Realized mixed–frequency factor models for vast dimensional covariance estimation

Karim Bannouh   Martin Martens   Roel Oomen
Dick van Dijk

24–10–2012

Abstract

We introduce a Mixed-Frequency Factor Model (MFFM) to estimate vast dimensional covariance matrices of asset returns. The MFFM uses high-frequency (intraday) data to estimate factor (co)variances and idiosyncratic risk and low-frequency (daily) data to estimate the factor loadings. We propose the use of highly liquid assets such as exchange traded funds (ETFs) as factors. Prices for these contracts are observed essentially free of microstructure noise at high frequencies, allowing us to obtain precise estimates of the factor covariances. The factor loadings instead are estimated from daily data to avoid biases due to market microstructure effects such as the relative illiquidity of individual stocks and non-synchronicity between the returns on factors and stocks. Our theoretical, simulation and empirical results illustrate that the performance of the MFFM is excellent, both compared to conventional factor models based solely on low-frequency data and to popular realized covariance estimators based on high-frequency data.

Keywords: high-frequency data; factor models; realized covariance; microstructure noise; non-synchronous trading
JEL codes:

ERIM Report Series *Research in Management*	
ERIM Report Series reference number	ERS–2012–017
Date of publication	2012–10–23
Version	2012–10–24
Number of pages	44
Persistent URL for paper	http://hdl.handle.net/1765/37470
Email address corresponding author	mmartens@ese.eur.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus University Rotterdam PO Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31104081182 Fax: +31104089640 Email: info@erim.eur.nl Internet: http://www.erim.eur.nl
Availability	The ERIM Report Series is distributed through the following platforms: RePub, the EUR institutional repository Social Science Research Network (SSRN) Research Papers in Economics (RePEc)
Classifications	The electronic versions of the papers in the ERIM Report Series contain bibliographic metadata from the following classification systems: Library of Congress Classification (LCC) Journal of Economic Literature (JEL) ACM Computing Classification System Inspec Classification Scheme (ICS)

Realized mixed-frequency factor models for vast dimensional covariance estimation

Karim Bannouh

Erasmus University Rotterdam

Martin Martens*

Erasmus University Rotterdam

Roel Oomen

Deutsche Bank and University of Amsterdam

Dick van Dijk

Erasmus University Rotterdam

October 2012

Abstract

We introduce a Mixed-Frequency Factor Model (MFFM) to estimate vast dimensional covariance matrices of asset returns. The MFFM uses high-frequency (intraday) data to estimate factor (co)variances and idiosyncratic risk and low-frequency (daily) data to estimate the factor loadings. We propose the use of highly liquid assets such as exchange traded funds (ETFs) as factors. Prices for these contracts are observed essentially free of microstructure noise at high frequencies, allowing us to obtain precise estimates of the factor covariances. The factor loadings instead are estimated from daily data to avoid biases due to market microstructure effects such as the relative illiquidity of individual stocks and non-synchronicity between the returns on factors and stocks. Our theoretical, simulation and empirical results illustrate that the performance of the MFFM is excellent, both compared to conventional factor models based solely on low-frequency data and to popular realized covariance estimators based on high-frequency data.

---

*Corresponding author, Dept. of Finance, Erasmus University Rotterdam, P.O. Box 1738, NL–3000 DR Rotterdam, The Netherlands, e-mail: `mmartens@ese.eur.nl`, phone:+31 10 - 40 88929, fax:+31 10 - 40 89162. We thank participants of the 2008 and 2009 Stanford Institute for Theoretical Economics workshop on "Econometric Analysis of High-Frequency Data and the Impact of Economic News", the Society for Financial Econometrics First European conference (Geneva, 2009), the 20th $(EC)^2$ conference "Real Time Econometrics" (Aarhus, 2009), CMAP Ecole Polytechnique (Paris, 2009) and CREATES (Aarhus, 2009) for useful comments and suggestions. The second and fourth author are grateful to Inquire Europe for providing financial support.

1 Introduction

Accurate measures and forecasts of asset return covariances are important for financial risk management and portfolio management. Recent academic research in these areas has focused on two different issues. First, intraday data has been shown to render more precise measures and forecasts of daily asset return volatilities and covariances. Second, for the practically relevant case of portfolios consisting of a large number of assets, factor structures have been found useful to tackle the "curse of dimensionality". In this paper we put forward a novel approach for accurate measurement and forecasting of the covariance matrix of vast dimensional portfolios by combining the use of high *and* low-frequency data with a linear factor structure. Specifically, we introduce a "mixed-frequency" factor model (MFFM), where high-frequency data on relatively liquid factors is used for precise estimation of the factor covariance matrix and idiosyncratic risk whereas the factor loadings are estimated from low-frequency data.

In recent years, a substantial body of literature has emerged on the use of high-frequency data for obtaining more accurate measures and forecasts of financial risk, see e.g. Andersen et al. (2006a) and McAleer and Medeiros (2008) for recent reviews. For the multivariate case Barndorff-Nielsen and Shephard (2004) introduced the so-called realized covariance, summing the cross-products of intraday returns. Market microstructure, however, poses two challenges: First, transactions take place against bid and ask prices, causing overestimation of the volatility. Second, non-synchronous trading of stocks biases covariance estimates towards zero. Several covariance estimators have been proposed that are robust to microstructure frictions. Focussing on the bi-variate case Bandi and Russell (2005) illustrate how to choose the optimal sampling frequency for the realized covariance in the presence of microstructure noise. Hayashi and Yoshida (2005) propose an "all overlapping" returns estimator that is robust to non-synchronous trading. Griffin and Oomen (2011), Martens (2006), and Voev and Lunde (2007) provide further insights into the properties of the Hayashi-Yoshida and lead-lag adjusted realized covariance estimators in the presence of non-trading and microstructure noise. Zhang (2011) extends the two-scales estimator of Zhang et al. (2005) to covariance estimation. Moving beyond a bi-variate setting, Barndorff-Nielsen et al. (2011) introduce multivariate realized kernels which deliver consistent and positive semi-definite covariance matrix estimates. For these multivariate kernels so-called refresh time-sampling discards a substantial part of the available high-frequency data, although Hautsch et al. (2012) propose a block approach to reduce this problem. None of the aforementioned approaches, however, can empirically cope with a universe consisting of hundreds or even thousands of stocks that make up most stock market indices used to benchmark fund managers.

Recently Fan et al. (2008) have revisited the use of factor models for covariance estimation in case of a large number of assets, in order to reduce the dimensionality of the problem. They show that the factor model approach improves over the sample covariance matrix (based on daily data) in particular when the portfolio optimization problem requires the inverse of the covariance matrix. The reason is that in the factor model approach only the factor covariance matrix needs to be inverted, which typically is of much lower dimension. In addition, using the covariance matrix based on a factor structure reduces the problem of error maximization for portfolio construction applications, see for example Jagannathan and Ma (2003).

With the MFFM we introduce a novel methodology that exploits the advantages of both high-frequency data and the factor model approach: It enables more efficient estimation of covariances whilst still being able to cope with a very large number of stocks. The covariance matrix based on the factor model requires three estimates: The covariance matrix of the factor returns, the factor loadings, and the stock-specific variances. Without compromising the consistency and positive-definiteness of the resulting covariance matrix we can choose different sampling frequencies for each of these three estimates.

First, in the MFFM approach we use realized covariances obtained from high-frequency intraday returns to estimate the daily factor covariance matrix. This is motivated by the fact that nowadays highly liquid financial contracts such as index futures and exchange-traded funds (ETFs) are available as proxies for the most commonly used factors. This further increases the added value of high-frequency data because microstructure frictions are relatively small. For this reason the factor covariance matrix can be estimated with high precision from intraday data. Second, we estimate the factor loadings using daily data for the reason that single-day betas based on high-frequency data are very noisy due to the non-synchronicity between factor returns and stock returns, see for example Andersen et al. (2006b), Todorov and Bollerslev (2010) or Hansen et al. (2010) for related discussions.

Finally, although intraday data is also available for individual stocks, these are generally less liquid than index futures and ETFs. Hence, we can use intraday data for stock-specific variances, but possibly at a lower frequency than the one used for the factor covariance matrix.

We provide theoretical, simulation-based and empirical evidence that the MFFM offers a useful approach for estimating vast dimensional covariance matrices. In the theoretical part of this paper we show that, assuming i.i.d. microstructure noise and a Poisson arrival process for non-synchronous trading, the covariance estimates of the MFFM are unbiased and we obtain a closed form expression for the variance of these covariance estimates. Based on analytical expressions for the variance of the estimators, we show that the MFFM improves substantially in terms of efficiency over that of the popular Hayashi and Yoshida (2005) and realized covariance (lead-lag) estimators.

Next, we use Monte Carlo simulations to show that also when we relax several of the assumptions underlying the theoretical results and move from the bi-variate case to a realistic setting of 500 assets, the MFFM estimator is superior to the realized covariance estimator.

We empirically evaluate the MFFM estimator by comparing its performance to the (sample) realized covariance and a factor model based on daily returns. We consider three stock universes: The S&P 500 (large caps, most liquid), the S&P 400 (mid caps), and the S&P 600 (small caps, illiquid). To the best of our knowledge, we are the first in the literature to consider such high dimensional problems involving high-frequency data. Of course, in the empirical case unlike for the theory and simulations we do not know the true covariances. For this reason we analyze two empirical applications. First, we use Mincer-Zarnowitz and forecast encompassing regressions to obtain insights in the ability of the MFFM to forecast the volatility of vast dimensional portfolios out-of-sample. Second, we evaluate the performance of minimum tracking error portfolios.[1] We find that in each of the three S&P universes the out-of-sample MFFM portfolio volatility forecasts improve upon realized covariance and daily factor model forecasts when we rank the forecasts on their Mincer-Zarnowitz R^2. Using encompassing regressions, in which we add the realized covariance or daily factor model forecasts to MFFM we find that the coefficient on realized covariance and the daily factor model is negative. Adding these forecasts to the MFFM forecasts improves the MFFM forecasts only marginally. When the objective is to track a benchmark using out-of-sample covariance matrix forecasts, the MFFM provides smaller tracking errors and much smaller portfolio turnover than the realized covariance. Conventional factor models based on daily data manage to achieve a similar tracking error as the MFFM, but only if a long historical data period is used. This is due to the fact that it needs a substantial amount of smoothing, whereas the MFFM can manage the same performance with a very short span of historical data. In addition, the portfolio turnover of the daily factor model is about three times larger than the MFFM turnover. For forecasting portfolio volatility and for minimizing the tracking error we find that differences between the MFFM and realized covariance increase as we move from the most liquid stock universe to the least liquid universe, as expected.

In recent work Hansen et al. (2010) and Noureldin et al. (2012) advocate the use of high-frequency data in a parametric GARCH framework. Related to our idea of using a mixed-frequency sampling approach for modeling vast dimensional covariance matrices several authors have recently implemented subcases and modifications of the mixed-frequency (factor model) methodology. Kyj et al. (2009) study a single-factor model, which is a special case of the MFFM, to forecast covariance matrices in the absence of noise and non-trading. Halbleib and Voev (2011) propose to use mixed-frequency sampling for predicting covariance matrices by using high-

[1] Chan et al. (1999) show that differences between covariance estimators are small for minimum variance portfolios because the market factor dominates.

frequency data for realized volatilities and low-frequency data for correlations. Hence, without using a factor structure, by using mixed-frequency sampling they successfully circumvent the issue of non-trading for estimating correlations. Combining the Hautsch et al. (2012) blocking and regularization kernel estimator with the MFFM, Hautsch et al. (2011) propose to select factors in a data driven way where mixed-sampling frequencies can be used for volatilities, correlation eigenvalues and eigenvectors. In contrast to our study they use a multi-time-scale approach for reducing the impacts of noise, non-trading and estimation error, rather than studying these frictions explicitly.

The remainder of this paper is structured as follows. In Section 2 we derive the theoretical properties of the MFFM and provide a theoretical comparison with the bi-variate Hayashi and Yoshida (2005), realized covariance and lead-lag estimators. Section 3 contains an extensive Monte Carlo study in which we replicate the S&P500 universe to evaluate the realized covariance and the MFFM covariance matrix estimates. In Section 4 we study the empirical performance of the MFFM and compare it to the realized covariance and a factor model based on daily data. We conclude in Section 5.

2 The Mixed-Frequency Factor Model

Consider a linear factor structure for the return on asset i, that is

$$r_i = \mu_i + \beta_i' f + \varepsilon_i \tag{1}$$

where f is a $K \times 1$ vector of common factors, β_i is a $K \times 1$ vector of factor loadings measuring the exposure to f, and ε_i is the idiosyncratic component. We assume that $\mathsf{E}[f] = 0$ and $\mathsf{E}[\varepsilon_i] = 0$, such that μ_i is the expected return. Furthermore, we assume that the idiosyncratic component is orthogonal to the common factors, i.e. $\varepsilon_i \perp f$. Under these assumptions the covariance between asset i and asset j can be expressed as

$$\gamma_{ij} \equiv \mathsf{Cov}[r_i, r_j] = \beta_i' \Lambda \beta_j + \sigma_{ij} \tag{2}$$

where $\Lambda = \mathsf{E}[ff']$ is the factor covariance matrix and $\sigma_{ij} = \mathsf{E}[\varepsilon_i \varepsilon_j]$ is the covariance between the assets' idiosyncratic components. Throughout, we consider a "strict" factor structure in the spirit of Ross (1976), i.e. we assume that the factor structure exhausts the dependence among the assets so that $\sigma_{ij} = 0$ for $i \neq j$. Approximate factor models where σ_{ij} can be non-zero but small are considered in Chamberlain and Rothschild (1983), Ingersoll (1984) and Connor and Korajczyk (1994).

Using hats to denote estimates of unknown quantities, the covariance estimator is given by

$$\widehat{\gamma}_{ij} = \widehat{\beta}_i' \widehat{\Lambda} \widehat{\beta}_j \qquad \text{for } i \neq j. \tag{3}$$

The properties of this generic covariance estimator are characterized in the theorem below, where we use the notation $\widehat{X} = X + X^{\varepsilon}$.

