# REGRESSION-BASED TESTS OF PREDICTIVE ABILITY

Kenneth D. West

Michael W. McCracken

July 1995
Last revised January 1998

We thank Frank Diebold, Neil Ericcson, Art Goldberger, Matthew Higgens, Yuichi Kitamura, three anonymous referees and seminar participants at the 1996 Midwest Economics Association Meetings, the Federal Reserve Board of Governors, the University of Maryland and the University of Wisconsin for helpful comments, and the National Science Foundation and the Graduate School of the University of Wisconsin for financial support.

**ABSTRACT**

We develop regression-based tests of hypotheses about out of sample prediction errors. Representative tests include ones for zero mean and zero correlation 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 error introduced by estimation of these parameters. We propose computationally convenient test statistics that properly account for such error. Simulations indicate that the procedures can work well in samples of size typically available, although there sometimes are substantial size distortions.

Kenneth D. West
Department of Economics
7458 Social Science Building
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
and NBER
kdwest@facstaff.wisc.edu

Michael W. McCracken
Department of Economics
6473A Social Science Building
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
mwmccrac@students.wisc.edu

## 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 (1982) competition.

If the predictions do not depend on estimated parameters, it follows from Diebold and Mariano (1996) 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 (1996) 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 (1996) procedure. More importantly, we show that such uncertainty sometimes is asymptotically non-negligible, 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)), Pagan and Schwert (1990)), and, we hope, may appear in still more papers upon development of techniques such as those proposed here.[1] 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 sections 3-5 to consider some common tests, when the underlying models are linear and exactly identified. Section 8 presents simulation evidence. Section 9 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 (kx1) unknown parameter vector $\beta^*$. 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 $T+\tau$ 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 $\beta^*$, 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 $\beta^*$ first with data from 1 to R, next with data from 1 to R+1, . . . , 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, $\beta^*$ is estimated first with data from 1 to R, next with data from 2 to R+1, . . . , 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 $\beta^*$ 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,... ,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, $\hat{\beta}_t=(\Sigma_{s=1}^{t}X_sX_s')^{-1}\Sigma_{s=1}^{t}X_sy_s$,
data from t-R+1 to t in the rolling scheme, $\hat{\beta}_t=(\Sigma_{s=t-R+1}^{t}X_sX_s')^{-1}\Sigma_{s=t-R+1}^{t}X_sy_s$,
data from 1 to R in the fixed scheme, $\hat{\beta}_t=(\Sigma_{s=1}^{R}X_sX_s')^{-1}\Sigma_{s=1}^{R}X_sy_s$.

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 fixed scheme a different regression estimate is used for each t. As well, 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, $P,R \rightarrow \infty$), with $\tau$ 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 ($\ell$x1) 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 $\ell$ is small, say $\ell=1$ or $\ell=2$. 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 $Ev_{t+\tau}g_{t+1}=0$. The obvious regression is one of $\hat{v}_{t,t+\tau}$ on $\hat{g}_{t+1}$ for $t=R, \ldots, R+P-1$, obtaining

$$(2.2) \quad \hat{v}_{t,t+\tau} = \hat{g}_{t+1}'\hat{\alpha} + \hat{\eta}_{t+\tau}, \quad \hat{\alpha} \equiv (\Sigma_{t=R}^{T}\hat{g}_{t+1}\hat{g}_{t+1}')^{-1}\Sigma_{t=R}^{T}\hat{g}_{t+1}\hat{v}_{t,t+\tau}, \quad \hat{\eta}_{t+\tau} \equiv \hat{v}_{t,t+\tau}-\hat{g}_{t+1}'\hat{\alpha}.$$

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(l) model $y_t=\beta^* y_{t-1}+u_t$, $|\beta^*|<1$.

1. <u>Mean prediction error</u>. Here, $g_{t+1}\equiv 1$ is a scalar. If $v_{t+\tau}$ is a $\tau$ 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. <u>Efficiency</u>. 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, 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(l) 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 (2x1) 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. Encomoassinq. 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 $\alpha$ are (2x1) and the null is $\alpha=(0,0)'$.

4. Serial correlation. If $v_{t+1}$ is the 1 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 $\alpha$ is a scalar, $\hat{\alpha}$ an estimate of the first order serial correlation coefficient, and $H_0$ is $\alpha=0$.[2]

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 iid 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 inference about $\alpha$. The first is uncertainty that would be present even if (counterfactually) $\beta^*$ were known and one could regress $v_{t+1}$ on $g_{t+1}$. The second results from use of $\hat{\beta}_t$ rather than the unknown $\beta^*$. 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) $\hat{v}_{t,t+\tau} = \hat{g}_{t+1}'\alpha + \hat{g}_{2t+1}'\alpha_2$ + disturbance $\equiv \hat{\hat{g}}_{t+1}'\tilde{\alpha}$ + disturbance,

where $\hat{g}_{2t+1}$ is a (rx1) set of extra variables included so that conventionally computed hypothesis tests on $\alpha$ are correctly sized accordingly to our asymptotic theory; $\hat{\hat{g}} \equiv (\hat{g}_{t+1}', \hat{g}_{2t+1}')'$ is $(\ell+r)$x1; $g_{t+1} \equiv \tilde{g}_{t+1}(\beta^*) \equiv (g_{t+1}(\beta^*)', g_{2t+1}(\beta^*)')' \equiv (g_{t+1}', g_{2t+1}')'$ and $\tilde{\alpha}$ are also $(\ell+r)$ x1.

## 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 the ones we consider in detail in sections 6 and 7 below. Second, 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 $n_t$, $\frac{\partial n_t}{\partial x}$ denotes the (sxm) matrix of partial derivatives of $n_t$; for any function $n_t$ whose domain is in $\mathbf{R}^k$, $n_{t\beta} \equiv \frac{\partial n_t}{\partial \beta}(\beta^*)$; 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≡R+P-1: for any variable x, $\Sigma x(t) \equiv \Sigma_{t=R}^{T} x(t)$, $\Sigma x_{t+\tau} \equiv \Sigma_{t=R}^{T} x_{t+\tau}$; 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, (c) 1 to R, for the fixed scheme: for any variable x, (a) $\Sigma x_s \equiv \Sigma_{s=1}^{t} x_s$ (recursive), (b) $\Sigma x_s \equiv \Sigma_{s=t-R+1}^{t} x_s$ (rolling), (c) $\Sigma_s x_s \equiv \Sigma_{s=1}^{R} x_s$ (fixed). Finally, let

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

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

<u>Assumption 1:</u> (a) In some neighborhood N around $\beta^*$, and with probability 1, $v_t(\beta)$

and $g_t(\beta)$ are measurable and twice continuously differentiable; (b) $Ev_{t+\tau}g_{t+1} = 0$; (c) $Ev_t v_{t\beta} = 0$; (d) $Ev_{t+\tau}g_{t+1,\beta} = 0$; (e) $Eg_t g_t'$ is of rank $\ell$.

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) $B(t) \overset{a.s.}{\rightarrow} B$, B a matrix of rank k; (b) $H(t)=t^{-1}\Sigma_s h_s(\beta^*)$ (recursive) or $H(t)=R^{-1}\Sigma_s h_s(\beta^*)$ (rolling or fixed) for a $(q\times 1)$ orthogonality condition $h_s(\beta^*)$; (c) $Eh_s(\beta^*)=0$; (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, $\sup_{\beta\in N} |\partial v_t(\beta)/\partial\beta\partial\beta'| < m_t$ for a measurable $m_t$ for which $Em_t^4<D$. The same holds when $v_t$ is replaced by an arbitrary element of $g_t$.

Assumption 4: Let $w_t \equiv (v_{t\beta}', vec(g_{t\beta})', v_t, g_t', h_t')'$. (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 $\Gamma_{ff}(j)=Ef_t f_{t-j}'$, $S_{ff}=\Sigma_{j=-\infty}^{\infty}\Gamma_{ff}(j)$. Then $S_{ff}$ is p.d..

Assumption 5: $R,P \rightarrow \infty$ as $T\rightarrow\infty$, and $\lim_{T\rightarrow\infty} \frac{P}{R} = \pi$, (a) $0\le\pi\le\infty$ for recursive ($\pi=\infty$ <==> $\lim_{T\rightarrow\infty} \frac{R}{P} = 0$), (b) $0\le\pi<\infty$ for rolling and fixed.

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

$$Ef_t=0,\quad F=Eg_{t+1}\left(\frac{\partial v_{t+\tau}}{\partial\beta}\right).$$

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 $\tau$.

Throughout, we maintain assumptions 1-5.

## 4. Basic Asymptotic Results

Let

(4.1) $\Gamma_{fh}(j) = Ef_t h_{t-j}'$, $S_{fh}=\Sigma_{j=-\infty}^{\infty}\Gamma_{fh}(j)$, $\Gamma_{hh}(j) = Eh_t h_{t-j}'$, $S_{hh}=\Sigma_{j=-\infty}^{\infty}\Gamma_{hh}(j)$, $V_\beta = BS_{hh}B'$.

$V_\beta$ 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 $\pi \equiv \lim_{T\to\infty} \frac{P}{R}$, as follows:

(4.2)

| Sampling scheme | $\lambda_{fh}$ | $\lambda_{hh}$ | $\lambda$ |
|---|---|---|---|
| recursive | $1-\pi^{-1}\ln(1+\pi)$ | $2[1-\pi^{-1}\ln(1+\pi)]$ | 1 |
| rolling, $\pi\le1$ | $\frac{\pi}{2}$ | $\pi-\frac{\pi^2}{3}$ | $1-\frac{\pi^2}{3}$ |
| rolling, $\pi>1$ | $1-\frac{1}{2\pi}$ | $1-\frac{1}{3\pi}$ | $\frac{2}{3\pi}$ |
| fixed | 0 | $\pi$ | $1+\pi$ |

Lemma 4.1: (a) $P^{-1/2}\Sigma\hat{g}_{t+1}\hat{v}_{t,t+\tau} = P^{-1/2}\Sigma g_{t+1}v_{t+\tau} + FB[P^{-1/2}\Sigma H(t)] + o_p(1)$.

(b) $P^{-1/2}\Sigma g_{t+1}v_{t+\tau} \sim_A N(0,S_{ff})$.

(c) $E[P^{-1}\Sigma H(t)\Sigma H(t)'] \to \lambda_{hh}S_{hh}$, $E[P^{-1}\Sigma g_{t+1}v_{t+\tau}\Sigma H(t)'] \to \lambda_{fh}S_{fh}$.

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: $P^{-1/2}\Sigma\hat{g}_{t+1}\hat{v}_{t,t+\tau} \sim_A N(0,\Omega)$, where $\Omega$ is the $(\ell x\ell)$ matrix

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

Lemma 4.3: $P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}' \to_p Eg_t g_t'$.

Theorem 4.1: Let $\hat{\alpha}$ be the least squares estimator of $\alpha$ (=0). Then $P^{1/2}\hat{\alpha} \sim_A N(0,V)$, $V \equiv (Eg_t g_t')^{-1}\Omega(Eg_t g_t')^{-1}$.

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 $(\ell x\ell)$ j'th sample autocovariance of $\hat{g}_{t+1}\hat{\eta}_{t+\tau}$:

(4.4) $\hat{\sigma}^2 \equiv (P-\ell)^{-1}\Sigma\hat{\eta}^2_{t+\tau} \equiv (P-\ell)^{-1}\Sigma(\hat{v}_{t,t+\tau}-\hat{g}_{t+1}'\hat{\alpha})^2$,

$\hat{\Gamma}_{ff}(j) \equiv P^{-1}\Sigma_{t=R+j}^{T}[(\hat{g}_{t+1}\hat{\eta}_{t+\tau})(\hat{g}_{t+1-j}\hat{\eta}_{t+\tau-j})']$ for $j\ge0$,

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

<u>**Theorem 4.2**</u>: (a) $\hat{\sigma}^2 \to_p \sigma^2 \equiv Ev_t^2$, $\hat{\sigma}^2(P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}')^{-1} \to_p \sigma^2(Eg_tg_t')^{-1}$,
(b) $\hat{\Gamma}_{ff}(j) \to_p$ **r,,(j)**, $(P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}')^{-1}\hat{\Gamma}_{ff}(0)(P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}')^{-1} \to_p$ $(Eg_tg_t')^{-1}\Gamma_{ff}(0)(Eg_tg_t')^{-1}$.

**(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** $\frac{M}{P^a}\to 0$ **and, if** $\pi=\infty$, $\frac{M}{R^a}\to 0$. **Then** $\hat{S}_{ff} \equiv \Sigma_{j=-P+1}^{P-1}K(j/M)\hat{\Gamma}_{ff}(j) \to_p S_{ff}$, **and** $(P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}')^{-1}\hat{S}_{ff}(P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}')^{-1} \to_p$ $(Eg_tg_t')^{-1}S_{ff}(Eg_tg_t')^{-1}$.

**Note that Theorem 4.2 assumes that the least squares residual** $\hat{\eta}_{t+\tau}$ **is used in estimating** $\hat{\sigma}^2$ **and i?,,(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, 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** $\hat{S}_{ff} = \hat{\Gamma}_{ff}(0)+\Sigma_{j=1}^{M}[(1-\frac{j}{M})[\hat{\Gamma}_{ff}(j)+\hat{\Gamma}_{ff}(j)']$ **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 $E(v_{t+1}|g_{t+1},v_t,g_t,v_{t-1},...)=0$, $E(v_{t+1}^2g_{t+1}g_{t+1}') = Ev_{t+1}^2Eg_{t+1}g_{t+1}' = S_{ff}$. **The least squares estimator of the regression covariance matrix is** $\hat{\sigma}^2(P^{-1}\Sigma\hat{g}_{t+1}\hat{g}_{t+1}')^{-1}$. **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 $\beta^*$ is known, and one regresses $v_{t+\tau}(\beta^*)$ on $g_{t+1}(\beta^*)$. But since $\beta^*$ 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 $\beta^*$. 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 $\beta^*$: 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 $\beta^*$ is known. Two simple conditions are sufficient to imply $S_{ff}=\Omega$. One is $F \equiv E\frac{\partial f_t}{\partial\beta}(\beta^*) \equiv E[g_{t+1}(\beta^*)\frac{\partial v_{t+\tau}}{\partial\beta}(\beta^*)] = 0$. This is essentially a condition that there is block diagonality in the asymptotic variance-covariance matrix for the estimators of $\beta^*$ and $Ef_{t+\tau}\equiv Eg_{t+1}v_{t+\tau}$. This conditions 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.'

A second condition sufficient for $\Omega=S_{ff}$ is $\pi\equiv\lim_{T\to\infty}\frac{P}{R} = 0$, 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 $\beta^*$ 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 .25, and that was not sufficiently small to always make it harmless to ignore error in estimation of $\beta^*$.

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

## 5. Obtaininq Asymptotically Valid Test Statistics

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

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

For the recursive scheme, $\hat{\lambda}=1$ for all $\pi$, for the fixed scheme $\hat{\lambda}=1+\frac{P}{R}$, and so on. Clearly, $\hat{\lambda} \rightarrow \lambda$.

<u>Corollary 5.1</u>: Suppose that

$$(5.1)\quad S_{ff} = -\frac{1}{2}(FBS_{fh}' + S_{fh}B'F') = FV_{\beta}F' .$$

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

Condition (5.1) implies that $\Omega$ (defined in (4.3)) is equal to $\lambda S_{ff}$, 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 $\beta^*$ simply introduces a factor of $\lambda$ into the asymptotic variance of $P^{1/2}\hat{\alpha}$. For the recursive sampling scheme, $\lambda \equiv 1$, so error in estimation of $\beta^*$ 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) $\beta^*$ 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, decreases them for the rolling scheme.

