REGRESSION-BASED TESTS OF PREDICTIVE ABILITY*

BY   KENNETH D. WEST AND MICHAEL W. MCCRACKEN[†1]

University of Wisconsin, U.S.A.
Louisiana State University, U.S.A.

We develop regression-based tests of hypotheses about out of sample predic-
tion errors. Representative tests include ones for zero mean and zero correla-
tion between a prediction error and a vector of predictors. The relevant
environments are ones in which predictions depend on estimated parameters.
We show that standard regression statistics generally fail to account for errors
introduced by estimation of these parameters. We propose computationally
convenient test statistics that properly account for such errors. Simulations
indicate that the procedures can work well in samples of size typically available,
although there sometimes are substantial size distortions.

1. INTRODUCTION

In this paper, we develop and simulate regression tests for properties of out of
sample prediction errors. Examples of such properties are: zero mean, zero serial
correlation (if the prediction is one-step ahead), zero correlation with the prediction,
and zero correlation with the prediction from another, non-nested model. Empirical
papers that examine these or related properties include Mincer and Zarnowitz
(1969), Nelson (1972), Howrey et al. (1974), Berger and Krane (1985), Meese and
Rogoff (1983, 1988), Akgiray (1989), Diebold and Nason (1990), Fair and Shiller
(1990), Pagan and Schwert (1990), West and Cho (1995) and some of the participants
in the Makridakis et al. (1982) competition.

If the predictions do not depend on estimated parameters, it follows from Diebold
and Mariano (1995) that under mild conditions standard regression statistics may be
used. For zero serial correlation in one step ahead prediction errors, for example,
one can simply regress the period $t + 1$ prediction error on the period t prediction
error, and use a standard t-test to test the null that the coefficient is zero.

But if the predictions do depend on estimated parameters, the results of Diebold
and Mariano (1995) need not apply. The usual tests do account for uncertainty that

would be present if (counterfactually) the underlying parameter vector were known rather than estimated, but ignore uncertainty resulting from error in estimation of that parameter vector. Using a conventional set of assumptions, we establish conditions under which this second type of uncertainty is asymptotically negligible, thereby validating the Diebold and Mariano (1995) procedure. More importantly, we show that such uncertainty sometimes is asymptotically nonnegligible, and then suggest computationally convenient ways to obtain test statistics that account for both types of uncertainty. Simulations indicate that failure to account for the second type of uncertainty sometimes results in poorly sized hypothesis tests, while our own adjusted tests usually but not always yield more accurately sized tests.

A vast literature has considered predictive accuracy. A distinguishing element of our work is explicit consideration of the role of estimation of parameters needed for prediction. We focus on test statistics produced by regression packages. These have appeared in a number of applied papers (e.g., Fair and Shiller 1990, and Pagan and Schwert 1990), and, we hope, may appear in still more papers upon development of techniques such as those proposed here.[2] We build on earlier work (especially West 1996) not only by developing computationally convenient procedures, but also by allowing additional sampling schemes (additional ways of dividing available data into estimation and prediction components), relaxing certain technical conditions that implicitly ruled out certain important tests (including zero correlation between a prediction error and a prediction), and supplying new simulation evidence.

Section 2 of the paper describes the environment. Sections 3 and 4 present technical assumptions and basic asymptotic results. Section 5 presents our computationally convenient adjustments to standard regression statistics. Sections 6 and 7 specialize on the discussions in Sections 3–5 to consider some common tests, when the underlying models are linear and exactly identified. Section 8 presents simulation evidence and concludes. The Appendix presents proofs. An additional Appendix available on request from the authors presents details of proofs and simulation results omitted from the paper to save space.

2. DESCRIPTION OF ENVIRONMENT

Let $\tau \geq 1$ be the prediction horizon of interest. There are P Predictions in all, which rely on estimates of a $(k \times 1)$ unknown parameter vector β^*. To avoid certain singularities we assume $k > 0$ and merely note that our results specialize in the obvious way when regression estimates are not required to make predictions.

The first prediction uses data from period R or earlier to predict a period $R + \tau$ event, the second from period $R + 1$ or earlier to predict a period $R + 1 + \tau$ event,..., the last from period $R + P - 1 \equiv T$ or earlier to predict a period $T + \tau$

---

[2] We hope our work will be useful even for the interpretation of completed papers. With the exception of one paper that came to our attention after this paper was written (Hoffman and Pagan 1989), to our knowledge all such papers have used standard regression statistics, without adjusting for dependence of predictions on estimated parameters. We establish conditions for the asymptotic validity of such statistics, and in some cases we are able to propose adjustments for such dependence that can be made even without access to the data. See Sections 4, 5 and 7.

event. The total sample size is $R + P - 1 + \tau = T + \tau$:
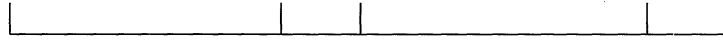
Observation: 1 R $R + \tau$ $R + P - 1 = T$ $T + \tau$

In estimating β^*, three different schemes to use available data are prominent in the forecasting literature. We consider the three explicitly because results vary for the three. The first scheme, which we call *recursive*, was used by, for example, Fair and Shiller (1990). This scheme uses all available data, estimating β^* first with data from 1 to R, next with data from 1 to $R + 1, \ldots$, and finally with data from 1 to T. The second scheme, which we call *rolling*, was used by, for example, Akgiray (1989). This scheme fixes the sample size, say at R, and drops distant observations as recent ones are added. Thus, β^* is estimated first with data from 1 to R, next with data from 2 to $R + 1, \ldots$, and finally with data from P to T. The third and final scheme, which we call *fixed*, was used by, for example, Pagan and Schwert (1990). This scheme estimates β^* just once, say on data from 1 to R, and uses the estimate in forming all P predictions; data realized subsequent to R are, however, used in forming predictions, as described in the previous paragraph and below.

For $t = R, \ldots, T$, let $\hat{\beta}_t$ be the regression vector used for prediction when data from period t and earlier are used. In the least squares model $y_t = X_t' \beta^* + u_t$, for example, $\hat{\beta}_t$ is estimated using

(2.1)   data from 1 to t in the recursive scheme,

$$\hat{\beta}_t = \left(\sum_{s=1}^{t} X_s X_s' \right)^{-1} \sum_{s=1}^{t} X_s y_s,$$

data from $t - R + 1$ to t in the rolling scheme,

$$\hat{\beta}_t = \left(\sum_{s=t-R+1}^{t} X_s X_s' \right)^{-1} \sum_{s=t-R+1}^{t} X_s y_s,$$

data from 1 to R in the fixed scheme,

$$\hat{\beta}_t = \left(\sum_{s=1}^{R} X_s X_s' \right)^{-1} \sum_{s=1}^{R} X_s y_s.$$

Note that for the fixed scheme, $\hat{\beta}_t$ is the same for all t, and depends only on R and not t, while in the recursive and rolling schemes a different regression estimate is used for each t. In addition, for the rolling and fixed schemes, $\hat{\beta}_t$ should properly be subscripted $\hat{\beta}_{t, R}$; the dependence on R is suppressed for notational simplicity. The asymptotic approximation assumes that both P and R are large (formally, $P, R \to \infty$), with τ fixed.

One is interested in the relationship between a scalar prediction error and a vector of variables—say, whether the prediction error is correlated with the vector of variables. As illustrated in example 2 below, we can limit the formal discussion to prediction errors and still yield results applicable to inference about predictions as

well; given the linearity of the procedures we analyze, results for predictions (= observed data point − prediction error) follow immediately. We limit the formal analysis to a scalar dependent variable to economize on notation; we comment occasionally on vector generalizations of our results.

Let $v_{t+\tau}(\beta^*) \equiv v_{t+\tau}$ be the scalar prediction error of interest, with $v_{t+\tau}(\hat{\beta}_t) \equiv \hat{v}_{t,t+\tau}$ ($\hat{v}_{t+1} \equiv \hat{v}_{t,t+1}$) the corresponding random variable evaluated at $\hat{\beta}_t$. As the dating suggests, $v_{t+\tau}$ typically relies on data realized in period $t + \tau$. One is interested in the linear relationship between $v_{t+\tau}$ and a vector function of period t data. Let $g_{t+1}(\beta^*) \equiv g_{t+1}$ denote this ($l \times 1$) vector function, with $g_{t+1}(\hat{\beta}_t) \equiv \hat{g}_{t+1}$ the sample counterpart evaluated at $\hat{\beta}_t$. In most applications l is small, say $l = 1$ or $l = 2$. Here, $g_{t+1}(\beta^*)$ depends on data observed in period t and earlier; the dating convention is used because g_{t+1} often depends on the predetermined variables available at time $t + 1$. See the examples below.

The aim is to use a least squares regression to test the null hypothesis that $E v_{t+\tau} g_{t+1} = 0$. The obvious regression is one of $\hat{v}_{t,t+\tau}$ on $\hat{g}_{t+1}$ for $t = R, \ldots, R + P - 1$, obtaining

$$(2.2) \qquad \hat{v}_{t,t+\tau} = \hat{g}'_{t+1}\hat{\alpha} + \hat{\eta}_{t+\tau}, \qquad \hat{\alpha} \equiv \left(\sum_{t=R}^{T} \hat{g}_{t+1}\hat{g}'_{t+1} \right)^{-1} \sum_{t=R}^{T} \hat{g}_{t+1}\hat{v}_{t,t+\tau},$$

$$\hat{\eta}_{t+\tau} \equiv \hat{v}_{t,t+\tau} - \hat{g}'_{t+1}\hat{\alpha}.$$

One then uses the estimate of $\hat{\alpha}$ and a suitable variance-covariance matrix to test the null.

To illustrate, here are four examples, illustrated with the simple zero mean AR(1) model $y_t = \beta^* y_{t-1} + u_t$, $|\beta^*| < 1$.

1. *Mean Prediction Error.* Here, $g_{t+1} \equiv 1$ is a scalar. If $v_{t+\tau}$ is a τ step ahead forecast error in the AR(1) model, then $\hat{v}_{t,t+\tau} = y_{t+\tau} - \hat{\beta}_t^\tau y_t$.

2. *Efficiency.* Here, one regresses $y_{t+\tau}$ on the period t prediction (= $\hat{\beta}_t^\tau y_t$, in the AR(1) model) and perhaps a constant and other possible predictors as well. The null is that the coefficient on the prediction is unity, and on any other included variables is zero. To analyze this regression using our framework, which presumes that the dependent variable is a prediction error, note that if one uses the prediction error (= $y_{t+\tau} - \hat{\beta}_t^\tau y_t$ in the AR(1) model) as the dependent variable the regression results are algebraically identical to those with $y_{t+\tau}$ on the left-hand side, except that the estimated coefficient on the prediction will be smaller by unity. Hence, for say $\tau = 1$, if $\hat{g}_{t+1}$ is (2×1) and includes a constant term as well as $\hat{\beta}_t y_t$, H_0 is $\alpha = (0,0)'$. Note the dating convention: $\hat{g}$ is dated $t + 1$, but depends on y_t, the regressor available for prediction at time $t + 1$.

3. *Encompassing.* Here, v_{t+1} is a one step ahead forecast error from a putatively encompassing model. The right-hand side variable $\hat{g}_{t+1}$ is the scalar prediction from a putatively encompassed model, and the null is $\alpha = 0$. More generally, the

right-hand side might include a constant, in which case $\hat{g}_{t+1}$ and α are (2×1) and the null is $\alpha = (0,0)'$.