Theorem 2.1 *Assuming (i) $\mathsf{E}[\sigma_{ij}] = 0$ for $i \neq j$, (ii) $\mathsf{E}[\beta^{\varepsilon}] = 0$, (iii) $\mathsf{E}[\Lambda^{\varepsilon}] = 0$, and (iv) $\beta^{\varepsilon} \perp \Lambda^{\varepsilon}$ element-by-element, then*

$$\mathsf{E}[\widehat{\gamma}_{ij}] = \gamma_{ij}, \tag{4}$$

for $i \neq j$ with

$$\begin{aligned} \mathsf{V}[\widehat{\gamma}_{ij}] = {} & \beta_i' \Lambda \Sigma_{\beta,j} \Lambda \beta_i + \beta_j' \Lambda \Sigma_{\beta,i} \Lambda \beta_j + tr(\Sigma_{\beta,i} \Lambda \Sigma_{\beta,j} \Lambda) \\ & + g(\beta_i \beta_i', \beta_j \beta_j', \Phi) + g(\beta_i \beta_i', \Sigma_{\beta,j}, \Phi) + g(\beta_j \beta_j', \Sigma_{\beta,i}, \Phi) + g(\Sigma_{\beta,i}, \Sigma_{\beta,j}, \Phi), \end{aligned} \tag{5}$$

where $\Sigma_{\beta,i} = \mathsf{V}[\widehat{\beta}_i]$ and $\Phi = \mathsf{E}[vech(\Lambda^{\varepsilon}) vech(\Lambda^{\varepsilon})']$ and

$$g(A, B, \Phi) = \sum_{m,n,p,q}^{N} A_{mp} B_{nq} \Phi_{f(p,n), f(q,m)},$$

and $f(p,q) = N(\min\{p,q\} - 1) + \frac{1}{2}(\min\{p,q\} - \min\{p,q\}^2) + \max\{p,q\}$.

Proof See Appendix A. ■

It is useful to note that the assumptions in this Theorem are not unreasonable for the mixed-frequency approach developed in this paper. Specifically, we propose to estimate betas using low-frequency data, such that it is plausible to assume that betas are unbiased, whereas the factor covariance matrix is estimated from high-frequency data. The factors are essentially free of microstructure noise since the ETFs we propose as factors are very liquid, see Table 1. This justifies the assumption that the factor covariance estimates are unbiased and that possible sources of noise in low-frequency betas and factors observed at high sampling frequencies are independent[2].

The linear factor decomposition of asset returns in (1) has a long and established history in the theoretical and empirical finance literature. Three types of factor models can be distinguished, depending on how the factors f and the associated exposures β are constructed. Specifically, the

[2]In Section 3 we analyze the impact of estimation errors in betas for the MFFM.

model in (1) can be categorized as (i) a statistical factor model (Ross 1976) when both β and f are unspecified and inferred from the panel of asset returns, (ii) a characteristic-based factor model (Rosenberg 1974) when β is fixed and determined by asset-specific characteristics while f is inferred from the data, or (iii) a macro-economic factor model (Chen et al. 1986) when f is observable and derived from macroeconomic or asset pricing theory while β is estimated from the data. See Grinold and Kahn (2000) or Connor et al. (2012) for further discussion.

The factor model we develop in this paper can be classified as a traditional macro-economic model in the sense that the factors are observable and their loadings are estimated from the data. However, its construction is designed to make efficient use of high-frequency data while simultaneously avoiding the potentially severe biases induced by market micro-structure noise. Specifically, our "*mixed-frequency factor model*" involves the use of liquid assets as factors for precise estimation of the factor covariance matrix using high-frequency data, while factor loadings are estimated using lower-frequency returns of the, possibly illiquid, individual assets. The use of liquid factors in the MFFM is motivated by the empirical observation that a growing number of highly liquid exchange traded funds (ETFs) and futures contracts are now available that proxy commonly used country, industry, and style factors. With minimal spreads and accurate real-time pricing for many of these contracts, the effects of market microstructure noise are of little concern and the use of high-frequency data is justified. Particularly for a large and heterogeneous asset universe, however, many of the individual assets may be illiquid and contaminated by market microstructure effects at high sampling frequencies. To support this point Table 1 shows statistics on the ETFs we use in our empirical application. The average number of observations for these ETFs is over 54,000 per day. In contrast the average number of observations for a constituent of the S&P500 is just over 19,000 per day, and this drops to about only 2,000 per day for the constituents of the S&P600, i.e. the small cap index.

– INSERT TABLE 1 ABOUT HERE –

We now specialize the rather general result in Theorem 2.1. to the MFFM setting to gain further insights into its properties. We define F and $\mathcal{F}$ as the matrices of low- and high-frequency factor return observations with dimensions $(T \times K)$ and $(M \times K)$. Similarly, R_i and $\mathcal{R}_i$ denote the vectors of low- and high-frequency returns of asset i of length T and N_i, and τ_i the $(N_i \times 1)$ vector of time-stamps associated with $\mathcal{R}_i$.

Assumption N The factor returns $\mathcal{F}$ are jointly normal with zero mean, serially uncorrelated and observed without friction[3]. The (integrated) factor covariance matrix is estimated using the high-frequency factor returns as $\widehat{\Lambda} = \mathcal{F}'\mathcal{F}$.

[3]Given the highly liquid ETFs we propose as factors, see Table 1, it is justified to assume that factor returns are serially uncorrelated and observed without friction.

Assumption O The asset return dynamics at low frequency are governed by a linear factor model as in (1) with i.i.d. normal residuals ε_i. The factor loadings are estimated by means of linear regression using the low-frequency returns $\widehat{\beta}_i = (F'F)^{-1}F'R_i$.

Corollary 2.2 *Let assumption* ***N***, ***O***, *and those in Theorem 2.1. hold. Then for* $i \neq j$

$$\mathsf{E}[\widehat{\gamma}_{ij}] = \gamma_{ij} \tag{6}$$

and

$$\mathsf{V}[\widehat{\gamma}_{ij}] = \frac{A}{T} + \frac{B}{M}, \tag{7}$$

where

$$\begin{aligned} A &= \sigma_j^2 \beta_i' \Lambda \beta_i + \sigma_i^2 \beta_j' \Lambda \beta_j + \sigma_i^2 \sigma_j^2 \frac{K}{T}, \\ B &= \sum_{m,n,p,q}^{K} \left(\beta_i(m)\beta_i(p) + \Sigma_{\beta,i}(m,p)\right)\left(\beta_j(n)\beta_j(q) + \Sigma_{\beta,j}(n,q)\right)\left(\Lambda_{pq}\Lambda_{nm} + \Lambda_{pm}\Lambda_{nq}\right) \end{aligned}$$

Proof See Appendix A. ■

The above corollary provides insights into the properties of the MFFM covariance estimator. In particular, it is unbiased with a variance that can be attributed to the measurement error in factor loadings (i.e. A/T) and to the measurement error in the factor covariance matrix (i.e. B/M).[4]

– INSERT FIGURE 1 and 2 ABOUT HERE –

To illustrate the efficiency of the MFFM in a bi-variate setting, we compare it to the (i) Hayashi and Yoshida (2005) estimator, (ii) realized covariance and (iii) realized covariance lead-lag estimator. For this purpose, we assume that intraday price observations for asset i (from which the returns $\mathcal{R}_i$ are computed) arrive according to a Poisson process with intensity $\lambda_i = \mathsf{E}[N_i]$. Further, we assume that prices are contaminated with i.i.d. microstructure noise with variance $\xi_i^2 = \pi_i \gamma_i^2 / \lambda_i$. We use closed-form expressions for the efficiency of the popular aforementioned estimators (see Griffin and Oomen (2011) for details) and compare these with the variance of the MFFM covariance estimator. To compute the variance of the MFFM covariance estimator, we need to make some assumptions about the underlying factor structure. Here, we use a setting

[4]Note that in some circumstances β is (assumed to be) known so that $V(\widehat{\gamma}_{ij}) = B/M$, see e.g. Grinold and Kahn (2000, Ch. 3).

with $K = 5$ factors, factor loadings $\beta_i = (0.5, -0.1, 0, 0.2, 0.6)'$, $\beta_j = (0.7, -0.2, -0.3, 0.4, 0.2)'$, and factor covariance matrix $\Lambda = I_K + \frac{1}{2}(1 - I_K)$. The specific or idiosyncratic risk component is $\sigma_h^2 = \beta_h' \Lambda \beta_h$ for $h \in \{i, j\}$ so that the R^2 of the factor regression is around 50% and the assets have a correlation of $\rho_{ij} \approx 40\%$ with:

$$V(r) = (\beta_i, \beta_j)' \Lambda (\beta_i, \beta_j) + \Sigma = \begin{pmatrix} 2.075 & 0.765 \\ 0.765 & 1.584 \end{pmatrix}$$

In Figures 1 and 2 the efficiency of the estimators is plotted against the number of returns an estimator has access to. Figure 1 displays the performance for asynchronously traded assets i and j that are observed without additive microstructure noise. Figure 2 shows the performance when the asynchronous returns are contaminated with microstructure noise.

From these graphs, we observe that for reasonable scenarios the MFFM comfortably outperforms the HY estimator unless a large number of intraday return observations on the individual assets is available. For instance, using 5-minute ($M = 78$) factor returns to estimate the 5×5 factor covariance matrix and 1 year ($T = 250$) of daily asset returns to estimate the 5×1 factor loading vector β, the MFFM delivers better estimates unless the HY estimator has access to more than 500 clean or 1250 noisy intraday (asynchronous) observations. The MFFM is also substantially more efficient than the realized covariance (lead–lag) estimator.

3 Monte Carlo Simulation

The theoretical results presented in the previous section demonstrate the superior properties of MFFM compared to existing covariance estimators in a bi-variate setting. An important additional feature of the MFFM is that its factor structure ensures stable and positive definite covariance matrices in higher dimensional settings. In this section we provide further insights into this property of the MFFM by means of an extensive simulation study. In addition to increasing the dimension of the covariance matrix to realistic magnitudes of several hundreds of assets, we relax some of the assumptions made in the previous section to study the effects of estimation errors in the factor exposures for individual stocks.

3.1 Simulation design

We simulate returns for asset i at high frequency as

$$\mathcal{R}_{i,t_j} = \mathcal{F}_{t_j} \beta_i + \varepsilon_{i,t_j} + \eta_{i,t_j} - \eta_{i,t_{j-1}}$$

where $i = 1, 2, \ldots, S$ (number of stocks), $j = 1, 2, \ldots, N_i$ (number of observations in a day), $0 \leq t_{j-1} < t_j \leq 1$, and $\mathcal{F}_{t_j}$ denotes the factor return between t_{j-1} and t_j. To ensure a realistic setup, we calibrate the data generating process (DGP) based on characteristics of the data used in the empirical application in Section 4. Specifically, the common factor $\mathcal{F}$ is a tri-variate Brownian motion with a covariance structure Λ as estimated for the daily Fama and French three-factor (market, size, and value) returns[5] over the period January 1998 through December 2007. The 3×1 vector of factor exposures β_i are obtained from regressing daily (corporate action adjusted) excess returns for each of the S&P500 constituents on the Fama and French three-factor returns, using the same sample period.

The idiosyncratic component $\varepsilon_{i,t_j} \sim$ i.i.d. $\mathcal{N}(0, \sigma_i^2(t_j - t_{j-1})/N_i)$ where σ_i^2 is the residual variance of the Fama and French regression for the i^{th} S&P500 constituent, the market microstructure noise component $\eta_{i,t_j} \sim$ i.i.d. $\mathcal{N}(0, \omega_i^2)$ where $\omega_i^2 = \frac{1}{4}(\beta_i' \Lambda \beta_i + \sigma_i^2)/N_i$,[6], and the observation times t_j are based on a Poisson process with intensity λ_i set to the average number of daily trades for the i^{th} S&P500 constituent, N_i.

This simulation setup ensures a realistic covariance structure of the 500-dimensional returns process at low frequency. At the same time, it incorporates non-synchronous trading and market microstructure noise at high frequency. We simulate second-by-second factor prices for a 6,5 hour trading day (23,400 seconds) and residuals to generate stock returns according to the DGP. The Poisson process in combination with the market microstructure noise then provide the simulated stock price paths.

3.2 Covariance models

For the simulated asset returns we estimate the covariance matrix using either MFFM or the realized covariance matrix.

3.2.1 Realized covariance

The realized covariance is a popular and efficient estimator of the latent integrated covariance. RC converges in probability to the integrated covariance in the absence of noise, see Barndorff-Nielsen and Shephard (2004). The RC is estimated as the cross-product of intraday returns:

$$RC = \mathcal{R}'\mathcal{R}$$

[5]Data available from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html.

[6]As shown in Oomen (2009), this level of noise is representative for the S&P500 universe.

where $\mathcal{R}$ is a $N \times S$ matrix of intraday returns. Here N is the number of non-overlapping intraday intervals where in each interval we take the last observed price. In case an interval has no price the last price of the previous interval is used, resulting in a zero return for that interval.

3.2.2 Mixed-frequency factor model

For the MFFM we need to estimate the factor loadings, the factor covariance matrix, and the residual variances. This will give us the MFFM-based covariance matrix as

$$\text{MFFM} = \tilde{\beta}'\widehat{\Lambda}\tilde{\beta} + \widehat{\Delta} \tag{8}$$

where $\widehat{\Lambda} = \mathcal{F}'\mathcal{F}$ is the estimated $K \times K$ realized factor covariance matrix, $\tilde{\beta}$ is the $K \times S$ matrix of factor loadings contaminated with i.i.d. measurement errors, and $\widehat{\Delta}$ is a $S \times S$ matrix with the estimated residual variances on the diagonal and zeroes elsewhere.

