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

Peter Reinhard Hansen[1,3] and Allan Timmermann[2,3]

[1]European University Institute
[2]University of California, San Diego, Rady
[3]CREATES

Outline

- Out-of-sample tests of predictive accuracy are used extensively throughout economics and finance.
- Regarded by many researchers as the "ultimate test of a forecasting model" to quote: Stock and Watson (2007).
- Frequently done with the approach by West (1996), McCracken (2007), and Clark & McCracken (2001,2005).
  - Linear Regression models, estimated with past data, e.g. recursively, or by rolling window.

Outline

- Out-of-sample tests of predictive accuracy are used extensively throughout economics and finance.
- Regarded by many researchers as the "ultimate test of a forecasting model" to quote: Stock and Watson (2007).
- Frequently done with the approach by West (1996), McCracken (2007), and Clark & McCracken (2001,2005).
  - Linear Regression models, estimated with past data, e.g. recursively, or by rolling window.

Outline

- Out-of-sample tests of predictive accuracy are used extensively throughout economics and finance.
- Regarded by many researchers as the "ultimate test of a forecasting model" to quote: Stock and Watson (2007).
- Frequently done with the approach by West (1996), McCracken (2007), and Clark & McCracken (2001,2005).
    - Linear Regression models, estimated with past data, e.g. recursively, or by rolling window.

Outline

- Out-of-sample tests of predictive accuracy are used extensively throughout economics and finance.
- Regarded by many researchers as the "ultimate test of a forecasting model" to quote: Stock and Watson (2007).
- Frequently done with the approach by West (1996), McCracken (2007), and Clark & McCracken (2001,2005).
  - Linear Regression models, estimated with past data, e.g. recursively, or by rolling window.

A Predictive Regression Model

- Predictive regression model for an h-period forecast horizon

$$y_{t+h} = \beta' X_t + \varepsilon_{t+h}, \qquad t = 1, \ldots, n$$

where $X_t \in \mathbb{R}^k$.

- Recursive least squares. Obtain $\hat{\beta}_t$ by regressing y_s on X_{s-h}, for $s = 1, \ldots, t$.
- Forecast

$$\hat{y}_{t+h|t}(\hat{\beta}_t) = \hat{\beta}_t' X_t.$$

A Predictive Regression Model

- Predictive regression model for an h-period forecast horizon

$$y_{t+h} = \beta' X_t + \varepsilon_{t+h}, \qquad t = 1, \ldots, n$$

  where $X_t \in \mathbb{R}^k$.
- Recursive least squares. Obtain $\hat{\beta}_t$ by regressing y_s on X_{s-h}, for $s = 1, \ldots, t$.
- Forecast

$$\hat{y}_{t+h|t}(\hat{\beta}_t) = \hat{\beta}_t' X_t.$$

A Predictive Regression Model

- Predictive regression model for an h-period forecast horizon

$$y_{t+h} = \beta' X_t + \varepsilon_{t+h}, \qquad t = 1, \ldots, n$$

where $X_t \in \mathbb{R}^k$.

- Recursive least squares. Obtain $\hat{\beta}_t$ by regressing y_s on X_{s-h}, for $s = 1, \ldots, t$.
- Forecast

$$\hat{y}_{t+h|t}(\hat{\beta}_t) = \hat{\beta}_t' X_t.$$

Another Predictive Regression Model

- Predictive regression model with fewer regressors

$$y_{t+h} = \delta' \tilde{X}_t + \eta_{t+h}, \qquad t = 1, \ldots, n,$$

$\tilde{X}_t \in \mathbb{R}^{\tilde{k}}$.

- Now

$$\hat{\delta}_t = \left(\sum_{s=1}^{t} \tilde{X}_{s-h} \tilde{X}'_{s-h} \right)^{-1} \sum_{s=1}^{t} \tilde{X}_{s-h} y_s$$

- and

$$\tilde{y}_{t+h|t}(\hat{\delta}_t) = \hat{\delta}'_t \tilde{X}_t.$$

Another Predictive Regression Model

- Predictive regression model with fewer regressors

$$y_{t+h} = \delta' \tilde{X}_t + \eta_{t+h}, \qquad t = 1, \ldots, n,$$

$\tilde{X}_t \in \mathbb{R}^{\tilde{k}}$.

- Now

$$\hat{\delta}_t = \left(\sum_{s=1}^{t} \tilde{X}_{s-h} \tilde{X}'_{s-h} \right)^{-1} \sum_{s=1}^{t} \tilde{X}_{s-h} y_s$$

- and

$$\tilde{y}_{t+h|t}(\hat{\delta}_t) = \hat{\delta}'_t \tilde{X}_t.$$

Another Predictive Regression Model

- Predictive regression model with fewer regressors

$$y_{t+h} = \delta' \tilde{X}_t + \eta_{t+h}, \qquad t = 1, \ldots, n,$$

$\tilde{X}_t \in \mathbb{R}^{\tilde{k}}$.