4. *Serial Correlation.* If v_{t+1} is the one step ahead forecast error in a model presumed to have serially uncorrelated errors $(= u_{t+1}$ in the AR(1) model), then $g_{t+1} = v_t$ is the previous period's forecast error. So α is a scalar, $\hat{\alpha}$ an estimate of the first-order serial correlation coefficient, and H_0 is $\alpha = 0$.[3]

One of our major aims is to develop computationally convenient procedures, which in our regression context means using standard errors produced by standard computer programs, or perhaps simple adjustments to those standard errors. As we shall see, conventional test statistics are not always asymptotically valid, even when $\tau = 1$ and $v_{t+1} \equiv v_{t+1}(\beta^*)$ is a zero mean i.i.d. variable that is independent of $g_{t+1} \equiv g_{t+1}(\beta^*)$. The reason is that in some applications, two sources of uncertainty affect asymptotic interence about α. The first is uncertainty that would be present even if (counterfactually) β^* were known and one could regress $v_{t+\tau}$ on g_{t+1}. The second results from use of $\hat{\beta}_t$ rather than the unknown β^*. According to our asymptotic approximation, standard regression statistics properly account for the first source of uncertainty but not necessarily the second. We show below that in some important examples, properly accounting for both sorts of uncertainty requires merely rescaling the least squares variance-covariance matrix by a certain function of P/R.

When such a simple adjustment does not suffice, one can sometimes obtain asymptotically valid test statistics by *augmenting* the regression (2.2) with a judiciously chosen set of variables $\hat{g}_{2t+1}$. In this case, one runs the regression

$$(2.3) \qquad \hat{v}_{t,t+\tau} = \hat{g}'_{t+1}\alpha + \hat{g}'_{2t+1}\alpha_2 + \text{disturbance} \equiv \hat{\tilde{g}}'_{t+1}\tilde{\alpha} + \text{disturbance},$$

where $\hat{g}_{2t+1}$ is a $(r \times 1)$ set of extra variables included so that conventionally computed hypothesis tests on α are correctly sized according to our asymptotic theory; $\hat{\tilde{g}} \equiv (\hat{g}'_{t+1}, \hat{g}'_{2t+1})'$ is $(l+r) \times 1$; $\tilde{g}_{t+1} \equiv \tilde{g}_{t+1}(\beta^*) \equiv (g_{t+1}(\beta^*)', g_{2t+1}(\beta^*)')'$ $\equiv (g'_{t+1}, g'_{2t+1})'$ and $\tilde{\alpha}$ are also $(l+r) \times 1$.

3. ASSUMPTIONS

This section presents assumptions relevant for the basic regression (2.2); Section 5 will present an extension for analysis of the augmented regression (2.3). Our assumptions are 'high level' ones. We use relatively abstract assumptions for two reasons. First, they allow us or others to verify that our results apply to tests and models other than the ones we consider in detail in Sections 6 and 7 below. Second,

---

[3] This test is most naturally run by regressing $y_{t+1} - \hat{\beta}_t y_t$ on $y_t - \hat{\beta}_{t-1} y_{t-1}$. Strictly speaking, our notation implies that $y_{t+1} - \hat{\beta}_{t-1} y_t$ rather than $y_{t+1} - \hat{\beta}_t y_t$ is on the left: we assume that both left- and right-hand side variables are constructed from the same estimate of β^*, and a rank condition presented below rules out simply defining parameters so that the population parameter of interest is 2×1 without a 2×1 period t estimate of $(\hat{\beta}_t, \hat{\beta}_{t-1})'$. But this rank condition is easily relaxed, and results may be generalized to allow the natural version of this test. To economize on notation, we do not explicitly do so in this paper.

they can be presented compactly. In the interest of concision and clarity, we also do not attempt to state each theorem using a minimal set of assumptions. For example, a weaker version of Assumption 3 applies in applications with parametric covariance matrix estimators.

Some notation: for any differentiable function $n_t : \mathbf{R}^m \to \mathbf{R}^s$ and for x in the domain of n_t, $\partial n_t / \partial x$ denotes the $(s \times m)$ matrix of partial derivatives of n_t; for any function n_t whose domain is in $\mathbf{R}^k$, $n_{t\beta} \equiv \partial n_t(\beta^*)/\partial \beta$; for any matrix $A = [a_{ij}]$, let $|A| \equiv \max_{i,j} |a_{ij}|$; summations of variables indexed by t or $t + \tau$ run from $t = R$ to $t = T \equiv R + P - 1$: for any variable x, $\sum x(t) \equiv \sum_{t=R}^{T} x(t)$, $\sum x_{t+\tau} \equiv \sum_{t=R}^{T} x_{t+\tau}$; summations of variables indexed by s run from (a) 1 to t, for the recursive scheme, (b) $t - R + 1$ to t, for the rolling scheme, and (c) 1 to R, for the fixed scheme: for any variable x, (a) $\sum x_s \equiv \sum_{s=1}^{t} x_s$ (recursive), (b) $\sum x_s \equiv \sum_{s=t-R+1}^{t} x_s$ (rolling), or (c) $\sum_s x_s \equiv \sum_{s=1}^{R} x_s$ (fixed). Finally, let

$$(3.1) \quad f_{t+\tau}(\beta^*) \equiv g_{t+1}(\beta^*) v_{t+\tau}(\beta^*), \qquad f_{t+\tau,\beta} \equiv \frac{\partial f_{t+\tau}}{\partial \beta}(\beta^*), \qquad F \equiv E f_{t+\tau,\beta}.$$

Here, $f_{t+\tau} : \mathbf{R}^k \to \mathbf{R}^l$; the $(l \times k)$ matrix F is not subscripted by t in accordance with a stationarity assumption about to be made.

ASSUMPTION 1. (a) In some neighborhood N around β^*, and with probability 1, $v_t(\beta)$ and $g_t(\beta)$ are measurable and twice continuously differentiable; (b) $E v_{t+\tau} g_{t+1} = 0$; (c) $E v_t v_{t\beta} = 0$; (d) $E v_{t+\tau} g_{t+1,\beta} = 0$; (e) $E g_t g_t'$ is of rank l.

ASSUMPTION 2. The estimate $\hat{\beta}_t$ satisfies $\hat{\beta}_t - \beta^* = B(t)H(t)$, where $B(t)$ is $(k \times q)$ and $H(t)$ is $(q \times 1)$, with (a) $B(t) \to_{\text{a.s.}} B$, B a matrix of rank k; (b) $H(t) = t^{-1}\sum_s h_s(\beta^*)$ (recursive) or $H(t) = R^{-1}\sum_s h_s(\beta^*)$ (rolling or fixed) for a $(q \times 1)$ orthogonality condition $h_s(\beta^*)$; (c) $E h_s(\beta^*) = 0$; (d) in the neighborhood N of assumption 1, h_t is measurable and continuously differentiable.

ASSUMPTION 3. In the neighborhood N of Assumption 1, there is a constant $D < \infty$ such that for all t, $\sup_{\beta \in N} |\partial^2 v_t(\beta)/\partial \beta \partial \beta'| < m_t$ for a measurable m_t for which $E m_t^4 < D$. The same holds when v_t is replaced by an arbitrary element of g_t.

ASSUMPTION 4. Let $w_t \equiv (v_{t\beta}', \text{vec}(g_{t\beta})', v_t, g_t', h_t')'$. (a) For some $d > 1$, $\sup_t E\|w_t\|^{8d} < \infty$, where $\|\cdot\|$ denotes Euclidean norm. (b) w_t is strong mixing, with mixing coefficients of size $-3d/(d - 1)$. (c) w_t is fourth-order stationary. (d) Let $\Gamma_{ff}(j) = E f_t f_{t-j}'$, $S_{ff} = \sum_{j=-\infty}^{\infty} \Gamma_{ff}(j)$. Then S_{ff} is positive definite.

ASSUMPTION 5. $R, P \to \infty$ as $T \to \infty$, and $\lim_{T \to \infty} P/R = \pi$, (a) $0 \le \pi \le \infty$ for recursive ($\pi = \infty \Leftrightarrow \lim_{T \to \infty} R/P = 0$), (b) $0 \le \pi < \infty$ for rolling and fixed.

Note that from Assumptions 1(b) and 1(d),

$$(3.2) \qquad\qquad E f_t = 0, \qquad F = E g_{t+1}\left(\frac{\partial v_{t+\tau}}{\partial \beta}\right).$$

In allowing not only for recursive but also rolling and fixed sampling schemes, Assumptions 2–5 generalize similar assumptions in West (1996), where some discussion of the assumptions may be found. To illustrate briefly here: The moment conditions in Assumptions 3 and 4 rule out unit autoregressive roots, but otherwise do not seem restrictive. Assumption 2 allows standard estimation techniques, including GMM and maximum likelihood. In the AR(1) model of Section 2, for example, $B = (Ey_{t-1}^2)^{-1}$, $h_s = y_{s-1} u_s$. Assumption 5 says that both P and R are large; in particular, they are large relative to the forecast horizon τ.

Throughout, we maintain Assumptions 1–5.

4. BASIC ASYMPTOTIC RESULTS

Let

$$(4.1) \qquad \Gamma_{fh}(j) = Ef_t h'_{t-j}, \qquad S_{fh} = \sum_{j=-\infty}^{\infty} \Gamma_{fh}(j), \qquad \Gamma_{hh}(j) = Eh_t h'_{t-j},$$

$$S_{hh} = \sum_{j=-\infty}^{\infty} \Gamma_{hh}(j), \qquad V_\beta = BS_{hh}B'.$$

V_β is the asymptotic variance-covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta^*)$.

Define $\lambda_{fh}, \lambda_{hh}$ and $\lambda \equiv 1 - 2\lambda_{fh} + \lambda_{hh}$, all of which are scalar functions of $\pi \equiv \lim_{T\to\infty} P/R$, as follows:

(4.2)

Sampling scheme	λ_{fh}	λ_{hh}	λ
recursive	$1 - \pi^{-1}\ln(1+\pi)$	$2[1 - \pi^{-1}\ln(1+\pi)]$	1
rolling, $\pi \le 1$	$\dfrac{\pi}{2}$	$\pi - \dfrac{\pi^2}{3}$	$1 - \dfrac{\pi^2}{3}$
rolling, $\pi > 1$	$1 - \dfrac{1}{2\pi}$	$1 - \dfrac{1}{3\pi}$	$\dfrac{2}{3\pi}$
fixed	0	π	$1 + \pi$

LEMMA 4.1. (a) $P^{-1/2}\sum \hat{g}_{t+1}\hat{v}_{t,t+\tau} = P^{-1/2}\sum g_{t+1}v_{t+\tau} + FB[P^{-1/2}\sum H(t)] + o_p(1)$. (b) $P^{-1/2}\sum g_{t+1}v_{t+\tau} \sim_A N(0, S_{ff})$. (c) $E[P^{-1}\sum H(t)\sum H(t)'] \to \lambda_{hh}S_{hh}$, $E[P^{-1}\sum g_{t+1}v_{t+\tau}\sum H(t)'] \to \lambda_{fh}S_{fh}$.