In empirical applications the factor covariances, residual variances and factor loadings can be estimated at different sampling frequencies. First, we propose to estimate the betas at the daily frequency. The main problem with estimating betas with intraday returns is that they can become severely biased towards zero due to the non-synchronicity of the relatively liquid factors and the considerably less liquid stocks. Also, Todorov and Bollerslev (2010) illustrate that jumps can cause single-day realized betas to exhibit erratic time-series behavior. We therefore propose a simple moving window history of 2.5 years of daily returns data that combined with OLS delivers betas that are smooth and by construction exhibit a much smaller variance than single-day realized betas while improving upon using monthly data.[7]

Second, the realized factor covariance matrix can be estimated at very high frequencies due to the high liquidity of ETF factor proxies. Third and finally, the residual variances can also be estimated using intraday data, but possibly at a lower frequency than the factor covariance matrix. This is to reduce the impact of the noise terms (η). We first compute the residuals, using $\varepsilon_{i,t_j} = \mathcal{R}_{i,t_j} - \mathcal{F}_{t_j}\tilde{\beta}_i$. Then we compute the variances of these residuals. While it is possible to use all intra-day returns for asset i for this purpose, due to market microstructure noise and the difference between the observation frequency for the factors and the stock prices these residual variances will be biased upwards. Below we examine to what extent lowering the sampling frequency to compute these residual variances reduces this bias.

[7] We have empirically experimented with the use of intraday data to estimate beta's. For sampling frequencies ranging from 15s to 65m we find that using intraday data to estimate beta's substantially increases the variance of the MFFM estimator. Aggregating realized betas to monthly or quarterly data and then applying EWMA smoothing helps to decrease the variance but the performance is inferior compared to using low-frequency beta's. Detailed results are available upon request.

3.3 Simulation results

As a measure of relative accuracy of the covariance estimates, we compute their distance to the true covariance matrix using the Frobenius norm. We do this separately for the diagonal and off-diagonal elements to disentangle the variance and covariance terms, i.e. we compute

$$\sum_{i=1}^{S} |\widehat{\Gamma}_{ii} - \Gamma_{ii}|^2 \qquad \text{and} \qquad 2\sum_{i=1}^{S}\sum_{j=i+1}^{S} |\widehat{\Gamma}_{ij} - \Gamma_{ij}|^2 \tag{9}$$

where $\Gamma = \beta'\Lambda\beta + \Sigma$, with Σ the diagonal matrix with the residual variances on the diagonal, and $\widehat{\Gamma}$ being either the MFFM or the RC covariance matrix estimate.

Non-synchronous prices, no noise
Figure 3 illustrates the performance of the RC and the MFFM when prices are non-synchronous but market microstructure noise is absent (i.e. $\omega_i = 0$). The covariance results illustrate that the MFFM has an excellent performance and is very robust across sampling frequencies. Furthermore, in contrast to RC, its performance is not affected by non-synchronicity.

– INSERT FIGURE 3 ABOUT HERE –

Non-synchronicity, however, does affect the MFFM variance estimates. This may seem counterintuitive at first as non-synchronicity usually affects the covariances and not so much the variances. The reason for the upward bias in the MFFM variances is caused by a mismatch between the very frequently observed factor returns and less frequently observed stock returns, which results in an additional quadratic bias term in the MFFM diagonal. The mismatch between liquid factors and less liquid stocks disappears when sampling at the 5-minute frequency or lower. Also note that with an increasing number of assets in a portfolio, the variance elements play a more limited role as the covariances become more dominant. For example, in the simulation with 500 stocks we only have 500 variances in contrast to 249,500 covariances. However, in some circumstances it may be interesting to introduce a third sampling frequency, that is, we can sample the residual returns at a lower sampling frequency than the sampling frequency for factor returns used to estimate the factor covariance matrix. We examine this possibility below.

Non-synchronous prices and market microstructure noise
Figure 4 illustrates the more practically relevant case where prices are non-synchronous but also contaminated by additive market microstructure noise.

– INSERT FIGURE 4 ABOUT HERE –

Market microstructure noise does not deteriorate the performance of both covariance estimators as the noise is (assumed to be) cross-sectionally independent. However, the noise does affect the variances computed with the MFFM (through the residual variances) and RC. For both estimators the diagonal elements perform fairly similar at the 5min and lower sampling frequencies while the MFFM covariances are substantially more efficient.

Lower frequency for residual variances to reduce MFFM variance bias
Finally, we examine the effects of introducing a third sampling frequency, that is, sampling the residual returns at a lower sampling frequency than the sampling frequency for factor returns used to estimate the factor covariance matrix.

Note that for the MFFM we use the assumption that the common factors fully capture the correlation among asset returns thus the residual returns only enter the MFFM by adding the diagonal residual variances to the systematic variances. Hence introducing a third frequency still delivers a well conditioned positive semi-definite covariance matrix.

Figure 5 illustrates how reducing the sampling frequency of the residual variances relative to the frequency used for the factor covariances can improve the efficiency of the variance elements in the MFFM. If the sampling frequency for the factor returns is ultra-high (sampling more frequently than once a minute) we use the 1-minute frequency to sample the residual variances to restore the efficiency of the variance elements in the MFFM. At sampling frequencies lower than the 1-minute frequency we use the same frequency for the factor covariances and residual variances. Using a lower sampling frequency than the 1-minute frequency to calculate residual risk is of course also possible to eliminate the bias but would deteriorate the performance of the MFFM as it also increases the variance of the estimates. This is the well-known trade-off in the efficiency of high-frequency data estimates between bias and precision.

– INSERT FIGURE 5 ABOUT HERE –

4 Empirical applications

We apply the MFFM approach to three universes of stocks with different levels of market capitalization to assess its empirical performance compared to the realized covariance and the factor model based on daily data. Whereas in the simulation experiments reported in the previous section we evaluated the (relative) accuracy of *measurements* of daily covariances, here we focus on the performance in terms of out-of-sample *forecasts*. In empirical applications the "true" covariances are unobservable. For this reason we focus on forecasts instead of covariance measurements. We do this in two ways. First, we evaluate the forecasting performance of the MFFM and RC for the volatility of vast dimensional equally-weighted portfolios. Second, we compare the out-of-sample performance by constructing minimum tracking error portfolios.

4.1 Data

Our data sets comprises the constituents of the S&P500 (large caps), S&P400 (mid caps) and S&P600 (small cap) indexes. For each index we only use those stocks that were included in the index during the complete sample period, which runs from May 1, 2004 until April 30, 2009. This leaves 442 large-caps, 342 mid-caps and 491 small-caps. We collect high-frequency data from November 1, 2006 onwards. Specifically, we sample National Best Bid Best Offer (NBBO) mid-points, originating from NYSE and NASDAQ only, at the 15-seconds sampling frequency. The first 2.5 years of the sample period are used only to obtain estimates of the factor loadings in the MFFM, for which we require only daily (close-to-close) returns.

4.2 Covariance estimators

Volatilities and correlations of stock returns typically are time-varying. We incorporate this feature explicitly in the methodology that is used to obtain covariance forecasts, as described in detail below.

4.2.1 Realized Covariance

In the portfolio volatility forecasting exercise with S stocks we use the traditional RC estimator to obtain an estimate of the covariance matrix on day t, that is,

$$RC_t = \mathcal{R}_t'\mathcal{R}_t, \tag{10}$$

where $\mathcal{R}_t$ is the $N \times S$ matrix of (intraday) stock returns on day t.

In the minimum tracking error application we employ intraday *excess* stock returns net of the relevant benchmark, which for each of the three universes is taken to be the corresponding S&P index. The *active* realized covariance estimator is then computed as

$$RC_t^A = (\mathcal{R}_t - \mathcal{R}_{Mt}e)'(\mathcal{R}_t - \mathcal{R}_{Mt}e), \tag{11}$$

where $\mathcal{R}_{Mt}$ is a $N \times 1$ vector of intraday returns on the corresponding index, and e is an $S \times 1$ vector of ones. In both cases we include overnight returns by adding the outer product of the vector of *close-to-open* (active) returns.

Finally, we consider the RC_t and RC_t^A estimators for a range of intra-day sampling frequencies, equal to 15 seconds, 1, 5, 15, 30, 65 and 130 minutes. We also include the sample realized covariance based on daily *close-to-close* returns.

4.2.2 Mixed-frequency factor models

For the MFFM approach we employ a 12-factor model based on the Fama and French (1993) size and value factors and ten industry factors. The motivation to use 10 industry factors is that many stocks have activities in (and thus exposure to) multiple sectors, see Grinold and Kahn (2000), page 60. We allow for time-varying factor loadings, which are estimated using a moving window of 2.5 years (632 days) of daily *close-to-close* returns[8] by means of the regression

$$R_{i,t-j} = F_{t-j}\beta_{i,t} + \varepsilon_{i,t-j}, \qquad \text{for } j = 0, 1, \dots, L-1, \tag{12}$$

where R_{it} is a vector of daily returns on stock i, $F_t = [SMB_t \quad HML_t \quad I_1 \quad \dots \quad I_{10}]$ is a matrix of factor returns on the size (Small-Minus-Big), value (High-Minus-Low) and industry factors, and L denotes the length of the moving window. The intraday residuals needed to compute idiosyncratic variances are obtained as

$$\hat{\varepsilon}_t = \mathcal{R}_t - \mathcal{F}_t\hat{\beta}_{t-1}, \tag{13}$$

Finally the MFFM covariance matrix estimate for day t is then computed as

$$\text{MFFM}_t = \hat{\beta}_{t-1}'\hat{\Lambda}_t\hat{\beta}_{t-1} + diag(\hat{\varepsilon}_t'\hat{\varepsilon}_t), \tag{14}$$

where $\hat{\Lambda}_t = \mathcal{F}_t'\mathcal{F}_t$ is the factor covariance matrix. The motivation to use 'lagged' factor loading estimates $\hat{\beta}_{t-1}$ (that is, based on the moving window that ends on day $t-1$) rather than $\hat{\beta}_t$ stems from assumption *(iv)* in Theorem 2.1. stating that the measurement errors in the factor loadings and in the factor covariance matrix are orthogonal, i.e. $\beta^\varepsilon \perp \Lambda^\varepsilon$. By lagging the beta estimates in (13) and (14) we avoid the possibility that measurement errors in factor loadings are correlated with the measurement errors in the factor covariance matrix.[9] For the minimum tracking error application we follow the same approach, except that we use stock returns in excess of the returns on the relevant market index. Hence, we obtain estimates of the factor loadings

[8] In earlier studies on factor models the number of observations used for estimating betas is usually 3 to 5 years. Here we use 2.5 years as using a longer history would limit the number of constituents that survived our sample period, thereby reducing the dimension of the covariance matrix.

[9] We have experimented with using $\hat{\beta}_t$ instead of $\hat{\beta}_{t-1}$, finding that this deteriorates the performance of the MFFM forecasts (although the differences are small).

from the regression

$$R_{i,t-j} - R_{M,t-j} = F_{t-j}\beta_t^A + \varepsilon_{i,t-j}^A, \qquad j = 0, 1, \ldots, L-1, \tag{15}$$

while we compute the *active* intraday residuals as

$$\hat{\varepsilon}_t^A = \mathcal{R}_t - \mathcal{R}_{M,t} - \mathcal{F}_t \hat{\beta}_{t-1}^A, \tag{16}$$

and the MFFM estimator for day t using (14). We include overnight returns in the factor covariance matrix $\hat{\Lambda}_t$ by adding the outer product of the vector of *close-to-open* factor returns, similar to including overnight stock returns in the realized covariance. For the idiosyncratic variances we also include the (active) residual overnight returns throughout the empirical analysis.

In the MFFM estimator in (14), we consider the same range of intra-day sampling frequencies for the factor covariance matrix and the idiosyncratic variances as used for the realized covariance estimator given in the previous subsection. Also, we include a conventional 'low-frequency' factor model where all parts of (14) are based on daily *close-to-close* returns.

4.3 Covariance matrix forecasts

We consider forecasts based on an exponentially weighted moving average (EWMA) scheme, motivated by the work of Foster and Nelson (1996) and Andreou and Ghysels (2002). In this framework, the covariance matrix forecast for day t, denoted $\Sigma_{t|t-1}$, is given by

$$\Sigma_{t|t-1} = \alpha \Sigma_{t-1|t-2} + (1-\alpha)\widehat{\Sigma}_{t-1}, \tag{17}$$

where the scalar α is a fixed decay parameter and $\widehat{\Sigma}_{t-1}$ is the covariance matrix estimate for day $t-1$ as given by either the RC estimator in (10) (or (11) in the minimum tracking error application) or the MFFM estimator in (14). We consider several weighting schemes with $\alpha \in \{0.94, 0.75, 0.50, 0.25\}$. The value of 0.94 for α is the optimal decay parameter for daily data documented by RiskMetrics (see e.g. Zumbach (2006)). The use of smaller decay parameters allows us to examine the effects on the forecasting performance when putting more weight on more recent data. Smaller levels of α are also closer to our simulation study where in fact $\alpha = 0$. Further, using smaller values of α provides more insight in the quality of the covariance estimator itself rather than the 'smoothed' forecast. Less smoothing can be important also from an economic point of view, as it enables the forecasts to adjust more rapidly to important changes in (co)variance dynamics, which for example occur at turning points between periods of high and low volatility.

We use the period from November 1, 2006 until December 31, 2006 as 'burn-in period' for the covariance dynamics in (17) and exclude these two months in the performance evaluations below. The out-of-sample period therefore runs from January 3, 2007 until April 30, 2009.

4.4 Equally-weighted portfolios

In our first forecasting exercise, we consider equally-weighted portfolios for the S&P500, S&P400 and S&P600 stock universes. As noted before, we only use the S constituents that were included in a single index during the complete sample period. For each universe the daily equally-weighted portfolio return is computed as $r_{p,t} = e'r_t$ where r_t is an $S \times 1$ vector of *close-to-close* returns on the individual stocks and e is the equal-weight vector with entries $1/S$. We obtain one-day ahead forecasts of the volatility of these equally-weighted portfolios as $\hat{\sigma}^2_{P,t|t-1} = e'\Sigma_{t|t-1}e$, using the MFFM- and RC-based covariance matrix forecasts from (17).

