Asymptotic Inference about Predictive Ability
Author(s): Kenneth D. West
Source: *Econometrica*, Vol. 64, No. 5 (Sep., 1996), pp. 1067-1084
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/2171956

Econometrica, Vol. 64, No. 5 (September, 1996), 1067–1084

ASYMPTOTIC INFERENCE ABOUT PREDICTIVE ABILITY

BY KENNETH D. WEST[1]

This paper develops procedures for inference about the moments of smooth functions of out-of-sample predictions and prediction errors, when there is a long time series of predictions and realizations. The aim is to provide tools for analysis of predictive accuracy and efficiency, and, more generally, of predictive ability. The paper allows for nonnested and nonlinear models, as well as for possible dependence of predictions and prediction errors on estimated regression parameters. Simulations indicate that the procedures can work well in samples of size typically available.

KEYWORDS: Forecasting, forecast evaluation, testing, hypothesis test, model comparison.

1. INTRODUCTION

THIS PAPER DEVELOPS PROCEDURES for asymptotic inference about the moments of smooth functions of predictions and out-of-sample prediction errors. The relevant environments are ones in which a long time series of predictions has been made from a sequence of base periods, and, if predictions are based on regression estimates, each of the sequence of regression estimates underlying the predictions has also been obtained from a long time series.

The aim is to provide tools that will be useful in two broad classes of applications. The first evaluates a model or a set of models on the basis of predictive accuracy or efficiency. Common measures of accuracy or efficiency include mean or mean squared prediction error, correlation between one model's prediction and another model's prediction error, correlation between prediction and realization, and serial correlation of prediction errors (e.g., Nelson (1972), Meese and Rogoff (1983), Fair and Shiller (1990), Stock and Watson (1993), West and Cho (1995)). The second, closely related, class of applications uses a series of predictions to make hypothetical decisions about, say, asset allocation, and then measures the quality of a model by the mean profit or utility yielded by use of its predictions (e.g., McCulloch and Rossi (1990)).

Diebold and Mariano (1994) provide a review and extension of procedures to perform inference about predictions when predictions do not rely on regression estimates. Typically, however, economic predictions rely on such estimates.

[1]I thank two anonymous referees, the Co-Editor, Patrick DeFontnouvelle, Steven Durlauf, Whitney Newey, Mark Watson, Ka-Fu Wong, Jeff Wooldridge, and seminar participants at the Federal Reserve Board of Governors, NBER, Northwestern University, and the Universities of Chicago, Iowa, Rochester, and Wisconsin for helpful comments and discussions. I also thank Chia-Yang Hueng, John Jones, Michael McCracken, and Ka-Fu Wong for excellent research assistance, and the NSF, the Sloan Foundation, and the University of Wisconsin Graduate School and College of Letters and Science for financial support.

Asymptotic tests that examine the correlation between realizations or prediction errors on the one hand and regression-based predictions on the other have been applied by a number of authors, one of the earliest being Nelson (1972). For asymptotic inference about mean squared prediction errors, Meese and Rogoff (1988), Christiano (1989), and Diebold and Rudebusch (1991) suggest applying standard asymptotic results (e.g., Hansen (1982)) in a fashion that, in the end, entails ignoring possible effects of error in estimation of parameters. These papers do not, however, justify their procedures with formal theory or with simulation evidence.

Simulation methods to compare the predictive accuracy of competing models have been proposed by Fair (1980). Such methods tend, however, to be computationally intensive, and may require specification of a null model to generate data, a specification that sometimes is not easy to come by when one is comparing competing, non-nested models. Nevertheless, simulation methods may be applied quite generally, and the asymptotic procedures proposed here should be viewed as complementary rather than competing.

These procedures allow for: a variety of linear and nonlinear techniques to estimate the models used to make the predictions, including maximum likelihood and generalized method of moments; serial correlation and conditional heteroskedasticity in regression disturbances and prediction errors; comparison of non-nested models; inference about moments of general, nonlinear functions of single or multi-period predictions and prediction errors. Using standard regularity conditions, I establish consistency and asymptotic normality of the estimators of the moments, and show that the asymptotic variance-covariance matrix may be estimated by familiar methods, including kernel techniques that allow for unknown forms of serial correlation and heteroskedasticity in prediction errors. A small set of simulations suggests that the procedure can yield accurately sized tests in samples of size typically available.

Whether and how one adjusts for error in estimation of regression parameters used to form predictions depends on a number of factors: the moment being examined, the regression technique, the fraction of the total sample used for out-of-sample estimation of the moments, and, of course, the probabilistic environment. Simulation evidence indicates that substantial size distortions may result if one ignores uncertainty about the regression vector that, according to this paper's asymptotic approximation, is asymptotically relevant for inference.[2]

Section 2 describes the basic environment, Section 3 presents technical assumptions, Section 4 derives asymptotic results, Section 5 presents Monte

[2] It should be acknowledged at the outset that the paper does not provide a formal statistical justification for the use of out-of-sample rather than in-sample analysis: the conventional statistical theory used here indicates that a model that does not fit well in-sample will not fit well out-of-sample. But out-of-sample comparisons sometimes bring surprising and important insights (e.g., Nelson (1972), Meese and Rogoff (1983)), perhaps because of inadvertent over-fitting that results from repeated profession-wide use of a limited body of data. This suggests that tests of predictive ability have good power, and, in any case, the tools developed here are relevant for work aimed at developing good methods to predict out-of-sample (e.g., Stock and Watson (1993)).

Carlo evidence, and Section 6 concludes. All proofs are in an Appendix. Additional proofs, simulation results, discussion, and interpretation have been omitted from the paper to save space, but are available in an additional Appendix and in two earlier versions of this paper that are collectively referenced as West (1994).

2. DESCRIPTION OF ENVIRONMENT

One is interested in an $(l \times 1)$ vector of moments Ef_t, where f_t depends as described below on the $(k \times 1)$ unknown parameter vector β^*, and the t subscript indicates that $f(\cdot)$ depends not only on β^* but also on observable data. If estimated serial correlation coefficients are used in prediction, these will be included along with regression parameters in β^*. If moments of predictions or prediction errors of competing sets of regressions are to be compared, the parameter vectors from the various regressions are stacked to form β^*. A special case of the analysis below is when predictions do not depend on estimated parameters, so that $k = 0$; to avoid repeated qualifications to statements concerning ranks of certain matrices, I assume $k > 0$ and note in Section 4 how the results specialize when $k = 0$.

Let $\tau \geq 1$ be the longest prediction horizon of interest. There are P *P*redictions or vectors of predictions in all. The first prediction is based on a parameter vector estimated using data from 1 to R (R as in *R*egress), the second on a parameter vector estimated using data from 1 to $R+1, \ldots,$ the last on a parameter vector estimated using data from 1 to $R + P - 1 \equiv T$.[3] Thus, $R + P - 1 + \tau = T + \tau$ is the size of available sample. Division of the available data into R, P, and τ is taken as given; when a model of interest was first estimated on data from $t = 1, \ldots, R$, there was a natural choice of R.

Let $\hat{\beta}_t$ denote an estimate of β^* that relies on data from period t and earlier. The estimate of Ef_t is constructed as

$$\bar{f} \equiv P^{-1} \sum_{t=R}^{T} f_{t+\tau}\left(\hat{\beta}_t\right).$$

To illustrate f_t, consider the scalar linear model $y_t = X_t'\beta^* + u_t$, and assume that the population one-step-ahead prediction of y_{t+1} is computed as $X_{t+1}'\beta^*$. The corresponding sample prediction would be $X_{t+1}'\hat{\beta}_t$, where, for example, $\hat{\beta}_t = (\sum_{s=1}^{t} X_s X_s')^{-1} \sum_{s=1}^{t} X_s y_s$ if ordinary least squares is the estimator.