The results for the recursive scheme follow from West (1996), and are repeated here for completeness. The results for the rolling and fixed schemes are new.

LEMMA 4.2. $P^{-1/2}\sum \hat{g}_{t+1}\hat{v}_{t,t+\tau} \sim_A N(0,\Omega)$, where Ω is the $(l \times l)$ matrix

$$(4.3) \qquad \Omega = S_{ff} + \lambda_{fh}(FBS'_{fh} + S_{fh}B'F') + \lambda_{hh}FV_\beta F'.$$

LEMMA 4.3. $P^{-1} \Sigma \hat{g}_{t+1} \hat{g}'_{t+1} \to_p E g_t g'_t$.

THEOREM 4.1. *Let* $\hat{\alpha}$ *be the least squares estimator of* α $(= 0)$. *Then* $P^{1/2} \hat{\alpha} \sim_A$ $N(0, V)$, $V \equiv (E g_t g'_t)^{-1} \Omega (E g_t g'_t)^{-1}$.

For inference, an estimate of V is required. To discuss this, we introduce some more notation. Let $\hat{\eta}_{t+\tau} \equiv \hat{v}_{t, t+\tau} - \hat{g}'_{t+1} \hat{\alpha}$ be the least squares regression residual, $\hat{\sigma}$ the usual scalar estimate of the standard error of the regression disturbance, and $\hat{\Gamma}_{ff}(j)$ the $(l \times l)$ j'th sample autocovariance of $\hat{g}_{t+1} \hat{\eta}_{t+\tau}$:

$$(4.4) \qquad \hat{\sigma}^2 \equiv (P - l)^{-1} \sum \hat{\eta}^2_{t+\tau} \equiv (P - l)^{-1} \sum \left(\hat{v}_{t, t+\tau} - \hat{g}'_{t+1} \hat{\alpha} \right)^2,$$

$$\hat{\Gamma}_{ff}(j) \equiv P^{-1} \sum_{t = R + j}^{T} \left[\left(\hat{g}_{t+1} \hat{\eta}_{t+\tau} \right) \left(\hat{g}_{t+1-j} \hat{\eta}_{t+\tau-j} \right)' \right] \quad \text{for } j \geq 0,$$

$$\hat{\Gamma}_{ff}(j) \equiv \hat{\Gamma}_{ff}(-j)' \quad \text{for } j < 0.$$

THEOREM 4.2. (a) $\hat{\sigma}^2 \to_p \sigma^2 \equiv E v_t^2$, $\hat{\sigma}^2 (P^{-1} \Sigma \hat{g}_{t+1} \hat{g}'_{t+1})^{-1} \to_p \sigma^2 (E g_t g'_t)^{-1}$, (b) $\hat{\Gamma}_{ff}(j) \to_p \Gamma_{ff}(j)$, $(P^{-1} \Sigma \hat{g}_{t+1} \hat{g}'_{t+1})^{-1} \hat{\Gamma}_{ff}(0) (P^{-1} \Sigma \hat{g}_{t+1} \hat{g}'_{t+1})^{-1} \to_p$ $(E g_t g'_t)^{-1} \Gamma_{ff}(0) (E g_t g'_t)^{-1}$. (c) *Let* $K(x)$ *be a kernel such that for all* x, $|K(x)| \leq 1$, $K(x) = K(-x)$, $K(0) = 1$, $K(x)$ *is continuous for all* x, *and* $\int_{-\infty}^{\infty} |K(x)| \, dx < \infty$. *For some bandwidth* M *and some constant* a, $0 < a < 1/2$, *suppose* $(M/P^a) \to 0$ *and, if* $\pi = \infty$, $(M/R^a) \to 0$. *Then* $\hat{S}_{ff} \equiv \Sigma_{j = -P+1}^{P-1} K(j/M) \hat{\Gamma}_{ff}(j) \to_p S_{ff}$, *and* $(P^{-1} \Sigma \hat{g}_{t+1} \hat{g}'_{t+1})^{-1} \hat{S}_{ff} (P^{-1} \Sigma \hat{g}_{t+1} \hat{g}'_{t+1})^{-1} \to_p (E g_t g'_t)^{-1} S_{ff} (E g_t g'_t)^{-1}$.

Note that Theorem 4.2 assumes that the least squares residual $\hat{\eta}_{t+\tau}$ is used in estimating $\hat{\sigma}^2$ and $\hat{\Gamma}_{ff}(j)$. Since $\alpha = 0$ the asymptotic results are unchanged if one replaces $\hat{\eta}_{t+\tau}$ with the left-hand side variable $\hat{v}_{t, t+\tau}$; our formal analysis and our simulation results below both use $\hat{\eta}_{t+\tau}$ because that is what will be used by standard computer programs.

Part (a) of Theorem 4.2 considers the textbook estimator of the least squares covariance matrix, and part (b) a heteroskedasticity consistent estimator that is sometimes referred to as the White (1980) covariance matrix estimator. In part (c), a nonparametric estimator is described, under conditions similar to those in Andrews (1991) or Newey and West (1994). So one can use kernels such as the Bartlett, in which $\hat{S}_{ff} = \hat{\Gamma}_{ff}(0) + \Sigma_{j=1}^{M} [1 - (j/M)][\hat{\Gamma}_{ff}(j) + \hat{\Gamma}_{ff}(j)']$ with $M \to \infty$ at a suitable rate, or the Quadratic Spectral. From part (b), if $\Gamma_{ff}(j) = 0$ for $j \geq \tau$, as will typically be the case, another estimator that is consistent for S_{ff} is the truncated estimator; here, $\hat{S}_{ff} = \hat{\Gamma}_{ff}(0) + \Sigma_{j=1}^{\tau-1} [\hat{\Gamma}_{ff}(j) + \hat{\Gamma}_{ff}(j)']$.

Theorem 4.2 says that some sample moments are consistent for the analogous population moments. But inspection of Theorem 4.1 indicates that use of these estimators may not produce a consistent estimate of V. To illustrate, consider a simple setup in which $\tau = 1$ and v_{t+1} is i.i.d. and independent of current and past g_{t+1}. Then $E(v_{t+1} | g_{t+1}, v_t, g_t, v_{t-1}, \ldots) = 0$, and $E(v_{t+1}^2 g_{t+1} g'_{t+1}) = E v_{t+1}^2 E g_{t+1} g'_{t+1}$ $= S_{ff}$. The least squares estimator of the regression covariance matrix is

$\hat{\sigma}^2(P^{-1}\sum \hat{g}_{t+1}\hat{g}'_{t+1})^{-1}$. From Theorem 4.2, this estimator converges in probability to $\sigma^2(Eg_tg'_t)^{-1} \equiv Ev_t^2(Eg_tg'_t)^{-1}$. From Lemma 4.1(b) and the proof of Lemma 4.3, this is the covariance matrix that is applicable in the counterfactual case in which β^* is known, and one regresses $v_{t+\tau}(\beta^*)$ on $g_{t+1}(\beta^*)$. But since β^* is not known, we see from Theorem 4.2 that the asymptotic variance of $P^{1/2}\hat{\alpha}$ is not $Ev_t^2(Eg_tg'_t)^{-1}$ but $(Eg_tg'_t)^{-1}\Omega(Eg_tg'_t)^{-1} = Ev_t^2(Eg_tg'_t)^{-1} + \{(Eg_tg'_t)^{-1}[\lambda_{fh}(FBS'_{fh} + S_{fh}B'F') + \lambda_{hh}FV_\beta F'](Eg_tg'_t)^{-1}\}$. The additional terms in braces are ones that result from uncertainty about β^*. In this example and more generally, use of the usual regression formulas may result in asymptotically invalid tests.

If these formulas are instead to result in asymptotically valid tests, we must have $S_{ff} = \Omega$. This condition implies that the asymptotic distribution of $\hat{\alpha}$ does not depend on uncertainty about β^*: the distribution of $P^{1/2}\hat{\alpha}$ is identical to that of the estimator obtained by regressing $v_{t+\tau}(\beta^*)$ on $g_{t+1}(\beta^*)$ in the hypothetical case in which β^* is known. Two simple conditions are sufficient to imply $S_{ff} = \Omega$. One is $F \equiv E\partial f_t(\beta^*)/\partial\beta \equiv E[g_{t+1}(\beta^*)\partial v_{t+\tau}(\beta^*)/\partial\beta] = 0$. This is essentially a condition that there is block diagonality in the asymptotic variance-covariance matrix for the estimators of β^* and $Ef_{t+\tau} \equiv Eg_{t+1}v_{t+\tau}$. This condition occasionally applies in practice, for example in testing for first-order serial correlation with strictly exogenous predictors. But since such examples are uncommon, we do not further discuss this condition.[4]

A second condition sufficient for $\Omega = S_{ff}$ is $\pi \equiv \lim_{T\to\infty} P/R = 0$, because this implies $\lambda_{fh} = \lambda_{hh} = 0$. When $\pi = 0$, the limiting ratio of the size of the prediction sample to that of the regression sample is zero. As noted informally by Chong and Hendry (1986) in the context of encompassing tests, one can then act as if β^* is known. The practical implication is that if P/R is small, it may be safe to use the usual regression statistics. How small P/R must be depends on the data and the tests; in our simple Monte Carlo experiment, the lowest value of P/R was $1/7$, and that was not sufficiently small to always make it harmless to ignore error in estimation of β^*.

The next section discusses ways to obtain asymptotically valid test statistics, even when $S_{ff} \neq \Omega$.

5. OBTAINING ASYMPTOTICALLY VALID TEST STATISTICS

Throughout this section, we assume that we have an estimator of S_{ff} that satisfies $\hat{S}_{ff} \to_p S_{ff}$. Theorem 4.2 describes how to obtain such an estimator. In addition, for $\lambda = \lambda(\pi)$ defined in (4.2), define

$$\hat{\lambda} \equiv \lambda(\hat{\pi}), \qquad \hat{\pi} \equiv P/R.$$

For the recursive scheme, $\hat{\lambda} = 1$ for all π, for the fixed scheme $\hat{\lambda} = 1 + (P/R)$, and so on. Clearly, $\hat{\lambda} \to \lambda$.

---

[4] See West (1996) and McCracken (1998) for further discussion of the conditions under which $F = 0$.

COROLLARY 5.1. *Suppose that*

(5.1) $$S_{ff} = -\frac{1}{2}\left(FBS'_{fh} + S_{fh}B'F'\right) = FV_{\beta}F'.$$

Then $\hat{\lambda}(P^{-1}\Sigma\,\hat{g}_{t+1}\hat{g}'_{t+1})^{-1}\hat{S}_{ff}(P^{-1}\Sigma\,\hat{g}_{t+1}\hat{g}'_{t+1})^{-1} \to_p V \equiv \lambda(Eg_t g'_t)^{-1}S_{ff}(Eg_t g'_t)^{-1}$, *where* $P^{1/2}\hat{\alpha} \sim_A N(0, V)$.

Condition (5.1) implies that Ω (defined in (4.3)) is equal to λS_{ff}, and Corollary 5.1 then follows directly from Theorem 4.1. Condition (5.1) might seem unlikely. But in fact, as detailed below, in certain linear models it holds for tests for: (1) mean prediction error and for efficiency, under general conditions, and (2) tests for encompassing and zero first-order serial correlation when the sampling scheme is recursive and the forecast error is conditionally homoskedastic.