- Now

$$\hat{\delta}_t = \left(\sum_{s=1}^{t} \tilde{X}_{s-h} \tilde{X}_{s-h}' \right)^{-1} \sum_{s=1}^{t} \tilde{X}_{s-h} y_s$$

- and

$$\tilde{y}_{t+h|t}(\hat{\delta}_t) = \hat{\delta}_t' \tilde{X}_t.$$

The Null Hypothesis

- 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:

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

- Note: In nested case, $\tilde{X}_t \subset X_t$, equivalent to testing $H_0' : \beta_2 = 0$ (where $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$).

The Null Hypothesis

- 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:

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

- Note: In nested case, $\tilde{X}_t \subset X_t$, equivalent to testing $H_0' : \beta_2 = 0$ (where $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$).

MSE Statistics

- Consider the difference of the resulting out-of-sample MSEs

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

where $n_\rho = \lfloor \rho n \rfloor$ with $0 < \rho < 1$, is the number of observation set aside for the initial estimation.

A Test Statistic for Nested Case

- In nested case, $\tilde{X}_t \subset X_t$, $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$.
- McCracken (2007) established the limit distribution of

$$T_n = \frac{\Delta \mathrm{MSE}_n}{\hat{\sigma}_\varepsilon}$$

for the case $h = 1$ and homoskedastic errors.

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

where $B_i(u)$ are mutually independent standard Brownian motions.

- $q = k - \tilde{k} = \dim(\beta_2)$ (number of extra regressors in larger model).
- McCracken tabulated critical values using simulations.

A Test Statistic for Nested Case

- In nested case, $\tilde{X}_t \subset X_t$, $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$.
- McCracken (2007) established the limit distribution of

$$T_n = \frac{\Delta\mathrm{MSE}_n}{\hat{\sigma}_\varepsilon}$$

for the case $h = 1$ and homoskedastic errors.

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

  where $B_i(u)$ are mutually independent standard Brownian motions.
- $q = k - \tilde{k} = \dim(\beta_2)$ (number of extra regressors in larger model).
- McCracken tabulated critical values using simulations.

A Test Statistic for Nested Case

- In nested case, $\tilde{X}_t \subset X_t$, $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$.
- McCracken (2007) established the limit distribution of

$$T_n = \frac{\Delta \mathrm{MSE}_n}{\hat{\sigma}_\varepsilon}$$

  for the case $h = 1$ and homoskedastic errors.
- $$T_n \xrightarrow{d} \sum_{i=1}^{q} \left[2 \int_\rho^1 u^{-1} B_i(u) \mathrm{d} B_i(u) - \int_\rho^1 u^{-2} B_i(u)^2 \mathrm{d} u \right],$$

  where $B_i(u)$ are mutually independent standard Brownian motions.
- $q = k - \tilde{k} = \dim(\beta_2)$ (number of extra regressors in larger model).
- McCracken tabulated critical values using simulations.

A Test Statistic for Nested Case

- In nested case, $\tilde{X}_t \subset X_t$, $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$.
- McCracken (2007) established the limit distribution of

$$T_n = \frac{\Delta \mathrm{MSE}_n}{\hat{\sigma}_\varepsilon}$$

for the case $h = 1$ and homoskedastic errors.

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

where $B_i(u)$ are mutually independent standard Brownian motions.

- $q = k - \tilde{k} = \dim(\beta_2)$ (number of extra regressors in larger model).
- McCracken tabulated critical values using simulations.

A Test Statistic for Nested Case

- In nested case, $\tilde{X}_t \subset X_t$, $\beta = (\beta_1', \beta_2')'$ and $\beta_1 = \delta$.
- McCracken (2007) established the limit distribution of

$$T_n = \frac{\Delta \mathrm{MSE}_n}{\hat{\sigma}_\varepsilon}$$

for the case $h = 1$ and homoskedastic errors.

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

where $B_i(u)$ are mutually independent standard Brownian motions.
- $q = k - \tilde{k} = \dim(\beta_2)$ (number of extra regressors in larger model).
- McCracken tabulated critical values using simulations.

More on the Nested Case

- Limit distribution for the general case ($h \geq 1$ and heteroskedastic error) derived by Clark and McCracken (2005).
- Their expression simplified by Stock and Watson (2003) to:

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

where λ_i are eigenvalues (to be defined).

More on the Nested Case

- Limit distribution for the general case ($h \geq 1$ and heteroskedastic error) derived by Clark and McCracken (2005).
- Their expression simplified by Stock and Watson (2003) to:

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

where λ_i are eigenvalues (to be defined).

