Baruník, Jozef; Čech, František

Working Paper

On the modelling and forecasting multivariate realized volatility: Generalized Heterogeneous Autoregressive (GHAR) model

IES Working Paper, No. 23/2014

Provided in Cooperation with:
Charles University, Institute of Economic Studies (IES)

Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

On the Modelling and Forecasting Multivariate Realized Volatility: General Heterogeneous Autoregressive (GHAR) Model

Jozef Baruník
František Čech

IES Working Paper: 23/2014

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Baruník J., Čech F. (2014). "On the modelling and forecasting multivariate realized volatility: Generalized Heterogeneous Autoregressive (GHAR) model" IES Working Paper 23/2014. IES FSV. Charles University.

On the Modelling and Forecasting Multivariate Realized Volatility: Generalized Heterogeneous Autoregressive (GHAR) Model

Jozef Baruník[a]
František Čech[b]

[a]Institute of Economic Studies, Charles University, Opletalova 26, 110 00, Prague, CR and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 00, Prague, Czech Republic. Phone: +420 776 259 273. Email: barunik@utia.cas.cz
[b]Institute of Economic Studies, Charles University, Opletalova 26, 110 00, Prague, CR and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 00, Prague, Czech Republic. Phone: +420 739 538 376 Email: fero.cech@gmail.com

August 2014

Abstract:

We introduce a methodology for dynamic modelling and forecasting of realized covariance matrices based on generalization of the heterogeneous autoregressive model (HAR) for realized volatility. Multivariate extensions of popular HAR framework leave substantial information unmodeled in residuals. We propose to employ a system of seemingly unrelated regressions to capture the information. The newly proposed generalized heterogeneous autoregressive (GHAR) model is tested against natural competing models. In order to show the economic and statistical gains of the GHAR model, portfolio of various sizes is used. We find that our modeling strategy outperforms competing approaches in terms of statistical precision, and provides economic gains in terms of mean-variance trade-off. Additionally, our results provide a comprehensive comparison of the performance when realized covariance and more efficient, noise-robust multivariate realized kernel estimator, is used. We study the contribution of both estimators across different sampling frequencies, and we show that the multivariate realized kernel estimator delivers further gains compared to realized covariance estimated on higher frequencies.

Keywords: GHAR, portfolio optimization, economic evaluation
JEL: C18, C58, G15

Acknowledgements:

The support from the Czech Science Foundation under the 13-32263S project and the support from the Grant Agency of Charles University under the 1198214 project is gratefully acknowledged.

1 Introduction

The risk of individual financial instruments is crucial for asset pricing, portfolio and risk management. Besides volatility of individual assets capturing the risk, knowledge of covariance structure between assets in portfolio is of great importance. Accurate forecasts of variance-covariance matrices are particularly important in asset allocation and portfolio management.

Nature of the financial data with dependencies in higher moments of the daily return series motivated the work of Engle (1982) and Bollerslev (1986), who developed widely used family of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. Subsequent research developed numerous multivariate extensions of the framework, most importantly constant conditional correlation GARCH of Bollerslev (1990) further generalized by Engle (2002) to capture dynamics in correlation structure, and BEKK model of Engle and Kroner (1995). Multivariate GARCH (MGARCH) models are popular in the literature although they suffer from curse of dimensionality problem. Detailed information about MGARCH specifications can be found in Bauwens et al. (2006) for example.

Increased availability of high-frequency data in the last decade resulted in development of the new non-parametric approach of treating volatility, which is an interesting alternative to traditional MGARCH models. Model-free estimator called "realized volatility" that makes volatility observable has been proposed by Andersen et al. (2001), and later rigorously studied by Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004). Barndorff-Nielsen and Shephard (2004) further completed the theory to multivariate "realized covariation". Estimates of variance-covariance matrix that are obtained by realized covariation do not have to be necessarily positive semi-definite due to market microstructure noise. Therefore Barndorff-Nielsen et al. (2011) introduced multivariate realized kernels estimator guaranteeing the positive semi-definiteness of the variance-covariance matrix.

Once the covariance matrix is estimated from the high-frequency data, it needs to be further modelled. The research dedicated to modeling the entire covariance matrices is still lively. From the already established methods, let us mention Wishart Autoregression (WAR) of Gouriéroux et al. (2009) with numerous extensions presented in Bonato (2009) and Bonato et al. (2013). Different approach of realized volatility modelling can be found in Bauer and Vorkink (2011) model realized stock market volatility using matrix-logarithm transformation and primarily concentrate on forecasting performance of factor model. More common approach of obtaining positive definite forecasts of covariance matrices is use of Cholesky decomposition. The use of Cholesky factors further estimated by Vector Autoregressive Fractionally Integrated Moving Average (VARFIMA), Heterogeneous Autoregression (HAR) or WAR-HAR can be found in recent work of Chiriac and Voev (2011).

In this paper, we contribute to this literature by proposing a new model for dynamic covariance matrix modelling and forecasting. We model Cholesky factors of realized covariance matrix as a system of seemingly unrelated heterogeneous autoregressions. The main motivation is that we may expect the residuals from simple HAR model to be contemporaneously correlated, and moreover heteroscedastic due to well known volatility in volatility effect (Corsi et al., 2008). Estimating the system of HAR equations using generalized least squares allows to capture these dependencies. Hence the generalizated HAR (GHAR) may provide more precise and more efficient forecasts, which will translate to economic gains directly. On the portfolios of various sizes, we show that GHAR model delivers significant economic gains and statistically is not substantially outperformed, when compared to natural benchmark models based on high frequency data (HAR, VARFIMA), as well as daily data (DCC-GARCH, RiskMetrics). In addition, we study the economic benefits of estimating the realized covariance with more efficient sub-sampled realized covariance and multivariate realized kernel estimators.

The rest of the paper is structured as follows. We provide background for estimation of realized covariation from high frequency data in section 2. Section 3 describes frameworks for modeling multivariate volatility, and it presents our GHAR model. Section 4 provides description of dataset, and research design including economic as well as statistical evaluation criteria. In Section 5, we discuss out-of-sample forecast evaluation, and Section 6 concludes.

2 Estimation of covariation form high frequency data

We assume that the q-dimensional efficient price process p_t evolves over time $0 \leq t \leq T$ according to the following dynamics

$$dp_t = \mu_t dt + \Sigma_t dW_t + dJ_t, \tag{1}$$

where μ_t is predictable component, Σ_t is real-values $q \times q$ volatility process, $W_1, \ldots, W_q$ is an q-dimensional Brownian motion, and dJ_t is a jump process. A central object of interest is the integrated covariation, which measures the covariance of asset returns over a particular period. Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) suggest to estimate the quadratic covariation matrix analogously to the realized variance, by taking outer product of the observed high-frequency return over the period. This estimation although assumes synchronised equidistant data.

In practice, trading is non-synchronous, delivering fresh prices at irregularly spaced times which differ across stocks. In order to estimate the covariance, the data need to be synchronized, meaning that the prices of the q assets need to be collected at the same time stamp. Research of non-synchronous trading has been an active field of financial econometrics in past years – see, for example, Hayashi and Yoshida (2005) and Voev and Lunde (2007). This practical issue induces bias in the estimators and may be partially responsible for the Epps effect (Epps, 1979), a phenomenon of decreasing empirical correlation between the returns of two different stocks with increasing data sampling frequency. Aït-Sahalia et al. (2010) compare various synchronization schemes and find that the estimates do not differ significantly from the estimates using the so called refresh time scheme when dealing with highly liquid assets. The data used further in our study consists of the most liquid U.S. stocks, hence we can restrict ourselves to the refresh time synchronization scheme in our work.

Let $N_{(q)t}$ be the counting process governing the number of observations in the q-th asset up to time t, with times of trades $t_{(q)1}, t_{(q)2}, \ldots$. Following Barndorff-Nielsen et al. (2011), we define the first refresh time as

$$\tau_1 = \max(t_{(1)1}, \ldots, t_{(d)1}), \tag{2}$$

for $d = 1, \ldots, q$ assets, and all subsequent refresh times as

$$\tau_{j+1} = \max(t_{(1)N_{(1)\tau_j}+1}, \ldots, t_{(d)N_{(d)\tau_j}+1}), \tag{3}$$

with the resulting refresh time sample being of length N. τ_1 is thus the first time that all assets record prices, while τ_2 is the first time that all asset prices are refreshed. In the following analysis, we will always set our clock time to τ_j when using the estimators.