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

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

Theorem 5.1: Let $\tilde{g}_{t+1}(\beta^*) = (g_{t+1}', g_{2t+1}')'$ for a (rxl) vector $g_{2t+1}$ defined as either (a) $g_{2t+1} = \frac{\partial v_{t+\tau}}{\partial \beta}(\beta^*)$ (==>r=k) or (b) $g_{2t+1}=Z_{t+1}$ for a vector of variables $Z_{t+1}$ that satisfies $\frac{\partial v_{t+\tau}}{\partial \beta}(\beta^*) = G_2(\beta^*)Z_{t+1}$, $G_2(\beta^*)$ a (kxr) 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 $f_{t+\tau}$ 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 $f_{t+\tau}$ replacing $f_{t+\tau}$. Let $\Omega = S_{\tilde{f}\tilde{f}} + \lambda_{fh}(\tilde{F}BS_{\tilde{f}h}' + S_{\tilde{f}h}B'F') + \lambda_{hh}\tilde{F}V_\beta\tilde{F}'$. 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 $\ell$ elements of $\hat{\tilde{\alpha}}$. Then $P^{1/2}\hat{\alpha} \sim_A N(0,V)$, V the $(\ell x \ell)$ 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 $\ell x \ell$ 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 l(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 $\alpha_2$, 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) $y_t = x_t'\beta^* + v_t$,

where $y_t$ and $v_t$ are scalars, $x_t$ and $\beta^*$ are (kx1). The sample counterpart of $v_{t+1}$ is computed as

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

For the encompassing test, we need to describe as well the encompassed model. This will require redefining $\beta^*$. Model "1" is the encompassing model, "2" the encompassed model. Let $\beta^*=(\beta_1^{*\prime},\beta_2^{*\prime})'$, where $\beta_i^*$ is $(k_i x1)$, $k\equiv k_1+k_2$, with the model i prediction dependent only on $\beta_i^*$. Let $x_{2t}$ be the vector of predetermined 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-no, $R\to\infty$), 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 test, $h_s=(x_s'v_s,x_{2s}'v_{2s})'$). The estimate $\hat{\beta}_t$ satisfies $\hat{\beta}_t-\beta^* = B(t)H(t)$, where B(t) is (kxk) and H(t) is (kx1), with B(t) and H(t) defined as follows. (i) $B(t) = (t^{-1}\Sigma_s x_s x_s')^{-1}$ (recursive), $B(t)=(R^{-1}\Sigma_s x_s x_s')^{-1}$ (rolling or fixed). For the encompassing test, B(t) is block diagonal with analogously defined $B_i(t)$ on the diagonals. (ii) $H(t)=t^{-1}\Sigma_s h_s(\beta^*)$ (recursive) or $H(t)=R^{-1}\Sigma_s h_s(\beta^*)$ (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, etc. 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 not lie in the linear span of the regressors from the encompassing model.[4]

## 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.

<u>Theorem 7.1</u>: (a) For $g_t$ defined as in one of the rows of Table 1, let $\hat{\alpha} = (\Sigma \hat{g}^2_{t+1})^{-1}(\Sigma \hat{g}_{t+1}\hat{v}_{t+1})$.

(i) For mean prediction error or efficiency, $P^{1/2}\hat{\alpha} \sim_A N(0,V)$, $V=\lambda(Eg^2_t)^{-2}Ev^2_t g^2_t$.

(ii) Let the sampling scheme be recursive, and suppose that the underlying disturbance $v_t$ is conditionally homoskedastic, $E(v^2_t|x_t) = Ev^2_t$ (for encompassing, assume $E(v^2_t|x_t,x_{2t})=Ev^2_t$ and $E(v^2_{2t}|x_t,x_{2t})=Ev^2_{2t}$). Then for any one of the four tests in the table, $P^{1/2}\hat{\alpha} \sim_A N(0,V)$, $V = \sigma^2(Eg^2_t)^{-1}$, $\sigma^2 \equiv Ev^2_t$.

(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 $(E\tilde{g}_t\tilde{g}_t')^{-1}Ev^2_t\tilde{g}_t\tilde{g}_t'(E\tilde{g}_t\tilde{g}_t')^{-1}$.

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 mis-sized 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 $\lambda Ev^2_t$. For encompassing and serial correlation, under conditional homoskedasticity the formula for V in part(b) reduces to $Ev^2_t(E\tilde{g}_t\tilde{g}_t')^{-1}$.

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 $\lambda\Sigma_{j=-\tau+1}^{\tau-1} E v_t v_{t-j}$ 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 $\lambda$.

7. Suppose that $\beta^*$ 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 non-linear least squares or an asymptotically equivalent technique. Then condition (5.1) (which underlies Theorem 7.1(a)) continues to hold for mean prediction error. So under suitable conditions the result in Theorem 7.1(a) will continue to hold as well.[5]

## 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 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 and 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 (Px1) vector of prediction errors used on the left hand side of the regression tests was $\{\hat{v}_{t+1}\}$, t=R, . . . ..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 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) $y_t = \beta^* y_{t-1} + v_t,\ \beta^*=0.5,\ v_t \sim N(0,1)$ .

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 $y_t = \beta y_{t-2} + v_{2t}$. This was estimated by least squares, so $\beta \equiv (Ey_{t-2}^2)^{-1} Ey_{t-2}y_t$. The introduction of the second lag meant that some regression samples were 1 observation smaller than the "R" reported in the table.

We report tests of nominal size .05. Tests of nominal size .01 and .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 3A presents results for mean prediction error. Tests for the recursive scheme work quite well, with nominal .05 tests having actual sizes between .046 and .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 $P/R \leq .5$.

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 $\beta^*$ 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 .058, the unadjusted a size of .081), but it is dramatic for large P/R (for P=175, R=25, our test statistic has a size of .099 vs. .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 > .05) while the unadjusted typically rejects too often (actual size < .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 .05 test have a probability of rejecting of say 7.4 percent (P=50, R=25, our test statistic) than of .3 percent (unadjusted test statistic), all other things 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 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%, while panel B1 indicates that use of the usual least squares test statistic yielded a size of 7.2%.

Tables 5 and 6 indicate that for the encompassing test and the test for zero first order serial correlation, the Table 4 results apply qualitatitively: 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, 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 reasonably accurate (actual size of .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 mis-sizing 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.

Footnotes

1. 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.

2. 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 $\beta^*$, and a rank condition presented below rules out simply defining parameters so that the population parameter of interest is 2x1 with a 2x1 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.

3. See West (1996) and McCracken (1997) for further discussion of the conditions under which $F=0$.

4. Note that this last condition rules out tests of nested (rather than non-nested) 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 non-nested models; to conserve space, we do not consider such models here.

5. It is, however, possible to construct examples in which the results of Theorem 7.1 fail. Let $\epsilon_t$ 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 $Ex_t^2 v_t^2 \neq Ex_t^2 Ev_t^2$.

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; the summation "$\Sigma$" means "$\Sigma_{t=R}^{T}$"; For notational simplicity, we consider throughout the case in which k=1 and $\ell$=1, so that $\beta^*$, $g_{t+1}$ and $f_{t+\tau}$ are scalars, and we let "$f_{t+\tau,\beta\beta}(\tilde{\beta}_t)$" mean "$\frac{\partial^2 f}{\partial\beta^2}{}_{t+\tau}(\tilde{\beta}_t)$". 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) $\sup_t|P^aH(t)|\rightarrow_p 0$; (b) $\sup_t|P^a(\hat{\beta}_t-\beta^*)|\rightarrow_p 0$.
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,...,T, $P^{-1}\Sigma|f_{t+\tau,\beta\beta}(\tilde{\beta}_t)|^2 = O_p(1)$.
Lemma A3: Let $\mathring{\Gamma}_{ff}(j) \equiv P^{-1}\Sigma_{t=R+j}^{T}[(\hat{g}_{t+1}\hat{v}_{t,t+\tau})(\hat{g}_{t+1-j}\hat{v}_{t-j,t+\tau-j})] \equiv P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}(\hat{\beta}_t)f_{t+\tau-j}(\hat{\beta}_{t-j})$. Then $\mathring{\Gamma}_{ff}(j)\rightarrow_p\Gamma_{ff}(j)$.
Lemma A4: Under the assumptions of Theorem 4.2, and with $\mathring{\Gamma}_{ff}(j)$ defined as in Lemma A3, $\mathring{S}_{ff} \equiv \Sigma_{j=-P+1}^{P-1}K(j/M)\mathring{\Gamma}_{ff}(j)\rightarrow_p S_{ff}$.
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, . . ..R+P-1. and sum the results, yielding

$$\text{(A2)}\quad P^{-1/2}\Sigma f_{t+\tau}(\hat{\beta}_t) = P^{-1/2}\Sigma f_{t+\tau} + P^{-1/2}\Sigma f_{t+\tau,\beta}B(t)H(t) + P^{-1/2}\Sigma w_{t+\tau},$$

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

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

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 $\gamma_j \equiv Ex_th_{t-j}$. For P $\le$ R (the P > R case is similar) we have $|EP^{-1/2}\Sigma x_tH(t)| =$

$(P^{1/2}/R)|\gamma_0+...+\gamma_{R-1}| \le (P/R)^{1/2}R^{-1/2}\Sigma_{j=0}^{\infty}|\gamma_j| \to 0$ 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 $\mathrm{var}[P^{-1/2}\Sigma x_t H(t)] = 0$, 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 $|ER^{-1}\Sigma f_{t+\tau}(\Sigma_{s=1}^{R}h_s')| =$ $|R^{-1}[(\gamma_{R-1}+...+\gamma_0)+...+(\gamma_{R+P-2}+...+\gamma_{P-1})]| \le R^{-1}\Sigma_{j=-\infty}^{\infty}|j||\gamma_j| \to 0$ since assumption 4 implies $\Sigma_{j=-\infty}^{\infty}|j||\gamma_j|<\infty$ (Andrews (1991)).

For the rolling scheme, we will sketch the result that $E[P^{-1}\Sigma H(t)\Sigma H(t)'] \to \lambda_{hh}S_{hh} \equiv (\pi-\frac{\pi^2}{3})S_{hh}$ for $\pi<1$. The proofs for $\pi\ge1$, 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, $\Sigma H(t) = A_1 + A_2 + A_3$, $A_1 \equiv R^{-1}[h_1+2h_2+...+(P-1)h_{P-1}]$, $A_2 \equiv PR^{-1}[h_P+...+h_R]$, $A_3 \equiv R^{-1}[(P-1)h_{R+1}+...+2h_{R+P-2}+h_{R+P-1}]$. It is easy to see that lim $\mathrm{var}(P^{-1/2}A_2) =$ lim $P(R-P+1)R^{-2}\Sigma_{|j|\le R-P+1}Eh_th_{t-j}' + o(1) \to (\pi-\pi^2)S_{hh}$. We will sketch the argument that shows lim $\mathrm{var}(P^{-1/2}A_1) = \frac{\pi^2}{3}S_{hh}$. That lim $\mathrm{var}(P^{-1/2}A_3) = \frac{\pi^2}{3}S_{hh}$ follows from a nearly identical argument. Since, finally, it can be shown that lim $\mathrm{cov}(P^{-1/2}A_i,P^{-1/2}A_j) = 0$ for $i\ne j$, the result will follow.

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

$$\mathrm{var}(A_1) = R^{-2}\Sigma_{j=-P+2}^{P-2}d_j\gamma_j = R^{-2}d_0\Sigma\gamma_j - R^{-2}\Sigma(d_0-d_j)\gamma_j,$$

We have $P^{-1}R^{-2}d_0 \sim [P^3/(3PR^2)] \to \frac{\pi^2}{3}$, 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 $d_0 \le \int_0^P x^2dx$, $d_j \ge \int_{j+1}^{P-1}(x-j)xdx \implies |d_0-d_j| \le |\int_0^P x^2dx-\int_{j+1}^{P-1}(x-j)xdx|$, 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 $\mathrm{var}[P^{-1/2}X(T)] = \Omega$. 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.

<u>Proof of Theorem 4.1</u>: Follows immediately from Lemmas 4.1, 4.2 and 4.3.

<u>Proof of Theorem 4.2</u>: <u>(a)</u> 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

$$(P-1)^{-1}\Sigma(\hat{v}_{t,t+\tau}-\hat{g}_{t+1}\hat{\alpha})^2 = (P-1)^{-1}\Sigma\hat{v}_{t,t+\tau}^2 + [(P-1)^{-1}\Sigma\hat{g}_{t+1}^2]\hat{\alpha}^2-2\hat{\alpha}[(P-1)^{-1}\Sigma\hat{g}_{t+1}\hat{v}_{t,t+\tau}].$$

That $(P-1)^{-1}\Sigma\hat{v}_{t,t+\tau}^2 \to_p Ev_{t+\tau}^2$ follows from Lemma A3. By Theorem 4.1, $\hat{\alpha} = o_p(1)$; by Lemmas 4.2 and 4.3, $(P-1)^{-1}\Sigma\hat{g}_{t+1}^2 = O_p(1)$, $(P-1)^{-1}\Sigma\hat{g}_{t+1}\hat{v}_{t,t+\tau} = O_p(1)$. The desired result now follows.

<u>(b)</u> 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 $\hat{\Gamma}_{ff}(j) \equiv P^{-1}\Sigma_{t=R+j}^T\hat{g}_{t+1}\hat{g}_{t+1-j}\hat{\eta}_{t+\tau}\hat{\eta}_{t+\tau-j} \to_p Eg_{t+1}g_{t+1-j}v_{t+\tau}v_{t+\tau-j} \equiv \Gamma_{ff}(j)$ for all j. For $\mathring{\Gamma}_{ff}(j)$ defined in Lemma A3, we have $\hat{\Gamma}_{ff}(j) = \mathring{\Gamma}_{ff}(j) +$ $P^{-1}\Sigma_{t=R+j}^T\hat{g}_{t+1}\hat{g}_{t+1-j}(\hat{\eta}_{t+\tau}\hat{\eta}_{t+\tau-j}-\hat{v}_{t,t+\tau}\hat{v}_{t-j,t+\tau-j})$. 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.

<u>(c)</u> 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 $\hat{S}_{ff} \equiv \Sigma_{j=-P+1}^{P-1}K(j/M)\hat{\Gamma}_{ff}(j) \to_p S_{ff}$. For $\mathring{S}_{ff}$ defined in Lemma A4, we have $\hat{S}_{ff} = \mathring{S}_{ff} +$ $\Sigma_{j=-P+1}^{P-1}K(j/M)[\hat{\Gamma}_{ff}(j)-\mathring{\Gamma}_{ff}(j)]$. 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.

<u>Proof of Corollary 5.1</u>: Follows immediately from Theorem 4.1.

<u>Proof of Theorem 5.1</u>: By definition, the $(\ell+r)\times k$ matrix $F \equiv E\tilde{g}_{t+1}v_{t+\tau,\beta}$; 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,

$$\begin{aligned} P^{1/2}\hat{\hat{\alpha}} &= (P^{-1}\Sigma\hat{\hat{g}}_{t+1}\hat{\hat{g}}_{t+1}')^{-1}(P^{-1/2}\Sigma\hat{\hat{g}}_{t+1}\hat{v}_{t,t+\tau}) \\ &= (E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}(P^{-1/2}\Sigma\hat{\hat{g}}_{t+1}\hat{v}_{t,t+\tau}) + o_p(1) \\ &= (E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}(P^{-1/2}\Sigma\tilde{g}_{t+1}v_{t+\tau}) + (E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}\tilde{F}B(P^{-1/2}\Sigma H(t)) + o_p(1), \end{aligned}$$

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 $\ell$ rows of the $(\ell+r)\times k$ matrix $(E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}F$ are identically zero. Since $\hat{\alpha}$ consists of the first $\ell$ components of $\hat{\hat{\alpha}}$, $P^{1/2}\hat{\alpha}$ equals the first $\ell$ rows of $(E\tilde{g}_{t+1}\tilde{g}_{t+1}')^{-1}(P^{-1/2}\Sigma\tilde{g}_{t+1}v_{t+\tau}) + o_p(1)$, and the proof is complete.

<u>Proof of Theorem 7.1</u>: <u>(a)</u>(i) From Corollary 5.1, condition (5.1) is sufficient

to guarantee the result. From assumption (*), we have $B=(Ex_tx_t')^{-1}$ and $S_{hh}=Ev_t^2x_tx_t'$. 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 $FB=-(1\ 0\ .\ .\ .\ 0)$; 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 $Ev_t^2g_t^2 = Ev_t^2Eg_t^2$ 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 $F=-Ev_{t-1}x_t'$, $B=(Ex_tx_t')^{-1}$, $S_{hh}=Ev_t^2B^{-1}$, and $S_{fh}=-Ev_t^2F$. For encompassing this follows since $F=(-\beta_2^{*\prime}Ex_{2t}x_t', 0')$, $B=\text{diag}[(Ex_tx_t')^{-1},(Ex_{2t}x_{2t}')^{-1}]$, the $(k_1xk_1)$ block in the upper left hand corner of $S_{hh}$ is $Ev_t^2Ex_tx_t'$, and $S_{fh}=\beta_2^{*\prime}(Ev_t^2Ex_{2t}x_t', Ev_tv_{2t}Ex_{2t}x_{2t}')$.

(b) Follows from Theorem 5.1.

