Check for updates

Econometrica, Vol. 83, No. 6 (November, 2015), 2485–2505

EQUIVALENCE BETWEEN OUT-OF-SAMPLE FORECAST COMPARISONS AND WALD STATISTICS

BY PETER REINHARD HANSEN AND ALLAN TIMMERMANN[1]

We demonstrate the asymptotic equivalence between commonly used test statistics for out-of-sample forecasting performance and conventional Wald statistics. This equivalence greatly simplifies the computational burden of calculating recursive out-of-sample test statistics and their critical values. For the case with nested models, we show that the limit distribution, which has previously been expressed through stochastic integrals, has a simple representation in terms of χ^2-distributed random variables and we derive its density. We also generalize the limit theory to cover local alternatives and characterize the power properties of the test.

KEYWORDS: Out-of-sample forecast evaluation, nested models, testing.

1. INTRODUCTION

OUT-OF-SAMPLE TESTS OF PREDICTIVE ACCURACY are used extensively throughout economics and finance and are regarded by many researchers as the "ultimate test of a forecasting model" (Stock and Watson (2007, p. 571)). Such tests are frequently undertaken using the approach of West (1996), which accounts for the effect of recursive updating in parameter estimates. This approach can be used to test the null of equal predictive accuracy of two nonnested regression models evaluated at the probability limits of the estimated parameters (West (1996)), and for comparisons of nested models (McCracken (2007) and Clark and McCracken (2001, 2005)). The nested case gives rise to a test statistic whose limiting distribution (and, hence, critical values) depends on integrals of Brownian motion. The test is burdensome to compute and depends on nuisance parameters such as the relative size of the initial estimation sample versus the out-of-sample evaluation period.

This paper shows that a recursively generated out-of-sample test of equal predictive accuracy is asymptotically equivalent to one based on simple Wald statistics and documents that the equivalence is reliable in finite samples. Our result has four important implications. First, it simplifies calculation of the test statistics, which no longer requires recursively updated parameter estimates. Second, for the case with nested models, it greatly simplifies the computation of critical values, which has so far relied on numerical approximation to integrals of Brownian motion but now reduces to simple convolutions of chi-

[1]Valuable comments were received from Ulrich Mueller, Jim Stock, Frank Diebold, Silvia Gonçalves, Hashem Pesaran, two anonymous referees, and seminar participants at University of Chicago, University of Pennsylvania, USC, the Triangle Econometrics Seminar, UC Riverside, University of Cambridge, the conference "Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions" in Honor of Halbert L. White Jr., and the NBER/NSF Summer Institute 2011. The authors acknowledge support from CREATES funded by the Danish National Research Foundation.

© 2015 The Econometric Society DOI: 10.3982/ECTA10581

squared random variables. Third, our asymptotic results also cover the case with local alternatives, thus shedding new light on the power properties of the test. Fourth, our result provides a new interpretation of out-of-sample tests of equal predictive accuracy which we show are equivalent to simple parametric hypotheses and so could be tested with greater power using conventional test procedures.

The paper is organized as follows. Section 2 establishes the equivalence between the out-of-sample statistics and conventional Wald statistics for any pair of regression models. Section 3 focuses on the comparison of nested models and establishes the simplifications of the limit distribution for a test of equal predictive accuracy, while Section 4 concludes. Proofs of the results are provided in the Appendix. A companion Supplemental Material document (Hansen and Timmermann (2015b)), available on the web, contains additional details and simulation results.

2. THEORY

Consider the predictive regression model for an h-period forecast horizon

$$y_t = \beta' X_{t-h} + \varepsilon_t, \quad t = 1, \dots, n. \tag{1}$$

To avoid "look-ahead" biases, out-of-sample forecasts generated by the regression model (1) are commonly based on recursively estimated parameter values. This can be done by regressing y_s on X_{s-h}, for $s = 1, \dots, t$, resulting in least squares estimates $\hat{\beta}_t = (\sum_{s=1}^t X_{s-h} X'_{s-h})^{-1} \sum_{s=1}^t X_{s-h} y_s$, and using $\hat{y}_{t+h|t}(\hat{\beta}_t) = \hat{\beta}'_t X_t$ to forecast y_{t+h}.[2]

The resulting forecast can be compared to that from an alternative regression model that uses $\tilde{X}_{t-h}$ as a regressor:

$$y_t = \delta' \tilde{X}_{t-h} + \eta_t, \tag{2}$$

whose forecasts are given by $\tilde{y}_{t+h|t}(\hat{\delta}_t) = \hat{\delta}'_t \tilde{X}_t$, where $\hat{\delta}_t = (\sum_{s=1}^t \tilde{X}_{s-h} \tilde{X}'_{s-h})^{-1} \times \sum_{s=1}^t \tilde{X}_{s-h} y_s$. We do not specify how $\tilde{X}_t$ is related to X_t. In particular, the two models may be nested, nonnested, or overlapping. We let k and $\tilde{k}$ denote the dimension of X_t and $\tilde{X}_t$, respectively.

West (1996) proposed to judge the merits of a prediction model through its expected loss evaluated at the population parameters. Under mean squared error (MSE) loss, a test of equal predictive performance takes the form[3]

$$H_0 : \mathrm{E}\big[y_t - \hat{y}_{t|t-h}(\beta)\big]^2 = \mathrm{E}\big[y_t - \tilde{y}_{t|t-h}(\delta)\big]^2, \tag{3}$$

[2]We assume that initial values $X_{-1}, \dots, X_{-h+1}$, are observed.

[3]Another approach considers $\mathrm{E}[y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h})]^2$ which typically depends on t; see Giacomini and White (2006).

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where β and δ are the probability limits of $\hat{\beta}_n$ and $\hat{\delta}_n$, respectively, as $n \to \infty$. This and related hypotheses motivate a test statistic based on the out-of-sample MSE loss differential

$$\Delta\text{MSE}_n = \sum_{t=n_\rho+1}^{n} \left(y_t - \tilde{y}_{t|t-h}(\hat{\delta}_{t-h})\right)^2 - \left(y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h})\right)^2,$$

where n_ρ is the number of observations set aside for initial estimation of β and δ, while $t = n_\rho + 1, \dots, n$ is the out-of-sample period used for forecast evaluation. This is taken to be a fraction $\rho \in (0, 1)$ of the full sample, n, that is, $n_\rho = \lfloor n\rho \rfloor$ (the integer part of $n\rho$). Test statistics based on ΔMSE_n appear in many studies, including Diebold and Mariano (1995), West (1996), McCracken (2007), and Clark and McCracken (2014), in comparisons of nested, nonnested, and overlapping regression models.

Our first result compares the MSE loss of $\hat{y}_{t+h|t}(\hat{\beta}_t)$ to the corresponding loss from the very simple model that has no predictors, that is, $\tilde{y}_{t+h|t}(\hat{\delta}_t) = 0$. Although the scope of this result is obviously limited, this no-change forecast has featured prominently in testing the random walk model in finance and has also been used as a benchmark in macroeconomic forecasting. Moreover, results for the general case can be derived from this simple case. We will show that ΔMSE_n can be expressed in terms of two pairs of standard Wald statistics, with one pair being based on the full sample, $t = 1, \dots, n$, while the other is based on the initial estimation sample, $t = 1, \dots, n_\rho$. In the case with nested regression models, the result simplifies further in a way that allows us to express ΔMSE_n as the difference between two Wald statistics.

To prove this result, we need assumptions ensuring that the recursive least squares estimates, $\hat{\beta}_{t-h}$, $t = n_\rho + 1, \dots, n$, and related objects converge at conventional rates in a uniform sense. We make the following assumption, where $\|\cdot\|$ denotes the Frobenius norm, that is, $\|A\| = \sqrt{\text{tr}\{A'A\}}$ for any matrix A.

