

Article

Forecasting Value-at-Risk Using High-Frequency Information

Huiyu Huang¹ and Tae-Hwy Lee^{2,*}

¹ Grantham, Mayo, Van Otterloo and Company LLC, 2150 Shattuck Ave, Suite 900, Berkeley, CA 94704, U.S.A.; E-Mail: evelyn.hhuang@gmail.com

² Department of Economics, University of California, Riverside, CA 92521-0427, USA

* Author to whom correspondence should be addressed; Phone: +1 (951) 827-1509.
E-Mail: taelee@ucr.edu.

Received: 7 March 2013 / Accepted: 22 April 2013 / Published: 21 June 2013

Abstract: in the prediction of quantiles of *daily* Standard & Poor's 500 (S&P 500) returns we consider how to use high-frequency 5-minute data. We examine methods that incorporate the high frequency information either indirectly, through combining forecasts (using forecasts generated from returns sampled at different intraday interval), or directly, through combining high frequency information into one model. We consider subsample averaging, bootstrap averaging, forecast averaging methods for the indirect case, and factor models with principal component approach, for both direct and indirect cases. We show that in forecasting the daily S&P 500 index return quantile (Value-at-Risk or VaR is simply the negative of it), using high-frequency information is beneficial, often substantially and particularly so, in forecasting downside risk. Our empirical results show that the averaging methods (subsample averaging, bootstrap averaging, forecast averaging), which serve as different ways of forming the ensemble average from using high-frequency intraday information, provide an excellent forecasting performance compared to using just low-frequency daily information.

Keywords: VaR; Quantiles; Subsample averaging; Bootstrap averaging; Forecast combination; High-frequency data

1. Introduction

Due to increasing fragility in financial markets and the extensive use of derivative products, effective management of financial risks has become ever more important. A risk measurement tool, namely

“Value-at-Risk” (VaR), has received great attention from both regulators and the academic world due to its simplicity, despite its deficiency in that it lacks coherence in the sense of [1]. Numerous papers have studied various aspects of VaR methodology. Typically the VaR is computed using *daily* frequency, such as for the 1 to 10-day forecasts of the tail quantiles. In this paper, we discuss how we can improve the accuracy of daily out-of-sample VaR forecasts by using high-frequency intraday information.

Consider a financial return series $\{r_t\}_{t=1}^T$, generated by the probability law $\Pr(r_t \leq r | \mathcal{F}_{t-1}) \equiv F_t(r)$ conditional on the information set $\mathcal{F}_{t-1}$ (σ -field) at time $t-1$. Suppose $\{r_t\}$ admits the stochastic process

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t, \quad (1)$$

where $\mu_t = E(r_t | \mathcal{F}_{t-1})$, $\sigma_t^2 = E(\varepsilon_t^2 | \mathcal{F}_{t-1})$, and $\{z_t\} \equiv \{\varepsilon_t / \sigma_t\}$ has the conditional distribution function $G_t(z) \equiv \Pr(z_t \leq z | \mathcal{F}_{t-1})$. The VaR with a given tail probability $\alpha \in (0, 1)$ is defined as the negative of conditional quantile (denoted $q_t(\alpha)$)

$$\Pr(r_t \leq q_t(\alpha) | \mathcal{F}_{t-1}) = F_t(q_t(\alpha)) = \alpha. \quad (2)$$

The quantile can be estimated either from inverting the distribution function

$$q_t(\alpha) = F_t^{-1}(\alpha) = \mu_t + \sigma_t G_t^{-1}(\alpha), \quad (3)$$

where the second equality holds if $F_t(r)$ belongs to location-scale family, or from the quantile regression. A quantile model involves the specification of the conditional distribution $F_t(\cdot)$ or the quantile regression function $q_t(\alpha)$. The former can also be estimated by its components, namely, μ_t , σ_t^2 , $G_t(\cdot)$. For instance, [2] makes distributional assumptions for future daily returns coupled with forecasts of realized volatility to get quantile forecasts. In this paper we focus on the latter, the quantile regression with the high frequency information incorporated.

The rich dynamics in ultra-high-frequency financial data have been used extensively in estimating quadratic variation (integrated variance) in so-called realized volatility literature, and also used in high-frequency duration models. There is a considerable amount of literature addressing the use of high-frequency data for realized volatility and duration models, whereas little work has been done for quantiles, which is practically more relevant and central to financial risk management. This paper contributes to the literature by examining whether it pays to incorporate the intraday data and, more importantly, how to incorporate them to achieve better performance for forecasting daily quantiles (VaR is simply the negative of it).

The existing literature on high-frequency data is mainly devoted to volatility issues: estimating quadratic variation, mixed data sampling (MIDAS) approach of [3], and the autoregressive conditional duration models of [4]. In this paper we use the high-frequency data for a different purpose. Motivated by the bagging (bootstrap aggregating) approach of [5], which replicates the true distribution by bootstrap distribution in computing the ensemble average, we consider returns computed from high frequency intraday observations as multiple replications of the true distribution of daily returns. Suppose there are hourly time series available while we are interested in generating daily forecasts (e.g., 1 day ahead or 10 days ahead). Assuming that each group of hourly information is generated from an identical distribution (without intraday pattern or diurnal cycles), we may consider the high-frequency (hourly) series as

multiple replications of lower frequency (daily) series. In that context, we regard the high-frequency data as *subsamples* of daily series taken at different times within a day.