Wald Statistic

- Consider quadratic form statistic

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

- Conventional Wald statistic ($H_0 : \beta = 0$) takes the form

$$W_n = \hat{\sigma}_{\varepsilon}^{-2} \hat{\beta}'_n \left(\sum_{t=1}^{n} X_{t-h} X'_{t-h} \right) \hat{\beta}_n \quad = \frac{S_n}{\hat{\sigma}_{\varepsilon}^2} .$$

Wald Statistic

- Consider quadratic form statistic

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

- Conventional Wald statistic ($H_0 : \beta = 0$) takes the form

$$W_n = \hat{\sigma}_\varepsilon^{-2} \hat{\beta}'_n \left(\sum_{t=1}^{n} X_{t-h} X'_{t-h} \right) \hat{\beta}_n \quad = \frac{S_n}{\hat{\sigma}_\varepsilon^2}.$$

Contributions I: Equivalence

- Test statistic

$$\Delta\mathrm{MSE}_n = S_n - S_{n_\rho} - \tilde{S}_n + \tilde{S}_{n_\rho} + \mathrm{constant} + o_p(1),$$

- where $\Delta\mathrm{MSE}_n = \sum_{t=n_\rho+1}^{n}(y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2$ and

$$\begin{aligned} S_n &= \sum_{t=1}^{n} y_t X_{t-h}' \left[\sum_{t=1}^{n} X_{t-h}X_{t-h}'\right]^{-1} \sum_{t=1}^{n} X_{t-h}y_t, \\ \tilde{S}_n &= \sum_{t=1}^{n} y_t \tilde{X}_{t-h}' \left[\sum_{t=1}^{n} \tilde{X}_{t-h}\tilde{X}_{t-h}'\right]^{-1} \sum_{t=1}^{n} \tilde{X}_{t-h}y_t. \end{aligned}$$

Contributions I: Equivalence

- Test statistic

$$\Delta \mathrm{MSE}_n = S_n - S_{n_\rho} - \tilde{S}_n + \tilde{S}_{n_\rho} + \mathrm{constant} + o_p(1),$$

- where $\Delta \mathrm{MSE}_n = \sum_{t=n_\rho+1}^{n} (y_t - \tilde{y}_{t|t-h})^2 - (y_t - \hat{y}_{t|t-h})^2$ and

$$S_n = \sum_{t=1}^{n} y_t X'_{t-h} \left[\sum_{t=1}^{n} X_{t-h} X'_{t-h} \right]^{-1} \sum_{t=1}^{n} X_{t-h} y_t,$$

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

Contributions I: Equivalence in Nested Case

- Test statistic

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

- Instead

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

  - where W_m is conventional Wald statistic for $H_0 : \beta = 0$ using observations $t = 1, \ldots, m$.
  - Just difference of two Wald tests (aside from constant)

Contributions I: Equivalence in Nested Case

- Test statistic

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

- Instead

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

  - where W_m is conventional Wald statistic for $H_0 : \beta = 0$ using observations $t = 1, \ldots, m$.
  - Just difference of two Wald tests (aside from constant)

Contributions I: Equivalence in Nested Case

- Test statistic

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

- Instead

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

  - where W_m is conventional Wald statistic for $H_0 : \beta = 0$ using observations $t = 1, \ldots, m$.
  - Just difference of two Wald tests (aside from constant)

Contributions I: Equivalence in Nested Case

- Test statistic

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

- Instead

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

  - where W_m is conventional Wald statistic for $H_0 : \beta = 0$ using observations $t = 1, \ldots, m$.
  - Just difference of two Wald tests (aside from constant)

Contributions II: Simplification

- Limit distribution which involves the stochastic integrals

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B(u)^2\mathrm{d}u = B^2(1) - \rho^{-1}B^2(\rho) + \log\rho.$$

- Simply difference of two (dependent) χ_1^2s — plus constant $\log\rho$.
- Moreover,

$$B^2(1) - \rho^{-1}B^2(\rho) = \sqrt{1-\rho}(Z_1^2 - Z_2^2)$$

  (two independent χ_1^2).
- Special case $q = 2: \; T_n(\rho) \xrightarrow{d}$ double-exponential.

Contributions II: Simplification

- Limit distribution which involves the stochastic integrals

$$2\int_\rho^1 u^{-1}B(u)\mathrm{d}B(u) - \int_\rho^1 u^{-2}B(u)^2\mathrm{d}u = B^2(1) - \rho^{-1}B^2(\rho) + \log\rho.$$

- Simply difference of two (dependent) χ_1^2s — plus constant $\log\rho$.
- Moreover,

$$B^2(1) - \rho^{-1}B^2(\rho) = \sqrt{1-\rho}(Z_1^2 - Z_2^2)$$

(two independent χ_1^2).

- Special case $q = 2$: $T_n(\rho) \xrightarrow{d}$ double-exponential.

Contributions II: Simplification

- Limit distribution which involves the stochastic integrals

$$2\int_\rho^1 u^{-1}B(u)\mathrm{d}B(u) - \int_\rho^1 u^{-2}B(u)^2\mathrm{d}u = B^2(1) - \rho^{-1}B^2(\rho) + \log\rho.$$

- Simply difference of two (dependent) χ_1^2s — plus constant $\log\rho$.
- Moreover,

$$B^2(1) - \rho^{-1}B^2(\rho) = \sqrt{1-\rho}(Z_1^2 - Z_2^2)$$

  (two independent χ_1^2).
- Special case $q = 2:\ T_n(\rho) \xrightarrow{d}$ double-exponential.

Contributions II: Simplification

- Limit distribution which involves the stochastic integrals

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B(u)^2\mathrm{d}u = B^2(1) - \rho^{-1}B^2(\rho) + \log\rho.$$

- Simply difference of two (dependent) χ_1^2s — plus constant $\log\rho$.
- Moreover,

$$B^2(1) - \rho^{-1}B^2(\rho) = \sqrt{1-\rho}(Z_1^2 - Z_2^2)$$

  (two independent χ_1^2).
- Special case $q = 2:\ T_n(\rho) \xrightarrow{d}$ double-exponential.

