehat{\theta};\ u_{n+s},\ s = 1, \ldots, S\}$ are randomly drawn according to those distributional assumptions, with each set of values generating a path of S forecasts. An estimate of the bias can then be constructed by subtracting δ_{n+s} from the average across simulations of the sth forecast.

A natural set of distributional assumptions for $\{\widehat{\theta};\ u_{n+s},\ s = 1, \ldots, S\}$ is that $\widehat{\theta}$ takes its asymptotic distribution in finite samples and that the future disturbances $\{u_{n+s},\ s = 1, \ldots, S\}$ have the same distribution as their in-sample equivalents $\{u_t,\ t = 1, \ldots, n\}$. Thus, for the kth of K total simulations, the drawing of $\widehat{\theta}$ (denoted θ_k) is generated as:

$$\theta_k \ \sim \ N(\widehat{\theta}, \widehat{\Psi}) \ , \tag{40}$$

where Ψ is the $c \times c$ asymptotic covariance matrix of $\widehat{\theta}$, and $\widehat{\Psi}$ is the empirical estimate of Ψ. Likewise, the kth drawing of u_{n+s} (denoted $u_{k,n+s}$) is generated as:

$$u_{k,n+s} \ \sim \ IN(0, \widehat{\Omega}) \ , \qquad s = 1, \ldots, S \ , \tag{41}$$

where Ω is the $m \times m$ covariance matrix of u_t and $\widehat{\Omega}$ is its empirical estimate. That is, sets of random numbers are drawn for $\{\theta_k;\ u_{k,n+s},\ s = 1, \ldots, S\}$, which satisfy the properties in (40) and (41). In practice, mean-zero random numbers $\theta_k - \widehat{\theta}$ are generated from the distribution $N(0, \widehat{\Psi})$, with $\widehat{\theta}$ added to them afterwards to obtain θ_k in (40). This nuance is important for the generation of antithetic variates below.

At the kth replication, S forecasts are solved sequentially via (2) by combining θ_k from (40), $\{u_{k,n+s},\ s = 1, \ldots, S\}$ from (41), the initial condition $y_{k,n} = y_n$, and the paths of the exogenous variables:

$$NX_{k,n+s} \ = \ P' g(y_{k,n+s-1}, z_{n+s}, \theta_k, u_{k,n+s}) \ , \qquad s = 1, \ldots, S \ , \tag{42}$$

where $NX_{k,n+s}$ is the (kth) simulated forecast of the actual trade balance NX_{n+s}. The full stochastic simulation produces a random sample of K forecasts, from which the mean forecast (μ_{n+s}) and variance of the simulated forecasts (σ^2_{n+s}) are estimated by:

$$\widetilde{\mu}_{n+s} \ = \ \frac{\sum_{k=1}^{K} NX_{k,n+s}}{K} \ , \quad \widetilde{\sigma}^2_{n+s} \ = \ \frac{\sum_{k=1}^{K} (NX_{k,n+s} - \widetilde{\mu}_{n+s})^2}{K} \ , \qquad s = 1, \ldots, S \ . \tag{43}$$

These Monte Carlo estimators are denoted by a tilde superscript $\widetilde{\ }$ in order to distinguish them from empirical estimators such as $\widehat{\theta}$. Paralleling analytical formulae for

linear models, both approximate and asymptotic Monte Carlo calculations are possible. The matrix generalizations of $\widetilde{\mu}_{n+s}$ and $\widetilde{\sigma}^2_{n+s}$ follow immediately; see Marquez and Ericsson (1993). For the current simulations, $K = 1000$ and the forecast period is 1985Q1 to 1987Q4 ($S = 12$).

By exploiting the symmetry of the distribution of $(\{u_{k,n+s}\}, \theta_k)$, antithetic variates can provide a much more precise estimate of μ_{n+s} than that offered by the naive Monte Carlo estimator $\widetilde{\mu}_{n+s}$. To start, note that $\theta_k = \widehat{\theta} + [\theta_k - \widehat{\theta}]$, and so the values $(\{u_{k,n+s}\}, \theta_k)$ and $(\{-u_{k,n+s}\}, \widehat{\theta} - [\theta_k - \widehat{\theta}])$ are equally likely by symmetry of the normal distribution. Because much of the computational expense for a given replication k arises from generating the random numbers $(\{u_{k,n+s}\}, \theta_k)$, calculating simulated trade balances from both $(\{u_{k,n+s}\}, \theta_k)$ and $(\{-u_{k,n+s}\}, \widehat{\theta} - [\theta_k - \widehat{\theta}])$ rather than from just $(\{u_{k,n+s}\}, \theta_k)$ requires little additional expense. Obtaining the associated antithetic-variate estimate of the bias involves three steps.

1. Generate a random sample of forecasts with drawings from (40)–(41), as in (42).

2. Generate a random sample of forecasts for the trade balance using the negative of the (mean-zero) random numbers $(\{u_{k,n+s}\}, \theta_k - \widehat{\theta})$ drawn for Step 1. These forecasts for the kth replication are:

$$NX^{\dagger}_{k,n+s} \;=\; P' g(y^{\dagger}_{k,n+s-1}, z_{n+s}, \widehat{\theta} - [\theta_k - \widehat{\theta}], -u_{k,n+s}) \;, \qquad s = 1, \ldots, S \;, \quad (44)$$

   where the dagger † indicates switching the signs of the random numbers.

