

# Equivalence Between Out-of-Sample Forecast Comparisons and Wald Statistics\*

Peter Reinhard Hansen

Allan Timmermann

European University Institute and CREATES

UCSD and CREATES

July 3, 2013

## Abstract

We demonstrate the equivalence between a commonly used out-of-sample test of equal predictive accuracy and the difference between two Wald statistics. This equivalence greatly simplifies the computational burden of calculating recursive out-of-sample tests and evaluating their critical values. Next, we show that the limit distribution, which has previously been expressed through stochastic integrals, has a simple representation in terms of  $\chi^2$ -distributed random variables and we derive the density of the limit distribution. We also generalize the limit theory to cover local alternatives and characterize the power properties of the test. Our results shed new light on the test and establish certain weaknesses associated with using out-of-sample forecast comparison tests to conduct inference about nested regression models.

*Keywords:* Out-of-Sample Forecast Evaluation, Nested Models, Testing.

*JEL Classification:* C12, C53, G17

---

\*Valuable comments were received from Ulrich Mueller, Jim Stock, Frank Diebold, two anonymous referees and seminar participants at University of Pennsylvania, the Triangle Econometrics Seminar, UC Riverside, University of Cambridge, the UCSD conference in Honor of Halbert White, and the NBER/NSF Summer Institute 2011. The authors acknowledge support from CREATES funded by the Danish National Research Foundation.

# 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), McCracken (2007) and Clark and McCracken (2001, 2005) 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 nested regression models evaluated at the probability limits of the estimated parameters and 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 equivalent to the difference between two simple Wald statistics based on the full sample and the initial estimation sample, respectively. Our result has four important implications. First, it greatly simplifies calculation of the critical values of the test statistic which has so far relied on numerical approximation to integrals of Brownian motion but now reduces to simple convolutions of chi-squared random variables. Second, our result simplifies computation of the test statistic itself which no longer requires recursively updated parameter estimates. Third, we greatly simplify the expressions of the asymptotic distribution and derive new results that cover local alternatives, thus shedding 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.

## 2 Theory

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

$$y_t = \beta_1' X_{1,t-h} + \beta_2' X_{2,t-h} + \varepsilon_t, \quad t = 1, \dots, n \quad (1)$$

where  $X_{1t} \in \mathbb{R}^k$  and  $X_{2t} \in \mathbb{R}^q$ . Also, define  $X_t = (X_{1,t}', X_{2,t}')'$  and  $\beta = (\beta_1', \beta_2')'$ .

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  $X_{s-h} = (X'_{1,s-h}, X'_{2,s-h})'$ , for  $s = 1, \dots, t$ , resulting in the least squares estimate  $\hat{\beta}_t = (\hat{\beta}'_{1t}, \hat{\beta}'_{2t})'$ , and using  $\hat{y}_{t+h|t}(\hat{\beta}_t) = \hat{\beta}'_{1t}X_{1t} + \hat{\beta}'_{2t}X_{2t}$  to forecast  $y_{t+h}$ . The resulting forecast can be compared to that of a smaller (nested) regression model,

$$y_t = \delta' X_{1,t-h} + \eta_t,$$

whose forecasts are given by  $\tilde{y}_{t+h|t}(\hat{\delta}_t) = \hat{\delta}'_t X_{1t}$ , where  $\hat{\delta}_t = \left( \sum_{s=1}^t X_{1,s-h} X'_{1,s-h} \right)^{-1} \sum_{s=1}^t X_{1,s-h} y_s$ .<sup>1</sup>

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, this suggests testing<sup>2</sup>

$$H_0 : E[y_t - \hat{y}_{t|t-h}(\beta)]^2 = E[y_t - \tilde{y}_{t|t-h}(\delta)]^2. \quad (2)$$

McCracken (2007) considered a test of this null based on the test statistic

$$T_n = \frac{\sum_{t=n\rho+1}^n (y_t - \tilde{y}_{t|t-h}(\hat{\delta}_{t-h}))^2 - (y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h}))^2}{\hat{\sigma}_\varepsilon^2}, \quad (3)$$

where  $\hat{\sigma}_\varepsilon^2$  is a consistent estimator of  $\sigma_\varepsilon^2 = \text{var}(\varepsilon_{t+h})$  and  $n_\rho$  is the number of observations set aside for the initial estimation of  $\beta$ . This is taken to be a fraction  $\rho \in (0, 1)$  of the full sample,  $n$ , i.e.,  $n_\rho = \lfloor n\rho \rfloor$ . Assuming homoskedastic forecast errors and  $h = 1$ , 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  $2 \int_\rho^1 u^{-1} B(u) dB(u) - \int_\rho^1 u^{-2} B(u)^2 du$ . Results for the case with  $h > 1$  and heteroskedastic errors were derived in Clark and McCracken (2005).

We will show that the test statistic,  $T_n$ , amounts to taking the difference between two Wald statistics, both testing the same null  $H_0 : \beta_2 = 0$ , but based on the full sample versus the initial estimation sample, respectively. To prove this result, we make an assumption which ensures that the least squares estimators and related objects converge at conventional rates in a uniform sense. For any matrix (including vectors and scalars) we use the notation  $\|A\| = \max_{i,j} |A_{ij}|$ .

**Assumption 1.** *For some positive definite matrix,  $\Sigma$ , we have*

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

<sup>1</sup>We assume that the initial values  $X_{-1}, \dots, X_{-h+1}$  are observed.

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

$$\sup_{u \in [0,1]} \left\| \sum_{t=1}^{\lfloor nu \rfloor} X_{1,t-h} \eta_t \right\| = O_p(n^{1/2}) \quad \text{and} \quad \sup_{u \in [0,1]} \left\| \sum_{t=1}^{\lfloor nu \rfloor} X_{t-h} \varepsilon_t \right\| = O_p(n^{1/2}).$$

Assumption 1 ensures that recursive estimates satisfy that  $\sup_t \|\hat{\beta}_{t-h} - \beta\|$  and  $\sup_t \|\hat{\delta}_{t-h} - \delta\|$  are  $O_p(n^{-1/2})$ , where the supremum is taken over  $t = n_\rho + 1, \dots, n$ .

For convenience we express  $\Sigma$ , implicitly defined by Assumption 1, as follows

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

where the blocks in  $\Sigma$  refer to  $X_{1t}$  and  $X_{2t}$ , respectively. Define the auxiliary regression variable

$$Z_t = X_{2t} - \Sigma_{21} \Sigma_{11}^{-1} X_{1t},$$

which captures the part of  $X_{2t}$  that is orthogonal to  $X_{1t}$ , and note that  $\Sigma_{zz} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$  is positive definite.

The autocovariances of  $\{Z_{t-h} \varepsilon_t\}$  play an important role when  $h > 1$ . We make the following assumption about these and the long-run variance of  $Z_{t-h} \varepsilon_t$ .