We evaluate the accuracy of the volatility forecasts in two ways. First, we run Mincer-Zarnowitz (MZ) regressions, in which the portfolio volatility proxy $\hat{\sigma}^2_{p,t|t}$ is regressed on a constant and one of the volatility forecasts, that is,

$$\hat{\sigma}^2_{p,t|t} = \gamma + \delta\hat{\sigma}^2_{p,t|t-1} + \varepsilon_t. \tag{18}$$

Here we use the squared daily return $r^2_{p,t}$ as the volatility proxy. Although this proxy is known to be noisy, at least it is unbiased. Obvious alternatives would be to use the RC or MFFM estimates of the covariance matrix for day t, but this might bias the MZ regression towards one of the forecasts. Using the squared daily return avoids this issue.[10]

In addition we report results for forecast encompassing regressions where the squared daily return is regressed on the MFFM-based forecast $\hat{\sigma}^2_{p,MFFM,t|t-1}$ and a competing forecast $\hat{\sigma}^2_{p,X,t|t-1}$, that is,

$$r^2_{p,t} = \gamma + \delta_1\hat{\sigma}^2_{p,MFFM,t|t-1} + \delta_2\hat{\sigma}^2_{p,X,t|t-1} + \varepsilon_t. \tag{19}$$

These regressions can be used to obtain insights in how well the MFFM approach empirically competes with existing forecast methods. We consider two competing forecasts X, namely the RC at the same intraday sampling frequency as used for the MFFM and the daily factor model,

[10] Using the RC and the MFFM estimator based on a 5 min sampling frequency as the volatility proxy does not alter the main conclusions as reported here. The main difference is that we obtain higher regression R^2's that are about 10 to 15% higher than the R^2 for the daily squared return. In addition, we have considered the MZ regression using the absolute return as dependent variable (which then is regressed on a constant and the square root of $\hat{\sigma}^2_{p,t|t-1}$. This also results in higher R^2 values than those reported here (by about 5%), mostly because the absolute return is more robust to outliers. However, using this transformation of the variance does not lead to consistent forecast rankings when the forecast target is the conditional variance, see Patton (2011).

denoted FM. Regression R^2's and coefficients are reported and statistically significant coefficients at the 5% level are displayed in bold fonts.

Figures 6–8 illustrates that the RC and MFFM provide very similar dynamics at the 5 min sampling frequency. In addition we observe that the estimates obtained with high-frequency data for the RC and MFFM are much more precise than their daily counterparts. The daily sample covariance and daily factor model estimates are "noisy". The equally-weighted portfolio volatility estimates are all plotted against the (scaled) daily absolute return.

We run Mincer-Zarnowitz and encompassing regression results with decay parameter $\alpha = \{0.94, 0.75, 0.50, 0.25, 0.00\}$. For space considerations we only report results for $\alpha = 0.94$ since the results for the other settings of α lead to similar conclusions. The only exception is that the performance of the daily factor model deteriorates rapidly for smaller α. This occurs because the daily factor model is based on only one observation per day and therefore require a longer history of covariance estimates to compete with the estimators based on higher sampling frequencies.

– INSERT FIGURES 6–8 ABOUT HERE –

Table 2 summarizes the results for the S&P500 Mincer-Zarnowitz and encompassing regressions. Based on the Mincer-Zarnowitz regressions we find that the EWMA forecasts for the volatility of the equally-weighted portfolio have statistically significant coefficients. The constants, frequently interpreted as forecast bias, are statistically insignificant across all sampling frequencies. From the regression R^2's we learn that the results for RC and MFFM are very close, indicating that our factor structure indeed does a good job, and for both the RC and the MFFM we find that using high-frequency data improves the R^2 by about 3%.[11]

– INSERT TABLE 2 ABOUT HERE –

For the relatively liquid S&P500 encompassing regressions we find that the RC and MFFM forecasts do not encompass each other. The bias and loadings on the forecast have statistically insignificant Newey-West t-statistics at the 5% level. Using similar encompassing regressions, but now for the MFFM sampled at each intraday frequency against the daily factor model (FM), we find that the daily factor model forecasts are encompassed by the MFFM forecasts at each intraday sampling frequency. The improvement in regression the regression R^2 compared to regressing on MFFM only (see Panel B in Table 2) is also small.

[11]By lowering α the forecasting performance of the daily counter parts of the RC and MFFM deteriorates rapidly and regression coefficients become close to zero if we do not apply EWMA to generate forecasts due to the high variance of estimators based on daily data as displayed in Figure 6. The differences in R^2 between daily and high-frequency data when using non-smoothed estimates ($\alpha = 0$) are about 15%.

– INSERT TABLE 3 ABOUT HERE –

For the S&P400 mid-cap universe Table 3 summarizes the forecast regression results. Using Mincer-Zarnowitz regressions we find that the forecasts based on RC and MFFM are statistically significant at each sampling frequency and the forecast bias is not significant. The regression R^2's for the MFFM regressions are higher than for the RC regressions. In the encompassing regression results for MFFM and RC (Panel C) we observe that, at high sampling frequencies, between 15 sec and 30 min, the MFFM forecasts encompass the RC forecasts if we use the squared daily return as proxy. Differences increase by moving from the relatively liquid S&P500 stocks to the less liquid S&P400 stocks where non-synchronicity plays a more important role. In line with the results for the S&P500 we find for the S&P400 universe that the daily factor model forecasts are encompassed by MFFM and this holds at every intraday sampling frequency, and adding the daily factor model forecasts to MFFM forecasts only improves the regression R^2 by about a half percent.

– INSERT TABLE 4 ABOUT HERE –

When we move to the relatively illiquid S&P600 constituents we observe in Table 4 that for the Mincer-Zarnowitz regressions the forecasts of RC and MFFM are significant at every frequency and the forecast bias is not. Similar to the S&P400 results we find that the regression R^2 for MFFM is higher than for RC. Using encompassing regressions we find that the MFFM forecasts are favored over the RC forecasts at very high frequencies between 15s and 1m. Consistent with the results for the S&P500 and S&P400 the MFFM forecasts obtained using intraday sampling encompass the factor model based on daily data.

4.5 Minimum tracking error portfolios

Given the one day ahead EWMA forecasts of the covariance matrices we construct minimum TE portfolios by calculating the standard fully-invested minimum variance portfolios (when using the *active* covariance matrix as we do here, then the minimum TE portfolio is the minimum variance portfolio):

$$w_t = \frac{\Sigma_{t|t-1}^{-1} e}{e' \Sigma_{t|t-1}^{-1} e} \tag{20}$$

where e is a $S \times 1$ vector of ones and Σ_t is the EWMA conditional covariance matrix forecast of RC or MFFM. The daily minimum TE portfolio returns are obtained by computing $R_{Pt} = w_t' r_t$ where r_t is the vector of daily stock returns. We calculate the *ex-post* tracking error using daily returns $\text{TE} = \text{Std}(R_P - R_M)$ and compare the results for the RC and the MFFM.

In this application we keep track of the daily turnover in the portfolio weights w_t which is directly associated with the transaction costs that an investor faces who wishes to re-balance his or her portfolio daily. We compute turnover by summing the absolute daily weight changes over the stock names,

$$TO_t = |w_t - w_{t-1}|'e. \tag{21}$$

We expect that a covariance estimator that is well-conditioned and numerically stable will result in smaller daily portfolio turnover. The daily turnover will also be related to the decay parameter α in (17) which is used to generate EWMA forecasts. A large decay parameter implies that more weight is assigned to historical estimates whereas a smaller decay parameter corresponds to assigning more weight to the most recent estimate(s). More weight on historical estimates will make an estimator more stable and cause less portfolio turnover but on the other hand, recent shifts in for example market volatility will be picked up at a slower pace.

4.6 Minimum tracking error results

Table 5 illustrates the performance in terms of annualized minimum tracking errors for the S&P500 large caps. Consistent with the simulation results we find that the MFFM covariance matrix estimator is remarkably robust across sampling frequencies indicating that, in contrast to RC, the factor covariance matrix can be estimated at very high frequencies as the level of market microstructure noise and non-synchronicity in the factors is relatively small compared to individual stocks. At almost each of the considered sampling frequencies and forecast weights α, the MFFM produces better results than RC. However, for the relatively liquid S&P500 universe the RC competes with the MFFM if we use a sampling frequency between 15s and 15m combined with $\alpha = 0.94$ for RC but deteriorates rapidly by putting more weight on the most recent estimates (lower α's). The difference between the MFFM and RC are small when we choose the best sampling frequency and forecasts weights for RC, but the differences are substantial on average across these settings. The covariance matrices considered here have a dimension of 442 and we find that at sampling frequencies of 30min and lower the RC is not well-conditioned and therefore not invertible, we indicate this with "NA". The naïvely diversified equally-weighted portfolio, advocated recently by DeMiguel et al. (2009), achieves a tracking error equal to 0.099 and is outperformed by the MFFM in each of the parameter settings and by most parameter settings for the RC given that these settings result in an invertible covariance matrix forecast.

– INSERT TABLE 5 ABOUT HERE –

Important differences in numerical stability of the covariance matrix forecasts are exemplified by the very large differences in portfolio turnover. At most of the sampling frequencies the difference is at least 8 times larger. This indicates, as noted by Fan et al. (2008), that using (sample) realized covariances for portfolio optimization can be "tricky" for vast dimensional portfolios. In contrast, due to its factor structure and the use of relatively liquid factors, the MFFM delivers exceptionally small levels of turnover associated with tracking errors that are at par with the best results for the realized covariance and outperform the realized covariance at all other settings without having to resort to putting a lot of weight on historical estimates. In fact the MFFM is found to be relatively insensitive to the choice of α and the sampling frequency. It is interesting to observe that on average the MFFM tracking errors are fairly constant across sampling frequencies and decay parameters. This is due to the factor structure which ensures stability of the covariance matrix. The level of turnover, however, does depend on the sampling frequency and decay parameter. The highest sampling frequencies produce very small levels of turnover because the (factor) covariance estimates are very precise. Applying a higher decay parameter, i.e. less weight on recent data, further smooths the covariance matrix and therefore reduces turnover. The lower sampling frequencies produce less precise (factor) covariance estimates and therefore higher levels of turnover. For lower sampling frequencies the covariance estimates are accurate on average due to the factor structure but they are less precise than when higher sampling frequencies are used.

For the S&P400, see Table 6, we decrease the naïve equally-weighted portfolio tracking error of 9.3% to 8.3% with RC and this result depends heavily on the forecast weighting scheme and sampling frequency. Using the MFFM further decreases the tracking error to 7.8% with results being robust. As expected, the tracking errors have increased for the S&P400 mid caps compared to the S&P500 large caps, see also Table 1 for the average number of trades per day in each S&P universe. Higher levels of non-synchronicity and microstructure noise in individual stocks explain this result. Similar to the portfolio turnover results for the S&P500 universe we find that the RC portfolios cause a daily turnover which is at least a factor 10 times larger than the turnover in the MFFM portfolios.

– INSERT TABLE 6 ABOUT HERE –

For the S&P600 small caps, where the level of non-synchronicity plays a more important role than for mid- and large-caps, we find larger tracking errors when using the RC because it is sensitive to market microstructure frictions and the increased portfolio dimension. The tracking errors for the RC are in fact larger, for every combination of sampling frequency and forecast weighting scheme, than that of the naïve $1/N$ portfolio which achieves a tracking error of 8.9%. The MFFM, however, achieves smaller tracking errors than it achieves for the S&P400 mid-caps, indicating that its performance is not compromised by the illiquidity of the S&P600 universe.

The MFFM comfortably decreases the best RC tracking error, being 9.1%, to 7.0% and the MFFM easily outperforms the naïve portfolio for all combinations of sampling frequencies and forecast weighting schemes. In line with the turnover results for the S&P500 and S&P400 we find that the RC portfolios have a turnover that is 12 times larger.

– INSERT TABLE 7 ABOUT HERE –

Note that outperforming the equally-weighted portfolio is not necessarily an easy task. DeMiguel et al. (2009) analyze various advanced methods consisting of Bayesian estimation, shrinkage, robust allocation etc. and find that none of the 14 models they implement can consistently outperform the $1/N$ portfolio. Hence, the fact that the MFFM consistently outperforms the $1/N$ and RC portfolios is encouraging support. Further, the results in Madhavan and Yang (2003) illustrate that using the sample (realized) covariance matrix for unrestricted optimization, results in a performance that is worse than the equally-weighted portfolio.

5 Conclusion

Recently there has been great interest in the use of high-frequency data to estimate variances and covariances. The advantage is that the use of high-frequency data results in more accurate covariance estimates, but on the other hand it also brings problems such as microstructure noise which reduces the efficiency of covariance estimators based on intraday data and non-synchronous trading leading to covariance estimates being biased towards zero. What so far has been lacking is to bring the merits of high-frequency data to factor models. With the introduction of exchange-traded funds important factors are now traded much more actively than individual stocks. For example the S&P500 ETFs (Spiders) have on average traded 18 times more frequently than the average individual stock in the S&P500. In this study we have proposed the Mixed Frequency Factor Model. In particular we can use ultra high-frequency data for ETFs to obtain a very accurate estimate of the factor covariance matrix, as prices are observed essentially free of noise. We use daily data to estimate the factor loadings conservatively to avoid problems inherent in the use high-frequency data for illiquid stocks and non-synchronicity biases between the returns on factors and stocks. Furthermore we take advantage of the facts that factor models can easily be applied to vast numbers of assets and that covariance matrices from factor models are less prone to error maximization in portfolio construction problems. Using Mincer-Zarnowitz and encompassing regressions we find that the MFFM portfolio volatility forecasts improve upon the daily factor and realized covariance forecasts when the forecasts are ranked on R^2 and as indicated by the positive weights on the MFFM versus negative weights on the RC and daily factor model. Adding a RC or daily factor model forecast to a MFFM forecast only improves the regression R^2 marginally. In a minimum tracking error application we reduce the tracking errors by using the MFFM rather than RC for computing the covariance matrix. The differences between RC and MFFM increase with the level of non-synchronicity between individual stocks, i.e. we find a larger difference when considering the S&P600 small caps than when we consider the S&P500 large caps. The RC outperforms the naïvely diversified equally-weighted $1/N$ portfolios when considering large- and mid-caps but fails by a substantial margin for the illiquid S&P600 small caps. The MFFM comfortably outperforms the $1/N$ portfolios regardless of the universe considered. For realized covariance the results in the tracking error applications depend severely on the sampling frequency and the weighting scheme applied to the past daily covariance matrices. In contrast, the performance of the MFFM is robust across sampling frequencies and weighting schemes and consistently outperforms RC and the naïve $1/N$ portfolios.

