Forecasting with Model Uncertainty: Representations and Risk Reduction

Keisuke Hirano and Jonathan H. Wright

University of Arizona and Johns Hopkins University

ECB, June 13 2014

Introduction

- Controversy between in-sample and OOS
- Considers forecasting with weak predictors
- Present paper highlights important effect of bagging
- Without bagging ordering is approximately:
    1. In-sample + AIC
    2. Out-of-sample
    3. Split sample

Introduction

- Controversy between in-sample and OOS
- Considers forecasting with weak predictors
- Present paper highlights important effect of bagging
- Without bagging ordering is approximately:
  1. In-sample + AIC
  2. Out-of-sample
  3. Split sample
- With bagging, it's generally reversed
- With alternate form of bagging, can prove that OOS and SS are dominated by bagging counterparts

Setup

Regression Model:

$$y_t = \beta' x_t + u_t$$

- k regressors (k fixed)
- $E[x_t x_t'] = \Sigma_{xx} = I_k$
- u_t IID, independent of x
- Local parametrization: $\beta = T^{-1/2} b$ (Inoue & Kilian (2006))

Forecast Assessment

Forecast: $\tilde{y}_{T+1} = \tilde{\beta}' x_{T+1}$.

- Unconditional MSPE

$$E[(y_{T+1} - \tilde{\beta}' x_{T+1})^2] = \sigma^2 + E\left[(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta)\right] + o_p(T^{-1})$$

- First term is $O(1)$ and same for all methods
- Second term is $O(T^{-1})$
- Normalized MSPE:

$$NMSPE = T(MSPE - \sigma^2) = E\left[T(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta)\right]$$

Forecasting Procedures

With k regressors, there are 2^k possible subsets.

- Big Model (OLS with all predictors)
- Small Model: $\tilde{\beta} = 0$.
- Positive-part James-Stein (shrinkage)
- Select model using AIC
- Out-of-sample forecasting
- Split-sample forecasting
- All methods with bagging

Bagging

Bagging = Bootstrap Aggregation (Breiman, 1996)

- Draw a bootstrap sample $\{x_t^*(i), y_t^*(i)\}$ from the original data $\{x_t, y_t\}$.
- Recompute estimator $\tilde{\beta}^*(i)$.
- Repeat for many bootstrap samples ($i = 1, \ldots, L$), average and generate the forecast
- Bühlmann and Yu (2002): bagging smooths hard-threshold estimators
- Inoue and Kilian (2008): application to forecasting CPI

Theorem 2: Limiting Distributions of Estimators

- OLS: $T^{1/2}\tilde{\beta} \to_d Y = N(b, \sigma^2)$
- JS: $T^{1/2}\tilde{\beta} \to_d S_1(Y) = Yw_1(Y)$
- AIC: $T^{1/2}\tilde{\beta} \to_d S_2(Y) = Yw_2(Y)$

Theorem 2: Limiting Distributions of Estimators

- OLS: $T^{1/2}\tilde{\beta} \to_d Y = N(b, \sigma^2)$
- JS: $T^{1/2}\tilde{\beta} \to_d S_1(Y) = Yw_1(Y)$
- AIC: $T^{1/2}\tilde{\beta} \to_d S_2(Y) = Yw_2(Y)$
- OOS: $T^{1/2}\tilde{\beta} \to_d S_3(Y, U_B)$
  where U_B is a Brownian bridge independent of Y and b
- SS: $T^{1/2}\tilde{\beta} \to_d S_4(Y, U_B)$

Representation of Partial Sums

All of the procedures we consider depend crucially on the partial sum process ($r \in [0,1]$): $T^{-1/2} \sum_{t=1}^{[Tr]} x_t y_t$

Theorem 1:

$$T^{-1/2} \sum_{t=1}^{[Tr]} x_t y_t \to_d rY + \sigma U_B(r)$$

where $Y \sim N(b, \sigma^2)$ and U_B is a Brownian bridge independent of Y and b

Adding Bagging Step

- Theorem 3: In the ith bootstrap step

$$T^{-1/2}\Sigma_{t=1}^{[Tr]} x_t^*(i) y_t^*(i) \to_d rY + \sigma V_i(r)$$

where V_i are independent Brownian *motions* (Park, 2002).