Viewing the high-frequency information in this manner leads us to consider the subsample averaging idea from the realized volatility (RV) literature and use it for forecasting quantiles instead. The original subsampling idea can be traced back to [6] and further studied by [7, 8]. References [9, 10] establish that RV, defined as the sum of squared intraday returns of small intervals, is an asymptotically unbiased estimator of the unobserved quadratic variation as the interval length approaches zero. However, in the presence of market microstructure noise, this helpful property of RV is contaminated. Recent innovative works investigating this issue include [8,11-15]. When the observed price process is the true underlying price process plus microstructure noise, it is shown that RV will be overwhelmed by the noise and explodes when the sampling frequency approaches infinity. Therefore, it may be optimal to sample less frequently than is the case in the absence of noise. Reference [6] proposes for the first time an unbiased data-driven estimator of volatility and a subsample averaging volatility estimator. References [8, 15] establish improved estimators for quadratic variation through subsampling. The bias-adjusted estimator of [8] based on the subsample averaging method is able to eventually push the estimation bias to zero. Reference [15] shows that subsampling is highly advantageous for RV estimators based on discontinuous kernels.

Motivated by this subsampling approach, which is shown to outperform those using directly the highest frequency series, and to avoid arbitrariness in choosing sampling frequencies, we propose a subsample averaging quantile forecast similar in construction to the bagging approach and to the simple average combination of forecasts. We also compare this approach with other forecasts utilizing the highest-frequency information directly. There is a vast amount of literature on predicting daily market returns using daily close data only. We deem that proper use of high-frequency intraday data should not only be helpful for achieving a more accurate estimation of volatility, but should also be beneficial for forecasting daily VaR/quantiles. The question is how to incorporate the vast amount of intraday high-frequency information for daily low-frequency modeling or forecasting of VaR/quantiles.

In [16], the relative advantages of the combination of forecasts (CF) over the combination of information (CI) are discussed. Under circumstances of highly correlated predictors, important predictors omitted, and/or low signal-to-noise ratio, CF is more likely to win over CI. In this paper, we proceed by considering these two alternate approaches of using intraday information and compare the CF approaches (combination of individual forecasts obtained from using one 5-minute intraday information block at a time) with the CI approaches (combination of all 5-minute intraday information blocks into one model) for their forecasting ability for daily market return VaR/quantiles prediction. It is well-known that the mean of market daily return is very hard to predict, whereas its quantiles may be predictable, particularly in tails. See [17] for some evidence collected by using bagging. Practically, quantile forecasts are very important for risk management purposes. See, for example, [18]. Therefore, we implement the quantile forecasts using high-frequency data through various CF and CI methods. To avoid being overly parameterized, we consider subsample averaging, some CF methods, and factor models with principal component approach. We compare their prediction errors and find that in daily S&P 500 return quantile forecasts, generally CF approaches with simple weighting schemes, subsample averaging in particular, are found to be superior to others.

The rest of the paper is organized as follows. Section 2 describes the data and its organization. Section 3 discusses the quantile forecasting methods we consider. Section 4 presents out-of-sample VaR/quantile forecasting results. Section 5 concludes.

2. Data

The data we use consists of S&P 500 index values at 5-minute intervals recorded between 9:35 a.m. and 4:00 p.m. from 9 June, 1997 to 30 May, 2003, a total of 1,501 days and 117,078 observations.¹ In cleaning the data, those periods of market closings are treated as no variation in index values, thus there exist 78 ticks each trading day. From this pool of 5-minute index data, we construct 78 “daily” returns, each having a time span of twenty-four hours but pointing at different times in a day. The sampling frequency used in [9] is also 5-minute when they analyze the distributional properties of RV. Many RV papers using high-frequency data have justified the use of 5-minute frequency.

Specifically, let us denote the daily close return by $r_t^{(0)} = p_t^{4:00pm} - p_{t-1}^{4:00pm}$, where $p_t^{4:00pm}$ denotes the logarithm of the S&P 500 index value at 4:00 p.m. on day t . $\{p_t^{4:00pm}\}_{t=1}^T$ is one subsample of the entire S&P 500 5-minute index data, obtained by systematic sampling at 4:00 p.m. each day, hence it represents one 78th of the entire 5-minute high-frequency information. Therefore, $r_t^{(0)}$ is the daily return based on this particular subsample, which is also the conventional daily close return.

We define a subsample daily return, $r_t^{(1)} = p_t^{9:35am} - p_{t-1}^{9:35am}$, and similarly other subsample daily returns $r_t^{(j)}$ for $j = 2, \dots, m = 77$, and we have $r_t^{(m=77)} = p_t^{3:55pm} - p_{t-1}^{3:55pm}$.

3. Forecasting Quantiles Using High-Frequency Information

In this section we describe various methods of using high-frequency information in quantile forecasting. These include Daily-Close quantile forecasts, Bagging-Daily-Close quantile forecasts, Subsample-Averaging (SA) quantile forecasts, Combining-Forecast (CF) quantile forecasts, Combining-Information (CI) quantile forecasts, and some variations of SA, CF, CI methods, as presented in Table 1.