References

Andersen, T., Bollerslev, T., Christoffersen, P. and Diebold, F.: 2006a, Volatility and correlation forecasting, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier, North-Holland, pp. 777–878.

Andersen, T., Bollerslev, T. and Wu, G.: 2006b, Realized beta: Persistence and predictability, *Advances in Econometrics* **20**, 1–39.

Andreou, E. and Ghysels, E.: 2002, Rolling-sample volatility estimators: Some new theoretical, simulation and empirical results, *Journal of Business and Economic Statistics* **20**, 363–376.

Bandi, F. and Russell, J.: 2005, Realized covariation, realized beta, and microstructure noise. University of Chicago, Working paper.

Barndorff-Nielsen, O. E. and Shephard, N.: 2004, Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics, *Econometrica* **72**, 885–925.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N.: 2011, Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* **162**, 149–169.

Chamberlain, G. and Rothschild, M.: 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* **51**(5), 1281–1304.

Chan, L., Karceski, J. and Lakonishok, J.: 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* **12**(Winter), 937–974.

Chen, N.-F., Roll, R. and Ross, S. A.: 1986, Economic forces and the stock market, *Journal of Business* **59**, 383–403.

Connor, G., Goldberg, L. and Korajczyk, R.: 2012, *Portfolio Risk Management*, Princeton University Press.

Connor, G. and Korajczyk, R.: 1994, A test for the number of factors in an approximate factor model, *Journal of Finance* **48**(4), 1263–1291.

DeMiguel, V., Garlappi, L. and Uppal, R.: 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* **22**, 1915–1953.

Fama, E. and French, K.: 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**, 3–56.

Fan, J., Fan, Y. and Lv, J.: 2008, High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* **147**, 186–197.

Foster, D. and Nelson, D.: 1996, Continuous record asymptotics for rolling sample variance estimators, *Econometrica* **64**(3), 139–174.

Griffin, J. E. and Oomen, R. C. A.: 2011, Covariance measurement in the presence of non-synchronous trading and market microstructure noise, *Journal of Econometrics* **160**, 58–68.

Grinold, R. and Kahn, R.: 2000, Active portfolio management, a quantitative approach for producing superior returns and controlling risk, *2nd Edition, McGraw-Hill* .

Halbleib, R. and Voev, V.: 2011, Forecasting covariance matrices: A mixed frequency approach. University of Brussels, Working Paper.

Hansen, P., Lunde, A. and Voev, V.: 2010, Realized beta garch: A multivariate garch model with realized measures of volatility and covolatility. University of Aarhus, Working Paper.

Hautsch, N., Kyj, L. and Malec, P.: 2011, The merit of high-frequency data in portfolio allocation. Humboldt-University Berlin, Working Paper.

Hautsch, N., Kyj, L. and Oomen, R.: 2012, A blocking and regularization approach to high dimensional realized covariance estimation, *Journal of Applied Econometrics* **27**, 625–645.

Hayashi, T. and Yoshida, N.: 2005, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* **11**, 359–379.

Ingersoll, J.: 1984, Some results in the theory of arbitrage pricing, *Journal of Finance* **39**(4), 1021–1039.

Jagannathan, R. and Ma, T.: 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* **58**(4), 1651–1684.

Kyj, L., Ostdiek, B. and Ensor, K.: 2009, Covariance estimation in dynamic portfolio optimization: A realized single factor model. Working paper.

Madhavan, A. and Yang, J.: 2003, Practical risk management, *Journal of Portfolio Management* **30**, 73–85.

Martens, M.: 2006, Estimating ubiased and precise realized covariances. Erasmus University Rotterdam, Working Paper.

McAleer, M. and Medeiros, M.: 2008, Realized volatility: a review, *Econometric Reviews* **27**, 10–45.

Noureldin, D., Shephard, N. and Sheppard, K.: 2012, Multivariate high-frequency-based volatility (heavy) models, *Journal of Applied Econometrics* . Forthcoming.

Oomen, R. C. A.: 2009, A universal scaling law of noise in financial markets. working paper, University of Amsterdam.

Patton, A.: 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* **160**(1), 246–256.

Rosenberg, B.: 1974, Extra-market components of covariance in security returns, *Journal of Financial and Quantitative Analysis* **9**, 263–274.

Ross, S.: 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* **13**, 341–360.

Todorov, V. and Bollerslev, T.: 2010, Jumps and betas: A new framework for disentangling and estimating systematic risks, *Journal of Econometrics* **157**(2), 220–235.

Voev, V. and Lunde, A.: 2007, Integrated covariance estimation using high-frequency data in the presence of noise, *Journal of Financial Econometrics* **5**(1), 68–104.

Zhang, L.: 2011, Estimating covariation: Epps effect and microstructure noise, *Journal of Econometrics* **160**(1), 33–47.

Zhang, L., Mykland, P. and Aït-Sahalia, Y.: 2005, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100**, 1394–1411.

Zumbach, G.: 2006, The riskmetrics 2006 methodology, *Available at SSRN: http://ssrn.com/abstract=1420185* .

A Proofs

Proof of Theorem 2.1 Using the notation $\widehat{X} = X + X^{\varepsilon}$, we have for $i \neq j$:

$$\widehat{\gamma}_{ij} = \beta_i'\Lambda\beta_j + \beta_i'\Lambda\beta_j^{\varepsilon} + \beta_i'\Lambda^{\varepsilon}\beta_j + \beta_i'\Lambda^{\varepsilon}\beta_j^{\varepsilon} + \beta_i^{\varepsilon\prime}\Lambda\beta_j + \beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon} + \beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j + \beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j^{\varepsilon}.$$

Assumption (i) implies $\beta_i^{\varepsilon} \perp \beta_j^{\varepsilon}$ so that $E(\beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon}) = 0$. All other terms, except $\beta_i'\Lambda\beta_j$, are zero in expectation by assumptions (ii-iv). Hence, we have unbiasedness. To compute the variance of this estimator, note that all terms are mutually uncorrelated, so that the variance of the sum is the sum of the variances.

$$\begin{aligned}
V(\beta_i'\Lambda\beta_j^{\varepsilon}) &= \beta_i'\Lambda\Sigma_{\beta,j}\Lambda'\beta_i \\
V(\beta_i'\Lambda^{\varepsilon}\beta_j) &= E(\beta_i'\Lambda^{\varepsilon}\beta_j\beta_j'\Lambda^{\varepsilon\prime}\beta_i) = E(\mathrm{tr}(\beta_i\beta_i'\Lambda^{\varepsilon}\beta_j\beta_j'\Lambda^{\varepsilon\prime})) = g(\beta_i\beta_i', \beta_j\beta_j', \Phi) \\
V(\beta_i'\Lambda^{\varepsilon}\beta_j^{\varepsilon}) &= E(\beta_i'\Lambda^{\varepsilon}\beta_j^{\varepsilon}\beta_j^{\varepsilon\prime}\Lambda^{\varepsilon\prime}\beta_i) = E(\beta_i'\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime}\beta_i) = E(\mathrm{tr}(\beta_i\beta_i'\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime})) = g(\beta_i\beta_i', \Sigma_{\beta,j}, \Phi) \\
V(\beta_i^{\varepsilon\prime}\Lambda\beta_j) &= \beta_j'\Lambda'\Sigma_{\beta,i}\Lambda\beta_j \\
V(\beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon}) &= E(\beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon}\beta_j^{\varepsilon\prime}\Lambda'\beta_i^{\varepsilon}) = E(\beta_i^{\varepsilon\prime}\Lambda\Sigma_{\beta,j}\Lambda'\beta_i^{\varepsilon}) = \mathrm{tr}(\Sigma_{\beta,i}\Lambda\Sigma_{\beta,j}\Lambda') \\
V(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j) &= E(\beta_j'\Lambda^{\varepsilon}\Sigma_{\beta,i}\Lambda^{\varepsilon\prime}\beta_j) = E(\mathrm{tr}(\beta_j\beta_j'\Lambda^{\varepsilon}\Sigma_{\beta,i}\Lambda^{\varepsilon\prime})) = g(\beta_j\beta_j', \Sigma_{\beta,i}, \Phi) \\
V(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j^{\varepsilon}) &= E(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j^{\varepsilon}\beta_j^{\varepsilon\prime}\Lambda^{\varepsilon\prime}\beta_i^{\varepsilon}) = E(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime}\beta_i^{\varepsilon}) = E(\mathrm{tr}(\Sigma_{\beta,i}\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime})) = g(\Sigma_{\beta,i}, \Sigma_{\beta,j}, \Phi)
\end{aligned}$$

All terms involving Λ^{ε} are of the form $E(\mathrm{tr}(AZBZ))$ where A, B, and Z are square symmetric matrices of equal dimension with A and B fixed and Z random with $E(\mathrm{vech}(Z)\mathrm{vech}(Z)') = \Phi$. Define $\overline{A} = AZ$ and $\overline{B} = BZ$ with

$$\overline{A}_{ij} = \sum_k A_{ik}Z_{kj} \qquad \text{and} \qquad \overline{B}_{ij} = \sum_m B_{im}Z_{mj}.$$

Then

$$E(\mathrm{tr}(AZBZ)) = \mathrm{tr}(E(\overline{A}\,\overline{B})) = \sum_{i,j} E(\overline{A}_{ij}\overline{B}_{ji}) = \sum_{i,j,k,m} A_{ik}B_{jm}E(Z_{kj}Z_{mi}) = \sum_{i,j,k,m} A_{ik}B_{jm}\Phi_{f(k,j),f(m,i)}.$$

■

Proof of Corollary 2.2 Given the assumptions, we have $\Sigma_{\beta,i} = \frac{1}{T}\sigma_i^2\Lambda^{-1}$ (asymptotically). Thus, $\beta_j'\Lambda\Sigma_{\beta,i}\Lambda'\beta_j = \frac{1}{T}\sigma_i^2\beta_j'\Lambda\beta_j$ and $\mathrm{tr}(\Sigma_{\beta,i}\Lambda\Sigma_{\beta,j}\Lambda') = \frac{1}{T^2}\sigma_i^2\sigma_j^2\mathrm{tr}(I_K) = \frac{K}{T^2}\sigma_i^2\sigma_j^2$. Combining this, gives term A. For term B, note that

$$\begin{aligned} g(A, B, \Phi) &= \sum_{m,n,p,q}^{N} A_{mp} B_{nq} E((\widehat{\Lambda}_{pn} - \Lambda_{pn})(\widehat{\Lambda}_{qm} - \Lambda_{qm})) \\ &= \sum_{m,n,p,q}^{N} A_{mp} B_{nq} (E(\widehat{\Lambda}_{pn} \widehat{\Lambda}_{qm}) - \Lambda_{pn} \Lambda_{qm}) \\ &= \frac{1}{M} \sum_{m,n,p,q}^{N} A_{mp} B_{nq} (\Lambda_{pq} \Lambda_{nm} + \Lambda_{pm} \Lambda_{nq}) \end{aligned}$$

using that for a multivariate normal random variable x with characteristic function $\ln \phi(\xi) = -\xi' \Sigma \xi / 2$ we have

$$E(\widehat{\sigma}_{mn} \widehat{\sigma}_{pq}) = \sigma_{mn} \sigma_{pq} + \frac{\sigma_{mp} \sigma_{nq} + \sigma_{mq} \sigma_{np}}{M}$$

where

$$\widehat{\sigma}_{mn} \equiv \frac{1}{M} \sum_{i=1}^{M} x_i^{(m)} x_i^{(n)}$$

B Tables

Table 1: Description of ETF contracts

ticker	description	sector / style classification	# trades per day
XLE.A	Energy Sector SPDR Fund	Energy	76,392
XLB.A	Materials Sector SPDR Fund	Materials	16,708
XLI.A	Industrial Sector SPDR Fund	Industrials	12,207
XLY.A	Consumer Discretionary Sector SPDR Fund	Consumer Discretionary	9,731
XLP.A	Consumer Staples Sector SPDR Fund	Consumer Staples	6,153
XLV.A	Health Care Sector SPDR Fund	Health Care	6,697
XLF.A	Financial Sector SPDR Fund	Financials	112,191
XLK.A	Technology Sector SPDR Fund	Information Technology	9,243
IYZ.N	iShares Telecommunications Sector Fund	Telecommunications	762
XLU.A	Utilities Sector SPDR Fund	Utilities	11,753
SPY.A	SPDR Trust Series 1	Large Cap	356,876
IWM.A	iShares Russell 2000 Index Fund	Small Cap	140,192
IVE.N	S&P 500 Value Index Fund	Value	3,030
IVW.N	S&P 500 Growth Index Fund	Growth	3,912
	Average across ETFs		54,703
	Average across S&P400 constituents		4,898
	Average across S&P500 constituents		19,395
	Average across S&P600 constituents		1,990

Note: This table lists the ETF contracts used in the empirical analysis, together with the average number of trades per day over the period January 2007 through April 2009. The "SMB" ("HML") factor is specified as IWM.A - SPY.A (IVE.N - IVW.N).