Assumption 1

- For some positive definite matrix, Σ, we have

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

Assumption 2

- "Scores" $X_{t-h}\varepsilon_t$ play an important role.
  Define
  $$u_{n,t} = n^{-1/2} X_{t-h}\varepsilon_t.$$

- $$\sup_{u\in[0,1]} \left\| n^{-1} \sum_{t=1}^{\lfloor nu \rfloor} u_{n,t} u'_{n,t-j} - u\Gamma_j \right\| = o_p(1),$$
  for some Γ_j, $j = 0, 1, \ldots, h-1$.
- Define (*nearly* long-run variance)
  $$\Omega = \sum_{j=-h+1}^{h-1} \Gamma_j.$$

Assumption 2

- "Scores" $X_{t-h}\varepsilon_t$ play an important role.
  Define
  $$u_{n,t} = n^{-1/2} X_{t-h}\varepsilon_t.$$

- $$\sup_{u\in[0,1]} \left\| n^{-1} \sum_{t=1}^{\lfloor nu \rfloor} u_{n,t} u'_{n,t-j} - u\Gamma_j \right\| = o_p(1),$$
  for some Γ_j, $j = 0, 1, \ldots, h-1$.

- Define (*nearly* long-run variance)
  $$\Omega = \sum_{j=-h+1}^{h-1} \Gamma_j.$$

Assumption 2

- "Scores" $X_{t-h}\varepsilon_t$ play an important role.
  Define
  $$u_{n,t} = n^{-1/2} X_{t-h}\varepsilon_t.$$
-
  $$\sup_{u\in[0,1]} \left\| n^{-1} \sum_{t=1}^{\lfloor nu \rfloor} u_{n,t} u'_{n,t-j} - u\Gamma_j \right\| = o_p(1),$$
  for some Γ_j, $j = 0, 1, \ldots, h-1$.
- Define (*nearly* long-run variance)
  $$\Omega = \sum_{j=-h+1}^{h-1} \Gamma_j.$$

Assumption 3

- Define

$$U_n(s) = \sum_{t=1}^{\lfloor ns \rfloor} u_{n,t} = n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} X_{t-h}\varepsilon_t.$$

- For some $U \in \mathbb{D}^k_{[0,1]}$, which is bounded in probability,
-

$$U_n(s) \Rightarrow U(s).$$

- (U is a Brownian motion in the canonical case).

Assumption 3

- Define

$$U_n(s) = \sum_{t=1}^{\lfloor ns \rfloor} u_{n,t} = n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} X_{t-h}\varepsilon_t.$$

- For some $U \in \mathbb{D}^k_{[0,1]}$, which is bounded in probability,
-

$$U_n(s) \Rightarrow U(s).$$

- (U is a Brownian motion in the canonical case).

Assumption 3

- Define

$$U_n(s) = \sum_{t=1}^{\lfloor ns \rfloor} u_{n,t} = n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} X_{t-h}\varepsilon_t.$$

- For some $U \in \mathbb{D}^k_{[0,1]}$, which is bounded in probability,
-

$$U_n(s) \Rightarrow U(s).$$

- (U is a Brownian motion in the canonical case).

Assumption 3

- Define

$$U_n(s) = \sum_{t=1}^{\lfloor ns \rfloor} u_{n,t} = n^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} X_{t-h}\varepsilon_t.$$

- For some $U \in \mathbb{D}^k_{[0,1]}$, which is bounded in probability,
-
$$U_n(s) \Rightarrow U(s).$$

- (U is a Brownian motion in the canonical case).

Assumption 4

- Define

$$M_t = \frac{1}{t}\sum_{s=1}^{t} X_t X_t'.$$

- We need

$$\sum_{t=n_\rho+1}^{n} U_{n,t-h}'(M_{t-h}^{-1} - \Sigma^{-1})u_{n,t} = o_p(1),$$

$$\frac{1}{n}\sum_{t=n_\rho+1}^{n} U_{n,t-h}'(M_{t-h}^{-1}X_{t-h}X_{t-h}'M_{t-h}^{-1} - \Sigma^{-1})U_{n,t-h} = o_p(1).$$

Assumption 4

- Define

$$M_t = \frac{1}{t}\sum_{s=1}^{t} X_t X_t'.$$

- We need

$$\begin{aligned}\sum_{t=n_\rho+1}^{n} U_{n,t-h}'(M_{t-h}^{-1} - \Sigma^{-1})u_{n,t} &= o_p(1),\\ \frac{1}{n}\sum_{t=n_\rho+1}^{n} U_{n,t-h}'(M_{t-h}^{-1}X_{t-h}X_{t-h}'M_{t-h}^{-1} - \Sigma^{-1})U_{n,t-h} &= o_p(1).\end{aligned}$$

Theorem 1: Simple No-Change Forecast

- Given Assumptions 1-4

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

- where

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

- True for any value of β.