ASSUMPTION 1:

(i) *For some positive definite matrix*, Σ,

$$\sup_{r\in[0,1]} \left\| \frac{1}{n} \sum_{t=1}^{\lfloor nr \rfloor} X_{t-h}X'_{t-h} - r\Sigma \right\| = o_p(1). \tag{4}$$

(ii) *Let* $u_{n,t} = n^{-1/2}X_{t-h}\varepsilon_t$. *For some* $\Gamma_j \in \mathbb{R}^{k\times k}$, $j = 0, \dots, h-1$, *we have*

$$\sup_{r\in[0,1]} \left\| \sum_{t=1}^{\lfloor nr \rfloor} u_{n,t}u'_{n,t-j} - r\Gamma_j \right\| = o_p(1). \tag{5}$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The autocovariances of $\{X_{t-h}\varepsilon_t\}$ play an important role when $h > 1$. Define $\Omega = \sum_{j=-h+1}^{h-1} \Gamma_j$ and note that Ω is closely related to the long-run variance, $\Omega_\infty := \text{plim}_{n\to\infty} \frac{1}{n}\sum_{s,t=1}^{n} X_{s-h}\varepsilon_s\varepsilon_t X'_{t-h}$, whenever it is well-defined. The two are obviously equal when the higher-order autocovariances are all zero, which would correspond to a type of unpredictability of the forecast errors beyond the forecast horizon, h; this can be tested by inspecting the autocorrelations.

Next, define

$$U_n(r) = \sum_{t=1}^{\lfloor nr \rfloor} u_{n,t} = n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} X_{t-h}\varepsilon_t \quad \text{for} \quad r \in [0, 1],$$

so $U_n \in \mathbb{D}^k_{[0,1]}$, where $\mathbb{D}^k_{[0,1]}$ denotes the space of cadlag mappings from the unit interval to $\mathbb{R}^k$. In the canonical case, U_n will converge to a Brownian motion. The Brownian limit leads to additional simplifications regarding the limit distribution, which we detail in Section 3. For now, we only need to make the following high-level assumption on $U_n(\frac{t}{n})$, as the Brownian limit is not needed to establish the equivalence of test statistics.

ASSUMPTION 2: *Let $M_t = \frac{1}{t+h}\sum_{s=1}^{t} X_{s-h}X'_{s-h}$. Then as $n \to \infty$,*

$$\sup_{r\in[\rho,1]} \left\| \sum_{t=n_\rho+1}^{\lfloor nr \rfloor} U'_n\left(\frac{t-h}{n}\right)\left(M^{-1}_{t-h} - \Sigma^{-1}\right)u_{n,t} \right\| = o_p(1), \tag{6}$$

$$\sup_{r\in[\rho,1]} \left\| \frac{1}{n} \sum_{t=n_\rho+1}^{\lfloor nr \rfloor} U'_n\left(\frac{t-h}{n}\right)\left(M^{-1}_{t-h}X_{t-h}X'_{t-h}M^{-1}_{t-h} - \Sigma^{-1}\right)U_n\left(\frac{t-h}{n}\right) \right\| = o_p(1), \tag{7}$$

and $\sup_{r\in[0,1]} \|U_n(r)\| = O_p(1)$.

The convergences in (6) and (7) were obtained by Clark and McCracken (2001) under mixing and moment assumptions that guarantee a Brownian limit of U_n; see also Clark and McCracken (2000) and McCracken (2007, pp. 745–746).

2.1. *Comparison With No-Change Forecast*

Consider first the simple case where the forecasts from the regression model (1) are compared to the trivial forecast $\tilde{y}_{t|t-h} = 0$. Define the quadratic form statistic

$$S_n = \sum_{t=1}^{n} y_t X'_{t-h}\left(\sum_{t=1}^{n} X_{t-h}X'_{t-h}\right)^{-1} \sum_{t=1}^{n} X_{t-h}y_t.$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This is similar to the explained sum-of-squares in regression analysis—the difference being that the explanatory variables, X_{t-h}, are not demeaned.

THEOREM 1: *Given Assumptions 1 and 2,*

$$\sum_{t=n_\rho+1}^{n}\left[y_t^2-\left(y_t-\hat{y}_{t|t-h}(\hat{\beta}_{t-h})\right)^2\right]=S_n-S_{n_\rho}+\kappa\log\rho+o_p(1),$$

where $\kappa=\mathrm{tr}\{\Sigma^{-1}\Omega\}$.

Next, consider

$$W_n=\hat{\sigma}_\varepsilon^{-2}\hat{\beta}_n'\left(\sum_{t=1}^{n}X_{t-h}X_{t-h}'\right)\hat{\beta}_n, \tag{8}$$

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimator of σ_ε^2. This is a simple Wald statistic associated with the hypothesis $H_0:\beta=0$. Since $W_n=\hat{\sigma}_\varepsilon^{-2}S_n$, Theorem 1 shows that, aside from the scaling by $\hat{\sigma}_\varepsilon^{-2}$, the first two terms on the right hand side in Theorem 1 are closely related to conventional Wald statistics—one based on the full sample of n observations, the other based on the initial n_ρ observations.

Note that the Wald statistic in (8) is "homoscedastic" although we have not assumed the underlying processes to be homoscedastic. Theorem 1 shows that $\Delta\mathrm{MSE}_n$ is related to the "homoscedastic" Wald statistics for testing $H_0:\beta=0$, regardless of whether the underlying process is homoscedastic and regardless of whether $\beta=0$ or not. As the reader may recall, if the underlying process is heteroscedastic, then, under the null hypothesis ($\beta=0$) and standard regularity conditions, $W_n\xrightarrow{d}\sum_{i=1}^{k}\lambda_i\chi_i^2$ as $n\to\infty$, where $\lambda_1,\ldots,\lambda_k$ are the eigenvalues of $\sigma_\varepsilon^{-2}\Sigma^{-1}\Omega_\infty$ and $\chi_1^2,\ldots,\chi_k^2$ are independent χ^2-distributed random variables with one degree of freedom; see, for example, White (1994, Theorem 8.10). Another interesting relation to notice is that if $\Omega=\Omega_\infty$, these eigenvalues are related to the constant in Theorem 1, κ, as

$$\kappa=\sigma_\varepsilon^2\sum_{i=1}^{k}\lambda_i.$$

The expression in Theorem 1 suggests the following estimator of κ: $\hat{\kappa}(\rho)=[\sum_{t=n_\rho+1}^{n}y_t^2-(y_t-\hat{y}_{t|t-h}(\hat{\beta}_{t-h}))^2-S_n+S_{n_\rho}]/\log\rho$. This estimator can be combined with a consistent estimator of σ_ε^2, to inspect whether the homoscedasticity assumption, which implies $k=\kappa(\rho)/\sigma_\varepsilon^2$ for all $\rho\in(0,1)$, is valid.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2. *Comparison of Arbitrary Pairs of Regression-Based Forecasts*

Next, consider general comparisons of pairs of regression models that could be nested, nonnested, or overlapping. Analogously to the definitions for model (1), introduce objects for model (2), σ_η^{-2}, $\tilde{\Sigma}$, $\tilde{\Omega}$, $\tilde{\kappa} = \mathrm{tr}\{\tilde{\Sigma}^{-1}\tilde{\Omega}\}$, and define

$$\tilde{S}_n = \sum_{t=1}^{n} y_t \tilde{X}'_{t-h} \left(\sum_{t=1}^{n} \tilde{X}_{t-h} \tilde{X}'_{t-h} \right)^{-1} \sum_{t=1}^{n} \tilde{X}_{t-h} y_t.$$

To simplify the exposition, we write $\tilde{y}_{t|t-h}$ and $\hat{y}_{t|t-h}$ in place of $\tilde{y}_{t|t-h}(\hat{\delta}_{t-h})$ and $\hat{y}_{t|t-h}(\hat{\beta}_{t-h})$.

COROLLARY 1: *Suppose that Assumptions 1 and 2 hold for both models. Then*

$$\sum_{t=n_\rho+1}^{n} (y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2 = S_n - S_{n_\rho} - (\tilde{S}_n - \tilde{S}_{n_\rho}) + (\kappa - \tilde{\kappa}) \log \rho + o_p(1). \tag{9}$$

The corollary shows that the difference in the MSE of the two regression models can be expressed in terms of linear combinations of two pairs of quadratic form statistics—one based on the full sample, the other based on the initial estimation sample—that test $\beta = 0$ and $\delta = 0$, respectively. This result holds regardless of the values of β and δ.

