Optimal Window Selection for Forecasting in The Presence of Recent Structural Breaks

Yongli Wang

University of Leicester

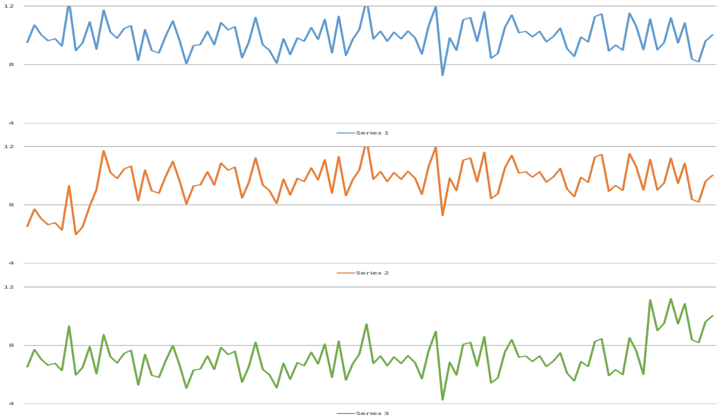
Econometric Research in Finance
Workshop on 15 September 2017
SGH Warsaw School of Economics

Motivation

- The presence of structure breaks is a crucial issue in forecasting
  - including pre-break data may lead to biased parameter estimates and biased forecasts
  - however reducing sample size increases the variance of the parameter estimates, which maps into the forecast errors
- Trade-off between the bias and variance $\rightleftharpoons$ Optimal window size (Pesaran and Timmermann, 2007, Journal of Econometrics)
- In other words, **how many observations should be used to estimate the parameter vector?**

Motivation

$$y_t = \alpha + \epsilon_t, \quad t = 1, 2, ..., 100$$



$$\hat{\mu}_1 = 9.90$$
$$\hat{\mu}_2 = 9.60, \qquad \tilde{\mu}_2 = 9.88$$
$$\hat{\mu}_3 = 7.20, \qquad \tilde{\mu}_3 = 9.40$$

Motivation

- Two most important papers on the optimal window selection
    - **Pesaran and Timmermann's (2007, Journal of Econometrics) cross-validation (PTCV) method**
        - selects the starting point of the window by partitioning data into two periods and comparing the recursive pseudo out-of-sample forecasts
        - requires strictly exogenous regressors and uncorrelated errors
        - suffers selection bias, when a break occurring shortly before the date of making forecasts distorts the ranking in the validation
    - **Inoue, Jin, and Rossi's (2017, Journal of Econometrics) algorithm (IJR)**
        - allows weak dependence and multi-step ahead forecasting
        - suffers selection bias, combining PTCV method

Contribution

- Propose two alternative algorithms developed from IJR's framework
  - Bootstrap Method
  - Simple Selection Method
- Keep the desired properties of the original method
  - Weak dependence
  - Multi-step ahead forecasting
  - Asymptotic validity

Model Framework

- Suppose we forecast y_{T+h} at time T
- The optimal forecast is given by

$$\hat{y}_{T+h} = x'_T \hat{\beta}_{\hat{R}} \tag{1}$$

- $\hat{\beta}_{\hat{R}}$ is the OLS estimates, using the most recent $\hat{R}$ observations (known as the window size)

Model Framework

- The optimal window size $\hat{R}$ is given by

$$\hat{R} \equiv \quad arg \min_{R \in \Theta_R} [\hat{\beta}_R(1) - \tilde{\beta}(1)]' x_T x_T' [\hat{\beta}_R(1) - \tilde{\beta}(1)] \qquad (2)$$

  where

$$\begin{bmatrix} \tilde{\beta}(1) \\ \tilde{\beta}^{(1)}(1) \end{bmatrix} = \begin{bmatrix} \sum x_t x_t' & \sum x_t x_t' (\frac{t-T}{T}) \\ \sum x_t x_t' (\frac{t-T}{T}) & \sum x_t x_t' (\frac{t-T}{T})^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t y_{t+h} \\ \sum x_t y_{t+h} (\frac{t-T}{T}) \end{bmatrix} \qquad (3)$$