**Assumption 2.** For some  $\Gamma_j \in \mathbb{R}^{q \times q}$ ,  $j = 0, \dots, h-1$ , we have

$$\sup_u \left\| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} Z_{t-h} \varepsilon_t \varepsilon_{t-j}' Z_{t-h-j}' - u \Gamma_j \right\| = o_p(1),$$

where  $\Omega = \sum_{j=-h+1}^{h-1} \Gamma_j$ , and  $\Omega := \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{s,t=1}^n Z_{s-h} \varepsilon_s \varepsilon_t' Z_{t-h}'$  is positive definite.

The last part of Assumption 2 imposes a type of unpredictability of the forecast errors beyond the forecast horizon,  $h$ ; this is easily tested by inspecting the autocorrelations of  $Z_{t-h} \varepsilon_t$ .

The null hypothesis  $H_0$  in (2) is equivalent to  $H'_0 : \beta_2 = 0$  which can be tested with conventional tests. To this end, consider the Wald statistic based on the first  $m$  observations,

$$W_m = m \hat{\beta}'_{2m} \left[ \hat{\sigma}_\varepsilon^2 \hat{\Sigma}_{zz}^{-1} \right]^{-1} \hat{\beta}_{2m},$$

where  $\hat{\sigma}_\varepsilon^2$  and  $\hat{\Sigma}_{zz}$  are consistent estimators of  $\sigma_\varepsilon^2$  and  $\Sigma_{zz}$ , respectively. This statistic is based on a ‘‘homoskedastic’’ estimator of the asymptotic variance, which causes the eigenvalues of  $\Xi = \sigma_\varepsilon^{-2} \Sigma_{zz}^{-1} \Omega$ ,  $\lambda_1, \dots, \lambda_q$ , to appear in the limit distribution. Specifically,  $W_m \xrightarrow{d} \sum_{i=1}^q \lambda_i \chi_{(1)}^2$

under the null hypothesis; see, e.g., White (1994, theorem 8.10). The reason we focus on the Wald statistic based on a “homoskedastic” variance estimator will be clear from our main result which appears in the following theorem.

**Theorem 1.** *Given Assumptions 1 and 2, the out-of-sample test statistic in (3) can be written as*

$$T_n = W_n - W_{n_\rho} + \kappa \log \rho + o_p(1),$$

where  $\kappa = \sum_{i=1}^q \lambda_i$ , in the following three cases: (i)  $\beta_2 = 0$  (null), (ii)  $\beta_2 = n^{-1/2}b$  (local alternative), and (iii)  $\beta_2 = b$  (fixed alternative) with  $k = 0$ , where  $b \in \mathbb{R}^q/\{0\}$  is fixed. When  $k > 0$  and  $\beta_2 = b$ , the remainder is  $O_p(1)$ .

Note that we have not assumed the underlying processes to be homoskedastic. The theorem shows that  $T_n$  is related to the “homoskedastic” Wald statistics, regardless of the underlying process being homoskedastic or not.

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  $t = 1, \dots, n_\rho$ .

The remainder term in Theorem 1 is  $o_p(1)$  except in the case where  $k > 0$  and  $\beta_2 = b \in \mathbb{R}^q/\{0\}$ . This may seem surprising but can be traced to a relationship between three Wald statistics. Let  $W''$  and  $W'$  denote the Wald statistics testing  $\beta = 0$  vs  $\beta \in \mathbb{R}^{k+q}$  and  $\beta = 0$  vs  $\beta_1 \in \mathbb{R}^k, \beta_2 = 0$ , respectively, i.e.,  $W''_m = \hat{\sigma}_\varepsilon^{-2} \hat{\beta}'_m [\sum_{t=1}^m X_{t-h} X'_{t-h}] \hat{\beta}_m$  and  $W'_m = \hat{\sigma}_\eta^{-2} \hat{\delta}'_m [\sum_{t=1}^m X_{1,t-h} X'_{1,t-h}] \hat{\delta}_m$ . Under the null hypothesis (and local alternatives) it is well known that  $W'' = W' + W + o_p(1)$ , provided standard regularity conditions hold. However, this identity does not hold under fixed alternatives unless  $k = 0$ , because  $k = 0$  trivially implies  $W'' = W$ . A more general, albeit less transparent, expression than that in Theorem 1 is stated in the following result.

**Theorem 2.** *Suppose that Assumption 1 holds while Assumption 2 holds for both  $X_{t-h}\varepsilon_t$  and  $X_{1,t-h}\eta_t$  (in place of  $Z_{t-h}\varepsilon_t$ ), and let  $\lambda''_i, i = 1, \dots, k + q$ , and  $\lambda'_j, j = 1, \dots, k$ , denote the resulting eigenvalues. Further, suppose that (i)  $\beta_2 = 0$ , (ii)  $\beta_2 = n^{-1/2}b$ , or (iii)  $\beta_2 = b$ , with  $b \in \mathbb{R}^q/\{0\}$  fixed. Then the out-of-sample test statistic in (3) can be written as*

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

where  $\kappa'' = \sum_{i=1}^{k+q} \lambda''_i$  and  $\kappa' = \sum_{i=1}^k \lambda'_i$ .

Table 1: Finite Sample Correlation of Test Statistics ( $n = 200$ )

$\rho$	$\pi = \frac{1-\rho}{\rho}$	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

Finite sample correlations between  $T_n$  and the expression based on Wald statistics in Theorem 1, for  $n = 200$ . The simulation design is based on Clark and McCracken (2005). DGP-1 and DGP-2 assume i.i.d. (DGP-1) and serially correlated (DGP-2) processes with homoskedastic, serially uncorrelated errors ; DGPs 3-4 assume heteroskedastic errors;, DGP-5 allows for serial correlation in the errors, while DGP-6 allows for both serial correlation and heteroskedasticity in the errors. The results are based on 10,000 replications.  $\pi = (1 - \rho)/\rho$  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 1 and Figure 1. The former is based on just  $n = 200$  observations and the latter is for  $n = 500$ . Thus the correlations reported in Table 1 are for out-of-sample statistics that are based on sums with as few as 67 terms. The simulations are computed using a design where the null hypothesis holds, so that the remainder term is  $o_p(1)$ . Correlations may be smaller under fixed alternatives where the remainder term is  $O_p(1)$ , albeit in this case both  $T_n$  and  $W_n - W_{n\rho}$  diverge at rate  $n$ , thereby dominating the  $O_p(1)$  term. For a general expression that is reliable under both the null and alternative, one can instead use the expression of Theorem 2 that involves four Wald statistics. Additional simulation results and details are available in an online appendix.

The equivalence between these test statistics holds without detailed distributional assumptions. This equivalence has interesting implications for the limit distributions which rely on the following additional assumption that typically holds under standard regularity conditions used in this literature, such as those in Hansen (1992) (mixing) or in De Jong and Davidson (2000) (near-epoch).

