

SUPPLEMENT TO “FORECASTING WITH MODEL UNCERTAINTY:  
REPRESENTATIONS AND RISK REDUCTION”  
(*Econometrica*, Vol. 85, No. 2, March 2017, 617–643)

BY KEISUKE HIRANO AND JONATHAN H. WRIGHT

This supplement introduces some alternative procedures to the ones considered in the main text, and provides extended numerical comparisons of local asymptotic risk among the various methods. It also conducts a small Monte Carlo study of finite-sample risk, and provides a comparison of shrinkage factors for a number of the procedures.

WE FIRST INTRODUCE three alternative procedures: the positive part James–Stein estimator; the Laplace estimator; and LASSO.

1. *Positive Part James–Stein Estimator.* Let  $\hat{\beta}$  be the unrestricted MLE and  $\hat{V}$  be an estimate of its asymptotic variance–covariance matrix. The positive part James–Stein estimator for  $k > 2$  is

$$\tilde{\beta} = \hat{\beta} \cdot \max \left\{ 1 - \frac{k-2}{T \hat{\beta}' \hat{V}^{-1} \hat{\beta}}, 0 \right\}.$$

2. *Laplace Estimator.* Let  $t$  denote the vector of  $t$ -statistics consisting of the elements of  $\hat{\beta}$  each divided by the  $j$ th diagonal element of  $\hat{V}$ . The Laplace estimator of  $\beta$  is

$$\tilde{\beta} = \hat{\beta} \circ \tilde{h}(t),$$

where  $\tilde{h}(x) = (1 - \frac{ch(x)}{x})$  (element-by-element),  $h(x) = \frac{e^{-cx}\Phi(x-c) - e^{cx}\Phi(-x-c)}{e^{-cx}\Phi(x-c) + e^{cx}\Phi(-x-c)}$ ,  $c = \ln(2)$ , and  $\Phi(\cdot)$  is the standard normal c.d.f. This is equivalent to the Bayesian posterior mean in the normal model when the prior is a product of Laplace distributions (Magnus (2002), Magnus, Powell, and Prüfer (2010)). Therefore, this estimator is asymptotically admissible under our local asymptotics.

3. *LASSO Estimator.* Set  $\tilde{\beta}$  to maximize  $l(\beta) - \lambda_T \sum_{j=1}^k |\beta_j|$ , where  $\lambda_T$  is a penalty term. Penalized maximum likelihood estimators of this sort have been considered by Tibshirani (1996) and many others.

If we use the James–Stein estimator, then under the conditions in Proposition 3.5, its limiting distribution will be

$$T^{1/2} \tilde{\beta} \rightarrow_d \Sigma^{-1} Y(1) \max \left\{ 1 - \frac{k-2}{Y(1)' \Sigma^{-2} Y(1)}, 0 \right\}.$$

If we incorporate a bagging step, then

$$T^{1/2} \tilde{\beta} \rightarrow_d E^* \left[ \Sigma^{-1} Y^*(1) \max \left\{ 1 - \frac{k-2}{Y^*(1)' \Sigma^{-2} Y^*(1)}, 0 \right\} \right].$$

If we use the standard Laplace estimator, then under the conditions in Proposition 3.5, its limiting distribution will be

$$T^{1/2} \tilde{\beta} \rightarrow_d Y(1) \circ \tilde{h}(WY(1)),$$

where  $W$  is a diagonal matrix with the  $j$ th diagonal element equal to the reciprocal of the square root of the  $j$ th diagonal element of  $\Sigma^{-1}$ .

If we incorporate a bagging step, then

$$T^{1/2}\tilde{\beta} \rightarrow_d E^*[Y^*(1) \circ \tilde{h}(WY^*(1))].$$

In the linear regression model (Section 3.2), the results of [Knight and Fu \(2000\)](#) apply to the LASSO estimator. They showed that for the LASSO estimator, if  $\max_t T^{-1}x'_t x_t \rightarrow 0$  and  $T^{-1/2}\lambda_T \rightarrow \lambda_0 \geq 0$ , then

$$T^{1/2}\tilde{\beta} \rightarrow_d \arg \min_v v' \Sigma v - 2v' Y(1) + \lambda_0 \sum_{j=1}^k |v_j|.$$

If we incorporate a bagging step, then

$$T^{1/2}\tilde{\beta} \rightarrow_d E^*\left[\arg \min_v v' \Sigma v - 2v' Y^*(1) + \lambda_0 \sum_{j=1}^k |v_j|\right].$$

RB makes no difference to the James–Stein, Laplace, or LASSO estimators.

In the remainder of this document, we report some further numerical results:

1. We first report the same computations of local asymptotic risk as in Figures 1–3 of the paper, but adding in LASSO, James–Stein, and Laplace estimators, in their standard

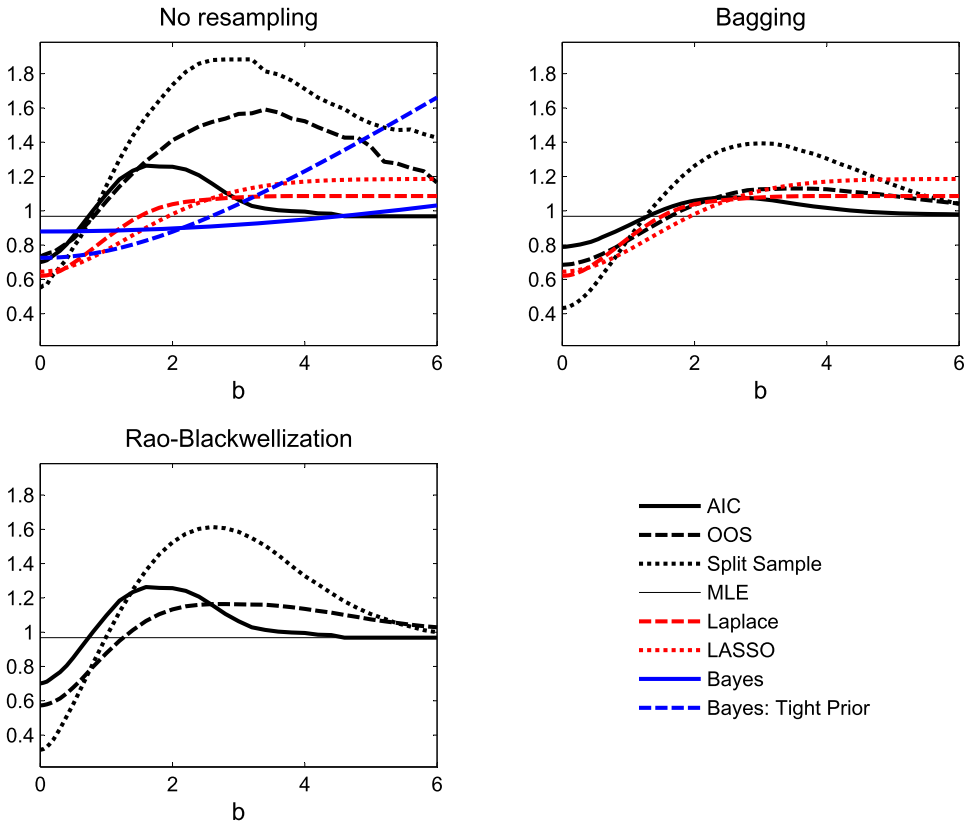


FIGURE A1.—Local asymptotic risk ( $k = 1$ ). These are the simulated local asymptotic risk values, Equation (5.1), for different procedures, plotted against  $b$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao–Blackwellization. AIC, Laplace, and LASSO are the same without any resampling or with RB.

