Bootstrap Joint Prediction Regions

Michael Wolf Dan Wunderli

Department of Economics
University of Zurich

Motivational Quote

. . . a central bank seeking to maximize its probability of achieving its goals is driven, I believe, to a risk-management approach to policy. By this I mean that policymakers need to consider not only the most likely future path for the economy but also the distribution of possible outcomes about that path.

Alan Greenspan (2003)

Outline

Outline

The Problem

Object of interest:

- Observed time series $\{y_1, \ldots, y_T\}$
- Interested in the future path $Y_{T,H} \equiv (y_{T+1}, \ldots, y_{T+H})'$, where H is the maximum forecast horizon

The Problem

Object of interest:

- Observed time series $\{y_1, \ldots, y_T\}$
- Interested in the future path $Y_{T,H} \equiv (y_{T+1}, \ldots, y_{T+H})'$, where H is the maximum forecast horizon

For starters:

- Denote a forecast h periods ahead by $\hat{y}_T(h)$
- Want a path-forecast $\hat{Y}_T(H) \equiv (\hat{y}_T(1), \ldots, \hat{y}_T(H))'$

The Problem

Object of interest:

- Observed time series $\{y_1, \ldots, y_T\}$
- Interested in the future path $Y_{T,H} \equiv (y_{T+1}, \ldots, y_{T+H})'$, where H is the maximum forecast horizon

For starters:

- Denote a forecast h periods ahead by $\hat{y}_T(h)$
- Want a path-forecast $\hat{Y}_T(H) \equiv (\hat{y}_T(1), \ldots, \hat{y}_T(H))'$

In the end:

- Also want a joint prediction region (JPR) that contains the entire future path $Y_{T,H}$ with prespecified probability $1 - \alpha$
- For purposes of interpretation, such a JPR should be of the form of simultaneous prediction intervals for y_{T+h}, for $h = 1, \ldots, H$

Restriction To Rectangular JPRs

In general:

- $Y_{T,H}$ is a H-dimensional vector
- In principle, a JPR can be any region in $\mathbb{R}^H$ that contains the vector $Y_{T,H}$ with probability $1 - \alpha$
- For example, an elliptical JPR based on the classical Scheffé method (details later)

Restriction To Rectangular JPRs

In general:

- $Y_{T,H}$ is a H-dimensional vector
- In principle, a JPR can be any region in $\mathbb{R}^H$ that contains the vector $Y_{T,H}$ with probability $1 - \alpha$
- For example, an elliptical JPR based on the classical Scheffé method (details later)

In practice:

- Want an implied 'prediction interval' for y_{T+h} at each horizon h
- So the JPR should represent simultaneous prediction intervals: in other words, one wants a rectangular JPR

Restriction To Rectangular JPRs

In general:

- $Y_{T,H}$ is a H-dimensional vector
- In principle, a JPR can be any region in $\mathbb{R}^H$ that contains the vector $Y_{T,H}$ with probability $1 - \alpha$
- For example, an elliptical JPR based on the classical Scheffé method (details later)

In practice:

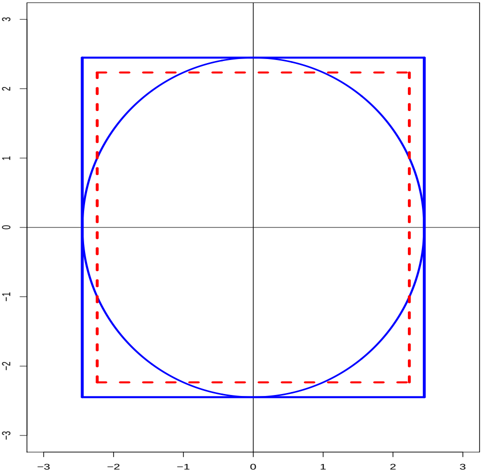
- Want an implied 'prediction interval' for y_{T+h} at each horizon h
- So the JPR should represent simultaneous prediction intervals: in other words, one wants a rectangular JPR

Note:

- One can always start with a JPR of arbitrary shape and then 'project' it onto the axes of $\mathbb{R}^H$ to obtain a rectangular JPR
- But, clearly, such a procedure is sub-optimal
- Instead, one should construct a 'direct' rectangular JPR

Restriction To Rectangular JPRs

An illustration of elliptical (and projected) JPR versus rectangular JPR:



The Non-Solution

How not to do it:

- Compute a marginal prediction interval for y_{T+h} at level $1 - \alpha$ for each $h = 1, \ldots, H$
- Then 'string together' these H intervals

The Non-Solution

How not to do it:

- Compute a marginal prediction interval for y_{T+h} at level $1 - \alpha$ for each $h = 1, \ldots, H$
- Then 'string together' these H intervals

Advantage:

- (Relatively) easy to do:
- How to compute reliable marginal prediction intervals has been worked out finally

Disadvantage:

- The joint coverage probability for the path $Y_{T,H}$ is less than $1 - \alpha$
- Furthermore, *ceteris paribus* this probability decreases in H

The Non-Solution

How not to do it:

- Compute a marginal prediction interval for y_{T+h} at level $1-\alpha$ for each $h = 1, \ldots, H$
- Then 'string together' these H intervals

Advantage:

- (Relatively) easy to do:
- How to compute reliable marginal prediction intervals has been worked out finally

Disadvantage:

- The joint coverage probability for the path $Y_{T,H}$ is less than $1-\alpha$
- Furthermore, *ceteris paribus* this probability decreases in H

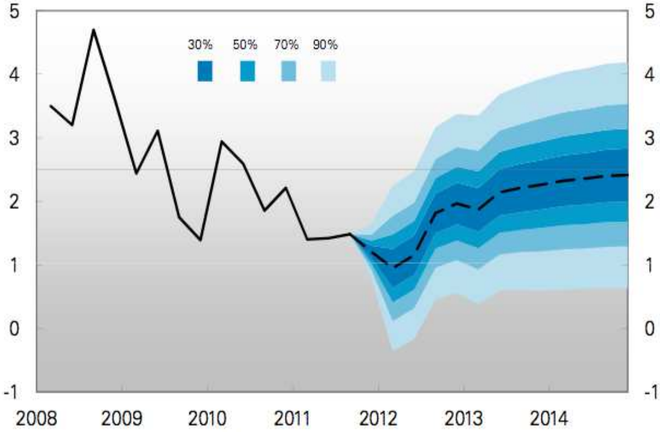
Amazingly:

- This method is still widely used
- For example, in fan charts published by the Bank of England and the Central Bank of Norway

The Non-Solution

An (unfortunate) example:

Chart 1.14c Projected CPI in the baseline scenario with fan chart.
Four-quarter change. Per cent. 2008 Q1 – 2014 Q4



Sources: Statistics Norway and Norges Bank

Outline

Some Notation

In the real world:

- Data $\{y_1, \ldots, y_T, y_{T+1}, \ldots y_{T+H}\}$ generated by mechanism $\mathbb{P}$
- Vector of prediction errors:
  $\hat{U}_T(H) \equiv (\hat{u}_T(1), \ldots, \hat{u}_T(H))' \equiv \hat{Y}_T(H) - Y_{T,H}$
- Prediction standard error for $\hat{u}_T(h)$ denoted by $\hat{\sigma}_T(h)$
- Vector of standardized prediction errors:
  $\hat{S}_T(H) \equiv (\hat{u}_T(1)/\hat{\sigma}_T(1), \ldots, \hat{u}_T(H)/\hat{\sigma}_T(H))'$

Some Notation

In the real world:

- Data $\{y_1, \ldots, y_T, y_{T+1}, \ldots y_{T+H}\}$ generated by mechanism $\mathbb{P}$
- Vector of prediction errors:
  $\hat{U}_T(H) \equiv (\hat{u}_T(1), \ldots, \hat{u}_T(H))' \equiv \hat{Y}_T(H) - Y_{T,H}$
- Prediction standard error for $\hat{u}_T(h)$ denoted by $\hat{\sigma}_T(h)$
- Vector of standardized prediction errors:
  $\hat{S}_T(H) \equiv (\hat{u}_T(1)/\hat{\sigma}_T(1), \ldots, \hat{u}_T(H)/\hat{\sigma}_T(H))'$

In the bootstrap world:

- Data $\{y^*_1, \ldots, y^*_T, y^*_{T+1}, \ldots y^*_{T+H}\}$ generated by mechanism $\hat{\mathbb{P}}_T$
- Vector of bootstrap prediction errors:
  $\hat{U}^*_T(H) \equiv (\hat{u}^*_T(1), \ldots, \hat{u}^*_T(H))' \equiv \hat{Y}^*_T(H) - Y^*_{T,H}$
- Prediction standard error for $\hat{u}^*_T(h)$ denoted by $\hat{\sigma}^*_T(h)$
- Vector of bootstrap standardized prediction errors:
  $\hat{S}^*_T(H) \equiv (\hat{u}^*_T(1)/\hat{\sigma}^*_T(1), \ldots, \hat{u}^*_T(H)/\hat{\sigma}^*_T(H))'$

Some Notation

In the real world:

- Data $\{y_1, \ldots, y_T, y_{T+1}, \ldots y_{T+H}\}$ generated by mechanism $\mathbb{P}$
- Vector of prediction errors:
  $\hat{U}_T(H) \equiv (\hat{u}_T(1), \ldots, \hat{u}_T(H))' \equiv \hat{Y}_T(H) - Y_{T,H}$
- Prediction standard error for $\hat{u}_T(h)$ denoted by $\hat{\sigma}_T(h)$
- Vector of standardized prediction errors:
  $\hat{S}_T(H) \equiv (\hat{u}_T(1)/\hat{\sigma}_T(1), \ldots, \hat{u}_T(H)/\hat{\sigma}_T(H))'$

In the bootstrap world:

- Data $\{y_1^*, \ldots, y_T^*, y_{T+1}^*, \ldots y_{T+H}^*\}$ generated by mechanism $\hat{\mathbb{P}}_T$
- Vector of bootstrap prediction errors:
  $\hat{U}_T^*(H) \equiv (\hat{u}_T^*(1), \ldots, \hat{u}_T^*(H))' \equiv \hat{Y}_T^*(H) - Y_{T,H}^*$
- Prediction standard error for $\hat{u}_T^*(h)$ denoted by $\hat{\sigma}_T^*(h)$
- Vector of bootstrap standardized prediction errors:
  $\hat{S}_T^*(H) \equiv (\hat{u}_T^*(1)/\hat{\sigma}_T^*(1), \ldots, \hat{u}_T^*(H)/\hat{\sigma}_T^*(H))'$

Note:

- The methodology is completely generic
- All implementation details are up to the applied researcher

High-Level Assumption

Relevant quantities:

- $\hat{J}_T$ denotes the probability law under $\mathbb{P}$ of $\hat{S}_T(H)|y_T, y_{T-1}, \ldots$
- $\hat{J}_T^*$ denotes the probability law under $\hat{\mathbb{P}}_T$ of $\hat{S}_T^*(H)|y_T^*, y_{T-1}^*, \ldots$

In the asymptotic framework, T tends to infinity and H remains fixed.

Assumption 2.1

- *$\hat{J}_T$ converges in distribution to a non-random continuous limit law $\hat{J}$.*
- *Furthermore, $\hat{J}_T^*$ consistently estimates this limit law: $\rho(\hat{J}_T, \hat{J}_T^*) \to 0$ in probability, for any metric ρ metrizing weak convergence.*

Flexible Criterion To Construct JPRs

Possible concern:

- When H is large, it may be deemed too strict that all elements of the future path must be contained in the JPR (with prob. $1 - \alpha$)

Flexible Criterion To Construct JPRs

Possible concern:

- When H is large, it may be deemed too strict that all elements of the future path must be contained in the JPR (with prob. $1 - \alpha$)

We thus adapt a concept from the multiple-testing literature to offer a flexible solution:

Generalized family-wise error rate (k-FWE)

- k-FWE $\equiv \mathbb{P}\{$At least k of the y_{T+h} not contained in the JPR$\}$

Flexible Criterion To Construct JPRs

Possible concern:

- When H is large, it may be deemed too strict that all elements of the future path must be contained in the JPR (with prob. $1 - \alpha$)

We thus adapt a concept from the multiple-testing literature to offer a flexible solution:

Generalized family-wise error rate (k-FWE)

- k-FWE $\equiv \mathbb{P}\{$At least k of the y_{T+h} not contained in the JPR$\}$

Implication:

- For $k = 1$, one wants to catch the entire future path in the JPR
- For $k > 1$, one is willing to miss up to $k - 1$ elements in the JPR, but is afforded a smaller region in return (see below)

Flexible Criterion To Construct JPRs

Possible concern:

- When H is large, it may be deemed too strict that all elements of the future path must be contained in the JPR (with prob. $1 - \alpha$)

We thus adapt a concept from the multiple-testing literature to offer a flexible solution:

Generalized family-wise error rate (k-FWE)

- k-FWE $\equiv \mathbb{P}\{$At least k of the y_{T+h} not contained in the JPR$\}$

Implication:

- For $k = 1$, one wants to catch the entire future path in the JPR
- For $k > 1$, one is willing to miss up to $k - 1$ elements in the JPR, but is afforded a smaller region in return (see below)

Goal:

- The applied researcher chooses the value of k, given his needs
- The JPR should then deliver k-FWE $\leq \alpha$, at least asymptotically

How To Make It Happen

Some further notation:

- Let $X \equiv (x_1, \ldots, x_H)'$ be a vector with H elements
- k-max(X) returns the k^{th}-largest value of X
- $|X|$ denotes the vector $(|x_1|, \ldots, |x_H|)'$

How To Make It Happen

Some further notation:

- Let $X \equiv (x_1, \ldots, x_H)'$ be a vector with H elements
- k-max(X) returns the k^{th}-largest value of X
- $|X|$ denotes the vector $(|x_1|, \ldots, |x_H|)'$

The ideal JPR, controlling the k-FWE in finite samples, is of the form:

$$\Big[.\Big] \times \ldots \times \Big[\hat{y}_T(h) \pm d^{max}_{|\cdot|,1-\alpha}(k) \cdot \hat{\sigma}_T(h)\Big] \times \ldots \times \Big[.\Big]$$

where $d^{max}_{|\cdot|,1-\alpha}(k)$ is the $1-\alpha$ quantile of random variable k-max$\Big(|\hat{S}_T(H)|\Big)$.

How To Make It Happen

Some further notation:

- Let $X \equiv (x_1, \ldots, x_H)'$ be a vector with H elements
- k-max(X) returns the k^{th}-largest value of X
- $|X|$ denotes the vector $(|x_1|, \ldots, |x_H|)'$

The ideal JPR, controlling the k-FWE in finite samples, is of the form:

$$\Big[.\Big] \times \ldots \times \Big[\hat{y}_T(h) \pm d^{max}_{|\cdot|,1-\alpha}(k) \cdot \hat{\sigma}_T(h)\Big] \times \ldots \times \Big[.\Big]$$

where $d^{max}_{|\cdot|,1-\alpha}(k)$ is the $1-\alpha$ quantile of random variable k-max$\Big(|\hat{S}_T(H)|\Big)$.

The feasible JPR, controlling the k-FWE asymptotically, is of the form:

$$\Big[.\Big] \times \ldots \times \Big[\hat{y}_T(h) \pm d^{max,*}_{|\cdot|,1-\alpha}(k) \cdot \hat{\sigma}_T(h)\Big] \times \ldots \times \Big[.\Big] \tag{1}$$

where $d^{max,*}_{|\cdot|,1-\alpha}(k)$ is the $1-\alpha$ quantile of random variable k-max$\Big(|\hat{S}^*_T(H)|\Big)$

Formal Result

Proposition 2.1

Under Assumption 2.1, the JPR (1) *for* $Y_{T,H}$ *satisfies*

$$\limsup_{T\to\infty}\ k\text{-}FWE \leq \alpha$$

where

$$k\text{-}FWE \equiv \mathbb{P}\{\textit{At least } k \textit{ of the } y_{T+h} \textit{ not contained in the JPR}\}\ .$$

Formal Result

Proposition 2.1

Under Assumption 2.1, the JPR (1) *for* $Y_{T,H}$ *satisfies*

$$\limsup_{T\to\infty} \ k\text{-}FWE \le \alpha$$

where

$$k\text{-}FWE \equiv \mathbb{P}\{\textit{At least } k \textit{ of the } y_{T+h} \textit{ not contained in the JPR}\} .$$

Alternative JPRs:

- The JPR (1) is two-sided
- Alternatively, lower and upper one-sided JPRs can be constructed in a similar fashion; see the paper for details

Bootstrap Details

Algorithm 2.1 (Computation of the JPR Multiplier)

1. Generate bootstrap data $\{y_1^*, \ldots, y_T^*, y_{T+1}^*, \ldots, y_{T+H}^*\}$ from $\hat{\mathbb{P}}_T$
2. Not making use of the stretch $\{y_{T+1}^*, \ldots, y_{T+H}^*\}$, compute forecasts $\hat{y}_T^*(h)$ and prediction standard errors $\hat{\sigma}_T^*(h)$
3. Compute bootstrap prediction errors $\hat{u}_T^*(h) \equiv \hat{y}_T^*(h) - y_{T+h}^*$
4. Compute standardized bootstrap prediction errors $\hat{s}_T^*(h) \equiv \hat{u}_T^*(h)/\hat{\sigma}_T^*(h)$ and let $\hat{S}_T^*(H) \equiv \big(\hat{s}_T^*(1), \ldots, \hat{s}_T^*(H)\big)'$
5. Compute $k\text{-max}_{|\cdot|}^* \equiv k\text{-max}\Big(\big|\hat{S}_T^*(H)\big|\Big)$
6. Repeat this process B times $\Longrightarrow \Big\{k\text{-max}_{|\cdot|,1}^*, \ldots, k\text{-max}_{|\cdot|,B}^*\Big\}$
7. $d_{|\cdot|,1-\alpha}^{max,*}(k)$ is the empirical $1-\alpha$ quantile of these B statistics

Multivariate Time Series

More general scenario:

- One observes a K-variate time series $\{Z_1, \ldots, Z_T\}$
- The goal is to predict the next stretch of H observations for a particular component of Z_t, say the first one w.l.o.g.
- Write $Z_t \equiv (y_t, z_{2,t}, \ldots, z_{K,t})'$
- The forecasts $\hat{y}_T(h)$ and the prediction standard errors $\hat{\sigma}_T(h)$ are computed from $\{Z_1, \ldots, Z_T\}$ rather than from $\{y_1, \ldots, y_T\}$ only
- Ditto in the bootstrap world

Multivariate Time Series

More general scenario:

- One observes a K-variate time series $\{Z_1, \ldots, Z_T\}$
- The goal is to predict the next stretch of H observations for a particular component of Z_t, say the first one w.l.o.g.
- Write $Z_t \equiv (y_t, z_{2,t}, \ldots, z_{K,t})'$
- The forecasts $\hat{y}_T(h)$ and the prediction standard errors $\hat{\sigma}_T(h)$ are computed from $\{Z_1, \ldots, Z_T\}$ rather than from $\{y_1, \ldots, y_T\}$ only
- Ditto in the bootstrap world

More general relevant quantities:

- $\hat{J}_T$ denotes the probability law under $\mathbb{P}$ of $\hat{S}_T(H)|Z_T, Z_{T-1}, \ldots$
- $\hat{J}^*_T$ denotes the probability law under $\hat{\mathbb{P}}_T$ of $\hat{S}^*_T(H)|Z^*_T, Z^*_{T-1}, \ldots$