3.1. Daily Close Quantile Forecasts

Suppose our objective is to predict the α -quantile of daily close return $r_{T+h}^{(0)}$ conditioning on the information up to time T : $Q_\alpha(r_{T+h}^{(0)}|X_T)$, where h denotes the forecast horizon. Typically, $X_T \subset \mathcal{F}_T$ is the vector of past values of daily returns in the same daily frequency as $\{r_t^{(0)}\}_{t=1}^T$. As in [17] we use a quantile regression model

$$Q_\alpha(r_{t+h}^{(0)}|X_t) = X_t^{(0)'}\beta^{(0)}(\alpha) + \varepsilon_{t+h}, \quad (4)$$

to predict the quantiles of daily S&P 500 return with $h = 1$, using [18]’s linear polynomial model (with the quadratic term capturing the volatility clustering effect on quantiles) such that

$$X_t^{(0)} = \left(1 \ r_t^{(0)} \ r_t^{(0)2}\right)'. \quad (5)$$

¹We are grateful to George Jiang who generously shared this high-frequency intra-day data with us. The data are extracted from the contemporaneous index levels recorded with the quotes of SPX options from the CBOE.

While more complicated nonlinear models such as the artificial neural network models may be considered, we follow [17] where the linear polynomial model is successfully used. The quantile forecast $\hat{Q}_\alpha \left(r_{T+h}^{(0)} | X_T \right) = X_T^{(0)'} \hat{\beta}^{(0)}(\alpha)$ from this quantile regression model is denoted “Daily-Close” in Table 1, which is estimated by minimizing the “tick” (or check) function of [19]

$$\rho_\alpha(\varepsilon) = [\alpha - 1(\varepsilon < 0)] \varepsilon. \quad (6)$$

To examine whether it pays to incorporate the intraday data, we use this Daily-Close as the benchmark model to be compared with other models incorporating intraday information in our empirical analysis in Section 4.

3.2. CI Quantile Forecasts

Besides the past realization of daily close return $r_t^{(0)}$, now we have other subsamples $r_t^{(j)}$, $j = 1, \dots, m$, constructed from the high-frequency intraday data. For the purpose of accommodating more information, it is natural to expand X_t to include the rich dynamics in all other intraday subsamples $\left(r_t^{(0)} r_t^{(1)} \dots r_t^{(m)} \right)$. As an example, for the linear polynomial model, to utilize high frequency information, the new regressors will be

$$X_t^m = \left(1 \ r_t^{(0)} \ r_t^{(0)^2} \ r_t^{(1)} \ r_t^{(1)^2} \ \dots \ r_t^{(m)} \ r_t^{(m)^2} \right)'. \quad (7)$$

By doing this, one follows the CI scheme. That is, combining all the high-frequency information into one model altogether to generate an ultimate forecast:

$$\hat{Q}_\alpha^m \left(r_{T+h}^{(0)} | X_T^m \right) = X_T^{m'} \hat{\beta}_m(\alpha), \quad (8)$$

where the coefficient $\hat{\beta}_m(\alpha)$ is estimated by running the quantile regression of $r_t^{(0)}$ on X_{t-h}^m . Denote this quantile forecast “CI-Unrestricted” in Table 1.

Now re-define $X_t^m = \left(r_t^{(0)} r_t^{(0)^2} r_t^{(1)} r_t^{(1)^2} \dots r_t^{(m)} r_t^{(m)^2} \right)'$ excluding the constant term. We note that this is a fairly large vector with dimension 156 ($m = 77$ in our empirical study). Moreover, consecutive subsample returns, $r_t^{(j)}$ and $r_t^{(j+1)}$, may be highly correlated. Therefore methods that can mitigate high dimension and multicollinearity problems are desirable in order to improve out-of-sample forecast performance. We consider factor model with principal component approach (see [20, 21]). The procedure here starts by assuming that $\left(X_t^m, r_{t+h}^{(0)} \right)$ admits a factor model representation with k common latent factors f_t

$$X_t^m = \Lambda f_t + e_t, \quad (9)$$

$$Q_\alpha(r_{t+h}^{(0)} | X_t^m) = (1 \ f_t') \hat{\beta}_f(\alpha) + \varepsilon_{t+h}, \quad (10)$$

and $\hat{\beta}_f(\alpha)$ is obtained by running quantile regression of $r_t^{(0)}$ on $\left(1 \ f_{t-h}' \right)'$. This quantile forecast is denoted as “CI-PC” in Table 1.²

²To better comprehend this factor representation, if we think of the stock market as a vast pool of information and that price is determined by aggregate behavior of market participants, there may exist a few fundamental shocks to the market in a given day that ultimately determine the daily return at close. This small set of shocks (main forces) may be captured by the latent factors in the factor model Equation (9), or technically, by a few principal components of the big explanatory variable set X_t^m , which contains as large as 156 variables capturing levels and volatilities of the market throughout the trading day.