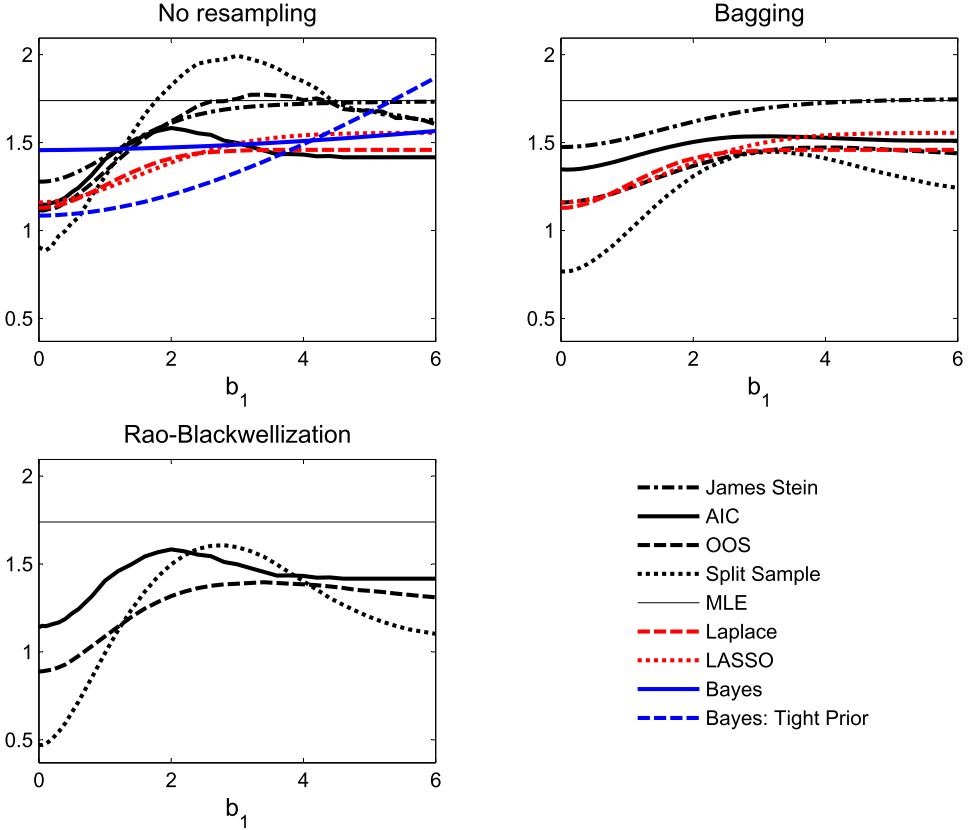


FIGURE A2.—Local asymptotic risk ( $k = 3$ ). These are the simulated local asymptotic risk values, Equation (5.1), for different procedures, plotted against  $b_1$ , where  $b = (b_1, 0, \dots, 0)'$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao–Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

form and with bagging (by construction, RB has no effect on them). LASSO is implemented with the penalty  $\lambda_T$  equal to the square root of sample size, using code from the SpaSM toolbox (Sjöstrand, Clemmensen, Larsen and Ersbøll (2012)).

These results are shown in Figures A1–A3. Bagging makes little difference to the LASSO or Laplace estimators. The LASSO and Laplace estimators have low risk for many values of  $b$  and are numerically quite similar to out-of-sample with bagging, except that for large  $b$ , LASSO and Laplace have higher risk. We also include a Bayesian estimator with a tight prior in Figures A1–A3. This prior is multivariate normal with mean zero and variance matrix  $\Omega = 3 \cdot \text{diag}\{1, 1/2, \dots, 1/k\}$ . As noted in the paper, by construction, neither bagging nor RB has any effect on the Bayesian estimator. Also, as noted in the paper, achieving the local asymptotic risk of this proper-prior Bayesian estimator would require using a sequence of shrinking priors in the original model. Using the tighter prior improves the performance of the Bayes estimator for small values of  $b$ , although its risk is higher than some of the other procedures for larger values of  $b$ .

2. We report further computations of local asymptotic risk with multiple predictors as in Section 5 of the paper, but where the associated coefficients are all equal. The LASSO and Laplace estimators are again included. In the notation of the paper, we specify that