Theorem 1: Simple No-Change Forecast

- Given Assumptions 1-4

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

- where

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

- True for any value of β.

Theorem 1: Simple No-Change Forecast

- Given Assumptions 1-4

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

- where

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

- True for any value of β.

Corollary 1: Compare Any Two

- Given Assumptions 1-4 (for both models)

$$\sum_{t=n_\rho+1}^{n} (y_t - \tilde{y}_{t|t-h}(\hat{\delta}_{t-h}))^2 - (y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h}))^2$$

equals

$$S_n - S_{n_\rho} - \tilde{S}_n + \tilde{S}_{n_\rho} + (\kappa - \tilde{\kappa}) \log \rho + o_p(1),$$

- where

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

Corollary 1: Compare Any Two

- Given Assumptions 1-4 (for both models)

$$\sum_{t=n_\rho+1}^{n} (y_t - \tilde{y}_{t|t-h}(\hat{\delta}_{t-h}))^2 - (y_t - \hat{y}_{t|t-h}(\hat{\beta}_{t-h}))^2$$

equals

$$S_n - S_{n_\rho} - \tilde{S}_n + \tilde{S}_{n_\rho} + (\kappa - \tilde{\kappa}) \log \rho + o_p(1),$$

- where

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

Nested Case

- Suppose

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

and

$$y_{t+h} = \delta' X_{1t} \qquad\qquad\qquad + \eta_{t+h}, \qquad t = 1, \ldots, n.$$

Auxiliary Regressor (infeasible)

- Write

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

and define

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

which captures the part of X_{2t} that is orthogonal to X_{1t}.

Auxiliary Regressor (feasible)

- Sample equivalent

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

- Used to compute

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

(variation of y_t explained by $X_{2,t-h}$, which is not explained by $X_{1,t-h}$).

Auxiliary Regressor (feasible)

- Sample equivalent

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

- Used to compute

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

(variation of y_t explained by $X_{2,t-h}$, which is not explained by $X_{1,t-h}$).

Theorem 2: Compare Nested Models

- Given Assumptions 1-4

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

equals

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

- where $\check{\kappa} = \kappa - \tilde{\kappa}$.

Theorem 2: Compare Nested Models

- Given Assumptions 1-4

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

equals

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

- where $\check{\kappa} = \kappa - \tilde{\kappa}$.

Theorem 2 (cont): Nested Local Alternative

- $\check{W}_n - \check{W}_{n_\rho} + \sigma_\varepsilon^{-2}\check{\kappa}\log\rho + o_p(1)$ with $\check{\kappa} = \kappa - \tilde{\kappa}$.
- If
$$\beta_2 = n^{-1/2}b$$
with $b \in \mathbb{R}^q$ fixed, then
$$\check{\kappa} = \mathrm{tr}\{\check{\Sigma}^{-1}\check{\Omega}\},$$
- where $\check{\Omega}$ is long-run variance of $\{Z_{n,t-h}\varepsilon_t\}$, and
$$\check{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Theorem 2 (cont): Nested Local Alternative

- $\check{W}_n - \check{W}_{n_\rho} + \sigma_\varepsilon^{-2}\check{\kappa}\log\rho + o_p(1)$ with $\check{\kappa} = \kappa - \tilde{\kappa}$.
- If

$$\beta_2 = n^{-1/2} b$$

with $b \in \mathbb{R}^q$ fixed, then

$$\check{\kappa} = \mathrm{tr}\{\check{\Sigma}^{-1}\check{\Omega}\},$$

- where $\check{\Omega}$ is long-run variance of $\{Z_{n,t-h}\varepsilon_t\}$, and

$$\check{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Theorem 2 (cont): Nested Local Alternative

- $\check{W}_n - \check{W}_{n_\rho} + \sigma_\varepsilon^{-2}\check{\kappa}\log\rho + o_p(1)$ with $\check{\kappa} = \kappa - \tilde{\kappa}$.
- If

$$\beta_2 = n^{-1/2}b$$

  with $b \in \mathbb{R}^q$ fixed, then

$$\check{\kappa} = \mathrm{tr}\{\check{\Sigma}^{-1}\check{\Omega}\},$$

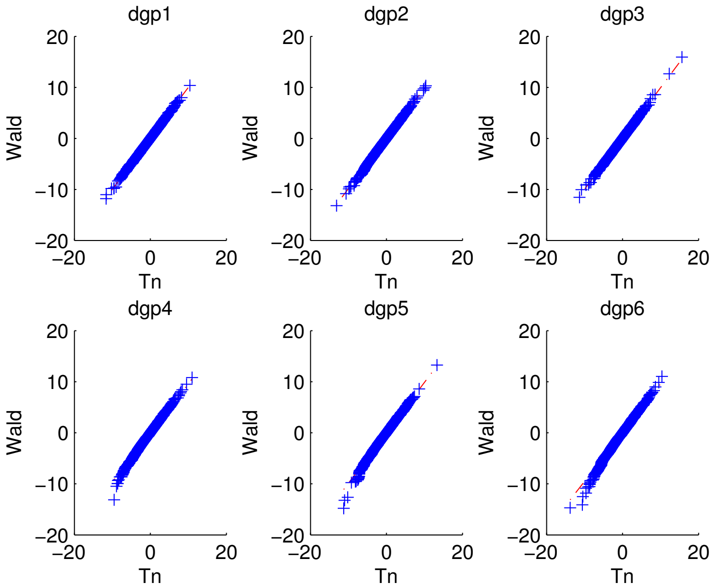
- where $\check{\Omega}$ is long-run variance of $\{Z_{n,t-h}\varepsilon_t\}$, and

$$\check{\Sigma} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Finite Sample Correlation (n=200)

