Bagging Binary and Quantile Predictors for Time Series: Further Issues

Tae-Hwy Lee*
Department of Economics
University of California, Riverside
Riverside, CA 92521-0427
Email: taelee@ucr.edu

Yang Yang
Wells Fargo Bank
420 Montgomery Street
San Francisco, CA 94104
Email: yang.yang@wellsfargo.com

First Version: April 15, 2006
Second Version: August 18, 2006
This Version: February 12, 2007

*Corresponding author. Tel: +1 (951) 827-1509. Fax: +1 (951) 827-5685.

ABSTRACT

Bagging (bootstrap aggregating) is a smoothing method to improve predictive ability under the presence of parameter estimation uncertainty and model uncertainty. In Lee and Yang (2006), we examined how (equal-weighted and BMA-weighted) bagging works for one-step ahead binary prediction with an asymmetric cost function for time series, where we considered simple cases with particular choices of a linlin tick loss function and an algorithm to estimate a linear quantile regression model. In the present paper, we examine how bagging predictors work with different aggregating (averaging) schemes, for multi-step forecast horizons, with a general class of tick loss functions, with different estimation algorithms, for nonlinear quantile regression models, and in different data frequencies. Bagging quantile predictors are constructed via (weighted) averaging over predictors trained on bootstrapped training samples, and bagging binary predictors are conducted via (majority) voting on predictors trained on the bootstrapped training samples. We find that median bagging and trimmed-mean bagging can alleviate the problem of extreme predictors from bootstrap samples and have better performance than equally-weighted bagging predictors; that bagging works more with longer forecast horizons; that bagging works well with highly nonlinear quantile regression models (e.g., artificial neural network), and with general tick loss functions. We also find that the performance of bagging may be affected by using different quantile estimation algorithms (in small sample, even if the estimation is consistent) and by using the different frequency of the time series data.

Keywords : Algorithm, Bagging, Median bagging, Binary prediction, Frequency, Majority voting, Multi-step prediction, Neural network, Quantile prediction, Time series.

JEL Classification : C3, C5, G0.

1 Introduction

To improve on unstable forecast, *b*ootstrap *agg*regat*ing* or bagging is introduced by Breiman (1996). In Lee and Yang (2006), we show how bagging, with equal-weight averaging and weighted averaging using Bayesian model averaging (BMA) method, works for one-step ahead binary prediction under an asymmetric cost function for time series. In that paper, we considered simple cases with particular choices of a loss function (linlin) and a regression model (linear).

We now consider the following extensions: (a) aggregating the bootstrap forecasts by other combination schemes as considered, e.g., by Stock and Watson (1999) and Timmermann (2007), (b) multi-step forecasts, (c) nonlinear models such as the neural network quantile model of White (1992), (d) different quantile estimation algorithms as discussed by Komunjer (2005), (e) a general class of the tick loss functions of Komunjer (2005) and Komunjer and Vuong (2005), and (f) using other macroeconomic and financial time series in various frequencies.

According to our experience in Monte Carlo and empirical experiments, some bootstrap predictors may generate extreme values that will seriously worsen the forecasts of equally weighted bagging predictors. To alleviate this problem of extreme forecasts, we consider alternative averaging schemes to generate bagging predictors (an idea borrowed from forecast combination literature). The first is the BMA-weighted bagging as used in Lee and Yang (2006). The second one is trimmed bagging, for which we remove extreme bootstrap forecasts in forming a bagging predictor. However, it will be very hard to decide which bootstrap predictors to keep and which to discard beforehand. In this paper, we simply trim a certain number of the largest and the smallest bootstrap predictors. We also use the median of the bootstrap predictors as our bagging predictor, which can be considered as an extreme case of trimmed bagging predictors. Hence we have the equal-weighted bagging, BMA-weighted bagging, trimmed-mean bagging, and median bagging. Our Monte Carlo and empirical experiments show that: when sample size is small and/or the predictors lies on the sparse parts of the density, median bagging and trimmed-mean bagging generally give better bagging forecasts than the equal-weighted bagging predictor (which is better than unbagged predictors); and when sample size is large and/or the predictor lies on the dense part of the data density, median bagging and trimmed bagging have no obvious advantage over the equal-weighted bagging (whose advantage over unbagged predictors is also weak in such a case).

We explore the performance of bagging predictors for multi-step forecast (for the conditional quantile) in this paper. As discussed by Brown and Mariano (1989) and Lin and Granger (1994), there are several ways to generate multi-step forecasts. These methods can be put into two groups: iteration of one-step ahead forecasts and direct multi-step forecasts. Among iterated multi-step forecasting methods, we can further classify them as the naive method, the exact method, the Monte Carlo method, and the bootstrap method. If the true forecast model is linear and known, all these methods should give same predictions. However, if the true forecast model is non-linear or unknown, different multi-step forecasting methods give quite different predictions. We use the direct multi-step forecast method for the conditional quantile prediction in our Monte Carlo experiments. It is found that, compared with unbagged predictors, the performance of a bagging predictor tends to get better with longer forecast horizons.

Lee and Yang (2006) attributed a part of success of the bagging predictors to the small sample estimation uncertainties. Therefore, a question that may arise is that whether the good performance of bagging predictors critically depends on algorithms we employ in our estimation. Lee and Yang (2006) used the interior point algorithm for quantile estimations as suggested by Portnoy and Koenker (1997). To examine how other algorithms may work for the bagging, we also use the minimax algorithm of Komunjer (2005) in this paper. The interior point algorithm for quantile estimation can be used for a linear quantile regression model under the standard linlin tick loss function while the minimax algorithm allows flexible function forms for quantile regressions such as a neural network model.

We use the minimax algorithm to estimate linear and nonlinear quantile regression model under a general class of tick functions, namely, the tick-exponential family defined by Komunjer (2005). Our simulation results show that the bagging works (i.e., better than the unbagged predictors) for quantiles almost equally well for the different tick functions in the tick-exponential family in small samples. Komunjer (2005) shows that QMLE under the tick-exponential family is consistent.

With the flexibility provided by the minimax algorithm, we check the performance of bagging predictors on highly non-linear quantile regression models – artificial neural network models. When the sample size is limited, it is usually hard to choose the number of hidden nodes and the number of inputs (lags), and to estimate the large number of parameters in neural network model. Therefore, a neural network model generate poor predictions with a small sample. In such cases, the bagging can do a wonderful job to improve the forecasting performance as shown later in our empirical experiments.

We finally investigate whether the performance of bagging can be affected by the frequency of the data.

The plan of this paper is as follows. Section 2 gives a brief introduction to bagging predictors. Section 3 explains different ways to aggregating bootstrap predictors. In Section 4, we examine how bagging works for the multi-step predictions of the conditional quantiles. In Section 5 we examine how the bagging works for quantile prediction under the different tick loss functions of the tick-exponential family. In Section 6, we consider whether the performance of bagging predictor will be affected by different estimation algorithms. In Section 7 we examine the bagging predictors on the (nonlinear) neural network quantile regression models. Section 8 examines the effect of the different data frequencies on the bagging performance. In Section 9 we discuss a potential extension with pretesting for bagging. Section 10 provides a brief field guide to bagging based on what we have learned in this paper. Section 11 concludes.

2 What is Bagging?

Bagging predictor is a combined predictor formed over a set of training sets to smooth out the "instability" caused by parameter estimation uncertainty and model uncertainty. A predictor is said to be "unstable" if a small change in the training set will lead to a significant change in the predictor (Breiman, 1996). In this section, we will show how bagging predictor may improve the predicting performance of its underlying predictor. Let

$$\mathcal{D}_t \equiv \{(Y_s, \mathbf{X}_{s-1})\}_{s=t-R+1}^{t} \qquad (t = R, \ldots, T)$$

be a training set at time t and let $\varphi(\mathbf{X}_t, \mathcal{D}_t)$ be a forecast of Y_{t+1} or of the binary variable $G_{t+1} \equiv \mathbf{1}(Y_{t+1} \geq 0)$ using this training set $\mathcal{D}_t$ and the explanatory variable vector $\mathbf{X}_t$. The optimal forecast $\varphi(\mathbf{X}_t, \mathcal{D}_t)$ for Y_{t+1} will be the conditional mean of Y_{t+1} given $\mathbf{X}_t$ if we have the squared error loss function, or the conditional quantile of Y_{t+1} on $\mathbf{X}_t$ if the loss is a tick function. Below we also consider the binary forecast for $G_{t+1} \equiv \mathbf{1}(Y_{t+1} \geq 0)$.

Suppose each training set $\mathcal{D}_t$ consists of R observations generated from the underlying probability distribution $\mathbf{P}$. The forecast $\{\varphi(\mathbf{X}_t, \mathcal{D}_t)\}_{t=R}^{T}$ can be improved if more training sets were able to be generated from $\mathbf{P}$ and the forecast can be formed from averaging the multiple forecasts obtained from the multiple training sets. Ideally, if $\mathbf{P}$ were known and multiple training sets $\mathcal{D}_t^{(j)}$ $(j = 1, \ldots, J)$ may be drawn from $\mathbf{P}$, an ensemble aggregating predictor $\varphi_A(\mathbf{X}_t)$ can be constructed by the weighted averaging of $\varphi(\mathbf{X}_t, \mathcal{D}_t^{(j)})$ over j, i.e.,

$$\varphi_A(\mathbf{X}_t) \equiv \mathbb{E}_{\mathcal{D}_t}\varphi(\mathbf{X}_t,\mathcal{D}_t) \equiv \sum_{j=1}^{J} w_{j,t}\varphi(\mathbf{X}_t, \mathcal{D}_t^{(j)}), \tag{1}$$

where $\mathbb{E}_{\mathcal{D}_t}(\cdot)$ denotes the expectation over $\mathbf{P}$, $w_{j,t}$ is the weight function with $\sum_{j=1}^{J} w_{j,t} = 1$, and the subscript A in φ_A denotes "aggregation".

Lee and Yang (2006, Propositions 1 and 4) show that the ensemble aggregating predictor $\varphi_A(X_t)$ has no larger expected loss than the original predictor $\varphi(X_t, \mathcal{D}_t)$. For any convex loss function $c(\cdot)$ on the forecast error z_{t+1}, we will have

$$\mathbb{E}_{\mathcal{D}_t,Y_{t+1},\mathbf{X}_t} c(z_{t+1}) \geq \mathbb{E}_{Y_{t+1},\mathbf{X}_t} c(\mathbb{E}_{\mathcal{D}_t}(z_{t+1})),$$

where $\mathbb{E}_{\mathcal{D}_t}(z_{t+1})$ is the aggregating forecast error, and $\mathbb{E}_{\mathcal{D}_t,Y_{t+1},\mathbf{X}_t}(\cdot) \equiv \mathbb{E}_{\mathbf{X}_t}[\mathbb{E}_{Y_{t+1}|\mathbf{X}_t}\{\mathbb{E}_{\mathcal{D}_t}(\cdot)\,|X_t\}]$ denotes the expectation $\mathbb{E}_{\mathcal{D}_t}(\cdot)$ taken over $\mathbf{P}$ (i.e., averaging over the multiple training sets generated from $\mathbf{P}$), then taking an expectation of Y_{t+1} conditioning on X_t, and then taking an expectation of X_t. Similarly we define the notation $\mathbb{E}_{Y_{t+1},\mathbf{X}_t}(\cdot) \equiv \mathbb{E}_{\mathbf{X}_t}[\mathbb{E}_{Y_{t+1}|\mathbf{X}_t}(\cdot)\,|X_t]$. Therefore, the aggregating predictor will always have no larger expected cost than the original predictor for a convex loss function $\varphi(X_t, D_t)$. The examples of the convex loss function includes the squared error loss and a tick (or check) loss of Koenker and Basset (1978)

$$\rho_\alpha(z) \equiv [\alpha - \mathbf{1}(z < 0)]z. \tag{2}$$

How much this aggregating predictor can improve depends on the distance between $\mathbb{E}_{\mathcal{D}_t,Y_{t+1},\mathbf{X}_t} c(z_{t+1})$ and $\mathbb{E}_{Y_{t+1},\mathbf{X}_t} c(\mathbb{E}_{\mathcal{D}_t}(z_{t+1}))$. We can define this distance by $\Delta \equiv \mathbb{E}_{\mathcal{D}_t,Y_{t+1},\mathbf{X}_t} c(z_{t+1}) - \mathbb{E}_{Y_{t+1},\mathbf{X}_t} c(\mathbb{E}_{\mathcal{D}_t}(z_{t+1}))$. Therefore, the effectiveness of the aggregating predictor depends on the *convexity* of cost function. The more convex is the cost function, the more effective this aggregating predictor can be. We will see the effect of the convexity on the performance of bagging later in this paper (Section 6). If the loss function is the squared error loss, then it can be shown that $\Delta = \mathbb{V}_{\mathcal{D}_t}\left[\varphi(\mathbf{X}_t, \mathcal{D}_t)\right]$ is the variance of the predictor, which measures the "instability" of the predictor. See Lee and Yang (2006, Proposition 1) and Breiman (1996). If the loss is the tick function, the effectiveness of bagging is also different for different quantile predictions: bagging works better for tail-quantile predictions than for mid-quantile predictions.

In practice, however, $\mathbf{P}$ is not known. In that case we may estimate $\mathbf{P}$ by its empirical distribution, $\hat{\mathbf{P}}(\mathcal{D}_t)$, for a given $\mathcal{D}_t$. Then, from the empirical distribution $\hat{\mathbf{P}}(\mathcal{D}_t)$, multiple training sets may be drawn by the bootstrap method. Bagging predictors, $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$, can then be computed by taking weighted average of the predictors trained over a set of bootstrap training sets. More specifically, the bagging predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be obtained in the following steps:

1. Given a training set of data at time t, $\mathcal{D}_t \equiv \{(Y_s, \mathbf{X}_{s-1})\}_{s=t-R+1}^{t}$, construct the jth bootstrap sample $\mathcal{D}_t^{*(j)} \equiv \{(Y_s^{*(j)}, \mathbf{X}_{s-1}^{*(j)})\}_{s=t-R+1}^{t}$, $j = 1, \ldots, J$, according to the empirical distribution of $\hat{\mathbf{P}}(\mathcal{D}_t)$ of $\mathcal{D}_t$.

2. Train the model (estimate parameters) from the jth bootstrapped sample $\mathcal{D}_t^{*(j)}$.

3. Compute the bootstrap predictor $\varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)})$ from the jth bootstrapped sample $\mathcal{D}_t^{*(j)}$.

4. Finally, for mean and quantile forecast, the bagging predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be constructed by averaging over J bootstrap predictors

$$\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*) \equiv \sum_{j=1}^{J} \hat{w}_{j,t} \varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)});$$

   and for binary forecast, the bagging binary predictor $\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*)$ can be constructed by majority voting over J bootstrap predictors:

$$\varphi^B(\mathbf{X}_t, \mathcal{D}_t^*) \equiv \mathbf{1}\left(\sum_{j=1}^{J} \hat{w}_{j,t} \varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)}) > 1/2\right)$$

   with $\sum_{j=1}^{J} \hat{w}_{j,t} = 1$ in both cases.

One concern of applying bagging to time series is whether a bootstrap can provide a sound simulation sample for dependent data, for which the bootstrap is required to be consistent. It has been shown that some bootstrap procedure (such as moving block bootstrap) can provide consistent densities for moment estimators and quantile estimators. See, e.g., Fitzenberger (1997).

3 Bagging with Different Averaging Schemes

There are several ways to generate the averaging weight $\hat{w}_{j,t}$ for bagging predictors introduced in the previous section. The most commonly used one is equal-weighting across all bootstrap samples, that is, $\hat{w}_{j,t} = 1/J$, $j = 1, \ldots, J$. However, one problem with equal weighted bagging is that some bootstrap samples could (and do) make extreme forecasts. Possible sources of these extreme forecasts include random procedures of generating bootstrap samples (especially from small samples), difficulties arising from multiple local optima for the nonlinear models, estimation difficulties for non-differentiable loss functions. In these cases, we may get some erratic values for the predictive parameter $\hat{\boldsymbol{\beta}}_t^{*(j)}(\mathcal{D}_t^{*(j)})$, and hence "crazy" bootstrap predictors $\varphi^{*(j)}(\mathbf{X}_t, \mathcal{D}_t^{*(j)})$. The extreme forecasts may happen more frequently for the conditional quantile predictions than for the conditional mean predictions. The effect may be large that such crazy bootstrap sample predictors may deteriorate performance of bagging predictors. By finding a way to alleviate or eliminate the effect of such crazy bootstrap predictors, we may improve the bagging predictors.

We consider several ways to solve these extreme forecasts problems. One is to estimate the combination weight based on in-sample performance of each predictor, for example, using Bayesian model averaging (BMA) weighting. By setting

$$\hat{w}_{j,t} \equiv \Pr\left[\hat{\boldsymbol{\beta}}_\alpha(\mathcal{D}_t^{*(j)})|\mathcal{D}_t\right], \qquad j = 1, \ldots, J,$$

a bootstrap predictor with better in-sample performance will be assigned a larger weight. Extreme-valued predictors are generated when parameters in the forecasting model are poorly estimated for bootstrap samples, in which case it is expected that the in-sample performance of the bootstrap estimators will not be good either. Therefore, by assigning the weights according to the in-sample performance, the BMA bagging predictors can alleviate the extreme-valued predictor problem to a certain extent. However, the BMA bagging predictors still put some positive weight on the extreme value predictors and could not completely eliminate the effect of these crazy forecasts.

Another way to deal with these extreme value predictors is to sort all the bootstrap predictors and trim a certain number of bootstrap predictors from both tails before the averaging procedure. This procedure will be called the trimmed bagging. The user can decide the number of bootstrap predictors to trim depending on the seriousness of the extreme value predictors problem. However, it is hard to decide a priori, and thus in our Monte Carlo and empirical analysis, we choose to trim a fixed number (e.g., 5 and 10) of bootstrap predictors on each tail of the sorted bootstrap predictors without checking whether they are extreme or not.

Alternatively, we can simply use the median of bootstrap predictors (instead of the mean or trimmed mean of the bootstrap predictors), which is the extreme case of the trimmed bagging by using only the middle one or two bootstrap predictors. In the median bagging, we can avoid the arbitrary choice of how many bootstrap predictors are to be discarded in the trimmed bagging predictor.

We use a set of Monte Carlo simulation to gain further insights on how these different bootstrap aggregating weight schemes work. For quantile predictions, we obtain the out-of-sample mean loss values for the unbagged predictors with $J = 1$ (S_1) and for the bagging predictors with $J = 50$ (S_a, $a \geq 2$). We consider nine quantile levels with left tail probability $\alpha = 0.01,\ 0.05,\ 0.1, 0.3, 0.5, 0.7,\ 0.9,\ 0.95,$ and 0.99. It will be said that bagging "works" if $S_1 > S_a$. To rule out the chance of pure luck by a certain criterion, we compute the following three summary performance statistics from 100 Monte Carlo replications ($r = 1, \ldots, 100$):

$$
\begin{aligned}
T_{1,a} &\equiv \frac{1}{100}\sum_{r=1}^{100} S_a^r, \\
T_{2,a} &\equiv \left(\frac{1}{100}\sum_{r=1}^{100}(S_a^r - T_{1,a})^2\right)^{1/2}, \\
T_{3,a} &\equiv \frac{1}{100}\sum_{r=1}^{100} \mathbf{1}(S_1^r > S_a^r),
\end{aligned}
$$

where $a = 1$ for the non-bagged predictor ($J = 1$), and $a \geq 2$ for various bagging predictors with different weighting (equally-weighted mean bagging, BMA bagging, median bagging, and trimmed-mean bagging). T_1 measures the Monte Carlo mean of the out-of-sample mean loss, T_2 measures the Monte Carlo standard deviation of the out-of-sample mean loss, T_3 measures the Monte Carlo frequency that bagging works. We present T_1, T_2 and T_3 in Table

1 (Panels A-F). To make the comparison of the bagging predictors and unbagged predictors easier, we also report in figures (Figure 1, Panels A-F) to show two *relative* performance statistics: $T_{1,a}/T_{1,1}$ and $T_{2,a}/T_{2,1}$. For both of them, a value smaller than 1 indicates bagging predictors work better than the unbagged predictor.

We generate the data from

$$\begin{aligned} Y_t &= \rho Y_{t-1} + \varepsilon_t, \\ \varepsilon_t &= z_t \left[(1-\theta) + \theta \varepsilon_{t-1}^2\right]^{1/2} \\ z_t &\sim \text{i.i.d. } MW_i \end{aligned} \tag{3}$$

where the i.i.d. innovation z_t is generated from the first eight mixture normal distributions of Marron and Wand (1992, p. 717), each of which will be denoted as MW_i $(i = 1, \ldots, 8)$.[1] In Table 1-Panel A and Figure 1-Panel A, we consider the data generating processes for ARCH-MW_1 with $\theta = 0.5$ (and $\rho = 0$), while in Table 1-Panels B-F and Figure 1-Panels B-F, we consider the data generating processes for AR-MW_i $(i = 1, \ldots, 5)$ with $\rho = 0.6$ (and $\theta = 0$). Therefore, our data generating processes fall into two categories: the (mean-unpredictable) martingale-difference ARCH(1) processes without AR structure and the mean-predictable AR(1) processes without ARCH structure.