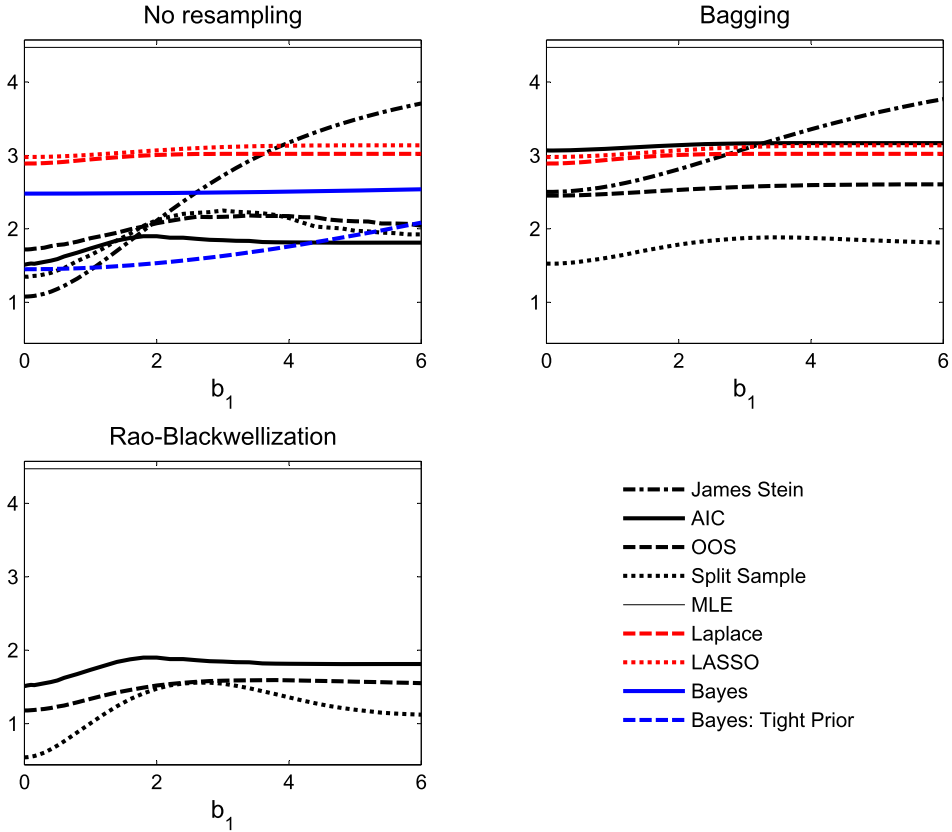


FIGURE A3.—Local asymptotic risk ( $k = 20$ ). These are the simulated local asymptotic risk values, Equation (5.1), for different procedures, plotted against  $b_1$ , where  $b = (b_1, 0, \dots, 0)$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao–Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

$b = b_1 k^{-1/2} i$ , where  $i$  denotes a  $k \times 1$  vector of ones. Figures A4 and A5 plot the risk for  $k = 3$  and  $k = 20$  against  $b_1$  for the different procedures. RB and bagging help the split-sample and out-of-sample methods a good deal.

3. We calculate the integrated local asymptotic risk for the various procedures, using a uniform weighting on  $[-\bar{b}, \bar{b}]^k$ , where  $k$  is the dimension of the parameter space,  $k = 1$  or 3, and  $\bar{b} = 3$  or 6. These are shown in Table A-I. The inefficiency of the out-of-sample and split-sample methods is apparent, but as noted in the main text, either bagging or Rao–Blackwellization improves them considerably and their integrated risk is similar, and sometimes lower, than the in-sample (AIC) procedure. We also include results for the Bayes estimator where the prior is flat on  $[-\bar{b}, \bar{b}]^k$ : the Bayes estimator of the localization parameter is then  $\max(\min(Y(1), \bar{b}), -\bar{b})$ . Not surprisingly, this gives lower integrated risk than any of the alternatives, although other methods can be quite close.

4. We consider the case where  $b$  has  $k$  elements and we do a grid search over possible values of  $\bar{k}$  of these elements, setting the remaining elements to zero. In Table A-II, we list the cases in which one method dominates another one uniformly over the nonzero elements of  $b$  in terms of local asymptotic risk for various pairs of possible forecasting

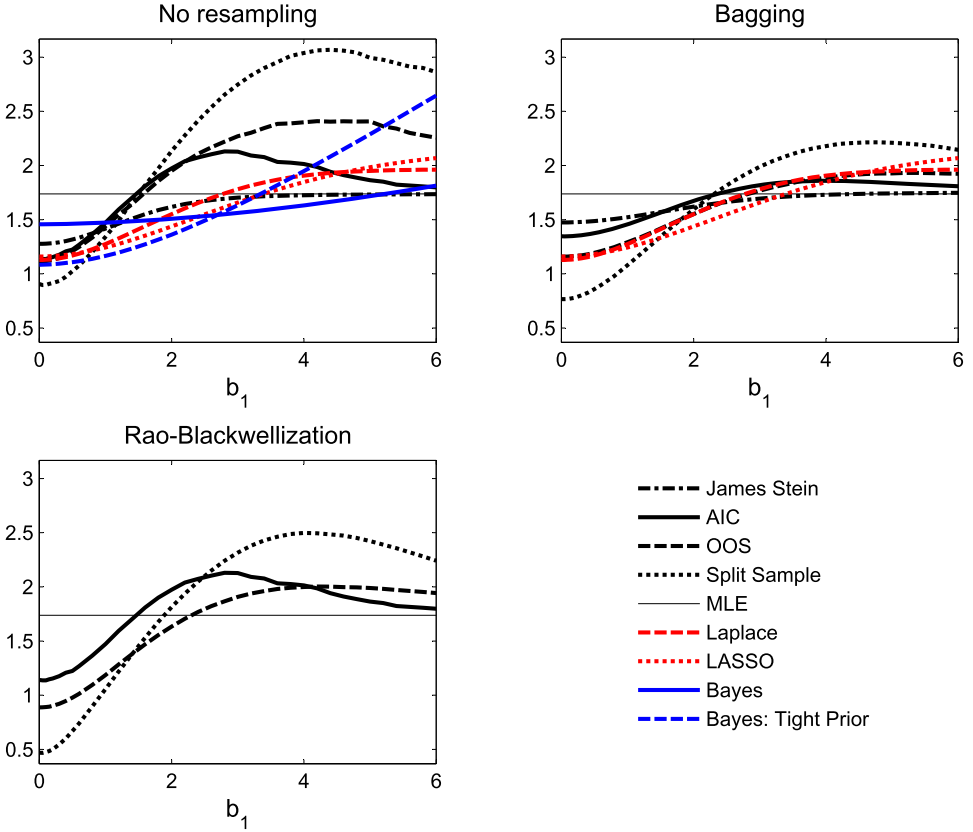


FIGURE A4.—Local asymptotic risk ( $k = 3$ , all coefficients equal). These are the simulated local asymptotic risk values, Equation (5.1), for different procedures, plotted against  $b_1$ , where  $b = b_1 k^{-1/2} i$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao–Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

methods. As this is done by grid search, it is only feasible for  $\bar{k} = 1, 2$ . We find that in all cases, the split-sample forecasts with RB or bagging dominate those without. For bagging, this is a numerical result, but for RB it is a theoretical one, as discussed above. The out-of-sample forecasts with RB dominate those without, though that is not true for bagging. In Table A-II, if  $k \geq 4$  and  $\bar{k} = 1$ , then the split-sample scheme with bagging or RB dominates in-sample forecasting (with or without bagging), the maximum-likelihood estimator, and the James–Stein estimator. It seems that the split-sample forecasting scheme with bagging or RB does best if the model is sparse—there are multiple coefficients, most of which are equal to zero. The out-of-sample scheme with RB dominates in-sample forecasting (with or without bagging), the maximum-likelihood estimator, and the James–Stein estimator if  $k \geq 3$  and  $\bar{k} = 1$ .

5. To assess whether the local asymptotic theory provides useful approximations to the finite-sample performance of the procedures we examine in the paper, we conducted a small Monte Carlo study. We simulated the linear regression model in Section 2.1 with  $t(5)$  errors<sup>1</sup> scaled to have unit variance, independent standard normal regressors, sample

<sup>1</sup>Results with normal errors were very similar.

