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

Peter Reinhard Hansen
European University Institute and CREATES

Allan Timmermann
UCSD and CREATES

November 13, 2012

Abstract

We establish 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. Our results shed new light on many aspects of the test and establishes 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 Frank Diebold, Jim Stock, 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 tests based on the full sample and the initial estimation sample, respectively. Our result has three 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 depends on a potentially very large set of recursively updated parameter estimates. Third, 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

$$y_{t+h} = \beta_1' X_{1t} + \beta_2' X_{2t} + \varepsilon_{t+h}, \qquad t = 1, \ldots, n \tag{1}$$

where $X_{1t} \in \mathbb{R}^k$ and $X_{2t} \in \mathbb{R}^q$.

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 $(X_{1,s-h}', X_{2,s-h}')'$, for $s = 1, \ldots, t$, resulting in the least squares estimate $\hat{\beta}_t = (\hat{\beta}_{1t}', \hat{\beta}_{2t}')'$, and using $\hat{y}_{t+h|t}(\hat{\beta}_t) = \hat{\beta}_{1t}' X_t + \hat{\beta}_{2t}' X_{2t}$ to forecast y_{t+h}. The resulting forecast can be compared

to that of a smaller (nested) regression model, $y_{t+h} = \beta_1' X_{1t} + \tilde{\varepsilon}_{t+h}$ say, whose forecasts are given by $\tilde{y}_{t+h|t}(\tilde{\beta}_{1t}) = \tilde{\beta}_{1t}' X_{1t}$, where $\tilde{\beta}_{1t} = \left(\sum_{s=1}^{t} X_{1,s-h} X_{1,s-h}'\right)^{-1} \sum_{s=1}^{t} X_{1,s-h} y_s$.

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[1]

$$H_0 : \mathrm{E}[y_t - \hat{y}_{t|t-h}(\beta)]^2 = \mathrm{E}[y_t - \tilde{y}_{t|t-h}(\beta_1)]^2. \tag{2}$$

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

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

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimator of $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{t+h})$ and n_ρ is the number of observations set aside for the initial estimation of β (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) shows that the asymptotic distribution of T_n is given as a convolution of q independent random variables, each with a distribution of $2\int_\rho^1 u^{-1} B(u) \mathrm{d}B(u) - \int_\rho^1 u^{-2} B(u)^2 \mathrm{d}u$. Results for the case with $h > 1$ and heteroskedastic errors are 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, define the vector of stacked variables $V_t = (y_t, X_{t-h}')'$. We make the following assumption:

Assumption 1. $\Sigma_{vv} = \mathrm{E}(V_t V_t')$ *is positive definite and does not depend on* t. *Moreover,*

$$\sup_{u \in (0,1]} \left| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} V_t V_t' - u \Sigma_{vv} \right| = o_p(1). \tag{4}$$

The first part of Assumption 1 ensures that the population predictive regression coefficients do not depend on t. For convenience, we express the block structure of Σ_{vv} as follows

$$\Sigma_{vv} = \begin{pmatrix} \Sigma_{yy} & \bullet \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \qquad \text{with} \quad \Sigma_{xx} = \begin{pmatrix} \Sigma_{11} & \bullet \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

[1] Another approach is to consider $\mathrm{E}[y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h})]^2$ which typically depends on t, see, e.g., Giacomini and White (2006).

where the blocks in Σ_{xx} refer to X_{1t} and X_{2t}, respectively. Similarly, define

$$\varepsilon_t = y_t - \Sigma_{yx}\Sigma_{xx}^{-1}X_{t-h}, \qquad Z_t = X_{2t} - \Sigma_{21}\Sigma_{11}^{-1}X_{1t},$$

as the "error" term from the large model and the auxiliary variables, respectively, so that Z_t is constructed to be the part of X_{2t} that is orthogonal to X_{1t}. Next, define the population objects, $\sigma_\varepsilon^2 = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$, $\Sigma_{zz} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, and $(\beta_1', \beta_2')' = \Sigma_{xx}^{-1}\Sigma_{xy}$. Then $\sigma_\varepsilon^2 > 0$ and Σ_{zz} is positive definite because Σ_{vv} is positive definite. Further, let $\Sigma = \sigma_\varepsilon^2\Sigma_{zz}$ and $\Omega := \text{plim}_{n\to\infty}\frac{1}{n}\sum_{s,t=1}^{n} Z_{s-h}\varepsilon_s\varepsilon_t Z_{t-h}'$, where the latter is the long-run variance of Ω. We make the following assumption about the partial sum of $Z_{t-h}\varepsilon_t$:[2]