For each series, 100 extra series is generated and then discarded to alleviate the effect of the starting values in random number generation. We consider one fixed out-of-sample size $P = 100$ and a range of estimation sample sizes $R = 200$ and 500. Our bagging predictors are generated by averaging over $J = 50$ bootstrap predictors.

We consider a group of simple univariate polynomial quantile regression function of Chernozhukov and Umantsev (2001) as our predictive methods:

$$Q_\alpha(Y_{t+h}|\mathbf{X}_t) = \tilde{\mathbf{X}}_t' \boldsymbol{\beta}_{\alpha,h}, \tag{4}$$

with h representing the forecast horizons, $\mathbf{X}_t = (Y_t \ \ldots \ Y_{t+h-1})$, $\tilde{\mathbf{X}}_t = (1 \ Y_t \ Y_t^2 \ \ldots \ Y_{t-h+1} \ Y_{t-h+1}^2)'$, and $\boldsymbol{\beta}_{\alpha,h} = [\boldsymbol{\beta}_{\alpha,h,0} \ \boldsymbol{\beta}_{\alpha,h,1} \ \boldsymbol{\beta}_{\alpha,h,2} \ \ldots \ \boldsymbol{\beta}_{\alpha,h,2h-1} \ \boldsymbol{\beta}_{\alpha,h,2h}]'$. For now we set $h = 1$ to generate one-step ahead forecast, and we will talk about multi-step forecast later in this paper.

[1] MW_1 is Gaussian, MW_2 is Skewed unimodal, MW_3 Strongly skewed, MW_4 Kurtoic unimodal, MW_5 Outlier, MW_6 Bimodal, MW_7 Separated bimodal, and MW_8 is Skewed bimodal. See Marron and Wand (1992, p. 717). To save space we report only for MW_i $(i = 1, \ldots, 5)$ in each panel of Table 1 and Figure 1. The other four results for $i = 5, \ldots, 8$ are basically similar in the pattern how the bagging works, and are available upon request.

We estimate $\boldsymbol{\beta}_{\alpha,h}$ recursively using the "rolling" samples of size $R - 2h + 1$. Suppose there are T ($\equiv R + P$) observations in total. We use the most recent $R - 2h + 1$ observations available at time t, $R \leq t < T - h$, as a training sample $\mathcal{D}_t \equiv \{(Y_s, \mathbf{X}_{s-h})\}_{s=t-R+2h}^{t}$. We then generate P ($= T - R$) h-step-ahead forecasts for the remaining forecast validation sample. For each time t in the P prediction periods, we use a rolling training sample $\mathcal{D}_t$ of size $R - 2h + 1$ to estimate model parameters:

$$\hat{\boldsymbol{\beta}}_{\alpha,h}(\mathcal{D}_t) \equiv \arg\min_{\boldsymbol{\beta}_{\alpha,h}} \sum_{s=t-R+h+1}^{t} \rho_\alpha(u_s), \qquad t = R, \ldots, T, \tag{5}$$

where $u_s \equiv Y_s - Q_\alpha(Y_s|\mathbf{X}_{s-h}) = Y_s - \tilde{\mathbf{X}}'_{s-h}\boldsymbol{\beta}_{\alpha,h}$. $\hat{\boldsymbol{\beta}}_{\alpha,h}(\mathcal{D}_t)$ is estimated using the interior-point algorithm suggested by Portnoy and Koenker (1997).

To generate bootstrap samples, we use the block bootstrap for both Monte Carlo experiments and empirical applications. We choose the block size that minimizes the in-sample average cost recursively and therefore we use a different block size at each forecasting time t and for each loss function with different α's.

TABLE 1 ABOUT HERE

The Monte Carlo results are reported in Table 1-Panels A-F and Figure 1-Panels A-F, where *mean*, BMA_k, *med*, $trim_k$ denote the equal weighted bagging predictors, BMA-weighted bagging predictors using k-most recent in-sample observations, median weighted bagging predictors, and k-trimmed on each tail weighted bagging predictors.

According to our Monte Carlo results on quantile predictions shown in Table 1-Panel A-F, we summarize our observations as follows. First, in most of cases, BMA weighted bagging predictors, median bagging predictors and trimmed bagging predictors have better predicting performance (smaller T_1 and T_2 and larger T_3) compared to the mean bagging predictor even when we have a relatively large sample size. Second, on average, the improvement brought by the median bagging is larger than the trimmed bagging predictors and the BMA weighted bagging predictors, so median bagging tends to give the smallest T_1 and largest T_3 among all predictors. Third, the outstanding performance of the median bagging predictors is most obvious when α values are close to 0 or 1, where the extreme value problem are most serious because there are fewer observations on tails and the parameters regression estimators are sensitive to the estimation sample. When sample size R is 200 and α values are close to 0 or

1, the median bagging predictors can further reduce the average loss (T_1) by about 1% on average and increase the percentage of bagging works (T_3) by about 4% average compared with the mean bagging predictors. However, the advantage of median bagging predictors are not so clear when α values are close to 0.5.

FIGURE 1 ABOUT HERE

From Figure 1-Panels A-E, we can see that different bagging predictors work in similar trends. First, bagging predictors work better when the sample size is smaller, so the $R = 200$ lines lie below the $R = 500$ lines in the figures, and both $R = 200$ and $R = 500$ lie below the unit line most of the time. Second, bagging predictors works better when α values are close to 0 or 1, so the bagging lines look like the letter "n" especially when $R = 200$. Third, bagging predictors work better when α-quantiles lie on the sparse part of the error distribution. Our explanation is that for the sparse part of the error distribution, there are fewer observations, therefore quantile predictions are sensitive to the estimation sample and bagging predictors work better for unstable predictions. For example, when the error term are left skewed as in Figure 1-Panel C, bagging predictors give larger loss reduction for the prediction of small α-quantiles than for large α-quantiles; when the error term are right skewed as in Figure 1-Panel D, bagging predictors give large loss reduction for the prediction of larger α-quantiles but do not work for small α-quantiles; and among Figure 1-Panels A-F, Panel F has the sparsest distribution on both tails among all DGPs and bagging predictors give best performance (the smallest T_1 and T_2 and largest T_3).

Our conclusions on the performance of BMA bagging predictors and median bagging predictors are further testified by empirical experiments. We make pseudo true real time forecast of the daily returns of six major U.S. stock indices and two major foreign exchange rates. We split the series into two parts: one for in-sample estimation with the size $R = 100$ and 300, and another for out-of-sample forecast validation with sample size $P = 250$ (fixed for both R's). We choose the most recent $P = 250$ days in the sample as a out-of-sample validation sample. We use a rolling-sample scheme, that is, the first forecast is based on observations $T - P - R + 1$ through $T - P$, the second forecast is based on observations $T - P - R + 2$ through $T - P + 1$, and so on. The eight series are Dow Jones Industrial Averages (Dow Jones), New York Stock Exchange Composite (NYSE), Standard and Poor's 500 (SP500), National Association of Securities Dealers Automated Quotations Composite

(NASDAQ), Russell 2000 index (Rusell2000), Pacific Exchange Technology (PET), US dollar per Euro (USD/EUR), and US dollar per Japanese Yen (USD/JPY). The total sample period and the out-of-sample forecasting period are summarized as follows:

	Total sample period	Out-of-sample period ($P = 250$)
Dow Jones	10/27/1998 ∼ 12/31/2000	01/05/2000 ∼ 12/31/2000
NYSE	10/27/1998 ∼ 12/31/2000	01/05/2000 ∼ 12/31/2000
SP500	10/27/1998 ∼ 12/31/2000	01/05/2000 ∼ 12/31/2000
NASDAQ	10/27/1998 ∼ 12/31/2000	01/05/2000 ∼ 12/31/2000
Rusell2000	10/27/1998 ∼ 12/31/2000	01/05/2000 ∼ 12/31/2000
PET	10/27/1998 ∼ 12/31/2000	01/05/2000 ∼ 12/31/2000
USD/EUR	10/10/2003 ∼ 04/11/2005	08/05/2004 ∼ 04/11/2005
USD/YEN	10/10/2003 ∼ 04/11/2005	08/05/2004 ∼ 04/11/2005

TABLE 2 ABOUT HERE

FIGURE 2 ABOUT HERE

We consider nine quantile parameter $\alpha = 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95$ and 0.99. The empirical experiments are reported in Table 2-Panels A-H and Figure 2 Panels A-H. Our findings are as follows. First, bagging predictors works better when the sample size is smaller. Second, bagging predictors work better for α values close to 0 or 1 than for α values close to 0.5. Third, for the six indices on stock returns, bagging predictors work better for α values close to 0 than α values close to 1 because the distribution of stock returns all have long left tails. However, for the two foreign exchange series, bagging works rather symmetrically for α values close to 0 and 1 because they have symmetric distributions.

4 Bagging Multi-step Quantile Forecasts

We will show how bagging works for multi-step predictions in this section. It is important to make multi-step forecasts in the real world. A group of users for time series predictions are policy makers. Since it takes a long time for monetary policies and fiscal policies to generate expected effect in economy, policy makers have to produce predictions more than one period ahead.

We check four multi-step horizons, $h = 1, 2, 3,$ and 4. If we have a simple linear model, then multi-step forecasts can be achieved by simple iteration of the one-step ahead predictors. However, we may not apply this naive iteration method to generate multi-step forecasts for non-linear models. We use polynomial quantile regression models to take account of the non-linear structures in the data. As mentioned by Tsay (1993), Lin and Tsay (1996) and Chevillon and Hendry (2004), "direct" multi-step method will suffer less from the model misspecification than the "iterated" multi-step methods, therefore the direct multi-step method is also called "adaptive estimation", and should be able to generate predictions much better or at lease as good as iterated methods in case of model uncertainty or misspecification.

There are few literature discussing how to make multi-step conditional quantile forecasts. We can either iterate one-step ahead forecast or model the relationship between $Q_\alpha(Y_{t+h}|\mathbf{X}_t)$ and $\mathbf{X}_t$ directly. To apply the iterated method for multi-step quantile, we need to set up a quantile regression model based on the lags of quantile itself, for example, CaViaR model of Engle and Manganelli (2004):

$$Q_\alpha(Y_{t+1}|\mathbf{X}_t) = b_0 + b_1 Q_\alpha(Y_t|\mathbf{X}_{t-1}) + e_{t+1}. \tag{6}$$

Even with CaViaR model, we can only use the "naive" iteration to get the multi-step quantile forecast. The naive iterated multistep quantile forecasts may generate poor forecasts. To be comparable with the results from other part of this paper, we model the relationship between $Q_\alpha(Y_{t+h}|\mathbf{X}_t)$ and $\mathbf{X}_t$ directly using the polynomial quantile regression model as discussed in equation (4) in Section 2. We use the same DGP as discuss in equation (3) for our Monte Carlo experiments. We will only make "direct" multi-step quantile forecast.

TABLE 3 ABOUT HERE

FIGURE 3 ABOUT HERE

According to our Monte Carlo results for quantile forecast reported in Table 3 and Figure 3, we find: loss level (T_1) for both unbagged predictors and bagging predictors increases as forecast horizon increase; the frequencies that bagging predictors out-perform unbagged predictors (T_3) also increase with the forecast horizons;. the relative average loss of bagging predictors compared to unbagged predictors ($T_{1,a}/T_{1,1}$) and the relative standard error of loss for the bagging predictors compared to unbagged predictors ($T_{2,a}/T_{2,1}$) decreases with the forecast horizon increases.

5 Bagging Quantile Forecasts with Different Tick Losses

Komunjer (2005) introduced a tick-exponential family defined by:

$$\begin{aligned} & \varphi_{t+h}^{\alpha}(Y_{t+h}, Q_\alpha(Y_{t+h}|\mathbf{X}_t)) \\ = \ & \exp(-(1-\alpha)\left[a_t\left(Q_\alpha(Y_{t+h}|\mathbf{X}_t)\right) - b_t\left(Y_{t+h}\right)\right]\mathbf{1}\left\{Y_{t+h} \leq Q_\alpha(Y_{t+h}|\mathbf{X}_t)\right\} \\ & +\alpha\left[a_t\left(Q_\alpha(Y_{t+h}|\mathbf{X}_t)\right) - c_t\left(Y_{t+h}\right)\right]\mathbf{1}\left\{Y_{t+h} > Q_\alpha(Y_{t+h}|\mathbf{X}_t)\right\}), \end{aligned} \tag{7}$$

where (i) a_t is continuously differentiable function, b_t and c_t are $\mathcal{F}_t$-measurable functions; (ii) φ_{t+h}^{α} is a probability density; (iii) $Q_\alpha(Y_{t+h}|\mathbf{X}_t)$ is the α-quantile of $Y_{t+h}|\mathbf{X}_t$.

A class of quasi-maximum likelihood estimators (QMLEs), $\hat{\boldsymbol{\beta}}_{\alpha,h}(\mathcal{D}_t)$, can be obtained by solving

$$\boldsymbol{\beta}_{\alpha,h}(\mathcal{D}_t) = \arg\max_{\boldsymbol{\beta}_{\alpha,h}} R^{-1} \sum_{s=t-R+h+1}^{t} \ln \varphi_s^{\alpha}(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})). \tag{8}$$

If $a_t\left(Q_\alpha(Y_{t+h}|\mathbf{X}_t)\right) = Q_\alpha(Y_{t+h}|\mathbf{X}_t)$, and $b_t\left(Y_{t+h}\right) = c_t\left(Y_{t+h}\right) = Y_{t+h}$, the maximizing problem in equation (8) is equivalent to the minimization problem of Koenker and Bassett (1978) as shown in equation (2). We also try another group of exponential tick family loss functions introduced by Komunjer (2005) by setting

$$a_t\left(\eta\right) = b_t\left(\eta\right) = c_t\left(\eta\right) = \frac{1}{\alpha\left(1-\alpha\right)} \operatorname{sgn}\left(\eta\right) \ln\left(1+|\eta|^p\right), \tag{9}$$

where $\operatorname{sgn}(\eta) \equiv \mathbf{1}\left\{\eta \geq 0\right\} - \mathbf{1}\left\{\eta < 0\right\}$.

TABLE 4 ABOUT HERE

Our empirical results of quantile prediction for S&P500 daily return during 01/13/2004 $\sim$ 01/07/2005 ($P = 250$) using the rolling estimation samples with $R = 100$ and 300 (10/31/2002 $\sim$ 01/07/2005) is shown in Table 4, where *tick* denotes $\boldsymbol{\beta}_{\alpha,h}(\mathcal{D}_t)$ is estimated using equation (2), $p = 1$, 2, and 3 denote $\boldsymbol{\beta}_{\alpha,h}(\mathcal{D}_t)$ is estimated using equation (9). We can see from the table that no matter which tick losses we use, bagging predictors have lower quantile cost when α is small, and have no obvious advantage over the unbagged predictors when α is large.

6 Bagging Quantile Forecasts with Different Estimation Algorithms

If we want to forecast the conditional mean, usually it is not a big problem to estimate the parameters for linear models and most of non-linear models. However, for quantile forecast, since the quantile loss function is not differentiable, it is very hard to estimate model parameters, especially when we use non-linear quantile regression models. The algorithms that can be used for the quantile estimation has been reviewed by Buchinsky (1998), Koenker and Park (1996), Frenk *et al.* (1994), Chernozhukov and Hong (2003), and Komunjer (2005) for both linear quantile models and nonlinear quantile models. We compare two different algorithms for quantile estimation in this paper in terms of the bagging. The two algorithms are the interior point algorithm introduced by Portnoy and Koenker (1997) and the minimax algorithm introduced by Komunjer (2005).

Portnoy and Koenker (1997) propose a statistical preprocessing for general quantile regression problems and combine it with "interior point" methods for solving linear programs. The following is a brief explanation on how to apply the interior algorithm for quantile estimation. If we put all the error term u_s for $s = t-R+h+1, t-R+h+2, \ldots, t$ in equation (5) into positive numbers, the quantile estimation problem can be rewritten as

$$\hat{\boldsymbol{\beta}}_{\alpha,h}(\mathcal{D}_t) = \arg\min_{\boldsymbol{\beta}_{\alpha,h}} (\alpha u^+ + (1-\alpha)u^- | Y_s = \tilde{\mathbf{X}}'_{s-h}\boldsymbol{\beta}_{\alpha,h} + u_s), \tag{10}$$

where u^+ a $R-h$-vector of positive errors or zeros, u^- a $R-h$-vector of absolute value of negative errors or zeros. Portnoy and Koenker (1997) show that the optimization program (10) can be rewritten into the following dual formulations

$$\omega = \arg\max_{\omega} \left(\sum\nolimits_{s=t-R+h+1}^{t} Y_s\omega_s | \sum\nolimits_{s=t-R+h+1}^{t} \tilde{\mathbf{X}}_{s-h}\omega_s = 0,\ \omega_s \in [-1,1] \right), \tag{11}$$

where $\omega_s = 1$ if $u_s > 0$, $\omega_s = -1$ if $u_s < 0$, $-1 < \omega_s < 1$ if $u_s = 0$; and ω_s is like Lagrange multipliers on the constraints, or marginal costs of relaxing the constraints. The optimization problem in (11) is the standard formulations of interior point methods for linear programs with bounded variables.

The interior point algorithm is easy to apply, runs fast and is embodied in most popular computer software, for example, GAUSS and MATLAB. However, the interior point algorithm can only be used for linear quantile models with the tick loss function. If we have non-linear quantile regression models or use the tick-exponential family introduced by Komunjer (2005) for the quantile estimation, we have to choose another algorithm for parameter estimation.

Komunjer (2005) introduces a new quantile regression algorithm – minimax algorithm, which is a more flexible method than the interior point algorithm and can be used for non-linear quantile regression models and for more general quantile loss functions. The idea is that: the function $\varphi^{\alpha}_{t+h}(Y_{t+h}, Q_\alpha(Y_{t+h}|\mathbf{X}_t))$ in (7) is twice continuously differentiable by parts and the optimization problem in (8) can be represented as a maximum of two separated branches which are both convex and twice continuously differentiable. Define

$$\psi^{\alpha}_s\left(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})\right) \equiv \exp\left\{\alpha\left[a_s\left(Q_\alpha(Y_s|\mathbf{X}_{s-h})\right) - c_t\left(Y_s\right)\right]\right\}$$

and

$$\phi^{\alpha}_s\left(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})\right) \equiv \exp\left\{-(1-\alpha)\left[a_s\left(Q_\alpha(Y_s|\mathbf{X}_{s-h})\right) - b_t\left(Y_s\right)\right]\right\},$$

the optimization problem in (8) becomes $\max \ln \psi^{\alpha}_1\left(Y_1, Q_\alpha(Y_1|\mathbf{X}_0)\right)$ in case of $t = 1$ and $h = 1$, i.e.

$$\max\min\{\ln \psi^{\alpha}_1\left(Y_1, Q_\alpha(Y_1|\mathbf{X}_0)\right), \ln \phi^{\alpha}_1\left(Y_1, Q_\alpha(Y_1|\mathbf{X}_0)\right)\},$$

or equivalently

$$-\min[\max\{-\ln \psi^{\alpha}_1\left(Y_1, Q_\alpha(Y_1|\mathbf{X}_0)\right), -\ln \phi^{\alpha}_1\left(Y_1, Q_\alpha(Y_1|\mathbf{X}_0)\right)\}].$$

Therefore, the maximization problem in (8) is transformed into a minimax problem.

Using the idea, Komunjer (2005, Theorem 6) shows that the QMLE estimator $\hat{\boldsymbol{\beta}}_{\alpha,h}(\mathcal{D}_t)$ from equation (8) can be written as a solution to a minimax problem

$$\min_{\boldsymbol{\beta}_{\alpha,h}(\mathcal{D}_t)}\left[\max_{t-R+h\le k\le t}\left\{-P_k\left(Y, Q_\alpha(Y|\mathbf{X})\right)\right\}\right],$$

where

$$P_k\left(Y, Q_\alpha(Y|\mathbf{X})\right) \equiv \begin{cases} (R-h)^{-1}\sum\limits_{s=t-R+h+1}^{t} \ln \psi^{\alpha}_s\left(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})\right), \text{ if } k = t-R+h, \\ (R-h)^{-1}\left[\sum\limits_{s=t-R+h+1}^{k} \ln \phi^{\alpha}_s\left(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})\right) + \sum\limits_{s=k+1}^{t} \ln \psi^{\alpha}_s\left(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})\right)\right], \\ \qquad \text{if } t-R+h < k < t, \\ (R-h)^{-1}\sum\limits_{s=t-R+h+1}^{t} \ln \phi^{\alpha}_s\left(Y_s, Q_\alpha(Y_s|\mathbf{X}_{s-h})\right), \text{ if } k = t. \end{cases}$$

Intuitively, Komunjer's minimax algorithm can be decomposed into two step. First, for a given set of parameters, we assign all the forecast errors proper costs to make sure all the forecast errors get positive punishment, i.e., maximize the punishment for a given set of parameters. The second step is to find out the set of parameters that can minimize the forecast cost.