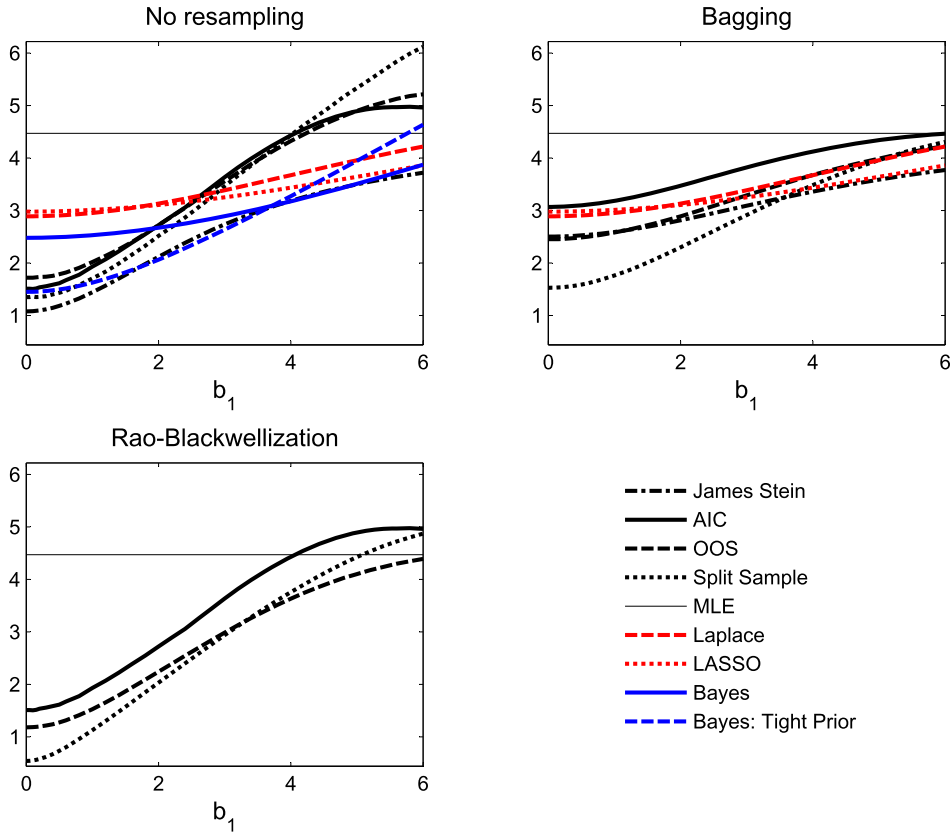


FIGURE A5.—Local asymptotic risk ( $k = 20$ , all coefficients equal). These are the simulated local asymptotic risk values, Equation (5.1), for different procedures, plotted against  $b_1$ , where  $b = b_1 k^{-1/2} i$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao–Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

size  $T = 100$ , and different values of  $k$ . In each simulation, we drew  $T + 1$  observations on  $y_t$  and  $x_t$ , and used the first  $T$  for model selection and parameter estimation according to one of the methods discussed in the paper or above. Then, given  $x_{T+1}$ , we worked out the prediction for  $y_{T+1}$ , and computed the mean squared prediction error (MSPE).

Figure A6 plots the simulated root normalized mean squared prediction errors ( $\sqrt{T} * (\text{MSPE} - 1)$ ) with  $k = 1$  against  $\beta$ . Figures A7 and A8 plot the simulated root normalized mean squared prediction errors with  $k = 3$  and  $k = 20$ , where  $\beta = (\beta_1, 0, \dots, 0)'$  against  $\beta_1$ , also including James–Stein estimators. Figures A9 and A10 do the same where  $\beta = \beta_1 k^{-1/2} i$ . The Monte Carlo results are generally consistent with local asymptotics. In the case  $k = 20$ , MSPEs for all methods tend to be a bit higher than would be expected under the local asymptotic approximations, because of estimation of variance–covariance matrices.

6. We consider Monte Carlo simulations where the data generating process consists of  $y_t = \beta_t x_t + u_t$ , where  $x_t$  are independent standard normal regressors,  $\beta_t$  is a Gaussian random walk with innovation standard deviation  $\omega$ , the sample size is  $T = 100$ , and  $u_t$  are  $t(5)$  errors with variance scaled to unity. Although the data generating process features time-varying parameters, we suppose that the researcher does not know this, and applies

TABLE A-I  
INTEGRATED LOCAL ASYMPTOTIC RISK<sup>a</sup>

| Method   | $k = 1$       | $k = 1$       | $k = 3$       | $k = 3$       |
|----------|---------------|---------------|---------------|---------------|
|          | $\bar{b} = 6$ | $\bar{b} = 3$ | $\bar{b} = 6$ | $\bar{b} = 3$ |
| MLE      | 1.00          | 1.00          | 3.00          | 3.00          |
| JS       | 1.00          | 1.00          | 2.95          | 2.76          |
| AIC      | 1.16          | 1.27          | 3.29          | 3.80          |
| Laplace  | 1.08          | 0.90          | 3.23          | 2.68          |
| LASSO    | 1.14          | 0.83          | 3.42          | 2.48          |
| Bayes    | 0.91          | 0.83          | 2.74          | 2.47          |
| OOS      | 1.75          | 1.52          | 4.54          | 4.57          |
| SS       | 2.43          | 2.08          | 7.05          | 6.51          |
| JSB      | 1.00          | 1.00          | 2.99          | 2.78          |
| AICB     | 1.05          | 1.01          | 3.11          | 3.03          |
| LaplaceB | 1.07          | 0.86          | 3.17          | 2.55          |
| LASSOB   | 1.13          | 0.83          | 3.37          | 2.45          |
| OOSB     | 1.12          | 0.93          | 3.26          | 2.84          |
| SSB      | 1.40          | 1.16          | 3.78          | 3.39          |
| OOSRB    | 1.17          | 1.02          | 3.36          | 3.13          |
| SSRB     | 1.63          | 1.55          | 4.20          | 4.42          |

<sup>a</sup>These are the simulated local asymptotic risk values,  $E[(T^{1/2}\bar{\beta} - b)'(T^{1/2}\bar{\beta} - b)]$  (the square of Equation (5.1)), for different procedures, integrated using a uniform weighting over the local parameter  $b$ , for  $b$  in  $[-\bar{b}, \bar{b}]^k$ . These represent simple average risk values and note that here we report the mean squared error, not the root mean squared error. Recall that MLE and Bayes are the same without any resampling, with bagging or with RB. AIC, LASSO, Laplace, and JS are the same without any resampling or with RB.

the methods discussed in the paper treating parameters as constant. Table A-III tabulates the simulated root normalized mean squared prediction errors for various values of  $\omega$ .