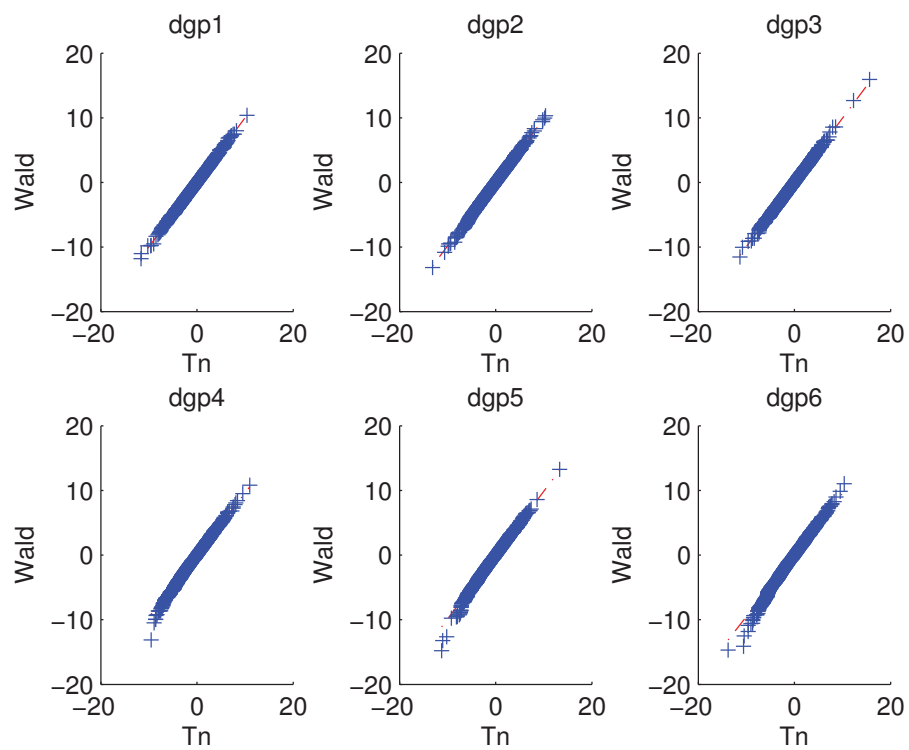


Figure 1: Q-Q plot for the Wald-based statistic against  $T_n$ , in a simulation study where  $n = 500$  and  $\rho = 0.5$ . Small discrepancies are noted for small values of the test statistics, whereas the two are almost indistinguishable for large realizations. The latter is the important region because the test is rejected for large values of  $T_n$ . See the caption of Table 1 for a description of the DGPs.

**Assumption 3.**

$$U_n(u) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} Z_{t-h} \varepsilon_t \Rightarrow \Omega^{1/2} B(u) \quad \text{on } \mathbb{D}_{[0,1]}^q,$$

where  $B(u)$  is a standard  $q$ -dimensional Brownian motion and  $\mathbb{D}_{[0,1]}^q$  denotes the space of cadlag mappings from the unit interval to  $\mathbb{R}^q$ .

The relationship between  $T_n$  and Wald statistics implies that the existing expression 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  $\Xi = \sigma_\varepsilon^{-2} \Sigma_{zz}^{-1} \Omega$  and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_q)$ .

**Theorem 3.** *Suppose that Assumptions 1-3 hold. Let  $\beta_2 = cn^{-1/2}b$  where  $\sigma_\varepsilon^{-2}b' \Sigma_{zz} b = \kappa$ , and  $c \in \mathbb{R}$ . Define  $a = b' \Sigma_{zz} \Omega^{1/2} Q' \in \mathbb{R}^q$ . Then*

$$T_n \xrightarrow{d} \sum_{i=1}^q \lambda_i \left[ 2 \int_\rho^1 u^{-1} B_i(u) dB_i(u) - \int_\rho^1 u^{-2} B_i^2(u) du + (1 - \rho)c^2 + ca_i \{B_i(1) - B_i(\rho)\} \right], \quad (5)$$

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

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

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  $c = 0$ .<sup>3</sup> Second, it simplifies the expression of the limit distribution from one 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 of two independent  $\chi^2$ -distributed random variables.

**Theorem 4.** *Let  $B$  be a univariate standard Brownian motion. The distribution of  $2 \int_\rho^1 u^{-1} B dB - \int_\rho^1 u^{-2} B^2 du$  is identical to that of  $\sqrt{1 - \rho}(Z_1^2 - Z_2^2) + \log \rho$ , where  $Z_i \sim \text{iid}N(0, 1)$ .*

Theorems 3 and 4 show that the limit distribution of  $T_n/\sqrt{1 - \rho}$  is invariant to  $\rho$  under the null hypothesis, whereas the non-centrality parameter,  $\sqrt{1 - \rho}c^2$ , and hence the power of the

<sup>3</sup>The expression in Clark and McCracken (2005) involves a  $q \times q$  matrix of nuisance parameters. This was simplified by Stock and Watson (2003) to that in (5) in the special case where  $c = 0$ .



test, is decreasing in  $\rho$ . This property of the test might suggest choosing  $\rho$  as small as possible to maximize power, although such a conclusion is unwarranted because the result relied on  $\rho$  being strictly greater than zero, e.g., 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, the result shows that to obtain the same power for  $\rho = 0.75$  as one has with  $\rho = 0.25$ , one would need a 73% greater sample size.

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

**Theorem 5.** *The density of  $\sum_{j=1}^q \left[ 2 \int_{\rho}^1 u^{-1} B_j(u) dB_j(u) - \int_{\rho}^1 u^{-2} B_j(u)^2 du \right]$  is given by*

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

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

$$f_1(x) = \frac{1}{2\pi\sqrt{1-\rho}} K_0\left(\frac{|x-\log \rho|}{2\sqrt{1-\rho}}\right) \quad \text{and} \quad f_2(x) = \frac{1}{4\sqrt{1-\rho}} \exp\left(-\frac{|x-2 \log \rho|}{2\sqrt{1-\rho}}\right),$$

respectively, where  $K_0(x) = \int_0^{\infty} \frac{\cos(xt)}{\sqrt{1+t^2}} dt$  is the modified Bessel function of the second kind.

So in the case  $q = 2$ , the limit distribution is simply the non-central Laplace distribution. The density for  $q = 1$  is also readily available, since  $K_0(x)$  is implemented in standard software.

### 3 Conclusion

We show that a test statistic that is widely used for out-of-sample forecast comparisons of nested regression models is equal in probability to the difference between two Wald statistics of the same null - one using the full sample and one using a subsample. This equivalence greatly simplifies both the computation of the test statistic and the expression for its limit distribution. In fact, the limit distribution can be expressed as a difference between two independent  $\chi^2$ -distributions, and convolutions thereof. We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point,  $\rho$ .

These results raise serious questions about testing the stated null hypothesis out-of-sample in this manner. Subtracting a subsample Wald statistic from the full sample Wald statistic dilutes the power of the test and does not lead to any obvious advantages, such as robustness to outliers. Moreover, the test statistic,  $T_n$ , is not robust to heteroskedasticity, 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 heteroskedastic case by using a robust estimator for the asymptotic variance of  $\hat{\beta}_{2,n}$ .