Limiting Distributions of Estimators with Bagging

- OLS $T^{1/2}\tilde{\beta}_i \rightarrow_d Y + V_i$
- JS: $T^{1/2}\tilde{\beta}_i \rightarrow_d S_1(Y, V_i)$
- AIC: $T^{1/2}\tilde{\beta}_i \rightarrow_d S_2(Y, V_i)$
- OOS: $T^{1/2}\tilde{\beta}_i \rightarrow_d S_3(Y, V_i)$
  where V_i is a Brownian motion independent of Y and b
- SS: $T^{1/2}\tilde{\beta}_i \rightarrow_d S_4(Y, V_i)$
- Repeating across different i and averaging means that all estimators eliminate V_i and are generalized shrinkage estimators.

Bagging Comments

- For OOS and SS, bagging replaces U_B with V_i and then eliminates by integration.
- Intuition: for SS, bagging randomizes over partitions of the data $\Rightarrow$ uses all obs for both model selection and estimation

Simpler Representations with $k = 1$

- AIC without bagging: $T^{1/2}\tilde{\beta} \to_d Y\mathbf{1}(Y > \sqrt{2}\sigma)$
- SS without bagging: $Z_1\mathbf{1}(|Z_2| > \sqrt{2/\pi\sigma})$

  where $Z_1 \sim N(b, \frac{\sigma^2}{1-\pi}) \perp \ Z_2 \sim N(b, \frac{\sigma^2}{\pi})$
- AIC with bagging:
  $Y - Y\Phi(\frac{\sqrt{2}\sigma - Y}{\sigma}) + \sigma\phi(\frac{\sqrt{2}\sigma - Y}{\sigma}) + Y\Phi(\frac{-\sqrt{2}\sigma - Y}{\sigma}) - \sigma\phi(\frac{-\sqrt{2}\sigma - Y}{\sigma})$
- SS with bagging: $Y - Y\Phi(\frac{\sqrt{2}\sigma - \sqrt{\pi}Y}{\sigma}) + Y\Phi(\frac{-\sqrt{2}\sigma - \sqrt{\pi}Y}{\sigma})$

Risk Reduction

- In the limit, OOS and SS are functionals of *both* $Y = Y(1)$ and $U = U_B$.
- But Y is sufficient.
- Marginalize out the random noise term U:

$$\tilde{S}(Y) = E\left[S(Y,U) \mid Y\right].$$

- By the Rao-Blackwell theorem,

$$MSPE(\tilde{S}, b) \leq MSPE(S, b) \quad \forall b$$

Risk Reduction

- Calculations indicate strict risk reduction for at least some b.
- Hence OOS and SS are asymptotically inadmissible.

Risk Reduction

- Calculations indicate strict risk reduction for at least some b.
- Hence OOS and SS are asymptotically inadmissible.
- Bagging is like Rao-Blackwellization wrt V instead of U.
- Might want to do Rao-Blackwellization or an alternative form of bagging that achieves this.

Alternative Form of Bagging

- All estimators are functions of $x_t x_t'$ and $x_t y_t$ alone.
- Let

$$z_t = x_t y_t = x_t x_t' \hat{\beta} + x_t e_t$$

and define the ith bootstrap draw of z_t as:

$$z_t^*(i) = x_t x_t' \hat{\beta} + \theta_t(i) x_t e_t - T^{-1} \Sigma_{s=1}^{T} \theta_s(i) x_s e_s$$

where $\theta_t(i)$ is the "wild" term.

Alternative Form of Bagging

- All estimators are functions of $x_t x_t'$ and $x_t y_t$ alone.
- Let

$$z_t = x_t y_t = x_t x_t' \hat{\beta} + x_t e_t$$

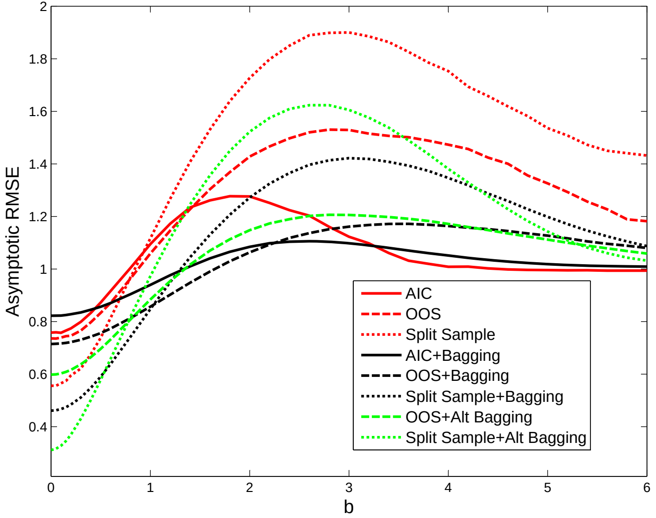
and define the ith bootstrap draw of z_t as:

$$z_t^*(i) = x_t x_t' \hat{\beta} + \theta_t(i) x_t e_t - T^{-1} \Sigma_{s=1}^{T} \theta_s(i) x_s e_s$$