3. Construct the pooled Monte Carlo estimates of the forecasts by averaging the antithetic variates from Steps 1 and 2:

$$\bar{\mu}_{n+s} \;=\; \frac{\sum_{k=1}^{K} (NX_{k,n+s} + NX^{\dagger}_{k,n+s})}{2K} \;=\; \frac{\widetilde{\mu}_{n+s} + \widetilde{\mu}^{\dagger}_{n+s}}{2} \;, \qquad s = 1, \ldots, S \;. \quad (45)$$

   Thus, for K sets of random numbers, $2K$ simulated values of the trade balance are generated and are then averaged to produce a forecast of the trade balance.

To illustrate how gains from antithetic variates arise, consider the relationship between the variance of the antithetic-variate estimator $\bar{\mu}_{n+s}$ and the variance of the standard Monte Carlo estimator $\widetilde{\mu}_{n+s}$:

$$\mathrm{var}(\bar{\mu}_{n+s}) \;=\; \frac{(1 + \rho_{n+s}) \cdot \sigma^2_{n+s}}{2K} \;\leq\; \frac{\sigma^2_{n+s}}{K} \;=\; \mathrm{var}(\widetilde{\mu}_{n+s}) \;, \quad (46)$$

where ρ_{n+s} is the correlation coefficient between $\widetilde{\mu}_{n+s}$ and $\widetilde{\mu}^{\dagger}_{n+s}$. The inequality in (46) is strict, provided that ρ_{n+s} is less than unity. This correlation is also the correlation between $NX_{k,n+s}$ and $NX^{\dagger}_{k,n+s}$.

In principle, even greater efficiency gains can be obtained by choosing unequal (normalized) weights on $NX_{k,n+s}$ and $NX^{\dagger}_{k,n+s}$ when calculating $\bar{\mu}_{n+s}$, rather than the equal weights imposed in (45). Davidson and MacKinnon (1992) suggest calculating the weights by the following regression:

$$NX_{k,n+s} \;=\; \kappa_0 + \kappa_1 NX^{\dagger}_{k,n+s} + \zeta_k, \qquad k = 1, \ldots, K \;, \tag{47}$$

where the coefficient on $NX^{\dagger}_{k,n+s}$ is κ_1, which should be approximately -1 from (45); $\kappa_0/2$ is an estimate of μ_{n+s}; and ζ_k is the remaining simulation error. Davidson and MacKinnon (1992) also propose a similar regression for estimating the optimal weight on a control variate when constructing a pooled estimator.

Equation (47) has a disadvantage. It treats $NX_{k,n+s}$ and $NX^{\dagger}_{k,n+s}$ asymmetrically, normalizing on the first and conditioning on the second, even though nothing favors normalizing on $NX_{k,n+s}$ rather than $NX^{\dagger}_{k,n+s}$. Reversing the roles of $NX_{k,n+s}$ and $NX^{\dagger}_{k,n+s}$ in (47) does not solve this anomaly. However, one regression does treat $NX_{k,n+s}$ and $NX^{\dagger}_{k,n+s}$ symmetrically, namely:

$$\frac{NX_{k,n+s} + NX^{\dagger}_{k,n+s}}{2} \;=\; \lambda_0 + \lambda_1 (NX_{k,n+s} - NX^{\dagger}_{k,n+s}) + \xi_k, \qquad k = 1, \ldots, K \;, \tag{48}$$

where λ_1 represents the deviation between the implied weight on $NX^{\dagger}_{k,n+s}$ in the regression-based antithetic variate and the imposed weight of $\frac{1}{2}$ in the standard antithetic variate (45); λ_0 is an estimate of μ_{n+s}; and ξ_k is the remaining simulation error. In effect, (48) starts with the standard antithetic variate formula as its dependent variable and then calculates the extent to which the equal weighting on $(NX_{k,n+s}, NX^{\dagger}_{k,n+s})$ might deviate from $(\frac{1}{2}, \frac{1}{2})$, while imposing that the weights sum to unity. By contrast, neither (47) nor the corresponding reverse regression ensure that the sum of the weights on $NX_{k,n+s}$ and $NX^{\dagger}_{k,n+s}$ is fixed.

To test whether or not the deterministic forecast is biased, compute the t-ratio ψ_{n+s} for the hypothesis $\mu_{n+s} = \delta_{n+s}$:

$$\psi_{n+s} \;=\; \frac{\bar{\mu}_{n+s} - \delta_{n+s}}{\sqrt{\widetilde{\mathrm{var}}(\bar{\mu}_{n+s})}} \;=\; \frac{K^{1/2} \cdot (\bar{\mu}_{n+s} - \delta_{n+s})}{\sqrt{\widetilde{\mathrm{var}}\left(\dfrac{NX_{k,n+s} + NX^{\dagger}_{k,n+s}}{2}\right)}} \;, \tag{49}$$

where the variance of the average $(NX_{k,n+s}+NX^{\dagger}_{k,n+s})/2$ is estimated as was the variance of $NX_{k,n+s}$ from $\widetilde{\sigma}^2_{n+s}$ in (43), but with $(NX_{k,n+s}+NX^{\dagger}_{k,n+s})/2$ and $\bar{\mu}_{n+s}$ replacing $NX_{k,n+s}$ and $\widetilde{\mu}_{n+s}$ in that formula. Selection of a large enough number of replications K will make any nonzero bias statistically significant, so it is also of interest to consider the numerical importance of the estimated bias, $\bar{\mu}_{n+s} - \delta_{n+s}$.

For discussions of antithetic variates, see Hammersley and Handscomb (1964) and Hendry (1984). For applications of antithetic variates to forecasts from econometric

models, see Calzolari (1979), Fisher and Salmon (1986), Mariano and Brown (1989), and Gallo and Don (1991), noting that they all treat the coefficient estimates as though they were known with certainty, whereas this section relaxes that assumption.