Having synchronised the data, let us denote $\Delta_k p_t = p_{t-1+\tau_k/N} - p_{t-1+\tau_{k-1}/N}$ a discretely sampled vector of k-th intraday log-returns in $[t-1, t]$, with N intraday observations available for each asset q. A simple estimator of realized covariance is then constructed as

$$\widehat{\Sigma}_t^{(RC)} = \sum_{k=1}^{N} (\Delta_k p_t)(\Delta_k p_t)'. \tag{4}$$

As shown by Barndorff-Nielsen and Shephard (2004), realized covariance is consisted estimator of integrated covariance and is asymptotically mixed normal. However, the estimator is biased and become inconsistent in case microstructure noise is present in data. Sparse sampling is used to mitigate the trade-off between the bias due to noise and variance of the estimator.

To effectively use all available high-frequency data Zhang et al. (2005) propose to use sub-sampling and averaging for realized variance calculation. In their set-up whole sample is divided into M non-overlapping sub-samples, in each sub-sample realized variance is calculated and average across the sub-sampled estimates form the final estimate.

$$\widehat{\Sigma}_t^{(RCSS)} = \frac{1}{M}\sum_{i=1}^{M} \widehat{\Sigma}_{t,i}^{(RC)} \tag{5}$$

In addition, covariance matrix estimated by realized covariance might not necessary be positive semi-definite. To overcome these problems, Barndorff-Nielsen et al. (2011) introduced multivariate realized kernels (MRK) estimator that guaranties covariance matrix to be positive semi-definite. Moreover, MRK is more efficient, and it is able to deal with noise. Following Barndorff-Nielsen et al. (2011) the MRK estimator is defined as

$$\widehat{\Sigma}_t^{(MRK)} = \sum_{h=-n}^{n} k\left(\frac{h}{H}\right)\Gamma_h \tag{6}$$

where Γ_h stands for h-th realized autocovariance and $k(x)$ is a non-stochastic weight function. In the empirical implementation, we need to choose the kernel function and bandwidth parameter. Following Barndorff-Nielsen et al. (2011), we use a Parzen kernel,[1] which satisfies the smoothness conditions, $K'(0) = K'(1) = 0$, and guarantees $\widehat{\Sigma}_t^{(MRK)}$ to be positive semi-definite. We use optimal bandwidth derived in Barndorff-Nielsen et al. (2011).

3 Modeling and forecasting multivariate volatility

Modeling and forecasting conditional covariance matrix of asset returns Σ_t is pivotal to asset allocation, risk management, and option pricing. In order to have a valid multivariate forecasting model, one needs to specify a model that produces symmetric and positive semi-definite covariance matrix predictions. Whereas it is still relatively scarce to use high frequency data in multivariate modeling, literature dealing with challenging issues is growing quickly. There are three types of approaches proposed recently: modeling the Cholesky factorisation of covariance matrix (Chiriac and Voev, 2011), its matrix-log transformation with the use of latent factors (Bauer and Vorkink, 2011), and direct modeling of the covariance dynamics as a Wishart autoregressive model (Bonato, 2009; Jin and Maheu, 2013).

To ensure positive semi-definiteness of covariance matrix forecasts, we adopt approach from Chiriac and Voev (2011) - we apply the Cholesky decomposition on covariance matrix. This approach is attractive, as it also helps to reduce the curse of dimensionality, especially in the model structures we are going to use in this study. Following Chiriac and Voev (2011), we model the lower triangular elements of the Cholesky factorization,

$$X_t = vech\left(P_t\right), \tag{7}$$

[1] The Parzen kernel function is given by $k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \le x \le 1/2 \\ 2(1-x)^3 & 1/2 \le x \le 1. \\ 0 & x > 1 \end{cases}$

where P_t are Cholesky factors $P_t'P_t = \Sigma_t$, and X_t is $m \times 1$ vector, with $m = \frac{q(q+1)}{2}$. Forecasts of the covariance matrix are then obtained by reverse tranformation.

3.1 Generalized heterogeneous autoregressive (GHAR) model

A simple approximate long-memory model for realized volatility, heterogeneous autoregression (HAR), has been introduced by Corsi (2009). Whereas the approach has been introduced for the univariate volatility modeling, its extension to multivariate volatility has been recently used in the literature (see e.g. Chiriac and Voev (2011) or Bauer and Vorkink (2011)). Original HAR model has an autoregressive structure, and combines volatilities measured at different frequencies (daily, weekly, monthly). Chiriac and Voev (2011) propose a multivariate extension of the HAR to model vector of Cholesky factors X_t, as

$$X_{t+1}^{(1)} = c + \beta^{(1)} X_t^{(1)} + \beta^{(5)} X_t^{(5)} + \beta^{(22)} X_t^{(22)} + \epsilon_t, \quad \epsilon_t \sim i.i.d. \tag{8}$$

where 1,5, and 22 stands for day, week (5 days) and month (22 days) respectively, c is $m \times 1$ vector of constants, $\beta^{(\cdot)}$ are scalar parameters, and $X_t^{(\cdot)}$ are averages of lagged daily volatility e.g. $X_t^{(5)} = \frac{1}{5} \sum_{i=0}^{4} X_{t-i}$. To obtain parameter estimates, ordinary least squares are used.

One of the disadvantages of this modeling strategy is that we are assuming the same structure for all elements of the Cholesky factors in X_t. Much more importantly, we are leaving significant amount of information in error term. One can expect error term to be heteroscedastic due to volatility of volatility (Corsi et al., 2008) present in the realized measures. More importantly, a common structure of X_t elements may be left unmodeled in residuals. Hence, it may be more natural to estimate the model in Eq. 15 as system of equations with some covariance structure of the error terms.

To deal with this problem, we propose to build a system of seemingly unrelated HAR regressions (Zellner, 1962) for all elements of X_t. The advantage of this approach is that we estimate a multivariate HAR model, which will capture the separate dynamics of the variances and covariances, but also possible common structure. Moreover, it will also yield more efficient estimates. As we know, error terms from HAR are heteroscedastic (Corsi et al., 2008), which makes the coefficient estimates less efficient. Moreover, in case there is no information about dependence between equations left in the residuals from regression Eq. 15, estimator will converge to a simple OLS estimates, as diagonal weighting matrix in generalized regression will reduce the estimates to OLS. On the other hand, the possible disadvantage is in larger number of parameters to be estimated, which may yield the model unreliable with highly dimensional portfolios.

Let us consider the system of $i = 1, ..., m$ equations, where $m = \frac{q(q+1)}{2}$

$$X_{i,t+1}^{(1)} = \beta_i^{(c)} + \beta_i^{(1)} X_{i,t}^{(1)} + \beta_i^{(5)} X_{i,t}^{(5)} + \beta_i^{(22)} X_{i,t}^{(22)} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim i.i.d. \tag{9}$$

There are m equations representing elements of the Cholesky factors, with T observations. Define the $mT \times 1$ vector of disturbances $\epsilon = (\epsilon_1', \dots, \epsilon_m')'$, and rewrite the model as

$$\begin{pmatrix} X_{1,t+1}^{(1)} \\ \vdots \\ X_{m,t+1}^{(1)} \end{pmatrix} = \begin{pmatrix} X_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{m,t} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{m,t} \end{pmatrix} \tag{10}$$

where $X_{i,t} = \begin{pmatrix} e & X^{(1)}_{i,t} & X^{(5)}_{i,t} & X^{(22)}_{i,t} \end{pmatrix}$ is i-th element of X_t and e being vector of ones, $\beta_i = \begin{pmatrix} \beta^{(c)}_i & \beta^{(1)}_i & \beta^{(5)}_i & \beta^{(22)}_i \end{pmatrix}'$ and $\beta^{(c)}_i$ being estimates of intercept. It is more convenient to work with this system in the following form

$$y = Z\beta + \epsilon, \tag{11}$$

where $y = \left(X^{(1)}_{1,t+1}, \dots, X^{(1)}_{m,t+1} \right)'$ and ϵ are $mT \times 1$ vectors, $Z = diag\{X_{1,t}, \dots, X_{m,t}\}$ is block diagonal matrix of dimension $(m \times 4) \times T$, and $\beta = (\beta_1, \dots, \beta_m)'$ is $(m \times 4) \times 1$ vector of parameters.