If k is unknown, it can be estimated by some conventional information criteria (IC) or ones tailored to quantile forecasting models. Reference [22] provides variants of the Schwarz Information Criterion for model selection in M-estimation, which include quantile regression as a special case. According to [23], for median regression, the criterion $SIC = \ln(\hat{\zeta}_T) + \frac{1}{2}p \ln T$, where $\hat{\zeta}_T \equiv \frac{1}{T-h} \sum_{t=1}^{T-h} \rho_\alpha(\varepsilon_{t+h})$, $\alpha = 0.5$, $\rho_\alpha(\cdot)$ is the tick loss function in Equation (6), and p is the dimension of the model, yields a consistent model selection procedure under certain assumptions. To our knowledge, however, a consistent model selection procedure and corresponding information criterion for *factor* quantile regressions have not been established yet. We leave the theoretical study of this procedure for our future research. In our empirical study, we consider conventional information criteria such as AIC and BIC where the estimated number of factors k is selected by $\min_{1 \leq k \leq k_{max}} IC_k = \ln(SSR(k)/T) + g(T)k$, where k_{max} is the hypothesized upper limit for the true number of factors and SSR is the sum of squared residuals. Here we simply extend the formulas for AIC (with $g(T) = 2/T$) and BIC (with $g(T) = \ln T/T$) straight into the quantile case, but substitute “residual” in the above-mentioned term SSR by tick loss $\left[\alpha - \mathbf{1} \left(r_{t+h}^{(0)} < \hat{Q}_\alpha \left(r_{t+h}^{(0)} | X_t^m \right) \right) \right] \left(r_{t+h}^{(0)} - \hat{Q}_\alpha \left(r_{t+h}^{(0)} | X_t^m \right) \right)$, for $t = 1, \dots, T-h$. Alternatively, we can fix k ex-ante at some small values such as 1, 2, or 3.

3.3. CF Quantile Forecasts

The conventional CF methodology with one fixed forecast target, i.e., the α -quantile of $r_{T+h}^{(0)}$, is implemented as follows.

1. CF Step 1: Compute quantile forecasts from regressing $r_t^{(0)}$ on each individual sub-sample,

$$\hat{Q}_\alpha^{(j)} \left(r_{T+h}^{(0)} | X_T^{(j)} \right) = X_T^{(j)'} \hat{\beta}^{(j)}(\alpha), \quad (11)$$

where $X_t^{(j)} = \left(1 \ r_t^{(j)} \ r_t^{(j)2} \right)'$, and $\hat{\beta}^{(j)}(\alpha)$ is obtained by estimating the quantile regression of $r_t^{(0)}$ on $X_{t-h}^{(j)}$, for $j = 0, 1, \dots, m$.

2. CF Step 2: Combine the quantile forecasts from Step 1 by some weighting methods for forecast combination. A simplest example of the combination methodology is the simple average:

$$\bar{Q}_\alpha^{CF-Mean} \left(r_{T+h}^{(0)} | X_T^m \right) = \frac{1}{m+1} \sum_{j=0}^m \hat{Q}_\alpha^{(j)} \left(r_{T+h}^{(0)} | X_T^{(j)} \right), \quad (12)$$

i.e., taking the mean of all the individual quantile forecasts (denoted “CF-Mean” in Table 1). One can also use the median of the individual quantile forecasts as a combined forecast (denoted “CF-Median” in Table 1).

In Step 1, given the high-frequency 5-minute data we have, there are in total 78 individual quantile forecasts to be combined. Besides the simple methods such as combining these 78 forecasts by simple average or median without estimating any weighting parameters, we also use the regression approach forecast combination [24] to explore more of the cross-sectional information contained in these 78 individual forecasts. Obviously, this regression approach is not working well because of the high

dimensionality. Therefore, we consider the principal component methodology for combining the quantile forecasts generated in Step 1 and form the VaR forecast (denoted “CF-PC” in Table 1) by the following factor model of the quantile forecasts

$$\hat{\mathbf{Q}}_{\alpha}^m \left(r_t^{(0)} | X_{t-h}^m \right) = \Gamma \hat{F}_t + v_t, \quad (13)$$

where

$$\hat{\mathbf{Q}}_{\alpha}^m \left(r_t^{(0)} | X_{t-h}^m \right) = \left[\hat{Q}_{\alpha}^{(0)} \left(r_t^{(0)} | X_{t-h}^{(0)} \right), \hat{Q}_{\alpha}^{(1)} \left(r_t^{(0)} | X_{t-h}^{(1)} \right), \dots, \hat{Q}_{\alpha}^{(m)} \left(r_t^{(0)} | X_{t-h}^{(m)} \right) \right]'. \quad (14)$$

Then we estimate $\hat{\beta}_F(\alpha)$ from

$$Q_{\alpha}^{CF-PC} \left(r_{t+h}^{(0)} | X_t^m \right) = \left(1 \ \hat{F}_{t+h}' \right) \beta_F(\alpha) + \eta_{t+h}, \quad (15)$$

to obtain the final CF-PC quantile forecast

$$\hat{Q}_{\alpha}^{CF-PC} \left(r_{T+h}^{(0)} | X_T^m \right) = \left(1 \ \hat{F}_{T+h}' \right) \hat{\beta}_F(\alpha). \quad (16)$$

The size of the common factor set $\hat{F}_t$, i.e., the number of principal components of $\hat{\mathbf{Q}}_{\alpha}^m \left(r_t^{(0)} | X_{t-h}^m \right)$, can be determined again by information criteria AIC or BIC as discussed in Section 3.2, or fixed ex-ante at some small values such as 1, 2, or 3.