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

Q-Q Plot (n=500)



Assumption 5

- $$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} Z_{t-h} \varepsilon_t \Rightarrow \check{\Omega}_{\infty}^{1/2} B(u),$$

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

Eigenvalues

- Consider
$$\Xi = \sigma_{\varepsilon}^{-2}\check{\Sigma}^{-1}\check{\Omega}_{\infty}.$$

- Diagonalize, so that
$$\Xi = Q'\Lambda Q,$$
where $Q'Q = I$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_q)$.

Eigenvalues

- Consider
$$\Xi = \sigma_{\varepsilon}^{-2}\check{\Sigma}^{-1}\check{\Omega}_{\infty}.$$

- Diagonalize, so that
$$\Xi = Q'\Lambda Q,$$
where $Q'Q = I$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_q)$.

Theorem 3 (Null)

- Under the null hypothesis ($\beta_2 = 0$)

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

where $B = (B_1, \ldots, B_q)'$ is a standard q-dimensional Brownian motion.

- This limit distribution is identical to

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

Theorem 3 (Null)

- Under the null hypothesis ($\beta_2 = 0$)

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

where $B = (B_1, \ldots, B_q)'$ is a standard q-dimensional Brownian motion.

- This limit distribution is identical to

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

Theorem 3 (all of it)

- Consider the local alternative

$$\beta_2 = cn^{-1/2}b,$$

(normalized so that $\sigma_\varepsilon^{-2} b' \check{\Sigma} b = \kappa$)

- $T_n \xrightarrow{d}$

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

where $a = b' \check{\Sigma} \check{\Omega}_\infty^{1/2} Q'$.

Theorem 3 (all of it)

- Consider the local alternative

$$\beta_2 = cn^{-1/2} b,$$

(normalized so that $\sigma_\varepsilon^{-2} b' \check{\Sigma} b = \kappa$)

- $T_n \xrightarrow{d}$

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

where $a = b' \check{\Sigma} \check{\Omega}_\infty^{1/2} Q'$.

Proof of Simplification

- Consider (for $u > 0$)

$$F(u) = \frac{1}{u} B^2(u) - \log u.$$

- By Ito stochastic calculus:

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

- So $\int_\rho^1 \frac{2}{u}B\mathrm{d}B - \int_\rho^1 \frac{1}{u^2}B^2\mathrm{d}u = \int_\rho^1 \mathrm{d}F(u)$ equals

$$F(1) - F(\rho) = B^2(1) - \log 1 - B^2(\rho)/\rho + \log \rho.$$

- Moreover. Same as

$$\sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho, \qquad Z_i \sim \mathrm{iid} N(0,1)$$

Proof of Simplification

- Consider (for $u > 0$)

$$F(u) = \frac{1}{u} B^2(u) - \log u.$$

- By Ito stochastic calculus:

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

- So $\int_\rho^1 \frac{2}{u}B\mathrm{d}B - \int_\rho^1 \frac{1}{u^2}B^2\mathrm{d}u = \int_\rho^1 \mathrm{d}F(u)$ equals

$$F(1) - F(\rho) = B^2(1) - \log 1 - B^2(\rho)/\rho + \log \rho.$$

- Moreover. Same as

$$\sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho, \qquad Z_i \sim \mathrm{iid} N(0,1)$$

Proof of Simplification

- Consider (for $u > 0$)

$$F(u) = \frac{1}{u} B^2(u) - \log u.$$

- By Ito stochastic calculus:

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

- So $\int_\rho^1 \frac{2}{u} B \mathrm{d}B - \int_\rho^1 \frac{1}{u^2} B^2 \mathrm{d}u = \int_\rho^1 \mathrm{d}F(u)$ equals

$$F(1) - F(\rho) = B^2(1) - \log 1 - B^2(\rho)/\rho + \log \rho.$$

- Moreover. Same as

$$\sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho, \qquad Z_i \sim \mathrm{iid} N(0,1)$$

Proof of Simplification

- Consider (for $u > 0$)

$$F(u) = \frac{1}{u}B^2(u) - \log u.$$

- By Ito stochastic calculus:

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

- So $\int_\rho^1 \frac{2}{u}B\mathrm{d}B - \int_\rho^1 \frac{1}{u^2}B^2\mathrm{d}u = \int_\rho^1 \mathrm{d}F(u)$ equals

$$F(1) - F(\rho) = B^2(1) - \log 1 - B^2(\rho)/\rho + \log \rho.$$

- Moreover. Same as

$$\sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho, \qquad Z_i \sim \mathrm{iid}N(0,1)$$

Theorem 4

- Let B be a univariate

$$2\int_{\rho}^{1} u^{-1} B \mathrm{d}B - \int_{\rho}^{1} u^{-2} B^2 \mathrm{d}u \overset{d}{=} \sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho,$$

where $Z_i \sim \mathrm{iid} N(0,1)$.