However, the minimax algorithm runs slower than the interior point algorithm.

TABLE 5 ABOUT HERE

To compare the two algorithms, we can check tick losses of quantile prediction of S&P500 daily returns reported in Table 4 (minimax algorithm) and Table 5 (interior point algorithm). The interior point algorithm and minimax algorithm give somewhat different results. Therefore, in small samples, bagging may work differently depending on the estimation algorithms.

7 Bagging Quantile Forecasts with Different Quantile Regression Models

With the flexibility provided by the minimax algorithm, we check the performance of bagging predictors on high non-linear quantile regression models – artificial neural network models. Given model uncertainty, when the sample size is limited, it is usually hard to choose the number of hidden nodes and the number of inputs (lags), and to estimate the large number of parameters in a neural network model. Therefore, a neural network model generate poor predictions with a small sample. In such cases, the bagging can do a wonderful job to improve the forecasting performance.

A non-linear quantile regression function we use in this section is the univariate single-layer feed-forward artificial neural network function of White (1992). Follow the definition in equation (2), the neural network models are set with $\mathbf{X}_t = (1\ Y_t\ Y_{t-1}\ \cdots\ Y_{t-l+1})'$, $\tilde{X}_{t,j} = [1+\exp(-\mathbf{X}_t'\gamma_j)]^{-1}$ $(j = 2, \ldots, k)$, $\tilde{\mathbf{X}}_t = (\mathbf{X}_t'\ \tilde{X}_{t,2}\ \cdots\ \tilde{X}_{t,k})'$, $\boldsymbol{\beta}_{\alpha,h} = [\boldsymbol{\beta}_1'\ \beta_2\ \cdots\ \beta_k]'$ is a $(l+k)$ vector, $\boldsymbol{\beta}_1$ is a $(l+1)$ vector, and $Q_\alpha(Y_{t+h}|\mathbf{X}_t) = \tilde{\mathbf{X}}_t'\boldsymbol{\beta}_{\alpha,h}$. We consider the number of nodes $k-1$ from 0 to 5 and the number of lags l from 1 to 3. Both l and k are selected for each estimation process using the SIC. We choose one combination of p and l from 18 candidates for each prediction. When $k = 1$, we have a linear regression model; and when $k \geq 2$, we have a non-linear regression model.

The neural network model has been widely used in modeling unknown nonlinearities in econometrics and finance. However, with the choice of explanatory variables, and number of nodes, the model uncertainty problem and parameter estimation problem can be very serious. Lee (2000) introduces a method called Bayesian Random Searching (BARS) to choose the optimal number of hidden nodes as well as the best subset of explanatory variables. Instead of choosing only one, he selects several best performed models and takes an average over them. He also provides the asymptotic consistency proof of the posterior neural network regression based on the i.i.d. normal error term assumption. The BARS method is built upon the model space searching work by Raftery, Madigan and Hoeting (1997) and is similar to the approach of Chipman, George, and McCulloch (1998) in their implementation of Bayesian classification and regression tree (CART). We find that the BARS method is simply the BMA weighted bagging when our basic model is the artificial neural network.

Because of the large number of parameters to be estimated and the highly non-linear structure, we can expect that the neural network model will generate poor predictions if we have a small sample size and we can expect that bagging process can play a crucial rule to save the neural network models. The only problem with bagging neural network models is that we need to choose the number of lags, number of nodes and estimate all the parameters for each combination of the lags and nodes, so it takes long computer time to generate predictions. Therefore, we only conduct one empirical experiment to give a rough idea on how bagging predictors work for neural network models..We make quantile predictions with $\alpha = 0.1, 0.3, 0.5, 0.7$, and 0.9 using SP500 monthly data which is summarized as follows:

	In-sample period	Out-of-sample period	$T+1$	P
S&P 500	October 1982 $\sim$ October 1995	November 1995 $\sim$ February 2004	257	100

TABLE 6 ABOUT HERE

FIGURE 4 ABOUT HERE

From Table 6 and Figure 4, we can see that even when in-sample size R is small, unbagged neural network predictors already show some advantage over the simple polynomial (PN) predictors because of flexibility of neural network (NN) models to capture non-linearities in the data. Bagging works well for both PN and NN models, and the improvement by bagging is quite substantial when R is small. When the sample size R is large, the neural network model show much clearer advantage over polynomial predictors by always generating better predictions, and yet bagging neural network predictors still make further improvement over unbagged neural network predictors. Therefore, using bagging predictors, we can save a more complicated prediction model which is more flexible to capture nonlinear structure but harder to estimate.

8 Bagging Binary and Quantile Forecasts in Different Frequencies

We concern about prediction in different frequencies because the predictability of time series may be different in different frequencies. As discussed by, e.g., Christofferson and Diebold (2006), the sign predictability of stock returns may depend on the frequency. The optimal binary prediction $G_{t,1}(\mathbf{X}_t)$ that minimizes $\mathbb{E}_{Y_{t+1}}(\rho_\alpha(G_{t+1} - G_{t,1}(\mathbf{X}_t))|\mathbf{X}_t)$ will be the α-quantile of G_{t+1} conditioning on $\mathbf{X}_t$, which can be achieved by an indicator function of the α-quantile of Y_{t+1} conditioning on $\mathbf{X}_t$ (Lee and Yang, 2006), i.e.,

$$G_{t,1}(\mathbf{X}_t) = Q_\alpha(G_{t+1}|\mathbf{X}_t) = \mathbf{1}(Q_\alpha(Y_{t+1}|\mathbf{X}_t) > 0),$$

where the second equation holds because the indicator function $\mathbf{1}(\cdot)$ is monotonic (Powell, 1986).

TABLE 7 ABOUT HERE

FIGURE 5 ABOUT HERE

We conduct bagging predictions for S&P500 binary and quantile prediction in both daily frequency (Table 2-Panel E, Figure 2-Panel E, Table 5, and Table 7) and monthly frequency (Table 6, Figure 5). We find that the bagging quantile prediction works in a similar pattern for both daily (Table 5) and monthly frequencies (Table 6). However, for binary predictions, bagging works much less with high frequency (daily) series, perhaps because daily signs may be too noisy and difficult to forecast anyway. See Figure 5 for bagging binary prediction on the monthly returns and Table 7 on the daily returns. The result is therefore consistent with Christofferson and Diebold (2006).

9 Pretesting and Bagging

In this section we discuss a potential extension of this paper, with pretesting as considered in Bühlmann and Yu (2002), Inoue and Kilian (2006), and Stock and Watson (2006). Bühlmann and Yu (2002) show that bagging works by smoothing the hard threshold function (e.g. an indicator function). To see this, suppose bootstrap works for $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$, and $Z_n \equiv n^{1/2}(\bar{Y}_n - \mu)/\sigma \to^d N(0,1)$ as $n \to \infty$. Let $y \equiv \mu + c\sigma n^{-1/2}$. Consider a binary model

$$\begin{aligned}
\hat{\theta}_n(y) &= \mathbf{1}(\bar{Y}_n > y) \\
&= \mathbf{1}(\bar{Y}_n > \mu + c\sigma n^{-1/2}) \\
&= \mathbf{1}(n^{1/2}(\bar{Y}_n - \mu)/\sigma > c) \\
&= \mathbf{1}(Z_n > c),
\end{aligned}$$

whose bagging predictor is

$$\begin{aligned}
\hat{\theta}_{n,B}(y) &= \mathbb{E}^*\hat{\theta}_n^*(y) \\
&= \mathbb{E}^*\mathbf{1}(\bar{Y}_n^* > y) \\
&= \mathbb{E}^*\mathbf{1}\left(n^{1/2}(\bar{Y}_n^* - \bar{Y}_n)/\sigma > n^{1/2}(y - \bar{Y}_n)/\sigma\right) \\
&\approx 1 - \Phi(n^{1/2}(y - \bar{Y}_n)/\sigma) \\
&= 1 - \Phi(n^{1/2}(\mu + c\sigma n^{-1/2} - \bar{Y}_n)/\sigma) \\
&= 1 - \Phi(c - Z_n),
\end{aligned}$$

where $\approx$ denotes the asymptotic equivalence when $n \to \infty$. When $y = \mu$, $\hat{\theta}_n(y)$ is most unstable. Let us compare the predictors at this value $y = \mu$ (or $c = 0$),. When $y = \mu$ ($c = 0$), $\hat{\theta}_n(\mu)$ has mean 1/2 and variance $(1/2)(1-1/2) = 1/4$. In comparison, when $y = \mu$ ($c = 0$), the bagging predictor $\hat{\theta}_{n,B}(\mu) \approx 1 - \Phi(-Z_n) = \Phi(Z_n) = U$ has mean 1/2 and variance 1/12. Hence, bagging reduces the variance of the predictor from 1/4 to 1/12.

Bühlmann and Yu (2002) use the above idea that bagging works via smoothing the hard-thresholding into soft-thresholding for the location model and regression model as well. Consider a location model with pretesting (PT)

$$PT = \hat{\theta}_n(y) = \hat{\beta}_{0,n}\mathbf{1}(\hat{\beta}_{0,n} > y) = \hat{\beta}_{0,n}\mathbf{1}(Z_n > c),$$

and its bagging predictor (BA)

$$BA = \hat{\theta}_{n,B}(y) = \mathbb{E}^*\hat{\theta}_n^*(y) = \mathbb{E}^*\hat{\beta}_{0,n}^*\mathbf{1}(\hat{\beta}_{0,n}^* > y) = \mathbb{E}^*\hat{\beta}_{0,n}^*\mathbf{1}(Z_n^* > c).$$

Here the location parameter is β_0 if $Z_n > c$, and zero otherwise. The PT model has hard thresholding around $Z_n = c$, while BA has a smooth soft-thresholing.

Bühlmann and Yu (2002) also consider the variable-selection in a regression model by pretesting

$$PT = \hat{\theta}_n(y) = \sum_{j=0}^{M} \hat{\beta}_{j,n} \mathbf{1}(\hat{\beta}_{j,n} > y) x_n^{(j)} = \sum_{j=0}^{M} \hat{\beta}_{j,n} \mathbf{1}(Z_{n,j} > c) x_n^{(j)},$$

where the jth variable $x_n^{(j)}$ is included if its coefficient is bigger than a given threshold c. The variable-selection conducted via pretesting introduces a hard-thresholding. The bagging can smooth the hard-thresholding in this case as follows

$$BA = \hat{\theta}_{n,B}(y) = \mathbb{E}^* \hat{\theta}_n^*(y) = \mathbb{E}^* \sum_{j=0}^{M} \hat{\beta}_{j,n}^* \mathbf{1}(Z_j^* > c) x_n^{(j)}.$$

Inoue and Kilian (2006) exploit this idea that bagging can reduce the variance of the predictor from a regression model when the predictors/regressors are selected by pretesting, to show how bagging works for forecasting inflation.

Breiman (1996) and Lee and Yang (2006) consider the case when $c = 0$. In other words, they did not consider pretesting, and bagging is applied to unrestricted regression (UR) with all the M predictors/regressors included (without selecting a subset of them by pretesting). In this case, bagging would still work especially when UR is bad (particularly in small sample). Certainly $c = 0$ is not optimal as bagging would work better with some larger values of c. If $c = 0$, bagging is *not asymptotically* admissible (Stock and Watson 2006). An example is shown in Lee and Yang (2006) for bagging binary prediction with majority-voting where bagging works well in small samples but does not work asymptotically with $c = 0$. The choice of c is like the choice of the shrinkage parameter as shown in Stock and Watson (2006) and also noted in Inoue and Kilian (2006). Stock and Watson show that $c = 1.96$ is too small for bagging to be comparable to the factor methods. As Stock and Watson note, $c = 2.58$ makes bagging work better. In this paper, we consider only $c = 0$ (no pretesting) for both binary and quantile prediction as in Lee and Yang (2006). With pretesting ($c > 0$) we expect that bagging would work more/better, based on the results of Bühlmann and Yu (2002), Inoue and Kilian (2006), and Stock and Watson (2006). Investigation of bagging with pretesting for the binary and quantile prediction is left for future work. I can be easily conjectured that pretesting would be more beneficial in improving bagging, particularly for longer multi-step forecasting.

10 Summary and Field Guide

This section is to provide a brief field guide to bagging based on what we have learned in this paper. Bagging is a smoothing method to improve predictive ability under the presence of parameter estimation uncertainty and model uncertainty. There are two ways of aggregating – Averaging or Voting. Bagging quantile predictors are constructed via weighted averaging over predictors trained on bootstrapped training samples. Bagging binary predictors are conducted via (majority) voting on predictors trained on the bootstrapped training samples.

To understand how bagging works various explanations have been made. It may be hard to understand the meaning of multiple training set $\mathcal{D}_t^{(j)}$ in the time series circumstances since time is not repeatable. However, considering an example of the estimation and forecast procedure with panel data may be helpful. Suppose we want to forecast consumption of a household in next period. When the historical observations of the interested household is very limited, our parameters estimated and the predictors will have rather large variances, especially for non-linear regression models. If we can find some other households that have similar consumption patterns (similar underlying probability distribution $\mathbf{P}$), it would be better to use historical observations from all similar households than just from this interested household in the estimation process, though we only use data of this interested household to do forecast. Therefore, the ensemble aggregating predictor is just like to find similar households, and the bootstrap aggregating predictor is just like to find similar bootstrapped (artificial) households.

What was done in Lee and Yang (2006) is to examine how bagging works (i) with equal-weighted and BMA-weighted averaging, (ii) for one-step ahead binary prediction (with voting) and for one-step ahead quantile prediction (with averaging), (iii) with particular choice of a loss function (linlin, check) and (iv) with particular choice of a regression model (linear, polynomial).

What we do in this paper ("Further Issues") is to consider (i) different aggregating schemes (trimmed mean bagging, median bagging), (ii) multi-step forecast horizons (to see how bagging performs with greater uncertainty), (iii) a more general class of loss functions, i.e., so called the tick-exponential family to examine the effect of the convexity of the loss (in addition to the check loss (lin-lin) for quantile estimation), (iv) different algorithms (the minimax algorithm vs the interior point algorithm for the estimation of the quantile model), (v) different regression models (polynomial quantile model and neural network quantile model), and (vi) different data frequencies (monthly and daily S&P500 returns).

What we find now is as follows. (i) Median bagging and trimmed-mean bagging can be more robust to extreme predictors from bootstrap samples and have better performance than equally weighted bagging predictors. (ii) Bagging works more with longer forecast horizons. (iii) Bagging works well under more general tick loss functions. (iv) Bagging may work differently with different quantile estimation algorithms. (v) Bagging works well with highly nonlinear quantile regression models (e.g., artificial neural network). (vi) Bagging quantile predictor is not affected by the frequency of the data, while bagging binary predictor is much affected when daily returns are considered instead of month returns.

From comparing different averaging schemes, we find that (i) the BMA-, median-, and trimmed-bagging predictors have better predicting performance than equal-weighted bagging predictors even when we have a relatively large sample size. (ii) The median bagging is generally the best. (iii) The outstanding performance of median bagging predictors is most obvious when α values are close to 0 or 1, where the extreme value problem are most serious because there are fewer observations on tails and the parameters regression estimators are sensitive to the estimation sample. However, the advantage of median bagging predictors are not so clear when α values are close to 0.5. (iv) Bagging works more when the sample size is smaller. (v) Bagging works more when α-quantiles lie on the sparse part of the error distribution. Our explanation is that for the sparse part of the error distribution, there are fewer observations, therefore quantile predictions are sensitive to the estimation sample and bagging predictors work better for unstable predictions.

From bagging multi-step quantile forecasts, we find that the performance of bagging relative to unbagged predictor gets better as the forecast horizon increases. From examining how other algorithms may work for the bagging, we find that the interior point algorithm and minimax algorithm give somewhat different results. Therefore in small samples, bagging may work differently depending on the estimation algorithms. From checking the performance of bagging predictors on high non-linear quantile regression models – artificial neural network models, we find that, given model uncertainty when the sample size is limited, it is usually hard to choose the number of hidden nodes and the number of inputs (lags), and to estimate the large number of parameters in a neural network model, in which cases, using bagging predictors, we can save the complicated model (more flexible to capture nonlinear structure but harder to estimate) for out-of-sample forecasting.

11 Conclusions

We have examined how bagging work for the binary prediction and the quantile prediction with different bagging weighting schemes, different forecast horizons, different loss functions, different estimation algorithms, different regression models – linear and nonlinear (polynomial, neural network), and different data frequencies for time series.

Bagging the conditional quantile predictors are constructed via weighted averaging over predictors trained on bootstrapped training samples, and bagging binary predictors are conducted via majority voting on predictors trained on the bootstrapped training samples.

We show that the median bagging, the trimmed bagging and the BMA bagging can alleviate the problem of extreme predictors from bootstrap sampling errors and further improve the performance of simple averaging bagging predictors. Interestingly, it is found that the performance of bagging predictors gets better with the increase the forecast horizons. This means that there is more room (due to more uncertainty) for bagging to operate for longer forecast horizon h.

Finally, as this paper contributes to a volume, *Forecasting in Presence of Structural Breaks and Model Uncertainty*, we conclude with some comments on how/why bagging may be useful in the presence of structural breaks and model uncertainty.

In the presence of structural breaks: In this paper we find that bagging may work more when the size of the training sample is small and the predictor is unstable. Bagging seems to smooth out the parameter estimation uncertainty due to a small sample to improve the forecasting performance. The potential advantage of bagging lies in areas where small sample is common. Bagging may be useful when structural breaks are frequent so that simply using as many observations as possible is not a wise choice for out-of-sample prediction. The forecast can fail under the presence of breaks. It is not clear whether using the samples after the breaks is optimal or not, as pointed in a recent paper by Pesaran and Timmermann (2007). When we are to use only the post-break samples, bagging can be of help. It is not clear whether using the samples after the breaks is optimal or not and whether/how bagging can be of help. It is very likely that the forecast will be affected by the breaks. While it can be easily done in a simple simulation exercise to illustrate how bagging works under breaks, it could be done much more carefully given the new results of Pesaran and Timmermann, and therefore we leave this for other work.

In the presence of model uncertainty: Bagging is a smoothing method to improve predictive ability under the presence of parameter estimation uncertainty and model uncertainty. For example, as we find in Section 4, bagging performs more for multi-step prediction with larger h, as there is more uncertainty for longer forecast horizon and more smoothing can operate. As a referee points out, bagging may improve forecasting when there is uncertainty concerning measurement of a variable, functional form and the exact proxy to use. We leave the investigation of these for a future study as the current results in the paper do not directly address the issues of measurement and proxy variables. For functional form, it is a form of model uncertainty and so bagging may smooth it out. However, this is also an issue that is not directly addresses in the current results of the paper. Many issues are still left for further work even after the "Further Issues" considered in the present paper.

Acknowledgements

We would like to thank the co-editors (David Rapach and Mark Wohar), an anonymous referee, and the seminar participants at the conference on "Forecasting in Presence of Structural Breaks and Model Uncertainty" at Saint Louis University for their useful comments. A part of the research was conducted while Lee was visiting the California Institute of Technology. Lee thanks for their hospitality and the financial support during the visit. Yang thanks for the Chancellor's Distinguished Fellowship from University of California, Riverside. All remaining errors are our own.

References

Brown, B.W. and R. S. Mariano (1989), Residual-Based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System, *Econometrica*, 52(1), 321-344.

Breiman, L. (1996), Bagging Predictors, *Machine Learning,* 24, 123-140.

Buchinsky, M. (1998), Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research, *The Journal of Human Resources*, 33(1), 88-126.

Chernozhukov, V. and Hong, H. (2003), An MCMC Approach to Classical Estimation, *Journal of Econometrics*, 115, 293-346.

Chernozhukov, V. and Len Umantsev (2001), Conditional Value-at-Risk: Aspects of Modeling and Estimation, *Empirical Economics*, 26, 271-292.

Chipman, H., E. George, and R. McCulloch (1998), Bayesian CART Model Search, *The Journal of the American Statistical Association*, 93, 935-960.

Chevillon, G. and D. Hendry (2004), Non-Parametric Direct Multi-step Estimation for Forecasting Economic Processes, *International Journal of Forecasting*, 21, 201-218.

Clements, M. P. and D. F. Hendry (1999), *Forecasting Non-Stationary Economic Time Series*, Cambridge, MA: The MIT Press.