If the object of interest is the one-step-ahead *mean squared prediction error* (*MSPE*), then $Ef_t = E(y_t - X_t'\beta^*)^2$ and $\bar{f} = P^{-1} \sum_{t=R}^{T} (y_{t+1} - X_{t+1}'\hat{\beta}_t)^2$. If the object of interest is one-step-ahead *mean prediction error* (*MPE*), then $Ef_t = E(y_t - X_t'\beta^*)$, and here, and in all subsequent examples but the last, $\bar{f}$ is constructed in the obvious way. If, further, $X_{t+1} = y_t$, so that $k = 1$ and the

[3] West (1994) considers how the results vary if instead: (a) a single estimate of β^* is used to form all P out of sample predictions, or (b) the sequence of P estimates of β^* is obtained from a series of rolling samples, each of size R.

model is a zero mean univariate autoregression, the τ-step-ahead MSPE is $E[y_{t+\tau}-X_{t+1}(\beta^*)^{\tau}]^2$. As indicated in this example, the dating of X_{t+1} is arbitrary, and X_{t+1} may be realized in period t or earlier. Finally, for the first autocovariance of u_t, $E(y_{t+1}-X'_{t+1}\beta^*)(y_t-X'_t\beta^*)$, we have $\bar{f}=P^{-1}\sum_{t=R}^{T}(y_{t+1}-X'_{t+1}\hat{\beta}_t)(y_t-X'_t\hat{\beta}_t)$. (The natural, and conventional, way to compute the corresponding sample moment is $P^{-1}\sum_{t=R}^{T}(y_{t+1}-X'_{t+1}\hat{\beta}_t)(y_t-X'_t\hat{\beta}_{t-1})$, which is asymptotically equivalent (West (1994)). To keep the notation relatively simple, in the theory I assume $\bar{f}$ depends only on $\hat{\beta}_t$, although I use the natural estimator in the simulations.)

It will rarely be the case in practice that $l=1$, as in these examples. But more realistic examples may be built from such simple ones by stacking the objects of interest into a vector f_t with $l>1$. See, for example, the experiment described in Section 5.2 below.

3. ASSUMPTIONS

The first assumption is that f_t is well approximated by a smooth quadratic in the neighborhood of the parameter vector. I use the following notation: for any differentiable function g_t: $R^m \to R^s$ and for x in the domain of g_t, let $\partial g_t/\partial x$ denote the $(s\times m)$ matrix of partial derivatives of g_t; 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 1 to t; for any variable x, $\sum x_s\equiv\sum_{s=1}^{t}x_s$.

ASSUMPTION 1: *In some open neighborhood N around β^*, and with probability one*:

(a) *$f_t(\beta)$ is measurable and twice continuously differentiable with respect to β.*

(b) *Let f_{it} be the ith element of f_t. For $i=1,\dots,l$ there is a constant $D<\infty$ such that for all t, $\sup_{\beta\in N}|\partial^2 f_{it}(\beta)/\partial\beta\,\partial\beta'|<m_t$ for a measurable m_t for which $Em_t<D$.*

Provided appropriate second moments exist, Assumption 1 holds trivially when the function is MPE, MSPE, or an autocovariance, the horizon is one period, and the model is linear. It also holds when these functions are based on: (a) multiperiod ARIMA forecasts and single or multiperiod reduced form forecasts for linear simultaneous equations models, all of which are nonlinear in the parameters; (b) predictions based on smooth transformations of forecasts constructed from regression estimates (e.g., Fair and Shiller (1990), West et al. (1993)).

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) *$B(t)\xrightarrow{a.s.}B$, B a matrix of rank k*; (b) *$H(t)=t^{-1}\sum_{s=1}^{t}h_s(\beta^*)$ for a $(q\times 1)$ orthogonality condition $h_s(\beta^*)$*; (c) *$Eh_s(\beta^*)=0$.*

$H(t)$ is a sample average of a $(q \times 1)$ orthogonality condition used to identify β^*, $B(t)$ a $(k \times q)$ matrix that selects a linear combination of these sample averages.

To illustrate, let $h_t \equiv h_t(\beta^*)$, and assume stationarity. If, say, the only regression equation is the scalar model $y_t = X_t'\beta^* + u_t$, and the regression technique is OLS, $B = (EX_tX_t')^{-1}$ and $h_t = X_t u_t$. Note that there is no presumption that u_t or $X_t u_t$ is serially uncorrelated (although it is of course required that $EX_t u_t = 0$). Suppose more generally that for y_t $(n \times 1)$ and a smooth function g: $R^k \to R^n$, the set of regression equations is $y_t = g(X_t, \beta^*) + u_t$. Let β^* be estimated by GMM: $\hat{\beta}_t$ is chosen to $\min(\Sigma \tilde{h}_s)'W(t)(\Sigma \tilde{h}_s)$, $\tilde{h}_s \equiv [y_s - g(X_s, \hat{\beta}_t)] \otimes Z_s$, Z_s a vector of instruments, $Eh_s \equiv Eu_s \otimes Z_s = 0$, and $W(t) \xrightarrow{p} W$, W p.d. and symmetric. Then

$$B(t) = -\left[A(t)t^{-1}\sum \frac{\partial \mathring{h}_s}{\partial \beta}\right]^{-1} A(t), \quad \text{where}$$

$$A(t) = \left[-t^{-1}\sum \frac{\partial h_s}{\partial \beta}(\hat{\beta}_t)\right]' W(t),$$

$$-\frac{\partial h_s}{\partial \beta} = \frac{\partial g_s}{\partial \beta} \otimes Z_s,$$

and $\partial \mathring{h}_s/\partial\beta$ is $\partial h_s/\partial\beta$ evaluated at points between $\hat{\beta}_t$ and β^*; optimal GMM estimation (again, allowed but not required) entails $W = \sum_{j=-\infty}^{\infty} Eh_t h_{t-j}'$. The analogous definition for maximum likelihood is that $B(t)$ is the inverse of the Hessian evaluated at a vector between the population parameter vector and its estimate, and $H(t)$ is the score. See Hansen (1982) or Andrews (1987) for primitive conditions that imply Assumption 2.

Let

$$f_t \equiv f_t(\beta^*), \quad f_{t\beta} \equiv \frac{\partial f_t}{\partial \beta}(\beta^*), \quad F \equiv Ef_{t\beta}.$$

ASSUMPTION 3: (a) *For some* $d > 1$, $\sup_t E\|[\mathrm{vec}(f_{t\beta})', f_t', h_t']'\|^{4d} < \infty$, *where* $\|\cdot\|$ *denotes Euclidean norm.* (b) $[\mathrm{vec}(f_{t\beta} - F)', (f_t - Ef_t)', h_t']'$ *is strong mixing, with mixing coefficients of size* $-3d/(d-1)$. (c) $[\mathrm{vec}(f_{t\beta})', f_t', h_t']'$ *is covariance stationary.* (d) *Let* $\Gamma_{ff}(j) = E(f_t - Ef_t)(f_{t-j} - Ef_t)'$, $S_{ff} = \sum_{j=-\infty}^{\infty} \Gamma_{ff}(j)$. *Then* S_{ff} *is p.d.*

Assumption 3 allows serial correlation and conditional heteroskedasticity in both $f_t - Ef_t$ and h_t. The assumption is rather stronger than is necessary, but is convenient.

ASSUMPTION 4: $R, P \to \infty$ *as* $T \to \infty$, *and* $\lim_{T\to\infty}(P/R) = \pi$, $0 \le \pi \le \infty$; $\pi = \infty \Leftrightarrow \lim_{T\to\infty}(R/P) = 0$.

$R \to \infty$ allows use of asymptotic theory to formalize how uncertainty about β^* affects the estimated moment; even in the absence of such uncertainty, $P \to \infty$ is obviously necessary to allow asymptotic analysis. The parameter π is taken as given. Given that the forecast horizon τ is fixed, the decision to design a symbol to the limit of P/R rather than of, say, $P/(T+\tau)$ is arbitrary. That τ is fixed also suggests that the relevant applications are those in which $\tau \ll R, P$.

Throughout, I maintain Assumptions 1–4.