Assumption 2. *Let* $U_n(u) := \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un \rfloor} Z_{t-h}\varepsilon_t$, *and assume that*

$$U_n(u) \Rightarrow U(u) = \Omega^{1/2}B(u), \qquad on \quad \mathbb{D}_{[0,1]}^q,$$

with $\det\Omega > 0$, *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* $\mathbb{R}^q$.

Finally, we make an assumption that imposes a type of unpredictability of the forecast errors beyond the forecasting horizon, h, and simplifies the expression for Ω because higher order autocovariances are all zero. This assumption is easily tested by inspecting the autocorrelations of $Z_{t-h}\varepsilon_t$.

Assumption 3. $\text{cov}(Z_{t-h}\varepsilon_t, Z_{t-h-j}\varepsilon_{t-j}) = 0$ *for* $|j| \geq h$.

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 σ_ε^2 and Σ_{zz}, respectively. This statistic is based on a "homoskedastic" estimator of the asymptotic variance, which causes the eigenvalues of $\sigma_\varepsilon^{-2}\Sigma_{zz}^{-1}\Omega$, $\lambda_1, \ldots, \lambda_q$, 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).

[2]This assumption can be shown to hold under standard regularity conditions often used in the literature, such as those in Hansen (1992) (mixing) or those in De Jong and Davidson (2000) (near-epoch).

With the above assumptions and Assumption A.1 from the Appendix, we can now formulate our main result.

Theorem 1. *Given Assumptions 1-3 and A.1, the out-of-sample test statistic in equation (3) can be written as* $T_n = W_n - W_{n_\rho} + \kappa \log \rho + o_p(1)$*, where* $\kappa = \sum_{i=1}^{q} \lambda_i$*.*

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

For the general case with $h \geq 1$ and heteroskedastic prediction errors, the limit distribution for T_n (under the null hypothesis) was derived in Clark and McCracken (2005).[3] It involves a $q \times q$ matrix of nuisance parameters, but was simplified by Stock and Watson (2003) to

$$\sum_{i=1}^{q} \lambda_i \left[2 \int_{\rho}^{1} u^{-1} B_i(u) \mathrm{d}B_i(u) - \int_{\rho}^{1} u^{-2} B_i(u) B_i(u) \mathrm{d}u \right], \tag{5}$$

where $B = (B_1, \ldots, B_q)'$ is a standard q-dimensional Brownian motion. Theorem 1 implies that this expression can be greatly simplified:

Corollary 1. *The distribution in equation (5) is identical to that of*

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

Next, we show that the limit distribution can be expressed in terms of differences between two independent χ^2-distributed random variables (as opposed to the dependent ones $B_i^2(1)$ and $\rho^{-1} B_i^2(\rho)$).

Theorem 2. *The distribution of* $2 \int_{\rho}^{1} u^{-1} B \mathrm{d}B - \int_{\rho}^{1} u^{-2} B^2 \mathrm{d}u$ *is identical to that of* $\sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho$*, where* $Z_i \sim iidN(0,1)$*.*

Because the distribution is expressed in terms of two independent χ^2-distributed random variables, in the homoskedastic 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 3. *The density of* $\sum_{j=1}^{q} \left[2 \int_{\rho}^{1} u^{-1} B_j(u) \mathrm{d}B_j(u) - \int_{\rho}^{1} u^{-2} B_j(u)^2 \mathrm{d}u \right]$ *is given by*

$$f_1(x) = \tfrac{1}{2\pi\sqrt{1-\rho}} \mathrm{K}_0\left(\tfrac{|x - \log \rho|}{2\sqrt{1-\rho}}\right),$$

[3] The standard Brownian motion, B, that appears in (5) need not be identical to that used in Assumption 2.

for $q = 1$*, 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* $q = 2$ *we have*

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

which is the non-central Laplace distribution.