The disturbances will satisfy strict exogeneity $E[\epsilon|Z] = 0$, but will be correlated across equations, $E[\epsilon'_i \epsilon_j|Z] = \sigma_{ij} I_T$ or

$$\Omega = \begin{pmatrix} \sigma_{11} I_T & \cdots & \sigma_{1m} I_T \\ \vdots & \ddots & \vdots \\ \sigma_{m1} I_T & \cdots & \sigma_{mm} I_T \end{pmatrix} = \Sigma \otimes I_T, \tag{12}$$

where $\Sigma = \sigma_{ij}$ for $i, j = 1, ..., m$, $\otimes$ is Kronecker product and I_T is identity matrix of dimension $T \times T$. The model parameters are estimated in two step feasible generalized least squares. We run ordinary least squares regression in the first step to obtain estimates $\hat{\sigma}_{ij}$ from residuals. In the second step, we run generalized least squares regression using variance matrix $\widehat{\Omega} = \widehat{\Sigma} \otimes I_T$ as

$$\widehat{\beta} = \left(Z' \widehat{\Omega}^{-1} Z \right)^{-1} Z' \widehat{\Omega}^{-1} y \tag{13}$$

The estimator $\widehat{\beta}$ is unbiased, and consistent estimator of β with asymptotically normal limiting distribution

$$\sqrt{T} \left(\widehat{\beta} - \beta \right) \xrightarrow{d} \mathcal{N} \left(0, \left(\frac{1}{T} Z' \widehat{\Omega}^{-1} Z \right)^{-1} \right) \tag{14}$$

While this is a standard estimation technique, we will refrain from discussing any further details about the properties of the generalized least squares estimator.

3.2 Competing models

To show the contribution of the GHAR model, we compare the forecasts to several competing alternatives. The first natural choice of benchmark model is multivariate extension of original HAR. By comparing these two models, we will see the portion of contribution brought by allowing for correlated residuals in the estimation. Another natural candidate is vector ARFIMA as Chiriac and Voev (2011) find it to outperform the HAR model slightly, but conclude that HAR performs reasonably well in comparison to VARFIMA. Hence we may have reason to believe that our approach will provide better results than VARFIMA model.

These three main models share the same framework of modeling elements of Cholesky factors from realized covariance matrix. Hence, we also contrast them to two benchmark models, namely popular DCC GARCH of Engle (2002) and risk metrics standard widely used in the business industry. These benchmark models operate on the daily data, so we will have a direct comparison of gains from high frequency data.

3.2.1 HAR

A first, natural competing model to our generalized HAR strategy is multivariate extension of an original HAR, which models vector of Cholesky factors X_t, as

$$X_{t+1}^{(1)} = c + \beta^{(1)} X_t^{(1)} + \beta^{(5)} X_t^{(5)} + \beta^{(22)} X_t^{(22)} + \epsilon_t, \quad \epsilon_t \sim i.i.d. \tag{15}$$

where 1,5, and 22 stands for day, week (5 days) and month (22 days) respectively, c is $m \times 1$ vector of constants, $\beta^{()}$ is $m \times 1$ vector of parameters and $X_t^{()}$ are averages of lagged daily volatility e.g. $X_t^{(5)} = \frac{1}{5} \sum_{i=0}^{4} X_{t-i}$. To obtain parameter estimates, ordinary least squares are used.

3.2.2 Vector ARFIMA model

Second competing model to HAR family is vector autoregressive fractionally integrated moving average (VARFIMA) model of Chiriac and Voev (2011), who use restricted VARFIMA (1,d,1) specification to model and forecast dynamics of X_t directly. Authors find that ARFIMA provides slightly better forecast in comparison to HAR model, which makes it natural candidate to our modeling strategy. We consider the vector ARFIMA model

$$(1 - \phi L)\, D(L)\, [X_t - c] = (1 - \theta L)\, \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma) \tag{16}$$

where ϕ and θ are scalars, c is $m \times 1$ vector of constants and $D(L) = (1 - L)^d I_m$ with common parameter of fractional integration d for all constituents of X_t. In our case we reject the hypothesis about equality of d thus we estimated each element of X_t using unique $d_t : D(L) = diag\left\{(1 - L)^{d_1}, \cdots, (1 - L)^{d_m}\right\}$. Hence, we use the model 1 in Chiriac and Voev (2011).

3.2.3 RiskMetrics

RiskMetrics of J.P. Morgan Chase, based on exponentially weighted moving average (EWMA), is a financial industry standard and common benchmark for any volatility model (univariate or multivariate). In our work we use specification from Longerstaey and Spencer (1996) with decay factor, λ set to 0.94. We assume a $q \times 1$ vector of daily returns $r_t = \sum_{k=1}^{n} (\Delta_k p_t)$ for $t = 1, ..., T$ such that $r_t \sim N\left(\mu_t, \sigma_t^2\right)$, where μ_t is conditional mean and σ_t^2 is conditional variance of daily returns. Moreover if we assume $\mu_t = 0$, conditional covariance has the form

$$\sigma_{i,j} = (1 - \lambda) \sum_{t=1}^{T} \lambda^{t-1} r_i r_j. \tag{17}$$

Previous equation can be rewritten into recursive form

$$\sigma_{i,j,t} = \lambda \sigma_{i,j,t-1} + (1 - \lambda) r_{i,t-1} r_{j,t-1} \tag{18}$$

where expression $\sigma_{i,j,t}$ stands for covariance between assets i and j in time t.

3.2.4 DCC-GARCH

Dynamic conditional correlation generalized autoregressive conditional heteroscedasticity (DCC-GARCH) of Engle (2002) is widely used multivariate GARCH model in practice. It is a generalization of Bollerslev (1990)'s constant conditional correlation GARCH, with time-varying correlation matrix R. The model is defined as

$$H_t = D_t R_t D_t, \tag{19}$$

where D_t is diagonal matrix of conditional time varying standard deviations, $D_t = diag\left(\sqrt{h_{i,t}}\right)$, and $h_{i,t}$ are univariate GARCH processes, $h_{i,t} = \omega_i + \sum_{p=1}^{P_i} \alpha_{i,p} r_{i,t-p}^2 + \sum_{q=1}^{Q_i} \beta_{i,q} h_{i,t-q}$. Dynamics of correlation matrix is given by transformation

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}, \tag{20}$$

where $Q_t = \left(1 - \sum_{m=1}^{M} \alpha_m - \sum_{n=1}^{N} \beta_n\right) \bar{Q} + \sum_{m=1}^{M} A_m \left(\epsilon_{t-m} \epsilon_{t-m}^T\right) + \sum_{n=1}^{N} B_n Q_{t-n}$,$\bar{Q}$ is the unconditional covariance matrix of the standardized residuals from the univariate GARCH processes and $Q_t^* = diag\left(\sqrt{q_{ii,t}}\right)$. In our work we use two stage estimator presented in Engle (2002) or Engle and Sheppard (2001).

4 Data and research design

The dataset consists of tick prices of 15 S&P 500 index constituents with highest liquidity and market capitalization. Final portfolio thus consists[2] of Apple Inc. (AAPL), Exxon Mobile Corp. (XOM), Google Inc. (GOOG), Wal-Mart Stores (WMT), Microsoft Corp. (MSFT), General Electric Co (GE), International Business Machines Corp. (IBM), Johnson & Johnson (JNJ), Chevron Corp. (CVX), Procter & Gamble (PG), Pfizer Inc. (PFE), AT&T Inc. (T), Wells Fargo & Co (WFC), JP Morgan Chase & Co (JPM) and Coca-Cola Co. (KO). We obtain 390, 78, 39, 26 and 19 time-synchronized intraday observations using refresh-time, resulting in 1, 5, 10, 15 and 20 minute intraday returns. Besides 1 to 20 minute returns we construct also open-to-close returns that are used for RiskMetrics and DCC-GARCH models. Moreover, we create sub-portfolios consisting of 5, 10, and 15 assets (assets chosen according to market capitalization). Hence in total, we study 18 different datasets.

Sample covers period from July, 1 2005 to January, 3 2012 (1623 trading days), and we consider trades between 9:30 to 16:00 EST time. To ensure sufficient liquidity on the market we explicitly exclude weekends and holidays (New Year's Day, Independence Day, Thanksgiving Day, Christmas). For estimation and forecasting purposes we divide our sample into in-sample spanning from July, 1 2005 to July, 9 2008 and out-of-sample July, 10 2008 to January, 3 2012. For the forecasting, we use rolling window estimation with fixed length of 750 days. Summary statistics of all returns are presented in the Appendix D.