On a constructive note, one could use the simplified expressions derived here to develop a test that is robust to potential mining over the sample split, analogous to the results derived in Rossi and Inoue (2012). In the present context, one can establish the weak convergence  $T_n(u) \Rightarrow B(1)' \Lambda B(1) - u^{-1} B(u)' \Lambda B(u) + \kappa \log u$  on  $\mathbb{D}_{[\underline{\rho}, \bar{\rho}]}^q$ , with  $0 < \underline{\rho} < \bar{\rho} < 1$ , where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_q)$ , which can be used to construct a robust inference procedure.

## References

- Clark, T. E., McCracken, M. W., 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105, 85–110.
- Clark, T. E., McCracken, M. W., 2005. Evaluating direct multi-step forecasts. *Econometric Reviews* 24, 369–404.
- De Jong, R. M., Davidson, J., 2000. The functional central limit theorem and convergence to stochastic integrals I: Weakly dependent processes. *Econometric Theory* 16, 621–642.
- Giacomini, R., White, H., 2006. Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Hansen, B., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
- McCracken, M. W., 2007. Asymptotics for out-of-sample tests of granger causality. *Journal of Econometrics* 140, 719–752.
- Rossi, B., Inoue, A., 2012. Out-of-sample forecast tests robust to the choice of window size. *Journal of Business and Economic Statistics* 30, 432–453.
- Stock, J. H., Watson, M. W., 2003. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41, 788–829.
- Stock, J. H., Watson, M. W., 2007. *Introduction to Econometrics*. 2nd Edition. AddisonWesley.
- West, K. D., 1996. Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.

## Appendix of Proofs

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

**Lemma A.1.** *Given Assumption 2, we have*

$$-\frac{1}{n} \sum_t \frac{n}{t} \varepsilon_t Z'_{t-h} \Sigma_{zz}^{-1} Z_{t-h-j} \varepsilon_{t-j} = \gamma_j \log \rho + o_p(1),$$

where  $\gamma_j = \text{tr}\{\Sigma_{zz}^{-1} \Gamma_j\}$ .

*Proof.* Let  $x_t = \varepsilon_t Z'_{t-h} \Sigma_{zz}^{-1} Z_{t-h-j} \varepsilon_{t-j} = \text{tr}\{\Sigma_{zz}^{-1} Z_{t-h} \varepsilon_t \varepsilon_{t-j}' Z'_{t-h-j}\}$ , since  $\Sigma_{zz}$  is symmetric. It follows by Assumption 2 that  $\sup_u \left| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} x_t - u \gamma_j \right| = o_p(1)$ . To establish the result in the lemma, we construct the subsamples,  $S_b = t_{b-1} + 1, \dots, t_b$ , for  $b = 1, \dots, B$ , where  $t_b = n_\rho + \lfloor b \frac{n-n_\rho}{B} \rfloor$ . It can be shown that there exists a  $K$ , so that  $\sup_{b=1, \dots, B} \left| \frac{n}{t_b} - \frac{n}{t_{b-1}} \right| \leq \frac{K}{B}$ . Let  $\zeta_n(u) = \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} x_t - \frac{\lfloor un \rfloor}{n} \gamma_j$ , and consider

$$\frac{1}{n} \sum_{t=n_\rho+1}^n \frac{n}{t} (x_t - \gamma_j) = \frac{1}{n} \sum_{b=1}^B \sum_{t \in S_b} \left( \frac{n}{t} - \frac{n}{t_b} \right) (x_t - \gamma_j) + \frac{1}{n} \sum_{b=1}^B \frac{n}{t_b} \sum_{t \in S_b} (x_t - \gamma_j).$$

The absolute value of the first term is bounded by

$$\sum_{b=1}^B \sup_{t \in S_b} \left| \frac{n}{t} - \frac{n}{t_b} \right| \left| \zeta_n\left(\frac{t_b}{n}\right) - \zeta_n\left(\frac{t_{b-1}}{n}\right) \right| \leq 2B \frac{K}{B} \sup_{b=0, \dots, B} \left| \zeta_n\left(\frac{t_b}{n}\right) \right| = o_p(1),$$

whereas the second term equals

$$\sum_{b=1}^B \frac{n}{t_b} \left[ \zeta_n\left(\frac{t_b}{n}\right) - \zeta_n\left(\frac{t_{b-1}}{n}\right) \right] = \zeta_n\left(\frac{n}{n}\right) - \left[ \frac{n}{t_B} - \frac{n}{t_{B-1}} \right] \zeta_n\left(\frac{t_{B-1}}{n}\right) - \dots - \left[ \frac{n}{t_2} - \frac{n}{t_1} \right] \zeta_n\left(\frac{t_1}{n}\right) - \frac{n}{t_1} \zeta_n\left(\frac{n_\rho}{n}\right),$$

whose absolute value is bounded by

$$\left| \zeta_n\left(\frac{n}{n}\right) \right| + \frac{B-1}{B} K \sup_{b=1, \dots, B-1} \left| \zeta_n\left(\frac{t_b}{n}\right) \right| + \frac{1}{\rho} \left| \zeta_n\left(\frac{n_\rho}{n}\right) \right| \leq \frac{3K}{\rho} \sup_t \zeta\left(\frac{t}{n}\right) = o_p(1).$$

This proves that  $\frac{1}{n} \sum_t \frac{n}{t} (\varepsilon_t Z'_{t-h} \Sigma_{zz}^{-1} Z_{t-h-j} \varepsilon_{t-j} - \gamma_j) = o_p(1)$ . The result now follows from  $\frac{1}{n} \sum_{t=n_\rho+1}^n \frac{n}{t} \gamma_j = \gamma_j \int_\rho^1 u^{-1} du + o(1) = \gamma_j \log \rho + o(1)$ .  $\square$

**Lemma A.2.** *Suppose  $U_t = U_{t-1} + u_t \in \mathbb{R}^q$  and let  $M$  be a symmetric  $q \times q$  matrix. Then  $2U'_{t-1} M u_t = U'_t M U_t - U'_{t-1} M U_{t-1} - u'_t M u_t$ .*

*Proof.*  $U'_{t-1} M u_t = (U_t - u_t)' M u_t = U'_t M (U_t - U_{t-1}) - u'_t M u_t$  equals

$$U'_t M U_t - (U_{t-1} + u_t)' M U_{t-1} - u'_t M u_t = U'_t M U_t - U'_{t-1} M U_{t-1} - u'_t M U_{t-1} - u'_t M u_t.$$

Rearranging terms and using  $u'_t M U_{t-1} = U'_{t-1} M u_t$  yields the result.  $\square$