5.3 Post-evaluation Analysis: Biases

Table 3 reports the estimated bias $\bar{\mu}_{n+s} - \delta_{n+s}$ and the estimated standard error of the estimated bias $[\widetilde{\text{var}}(\bar{\mu}_{n+s})]^{1/2}$ for both one-step and s-step ahead forecasts, all using the approximate formula. Their ratios are the corresponding t-ratios. Estimated biases, standard errors, and t-ratios using the asymptotic formula appear in Marquez and Ericsson (1990) and are qualitatively the same, with a few minor exceptions noted below.

Some deterministic forecasts from all models have statistically significant nonlinearity biases, summarized as follows. First, the estimated biases from all models are usually highly statistically significant, although not as often or as much for M3 (one-step ahead) or M1 (s-step ahead). Overall, 90% of the one-step ahead forecasts have t-ratios for biases that are larger than 3 in absolute value (80% for s-step ahead forecasts), and half have t-ratios greater than 10 (one-third for s-step ahead forecasts). Second, the size of the bias changes with the forecast period, even switching sign in some cases. Third, for Helkie and Hooper's model, the choice of estimation method hardly affects either the magnitude or the profile of the bias.

Despite their statistical significance, the nonlinearity biases need not be economically significant. For models M1–M3, the deterministic and stochastic forecasts differ by only \$2 billion at most, although often statistically significantly so. By contrast, models M4–M6 have much larger and economically potentially important biases, e.g., of approximately \$10 billion at twelve periods ahead. The sometime difference between economically significant and statistically significant biases stems from the considerable accuracy with which antithetic variates estimate the mean of the forecast for the trade balance, making numerically slight departures from this mean statistically significant.

The numerically large biases for s-step ahead forecasts from models M4, M5, and M6 appear to arise for distinctly different reasons, e.g., dynamics and the value of y_n for M5 and M6, and disaggregation for M4. Model M6 has (analytically) zero bias using the asymptotic formula (no uncertainty from estimation) but substantial bias with the approximate formula. That bias arises from the nonlinear way in which dynamics affects multi-step ahead forecasts, in combination with the forecasts being conditional on NX_n. Specifically, the s-step ahead forecast error for model M6 is (19), where $y = NX$ and the intercept is ignored for expositional simplicity. The errors are assumed normal, so the linear combination of them (the first term on the right-hand side of (19)) is also normal. The asymptotic formula sets $\beta = \widehat{\beta}$, ignoring the uncertainty from estimation (the second term on the right-hand side of (19)) and making

Table 3. Nonlinearity Biases in Deterministic Forecasts of the U.S. Trade Balance
Approximate Formula (billion US$)

Model	Forecast Period											
	1985				1986				1987			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
					Historical Trade Balance							
	–99.5	–119.8	–124.8	–144.5	–141.7	–135.4	–146.9	–154.1	–159.5	–158.2	–158.7	–164.8
					One-step Ahead Forecasts							
					Estimated Bias							
M1	–0.2	–0.2	–0.2	–0.3	–0.4	–0.5	–0.6	–0.8	–1.0	–1.0	–1.2	–1.3
M2	–0.3	–0.2	–0.3	–0.4	–0.6	–0.5	–0.6	–0.7	–0.8	–1.0	–1.1	–1.4
M3	–0.1	–0.1	–0.0	–0.1	–0.1	–0.0	–0.0	–0.0	–0.1	–0.1	–0.1	–0.1
M4	3.3	–1.1	–1.1	–4.9	–0.7	–5.0	–2.5	1.1	0.1	7.9	9.7	7.7
M5	–0.2	–0.2	–0.2	–0.2	–0.3	–0.2	–0.2	–0.4	–0.4	–0.4	–0.4	–0.4
M6	0	0	0	0	0	0	0	0	0	0	0	0
					Estimated Standard Error of the Bias							
M1	0.040	0.043	0.047	0.053	0.057	0.062	0.066	0.072	0.076	0.081	0.090	0.100
M2	0.036	0.039	0.044	0.049	0.053	0.053	0.054	0.060	0.067	0.075	0.088	0.101
M3	0.016	0.016	0.016	0.017	0.018	0.018	0.021	0.020	0.021	0.022	0.023	0.026
M4	0.039	0.045	0.047	0.050	0.057	0.062	0.062	0.061	0.069	0.078	0.082	0.094
M5	0.019	0.019	0.020	0.021	0.023	0.022	0.022	0.030	0.037	0.032	0.034	0.038
M6	0	0	0	0	0	0	0	0	0	0	0	0
					s-step Ahead Forecasts							
					Estimated Bias							
M1	–0.2	–0.0	–0.1	–0.0	–0.1	0.0	–0.1	–0.1	–0.4	–0.7	–0.8	–1.0
M2	–0.3	–0.2	–0.1	–0.3	–0.3	–0.3	–0.1	–0.1	–0.1	–0.4	–0.8	–1.3
M3	–0.1	–0.2	–0.2	–0.3	–0.4	–0.6	–0.7	–0.8	–0.7	–0.8	–0.7	–0.7
M4	3.3	–1.2	–1.0	–4.5	–0.0	–4.4	–2.3	0.8	–0.7	7.2	9.8	7.7
M5	–0.2	–0.7	–0.6	–0.8	–1.6	–2.5	–3.1	–3.7	–4.5	–5.0	–5.6	–6.8
M6	0	–0.1	–0.4	–0.8	–1.3	–2.1	–3.1	–4.3	–5.8	–7.6	–9.8	–12.4
					Estimated Standard Error of the Bias							
M1	0.040	0.064	0.084	0.099	0.104	0.111	0.118	0.123	0.129	0.137	0.148	0.150
M2	0.036	0.057	0.070	0.082	0.090	0.096	0.102	0.114	0.120	0.128	0.136	0.146
M3	0.016	0.041	0.058	0.076	0.089	0.096	0.106	0.118	0.135	0.149	0.164	0.178
M4	0.039	0.066	0.083	0.100	0.115	0.136	0.145	0.155	0.171	0.202	0.212	0.240
M5	0.019	0.120	0.235	0.341	0.446	0.546	0.670	0.812	0.952	1.078	1.218	1.349
M6	0	0.011	0.027	0.051	0.082	0.124	0.177	0.243	0.323	0.418	0.531	0.665