The equivalence stated by Corollary 1 is demonstrated by the scatter plots in Figure 1, where the expression based on the S-statistics on the right hand side of (9) is plotted against $\Delta\mathrm{MSE}_n$ for a number of data generating processes (DGPs). Additional simulation results are presented in the Supplemental Material; see Hansen and Timmermann (2015b).

2.3. *Nested Regression Models*

Sharper results can be established for the special case in which one of the regression models is nested by the other. This case arises when $\tilde{X}_t = X_{1t}$, where $X_t = (X'_{1,t}, X'_{2,t})'$ with $X_{1t} \in \mathbb{R}^{\tilde{k}}$ and $X_{2t} \in \mathbb{R}^q$, so that $k = \tilde{k} + q$. We decompose β accordingly, that is, $\beta = (\beta'_1, \beta'_2)'$. The case with nested models was studied by McCracken (2007), who considered the test statistic

$$T_n = \frac{\sum_{t=n_\rho+1}^{n} (y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2}{\hat{\sigma}_\varepsilon^2}, \tag{10}$$

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimator of $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{t+h})$.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

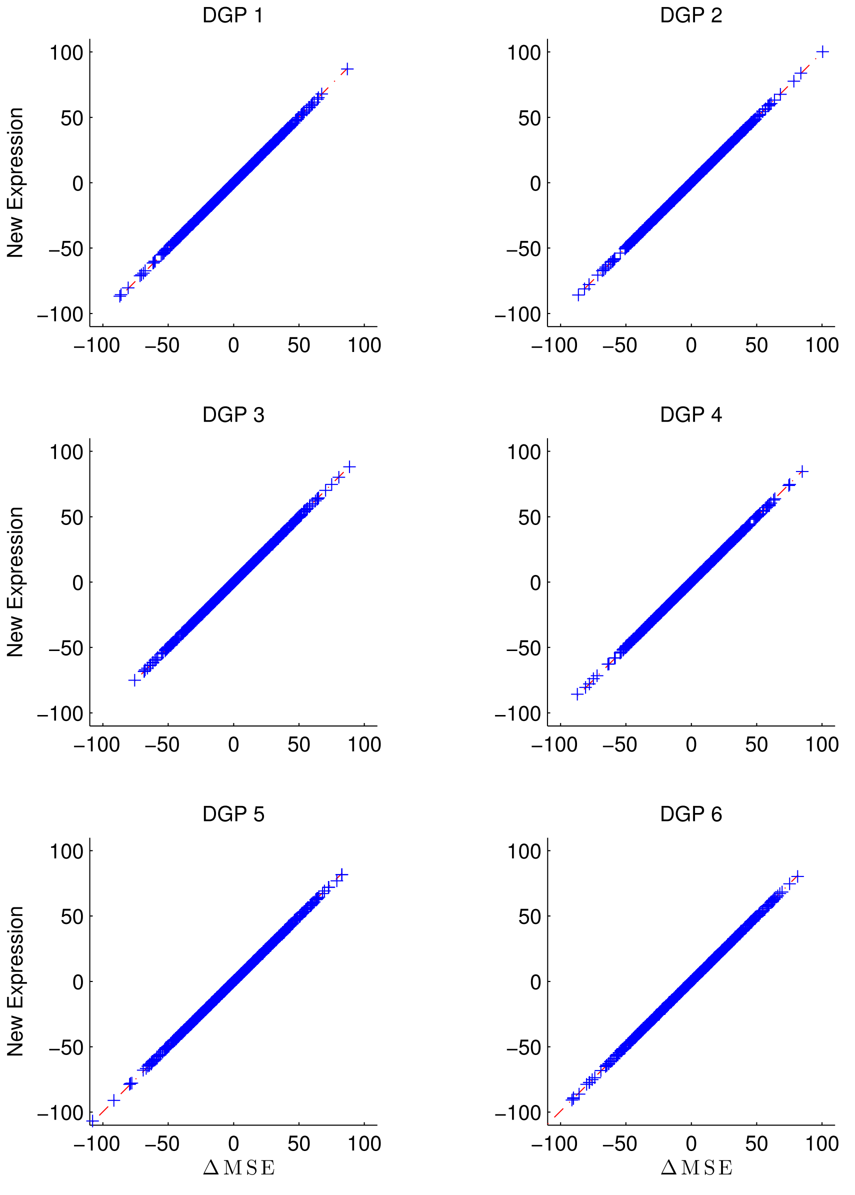

FIGURE 1.—Scatter plots of the terms on the right hand side in (9) (excluding the $o_p(1)$ term and using $q = \kappa - \tilde{\kappa}$) against ΔMSE_n. The plots are based on 1,000 simulations and assume $n = 500$ and $\rho = 0.5$. The six DGPs are based on those in Clark and McCracken (2005) and include cases with homoscedastic (DGP 1 and DGP 2), heteroscedastic (DGP 3 and DGP 4), and serially dependent errors (DGP 5 and DGP 6). See the Supplemental Material for details.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Corollary 1 is directly applicable to this statistic. However, we can use a well-known identity for Wald statistics involving nested hypotheses to simplify the expression. To this end, we partition Σ into blocks

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \bullet \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where Σ_{22} is a $q \times q$ matrix. Define $\check{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. This matrix is positive definite as a consequence of Assumption 1. Next, define the auxiliary variables

$$Z_t = X_{2,t} - \Sigma_{21}\Sigma_{11}^{-1}X_{1,t}, \quad t+h = 1, \dots, n.$$

The variable Z_t captures that part of X_{2t} that is orthogonal to X_{1t}. Also define

$$\check{\Omega} = \sum_{j=-h+1}^{h-1} \check{\Gamma}_j, \quad \text{with} \quad \check{\Gamma}_j = \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} Z_{t-h}\varepsilon_t\varepsilon_{t-j}Z'_{t-h-j}.$$

The residuals obtained from regressing $X_{2,t-h}$ on $X_{1,t-h}$ are given by

$$Z_{n,t-h} = X_{2,t-h} - \sum_{t=1}^{n} X_{2,t-h}X'_{1,t-h}\left(\sum_{t=1}^{n} X_{1,t-h}X'_{1,t-h}\right)^{-1} X_{1,t-h},$$
$$t = 1, \dots, n.$$

These can be used to compute the statistic

$$\check{S}_n = \sum_{t=1}^{n} y_t Z'_{n,t-h}\left(\sum_{t=1}^{n} Z_{n,t-h}Z'_{n,t-h}\right)^{-1} \sum_{t=1}^{n} Z_{n,t-h}y_t.$$

$\check{S}_n$ measures that part of the variation in y_t that is explained by $X_{2,t-h}$, but unexplained by $X_{1,t-h}$. It is straightforward to verify that $\check{W}_n = \check{S}_n/\hat{\sigma}_\varepsilon^2$ is a conventional (homoscedastic) Wald statistic associated with the hypothesis $\beta_2 = 0$.

THEOREM 2: *Given Assumptions 1 and 2, the out-of-sample test statistic in (10) can be written as*

$$T_n = \check{W}_n - \check{W}_{n_\rho} + \sigma_\varepsilon^{-2}\check{\kappa}\log\rho + o_p(1),$$

where $\check{\kappa} = \kappa - \tilde{\kappa}$, *which simplifies to* $\check{\kappa} = \operatorname{tr}\{\check{\Sigma}^{-1}\check{\Omega}\}$ *if* $\beta_2 = n^{-1/2}b$ *with* $b \in \mathbb{R}^q$ *fixed*.