Multivariate Time Series

More general scenario:

- One observes a K-variate time series $\{Z_1, \ldots, Z_T\}$
- The goal is to predict the next stretch of H observations for a particular component of Z_t, say the first one w.l.o.g.
- Write $Z_t \equiv (y_t, z_{2,t}, \ldots, z_{K,t})'$
- The forecasts $\hat{y}_T(h)$ and the prediction standard errors $\hat{\sigma}_T(h)$ are computed from $\{Z_1, \ldots, Z_T\}$ rather than from $\{y_1, \ldots, y_T\}$ only
- Ditto in the bootstrap world

More general relevant quantities:

- $\hat{J}_T$ denotes the probability law under $\mathbb{P}$ of $\hat{S}_T(H)|Z_T, Z_{T-1}, \ldots$
- $\hat{J}_T^*$ denotes the probability law under $\hat{\mathbb{P}}_T$ of $\hat{S}_T^*(H)|Z_T^*, Z_{T-1}^*, \ldots$

Unchanged methodology:

- Given the modifications above, the bootstrap methodology to construct JPRs remains unchanged
- Proposition 2.1 continues to hold

Outline

(Modified) Scheffé JPR

Jordà and Marcellino (2010) propose an 'asymptotic' JPR based on

Assumption 3.1

$$\sqrt{T}\Big(\hat{Y}_T(H) - Y_{T,H}|Z_T, Z_{T-1}, \ldots\Big) \xrightarrow{d} N(0, \Xi_H) \quad \textit{and} \quad \hat{\Xi}_H \xrightarrow{\mathbb{P}} \Xi_H \,.$$

(Modified) Scheffé JPR

Jordà and Marcellino (2010) propose an 'asymptotic' JPR based on

Assumption 3.1

$$\sqrt{T}\Big(\hat{Y}_T(H) - Y_{T,H}|Z_T, Z_{T-1}, \ldots\Big) \xrightarrow{d} N(0, \Xi_H) \quad \textit{and} \quad \hat{\Xi}_H \xrightarrow{\mathbb{P}} \Xi_H \, .$$

Furthermore, let P be the lower-triangular Cholesky decomposition of $\hat{\Xi}_H/T$, satisfying $PP' = \hat{\Xi}_H/T$.

(Modified) Scheffé JPR

Jordà and Marcellino (2010) propose an 'asymptotic' JPR based on

Assumption 3.1

$$\sqrt{T}\Big(\hat{Y}_T(H) - Y_{T,H}|Z_T, Z_{T-1}, \ldots\Big) \xrightarrow{d} N(0, \Xi_H) \quad and \quad \hat{\Xi}_H \xrightarrow{\mathbb{P}} \Xi_H \,.$$

Furthermore, let P be the lower-triangular Cholesky decomposition of $\hat{\Xi}_H/T$, satisfying $PP' = \hat{\Xi}_H/T$.

The proposed Scheffé JPR is obtained in three steps:

(S1) $\Big\{\widetilde{Y} : T(\hat{Y}_T(H) - \widetilde{Y})'\hat{\Xi}_H^{-1}(\hat{Y}_T(H) - \widetilde{Y}) \leq \chi^2_{H,1-\alpha}\Big\}$ (classical Scheffé JPR)

(Modified) Scheffé JPR

Jordà and Marcellino (2010) propose an 'asymptotic' JPR based on

Assumption 3.1

$$\sqrt{T}\Big(\hat{Y}_T(H) - Y_{T,H}|Z_T, Z_{T-1}, \ldots\Big) \xrightarrow{d} N(0, \Xi_H) \quad and \quad \hat{\Xi}_H \xrightarrow{\mathbb{P}} \Xi_H \ .$$

Furthermore, let P be the lower-triangular Cholesky decomposition of $\hat{\Xi}_H/T$, satisfying $PP' = \hat{\Xi}_H/T$.

The proposed Scheffé JPR is obtained in three steps:

(S1) $\left\{\widetilde{Y} : T(\hat{Y}_T(H) - \widetilde{Y})'\hat{\Xi}_H^{-1}(\hat{Y}_T(H) - \widetilde{Y}) \leq \chi^2_{H,1-\alpha}\right\}$ (classical Scheffé JPR)

(S2) $\hat{Y}_T(H) \pm P\left[\sqrt{\frac{\chi^2_{H,1-\alpha}}{H}}\,\mathbf{1}_H\right]$ (by Bowden's (1970) Lemma . . .)

(Modified) Scheffé JPR

Jordà and Marcellino (2010) propose an 'asymptotic' JPR based on

Assumption 3.1

$$\sqrt{T}\Big(\hat{Y}_T(H) - Y_{T,H}|Z_T, Z_{T-1}, \ldots\Big) \xrightarrow{d} N(0, \Xi_H) \quad \textit{and} \quad \hat{\Xi}_H \xrightarrow{\mathbb{P}} \Xi_H \, .$$

Furthermore, let P be the lower-triangular Cholesky decomposition of $\hat{\Xi}_H/T$, satisfying $PP' = \hat{\Xi}_H/T$.

The proposed Scheffé JPR is obtained in three steps:

(S1) $\left\{\widetilde{Y} : T(\hat{Y}_T(H) - \widetilde{Y})'\hat{\Xi}_H^{-1}(\hat{Y}_T(H) - \widetilde{Y}) \le \chi^2_{H,1-\alpha}\right\}$ (classical Scheffé JPR)

(S2) $\hat{Y}_T(H) \pm P\left[\sqrt{\frac{\chi^2_{H,1-\alpha}}{H}}\,\mathbf{1}_H\right]$ (by Bowden's (1970) Lemma . . .)