3.4. Subsample-Averaging Quantile Forecasts

Inspired from the RV literature and as discussed in Section 1, here we propose a new methodology that extends the subsample idea for RV estimation (using intraday high-frequency information) to daily quantile forecasting. We call it *subsample average* (SA), which is unique to forecasting using high-frequency data. This approach is similar to the CF-Mean procedure discussed in Section 3.3 but it differs from CF-Mean in Step 1: instead of fixing left-hand-side (LHS) variable in Equation (11) as the 4 p.m. daily close return $r_t^{(0)}$, we modify Equation (11) by setting LHS variable as the corresponding subsample return $r_t^{(j)}$ for each $j = 0, 1, \dots, m$.

1. SA Step 1:

$$\tilde{Q}_{\alpha}^{(j)} \left(r_{t+h}^{(j)} | X_t^{(j)} \right) = X_t^{(j)'} \tilde{\beta}^{(j)}(\alpha), \quad (17)$$

which give the SA quantile forecasts: $\tilde{Q}_{\alpha}^{(j)} \left(r_{T+h}^{(j)} | R_T^{(j)} \right) = X_T^{(j)'} \tilde{\beta}^{(j)}(\alpha)$ for $j = 0, 1, \dots, m$.

2. SA Step 2: Taking simple average, we get the quantile forecast

$$\bar{Q}_{\alpha, T+h}^{SA-Mean} \left(r_{T+h}^{(j)} | X_T^m \right) = \frac{1}{m+1} \sum_{j=0}^m \tilde{Q}_{\alpha}^{(j)} \left(r_{T+h}^{(j)} | X_T^{(j)} \right), \quad (18)$$

which is denoted “SA-Mean” in Table 1. Similarly, “SA-Median” is to take median instead of average.

We compare the out-of-sample tick loss of SA-Mean and SA-Median against $r_{T+h}^{(0)}$, i.e., $\varepsilon = r_{T+h}^{(0)} - \bar{Q}_{\alpha, T+h}^{SA} \left(r_{T+h}^{(j)} | X_T^m \right)$ in Equation (6). Although the benchmark Daily-Close, CF, and CI quantile forecasts are all computed for $r_{T+h}^{(0)}$ by design, we consider returns computed from high-frequency intraday subsample observations as multiple replications of the true distribution of daily returns (even if it is at a different time of a day). This is motivated by the bagging approach, which replicates the true distribution by bootstrap distribution in computing the ensemble average. In this sense, we consider the high-frequency subsamples just like the bootstrap samples, which both permit us to estimate the ensemble average in daily frequency. Thus we treat the subsample averaging VaR/quantile forecasts as daily close VaR/quantile forecasts.

The main idea of SA is based on the assumption that subsamples from high frequency data may replicate the true distribution, which may permit computing the ensemble average of a stochastic process. The subsamples from the high-frequency series may be considered as multiple replications of lower-frequency (daily) series. Under the assumption of the strict stationarity we use the high-frequency data to generate subsamples of daily series in different times within a day. In this section we consider high-frequency observations as multiple replications of the true distribution of daily returns. The subsample returns, $r_t^{(j)}$ ($j = 0, 1, \dots, m$), are considered as $(m + 1)$ replications (draws) of the daily return series.

The assumption that the daily returns measured at different points of time during a trading day are originating from the same distribution is relevant only for SA, which uses high-frequency data. For bagging (next section), we use daily data and do not use the high-frequency data, and thus we do not need to make such assumptions for bagging. However, as a referee pointed out, we do not need to exclude the intradaily or diurnal patterns. All we need to assume is that the diurnal patterns are the same from day to day. We can allow diurnal patterns as long as they are the same from day to day.³

3.5. Bagging Daily Close Quantile Forecasts

Motivated from the subsample averaging, which replicates the true distribution by using returns computed from high frequency data in computing the ensemble average of a stochastic process, we consider bagging in this section.

Bootstrap averaging (bagging) replicates the true distribution by bootstrap distribution in computing the ensemble average of a stochastic process. 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 [5]. Reference [25] shows that the indicator functions are unstable. The check loss function of [19] for the regression quantile involves an indicator function as shown in Equation (6). The mechanism of bagging has been explained in various ways. Reference [5] uses the Cauchy-Schwarz inequality for a squared error loss. Reference [17] extends it to a convex loss (e.g., a tick function for quantiles) using the Jensen’s inequality. Reference [25] shows that, for a nonsmooth unstable predictor, bagging reduces variance of the first order term. In particular, they show that bagging can reduce the mean squared error of forecasts by averaging over the randomness of variable selection. References [26, 27] expand a smooth unstable function into linear and higher order terms, and show

³We thank a referee for this point.

bagging reduces the variance of the higher order terms. Bagging also stabilizes prediction by equalizing the influence of outlying training observations. Reference [21] shows that bagging is asymptotically Bayesian shrinkage. Applications of bagging include inflation [28], financial volatility [29], equity premium [16], short-term interest rates [30], and employment data [31].

Bagging for the α -quantile of daily close return $r_t^{(0)}$ is implemented as follows.