**Lemma A.3.** *Suppose that Assumption 1 holds and let  $\hat{\theta}_t = M_{11,t}^{-1} M_{12,t}$  and  $\theta = \Sigma_{11}^{-1} \Sigma_{12}$ , where  $M_{ij,t} = n^{-1} \sum_{s=1}^t X_{i,s} X'_{j,s}$ ,  $i, j = 1, 2$ . Then (i)  $\sup_t \left\| \sum_{s=1}^t X_{1,s-h} Z'_{s-h} \beta_2 \right\| = O_p(n^{1/2})$ , (ii)  $\sup_t \left\| \hat{\beta}_{t-h} - \beta \right\| = O_p(n^{-1/2})$ , (iii) if  $\beta_2 \neq 0$ , then  $\sup_t \left\| \hat{\theta}_{t-h} - \theta \right\| = O_p(n^{-1/2})$ , while  $\sup_t \left\| \hat{\theta}_{t-h} - \theta \right\| = o_p(1)$  if  $\beta_2 = 0$ , and (iv)*

$$\sup_t \left[ n^{1/2} (\hat{\beta}_{2,t-h} - \beta_2) - \frac{n}{t} \Sigma_{zz}^{-1} U_{n,t} \right] = o_p(1),$$

where  $U_{n,t} = n^{-1/2} \sum_{s=1}^t Z_{s-h} \varepsilon_s$ .

*Proof.* Since  $X_{1,t-h} \eta'_t = X_{1,t-h} (Y_t - X'_{1,t-h} \delta) = X_{1,t-h} (Y_t - X'_{1,t-h} \delta - Z'_{t-h} \beta_2 + Z'_{t-h} \beta_2) = X_{1,t-h} (\varepsilon_t +$

$Z'_{t-h}\beta_2$ ), (i) follows by Assumption 1 and (ii) follows directly from the same Assumption. Next,  $\sum_{s=1}^t X_{1,s-h}X'_{2,s-h} = \sum_{s=1}^t X_{1,s-h}(X'_{1,s-h}\theta + X'_{2,s-h} - X'_{1,s-h}\theta) = \sum_{s=1}^t X_{1,s-h}(X'_{1,s-h}\theta + Z'_{s-h})$ , which implies

$$\hat{\theta}_{t-h} - \theta = \left( \sum_{s=1}^t X_{1,s-h}X'_{1,s-h} \right)^{-1} \sum_{s=1}^t X_{1,s-h}Z'_{s-h},$$

so that for  $\beta_2 \neq 0$  (iii) follows by using (i) and Assumption 1, while for  $\beta_2 = 0$ , (iii) follows directly from Assumption 1. Finally, we establish (iv) by noting that, from Assumption 1,

$$\sup_t \left[ \left( n^{-1} \sum_{s=1}^t \hat{Z}_{s-h}\hat{Z}'_{s-h} \right)^{-1} - \frac{n}{t} \Sigma_{zz}^{-1} \right] = o_p(1).$$

Next, since  $Z_{t-h} - \hat{Z}_{t-h} = X_{2,t-h} - \theta'X_{1,t-h} - X_{2,t-h} + \hat{\theta}'_{t-h}X_{1,t-h} = (\hat{\theta}_{t-h} - \theta)'X_{1,t-h}$  so that  $\hat{Z}_{t-h} = Z_{t-h} - (\hat{\theta}_{t-h} - \theta)'X_{1,t-h}$ , we have

$$n^{-1/2} \sum_{s=1}^t \hat{Z}_{s-h}\varepsilon_{s-h} = U_{n,t} - n^{-1/2} \sum_{s=1}^t (\hat{\theta}_{s-h} - \theta)'X_{1,s-h}\varepsilon_{s-h},$$

where (as a consequence of (iii) and Assumption 1) we can conclude that the last term is  $o_p(1)$  even if one takes the supremum over  $t = n_\rho + 1, \dots, n$ .  $\square$

Define

$$\begin{aligned} \xi_{1,t} &= (\hat{\delta}_t - \delta)'X_{1,t}, \\ \xi_{2,t} &= \beta_2'Z_t, \\ \xi_{3,t} &= (\hat{\beta}_{2,t} - \beta_2)'Z_t, \\ \xi_{4,t} &= \hat{\beta}'_{2,t}(Z_t - \hat{Z}_t), \end{aligned}$$

where  $\hat{Z}_t = X_{2,t} - M_{21,t}M_{11,t}^{-1}X_{1,t}$ .

In the following we simplify notations and 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  $\hat{y}_{t|t-h}(\hat{\beta}_{t-h})$ .

**Lemma A.4.** *We have*

$$y_{t+h} - \tilde{y}_{t+h|t} = \varepsilon_{t+h,t} - \xi_{1,t} + \xi_{2,t} \quad \text{and} \quad y_{t+h} - \hat{y}_{t+h|t} = \varepsilon_{t+h,t} - \xi_{1,t} - \xi_{3,t} + \xi_{4,t},$$

so that

$$\begin{aligned} (y_{t+h} - \tilde{y}_{t+h|t})^2 - (y_{t+h} - \hat{y}_{t+h|t})^2 &= \xi_{2,t}^2 + 2\xi_{2,t}\varepsilon_{t+h,t} + 2\xi_{3,t}\varepsilon_{t+h,t} - \xi_{3,t}^2 \\ &\quad - 2\xi_{4,t}\varepsilon_{t+h,t} - 2\xi_{1,t}(\xi_{2,t} + \xi_{3,t} - \xi_{4,t}) - \xi_{4,t}^2 + 2\xi_{3,t}\xi_{4,t} \quad (\text{A.1}) \end{aligned}$$

*Proof.* First, write the forecast error from the smaller model as

$$\begin{aligned} y_t - \tilde{y}_{t|t-h} &= y_t - \hat{\delta}'_{t-h} X_{1,t-h} = y_t - \delta' X_{1,t-h} - \beta_2 Z_{t-h} - (\hat{\delta}_{1,t-h} - \delta)' X_{1,t-h} + \beta_2 Z_{t-h} \\ &= \varepsilon_{t,t-h} - (\hat{\delta}_{1,t-h} - \delta)' X_{1,t-h} + \beta_2' Z_{t-h} = \varepsilon_{t,t-h} - \xi_{1,t-h} + \xi_{2,t-h}. \end{aligned}$$

Next, write the forecast error from the large model as

$$\begin{aligned} y_t - \hat{y}_{t|t-h} &= y_t - \hat{\beta}'_{1,t-h} X_{1,t-h} - \hat{\beta}'_{2,t-h} X_{2,t-h} = y_t - \tilde{\beta}'_{1,t-h} X_{1,t-h} - \hat{\beta}'_{2,t-h} \hat{Z}_{t,t-h} \\ &= y_t - \delta' X_{1,t-h} - \beta_2' Z_{t-h} - (\hat{\delta}_{t-h} - \delta)' X_{1,t-h} - (\hat{\beta}_{2,t-h} - \beta_2)' Z_{t-h} + \hat{\beta}'_{2,t-h} (Z_{t-h} - \hat{Z}_{t,t-h}) \\ &= \varepsilon_{t,t-h} - \xi_{1,t-h} - \xi_{3,t-h} + \xi_{4,t-h}. \end{aligned}$$