# Regression Based Tests of Predictive Ability

Kenneth D. West
Michael W. McCracken

May, 1997
Last revised January 1998

## Additional Appendix

This not-for-publication additional appendix contains material omitted from the body of the paper to save space:

I. Additional simulation results:

A. Plots of Actual versus Nominal Sizes of Hypothesis Tests:

| | |
|---|---|
| Mean Prediction Error | |
| Recursive | Figure AA1 |
| Rolling | Figure AA2 |
| Fixed | Figure AA3 |
| Efficiency | |
| Recursive | Figure AA4 |
| Rolling | Figure AA5 |
| Fixed | Figure AA6 |
| Encompassing | |
| Recursive | Figure AA7 |
| Rolling | Figure AA8 |
| Fixed | Figure AA9 |
| First Order Serial Correlation | |
| Recursive | Figure AA10 |
| Rolling | Figure AA11 |
| Fixed | Figure AA12 |

B. Efficiency Test with large sample sizes

| | |
|---|---|
| Rolling: P/R = 2 | Figure AA13-a |
| Rolling: P/R = 4 | Figure AA13-b |

C. Tables for large sample sizes

| | |
|---|---|
| Mean Prediction Error | |
| R+P≤800 | Table AA1-A |
| R+P≤1600 | Table AA1-B |
| Efficiency | |
| R+P≤800 | Table AA2-A |
| R+P≤1600 | Table AA2-B |
| Encompassing | |
| R+P≤800 | Table AA3-A |
| R+P≤1600 | Table AA3-B |
| First Order Serial Correlation | |
| R+P≤800 | Table AA4-A |
| R+P≤1600 | Table AA4-B |

II. Proofs:

Lemmas Al-A4
Lemma 4.1 (for fixed scheme)
Lemma 4.2 (additional detail)
Lemma 4.3
Theorem 4.2(b), (c) (additional detail)

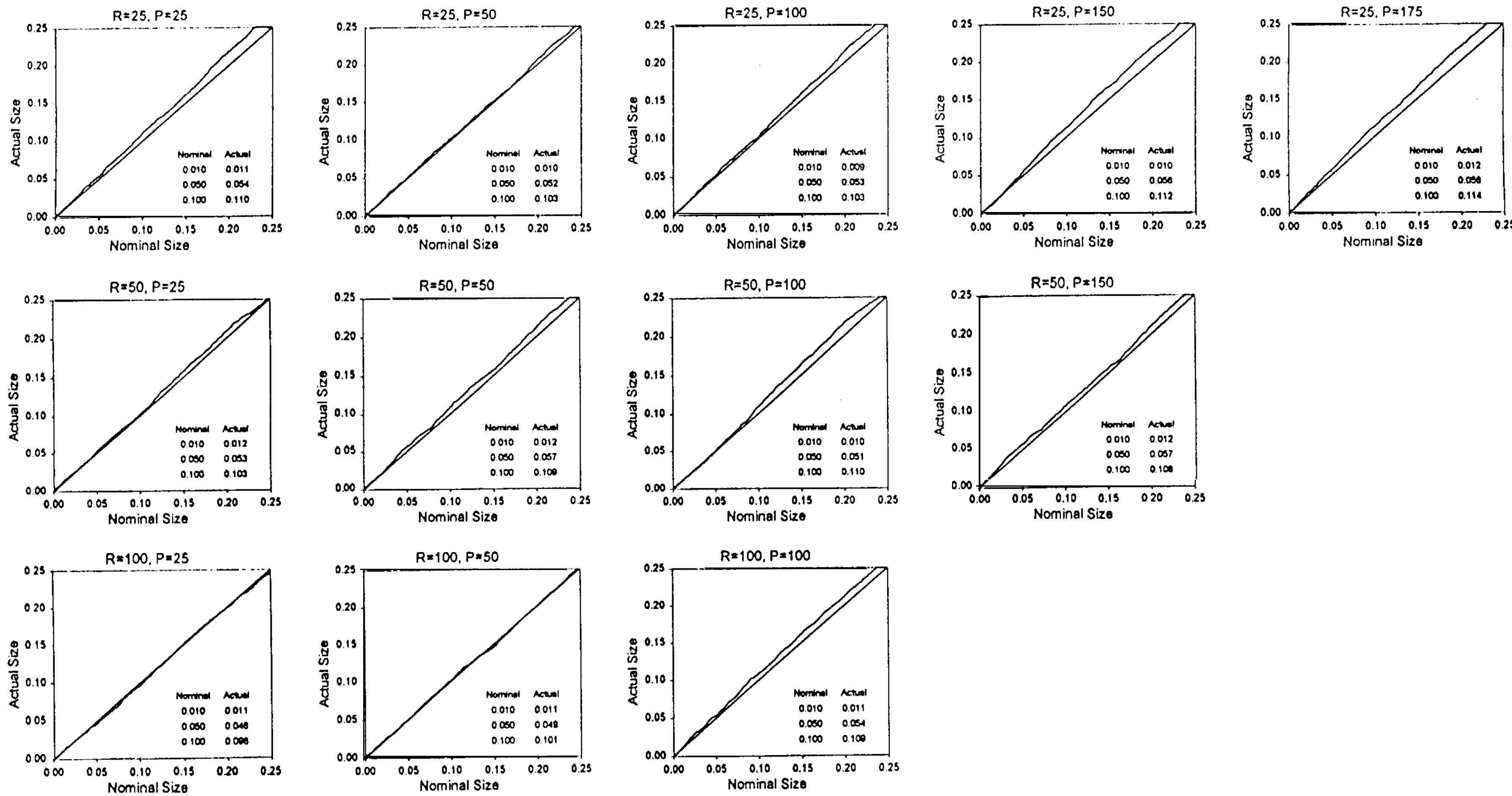


Figure AA1
Actual versus Nominal Sizes of Tests for Mean Prediction Error
Recursive scheme, R+P ≤ 200

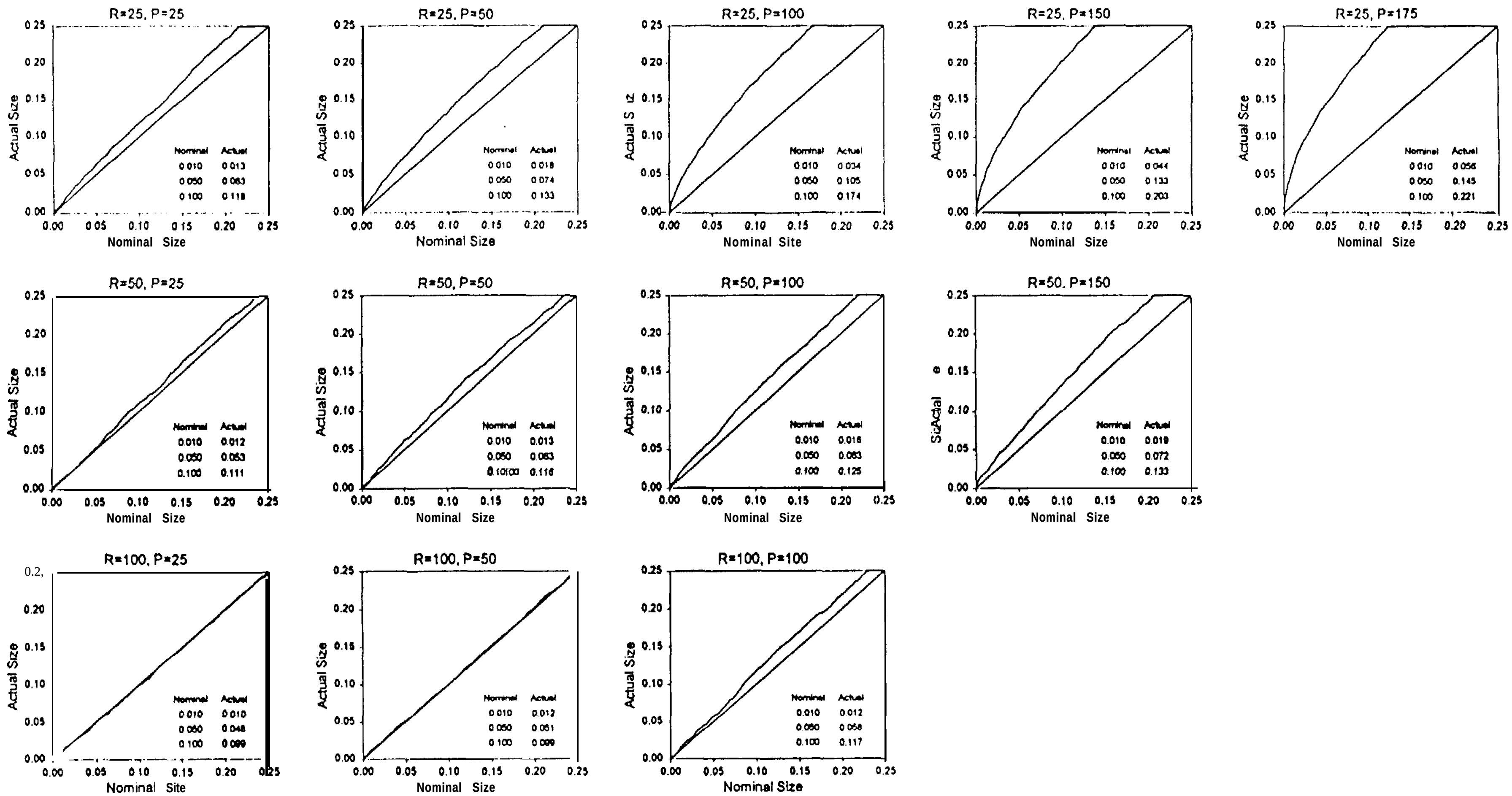


**Figure AA2**
**Actual versus Nominal Sizes of Tests for Mean Prediction Error**
**Rolling scheme, R+P ≤ 200**

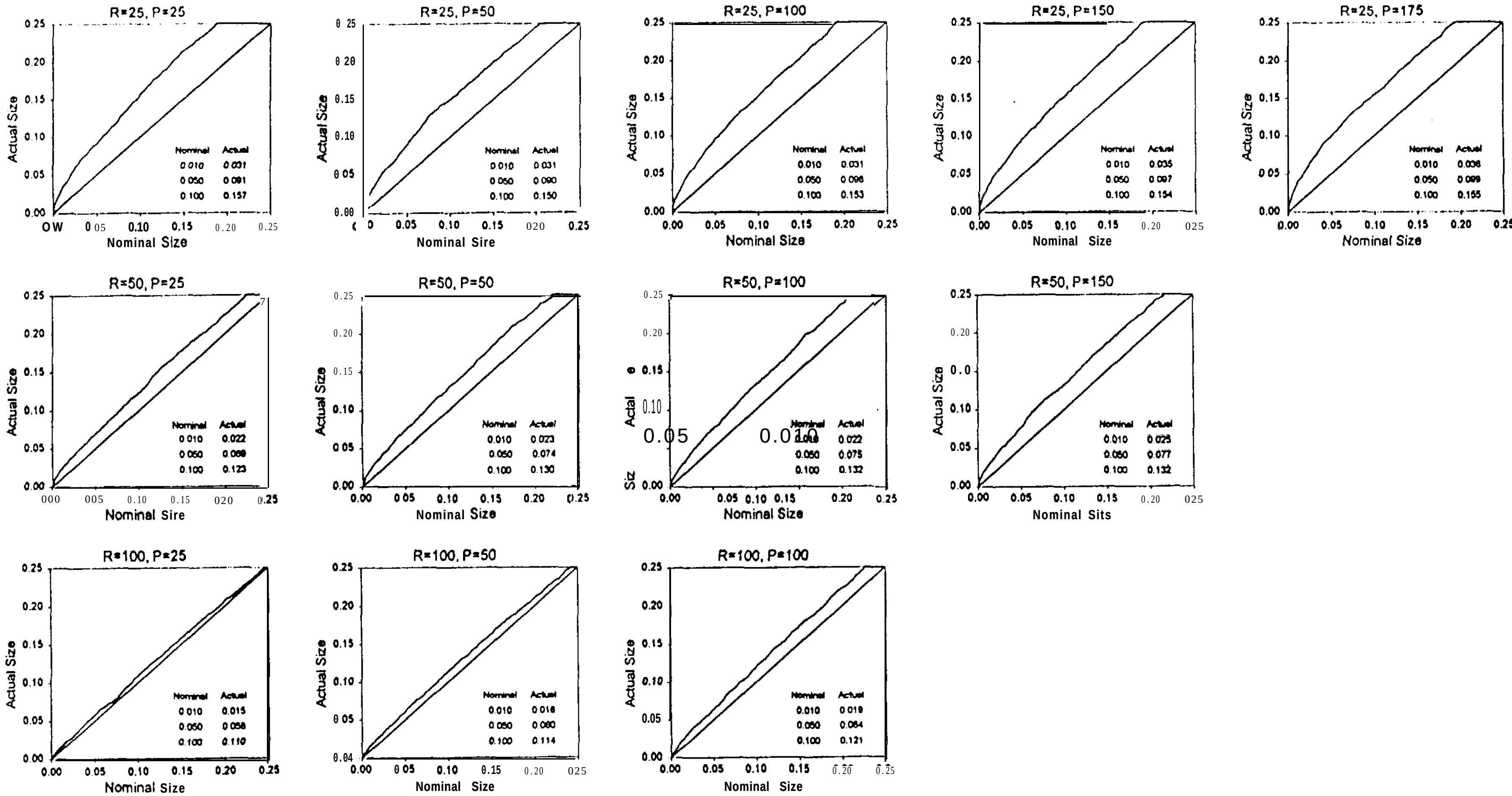


**Figure AA3**
**Actual versus Nominal Sizes of Tests for Mean Prediction Error**
**Fixed, R+P s 200**

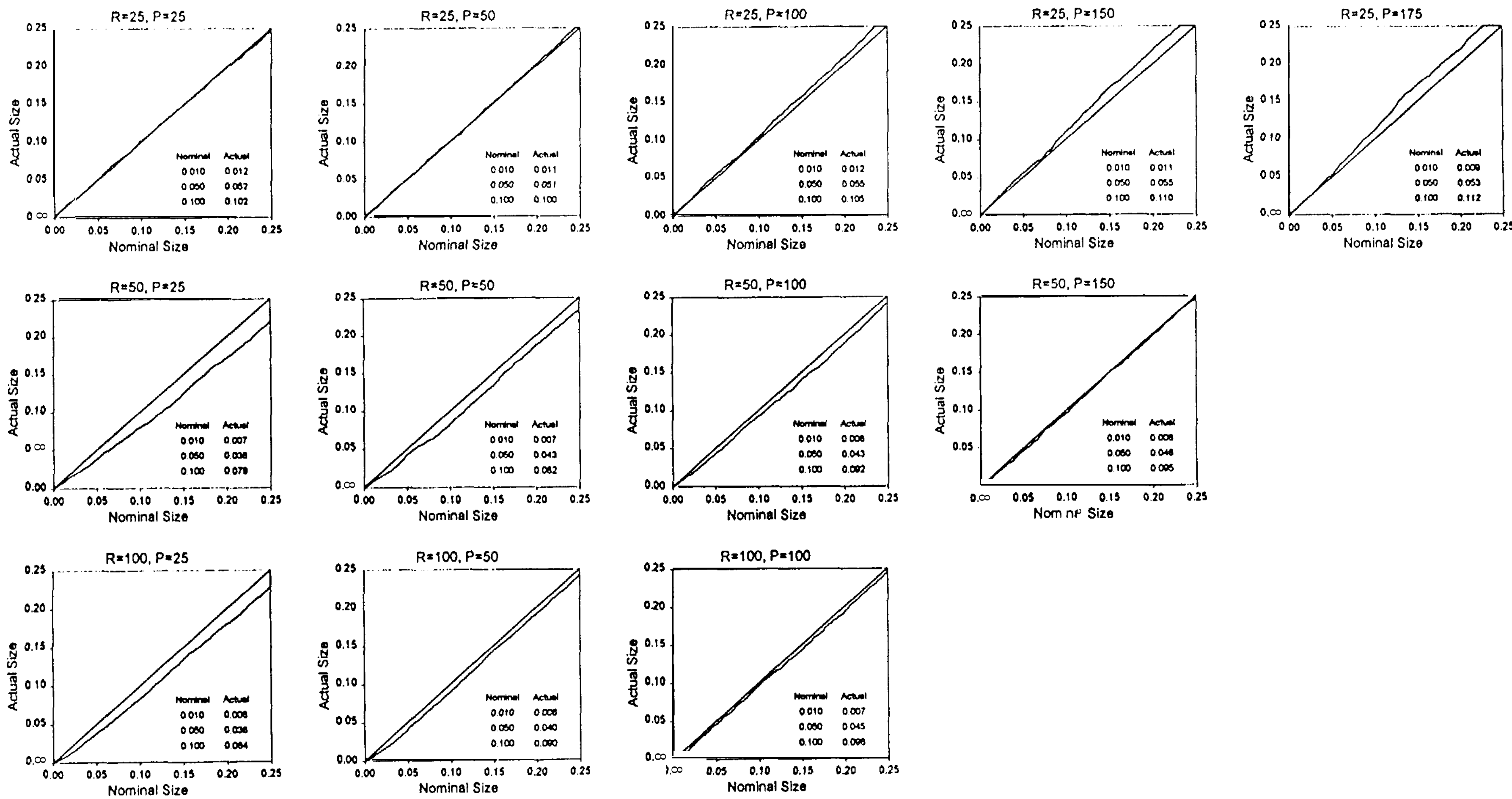


Figure AA4
Actual versus Nominal Sizes of Tests for Efficiency
Recursive scheme, R+P ≤ 200