Christofferson, P.F. and F.X. Diebold (2006), Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics, *Management Science*, 52, 1273-1287.

Engle, R.F. and S. Manganelli (2004), CaViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and Economic Statistics*, 22(4), 367-381.

Frenk, J.B.G., Gromicho, J. and Zhang, S. (1994), A Deep Cut Ellipsoid Algorithm for Convex Programming: Theory and Applications, *Mathematical Programming*, 163, 83-108.

Granger, C.W.J. and P. Newbold (1976), Forecasting Transformed Variables, *Journal of the Royal Statistical Society Series B*, 38, 189-203.

Granger, C.W.J. and M.H. Pesaran (2000), Economic and Statistical Measures of Forecast Accuracy, *Journal of Forecasting,* 19, 537-560.

Koenker, R and G. Basset (1978), Asymptotic Theory of Least Absolute Error Regression, *The Journal of the American Statistical Association*, 73, 618-622.

Koenker, R. and B.J. Park (1996), An Interior Point Algorithm for Nonlinear Quantile Regression, *Journal of Econometrics*, 71, 265-283.

Komunjer, I. (2005), Quasi-Maximum Likelihood Estimation for Conditional Quantiles, *Journal of Econometrics*, 128(1), 137-164.

Komunjer, I and Q. Vuong (2005), Efficient Conditional Quantile Estimation: The Time Series Case, UCSD and Penn State University.

Lee, H. (2000), Consistency of Posterior Distributions for Neural Networks, *Neural Networks*, 13, 629-642.

Lee, T.-H. and Y. Yang (2006), Bagging Binary and Quantile Predictors for Time Series, *Journal of Econometrics*, 135, 465-497.

Lin, J.L and C.W.J. Granger (1994), Forecasting from Non-Linear Models in Practice, *Journal of Forecasting*, 13, 1-9.

Lin, J. L. and R. S. Tsay (1996), Co-integration Constraint and Forecasting: An Empirical Examination, *Journal of Applied Econometrics*, 11, 519-538.

Marron, J.S. and M.P. Wand (1992), Exact Mean Integrated Squared Error, *Annals of Statistics*, 20, 712–736.

Pesaran, M.H. and A. Timmermann (2007), Selection of Estimation Window in the Presence of Breaks, *Journal of Econometrics*, forthcoming.

Portnoy, S. and R. Koenker (1997), The Gaussian Hare and the Laplacean Tortoise: Computability of l_1 vs l_2 Regression Estimators, *Statistical Science*, 12, 279-300.

Powell, J.L. (1986), Censored Regression Quantiles, *Journal of Econometrics,* 32, 143-155.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997), Bayesian Model Averaging for Linear Regression Models, *Journal of the American Statistical Association*, 92, 179-191.

Stock, J.H. and M.W. Watson (1999), A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series, *Cointegration, Causality, and Forecasting, A Festschrift in Honor of C.W.J. Granger*, edited by R.F. Engle and H. White, Oxford University Press: London, pp. 1-44.

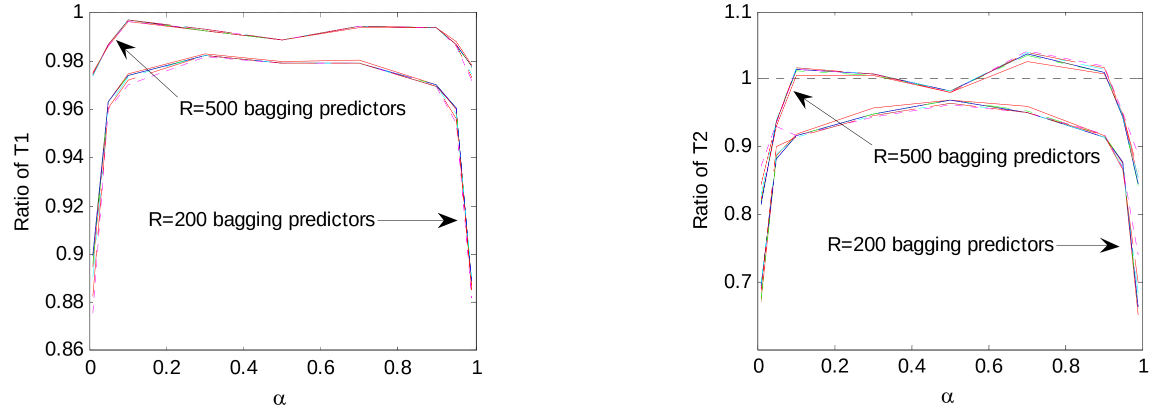
Timmermann, A. (2007), Forecast Combinations, *Handbook of Economic Forecasting*, edited by G. Elliott, C.W.J. Granger, and A. Timmermann, Amsterdam: North-Holland, forthcoming.

Tsay, R. S. (1993), Comment: Adaptive Forecasting, *Journal of Business and Economic statistics*, 11(2), 140-142.

White, H. (1992), Nonparametric Estimation of Conditional Quantiles Using Neural Networks, *Proceedings of the Symposium on the Interface*, New York: Springer-Verlag, 190-199.

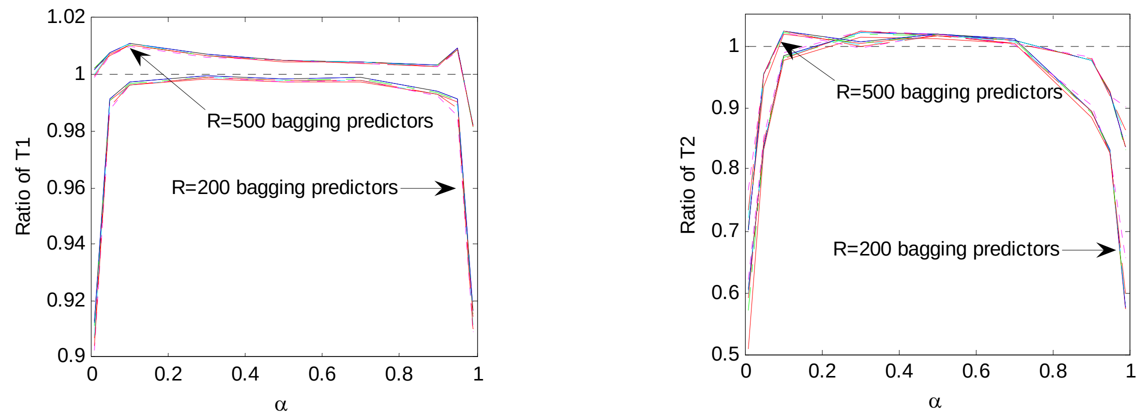
Figure 1. Bagging Quantile Prediction for AR-ARCH Models

Panel (A): AR(0)-ARCH(1)-Gaussian



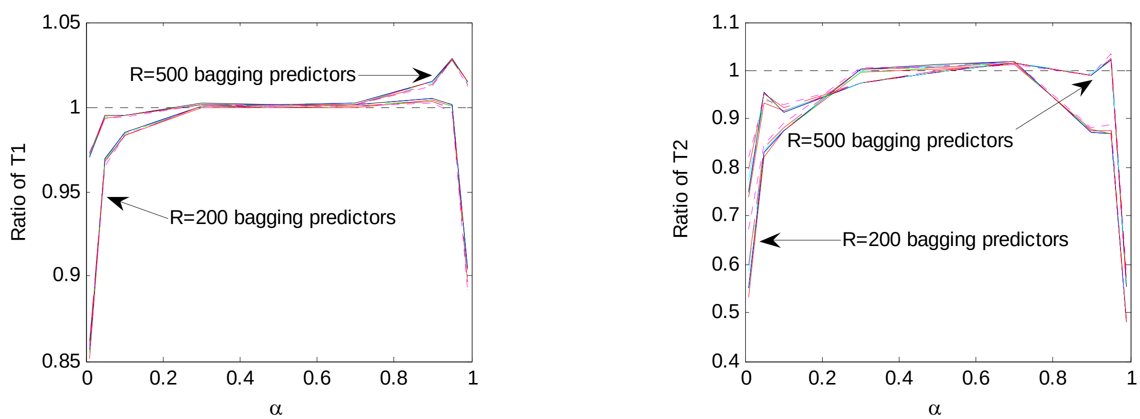
Note: The ARCH(1) parameter in Equation (1) is $\theta = 0.5$. The two figures report the tick loss ratio and standard error ratio of bagging predictors over unbagged predictors for 100 Monte Carlo replications.

Panel B. AR(1)-ARCH(0)-Gaussian

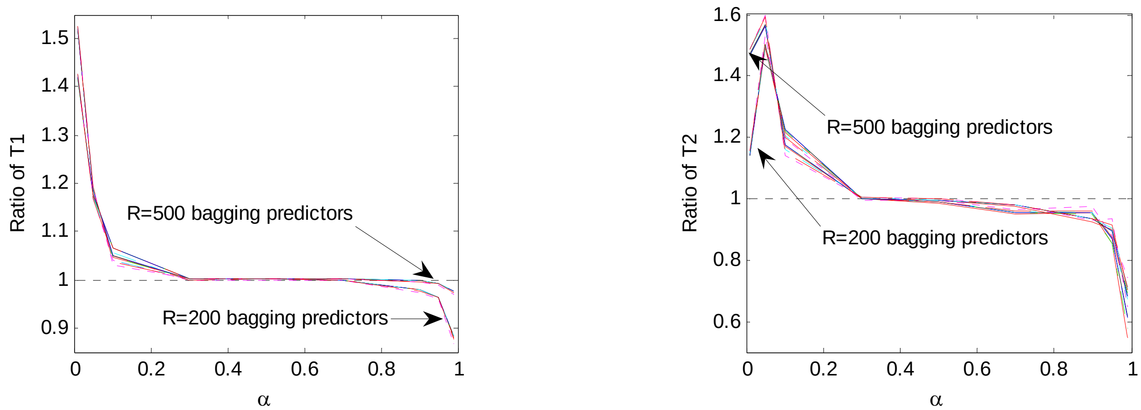


Note: The AR(1) parameter in Equation (1) is $\rho = 0.6$.

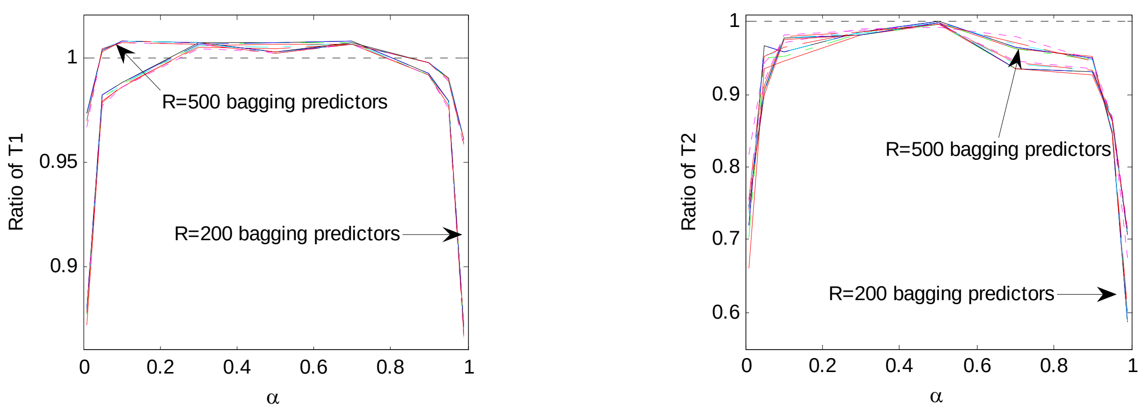
Panel C. AR(1)-ARCH(0)-Skewed unimodal



Panel D. AR(1)-ARCH(0)-Strongly skewed



Panel E. AR(1)-ARCH(0)-Kurtotic unimodal



Panel F. AR(1)-ARCH(0)-Outlier

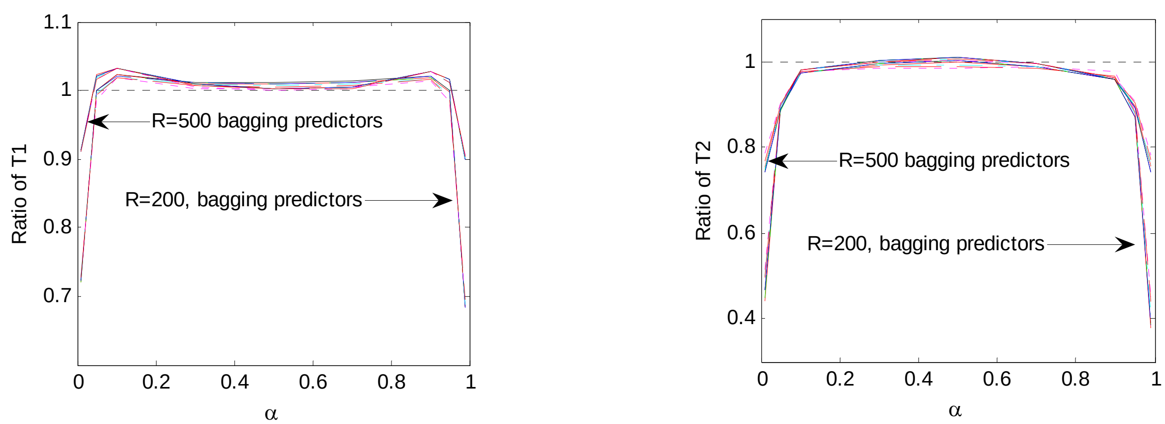
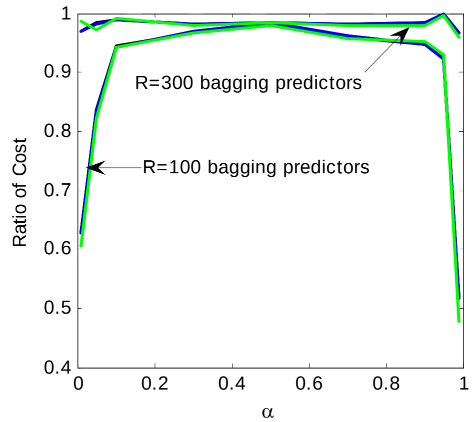
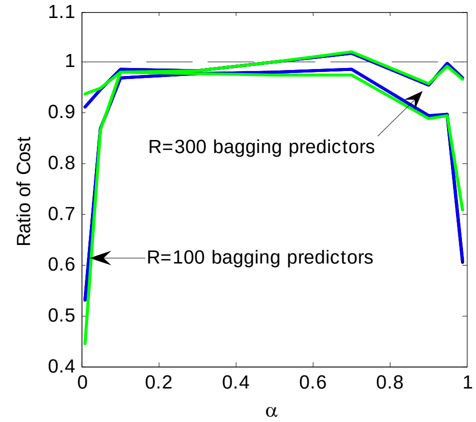


Figure 2. Empirical Applications of Bagging Quantile Prediction

Panel A. USD/EUR Daily Returns

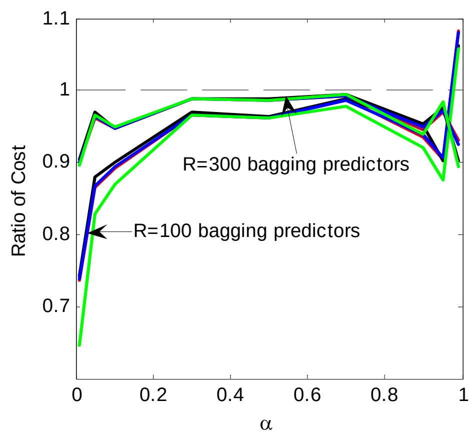


Panel B. USD/JPY Daily Returns

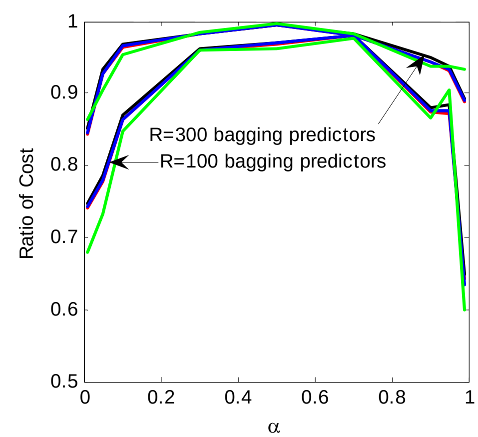


Note: Panels A and B report the tick loss ratio of bagging predictors over unbagged predictors during 08/05/04 ~ 04/11/05.

Panel C. Dow Jones Industrial Averages Daily Returns

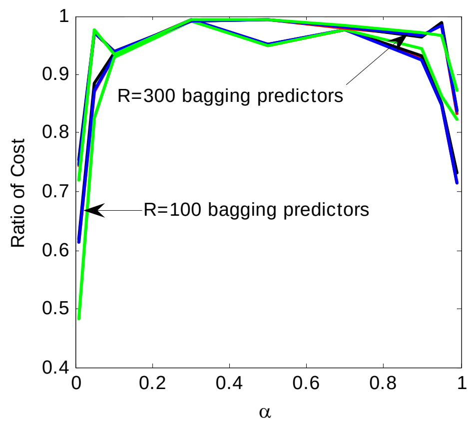


Panel D. New York Stock Exchange Composite Daily Returns

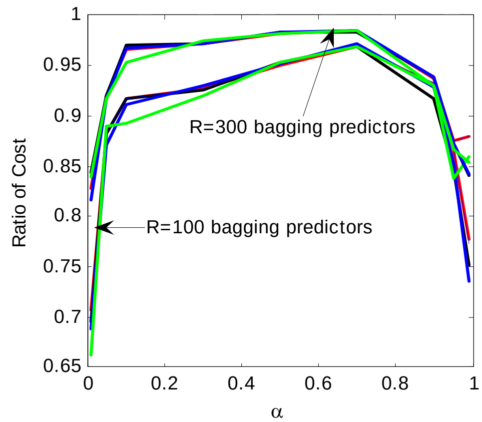


Note: Panels C and D report the tick loss ratio of bagging predictors over unbagged predictors over 01/05/00~12/31/00.

Panel E. Standard and Poor's 500 Daily Returns

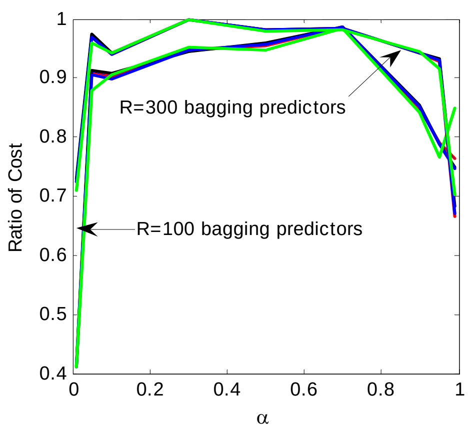


Panel F. NASDAQ Daily Returns

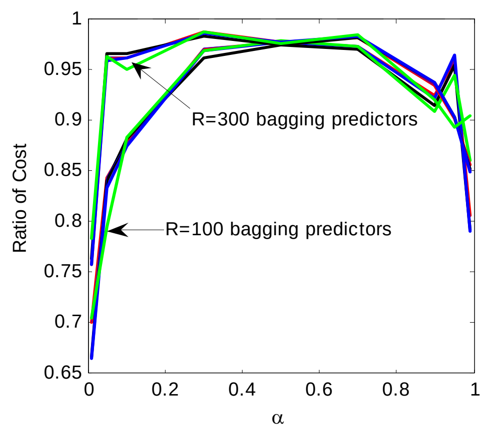


Note: Panels E and F report the tick loss ratio of bagging predictors over unbagged predictors over 01/05/00~12/31/00.

