Econometric Reviews ›

Volume 27, 2008 - Issue 1-3

Original Articles

Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data—But Which Frequency to Use?

Michiel de Pooter, Martin Martens & Dick van Dijk

Pages 199-229 | Received 21 Sep 2005, Accepted 04 Apr 2006, Published online: 07 Mar 2008

Cite this article https://doi.org/10.1080/07474930701873333

Abstract

This article investigates the merits of high-frequency intraday data when forming mean-variance efficient stock portfolios with daily rebalancing from the individual constituents of the S&P 100 index. We focus on the issue of determining the optimal sampling frequency as judged by the performance of these portfolios. The optimal sampling frequency ranges between 30 and 65 minutes, considerably lower than the popular five-minute frequency, which typically is motivated by the aim of striking a balance between the variance and bias in covariance matrix estimates due to market microstructure effects such as non-synchronous trading and bid-ask bounce. Bias-correction procedures, based on combining low-frequency and high-frequency covariance matrix estimates and on the addition of leads and lags do not substantially affect the optimal sampling frequency or the portfolio performance. Our findings are

also robust to the presence of transaction costs and to the portfolio rebalancing frequency.

Keywords:

Bias-correction High-frequency data Mean-variance analysis Realized volatility Tracking error Volatility timing

JEL Classification:

G11

ACKNOWLEDGMENTS

We thank Federico Bandi, Jeffrey Russell, Valeri Voev, the editors Michael McAleer and Essie Maasoumi, and two anonymous referees for helpful comments and suggestions. Any remaining errors are ours alone.

Notes

[1]Zhang et al. (2005) focus solely on estimating the variance but here we apply their approach to covariances as well, as suggested by Zhang (2005).

[2]We should note that the two time-scales estimator of Zhang et al. (2005) is derived under the same model assumptions (see Equation (4) in Zhang et al., 2005). By using it in the same format for covariances we acknowledge that for example the weights for the covariance matrices at two frequencies may be suboptimal and that the estimator may be biased. Of course in our empirical work we can see whether despite these reservations the idea itself is useful for predicting covariances.

[3] http://www.price-data.com/

[4]For obvious reasons the overnight return from 10 to 17 September, 2001 (the first trading day after 9/11) has been dropped.

Notes: The table shows mean and variance of the realized (co-)variances at various sampling frequencies for 78 constituents of the S&P 100 index from April 16, 1997, through May 27, 2004 (1788 trading days). For the realized variance, the mean reflects the average taken over all 78 stocks and over all 1788 trading days. The variance is the average taken over the 78 sample variances of the realized variances. For the realized covariance the mean reflects the average taken over all 3003 pairs of stocks and over all 1788 trading days. The variance is the average taken over the 3003 sample variances of the realized covariances. In Panel A the “standard” realized covariance matrix $V_{t-1,\,h}$ given in (1) is used. Panel B is based on the two time-scales estimator given in (2), while Panel C shows results for the lead-lag corrected estimator given in (3), with Bartlett-kernel weights $d_l = 1 - l/(q + 1)$ and $q = 1$.

[5]An exception is the realized variance at the one- and two-minute frequencies, where also the variance increases due to the increased importance of bid-ask bounce.

[6]We experimented with alternative values for q, which led to qualitatively similar findings. Detailed results are available upon request. The issue of determining the optimal value of q is beyond the scope of this article and is left for future research.

[7]As explained below, we require part of the sample period to initialize the conditional covariance matrix estimates, which in our case equals 122 trading days. This implies that the effective sample period available for portfolio construction and evaluation runs from October 8, 1997 until May 27, 2004 (1666 trading days).

[8]Fleming et al. (2003) use a similar approach when assessing the effect of transaction costs.

[9]Note that our approach here differs from Fleming et al. (2003). Our method of holding multiple portfolios simultaneously is commonly applied in the literature on stock selection, see Jegadeesh and Titman (1993) and Rouwenhorst (1998), among many others.

[10]Fleming et al. (2003) show that actually using the (unrestricted) multivariate GARCH model leads to a better fit of the data as expected, but the covariance matrix forecasts result in worse portfolios than those obtained from the rolling covariance estimator. They cite the smoothness of the rolling estimator as the main reason for this.