Table 2: S&P500 Portfolio volatility, Mincer-Zarnowitz and encompassing regressions

	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RC	**1.599**	**1.350**	**1.233**	**1.226**	**1.154**	**1.136**	**1.010**	**0.906**
R^2	0.241	0.242	0.244	0.246	0.244	0.245	0.229	0.205
Panel B: MFFM Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**1.490**	**1.366**	**1.299**	**1.311**	**1.253**	**1.233**	**1.120**	**1.036**
R^2	0.242	0.244	0.247	0.249	0.248	0.247	0.232	0.208
Panel C: MFFM + RC Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	1.236	4.353	4.179	3.745	3.591	2.091	1.841	2.230
RC	0.272	-2.968	-2.752	-2.296	-2.174	-0.795	-0.656	-1.056
R^2	0.242	0.246	0.250	0.252	0.250	0.248	0.232	0.209
Panel D: MFFM + FM Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
MFFM	**3.315**	**3.284**	**3.288**	**3.470**	**3.157**	**2.445**	**1.656**	
FM	-1.396	-1.609	-1.762	-1.904	-1.753	-1.141	-0.536	
R^2	0.257	0.264	0.271	0.277	0.272	0.261	0.234	

Note: This Table summarizes the results for Mincer-Zarnowitz and encompassing regressions using the daily squared portfolio return as unbiased proxy for the latent portfolio variance. The evaluation is based on 442 of the S&P500 constituents to forecast the variance of the equally-weighted portfolio one day ahead using EWMA covariance matrix forecasts with decay parameter $\alpha = 0.94$. Compared are the volatility forecasts generated with the MFFM, RC and the daily factor model. The out-of-sample period is Jan. 2007 – Apr. 2009. Coefficients that are statistically significant at the 5% level, based on Newey-West standard errors with 20 lags, are displayed in bold fonts.

Table 3: S&P400 Portfolio volatility, Mincer-Zarnowitz and encompassing regressions

	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RC	**2.105**	**1.594**	**1.369**	**1.365**	**1.257**	**1.220**	**1.064**	**0.915**
R^2	0.250	0.249	0.255	0.258	0.254	0.259	0.244	0.221
Panel B: MFFM Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**1.488**	**1.363**	**1.283**	**1.304**	**1.231**	**1.196**	**1.081**	**0.993**
R^2	0.259	0.259	0.262	0.266	0.266	0.266	0.252	0.230
Panel C: MFFM + RC Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**9.158**	**8.508**	**5.035**	**4.407**	**4.207**	2.469	2.362	3.429
RC	-11.051	**-8.535**	-4.070	-3.310	-3.119	-1.322	-1.287	-2.300
R^2	0.279	0.284	0.272	0.275	0.277	0.268	0.255	0.240
Panel D: MFFM + FM Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
MFFM	**2.355**	**2.333**	**2.324**	**2.527**	**2.457**	**1.923**	**1.355**	
FM	-0.630	-0.768	-0.881	-1.027	-1.087	-0.669	-0.271	
R^2	0.263	0.266	0.271	0.278	0.278	0.272	0.253	

Note: This Table summarizes the results for Mincer-Zarnowitz and encompassing regressions using the daily squared portfolio return as unbiased proxy for the latent portfolio variance. The evaluation is based on 342 of the S&P400 constituents to forecast the variance of the equally-weighted portfolio one day ahead using EWMA covariance matrix forecasts with decay parameter $\alpha = 0.94$. The out-of-sample period is Jan. 2007 – Apr. 2009. Coefficients that are statistically significant at the 5% level, based on Newey-West standard errors with 20 lags, are displayed in bold fonts.

Table 4: S&P600 Portfolio volatility, Mincer-Zarnowitz and encompassing regressions

	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RC	**2.878**	**2.080**	**1.656**	**1.590**	**1.406**	**1.306**	**1.125**	**0.916**
R^2	0.229	0.232	0.235	0.237	0.236	0.240	0.225	0.209
Panel B: MFFM Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**1.188**	**1.191**	**1.167**	**1.198**	**1.114**	**1.068**	**0.957**	**0.924**
R^2	0.241	0.241	0.241	0.244	0.243	0.241	0.228	0.216
Panel C: MFFM + RC Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**2.226**	**3.397**	2.275	1.875	1.475	0.686	0.682	**2.254**
RC	-2.601	-3.942	-1.599	-0.918	-0.466	0.475	0.329	-1.345
R^2	0.244	0.247	0.243	0.245	0.243	0.242	0.228	0.220
Panel D: MFFM + FM Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
MFFM	**1.523**	**1.543**	**1.442**	**1.576**	**1.497**	**1.184**	0.783	
FM	-0.284	-0.298	-0.239	-0.322	-0.348	-0.111	0.180	
R^2	0.242	0.242	0.242	0.246	0.245	0.241	0.229	

Note: This Table summarizes the results for Mincer-Zarnowitz and encompassing regressions using the daily squared portfolio return as unbiased proxy for the latent portfolio variance. The evaluation is based on 491 of the S&P600 constituents to forecast the variance of the equally-weighted portfolio one day ahead using EWMA covariance matrix forecasts with decay parameter $\alpha = 0.94$. The out-of-sample period is Jan. 2007 – Apr. 2009. Coefficients that are statistically significant at the 5% level, based on Newey-West standard errors with 20 lags, are displayed in bold fonts.

Table 5: Annualized tracking errors S&P500 (large cap) universe

α	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC tracking error								
0.94	0.060	0.058	0.056	0.060	0.066	0.211	NA	NA
0.75	0.062	0.062	0.061	0.075	0.088	0.178	NA	NA
0.50	0.064	0.068	0.077	0.102	0.122	NA	NA	NA
0.25	0.068	0.077	0.099	0.127	NA	NA	NA	NA
Panel B: MFFM tracking error								
0.94	0.059	0.059	0.058	0.058	0.059	0.059	0.058	0.058
0.75	0.059	0.058	0.058	0.058	0.058	0.059	0.058	0.058
0.50	0.059	0.058	0.058	0.057	0.058	0.058	0.058	0.059
0.25	0.059	0.058	0.058	0.057	0.058	0.058	0.058	0.064
Panel C: RC turnover								
0.94	0.236	0.292	0.357	0.458	0.582	7.922	NA	NA
0.75	0.844	1.041	1.379	1.875	2.435	6.842	NA	NA
0.50	1.728	2.222	3.115	4.229	5.450	NA	NA	NA
0.25	2.861	3.960	5.748	7.668	NA	NA	NA	NA
Panel D: MFFM turnover								
0.94	0.028	0.032	0.037	0.042	0.046	0.051	0.057	0.079
0.75	0.089	0.101	0.120	0.141	0.158	0.182	0.209	0.297
0.50	0.162	0.185	0.225	0.269	0.307	0.361	0.420	0.603
0.25	0.239	0.275	0.339	0.414	0.478	0.573	0.675	0.986

Note: This table reports the *ex-post* annualized minimum tracking errors in percentages and the daily average portfolio turnover using 442 of the S&P500 constituents. The results are based on RC and MFFM to forecast the active covariance matrix one day ahead using EWMA forecasts over the sample period 3/1/2007 - 30/4/2009 with decay parameter α. For the MFFM we use a 12-factor model specification (size, value, and 10 industry factors). The "NA" entries indicate that the RC is not-invertible at certain combinations of sampling frequencies and weighting schemes.

Table 6: Annualized tracking error S&P400 (mid cap) universe

α	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC tracking error								
0.94	0.089	0.085	0.083	0.087	0.091	0.275	NA	NA
0.75	0.095	0.095	0.096	0.114	0.122	0.305	NA	NA
0.50	0.102	0.104	0.118	0.156	0.174	NA	NA	NA
0.25	0.107	0.120	0.154	0.211	NA	NA	NA	NA
Panel B: MFFM tracking error								
0.94	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078
0.75	0.078	0.078	0.078	0.078	0.078	0.077	0.078	0.079
0.50	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.081
0.25	0.078	0.078	0.078	0.078	0.079	0.078	0.079	0.086
Panel C: RC turnover								
0.94	0.251	0.341	0.411	0.484	0.597	3.397	NA	NA
0.75	0.915	1.183	1.520	1.939	2.427	6.485	NA	NA
0.50	1.869	2.429	3.403	4.364	5.357	NA	NA	NA
0.25	3.071	4.197	6.217	8.000	NA	NA	NA	NA
Panel D: MFFM turnover								
0.94	0.023	0.025	0.027	0.031	0.033	0.037	0.040	0.075
0.75	0.083	0.089	0.101	0.115	0.126	0.141	0.160	0.282
0.50	0.160	0.172	0.198	0.229	0.254	0.291	0.335	0.580
0.25	0.245	0.265	0.309	0.363	0.407	0.475	0.560	0.963

Note: This table reports the *ex-post* annualized minimum tracking errors in percentages and the daily average portfolio turnover using 342 of the S&P400 constituents. The tracking errors are based on RC and MFFM to forecast the active covariance matrix one day ahead using EWMA forecasts over the sample period 3/1/2007 - 30/4/2009 with decay parameter α. For the MFFM we use a 12-factor model specification (size, value, and 10 industry factors). The "NA" entries indicate that the RC is not-invertible at certain combinations of sampling frequencies and weighting schemes.

Table 7: Annualized tracking error S&P600 (small-cap) universe

α	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC tracking error								
0.94	0.098	0.091	0.091	0.095	0.105	NA	NA	NA
0.75	0.109	0.114	0.126	0.157	0.200	NA	NA	NA
0.50	0.121	0.135	0.183	0.249	NA	NA	NA	NA
0.25	0.136	0.162	NA	NA	NA	NA	NA	NA
Panel B: MFFM tracking error								
0.94	0.073	0.072	0.072	0.071	0.071	0.071	0.071	0.070
0.75	0.072	0.072	0.071	0.070	0.070	0.071	0.071	0.070
0.50	0.072	0.071	0.071	0.070	0.070	0.070	0.071	0.070
0.25	0.072	0.071	0.071	0.070	0.070	0.071	0.072	0.077
Panel C: RC turnover								
0.94	0.367	0.503	0.658	0.806	1.018	NA	NA	NA
0.75	1.393	1.885	2.649	3.389	4.219	NA	NA	NA
0.50	2.901	4.043	6.035	7.636	NA	NA	NA	NA
0.25	4.842	7.230	NA	NA	NA	NA	NA	NA
Panel D: turnover								
0.94	0.030	0.033	0.039	0.044	0.049	0.056	0.062	0.085
0.75	0.090	0.100	0.121	0.145	0.164	0.192	0.221	0.308
0.50	0.164	0.181	0.224	0.274	0.315	0.374	0.435	0.613
0.25	0.244	0.270	0.338	0.421	0.490	0.589	0.690	1.000

Note: This table reports the *ex-post* annualized minimum tracking errors in percentages and the daily average portfolio turnover using 491 of the S&P600 constituents. The tracking errors are based on RC and MFFM to forecast the active covariance matrix one day ahead using EWMA forecasts over the sample period 3/1/2007 - 30/4/2009 with decay parameter α. For the MFFM we use a 12-factor model specification (size, value, and 10 industry factors). The "NA" entries indicate that the RC is not-invertible at certain combinations of sampling frequencies and weighting schemes.

C Figures

Figure 1: Comparison of MFFM to Hayashi-Yoshida, RC and RC-LL in terms of ln MSE without microstructure noise