  - $\sum$ represents $\sum_{t=T-S+1}^{t=T-h}$
  - $S \geqslant 2k$ is an arbitrary number

- **The choice of S matters!**
- IJR chooses S using PTCV method
  - it may suffer from selection bias
  - its forecasting performance can be improved furthermore

Proposed Bootstrap Method

- Consider an optimization problem

$$S^{\star} \equiv arg \min_{S \in \Psi} \sum_{m=1}^{B} (y_{T+h}^{(m)} - \hat{y}_{T+h|T,S}^{(m)})^2 \quad (4)$$

  where
  - $y_{T+h}^{(m)}$ is the outcome at time $T+h$ for the m-th replication
  - $\hat{y}_{T+h|T,S}^{(m)}$ is the h-step ahead forecast at time T under S for the m-th replication
  - $\Psi = \{s\}_{s=2k}^{T}$ is the set of S
- B is the number of bootstrap re-sampling

Proposed Bootstrap Method

1. Partition the data into two periods according to the break date T_b as $\{y_t, x_t\}_{t=1}^{T_b}$ and $\{y_t, x_t\}_{t=T_b+1}^{T}$
2. Estimate parameter vectors $\hat{\beta}_1$ and $\hat{\beta}_2$ by OLS
3. Compute residuals $\{\hat{\epsilon}_{1,t}\}_{t=1+h}^{T_b}$ and $\{\hat{\epsilon}_{2,t}\}_{t=T_b+1+h}^{T}$
4. Centre estimated residuals as the empirical distribution function (EDF) E_1 and E_2
5. Resample residuals with replacement from the EDFs
    a. resample T_b residuals $\{\epsilon^{\star}_{1,t}\}_{t=1+h}^{T_b+h}$ from E_1
    b. resample $(T - T_b)$ residuals $\{\epsilon^{\star}_{2,t}\}_{t=T_b+1+h}^{T+h}$ from E_2

Proposed Bootstrap Method

6. Generate a bootstrap sample $\{y_t^\star\}_{t=1}^{T+h}$ with updates
   a. $y_{t+h}^\star = \hat{\beta}_1' x_t^\star + \epsilon_{1,t+h}^\star, \quad t = 1, 2, \cdots, T_b$
   b. $y_{t+h}^\star = \hat{\beta}_2' x_t^\star + \epsilon_{2,t+h}^\star, \quad t = T_b + 1, T_b + 2, \cdots, T$
7. Repeat steps 5-6, and generate B bootstrap samples, containing the information of the break in the original series
8. Apply (4) to choose the estimation window size for $\tilde{\beta}(1)$, S
9. Using S in step 8, apply (2) and (3) to select the optimal window size for forecasting

Proposed Simple Selection Method

- Concerning the computation burden of introducing the bootstrap, simplify the decision rule
- Estimate $\tilde{\beta}(1)$ using only post-break data
- $S = T - T_b$
- In practice, the break dates can be estimated by using the Sup-F test in Bai and Perron (1998, Econometrica)

Table: Comparison of Four Methods

Method	PTCV	IJR	Bootstrap	Simple Selection
Lagged Dependent Variables	No	Allowed	Allowed	Allowed
Correlated Error Terms	No	Allowed	NA	Allowed
Multi-step Ahead Forecasts	No	Allowed	Allowed	Allowed
Computation Burden	Medium	Heavy	Extremely Heavy	Medium

Monte-Carlo Study

- **Object**
  Test the forecasting performance of the proposed methods against that of existing methods under a structural break
- **Experiment Design**
  - Data Generating Process (DGP)

$$\begin{bmatrix} y_{t+1} \\ w_{t+1} \end{bmatrix} = \begin{bmatrix} a_t & b_t \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} y_t \\ w_t \end{bmatrix} + \begin{bmatrix} \mu_{t+1} \\ \upsilon_{t+1} \end{bmatrix} \tag{5}$$

where

$$\begin{bmatrix} \mu_{t+1} \\ \upsilon_{t+1} \end{bmatrix} \sim i.i.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

  - A break on either a_t or b_t at time T_b is engaged
  - Various setups on break size and break date (T_b) are used