Notes: The table shows the decay rates (α) that maximize the likelihood of the model in (13) and (12) for daily data and (13) and (14) for intraday data. In Panel A the model is estimated for total returns, whereas in Panel B the model is estimated for excess returns (stock returns minus S&P 500 returns). The second and third column show the optimal decay rates and accompanying log-likelihood values when the covariance updates are based on the standard realized (co-)variances, the fourth and fifth column when the updates are based on the two time-scales estimator, and the final two columns when 1 lead and 1 lag of the (co-)variances are added to the contemporaneous (realized) covariances.

Notes: The table shows the out-of-sample performance of the overall minimum variance portfolio, with weights given in (6), and the minimum variance portfolio given a target level of return of 10%, with weights given in (8), constructed using rolling covariance matrix forecasts based on various sampling frequencies and based on different ways of measuring the realized covariance matrix (standard, two time-scales, and 1 lead and 1 lag). For the target return portfolios, we report the mean return (μ_P) and standard deviation (σ_P) in annualized percentage points, the Sharpe ratio (SR), the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios, and average daily turnover (TO) in percentage points. For the minimum variance portfolios, we report the standard deviation (α_{MVP}) in annualized percentage points average daily turnover (TO) in percentage points.

[11]We examined the sensitivity of our results to the target return level by varying μ_P between 2% and 18%. These alternative target return levels led to qualitatively similar conclusions as those reported below. Detailed results are therefore not shown here, but are available on request.

Notes: The table shows the out-of-sample performance of the overall minimum tracking error portfolio, with weights given in (6), and the minimum tracking error portfolio given a target level of return of 1%, with weights given in (8), constructed using rolling covariance matrix forecasts based on various sampling frequencies and based on different ways of measuring the realized covariance matrix (standard, two time-scales, and 1 lead and 1 lag). For the target active return portfolios, we report the mean active return (μ_P) and tracking error (TE_P) in annualized percentage points, the information ratio (IR), the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios, and average daily turnover (TO) in percentage points. For the minimum tracking error portfolios, we report the tracking error (TE_{MTE}) in annualized percentage points average daily turnover (TO) in percentage points.

Notes: The table shows the out-of-sample performance of the minimum variance portfolio given a target level of return of 10%, with weights given in (8), constructed using rolling covariance matrix forecasts based on daily returns and on intraday returns at the sampling frequency that maximized the information ratio, based on the "standard" way of measuring the realized covariance matrix. We report the mean return (μ_P) and standard deviation (α_P) in annualized percentage points, the Sharpe ratio (SR), and the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios. The column headed c indicates the level of transaction costs, expressed in annualized percentage points, which correspond with the reduction in the annualized portfolio return if the entire portfolio would have to be traded every day during the whole year. The column headed h indicates the optimal sampling frequency, expressed as the length of the corresponding return interval in minutes.

Notes: The table shows the out-of-sample performance of the minimum tracking error portfolio given a target level of return of 1%, with weights given in (8), constructed using rolling covariance matrix forecasts based on daily returns and on intraday returns at the sampling frequency that maximized the information ratio, based on the 'standard' way of measuring the realized covariance matrix. We report the mean active return (μ_P) and tracking error (TE_P) in annualized percentage points, the information ratio (IR), the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios. The column headed c indicates the level of transaction costs, expressed in annualized percentage points, which correspond with the reduction in the annualized portfolio return if the entire portfolio would have to be traded every day during the whole year. The column headed h indicates the optimal sampling frequency, expressed as the length of the corresponding return interval in minutes.

Notes: The table shows the out-of-sample performance of the overall minimum volatility (tracking error) portfolio, with weights given in (6), and the minimum variance portfolio given an annualized target level of (active) return of 10% (1%), with weights given in

(8), constructed using rolling co-variance matrix forecasts based on various sampling frequencies and based on the 'standard' realized covariance matrix. Panel A shows results for total returns and Panel B for excess returns (stock returns minus S&P 500 returns). The optimal decay parameters are determined by optimizing portfolio performance using an expanding window period (starting with 250 days). Columns 2 and 3, and 6 and 7, report the mean and standard deviation of the resulting estimates of α_h. Columns 4 and 8, headed 'Perf.', show the Sharpe ratio and volatility (panel A) or the information ratio and tracking error (panel (B) for the resulting portfolios. Columns 5 and 9, headed 'LogL', show the SR/IR and σ_P/TE_P for portfolios constructed with decay parameters for the conditional covariance matrix that are estimated by maximizing the log-likelihood over the complete out-of-sample period.