The complex out-of-sample test statistic for equal predictive accuracy, T_n, depends on sequences of recursive estimates. It is surprising that this is equivalent to the difference between two Wald statistics, one using the full sample, the other using the subsample $t = 1, \dots, n_\rho$.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE I

FINITE SAMPLE CORRELATION OF TEST STATISTICS ($n = 200$)[a]

ρ	$\pi = \frac{1-\rho}{\rho}$	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6
0.833	0.2	0.962	0.972	0.959	0.954	0.969	0.955
0.714	0.4	0.975	0.980	0.971	0.963	0.971	0.956
0.625	0.6	0.977	0.979	0.975	0.960	0.973	0.943
0.556	0.8	0.979	0.98	0.977	0.955	0.971	0.947
0.500	1.0	0.980	0.978	0.975	0.96	0.969	0.941
0.455	1.2	0.980	0.976	0.975	0.954	0.967	0.935
0.417	1.4	0.979	0.974	0.976	0.954	0.962	0.934
0.385	1.6	0.978	0.973	0.974	0.948	0.959	0.936
0.357	1.8	0.977	0.973	0.975	0.948	0.959	0.926
0.333	2.0	0.975	0.972	0.975	0.948	0.958	0.927

[a] Finite sample correlations between T_n and the expression based on Wald statistics in Theorem 2. The sample size is $n = 200$, but the simulation design is otherwise identical to that in Figure 1. The results are based on 10,000 replications. The parameter, $\pi = (1 - \rho)/\rho$, is the notation used in Clark and McCracken (2005).

The results in Theorems 1 and 2 are asymptotic in nature, but the relationship is very reliable in finite samples, as is evident from the simulations reported in Table I which use $n = 200$ observations. Thus the correlations reported in Table I are for out-of-sample statistics that are based on sums with as few as 34 terms. The main source of differences between the recursive MSE differences and the Wald statistics is estimation error in $\hat{\sigma}_\varepsilon^2$, because the two Wald statistics employ sample variances based on different sample sizes, n_ρ and n, respectively. In fact, the correlations between the expressions on the two sides of Equation (9) in Corollary 1 exceed 0.995 across each of the simulation experiments shown in Figure 1. Additional simulation results are presented in the Supplemental Material.

3. SIMPLIFIED LIMIT DISTRIBUTION FOR NESTED COMPARISONS

This section turns to the limit distribution of T_n for comparisons of nested models. The equivalence between the test statistics established above holds without detailed distributional assumptions. Under standard assumptions used to establish the limit distribution of T_n, the equivalence between T_n and Wald statistics has interesting implications for the limit distribution and results in a simplified expression.

For the asymptotic limit results, we shall rely on the following additional assumption that is known to hold under standard regularity conditions used in this literature, such as those in Hansen (1992) (mixing) or in De Jong and Davidson (2000) (near-epoch).

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ASSUMPTION 3:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} Z_{t-h}\varepsilon_t \Rightarrow \check{\Omega}_\infty^{1/2}B(r) \quad on \quad \mathbb{D}^q_{[0,1]},$$

where $B(r)$ is a standard q-dimensional Brownian motion.

Assumption 3 requires that certain linear combinations of $U_n(r)$ converge to a Brownian motion with covariance matrix $\check{\Omega}_\infty$, which is defined analogously to Ω_∞ as the long-run variance of $\{Z_{t-h}\varepsilon_t\}$.

For the special case where $h = 1$ and forecast errors are homoscedastic, McCracken (2007) showed that the asymptotic distribution of T_n is given as a convolution of q independent random variables, each with a distribution of $2\int_\rho^1 u^{-1}B(r)\,\mathrm{d}B(r) - \int_\rho^1 u^{-2}B(r)^2\,\mathrm{d}r$. Results for the case with $h > 1$ and heteroscedastic errors were derived in Clark and McCracken (2005).

The relation between T_n and Wald statistics implies that existing expressions for the limit distribution of T_n can be greatly simplified and generalized to cover the case with local alternatives. To this end, we need to introduce Q, defined by $Q'\Lambda Q = \Xi$, $Q'Q = I$, where $\Xi = \sigma_\varepsilon^{-2}\check{\Omega}_\infty^{1/2}\check{\Sigma}^{-1}\check{\Omega}_\infty^{1/2}$ and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_q)$.

THEOREM 3: *Suppose that Assumptions 1–3 hold. Let $\beta_2 = cn^{-1/2}b$ for some vector, b, normalized by $b'\check{\Sigma}b = \sigma_\varepsilon^2\kappa$, and $c \in \mathbb{R}$. Define $a = b'\check{\Sigma}\check{\Omega}_\infty^{-1/2}Q' \in \mathbb{R}^q$. Then*

$$T_n \xrightarrow{d} \sum_{i=1}^q \lambda_i\Big[2\int_\rho^1 r^{-1}B_i(r)\,\mathrm{d}B_i(r) - \int_\rho^1 r^{-2}B_i^2(r)\,\mathrm{d}r + (1-\rho)c^2 + 2ca_i\big\{B_i(1) - B_i(\rho)\big\}\Big], \tag{11}$$

where $B = (B_1, \dots, B_q)'$ is a standard q-dimensional Brownian motion. Moreover, the limit distribution is identical to that of

$$\sum_{i=1}^q \lambda_i\big[B_i^2(1) - \rho^{-1}B_i^2(\rho) + \log\rho + (1-\rho)c^2 + 2a_ic\big\{B_i(1) - B_i(\rho)\big\}\big].$$

The contributions of Theorem 3 are twofold. First, the theorem establishes the asymptotic distribution of T_n under local alternatives ($c \neq 0$), thereby generalizing the results in Clark and McCracken (2005) who showed results for $c = 0$.[4] Second, it simplifies the expression of the limit distribution from one

[4]The expression in Clark and McCracken (2005) involves a $q \times q$ matrix of nuisance parameters. For the case $c = 0$, this expression was simplified by Stock and Watson (2003) to that in (11).

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

involving stochastic integrals to one involving (dependent) $\chi^2(1)$-distributed random variables, $B_i^2(1)$ and $\rho^{-1}B_i^2(\rho)$. Below, we further simplify the limit distribution under the null hypothesis to an expression involving differences between two independent χ^2-distributed random variables.

THEOREM 4: *Let B be a univariate standard Brownian motion. The distribution of $2\int_\rho^1 r^{-1}B\,\mathrm{d}B - \int_\rho^1 r^{-2}B^2\,\mathrm{d}r$ is identical to that of $\sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log\rho$, where $Z_i \sim i.i.d.\ N(0,1)$.*

Theorems 3 and 4 show that the limit distribution of $T_n/\sqrt{1-\rho}$ is invariant to ρ under the null hypothesis, whereas the noncentrality parameter, $\sqrt{1-\rho}c^2$, and hence the power of the test, is decreasing in ρ. This property of the test might suggest choosing ρ as small as possible to maximize power, although such a conclusion is unwarranted because the result relied on ρ being strictly greater than zero to ensure that $(n^{-1}\sum_{t=1}^{n_\rho} X_{t-h}X_{t-h})^{-1}$ is bounded in probability and $\hat{\beta}_t$ is well behaved. Still, comparing the test with $\rho = 0.75$ to the test using $\rho = 0.25$, the noncentrality parameter reveals that the former amounts to the same loss in asymptotic power as discarding $(1 - 1/\sqrt{3}) \simeq 42\%$ of the sample, a substantial loss of power.

The asymptotic results in Theorems 1–4 take the sample split, ρ, to be fixed, but could be generalized to be uniform in ρ over some interval $(a,b) \subset [0,1]$. Such results could be used to develop a test that is robust to mining over the sample split, analogously to the results derived in Rossi and Inoue (2012).

Because the distribution is expressed in terms of two independent χ^2-distributed random variables, in the homoscedastic case where $\lambda_1 = \cdots = \lambda_q = 1$ it is possible to obtain relatively simple closed-form expressions for the limit distribution of T_n:

THEOREM 5: *The density of $\sum_{j=1}^q[2\int_\rho^1 r^{-1}B_j(r)\,\mathrm{d}B_j(r) - \int_\rho^1 r^{-2}B_j(r)^2\,\mathrm{d}r]$ is given by*

$$f_q(x) = \frac{1}{\sqrt{1-\rho}2^q\Gamma\left(\frac{q}{2}\right)^2} e^{-|x-q\log\rho|/(2\sqrt{1-\rho})} \times \int_0^\infty \left(u\left(u + \frac{|x-q\log\rho|}{\sqrt{1-\rho}}\right)\right)^{q/2-1} e^{-u}\,\mathrm{d}u.$$

For $q = 1$ and $q = 2$, the expression simplifies to

$$f_1(x) = \frac{1}{2\pi\sqrt{1-\rho}}\mathrm{K}_0\left(\frac{|x-\log\rho|}{2\sqrt{1-\rho}}\right) \quad \textit{and}$$

$$f_2(x) = \frac{1}{4\sqrt{1-\rho}}\exp\left(-\frac{|x-2\log\rho|}{2\sqrt{1-\rho}}\right),$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

respectively, where $\mathrm{K}_0(x) = \int_0^\infty \frac{\cos(xt)}{\sqrt{1+t^2}}\,\mathrm{d}t$ *is the modified Bessel function of the second kind.*

For $q = 2$, the limit distribution is simply the noncentral Laplace distribution. The density for $q = 1$ is also readily available since $\mathrm{K}_0(x)$ is implemented in standard software.