7. We lastly plot the implicit local asymptotic shrinkage functions for the AIC, out-of-sample, and split-sample estimators, in their standard form, with bagging and with RB in the case  $k = 1$  ( $g(Y)$  in the notation of Appendix B). These are shown in Figure A11. The out-of-sample procedure with RB has a function  $g(Y)$  that is apparently continuous but that is equal to zero for small values of  $|Y|$ . It thus combines shrinkage with variable selection, in common with LASSO. The split-sample procedure with RB has a function  $g(Y)$  that goes negative for small values of  $|Y|$ .

TABLE A-II  
DOMINANCE RELATIONS IN LOCAL ASYMPTOTIC RISK<sup>a</sup>

| $k$           | 1       | 2       | 3       | 4       | 5       | 6       | 2       | 3       | 4       | 5       | 6       |
|---------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| $\bar{k}$     | 1       | 1       | 1       | 1       | 1       | 1       | 2       | 2       | 2       | 2       | 2       |
| MLE v. AICB   | “_”     | “AICB”  | “AICB”  | “AICB”  | “AICB”  | “AICB”  | “_”     | “AICB”  | “AICB”  | “AICB”  | “AICB”  |
| MLE v. OOSB   | “_”     | “OOSB”  | “OOSB”  | “OOSB”  | “OOSB”  | “OOSB”  | “_”     | “OOSB”  | “OOSB”  | “OOSB”  | “OOSB”  |
| MLE v. SSB    | “_”     | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “_”     | “_”     | “SSB”   | “SSB”   | “SSB”   |
| MLE v. OOSRB  | “_”     | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “_”     | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” |
| MLE v. SSRB   | “_”     | “_”     | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “_”     | “_”     | “SSRB”  | “SSRB”  | “SSRB”  |
| JS v. AICB    | “NA”    | “NA”    | “_”     | “_”     | “_”     | “_”     | “NA”    | “_”     | “_”     | “_”     | “_”     |
| JS v. OOSB    | “NA”    | “NA”    | “OOSB”  | “_”     | “_”     | “_”     | “NA”    | “OOSB”  | “_”     | “_”     | “_”     |
| JS v. SSB     | “NA”    | “NA”    | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “NA”    | “_”     | “SSB”   | “SSB”   | “SSB”   |
| JS v. OOSRB   | “NA”    | “NA”    | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “NA”    | “_”     | “OOSRB” | “OOSRB” | “OOSRB” |
| JS v. SSRB    | “NA”    | “NA”    | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “NA”    | “_”     | “_”     | “_”     | “SSRB”  |
| AIC v. AICB   | “_”     | “_”     | “_”     | “_”     | “AIC”   | “AIC”   | “_”     | “_”     | “_”     | “_”     | “_”     |
| AIC v. OOSB   | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     |
| AIC v. SSB    | “_”     | “_”     | “_”     | “SSB”   | “SSB”   | “SSB”   | “_”     | “_”     | “_”     | “_”     | “SSB”   |
| AIC v. OOSRB  | “_”     | “_”     | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “_”     | “_”     | “OOSRB” | “OOSRB” | “OOSRB” |
| AIC v. SSRB   | “_”     | “_”     | “_”     | “SSRB”  | “SSRB”  | “SSRB”  | “_”     | “_”     | “_”     | “_”     | “_”     |
| OOS v. AICB   | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     |
| OOS v. OOSB   | “OOSB”  | “OOSB”  | “_”     | “_”     | “_”     | “_”     | “OOSB”  | “_”     | “_”     | “_”     | “_”     |
| OOS v. SSB    | “_”     | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   |
| OOS v. OOSRB  | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” |
| OOS v. SSRB   | “_”     | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “_”     | “_”     | “SSRB”  | “SSRB”  | “SSRB”  |
| SS v. AICB    | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     |
| SS v. OOSB    | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     | “_”     |
| SS v. SSB     | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “SSB”   |
| SS v. OOSRB   | “_”     | “_”     | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “_”     | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” |
| SS v. SSRB    | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  | “SSRB”  |
| AICB v. OOSB  | “_”     | “_”     | “OOSB”  | “OOSB”  | “OOSB”  | “OOSB”  | “_”     | “_”     | “OOSB”  | “OOSB”  | “OOSB”  |
| AICB v. SSB   | “_”     | “_”     | “SSB”   | “SSB”   | “SSB”   | “SSB”   | “_”     | “_”     | “SSB”   | “SSB”   | “SSB”   |
| AICB v. OOSRB | “_”     | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “_”     | “_”     | “OOSRB” | “OOSRB” | “OOSRB” |

(Continues)



TABLE A-II—Continued

| $k$           | 1   | 2   | 3       | 4       | 5       | 6       | 2   | 3   | 4   | 5       | 6       |
|---------------|-----|-----|---------|---------|---------|---------|-----|-----|-----|---------|---------|
| $\tilde{k}$   | 1   | 1   | 1       | 1       | 1       | 1       | 2   | 2   | 2   | 2       | 2       |
| AICB v. SSRB  | “_” | “_” | “_”     | “SSRB”  | “SSRB”  | “SSRB”  | “_” | “_” | “_” | “_”     | “SSRB”  |
| OOSB v. SSB   | “_” | “_” | “_”     | “SSB”   | “SSB”   | “SSB”   | “_” | “_” | “_” | “_”     | “SSB”   |
| OOSB v. OOSRB | “_” | “_” | “OOSRB” | “OOSRB” | “OOSRB” | “OOSRB” | “_” | “_” | “_” | “OOSRB” | “OOSRB” |
| OOSB v. SSRB  | “_” | “_” | “_”     | “_”     | “_”     | “SSRB”  | “_” | “_” | “_” | “_”     | “_”     |
| SSB v. OOSRB  | “_” | “_” | “_”     | “_”     | “_”     | “_”     | “_” | “_” | “_” | “_”     | “_”     |
| SSB v. SSRB   | “_” | “_” | “_”     | “_”     | “_”     | “_”     | “_” | “_” | “_” | “_”     | “_”     |
| OOSRB v. SSRB | “_” | “_” | “_”     | “_”     | “_”     | “_”     | “_” | “_” | “_” | “_”     | “_”     |

<sup>a</sup>This table reports comparisons of local asymptotic risk between pairs of methods: maximum likelihood (MLE), positive part James–Stein estimator (JS) applicable if  $k > 2$ , in-sample with AIC (AIC), the counterpart with bagging (AICB), out-of-sample (OOS), the counterpart with bagging/Rao–Blackwellization (OOSB/OOSRB), the split-sample method (SS), and the counterpart with bagging/Rao–Blackwellization (SSB/SSRB). Results are shown for different numbers of predictors  $k$ . For each pairwise comparison, the table lists which method is uniformly dominant when only  $k$  of the predictors are in fact nonzero. If neither is dominant, then the entry in the table is “\_”.