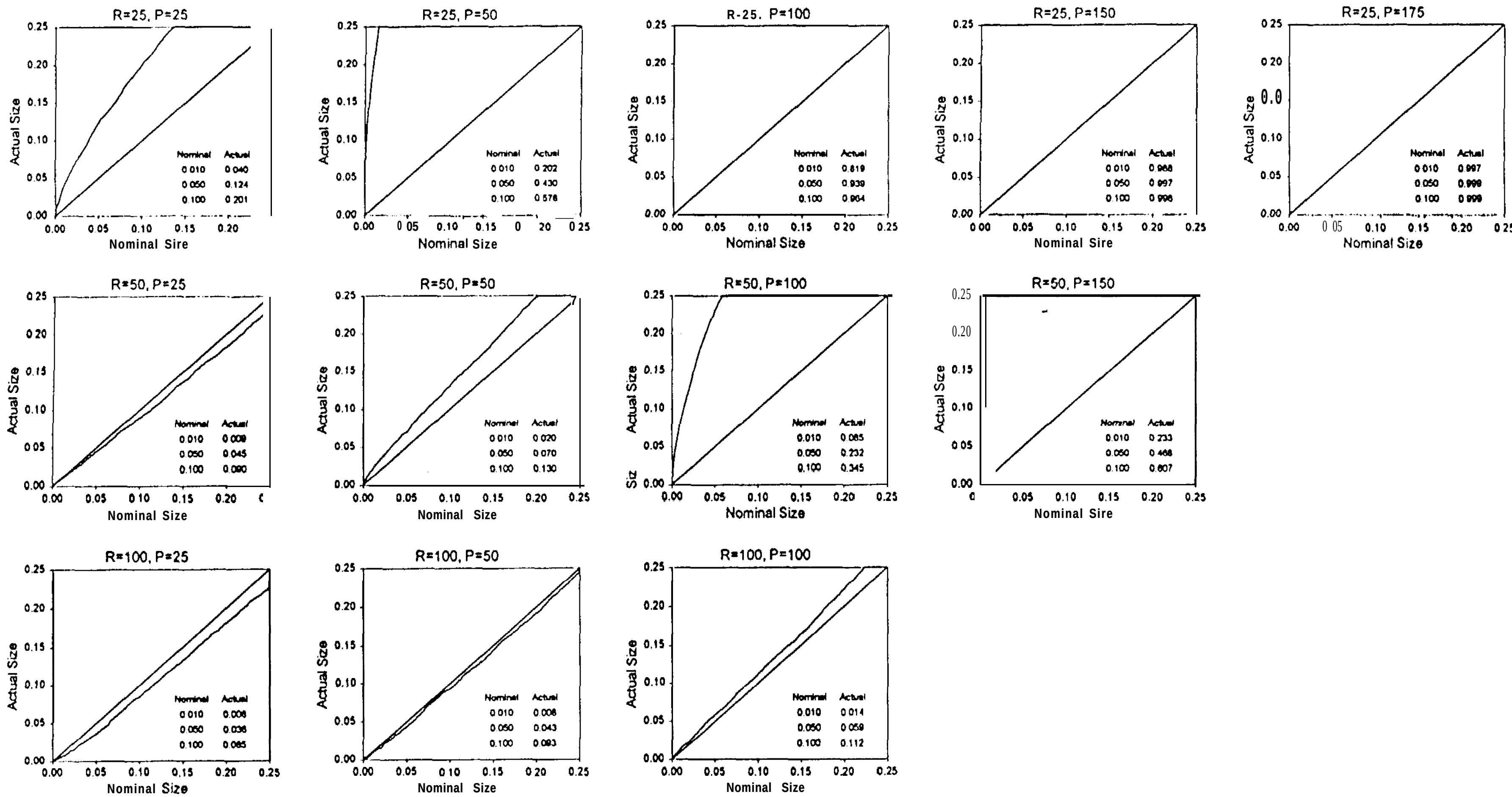


**Figure AA5**
**Actual versus Nominal Sizes of Tests for Efficiency**
**Rolling scheme, R+P ≤ 200**

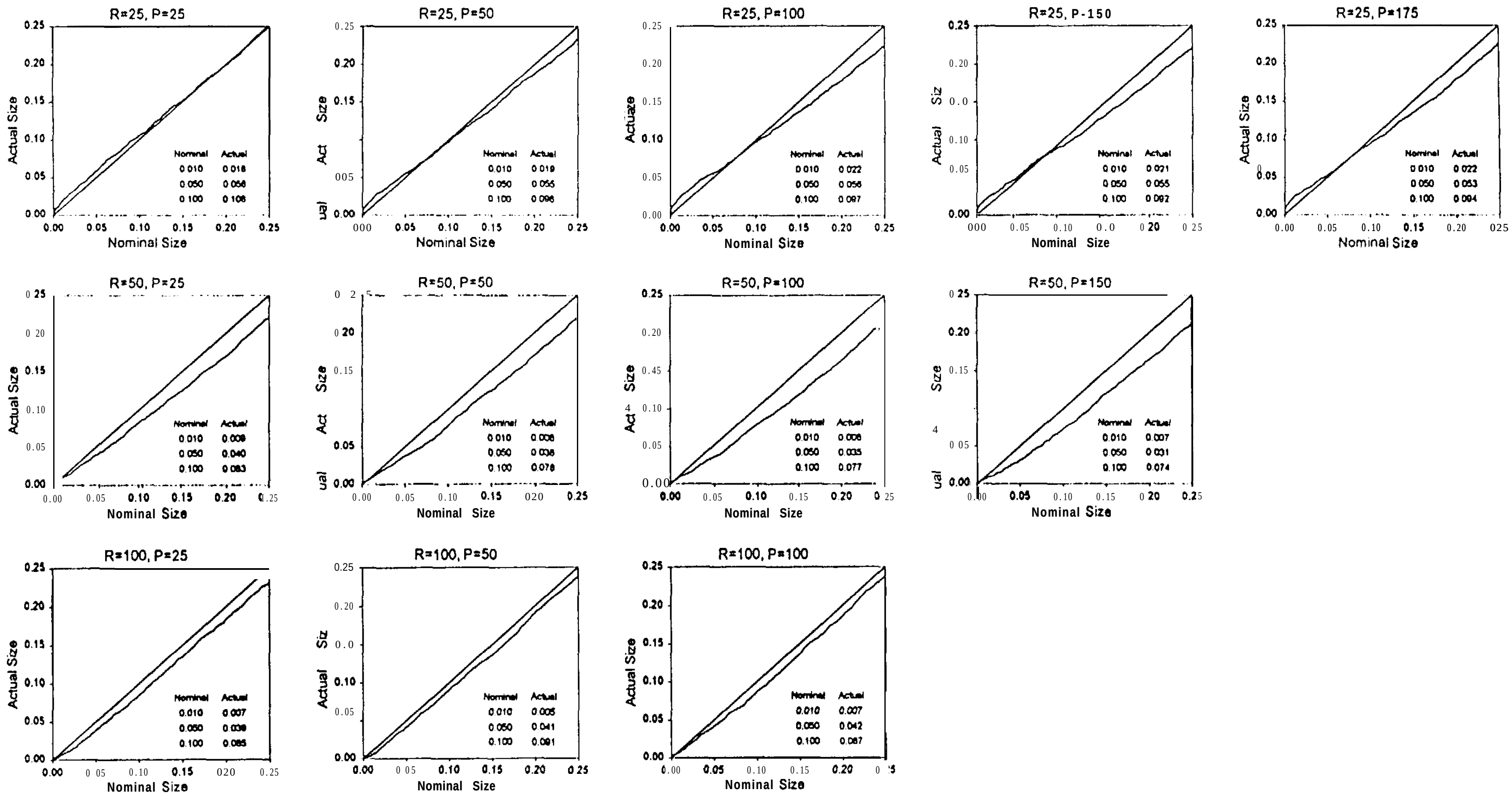


**Figure AA6**
**Actual versus Nominal Sizes of Tests for Efficiency**
**Fixed, R+P ≤ 200**

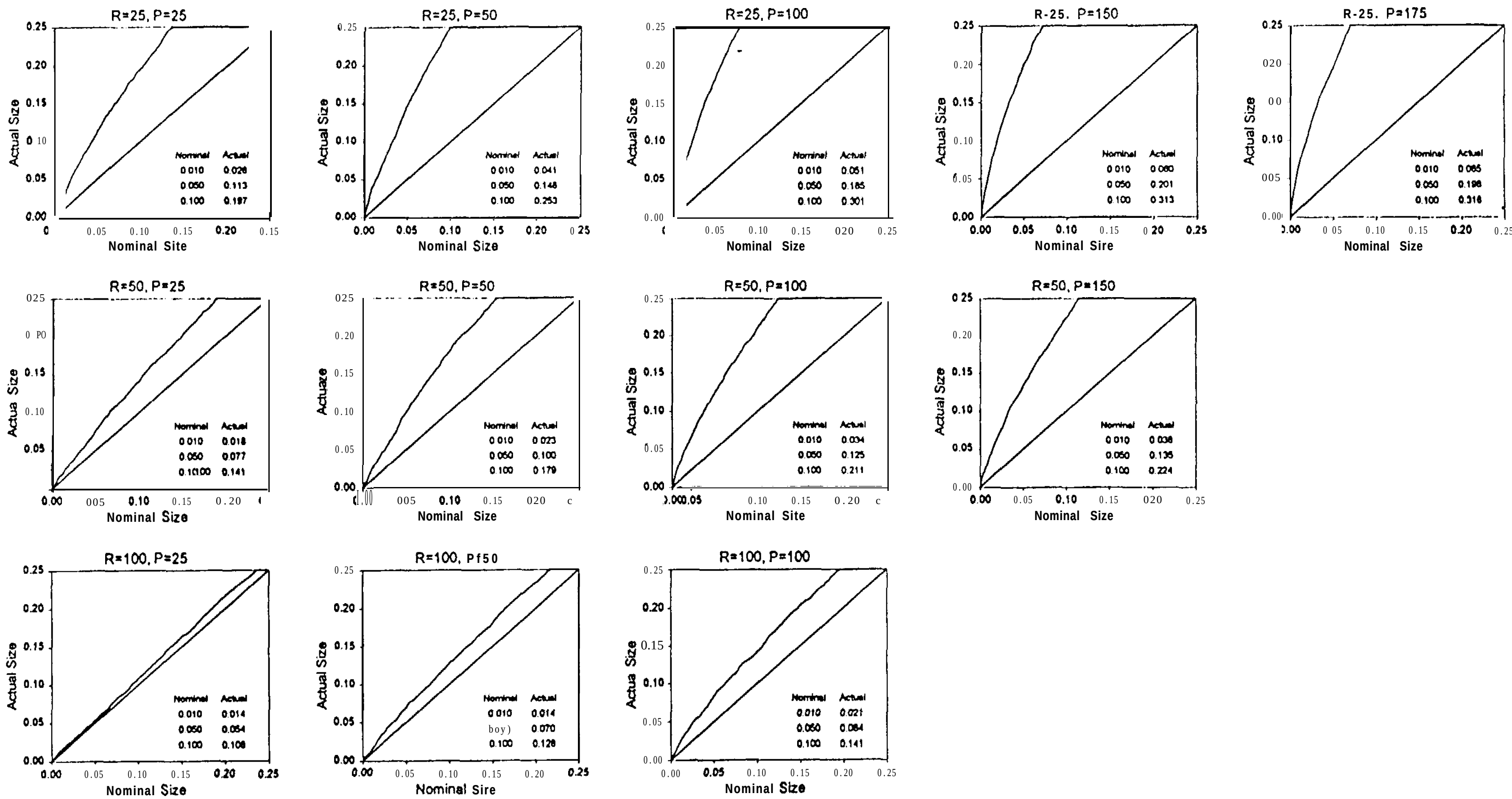


**Figure AA7**
**Actual versus Nominal Sizes of Tests for Encompassing**
**Recursive scheme, R+P ≤ 200**

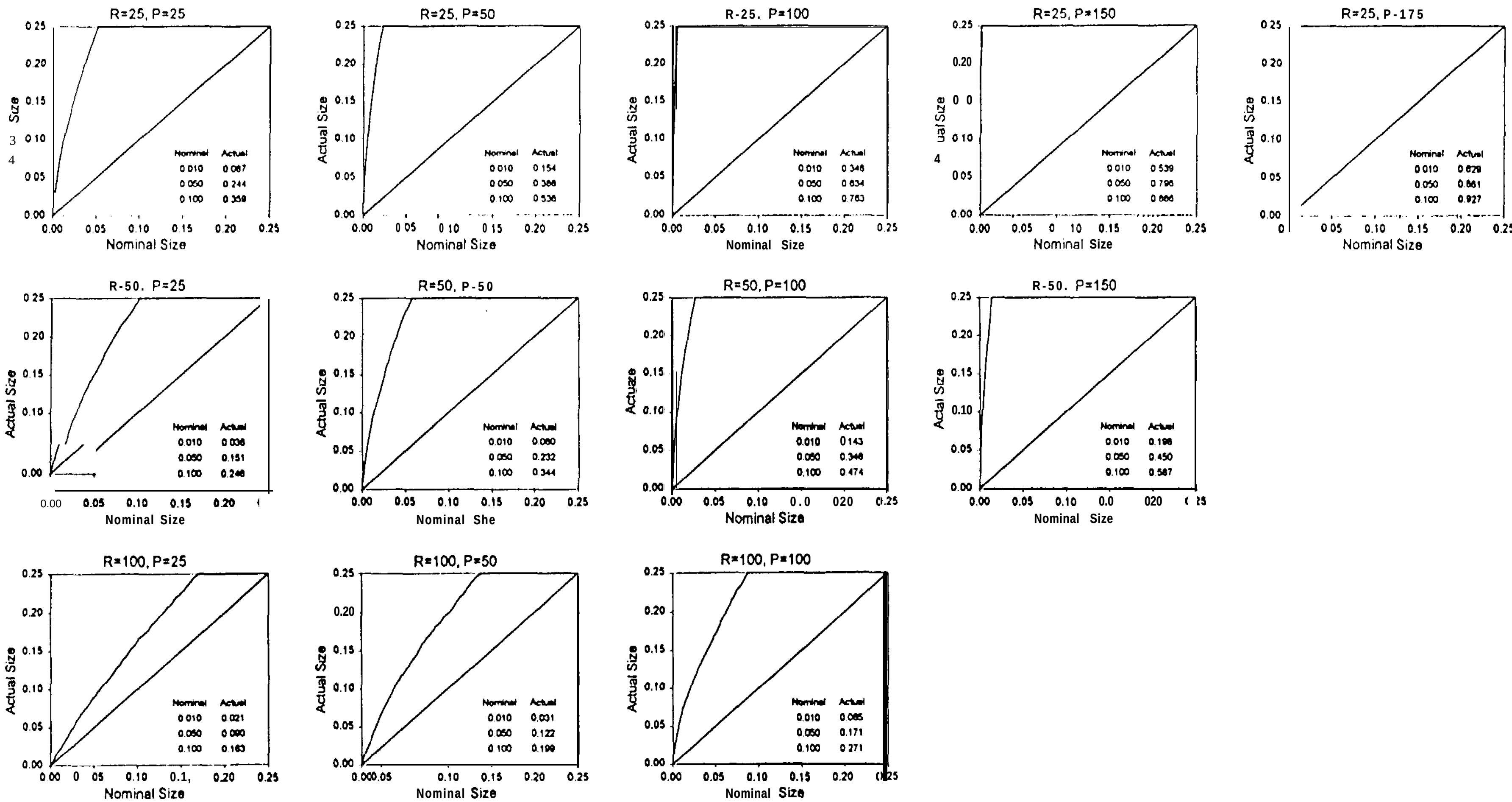


**Figure AA8**
**Actual versus Nominal Sizes of Tests for Encompassing**
**Rolling scheme, R+P $\leq$ 200**

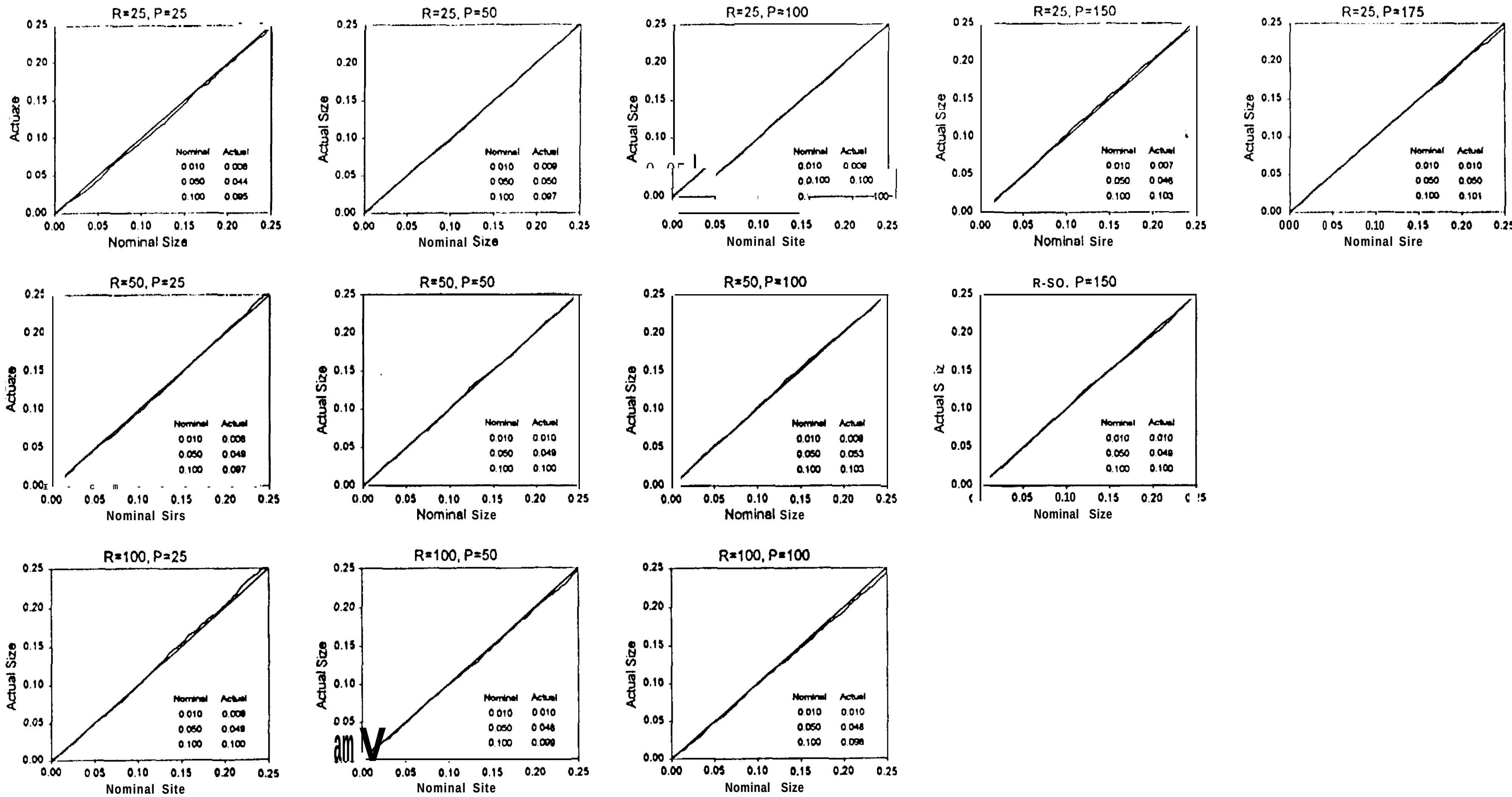


Figure AA9
Actual versus Nominal Sizes of Tests for Encompassing
Fixed, R+P ≤ 200