Accuracy of the forecasts is evaluated primarily according to economic criteria. Rationale behind is importance of well-conditioned and invertible forecasts rather than focus on unbiasedness, as unbiased forecast does not necessarily translate into unbiased inverse (Bauwens et al., 2012). As a robustness check we also provide ranking of the models based on statistical loss functions.

4.1 Economic forecasts evaluation

For economic evaluation of volatility forecasts, we use approach of Markowitz (1952). There are two possibilities of constructing optimal portfolio. In the first one we specify expected portfolio return and try to find assets weights minimizing the risk. In the second one expected return of portfolio is maximized according to certain risk. Asset weights, $w = (w_1, \ldots, w_q)'$, maximizing

[2] Assets are ordered according to market capitalization.

utility of risk averse investor can be found by solving following problem

$$\begin{aligned} \min_{w_{t+1}} \quad & w'_{t+1}\widehat{\Sigma}_{t+1|t}w_{t+1} \\ \text{s.t.} \quad & l'w_{t+1} = 1 \\ & w'_{t+1}\widehat{\mu}_{t+1|t} = \mu_P \end{aligned} \tag{21}$$

where w_{t+1} is $q \times 1$ vector of assets weights, $\widehat{\Sigma}_{t+1|t}$ represents a covariance matrix forecast, l denotes a $q \times 1$ vector of ones, $\widehat{\mu}_{t+1|t}$ is a vector of mean forecasts and μ_P stands for portfolio return. Once the optimization problem is solved for different risk levels we are able to construct efficient frontier. Markowitz-type portfolio relies heavily on mean forecasts. As these forecasts might be noisy, portfolio weights and variance can become notably sensitive to changes in assets mean. To overcome these difficulties we also consider problem of finding Global Minimum Variance Portfolio (GMVP). Specification of the optimization problem is similar to Markowitz set-up

$$\begin{aligned} \min_{w_{t+1}} \quad & w'_{t+1}\widehat{\Sigma}_{t+1|t}w_{t+1} \\ \text{s.t.} \quad & l'w_{t+1} = 1 \end{aligned} \tag{22}$$

which can be solved analytically[3]

$$w_{t+1}^{GMV} = \frac{\widehat{\Sigma}_{t+1|t}^{-1}l}{l'\widehat{\Sigma}_{t+1|t}^{-1}l}, \tag{23}$$

with expected return variance being

$$\sigma^{2\,GMV}_{t+1} = w_{t+1}^{GMV\prime}\widehat{\Sigma}_{t+1|t}w_{t+1}^{GMV} = \frac{1}{l'\widehat{\Sigma}_{t+1|t}^{-1}l}. \tag{24}$$

4.2 Statistical forecasts evaluation

For statistical evaluation of covariance forecasts, we employ Root Mean Squared Error (RMSE) loss functions based on the Frobenius norm [4]. As a volatility proxy we use Realized Covariance, Sub-Sampled Realized Covariance (RCOV SS) and Multivariate Realized Kernels estimates at given frequencies i.e. to calculate loss function for forecasts based on 5 minutes Realized Covariance we use Realized Covariance estimates based on 5 minutes data as a benchmark. In case of DCC-GARCH and RiskMetrics forecasts we calculate loss functions using all RCOV, RCOV SS and MRK estimates at all frequencies. The measures are calculated for the $t = 1, \ldots, T$ forecasts as

$$e_{t,t+h} = \Sigma_{t+h} - \widehat{\Sigma}_{t+h|t} \tag{25}$$

$$\mathcal{L}^{RMSE} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\sum_{i,j}\left|e_{t_{i,j}}\right|^2} \tag{26}$$

where $\widehat{\Sigma}_{t+h|t}$ is a covariance matrix forecast and Σ_{t+h} is the volatility proxy.

To test significant differences of competing models, we use the Model Confidence Set (MCS) methodology of Hansen et al. (2011). Given a set of forecasting models, $\mathcal{M}_0$, we identify the

[3] Kempf and Memmel (2006)

[4] Frobenius norm of $m \times n$ matrix A is defined as $\|A\|_F{}^2 = \sum_{i,j} |a_{i,j}|^2$

model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha} \subset \mathcal{M}_0$, which is the set of models that contain the "best" forecasting model given a level of confidence α. For a given model $i \in \mathcal{M}_0$, the p-value is the threshold confidence level. Model i belongs to the MCS only if $\widehat{p}_i \geq \alpha$. MCS methodology repeatedly tests the null hypothesis of equal forecasting accuracy

$$H_{0,\mathcal{M}} : E[\mathcal{L}_{i,t} - \mathcal{L}_{j,t}] = 0, \qquad \text{for all } i, j \in \mathcal{M}$$

with $L_{i,t}$ being an appropriate loss function of the i-th model. Starting with the full set of models, $\mathcal{M} = \mathcal{M}_0$, this procedure sequentially eliminates the worst-performing model from $\mathcal{M}$ when the null is rejected. The surviving set of models then belong to the model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. Following Hansen et al. (2011), we implement the MCS using a stationary bootstrap with an average block length of 10 days.[5]

5 Results

For the clarity of presentation, we begin with the discussion of the results of 1-step ahead forecasts for the portfolio of 5 stocks (AAPL, XOM, GOOG, WMT, MSFT), whereas we leave portfolios of 10 and 15 stocks and also 5-step and 10-step ahead forecasts as a robustness check showing that the methodology works well also in larger dimensions and different forecasting horizons. Focusing on the economic evaluation, we first discuss the results from GMVP,[6] followed by Markowitz approach and statistical evaluation.

We present GMVP comparison through cumulative and annualized risk. In cumulative approach we use covariance forecasts for daily rebalancing of our portfolio – at each step we calculate optimal asset weights and using these weights we calculate corresponding daily portfolio risk. Results presented in the Table 1 are sums of portfolio risk $\sigma_{cum.}$ for whole out-of sample period. Table 1 is divided into 7 parts according to realized measures and frequencies used for the calculation. For RiskMetrics and DCC-GARCH corresponding $\sigma_{cum.}$ are constant for all frequencies because they are calculated using open-close returns. We present results of DCC-GARCH and RiskMetrics in all columns of Table 1 so we can compare performance of covariance based models estimated on different frequencies with daily data based models.

From the Table 1 we can see that the model with lowest level of risk is GHAR estimated on 20 min. Sub-Sampled RCOV followed by GHAR estimated on 15 min. Sub-Sampled RCOV covariance matrix. For the remaining frequencies DCC-GARCH outperformed covariance based models. We can also observe that for RCOV and Sub-Sampled RCOV the lower the frequency the lower the portfolio risk for all covariance based models.

Disadvantage of model comparison according to cumulative risk is daily rebalancing implying high transaction costs. More comprehensive way of model comparison is to use annualized annualized portfolio risks. In Table 2 we present results for annualized version of GMVP.

Similar to cumulative GMVP, model with the lowest achievable risk is GHAR estimated on 20 min. Sub-Sampled RCOV covariance matrix. Remaining results from the Table 1 and Table 2 partly match the results presented in Chiriac and Voev (2011). Model that scored second is VARFIMA followed by HAR for Sub-Sampled RCOV estimated at 15 min. and 20 min. frequency. For the remaining frequencies and realized measures DCC-GARCH outperform

[5]We have used different block lengths, including the ones depending on the forecasting horizons, to assess the robustness of the results, without any change in the final results. These results are available from the authors upon request.

[6]With shortselling allowed.

Table 1: Cumulative version of GMVP - portfolio of 5 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	30,50	30,50	30,50	30,50	30,50	30,50	30,50
RiskMetrics	40,64	40,64	40,64	40,64	40,64	40,64	40,64
ARFIMA	30,76	34,47	32,44	32,84	31,04	29,86	29,31
GHAR	30,60	34,14	32,22	32,53	30,83	29,65	29,08
HAR	31,42	34,84	33,05	33,35	31,61	30,50	29,99

Note: Model with the lowest risk for given frequency is highlighted

Table 2: Annualized version of GMVP - portfolio of 5 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	17,38	17,38	17,38	17,38	17,38	17,38	17,38
RiskMetrics	23,13	23,13	23,13	23,13	23,13	23,13	23,13
VARFIMA	17,62	19,39	18,44	18,61	17,68	17,04	16,77
GHAR	17,32	19,08	18,08	18,27	17,36	16,69	16,38
HAR	18,01	19,61	18,79	18,91	18,01	17,40	17,14

Note: Model with the lowest risk for given frequency is highlighted

covariance based models. Overall we can say that covariance based models with proper choice of realized measure outperform return based models.