4. ASYMPTOTIC DISTRIBUTION OF THE OUT-OF-SAMPLE MEAN OF f_t

The Appendix applies a mean value expansion of $f_{t+\tau}(\hat{\beta}_t)$ around β^* to show

$$P^{1/2}\left(\bar{f} - Ef_t\right) = P^{-1/2}\sum (f_{t+\tau} - Ef_t) + FB\left[P^{-1/2}\sum H(t)\right] + o_p(1). \tag{4.1}$$

The first term on the right-hand side represents uncertainty about $\bar{f}$ that is present even when β^* is known. The second term represents uncertainty about β^*. To state precisely how the asymptotic variance of $P^{1/2}(\bar{f} - Ef_t)$ reflects both forms of uncertainty requires some additional notation. Let

$$\Gamma_{fh}(j) = E(f_t - Ef_t)h'_{t-j}, \quad S_{fh} = \sum_{j=-\infty}^{\infty} \Gamma_{fh}(j), \quad \Gamma_{hh}(j) = Eh_t h'_{t-j}, \tag{4.2}$$

$$S_{hh} = \sum_{j=-\infty}^{\infty} \Gamma_{hh}(j), \quad S = \begin{pmatrix} S_{ff} & S_{fh}B' \\ BS'_{fh} & BS_{hh}B' \end{pmatrix} \equiv \begin{pmatrix} S_{ff} & S_{fh}B' \\ BS'_{fh} & V_\beta \end{pmatrix},$$

$$\Pi = \Pi(\pi) = 1 - \pi^{-1}\ln(1+\pi) \quad \text{for} \quad 0 < \pi < \infty,$$

$$\Pi = 0 \quad \text{for} \quad \pi = 0, \quad \Pi = 1 \quad \text{for} \quad \pi = \infty.$$

Apart from a scale factor, S is the $(l+k)\times(l+k)$ spectral density of $[(f_t - Ef_t)', h'_t B']'$ at frequency zero. The $(k \times k)$ matrix $V_\beta \equiv BS_{hh}B'$ is the asymptotic variance-covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta^*)$. The scalar function Π is discussed in comment 4 below.

LEMMA 4.1: (a) $P^{-1/2}\sum(f_{t+\tau} - Ef_t) \overset{A}{\sim} N(0, S_{ff})$, $E[P^{-1}\sum H(t) \sum H(t)'] \to 2\Pi S_{hh}$, $E[P^{-1}\sum(f_{t+\tau} - Ef_t)\sum H(t)'] \to \Pi S_{fh}$.

(b) *If S is p.d.*, $P^{-1/2}[\sum(f_{t+\tau} - EF_t)', \sum H(t)'B']' \overset{A}{\sim} N(0, V)$, *where*

$$V = \begin{pmatrix} S_{ff} & \Pi S_{fh}B' \\ \Pi BS'_{fh} & 2\Pi V_\beta \end{pmatrix};$$

V has full rank if $\pi \neq 0$.

THEOREM 4.1: (a) *If $\pi = 0$ or $F = 0$*, $P^{1/2}(\bar{f} - Ef_t) \overset{A}{\sim} N(0, \Omega)$, $\Omega = S_{ff}$.

(b) *If S is p.d.*, $P^{1/2}(\bar{f} - Ef_t) \overset{A}{\sim} N(0, \Omega)$, $\Omega = S_{ff} + \Pi(FBS'_{fh} + S_{fh}B'F') + 2\Pi FV_\beta F'$.

The two parts of Theorem 4.1 of course are not mutually exclusive; the formula for the rank l matrix Ω in Theorem 4.1(b) reduces to S_{ff} under the conditions of Theorem 4.1(a). Primitive conditions that ensure that S is p.d. may be found in, for example, Durbin (1970) when Ef_t equals the first autocovariance of a disturbance, and Hamilton (1994, p. 301) when $Ef_t = MSPE$ in a vector autoregression. Consistency ($\bar{f} \xrightarrow{p} Ef_t$) obviously follows from Theorem 4.1. If β^* is known (no regression estimation is required for prediction), $B = 0$ and, as noted by Diebold and Mariano (1994), $P^{1/2}(\bar{f} - Ef_t) \overset{A}{\sim} N(0, S_{ff})$.

The following six comments on Theorem 4.1 are relevant. The first three and the fifth discuss cases in which $P^{1/2}(\bar{f} - Ef_t) \overset{A}{\sim} N(0, S_{ff})$ and uncertainty about β^* is asymptotically irrelevant for inference about Ef_t; I make a point of noting these because in such cases inference is greatly simplified.

1. From part (a) of Theorem 4.1, one condition for asymptotic irrelevance is when $0 = \pi \equiv \lim_{T \to \infty}(P/R)$. The intuition is that one may treat the estimate of β^* as known, if the asymptotic distribution of $\bar{f}$ is derived under the assumption that for arbitrarily large T there will be an arbitrarily large number of observations used for estimation of β^* relative to the number used for estimation of Ef_t. This point was argued informally by Chong and Hendry (1986) in the context of tests of forecast encompassing.

2. Theorem 4.1(a) also indicates that asymptotic irrelevance holds when $0 = F \equiv E[(\partial f_t/\partial\beta)(\beta^*)]$. The leading case here is MSPE when the predictors are uncorrelated with the prediction error. In the scalar linear model $y_t = X_t'\beta^* + u_t$ with one step ahead population prediction $X_{t+1}'\beta^*$; for example, one has $f_t(\beta^*) = (y_t - X_t'\beta^*)^2$,

$$\frac{\partial f_t}{\partial \beta}(\beta^*) = -2u_t X_t' \Rightarrow F = -2Eu_t X_t' \quad \text{and} \quad F = 0 \quad \text{if} \quad Eu_t X_t' = 0.$$

More generally, asymptotic irrelevance applies for the MSPE of single or multiperiod predictions from, for example, the reduced form of simultaneous equations models. When one is comparing non-nested models, the appropriate asymptotic variance is still S_{ff} as long as each prediction error is uncorrelated with the corresponding predictors, even if the competing models use different information sets so that one model's prediction error is correlated with another model's prediction.

3. On the other hand, asymptotic irrelevance sometimes applies to an out-of-sample test when it does not apply to the corresponding in-sample test. The technical condition here is that $\Pi(FBS_{fh}' + S_{fh}B'F') + 2\Pi FV_\beta F' = 0$, so that the variance induced by error in estimation of β^* is exactly offset by the covariance between such terms and terms that would be present even if β^* were known.

One such case is in testing for first order serial correlation in one step ahead prediction errors in certain models. To illustrate, consider once again the scalar linear model, now specialized for simplicity so that X_t is a scalar and $X_t = y_{t-1}$ $\Rightarrow y_t = y_{t-1}\beta^* + u_t$. Assume $E(u_t | y_{t-1}, y_{t-2}, \ldots) = 0$ and let the object of interest be $\rho \equiv (Eu_t^2)^{-1}Eu_tu_{t-1} \equiv \gamma_0^{-1}\gamma_1 \equiv \gamma_0^{-1}Ef_t(=0)$. Let the estimator be OLS, let $\hat{u}_t$ be the one-step-ahead prediction error, $\hat{\gamma}_j = P^{-1}\sum \hat{u}_t\hat{u}_{t-j}$, $\bar{f} = \hat{\gamma}_1$, $\hat{\rho} \equiv \hat{\gamma}_0^{-1}\bar{f}$. Theorem 4.1 indicates that the asymptotic variance of $P^{1/2}\hat{\rho}$ is

$$(4.3) \qquad \gamma_0^{-2}\Omega = \gamma_0^{-2}\Big\{Eu_t^2u_{t-1}^2 - 2\Pi\left[(Eu_ty_t)(Ey_t^2)^{-1}Eu_t^2u_{t-1}y_{t-1}\right] + 2\Pi\left[(Eu_ty_t)^2(Ey_t^2)^{-2}(Ey_{t-1}^2u_t^2)\right]\Big\}.$$

It may be seen that if $E(u_t^2|y_{t-1}, y_{t-2}, \ldots) = Eu_t^2 \equiv \gamma_0$, the right-hand side of (4.3) reduces to $\gamma_0^{-2}\gamma_0^2 = 1$: the standard error to be used in hypothesis testing is the familiar $1/P^{1/2}$ that applies as well when β^* is known. In contrast, it is well known (e.g., Durbin (1970)) that if in-sample residuals are used, the relevant asymptotic variance includes a term reflecting uncertainty about β^*. The result illustrated in (4.3) holds generally when predicting from the reduced form of a linear simultaneous equations model estimated by 3SLS (a VAR estimated by OLS is a special case), as long as the disturbance vector is a conditionally homoskedastic martingale difference.

4. Such asymptotic irrelevance of course does not always apply. One example is MSPE when the predictors are correlated with the prediction error (see Section 5 below). Also, in the tests for serial correlation and for zero covariance just described in comment 3, asymptotic irrelevance typically fails if the disturbance is conditionally heteroskedastic and there are lagged dependent variables, even if the disturbance is a martingale difference. This may be verified by working through (4.3) with a specific parametric model for the conditional heteroskedasticity.