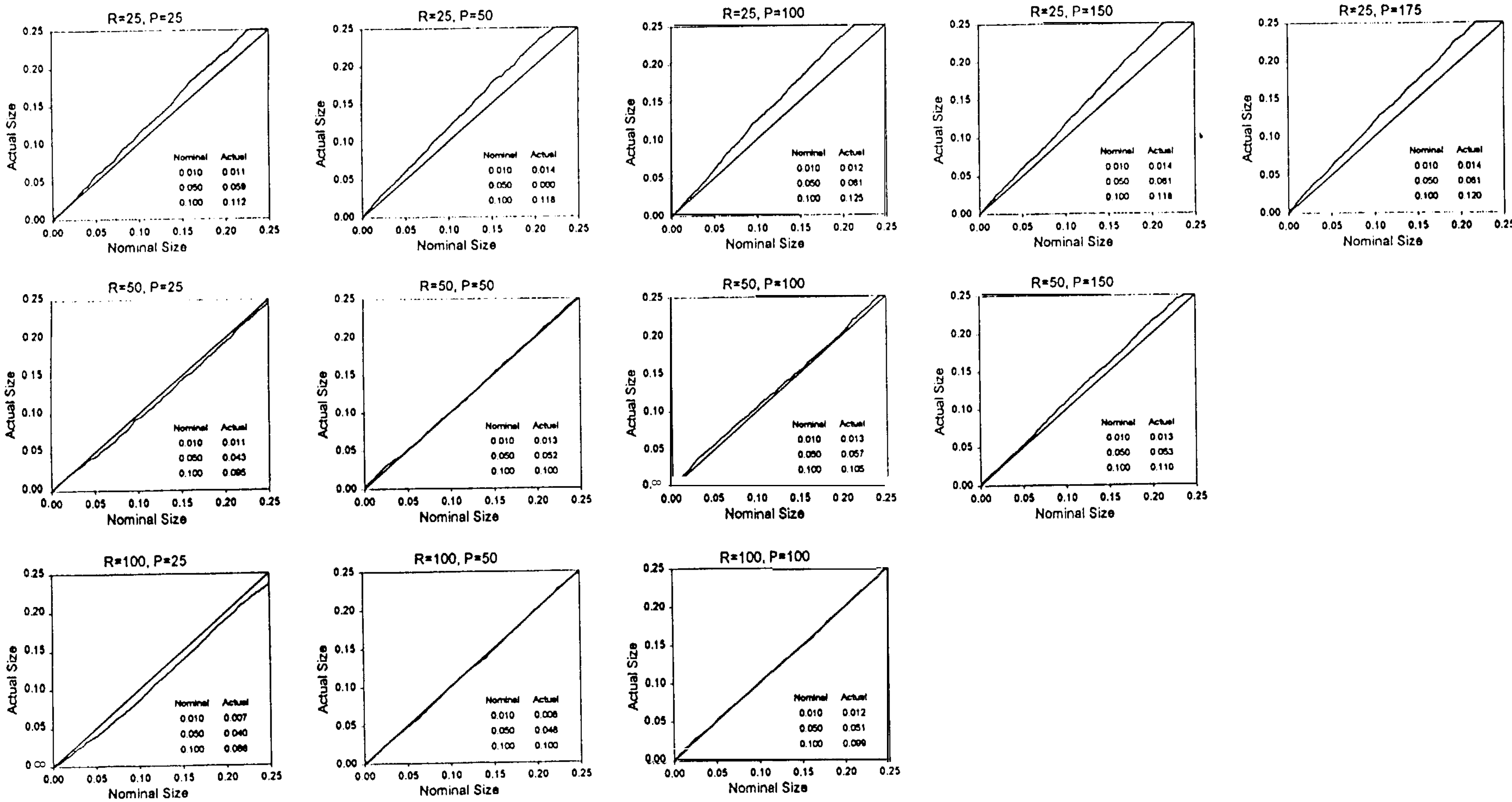


Figure AA10
Actual versus Nominal Sizes of Tests for First Order Serial Correlation
Recursive scheme, R+P ≤ 200

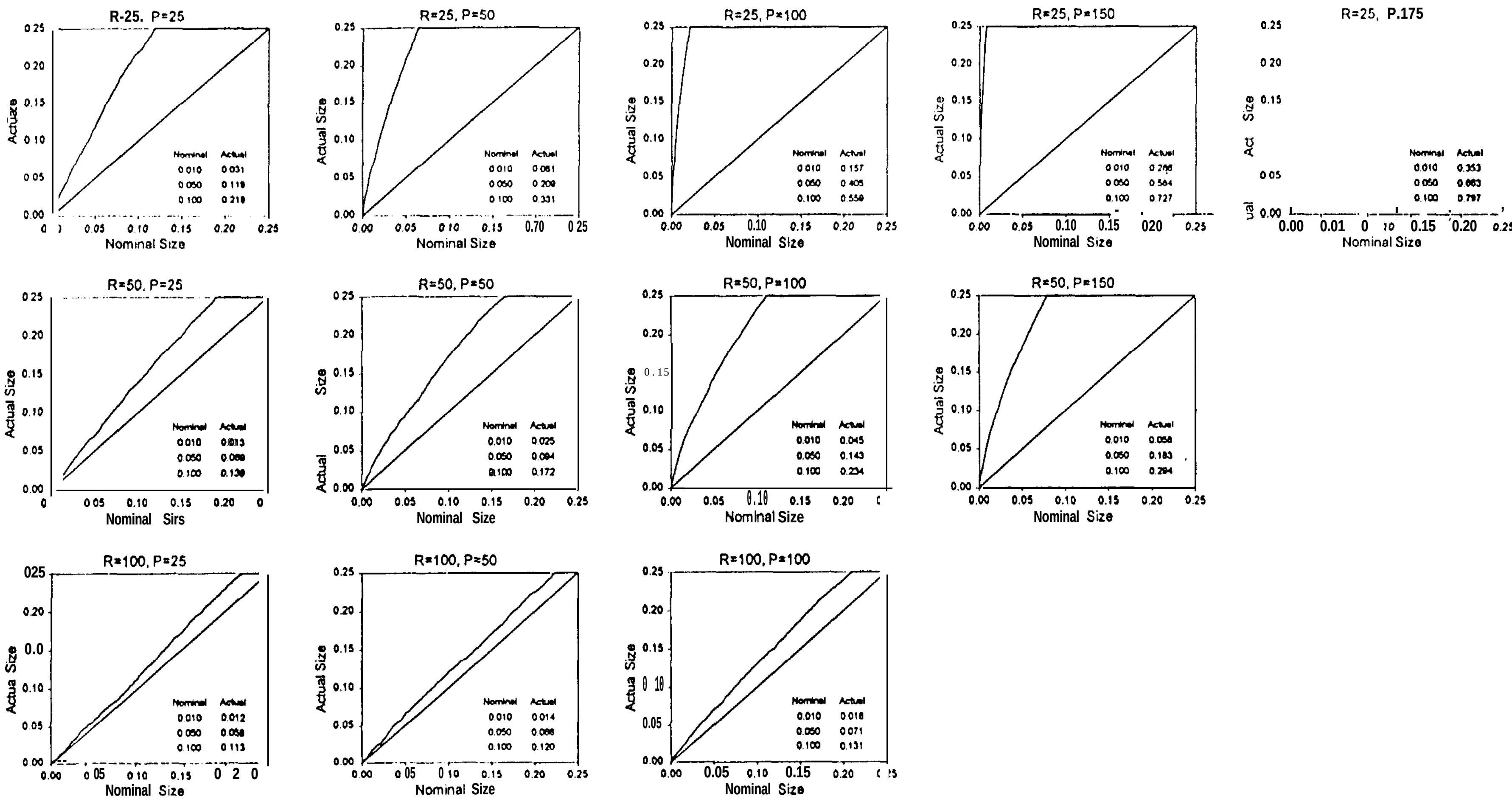


Figure AA11
Actual versus Nominal Sizes of Tests for First Order Serial Correlation
Rolling scheme, R+P ≤ 200

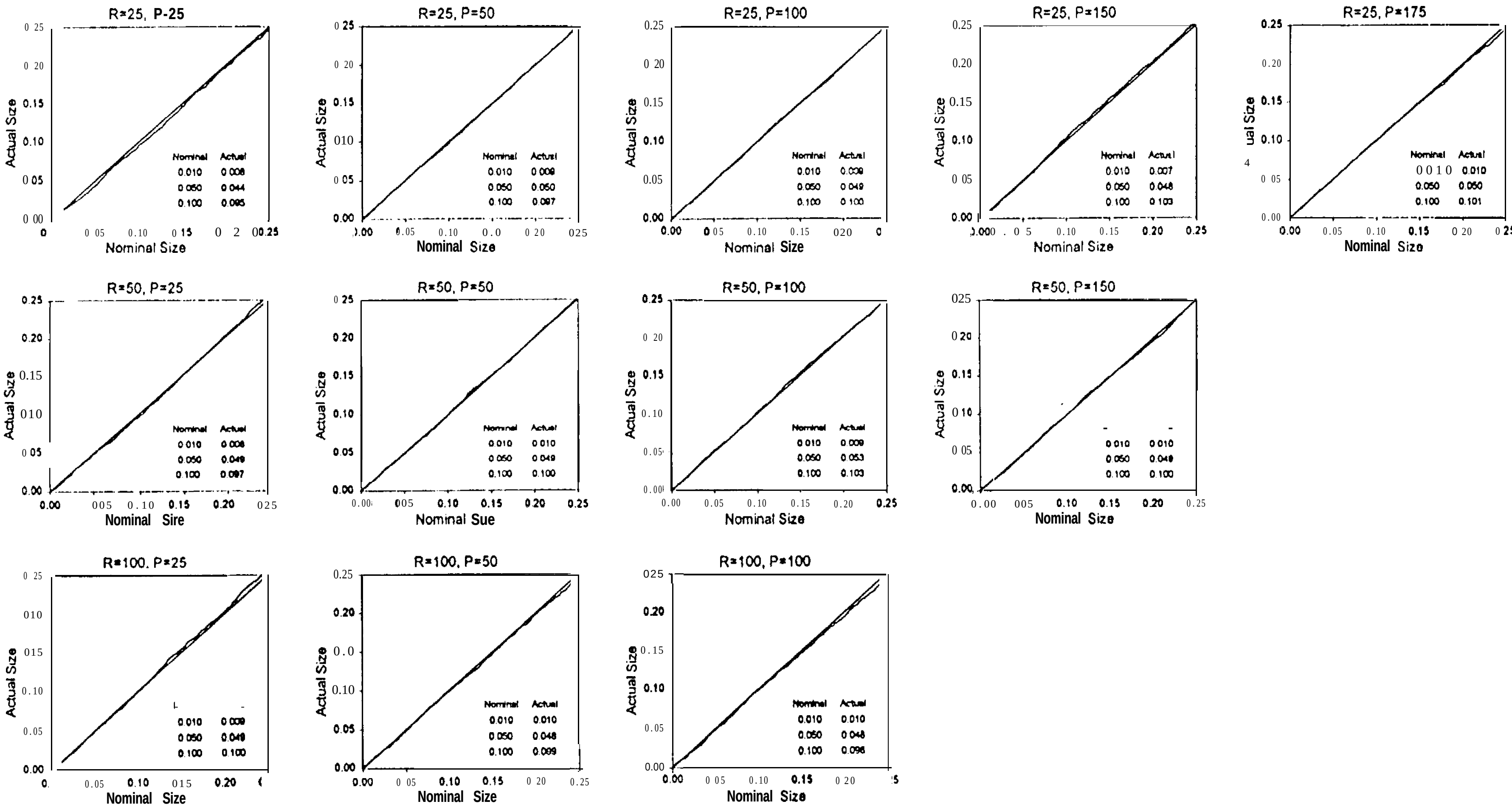


**Figure AA12**
**Actual versus Nominal Sizes of Tests for First Order Serial Correlation**
**Fixed, R+P ≤ 200**

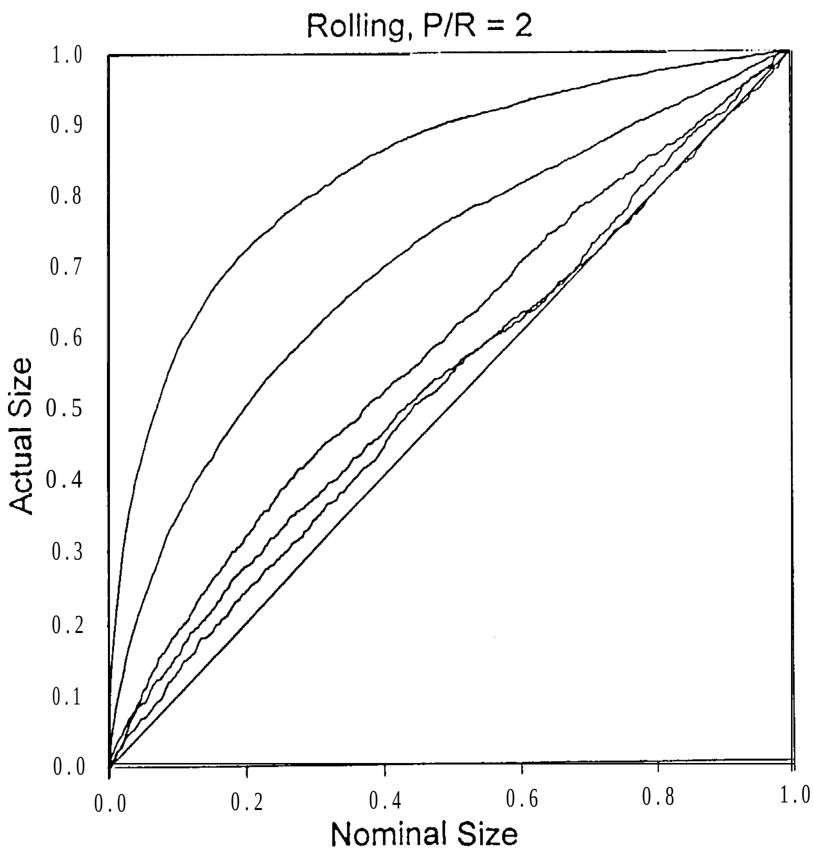


**Figure** AA13-a
**Actual versus Nominal Sizes of Tests of Efficiency**
**Rolling scheme**
**P/R** = 50/25, 100/50, 200/100, 400/200, 800/400