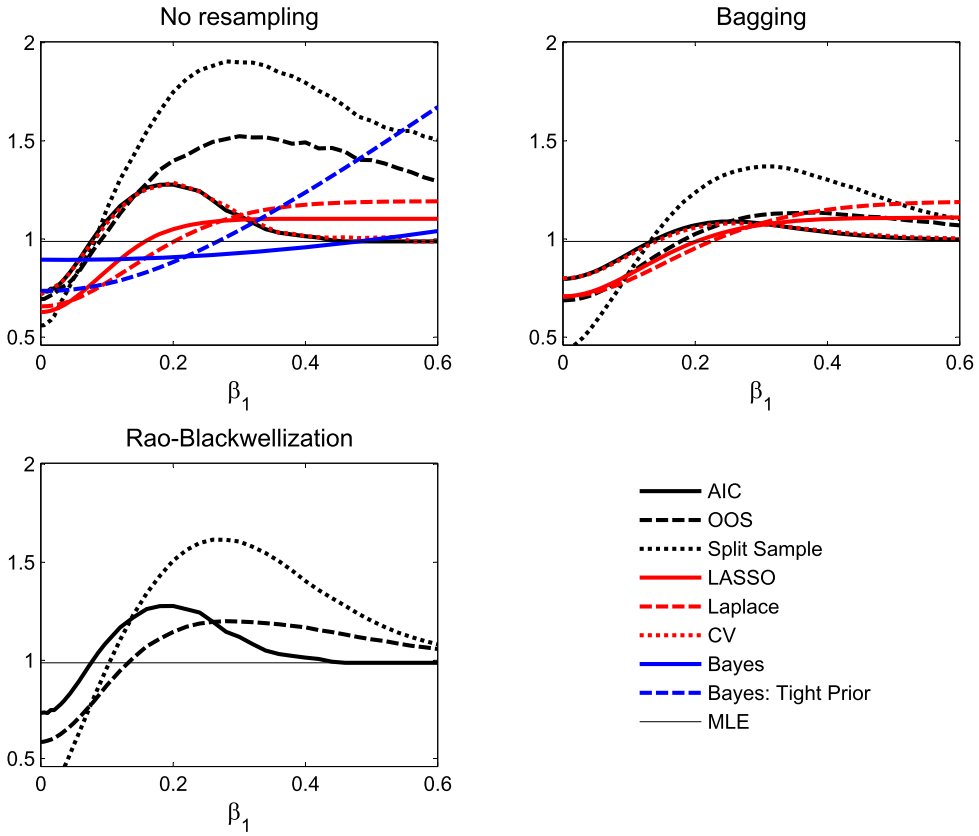


FIGURE A6.—Root normalized mean squared prediction errors ( $k = 1$ ). These are the simulated root normalized mean squared prediction errors using different procedures, plotted against  $\beta$ . There is one possible predictor. The Monte Carlo simulation design is as described in Section 6, and the sample size is  $T = 100$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao-Blackwellization. AIC, Laplace, and LASSO are the same without any resampling or with RB.

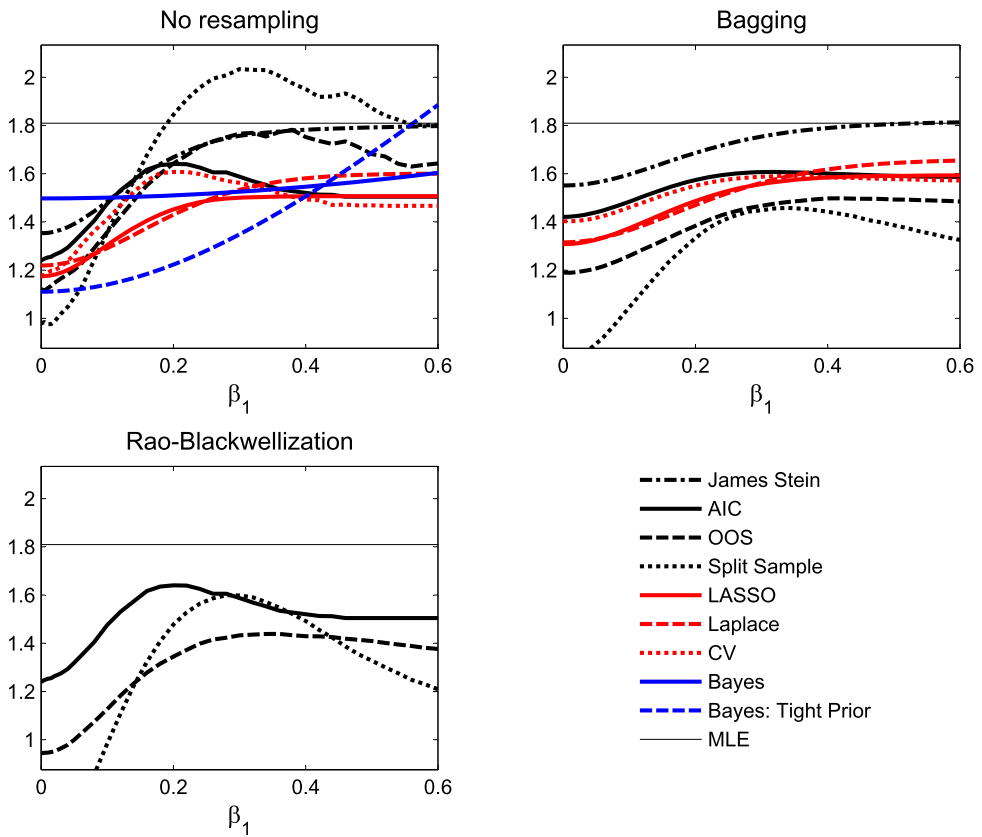


FIGURE A7.—Root normalized mean squared prediction errors ( $k = 3$ ). These are the simulated root normalized mean squared prediction errors using different procedures, where  $\beta = (\beta_1, 0, 0)'$ , plotted against  $\beta_1$ . The sample size is  $T = 100$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao-Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