The densities for $q = 3, 4, 5, \ldots$ can be obtained by convolution of those stated in the Corollary. Fortunately, $\mathrm{K}_0(x)$ is implemented in standard software and is easy to compute.

3 Conclusion

We show that a test statistic, which 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.

Our result raises 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, etc. Moreover, the conventional full sample Wald test can easily be adapted to the heteroskedastic case by using a robust estimator for the asymptotic variance of β_2.

On a more constructive note, one could use the simplified expressions derived here to develop a test that is robust to potential mining over the sample split. By strengthening the convergence results in Assumption A.1 to be uniform in ρ over the range $\rho \in [\underline{\rho}, \bar{\rho}]$, with $0 < \underline{\rho} < \bar{\rho} < 1$, one achieves $T_n(u) \xrightarrow{d} G(u) = B(1)'\Lambda B(1) - u^{-1}B(u)'\Lambda B(u) + \kappa \log u$ with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_q)$, which can be used to derive a test whose test statistic is constructed from a range of sample splits; see Rossi and Inoue (2012) and Hansen and Timmermann (2012).

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.

Hansen, P. R., Timmermann, A., 2012. Choice of sample split in out-of-sample forecast evaluation. working.

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 61, 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

Assumption A.1. *Let* $\gamma_j = \mathrm{E}(\varepsilon_t Z'_{t-h}\Sigma_{zz}^{-1} Z_{t-h-j}\varepsilon_{t-j})$. *We assume that as* $n \to \infty$

$$\sum_{t=n_\rho+1}^{n} (\hat{\beta}_{2,t-h} - \beta_2)'(\Sigma_{zz} - Z_{t-h}Z'_{t-h})(\hat{\beta}_{2,t-h} - \beta_2) \quad \xrightarrow{p} \quad 0, \tag{A.1}$$

$$\tfrac{1}{n} \sum_{t=n_\rho+1}^{n} \tfrac{n}{t}(\varepsilon_t Z'_{t-h}\Sigma_{zz}^{-1} Z_{t-h-j}\varepsilon_{t-j} - \gamma_j) \quad \xrightarrow{p} \quad 0. \tag{A.2}$$

Convergence in probability holds under suitable regularity conditions and is, in fact, uniform in ρ for $\rho \in (a, b)$ where $0 < a < b < 1$, under suitable mixing conditions by applying Hansen (1992, theorem 3.3), see Hansen and Timmermann (2012). (A.2) implies that $-\frac{1}{n}\sum_{t=n_\rho+1}^{n} \frac{n}{t}\varepsilon_t Z'_{t-h}\Sigma_{zz}^{-1} Z_{t-h-j}\varepsilon_{t-j} = \gamma_j \int_\rho^1 u^{-1}\mathrm{d}u + o_p(1) = \gamma_j \log\rho + o_p(1)$.

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

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

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

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

Proof of Theorem 1. Without loss of generality we consider the case where $k = 0$, so that $Z_t = X_{2t}$. The general case with $k > 0$ results in additional terms (involving cross products of $X_{1,t-h}Z'_{t-h}$) that all vanish in probability in this analysis, see Hansen and Timmermann (2012, Lemma A.2). We decompose

the loss differential $\sum_{t=n_\rho+1}^{n}(y_t-\tilde{y}_{t|t-h})^2-(y_t-\hat{y}_{t|t-h})^2$ as follows:

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

Let $U_{n,t}=n^{-1/2}\sum_{t=1}^{t}Z_{t-h}\varepsilon_t$ and $u_{n,t}=n^{-1/2}Z_{t-h}\varepsilon_t$. By (A.1) $D=\frac{1}{n}\sum_{t=n_\rho+1}^{n}(\frac{n}{t})^2U_{n,t}'\Sigma_{zz}^{-1}U_{n,t}+o_p(1)$ and

$$\begin{aligned} C &= \sum_{t=n_\rho+1}^{n} 2\tfrac{n}{t}U_{n,t-h}'\Sigma_{zz}^{-1}u_{n,t}+o_p(1) \\ &= \sum_{t=n_\rho+1}^{n} 2\tfrac{n}{t}U_{n,t-1}'\Sigma_{zz}^{-1}u_{n,t} - 2\sum_{t=n_\rho+1}^{n}\tfrac{n}{t}\sum_{i=1}^{h-1}u_{n,t-i}'\Sigma_{zz}^{-1}u_{n,t}+o_p(1) \\ &= \sum_{t=n_\rho+1}^{n} 2\tfrac{n}{t}U_{n,t-1}'\Sigma_{zz}^{-1}u_{n,t}+\xi+o_p(1), \end{aligned}$$

where $\xi=2(\gamma_1+\cdots+\gamma_{h-1})\log\rho$, using (A.2). Now apply Lemma A.1

$$\begin{aligned} C &= \sum_{t=n_\rho+1}^{n}\tfrac{n}{t}(U_{n,t}'\Sigma_{zz}^{-1}U_{n,t}-U_{n,t-1}'\Sigma_{zz}^{-1}U_{n,t-1}-u_{n,t}'\Sigma_{zz}^{-1}u_{n,t})+\xi+o_p(1), \\ &= U_{n,n}'\Sigma_{zz}^{-1}U_{n,n}-\tfrac{n}{n_\rho}U_{n,n_\rho}'\Sigma_{zz}^{-1}U_{n,n_\rho}+\tfrac{1}{n}\sum_{t=n_\rho+1}^{n}(\tfrac{n}{t})^2U_{n,t}'\Sigma_{zz}^{-1}U_{n,t}+\sigma_\varepsilon^2\kappa\log\rho+o_p(1), \quad \text{(A.3)} \end{aligned}$$

where we used that $\sigma_\varepsilon^{-2}\kappa=\mathrm{tr}\{\Sigma_{zz}^{-1}\Omega\}=\sum_{j=-h+1}^{h-1}\mathrm{tr}\{\Sigma_{zz}^{-1}\mathrm{E}[Z_{t-h-j}\varepsilon_{t-j}\varepsilon_tZ_{t-h}']\}=\sum_{j=-h+1}^{h-1}\gamma_j$ under Assumption 3. The penultimate term in (A.3) offsets the contributions from D, whereas $A+B$ equals

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

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

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

and the result now follows.□

Proof of Corollary 2. Let $U=\frac{B(1)-B(\rho)}{\sqrt{1-\rho}}$ and $V=\frac{B(\rho)}{\sqrt{\rho}}$, so that $B(1)=\sqrt{1-\rho}U+\sqrt{\rho}V$, and note that U and V are independent standard Gaussian random variables. Expressing the distribution we

seek 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 decomposing the 2×2 symmetric matrix into $Q'\Lambda Q$, where $Q'Q = I$ and $\Lambda = \mathrm{diag}(\sqrt{1-\rho}, -\sqrt{1-\rho})$ (the eigenvalues) the expression simplifies to $\sqrt{1-\rho}(Z_1^2 - Z_2^2)$ where $Z = Q(U,V)' \sim N_2(0,I)$. □

Proof of Corollary 3. Let Z_{1i}, Z_{2i}, $i = 1,\ldots,q$ 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 we seek is given by the convolution,

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

so we seek the distribution of $S = X - Y$ where X and Y are independent χ_q^2-distributed random variables. The density of a χ_q^2 is

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

and we are interested in the convolution of X and $-Y$

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

For $s < 0$ the density is $2^{-q}\Gamma(\frac{q}{2})^{-2}e^{s/2}\int_0^\infty \left(u(u-s)\right)^{q/2-1} e^{-u}\mathrm{d}u$. By taking advantage of the symmetry about zero, we obtain the expression

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

When $q = 1$ this simplifies to $f_1(s) = \frac{1}{2\pi}\mathrm{B}_0(\frac{|s|}{2})$ where $\mathrm{B}_k(x)$ denotes the modified Bessel function of the second kind. For $q = 2$ we have the simpler expression $f_2(x) = \frac{1}{4}e^{-\frac{|s|}{2}}$. □