Upon comparing Corollary 5.1 and Lemma 4.1(b), we see that when the conditions of Corollary 5.1 hold, uncertainty about β^* simply introduces a factor of λ into the asymptotic variance of $P^{1/2}\hat{\alpha}$. For the recursive sampling scheme, $\lambda \equiv 1$, so error in estimation of β^* is asymptotically irrelevant: the variance of such estimation error ($= \lambda_{hh}FV_{\beta}F'$) is exactly offset by $-\lambda_{fh}(FBS'_{fh} + S_{fh}B'F')$, which is the covariance between (1) such error, and (2) error that would be present even if (counterfactually) β^* were known. For the fixed scheme, $\lambda > 1$, so failure to adjust will result asymptotically in t- and chi-squared statistics that are too small and thus in too many rejections at any specified significance level. For the rolling scheme, $\lambda < 1$, so failure to adjust will result asymptotically in too few rejections at any specified significance level. Further, in any finite sample, the adjustment by $\hat{\lambda}$ by construction increases t- and chi-squared statistics for the fixed scheme, and decreases them for the rolling scheme.

When condition (5.1) does not hold, uncertainty about β^* usually results in greater complications. To handle these, we propose the augmented regression (2.3), which we repeat here for convenience:

(2.3) $$\hat{v}_{t,t+\tau} = \hat{g}'_{t+1}\alpha + \hat{g}'_{2t+1}\alpha_2 + \text{disturbance} \equiv \hat{\tilde{g}}'_{t+1}\alpha_0 + \text{disturbance},$$

THEOREM 5.1. *Let* $\tilde{g}_{t+1}(\beta^*) = (g'_{t+1}, g'_{2t+1})'$ *for a* $(r \times 1)$ *vector* g_{2t+1} *defined as either* (a) $g_{2t+1} = \partial v_{t+\tau}(\beta^*)/\partial\beta$ ($\Rightarrow r = k$) *or* (b) $g_{2t+1} = Z_{t+1}$ *for a vector of variables* Z_{t+1} *that satisfies* $\partial v_{t+\tau}(\beta^*)/\partial\beta = G_2(\beta^*)Z_{t+1}$, $G_2(\beta^*)$ *a* $(k \times r)$ *nonstochastic matrix. Define* $\tilde{f}_{t+\tau} \equiv \tilde{g}_{t+1}v_{t+\tau}$. *Suppose that for one of the definitions of* g_{2t+1}, *Assumptions 1, 2, and 4 are satisfied when* $\tilde{f}_{t+\tau}$ *and* $\tilde{g}_{t+1}$ *replace* $f_{t+\tau}$ *and* g_{t+1}. *Continue to maintain Assumptions 3 and 5 as well. Let* $S_{\tilde{f}\tilde{f}}$ *and* $S_{\tilde{f}h}$ *be defined as in equation (4.1),* $\tilde{F}$ *as in equation (3.2), with* $\tilde{f}_{t+\tau}$ *replacing* $f_{t+\tau}$. *Let* $\tilde{\Omega} = S_{\tilde{f}\tilde{f}} + \lambda_{fh}(\tilde{F}BS'_{\tilde{f}h} + S_{\tilde{f}h}B'\tilde{F}') + \lambda_{hh}\tilde{F}V_{\beta}\tilde{F}'$. *Let* $\hat{\tilde{\alpha}} = (\Sigma\,\hat{\tilde{g}}_{t+1}\hat{\tilde{g}}'_{t+1})^{-1}(\Sigma\,\hat{\tilde{g}}_{t+1}\hat{v}_{t,t+\tau})$ *be the result of a regression of* $\hat{v}_{t,t+\tau}$ *on* $\hat{\tilde{g}}_{t+1}$, *with* $\hat{\alpha}$ *the first* l *elements of* $\hat{\tilde{\alpha}}$. *Then* $P^{1/2}\hat{\alpha} \sim_A N(0, V)$, V *the* $(l \times l)$ *matrix in the upper-left-hand corner of* $(E\tilde{g}_t\tilde{g}'_t)^{-1}S_{\tilde{f}\tilde{f}}(E\tilde{g}_t\tilde{g}'_t)^{-1}$.

For in-sample tests, similar augmentation is proposed by Pagan and Hall (1983), Davidson and MacKinnon (1984, 1989), and Wooldridge (1990, 1991).

Theorem 5.1 states that conventional regression output can be used. From Theorem 4.2, conventional regression programs consistently estimate $S_{\tilde{f}\tilde{f}}$. So, for example, if $\tau = 1$ and v_{t+1} is a textbook error—conditionally homoskedastic and serially uncorrelated—for inference one can use the $l \times l$ matrix in the upper-left-hand corner of $\hat{\sigma}^2 (P^{-1} \Sigma \, \hat{\tilde{g}}_{t+1} \hat{\tilde{g}}'_{t+1})^{-1}$, $\hat{\sigma}$ the usual least squares estimate of the standard error of the regression disturbance that is defined in Theorem 4.2(a). More generally, if $\tau > 1$ or there is conditional heteroskedasticity, heteroskedasticity and autocorrelation consistent covariance matrix estimators may be used.

It should be noted that one of the assumptions of the theorem, that $E\tilde{g}_t \tilde{g}'_t$ is of full rank (this is Assumption 1(e)) is not always innocuous. With tests of mean prediction error or of efficiency in linear models, for example, the rank condition will fail for either definition of $\tilde{g}_t$. For these tests, the computationally convenient test that we propose is the one described in Corollary 5.1.

On the other hand, the condition typically is satisfied in tests for zero serial correlation of one step ahead prediction errors and for encompassing tests. For univariate ARMA models, one will augment with $\partial v_{t+\tau} / \partial \beta$ evaluated at $\hat{\beta}_t$, for linear simultaneous equations models with the vector of predetermined variables.

To prevent confusion, we emphasize that Theorem 5.1 does *not* say that one can use the usual regression output for inference about α_2, the coefficients on g_{2t+1}. It is true that $\hat{\alpha}_2$ converges in probability to zero. But in general the usual regression output will not consistently estimate the asymptotic variance-covariance matrix, as discussed in Section 4.

6. FOUR COMMON TESTS

In this section and the next, we consider the four common tests listed in Section 2: mean prediction error, efficiency, first-order serial correlation, and encompassing. For conciseness and clarity, we limit our formal statements to one step ahead prediction errors ($\tau = 1$) in a model estimated by least squares. We comment in Section 7 on generalizations to predictions from the reduced form of linear simultaneous equations models or from univariate ARMA models, and to multiperiod predictions. This section lays out the setup. The next section presents results.

The model is

$$(6.1) \qquad y_t = x'_t \beta^* + v_t,$$

where y_t and v_t are scalars, x_t and β^* are $(k \times 1)$. The sample counterpart of v_{t+1} is computed as

$$(6.2) \qquad \hat{v}_{t+1} = y_{t+1} - x'_{t+1} \hat{\beta}_t.$$

For the encompassing test, we need to describe as well the encompassed model. This will require redefining β^*. Model '1' is the encompassing model, and '2' the encompassed model. Let $\beta^* = (\beta_1^{*\prime}, \beta_2^{*\prime})'$, where β_i^* is $(k_i \times 1)$, $k \equiv k_1 + k_2$, with the model i prediction dependent only on β_i^*. Let x_{2t} be the vector of predeter-

mined variables in model 2, $y_t = x'_{2t} \beta_2^* + v_{2t}$. The null is that v_{t+1} is uncorrelated with $x'_{2t+1} \beta_2^*$, the forecast from model 2.

Along with Assumption 5 (i.e., $P \to \infty$, $R \to \infty$), we assume

ASSUMPTION (*). (a) x_t includes a constant. (b) $E(v_t | x_t, x_{t-1}, \ldots, v_{t-1}, v_{t-2}, \ldots)$ $= 0$ (for the encompassing test, $E(v_t | x_t, x_{2t}, x_{t-1}, x_{2t-1}, \ldots, v_{t-1}, v_{2t-1}, v_{t-2}, v_{2t-2} \ldots) = 0$). (c) For g_t and $\tilde{g}_t$ defined in Table 1, $Eg_t^2 > 0$ and $E\tilde{g}_t \tilde{g}'_t$ is of full rank. (d) Let $h_s(\beta^*) = x_s v_s$ (for the encompassing text, $h_s = (x'_s v_s, x'_{2s} v_{2s})'$). The estimate $\hat{\beta}_t$ satisfies $\hat{\beta}_t - \beta^* = B(t) H(t)$, where $B(t)$ is $(k \times k)$ and $H(t)$ is $(k \times 1)$, with $B(t)$ and $H(t)$ defined as follows. (i) $B(t) = (t^{-1} \sum_s x_s x'_s)^{-1}$ (recursive), $B(t) = (R^{-1} \sum_s x_s x'_s)^{-1}$ (rolling or fixed). For the encompassing test, $B(t)$ is block diagonal with analogously defined $B_i(t)$ on the diagonals. (ii) $H(t) = t^{-1} \sum_s h_s(\beta^*)$ (recursive) or $H(t) = R^{-1} \sum_s h_s(\beta^*)$ (rolling or fixed). (iii) $Ev_t^2 > 0$, and $Ex_t x'_t$ and $Ex_t x'_t v_t^2$ are positive definite (for the encompassing test, the same holds for model 2). (e) (i) Let $w_t = (x'_t, v_t)'$. For some $d > 1$, $\sup_t E\|w_t\|^{8d} < \infty$. (ii) w_t is strong mixing, with mixing coefficients of size $-3d/(d-1)$. (iii) w_t is fourth-order stationary. For the encompassing test, the same holds for $w_t = (x'_t, v_t, x'_{2t}, v_{2t})'$.

The 'low level' assumption (*) may be shown to imply the 'high level' Assumptions 1–4, as well as the validity of the null hypotheses of zero mean prediction error, zero serial correlation, and so forth. As well, part (c) of Assumption (*) follows from the other parts for mean prediction error and serial correlation; as long as $\beta^* \neq 0$ part (c) follows as well for efficiency. For encompassing tests, part (c) follows from the mild additional condition that the prediction from the encompassed model does not lie in the linear span of the regressors from the encompassing model.[5]

7. OBTAINING REGRESSION-BASED TEST STATISTICS FOR THE FOUR COMMON TESTS

Column (2) of Table 1 lists the scalar right-hand side variable in the simplest version of these tests.

THEOREM 7.1. (a) For g_t defined as in one of the rows of Table 1, let $\hat{\alpha} = (\sum \hat{g}_{t+1}^2)^{-1} (\sum \hat{g}_{t+1} \hat{v}_{t+1})$. (i) For mean prediction error or efficiency, $P^{1/2} \hat{\alpha} \sim_A N(0, V)$, $V = \lambda (Eg_t^2)^{-2} Ev_t^2 g_t^2$. (ii) Let the sampling scheme be recursive, and suppose that the underlying disturbance v_t is conditionally homoskedastic, $E(v_t^2 | x_t) = Ev_t^2$ (for encompassing, assume $E(v_t^2 | x_t, x_{2t}) = Ev_t^2$ and $E(v_{2t}^2 | x_t, x_{2t}) = Ev_{2t}^2$). Then for any one of the four tests in the table, $P^{1/2} \hat{\alpha} \sim_A N(0, V)$, $V = \sigma^2 (Eg_t^2)^{-1}$, $\sigma^2 \equiv Ev_t^2$. (b) For encompassing or first-order serial correlation, augment the regression as indicated in Table 1, and regress $\hat{v}_{t+1}$ on $\hat{g}_{t+1}$ and $\hat{g}_{2t+1}$. Let $\hat{\alpha}$ be the first element of the resulting coefficient vector. Then $P^{1/2} \hat{\alpha} \sim_A N(0, V)$, V the $(1,1)$ element in $(E\tilde{g}_t \tilde{g}'_t)^{-1} Ev_t^2 \tilde{g}_t \tilde{g}'_t (E\tilde{g}_t \tilde{g}'_t)^{-1}$.