Panel G. Russell 2000 Daily Returns

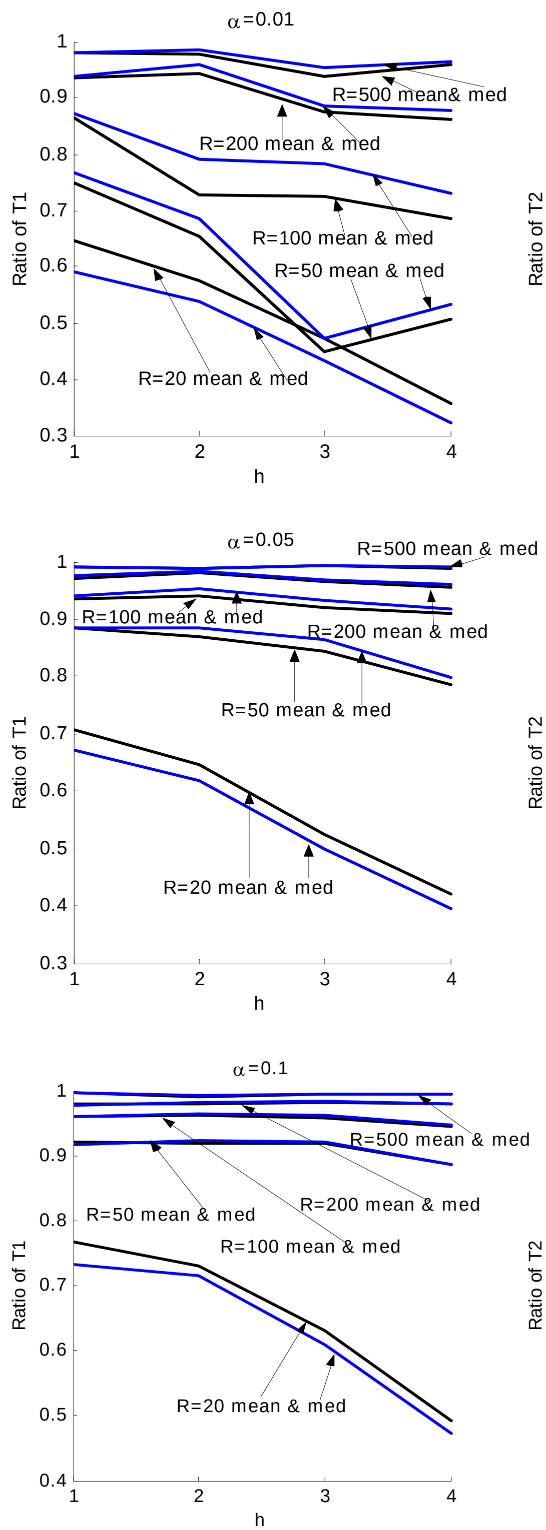


Panel H. Pacific Exchange Technology Daily Returns



Note: Panels G and H report the tick loss ratio of bagging predictors over unbagged predictors over 01/05/00~12/31/00.

Figure 3. Bagging Multi-step Quantile Forecast for AR(0)-ARCH(1)-Gaussian Model



Note: The two figures in each row report the tick loss ratio and standard error ratio of bagging predictors over unbagged predictors over 100 Monte Carlo replications (see detail explanations in the main text).

Figure 3 (Continued). Bagging Multi-step Quantile Forecast for AR(0)-ARCH(1)-Gaussian Model

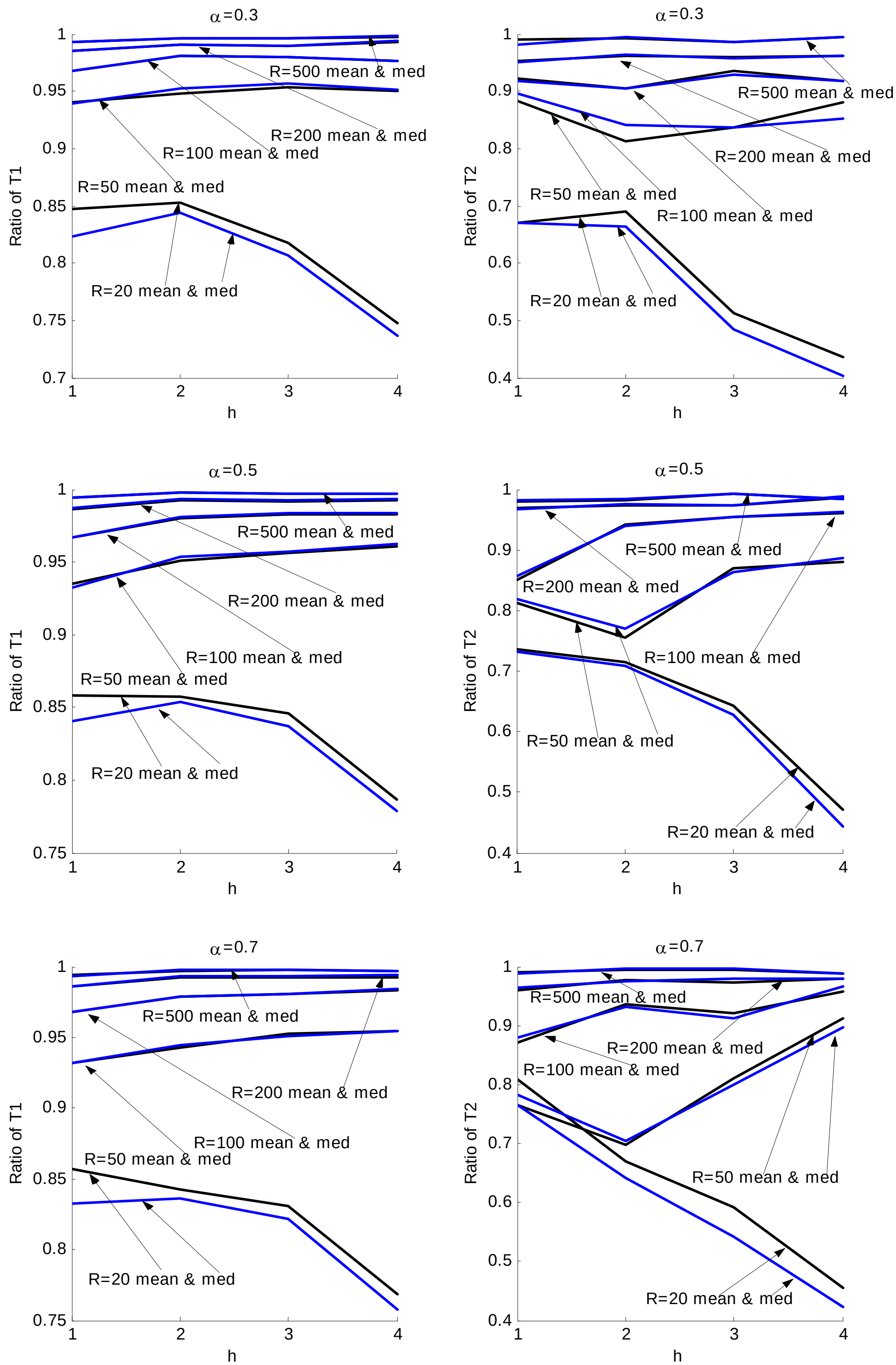


Figure 3 (Continued). Bagging Multi-step Quantile Forecast for AR(0)-ARCH(1)-Gaussian Model

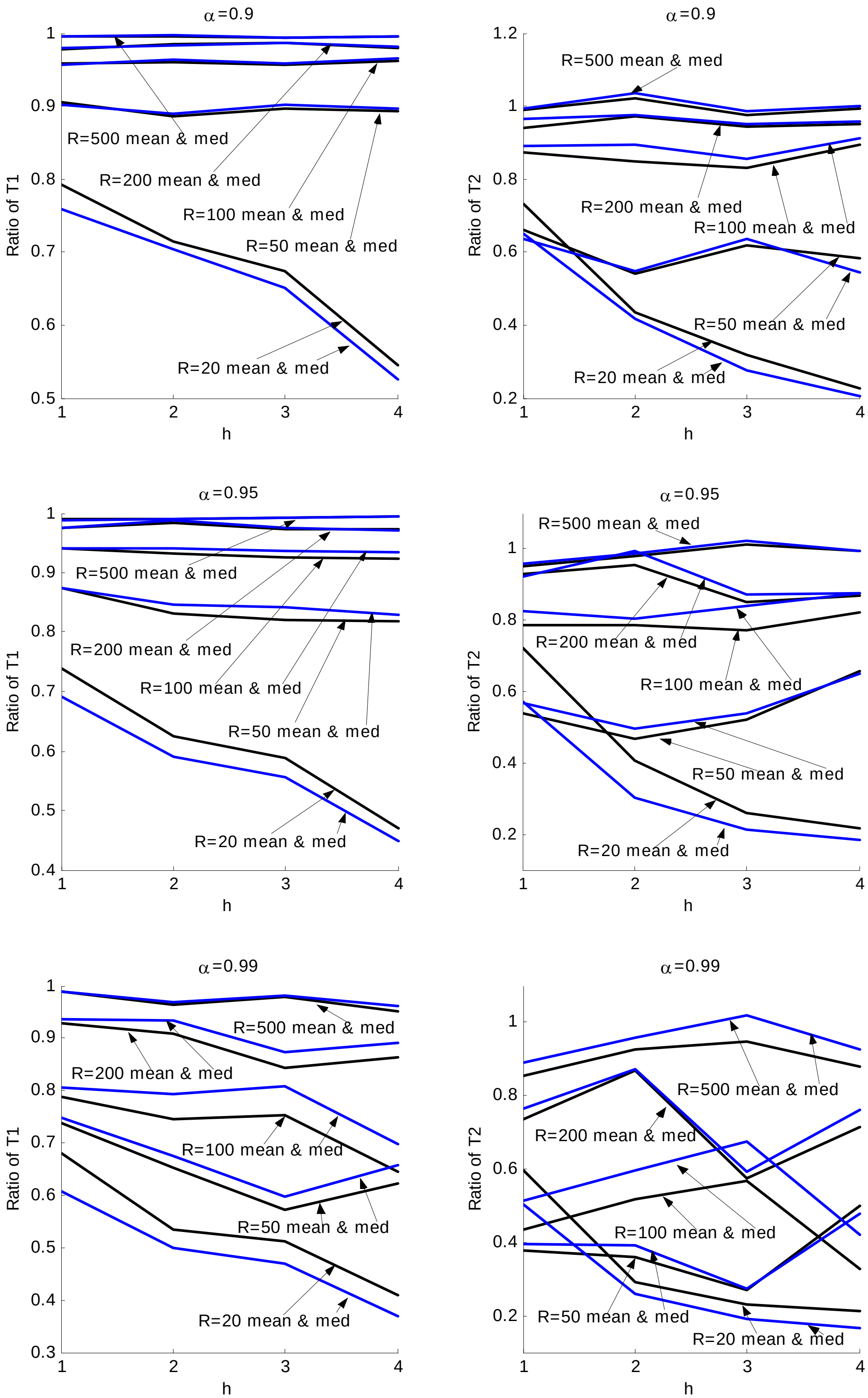
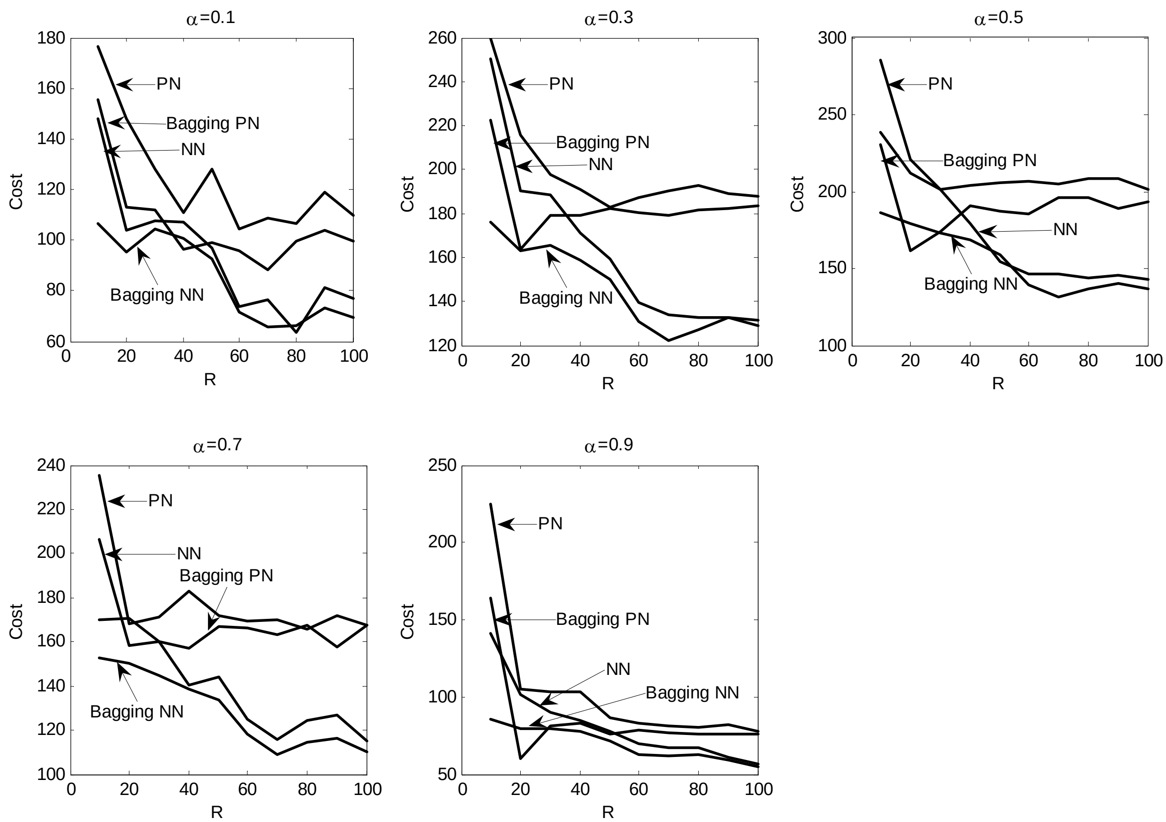
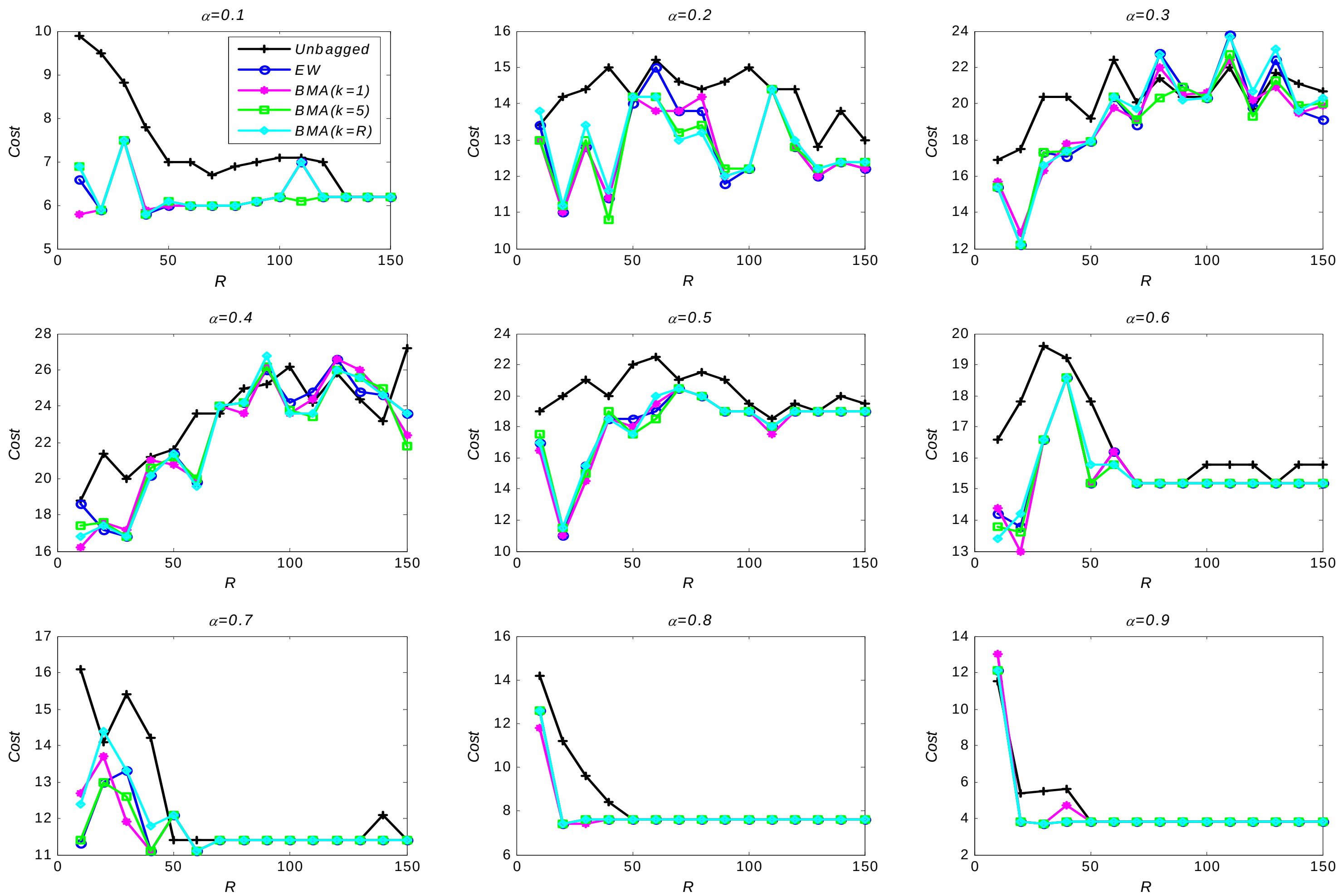


Figure 4. Bagging Quantile Predictions with Different Regression Models



Note: The five figures report the tick losses of quantile predictors of SP500 monthly returns over the period November 1995 ~ February 2004 using polynomial and neural network quantile regression models. PN represents the forecast loss from polynomial model and NN represents the forecast loss from neural network model.

Figure 5: Bagging Binary Prediction for SP 500 Monthly Return



Note: The nine graphs above report the asymmetric losses of binary predictors of SP500 monthly returns over the period November 1995 ~ February 2004.

Table 1. Bagging Quantile Prediction for AR-ARCH Models

Panel A. AR(0)-ARCH(1)-Gaussian

		R = 200								R = 500							
		J=1	J=50							J=1	J=50						
			mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$		mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$
α=.01	T_1	2.92	2.62	2.61	2.61	2.62	2.55	2.60	2.57	2.64	2.57	2.57	2.57	2.57	2.57	2.57	2.57
	T_2	1.04	0.72	0.70	0.70	0.72	0.85	0.72	0.71	0.57	0.46	0.47	0.47	0.46	0.50	0.47	0.48
	T_3		0.77	0.78	0.78	0.78	0.85	0.78	0.81		0.58	0.58	0.58	0.58	0.61	0.61	0.63
α=.05	T_1	9.97	9.60	9.60	9.60	9.60	9.57	9.58	9.57	9.97	9.84	9.83	9.84	9.84	9.83	9.84	9.84
	T_2	1.83	1.61	1.62	1.61	1.61	1.70	1.62	1.64	1.53	1.43	1.42	1.43	1.43	1.43	1.43	1.43
	T_3		0.83	0.82	0.83	0.83	0.92	0.85	0.87		0.68	0.71	0.69	0.68	0.76	0.72	0.73
α=.10	T_1	16.73	16.30	16.30	16.30	16.30	16.23	16.28	16.26	16.74	16.69	16.68	16.69	16.69	16.68	16.69	16.69
	T_2	2.73	2.50	2.51	2.50	2.50	2.50	2.50	2.49	2.28	2.32	2.30	2.31	2.32	2.32	2.32	2.32
	T_3		0.81	0.82	0.83	0.81	0.86	0.83	0.85		0.71	0.74	0.74	0.71	0.75	0.72	0.73
α=.30	T_1	32.86	32.27	32.31	32.27	32.27	32.25	32.27	32.27	32.52	32.30	32.29	32.30	32.30	32.30	32.30	32.30
	T_2	4.51	4.27	4.31	4.28	4.27	4.26	4.27	4.27	3.86	3.88	3.88	3.88	3.88	3.89	3.89	3.89
	T_3		0.88	0.89	0.88	0.88	0.89	0.90	0.91		0.70	0.70	0.70	0.70	0.70	0.69	0.69
α=.50	T_1	37.53	36.76	36.77	36.76	36.76	36.76	36.76	36.76	37.12	36.70	36.69	36.70	36.70	36.70	36.69	36.69
	T_2	4.90	4.74	4.75	4.75	4.74	4.72	4.73	4.72	4.49	4.41	4.40	4.40	4.41	4.40	4.40	4.40
	T_3		0.90	0.92	0.91	0.90	0.91	0.91	0.90		0.86	0.86	0.86	0.86	0.88	0.86	0.87
α=.70	T_1	32.70	32.01	32.05	32.02	32.01	32.02	32.02	32.02	32.48	32.30	32.28	32.30	32.30	32.31	32.30	32.30
	T_2	4.42	4.20	4.24	4.21	4.20	4.19	4.20	4.20	4.07	4.22	4.17	4.21	4.22	4.24	4.23	4.23
	T_3		0.92	0.93	0.92	0.92	0.94	0.91	0.93		0.77	0.76	0.77	0.77	0.75	0.75	0.75
α=.90	T_1	16.86	16.36	16.36	16.36	16.36	16.35	16.35	16.35	16.62	16.51	16.52	16.51	16.51	16.52	16.52	16.52
	T_2	2.58	2.36	2.37	2.36	2.36	2.37	2.36	2.36	2.36	2.38	2.37	2.38	2.38	2.40	2.39	2.39
	T_3		0.88	0.89	0.89	0.88	0.90	0.92	0.93		0.74	0.71	0.74	0.74	0.79	0.73	0.72
α=.95	T_1	10.12	9.72	9.72	9.72	9.72	9.66	9.70	9.68	9.84	9.71	9.72	9.71	9.71	9.70	9.70	9.70
	T_2	1.74	1.53	1.53	1.53	1.53	1.51	1.52	1.51	1.67	1.58	1.59	1.58	1.58	1.59	1.58	1.59
	T_3		0.87	0.85	0.87	0.86	0.90	0.88	0.88		0.71	0.73	0.72	0.71	0.76	0.75	0.79
α=.99	T_1	2.96	2.63	2.62	2.63	2.63	2.61	2.64	2.63	2.65	2.59	2.59	2.59	2.59	2.57	2.58	2.58
	T_2	1.05	0.69	0.68	0.69	0.69	0.78	0.72	0.73	0.76	0.64	0.65	0.64	0.64	0.68	0.65	0.66
	T_3		0.74	0.77	0.75	0.74	0.83	0.81	0.80		0.58	0.58	0.58	0.58	0.66	0.60	0.63

Note: The ARCH(1) parameter is $\theta = 0.5$ in Equation (1). See the definition of T_1, T_2, and T_3 in the text, which are computed from 100 Monte Carlo replications.