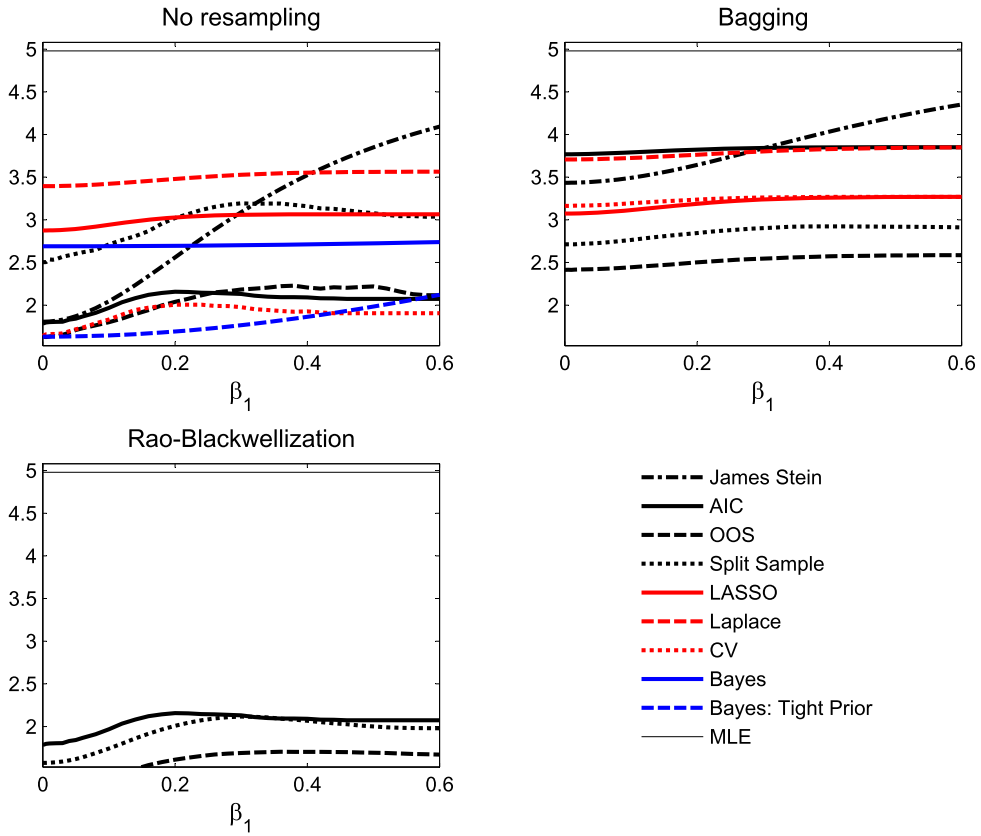


FIGURE A8.—Root normalized mean squared prediction errors ( $k = 20$ ). These are the simulated root normalized mean squared prediction errors using different procedures, where  $\beta = (\beta_1, 0, 0)'$ , plotted against  $\beta_1$ . The Monte Carlo simulation design is as described in Section 6, and the sample size is  $T = 100$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao-Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

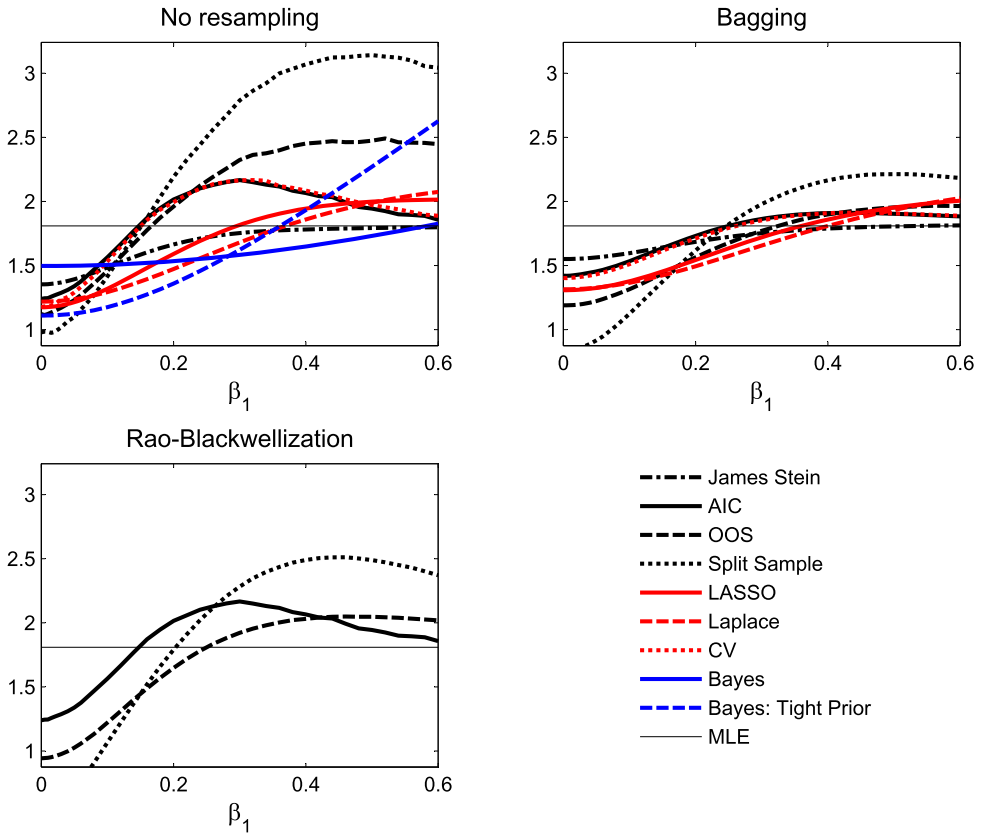


FIGURE A9.—Root normalized mean squared prediction errors ( $k = 3$ , all coefficients equal). These are the simulated root normalized mean squared prediction errors using different procedures, where  $\beta = \beta_1 k^{-1/2} i$ , plotted against  $\beta_1$ . The Monte Carlo simulation design is as described in Section 6, and the sample size is  $T = 100$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao-Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