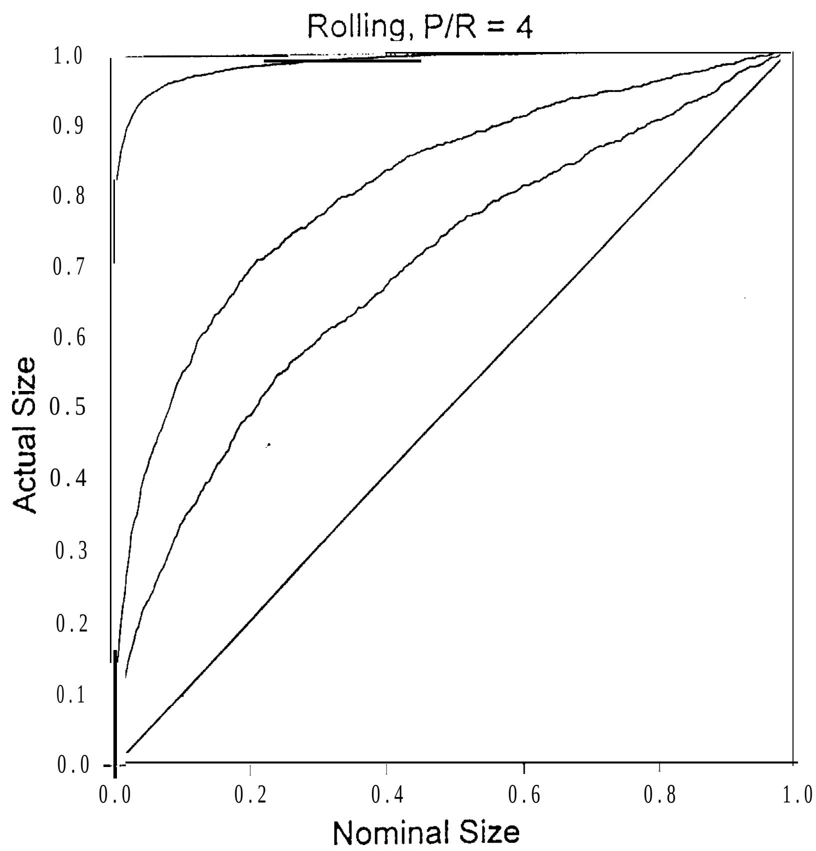


**Figure** AA13-b
**Actual versus Nominal Sizes of Tests of Efficiency**
**Rolling Scheme**
**P/R** = 100/25, 400/100, 800/200

## Table AA1

### Size of Nominal .05 Tests, Mean Prediction Error

### A. Accounting for Error in Estimation of $\beta^*$, R+P≤800

| Sampling Scheme | R | P: 100 | 200 | 400 | 600 | 700 |
|---|---|---|---|---|---|---|
| 1.Recursive | 100 | .054 | .055 | .059 | .050 | .051 |
| | 200 | .055 | .050 | .048 | .055 | |
| | 400 | .053 | .053 | .048 | | |
| 2.Rolling | 100 | .055 | .071 | .079 | .070 | .072 |
| | 200 | .050 | .050 | .054 | .055 | |
| | 400 | .052 | .041 | .046 | | |
| 3.Fixed | 100 | .067 | .072 | .068 | .070 | .070 |
| | 200 | .062 | .057 | .060 | .054 | |
| | 400 | .054 | .051 | .048 | | |

### B. Accounting for Error in Estimation of $\beta^*$, R+P≤1600

| Sampling Scheme | R | P: 200 | 400 | 800 | 1200 | 1400 |
|---|---|---|---|---|---|---|
| 1.Recursive | 200 | .050 | .048 | .050 | .049 | .046 |
| | 400 | .053 | .048 | .057 | .053 | |
| | 800 | .047 | .055 | .054 | | |
| 2.Rolling | 200 | .050 | .054 | .054 | .056 | .053 |
| | 400 | .041 | .046 | .055 | .049 | |
| | 800 | . 049 | .056 | .042 | | |
| 3.Fixed | 200 | .072 | .068 | .062 | .056 | .060 |
| | 400 | .057 | .060 | .066 | .062 | |
| | 800 | .055 | .058 | .053 | | |

Notes: 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,...,R+P-1. Panels Al-A3 report the fraction of 1000 simulations in which the conventionally computed t-statistic, divided by the square root of $\hat{\lambda}$, was greater than 1.96 in absolute value. Panels Bl-B3 report the same, but for larger values of P and R.

## Table AA2

### Size of Nominal .05 Tests, Efficiency Test

#### A. Accounting for Error in Estimation of $\beta^*$, R+P≤800

| Sampling Scheme | R | P: 100 | 200 | 400 | 600 | 700 |
|---|---|---|---|---|---|---|
| 1.Recursive | 100 | .039 | .035 | .040 | .049 | .046 |
| | 200 | .050 | .042 | .040 | .045 | |
| | 400 | .040 | .049 | .044 | | |
| 2.Rolling | 100 | .049 | .108 | .409 | .705 | .869 |
| | 200 | .042 | .051 | .091 | .145 | |
| | 400 | .051 | .046 | .046 | | |
| 3.Fixed | 100 | .032 | .042 | .031 | .038 | .037 |
| | 200 | .045 | .044 | .043 | .041 | |
| | 400 | .042 | .050 | .047 | | |

#### B. Accounting for Error in Estimation of $\beta^*$, R+P≤1600

| Sampling Scheme | R | P: 200 | 400 | 800 | 1200 | 1400 |
|---|---|---|---|---|---|---|
| 1.Recursive | 200 | .042 | .040 | .045 | .047 | .051 |
| | 400 | .049 | .044 | .061 | .061 | |
| | 800 | .048 | .038 | .047 | | |
| 2.Rolling | 200 | .051 | .091 | .229 | .424 | .566 |
| | 400 | .046 | .046 | .069 | .111 | |
| | 800 | .048 | .038 | .053 | | |
| 3.Fixed | 200 | .044 | .043 | .037 | .041 | .038 |
| | 400 | .050 | .047 | .058 | .079 | |
| | 800 | .053 | .041 | .046 | | |

Notes: 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,...,R+P-1. Panels A1-A3 report the fraction of 1000 simulations in which the conventionally computed t-statistic on the coefficient on $y_t\hat{\beta}_t$, divided by the square root of $\hat{\lambda}$, was greater than 1.96 in absolute value. Panels B1-B3 report the same, but for larger values of P and R.

## Table AA3

### Size of Nominal .05 Tests, Encompassing Test

#### A. Accounting for Error in Estimation of $\beta^*$, $R+P\leq 800$

| Sampling Scheme | R | P: 100 | 200 | 400 | 600 | 700 |
|---|---|---|---|---|---|---|
| 1.Recursive | 100 | .070 | .088 | .090 | .088 | .078 |
| | 200 | .052 | .060 | .072 | .073 | |
| | 400 | .040 | .059 | .052 | | |
| 2.Rolling | 100 | .185 | .271 | .398 | .538 | .618 |
| | 200 | .083 | .115 | .167 | .206 | |
| | 400 | .060 | .074 | .078 | | |
| 3.Fixed | 100 | .050 | .054 | .050 | .048 | .042 |
| | 200 | .054 | .044 | .060 | .055 | |
| | 400 | .047 | .061 | .052 | | |

#### B. Accounting for Error in Estimation of $\beta^*$, $R+P\leq 1600$

| Sampling Scheme | R | P: 200 | 400 | 800 | 1200 | 1400 |
|---|---|---|---|---|---|---|
| 1.Recursive | 200 | .060 | .072 | .068 | .073 | .072 |
| | 400 | .059 | .052 | .067 | .064 | |
| | 800 | .052 | .039 | .063 | | |
| 2.Rolling | 200 | .115 | .167 | .257 | .359 | .395 |
| | 400 | .074 | .078 | .113 | .142 | |
| | 800 | .061 | .054 | .082 | | |
| 3.Fixed | 200 | .044 | .060 | .043 | .049 | .042 |
| | 400 | .061 | .052 | .065 | .052 | |
| | 800 | .052 | .050 | .049 | | |

Notes: The DGP is a univariate AR(1); see text for details. Let $\hat{\beta}_{2t}$ denote the least squares estimate of a regression of $y_s$ on $y_{s-2}$ 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 $y_{t-2}\hat{\beta}_{2t}$ for t=R,...,R+P-1. Panel A1 reports the fraction of 1000 simulations in which the conventionally computed t-statistic on the coefficient on $y_{t-2}\hat{\beta}_{2t}$ that was greater than 1.96 in absolute value. Panels A2 and A3 report the same, when $y_t$ was included as a third regressor. Panels B1-B3 report the same, but for larger values of P and R.

## Table AA4

### Size of Nominal .05 Tests, Test for Zero First Order Serial Correlation

### A. Accounting for Error in Estimation of $\beta^*$, R+P≤800

| Sampling Scheme | R | P: 100 | 200 | 400 | 600 | 700 |
|---|---|---|---|---|---|---|
| 1.Recursive | 100 | .049 | .053 | .047 | .052 | .049 |
| | 200 | .050 | .038 | .056 | .051 | |
| | 400 | .055 | .054 | .044 | | |
| 2.Rolling | 100 | .071 | .096 | .137 | .178 | .209 |
| | 200 | .059 | .051 | .075 | .082 | |
| | 400 | .048 | .063 | .052 | | |
| 3.Fixed | 100 | .050 | .054 | .050 | .048 | .042 |
| | 200 | .054 | .044 | .047 | .055 | |
| | 400 | .047 | .061 | .052 | | |

### B. Accounting for Error in Estimation of $\beta^*$, R+P≤1600

| Sampling Scheme | R | P: 200 | 400 | 800 | 1200 | 1400 |
|---|---|---|---|---|---|---|
| 1.Recursive | 200 | .038 | .056 | .056 | .051 | .048 |
| | 400 | .054 | .044 | .059 | .058 | |
| | 800 | .050 | .049 | .052 | | |
| 2.Rolling | 200 | .051 | .075 | .096 | .122 | .136 |
| | 400 | .063 | .052 | .069 | .069 | |
| | 800 | .054 | .048 | .059 | | |
| 3.Fixed | 200 | ,044 | .047 | .043 | .049 | .042 |
| | 400 | .061 | .052 | .065 | .052 | |
| | 800 | .052 | .050 | .049 | | |

Notes: The DGP is a univariate AR(1); see text for details. In panel A, $\hat{v}_{t+1}$ was regressed on a constant and $\hat{v}_t$ for t=R,...,R+P-1, for the indicated values of P and R,. Panel A1 reports the fraction of 1000 simulations in which the conventionally computed t-statistic on the coefficient on $v_t$ was greater than 1.96 in absolute value. Panels A2 and A3 report the same, when $y_t$ was included as a third regressor. Panels B1-B3 report the same, but for larger values of P and R.

## II. Proofs

Proof of Lemma A1: The results for the recursive and fixed schemes follow from Lemma A3 of West (1996). For the rolling scheme, we show (a) (given (a), the proof of (b) is similar to that of Lemma A3(b) of West (1996)) : we have

$$P^aH(t) \equiv P^aR^{-1}(h_{t-R+1}+...+h_t) = P/R(P/t)^{a-1}[t^{a-1}(h_1+...+h_t)] - P/R[P^{a-1}(h_1+...+h_{t-R})]$$

$$\Longrightarrow \quad \sup_t|P^aH(t)| \le \sup_t|(P/R)(P/t)^{a-1}[t^{a-1}(h_1+...+h_t)]| +$$
$$(P/R)\sup_t|P^{a-1}(h_1+...+h_{t-R})|$$
$$\le (P/R)\,{}^a\sup_t|t^{a-1}(h_1+...+h_t)| + (P/R)\sup_t|P^{a-1}(h_1+...+h_{t-R})|.$$

Since $\pi < \infty$, it suffices to show $\sup_t|t^{a-1}(h_1+...+h_t)| \to_p 0$, $\sup_{1\le s\le P-1}|P^{a-1}(h_1+...+h_s)| \to_p 0$. The first follows from Lemma A3(a) in West (1996). The second: Let q=1 for notational simplicity. From Hall and Heyde (1980, p20) and the proof of Lemma A3(a) in West (1996), $h_t$ is a mixingale satisfying $E[\sup_{1\le s\le P-1}(h_1+...+h_s)^2] \le cP$ for a certain constant c. So $E[\sup_{1\le s\le P-1}|P^{2a-2}(h_1+...+h_s)^2|] \le cP^{2a-1} \to 0$ and the result follows from Markov's inequality.

Proof of Lemma A2: For (a), we have $P^{-1}\Sigma f_{t+\tau}^2 \equiv P^{-1}\Sigma v_{t+\tau}^2 g_{t+1}^2 \le (P^{-1}\Sigma v_{t+\tau}^4)^{1/2}(P^{-1}\Sigma g_{t+1}^4)^{1/2}$ $= O_p(1)$ by assumption 4 and Markov's inequality.

(b) By definition, $f_{t+\tau,\beta} = v_{t+\tau}g_{t+1,\beta} + g_{t+1}v_{t+\tau,\beta}$. Hence by assumption 4 and Markov's inequality,

$$P^{-1}\Sigma f_{t+\tau,\beta}^2 \le P^{-1}\Sigma v_{t+\tau}^2 g_{t+1,\beta}^2 + P^{-1}\Sigma g_{t+1}^2 v_{t+\tau,\beta}^2 + 2P^{-1}\Sigma|v_{t+\tau}g_{t+1,\beta}||g_{t+1}v_{t+\tau,\beta}|$$
$$\le (P^{-1}\Sigma v_{t+\tau}^4)^{1/2}(P^{-1}\Sigma g_{t+1,\beta}^4)^{1/2} + (P^{-1}\Sigma v_{t+\tau,\beta}^4)^{1/2}(P^{-1}\Sigma g_{t+1}^4)^{1/2} +$$
$$2(P^{-1}\Sigma v_{t+\tau}^4)^{1/4}(P^{-1}\Sigma g_{t+1,\beta}^4)^{1/4}(P^{-1}\Sigma v_{t+\tau,\beta}^4)^{1/4}(P^{-1}\Sigma g_{t+1}^4)^{1/4} = O_p(1).$$

(c) In this proof and this proof only, for any function n, let $n(\beta_t) \equiv \tilde{n}$. By definition, $\tilde{f}_{t+\tau,\beta\beta} = \tilde{v}_{t+\tau}\tilde{g}_{t+1,\beta\beta} + 2\tilde{v}_{t+\tau,\beta}\tilde{g}_{t+1,\beta} + \tilde{g}_{t+1}\tilde{v}_{t+\tau,\beta\beta}$. Now since $v_{t+\tau}$ and $g_{t+1}$ are twice continuously differentiable, $\tilde{f}_{t+\tau,\beta\beta}$ can be written as

$$\{v_{t+\tau}\tilde{g}_{t+1,\beta\beta} + \tilde{g}_{t+1,\beta\beta}v_{t+\tau,\beta}(\tilde{\beta}_t-\beta^*) + .5\tilde{g}_{t+1,\beta\beta}\tilde{\tilde{v}}_{t+\tau,\beta\beta}(\tilde{\beta}_t-\beta^*)^2\} + \{2v_{t+\tau,\beta}g_{t+1,\beta} +$$
$$2v_{t+\tau,\beta}\tilde{\tilde{g}}_{t+1,\beta\beta}(\tilde{\beta}_t-\beta^*) + 2g_{t+1,\beta}\tilde{\tilde{v}}_{t+\tau,\beta\beta}(\tilde{\beta}_t-\beta^*) + 2(\tilde{\beta}_t-\beta^*)^2\tilde{\tilde{g}}_{t+1,\beta\beta}\tilde{\tilde{v}}_{t+\tau,\beta\beta}\} +$$
$$\{g_{t+1}\tilde{v}_{t+\tau,\beta\beta} + \tilde{v}_{t+\tau,\beta\beta}g_{t+1,\beta}(\tilde{\beta}_t-\beta^*) + .5\tilde{v}_{t+1,\beta\beta}\tilde{\tilde{g}}_{t+1,\beta\beta}(\tilde{\beta}_t-\beta^*)^2\}$$

$\equiv w_{1t} + w_{2t} + w_{3t}$,

where for notational simplicity we are assuming that we have the same $\tilde{\tilde{\beta}}_t$ on the line between $\tilde{\beta}_t$ and $\beta^*$ for each expansion. To show that $P^{-1}\Sigma f^2_{t+\tau,\beta\beta}(\tilde{\beta}_t) = O_p(1)$ it suffices to show that $P^{-1}\Sigma w^2_{it} = O_p(1)$ for each i. We will show this for $w_{1t}$, the others follow from similar arguments. Squaring out $w_{1t}$ we have

$$P^{-1}\Sigma w^2_{1t} \le P^{-1}\Sigma v^2_{t+\tau}\tilde{g}^2_{t+1,\beta\beta} + P^{-1}\Sigma\tilde{g}^2_{t+1,\beta\beta}v^2_{t+\tau,\beta}(\tilde{\beta}_t-\beta^*)^2 + P^{-1}\Sigma\tilde{g}^2_{t+1,\beta\beta}\tilde{\tilde{v}}^2_{t+\tau,\beta\beta}(\tilde{\beta}_t-\beta^*)^4 +$$
$$2P^{-1}\Sigma|v_{t+\tau}||v_{t+\tau,\beta}|\tilde{g}^2_{t+1,\beta\beta}|\tilde{\beta}_t-\beta^*| + 2P^{-1}\Sigma\tilde{g}^2_{t+1,\beta\beta}|v_{t+\tau}||\tilde{\tilde{v}}_{t+\tau,\beta\beta}|(\tilde{\beta}_t-\beta^*)^2 +$$
$$2P^{-1}\Sigma\tilde{g}^2_{t+1,\beta\beta}|v_{t+\tau,\beta}||\tilde{\tilde{v}}_{t+\tau,\beta\beta}||\tilde{\beta}_t-\beta^*|^3$$

$$\le (P^{-1}\Sigma v^4_{t+\tau})^{1/2}(P^{-1}\Sigma\tilde{g}^4_{t+1,\beta\beta})^{1/2} + \qquad \text{(AA1)}$$
$$(\sup_t|\tilde{\beta}_t-\beta^*|)^2(P^{-1}\Sigma v^4_{t+\tau,\beta})^{1/2}(P^{-1}\Sigma\tilde{g}^4_{t+1,\beta\beta})^{1/2} +$$
$$(\sup_t|\beta_t-\beta^*|)^4(P^{-1}\Sigma\tilde{\tilde{v}}^4_{t+\tau,\beta\beta})^{1/2}(P^{-1}\Sigma\tilde{g}^4_{t+1,\beta\beta})^{1/2} +$$
$$2(\sup_t|\tilde{\beta}_t-\beta^*|)(P^{-1}\Sigma v^4_{t+\tau})^{1/4}(P^{-1}\Sigma\tilde{g}^4_{t+1,\beta\beta})^{1/2}(P^{-1}\Sigma v^4_{t+\tau,\beta})^{1/4} +$$
$$(\sup_t|\beta_t-\beta^*|)^2(P^{-1}\Sigma\tilde{\tilde{v}}^4_{t+\tau,\beta\beta})^{1/4}(P^{-1}\Sigma\tilde{g}^4_{t+1,\beta\beta})^{1/2}(P^{-1}\Sigma v^4_{t+\tau})^{1/4} +$$
$$2(\sup_t|\beta_t-\beta^*|)^3(P^{-1}\Sigma\tilde{\tilde{v}}^4_{t+\tau,\beta\beta})^{1/4}(P^{-1}\Sigma\tilde{g}^4_{t+1,\beta\beta})^{1/2}(P^{-1}\Sigma v^4_{t+\tau,\beta})^{1/4}.$$

The first term on the rhs of (AA1) is $O_p(1)$ by assumptions 3 and 4, and Markov's inequality; the remaining terms on the rhs of (AA1) are $O_p(1)$ by Lemma A1, assumptions 3 and 4, and Markov's inequality.

