Bootstrapping joint prediction regions

Michael Wolf and Dan Wunderli

Discussion by Jörg Breitung

University of Bonn, Germany

Forecasting workshop, Eltville

June 1, 2012

General assessment

- fundamental problem in statistical inference
- here: joint prediction regions (JPR)
- Simple and robust bootstrap approach based on the maximum statistic
- Most reliable procedure so far
- What is the "best JPR" ?

$$Pr\{d_h^- \leq Y_{T+h} \leq d_h^+ \; \forall \; h = 1, \ldots, H\} = 1 - \alpha$$

- uniform (balanced) boundaries (Anaolyev/Kosenok 2011)

$$Pr\left(d_h^- \leq Y_{T+h} \leq d_h^+ \text{ for } h \in \{1, \cdots, H\}\right) \;=\; 1 - \alpha^*$$

  for $\alpha/H \leq \alpha^* \leq \alpha$ and $\alpha^* = 1 - (1-\alpha)^H$ for uncorrelated forecasts
- Remark 3.3 argues that the JPR is balanced

Simple example

Remark 2.1 considers the following example:

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

1. **Projecting Scheffe's region**

$$\begin{aligned} U_1^2 + U_2^2 &\sim \chi_2^2 \\ U_1^2 + U_2^2 &\leq 5.99 \end{aligned}$$

$\Rightarrow$ Prob($U_1^2 \leq 5.99$ and $U_2^2 \leq 5.99$) $= 0.9715$

2. **Bootstrap method of Staszewska-Bystrova** (2010):
   - Generate paths $(U_1^*, U_2^*)' \sim \mathcal{N}(0, I)$
   - Delete the 5% extreme paths measured by $\mathcal{D} = U_1^2 + U_2^2$ (here: $\mathcal{D} \leq 5.99$)
   - Compute the envelope of the paths

This method is equivalent to Scheffe's projections:
The envelope of U_1 is obtained by:

$$\max(U_1^*)^2 \leq 5.99 - (U_2^*)^2$$

Since U_2^* may be arbitrarily close to 0, it follows $\max(U_1^*)^2 \to 5.99$.

3. **Rectangular region based on the maximum** (Wolf/Wunderli 2012)
   - $\max(U_1^2, U_2^2) < d_{1-\alpha}$ with probability $1-\alpha$ implies that

     $$\{U_1^2 < d_{1-\alpha}\} \bigcup \{U_2^2 < d_{1-\alpha}\} \quad \text{with prob. } 1-\alpha$$

     $\Rightarrow$ 0.95-quantile computed by simulation ($= 2.24$)
   - Quantile can be determined by noting that

     $$P(U_1^2 < d^*) = \sqrt{0.95} = 0.9747 \;\Rightarrow\; d^* = 2.2368$$

- If the distribution of U_1, U_2 is skewed a symmetric interval (based on the max-statistic is not optimal
- Separate computation of lower and upper bounds:

$$Pr\Big(\max(U_1, U_2) < d^{\max}_{1-\alpha/2}\Big) = 1 - \alpha/2$$
$$Pr\Big(\min(U_1, U_2) > d^{\min}_{1-\alpha/2}\Big) = 1 - \alpha/2$$

Example of an alternative statistic:

- the mean statistic:

$$\begin{aligned} \frac{1}{2}\ (U_1 + U_2)^2 &\leq 3.84 \\ |U_1 + U_2| &\leq 2.77 \end{aligned}$$

- smaller region if both errors are positive
- larger region if errors have different sign
- Projection on axes yield very conservative JPR

4. **The approach for of Jorda/Marcellino** (2008)

Original limits: (note that $\sqrt{5.99} = 2.45$)

$$\begin{bmatrix} d_1^*(1-\alpha) \\ \\ d_2^*(1-\alpha) \end{bmatrix} = P \begin{bmatrix} \sqrt{\chi^2_{H,1-\alpha}/H} \\ \\ \sqrt{\chi^2_{H,1-\alpha}/H} \end{bmatrix} = \begin{bmatrix} 1.73 \\ \\ 1.73 \end{bmatrix}$$

Refined limits

$$\begin{bmatrix} \widetilde{d}_1^*(1-\alpha) \\ \\ \widetilde{d}_2^*(1-\alpha) \end{bmatrix} = P \begin{bmatrix} \sqrt{\chi^2_{1,1-\alpha}} \\ \\ \sqrt{\chi^2_{2,1-\alpha}/2} \end{bmatrix} = \begin{bmatrix} 1.96 \\ \\ 1.73 \end{bmatrix}$$

- simulation results for

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

- nominal confidence level: 0.95

Table: Actual coverage rates for the JM limits

ρ :	0.8	0.4	0	–0.4	–0.8
original	0.914	0.900	0.836	0.582	0.000
refined	0.948	0.937	0.873	0.558	0.000

$\Rightarrow$ similar findings in Wolf/Wunderli (2012)

Explanation of the results:

Limits for the orthogonalized random variables:

$$Z \;=\; P^{-1}U \;\equiv\; \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \le \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Multiplying with P yields

$$\begin{aligned} p_{11}(Z_1 - c_1) \;&\le\; 0 \\ {\color{red}p_{21}}\underbrace{(Z_1 - c_1)}_{\le 0} + p_{22}\underbrace{(Z_2 - c_2)}_{\le 0} \;&\le\; 0 \end{aligned}$$

Further issues

(i) Generalization of “familywise error rate”

$$\text{FWE} = Pr(\text{ At least one of the } y_h \text{ not contained in the JPR})$$

is generalized as

$$k\text{−FWE} = Pr(\text{ At least } k \text{ of the } y_h \text{ not contained in the JPR})$$

I do not think such a generalization is very attractive in empirical practice:

- Why should we ignore a (single) dramatic forecast failure?
- How to choose k?
- difficult to interpret

(ii) Is the bias correction really necessary‘?

- not often used it in practice
- bias is small $O(T^{-1})$ relative to the forecast error $O_p(1)$
- no SE and t-statistics for bias-corrected estimators
- White’s approximation is for AR(1)

(iii) How to generate bootstrap samples?

- As mentioned many possible ways to bootstrap (V)AR processes
- It would be helpful to study the relative performance of alternative approaches
- Since the estimation error is $O_p(T^{-1/2})$ and the forecast error $O_p(1)$ is may be sufficient to simplify the bootstrap procedure by just drawing from the forecast errors at least if T is large

Conclusion

- The paper is very well written and path breaking
- The (somewhat harsh) critique on the JM approach is sound and justified
- the suggested approach is a benchmark difficult to improve upon
- really great pleasure to read and comment the paper