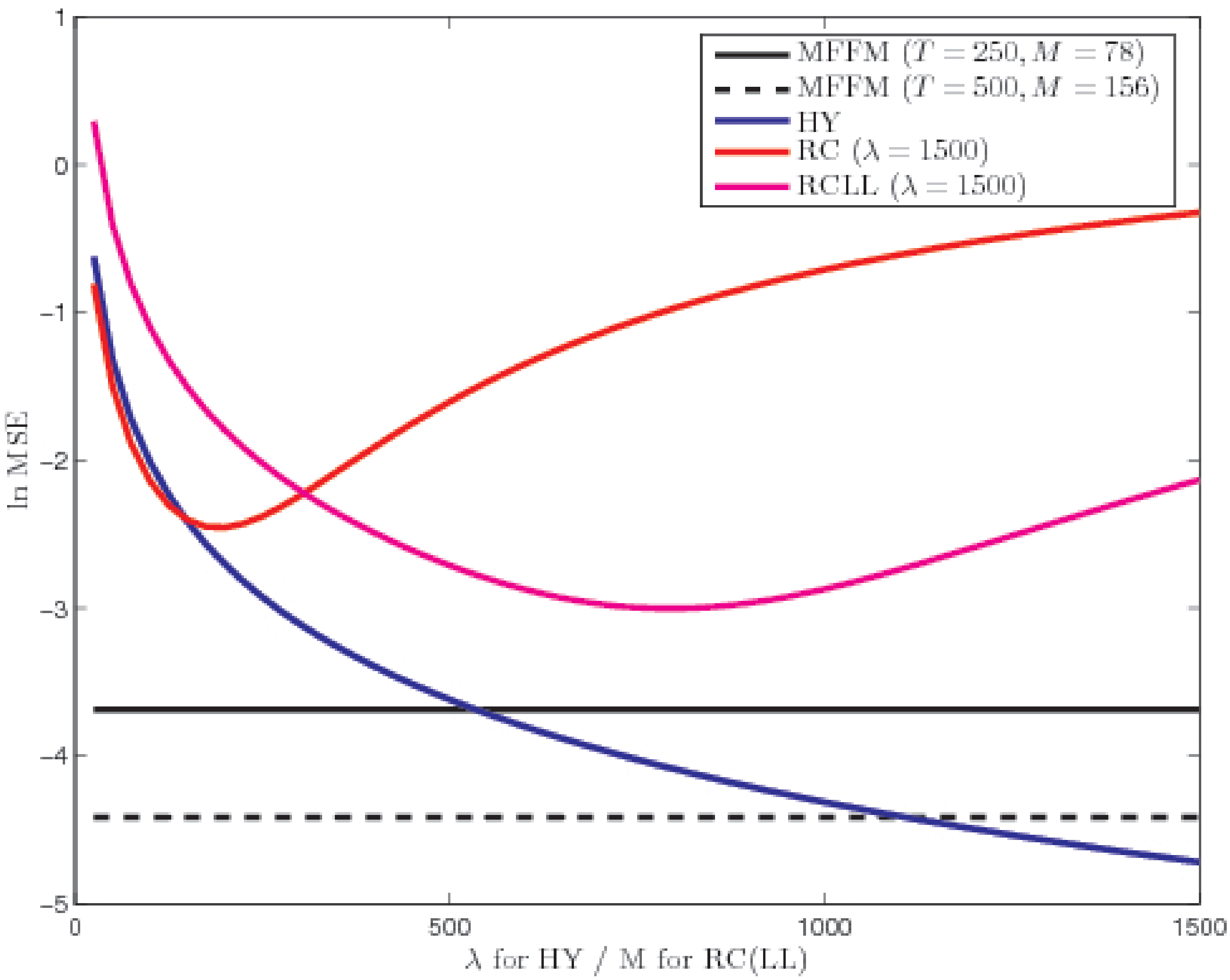
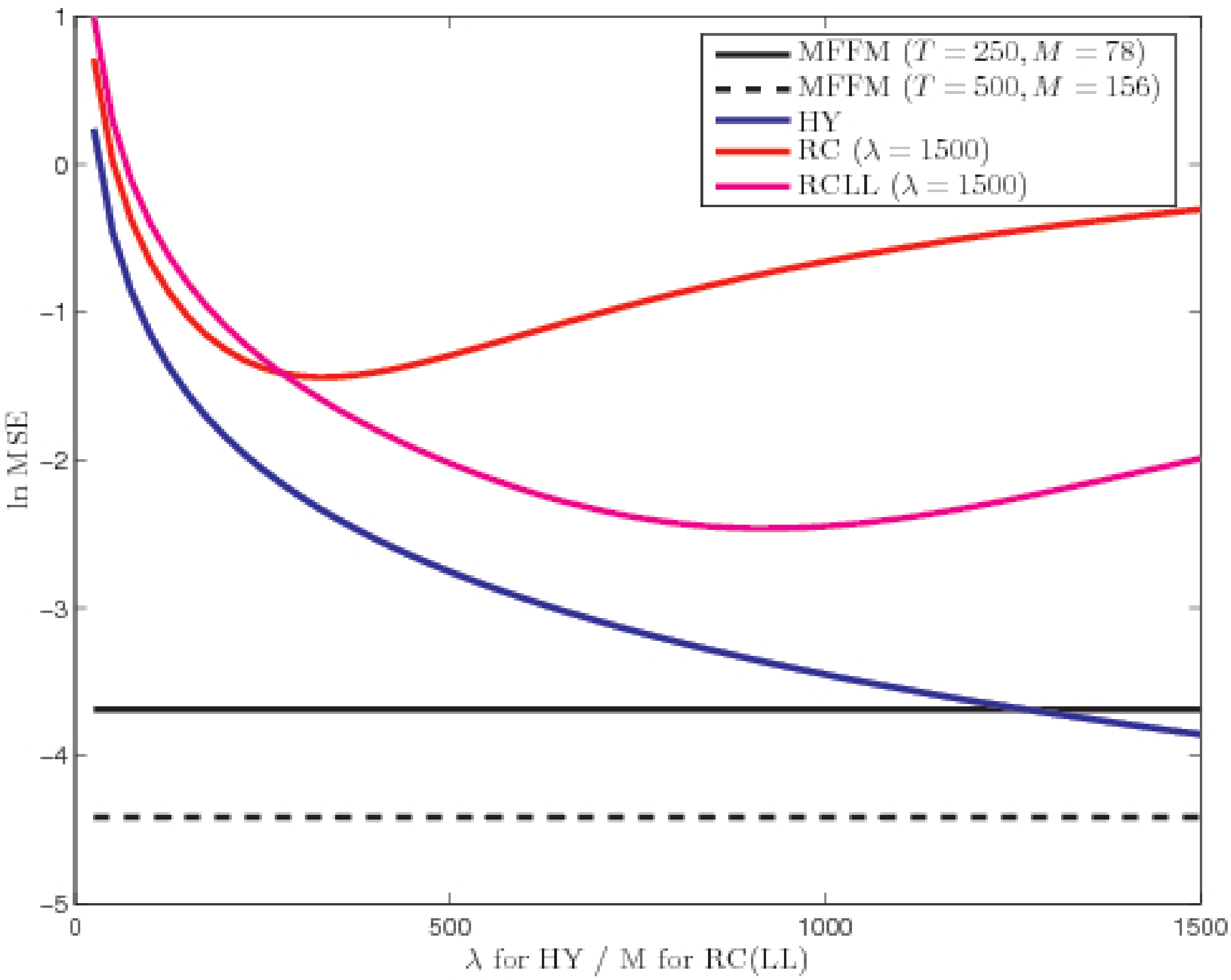


Figure 2: Comparison of MFFM to Hayashi-Yoshida, RC and RC-LL in terms of ln MSE with microstructure noise



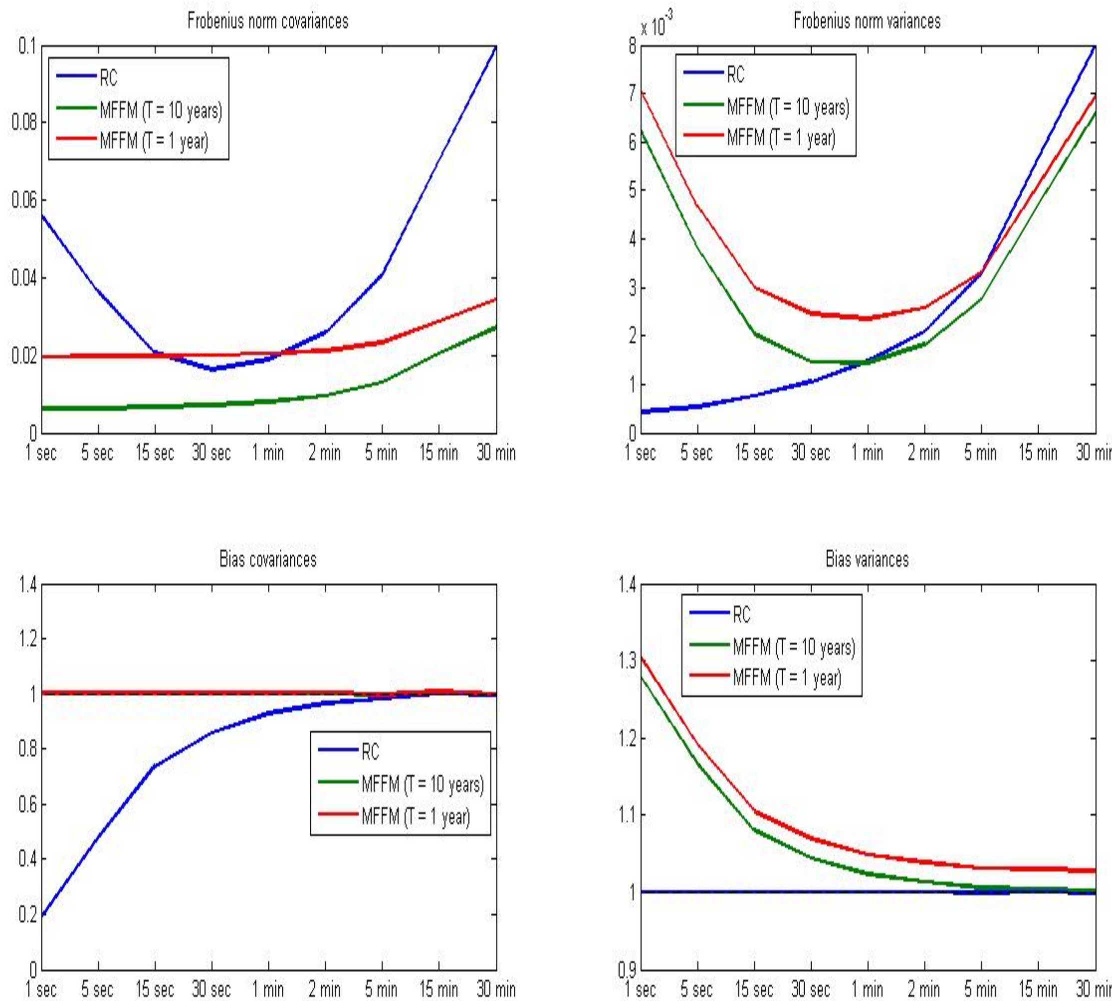


Figure 3: This Figure displays the Frobenius Norm (Equation (9)) of the (co)variance estimation errors for the MFFM (Equation (8)) and RC based on 10,000 repetitions in a setting with high liquidity for the individual stocks, the prices are non-synchronous and not contaminated by microstructure noise.

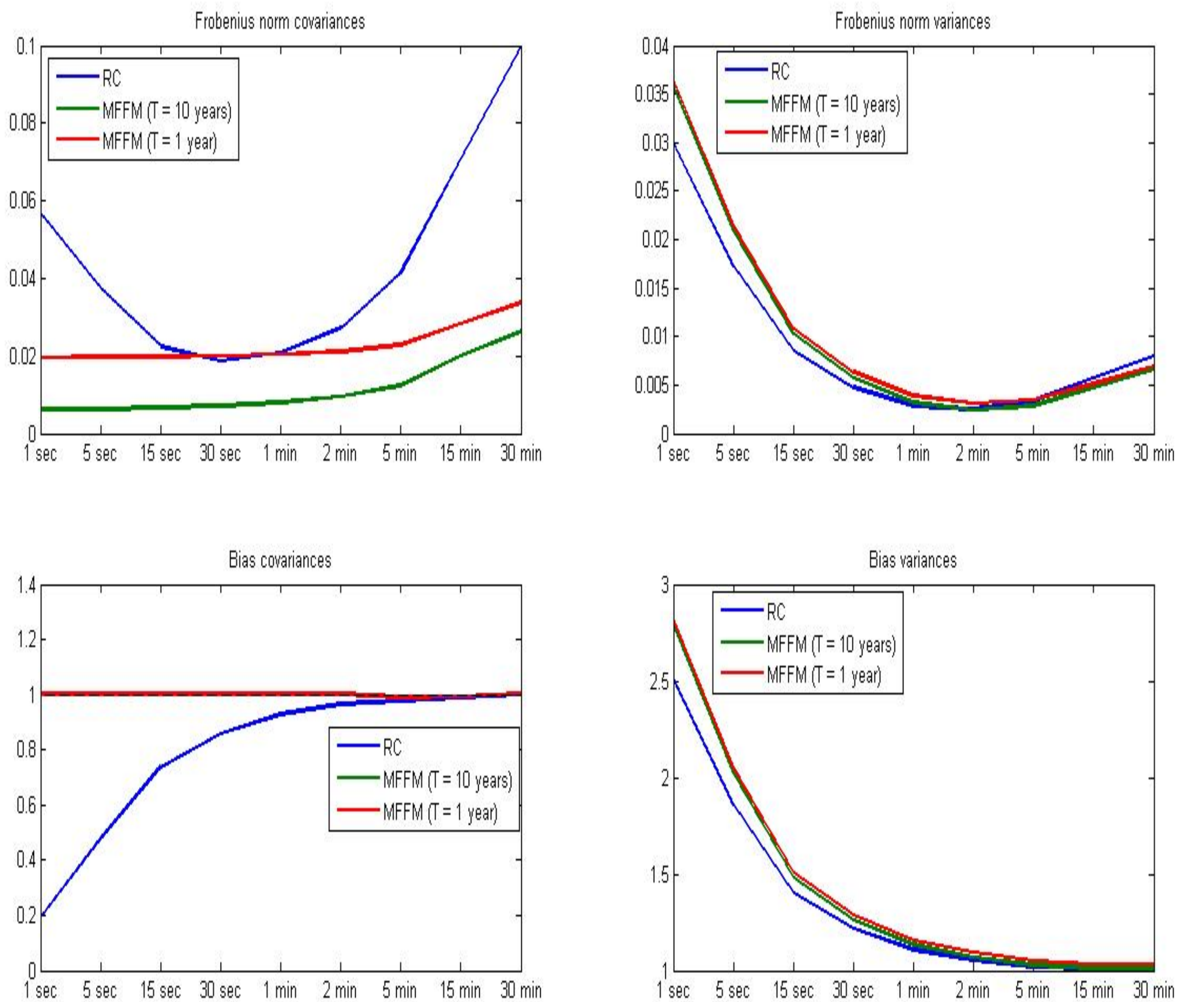


Figure 4: This Figure displays the Frobenius Norm (Equation (9)) of the (co)variance estimation errors for the MFFM (Equation (8)) and RC based on 10,000 repetitions in a setting with high liquidity for the individual stocks, the prices are non-synchronous and contaminated by microstructure noise.