Proof of Lemma A3: Consider $\mathring{\Gamma}_{ff}(0) \equiv P^{-1}\Sigma f^2_{t+\tau}(\hat{\beta}_t)$ ; other autocovariances may be handled similarly. A mean value expansion of $f_{t+\tau}(\hat{\beta}_t)$ around $f_{t+\tau}(\beta^*)\equiv f_{t+\tau}$ yields $f_{t+\tau}(\hat{\beta}_t) = f_{t+\tau} + r_{t+\tau}$, $r_{t+\tau} \equiv [f_{t+\tau,\beta}(\hat{\beta}_t-\beta^*)] + w_{t+\tau}$, $w_{t+\tau} \equiv .5f_{t+\tau,\beta\beta}(\tilde{\beta}_t)(\hat{\beta}_t-\beta^*)^2$, $\tilde{\beta}_t$ on the line between $\hat{\beta}_t$ and $\beta^*$. Hence

$$\mathring{\Gamma}_{ff}(0) = P^{-1}\Sigma f^2_{t+\tau} + 2P^{-1}\Sigma[f_{t+\tau}r_{t+\tau}] + P^{-1}\Sigma r^2_{t+\tau} . \qquad \text{(AA2)}$$

The first term on the rhs of (AA2) converges in probability to I',,(O) by White (1984, Corollary 3.48). For the second term, the triangle and Cauchy-Schwarz inequalities yield

$$|P^{-1}\Sigma(f_{t+\tau}r_{t+\tau})| = |P^{-1}\Sigma[f_{t+\tau}f_{t+\tau,\beta}(\hat{\beta}_t-\beta^*)] + P^{-1}\Sigma f_{t+\tau}w_{t+\tau}|$$

$$s(\sup_t|\hat{\beta}_t-\beta^*|)(P^{-1}\Sigma f_{t+\tau}^2)^{1/2}(P^{-1}\Sigma f_{t+\tau,\beta}^2)^{1/2} + .5(\sup_t|\hat{\beta}_t-\beta^*|)^2(P^{-1}\Sigma f_{t+\tau}^2)^{1/2}[P^{-1}\Sigma f_{t+\tau,\beta\beta}^2(\tilde{\beta}_t)]^{1/2}\rightarrow_p 0$$

by Lemmas Al and A2. For the third term, it is straightforward to show that $P^{-1/2}\Sigma r_{t+\tau}^2$ is less than or equal to

$$(\sup_t|\hat{\beta}_t-\beta^*|)^2(P^{-1}\Sigma f_{t+\tau,\beta}^2) + (\sup_t|\hat{\beta}_t-\beta^*|)^3(P^{-1}\Sigma f_{t+\tau,\beta}^2)^{1/2}(P^{-1}\Sigma f_{t+\tau,\beta\beta}^2(\tilde{\beta}_t))^{1/2} + (\sup_t|\hat{\beta}_t-\beta^*|)^4(P^{-1}\Sigma f_{t+\tau,\beta\beta}^2(\beta_t)).$$

Since $\sup_t|\hat{\beta}_t-\beta^*| = o,(l)$ by Lemma Al, the result follows from Lemma A2.

<u>Proof of Lemma A4</u>: Let $K_j \equiv K(j/M)$, suppressing for simplicity the dependence of $K_j$ on M and thus P. Furthermore, define $r_{t+\tau}$ and $w_{t+\tau}$ as in (AA2). An expansion such as in the proof of Lemma A3 then yields

$$\Sigma_{j=-P+1}^{P-1}K_j\mathring{\Gamma}_{ff}(j) = \Sigma_{j=-P+1}^{P-1}K_j(P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}f_{t+\tau-j}) + \Sigma_{j=-P+1}^{P-1}K_j(P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}r_{t+\tau-j}) + \Sigma_{j=-P+1}^{P-1}K_j(P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau-j}r_{t+\tau}) + \Sigma_{j=-P+1}^{P-1}K_j(P^{-1}\Sigma_{t=R+j}^{T}r_{t+\tau}r_{t+\tau-j}).$$

It follows from Andrews (1991) that $\Sigma_{j=-P+1}^{P-1}K_j(P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}f_{t+\tau-j}) \rightarrow_p \Sigma_{j=-\infty}^{\infty}\Gamma_{ff}(j)$, so it suffices to show that the other three double summations converge in probability to zero. We will show this for the second double summation; the arguments for the third and fourth double summations are similar. Note that

$$|P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}f_{t+\tau-j,\beta}| \le (P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}^2)^{1/2}(P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau,\beta}^2)^{1/2} \le (P^{-1}\Sigma f_{t+\tau}^2)^{1/2}(P^{-1}\Sigma f_{t+\tau,\beta}^2)^{1/2},$$

$$|P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}w_{t+\tau-j}| \le .5(\sup_t|\hat{\beta}_t-\beta^*|)^2(P^{-1}\Sigma f_{t+\tau}^2)^{1/2}(P^{-1}\Sigma f_{t+\tau,\beta\beta}^2(\tilde{\beta}_t))^{1/2}.$$

Let "a" be defined as in the statement of the theorem, $0<a<.5$. Then for $\pi<\infty$

$$|\Sigma_{j=-P+1}^{P-1}K_j(P^{-1}\Sigma_{t=R+j}^{T}f_{t+\tau}r_{t+\tau-j})| \le (M/P^a)(M^{-1}\Sigma_{j=-P+1}^{P-1}|K_j|)\{(P^a\sup_t|\hat{\beta}_t-\beta^*|)(P^{-1}\Sigma f_{t+\tau}^2)^{1/2}(P^{-1}\Sigma f_{t+\tau,\beta}^2)^{1/2} + .5(P^{a/2}\sup_t|\hat{\beta}_t-\beta^*|)^2(P^{-1}\Sigma f_{t+\tau}^2)^{1/2}[P^{-1}\Sigma f_{t+\tau,\beta\beta}^2(\tilde{\beta}_t)]^{1/2}\}.$$

By Lemma Al, $P^a\sup_t|\hat{\beta}_t-\beta^*| \rightarrow_p 0$ and by Lemma A2 each of the summations inside the braces is $O_p(1)$. Since assumption $(M/P^a) = O(1)$ and $M^{-1}\Sigma_{j=-P+1}^{P-1}|K_j| \rightarrow \int_{-\infty}^{\infty}|K(x)|dx < \infty$ the desired result follows. For $\pi=\infty$, the logic is the same except that $R^a$ replaces $P^a$.

<u>Additional detail on proof of Lemma 4.1</u>:

For the fixed scheme, for which $\hat{\beta}_t \equiv \hat{\beta}_R$ and H(t) $\equiv$ H(R), a simple mean value expansion of $P^{-1/2}\Sigma g_{t+1}(\hat{\beta}_R)v_{t+\tau}(\hat{\beta}_R) \equiv P^{-1/2}\Sigma f_{t+\tau}(\hat{\beta}_R)$ around $P^{-1/2}\Sigma f_{t+\tau}(\beta^*)$ yields

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

for $w_{t+\tau}$ defined in (AA2). From assumption 2, $B(t) \to_p B$; from assumption 4, $P^{-1}\Sigma f_{t+\tau\ \beta} \to_p F$ (White (1984, Corollary 3.48)), and $P^{1/2}H(R) = O_p(1)$. So the result follows if $P^{-1/2}\Sigma w_{t+\tau} \to_p 0$. We have

$$|P^{-1/2}\Sigma w_{t+\tau}| \le (P/R)^{1/2}(R^{1/4}|\hat{\beta}_R-\beta^*|)^2(P^{-1}\Sigma|f_{t+\tau,\beta\beta}(\tilde{\beta}_R)|).$$

Since $\pi \equiv \lim P/R < \infty$ and $R^{1/4}H(R) \to_p 0$, the result follows from Lemma A2(c).

<u>Additional detail on proof of Lemma 4.2</u>:

Note: We will show the result for the recursive scheme. The other sampling schemes follow from a similar argument. In particular, the only change occurs in the definition of $b_{j,T}$.

Given Lemma 4.1 (a), it suffices to show that $P^{-1/2}\Sigma(g_{t+1}v_{t+\tau} + FBH(t)) \sim_A N(0,\Omega)$. Let $b_{R,T} = (1/R + ...+ 1/T)$. Then

$$FB\Sigma H(t) = b_{R,T}FBh_1 + ...+ b_{R,T}FBh_R + b_{R+1,T}FBh_{R+1} + ...+ b_{T,T}FBh_T.$$

Now define $Z_{T,t} \equiv P^{-1/2}(b_{R,T}FBh_t)$ for $1 \le t \le R$ and $Z_{T,t} \equiv P^{-1/2}(g_{t+1}v_{t+\tau} + b_{t,T}FBh_t)$ for $R+1 \le t \le T$. Using assumption 4 and Lemma 4.1 we know that $\Omega$ is p.d.. Hence for large enough T, $\Omega_T \equiv Var(\Sigma_{t=1}^{T}Z_{T,t})$ is invertible. If we define $X_{T,t} \equiv \Omega_T^{-1/2}Z_{T,t}$, then Theorem 3.1 of Wooldridge and White (1989) implies that

$$\Omega_T^{-1/2}P^{-1/2}\Sigma(g_{t+1}v_{t+\tau} + FBH(t)) = \Sigma X_{T,t} \sim_A N(0,I_\ell).$$

Then since $\Omega$ is p.d., we know that $P^{-1/2}\Sigma(g_{t+1}v_{t+\tau} + FBH(t)) \sim_A N(0,\Omega)$.

<u>Additional detail on proof of Lemma 4.3</u>:

Since $g_{t+1}(\beta)$ is twice continuously differentiable, it admits a mean value expansion

$$g_{t+1}(\hat{\beta}_t) = g_{t+1} + g_{t+1,\beta}(\hat{\beta}_t-\beta^*) + .5g_{t+1,\beta\beta}(\ddot{\beta}_t)(\hat{\beta}_t-\beta^*)^2$$

and hence

$$P^{-1}\Sigma g^2_{t+1}(\hat\beta_t) = P^{-1}\Sigma g^2_{t+1} + P^{-1}\Sigma\{g^2_{t+1,\beta}(\hat\beta_t-\beta^*)^2 + .25g^2_{t+1,\beta\beta}(\beta_t)(\hat\beta_t-\beta^*)^4 +$$
$$2g_{t+1}g_{t+1,\beta}(\hat\beta_t-\beta^*) + g_{t+1}g_{t+1,\beta\beta}(\tilde\beta_t)(\hat\beta_t-\beta^*)^2 + 2g_{t+1,\beta}g_{t+1,\beta\beta}(\tilde\beta_t)(\hat\beta_t-\beta^*)^3\}$$
$$\equiv P^{-1}\Sigma g^2_{t+1} + P^{-1}\Sigma w_{t+1}.$$

Now since $g_{t+1}$ is fourth order stationary and has 8d moments, Chebyshev's inequality implies that $P^{-1}\Sigma g^2_{t+1} \rightarrow_p Eg^2_{t+1}$.

To show that $P^{-1}\Sigma w_{t+1}$ = o,(l) we use the triangle and Cauchy-Schwarz inequalities to obtain

$$|P^{-1}\Sigma w_{t+1}| \le (\sup_t|\hat\beta_t-\beta^*|)^2(P^{-1}\Sigma g^2_{t+1,\beta}) +$$
$$(\sup_t|\hat\beta_t-\beta^*|)^4(P^{-1}\Sigma g^2_{t+1,\beta\beta}(\tilde\beta_t)) +$$
$$2(\sup_t|\hat\beta_t-\beta^*|)(P^{-1}\Sigma g^2_{t+1})^{1/2}(P^{-1}\Sigma g^2_{t+1,\beta})^{1/2} +$$
$$(\sup_t|\hat\beta_t-\beta^*|)^2(P^{-1}\Sigma g^2_{t+1})^{1/2}(P^{-1}\Sigma g^2_{t+1,\beta\beta}(\tilde\beta_t))^{1/2} +$$
$$2(\sup_t|\hat\beta_t-\beta^*|)^3(P^{-1}\Sigma g^2_{t+1,\beta})^{1/2}(P^{-1}\Sigma g^2_{t+1,\beta\beta}(\beta_t))^{1/2}$$

Since $\sup_t|\hat\beta_t-\beta^*|$ = o,(l) by Lemma Al, and $P^{-1}\Sigma g^2_{t+1,\beta\beta}(\tilde\beta_t) = O_p(1)$ by assumption 3, and the remaining terms are $O_p(1)$ by assumption 4 and Markov's inequality, the result is established.

<u>Additional detail on proof of Theorem 4.2(b) :</u>

Expanding $\hat\eta_{t+\tau}$ we have

$$P^{-1}\Sigma^T_{t=R+j}\hat g_{t+1}\hat g_{t+1-j}(\hat\eta_{t+\tau}\hat\eta_{t+\tau-j} - \hat v_{t-j,t+\tau}\hat v_{t,t+\tau-j}) = \quad \text{(AA3)}$$
$$P^{-1}\Sigma^T_{t=R+j}\{-\hat g_{t+1-j}\hat g^2_{t+1}\hat v_{t-j,t+\tau-j}\hat\alpha - \hat g_{t+1}\hat g^2_{t+1-j}\hat v_{t,t+\tau}\hat\alpha + \hat g^2_{t+1}\hat\alpha^2\hat g^2_{t+1-j}\}$$
$$\equiv P^{-1}\Sigma^T_{t=R+j}w_{t+1-j}$$

Via the triangle and Cauchy-Schwarz inequalities we have

$$|P^{-1}\Sigma^T_{t=R+j}w_{t+1-j}| \le (P^{-1}\Sigma^T_{t=R+j}\hat g^2_{t+1-j}\hat v^2_{t-j,t+\tau-j})^{1/2}(P^{-1}\Sigma^T_{t=R+j}\hat\alpha^2\hat g^4_{t+1})^{1/2} +$$
$$(P^{-1}\Sigma^T_{t=R+j}\hat g^2_{t+1}\hat v^2_{t,t+\tau})^{1/2}(P^{-1}\Sigma^T_{t=R+j}\hat\alpha^2\hat g^4_{t+1-j})^{1/2} + (P^{-1}\Sigma^T_{t=R+j}\hat\alpha^4\hat g^4_{t+1})^{1/2}(P^{-1}\Sigma^T_{t=R+j}\hat g^4_{t+1-j})^{1/2}$$

$$\le |\hat\alpha|(P^{-1}\Sigma\hat g^2_{t+1}\hat v^2_{t,t+\tau})^{1/2}(P^{-1}\Sigma\hat g^4_{t+1})^{1/2} +$$
$$|\hat\alpha|(P^{-1}\Sigma\hat g^2_{t+1}\hat v^2_{t,t+\tau})^{1/2}(P^{-1}\Sigma\hat g^4_{t+1})^{1/2} + \hat\alpha^2(P^{-1}\Sigma\hat g^4_{t+1})^{1/2}(P^{-1}\Sigma\hat g^4_{t+1})^{1/2}$$

The last inequality follows by adding non-negative terms, and pulling the $\hat{\alpha}$ terms out of the summations. By Theorem 4.1, $\hat{\alpha} = o_p(1)$, hence we need to show that the remaining terms are $O_p(1)$. That $P^{-1}\Sigma\hat{g}^2_{t+1}\hat{v}^2_{t,t+\tau} = O_p(1)$ follows from Lemma A3.