Notes: See Table 4.

the corresponding forecast unbiased. In the approximate formula, $\widehat{\beta}$ is assumed normally distributed, unbiased for β. Under that assumption, $\widehat{\beta}^s$ is generally biased for β^s $(s > 1)$. In Table 3, the nonzero nonlinearity biases for M6 reflect that nonlinearity in $\widehat{\beta}$, multiplied by a given y_n. Model M5 involves the same issues at a multivariate level, and also includes the nonlinear transformations from logs to levels and from levels to products of levels. Although dynamics may contribute to the nonlinearity biases for M4, other explanations could be more important. Model M4 is much more disaggregated (by country) than the other models and, if the relative fluctuations of the disaggregated variables are larger than of the aggregate, nonlinearity biases could well be more pronounced for the disaggregated model (M4).

Overall, the sensitivity of the nonlinearity bias to the forecast period, estimation method, level of aggregation, and dynamic specification suggests that no simple correction exists for the bias, other than simulation. When a substantive nonlinearity bias is present, deterministic forecasts will be at an inherent disadvantage in terms of forecasting performance relative to (unbiased) stochastic forecasts. All of the models examined have statistically significant nonlinearity biases, of which three have numerically substantive biases. More generally, since nonlinearity is a feature common to most macro-econometric models, deterministic forecasts from them should be interpreted with caution, at least until the magnitude of the bias is assessed.

For the structural trade models M1–M4, extreme values for forecasts may arise because the simulated structural coefficients are assumed normally distributed whereas the forecasts are obtained from the reduced form, which involves inverting a function of those normally distributed coefficient simulations; cf. McCarthy (1972). While important in principle, such extreme values do not appear to be a problem for the models involved. Marquez and Ericsson (1993, Tables 14A–6 and 14A–7) calculate skewness, excess kurtosis, and the Jarque-Bera statistic for all the model's forecast simulations and find little evidence of nonnormality, contradicting a basic implication of the extreme values hypothesis. Histograms and estimated densities for all the models' forecasts (available from the authors) provide further evidence that wild values are not a particular problem for these forecasts.

5.4 Post-evaluation Analysis: Efficiency Gains

This subsection calculates the efficiency gain from the use of antithetic variates in estimating the nonlinearity bias of the deterministic forecasts. The efficiency gain is commonly measured as $\widetilde{\text{var}}(\widetilde{\mu}_{n+s})/\widetilde{\text{var}}(\bar{\mu}_{n+s})$, the ratio of the estimated variances of the naive and antithetic-variate estimates of the bias. As that ratio, the efficiency gain measures the reduction in variance obtained by antithetic variates. Conversely, the efficiency gain represents the $\times$-fold number of replications that would be required to obtain a naive mean forecast with the same variance as that of the mean forecast using antithetic variates.

Table 4 lists the estimated efficiency gains for one-step and s-step ahead forecasts of all the models, both for the asymptotic formula and for the approximate formulae. For models M1–M4, the efficiency gains range from around 20 to over 2000. In those instances with very high efficiency gains, a single replication with antithetic variates would have produced a mean forecast with a smaller variance than that of the naive mean forecast obtained from the 1000 replications in the current simulation study. Conversely, some of the mean forecasts obtained with antithetic variates have a smaller variance than that of a naive mean forecast using *two million* replications.

The efficiency gains for the VAR (M5) typically are between 100 and 1000, but drop to as low as 4 for s-step ahead forecasts with the approximate formula. For model M6, the estimated gains for s-step ahead forecasts ($s > 1$) with the approximate formula range from 17 to almost 2000. For this model, efficiency gains are infinite for all s-step ahead forecasts with the asymptotic formula and for all one-step ahead forecasts with either formula because these forecasts are linear in the terms simulated. This has the following explanation.

In linear models, the one-step ahead forecast is linear in the parameters (and so in $\theta_k - \widehat{\theta}$) and is not a function of the disturbances, so a single replication provides a completely accurate estimate of the expected outcome, i.e., of the expectation μ_{n+s} in (3). Specifically, suppose that the model is $y_t = z_t'\theta + u_t$ with a symmetrically distributed estimator of θ denoted $\widehat{\theta}$ and that the one-step ahead forecast of y_{n+1} is $z_{n+1}'\widehat{\theta}$. Then $\widehat{\theta} + [\theta_k - \widehat{\theta}]$ and $\widehat{\theta} - [\theta_k - \widehat{\theta}]$ are equally likely to occur. The two resulting forecasts are $z_{n+1}'(\widehat{\theta} + [\theta_k - \widehat{\theta}])$ and $z_{n+1}'(\widehat{\theta} - [\theta_k - \widehat{\theta}])$, whose average is $z_{n+1}'\widehat{\theta}$. This is the average for any and all $\theta_k - \widehat{\theta}$ drawn, so the expectation in (3) is $z_{n+1}'\widehat{\theta}$. Here, the gains from antithetic variates are infinite, since no replications were actually necessary to find the exact forecast. See Hendry and Trivedi (1972, p. 120, footnote 4), who consider a similar situation of the bias of OLS with fixed regressors and who use antithetic variates for analyzing the standard deviation and bias of one-step ahead forecasts.