4. CONCLUSION

We show that a test statistic that is widely used for out-of-sample comparisons of regression-based forecasts is asymptotically equivalent to a linear combination of Wald statistics. This equivalence greatly simplifies the computation of the test statistic based on recursively estimated parameters, regardless of whether the models being compared are nested, overlapping, or nonnested.

For the case where the forecasts are based on nested regression models, we provide further simplifications. In this case, the test statistics can be expressed as the difference between two Wald statistics of the same null—one using the full sample, the other using a subsample. Moreover, in the nested case, the limit distribution can be expressed as a difference between two independent χ^2-distributions and convolutions thereof. We also derive local power properties for the test which establish that the power of the test is decreasing in the sample split fraction, ρ.

These results raise serious questions about testing the stated null hypothesis for nested comparisons through out-of-sample forecasting performance. Subtracting a subsample Wald statistic from the full sample Wald statistic dilutes the power of the test. Moreover, the test statistic, T_n, is not robust to heteroscedasticity, which causes nuisance parameters to show up in its limit distribution. In contrast, the conventional full-sample Wald test can easily be adapted to the heteroscedastic case by using a robust estimator for the asymptotic variance of $\hat{\beta}_{2,n}$.

While tests of equal out-of-sample forecasting performance are not well suited for testing simple parametric hypotheses such as (3), there may be other reasons for using such tests. Specifically, interest may lie in testing which model (or method) produces the best forecasting performance. Whether one forecasting model outperforms another model relies on the extent to which the forecasts are influenced by parameter estimation error, and entails a different null hypothesis than (3); see, for example, Hendry (1997) and Giacomini and White (2006). Out-of-sample forecast evaluation provides insights into the effect of estimation error on "real-time" forecasting performance in a manner that is not reflected in conventional full-sample tests. Moreover, out-of-sample methods can be useful in multiple comparison problems because spurious evidence of significance due to "data mining" (overfitting) is less likely to arise for out-of-sample than for in-sample comparisons; see Hansen and Timmermann (2015a). In both cases, our analysis shows that such properties of out-of-

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sample tests come at the expense of power. Our results help econometricians better decide which tests to use in a particular situation.

APPENDIX: PROOFS

We first prove a number of auxiliary results. To simplify the exposition, we will occasionally write $\sum_t$, $\sup_t$, and $\sup_r$ as short for $\sum_{t=n_\rho+1}^{n}$, $\sup_{n_\rho+1\leq t\leq n}$, and $\sup_{r\in[\rho,1]}$, respectively.

LEMMA A.1: *Let a_t and b_t be matrices whose dimensions are such that the product, $a_t b_t$, is well-defined. Then, for $l \leq m \leq n$,*

$$\sum_{t=m+1}^{n} a_t b_t = \sum_{t=m}^{n-1}(a_t - a_{t+1})\sum_{s=l}^{t} b_s + a_n \sum_{s=l}^{n} b_s - a_m \sum_{s=l}^{m} b_s$$

and

$$\sum_{t=m+1}^{n} a_t(b_t - b_{t-1}) = \sum_{t=m}^{n-1}(a_t - a_{t+1})b_t + a_n b_n - a_m b_m.$$

PROOF:

$$\begin{aligned}\sum_{t=m+1}^{n} a_t b_t &= \sum_{t=m+1}^{n} a_t\left(\sum_{s=l}^{t} b_s - \sum_{s=l}^{t-1} b_s\right)\\ &= \sum_{t=m+1}^{n} a_t \sum_{s=l}^{t} b_s - \sum_{t=m+1}^{n} a_t \sum_{s=l}^{t-1} b_s\\ &= \sum_{t=m+1}^{n} a_t \sum_{s=l}^{t} b_s - \sum_{t=m}^{n-1} a_{t+1} \sum_{s=l}^{t} b_s\\ &= \sum_{t=m}^{n-1}(a_t - a_{t+1})\sum_{s=l}^{t} b_s + a_n \sum_{s=l}^{n} b_s - a_m \sum_{s=l}^{m} b_s.\end{aligned}$$

The second result follows by

$$\begin{aligned}\sum_{t=m+1}^{n} a_t(b_t - b_{t-1}) &= \sum_{t=m}^{n-1}(a_t - a_{t+1})\sum_{s=l}^{t}(b_s - b_{s-1})\\ &\quad + a_n \sum_{s=l}^{n}(b_s - b_{s-1}) - a_m \sum_{s=l}^{m}(b_s - b_{s-1})\end{aligned}$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$= \sum_{t=m}^{n-1}(a_t - a_{t+1})(b_t - b_{l-1}) + a_n(b_n - b_{l-1})$$
$$- a_m(b_n - b_{l-1}),$$

because all terms involving b_{l-1} cancel out. *Q.E.D.*

LEMMA A.2: *Suppose that* $\sup_{\rho \le r \le 1} \|\frac{1}{n}\sum_{t=1}^{\lfloor rn \rfloor}(\zeta_{n,t} - \zeta)\| = o_p(1)$ *and that* $a_{n,t}$, $t = n_\rho, \ldots, n$, *is such that* $a_{n,n}$ *and* $\sum_{t=n_\rho}^{n-1}|a_{n,t} - a_{n,t+1}|$ *are both* $O(1)$. *Then*

$$\frac{1}{n}\sum_{t=n_\rho+1}^{n} a_{n,t}\zeta_{n,t} = \left(\frac{1}{n}\sum_{t=n_\rho+1}^{n} a_{n,t}\right)\zeta + o_p(1).$$

PROOF: Let $\tilde{\zeta}_{n,t} = (\zeta_{n,t} - \zeta)/n$ and apply Lemma A.1 with $l = 1$, $m = n_\rho$,

$$\sum_{t=n_\rho+1}^{n} a_{n,t}\tilde{\zeta}_{n,t} = \sum_{t=n_\rho}^{n-1}(a_{n,t} - a_{n,t+1})\sum_{s=1}^{t}\tilde{\zeta}_{n,s} + a_{n,n}\sum_{s=1}^{n}\tilde{\zeta}_{n,s} - a_{n,n_\rho}\sum_{s=1}^{n_\rho}\tilde{\zeta}_{n,s}.$$

The second term is by assumption $O(1)o_p(1) = o_p(1)$, and similarly for the last term because

$$|a_{n,n_\rho}| = \left|a_{n,n} + \sum_{t=n_\rho}^{n-1}(a_{n,t} - a_{n,t+1})\right| \le |a_{n,n}| + \sum_{t=n_\rho}^{n-1}|a_{n,t} - a_{n,t+1}|$$
$$= O(1) + O(1).$$

The first term is bounded by $\sum_{t=n_\rho}^{n-1}|a_{n,t} - a_{n,t+1}|\sup_{n_\rho \le t < n}\|\sum_{s=1}^{t}\tilde{\zeta}_{n,s}\|$, which is $O(1)o_p(1) = o_p(1)$ by assumption. This completes the proof. *Q.E.D.*

In the present context, we often have $a_{n,t} \simeq (\frac{t}{n})^b$, for some $b \in \mathbb{R}$, with $a_{n,t}$ monotonic in t so that $\sum_{t=n_\rho}^{n-1}|a_{n,t} - a_{n,t+1}| = |a_{n,n_\rho} - a_{n,n}| = O(1)$, and

$$\frac{1}{n}\sum_{t=n_\rho+1}^{n} a_{n,t} = \int_\rho^1 r^b\,\mathrm{d}r = \begin{cases} 1 - \rho^{b+1} & \text{if } b \ne -1, \\ -\log\rho & \text{if } b = -1. \end{cases}$$

This is illustrated in the following corollary.