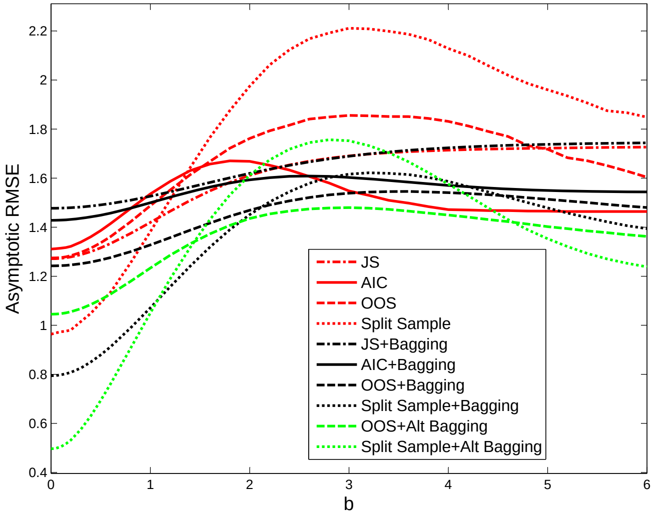
where $\theta_t(i)$ is the "wild" term.

- Theorem 4: Limiting distributions same as Theorem 2 but with $Y(r) = rY + \sigma U_B(r)$ replaced by $rY + \sigma U_B^i(r)$

Asymptotic Root NMSPE (k=1)



Asymptotic Root NMSPE (k=3)



Dominance Relations (1 nonzero coefficient)

k	1	2	3	4	5	6
AIC v OOS						
AIC v SS						
AIC v AICB						
AIC v OOSB	OOSB	OOSB	OOSB	OOSB	OOSB	OOSB
AIC v SSB				SSB	SSB	SSB
OOS v SS						
OOS v AICB						
OOS v OOSB	OOSB	OOSB	OOSB	OOSB	OOSB	OOSB
OOS v SSB	SSB	SSB	SSB	SSB	SSB	SSB
SS v AICB						
SS v OOSB						
SS v SSB	SSB	SSB	SSB	SSB	SSB	SSB
AICB v OOSB		OOSB	OOSB	OOSB	OOSB	OOSB
AICB v SSB			SSB	SSB	SSB	SSB
OOSB v SSB						

Dominance Relations (2 nonzero coefficients)

k	1	2	3	4	5	6
AIC v OOS						
AIC v SS						
AIC v AICB						
AIC v OOSB						
AIC v SSB						
OOS v SS						
OOS v AICB						
OOS v OOSB	OOSB	OOSB	OOSB	OOSB	OOSB	OOSB
OOS v SSB	SSB	SSB	SSB	SSB	SSB	SSB
SS v AICB						
SS v OOSB						
SS v SSB	SSB	SSB	SSB	SSB	SSB	SSB
AICB v OOSB					OOSB	OOSB
AICB v SSB					SSB	SSB
OOSB v SSB						

Comparison of Bayes Risk

- Prior:
  - Each regressor is included in the model with probability p.
  - Conditional on inclusion, prior for that element of b is $N(0, \phi)$.
- Can work out local asymptotic Bayes risk: limit of

$$E[(T^{1/2}\tilde{\beta} - b)'(T^{1/2}\tilde{\beta} - b)]$$

Comparison of Bayes Risk

- Prior:
  - Each regressor is included in the model with probability p.
  - Conditional on inclusion, prior for that element of b is $N(0, \phi)$.
- Can work out local asymptotic Bayes risk: limit of

$$E[(T^{1/2}\tilde{\beta} - b)'(T^{1/2}\tilde{\beta} - b)]$$

- OOS/SS with bagging do well
- But BMA always does better, and can do much better

h-step ahead forecasting

- Setup:

$$y_{t+h} = \beta' x_t + u_t$$

- Serial correlation in u_t could be exploited but isn't.

h-step ahead forecasting

- Setup:

$$y_{t+h} = \beta' x_t + u_t$$

- Serial correlation in u_t could be exploited but isn't.
- Without bagging

$$T^{-1/2} \Sigma_{t=1}^{[Tr]} x_t(i) y_t(i) \to_d rN(b, \omega^2 I) + \omega U_B(r)$$

- With bagging

$$T^{-1/2} \Sigma_{t=1}^{[Tr]} x_t^*(i) y_t^*(i) \to_d rN(b, \omega^2 I) + \sigma V_i(r)$$

h-step ahead forecasting

- Could get bagging to “mimic” serial dependence in the data.
  - Draw blocks of data of length that goes to infinity slowly.

h-step ahead forecasting

- Could get bagging to “mimic” serial dependence in the data.
  - Draw blocks of data of length that goes to infinity slowly.
- Easy to do Rao-Blackwellization with serial correlation

Forecasting in a VAR

- A p-variable stationary VAR with k lags and intercept:

$$y_t = Bx_t + \varepsilon_t$$

- Suppose that $B = CT^{-1/2}$.
- Each model consists of a set of zero restrictions on B.