With s-step ahead forecasts from possibly nonlinear dynamic models, antithetic variates reduce or eliminate the Monte Carlo variation from the weighted sum of future disturbances, and reduce the variation from coefficient uncertainty (which is nonlinear in the estimated coefficients). Equation (1) partitions these two components clearly. Because (1) arises from a linear model, antithetic variates would eliminate the variation from the weighted sum of future shocks.

The incremental improvements from the regression-based antithetic variates in (48) are marginal. Out of 251 regressions, the absolute value of the estimated λ_1 exceeds 0.01 in only 5 instances and never exceeds 0.02. Only 14 instances have incremental efficiency gains of more than 1%, and none have incremental gains exceeding 3%. E.g., for a 1000-fold efficiency gain reported in Table 4, regression-based antithetic variates would at best increase that to 1030-fold. These minor incremental gains are consistent with Davidson and MacKinnon (1992, pp. 297–298).

Table 4.
Efficiency Gains from Antithetic Variates
In Estimating the Nonlinearity Bias of Deterministic Forecasts

Model		Forecast Period											
Horizon:	Formula	1985				1986				1987			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
M1													
1-step:	Asymptotic	1783	1821	2049	1890	1905	1886	1771	1875	1819	1827	1828	1716
	Approximate	109	92	83	72	64	53	53	50	52	54	50	46
s-step:	Asymptotic	1783	1526	1379	1268	1469	1286	1234	1392	1340	1235	1290	1319
	Approximate	109	59	42	38	36	31	31	34	39	43	44	51
M2													
1-step:	Asymptotic	1833	1938	2002	1880	1627	1672	1561	1835	1801	1769	1791	1606
	Approximate	127	112	86	81	74	64	64	58	55	52	47	46
s-step:	Asymptotic	1833	1538	1225	1076	1078	914	770	790	759	818	842	772
	Approximate	127	70	51	49	45	36	36	33	39	42	48	53
M3													
1-step:	Asymptotic	1546	1564	1603	1547	1417	1398	1298	1447	1508	1477	1549	1353
	Approximate	724	689	697	660	589	565	490	545	551	527	544	465
s-step:	Asymptotic	1546	1238	930	815	807	722	640	702	677	686	676	658
	Approximate	724	169	103	75	58	47	42	37	37	32	31	29
M4													
1-step:	Asymptotic	651	515	555	574	501	485	531	473	577	580	505	555
	Approximate	217	183	179	196	166	125	133	129	145	125	124	108
s-step:	Asymptotic	651	484	444	391	402	360	350	361	403	382	356	331
	Approximate	217	91	66	58	49	37	34	30	32	27	26	23
M5													
1-step:	Asymptotic	880	715	870	802	889	939	774	906	889	860	893	830
	Approximate	631	592	557	513	549	519	539	413	332	414	395	347
s-step:	Asymptotic	880	341	251	208	138	115	95	78	67	58	51	48
	Approximate	631	39	15	10	8	8	8	7	6	5	5	4
M6													
1-step:	Asymptotic	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
	Approximate	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
s-step:	Asymptotic	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
	Approximate	∞	1750	460	210	125	82	56	41	31	25	20	17

Notes:
Model M1: Helkie and Hooper (1988).
Model M2: Model M1 by FIML.
Model M3: Model M1 by FIML, with partial adjustment, without non-price rationing.
Model M4: Bilateral trade model by FIML.
Model M5: VAR(4) of import and export volumes and price indices.
Model M6: AR(1) of the nominal trade balance.

Mean forecasts using antithetic variates often result in markedly improved estimates of the nonlinearity bias from deterministic forecasts, requiring virtually no additional programming and at most twice the computational time. Indeed, little justification remains for using naive Monte Carlo rather than antithetic variates, in so far as estimating the nonlinearity bias is concerned. Antithetic variates as a technique is applicable to calculating the deterministic bias arising either from nonzero future disturbances only ("asymptotic") or from nonzero future disturbances and coefficient uncertainty ("approximate"). The simulation above appears to be the first analyzing the bias from both sources by means of antithetic variates.

The precise gain from antithetic variates depends upon the forecast period, the nonlinearity of the process, and whether the asymptotic or approximate formula is used. However, the large gains in Table 4 are in line with other authors' results; cf. Hammersley and Handscomb (1964, p. 65), Calzolari (1979) and Fisher and Salmon (1986). Interestingly, efficiency gains for the nonlinearity bias of deterministic forecasts can be substantial for purely dynamic processes: that contrasts with the zero incremental gains from antithetic variates when measuring biases in the estimation of parameters from linear autoregressive processes. However, antithetic variates probably would obtain only small gains for estimates of forecast variances because those variances are nearly invariant to the sign of the random numbers drawn for (40) and (41). For one-step ahead forecasts from a linear model, the forecast variance is invariant to the sign, in which case the correlation ρ_{n+s} is unity and no gains accrue.