To assess the performance of the models not only from the risk minimizing point of view but also return maximization, we present efficient frontiers. In contrast to GMVP we do not allow here the short selling.[7] For the calculation of the efficient frontiers we use annualized forecasts of covariance matrices and annualized returns.

Similar to the results from GMVP evaluation model with the best risk-return tradeoff is model proposed in this paper, GHAR. The second best performing model is VARFIMA, followed by HAR. From the Figure 1 we can see that for estimates at 1 minutes RCOV and 5 minutes RCOV score of DCC-GARCH is better than all covariance based models, which is not in line with results presented in Chiriac and Voev (2011) where DCC-GARCH ended penultimate. We can address this difference to different dataset and period that include financial crisis during which periods of high intraday volatility are observable. For Sub-Sampled RCOV we observe decreasing risk and increasing returns with increasing frequency of realized measure.

As a robustness check to the economic evaluation, we provide also results from statistical comparison of forecasting performance of the competing models. In the Table 3 comparison based on the RMSE loss function is presented.

From the RMSE perspective lowest error has HAR model followed by VARFIMA and GHAR. These models always belong to 5% MCS irrespective of realized measure used for comparison. The worst performance has RiskMetrics which does not belong to 5% MCS in two cases and it has the highest RMSE in 5 out of 7 cases.

[7] In case the short-selling is allowed the ranking of the models is unchanged only the magnitude differ.

Table 3: RMSE – portfolio of 5 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	1.593	1.730	1.914	1.707	1.547	1.481	1.474
RiskMetrics	1.668	1.728	1.866	1.709	1.646	1.636	1.633
VARFIMA	1.406	1.537	1.682	1.473	1.363	1.331	1.328
GHAR	1.490	1.401	1.740	1.509	1.438	1.430	1.445
HAR	1.190	1.100	1.380	1.162	1.125	1.144	1.158

Note: Values are scaled by 10^{-3}; highlighted cells belongs to 5% MCS

5.1 Robustness check

Having discussed the results of 1-step ahead forecasts for portfolio consisting of 5 stocks we now turn to evaluation of 1-step ahead forecasts for portfolio consisting of 10 (AAPL, XOM, GOOG, WMT, MSFT, GE, IBM, JNJ, CVX, PG), and 15 (AAPL, XOM, GOOG, WMT, MSFT, GE, IBM, JNJ, CVX, PG, PFE, T, WFC, JPM, KO) stocks and 5 and 10-step ahead forecasts for portfolios consisting of 5,10 and 15 stocks. We will concentrate on main differences compared to smaller portfolio, as we use these results as a robustness check. We also relegate the Tables and Figures to Appendix A: 1 step ahead forecasts, Appendix B: 5 step ahead forecasts and Appendix C: 10 step ahead forecasts.

5.1.1 Portfolio of 10 and 15 stocks

According to GMVP criteria for portfolio consisting of 10 stocks, results do not differ from results obtained in portfolio of 5 stocks. Model with the lowest cumulative and annualized risk is GHAR estimated on 20 min. Sub-Sampled RCOV. In case of the portfolio consisting of 15 stocks only difference is that GHAR estimated on MRK covariance matrices outperformed DCC-GARCH.

From the risk-return tradeoff point of view there is notable difference for portfolio consisting of 10 stocks when the data of higher frequencies (1,5 and 10 minutes) are used. In case of these frequencies, model with the best risk-return tradeoff is DCC-GARCH. Remaining order of the models is identical to portfolio of 5 stocks – GHAR followed by VARFIMA and HAR. If the 15 minute data are used for optimization, GHAR share the first place with DCC-GARCH. These two models are closely followed by VARFIMA and HAR. For the 20 minutes data ordering of the models is similar to portfolio consisting of 5 stocks.

Concentrating on statistical evaluation, results of RMSE model comparison for portfolio consisting of 10 stocks are almost identical to results in portfolio of 5 stocks – only difference is that RiskMetrics does not belong to 5% MCS in any of the cases. On the other hand notable difference occurs in comparison of portfolio consisting of 15 stocks where GHAR belongs to 5% MCS only in one case (estimated at 5 min RCOV) and DCC-GARCH and RiskMetrics do not belong to 5% MCS at all. We address unambiguous results of statistical evaluation to problem with selecting “correct” proxy. These results are also consistent with findings in Kyj et al. (2010), who show that for large portfolios, the pure high frequency based covariance forecasts need to be conditioned in order to achieve the benefits of the high frequency data.

This points us to the result, that unmodelled dependence from HAR and VARFIMA models is increasing with increasing dimension of the portfolio. Hence GHAR model delivers significant economic gains with increasing dimension of portfolio.

5.1.2 5-step & 10-step ahead forecasts [8]

Extension of forecasting horizon from one to five/ten days does not substantially change results of our analysis. Only notable difference is absence of GHAR in 5% MCS in case of 10-step ahead forecasts of portfolio consisting of 15 stocks. Remaining results supports our previous findings that application of seemingly unrelated regression for HAR estimation delivers significant economic gains regardless the size of the portfolio and/or forecasting horizon.

6 Conclusion

In this paper we propose to employ seemingly unrelated regression of Zellner (1962) to estimate multivariate extension of heterogeneous autoregression model in order to improve the variance matrix forecasts. Resulting model, generalized HAR (GHAR), inherit all the favourable properties of HAR, and provides us with more efficient estimator that accounts for otherwise hidden dependencies among variables.

In our setup we closely follow Chiriac and Voev (2011) and model elements of Cholesky decomposed covariance matrices to test the economic and statistical value of the proposed modelling strategy. Moreover, we perform our analysis on portfolios consisting of 5, 10 and 15 assets, we include three covariance matrix estimators (realized covariation, sub-sampled realized covariation and multivariate realized kernels), and we obtain covariance matrix estimates using high-frequency data of five different frequencies (1,5,10,15 and 20 minutes). Overall we test performance of GHAR estimator on 15 different high-frequency datasets. Resulting forecasts of GHAR prove to perform significantly better than benchmark models according to Global Minimum Variance Portfolio and Mean-Variance evaluation criteria irrespective of frequency or size of the portfolio. If the statistical evaluations are used for models comparison, we find that GHAR is not systematically dominated by any benchmark model.

References

Aït-Sahalia, Y., J. Fan, and D. Xiu (2010). High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association 105*, 1504–1517.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (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*(2), 149–169.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

Bauer, G. H. and K. Vorkink (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics 160*(1), 93–101.

[8] To make the results comparable we scale them according to forecasting horizon

Bauwens, L., C. M. Hafner, and S. Laurent (2012). *Handbook of volatility models and their applications*, Volume 3. John Wiley & Sons.

Bauwens, L., S. Laurent, and J. V. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of applied econometrics 21*(1), 79–109.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Rev Econ Stat 72*, 498–505.

Bonato, M. (2009). Estimating the degrees of freedom of the realized volatility Wishart autoregressive model. *Available at SSRN 1357044*.

Bonato, M., M. Caporin, and A. Ranaldo (2013). Risk spillovers in international equity portfolios. *Journal of Empirical Finance 24*, 121–137.

Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics 26*(6), 922–947.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008). The volatility of realized volatility. *Econometric Reviews 27*(1-3), 46–78.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric theory 11*(01), 122–150.

Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Technical report, National Bureau of Economic Research.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association* (74), 291–296.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*(2), 167–181.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

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

Jin, X. and J. M. Maheu (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics 11*(2), 335–369.

Kempf, A. and C. Memmel (2006). ESTIMATING THE GLOBAL MINIMUM VARIANCE PORTFOLIO. *Schmalenbach Business Review (sbr) 58*(4).

Kyj, L. M., B. Ostdiek, and K. Ensor (2010). Covariance estimation in dynamic portfolio optimization: a realized single factor model. In *American Finance Association 2010 Atlanta Meetings Paper*.

Longerstaey, J. and M. Spencer (1996). RiskMetricsTM—Technical Document. *Morgan Guaranty Trust Company of New York: New York*.