1. Bagging Step 1: Construct a bootstrap sample $\{r_t^{(0)*}\}$ according to the empirical distribution of daily close returns. Run the quantile regression in Equation (4) by regressing $r_t^{(0)*}$ on $X_{t-h}^{(0)*} = (1 \ r_{t-h}^{(0)*} \ r_{t-h}^{(0)*2})'$, obtain $\hat{\beta}^{(0)*}(\alpha)$, and compute the bootstrap Daily Close quantile forecast

$$\hat{Q}_\alpha^{(0)*} \left(r_{T+h}^{(0)*} | X_T^{(0)*} \right) = X_T^{(0)*'} \hat{\beta}^{(0)*}(\alpha). \quad (19)$$

Repeat this B times.

2. Bagging Step 2: Combine the quantile forecasts from Step 1 by some weighting methods for forecast combination. A simplest example of the combination is the simple average:

$$\bar{Q}_\alpha^{Bagging} \left(r_{T+h}^{(0)} | X_T^{(0)} \right) = \frac{1}{B} \sum_{b=1}^B \hat{Q}_\alpha^{(0)*}(b) \left(r_{T+h}^{(0)*}(b) | X_T^{(0)*}(b) \right), \quad (20)$$

i.e., taking the simple mean of all the bootstrap Daily Close quantile forecasts (denoted “Bagging-Mean”). One can also use the median of the individual quantile forecasts as a combined forecast (denoted “Bagging-Median”). As they are very similar, we only report the former in Table 1, denoted as “Bagging Daily-Close.”

A 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 references [32, 33]. Therefore we use block bootstrapping in our empirical study as stock market returns are likely to exhibit time dependence. In the next section, we use the number of bootstrap samples $B = 50$ and the bootstrap block size fixed at 4.

4. Out-of-Sample Quantile Forecasting Results

Table 1 presents the performance of each forecasting method for predicting one day ahead ($h = 1$) daily close S&P 500 return quantiles. The 78 daily returns introduced in Section 2 are calculated by logarithm difference of corresponding index values and multiplied by 100. The size of out-of-sample period is $P = 500$ days from 29 May, 2001 to 30 May, 2003. The in-sample size is $R = 1000$ days, from 10 June, 1997 to 25 May, 2001. We use the rolling window scheme to estimate the parameters and set the size of estimation window at $R = 1000$. Quantile regressions are estimated using the interior-point algorithm by [34]. We report the out-of-sample mean tick loss ratios of each chosen forecasting scheme over Daily Close benchmark for the left-tail probability parameter $\alpha = 0.01, 0.05, \dots, 0.99$, as defined in Equation (2).

The results in Table 1 show that, except for CI-Unrestricted which obviously does not work well due to its dimensionality problem, in general we indeed gain by incorporating high-frequency intraday information into predicting daily close return quantiles. Evidently, we observe that generally the “averaging” methods (SA-Mean, SA-Median, CF-Mean, CF-Median, and Bagging) work better than various CI methods, for most values of α . All SA and CF methods with simple weighting schemes improve upon Daily-Close benchmark (less than 1 in loss ratios), sometimes quite substantially. This indicates the stability of the averaging methods with simple weighting schemes (mean or median).

For those with factor model approaches, the maximum hypothesized number of factors, k_{max} , is set at 15, so number of factors k is chosen within interval $[1, 15]$ for AIC and BIC or fixed at 1, 2, or 3. We see that generally BIC performs better than AIC but usually worse than fixing factor numbers (at 1, 2 or 3). Therefore, the simple rule-of-thumb of fixing k at a small number, that is, using a few fixed number of factors to summarize the entire high-frequency information, works considerably better than estimating k by AIC or BIC.

We also find that CF is more robust than CI. This highlights the merits of CF that are illustrated in [16], and consistent with what is often found in the literature about CF.

The percentage loss reductions at tails, especially at the left tails, are much larger than those of middle quantiles. It shows that there is more room for the various forecasting schemes to improve upon Daily-Close model at tails by incorporating the high frequency information to forecast daily quantiles. Let us look at the results for the left tail with $\alpha = 0.01$, which financial risk management is usually concerned with. Relative to the benchmark quantile forecast using Daily-Close returns, the Bagging Daily-Close quantile forecast reduces the relative out-of-sample mean forecast tick loss ratio to 91.39% (improving by 8.61%). The CF-Mean quantile forecast and CF-Median quantile forecast reduce the loss ratio to 92.26% and 93.39% respectively, which is slightly worse than Bagging Daily-Close quantile Forecast. The subsample averaging methods (SA-Mean and SA-Median) are generating the best daily quantile forecasts with the out-of-sample mean tick loss ratios at 88% and 89%. That is an astounding 11-12% reduction of tick loss for daily VaR forecasting with $\alpha = 0.01$, which is achieved by incorporating the intraday 5-minute high-frequency information into modeling the lower-frequency daily series.