6 Conclusions

This paper advances a tripartite framework of design, evaluation, and post-evaluation analysis for generating and interpreting economic forecasts. These concepts are not new in themselves, and they have been previously discussed in a narrower context for Monte Carlo analysis. The current paper generalizes these concepts and applies them to forecasting, re-interpreting the various aspects of forecasting as parts of a whole instead of as a collection of loosely related tasks.

This framework is illustrated by the Schmidt-Baillie formula for the MSFE of a VAR, with specific applications to the AR(1) model as studied by Orcutt and Winokur (1969), Hoque, Magnus, and Pesaran (1988), and ourselves. This framework is also applied to forecasts from empirical models of U.S. external trade. Both classes of models are characterized by properties of economic interest: the possible nonmonotonicity and nonexistence of the MSFE, and the nonlinearity bias of deterministic forecasts. Each property has been studied before, but in isolation from other aspects of the forecasting process. The framework advanced herein reveals how a clearer understanding of each property results from integrating all the activities generating the respective forecasts.

References

Anderson, T. W. (1971) *The Statistical Analysis of Time Series*, John Wiley, New York.

Ansley, C. F., and P. Newbold (1980) "Finite Sample Properties of Estimators for Autoregressive Moving Average Models", *Journal of Econometrics*, 13, 2, 159–183.

Baillie, R. T. (1979a) "The Asymptotic Mean Squared Error of Multistep Prediction from the Regression Model with Autoregressive Errors", *Journal of the American Statistical Association*, 74, 365, 175–184.

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

Baillie, R. T. (1981) "Prediction from the Dynamic Simultaneous Equation Model with Vector Autoregressive Errors", *Econometrica*, 49, 5, 1331–1337.

Beach, C. M., and J. G. MacKinnon (1978) "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors", *Econometrica*, 46, 1, 51–58.

Bianchi, C. (1990) "Ricampionamento e sottostima del 'Delta Method': alcuni controesempi econometrici", *Quaderni di Statistica e Econometria*, 12, 87–104.

Bianchi, C., G. Calzolari, and J.-L. Brillet (1987) "Measuring Forecast Uncertainty: A Review with Evaluation Based on a Macro Model of the French Economy", *International Journal of Forecasting*, 3, 2, 211–227.

Borland (1993) *Quattro Pro for Windows Version 5.0: User's Guide*, Borland International, Scotts Valley, California.

Box, G. E. P., and G. M. Jenkins (1970) *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Box, G. E. P., and M. E. Muller (1958) "A Note on the Generation of Random Normal Deviates", *Annals of Mathematical Statistics*, 29, 2, 610–611.

Brown, B. W., and R. S. Mariano (1989) "Predictors in Dynamic Nonlinear Models: Large-sample Behavior", *Econometric Theory*, 5, 3, 430–452.

Calzolari, G. (1979) "Antithetic Variates to Estimate the Simulation Bias in Nonlinear Models", *Economics Letters*, 4, 4, 323–328.

Calzolari, G. (1981) "A Note on the Variance of Ex-post Forecasts in Econometric Models", *Econometrica*, 49, 6, 1593–1595.

Calzolari, G. (1987) "Forecast Variance in Dynamic Simulation of Simultaneous Equation Models", *Econometrica*, 55, 6, 1473–1476.

Calzolari, G., and F. P. Sterbenz (1986) "Control Variates to Estimate the Reduced Form Variances in Econometric Models", *Econometrica*, 54, 6, 1483–1490.

Campos, J. (1992) "Confidence Intervals for Linear Combinations of Forecasts from Dynamic Econometric Models", *Journal of Policy Modeling*, 14, 4, 535–560.

Carrier, N., E. Atkins, and C. Taylor (1969) "Report on the London Atlas Random Number Generator", mimeo, University of London Computing Centre, London, December.

Chong, Y. Y., and D. F. Hendry (1986) "Econometric Evaluation of Linear Macroeconomic Models", *Review of Economic Studies*, 53, 4, 671–690.

Chow, G. C. (1960) "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 3, 591–605.

Clements, M. P., and D. F. Hendry (1993) "On the Limitations of Comparing Mean Square Forecast Errors", *Journal of Forecasting*, 12, 8, 617–637 (with discussion).

Clements, M. P., and D. F. Hendry (1995) "Forecasting in Cointegrated Systems", *Journal of Applied Econometrics*, 10, 2, 127–146.

Clements, M. P., and D. F. Hendry (1996) "Intercept Corrections and Structural Change", *Journal of Applied Econometrics*, 11, 5, 475–494.

Clements, M. P., and D. F. Hendry (1998a) *Forecasting Economic Time Series: The Marshall Lectures on Economic Forecasting*, Cambridge University Press, Cambridge, England (forthcoming).

Clements, M. P., and D. F. Hendry (1998b) *The Zeuthen Lectures on Economic Forecasting*, MIT Press, Cambridge, Massachusetts (forthcoming).

Davidson, R., and J. G. MacKinnon (1981) "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, 49, 3, 781–793.

Davidson, R., and J. G. MacKinnon (1992) "Regression-based Methods for Using Control Variates in Monte Carlo Experiments", *Journal of Econometrics*, 54, 1–3, 203–222.

Doornik, J. A., and D. F. Hendry (1996) *PcGive Professional 9.0 for Windows*, International Thomson Business Press, London.

Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983) "Exogeneity", *Econometrica*, 51, 2, 277–304.

Ericsson, N. R. (1986) "Post-simulation Analysis of Monte Carlo Experiments: Interpreting Pesaran's (1974) Study of Non-nested Hypothesis Test Statistics", *Review of Economic Studies*, 53, 4, 691–707.