Panel B. AR(1)-ARCH(0)-Gaussian

		R = 200								R = 500							
		J=1	J=50							J=1	J=50						
			mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$		mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$
α=.01	T_1	3.07	2.81	2.79	2.80	2.80	2.77	2.79	2.78	2.78	2.78	2.79	2.79	2.78	2.78	2.78	2.78
	T_2	0.85	0.51	0.43	0.48	0.51	0.53	0.51	0.50	0.53	0.37	0.37	0.37	0.37	0.40	0.38	0.39
	T_3		0.64	0.67	0.66	0.64	0.64	0.64	0.65		0.37	0.37	0.37	0.37	0.37	0.39	0.37
α=.05	T_1	10.64	10.55	10.54	10.55	10.55	10.51	10.53	10.52	10.59	10.67	10.66	10.67	10.67	10.66	10.67	10.66
	T_2	1.23	1.03	1.02	1.03	1.03	1.05	1.03	1.04	1.33	1.27	1.24	1.26	1.27	1.26	1.27	1.27
	T_3		0.55	0.54	0.55	0.55	0.60	0.59	0.58		0.43	0.44	0.43	0.43	0.39	0.4	0.38
α=.10	T_1	17.90	17.85	17.83	17.84	17.85	17.83	17.83	17.83	17.86	18.05	18.04	18.05	18.05	18.04	18.05	18.05
	T_2	1.76	1.73	1.71	1.73	1.73	1.74	1.72	1.72	1.80	1.85	1.84	1.84	1.85	1.84	1.85	1.85
	T_3		0.58	0.58	0.58	0.58	0.55	0.58	0.59		0.32	0.32	0.32	0.32	0.32	0.32	0.32
α=.30	T_1	35.29	35.27	35.22	35.25	35.26	35.25	35.25	35.25	34.83	35.07	35.05	35.06	35.07	35.04	35.06	35.05
	T_2	2.92	2.98	2.96	2.97	2.98	2.99	2.98	2.98	2.55	2.56	2.56	2.56	2.56	2.54	2.55	2.54
	T_3		0.58	0.58	0.58	0.58	0.55	0.59	0.57		0.33	0.32	0.33	0.33	0.34	0.32	0.33
α=.50	T_1	40.23	40.16	40.12	40.14	40.16	40.12	40.15	40.14	39.81	40.01	39.99	40.01	40.01	40.01	40.00	40.00
	T_2	3.11	3.17	3.15	3.16	3.17	3.17	3.16	3.16	2.79	2.84	2.83	2.84	2.84	2.84	2.83	2.83
	T_3		0.58	0.61	0.60	0.59	0.61	0.59	0.59		0.43	0.43	0.43	0.43	0.44	0.44	0.43
α=.70	T_1	34.93	34.89	34.85	34.87	34.89	34.86	34.87	34.86	34.69	34.86	34.84	34.85	34.85	34.83	34.84	34.84
	T_2	2.65	2.68	2.66	2.67	2.68	2.66	2.67	2.66	2.53	2.55	2.55	2.55	2.55	2.53	2.55	2.54
	T_3		0.58	0.62	0.58	0.58	0.61	0.61	0.61		0.45	0.44	0.45	0.45	0.44	0.45	0.46
α=.90	T_1	17.80	17.69	17.67	17.68	17.69	17.66	17.68	17.67	17.61	17.67	17.67	17.67	17.67	17.66	17.67	17.67
	T_2	1.60	1.43	1.42	1.43	1.43	1.45	1.43	1.43	1.66	1.62	1.62	1.62	1.62	1.62	1.62	1.62
	T_3		0.54	0.54	0.54	0.54	0.57	0.57	0.57		0.51	0.5	0.5	0.51	0.5	0.49	0.49
α=.95	T_1	10.55	10.46	10.45	10.45	10.46	10.39	10.44	10.42	10.33	10.42	10.42	10.42	10.42	10.41	10.42	10.42
	T_2	1.13	0.94	0.93	0.94	0.94	0.93	0.93	0.93	1.22	1.13	1.13	1.13	1.13	1.13	1.13	1.12
	T_3		0.53	0.53	0.53	0.53	0.61	0.55	0.57		0.4	0.4	0.4	0.4	0.4	0.41	0.42
α=.99	T_1	3.01	2.76	2.75	2.75	2.76	2.73	2.74	2.74	2.81	2.76	2.76	2.76	2.76	2.77	2.77	2.76
	T_2	0.67	0.39	0.39	0.39	0.39	0.44	0.39	0.40	0.57	0.48	0.48	0.47	0.48	0.51	0.49	0.49
	T_3		0.65	0.66	0.65	0.65	0.67	0.66	0.66		0.48	0.48	0.48	0.48	0.48	0.48	0.48

Note: The AR(1) parameter is $\rho = 0.6$ in Equation (1). See the definition of T_1, T_2, and T_3 in the text, which are computed from 100 Monte Carlo replications.

Panel C. AR(1)-ARCH(0)-Skewed unimodal

		R = 200								R = 500							
		J=1	J=50							J=1	J=50						
			mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$		mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$
α=.01	T_1	4.15	3.55	3.53	3.55	3.55	3.56	3.57	3.57	3.50	3.40	3.40	3.40	3.40	3.41	3.40	3.40
	T_2	1.36	0.75	0.73	0.75	0.75	0.92	0.80	0.82	0.80	0.60	0.59	0.60	0.60	0.66	0.62	0.64
	T_3		0.79	0.81	0.79	0.79	0.80	0.76	0.81		0.53	0.54	0.53	0.53	0.52	0.54	0.53
α=.05	T_1	13.13	12.73	12.71	12.72	12.73	12.67	12.71	12.70	12.67	12.61	12.58	12.60	12.61	12.59	12.61	12.60
	T_2	2.00	1.67	1.65	1.66	1.67	1.70	1.67	1.68	1.80	1.71	1.67	1.70	1.71	1.69	1.72	1.71
	T_3		0.74	0.77	0.74	0.74	0.78	0.76	0.77		0.57	0.59	0.58	0.57	0.53	0.58	0.58
α=.10	T_1	21.03	20.71	20.67	20.70	20.71	20.67	20.69	20.67	20.61	20.50	20.49	20.50	20.50	20.48	20.50	20.50
	T_2	2.89	2.54	2.53	2.53	2.54	2.58	2.54	2.55	2.58	2.36	2.37	2.36	2.36	2.40	2.37	2.39
	T_3		0.65	0.67	0.65	0.65	0.72	0.69	0.70		0.57	0.56	0.57	0.57	0.61	0.59	0.60
α=.30	T_1	36.34	36.37	36.32	36.35	36.37	36.35	36.36	36.36	36.10	36.17	36.16	36.16	36.17	36.16	36.16	36.15
	T_2	3.80	3.81	3.78	3.80	3.81	3.83	3.81	3.81	3.57	3.48	3.47	3.47	3.48	3.47	3.48	3.48
	T_3		0.51	0.55	0.53	0.51	0.53	0.53	0.53		0.47	0.48	0.48	0.47	0.48	0.49	0.48
α=.50	T_1	38.53	38.57	38.52	38.55	38.57	38.50	38.54	38.53	38.36	38.40	38.40	38.40	38.40	38.38	38.40	38.39
	T_2	3.52	3.56	3.54	3.55	3.56	3.54	3.55	3.55	3.12	3.11	3.11	3.11	3.11	3.12	3.11	3.12
	T_3		0.49	0.51	0.50	0.49	0.58	0.50	0.51		0.49	0.50	0.50	0.49	0.52	0.49	0.50
α=.70	T_1	31.59	31.63	31.58	31.61	31.63	31.60	31.61	31.60	31.66	31.72	31.72	31.72	31.72	31.70	31.71	31.71
	T_2	2.78	2.83	2.81	2.82	2.83	2.83	2.82	2.82	2.40	2.44	2.44	2.44	2.44	2.43	2.44	2.44
	T_3		0.50	0.55	0.50	0.50	0.50	0.50	0.49		0.46	0.46	0.46	0.46	0.49	0.46	0.46
α=.90	T_1	15.32	15.39	15.38	15.38	15.39	15.35	15.37	15.36	15.11	15.33	15.33	15.33	15.33	15.30	15.32	15.31
	T_2	1.64	1.43	1.44	1.43	1.43	1.44	1.44	1.44	1.29	1.28	1.28	1.28	1.28	1.27	1.28	1.28
	T_3		0.47	0.47	0.47	0.47	0.46	0.47	0.48		0.30	0.30	0.30	0.30	0.31	0.30	0.31
α=.95	T_1	9.01	9.02	9.02	9.02	9.02	8.98	9.01	9.00	8.78	9.03	9.03	9.03	9.03	9.02	9.03	9.02
	T_2	1.09	0.94	0.95	0.94	0.94	0.96	0.95	0.95	0.87	0.89	0.89	0.89	0.89	0.90	0.89	0.90
	T_3		0.50	0.51	0.51	0.51	0.51	0.51	0.50		0.22	0.22	0.22	0.22	0.22	0.20	0.21
α=.99	T_1	2.60	2.35	2.35	2.35	2.35	2.32	2.34	2.33	2.34	2.37	2.37	2.37	2.37	2.36	2.37	2.37
	T_2	0.73	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.58	0.32	0.32	0.32	0.32	0.33	0.32	0.33
	T_3		0.58	0.58	0.58	0.58	0.64	0.60	0.63		0.36	0.36	0.36	0.36	0.36	0.34	0.36

Note: The AR(1) parameter is $\rho = 0.6$ in Equation (1). See the definition of T_1, T_2, and T_3 in the text, which are computed from 100 Monte Carlo replications.

Panel D. AR(1)-ARCH(0)-Strongly skewed

		R = 200								R = 500							
		J=1	J=50							J=1	J=50						
			mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$		mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$
α=.01	T_1	1.08	1.53	1.53	1.53	1.53	1.54	1.54	1.54	1.04	1.58	1.58	1.58	1.58	1.59	1.59	1.59
	T_2	0.16	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.15	0.22	0.22	0.22	0.22	0.23	0.23	0.23
	T_3		0.02	0.02	0.02	0.02	0.02	0.02	0.02		0.02	0.02	0.02	0.02	0.01	0.02	0.01
α=.05	T_1	4.79	5.64	5.63	5.63	5.64	5.61	5.60	5.60	4.76	5.66	5.65	5.65	5.66	5.69	5.66	5.68
	T_2	0.45	0.68	0.68	0.68	0.68	0.69	0.68	0.68	0.51	0.79	0.79	0.79	0.79	0.81	0.80	0.81
	T_3		0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00
α=.10	T_1	9.25	9.87	9.85	9.85	9.86	9.61	9.77	9.70	9.18	9.63	9.62	9.63	9.63	9.46	9.56	9.52
	T_2	0.92	1.12	1.11	1.12	1.12	1.10	1.11	1.10	0.98	1.15	1.15	1.15	1.15	1.11	1.13	1.12
	T_3		0.01	0.01	0.01	0.01	0.07	0.01	0.03		0.00	0.00	0.00	0.00	0.05	0.01	0.02
α=.30	T_1	25.27	25.31	25.29	25.30	25.31	25.24	25.27	25.26	24.97	25.01	25.01	25.01	25.01	25.01	25.01	25.01
	T_2	2.68	2.68	2.68	2.68	2.68	2.67	2.67	2.67	2.74	2.75	2.75	2.75	2.75	2.75	2.75	2.75
	T_3		0.50	0.50	0.51	0.50	0.56	0.52	0.52		0.46	0.46	0.46	0.46	0.46	0.45	0.45
α=.50	T_1	36.90	36.96	36.94	36.95	36.96	36.90	36.94	36.93	36.33	36.39	36.38	36.39	36.39	36.38	36.38	36.38
	T_2	3.97	3.93	3.92	3.93	3.93	3.94	3.94	3.94	3.93	3.91	3.91	3.91	3.91	3.93	3.92	3.93
	T_3		0.47	0.51	0.49	0.47	0.55	0.51	0.52		0.45	0.43	0.45	0.45	0.46	0.46	0.46
α=.70	T_1	39.40	39.36	39.31	39.33	39.35	39.29	39.33	39.32	38.80	38.88	38.87	38.88	38.88	38.87	38.87	38.87
	T_2	4.40	4.21	4.18	4.20	4.21	4.23	4.22	4.22	4.22	4.13	4.12	4.13	4.13	4.15	4.14	4.14
	T_3		0.56	0.54	0.55	0.56	0.58	0.55	0.56		0.53	0.53	0.53	0.53	0.51	0.51	0.51
α=.90	T_1	24.17	23.68	23.66	23.66	23.68	23.55	23.65	23.62	23.23	23.18	23.16	23.17	23.18	23.12	23.17	23.15
	T_2	2.70	2.58	2.57	2.57	2.58	2.63	2.58	2.59	2.56	2.40	2.37	2.39	2.40	2.39	2.40	2.40
	T_3		0.72	0.73	0.74	0.72	0.81	0.75	0.78		0.59	0.57	0.58	0.59	0.58	0.58	0.59
α=.95	T_1	14.94	14.42	14.39	14.39	14.42	14.37	14.40	14.39	14.14	14.03	14.02	14.03	14.03	14.00	14.02	14.02
	T_2	1.91	1.67	1.63	1.64	1.66	1.68	1.68	1.68	1.72	1.54	1.55	1.54	1.54	1.61	1.56	1.57
	T_3		0.76	0.79	0.78	0.76	0.83	0.79	0.82		0.60	0.62	0.61	0.60	0.59	0.61	0.59
α=.99	T_1	4.35	3.85	3.83	3.84	3.85	3.77	3.84	3.81	3.75	3.66	3.67	3.66	3.66	3.63	3.66	3.65
	T_2	1.23	0.76	0.68	0.72	0.76	0.92	0.83	0.84	0.82	0.56	0.58	0.57	0.56	0.53	0.58	0.58
	T_3		0.76	0.76	0.76	0.76	0.82	0.77	0.83		0.52	0.52	0.52	0.52	0.60	0.59	0.63

Note: The AR(1) parameter is $\rho = 0.6$ in Equation (1). See the definition of T_1, T_2, and T_3 in the text, which are computed from 100 Monte Carlo replications.

Panel E. AR(1)-ARCH(0)-Kurtotic unimodal

		R = 200								R = 500							
		J=1	J=50							J=1	J=50						
			mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$		mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$
α=.01	T_1	3.59	3.16	3.13	3.15	3.16	3.16	3.16	3.15	3.20	3.12	3.12	3.12	3.12	3.10	3.11	3.10
	T_2	1.02	0.73	0.67	0.72	0.73	0.83	0.75	0.77	0.56	0.42	0.42	0.42	0.42	0.42	0.41	0.41
	T_3		0.71	0.73	0.73	0.72	0.71	0.74	0.73		0.57	0.57	0.57	0.57	0.57	0.57	0.57
α=.05	T_1	11.82	11.61	11.57	11.59	11.61	11.55	11.58	11.56	11.61	11.66	11.65	11.65	11.66	11.66	11.66	11.66
	T_2	1.68	1.63	1.57	1.60	1.63	1.59	1.61	1.60	1.44	1.31	1.30	1.30	1.31	1.33	1.32	1.32
	T_3		0.63	0.63	0.64	0.63	0.64	0.63	0.62		0.48	0.48	0.48	0.48	0.45	0.47	0.47
α=.10	T_1	19.33	19.10	19.06	19.08	19.10	19.05	19.08	19.07	18.93	19.08	19.07	19.08	19.08	19.07	19.08	19.07
	T_2	2.44	2.34	2.31	2.32	2.34	2.37	2.34	2.35	2.00	1.95	1.95	1.95	1.95	1.96	1.96	1.96
	T_3		0.58	0.59	0.58	0.58	0.60	0.58	0.60		0.37	0.38	0.38	0.37	0.37	0.38	0.38
α=.30	T_1	32.65	32.88	32.83	32.86	32.87	32.78	32.83	32.81	32.47	32.71	32.70	32.70	32.71	32.69	32.69	32.69
	T_2	3.85	3.79	3.78	3.78	3.79	3.81	3.79	3.80	3.55	3.49	3.48	3.48	3.49	3.51	3.49	3.50
	T_3		0.42	0.44	0.42	0.42	0.45	0.43	0.43		0.35	0.35	0.35	0.35	0.35	0.35	0.34
α=.50	T_1	34.17	34.43	34.40	34.41	34.43	34.28	34.35	34.32	34.31	34.40	34.40	34.40	34.40	34.38	34.39	34.38
	T_2	3.74	3.74	3.74	3.74	3.74	3.71	3.73	3.73	3.79	3.78	3.78	3.78	3.78	3.78	3.78	3.78
	T_3		0.29	0.29	0.29	0.29	0.37	0.32	0.35		0.41	0.40	0.41	0.41	0.44	0.43	0.43
α=.70	T_1	32.75	33.01	32.96	32.99	33.01	32.96	32.98	32.96	32.71	32.93	32.92	32.92	32.93	32.91	32.92	32.92
	T_2	3.67	3.53	3.53	3.53	3.54	3.59	3.55	3.56	4.29	4.02	4.02	4.02	4.02	4.06	4.04	4.05
	T_3		0.38	0.40	0.38	0.38	0.37	0.38	0.39		0.37	0.38	0.37	0.37	0.37	0.38	0.38
α=.90	T_1	19.51	19.37	19.35	19.35	19.37	19.36	19.36	19.36	19.38	19.34	19.33	19.34	19.34	19.33	19.34	19.34
	T_2	2.30	2.18	2.19	2.18	2.18	2.18	2.18	2.17	2.54	2.36	2.35	2.36	2.36	2.37	2.37	2.37
	T_3		0.60	0.63	0.63	0.60	0.60	0.60	0.60		0.55	0.52	0.54	0.55	0.52	0.50	0.50
α=.95	T_1	12.11	11.86	11.85	11.85	11.86	11.81	11.84	11.83	11.84	11.72	11.71	11.71	11.72	11.70	11.71	11.71
	T_2	1.80	1.53	1.53	1.53	1.53	1.56	1.53	1.54	1.83	1.58	1.58	1.58	1.58	1.59	1.59	1.59
	T_3		0.62	0.60	0.62	0.62	0.65	0.63	0.64		0.58	0.58	0.58	0.58	0.57	0.57	0.58
α=.99	T_1	3.76	3.27	3.28	3.27	3.27	3.26	3.27	3.26	3.27	3.14	3.14	3.14	3.14	3.13	3.14	3.13
	T_2	1.12	0.66	0.67	0.66	0.66	0.76	0.67	0.68	0.81	0.57	0.57	0.57	0.57	0.58	0.57	0.57
	T_3		0.73	0.74	0.73	0.73	0.73	0.73	0.74		0.49	0.49	0.49	0.49	0.54	0.51	0.52

Note: The AR(1) parameter is $\rho = 0.6$ in Equation (1). See the definition of T_1, T_2, and T_3 in the text, which are computed from 100 Monte Carlo replications.