Table 1. Forecasting Quantiles Using High-Frequency Information

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
Daily-Close (mean tick lossx100)	4.5816	15.1202	24.7583	48.3697	54.5713	48.6378	26.7932	16.5165	5.0059
Bagging Daily-Close	0.9139	0.9826	0.9976	0.9935	0.9971	0.9980	0.9934	0.9938	0.9131
SA-Mean	0.8803	0.9723	0.9877	0.9873	0.9990	0.9967	0.9882	0.9764	0.9513
SA-Median	0.8903	0.9731	0.9916	0.9885	0.9994	0.9954	0.9905	0.9818	0.9699
CF-Mean	0.9226	0.9718	0.9928	0.9938	0.9992	0.9943	0.9912	0.9750	0.9403
CF-Median	0.9339	0.9746	0.9952	0.9939	0.9997	0.9952	0.9940	0.9805	0.9445
CF-PC (AIC)	1.1006	1.0221	1.0273	1.0025	1.0097	1.0047	1.0224	0.9772	2.2536
CF-PC (BIC)	1.0746	1.0103	0.9982	0.9948	0.9985	0.9944	1.0064	0.9938	2.0761
CF-PC ($k = 1$)	0.9253	0.9726	0.9892	0.9962	0.9985	0.9982	0.9993	0.9832	0.9562
CF-PC ($k = 2$)	0.9356	0.9792	0.9894	0.9973	0.9985	0.9983	0.9916	1.0034	0.9896
CF-PC ($k = 3$)	0.9853	0.9823	1.0012	0.9983	0.9992	1.0008	0.9884	0.9985	0.9607
CI-Unrestricted	5.8885	2.0387	1.5619	1.1643	1.1971	1.2420	1.4362	1.8454	4.8916
CI-PC (AIC)	1.0956	1.0340	1.0337	0.9914	1.0006	0.9980	0.9888	1.0278	0.9905
CI-PC (BIC)	1.0456	1.0291	1.0284	0.9951	1.0004	0.9871	0.9891	0.9945	0.9967
CI-PC ($k = 1$)	0.9471	0.9675	1.0121	0.9983	1.0004	1.0018	1.0030	0.9952	0.9167
CI-PC ($k = 2$)	0.9589	0.9817	1.0132	0.9977	1.0033	1.0040	1.0095	0.9966	0.8947
CI-PC ($k = 3$)	0.9974	0.9752	1.0014	0.9984	1.0025	0.9955	1.0018	0.9879	1.0224

Notes: This table presents the relative performance of each forecasting scheme for one-day ahead daily close S&P 500 return quantile prediction. We use 5-minute high-frequency S&P 500 index data from 9 June, 1997 to 30 May, 2003, a total of 117,078 observations on 1,501 days. There are 78 5-minute price index observations during a day. We generate the 78 subsample “daily” return series from a time in a day to the same time in the following trading day. Each of the daily return series is thus with length 1,500 days. The daily returns are calculated by log difference of corresponding index values and multiplied by 100. Out-of-sample size P is 500 days, from 29 May, 2001 to 30 May, 2003; in-sample size R is 1,000 days, from 10 June, 1997 to 25 May, 2001. We use rolling window scheme to estimate the parameters and set the size of estimation window at $R = 1000$. Here we report the out-of-sample mean forecast tick loss ratios of each chosen forecasting scheme relative to the Daily Close benchmark model. We report for different values of the left-tail probability parameter α , including $\alpha = 0.5$ for the median. A less-than-one number indicates that the chosen model outperforms the benchmark Daily Close, and bolded number indicates the smallest average tick loss ratio of a model among all methods for each α in each column. For CF-PC and CI-PC, the information criteria AIC and BIC in selecting number of factors are modified for the mean tick forecast errors instead of the mean squared errors. In all factor model approaches, the maximum hypothesized number of factors, k_{max} , is set at 15, and so the number of factors k is chosen within interval $[1, 15]$. In Bagging the number of bootstrap samples is 50 and bootstrap block size is fixed at 4.

5. Conclusions

In this paper, we present and propose several methods of incorporating high-frequency information in forecasting daily quantiles, which are shown to be particularly useful for lower tail daily Value-at-Risk forecasting. We predict that one-day-ahead Value-at-Risk forecasting will become an essential daily risk management tool. Long horizon Value-at-Risk forecasting would also be interesting; we leave this for future work. The aftermath of the global financial crisis of 2007-08 makes daily risk management more relevant. Tighter banking regulations also lifted the bar on stress testing. Central banks, the IMF, and the Basel Committee call for tighter regulations to fix the causes of the crisis. Many financial institutions are required to report daily exposures and downside risks to regulators.

The methods considered in this paper are based on “averaging.” Unlike a typical method of model-averaging whose average is taken over different models, our averaging involves only one model but is taken over multiple forecasts that are generated from using multiple lower-frequency (daily) datasets constructed from higher-frequency (intra-day) data. Hence, it is not model averaging, but high-frequency data averaging or combining. We suggest three such data averaging/combining methods – combining forecast (CF), subsample averaging (SA), and bootstrap averaging (Bagging) – as explained in Section 3.

Each of these three averaging methods is designed to incorporate high-frequency (intraday) information into lower-frequency (daily) modelling/forecasting.

As demonstrated in Section 4 in forecasting VaR/quantiles of the S&P 500 index return, using high-frequency information is beneficial, often substantially and particularly so in forecasting downside risk. For daily S&P 500 return lower tail out-of-sample VaR forecasts, our empirical results show that the averaging methods of SA, Bagging, and CF (which serve as different ways of forming the ensemble average), which use high-frequency intraday information, have excellent forecasting performance when compared to using only low-frequency daily close information.