COROLLARY 2: *Given* (5) *of Assumption* 1, *we have*

$$\frac{1}{n}\sum_{t=n_\rho+1}^{n}\frac{n}{t}\varepsilon_{t-j}X'_{t-h-j}\Sigma^{-1}X_{t-h}\varepsilon_t = -\gamma_j\log\rho + o_p(1),$$

where $\gamma_j = \mathrm{tr}\{\Sigma^{-1}\Gamma_j\}$.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PROOF: We have

$$\frac{1}{n}\sum_t \varepsilon_{t-j}X'_{t-h-j}\Sigma^{-1}X_{t-h}\varepsilon_t = \operatorname{tr}\left\{\Sigma^{-1}\frac{1}{n}\sum_t X_{t-h}\varepsilon_t\varepsilon_{t-j}X'_{t-h-j}\right\}$$
$$= \operatorname{tr}\left\{\Sigma^{-1}\sum_t u_{n,t}u'_{n,t-j}\right\}$$
$$= \operatorname{tr}\left\{\Sigma^{-1}\left(\Gamma_j + o_p(1)\right)\right\},$$

where the last equality follows by Assumption 1(ii). The result now follows by Lemma A.2 with $\zeta_{n,t} = \varepsilon_{t-j}X'_{t-h-j}\Sigma^{-1}X_{t-h}\varepsilon_t = n\operatorname{tr}\{\Sigma^{-1}u_{n,t}u'_{n,t-h}\}$, $\zeta = \gamma_j = \operatorname{tr}\{\Sigma^{-1}\Gamma_j\}$, and $a_{n,t} = \frac{n}{t}$, which meets the conditions of Lemma A.2 and is such that $\frac{1}{n}\sum_{n_\rho+1}^{n} a_{n,t} = \int_\rho^1 r^{-1}\,\mathrm{d}r + o(1) = -\log\rho + o(1)$. *Q.E.D.*

LEMMA A.3: *Suppose* $U_t = U_{t-1} + u_t \in \mathbb{R}^q$ *and let* M *be a symmetric* $q\times q$ *matrix. Then* $2U'_{t-1}Mu_t = U'_tMU_t - U'_{t-1}MU_{t-1} - u'_tMu_t$.

PROOF: Rearranging the nonvanishing terms in

$$U'_tMU_t - U'_{t-1}MU_{t-1} = (U_{t-1}+u_t)'M(U_{t-1}+u_t)' - U'_{t-1}MU_{t-1}$$

and using $u'_tMU_{t-1} = U'_{t-1}Mu_t$ yields the result. *Q.E.D.*

LEMMA A.4: *The following identity holds for* ΔMSE:

$$\sum_{t=n_\rho+1}^{n} y_t^2 - \left(y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h})\right)^2 = A + 2B + 2C - D,$$

where

$$A = \sum_t \beta' X_{t-h}X'_{t-h}\beta,$$
$$B = \beta'\sum_t X_{t-h}\varepsilon_t,$$
$$C = \sum_t(\hat{\beta}_{t-h}-\beta)'X_{t-h}\varepsilon_t,$$
$$D = \sum_t(\hat{\beta}_{t-h}-\beta)'X_{t-h}X'_{t-h}(\hat{\beta}_{t-h}-\beta).$$

PROOF: Let $\xi_t = \beta'X_t$ and $\vartheta_t = (\hat{\beta}_t - \beta)'X_t$, so that $y_{t+h} = \varepsilon_{t+h} + \beta'X_t = \varepsilon_{t+h} + \xi_t$ and $y_{t+h} - \hat{y}_{t+h|t} = \varepsilon_{t+h} + \beta'X_t - \hat{\beta}'_tX_t = \varepsilon_{t+h} - \vartheta_t$. It follows that

$$y^2_{t+h} - (y_{t+h} - \hat{y}_{t+h|t})^2 = (\varepsilon_{t+h} + \xi_t)^2 - (\varepsilon_{t+h} - \vartheta_t)^2$$
$$= \xi_t^2 + 2\xi_t\varepsilon_{t+h} + 2\vartheta_t\varepsilon_{t+h} - \vartheta_t^2,$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

which are the terms in the sums that define A, B, C, and D, respectively. *Q.E.D.*

PROOF OF THEOREM 1: From the identity of Lemma A.4, the theorem follows by showing that

$$A+2B+2C-D=S_n-S_{n_\rho}+\kappa\log\rho+o_p(1).$$

We first consider C, which is the most interesting term, and simplify by writing $U_{n,t}$ in place of $U_n(\frac{t}{n})$. It follows from (6) and Lemma A.2 with $a_{n,t}=\frac{n}{t}$ that

$$\text{(A.1)}\qquad C=\sum_{t=n_\rho+1}^{n}\frac{n}{t}U'_{n,t-h}\Sigma^{-1}u_{n,t}+o_p(1)$$

$$=\sum_{t=n_\rho+1}^{n}\frac{n}{t}U'_{n,t-1}\Sigma^{-1}u_{n,t}-\sum_{j=1}^{h-1}\sum_{t=n_\rho+1}^{n}\frac{n}{t}u'_{n,t-j}\Sigma^{-1}u_{n,t}+o_p(1).$$

Applying Lemma A.3 to $2U'_{n,t-1}\Sigma^{-1}u_{n,t}$, we find

$$\text{(A.2)}\qquad 2C=\sum_{t=n_\rho+1}^{n}\frac{n}{t}\big(U'_{n,t}\Sigma^{-1}U_{n,t}-U'_{n,t-1}\Sigma^{-1}U_{n,t-1}\big)$$

$$-\sum_{j=-h+1}^{h-1}\sum_{t=n_\rho+1}^{n}\frac{n}{t}u'_{n,t-|j|}\Sigma^{-1}u_{n,t}+o_p(1).$$

From Corollary 2, we have that $-\sum_{t=n_\rho+1}^{n}\frac{n}{t}u'_{n,t-j}\Sigma^{-1}u_{n,t}=\gamma_j\log\rho+o_p(1)$, $j=1,\dots,h-1$, and since

$$\kappa=\operatorname{tr}\{\Sigma^{-1}\Omega\}$$

$$=\sum_{j=-h+1}^{h-1}\operatorname{tr}\left\{\Sigma^{-1}\sum_{t=1}^{n}u_tu'_{t-h}\right\}+o_p(1)=\sum_{j=-h+1}^{h-1}\gamma_j+o_p(1),$$

and $\gamma_j=\gamma_{-j}$, it follows that the contribution from the last term in (A.2) is simply $\kappa\log\rho+o_p(1)$.

For the remainder of (A.2), we have

$$\text{(A.3)}\qquad \sum_{t=n_\rho+1}^{n}\frac{n}{t}\big(U'_{n,t}\Sigma^{-1}U_{n,t}-U'_{n,t-1}\Sigma^{-1}U_{n,t-1}\big)$$

$$=U'_{n,n}\Sigma^{-1}U_{n,n}-\frac{n}{n_\rho}U'_{n,n_\rho}\Sigma^{-1}U_{n,n_\rho}+\frac{1}{n}\sum_{t=n_\rho}^{n-1}\frac{n}{t}\frac{n}{t+1}U'_{n,t}\Sigma^{-1}U_{n,t},$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where we used the second identity of Lemma A.1 with $a_t = \frac{n}{t}$ and $b_t = U'_{n,t}\Sigma^{-1}U_{n,t}$, so that

$$a_t - a_{t+1} = \frac{n}{t} - \frac{n}{t+1} = \frac{n}{t(t+1)}.$$

The last term of (A.3) offsets the contributions from $-D$, because

$$\begin{aligned} D &= \frac{1}{n}\sum_{t=n_\rho+1}^{n}\left(\frac{n}{t}\right)^2 U'_{n,t-h}M_{t-h}^{-1}X_{t-h}X'_{t-h}M_{t-h}^{-1}U_{n,t-h} \\ &= \frac{1}{n}\sum_{t=n_\rho+1}^{n}\left(\frac{n}{t}\right)^2 U'_{n,t-h}\Sigma^{-1}U_{n,t-h} + o_p(1), \end{aligned}$$

by (7) and Lemma A.2 with $a_{n,t} = (\frac{n}{t})^2$. That the two terms offset each other relies on

$$\frac{1}{n}\sum_{t=n_\rho+1}^{n}\left[\frac{n}{t}\frac{n}{t+1} - \left(\frac{n}{t+h}\right)^2\right]U'_{n,t}\Sigma^{-1}U_{n,t} = o_p(1),$$

using that $\sup_t |U'_{n,t}\Sigma^{-1}U_{n,t}| = O_p(1)$ by the last part of Assumption 2 and that $\frac{1}{n}\sum_{t=n_\rho+1}^{n}|\frac{n}{t}\frac{n}{t+1} - (\frac{n}{t+h})^2| = O(n^{-1})$.