Using these expressions, the difference between the squared forecast error takes the form

$$\begin{aligned} &(y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2 \\ &= (\varepsilon_{t,t-h} - \xi_{1,t-h} + \xi_{2,t-h})^2 - (\varepsilon_{t,t-h} - \xi_{1,t-h} - \xi_{3,t-h} + \xi_{4,t-h})^2 \\ &= \xi_{2,t-h}^2 + 2(\varepsilon_{t,t-h} - \xi_{1,t-h})(\xi_{2,t-h} + \xi_{3,t-h} - \xi_{4,t-h}) - (\xi_{3,t-h} - \xi_{4,t-h})^2 \\ &= \xi_{2,t-h}^2 + 2\xi_{2,t-h}\varepsilon_{t,t-h} + 2\xi_{3,t-h}\varepsilon_{t,t-h} - \xi_{3,t-h}^2 \\ &\quad - 2\xi_{4,t-h}\varepsilon_{t,t-h} - 2\xi_{1,t-h}(\xi_{2,t-h} + \xi_{3,t-h} - \xi_{4,t-h}) - \xi_{4,t-h}^2 + 2\xi_{3,t-h}\xi_{4,t-h} \end{aligned}$$

□

The lemma shows that the loss differential involves eight terms. We next derive results for these.

**Lemma A.5.** *Let*

$$\begin{aligned} A &= \sum_t \xi_{2,t-h}^2 = \sum_t \beta_2' Z_{t-h} Z_{t-h}' \beta_2, \\ B &= \sum_t \xi_{2,t-h} \varepsilon_t = \beta_2' \sum_t Z_{t-h} \varepsilon_t, \\ C &= \sum_t \xi_{3,t-h} \varepsilon_t = \sum_t (\hat{\beta}_{2,t-h} - \beta_2)' Z_{t-h} \varepsilon_t, \\ D &= \sum_t \xi_{3,t-h}^2 = \sum_t (\hat{\beta}_{2,t-h} - \beta_2)' Z_{t-h} Z_{t-h}' (\hat{\beta}_{2,t-h} - \beta_2). \end{aligned}$$

*Then*

$$A + 2B + 2C - D = \sigma_\varepsilon^2 (W_n - W_{n_\rho} + \kappa \log \rho) + o_p(1).$$

*Proof.* First consider  $D$ . From Lemma A.3 and Assumption 1 it follows that

$$\sup_t \left| \sum_t (\hat{\beta}_{2,t-h} - \beta_2)' [Z_{t-h} Z_{t-h}' - \Sigma_{zz}] (\hat{\beta}_{2,t-h} - \beta_2) \right| = o_p(1),$$

so that

$$D = \frac{1}{n} \sum_{t=n_\rho+1}^n \left(\frac{n}{t}\right)^2 (\hat{\beta}_{2,t-h} - \beta_2)' \Sigma_{zz}^{-1} (\hat{\beta}_{2,t-h} - \beta_2) + o_p(1) = \frac{1}{n} \sum_{t=n_\rho+1}^n \left(\frac{n}{t}\right)^2 U'_{n,t} \Sigma_{zz}^{-1} U_{n,t} + o_p(1),$$

where we used Lemma A.3.iv. Next, define  $u_{n,t} = n^{-1/2} Z_{t-h} \varepsilon_t$ , and apply Lemma A.3.iv to establish

$$\begin{aligned} C &= \sum_{t=n_\rho+1}^n \frac{n}{t} U'_{n,t-h} \Sigma_{zz}^{-1} u_{n,t} + o_p(1) \\ &= \sum_{t=n_\rho+1}^n \frac{n}{t} U'_{n,t-1} \Sigma_{zz}^{-1} u_{n,t} - \sum_{t=n_\rho+1}^n \frac{n}{t} \sum_{i=1}^{h-1} u'_{n,t-i} \Sigma_{zz}^{-1} u_{n,t} + o_p(1) \\ &= \sum_{t=n_\rho+1}^n \frac{n}{t} U'_{n,t-1} \Sigma_{zz}^{-1} u_{n,t} + \xi + o_p(1), \end{aligned}$$

where  $\xi = (\gamma_1 + \dots + \gamma_{h-1}) \log \rho$ , using Assumption 2 and Lemma A.1.

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

$$\begin{aligned} 2C &= \sum_{t=n_\rho+1}^n \frac{n}{t} (U'_{n,t} \Sigma_{zz}^{-1} U_{n,t} - U'_{n,t-1} \Sigma_{zz}^{-1} U_{n,t-1} - u'_{n,t} \Sigma_{zz}^{-1} u_{n,t}) + 2\xi + o_p(1) \\ &= U'_{n,n} \Sigma_{zz}^{-1} U_{n,n} - \frac{n}{n_\rho} U'_{n,n_\rho} \Sigma_{zz}^{-1} U_{n,n_\rho} + \frac{1}{n} \sum_{t=n_\rho+1}^n \left(\frac{n}{t}\right)^2 U'_{n,t} \Sigma_{zz}^{-1} U_{n,t} + \sigma_\varepsilon^2 \kappa \log \rho + o_p(1). \quad (\text{A.2}) \end{aligned}$$

Here we used  $\sigma_\varepsilon^{-2} \kappa = \text{tr}\{\Sigma_{zz}^{-1} \Omega\} = \sum_{j=-h+1}^{h-1} \text{tr}\{\Sigma_{zz}^{-1} n^{-1} \sum_t Z_{t-h-j} \varepsilon_{t-j} \varepsilon_t Z'_{t-h}\} + o_p(1) = \sum_{j=-h+1}^{h-1} \gamma_j + o_p(1)$ . The penultimate term in (A.2) offsets the contributions from  $-D$ , whereas  $A + 2B$  equals

$$\beta_2' \sum_{t=1}^n Z_{t-h} Z'_{t-h} \beta_2 - \beta_2' \sum_{t=1}^{n_\rho} Z_{t-h} Z'_{t-h} \beta_2 + 2n^{1/2} \beta_2' U_{n,n} - 2n^{1/2} \beta_2' U_{n,n_\rho}.$$

With  $W_m = \hat{\sigma}_\varepsilon^{-2} \hat{\beta}'_{2,m} [\sum_{t=1}^m Z_{t-h} Z'_{t-h}] \hat{\beta}_{2,m} = \hat{\sigma}_\varepsilon^{-2} (\hat{\beta}_{2,m} - \beta_2 + \beta_2)' [\sum_{t=1}^m Z_{t-h} Z'_{t-h}] (\hat{\beta}_{2,m} - \beta_2 + \beta_2)$ , we have