To interpret Theorem 4.1(b) in these and other cases, note first that $0 \le \Pi \le 1$ and $\Pi'(\pi) > 0$. Uncertainty about β^* adds $\Pi[FBS'_{fh} + S_{fh}B'F'] + 2\Pi FV_\beta F'$ to the asymptotic variance of $P^{1/2}(\bar{f} - Ef_t)$. In comparison to the hypothetical case when β^* is known, uncertainty about β^* may cause an asymptotic standard error to be larger, smaller, or, as noted in comment 3, no different. In those cases when uncertainty about β^* causes the standard error of interest to be larger (smaller), the standard error increases (decreases) with the fraction of the sample set aside for prediction, that is, with π.

5. This comment concerns MPE. Theorem 4.1 may not be applicable because S may not be p.d. despite positive definiteness of S_{ff} and of V_β. West (1994) shows, however, it is generally the case that a vector of mean prediction errors is asymptotically normal with variance-covariance matrix $S_{ff}(= \sum_{j=-\infty}^{\infty} Eu_tu_{t-j}$ in the model $y_t = X_t'\beta^* + u_t)$. In contrast to the serial correlation example (comment 3), such asymptotic irrelevance holds even if the disturbance vector is conditionally heteroskedastic and serially correlated.

6. The final comment concerns consistent estimation of Ω. Inspection of Theorem 4.1 indicates that one must estimate S_{ff} and, when uncertainty about β^* matters, one must also supply a value for Π and estimate B, F, S_{fh}, and S_{hh}. Sample analogues for the population quantities may be used (West (1994)): One

can set

$$\Pi = 1 - \left(\frac{P}{R}\right)^{-1} \ln\left(1 + \frac{P}{R}\right).$$

By assumption 2, $\hat{B} \equiv B(T) \xrightarrow{p} B$. Next,

$$\hat{F} \equiv P^{-1} \sum_{t=R}^{T} \frac{\partial f_{t+\tau}}{\partial \beta}(\hat{\beta}_t) \xrightarrow{p} F$$

and sample autocovariances of f_t and h_t are consistent (e.g., letting $\hat{f}_{t+\tau} \equiv f_{t+\tau}(\hat{\beta}_t)$ and $\hat{h}_t \equiv h_t(\hat{\beta}_t)$, $P^{-1} \sum_{t=R}^{T} (\hat{f}_{t+\tau} - \bar{f})(\hat{f}_{t+\tau} - \bar{f})' \xrightarrow{p} \Gamma_{ff}(0)$, $P^{-1} \sum_{t=R}^{T} (\hat{f}_{t+\tau} - \bar{f})\hat{h}_t' \xrightarrow{p} \Gamma_{fh}(0)$, $P^{-1} \sum_{t=R}^{T} \hat{h}_t \hat{h}_t' \xrightarrow{p} \Gamma_{hh}(0)$, with analogous results for auto- and cross-covariances at nonzero lags). Often, this suffices for inference, as in the simulations below. But if not, upon mild strengthening of the assumptions about h_t along the lines of Andrews (1991), conventional nonparametric kernel estimators, such as the Bartlett or quadratic spectral, may be used. When applied to the sample auto- and cross-covariances of f_t and h_t, these yield estimates $\hat{S}_{ff}$, $\hat{S}_{fh}$, and $\hat{S}_{hh}$ that are consistent for S_{ff}, S_{fh}, and S_{hh}, and, as well, a consistent p.s.d. estimate $\hat{\Omega} \equiv \hat{S}_{ff} + \Pi(\hat{F}\hat{B}\hat{S}_{fh}' + \hat{S}_{fh}\hat{B}'\hat{F}') + 2\Pi\hat{F}\hat{B}\hat{S}_{hh}\hat{B}'\hat{F}'$.

5. MONTE CARLO EVIDENCE

Here I present two small Monte Carlo experiments, aimed at giving a feel for whether Theorem 4.1's asymptotic approximation might yield well-sized test statistics. Both involve tests on various moments of one-step-ahead prediction errors in models estimated by instrumental variables. The first does not require accounting for uncertainty about the regression vector, the second does.

In the experiments, each of 5000 artificial samples of size 200 was split into 12 different regression (R) and prediction (P) samples: $R = 25$, $P = 25$, 50, 100, 150, 175; $R = 50$, $P = 25$, 50, 100, 150; $R = 100$, $P = 25$, 50, 100. This range for P/R (from .25 to 7), as well as the values of $T = P + R - 1$, seem broad enough to include most relevant empirical work. 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. I report tests of nominal size .05. Tests of nominal size .01 and .10 worked equally well (West (1994)).

5.1. *Predictions Using Predetermined Variables in a Simultaneous Equations Model*

Here I consider a bivariate model, borrowed from Schmidt (1977):

$$\begin{aligned} y_{1t} &= \beta_{11} + \beta_{12} y_{2t} + \beta_{13} y_{1t-1} + \beta_{14} x_{1t} + u_{1t} \equiv X_{1t}' \beta_1 + u_{1t}, \\ y_{2t} &= \beta_{21} + \beta_{22} y_{1t} + \beta_{23} y_{2t-1} + \beta_{24} x_{2t} + u_{2t} \equiv X_{2t}' \beta_2 + u_{2t}. \end{aligned} \tag{5.1}$$

In (5.1), $(u_{1t}, u_{2t})' \equiv u_t$ is an iid normal vector. The exogenous variables x_{1t} and x_{2t} were assumed to follow univariate AR(1)'s with common parameter ϕ and iid standard normal disturbances that are independent of one another and of u_t. The 5×1 vector of instruments is the list of predetermined variables in the reduced form, $Z_t \equiv (1, y_{1t-1}, y_{2t-1}, x_{1t}, x_{2t})'$.

Most of the parameters were borrowed from Schmidt (1977): $\beta^* \equiv (\beta_1', \beta_2')' = (0, .2, .5, 1, 0, 1, .5, 2)'$; $\phi = 0.8$; $Eu_{1t}^2 = Eu_{2t}^2 = 1$, $Eu_{1t}u_{2t} = 0.5$. For 5000 samples, an initial value of the (4×1) vector $(y_{10}, y_{20}, x_{10}, x_{20})'$ was drawn from the unconditional normal distribution of $(y_{1t}, y_{2t}, x_{1t}, x_{2t})'$, and samples of size 200 were generated. For each value of R and P, and for $t = R$ to $t = R + P - 1$, 3SLS was used to estimate the structural equations (5.1) using the 10×1 $(q = 10)$ orthogonality condition $h_t = u_t \otimes Z_t$. The implied reduced form coefficients were computed, y_{1t+1} was predicted from the reduced form, and a $(P \times 1)$ vector of one step ahead prediction errors $\{\hat{v}_{1t+1}\}$ was constructed.

This vector was used to perform the following three hypothesis tests: (a) Zero MPE: $\bar{f} = P^{-1} \sum \hat{v}_{1t+1}$, $\hat{\Omega} = P^{-1} \sum \hat{v}_{1t+1}^2$; from comments 5 and 6 in Section 4, $\hat{\Omega} \xrightarrow{p} \Omega = Ev_{1t}^2$. (b) MSPE equals the population value of $Ev_{1t+1}^2 (= 1.9375)$: $\bar{f} = P^{-1} \sum \hat{v}_{1t+1}^2$, $\hat{\Omega} = P^{-1} \sum \hat{v}_{1t+1}^4 - (P^{-1} \sum \hat{v}_{1t+1}^2)^2$; from comments 2 and 6 in Section 4, $\hat{\Omega} \xrightarrow{p} \Omega = Ev_{1t}^4 - (Ev_{1t}^2)^2 (= 2(Ev_{1t}^2)^2)$. (c) Zero first order serial correlation coefficient: the object of interest is $[P^{-1} \sum (\hat{v}_{1t}\hat{v}_{1t+1})]/(P^{-1} \sum \hat{v}_{1t+1}^2)$, and, in accordance with comment 3 in Section 4, the standard error used in hypothesis testing was simply $P^{-1/2}$.

Table I has the results for nominal .05 tests. There is a mild tendency to overreject; of the 36 entries in the Table, only 8 are less than .05. (The entries are not, however, independent. If a test happens to overreject for a given R and P, it is likely to do so for the same R and slightly larger P as well.) But the figures, which range between .034 and .100, are also close to .05.