Ericsson, N. R. (1992) "Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration", *Journal of Policy Modeling*, 14, 4, 465–495.

Ericsson, N. R., and J. Marquez (1989) "Exact and Approximate Multi-period Mean-square Forecast Errors for Dynamic Econometric Models", International Finance Discussion Paper No. 348, Board of Governors of the Federal Reserve System, Washington, D.C., April.

Ericsson, N. R., and J. Marquez (1993) "Encompassing the Forecasts of U.S. Trade Balance Models", *Review of Economics and Statistics*, 75, 1, 19–31.

Fair, R. C. (1980) "Estimating the Expected Predictive Accuracy of Econometric Models", *International Economic Review*, 21, 2, 355–378.

Fair, R. C. (1984) *Specification, Estimation, and Analysis of Macroeconometric Models*, Harvard University Press, Cambridge, Massachusetts.

Fair, R. C. (1988) "Sources of Economic Fluctuations in the United States", *Quarterly Journal of Economics*, 103, 2, 313–332.

Fair, R. C., and R. J. Shiller (1989) "The Informational Content of Ex Ante Forecasts", *Review of Economics and Statistics*, 71, 2, 325–331.

Fisher, P. G., and M. Salmon (1986) "On Evaluating the Importance of Nonlinearity in Large Macroeconometric Models", *International Economic Review*, 27, 3, 625–646.

Fisher, P. G., and K. F. Wallis (1990) "The Historical Tracking Performance of UK Macroeconometric Models 1978–85", *Economic Modelling*, 7, 2, 179–197.

Fuller, W. A., and D. P. Hasza (1980) "Predictors for the First-order Autoregressive Process", *Journal of Econometrics*, 13, 2, 139–157.

Gallo, G. M. (1991) "Forecast Error Decomposition in a Nonlinear Model with Provisional Data", *Annales d'Économie et de Statistique*, 22, April–June, 103–128.

Gallo, G. M. (1996) "Forecast Uncertainty Reduction in Nonlinear Models", *Journal of the Italian Statistical Society*, 5, 1, 73–98.

Gallo, G. M., and F. J. H. Don (1991) "Forecast Uncertainty due to Unreliable Data", *Economic and Financial Computing*, 1, Spring, 49–69.

Goldberger, A. S., A. L. Nagar, and H. S. Odeh (1961) "The Covariance Matrices of Reduced-form Coefficients and of Forecasts for a Structural Econometric Model", *Econometrica*, 29, 4, 556–573.

Haitovsky, Y., and N. Wallace (1972) "A Study of Discretionary and Nondiscretionary Monetary and Fiscal Policies in the Context of Stochastic Macroeconometric Models", in V. Zarnowitz (ed.) *The Business Cycle Today*, National Bureau of Economic Research, New York, 261–309 (Fiftieth Anniversary Colloquium I).

Hall, P. (1994) "Methodology and Theory for the Bootstrap", Chapter 39 in R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics*, Volume 4, North-Holland, Amsterdam, 2341–2381.

Hammersley, J. M., and D. C. Handscomb (1964) *Monte Carlo Methods*, Chapman and Hall, London.

Helkie, W. L., and P. Hooper (1988) "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*, Brookings Institution, Washington, D.C., 10–56.

Hendry, D. F. (1979) "Predictive Failure and Econometric Modelling in Macroeconomics: The Transactions Demand for Money", Chapter 9 in P. Ormerod (ed.) *Economic Modelling: Current Issues and Problems in Macroeconomic Modelling in the UK and the US*, Heinemann Education Books, London, 217–242.

Hendry, D. F. (1984) "Monte Carlo Experimentation in Econometrics", Chapter 16 in Z. Griliches and M. D. Intriligator (eds.) *Handbook of Econometrics*, Volume 2, North-Holland, Amsterdam, 937–976.

Hendry, D. F. (1997) "The Econometrics of Macroeconomic Forecasting", *Economic Journal*, 107, 444, 1330–1357.

Hendry, D. F., and R. W. Harrison (1974) "Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-stage Least Squares", *Journal of Econometrics*, 2, 2, 151–174.

Hendry, D. F., and P. K. Trivedi (1972) "Maximum Likelihood Estimation of Difference Equations with Moving Average Errors: A Simulation Study", *Review of Economic Studies*, 39, 2, 117–145.

Hoque, A., J. R. Magnus, and B. Pesaran (1988) "The Exact Multi-period Mean-square Forecast Error for the First-order Autoregressive Model", *Journal of Econometrics*, 39, 3, 327–346.

Howrey, P., and H. H. Kelejian (1969) "Simulation Versus Analytical Solutions", in T. H. Naylor (ed.) *The Design of Computer Simulation Experiments*, Duke University Press, Durham, North Carolina, 207–231.

Intex Solutions (1989) *TROLL Version 13, Reference Manuals*, Intex Solutions, Needham, Massachusetts (Volumes 1–6 and Updates).

Koopmans, T. C. (1942) "Serial Correlation and Quadratic Forms in Normal Variables", *Annals of Mathematical Statistics*, 13, 1, 14–34.

Lu, M., and G. E. Mizon (1991) "Forecast Encompassing and Model Evaluation", Chapter 9 in P. Hackl and A. H. Westlund (eds.) *Economic Structural Change: Analysis and Forecasting*, Springer-Verlag, Berlin, 123–138.

Maekawa, K. (1987) "Finite Sample Properties of Several Predictors from an Autoregressive Model", *Econometric Theory*, 3, 3, 359–370.