Markowitz, H. (1952). Portfolio selection. *The journal of finance 7*(1), 77–91.

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

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association 57*(298), 348–368.

Zhang, L., P. A. Mykland, and Y. Ait-Sahalia (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association 100*.

Appendix A: 1 step ahead forecasts

Table 4: GMVP - portfolio of 10 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	22,14	22,14	22,14	22,14	22,14	22,14	22,14
RiskMetrics	42,15	42,15	42,15	42,15	42,15	42,15	42,15
VARFIMA	23,34	27,70	24,75	25,64	23,82	22,52	21,85
GHAR	22,50	26,71	23,90	24,79	22,98	21,66	20,98
HAR	24,28	28,30	25,66	26,40	24,63	23,39	22,79
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	13,12	13,12	13,12	13,12	13,12	13,12	13,12
RiskMetrics	24,32	24,32	24,32	24,32	24,32	24,32	24,32
VARFIMA	13,74	15,76	14,40	14,84	13,90	13,21	12,88
GHAR	12,82	15,00	13,53	14,04	13,03	12,30	11,91
HAR	14,31	16,14	14,96	15,31	14,40	13,74	13,43

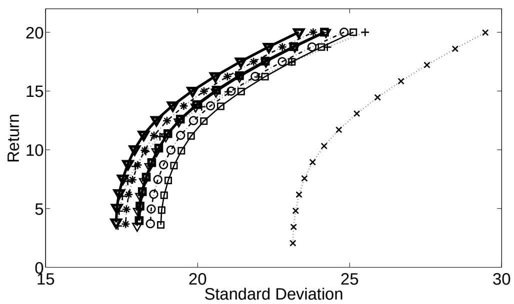
Note: Model with the lowest risk for given frequency is highlighted

Table 5: RMSE – portfolio of 10 stocks

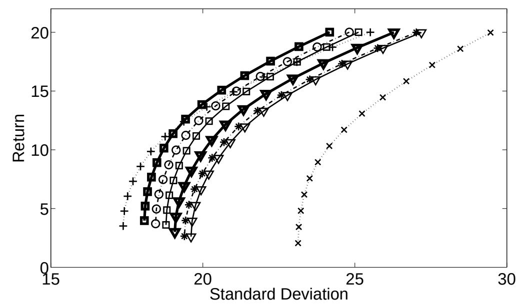
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	3.242	3.624	3.896	3.600	3.162	3.044	3.085
RiskMetrics	3.808	4.006	4.167	3.949	3.803	3.822	3.846
VARFIMA	2.592	3.028	3.228	2.903	2.551	2.494	2.539
GHAR	3.101	3.109	3.639	3.237	2.988	2.965	3.057
HAR	2.295	2.271	2.837	2.405	2.181	2.213	2.307

Note: Values are scaled by 10^{-3}; highlighted cells belongs to 5% MCS

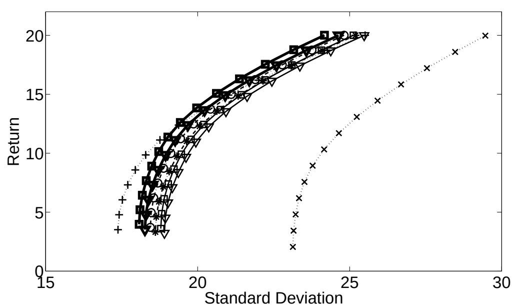
(a) RCOV 5min vs. MRK



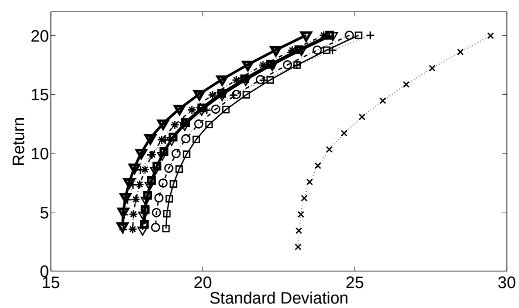
(b) RCOV 5 min vs. RCOV 1 min



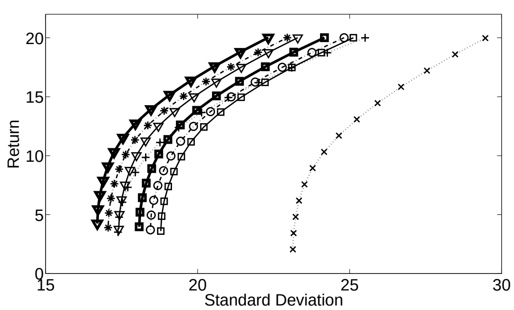
(c) RCOV 5 min vs. RCOV SS 5 min



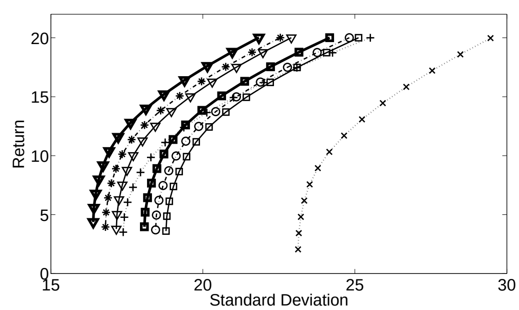
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

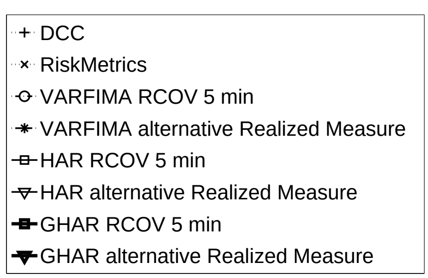


Figure 1: Efficient frontiers - portfolio of 5 stocks

Table 6: GMVP - portfolio of 15 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
Cumulative							
DCC	20,72	20,72	20,72	20,72	20,72	20,72	20,72
RiskMetrics	56,67	56,67	56,67	56,67	56,67	56,67	56,67
VARFIMA	21,34	25,63	22,71	23,71	21,91	20,62	19,93
GHAR	20,37	24,46	21,75	22,59	20,90	19,66	18,97
HAR	22,25	26,21	23,52	24,42	22,69	21,47	20,83
Annualized							
DCC	12,64	12,64	12,64	12,64	12,64	12,64	12,64
RiskMetrics	32,19	32,19	32,19	32,19	32,19	32,19	32,19
VARFIMA	12,88	14,80	13,52	13,99	13,06	12,40	12,07
GHAR	11,64	13,82	12,39	12,86	11,91	11,22	10,83
HAR	13,43	15,21	14,06	14,45	13,56	12,92	12,62

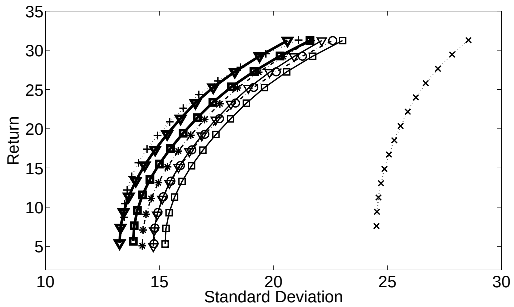
Note: Model with the lowest risk for given frequency is highlighted

Table 7: RMSE – portfolio of 15 stocks

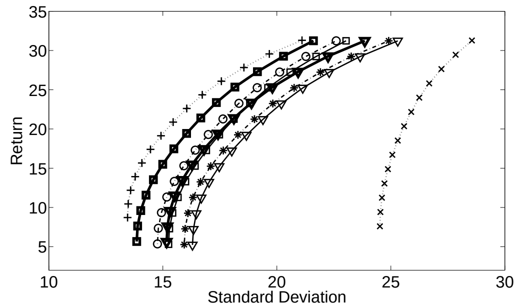
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	5.323	5.601	6.064	5.793	5.158	5.023	5.058
RiskMetrics	11.905	11.881	12.030	11.902	11.952	12.044	12.030
VARFIMA	4.555	4.809	5.207	4.900	4.374	4.276	4.323
GHAR	5.881	5.352	6.342	5.918	5.565	5.521	5.677
HAR	4.285	3.599	4.832	4.226	3.948	4.005	4.150