- So simple a difference between two independent chi-squares

Theorem 4

- Let B be a univariate

$$2\int_{\rho}^{1} u^{-1} B \mathrm{d}B - \int_{\rho}^{1} u^{-2} B^2 \mathrm{d}u \overset{d}{=} \sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho,$$

where $Z_i \sim \mathrm{iid} N(0,1)$.

- So simple a difference between two independent chi-squares

Analytical Results Beats Simulations

- McCracken:

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B^2(u)\mathrm{d}u.$$

- Critical values requires extensive simulations.
  - Brownian motion $B(u)$ discretized by $n^{-1/2}\sum_{i=1}^{\lfloor un \rfloor}\varepsilon_i$ with $\varepsilon_i \sim \mathrm{iid}N(0,1)$
  - N repetitions
- With $n = 5,000$ and $N = 10,000$...

  ... takes $50,000,000$ random variables to compute a critical value.

Analytical Results Beats Simulations

- McCracken:

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B^{2}(u)\mathrm{d}u.$$

- Critical values requires extensive simulations.
  - Brownian motion $B(u)$ discretized by $n^{-1/2}\sum_{i=1}^{\lfloor un \rfloor} \varepsilon_i$ with $\varepsilon_i \sim \mathrm{iid}N(0,1)$
  - N repetitions
- With $n = 5,000$ and $N = 10,000$...

  ... takes $50,000,000$ random variables to compute a critical value.

Analytical Results Beats Simulations

- McCracken:

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B^2(u)\mathrm{d}u.$$

- Critical values requires extensive simulations.
  - Brownian motion $B(u)$ discretized by $n^{-1/2}\sum_{i=1}^{\lfloor un \rfloor} \varepsilon_i$ with $\varepsilon_i \sim \mathrm{iid}N(0,1)$
  - N repetitions
- With $n = 5,000$ and $N = 10,000$...

  ... takes $50,000,000$ random variables to compute a critical value.

Analytical Results Beats Simulations

- McCracken:

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B^2(u)\mathrm{d}u.$$

- Critical values requires extensive simulations.
  - Brownian motion $B(u)$ discretized by $n^{-1/2}\sum_{i=1}^{\lfloor un \rfloor} \varepsilon_i$ with $\varepsilon_i \sim \mathrm{iid}N(0,1)$
  - N repetitions
- With $n = 5,000$ and $N = 10,000$...

  ... takes $50,000,000$ random variables to compute a critical value.

Analytical Results Beats Simulations

- McCracken:

$$2\int_{\rho}^{1} u^{-1}B(u)\mathrm{d}B(u) - \int_{\rho}^{1} u^{-2}B^{2}(u)\mathrm{d}u.$$

- Critical values requires extensive simulations.
  - Brownian motion $B(u)$ discretized by $n^{-1/2}\sum_{i=1}^{\lfloor un \rfloor} \varepsilon_i$ with $\varepsilon_i \sim \mathrm{iid}N(0,1)$
  - N repetitions
- With $n = 5,000$ and $N = 10,000$...
  ... takes $50,000,000$ random variables to compute a critical value.

Critical Values: Simulated vs Analytical (q=2)

$\rho =$	0.909	0.833	0.625	0.500	0.417	0.357	0.333
$\pi =$	0.1	0.2	0.6	1	1.4	1.8	2
$\alpha =$	2.168	2.830	3.851	4.146	4.225	4.214	4.191
0.99	*1.996*	*2.691*	*3.907*	*4.200*	*4.304*	*4.278*	*4.250*
$\alpha =$	1.198	1.515	1.880	1.870	1.766	1.633	1.563
0.95	*1.184*	*1.453*	*1.891*	*1.802*	*1.752*	*1.692*	*1.706*

1nd row: Analytical using non-central Laplace distribution.

2st row: Simulated critical values from McCracken (2007). $\pi = (1 - \rho)/\rho$.
(Discrepancies have little practical relevance, as the size distortions are very small).

Conclusion

- Equivalence of commonly used test statistic and Wald statistics.
  - Greatly simplifies both the computation of the test statistic and the expression for its limit distribution.
  - We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point, ρ.
- 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 (the conventional full sample Wald test can easily be adapted to the heteroskedastic case).

Conclusion

- Equivalence of commonly used test statistic and Wald statistics.
  - Greatly simplifies both the computation of the test statistic and the expression for its limit distribution.
  - We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point, ρ.
- 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 (the conventional full sample Wald test can easily be adapted to the heteroskedastic case).

Conclusion

- Equivalence of commonly used test statistic and Wald statistics.
  - Greatly simplifies both the computation of the test statistic and the expression for its limit distribution.
  - We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point, ρ.
- 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 (the conventional full sample Wald test can easily be adapted to the heteroskedastic case).