(S3) $\hat{Y}_T(H) \pm P\left[\sqrt{\frac{\chi^2_{h,1-\alpha}}{h}}\right]_{h=1}^{H}$ (by some 'stepwise' method)

(Modified) Scheffé JPR

Criticisms:

- Assumption 3.1 is reasonable in the context of estimation but not in the context of prediction

(Modified) Scheffé JPR

Criticisms:

- Assumption 3.1 is reasonable in the context of estimation but not in the context of prediction
- The way from (S1) to (S3) is not exactly paved with theoretical justification

(Modified) Scheffé JPR

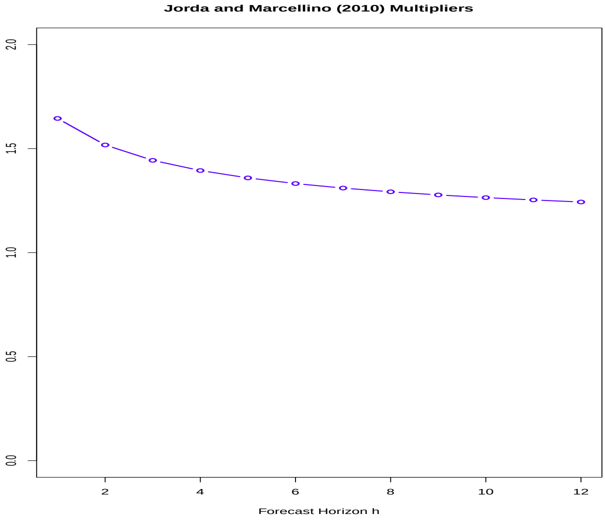
Criticisms:

- Assumption 3.1 is reasonable in the context of estimation but not in the context of prediction
- The way from (S1) to (S3) is not exactly paved with theoretical justification
- The width of the proposed JPR (S3) at forecast horizon h may not be (weakly) monotonically increasing in h:

  this can happen, since the multipliers $\sqrt{\chi^2_{h,1-\alpha}/h}$ are strictly decreasing in h (for commonly used values of α)

(Modified) Scheffé JPR

Multipliers of the (modified) Scheffé JPR for $H = 12$ and $\alpha = 0.1$:



NP Heuristic JPR

Staszewska-Bystrova (2010) proposes the following alternative bootstrap JPR:

NP Heuristic JPR

Staszewska-Bystrova (2010) proposes the following alternative bootstrap JPR:

- Generate B bootstrap path-forecasts $\hat{Y}_T^{*,b}(H)$, for $b = 1, \ldots, B$
- Discard αB of these bootstrap path-forecasts: those $\hat{Y}_T^{*,b}(H)$ that are 'furthest' away from the original path-forecast $\hat{Y}_T(H)$ (where distance is measured by the Euclidian distance, say)
- The neighboring-paths (NP) JPR is defined as the envelope of the remaining $(1-\alpha)B$ bootstrap path-forecasts $\hat{Y}_T^{*,b}(H)$

NP Heuristic JPR

Criticisms:

- The method is purely heuristic: no proof of asymptotic validity, under some suitable high-level assumption, is given

NP Heuristic JPR

Criticisms:

- The method is purely heuristic: no proof of asymptotic validity, under some suitable high-level assumption, is given
- The method seems to restricted to (V)AR models, since it uses the backward representation of a (V)AR model to generate the bootstrap path-forecasts $\hat{Y}_T^*(H)$

NP Heuristic JPR

Criticisms:

- The method is purely heuristic: no proof of asymptotic validity, under some suitable high-level assumption, is given
- The method seems to restricted to (V)AR models, since it uses the backward representation of a (V)AR model to generate the bootstrap path-forecasts $\hat{Y}_T^*(H)$
- The shape of the JPR can be jagged, which is unattractive

Property of Balance

Under the additional assumption that

$$\text{the marginal distribution of } \frac{\hat{y}_T(h) - y_{T+h}}{\hat{\sigma}_T(h)} \text{ is independent of } h$$

asymptotically, it is easily seen that our bootstrap JPR (1) has the property of being balanced, asymptotically:

$$\mathbb{P}\Big\{y_{T+h} \in \Big[\hat{y}_T(h) \pm d^{max,*}_{|\cdot|,1-\alpha}(k) \cdot \hat{\sigma}_T(h)\Big]\Big\} \quad \text{is independent of } h$$

Property of Balance

Under the additional assumption that

$$\text{the marginal distribution of } \frac{\hat{y}_T(h) - y_{T+h}}{\hat{\sigma}_T(h)} \text{ is independent of } h$$

asymptotically, it is easily seen that our bootstrap JPR (1) has the property of being balanced, asymptotically:

$$\mathbb{P}\Big\{y_{T+h} \in \Big[\hat{y}_T(h) \pm d^{max,*}_{|\cdot|,1-\alpha}(k) \cdot \hat{\sigma}_T(h)\Big]\Big\} \quad \text{is independent of } h$$

All forecasts $\hat{y}_T(h)$ are treated as equally important: the probability of violating the k-FWE criterion is spread out evenly over all horizons h.

Property of Balance

Under the additional assumption that

$$\text{the marginal distribution of } \frac{\hat{y}_T(h) - y_{T+h}}{\hat{\sigma}_T(h)} \text{ is independent of } h$$

asymptotically, it is easily seen that our bootstrap JPR (1) has the property of being balanced, asymptotically:

$$\mathbb{P}\Big\{y_{T+h} \in \Big[\hat{y}_T(h) \pm d^{max,*}_{|\cdot|,1-\alpha}(k) \cdot \hat{\sigma}_T(h)\Big]\Big\} \quad \text{is independent of } h$$

All forecasts $\hat{y}_T(h)$ are treated as equally important: the probability of violating the k-FWE criterion is spread out evenly over all horizons h.

Another way to argue that balance is a desirable property is by considering the following (extremely) unbalanced JPR:

$$\text{PI}_T(1) \times (-\infty, \infty) \times \ldots \times (-\infty, \infty)$$

where $\text{PI}_T(1)$ is a marginal prediction interval for y_{T+1}.

Outline

Preliminaries

We consider the general AR(p) model

$$y_t = \nu + \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + \epsilon_t \tag{2}$$

which can be alternatively expressed as

$$y_t = \nu + \rho y_{t-1} + \psi_1 \Delta y_{t-1} + \ldots + \psi_{p-1} \Delta y_{t-p+1} + \epsilon_t \tag{3}$$

to bring out the role of the largest autoregressive root $\rho \equiv \rho_1 + \ldots + \rho_p$.

Preliminaries

We consider the general AR(p) model

$$y_t = \nu + \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + \epsilon_t \tag{2}$$

which can be alternatively expressed as

$$y_t = \nu + \rho y_{t-1} + \psi_1 \Delta y_{t-1} + \ldots + \psi_{p-1} \Delta y_{t-p+1} + \epsilon_t \tag{3}$$

to bring out the role of the largest autoregressive root $\rho \equiv \rho_1 + \ldots + \rho_p$.

Estimation strategy:

- Estimate formulation (3) by OLS, yielding $\hat{\rho}_{OLS}$
- Transform to the bias-corrected estimator (e.g., see White, 1961)

$$\hat{\rho}_{BC} \equiv \hat{\rho}_{OLS} + \frac{1 + 3\,\hat{\rho}_{OLS}}{T} \tag{4}$$

Preliminaries

We consider the general AR(p) model

$$y_t = \nu + \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + \epsilon_t \quad (2)$$

which can be alternatively expressed as

$$y_t = \nu + \rho y_{t-1} + \psi_1 \Delta y_{t-1} + \ldots + \psi_{p-1} \Delta y_{t-p+1} + \epsilon_t \quad (3)$$

to bring out the role of the largest autoregressive root $\rho \equiv \rho_1 + \ldots + \rho_p$.