Note: Values are scaled by 10^{-3}; highlighted cells belongs to 5% MCS

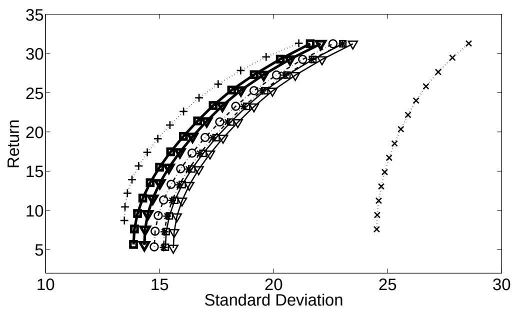
(a) RCOV 5min vs. MRK



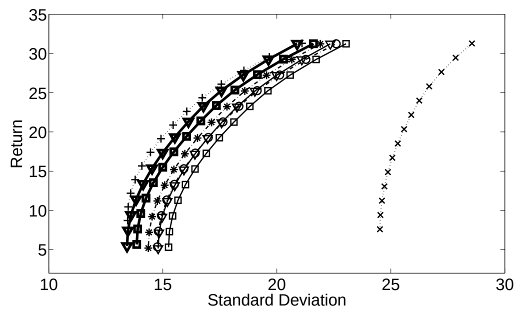
(b) RCOV 5 min vs. RCOV 1 min



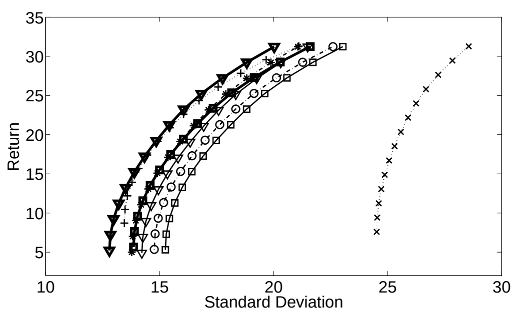
(c) RCOV 5 min vs. RCOV SS 5 min



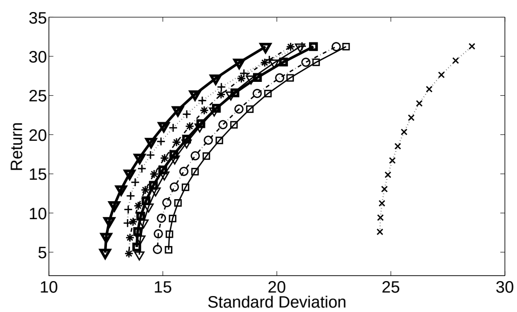
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

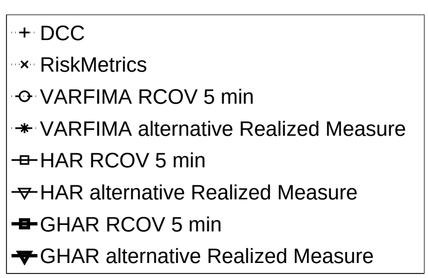
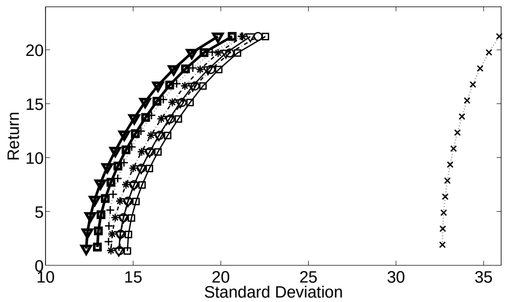
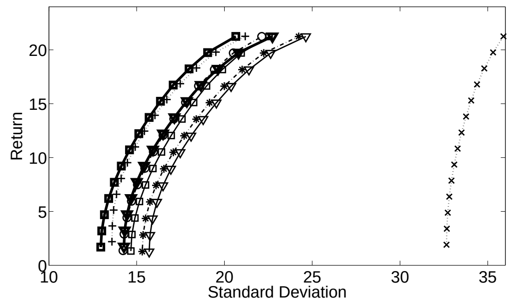


Figure 2: Efficient frontiers - portfolio of 10 stocks

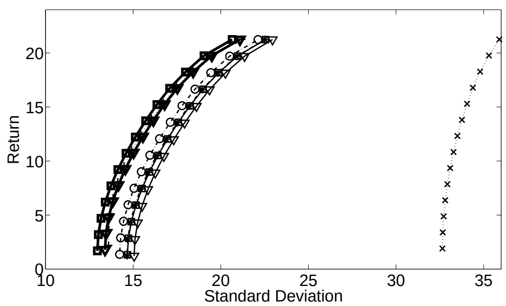
(a) RCOV 5 min vs. MRK



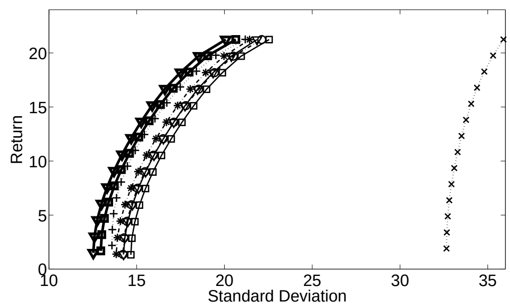
(b) RCOV 5 min vs. RCOV 1 min



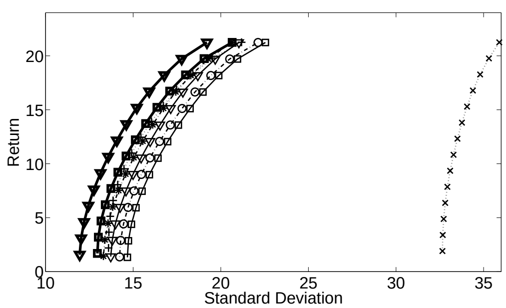
(c) RCOV 5 min vs. RCOV SS 5 min



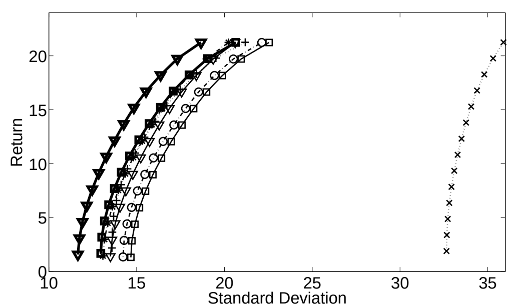
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

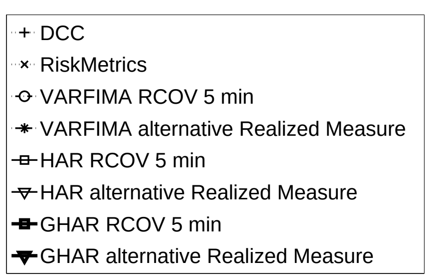


Figure 3: Efficient frontiers - portfolio of 15 stocks

Appendix B: 5 step ahead forecasts

Table 8: GMVP - portfolio of 5 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	30.50	30.50	30.50	30.50	30.50	30.50	30.50
RiskMetrics	40.61	40.61	40.61	40.61	40.61	40.61	40.61
VARFIMA	30.53	34.06	32.09	32.49	30.78	29.64	29.10
GHAR	30.49	33.88	32.07	32.36	30.72	29.54	28.96
HAR	31.30	34.62	32.86	33.19	31.47	30.38	29.87
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	17.38	17.38	17.38	17.38	17.38	17.38	17.38
RiskMetrics	23.17	23.17	23.17	23.17	23.17	23.17	23.17
VARFIMA	17.28	19.02	18.06	18.24	17.35	16.73	16.73
GHAR	17.18	18.87	17.93	18.12	17.23	16.57	16.57
HAR	17.85	19.45	18.63	18.75	17.86	17.25	17.25

Note: Model with the lowest risk for given frequency is highlighted; values are scaled by forecasting horizon

Table 9: RMSE – portfolio of 5 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	1.193	1.293	1.376	1.288	1.152	1.081	1.079
RiskMetrics	1.296	1.317	1.330	1.314	1.290	1.288	1.285
VARFIMA	1.043	1.023	1.153	1.055	0.993	0.968	0.978
GHAR	1.261	1.195	1.382	1.273	1.206	1.174	1.189
HAR	1.024	0.980	1.100	1.028	0.968	0.951	0.966

Note: Values are scaled by 10^{-3} and by forecasting horizon; highlighted cells belongs to 5% MCS