Next, $A + 2B$ equals

$$\beta'\sum_{t=1}^{n}X_{t-h}X'_{t-h}\beta - \beta'\sum_{t=1}^{n_\rho}X_{t-h}X'_{t-h}\beta + 2n^{1/2}\beta'U_{n,n} - 2n^{1/2}\beta'U_{n,n_\rho}.$$

With $S_m = \hat{\beta}'_m[\sum_{t=1}^{m}X_{t-h}X'_{t-h}]\hat{\beta}_m = (\hat{\beta}_m - \beta + \beta)'[\sum_{t=1}^{m}X_{t-h}X'_{t-h}](\hat{\beta}_m - \beta + \beta)$, we have

$$\begin{aligned} (S_n - S_{n_\rho}) &= U'_{n,n}\Sigma^{-1}U_{n,n} - \frac{n}{n_\rho}U'_{n,n_\rho}\Sigma^{-1}U_{n,n_\rho} + o_p(1) \\ &\quad + \beta'\sum_{t=n_\rho+1}^{n}X_{t-h}X'_{t-h}\beta + 2n^{1/2}\beta'(U_{n,n} - U_{n,n_\rho}). \end{aligned} \tag{A.4}$$

Q.E.D.

PROOF OF COROLLARY 1: This follows from writing

$$\begin{aligned} &(y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2 \\ &\quad = \{y_t^2 - (y_t - \hat{y}_{t|t-h})^2\} - \{y_t^2 - (y_t - \tilde{y}_{t|t-h})^2\}, \end{aligned}$$

where y_t^2 is the squared prediction error from the simple auxiliary (zero) forecast. *Q.E.D.*

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

PROOF OF THEOREM 2: The first result follows from Corollary 1 and the identity $Q_n = \tilde{Q}_n + \check{Q}_n$. Let

$$A = \begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{pmatrix}.$$

Consider

$$\kappa = \mathrm{tr}\{\Sigma^{-1}\Omega\} = \mathrm{tr}\{(A'\Sigma A)^{-1} A'\Omega A\} = \mathrm{tr}\left\{\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \check{\Sigma} \end{pmatrix} A'\Omega A\right\},$$

so that

$$\kappa = \mathrm{tr}\{\Sigma_{11}^{-1}\Omega_{11}\} + \mathrm{tr}\{\check{\Sigma}^{-1}\Omega_{22\cdot 1}\},$$

where $\Omega_{22\cdot 1} = (-\Sigma_{21}\Sigma_{11}^{-1}, I)\Omega(-\Sigma_{21}\Sigma_{11}^{-1}, I)'$. Now recall that $\Omega = \sum_{j=-h+1}^{h-1} \Gamma_j$ where $\Gamma_j = \mathrm{plim}\, \frac{1}{n}\sum_{t=1}^{n} X_{t-h}\varepsilon_t\varepsilon_{t-j}X'_{t-h-j}$, so that the terms that make up $\Omega_{22\cdot 1}$ are given from $\mathrm{plim}\, \frac{1}{n}\sum_{t=1}^{n} Z_{t-h}\varepsilon_t\varepsilon_{t-j}Z'_{t-h-j}$, proving that $\Omega_{22\cdot 1} = \check{\Omega}$. Hence, the result holds provided that

$$\mathrm{plim}\, \frac{1}{n}\sum_{t=1}^{n} X_{1,t-h}\varepsilon_t\varepsilon_{t-j}X'_{1,t-h-j} = \mathrm{plim}\, \frac{1}{n}\sum_{t=1}^{n} X_{1,t-h}\eta_t\eta_{t-j}X'_{1,t-h-j},$$
$$j = 0, \dots, h-1,$$

which would imply $\Omega_{11} = \tilde{\Omega}$. Since $\eta_t = \varepsilon_t + \beta_2' Z_{t-h}$, the result follows when $\beta_2 = n^{-1/2}b$, with b fixed. *Q.E.D.*

PROOF OF THEOREM 3: We establish the result by showing that the two expressions for the limit distribution are identical. Then we derive the limit distribution for the difference between the two Wald statistics and use their relation with T_n.

Consider $F(r) = \frac{1}{r}B^2(r) - \log r$ (for $r > 0$). By Ito stochastic calculus,

$$\mathrm{d}F = \frac{\partial F}{\partial B}\,\mathrm{d}B + \left[\frac{\partial F}{\partial u} + \frac{1}{2}\frac{\partial^2 F}{(\partial B)^2}\right]\mathrm{d}u$$
$$= \frac{2}{r}B\,\mathrm{d}B - \frac{1}{r^2}B^2\,\mathrm{d}r,$$

so $\int_\rho^1 \frac{2}{r}B\,\mathrm{d}B - \int_\rho^1 \frac{1}{r^2}B^2\,\mathrm{d}r = \int_\rho^1 \mathrm{d}F(r)$ equals $F(1) - F(\rho) = B^2(1) - \log 1 - B^2(\rho)/\rho + \log\rho$.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Next, consider $\check{W}_n - \check{W}_{n_\rho}$ where, analogously to (A.4), $\hat{\sigma}_\varepsilon^2(\check{W}_n - \check{W}_{n_\rho})$ is equal to

$$\begin{aligned}\check{S}_n - \check{S}_{n_\rho} &= \check{U}'_{n,n}\check{\Sigma}^{-1}\check{U}_{n,n} - \frac{n}{n_\rho}\check{U}'_{n,n_\rho}\check{\Sigma}^{-1}\check{U}_{n,n_\rho} + o_p(1)\\ &\quad + \beta_2'\sum_{t=n_\rho+1}^{n} Z_{t-h}Z'_{t-h}\beta_2 + 2n^{1/2}\beta_2'(\check{U}_{n,n} - \check{U}_{n,n_\rho})\\ &= B(1)'\check{\Omega}_\infty^{1/2}\check{\Sigma}^{-1}\check{\Omega}_\infty^{1/2}B(1) - \rho^{-1}B(\rho)'\check{\Omega}_\infty^{1/2}\check{\Sigma}^{-1}\check{\Omega}_\infty^{1/2}B(\rho)\\ &\quad + (1-\rho)c^2 b'\check{\Sigma}b + 2cb'\check{\Omega}_\infty^{1/2}\big[B(1) - B(\rho)\big] + o_p(1).\end{aligned}$$

Under Assumption 3, we have $\check{U}_{n,\lfloor nr\rfloor} = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} Z_{t-h}\varepsilon_t \Rightarrow \check{\Omega}_\infty^{1/2}B(r)$.