5.2. *MSPE with Endogenous Right-Hand-Side Variables as Predictors*

Here I consider the system

$$y_t = \beta_{11} + \beta_{12} w_{1t} + u_{1t} \equiv X_{1t}' \beta_1 + u_{1t}, \tag{5.2a}$$

$$y_t = \beta_{21} + \beta_{22} w_{2t} + u_{2t} \equiv X_{2t}' \beta_2 + u_{2t}, \tag{5.2b}$$

where the w_{it}'s are correlated with the u_{it}'s and $Eu_{1t}^2 = Eu_{2t}^2$. The idea is that (5.2a) and (5.2b) are competing models for y_t. Write (5.2) as

$$\begin{pmatrix} y_t \\ y_t \end{pmatrix} \equiv Y_t = X_t' \beta^* + u_t \equiv \begin{pmatrix} 1 & w_{1t} & 0 & 0 \\ 0 & 0 & 1 & w_{2t} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Let $(z_{1t}, z_{2t}, v_t)'$ be an iid normal random vector with an identity covariance matrix. Let $w_{1t} = z_{1t} + v_t$, $w_{2t} = z_{2t} + v_t$. The data are generated as $y_t = w_{1t} + w_{2t} + v_t$. So in (5.2a), $\beta_{11} = \beta_{21} = 0$, $\beta_{12} = \beta_{22} = 1$. Each equation was estimated by 2SLS, with a vector of instruments $(1, z_{1s})' \equiv Z_{1s}$ (equation (5.2a)) or $(1, z_{2s})' \equiv$

TABLE I

SIZE OF NOMINAL .05 TESTS, EXPERIMENT 1

	A. Mean Prediction Error				
	P				
R	25	50	100	150	175
25	.059	.064	.060	.057	.060
50	.059	.063	.057	.059	
100	.055	.058	.058		
	B. Mean Squared Prediction Error				
	P				
R	25	50	100	150	175
25	.076	.066	.061	.063	.066
50	.081	.060	.057	.056	
100	.100	.075	.055		
	C. First Order Serial Correlation				
	P				
R	25	50	100	150	175
25	.037	.046	.050	.049	.050
50	.038	.045	.053	.052	
100	.034	.040	.043		

Notes: 1. As detailed in Section 5.1, the following was done 5000 times: a. 200 observations on the variables in a bivariate simultaneous equations model were generated. b. The model was estimated 175 times by 3SLS, using observations 1–25, 1–26, ..., and 1–199. The implied reduced form coefficients were used to make one-period-ahead predictions of one of the two endogenous variables. c. For the values of R and P given above, a $(P \times 1)$ vector of prediction errors was formed using predictions dated $R, R+1, \ldots, R+P-1$. d. The vector of sample prediction errors was used to test: zero mean (panel A), variance equal to the underlying population value (panel B), zero serial correlation (panel C).

2. According to this paper's theory, each statistic is asymptotically $\chi^2(1)$. Each panel reports the fraction of the 5000 statistics that were greater than 3.84, which is the .05 critical value for a $\chi^2(1)$ random variable.

Z_{2s} (5.2b). Let $\hat{\beta}_{it} = (t^{-1} \sum_{s=1}^{t} Z_{is} X_{is}')^{-1}(t^{-1} \sum_{s=1}^{t} Z_{is} y_s)$ be the 2SLS estimate of β_i. For each value of R and P, and for $t = R$ to T, one step ahead prediction errors were obtained as $\hat{u}_{it+1} \equiv y_{t+1} - X_{it+1}' \hat{\beta}_{it}$. The hypothesis tested is that $Eu_{1t}^2 = Eu_{2t}^2 \Rightarrow \bar{f} = (P^{-1} \sum \hat{u}_{1t+1}^2, P^{-1} \sum \hat{u}_{2t+1}^2)'$.

Note that realized right-hand-side endogenous variables are used for "prediction." In the forecasting literature such "conditional" or "ex-post" forecasts are made when one is not interested in ex-ante prediction but evaluation of predictive ability of a model given a path for some unmodelled (but endogenous, in a larger system) set of variables. Examples include Meese and Rogoff (1983, 1988) and Oliner et al. (1993). In such cases, F is nonzero and sampling error in estimation of β^* is asymptotically relevant (except when $\pi = 0$).

In particular, in the system under consideration here,

$$F = -2E\begin{pmatrix} u_{1t} & 0 \\ 0 & u_{2t} \end{pmatrix} X_t' \equiv -2EU_t X_t' \Rightarrow F(1,2) = -2Eu_{1t}w_{1t} \neq 0,$$

$$F(2,4) = -2Eu_{2t}w_{2t} \neq 0.$$

Since $(u_{1t}, z_{1t}, u_{2t}, z_{2t})$ is normally distributed, $S_{fh} = 0 \Rightarrow \Omega = S_{ff} + 8\Pi(EU_tX_t')V_\beta(EX_tU_t')$. Thus use of S_{ff}, which abstracts from sampling error in estimation of β^*, will yield standard errors and confidence intervals that are too small, and will lead to too many rejections at any specified significance level. The lower the asymptotic ratio of observations used for prediction to those used for regression (the lower is π), the less will sampling error in estimation of β^* affect inference, and the closer will S_{ff} be to Ω.

$\hat{\Omega}$ was constructed as follows. (i) In a computationally convenient variant of the estimator suggested in comment 6 in Section 4, $V_\beta \equiv B'S_{hh}B$ was estimated in conventional in-sample fashion, using the largest regression sample: For $s = 1, \ldots, T$, let $\tilde{u}_s = Y_s - X_s'\hat{\beta}_T$, $\hat{h}_s \equiv (Z_{1s}'\tilde{u}_{1s}, Z_{2s}'\tilde{u}_{2s})'$. Then $\hat{S}_{hh} = T^{-1}\sum_{s=1}^T \hat{h}_s\hat{h}_s'$. For $i = 1,2$, the (i,i) block of $\hat{B}$ is $(T^{-1}\sum_{s=1}^T Z_{is}X_{is}')^{-1}$; the off-diagonal blocks are zero. (ii) As suggested in comment 6 in Section 4, S_{ff} and F were computed using the out of sample residuals. Let $\hat{f}_{t+1} = (\hat{u}_{1t+1}^2, \hat{u}_{2t+1}^2)'$. Then $\hat{S}_{ff} = P^{-1}\sum_{t=R}^T(\hat{f}_{t+1} - \bar{f})(\hat{f}_{t+1} - \bar{f})'$,

$$\hat{F} = \begin{pmatrix} 0 & -2P^{-1}\sum_{t=R}^T w_{1t+1}\hat{u}_{1t+1} & 0 & 0 \\ 0 & 0 & 0 & -2P^{-1}\sum_{t=R}^T w_{2t+1}\hat{u}_{2t+1} \end{pmatrix}.$$

Let $\alpha = (1\ \ -1)'$. The test statistic is $P(\alpha'\bar{f})^2/(\alpha'\hat{\Omega}\alpha) \overset{A}{\sim} \chi^2(1)$, where $\hat{\Omega} \equiv \hat{S}_{ff} + 2\Pi\hat{F}\hat{V}_\beta\hat{F}'$, $\hat{V}_\beta \equiv \hat{B}'\hat{S}_{hh}\hat{B}'$. A test statistic that ignores uncertainty about β^* is $P(\alpha'\bar{f})^2/(\alpha'\hat{S}_{ff}\alpha)$; it seems of interest to evaluate such a statistic since it seems to be one used by Meese and Rogoff (1988) and suggested by Diebold and Mariano (1994). For such an evaluation to be empirically relevant, one would want an empirically plausible ratio of $\alpha'FV_\beta F'\alpha$ (the term due to uncertainty about β^*) to $\alpha'S_{ff}\alpha$ (the term that would be present even if β^* were known): the test statistic that ignores $\alpha'FV_\beta F'\alpha$ will behave arbitrarily poorly for an arbitrarily high ratio of $\alpha'FV_\beta F'\alpha$ to $\alpha'S_{ff}\alpha$. Since predictive accuracy figures prominently in the exchange rate literature, I fit a convenient model (Meese (1986)). As described in a footnote, the resulting estimates are consistent with this experiment's ratio of $\alpha'FV_\beta F'\alpha$ to $\alpha'S_{ff}\alpha$.[4]

In Table II, panel A presents results when a consistent estimator of Ω is used, with π set to P/R, while panel B presents results when the term due to uncertainty about β^* is ignored. In panel A, tests are slightly more poorly sized than in the previous experiment when P and R are both less than 50. The remaining entries are all between .049 and .075.