---

[5] Note that this last condition rules out tests of nested (rather than nonnested) models. Such tests are in Ashley et al., (1980) and Clark (1997). An insightful referee has pointed out that some of our results do extend to nonnested models; to conserve space, we do not consider such models here.

TABLE 1
REGRESSORS FOR FOUR COMMON TESTS, LINEAR MODEL*

(1)	(2)	(3)	(4)
Test	$\hat{g}_{t+1}$	$\hat{g}_{2t+1}$	$\hat{\tilde{g}}_{t+1}$
(1) Mean Prediction Error	1	n.a.	n.a.
(2) Efficiency	$x'_{t+1}\hat{\beta}_t$	n.a.	n.a.
(3) Encompassing	$x'_{2t+1}\hat{\beta}_{2t}$	x_{t+1}	$(x'_{2t+1}\hat{\beta}_{2t}, x'_{t+1})$
(4) First Order Serial Correlation	$\hat{v}_t$	x_{t+1}	$(\hat{v}_t, x'_{t+1})'$

* The model is $y_{t+1} = x'_{t+1}\beta^* + v_{t+1}$, where y_{t+1} and v_{t+1} are scalars, x_{t+1} is a vector, and β^* is the unknown parameter vector. In the AR(1) example of Section 2, this specializes to $y_{t+1} = y_t \beta^* + v_{t+1}$. The left-hand side variable is a one step ahead prediction error, $\hat{v}_{t+1} \equiv y_{t+1} - x'_{t+1}\hat{\beta}_t$. The simpler regression analyzed in Sections 6 and 7 is one in which $\hat{g}_{t+1} \equiv g_{t+1}(\hat{\beta}_t)$ (column (2)) is the sole regressor; the more complicated regression is one in which $\hat{\tilde{g}}$ (column (4)) is the vector of regressors. See Sections 6 and 7 of the paper for more detail.

Table 2 summarizes when and how to adjust.

Comments. 1. In part (a)(i), asymptotically valid test statistics require scaling the usual covariance matrix by $\hat{\lambda}$ (which means no adjustment for the recursive scheme, for which $\hat{\lambda} \equiv 1$). In parts (a)(ii) and (b), no special adjustment is needed.

2. For the recursive scheme, the difference between the assumptions in (a)(i) and (a)(ii) is that (a)(i) allows conditional heteroskedasticity of the prediction error, (a)(ii) does not. The covariance matrix in part (i) reduces to that in part (ii) if there is no conditional heteroskedasticity. If there is conditional heteroskedasticity, tests for encompassing and first-order serial correlation will be missized if the inference is based on the covariance matrix given in part (a)(i).

3. While not stated formally, the results in part (a) continue to apply when a constant is included in the regression. Valid *t*- and chi-squared tests require merely rescaling the usual covariance matrix.

4. For mean prediction error, the formula for V in part (a) (i) simplifies to $\lambda E v_t^2$. For encompassing and serial correlation, under conditional homoskedasticity the formula for V in part (b) reduces to $E v_t^2 (E \tilde{g}_t \tilde{g}'_t)^{-1}$.

5. Zero mean prediction error seems to be the only one of these tests that is often done for multistep horizons (e.g., Meese and Rogoff 1983). For a reduced form which is a first-order VAR, we have established that the results in part (a) still apply, with $\lambda \sum_{j=-\tau+1}^{\tau-1} E v_t v_{t-j}$ replacing $\lambda E v_t^2$ as the asymptotic variance covariance matrix.

6. A vector of sample mean prediction errors is also asymptotically normal with the variance-covariance matrix being the usual one, multiplied by λ.

7. Suppose that β^* is estimated from the structural equations of a linear simultaneous equations model, with the reduced form used for predictions and prediction errors. Under some additional conditions, the results in Theorem 7.1 still obtain.

8. Suppose predictions are made from a univariate ARMA model that is estimated by nonlinear least squares or an asymptotically equivalent technique. Then condition (5.1), which underlies Theorem 7.1(a), continues to hold for mean predic-

TABLE 2
ADJUSTMENTS FOR FOUR COMMON TESTS, LINEAR MODEL*

Sampling Scheme	Correction Needed?	How to Correct the t-statistic
A. Zero Mean Prediction Error		
1. Recursive	no	n.a.
2. Rolling	yes	divide t-statistic by $\hat{\lambda}^{1/2}$
3. Fixed	yes	divide t-statistic by $\hat{\lambda}^{1/2}$
B. Efficiency		
1. Recursive	no	n.a.
2. Rolling	yes	divide t-statistic by $\hat{\lambda}^{1/2}$
3. Fixed	yes	divide t-statistic by $\hat{\lambda}^{1/2}$
C. Encompassing		
1. Recursive	no: v_{t+1} conditionally homoskedastic	n.a.
	yes: v_{t+1} conditionally heteroskedastic	augmented regression
2. Rolling	yes	augmented regression
3. Fixed	yes	augmented regression
D. Zero First-Order Serial Correlation		
1. Recursive	no: v_{t+1} conditionally homoskedastic	n.a.
	yes: v_{t+1} conditionally heteroskedastic	augmented regression
2. Rolling	yes	augmented regression
3. Fixed	yes	augmented regression

* (i) The model is $y_{t+1} = x'_{t+1}\beta^* + v_{t+1}$, with v_{t+1} serially uncorrelated. The prediction horizon is one period ($\tau = 1$). The regression run is one with $\hat{v}_{t+1}$ on the left-hand side, as described in Table 1. This table describes how and when to adjust the usual least squares standard errors to account for uncertainty about β^*.

(ii) The table assumes $\pi > 0$. π is the limiting value of P/R, where P is the number of predictions, R the size of the smallest regression sample. When $\pi = 0$, no adjustment is needed, for any of the tests in the table.

(iii) In panels C and D, 'v_{t+1} conditionally homoskedastic' means $Ev_{t+1}^2 x_{t+1} x'_{t+1} = Ev_{t+1}^2 Ex_{t+1} x'_{t+1}$; '$v_{t+1}$ conditionally heteroskedastic' allows the possibility that $Ev_{t+1}^2 x_{t+1} x'_{t+1} \neq Ev_{t+1}^2 Ex_{t+1} x'_{t+1}$.

(iv) Panel D allows $Ev_t x_{t+1} \neq 0$, as is typically the case in time series applications. If $Ev_t x_{t+1} = 0$, no correction is needed, for any of the schemes, and whether or not v_{t+1} is conditionally heteroskedastic.

tion error. So under suitable conditions the result in Theorem 7.1(a) will continue to hold as well.[6]

8. MONTE CARLO EVIDENCE

Here we present a simple Monte Carlo experiment. Our aim is to get a feel for whether our proposed adjustments to the usual least squares statistics are likely to be useful in practice, and, more generally, whether our asymptotic approximation

[6] It is, however, possible to construct examples in which the results of Theorem 7.1 fail. Let ϵ_t and u_t be independent standard normals, $v_t = \epsilon_t^2 u_t$, $x_t = (Ex_t) + \epsilon_t$ with $Ex_t \neq 0$, where all variables are scalars. Let a regression model be $y_t = x_t\beta^* + v_t$, with estimation by OLS. Then $S_{ff} = Ev_t^2 (= E\epsilon_t^4 Eu_t^2)$, $S_{fh} = Ex_t v_t^2 = Ex_t Ev_t^2$, $S_{hh} = Ex_t^2 v_t^2$, $F = Ex_t$. This violates Theorem 7.1's assumption that there is a constant term in the equation. Consider mean prediction error. Theorem 4.1 indicates that $P^{1/2}\hat{\alpha} \equiv P^{-1/2}\Sigma(y_{t+1} - x_{t+1}\hat{\beta}_t)$ is asymptotically normal with asymptotic variance $[Ev_t^2 - 2\lambda_{fh}Ex_t(Ex_t^2)^{-1}Ex_t Ev_t^2 + \lambda_{hh}(Ex_t)^2(Ex_t^2 v_t^2)(Ex_t^2)^{-2}]$. This does not reduce to $\lambda Ev_t^2 \equiv \lambda S_{ff}$ since $Ex_t \neq 0$ and $Ex_t^2 v_t^2 \neq Ex_t^2 Ev_t^2$.

might yield well-sized test statistics. It turns out that while our approximation does usually work well, the rolling sampling scheme does sometimes require unusually large samples sizes to generate accurate test statistics.

The experiment we present involved 5000 repetitions. Each repetition required generating 201 data points (200 excluding an initial condition). (Some additional experiments reported briefly in Table 6 and in detail in the additional Appendix involved 1000 repetitions of samples of size 1601.) Each of these 5000 artificial samples of size 200 were split into 15 different regression (R) and prediction (P) samples. The values of P and R were: $R = 25, P = 25, 50, 100, 150, 175$; $R = 50$, $P = 25, 50, 100, 150$; $R = 100, P = 25, 50, 100$; $R = 150, P = 25, 50$; $R = 175, P = 25$—15 combinations in all. This range for P/R (from $1/7$ to 7), as well as the values of $T = P + R - 1$, seem broad enough to include most relevant empirical work. For a given (P, R) pair, the $(P \times 1)$ vector of prediction errors used on the left-hand side of the regression tests was $\{\hat{v}_{t+1}\}$, $t = R, \ldots, R + P - 1$.

For each pair of R and P, the first $R + P$ observations of each sample of size 200 were used. So $R = 50/P = 100$ and $R = 100/P = 50$, for example, used the same 150 observations, but began the out-of-sample exercise at different points. This means, for example, that for the recursive scheme the 50 prediction errors used in the $R = 100/P = 50$ sample were identical to the last 50 in the $R = 50/P = 100$ sample.

A recent literature has emphasized the inaccuracy of conventional asymptotic approximations in some time series environments. Examples from our own work include Newey and West (1994) and West and Wilcox (1996). We suspect that our out-of-sample procedures will also work poorly in such environments. To give as clear as possible a sense for whether our procedures might work well, we consider a data generating process and regression that to our knowledge has in-sample behavior that is reasonably well approximated by conventional asymptotic theory. This process is a zero-mean AR(1) with i.i.d. normal disturbances and an autoregressive parameter that is not close to the unit circle,

$$(8.1) \qquad y_t = \beta^* y_{t-1} + v_t, \qquad \beta^* = 0.5, \qquad v_t \sim N(0, 1).$$

In each of the 5000 samples, y_0 was drawn from its unconditional $N(0, (1 - \beta^{*2})^{-1})$ distribution, and $y_1, \ldots, y_{200}$ were generated recursively using (8.1) and pseudo-random draws of v_t.