Estimation strategy:

- Estimate formulation (3) by OLS, yielding $\hat{\rho}_{OLS}$
- Transform to the bias-corrected estimator (e.g., see White, 1961)

$$\hat{\rho}_{BC} \equiv \hat{\rho}_{OLS} + \frac{1 + 3\,\hat{\rho}_{OLS}}{T} \quad (4)$$

- Regress $y_t - \hat{\rho}_{BC} y_{t-1}$ on $(1, \Delta y_{t-1}, \ldots, \Delta y_{t-p-1})$ by OLS to get corresponding estimators of $(\nu, \psi_1, \ldots, \psi_{p-1})$
- Use the one-to-one relations between the formulations (2)–(3) to get set of estimators $(\hat{\nu}, \hat{\rho}_1, \ldots, \hat{\rho}_p)$ and (centered) residuals $\{\hat{\epsilon}_t\}$

Preliminaries

Computation of the forecasts $\hat{y}_T(h)$: in the usual way.

Preliminaries

Computation of the forecasts $\hat{y}_T(h)$: in the usual way.

Computation of the prediction standard errors:

- Convert the $\widehat{\text{AR}}(p)$ model $(\hat{\nu}, \hat{\rho}_1, \ldots, \hat{\rho}_p)$ to an $\widehat{\text{MA}}(\infty)$ model $\{\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots\}$, with $\hat{\theta}_0 = 1$
- Then $\hat{\sigma}_T(h) \equiv \hat{\sigma}_\epsilon \sqrt{\hat{\theta}_0^2 + \ldots + \hat{\theta}_{h-1}^2}$

Preliminaries

Computation of the forecasts $\hat{y}_T(h)$: in the usual way.

Computation of the prediction standard errors:

- Convert the $\widehat{\text{AR}}(p)$ model $(\hat{\nu}, \hat{\rho}_1, \ldots, \hat{\rho}_p)$ to an $\widehat{\text{MA}}(\infty)$ model $\{\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots\}$, with $\hat{\theta}_0 = 1$
- Then $\hat{\sigma}_T(h) \equiv \hat{\sigma}_\epsilon \sqrt{\hat{\theta}_0^2 + \ldots + \hat{\theta}_{h-1}^2}$

Generation of the bootstrap data $\{y_1^*, \ldots, y_T^*\}$:

- Conditional on $\{y_1, \ldots, y_p\}$, using the $\widehat{\text{AR}}(p)$ model
- The $\widehat{\text{AR}}^*(p)$ model and the $\hat{\sigma}_T^*(h)$ are obtained as in the real world

Preliminaries

Computation of the forecasts $\hat{y}_T(h)$: in the usual way.

Computation of the prediction standard errors:

- Convert the $\widehat{\text{AR}}(p)$ model $(\hat{\nu}, \hat{\rho}_1, \ldots, \hat{\rho}_p)$ to an $\widehat{\text{MA}}(\infty)$ model $\{\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots\}$, with $\hat{\theta}_0 = 1$
- Then $\hat{\sigma}_T(h) \equiv \hat{\sigma}_\epsilon \sqrt{\hat{\theta}_0^2 + \ldots + \hat{\theta}_{h-1}^2}$

Generation of the bootstrap data $\{y_1^*, \ldots, y_T^*\}$:

- Conditional on $\{y_1, \ldots, y_p\}$, using the $\widehat{\text{AR}}(p)$ model
- The $\widehat{\text{AR}}^*(p)$ model and the $\hat{\sigma}_T^*(h)$ are obtained as in the real world

Generation of the bootstrap data $\{y_{T+1}^*, \ldots, y_{T+H}^*\}$:

- Conditional on $\{y_{T-H+1}, \ldots, y_T\}$, using the $\widehat{\text{AR}}(p)$ model

Preliminaries

Computation of the forecasts $\hat{y}_T(h)$: in the usual way.

Computation of the prediction standard errors:

- Convert the $\widehat{\text{AR}}(p)$ model $(\hat{\nu}, \hat{\rho}_1, \ldots, \hat{\rho}_p)$ to an $\widehat{\text{MA}}(\infty)$ model $\{\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots\}$, with $\hat{\theta}_0 = 1$
- Then $\hat{\sigma}_T(h) \equiv \hat{\sigma}_\epsilon \sqrt{\hat{\theta}_0^2 + \ldots + \hat{\theta}_{h-1}^2}$

Generation of the bootstrap data $\{y_1^*, \ldots, y_T^*\}$:

- Conditional on $\{y_1, \ldots, y_p\}$, using the $\widehat{\text{AR}}(p)$ model
- The $\widehat{\text{AR}}^*(p)$ model and the $\hat{\sigma}_T^*(h)$ are obtained as in the real world

Generation of the bootstrap data $\{y_{T+1}^*, \ldots, y_{T+H}^*\}$:

- Conditional on $\{y_{T-H+1}, \ldots, y_T\}$, using the $\widehat{\text{AR}}(p)$ model

Generation of the bootstrap path-forecast $\hat{Y}_T^*(H)$:

- Conditional on $\{y_{T-H+1}, \ldots, y_T\}$, using the $\widehat{\text{AR}}^*(p)$ model

$\Longrightarrow$ Employ the bootstrap approach of Pascual et al. (2001).