Panel F. AR(1)-ARCH(0)-Outlier

		R = 200								R = 500							
		J=1	J=50							J=1	J=50						
			mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$		mean	BMA_1	BMA_5	BMA_R	med	$trim_5$	$trim_{10}$
α=.01	T_1	7.93	5.73	5.72	5.71	5.73	5.79	5.78	5.77	5.91	5.39	5.40	5.39	5.39	5.38	5.38	5.39
	T_2	4.73	2.21	2.08	2.12	2.21	2.43	2.36	2.35	2.67	1.98	2.01	1.99	1.98	2.10	2.00	2.05
	T_3		0.84	0.84	0.86	0.84	0.88	0.86	0.85		0.69	0.69	0.69	0.69	0.74	0.71	0.73
α=.05	T_1	13.41	13.43	13.42	13.42	13.42	13.22	13.36	13.32	12.86	13.13	13.15	13.12	13.13	13.07	13.09	13.08
	T_2	5.34	4.74	4.79	4.75	4.74	4.78	4.79	4.81	5.42	4.81	4.87	4.80	4.82	4.90	4.85	4.87
	T_3		0.49	0.48	0.49	0.49	0.49	0.49	0.49		0.36	0.36	0.36	0.37	0.37	0.36	0.38
α=.10	T_1	15.83	16.18	16.18	16.18	16.17	16.12	16.12	16.11	15.49	16.00	16.00	16.00	16.00	16.00	16.00	16.00
	T_2	5.45	5.30	5.31	5.30	5.30	5.33	5.31	5.32	5.32	5.21	5.23	5.22	5.21	5.21	5.22	5.21
	T_3		0.27	0.26	0.27	0.27	0.29	0.29	0.31		0.17	0.18	0.16	0.17	0.17	0.16	0.15
α=.30	T_1	21.62	21.88	21.85	21.86	21.88	21.69	21.81	21.76	21.34	21.53	21.52	21.52	21.53	21.46	21.49	21.47
	T_2	4.78	4.75	4.74	4.75	4.75	4.70	4.74	4.72	4.66	4.66	4.66	4.66	4.66	4.65	4.65	4.65
	T_3		0.28	0.31	0.31	0.31	0.43	0.31	0.40		0.29	0.29	0.29	0.29	0.33	0.29	0.32
α=.50	T_1	23.19	23.44	23.39	23.42	23.43	23.25	23.35	23.31	23.15	23.22	23.21	23.21	23.22	23.17	23.19	23.19
	T_2	4.48	4.49	4.47	4.48	4.49	4.40	4.45	4.43	4.45	4.50	4.50	4.50	4.50	4.46	4.48	4.48
	T_3		0.27	0.28	0.28	0.27	0.45	0.36	0.39		0.40	0.40	0.40	0.40	0.44	0.41	0.40
α=.70	T_1	21.38	21.66	21.61	21.64	21.65	21.48	21.57	21.53	21.48	21.58	21.57	21.58	21.58	21.53	21.55	21.54
	T_2	4.85	4.80	4.80	4.80	4.80	4.79	4.79	4.78	4.80	4.78	4.78	4.78	4.78	4.77	4.77	4.77
	T_3		0.24	0.25	0.25	0.24	0.39	0.30	0.36		0.31	0.33	0.32	0.32	0.44	0.35	0.38
α=.90	T_1	15.54	15.85	15.83	15.83	15.85	15.76	15.79	15.77	15.54	15.98	15.98	15.98	15.98	15.99	15.98	15.98
	T_2	5.47	5.25	5.25	5.24	5.25	5.33	5.26	5.28	5.34	5.13	5.14	5.13	5.13	5.14	5.14	5.14
	T_3		0.31	0.33	0.31	0.32	0.37	0.35	0.35		0.19	0.21	0.21	0.19	0.21	0.23	0.21
α=.95	T_1	13.01	13.02	12.99	13.00	13.02	12.79	12.92	12.87	13.02	13.22	13.22	13.21	13.21	13.16	13.18	13.17
	T_2	5.12	4.45	4.45	4.46	4.46	4.54	4.50	4.51	5.21	4.65	4.67	4.66	4.66	4.72	4.68	4.70
	T_3		0.47	0.48	0.47	0.49	0.54	0.49	0.51		0.38	0.39	0.38	0.38	0.41	0.40	0.42
α=.99	T_1	7.84	5.36	5.36	5.36	5.36	5.42	5.40	5.45	6.24	5.61	5.64	5.62	5.62	5.65	5.63	5.65
	T_2	4.82	1.87	1.83	1.88	1.87	2.18	1.98	2.14	2.71	2.01	2.05	2.01	2.01	2.12	2.06	2.09
	T_3		0.89	0.89	0.89	0.89	0.94	0.90	0.93		0.71	0.71	0.71	0.71	0.69	0.70	0.73

Note: The AR(1) parameter is $\rho = 0.6$ in Equation (1). See the definition of T_1, T_2, and T_3 in the text, which are computed from 100 Monte Carlo replications.

Table 2. Empirical Applications of Bagging Quantile Prediction

Panel A. USD/EUR Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	5.59	3.51	3.51	3.51	3.51	3.39	4.07	3.94	3.94	3.94	3.94	4.01
α=.05	14.62	12.21	12.23	12.21	12.21	12.01	14.18	13.95	13.94	13.93	13.95	13.77
α=.10	23.35	22.01	22.03	22.01	22.02	21.98	22.95	22.70	22.68	22.69	22.70	22.74
α=.30	37.76	36.58	36.55	36.57	36.58	36.50	36.80	36.10	36.09	36.10	36.10	36.04
α=.50	40.90	40.22	40.20	40.21	40.22	40.06	40.38	39.74	39.73	39.74	39.74	39.70
α=.70	38.43	36.92	36.90	36.93	36.93	36.80	38.29	37.59	37.58	37.59	37.59	37.50
α=.90	23.80	22.53	22.53	22.54	22.53	22.67	23.24	22.86	22.86	22.87	22.86	22.74
α=.95	15.15	13.96	13.99	13.97	13.96	14.09	14.37	14.35	14.35	14.35	14.35	14.32
α=.99	7.18	3.71	3.71	3.71	3.71	3.43	4.32	4.18	4.18	4.18	4.18	4.14

Note: Each cell gives the tick loss of quantile prediction over the period 08/05/04~ 04/11/05.

Panel B. USD/JPY Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	7.20	3.81	3.81	3.81	3.82	3.21	4.36	3.98	3.98	3.98	3.98	4.09
α=.05	15.74	13.68	13.67	13.67	13.68	13.64	15.23	14.41	14.4	14.4	14.41	14.47
α=.10	23.63	22.91	22.89	22.91	22.91	23.17	22.59	22.3	22.3	22.3	22.3	22.17
α=.30	39.43	38.58	38.58	38.58	38.58	38.54	38.12	37.49	37.49	37.49	37.49	37.53
α=.50	40.60	39.86	39.86	39.86	39.86	39.65	40.79	40.83	40.84	40.83	40.83	40.82
α=.70	37.17	36.66	36.65	36.66	36.66	36.23	37.35	38.08	38.08	38.08	38.08	38.16
α=.90	24.66	22.09	22.05	22.09	22.09	21.96	23.06	22.01	22	22.01	22.01	22.07
α=.95	16.10	14.47	14.40	14.46	14.47	14.41	13.91	13.88	13.87	13.88	13.88	13.8
α=.99	6.88	4.18	4.17	4.18	4.18	4.87	3.75	3.64	3.64	3.64	3.64	3.62

Note: Each cell gives the tick loss of quantile prediction over the period 08/05/04~ 04/11/05.

Panel C. Dow Jones Industrial Averages Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	14.17	10.43	10.53	10.47	10.43	9.15	11.30	10.17	10.22	10.18	10.17	10.12
α=.05	38.67	33.43	33.98	33.51	33.46	32.07	39.10	37.60	37.89	37.65	37.60	37.76
α=.10	65.82	58.75	59.25	58.78	58.78	57.27	62.58	59.27	59.34	59.28	59.28	59.41
α=.30	114.68	110.87	111.26	110.84	110.88	110.72	110.42	109.17	109.18	109.18	109.17	109.14
α=.50	129.14	124.29	124.52	124.24	124.33	124.17	125.10	123.48	123.54	123.49	123.48	123.49
α=.70	109.85	108.48	108.59	108.39	108.47	107.43	106.85	106.09	106.25	106.14	106.09	106.36
α=.90	61.60	57.59	58.56	57.82	57.59	56.72	57.81	54.69	55.12	54.83	54.69	54.26
α=.95	37.73	34.09	34.07	34.19	34.11	33.03	32.44	31.45	31.67	31.52	31.45	31.90
α=.99	13.34	14.45	14.18	14.42	14.40	14.11	9.30	8.66	8.40	8.61	8.66	8.31

Note: Each cell gives the tick loss of quantile prediction over the period 01/05/00~12/31/00.

Panel D. New York Stock Exchange Composite Daily Returns

	R = 100						*R* = 300					
	J=1	*J* = 50					*J*=1	*J* = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	11.39	8.43	8.50	8.45	8.43	7.73	9.31	7.85	7.93	7.87	7.85	8.05
α=.05	36.33	28.23	28.56	28.32	28.24	26.65	32.25	29.88	30.08	29.93	29.88	29.20
α=.10	55.37	47.77	48.21	47.86	47.78	46.96	49.57	47.83	47.95	47.86	47.84	47.27
α=.30	94.59	90.80	91.00	90.81	90.82	90.72	91.18	89.62	89.63	89.62	89.63	89.69
α=.50	105.49	102.21	102.37	102.27	102.23	101.42	103.02	102.45	102.56	102.49	102.46	102.67
α=.70	94.18	92.39	92.34	92.37	92.39	91.97	92.73	90.92	91.03	90.96	90.92	91.12
α=.90	56.58	49.43	49.77	49.61	49.46	49.03	52.01	49.05	49.43	49.12	49.05	48.77
α=.95	34.57	30.16	30.54	30.31	30.18	31.27	31.68	29.49	29.71	29.57	29.49	29.73
α=.99	15.57	10.01	10.11	9.86	9.99	9.34	9.34	8.29	8.34	8.31	8.29	8.72

Note: Each cell gives the tick loss of quantile prediction over the period 01/05/00~12/31/00.

Panel E. Standard and Poor's 500 Daily Returns

	R = 100						*R* = 300					
	J=1	*J* = 50					*J*=1	*J* = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	16.54	10.13	10.21	10.16	10.13	7.99	13.18	9.82	9.94	9.86	9.82	9.49
α=.05	43.20	37.60	38.29	37.67	37.62	35.67	36.58	35.53	35.56	35.53	35.53	35.69
α=.10	67.41	62.86	63.10	62.83	62.87	62.66	62.62	58.85	58.84	58.82	58.85	58.58
α=.30	113.91	113.15	113.16	113.09	113.15	112.82	112.75	111.89	111.89	111.88	111.89	112.04
α=.50	135.85	129.14	129.27	129.16	129.17	128.88	131.30	130.63	130.63	130.62	130.63	130.40
α=.70	120.03	117.25	117.23	117.16	117.26	117.11	118.37	116.01	116.04	116.05	116.01	116.35
α=.90	71.74	66.46	66.82	66.39	66.48	67.83	67.71	65.53	65.30	65.49	65.53	65.82
α=.95	46.60	39.49	39.59	39.51	39.50	40.25	39.98	39.30	39.58	39.38	39.31	38.67
α=.99	18.44	13.18	13.49	13.19	13.14	15.19	13.66	11.40	11.45	11.45	11.40	11.92

Note: Each cell gives the tick loss of quantile prediction over the period 01/05/00~12/31/00.

Panel F. NASDAQ Daily Returns

	R = 100						*R* = 300					
	J=1	*J* = 50					*J*=1	*J* = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	40.75	28.79	28.33	28.07	28.75	27.01	30.00	24.83	25.30	24.49	24.84	25.20
α=.05	100.35	88.50	88.61	87.38	88.51	89.28	90.40	83.17	83.21	82.92	83.18	82.97
α=.10	144.47	132.42	132.46	131.75	132.54	129.04	143.22	138.43	138.91	138.48	138.44	136.49
α=.30	274.40	254.31	253.92	255.04	254.56	252.35	277.28	269.61	269.51	269.44	269.64	270.23
α=.50	310.14	294.82	295.50	295.34	295.09	295.45	309.47	303.89	304.19	304.16	303.91	304.09
α=.70	270.05	261.54	261.76	262.32	261.62	261.50	267.21	262.99	262.75	263.18	263.00	263.09
α=.90	153.03	142.64	140.41	142.08	142.73	142.44	143.95	134.93	135.16	135.11	134.96	133.81
α=.95	96.11	83.19	81.37	81.95	83.22	80.51	96.98	84.92	84.67	84.67	84.95	84.00
α=.99	40.68	31.65	30.55	29.94	31.44	34.98	33.91	29.85	28.49	28.56	29.84	28.95

Note: Each cell gives the tick loss of quantile prediction over the period 01/05/00~12/31/00.

Panel G. Russell 2000 Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	39.12	16.12	16.18	16.05	16.14	1605	22.33	16.17	16.28	16.19	16.18	15.84
α=.05	55.79	50.81	50.92	50.51	50.82	48.96	54.96	53.36	53.59	53.24	53.36	52.77
α=.10	94.53	85.34	85.74	85.01	85.37	85.57	92.39	86.78	87.12	86.80	86.79	87.01
α=.30	170.80	161.92	161.54	161.87	162.03	162.53	165.96	165.76	165.72	165.75	165.76	165.63
α=.50	192.64	184.05	184.83	184.58	184.11	182.66	187.40	183.92	184.03	184.00	183.92	183.65
α=.70	165.53	163.11	163.30	163.36	163.11	162.96	164.40	161.70	161.79	161.76	161.70	161.37
α=.90	101.89	86.55	86.96	86.67	86.61	85.78	97.04	91.45	91.44	91.43	91.46	91.76
α=.95	65.35	51.60	51.47	51.50	51.63	50.06	60.60	56.25	56.46	56.30	56.25	55.55
α=.99	29.24	22.36	21.93	21.82	22.31	24.83	21.75	14.49	14.86	14.56	14.51	15.26

Note: Each cell gives the tick loss of quantile prediction over the period 01/05/00~12/31/00.

Panel H. Pacific Exchange Technology Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	30.17	21.15	20.12	20.08	21.13	21.27	26.16	19.83	19.99	19.81	19.83	20.49
α=.05	83.32	70.34	70.04	69.47	70.35	66.23	74.15	71.39	71.67	71.14	71.40	71.44
α=.10	130.88	114.82	115.41	114.58	114.91	115.68	126.60	121.82	122.29	121.77	121.82	120.31
α=.30	257.32	249.79	247.66	249.31	249.99	249.57	256.27	253.02	252.03	252.96	253.04	253.02
α=.50	289.98	283.59	282.72	283.59	283.66	283.96	288.93	282.36	281.78	282.61	282.38	282.32
α=.70	245.96	239.35	238.88	239.53	239.40	239.37	249.67	245.34	245.17	245.49	245.35	245.76
α=.90	145.41	134.41	132.97	134.11	134.39	132.29	134.88	126.13	126.57	126.45	126.17	124.15
α=.95	83.24	79.86	79.50	80.31	79.83	78.72	87.76	79.15	79.28	79.30	79.17	78.42
α=.99	29.51	23.79	23.37	23.35	23.75	25.42	25.86	22.16	22.04	21.97	22.17	23.42

Note: Each cell gives the tick loss of quantile prediction over the period 01/05/00~12/31/00.

Table 3. Bagging Multi-step Quantile Forecast for AR(0)-ARCH(1)-Gaussian Model

R=20, *P*=100		*h*=1			*h*=2			*h*=3			*h*=4		
		J=1	*J*= 50		*J*=1	*J*= 50		*J*=1	*J*= 50		*J*=1	*J*= 50	
			Mean	Med		Mean	Med		Mean	Med		Mean	Med
α=.01	T_1	14.73	9.54	8.72	21.38	12.29	11.49	30.90	14.56	13.37	42.75	15.26	13.81
	T_2	6.90	2.89	2.84	13.25	4.15	3.61	23.99	6.42	5.00	53.20	6.12	4.37
	T_3		0.93	0.96		0.92	0.96		0.97	0.96		0.99	1.00
α=.05	T_1	19.97	14.14	13.39	26.44	17.10	16.32	35.68	18.74	17.82	46.80	19.62	18.47
	T_2	7.11	3.36	3.17	13.16	4.87	3.90	23.54	6.46	4.97	51.51	6.06	4.85
	T_3		0.97	0.98		0.98	0.97		0.98	0.98		0.99	0.99
α=.10	T_1	25.46	19.52	18.66	30.83	22.52	22.05	37.88	23.89	23.06	50.88	25.05	24.05
	T_2	6.75	3.39	3.32	11.72	5.20	4.24	17.48	5.94	4.92	48.86	6.55	5.19
	T_3		0.98	1.00		0.99	1.00		0.99	0.99		1.00	1.00
α=.30	T_1	40.75	34.52	33.55	43.05	36.70	36.36	45.89	37.50	37.00	51.27	38.32	37.78
	T_2	7.03	4.72	4.72	7.90	5.44	5.24	11.32	5.81	5.48	14.26	6.22	5.76
	T_3		1.00	1.00		0.99	0.99		0.98	1.00		1.00	1.00
α=.50	T_1	45.31	38.88	38.08	47.61	40.81	40.67	49.12	41.54	41.11	53.58	42.16	41.70
	T_2	7.12	5.25	5.22	8.27	5.91	5.86	9.02	5.80	5.67	13.66	6.43	6.07
	T_3		1.00	1.00		1.00	1.00		0.99	1.00		1.00	1.00
α=.70	T_1	40.10	34.36	33.38	42.91	36.14	35.89	44.81	37.24	36.83	49.11	37.74	37.21
	T_2	6.32	5.11	4.83	8.27	5.54	5.31	9.67	5.72	5.24	13.28	6.05	5.61
	T_3		0.99	1.00		0.99	0.99		0.99	0.99		0.98	0.99
α=.90	T_1	24.64	19.55	18.70	30.02	21.46	21.12	34.45	23.22	22.45	43.77	23.87	23.00
	T_2	5.18	3.79	3.37	10.98	4.80	4.59	17.75	5.67	4.91	24.49	5.53	5.01
	T_3		0.97	0.99		1.00	0.99		0.99	0.98		1.00	1.00
α=.95	T_1	19.04	14.07	13.17	26.34	16.44	15.54	31.06	18.29	17.27	39.54	18.62	17.72
	T_2	5.48	3.97	3.14	13.45	5.46	4.06	21.97	5.74	4.71	25.76	5.54	4.79
	T_3		0.96	0.97		0.96	0.98		0.95	0.95		1.00	0.98
α=.99	T_1	14.00	9.51	8.49	21.31	11.40	10.63	26.30	13.49	12.35	35.31	14.48	13.07
	T_2	5.42	3.24	2.73	13.56	3.97	3.53	22.49	5.23	4.32	26.23	5.63	4.36
	T_3		0.93	0.94		0.98	0.98		0.92	0.94		0.98	0.99
R=50, *P*=100		*h*=1			*h*=2			*h*=3			*h*=4		
α=.01	T_1	5.74	4.30	4.41	8.16	5.34	5.59	12.22	5.50	5.77	12.16	6.15	6.48
	T_2	2.45	1.39	1.49	6.66	2.32	2.38	41.34	2.53	2.72	17.53	3.10	3.51
	T_3		0.83	0.91		0.89	0.86		0.93	0.87		0.87	0.84
α=.05	T_1	12.31	10.89	10.89	13.98	12.15	12.36	14.68	12.37	12.67	16.44	12.91	13.12
	T_2	2.38	1.81	1.82	3.97	2.86	2.80	4.57	3.13	3.38	8.73	3.68	3.74
	T_3		0.92	0.94		0.89	0.83		0.94	0.89		0.90	0.88
α=.10	T_1	19.27	17.73	17.65	20.51	18.85	18.92	21.00	19.28	19.36	22.20	19.65	19.66
	T_2	3.02	2.57	2.61	4.52	3.25	3.27	5.14	3.81	3.86	6.97	4.29	3.91
	T_3		0.94	0.98		0.89	0.90		0.87	0.85		0.92	0.91
α=.30	T_1	35.74	33.59	33.56	36.02	34.13	34.29	36.10	34.42	34.54	36.36	34.57	34.58
	T_2	4.83	4.26	4.32	5.42	4.41	4.56	5.59	4.68	4.69	5.47	4.82	4.66
	T_3		0.98	0.99		0.94	0.94		0.91	0.88		0.94	0.94
α=.50	T_1	40.62	37.98	37.88	40.25	38.29	38.38	40.27	38.52	38.53	40.03	38.45	38.54
	T_2	5.43	4.42	4.45	6.03	4.56	4.64	5.44	4.72	4.70	5.31	4.67	4.70
	T_3		0.99	1.00		0.90	0.93		0.95	0.96		0.91	0.91
α=.70	T_1	35.56	33.14	33.12	35.75	33.71	33.77	35.59	33.89	33.84	35.61	33.98	34.00
	T_2	5.12	3.91	4.01	6.13	4.28	4.32	5.22	4.23	4.17	4.86	4.44	4.35
	T_3		0.98	0.99		0.94	0.92		0.92	0.93		0.90	0.90
α=.90	T_1	19.19	17.37	17.29	20.87	18.51	18.58	20.91	18.76	18.84	21.39	19.10	19.17
	T_2	3.96	2.63	2.52	6.46	3.50	3.54	5.32	3.29	3.39	7.79	4.56	4.24
	T_3		0.95	0.96		0.95	0.90		0.90	0.89		0.91	0.92
α=.95	T_1	12.15	10.62	10.61	14.29	11.89	12.10	14.89	12.21	12.53	15.27	12.50	12.67
	T_2	3.57	1.92	2.03	6.16	2.88	3.05	5.69	2.96	3.07	5.63	3.69	3.67
	T_3		0.88	0.91		0.90	0.89		0.89	0.85		0.94	0.91
α=.99	T_1	5.49	4.05	4.10	7.66	5.00	5.17	9.32	5.34	5.58	9.19	5.72	6.04
	T_2	3.96	1.49	1.56	5.98	2.16	2.34	9.26	2.50	2.54	6.27	3.14	2.99
	T_3		0.85	0.88		0.89	0.84		0.87	0.87		0.87	0.83

Note: The ARCH parameter is $\theta = 0.5$ as defined in equation (1). The three rows of each multi-step forecast method report the average, the standard error and the frequency of better performance of bagging predictors in terms of tick loss computed from 100 Monte Carlo replications. See the definition of T_1, T_2, and T_3 in the text.