Panel B indicates that ignoring uncertainty about β^* can result in substantial overrejection, with nominal .05 tests having actual sizes larger than .50. As predicted by the theory, for given P the size distortions are smaller for larger R

[4] I fit Meese's (1986) model to monthly Deutschemark-dollar data 1974:3–1988:8, using a lag of Meese's scalar measure of fundamentals as the instrument. The estimates of the scalars S_{ff} and $FV_\beta F'$ yielded $FV_\beta F' \approx 5S_{ff}$. With the DGP described above, it may be shown that $\alpha'FV_\beta F'\alpha \approx 4\alpha'S_{ff}\alpha$. So this experiment is, if anything, a little conservative about uncertainty about β^* (since $4 < 5$).

TABLE II

SIZE OF NOMINAL .05 TESTS ON EQUALITY OF MSPE, EXPERIMENT B

	A. $\pi = P/R$				
	P				
R	25	50	100	150	175
25	0.022	0.035	0.057	0.071	0.075
50	0.027	0.036	0.058	0.064	
100	0.049	0.056	0.063		

	B. $\pi = 0$				
	P				
R	25	50	100	150	175
25	0.378	0.446	0.489	0.508	0.513
50	0.276	0.365	0.421	0.453	
100	0.198	0.269	0.349		

Notes: 1. As detailed in Section 5.2, the following was done 5000 times: a. 200 observations of the variables in a pair of linear regressions were generated. b. Each of these regressions was estimated 175 times by 2SLS, using observations $1-25, 1-26, \ldots$, and $1-199$. The regressions have the same left-hand-side variable y_t. The resulting structural coefficients and the realized values of the right-hand-side endogenous variables were used to predict y_t one step ahead. c. For the values of R and P given above, and for each of the two competing models, a $(P \times 1)$ vector of prediction errors were formed using predictions dated $R, R+1, \ldots, R+P-1$. d. Equality of mean squared prediction error was tested, using the two vectors of sample prediction errors.
2. According to this paper's theory, the statistic in panel A is asymptotically $\chi^2(1)$, that in panel B is not. The panel A statistic accounts for the fact that the predictions were made using estimated parameters; the panel B statistic is one that would be $\chi^2(1)$ if the predictions were made using the population regression parameters rather than estimates. Each panel reports the fraction of the 5000 statistics that were greater than 3.84, which is the .05 critical value for a $\chi^2(1)$ random variable.

(i.e., for a smaller ratio of P to R; recall that as $\pi \equiv \lim(P/R) \to 0$, the term in Ω due to uncertainty about $\beta^* \to 0$).

One cautionary note: Recent literature suggests circumstances under which it may be difficult to obtain accurately sized in-sample tests (e.g., Newey and West (1994), Nelson and Startz (1991)). Such circumstances no doubt lead to poorly sized out-of-sample tests as well. In that sense, the simulation evidence presented here probably is unduly supportive of the asymptotic approximation.

6. CONCLUSION

Priorities for future research include: allowing for functions that are not differentiable, such as mean absolute prediction error and functions of the predicted sign of a left-hand-side variable; allowing for nonparametric estimators of regression parameters; application of the results in this paper to computationally convenient regression tests of predictive ability; and analysis of the power of tests applied to the moments of out-of-sample prediction errors, in a framework broad enough to include models that are overfit in-sample.

Dept. of Economics, University of Wisconsin, 1180 Observatory Dr., Madison, WI 53706-1393, U.S.A.

Manuscript received May, 1994; final revision received October, 1995.

APPENDIX

NOTATION: "$\sup_t$" means "$\sup_{R \le t \le T}$"; "var", "cov" denote variance and covariance; all limits are taken as the sample size T goes to infinity; summations of variables indexed by t, $t+\tau$, or $t+j$ run from $t=R$ to $t=R+P-1 \equiv T$; summations of variables indexed by s run from $s=1$ to $s=t$.

LEMMA A1: (a) *For* $0 \le a < .5$, $P^{-1/2}\sum t^{-1+a} \to 0$; (b) $P^{-1/2}\sum t^{-1/2} = O(1)$.

PROOF: (a) Choose a, $0 < a < .5$. For $\pi < \infty$, use $P^{-1/2}\sum t^{-1+a} \le P^{-1/2}(PR^{-1+a}) = P^{-1/2+a}(P/R)^{1-a} \to 0$. For $\pi = \infty$, use $P^{-1/2}\sum t^{-1+a} \le P^{-1/2}\int_{R-1}^{T} x^{-1+a}\,dx \le a^{-1}P^{-1/2}(R+P-1)^a \to 0$. For $a=0$ the result follows since $\sum t^{-1+a}$ is increasing in a. (b) Follows by setting $a=.5$ in the proof of part (a).

LEMMA A2: *Let* $x_t = [\operatorname{vec}(f_{t\beta}-F)', (f_t - Ef_t)', h_t']'$, $\gamma_j \equiv Ex_j x_{t-j}'$. *Then* (a) $\sum_{j=-\infty}^{\infty}|j||\gamma_j| < \infty$, (b) $[P^{-1/2}\sum t^{-1}]^2 \sum_{j=-\infty}^{\infty}|j||\gamma_j| \to 0$.

PROOF: (a) In light of Ibragimov and Linnik (1971, p. 307), (a) follows from Assumption 3. (b) follows from part (a) and Lemma A1(a).

LEMMA A3: *For* $0 \le a < .5$: (a) $\sup_t|t^a H(t)| \xrightarrow{p} 0$; (b) $\sup_t|t^a(\hat\beta_t - \beta^*)| \xrightarrow{p} 0$.

PROOF: (a) In this proof only, $[\cdot]$ denotes "integer part of," $q=1$, and k_i is an unimportant constant. Let $\{\alpha_j\}$ be the mixing coefficients. Note that by Assumption 3, α_j is of size $-3d/(d-1)$ $\Rightarrow \alpha_j$ is also of size $-2d/(d-1) \Rightarrow h_t$ is a mixingale with Hall and Heyde's (1980, p. 19) symbols c_n and ψ_j defined as $c_n \equiv \max\{1, \sup_{t \ge 0} E|h_t|^{4d/(d+1)}\}$, $\psi_j \equiv 5(\alpha_{[j/2]})^{(1/2)-\{(d+1)/4d\}} = 5(\alpha_{[j/2]})^{(d-1)/4d}$ (McLeish (1975, p. 837)). Thus, $\psi_j < k_1([j/2]^{-3d/(d-1)})^{(d-1)/4d} = k_1[j/2]^{-3/4} \le k_1\{(j-1)/2\}^{-3/4}$ $\le k_2 j^{-3/4} \Rightarrow \psi_j < k_3 j^{-1/2}(\log j)^{-2}$. Since, finally, $\sum_{t=0}^{\infty}(t^{-1+a})^2 < \infty$ for $0 \le a < .5$, the result follows from Hall and Heyde (1980, Theorem 2.21). (b) By Assumption 2, $\sup_t|B(t)-B| \xrightarrow{p} 0$;

$$
\begin{aligned}
\sup_t|t^a(\hat\beta_t - \beta^*)| &\equiv \sup_t|t^a B(t)H(t)| \\
&\le \sup_t|t^a[B(t)-B]H(t)| + q|B|\left[\sup_t|t^a H(t)|\right] \\
&\le q\left[\sup_t|B(t)-B|\right]\left[\sup_t t^a|H(t)|\right] + q|B|\left[\sup_t|t^a H(t)|\right] \xrightarrow{p} 0.
\end{aligned}
$$

LEMMA A4: (a) $P^{-1/2}\sum(f_{t+\tau,\beta}-F)BH(t) = o_p(1)$; (b) $P^{-1/2}\sum|(f_{t+\tau,\beta}-F)[B(t)-B]H(t)| = o_p(1)$; (c) $P^{-1/2}\sum[B(t)-B]H(t) = o_p(1)$.