Conclusion

- Equivalence of commonly used test statistic and Wald statistics.
    - Greatly simplifies both the computation of the test statistic and the expression for its limit distribution.
    - We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point, ρ.
- 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 (the conventional full sample Wald test can easily be adapted to the heteroskedastic case).

Conclusion

- Equivalence of commonly used test statistic and Wald statistics.
  - Greatly simplifies both the computation of the test statistic and the expression for its limit distribution.
  - We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point, ρ.
- 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 (the conventional full sample Wald test can easily be adapted to the heteroskedastic case).

Conclusion

- Equivalence of commonly used test statistic and Wald statistics.
  - Greatly simplifies both the computation of the test statistic and the expression for its limit distribution.
  - We also establish local power properties of the test. These show that the power of the test is decreasing in the sample split point, ρ.
- 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 (the conventional full sample Wald test can easily be adapted to the heteroskedastic case).

Proof of Simplification

Note that

$$B(1) = B(1) - B(\rho) + B(\rho) = \sqrt{1-\rho}\frac{B(1) - B(\rho)}{\sqrt{1-\rho}} + \sqrt{\rho}\frac{B(\rho)}{\sqrt{\rho}},$$

where

$$U = \frac{B(1) - B(\rho)}{\sqrt{1-\rho}} \qquad \text{and} \qquad V = \frac{B(\rho)}{\sqrt{\rho}},$$

are independent standard Gaussian random variables. Thus the distribution we seek is that of

$$W = \left(\sqrt{1-\rho}U + \sqrt{\rho}V\right)^2 - V^2 + \log\rho,$$

where $U, V \sim \mathrm{iid} N(0,1)$.

Proof of Simplification

Expressed in a quadratic from:

$$W = \begin{pmatrix} U \\ V \end{pmatrix}' \begin{pmatrix} 1-\rho & \sqrt{\rho(1-\rho)} \\ \sqrt{\rho(1-\rho)} & \rho - 1 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} + \log \rho.$$

We can now use the fact that any real symmetric matrix, A, can decomposed into $A = Q\Lambda Q'$ where $Q'Q = I$ and Λ is a diagonal matrix with the eigenvalues of A in the diagonal. This leads to

$$W = Z' \begin{pmatrix} \sqrt{1-\rho} & 0 \\ 0 & -\sqrt{1-\rho} \end{pmatrix} Z + \log \rho,$$

where $Z \sim N_2(0, I)$ (a simple rotation of U, V). So it now follows that

$$W = \sqrt{1-\rho}(Z_1^2 - Z_2^2) + \log \rho.$$

I.e. a scaled difference between to independent chi-squares plus $\log \rho$.

Proof of Simplification

Let X and Y be independent χ^2_q and consider $S = X - Y$. The density is of a χ^2_q is

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

and we seek the convolution between X and $-Y$

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

Proof of Simplification

Simplest if $q = 2$

$$q = 2 \Rightarrow \tfrac{1}{4} e^{s/2} \int_{0 \vee s}^{\infty} \left(u(u-s)\right)^0 e^{-u} \mathrm{d}u = \tfrac{1}{4} e^{-\frac{|s|}{2}}$$

which is the double exponential distribution with 2 as scale parameter.

$$\begin{aligned} q = 4 \Rightarrow f_4(s) &= \tfrac{1}{16} e^{s/2} \int_{0 \vee s}^{\infty} (u^2 - us) e^{-u} \mathrm{d}u \\ &= \tfrac{1}{16} e^{s/2} \left(\int_{0 \vee s}^{\infty} u^2 e^{-u} \mathrm{d}u - s \int_{0 \vee s}^{\infty} u e^{-u} \mathrm{d}u \right) \\ &= \tfrac{1}{16} e^{s/2} \left(\Gamma(3, 0 \vee s) - s\Gamma(2, 0 \vee s) \right), \end{aligned}$$

where $\Gamma(a, b)$ is the incomplete gamma function. By the symmetry of the distribution we can just derive the distribution for negative values of s. For $s < 0$ we have $f_4(s) = \tfrac{1}{16} e^{s/2} \left(\Gamma(3) - s\Gamma(2) \right) = \tfrac{1}{16} e^{s/2} (2 - s)$, so that

$$f_4(s) = 2^{-4} (2 + |s|) e^{-\frac{|s|}{2}}.$$

Proof of Simplification

Exploiting the symmetry in general leads to:

$$f_q(s) = \frac{1}{2^q \Gamma(\frac{q}{2})\Gamma(\frac{q}{2})} e^{\frac{-|s|}{2}} H(s),$$

where

$$H(s) = \int_0^\infty (u(u+|s|))^{\frac{q-2}{2}} e^{-u} \mathrm{d}u.$$

We also have a general expression for the mode of the distribution... because

$$H(0) = \int_0^\infty u^{q-2} e^{-u} \mathrm{d}u = \Gamma(q+1),$$

so that

$$f_q(0) = \frac{\Gamma(q+1)}{2^q \Gamma(\frac{q}{2})\Gamma(\frac{q}{2})} = \frac{q}{2^q B(\frac{q}{2}, \frac{q}{2})}.$$