Magnus, J. R., and B. Pesaran (1989) "The Exact Multi-period Mean-square Forecast Error for the First-order Autoregressive Model with an Intercept", *Journal of Econometrics*, 42, 2, 157–179.

Magnus, J. R., and B. Pesaran (1991) "The Bias of Forecasts from a First-order Autoregression", *Econometric Theory*, 7, 2, 222–235.

Mann, H. B., and A. Wald (1943) "On Stochastic Limit and Order Relationships", *Annals of Mathematical Statistics*, 14, 3, 217–226.

Mariano, R. S., and B. W. Brown (1983) "Asymptotic Behavior of Predictors in a Nonlinear Simultaneous System", *International Economic Review*, 24, 3, 523–536.

Mariano, R. S., and B. W. Brown (1989) "Stochastic Simulation, Prediction and Validation of Nonlinear Models", Chapter 2 in L. R. Klein and J. Marquez (eds.) *Economics in Theory and Practice: An Eclectic Approach*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 17–36.

Marquez, J. (1989) "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*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 129–176.

Marquez, J. (1991) "The Dynamics of Uncertainty or the Uncertainty of Dynamics: Stochastic J-Curves", *Review of Economics and Statistics*, 73, 1, 125–133.

Marquez, J., and N. R. Ericsson (1990) "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., March.

Marquez, J., and N. R. Ericsson (1993) "Evaluating Forecasts of the U.S. Trade Balance", Chapter 14 in R. C. Bryant, P. Hooper, and C. L. Mann (eds.) *Evaluating Policy Regimes: New Research in Empirical Macroeconomics*, Brookings Institution, Washington, D.C., 671–732.

McCarthy, M. D. (1972) "A Note on the Forecasting Properties of Two Stage Least Squares Restricted Reduced Forms — The Finite Sample Case", *International Economic Review*, 13, 3, 757–761.

Mizon, G. E., and J.-F. Richard (1986) "The Encompassing Principle and its Application to Testing Non-nested Hypotheses", *Econometrica*, 54, 3, 657–678.

Nagar, A. L. (1969) "Stochastic Simulation of the Brookings Econometric Model", Chapter 12 in J. S. Duesenberry, G. Fromm, L. R. Klein, and E. Kuh (eds.) *The Brookings Model: Some Further Results*, North-Holland, Amsterdam, 425–456.

Neave, H. R. (1973) "On Using the Box-Muller Transformation with Multiplicative Congruential Pseudo-random Number Generators", *Applied Statistics*, 22, 1, 92–97.

Numerical Algorithms Group (1984) *Handbook for the NAG Fortran PC50 Library — Release 1*, Numerical Algorithms Group, Oxford (Edition 2, November).

Orcutt, G. H., and H. S. Winokur, Jr. (1969) "First Order Autoregression: Inference, Estimation, and Prediction", *Econometrica*, 37, 1, 1–14.

Pagan, A. (1989) "On the Role of Simulation in the Statistical Evaluation of Econometric Models", *Journal of Econometrics*, 40, 1, 125–139.

Peters, S. C., and D. A. Freedman (1985) "Using the Bootstrap to Evaluate Forecasting Equations", *Journal of Forecasting*, 4, 3, 251–262.

Phillips, P. C. B. (1979) "The Sampling Distribution of Forecasts from a First-order Autoregression", *Journal of Econometrics*, 9, 3, 241–261.

Sargan, J. D. (1975) "Asymptotic Theory and Large Models", *International Economic Review*, 16, 1, 75–91.

Sargan, J. D. (1982) "On Monte Carlo Estimates of Moments That Are Infinite", in R. L. Basmann and G. F. Rhodes, Jr. (eds.) *Advances in Econometrics: A Research Annual*, Volume 1, JAI Press, Greenwich, Connecticut, 267–299.

Schmidt, P. (1974) "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model", *Econometrica*, 42, 2, 303–309.

Shenton, L. R., and W. L. Johnson (1965) "Moments of a Serial Correlation Coefficient", *Journal of the Royal Statistical Society, Series B*, 27, 2, 308–320.

Sims, C. A. (1974) "Distributed Lags", Chapter 5 in M. D. Intriligator and D. A. Kendrick (eds.) *Frontiers of Quantitative Economics*, Volume 2, North-Holland, Amsterdam, 289–332 (with discussion).

Sterbenz, F. P., and G. Calzolari (1990) "Alternative Specifications of the Error Process in the Stochastic Simulation of Econometric Models", *Journal of Applied Econometrics*, 5, 2, 137–150.

Trivellato, U., and E. Rettore (1986) "Preliminary Data Errors and Their Impact on the Forecast Error of Simultaneous-equations Models", *Journal of Business and Economic Statistics*, 4, 4, 445–453.

Tufte, E. R. (1983) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.

Tufte, E. R. (1990) *Envisioning Information*, Graphics Press, Cheshire, Connecticut.

Tufte, E. R. (1997) *Visual Explanations: Images and Quantities, Evidence and Narrative*, Graphics Press, Cheshire, Connecticut.

Wallis, K. F. (1984) "Comparing Time-series and Nonlinear Model-based Forecasts", *Oxford Bulletin of Economics and Statistics*, 46, 4, 383–389.

Wallis, K. F. (1995) "Large-scale Macroeconometric Modeling", Chapter 6 in M. H. Pesaran and M. R. Wickens (eds.) *Handbook of Applied Econometrics: Macroeconomics*, Blackwell Publishers, Oxford, 312–355.

West, K. D. (1996) "Asymptotic Inference About Predictive Ability", *Econometrica*, 64, 5, 1067–1084.