PROOF: (a) Let $v_t = (f_{t+\tau,\beta}-F)$, redefine Bh_t as h_t and let $\gamma_j \equiv Ev_t h_{t-j}'$. We have $|EP^{-1/2}\sum v_t H(t)| = P^{-1/2}|[R^{-1}(\gamma_0 + \cdots + \gamma_{R-1}) + \cdots + (R+P-1)^{-1}(\gamma_0 + \cdots + \gamma_{R+P-2})]| \le P^{-1/2}[R^{-1} + \cdots + (R+P-1)^{-1}]\sum_{j=0}^{\infty}|\gamma_j| \to 0$ by Lemma A1(a). Assumption 3(a) bounds the fourth moments of $[\operatorname{vec}(v_t)', h_t']'$ in such a way that $\lim \operatorname{var}[P^{-1/2}\sum v_t H(t)] = 0$ (West (1994)) and the result follows from Chebyshev's inequality. (b) Let $v_t = (f_{t+\tau,\beta}-F)$ and for simplicity assume $l=k=q=1$. We have $|P^{-1/2}\sum v_t[B(t)-B]H(t)| \le [\sup_t|B(t)-B|][P^{-1/2}\sum|v_t H(t)|]$ and since $[\sup_t|B(t)-B|] \xrightarrow{p} 0$ by Assumption 2 it suffices to show $P^{-1/2}\sum|v_t H(t)| = O_p(1)$. From Assumption 3, $E[v_t H(t)]^2 \le t^{-1}c$ for a constant c that does not depend on t; in the fourth order stationary case, for example, $c = Ev_t^2\sum_{j=-\infty}^{\infty}|Eh_t h_{t-j}| + 2(\sum_{j=-\infty}^{\infty}|Ev_t h_{t-j}|)^2 + \sum\sum_{i,j=-\infty}^{\infty}|\kappa(0,i,j)|$, where κ is the fourth cumulant. Then $P^{-1/2}\sum E|v_t H(t)| \le P^{-1/2}\sum\{E[v_t H(t)]^2\}^{1/2} \le cP^{-1/2}\sum t^{-1/2} = O(1)$ by Lemma A1(b) and the result follows from Markov's inequality. (c) By logic such as that in the proof of Lemma A4(b), $P^{-1/2}\sum|H(t)| = O_p(1)$. Then $|P^{-1/2}\sum[B(t)-B]H(t)| \le q[\sup_t|B(t)-B|][P^{-1/2}\sum|H(t)|] \xrightarrow{p} 0$ by Assumption 2.

PROOF OF EQUATION (4.1): A second order mean value expansion of $f_{t+\tau}(\hat{\beta}_t)$ around β^* for $t=R,\ldots,T$ yields $f_{t+\tau}(\hat{\beta}_t)=f_{t+\tau}+f_{t+\tau,\beta}(\hat{\beta}_t-\beta^*)+w_{t+\tau}$, where the ith element of $w_{t+\tau}$ is $w_{it+\tau}\equiv .5(\hat{\beta}_t-\beta^*)'[\partial^2 f_{it+\tau}(\tilde{\beta}_{it})/\partial\beta\partial\beta'](\hat{\beta}_t-\beta^*)$ and $\tilde{\beta}_{it}$ lies between $\hat{\beta}_t$ and β^*. Then $P^{1/2}(\bar{f}-Ef_t)=P^{-1/2}\sum(f_{t+\tau}-Ef_t)+P^{-1/2}\sum f_{t+\tau,\beta}(\hat{\beta}_t-\beta^*)+P^{-1/2}\sum w_{t+\tau}$ and (4.1) will follow if (a) $P^{-1/2}\sum f_{t+\tau,\beta}(\hat{\beta}_t-\beta^*)=FB[P^{-1/2}\sum H(t)]+o_p(1)$, and (b) $P^{-1/2}\sum w_{t+\tau}=o_p(1)$.

PROOF OF (a): We have

$$\begin{aligned} P^{-1/2}\sum f_{t+\tau,\beta}(\hat{\beta}_t-\beta^*) &\equiv P^{-1/2}\sum f_{t+\tau,\beta}B(t)H(t)\\ &= FB\Big[P^{-1/2}\sum H(t)\Big]+P^{-1/2}\sum(f_{t+\tau,\beta}-F)BH(t)\\ &\quad +P^{-1/2}F\sum[B(t)-B]H(t)\\ &\quad +P^{-1/2}\sum(f_{t+\tau,\beta}-F)[B(t)-B]H(t). \end{aligned}$$

The second term after the equality is $o_p(1)$ by Lemma A4(a), the third by Lemma A4(c), the fourth by Lemma A4(b).

PROOF OF (b): Choose a, $0<a<.5$ and define m_t as in Assumption 1. Then for the ith element of $w_{t+\tau}$,

$$\begin{aligned} &\Big|P^{-1/2}\sum w_{it+\tau}\Big|\\ &= .5\left|P^{-1/2}\sum\left\{[t^{1-a}(\hat{\beta}_t-\beta^*)']\left[\frac{\partial^2 f_{it+\tau}(\tilde{\beta}_{it})}{\partial\beta\partial\beta'}t^{-1+a}\right](\hat{\beta}_t-\beta^*)\right\}\right|\\ &\le .5k^2\sup_t|t^{.5-.5a}(\hat{\beta}_t-\beta^*)|^2P^{-1/2}\sum\{|\partial^2 f_{it+\tau}(\tilde{\beta}_{it})\partial\beta\partial\beta'|t^{-1+a}\}\\ &\le .5k^2\sup_t|t^{.5-.5a}(\hat{\beta}_t-\beta^*)|^2P^{-1/2}\sum m_{t+\tau}t^{-1+a}\xrightarrow{P}0, \end{aligned}$$

by Lemma A3(b), Lemma A1(a), Assumption 1, and Markov's inequality.

LEMMA A5: $\lim E[P^{-1}\sum H(t)\sum H(t)']=2[1-\pi^{-1}\ln(1+\pi)]S_{hh}\equiv 2\Pi S_{hh}$.

PROOF: Set $q=1$ for notational simplicity. Let

$$\gamma_j=\Gamma_{hh}(j),\quad a_{R,s}=(R+s)^{-1}+\cdots+(R+P-1)^{-1}\quad\text{for}\quad 0\le s\le P-1,$$

where the dependence of $a_{R,s}$ on P is suppressed for notational convenience. L'Hôpital's rule indicates that for $\pi=0$, $1-\pi^{-1}\ln(1+\pi)=0$. To establish the lemma in this case, proceed as follows: Since

$$\sum H(t)=a_{R,0}(h_1+\cdots+h_R)+a_{R,1}h_{R+1}+\cdots+a_{R,P-1}h_{R+P-1},$$

$$\operatorname{var}\Big(P^{-1/2}\sum H(t)\Big)\le\sum_{j=-\infty}^{\infty}|\gamma_j|P^{-1}(Ra_{R,0}^2+a_{R,1}^2+\cdots+a_{R,P-1}^2);$$

since $a_{R,j} \leq (P-j)/R$,

$$\begin{aligned} &P^{-1}[Ra_{R,0}^2 + a_{R,1}^2 + \cdots + a_{R,P-1}^2] \\ &\quad \leq (R/P)(P/R)^2 + (PR^2)^{-1}[1 + 2^2 + \cdots + (P-1)^2] \\ &\quad = (P/R) + (PR^2)^{-1}O(P^3) \\ &\quad = O(P/R) \to 0 \\ &\quad \Rightarrow \lim \operatorname{var}\left(P^{-1/2}\sum H(t)\right) = 0. \end{aligned}$$

For $\pi > 0$, write

$$\begin{aligned} &P^{-1/2}\sum H(t) = A_1 + A_2, \\ &A_1 = P^{-1/2}a_{R,0}(h_1 + \cdots + h_R), \\ &A_2 = P^{-1/2}[a_{R,1}h_{R+1} + \cdots + a_{R,P-1}h_{R+P-1}] \\ &\quad \Rightarrow \operatorname{var}\left[P^{-1/2}\sum H(t)\right] = \operatorname{var}(A_1) + \operatorname{var}(A_2) + 2\operatorname{cov}(A_1, A_2). \end{aligned}$$

It will be shown that when $\pi > 0$,

(A-1a) $\quad \lim \operatorname{var}(A_1) = \pi^{-1}\ln^2(1+\pi) \sum_{j=-\infty}^{\infty} \gamma_j,$

(A-1b) $\quad \lim \operatorname{var}(A_2) = \{2[1 - \pi^{-1}\ln(1+\pi)] - \pi^{-1}\ln^2(1+\pi)\} \sum_{j=-\infty}^{\infty} \gamma_j,$

(A-1c) $\quad \lim \operatorname{cov}(A_1, A_2) = 0.$

From the definition of A_1, $\operatorname{var}(A_1) = P^{-1}a_{R,0}^2 \sum_{j=-R+1}^{R-1}(R-|j|)\gamma_j = (R/P)a_{R,0}^2 \sum_{j=-R+1}^{R-1}\gamma_j - P^{-1}a_{R,0}^2 \sum_{j=-R+1}^{R-1}|j|\gamma_j$. In light of Lemma A2(b), (A-1a) will follow if $(R/P)a_{R,0}^2 \to \pi^{-1}\ln^2(1+\pi)$. We have

$$\begin{aligned} \ln[(R+P-1)/R] = \int_0^{P-1}(R+x)^{-1}\,dx \leq a_{R,0} &\leq \int_{-1}^{P-1}(R+x)^{-1}\,dx \\ &= \ln[(R+P-1)/(R-1)] \\ \Rightarrow (R/P)\ln^2[(R+P-1)/R] \leq (R/P)&a_{R,0}^2 \leq (R/P)\ln^2[(R+P-1)/(R-1)] \\ \Rightarrow (R/P)a_{R,0}^2 \to \pi^{-1}&\ln^2(1+\pi), \end{aligned}$$

where the result for $\pi = \infty$ follows since L'Hôpital's rule indicates that $x^{-1}\ln^2(1+x) \to 0$ as $x \to \infty$.