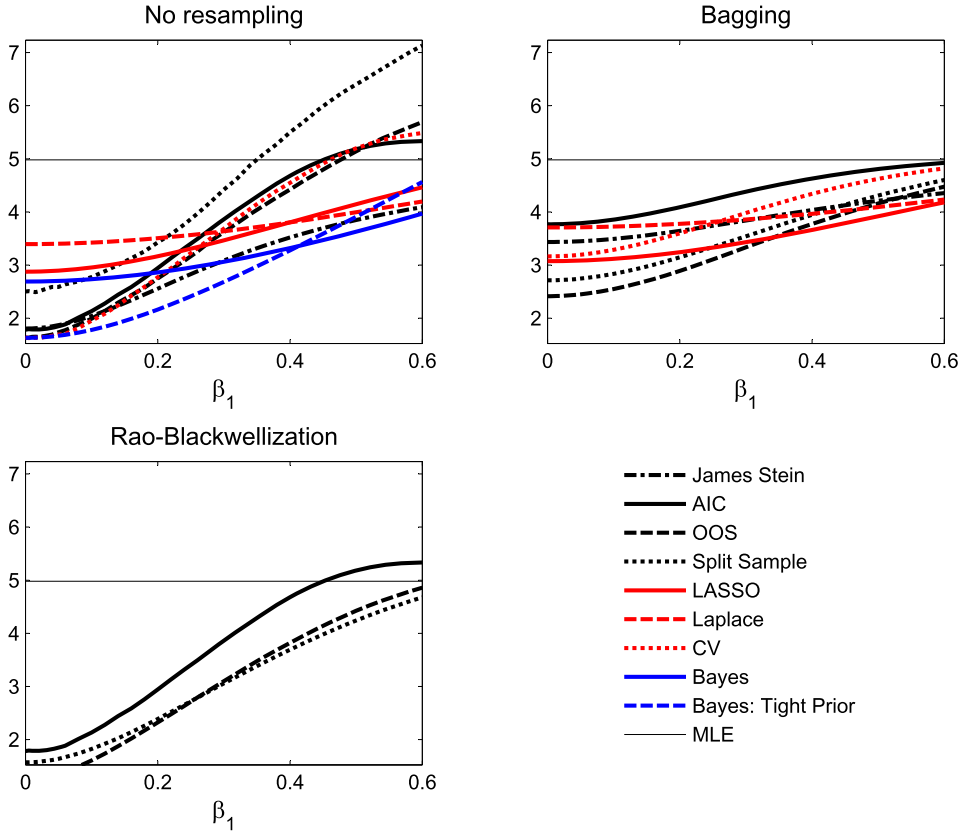


FIGURE A10.—Root normalized mean squared prediction errors ( $k = 20$ , all coefficients equal). These are the simulated root normalized mean squared prediction errors using different procedures, where  $\beta = \beta_1 k^{-1/2} i$ , plotted against  $\beta_1$ . The Monte Carlo simulation design is as described in Section 6, and the sample size is  $T = 100$ . Note that MLE and Bayes are the same without any resampling, with bagging or with Rao–Blackwellization. AIC, Laplace, LASSO, and JS are the same without any resampling or with RB.

TABLE A-III  
ROOT NORMALIZED MEAN SQUARED PREDICTION ERRORS WITH TIME-VARYING PARAMETERS<sup>a</sup>

| $\omega$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 |
|----------|------|------|------|------|------|
| AIC      | 0.76 | 1.07 | 1.55 | 2.02 | 2.51 |
| OOS      | 0.73 | 1.02 | 1.48 | 1.96 | 2.45 |
| SS       | 0.56 | 1.05 | 1.80 | 2.54 | 3.15 |
| AICB     | 0.81 | 1.01 | 1.46 | 1.95 | 2.45 |
| OOSB     | 0.70 | 0.95 | 1.45 | 1.99 | 2.52 |
| SSB      | 0.45 | 0.90 | 1.54 | 2.11 | 2.64 |
| OOSRB    | 0.59 | 0.93 | 1.48 | 2.02 | 2.55 |
| SSRB     | 0.30 | 0.93 | 1.64 | 2.23 | 2.73 |

<sup>a</sup>This table reports the simulated root normalized mean squared prediction errors using different procedures where there is a scalar parameter  $\beta_t$  that is a Gaussian random walk with innovation standard deviation  $\omega$ , the sample size is  $T = 100$ , and the procedures are applied treating the parameter as though it were constant.

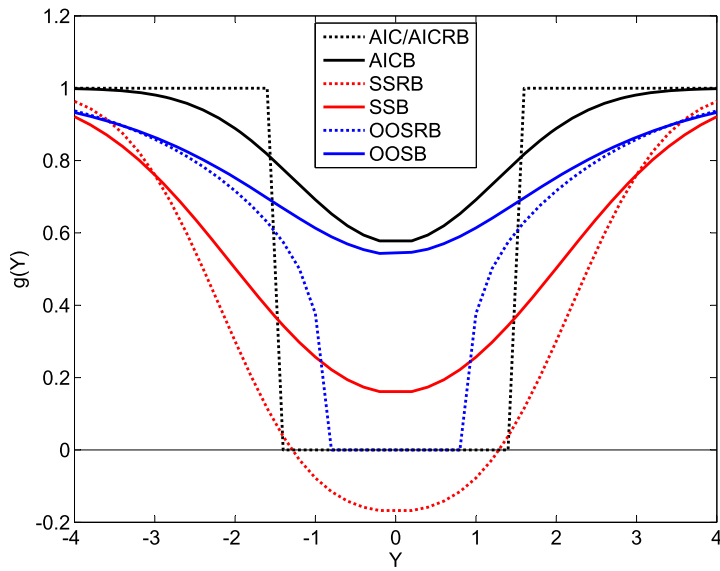


FIGURE A11.—Local asymptotic shrinkage functions. This plots the local asymptotic shrinkage functions for the case  $k = 1$ . Each estimator has a limit of the form  $Yg(Y)$  and this figure plots  $g(Y)$  against  $Y$ , as described in Appendix B. AIC is the same without any resampling or with RB.

## REFERENCES

- KNIGHT, K., AND W. FU (2000): "Asymptotics for LASSO-Type Estimators," *Annals of Statistics*, 28, 1356–1378. [2]
- MAGNUS, J. R. (2002): "Estimation of the Mean of a Univariate Normal Distribution With Known Variance," *Econometrics Journal*, 5, 225–236. [1]
- MAGNUS, J. R., O. POWELL, AND P. PRÜFER (2010): "A Comparison of Two Model Averaging Techniques With an Application to Growth Empirics," *Journal of Econometrics*, 154, 139–153. [1]
- SJÖSTRAND, K., L. H. CLEMMENSEN, R. LARSEN, AND B. ERSBØLL (2012): "SpaSM: A Matlab Toolbox for Sparse Statistical Modeling," Working paper, Technical University of Denmark. [3]
- TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267–288. [1]

*Dept. of Economics, Pennsylvania State University, 602 Kern Graduate Building, University Park, PA 16802, U.S.A.; [hirano@psu.edu](mailto:hirano@psu.edu)*

*and*

*Dept. of Economics, Johns Hopkins University, 3400 North Charles St., Baltimore, MD 21218, U.S.A.; [wrightj@jhu.edu](mailto:wrightj@jhu.edu).*

*Co-editor Elie Tamer handled this manuscript.*

*Manuscript received 13 April, 2015; final version accepted 28 November, 2016; available online 9 December, 2016.*