In each sample, and for each P and R, four hypothesis tests were conducted for one step ahead $(\tau = 1)$ predictions: mean prediction error, efficiency, zero serial correlation, and encompassing. For the last test the alternative model was $y_t = \beta_2 y_{t-2} + v_{2t}$. This was estimated by least squares, so $\beta_2 \equiv (E y_{t-2}^2)^{-1} E y_{t-2} y_t$. The introduction of the second lag meant that some regression samples were one observation smaller than the 'R' reported in the table.

We report tests of nominal size 0.05. Tests of nominal size 0.01 and 0.10 worked equally well, and tests with larger sample sizes worked better; see the additional Appendix. All regression tests included a constant term, since these typically would be included in practice. Apart from adjustment by a factor of $\hat{\lambda}$ in regressions in which our theory calls for such an adjustment, the usual least squares covariance matrix was used—that is, we did not use a heteroskedasticity consistent covariance matrix estimator.

TABLE 3
SIZE OF NOMINAL 0.05 TESTS, MEAN PREDICTION ERROR*

A. Accounting for Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Recursive	25	0.054	0.052	0.053	0.056	0.056
	50	0.053	0.057	0.051	0.057	
	100	0.046	0.049	0.054		
	150	0.056	0.056			
	175	0.052				
2. Rolling	25	0.063	0.074	0.105	0.133	0.145
	50	0.053	0.063	0.063	0.072	
	100	0.048	0.051	0.058		
	150	0.054	0.055			
	175	0.053				
3. Fixed	25	0.091	0.090	0.096	0.097	0.099
	50	0.069	0.074	0.075	0.077	
	100	0.058	0.060	0.064		
	150	0.062	0.050			
	175	0.058				

B. Ignoring Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Rolling	25	0.025	0.003	0.000	0.000	0.000
	50	0.043	0.021	0.002	0.000	
	100	0.046	0.044	0.021		
	150	0.054	0.052			
	175	0.052				
2. Fixed	25	0.220	0.297	0.421	0.498	0.523
	50	0.129	0.195	0.293	0.354	
	100	0.081	0.121	0.186		
	150	0.078	0.106			
	175	0.073				

* The DGP is a univariate AR(1); see text for details. For the indicated values of P and R, $\hat{v}_{t+1}$ (the one step ahead prediction error) was regressed on a constant for $t = R, \ldots, R + P - 1$. Panels B1 and B2 report the fraction of the 5000 simulations in which the conventionally computed t-statistic on the coefficient on the constant term was greater than 1.96 in absolute value. Panels A1–A3 report the same, when the conventionally computed t-statistic is divided by the square root of $\hat{\lambda}$.

Table 3A presents results for mean prediction error. Tests for the recursive scheme work quite well, with nominal 0.05 tests having actual sizes between 0.046 and 0.057. Our approximation does not work as well for the rolling and fixed schemes, although performance is perhaps tolerable for $P/R \leq 1$, and is quite good for $P/R \leq 0.5$.

Table 3B presents results when the least squares t-statistic is used, without dividing as we suggest by $\sqrt{\hat{\lambda}}$. Recall that by construction: (1) the rolling scheme must have lower actual size and the fixed scheme higher actual size when our adjustment for error in estimation of β^* is ignored; (2) the adjustment is smaller the smaller is P/R. Panels A3 and B2 indicate that for the fixed scheme, our adjustment improves the size for all P/R. The difference is perhaps not large for small P/R (e.g., for $P = 25$, $R = 100$, our test statistic yields a size of 0.058, the unadjusted a size of

0.081), but it is dramatic for large P/R (for $P = 175$, $R = 25$, our test statistic has a size of 0.099 versus 0.523 for the unadjusted test statistic).

For the rolling scheme, the comparison is not as clear cut, since our test statistic typically rejects too infrequently (actual size > 0.05), while the unadjusted typically rejects too often (actual size < 0.05). While we do not have a precise loss function for under- versus over-rejection, our own gut feeling is that we would rather have a nominal 0.05 test have a probability of rejecting of say 7.4 per cent ($P = 50$, $R = 25$, our test statistic) than of 0.3 per cent (unadjusted test statistic), all other things being equal. In this sense, our test statistics perform better for the rolling scheme as well. But we recognize that other researchers may have different loss functions, at least in some applications.

Table 4 has the results for the efficiency test. For the recursive and fixed schemes, our procedure seems to be a little more accurately sized than it was for mean

TABLE 4
SIZE OF NOMINAL 0.05 TESTS, EFFICIENCY TEST*

A. Accounting for Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Recursive	25	0.052	0.051	0.055	0.055	0.053
	50	0.038	0.043	0.043	0.046	
	100	0.038	0.040	0.045		
	150	0.041	0.047			
	175	0.042				
2. Rolling	25	0.124	0.430	0.939	0.997	0.999
	50	0.045	0.070	0.232	0.468	
	100	0.036	0.043	0.059		
	150	0.042	0.045			
	175	0.041				
3. Fixed	25	0.058	0.055	0.056	0.055	0.053
	50	0.040	0.038	0.035	0.031	
	100	0.039	0.041	0.042		
	150	0.041	0.036			
	175	0.042				

B. Ignoring Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Rolling	25	0.059	0.072	0.152	0.331	0.450
	50	0.037	0.030	0.016	0.013	
	100	0.034	0.034	0.021		
	150	0.042	0.041			
	175	0.040				
2. Fixed	25	0.158	0.234	0.355	0.434	0.456
	50	0.087	0.135	0.220	0.286	
	100	0.063	0.097	0.152		
	150	0.060	0.083			
	175	0.057				

*The DGP is a univariate AR(1); see text for details. For the indicated values of P and R, $\hat{v}_{t+1}$ (the one step ahead prediction) was regressed on a constant and $y_t \hat{\beta}_t$ (the one step ahead prediction) for $t = R, \ldots, R + P - 1$. Panels B1 and B2 report the fraction of the 5000 simulations in which the conventionally computed t-statistic on the coefficient on $y_t \hat{\beta}_t$ was greater than 1.96 in absolute value. Panels A1–A3 report the same, when the conventionally computed t-statistic is divided by the square root of $\hat{\lambda}$.

TABLE 5
SIZE OF NOMINAL 0.05 TESTS, ENCOMPASSING TEST*

A. Accounting for Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Recursive	25	0.113	0.148	0.185	0.201	0.196
	50	0.077	0.100	0.125	0.135	
	100	0.054	0.070	0.084		
	150	0.045	0.055			
	175	0.046				
2. Rolling	25	0.244	0.388	0.634	0.796	0.861
	50	0.151	0.232	0.346	0.450	
	100	0.090	0.122	0.171		
	150	0.075	0.088			
	175	0.071				
3. Fixed	25	0.044	0.050	0.049	0.048	0.050
	50	0.049	0.049	0.053	0.049	
	100	0.049	0.048	0.048		
	150	0.049	0.051			
	175	0.054				

B. Ignoring Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Rolling	25	0.130	0.250	0.508	0.713	0.790
	50	0.084	0.122	0.211	0.312	
	100	0.059	0.078	0.105		
	150	0.044	0.057			
	175	0.047				
2. Fixed	25	0.191	0.270	0.399	0.486	0.508
	50	0.103	0.160	0.240	0.300	
	100	0.063	0.093	0.133		
	150	0.055	0.068			
	175	0.052				

* The DGP is a univariate AR(1); see text for details. Let $\hat{\beta}_{2t}$ denote the least squares estimate of a regression of y_t on y_{t-2} using the same sample as that used to obtain $\hat{\beta}_t$. For the indicated values of P and R, $\hat{v}_{t+1}$ (the one step ahead prediction error) was regressed on a constant and $y_{t-1}\hat{\beta}_{2t}$ for $t = R, \ldots, R + P - 1$. Panels A1, B1, and B2 report the fraction of the 5000 simulations in which the conventionally computed t-statistic on the coefficient on $y_{t-1}\hat{\beta}_{2t}$ was greater than 1.96 in absolute value. Panels A2 and A3 report the same, when y_t was included as a third regressor.

prediction error. But for these two schemes the remarks made in connection with Table 3 generally apply here as well.

The rolling scheme, however, performs quite poorly for $P/R > 1$. In fact, for $P/R > 1$, the over-rejection is so extreme that failure to adjust generally improves the test statistic. For example, for $P = 50$, $R = 25$, panel A2 indicates that our procedure had an actual size of 43 per cent, while panel B1 indicates that use of the usual least squares test statistic yielded a size of 7.2 per cent.

Tables 5 and 6 indicate that for the encompassing test and the test for zero first-order serial correlation, the Table 4 results apply qualitatively: For the recursive and the fixed schemes, our test statistics work adequately, and dominate the unadjusted test statistic. But for the rolling scheme our test statistic works poorly.

In Tables 4–6, the rolling scheme worked quite poorly for $P/R > 1$. To see how large a sample is required for tolerable accuracy of the asymptotic approximation,

TABLE 6
SIZE OF NOMINAL 0.05 TESTS, TEST FOR ZERO FIRST-ORDER SERIAL CORRELATION*

A. Accounting for Error in Estimation of β^*

Sampling Scheme	R	P				
		25	50	100	150	175
1. Recursive	25	0.059	0.060	0.061	0.061	0.061
	50	0.043	0.052	0.057	0.053	
	100	0.040	0.048	0.051		
	150	0.045	0.054			
	175	0.045				
2. Rolling	25	0.119	0.209	0.405	0.584	0.663
	50	0.069	0.094	0.143	0.183	
	100	0.058	0.066	0.071		
	150	0.054	0.058			
	175	0.057				
3. Fixed	25	0.044	0.050	0.049	0.048	0.050
	50	0.049	0.049	0.053	0.049	
	100	0.049	0.048	0.048		
	150	0.049	0.051			
	175	0.053				
B. Ignoring Error in Estimation of β^*						
1. Rolling	25	0.034	0.022	0.027	0.052	0.067
	50	0.040	0.029	0.014	0.013	
	100	0.039	0.044	0.026		
	150	0.045	0.048			
	175	0.045				
2. Fixed	25	0.214	0.319	0.447	0.512	0.546
	50	0.111	0.177	0.269	0.335	
	100	0.066	0.105	0.156		
	150	0.062	0.086			
	175	0.058				

*The DGP is a univariate AR(1); see text for details. In panel B, $\hat{v}_{t+1}$ was regressed on a constant and $\hat{v}_t$ for $t = R, \ldots, R + P - 1$, for the indicated values of P and R. Panels A1, B1, and B2 report the fraction of the 5000 simulations in which the conventionally computed t-statistic on the coefficient on $\hat{v}_t$ was greater than 1.96 in absolute value. Panels A2 and A3 report the same, when y_t was included as a third regressor.

we generated 1000 samples of size 1601; we report here certain results with samples of size up to 1201 (full details are in the additional Appendix). We controlled the seed to the random number generator so that the first 201 observations in each sample were the same as in Tables 3–6. We then conducted the efficiency test for some larger sample sizes, holding P/R fixed at 2 and at 4. The results are in Table 7. As may be seen, by the time the sample size hits 1200, the result for $P/R = 2$ is

TABLE 7
SIZE OF NOMINAL 0.05 TESTS, EFFICIENCY TEST, LARGER SAMPLE SIZES*

A. $(P/R) = 2$				
$P = 50, R = 25$	$P = 100, R = 50$	$P = 200, R = 100$	$P = 400, R = 200$	$P = 800, R = 400$
0.430	0.232	0.108	0.091	0.069
B. $(P/R) = 4$				
	$P = 100, R = 25$	$P = 400, R = 100$	$P = 800, R = 200$	
	0.939	0.409	0.229	

*See notes to Table 4. The tests account for error in estimation of β^*. The figures for $P + R < 200$ are repeated from Table 4.

reasonably accurate (actual size of 0.069), at least by the standards of Tables 3–6 and much other work on hypothesis testing in time series models. For $P/R = 4$, however, substantial missizing still remains.