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 (W_n - W_{n_\rho}) &= U'_{n,n} \Sigma_{zz}^{-1} U_{n,n} - \frac{n}{n_\rho} U'_{n,n_\rho} \Sigma_{zz}^{-1} U_{n,n_\rho} + o_p(1) \\ &\quad + \beta_2' \sum_{t=n_\rho+1}^n Z_{t-h} Z'_{t-h} \beta_2 + 2n^{1/2} \beta_2' (U_{n,n} - U_{n,n_\rho}), \end{aligned}$$

and the result now follows.  $\square$

**Lemma A.6.** *Let*

$$\begin{aligned}
E &= \sum_t \xi_{4,t-h} \varepsilon_t = \sum_t \hat{\beta}'_{2,t-h} (Z_{t-h} - \hat{Z}_{t-h}) \varepsilon_t, \\
F &= \sum_t \xi_{1,t-h} (\xi_{2,t-h} + \xi_{3,t-h} - \xi_{4,t-h}) = \sum_t (\hat{\delta}_{t-h} - \delta)' X_{1,t-h} \hat{Z}'_{t-h} \hat{\beta}_{2,t-h}, \\
G &= \sum_t \xi_{4,t-h}^2 = \sum_t \hat{\beta}'_{2,t-h} (Z_{t-h} - \hat{Z}_{t-h}) (Z_{t-h} - \hat{Z}_{t-h})' \hat{\beta}_{2,t-h}, \\
H &= \sum_t \xi_{3,t-h} \xi_{4,t-h} = \sum_t (\hat{\beta}_{2,t-h} - \beta_2)' Z_{t-h} (Z_{t-h} - \hat{Z}_{t-h})' \hat{\beta}_{2,t-h}.
\end{aligned}$$

If  $\beta_2 = n^{-1/2}b$  (including the case  $\beta_2 = 0$ ) then  $E, F, G = o_p(1)$  and  $H = o_p(n^{-1/2})$ . If  $\beta_2 = b$  with  $b \neq 0$  then  $E, F, G$  are  $O_p(1)$ , whereas  $H = O_p(n^{-1/2})$ .

*Proof.* To simplify the notation, we write  $\leq_o$  to mean inequality in terms of orders of probability. First consider  $E$ . Recall that  $Z_{t-h} - \hat{Z}_{t-h} = (\hat{\theta}_{t-h} - \theta)' X_{1,t-h}$ , so that

$$\sum_t \hat{\beta}'_{2,t-h} (Z_{t-h} - \hat{Z}_{t-h}) \varepsilon_t = \sum_t \hat{\beta}'_{2,t-h} (\hat{\theta}_{t-h} - \theta)' X_{1,t-h} \varepsilon_t.$$

The order of this expression can be determined from

$$\beta'_2 \sum_t (\hat{\theta}_{t-h} - \theta)' X_{1,t-h} \varepsilon_t + \sum_t (\hat{\beta}_{2,t-h} - \beta_2)' (\hat{\theta}_{t-h} - \theta)' X_{1,t-h} \varepsilon_t.$$

If  $\beta_2 \neq 0$ , these two terms are  $\beta'_2 O_p(1) + O_p(n^{-1/2})$ , while if  $\beta_2 = 0$  the first term vanishes whereas the second term is  $o_p(1)$ . Turning to  $F$ , we have

$$\begin{aligned}
\sum_t (\hat{\delta}_{t-h} - \delta)' X_{1,t-h} \hat{Z}'_{t-h} \hat{\beta}_{2,t-h} &= \sum_t (\hat{\delta}_{t-h} - \delta)' X_{1,t-h} \hat{Z}'_{t-h} \beta_2 + \sum_t (\hat{\delta}_{t-h} - \delta)' X_{1,t-h} \hat{Z}'_{t-h} (\hat{\beta}_{2,t-h} - \beta_2) \\
&\leq_o \sup_t \left| \hat{\delta}_{t-h} - \delta \right|' \sum_t X_{1,t-h} \hat{Z}'_{t-h} \beta_2 \\
&\quad - \sup_t \left| \hat{\delta}_{t-h} - \delta \right|' \sum_t X_{1,t-h} X'_{1,t-h} (\hat{\theta}_{t-h} - \theta) \beta_2 \\
&\quad + \sup_t \left| \hat{\delta}_{t-h} - \delta \right|' \sum_t X_{1,t-h} \hat{Z}'_{t-h} (\hat{\beta}_{2,t-h} - \beta_2) \\
&\quad - \sup_t \left| \hat{\delta}_{t-h} - \delta \right|' \sum_t X_{1,t-h} X'_{1,t-h} (\hat{\theta}_{t-h} - \theta) (\hat{\beta}_{2,t-h} - \beta_2) \\
&\leq_o O_p(1) \beta_2 + O_p(1) \beta_2 + O_p(n^{-1/2}) + o_p(1).
\end{aligned}$$

Similarly, for  $G$  we have

$$\begin{aligned}
\sum_t \hat{\beta}'_{2,t-h} (Z_{t-h} - \hat{Z}_{t-h})(Z_{t-h} - \hat{Z}_{t-h})' \hat{\beta}_{2,t-h} &= \sum_t \hat{\beta}'_{2,t-h} (\hat{\theta}_{t-h} - \theta)' X_{1,t-h} X'_{1,t-h} (\hat{\theta}_{t-h} - \theta) \hat{\beta}_{2,t-h} \\
&\leq_o \beta'_2 \sup_t \left| \hat{\theta}_{t-h} - \theta \right|' \sum_t X_{1,t-h} X'_{1,t-h} \sup_t \left| \hat{\theta}_{t-h} - \theta \right| \beta_2 \\
&\quad + 2\beta'_2 \sup_t \left| \hat{\theta}_{t-h} - \theta \right|' \sum_t X_{1,t-h} X'_{1,t-h} \\
&\quad \times \sup_t \left| \hat{\theta}_{t-h} - \theta \right| \sup_t \left| \hat{\beta}_{2,t-h} - \beta_2 \right| \\
&\quad + \sup_t \left| \hat{\beta}_{2,t-h} - \beta_2 \right|' \sup_t \left| \hat{\theta}_{t-h} - \theta \right|' \sum_t X_{1,t-h} X'_{1,t-h} \\
&\quad \times \sup_t \left| \hat{\theta}_{t-h} - \theta \right| \sup_t \left| \hat{\beta}_{2,t-h} - \beta_2 \right| \\
&= \beta'_2 O_p(1) \beta_2 + 2\beta'_2 O_p(n^{-1/2}) + o_p(1).
\end{aligned}$$