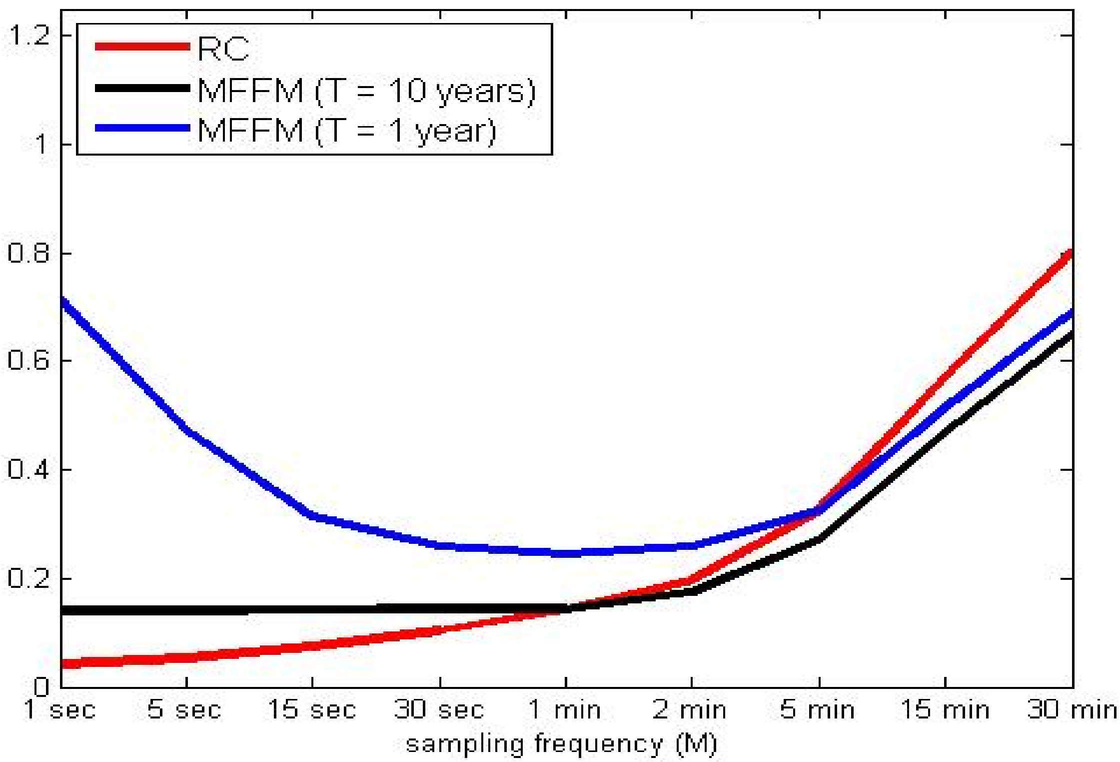
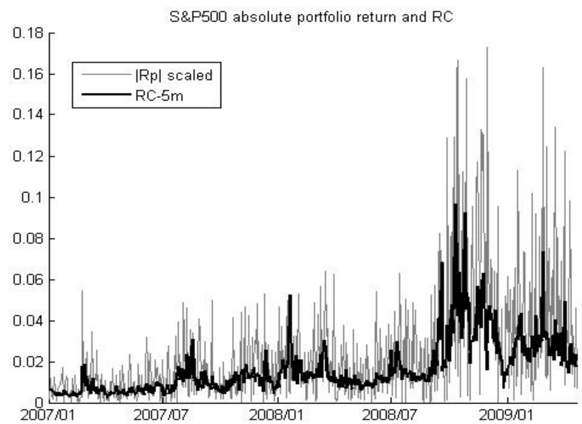
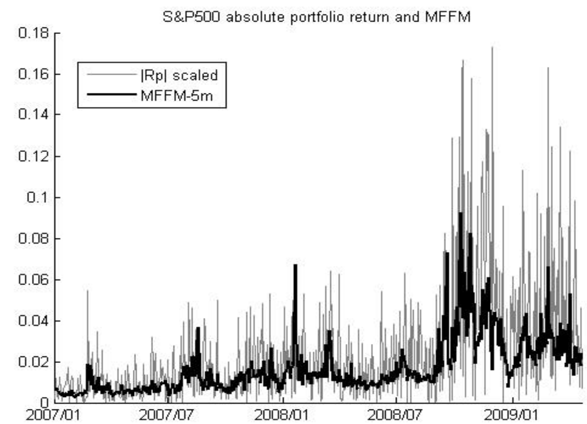


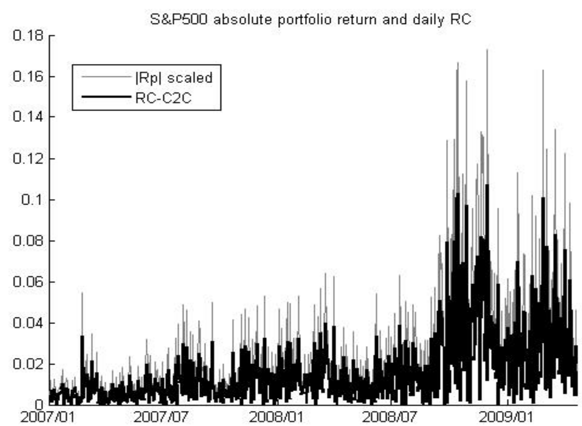
Figure 5: This Figure displays the Frobenius norm for the variance elements of the MFFM when a bias adjustment is used by introducing a lower 3rd sampling frequency for calculating idiosyncratic risk. The MFFM with small measurement errors in the betas (T = 10 years) is bias-adjusted while the case with larger measurement errors (T = 1 year) is not bias-adjusted. The sampling frequency used for residual risk is the 1m frequency if we use factor covariances sampled at higher frequencies. When the sampling frequency for the factor covariances is 1m or lower, then we use the same sampling frequency for residual risk.



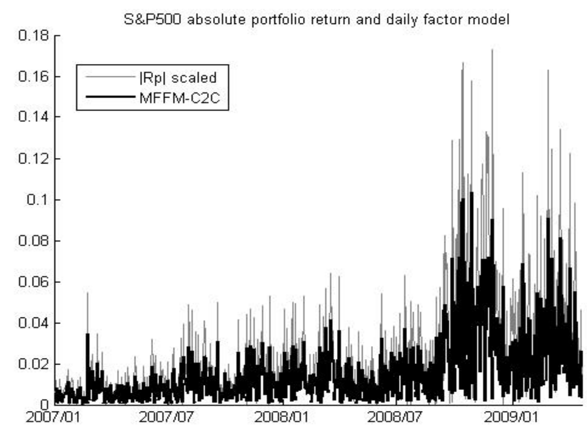
(a) RC-5m



(b) MFFM-5m

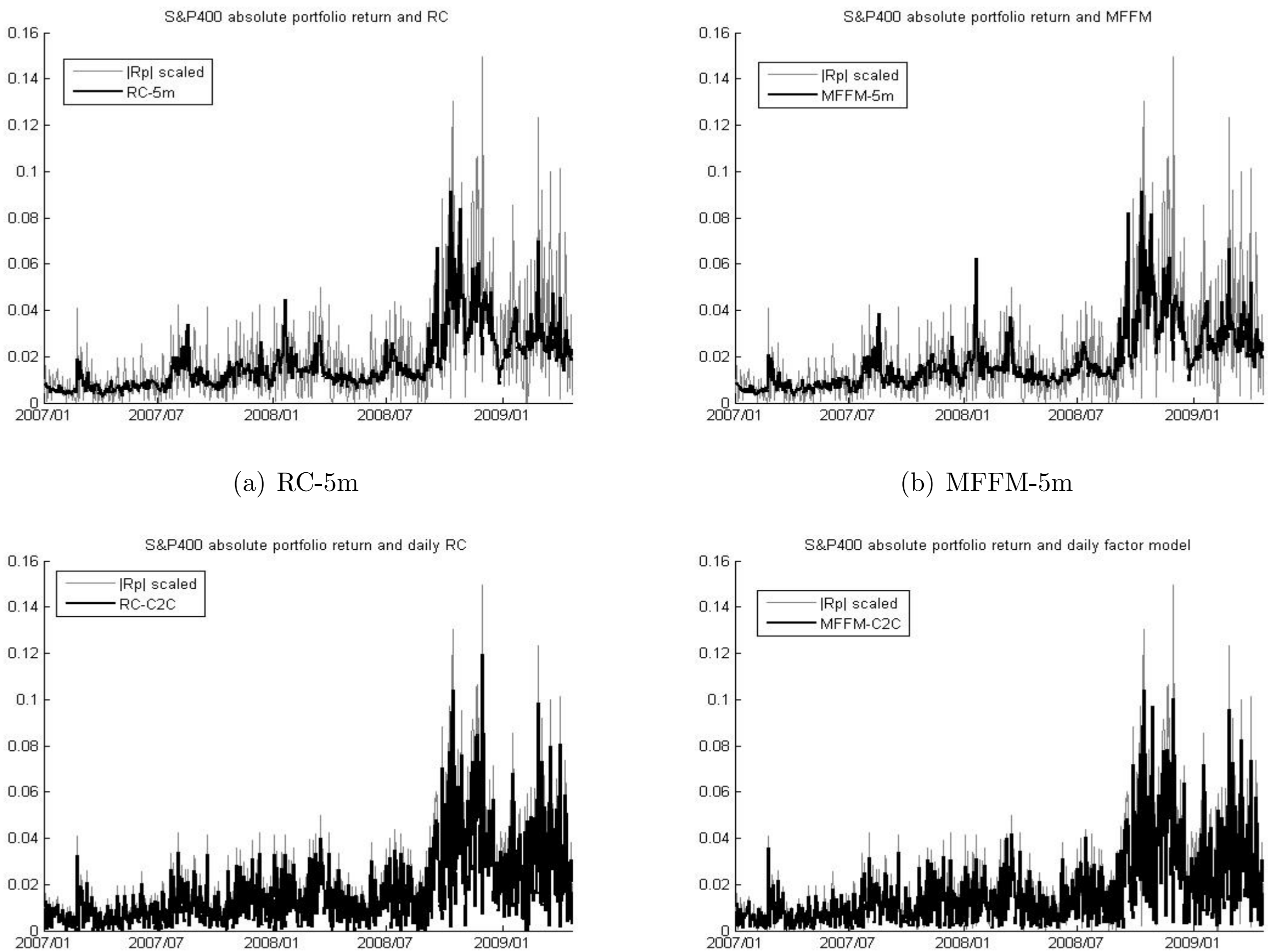


(c) RC-day



(d) FM-day

Figure 6: This Figure displays plots of the (scaled) absolute portfolio return of the equally-weighted S&P500 portfolio against the non-smoothed ($\alpha = 0$) volatility estimates obtained using the (a) realized covariance 5min (b) MFFM-5min (c) daily sample covariance and (d) daily factor model.



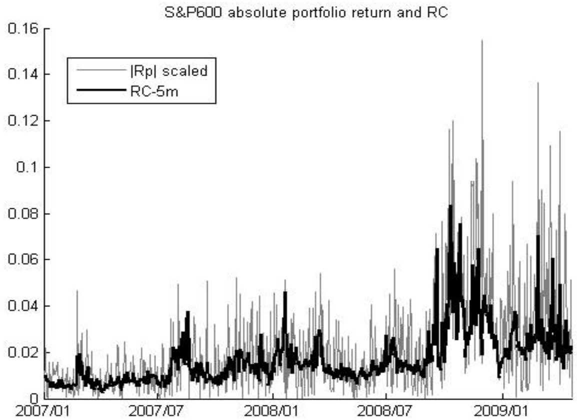
(a) RC-5m

(b) MFFM-5m

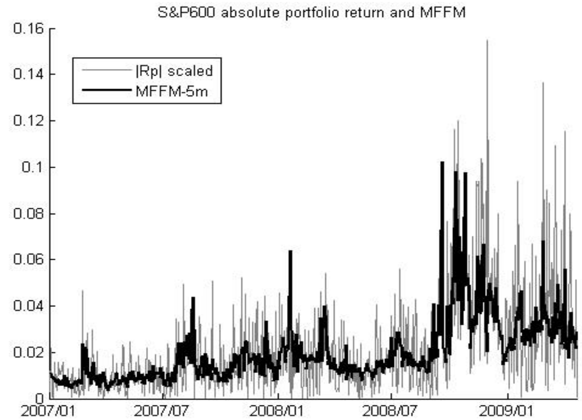
(c) RC-day

(d) FM-day

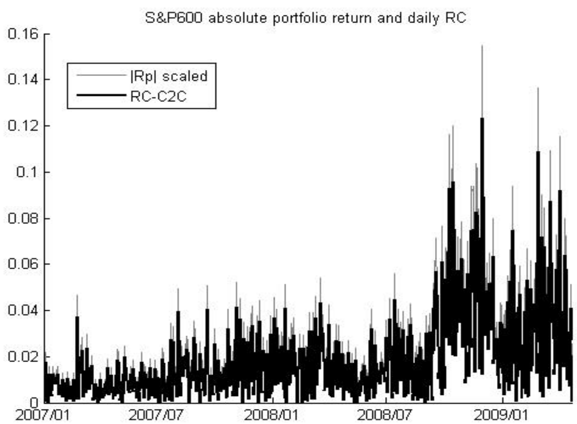
Figure 7: This Figure displays plots of the (scaled) absolute portfolio return of the equally-weighted S&P400 portfolio against the non-smoothed ($\alpha = 0$) volatility estimates obtained using the (a) realized covariance 5min (b) MFFM-5min (c) daily sample covariance and (d) daily factor model.



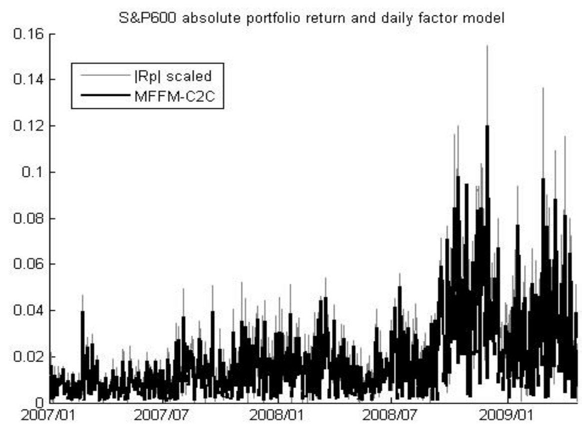
(a) RC-5m



(b) MFFM-5m



(c) RC-day



(d) FM-day

Figure 8: This Figure displays plots of the (scaled) absolute portfolio return of the equally-weighted S&P600 portfolio against the non-smoothed ($\alpha = 0$) volatility estimates obtained using the (a) realized covariance 5min (b) MFFM-5min (c) daily sample covariance and (d) daily factor model.