We conclude that our asymptotic approximation usually works reasonably well, but that for the rolling sampling scheme relatively large sample size sometimes are required.

<div align="center">APPENDIX</div>

Notation: '$\sup_t$' means '$\sup_{R \le t \le T}$'; 'var', and 'cov' denote variance and covariance; all limits are taken as the sample size T goes to infinity; the summation 'Σ' means '$\Sigma_{t=R}^T$'; For notational simplicity, we consider throughout the case in which $k = 1$ and $l = 1$, so that β^*, g_{t+1} and $f_{t+\tau}$ are scalars, and we let '$f_{t+\tau, \beta\beta}(\tilde{\beta}_t)$' mean '$\partial^2 f_{t+\tau}(\tilde{\beta}_t)/\partial \beta^2$'. To save space, proofs of Lemmas A1 to A4 and parts of other proofs are put in an additional Appendix available on request from the authors.

LEMMA A1. *Suppose* $\pi < \infty$. *For* $0 \le a < .5$: (a) $\sup_t |P^a H(t)| \to_p 0$; (b) $\sup_t |P^a(\hat{\beta}_t - \beta^*)| \to_p 0$.

LEMMA A2. (a) $P^{-1}\Sigma |f_{t+\tau}|^2 = O_p(1)$, (b) $P^{-1} \Sigma |f_{t+\tau, \beta}|^2 = O_p(1)$, (c) *For* $\tilde{\beta}_t$ *satisfying* $|\tilde{\beta}_t - \beta^*| \le |\hat{\beta}_t - \beta^*|$ *for* $t = R, \ldots, T$, $P^{-1}\Sigma |f_{t+\tau, \beta\beta}(\tilde{\beta}_t)|^2 = O_p(1)$.

LEMMA A3. *Let* $\overset{\circ}{\Gamma}_{ff}(j) \equiv P^{-1} \Sigma_{t=R+j}^T [(\hat{g}_{t+1}\hat{v}_{t,t+\tau})(\hat{g}_{t+1-j}\hat{v}_{t-j,t+\tau-j})] \equiv P^{-1} \Sigma_{t=R+j}^T f_{t+\tau}(\hat{\beta}_t)f_{t+\tau-j}(\hat{\beta}_{t-j})$. *Then* $\overset{\circ}{\Gamma}_{ff}(j) \to_p \Gamma_{ff}(j)$.

LEMMA A4. *Under the assumptions of Theorem* 4.2, *and with* $\overset{\circ}{\Gamma}_{ff}(j)$ *defined as in Lemma A*3, $\overset{\circ}{S}_{ff} \equiv \Sigma_{j=-P+1}^{P-1} K(j/M)\overset{\circ}{\Gamma}_{ff}(j) \to_p S_{ff}$.

PROOF OF LEMMA 4.1. (a) For the recursive scheme, this follows from Lemma 4.1 of West (1996). The relatively simple argument for the fixed scheme is in the additional Appendix. For the rolling scheme, expand $f_{t+\tau}(\hat{\beta}_t)$ around $f_{t+\tau}(\beta^*)$ for $t = R, \ldots, R + P - 1$, and sum the results, yielding

$$(A1) \quad P^{-1/2} \sum f_{t+\tau}(\hat{\beta}_t)$$

$$= P^{-1/2} \sum f_{t+\tau} + P^{-1/2} \sum f_{t+\tau, \beta} B(t)H(t) + P^{-1/2} \sum w_{t+\tau},$$

for $w_{t+\tau} \equiv .5 f_{t+\tau, \beta\beta}(\tilde{\beta}_t)(\hat{\beta}_t - \beta^*)^2$. We have $|P^{-1/2}\sum w_{t+\tau}| \le .5(P^{1/4} \sup_t |\hat{\beta}_t - \beta^*|)^2 (P^{-1}\Sigma |f_{t+\tau, \beta\beta}(\tilde{\beta}_t)|) \to_p 0$ by Lemmas A1 and A2. The second term in (A1) can be written

$$P^{-1/2}FB \sum H(t) + P^{-1/2} \sum \{F[B(t) - B]H(t)\} + P^{-1/2} \sum \left[(f_{t+\tau, \beta} - F)BH(t) \right]$$

$$+ P^{-1/2} \sum \{(f_{t+\tau, \beta} - F)[B(t) - B]H(t)\}$$

and hence we need show that the last three terms in the above expression are $o_p(1)$.

We will show the result for $P^{-1/2}\Sigma[(f_{t+\tau,\beta}-F)BH(t)]$; the others follow from arguments similar to those for the recursive scheme (West 1996).

For notational simplicity, let $x_t \equiv (f_{t+\tau,\beta}-F)$, redefine Bh_t as h_t, and let $\gamma_j \equiv Ex_t h_{t-j}$. For $P \le R$ (the $P > R$ case is similar) we have $|EP^{-1/2}\Sigma x_t H(t)| = (P^{1/2}/R)|\gamma_0 + \cdots + \gamma_{R-1}| \le (P/R)^{1/2}R^{-1/2}\Sigma_{j=0}^{\infty}|\gamma_j| \to 0$ since $\pi < \infty$ and $\Sigma_{j=0}^{\infty}|\gamma_j| < \infty$. Then since it can be shown that Assumption 4 bounds the fourth moments of $(x_t, h_t)'$ in such a way that $\lim \text{var}[P^{-1/2}\Sigma x_t H(t)] = 0$, the result follows from Chebyshev's inequality. (b) Follows from Theorem 3.1 of Wooldridge and White (1989). (c) For the recursive scheme the results are in West (1996). For the fixed scheme, $EP^{-1}\Sigma H(R)\Sigma H(R) = (P/R)E[(R^{-1/2}\Sigma_{s=1}^{R} h_s)(R^{-1/2}\Sigma_{s=1}^{R} h_s)'] \to \pi S_{hh}$. To show that $\lambda_{fh} = 0$, let $\gamma_j = Ef_{t+\tau}h'_{t-j}$. Then $|ER^{-1}\Sigma f_{t+\tau}(\Sigma_{s=1}^{R} h'_s)| = |R^{-1}[(\gamma_{R-1} + \cdots + \gamma_0) + \cdots + (\gamma_{R+P-2} + \cdots + \gamma_{P-1})]| \le R^{-1}\Sigma_{j=-\infty}^{\infty}|j||\gamma_j| \to 0$ since Assumption 4 implies $\Sigma_{j=-\infty}^{\infty}|j||\gamma_j| < \infty$ (Andrews 1991).

For the rolling scheme, we will sketch the result that $E[P^{-1}\Sigma H(t)\Sigma H(t)'] \to \lambda_{hh}S_{hh} \equiv [\pi - (\pi^2/3)]S_{hh}$ for $\pi < 1$. The proofs for $\pi \ge 1$, and for $E[P^{-1}\Sigma g_{t+1}v_{t+\tau}\Sigma H(t)'] \to \lambda_{fh}S_{fh}$, are conceptually similar.

With $P < R$, $\Sigma H(t)$ may be written as the sum of three terms, $\Sigma H(t) = A_1 + A_2 + A_3$, $A_1 \equiv R^{-1}[h_1 + 2h_2 + \cdots + (P-1)h_{P-1}]$, $A_2 \equiv PR^{-1}[h_P + \cdots + h_R]$, $A_3 \equiv R^{-1}[(P-1)h_{R+1} + \cdots + 2h_{R+P-2} + h_{R+P-1}]$. It is easy to see that $\lim \text{var}(P^{-1/2}A_2) = \lim P(R-P+1)R^{-2}\Sigma_{|j| \le R-P+1}Eh_t h'_{t-j} + o(1) \to (\pi - \pi^2)S_{hh}$. We will sketch the argument that shows $\lim \text{var}(P^{-1/2}A_1) = (\pi^2/3)S_{hh}$. That $\lim \text{var}(P^{-1/2}A_3) = (\pi^2/3)S_{hh}$ follows from a nearly identical argument. Since, finally, it can be shown that $\lim \text{cov}(P^{-1/2}A_i, P^{-1/2}A_j) = 0$ for $i \ne j$, the result will follow.

For simplicity, assume $q = 1$. Redefine γ_j as $\gamma_j \equiv Eh_t h_{t-j}$, and for $|j| \le P-2$ define $d_j = \Sigma_{i=1}^{P-1-|j|}[i(i+|j|)]$. Then

$$\text{var}(A_1) = R^{-2}\sum_{j=-P+2}^{P-2} d_j\gamma_j = R^{-2}d_0\sum\gamma_j - R^{-2}\sum(d_0 - d_j)\gamma_j.$$

We have $P^{-1}R^{-2}d_0 \sim [P^3/(3PR^2)] \to \pi^2/3$, and the result will follow if $P^{-1}R^{-2}\Sigma(d_0 - d_j)\gamma_j \to 0$. This result may be established using $d_0 \le \int_0^P x^2 dx$, $d_j \ge \int_{j+1}^{P-1}(x-j)x dx \Rightarrow |d_0 - d_j| \le |\int_0^P x^2 dx - \int_{j+1}^{P-1}(x-j)x dx|$, solving the integrals and manipulating the result to obtain $P^{-1}R^{-2}|\Sigma(d_0 - d_j)\gamma_j| \le (1/3P)\Sigma|j||\gamma_j| + o(1) \to 0$.

PROOF OF LEMMA 4.2. Let $X(T) \equiv \Sigma[g_{t+1}v_{t+\tau} + FBH(t)]$. From Lemma 4.1, $P^{-1/2}\Sigma \hat{g}_{t+1}\hat{v}_{t,t+\tau} = P^{-1/2}X(T) + o_p(1)$, with $\lim \text{var}[P^{-1/2}X(T)] = \Omega$. Asymptotic normality then follows from Theorem 3.1 of Wooldridge and White (1989). Details are in the additional Appendix.

PROOF OF LEMMA 4.3. Follows from a mean value expansion of $g_{t+1}(\hat{\beta}_t)$ around $g_{t+1}(\beta^*)$. Details are in the additional Appendix.

PROOF OF THEOREM 4.1. Follows immediately from Lemmas 4.1, 4.2 and 4.3.