Table 3 (Continued).

R=100, *P*=100		*h*=1			*h*=2			*h*=3			*h*=4		
		J=1	*J*= 50		*J*=1	*J*= 50		*J*=1	*J*= 50		*J*=1	*J*= 50	
			Mean	Med		Mean	Med		Mean	Med		Mean	Med
α=.01	T_1	3.69	3.19	3.22	5.32	3.87	4.20	5.41	3.92	4.23	6.01	4.12	4.40
	T_2	1.58	1.11	1.21	5.43	1.70	2.20	3.37	1.79	2.20	4.68	2.17	2.40
	T_3		0.78	0.83		0.83	0.75		0.82	0.78		0.89	0.86
α=.05	T_1	10.97	10.27	10.31	12.23	11.51	11.65	12.60	11.59	11.75	12.84	11.68	11.79
	T_2	2.23	1.89	1.99	3.22	2.84	2.97	3.63	2.99	3.10	4.33	3.25	3.27
	T_3		0.83	0.91		0.89	0.83		0.90	0.85		0.92	0.88
α=.10	T_1	17.83	17.10	17.12	18.98	18.26	18.32	19.09	18.30	18.38	19.44	18.36	18.40
	T_2	3.27	2.72	2.79	4.15	3.52	3.58	3.88	3.52	3.60	4.69	3.76	3.81
	T_3		0.85	0.87		0.87	0.85		0.89	0.86		0.92	0.92
α=.30	T_1	34.28	33.18	33.17	34.29	33.62	33.65	34.29	33.60	33.61	34.43	33.64	33.63
	T_2	5.12	4.73	4.70	5.41	4.90	4.90	5.26	4.92	4.88	5.47	5.03	5.03
	T_3		0.95	0.95		0.84	0.86		0.84	0.86		0.84	0.86
α=.50	T_1	38.87	37.58	37.59	38.59	37.84	37.86	38.58	37.91	37.94	38.57	37.90	37.93
	T_2	6.13	5.22	5.26	5.68	5.35	5.35	5.63	5.37	5.37	5.60	5.38	5.39
	T_3		0.98	0.99		0.87	0.89		0.91	0.90		0.85	0.83
α=.70	T_1	33.86	32.78	32.77	34.04	33.32	33.33	33.97	33.30	33.32	33.94	33.37	33.40
	T_2	5.53	4.82	4.86	5.30	4.96	4.93	5.48	5.04	5.00	5.20	4.99	5.03
	T_3		0.90	0.94		0.80	0.85		0.88	0.84		0.84	0.84
α=.90	T_1	17.58	16.86	16.83	18.66	17.92	18.00	18.86	18.04	18.07	18.89	18.18	18.22
	T_2	3.13	2.73	2.79	4.10	3.47	3.66	4.41	3.67	3.78	4.31	3.85	3.93
	T_3		0.87	0.89		0.85	0.85		0.84	0.86		0.86	0.87
α=.95	T_1	10.79	10.14	10.16	12.13	11.32	11.42	12.30	11.39	11.52	12.55	11.58	11.71
	T_2	2.43	1.91	2.01	3.58	2.81	2.88	3.56	2.75	2.99	3.90	3.21	3.41
	T_3		0.82	0.90		0.79	0.79		0.88	0.82		0.86	0.87
α=.99	T_1	4.07	3.21	3.28	5.20	3.87	4.13	5.20	3.92	4.21	6.34	4.09	4.43
	T_2	2.75	1.20	1.41	3.42	1.76	2.04	3.02	1.71	2.04	6.29	2.07	2.65
	T_3		0.85	0.89		0.80	0.78		0.84	0.77		0.91	0.90
R=200 , *P*=100		*h*=1			*h*=2			*h*=3			*h*=4		
α=.01	T_1	2.92	2.72	2.73	3.44	3.24	3.30	3.85	3.37	3.41	3.97	3.41	3.48
	T_2	1.04	0.89	1.03	1.51	1.37	1.44	1.95	1.53	1.58	2.48	1.82	1.81
	T_3		0.72	0.73		0.67	0.66		0.82	0.83		0.78	0.78
α=.05	T_1	9.97	9.68	9.72	10.78	10.56	10.60	11.05	10.67	10.68	11.19	10.70	10.75
	T_2	1.83	1.63	1.69	2.31	2.19	2.28	2.76	2.40	2.41	3.07	2.42	2.47
	T_3		0.80	0.83		0.76	0.73		0.75	0.74		0.82	0.80
α=.10	T_1	16.73	16.38	16.35	17.66	17.31	17.33	17.67	17.34	17.39	17.74	17.36	17.39
	T_2	2.73	2.50	2.51	3.21	3.02	3.08	3.38	3.08	3.11	3.60	3.13	3.14
	T_3		0.75	0.82		0.75	0.69		0.67	0.68		0.73	0.77
α=.30	T_1	32.86	32.37	32.37	33.02	32.72	32.73	33.09	32.73	32.74	32.98	32.75	32.77
	T_2	4.51	4.30	4.29	4.63	4.46	4.46	4.70	4.50	4.49	4.73	4.54	4.55
	T_3		0.80	0.83		0.66	0.67		0.74	0.75		0.68	0.65
α=.50	T_1	37.53	37.02	37.04	37.44	37.16	37.18	37.53	37.21	37.24	37.50	37.22	37.24
	T_2	4.90	4.75	4.74	5.08	4.95	4.95	5.16	5.02	5.03	5.03	4.96	4.97
	T_3		0.86	0.88		0.66	0.67		0.81	0.83		0.74	0.79
α=.70	T_1	32.70	32.23	32.23	32.96	32.71	32.73	32.95	32.70	32.72	32.93	32.69	32.73
	T_2	4.42	4.24	4.26	4.69	4.58	4.57	4.73	4.60	4.63	4.68	4.58	4.58
	T_3		0.83	0.87		0.80	0.75		0.74	0.76		0.74	0.71
α=.90	T_1	16.86	16.48	16.53	17.77	17.49	17.48	17.75	17.50	17.51	17.95	17.60	17.61
	T_2	2.58	2.42	2.49	3.31	3.23	3.23	3.51	3.31	3.35	3.50	3.33	3.35
	T_3		0.76	0.80		0.76	0.78		0.74	0.79		0.79	0.82
α=.95	T_1	10.12	9.87	9.88	11.15	10.96	11.02	11.30	10.99	11.03	11.39	11.07	11.06
	T_2	1.74	1.62	1.61	2.56	2.44	2.54	2.97	2.53	2.59	3.05	2.65	2.67
	T_3		0.78	0.80		0.70	0.66		0.72	0.74		0.71	0.75
α=.99	T_1	2.96	2.75	2.77	3.73	3.39	3.48	4.13	3.48	3.61	3.97	3.43	3.54
	T_2	1.05	0.77	0.80	1.54	1.34	1.34	2.53	1.45	1.50	2.00	1.43	1.52
	T_3		0.70	0.73		0.71	0.65		0.82	0.77		0.87	0.85

Table 4. Bagging Quantile Prediction for SP 500 Daily Return Using Tick-exponential Losses

		R = 100						R = 200						R = 300					
		J=1	J=50					J=1	J=50					J=1	J=50				
			mean	BMA_1	BMA_5	BMA_R	med		mean	BMA_1	BMA_5	BMA_R	med		mean	BMA_1	BMA_5	BMA_R	med
α=.01	*tick*	5.59	5.45	5.45	5.45	5.45	5.81	4.66	4.74	4.74	4.74	4.74	4.72	4.78	4.77	4.77	4.77	4.77	4.77
	p=1	5.95	5.30	5.30	5.30	5.30	5.21	4.84	4.81	4.81	4.81	4.81	4.74	4.73	4.84	4.84	4.84	4.84	4.73
	p=2	4.90	4.22	4.22	4.22	4.22	4.22	4.28	4.54	4.43	4.42	4.43	4.22	5.06	5.06	5.01	5.02	5.02	4.84
	p=3	4.51	4.26	4.26	4.26	4.26	4.29	4.95	4.64	4.64	4.64	4.64	4.31	5.44	5.71	5.71	5.71	5.71	4.92
α=.05	*tick*	20.49	19.46	19.46	19.46	19.46	19.75	19.33	18.76	18.76	18.76	18.76	19.09	18.51	18.40	18.40	18.40	18.40	18.45
	p=1	20.61	19.66	19.66	19.66	19.66	19.52	19.36	18.81	18.81	18.81	18.81	19.01	18.48	18.27	18.27	18.27	18.27	18.20
	p=2	19.40	19.16	19.16	19.16	19.16	19.22	18.71	18.63	18.63	18.63	18.63	18.55	19.41	18.26	18.26	18.26	18.26	18.29
	p=3	19.17	19.20	19.20	19.20	19.20	19.19	18.51	18.41	18.41	18.41	18.41	18.53	18.39	18.28	18.28	18.28	18.28	18.23
α=.10	*tick*	33.95	33.08	33.08	33.08	33.08	33.25	33.68	33.03	33.03	33.03	33.03	33.05	32.62	31.95	31.95	31.95	31.95	31.85
	p=1	33.98	33.19	33.19	33.19	33.19	33.26	33.55	32.99	32.99	32.99	32.99	32.98	32.47	31.97	31.97	31.97	31.97	31.86
	p=2	33.60	33.12	33.11	33.12	33.12	33.03	33.13	32.56	32.56	32.56	32.56	32.55	32.14	32.57	32.57	32.57	32.57	32.71
	p=3	33.01	33.20	33.20	33.20	33.20	33.04	32.50	32.53	32.53	32.53	32.53	32.58	32.28	32.15	32.15	32.15	32.15	32.30
α=.30	*tick*	65.93	63.52	63.52	63.52	63.52	63.90	63.51	62.66	62.66	62.66	62.66	62.64	63.18	62.84	62.84	62.84	62.84	62.72
	p=1	65.23	63.60	63.60	63.60	63.60	63.85	63.46	62.65	62.65	62.65	62.65	62.62	63.02	62.87	62.87	62.87	62.87	62.72
	p=2	64.00	62.93	62.93	62.93	62.93	62.71	62.69	62.74	62.74	62.74	62.74	62.80	62.84	62.72	62.72	62.72	62.72	62.90
	p=3	63.60	63.65	63.65	63.65	63.65	63.56	63.13	63.04	63.04	63.04	63.04	62.97	62.88	62.80	62.80	62.80	62.80	62.74
α=.50	*tick*	69.99	68.66	68.66	68.66	68.66	68.65	68.74	68.12	68.12	68.12	68.12	68.02	68.48	67.69	67.69	67.69	67.69	67.65
	p=1	70.03	68.66	68.66	68.66	68.66	68.67	68.65	68.17	68.17	68.17	68.17	68.16	68.39	67.70	67.70	67.70	67.70	67.55
	p=2	68.81	68.25	68.25	68.25	68.25	68.26	68.55	68.05	68.05	68.05	68.05	68.18	68.85	67.91	67.91	67.91	67.91	67.88
	p=3	74.22	73.62	73.62	73.62	73.62	73.87	72.07	71.93	71.93	71.93	71.93	72.03	71.44	71.79	71.79	71.79	71.79	71.89
α=.70	*tick*	59.81	59.85	59.85	59.85	59.85	59.87	60.08	59.49	59.49	59.49	59.49	59.41	60.54	60.10	60.10	60.10	60.10	60.20
	p=1	60.44	59.84	59.84	59.84	59.84	59.99	60.10	59.48	59.48	59.48	59.48	59.44	60.52	60.08	60.08	60.08	60.08	60.08
	p=2	59.48	59.55	59.55	59.55	59.55	59.69	59.59	59.77	59.77	59.77	59.77	59.96	60.39	59.60	59.60	59.60	59.60	59.60
	p=3	69.18	64.53	64.53	64.53	64.53	67.42	65.21	60.55	60.55	60.55	60.55	60.79	68.61	59.50	59.50	59.50	59.50	61.56
α=.90	*tick*	32.37	31.95	31.95	31.95	31.95	32.13	32.19	31.45	31.45	31.45	31.45	31.34	33.31	32.75	32.75	32.75	32.75	32.62
	p=1	31.55	31.91	31.91	31.91	31.91	32.07	31.64	31.43	31.43	31.43	31.43	31.33	33.29	32.74	32.74	32.74	32.74	32.64
	p=2	31.72	31.87	31.87	31.87	31.87	32.03	32.48	31.93	31.93	31.93	31.93	32.07	33.32	32.78	32.78	32.78	32.78	32.74
	p=3	32.68	31.83	31.83	31.83	31.83	32.19	33.71	31.11	31.11	31.11	31.11	31.77	33.25	31.69	31.69	31.69	31.69	32.28
α=.95	*tick*	19.68	18.94	18.94	18.94	18.94	19.36	18.99	18.84	18.84	18.84	18.84	18.80	20.03	19.40	19.40	19.40	19.40	19.33
	p=1	19.87	18.82	18.82	18.82	18.82	19.26	18.94	18.92	18.92	18.92	18.92	18.84	19.89	19.38	19.38	19.38	19.38	19.36
	p=2	18.57	18.43	18.43	18.43	18.43	18.23	19.01	18.55	18.55	18.55	18.55	18.40	19.74	19.45	19.45	19.45	19.45	19.45
	p=3	18.02	18.41	18.41	18.41	18.41	18.25	18.58	18.99	18.99	18.99	18.98	18.44	20.96	19.28	19.28	19.28	19.28	19.44
α=.99	*tick*	5.44	5.33	5.33	5.33	5.33	5.52	4.75	4.41	4.41	4.41	4.41	4.42	5.96	5.30	5.30	5.30	5.30	5.46
	p=1	5.19	5.38	5.39	5.38	5.38	5.49	4.44	4.45	4.45	4.45	4.45	4.44	5.95	5.52	5.52	5.52	5.52	5.60
	p=2	4.66	4.69	4.69	4.69	4.69	4.55	4.68	4.34	4.34	4.34	4.34	4.22	6.20	5.69	5.69	5.69	5.69	5.42
	p=3	4.52	4.26	4.25	4.26	4.25	4.16	5.74	4.98	4.97	4.97	4.97	4.23	11.94	5.61	5.55	5.55	5.54	4.85

Note: The four rows for each α report the tick loss of quantile predictors estimated by four different tick-exponential loss function as defined in text. The out-of-sample evaluation period is 01/13/2004 ~ 01/07/2005.

Table 5. Bagging Quantile Prediction for SP500 Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		mean	BMA_1	BMA_5	BMA_R	Med		mean	BMA_1	BMA_5	BMA_R	Med
α=.01	6.72	5.04	5.01	5.03	5.04	4.59	4.71	4.80	4.80	4.80	4.80	4.62
α=.05	20.63	19.37	19.28	19.33	19.36	19.13	18.66	18.41	18.42	18.41	18.41	18.24
α=.10	34.69	31.95	31.95	31.95	31.95	31.47	32.64	32.00	32.00	32.00	32.00	32.00
α=.30	66.22	62.68	62.84	62.75	62.70	62.64	63.28	61.72	61.70	61.72	61.72	61.77
α=.50	70.19	68.54	68.6	68.59	68.55	68.61	68.40	67.96	67.95	67.97	67.96	67.96
α=.70	59.95	59.54	59.51	59.53	59.54	59.16	60.45	60.19	60.19	60.19	60.19	60.22
α=.90	31.96	31.38	31.39	31.39	31.38	31.33	33.15	32.40	32.41	32.40	32.40	32.37
α=.95	19.98	18.38	18.41	18.4	18.38	18.86	19.91	19.69	19.68	19.68	19.69	19.57
α=.99	5.36	5.55	5.56	5.56	5.55	6.01	5.97	5.40	5.40	5.40	5.40	5.41

Note: Each cell gives the tick loss of quantile prediction over the period 01/13/2004 ~ 01/07/2005.

Table 6. Bagging Quantile Predictions for SP500 Monthly Returns with Different Regression Models

			R=10	*R*=20	*R*=30	*R*=40	*R*=50	*R*=60	*R*=70	*R*=80	*R*=90	*R*=100
α=.10	PN	*J*=1	176.40	148.05	127.88	110.75	128.19	104.49	108.95	106.48	118.67	109.68
		J=50	155.54	113.05	111.78	96.33	98.96	95.82	88.37	99.42	103.92	99.36
	NN	*J*=1	147.72	103.77	107.56	106.82	96.68	73.58	76.52	63.29	81.39	76.71
		J=50	106.78	95.21	104.30	100.84	92.39	71.48	65.61	66.29	73.10	69.59
α=.30	PN	*J*=1	259.51	215.73	197.99	190.70	182.65	187.37	190.61	192.54	188.85	187.91
		J=50	222.54	163.38	179.19	179.46	182.48	180.13	178.97	181.44	182.24	183.60
	NN	*J*=1	250.14	190.15	188.59	170.97	159.59	139.24	134.02	132.63	133.01	131.44
		J=50	176.34	163.32	165.44	158.73	150.32	130.68	122.14	127.02	132.52	129.04
α=.50	PN	*J*=1	285.33	221.18	201.47	203.87	205.49	206.28	204.95	208.30	208.57	201.01
		J=50	230.49	161.15	173.50	190.35	187.59	185.38	196.26	196.06	188.81	193.78
	NN	*J*=1	238.24	212.11	201.32	178.79	154.36	146.44	146.73	143.45	145.34	142.54
		J=50	186.58	179.40	172.79	168.43	158.84	139.33	131.57	136.56	140.16	136.60
α=.70	PN	*J*=1	235.22	167.92	171.43	183.12	171.77	169.55	170.17	165.58	171.91	167.27
		J=50	170.15	170.60	159.98	157.00	167.01	166.43	163.22	167.65	157.78	167.29
	NN	*J*=1	206.36	158.48	160.24	140.66	144.30	125.15	115.57	124.37	126.95	115.15
		J=50	152.51	150.00	144.50	138.59	133.36	118.25	109.09	114.22	116.41	110.17
α=.90	PN	*J*=1	224.54	105.44	103.49	102.94	86.34	82.77	80.88	80.34	82.36	77.69
		J=50	164.33	60.45	81.57	83.29	75.61	78.54	76.53	75.60	75.68	76.43
	NN	*J*=1	141.10	101.57	90.12	84.99	78.12	69.39	67.28	67.45	61.12	56.61
		J=50	85.78	79.45	79.51	78.07	72.02	62.39	62.00	62.40	59.62	55.13

Note: The four rows for each α report the tick losses of quantile predictors of SP500 monthly returns over the period November 1995 ~ February 2004 using polynomial (PN), mean bagging PN, neural network (NN), and mean bagging NN predictors.

Table 7. Bagging Binary Prediction for SP500 Daily Returns

	R = 100						R = 300					
	J=1	J = 50					J=1	J = 50				
		Mean	BMA_1	BMA_5	BMA_R	Med		Mean	BMA_1	BMA_5	BMA_R	Med
α=.01	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.37
α=.05	6.85	6.85	6.85	6.85	6.85	6.85	6.85	6.85	6.85	6.85	6.85	6.85
α=.10	13.70	13.70	13.70	13.70	13.70	13.70	13.70	13.70	13.70	13.70	13.70	13.70
α=.30	43.60	40.80	41.10	40.80	40.80	40.80	41.20	41.10	41.10	41.10	41.10	41.10
α=.50	60.50	58.00	58.00	58.00	59.00	58.00	57.00	56.00	56.00	55.50	56.00	56.00
α=.70	35.70	34.30	34.30	34.30	34.30	34.30	33.90	33.90	33.90	33.90	33.90	33.90
α=.90	11.30	11.30	11.30	11.30	11.30	11.30	11.30	11.30	11.30	11.30	11.30	11.30
α=.95	5.65	5.65	5.65	5.65	5.65	5.65	5.65	5.65	5.65	5.65	5.65	5.65
α=.99	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13	1.13

Note: Each cell reports the asymmetric binary prediction loss with parameter α over the period 01/13/2004 ~ 01/07/2005.