Monte-Carlo Study

- **Forecast Methods**
  - Post-break Method ("PB")
  - PT's CV Method ("PTCV")
  - IJR Method ("IJR")
  - Proposed Bootstap Method ("My1")
  - Proposed Simple Selection Method ("My2")

Results

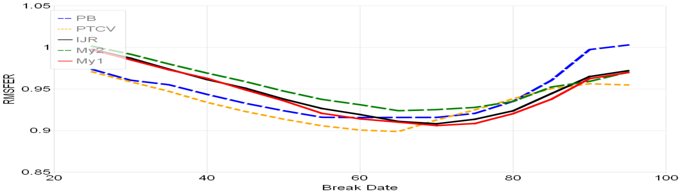
- Sample size $T = 100$
- One-step ahead forecasting practice $h = 1$
- 5000 Monte-Carlo simulations
- Benchmark: forecasts using the whole sample
- Criterion of forecast performance: ratio of square roots of MSFE (RMSFER)

$$\sqrt{\frac{\sum_{m=1}^{5000}(y_{T+1}^{(m)} - \hat{y}_{T+1}^{(m)})^2}{\sum_{m=1}^{5000}(y_{T+1}^{(m)} - \tilde{y}_{T+1}^{(m)})^2}}, \tag{6}$$

Results

- A small break on AR parameter with varying break date

Figure: RMSFER against break date

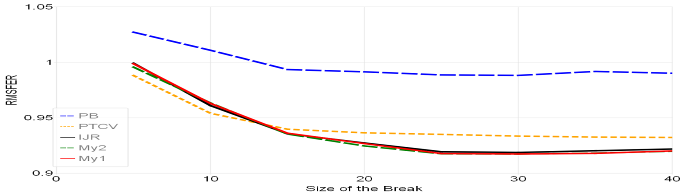


- "PTCV" dominates when the break date is before $0.65T$
- "My1" dominates when the break date is at $0.7T \sim 0.85T$
- "PTCV" dominates again when the break date is after $0.9T$

Results

- A break on AR parameter with varying break size at $T_b = 90$

Figure: RMSFER against break size

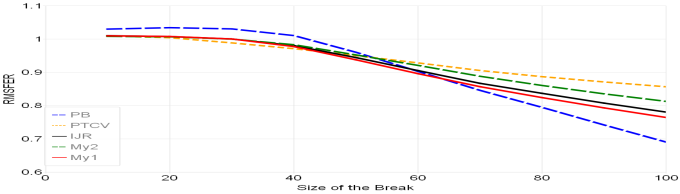


- Proposed "My1" and "My2" dominate others when the AR parameter shifts down by $0.15 \sim 0.4$

Results

- A break on marginal coefficient with varying break size at $T_b = 85$

Figure: RMSFER against break size



- "PTCV" dominates when the break size is small
- "My1" dominates when the break size is medium
- "PB" dominates when the break size is large

Conclusion

- The proposed bootstrap method outperforms IJR's original method in almost all cases
- The proposed bootstrap method performs best when there is a medium break close to the date of making forecasts
- If the break date is close to the forecast date, a small trimming value (e.g. 0.05) in Bai and Perron's (1998, Econometrica) test is preferred when using my bootstrap method.
- The proposed simple selection method performs well when the break occurs very close to the date of making forecasts
- When the break size is significant and the break date is far from the date of making forecasts, using post-break data only is almost always the best strategy

Discussion

- Caveats
  - What if there are more than one break (multiple breaks)
  - What if the parameter is time-varying
  - Extension to asymmetric loss function
  - When there exists weak dependence, the bootstrap may not be valid
    - Residual autocorrelation
    - Heteroscedasticity
  - Neither I or IJR investigated the ratio of the shift in mean and the variance

Thank you!

References I

BAI, J., AND P. PERRON (1998): "Estimating and testing linear models with multiple structural changes," *Econometrica*, pp. 47–78.

INOUE, A., L. JIN, AND B. ROSSI (2017): "Rolling window selection for out-of-sample forecasting with time-varying parameters," *Journal of Econometrics*, 196(1), 55–67.

PESARAN, M. H., AND A. TIMMERMANN (2007): "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137(1), 134–161.