PROOF OF THEOREM 4.2. (a) That $P^{-1}\Sigma \hat{g}_{t+1}^2 \to_p Eg_{t+1}^2$ follows from Lemma 4.3. Hence we need only show that $(P-1)^{-1}\Sigma(\hat{v}_{t,t+\tau} - \hat{g}_{t+1}\hat{\alpha})^2 \to_p Ev_{t+\tau}^2$. We have

$(P-1)^{-1}\sum(\hat{v}_{t,t+\tau}-\hat{g}_{t+1}\hat{\alpha})^2 = (P-1)^{-1}\sum\hat{v}_{t,t+\tau}^2 + [(P-1)^{-1}\sum\hat{g}_{t+1}^2]\hat{\alpha}^2 - 2\hat{\alpha}[(P-1)^{-1}\sum\hat{g}_{t+1}\hat{v}_{t,t+\tau}]$. That $(P-1)^{-1}\sum\hat{v}_{t,t+\tau}^2 \to_p Ev_{t+\tau}^2$ follows from Lemma A3. By Theorem 4.1, $\hat{\alpha} = o_p(1)$; by Lemmas 4.2 and 4.3, $(P-1)^{-1}\sum\hat{g}_{t+1}^2 = O_p(1)$, $(P-1)^{-1}\sum\hat{g}_{t+1}\hat{v}_{t,t+\tau} = O_p(1)$. The desired result now follows. (b) That $P^{-1}\sum\hat{g}_{t+1}^2 \to_p Eg_{t+1}^2$ follows from Lemma 4.3. Hence we need only show that $\hat{\Gamma}_{ff}(j) \equiv P^{-1}\sum_{t=R+j}^{T}\hat{g}_{t+1}\hat{g}_{t+1-j}\hat{\eta}_{t+\tau}\hat{\eta}_{t+\tau-j} \to_p Eg_{t+1}g_{t+1-j}v_{t+\tau}v_{t+\tau-j} \equiv \Gamma_{ff}(j)$ for all j. For $\overset{\circ}{\Gamma}_{ff}(j)$ defined in Lemma A3, we have

$$\hat{\Gamma}_{ff}(j) = \overset{\circ}{\Gamma}_{ff}(j) + P^{-1}\sum_{t=R+j}^{T}\hat{g}_{t+1}\hat{g}_{t+1-j}\Big(\hat{\eta}_{t+\tau}\hat{\eta}_{t+\tau-j} - \hat{v}_{t,t+\tau}\hat{v}_{t-j,t+\tau-j}\Big).$$

Lemma A3 shows that the first term converges in probability to $\Gamma_{ff}(j)$; the additional Appendix shows that the second term converges in probability to zero. (c) That $P^{-1}\sum\hat{g}_{t+1}^2 \to_p Eg_{t+1}^2$ follows from Lemma 4.3. Hence we need only show that $\hat{S}_{ff} \equiv \sum_{j=-P+1}^{P-1}K(j/M)\hat{\Gamma}_{ff}(j) \to_p S_{ff}$. For $\overset{\circ}{S}_{ff}$ defined in Lemma A4, we have $\hat{S}_{ff} = \overset{\circ}{S}_{ff} + \sum_{j=-P+1}^{P-1}K(j/M)[\hat{\Gamma}_{ff}(j) - \overset{\circ}{\Gamma}_{ff}(j)]$. Lemma A4 shows that the first term converges in probability to S_{ff}; the additional Appendix shows that the second term converges in probability to zero.

PROOF OF COROLLARY 5.1.   Follows immediately from Theorem 4.1.

PROOF OF THEOREM 5.1.   By definition, the $(l+r)\times k$ matrix $\tilde{F} \equiv E\tilde{g}_{t+1}v_{t+\tau,\beta}$; if $g_{2t+1} = v_{t+\tau,\beta}$, then, $\tilde{F} = (Eg_{2t+1}g_{t+1}', Eg_{2t+1}g_{2t+1}')$, while if $g_{2t+1} = Z_{t+1}$, $\tilde{F} = (Eg_{2t+1}g_{t+1}', Eg_{2t+1}g_{2t+1}')'G_2'$. From Lemmas 4.1 and 4.3 and Theorem 4.1,

$$P^{1/2}\hat{\tilde{\alpha}} = \Big(P^{-1}\sum\hat{\tilde{g}}_{t+1}\hat{\tilde{g}}_{t+1}'\Big)^{-1}\Big(P^{-1/2}\sum\hat{\tilde{g}}_{t+1}\hat{v}_{t,t+\tau}\Big)$$

$$= \Big(E\tilde{g}_{t+1}\tilde{g}_{t+1}'\Big)^{-1}\Big(P^{-1/2}\sum\hat{\tilde{g}}_{t+1}\hat{v}_{t,t+\tau}\Big) + o_p(1)$$

$$= \Big(E\tilde{g}_{t+1}\tilde{g}_{t+1}'\Big)^{-1}\Big(P^{-1/2}\sum\tilde{g}_{t+1}v_{t+\tau}\Big)$$

$$\quad + \Big(E\tilde{g}_{t+1}\tilde{g}_{t+1}'\Big)^{-1}\tilde{F}B\Big(P^{-1/2}\sum H(t)\Big) + o_p(1),$$

Upon partitioning $E\tilde{g}_{t+1}\tilde{g}_{t+1}'$ conformably with g_{t+1} and $g_{2,t+1}$ and using the formula for the inverse of a partitioned matrix, we find that the first l rows of the $(l+r)\times k$ matrix $(E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}\tilde{F}$ are identically zero. Since $\hat{\alpha}$ consists of the first l components of $\hat{\tilde{\alpha}}$, $P^{1/2}\hat{\alpha}$ equals the first l rows of $(E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}(P^{-1/2}\sum\tilde{g}_{t+1}v_{t+\tau}) + o_p(1)$, and the proof is complete.

PROOF OF THEOREM 7.1.   (a) (i) From Corollary 5.1, condition (5.1) is sufficient to guarantee the result. From Assumption (*), we have $B = (Ex_t x_t')^{-1}$ and $S_{hh} = Ev_t^2 x_t x_t'$. For mean prediction error, recall that x_t contains a constant. Without loss of generality define $x_t \equiv (1, \bar{x}_t')'$ where $\bar{x}_t$ is a vector of nonconstant regressors. Then $F = -(1, E\bar{x}_t')$, which implies $FB = -(1\ 0\ \cdots\ 0)$; the result follows since the $(1,1)$ elements of both S_{fh} and S_{hh} are $S_{ff} \equiv Ev_t^2$. For efficiency, note that $F = -\beta^{*\prime}B^{-1}$

and hence $FB = -\beta^{*\prime}$. The result then follows since $S_{ff} = \beta^{*\prime}S_{hh}\beta^*$ and $S_{fh} = \beta^{*\prime}S_{hh}$.

(a) (ii) For mean prediction error or efficiency, the conditional homoskedasticity assumption implies $Ev_t^2 g_t^2 = Ev_t^2 Eg_t^2$ and the result follows from part (i). For the other two tests, recall that for the recursive scheme $\lambda_{hh} = 2\lambda_{fh}$ and thus $\Omega = S_{ff} + \lambda_{fh}(FBS_{fh}' + S_{fh}B'F') + 2\lambda_{fh}FV_\beta F'$. Hence it suffices to show $-FBS_{fh}' = FV_\beta F'$. For serial correlation, this follows since $F = -Ev_{t-1}x_t'$, $B = (Ex_t x_t')^{-1}$, $S_{hh} = Ev_t^2 B^{-1}$, and $S_{fh} = -Ev_t^2 F$. For encompassing this follows since $F = (-\beta_2^{*\prime}Ex_{2t}x_t', 0')$, $B = \text{diag}[(Ex_t x_t')^{-1}, (Ex_{2t}x_{2t}')^{-1}]$, the $(k_1 \times k_1)$ block in the upper left hand corner of S_{hh} is $Ev_t^2 Ex_t x_t'$, and $S_{fh} = \beta_2^{*\prime}(Ev_t^2 Ex_{2t}x_t', Ev_t v_{2t}Ex_{2t}x_{2t}')$. (b) Follows from Theorem 5.1

REFERENCES

AKGIRAY, V., "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts," *Journal of Business* 62 (1989), 55–80.

ANDREWS, D.W.K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59 (1991), 1465–1471.

ASHLEY, R., GRANGER, C.W.J. AND R. SCHMALANSEE, "Advertising and Aggregate Consumption: An Analysis of Causality," *Econometrica* 48 (1980), 1149–1167.

BERGER, A. AND S. KRANE, "The Informational Efficiency of Econometric Model Forecasts," *Review of Economics and Statistics* 67 (1985), 128–134.

CHONG, Y.Y. AND D.F. HENDRY, "Econometric Evaluation of Linear Macro-Economic Models," *Review of Economic Studies* 53 (1986), 671–690.

CLARK, T., "Finite-Sample Properties of Tests for Forecast Equivalence," manuscript, Federal Reserve Bank of Kansas City, 1997.

DAVIDSON, R. AND J.G. MACKINNON, "Model Specification Tests Based on Artificial Linear Regressions," *International Economic Review* 25 (1984), 485–502.

——— AND ———, "Testing for Consistency Using Artificial Regressions," *Econometric Theory* 5 (1989), 363–384.

DIEBOLD, F.X. AND R.S. MARIANO, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–263.

——— AND J. NASON, "Nonparametric Exchange Rate Prediction?," *Journal of International Economics* 28 (1990), 315–322.

FAIR, R.C. AND R. SHILLER, "Comparing Information in Forecasts from Econometric Models," *American Economic Review* 80 (1990), 375–389.

HOFFMAN, D. AND A. PAGAN, "Practitioners Corner: Post-Sample Prediction Tests for Generalized Method of Moments Estimators," *Oxford Bulletin of Economics and Statistics* 51 (1989), 333–343.

HOWREY, P.E., L.R. KLEIN AND M.D. MCCARTHY, "Notes on Testing the Predictive Performance of Econometric Models," *International Economic Review* 15 (1974), 366–383.

MAKRIDAKIS, S. ET AL., "The Accuracy of Time Series Methods: The Results from a Forecasting Competition," *Journal of Forecasting* 1 (1982), 111–153.

MCCRACKEN, M.W., "Out of Sample Inference for Moments of Non-Differentiable Functions," manuscript, University of Wisconsin, 1998.

MEESE, R.A. AND K. ROGOFF, "Empirical Exchange Rate Models of the Seventies: Do they Fit Out of Sample?" *Journal of International Economics* 14 (1983), 3–24.

——— AND ———, "Was It Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating-Rate Period," *Journal of Finance* 43 (1988), 933–948.

MINCER, J. AND V. ZARNOWITZ, "The Evaluation of Economic Forecasts," in J. Mincer, ed., *Economic Forecasts and Expectations* (New York: National Bureau of Economic Research, 1969, pp. 3–46).

NELSON, C.R., "The Predictive Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review* 62 (1972), 902–917.

NEWEY, W.K. AND K.D. WEST, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies* 61 (1994), 631–654.

PAGAN, A.R. AND A.D. HALL, "Diagnostic Tests as Residual Analysis," *Econometric Reviews* 2 (1983), 159–218.

——— AND G.W. SCHWERT, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics* 45 (1990), 267–290.

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

——— AND D. CHO, "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics* 69 (1995), 367–391.

——— AND D.W. WILCOX, "A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model," *Journal of Business and Economic Statistics* 14 (1996), 281–293.

WHITE, H., "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48 (1980), 817–838.

WOOLDRIDGE, J.M., "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* 6 (1990), 17–43.

———, "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics* 47 (1991), 5–46.

——— AND H. WHITE, "Central Limit Theorems for Dependent, Heterogeneous Processes with Trending Moments," manuscript, Michigan State University, 1989.