Monte Carlo Details

The model:

- Use AR(2) model with various parameters and normal errors
- The sample size is $T \in \{100, 400\}$
- Estimate the lag order from the (bootstrap) data by the BIC

Competing methods:

- Joint Marginals
- Scheffé (S3)
- NP Heuristic
- k-FWE JPR (1) with $k \in \{1, 2, 3\}$

Nominal coverage level:

- $1 - \alpha = 90\%$

Monte Carlo Details

The model:

- Use AR(2) model with various parameters and normal errors
- The sample size is $T \in \{100, 400\}$
- Estimate the lag order from the (bootstrap) data by the BIC

Competing methods:

- Joint Marginals
- Scheffé (S3)
- NP Heuristic
- k-FWE JPR (1) with $k \in \{1, 2, 3\}$

Nominal coverage level:

- $1 - \alpha = 90\%$

Note:

- A much wider set of simulation results, including non-normal errors, are reported in the paper

Monte Carlo Results I

	$T = 100$			$T = 400$		
$(\rho_1, \rho_2) = (1.75, -0.85)$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Joint Marginals	72.1	61.8	49.2	76.2	64.5	48.0
Scheffé	87.9	86.0	64.4	89.2	88.8	66.1
NP Heuristic	89.2	91.5	93.1	89.8	90.7	90.5
1-FWE JPR	90.4	90.5	89.6	89.8	89.7	89.7
2-FWE JPR	90.4	89.8	89.7	89.9	89.8	89.7
3-FWE JPR	90.0	90.3	89.0	90.0	89.7	89.6
$(\rho_1, \rho_2) = (1.25, -0.75)$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Joint Marginals	63.6	46.1	27.0	65.3	47.1	25.5
Scheffé	63.7	23.2	07.5	66.5	21.6	04.2
NP Heuristic	87.9	86.7	85.8	88.8	87.8	86.0
1-FWE JPR	90.0	89.4	89.3	89.9	89.8	89.9
2-FWE JPR	90.2	89.5	89.5	89.9	89.9	89.8
3-FWE JPR	89.8	89.5	89.3	89.9	89.8	89.7

Monte Carlo Results II

	$T = 100$			$T = 400$		
$(\rho_1, \rho_2) = (-0.65, 0.15)$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Joint Marginals	65.1	48.9	30.4	64.5	47.2	26.2
Scheffé	02.6	00.2	00.0	02.9	00.1	00.0
NP Heuristic	88.8	87.9	86.8	89.1	88.0	86.1
1-FWE JPR	90.4	90.1	89.7	90.0	90.0	89.7
2-FWE JPR	90.5	89.9	89.8	90.1	90.0	90.0
3-FWE JPR ($k=3$)	89.7	89.7	89.6	90.0	89.8	89.8
$(\rho_1, \rho_2) = (-0.7, -0.2)$	$H=6$	$H=12$	$H=24$	$H=6$	$H=12$	$H=24$
Joint Marginals	59.9	39.5	18.2	59.6	37.3	14.9
Scheffé	03.0	00.1	00.0	01.9	00.1	00.0
NP Heuristic	87.8	86.9	85.3	88.7	87.7	85.5
1-FWE JPR	89.4	89.3	88.7	89.9	89.8	89.8
2-FWE JPR	89.2	89.4	89.8	90.0	90.0	90.0
3-FWE JPR	89.4	89.7	89.8	90.0	90.1	89.9

Monte Carlo Results: Summary

Joint Marginals:

- As expected, the performance decreases in H and is poor
- Stringing together marginal prediction intervals does not yield a proper JPR

Monte Carlo Results: Summary

Joint Marginals:

- As expected, the performance decreases in H and is poor
- Stringing together marginal prediction intervals does not yield a proper JPR

Scheffé:

- The performance ranges from acceptable to horrible
- It decreases strongly in $\rho \equiv \rho_1 + \rho_2$ and in H

Monte Carlo Results: Summary

Joint Marginals:

- As expected, the performance decreases in H and is poor
- Stringing together marginal prediction intervals does not yield a proper JPR

Scheffé:

- The performance ranges from acceptable to horrible
- It decreases strongly in $\rho \equiv \rho_1 + \rho_2$ and in H

NP Heuristic:

- The performance ranges from good to acceptable
- It decreases slightly in H

Monte Carlo Results: Summary

Joint Marginals:

- As expected, the performance decreases in H and is poor
- Stringing together marginal prediction intervals does not yield a proper JPR

Scheffé:

- The performance ranges from acceptable to horrible
- It decreases strongly in $\rho \equiv \rho_1 + \rho_2$ and in H

NP Heuristic:

- The performance ranges from good to acceptable
- It decreases slightly in H

k-FWE JPR:

- The performance ranges from very good to good
- It is remarkably stable over both H and the value of k

Outline

Data Set & Methodology

Data set:

- Quarterly data on US real GDP from Q1/1947 until Q3/2011
- The data are seasonally adjusted and expressed in billions of chained 2005 dollars
- We focus on the first differences of the log-series (in percent), which correspond to log quarter-to-quarter growth
- There are a total of 258 observations
- We choose $H = 12$, which corresponds to a period of three years
- The nominal coverage is $1 - \alpha = 90\%$

Data Set & Methodology

Data set:

- Quarterly data on US real GDP from Q1/1947 until Q3/2011
- The data are seasonally adjusted and expressed in billions of chained 2005 dollars
- We focus on the first differences of the log-series (in percent), which correspond to log quarter-to-quarter growth
- There are a total of 258 observations
- We choose $H = 12$, which corresponds to a period of three years
- The nominal coverage is $1 - \alpha = 90\%$

Methodology:

- We use the same AR(p) methodology used in the Monte Carlo study (with the lag order p estimated by the BIC)
- More complex approaches could be used alternatively:
  - A nonlinear (SE)TAR model as in Potter (1995)
  - A VAR model, using extra variables, as in Stock and Watson (2001)
  - Others . . .
- However, our goal is to keep it (acceptably) simple and focus on the relative performances of the various JPRs

Data Set

Quarterly US real GDP: original series and ∇ log-series:

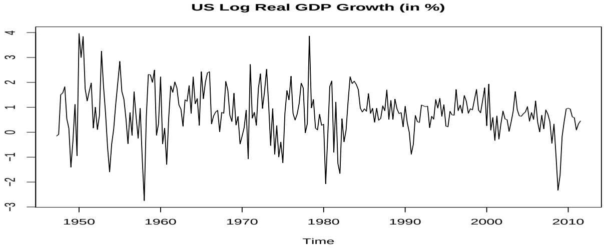


Illustration Exercise

To illustrate the salient features of the various JPRs:

- Use the last $T = 120$ observations to forecast the future path from Q4/2011 until Q3/2014
- Then compute corresponding JPRs

Illustration Exercise

To illustrate the salient features of the various JPRs:

- Use the last $T = 120$ observations to forecast the future path from Q4/2011 until Q3/2014
- Then compute corresponding JPRs

Fitting the model:

- The lag order chosen by the BIC is $\hat{p} = 1$
- The model fitted by OLS is

$$\hat{y}_{t+1} = 0.318 + 0.542 \cdot y_t$$

- The bias correction (4) yields the final fitted model

$$\hat{y}_{t+1} = 0.304 + 0.564 \cdot y_t$$

Illustration Exercise

First set of comparisons:

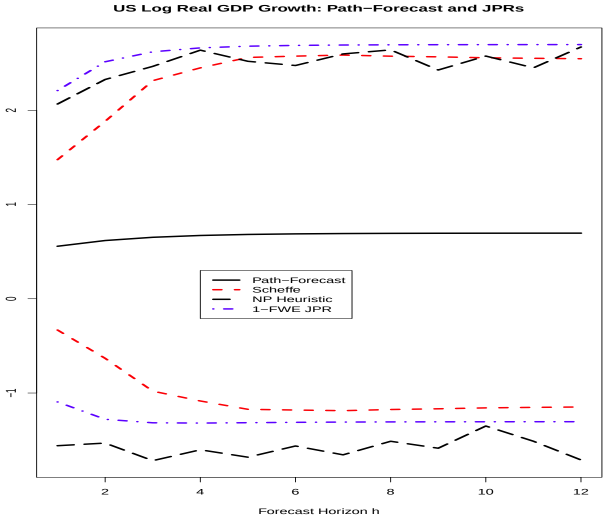


Illustration Exercise

Major findings:

- Scheffé has a smaller volume than the other two JPRs

Illustration Exercise

Major findings:

- Scheffé has a smaller volume than the other two JPRs
- The width of Scheffé at horizon h monotonically decreases from $h = 7$ to $h = 12$, if only slightly

Illustration Exercise

Major findings:

- Scheffé has a smaller volume than the other two JPRs
- The width of Scheffé at horizon h monotonically decreases from $h = 7$ to $h = 12$, if only slightly
- NP Heuristic and 1-FWE JPR have a comparable volume, but the shape of NP Heuristic is unattractively jagged (which cannot be blamed on a small number of bootstrap repetitions, since we used $B = 10,000$)

Illustration Exercise

Second set of comparisons:

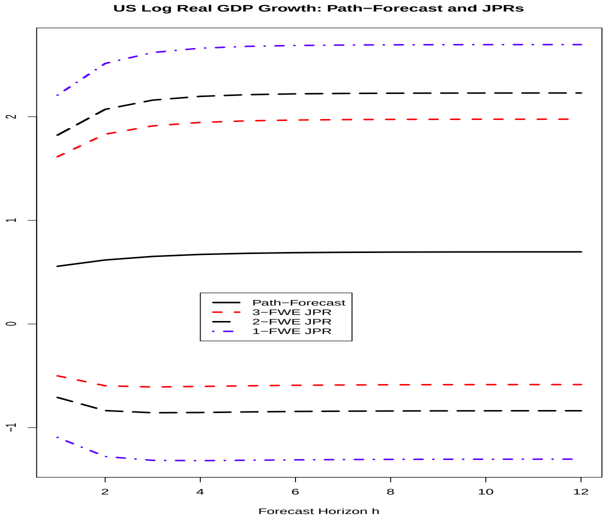


Illustration Exercise

Major finding:

- The volume of k-FWE JPR decreases in the value of k
- If the applied researcher is willing to miss up to one (or two) elements of the future path in the JPR (with probability 90%), he obtains a smaller and more informative region in return

Backtest Exercise

To get a feel for the out-of-sample performance of the various JPRs:

- Using the stretch $\{y_t, \ldots, y_{t+119}\}$ only, compute the JPR for the next $H = 12$ periods
- Compare the computed JPR against the path $(y_{t+120}, \ldots, y_{t+131})'$ to evaluate the 'success' in terms of the k-FWE criterion
- Do this for $t = 1, \ldots, 258 - 120 - 12 = 126$
- Then report the empirical coverage probability as the fraction of the 'successes' out of these 126 'trials'

Backtest Exercise

To get a feel for the out-of-sample performance of the various JPRs:

- Using the stretch $\{y_t, \ldots, y_{t+119}\}$ only, compute the JPR for the next $H = 12$ periods
- Compare the computed JPR against the path $(y_{t+120}, \ldots, y_{t+131})'$ to evaluate the 'success' in terms of the k-FWE criterion
- Do this for $t = 1, \ldots, 258 - 120 - 12 = 126$
- Then report the empirical coverage probability as the fraction of the 'successes' out of these 126 'trials'

Using this rolling-window approach, we get a fair, if not overly accurate, assessment of the out-of-sample performance.

Backtest Exercise

Empirical out-of-sample coverages for US log real GDP growth:

Method	Coverage
Joint Marginals	64.6
Scheffé	73.2
NP Heuristic	89.7
1-FWE JPR	89.9
2-FWE JPR	85.1
3-FWE JPR	87.3

Outline

Conclusions

Constructing joint prediction regions (JPRs) for a future path, along with a path-forecast, has not received the deserved attention so far.

Conclusions

Constructing joint prediction regions (JPRs) for a future path, along with a path-forecast, has not received the deserved attention so far.

We offer generic bootstrap JPRs that allow the applied researcher to determine the implementation details as he sees them most fit, given the application at hand.

Conclusions

Constructing joint prediction regions (JPRs) for a future path, along with a path-forecast, has not received the deserved attention so far.

We offer generic bootstrap JPRs that allow the applied researcher to determine the implementation details as he sees them most fit, given the application at hand.

Compared to two previous proposals, our bootstrap JPRs are shown to be asymptotically consistent, under a mild high-level assumption, and they also enjoy better finite-sample performance.

Conclusions

Constructing joint prediction regions (JPRs) for a future path, along with a path-forecast, has not received the deserved attention so far.

We offer generic bootstrap JPRs that allow the applied researcher to determine the implementation details as he sees them most fit, given the application at hand.

Compared to two previous proposals, our bootstrap JPRs are shown to be asymptotically consistent, under a mild high-level assumption, and they also enjoy better finite-sample performance.

In addition, we go beyond previous proposals by offering the more flexible k-FWE criterion: if the applied researcher is willing to miss a small number of elements of the future path, he is afforded a smaller, more informative region in return.

Bowden, D. C. (1970). Simultaneous confidence bands for linear regression models. *Journal of the American Statistical Association*, 65(329):413–421.

Greenspan, A. (2003). Remarks at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming on August 29, 2003. Available at `http://www.federalreserve.gov/boarddocs/speeches/2003/20030829/`.

Jordà, Ò. and Marcellino, M. G. (2010). Path-forecast evaluation. *Journal of Applied Econometrics*, 25:635–662.

Pascual, L., Romo, J., and Ruiz, E. (2001). Effects of parameter estimation on prediction densities: a bootstrap approach. *International Journal of Forecasting*, 17(1):83–103.

Potter, S. M. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, 2:109–125.

Staszewska-Bystrova, A. (2010). Bootstrap prediction bands for forecast paths from vector autoregressive models. *Journal of Forecasting*, Online Version, DOI:10.1002/for.1205.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

White, J. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika*, 48:85–95.