Now consider (A-1b). Let

$$\begin{aligned} &d_j = a_{R,1}a_{R,j+1} + \cdots + a_{R,P-j-1}a_{R,P-1} \quad \text{for} \quad 0 \leq j \leq P-2; \\ &d_j \equiv d_{-j} \quad \text{for} \quad -P+2 \leq j < 0; \\ &\quad \Rightarrow \operatorname{var}(A_2) = P^{-1}\sum_{j=-P+2}^{P-2} d_j\gamma_j = P^{-1}d_0 \sum_{j=-P+2}^{P-2}\gamma_j + P^{-1}\sum_{j=-P+2}^{P-2}(d_j - d_0)\gamma_j, \end{aligned}$$

where the dependence of d_j on P and R is suppressed for notational simplicity. Using logic such as that in the proof of (A-1a), we have $P^{-1}d_0 = P^{-1}\int_1^{P-1}\ln^2[(R+P-1)/(R+y)]\,dy + o(1)$, where "$o(1)$" denotes a sequence whose limit is zero. Routine manipulations then yield

$$\begin{aligned} P^{-1}d_0 = P^{-1}\{&2(P-2) - 2(R+1)\ln[(R+P-1)/(R+1)] \\ &-(R+1)\ln^2[(R+P-1)/(R+1)]\} + o(1) \\ \Rightarrow P^{-1}d_0 &\to 2[1 - \pi^{-1}\ln(1+\pi)] - \pi^{-1}\ln^2(1+\pi) \end{aligned}$$

again using L'Hôpital's rule for $\pi = \infty$.

So (A-1b) will follow if $P^{-1}\sum_{j=-P+2}^{P-2}(d_j-d_0)\gamma_j \to 0$. To show this, let $m=\min\{(P/2)-1, R\}$. Write $P^{-1}\sum_{j=-P+2}^{P-2}(d_j-d_0)\gamma_j = P^{-1}\sum_{|j|\le m}(d_j-d_0)\gamma_j + P^{-1}\sum_{|j|>m}(d_j-d_0)\gamma_j$. By the Cauchy-Schwarz inequality, $0\le d_j \le d_0 \Rightarrow |d_j-d_0|\le d_0$, so $P^{-1}\sum_{|j|>m}(d_j-d_0)\gamma_j\to 0$ since $P^{-1}d_0$ is bounded, $m\to\infty$, and $\sum_{j=-\infty}^{\infty}|\gamma_j|<\infty$. To analyze $P^{-1}\sum_{|j|\le m}(d_j-d_0)\gamma_j$, note that for $0\le j\le m$, we have

$$\begin{aligned} d_0-d_j &= .5\left[(a_{R,j+1}-a_{R,1})^2+\cdots+(a_{R,P-1}-a_{R,P-1-j})^2\right] \\ &\quad + .5[a_{R,1}^2+\cdots+a_{R,j}^2+a_{R,P-j}^2+\cdots+a_{R,P-1}^2] \\ &\le .5(P-j-1)[j/(R+1)]^2 + a_{R,0}^2 j \le .5P[j/(R+1)] + a_{R,0}^2 j, \end{aligned}$$

and $P^{-1}\sum_{|j|\le m}(d_j-d_0)\gamma_j\to 0$ now follows from Lemmas A2(a) and A2(b).

Finally, consider (A-1c). Since $a_{R,j}\le a_{R,0}$, a direct calculation yields $\mathrm{cov}(A_1, A_2)\le P^{-1}a_{R,0}^2\sum_{|j|\le R+P-2}|j||\gamma_j| \le (P^{-1/2}a_{R,0})^2\sum_{j=-\infty}^{\infty}|j||\gamma_j|\to 0$ by Lemma A2(b).

LEMMA A6: $\lim E[P^{-1}\sum(f_{t+\tau}-Ef_t)\sum H(t)'] = [1-\pi^{-1}\ln(1+\pi)]\sum_{j=-\infty}^{\infty}\Gamma_{fh}(j)\equiv \Pi S_{fh}$.

PROOF: Follows from an argument similar to that for Lemma A5 (West (1994)).

PROOF OF LEMMA 4.1: That the limiting variance of $P^{-1/2}[\sum(f_{t+\tau}-Ef_t)', \sum H(t)']'$ is as indicated follows from Lemmas A5 and A6. That $V(\pi)$ is of full rank when $\pi\neq 0$ and S is p.d. is straightforward to establish. Asymptotic normality follows from McLeish (1977) as shown in Theorem 3.1 of Wooldridge and White (1989).

PROOF OF THEOREM 4.1: Follows from equation (4.1) and Lemma 4.1.

REFERENCES

ANDREWS, D. W. K. (1987): "Consistency in Nonlinear Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 56, 1465–1471.

——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

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

CHRISTIANO, L. J. (1989): "*P**: Not the Inflation Forecaster's Holy Grail," *Federal Reserve Bank of Minneapolis Quarterly Review*, 13, 3–18.

DIEBOLD, F. X., AND R. S. MARIANO (1994): "Comparing Predictive Accuracy," NBER Technical Working Paper No. 169.

DIEBOLD, F. X., AND G. D. RUDEBUSCH (1991): "Forecasting Output with The Composite Leading Index: A Real-Time Analysis," *Journal of the American Statistical Association*, 86, 603–610.

DURBIN, J. (1970): "Testing for Serial Correlation in Least Squares Regression Models When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410–421.

FAIR, R. C. (1980): "Evaluating the Predictive Accuracy of Econometric Models," *International Economic Review*, 21, 355–378.

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

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press.

HAMILTON, J. S. (1994): *Time Series Analysis*. Princeton: Princeton University Press.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

IBRAGIMOV, I. A., AND Y. V. LINNIK (1971): *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

McCulloch, R., and P. E. Rossi (1990): "Posterior, Predictive and Utility Based Approaches to Testing the Arbitrage Pricing Theory," *Journal of Financial Economics*, 28, 7–38.

McLeish, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 3, 829–839.

——— (1977): "On the Invariance Principle for Nonstationary Mixingales," *Annals of Statistics*, 5, 616–621.

Meese, R. A. (1986): "Testing for Bubbles in Exchange Markets: A Case of Sparkling Rates?" *Journal of Political Economy*, 94, 345–373.

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

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

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

Nelson, C. R., and Richard Startz (1991): "The Distribution of the Instrumental Variables Estimator and its *t*-ratio When the Instrument is a Poor One," *Journal of Business*, 63, 967–976.

Newey, W. K., and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–654.

Oliner, S., G. D. Rudebusch, and D. Sichel (1993): "New and Old Models of Business Investment: A Comparison of Forecasting Performance," Federal Reserve Board of Governors Economic Activity Working Paper No. 141.

Schmidt, P. (1977): "Some Small Sample Evidence on the Distribution of Dynamic Simulation Forecasts," *Econometrica*, 45, 997–1005.

Stock, J. H., and M. W. Watson (1993): "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience," in *Business Cycles, Indicators and Forecasting*, ed. by J. H. Stock and M. W. Watson. Chicago: University of Chicago Press, 95–153.

West, K. D. (1994): "Asymptotic Inference About Predictive Ability," manuscript, University of Wisconsin.

West, K. D., and D. Cho (1994): "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, 69, 367–391.

West, K. D., H. J. Edison, and D. Cho (1993): "A Utility Based Evaluation of Some Models of Exchange Rate Variability," *Journal of International Economics*, 35, 23–46.

Wooldridge, J. M., and H. White (1989): "Central Limit Theorems for Dependent, Heterogeneous Processes with Trending Moments," Manuscript, Michigan State University.