Now, define $\tilde{B}(r) = QB(r)$, another q-dimensional standard Brownian motion, and use that $\sigma_\varepsilon^{-2}b'\Sigma_{zz}b = \kappa$ to arrive at

$$\begin{aligned}&\tilde{B}(1)'\Lambda\tilde{B}(1) - \rho^{-1}\tilde{B}(\rho)'\Lambda\tilde{B}(\rho) + (1-\rho)c^2\kappa\\ &\quad + 2c\sigma_\varepsilon^{-2}b'\Omega^{1/2}Q'\big[\tilde{B}(1) - \tilde{B}(\rho)\big]\\ &= \sum_{i=1}^{q}\lambda_i\big[\tilde{B}_i^2(1) - \rho^{-1}\tilde{B}_i^2(\rho) + (1-\rho)c^2 + 2ca_i\big[\tilde{B}(1) - \tilde{B}(\rho)\big]\big],\end{aligned}$$

where we used that $\sigma_\varepsilon^{-2}b'\check{\Omega}_\infty^{1/2}Q' = b'\check{\Sigma}\check{\Omega}_\infty^{-1/2}\sigma_\varepsilon^{-2}\check{\Omega}_\infty^{1/2}\check{\Sigma}^{-1}\check{\Omega}_\infty^{1/2}Q' = b'\check{\Sigma}\check{\Omega}_\infty^{-1/2} \times \Xi Q' = b'\check{\Sigma}\check{\Omega}_\infty^{-1/2}Q'\Lambda = (a_1\lambda_1, \ldots, a_q\lambda_q)$. Since $\tilde{B}$ and B are identically distributed, the limit distribution may be expressed in terms of B instead of $\tilde{B}$. *Q.E.D.*

Proof of Theorem 4: Let $B(r)$ be a standard one-dimensional Brownian motion and define $U = \frac{B(1)-B(\rho)}{\sqrt{1-\rho}}$ and $V = \frac{B(\rho)}{\sqrt{\rho}}$, so that $B(1) = \sqrt{1-\rho}U + \sqrt{\rho}V$. Note that U and V are independent standard Gaussian random variables. Express the random variable $B^2(1) - B^2(\rho)/\rho$ as a quadratic form:

$$\begin{aligned}&(\sqrt{1-\rho}U + \sqrt{\rho}V)^2 - V^2\\ &= \begin{pmatrix}U\\V\end{pmatrix}'\begin{pmatrix}1-\rho & \sqrt{\rho(1-\rho)}\\ \sqrt{\rho(1-\rho)} & \rho-1\end{pmatrix}\begin{pmatrix}U\\V\end{pmatrix},\end{aligned}$$

and decompose the 2×2 symmetric matrix into $Q'\Lambda Q$, where $\Lambda = \mathrm{diag}(\sqrt{1-\rho}, -\sqrt{1-\rho})$ (the eigenvalues) and

$$Q = \frac{1}{\sqrt{2}}\begin{pmatrix}\sqrt{1+\sqrt{1-\rho}} & \sqrt{1-\sqrt{1-\rho}}\\ -\sqrt{1-\sqrt{1-\rho}} & \sqrt{1+\sqrt{1-\rho}}\end{pmatrix},$$

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

so that $Q'Q = I$. Then the expression simplifies to $\sqrt{1-\rho}(Z_1^2 - Z_2^2)$ where $Z = Q(U, V)' \sim N_2(0, I)$. *Q.E.D.*

PROOF OF THEOREM 5: Let $Z_{1,i}, Z_{2,i}$, $i = 1, \dots, q$ be i.i.d. $N(0, 1)$, so that $X = \sum_{i=1}^q Z_{1,i}^2$ and $Y = \sum_{i=1}^q Z_{2,i}^2$ are both χ_q^2-distributed and independent. The distribution is given by the convolution

$$\sum_{i=1}^{q}\bigl[\sqrt{1-\rho}\bigl(Z_{1,i}^2 - Z_{2,i}^2\bigr) + \log\rho\bigr] = \sqrt{1-\rho}(X - Y) + q\log\rho.$$

To derive the distribution of $S = X - Y$, where X and Y are independent χ_q^2-distributed random variables, note that the density of a χ_q^2 is

$$\psi_q(u) = 1_{\{u\geq 0\}}\frac{1}{2^{q/2}\Gamma\left(\frac{q}{2}\right)}u^{q/2-1}e^{-u/2}.$$

We are interested in the convolution of X and $-Y$, whose density is given by

$$\begin{aligned}
f_q(s) &= \int 1_{\{u\geq 0\}}\psi_q(u)1_{\{u-s\geq 0\}}\psi_q(u-s)\,\mathrm{d}u \\
&= \int_{0\vee s}^{\infty}\psi_q(u)\psi_q(u-s)\,\mathrm{d}u, \\
&= \int_{0\vee s}^{\infty}\frac{1}{2^{q/2}\Gamma\left(\frac{q}{2}\right)}u^{q/2-1}e^{-u/2}\frac{1}{2^{q/2}\Gamma\left(\frac{q}{2}\right)}(u-s)^{q/2-1}e^{-(u-s)/2}\,\mathrm{d}u \\
&= \frac{1}{2^q\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{q}{2}\right)}e^{s/2}\int_{0\vee s}^{\infty}\bigl(u(u-s)\bigr)^{q/2-1}e^{-u}\,\mathrm{d}u.
\end{aligned}$$

For $s < 0$, the density is $2^{-q}\Gamma(\frac{q}{2})^{-2}e^{s/2}\int_0^\infty (u(u-s))^{q/2-1}e^{-u}\,\mathrm{d}u$. Using the symmetry about zero, we arrive at the expression

$$f_q(s) = \frac{1}{2^q\Gamma\left(\frac{q}{2}\right)^2}e^{-|s|/2}\int_0^\infty\bigl(u\bigl(u+|s|\bigr)\bigr)^{q/2-1}e^{-u}\,\mathrm{d}u.$$

When $q = 1$, this simplifies to $f_1(s) = \frac{1}{2\pi}\mathrm{K}_0(\frac{|s|}{2})$, where $\mathrm{K}_k(x)$ denotes the modified Bessel function of the second kind. For $q = 2$, the expression for the density reduces to the simpler expression, $f_2(s) = \frac{1}{4}e^{-|s|/2}$, which is the density of the Laplace distribution with scale parameter 2. *Q.E.D.*

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

CLARK, T. E., AND M. W. MCCRACKEN (2000): "Not-for-Publication Appendix to 'Tests of Equal Forecast Accuracy and Encompassing for Nested Models'." [2488]

——— (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110. [2485,2488]

——— (2005): "Evaluating Direct Multi-Step Forecasts," *Econometric Reviews*, 24, 369–404. [2485,2491,2493,2494]

——— (2014): "Tests of Equal Forecast Accuracy for Overlapping Models," *Journal of Applied Econometrics*, 29, 415–430. [2487]

DE JONG, R. M., AND J. DAVIDSON (2000): "The Functional Central Limit Theorem and Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory*, 16, 621–642. [2493]

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263. [2487]

GIACOMINI, R., AND H. WHITE (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578. [2486,2496]

HANSEN, B. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500. [2493]

HANSEN, P. R., AND A. TIMMERMANN (2015a): "Discussion of 'Comparing Predictive Accuracy, Twenty Years Later'," *Journal of Business & Economic Statistics*, 33, 17–21. [2496]

——— (2015b): "Supplement to 'Equivalence Between Out-of-Sample Forecast Comparisons and Wald Statistics'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA10581. [2486,2490]

HENDRY, D. F. (1997): "The Econometrics of Macroeconomic Forecasting," *The Economic Journal*, 107, 1330–1357. [2496]

MCCRACKEN, M. W. (2007): "Asymptotics for Out-of-Sample Tests of Granger Causality," *Journal of Econometrics*, 140, 719–752. [2485,2487,2488,2490,2494]

ROSSI, B., AND A. INOUE (2012): "Out-of-Sample Forecast Tests Robust to the Choice of Window Size," *Journal of Business & Economic Statistics*, 30, 432–453. [2495]

STOCK, J. H., AND M. W. WATSON (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 61, 788–829. [2494]

——— (2007): *Introduction to Econometrics* (Second Ed.). Reading: Addison-Wesley. [2485]

WEST, K. D. (1996): "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084. [2485-2487]

WHITE, H. (1994): *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press. [2489]

Dept. of Economics, European University Institute, Via della Piazzuola 43, Firenze, FI 50133, Italy, Dept. of Economics, University of North Carolina, Chapel Hill, 107 Gardner Hall, CB 3305, Chapel Hill, NC 27599-3305, U.S.A., and CREATES; peter.hansen@eui.eu

and

Rady School of Management and Dept. of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0553, U.S.A. and CREATES; atimmermann@ucsd.edu.

Manuscript received February, 2012; final revision received February, 2015.

14680262, 2015, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/ECTA10581 by Duke University Libraries, Wiley Online Library on [04/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License