Since $g_{t+1}(\beta)$ is twice continuously differentiable, it admits a mean value expansion such that, $P^{-1}\Sigma g^4_{t+1}(\hat{\beta}_t) = P^{-1}\Sigma g^4_{t+1} + P^{-1}\Sigma r_{t+1}$, where

$$\begin{aligned} r_{t+1} \equiv\ & [g_{t+1} + g_{t+1,\beta}(\hat{\beta}_t-\beta^*)]^4 + [.5g_{t+1,\beta\beta}(\tilde{\beta}_t)(\hat{\beta}_t-\beta^*)^2]^4 + \\ & 4[g_{t+1} + g_{t+1,\beta}(\hat{\beta}_t-\beta^*)]^3[.5g_{t+1,\beta\beta}(\tilde{\beta}_t)(\hat{\beta}_t-\beta^*)] + \\ & 12[g_{t+1} + g_{t+1,\beta}(\hat{\beta}_t-\beta^*)]^2[.5g_{t+1,\beta\beta}(\beta_t)(\hat{\beta}_t-\beta^*)^2]^2 + \\ & 4[g_{t+1} + g_{t+1,\beta}(\hat{\beta}_t-\beta^*)][.5g_{t+1,\beta\beta}(\tilde{\beta}_t)(\hat{\beta}_t-\beta^*)^2]^3. \end{aligned} \tag{AA4}$$

Now since $g_{t+1}$ has 8d moments, Markov's inequality implies that $P^{-1}\Sigma g^4_{t+1} = O_p(1)$. It then suffices to show that $P^{-1}\Sigma r_{t+1}$ is $o_p(1)$. To do so we will show that the absolute value of the fifth term on the rhs of (AA4) is $o_p(1)$, the other terms follow from similar arguments. Using the triangle and Holder inequalities it follows that

$$\begin{aligned} & |P^{-1}\Sigma 4[g_{t+1} + g_{t+1,\beta}(\hat{\beta}_t-\beta^*)][.5g_{t+1,\beta\beta}(\tilde{\beta}_t)(\hat{\beta}_t-\beta^*)^2]^3| \leq \\ & (\sup_t|\hat{\beta}_t-\beta^*|)^6(P^{-1}\Sigma g^4_{t+1})^{1/4}(P^{-1}\Sigma g^4_{t+1,\beta\beta}(\tilde{\beta}_t))^{3/4} + \\ & (\sup_t|\hat{\beta}_t-\beta^*|)^7(P^{-1}\Sigma g^4_{t+1,\beta})^{1/4}(P^{-1}\Sigma g^4_{t+1,\beta\beta}(\beta_t))^{3/4} \end{aligned}$$

From Lemma A1, $\sup_t|\hat{\beta}_t-\beta^*| = o_p(1)$ and by assumption 4 and Markov's inequality, $P^{-1}\Sigma g^4_{t+1} = O_p(1)$ and $P^{-1}\Sigma g^4_{t+1,\beta} = O_p(1)$. Since by assumption 3 and Markov's inequality, $P^{-1}\Sigma g^4_{t+1,\beta\beta}(\tilde{\beta}_t) = O_p(1)$, the result is established.

<u>Additional detail on the proof of Theorem 4.2(c):</u>

Expanding $\hat{\eta}_{t+\tau}$ as in part (b) of this proof, we have $\hat{S}_{ff} - \mathring{S}_{ff} = \Sigma^{P-1}_{j=-P+1}K(j/M)\{P^{-1}\Sigma^{T}_{t=R+j}w_{t+1-j}\}$ for $w_{t+1-j}$ defined in (AA3). Since the kernel is nonnegative, we can use the same inequalities as in part (b) to obtain

$$\begin{aligned} & |\Sigma^{P-1}_{j=-P+1}K(j/M)\{P^{-1}\Sigma^{T}_{t=R+j}w_{t+1-j}\}| \leq \Sigma^{P-1}_{j=-P+1}|K(j/M)|\{|\hat{\alpha}|(P^{-1}\Sigma\hat{g}^2_{t+1}\hat{v}^2_{t,t+\tau})^{1/2}(P^{-1}\Sigma\hat{g}^4_{t+1})^{1/2} + \\ & |\hat{\alpha}|(P^{-1}\Sigma\hat{g}^2_{t+1}\hat{v}^2_{t,t+\tau})^{1/2}(P^{-1}\Sigma\hat{g}^4_{t+1})^{1/2} + \hat{\alpha}^2(P^{-1}\Sigma\hat{g}^4_{t+1})^{1/2}(P^{-1}\Sigma\hat{g}^4_{t+1})^{1/2}\}. \end{aligned}$$

The bracketed term *on* the RHS of the previous inequality does not depend upon j and so it is unaffected by the outer summand. The RHS of the inequality is then less than or equal to

$$(M/P^{a})(M^{-1}\Sigma_{j=-P+1}^{P-1}|K(j/M)|)\{|P^{a}\hat{\alpha}|(P^{-1}\Sigma\hat{g}_{t+1}^{2}\hat{v}_{t,t+\tau}^{2})^{1/2}(P^{-1}\Sigma\hat{g}_{t+1}^{4})^{1/2} +$$
$$|P^{a}\hat{\alpha}|(P^{-1}\Sigma\hat{g}_{t+1}^{2}\hat{v}_{t,t+\tau}^{2})^{1/2}(P^{-1}\Sigma\hat{g}_{t+1}^{4})^{1/2} + (P^{a/2}\hat{\alpha})^{2}(P^{-1}\Sigma\hat{g}_{t+1}^{4})^{1/2}(P^{-1}\Sigma\hat{g}_{t+1}^{4})^{1/2}\}$$

where "a" is defined as in the statement of the theorem, $0<a<.5$. By assumption $(M/P^{a}) = O(1)$ and $(M^{-1}\Sigma_{j=-P+1}^{P-1}|K(j/M)|) \to \int_{-\infty}^{\infty}|K(x)|dx < \infty$. Notice also that by Theorem 4.1, $P^{a}\hat{\alpha} = o_{p}(1)$. That the RHS is $o_{p}(1)$ then follows from the same argument used in (b).

## Additional References


Hall, Peter and Christopher C. Heyde (1980): Martingale Limit Theory and Its Application, New York: Academic Press.

White, Halbert (1984): Asymptotic Theory for Econometricians, New York: Academic Press.

## References


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

Andrews, Donald W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica,* 59, 1465-1471.

Ashley, R., Granger, Clive W. J. and Richard Schmalansee (1980): "Advertising and Aggregate Consumption: An Analysis of Causality," Econometrica 48, 1149-1167.

Berger, Allen and Spencer Krane (1985): "The Informational Efficiency of Econometric Model Forecasts," *Review of* Economics and Statistics, 67, 128-134.

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

Clark, Todd (1997): "Finite-Sample Properties of Tests for Forecast Equivalence", manuscript, Federal Reserve Bank of Kansas City.

Davidson, Russell and James G. MacKinnon (1984): "Model Specification Tests Based on Artificial Linear Regressions," *International Economic Review, 25,* 485-502.

----------- (1989) : "Testing for Consistency Using Artificial Regressions," *Econometric Theory, 5, 363-384.*

Diebold, Francis X. and Robert S. Mariano (1996) : "Comparing Predictive Accuracy," Journal *of Business and* Economic Statistics, 13, 253-263.

Diebold, Francis X. and James Nason (1990): "Nonparametric Exchange Rate Prediction?", *Journal of International Economics, 28,* 315-322.

Fair, Ray C. and Robert Shiller (1990): "Comparing Information in Forecasts from Econometric Models," *American Economic* Review, 80, 375-389.

Hoffman, Dennis and Adrian Pagan (1989): "Practitioners Corner: Post-Sample Prediction Tests for Generalized Method of Moments Estimators," *Oxford Bulletin of Economics* and Statistics 51, 333-343.

Howrey, Philip E., Lawrence R. Klein and Michael D. McCarthy (1974) : "Notes on Testing the Predictive Performance of Econometric Models," *International* Economic Review, 15, 366-383.

Makridakis, Spiros et. al. (1982): "The Accuracy of Time Series Methods: The Results from a Forecasting Competition," *Journal of* Forecasting, 1, 111-153.

McCracken, Michael W. (1996): "Data Mining and Out of Sample Inference," manuscript, University of Wisconsin.

Meese, Richard A. and Kenneth 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 Differential Relation Over the Modern Floating-Rate Period ", Journal *of* Finance, 43, 933-948.

Mincer, Jacob and Victor Zarnowitz (1969): "The Evaluation of Economic Forecasts," in J. Mincer (ed.) Economic Forecasts and Expectations, New York: National Bureau of Economic Research.

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

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

Pagan, Adrian R. and Anthony D. Hall (1983): "Diagnostic Tests as Residual Analysis," *Econometric Reviews, 2, 159-218.*

Pagan, Adrian R. and G. William Schwert (1990) : "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics, 45, 267-290.*

West, Kenneth D. (1996): "Asymptotic Inference About Predictive Ability," Econometrica, 64, 1067-1084.

West, Kenneth D. and Dongchul Cho (1995): "The Predictive Ability of Several Models of Exchange Rate Volatility," Journal of Econometrics, 69, 367-391.

West, Kenneth D. and David W. Wilcox (1996): "A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model," *Journal of Business and Economic Statistics, 14, 281-293.*

White, Halbert (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica, 48, 817-838.*

Wooldridge, Jeffrey M. (1990): "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory, 6, 17-43.*

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

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

Table 1

Regressors for Four Common Tests, Linear Model

| (1) Test | (2) $\hat{g}_{t+1}$ | (3) $g_{2t+1}$ | (4) $\hat{g}_{t+1}$ |
|---|---|---|---|
| (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 $\beta^*$ 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 $\hat{g}$ (column (4)) is the vector of regressors. See sections 6 and 7 of the paper for more detail.

## 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 |

**Notes:**

1. 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 $\beta^*$.

2. The table assumes $\pi>0$. $\pi$ 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.

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

4. Panel D allows $Ev_{t+1}x_t\neq0$, as is typically the case in time series applications. If $Ev_{t+1}x_t=0$, no correction is needed, for any of the schemes, and whether or not $v_{t+1}$ is conditionally heteroskedastic.

## Table 3

### Size of Nominal .05 Tests, Mean Prediction Error

### A. Accounting for Error in Estimation of $\beta$*

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Recursive | 25 | .054 | .052 | .053 | .056 | .056 |
| | 50 | .053 | .057 | .051 | .057 | |
| | 100 | .046 | .049 | .054 | | |
| | 150 | .056 | .056 | | | |
| | 175 | .052 | | | | |
| 2.Rolling | 25 | .063 | .074 | .105 | .133 | .145 |
| | 50 | .053 | .063 | .063 | .072 | |
| | 100 | .048 | .051 | .058 | | |
| | 150 | .054 | .055 | | | |
| | 175 | .053 | | | | |
| 3.Fixed | 25 | .091 | .090 | .096 | .097 | .099 |
| | 50 | .069 | .074 | .075 | .077 | |
| | 100 | .058 | .060 | .064 | | |
| | 150 | .062 | .050 | | | |
| | 175 | .058 | | | | |

### B. Ignoring Error in Estimation of $\beta$*

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Rolling | 25 | .025 | .003 | .000 | 000 | .000 |
| | 50 | .043 | .021 | .002 | :000 | |
| | 100 | .046 | .044 | .021 | | |
| | 150 | .054 | .052 | | | |
| | 175 | .052 | | | | |
| 2.Fixed | 25 | .220 | .297 | .421 | .498 | .523 |
| | 50 | .129 | .195 | .293 | .354 | |
| | 100 | .081 | .121 | ,186 | | |
| | 150 | .078 | .106 | | | |
| | 175 | .073 | | | | |

Notes: 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,...,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 AI-A3 report the same, when the conventionally computed t-statistic is divided by the square root of $\hat{\lambda}$.

## Table 4

### Size of Nominal .05 Tests, Efficiency Test

### A. Accounting for Error in Estimation of $\beta^*$

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Recursive | 25 | .052 | .051 | .055 | .055 | .053 |
| | 50 | .038 | .043 | .043 | .046 | |
| | 100 | .038 | .040 | .045 | | |
| | 150 | .041 | .047 | | | |
| | 175 | .042 | | | | |
| 2.Rolling | 25 | .124 | .430 | .939 | .997 | .999 |
| | 50 | .045 | .070 | .232 | .468 | |
| | 100 | .036 | .043 | .059 | | |
| | 150 | .042 | .045 | | | |
| | 175 | .041 | | | | |
| 3.Fixed | 25 | .058 | .055 | .056 | .055 | .053 |
| | 50 | .040 | .038 | .035 | .031 | |
| | 100 | .039 | .041 | .042 | | |
| | 150 | .041 | .036 | | | |
| | 175 | .042 | | | | |

### B. Ignoring Error in Estimation of $\beta^*$

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Rolling | 25 | .059 | .072 | .152 | .331 | .450 |
| | 50 | .037 | .030 | .016 | .013 | |
| | 100 | .034 | .034 | .021 | | |
| | 150 | .042 | .041 | | | |
| | 175 | .040 | | | | |
| 2.Fixed | 25 | .158 | .234 | .355 | .434 | .456 |
| | 50 | .087 | .135 | .220 | .286 | |
| | 100 | .063 | .097 | .152 | | |
| | 150 | .060 | .083 | | | |
| | 175 | .057 | | | | |

Notes: 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, . . ..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 $y_t\hat{\beta}_t$ that 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 $\hat{A}$.

**Table 5**

**Size of Nominal .05 Tests, Encompassing Test**

**A. Accounting for Error in Estimation of $\beta^*$**

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Recursive | 25 | .113 | .148 | .185 | .201 | .196 |
| | 50 | .077 | .100 | .125 | .135 | |
| | 100 | [illegible] | | .084 | | |
| | 150 | [illegible] | .055 | | | |
| | 175 | .046 | | | | |
| 2.Rolling | 25 | .244 | .388 | .634 | .796 | .861 |
| | 50 | .151 | .232 | .346 | .450 | |
| | 100 | .090 | .122 | .171 | | |
| | 150 | .075 | .088 | | | |
| | 175 | .071 | | | | |
| 3.Fixed | 25 | .044 | .050 | .049 | .048 | .050 |
| | 50 | .049 | .049 | .053 | .049 | |
| | 100 | [illegible] | [illegible] | .048 | | |
| | 150 | | | | | |
| | 175 | .054 | | | | |

**B. Ignoring Error in Estimation of $\beta^*$**

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Rolling | 25 | .130 | .250 | .508 | .713 | .790 |
| | 50 | .084 | .122 | .211 | .312 | |
| | 100 | .059 | .078 | .105 | | |
| | 150 | .044 | .057 | | | |
| | 175 | .047 | | | | |
| 2.Fixed | 25 | .191 | .270 | .399 | .486 | .508 |
| | 50 | [illegible] | [illegible] | .240 | .300 | |
| | 100 | [illegible] | [illegible] | .133 | | |
| | 150 | .055 | .068 | | | |
| | 175 | .052 | | | | |

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

## Table 6

**Size of Nominal .05 Tests, Test for Zero First Order Serial Correlation**

**A. Accounting for Error in Estimation of $\beta^*$**

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Recursive | 25 | .059 | .060 | .061 | .061 | .061 |
| | 50 | .043 | .052 | .057 | .053 | |
| | 100 | .040 | .048 | .051 | | |
| | 150 | .045 | .054 | | | |
| | 175 | .045 | | | | |
| 2.Rolling | 25 | .119 | .209 | .405 | .584 | .663 |
| | 50 | .069 | .094 | .143 | .183 | |
| | 100 | .058 | .066 | .071 | | |
| | 150 | .054 | .058 | | | |
| | 175 | .057 | | | | |
| 3.Fixed | 25 | .044 | .050 | .049 | .048 | .050 |
| | 50 | .049 | .049 | .053 | .049 | |
| | 100 | .049 | .048 | .048 | | |
| | 150 | .049 | .051 | | | |
| | 175 | .053 | | | | |

**B. Ignoring Error in Estimation of $\beta^*$**

| Sampling Scheme | R | P: 25 | 50 | 100 | 150 | 175 |
|---|---|---|---|---|---|---|
| 1.Rolling | 25 | .034 | .022 | .027 | .052 | .067 |
| | 50 | .040 | .029 | .014 | .013 | |
| | 100 | .039 | .044 | .026 | | |
| | 150 | .045 | .048 | | | |
| | 175 | .045 | | | | |
| 2.Fixed | 25 | .214 | .319 | .447 | .512 | .546 |
| | 50 | .111 | .177 | .269 | .335 | |
| | 100 | .066 | .105 | .156 | | |
| | 150 | .062 | .086 | | | |
| | 175 | .058 | | | | |

Notes: 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,...,R+P-1, for the indicated values of P and R,. Panels Al, B1 and B2 report the fraction of the 5000 simulations in which the conventionally computed t-statistic on the coefficient on $\hat{v}_t$ that were greater than 1.96 in absolute value. Panels A2 and A3 report the same, when $y_t$ was included as a third regressor.

**Table 7**

**Size of Nominal .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 |
|---|---|---|---|---|
| .430 | .232 | .108 | .091 | .069 |

**B. (P/R)=4**

| P=100,R=25 | P=400,R=100 | P=800,R=200 |
|---|---|---|
| .939 | .409 | .229 |

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