Forecasting in a VAR

- A p-variable stationary VAR with k lags and intercept:

$$y_t = Bx_t + \varepsilon_t$$

- Suppose that $B = CT^{-1/2}$.
- Each model consists of a set of zero restrictions on B.
- All estimators depend on:
  - $T^{-1}\Sigma_{t=1}^{[Tr]} x_t x_t' \ \to_r r\Omega_{xx}$ where $\Omega_{xx} = E(x_t x_t')$
  - $T^{-1/2}\Sigma_{t=1}^{[Tr]} y_t x_t' \to_d [rC + B(r)]\Omega_{xx}$
- Estimators other than OOS or SS are functions of $Y \equiv C + B(1)$ alone
- OOS and SS are functions of Y and $U_B(r)$.

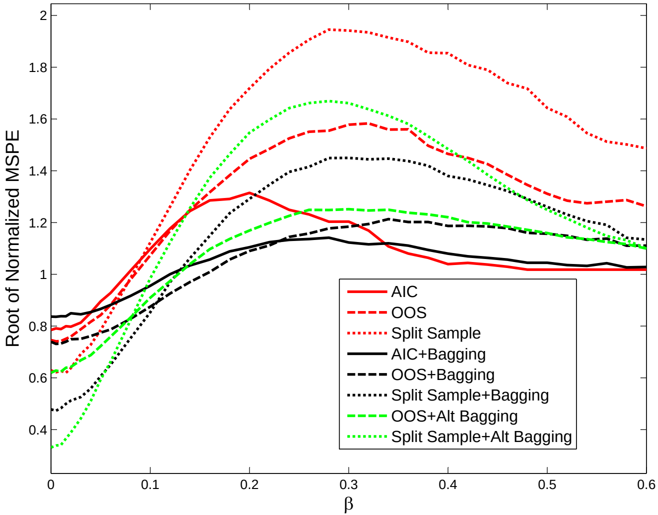
Extension to general likelihood framework

- Parameter θ and likelihood $l(\theta) = \Sigma_{t=1}^{T} l_t(\theta)$
- True value is $\theta_0 = cT^{-1/2}$
- Model selection amounts to imposing zeros on θ
- Need $T^{-1/2}\Sigma_{t=1}^{[Tr]} l_t'(\theta_0) \to B(r)$

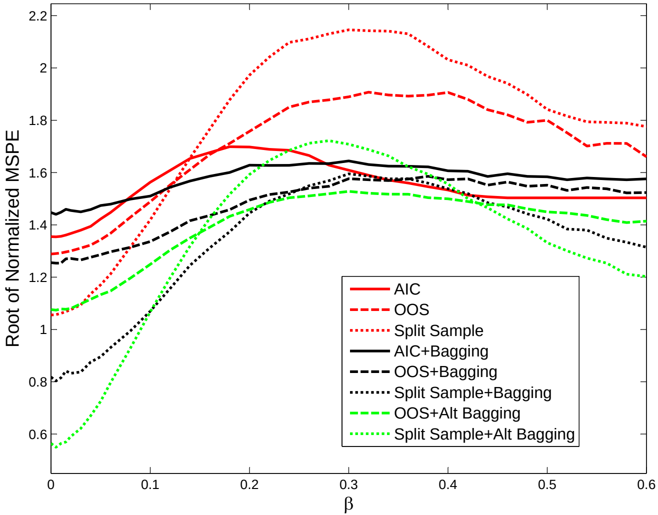
Monte-Carlo Simulation

- Monte-Carlo simulation with Gaussian shocks and $T = 100$
- Evaluated normalized root mean square prediction error $\sqrt{T * (MSPE - 1)})$

Monte-Carlo Root NMSPE (k=1)



Monte-Carlo Root NMSPE (k=3)



Conclusion

- Representation highlights dependence of OOS and SS "noise"
- This can be eliminated by bagging
- Or by Rao-Blackwellization (alternative bagging)
- Standard and alternative bagging on OOS/SS compares favorably with existing methods

Recap (in haiku)

Out of sample is
Inadmissible, but the
Future's in the bag.