6. Acknowledgements

We thank George Jiang, Eric Hillebrand, Aman Ullah and two referees for their useful comments. The opinions expressed in this paper are those of the authors and do not necessarily reflect those of Grantham, Mayo, Van Otterloo and Company LLC.

References

1. Artzner, P.; Delbaen, F.; Eber, J.-M.; Heath, D. Coherent Measures of Risk. *Mathematical Finance* **1999**, *9*(3), 203–228.
2. Clements, M.P.; Galvao, A.B.; Kim, J.H. Quantile Forecasts of Daily Exchange Rate Returns from Forecasts of Realized Volatility. *Journal of Empirical Finance* **2008**, *15*, 729–750.
3. Ghysels, E.; Sinko, A.; Valkanov, R. MIDAS Regressions: Further Results and New Directions. *Econometric Reviews* **2006**, *26*, 53–90.
4. Engle, R.F.; Russell, J.R. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* **1998**, *66*(5), 1127–1162.
5. Breiman, L. Bagging Predictors. *Machine Learning* **1996**, *36*, 105–139.
6. Zhou, B. High-Frequency Data and Volatility in Foreign-Exchange Rates. *Journal of Business and Economic Statistics* **1996**, *14*, 45–52.
7. Barndorff-Nielsen, O.E.; Hansen, P.R.; Lunde, A.; Shephard, N. Subsampling Realised Kernels. *Journal of Econometrics* **2011**, *160*, 204–219.
8. Zhang, L.; Mykland, P.A.; Aït-Sahalia, Y. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association* **2005**, *100*, 1394–1411.
9. Andersen, T.G.; Bollerslev, T.; Diebold, F.X.; Labys, P. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* **2001**, *96*, 42–55.
10. Barndorff-Nielsen, O.E.; Shephard, N. Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society Series B* **2002**, *64*, 853–223.
11. Aït-Sahalia, Y.; Mykland, P.A.; Zhang, L. How Often to Sample a Continuous-time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* **2005**, *10*, 351–416.
12. Bandi, F.M.; Russell, J.R. Microstructure Noise, Realized Variance, and Optimal Sampling. *Review of Economic Studies* **2008**, *75*, 339–369.

13. Hansen, P.R.; Lunde, A. Realized Variance and Market Microstructure Noise. *Journal of Business and Economic Statistics* **2006**, *24*, 127-218.
14. Barndorff-Nielsen, O.E.; Hansen, P.R.; Lunde, A.; Shephard, N. Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. *Econometrica* **2008**, *76*(6), 1481-1536.
15. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., and Shephard, N. , “Subsampling Realised Kernels,” *Journal of Econometrics*, **2011**, *160*, 204-219.
16. Huang, H.; Lee, T.-H. To Combine Forecasts or to Combine Information? *Econometric Reviews* **2010**, *29*(5), 534-570.
17. Lee, T.-H.; Yang, Y. Bagging Binary and Quantile Predictors for Time Series. *Journal of Econometrics* **2006**, *135*, 465-497.
18. Chernozhukov, V.; Umantsev, L. Conditional Value-at-Risk: Aspects of Modeling and Estimation. *Empirical Economics* **2001**, *36*, 221-292.
19. Koenker, R.; Basset, G. Asymptotic Theory of Least Absolute Error Regression. *The Journal of the American Statistical Association* **1978**, *73*, 618-622.
20. Stock, J.H.; Watson, M.W. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* **2002**, *97*, 1167-1179.
21. Stock, J.H.; Watson, M.W. Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business and Economic Statistics* **2012**, *30*, 481-493.
22. Machado, J.A.F. Robust Model Selection and M-Estimaion. *Econometric Theory* **1993**, *9*, 479-990.
23. Koenker, R. *Quantile Regression*, Cambridge University Press: Cambridge, UK, 2005.
24. Granger, C.W.J.; Ramanathan, R. Improved Methods of Combining Forecasts. *Journal of Forecasting* **1984**, *3*, 197-204.
25. Bühlmann, P.; Yu, B. Analyzing Bagging. *Annals of Statistics* **2002**, *30*, 927-961.
26. Buja, A.; Stuetzle, W. Observations on Bagging. *Statistica Sinica* **2006**, *16*, 323-352.
27. Friedman, J.H.; Hall, P. On Bagging and Nonlinear Estimation. *Journal of Statistical Planning and Inference* **2007**, *137*, 669-683.
28. Inoue, A.; Kilian, L. How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation. *Journal of the American Statistical Association* **2008**, *103*(482), 511-522.
29. Hillebrand, E.; Medeiros, M.C. The Benefits of Bagging for Forecast Models of Realized Volatility. *Econometric Reviews* **2010**, *29*(5), 571-593.
30. Audrino, F.; Medeiros, M.C. Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics* **2011**, *26*(6), 999-1022.
31. Rapach, D.; Strauss, J. Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth. *Econometric Reviews* **2010**, *29*(5), 511-533.
32. Hall, P.; Horowitz, J.L.; Jing, B.-Y. On Block Rules for the Bootstrap with Dependent Data. *Biometrika* **1995**, *82*(3), 561-574.
33. Fitzenberger, B. The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions. *Journal of Econometrics* **1997**, *82*, 235-287.

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

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).