Finally, for  $H$  we observe

$$\begin{aligned}
\sum_t (\hat{\beta}_{2,t-h} - \beta_2)' Z_{t-h} (Z_{t-h} - \hat{Z}_{t-h})' \hat{\beta}_{2,t-h} &\leq_o \sup_t \left| \hat{\beta}_{2,t-h} - \beta_2 \right|' \sum_t Z_{t-h} X'_{1,t-h} \sup_t \left| \hat{\theta}_{t-h} - \theta \right| \beta_2 \\
&\quad + \sup_t \left| \hat{\beta}_{2,t-h} - \beta_2 \right|' \sum_t Z_{t-h} X'_{1,t-h} \sup_t \left| \hat{\theta}_{t-h} - \theta \right| \\
&\quad \times \sup_t \left| \hat{\beta}_{2,t-h} - \beta_2 \right| \\
&= O_p(n^{-1/2}) \beta_2 + o_p(n^{-1/2}).
\end{aligned}$$

□

**Proof of Theorem 1.** From Lemma A.4 it follows that

$$T_n = \frac{A + 2B + 2C - D - 2E - 2F - G + 2H}{\hat{\sigma}_\varepsilon^2}.$$

The theorem now follows by combining the results in Lemmas A.5 and A.6, where the  $O_p(1)$  term that arises when  $k > 0$  and  $\beta_2 = b \neq 0$  stems from  $E$ ,  $F$ , and  $G$ . □

**Proof of Theorem 2.** First, note that

$$(y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2 = \{y_t^2 - (y_t - \hat{y}_{t|t-h})^2\} - \{y_t^2 - (y_t - \tilde{y}_{t|t-h})^2\},$$

where  $y_t^2$  can be viewed as the squared prediction error from an auxiliary prediction model that always predicts  $y_t$  to be zero. We proceed by summing over  $t = n_\rho + 1, \dots, n$ , and apply Lemma A.5 separately to the two sums. Specifically, it follows that

$$\sum_t \{y_t^2 - (y_t - \hat{y}_{t|t-h})^2\} = A'' + 2B'' + 2C'' - D'',$$



where  $A'' = \sum_t \beta' X_{t-h} X'_{t-h} \beta$ ,  $B'' = \beta' \sum_t X_{t-h} \varepsilon_t$ , etc. and the corresponding  $E''$ ,  $F''$ ,  $G''$ , and  $H''$  terms are all zero since there are no parameters to estimate in the smallest prediction model. It follows directly from Lemma A.5 that the sum equals  $\sigma_\varepsilon^2 \left\{ W_n'' - W_{n_\rho}'' + \kappa'' \log \rho \right\} + o_p(1)$ , where  $\kappa'' = \sum_{i=1}^{k+q} \lambda_i''$ , with  $\lambda_i''$  being the eigenvalues of  $\sigma_\varepsilon^{-2} \Sigma^{-1} \Omega''$  and  $\Omega'' = \sum_{j=-h+1}^{h-1} \Gamma_j''$ , with  $\Gamma_j'' = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_{t-h} \varepsilon_t \varepsilon_{t-j}' X'_{t-h-j}$ . Similarly, the sum over  $y_t^2 - (y_t - \tilde{y}_{t|t-h})^2$  equals  $\sigma_\eta^2 \{ W_n' - W_{n_\rho}' + \kappa' \log \rho \} + o_p(1)$ , where we note that  $\sigma_\eta^2 > \sigma_\varepsilon^2$  unless  $\beta_2 = 0$ , and  $\kappa' = \sum_{i=1}^k \lambda_i'$ , with  $\lambda_i'$  being the eigenvalues of  $\sigma_\eta^{-2} \Sigma_{11}^{-1} \Omega'$  and  $\Omega' = \sum_{j=-h+1}^{h-1} \Gamma_j'$ , with  $\Gamma_j' = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n X_{1,t-h} \eta_t \eta_{t-j}' X'_{1,t-h-j}$ . This completes the proof.  $\square$

**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  $F(u) = \frac{1}{u} B^2(u) - \log u$  (for  $u > 0$ ). By Ito stochastic calculus:

$$dF = \frac{\partial F}{\partial B} dB + \left[ \frac{\partial F}{\partial u} + \frac{1}{2} \frac{\partial^2 F}{(\partial B)^2} \right] du = \frac{2}{u} B dB - \frac{1}{u^2} B^2 du,$$

so  $\int_\rho^1 \frac{2}{u} B dB - \int_\rho^1 \frac{1}{u^2} B^2 du = \int_\rho^1 dF(u)$  equals  $F(1) - F(\rho) = B^2(1) - \log 1 - B^2(\rho)/\rho + \log \rho$ .

Next, consider  $W_n - W_{n_\rho}$ . Under Assumption 3, we have  $U_{n, \lfloor un \rfloor} = n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} Z_{t-h} \varepsilon_t \Rightarrow U(u) = \Omega^{1/2} B(u)$ , so that

$$\begin{aligned} W_n - W_{n_\rho} &= \hat{\sigma}_\varepsilon^{-2} [U'_{n,n} \Sigma_{zz}^{-1} U_{n,n} - \frac{n}{n_\rho} U'_{n,n_\rho} \Sigma_{zz}^{-1} U_{n,n_\rho}] \\ &\quad + \hat{\sigma}_\varepsilon^{-2} [\beta_2' \sum_{t=n_\rho+1}^n Z_{t-h} Z'_{t-h} \beta_2 + 2cn^{1/2} \beta_2' (U_{n,n} - U_{n,n_\rho})] + o_p(1), \\ &= B(1)' \Xi B(1) - \rho^{-1} B(\rho)' \Xi B(\rho) \\ &\quad + (1-\rho)c^2 \sigma_\varepsilon^{-2} b' \Sigma_{zz} b + 2c \sigma_\varepsilon^{-2} b' \Omega^{1/2} [B(1) - B(\rho)] + o_p(1). \end{aligned}$$

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

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

where we used that  $\sigma_\varepsilon^{-2} b' \Omega^{1/2} Q' = b' \Sigma_{zz} \Omega^{1/2} \sigma_\varepsilon^{-2} \Omega^{1/2} \Sigma_{zz}^{-1} \Omega^{1/2} Q' = b' \Sigma_{zz} \Omega^{1/2} \Xi Q' = b' \Sigma_{zz} \Omega^{1/2} Q' \Lambda = (a_1 \lambda_1, \dots, a_q \lambda_q)$ . Since  $\tilde{B}$  and  $B$  are identically distributed, the limit distribution may be expressed in terms of  $B$  instead of  $\tilde{B}$ .  $\square$

**Proof of Theorem 4.** Let  $B(u)$  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:

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

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

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

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

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

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

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

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

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

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

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

$$f_q(s) = \frac{1}{2^q \Gamma(\frac{q}{2})^2} e^{-|s|/2} \int_0^{\infty} (u(u+|s|))^{q/2-1} e^{-u} du.$$

When  $q = 1$  this simplifies to  $f_1(s) = \frac{1}{2\pi} K_0(\frac{|s|}{2})$ , where  $K_k(x)$  denotes the modified Bessel function of the second kind. For  $q = 2$  the expression for the density reduces to the simpler expression,  $f_2(s) = \frac{1}{4} e^{-\frac{|s|}{2}}$ , which is the density of the Laplace distribution with scale parameter 2.  $\square$