Table 10: GMVP - portfolio of 10 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	22.10	22.10	22.10	22.10	22.10	22.10	22.10
RiskMetrics	42.12	42.12	42.12	42.12	42.12	42.12	42.12
VARFIMA	23.11	27.25	24.44	25.27	23.55	22.30	21.65
GHAR	22.33	26.45	23.72	24.59	22.80	21.50	20.82
HAR	24.25	28.14	25.56	26.30	24.57	23.35	22.75
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	13.07	13.07	13.07	13.07	13.07	13.07	13.07
RiskMetrics	24.36	24.36	24.36	24.36	24.36	24.36	24.36
VARFIMA	13.38	15.36	14.03	14.44	13.54	12.88	12.54
GHAR	12.67	14.80	13.38	13.87	12.88	12.16	11.78
HAR	14.15	15.99	14.81	15.15	14.24	13.59	13.28

Note: Model with the lowest risk for given frequency is highlighted; values are scaled by forecasting horizon

Table 11: RMSE – portfolio of 10 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	2.487	2.683	2.773	2.690	2.402	2.290	2.309
RiskMetrics	3.232	3.250	3.217	3.222	3.242	3.278	3.267
VARFIMA	1.952	1.966	2.166	2.024	1.867	1.833	1.872
GHAR	2.598	2.480	2.759	2.611	2.481	2.445	2.501
HAR	1.950	1.881	2.103	1.984	1.845	1.826	1.877

Note: Values are scaled by 10^{-3} and by forecasting horizon; highlighted cells belongs to 5% MCS

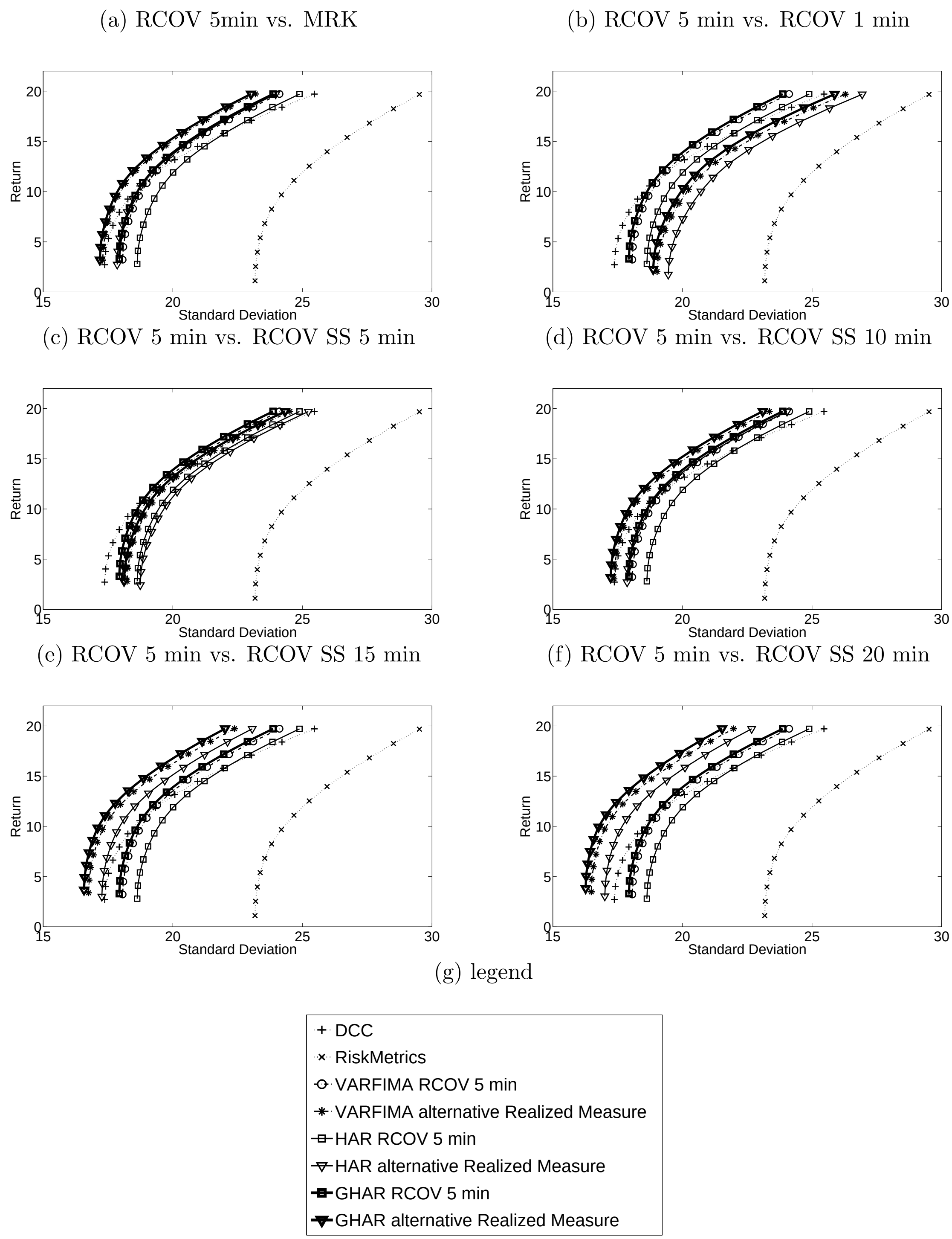


Figure 4: Efficient frontiers - portfolio of 5 stocks

Table 12: GMVP - portfolio of 15 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	20.70	20.70	20.70	20.70	20.70	20.70	20.70
RiskMetrics	56.64	56.64	56.64	56.64	56.64	56.64	56.64
VARFIMA	21.23	25.28	22.52	23.44	21.75	20.52	19.86
GHAR	20.31	24.30	21.65	22.45	20.83	19.62	18.92
HAR	22.31	26.13	23.51	24.40	22.72	21.51	20.89
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	12.60	12.60	12.60	12.60	12.60	12.60	12.60
RiskMetrics	32.25	32.25	32.25	32.25	32.25	32.25	32.25
VARFIMA	12.53	14.43	13.17	13.60	12.72	12.07	11.74
GHAR	11.53	13.66	12.26	12.70	11.79	11.12	10.73
HAR	13.29	15.07	13.94	14.31	13.42	12.78	12.48

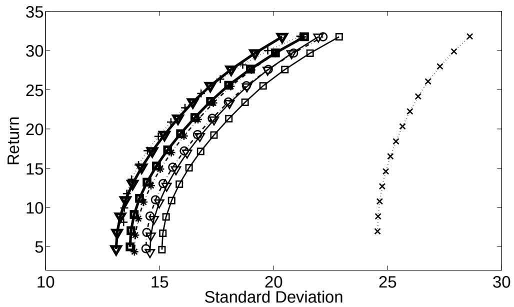
Note: Model with the lowest risk for given frequency is highlighted; values are scaled by forecasting horizon

Table 13: RMSE – portfolio of 15 stocks

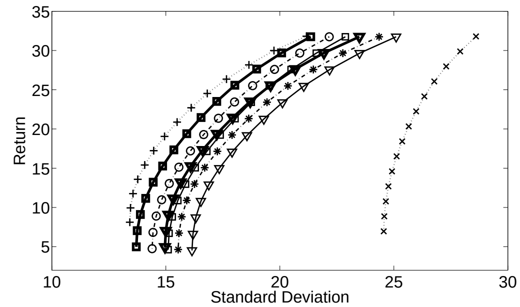
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	4.110	4.251	4.384	4.329	3.992	3.919	3.949
RiskMetrics	11.404	11.318	11.262	11.260	11.487	11.599	11.573
VARFIMA	3.453	3.223	3.596	3.422	3.239	3.201	3.283
GHAR	4.913	4.490	4.961	4.821	4.644	4.590	4.706
HAR	3.575	3.216	3.644	3.489	3.331	3.314	3.421

Note: Values are scaled by 10^{-3} and by forecasting horizon; highlighted cells belongs to 5% MCS

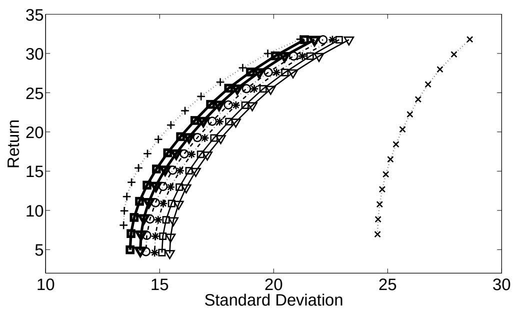
(a) RCOV 5min vs. MRK



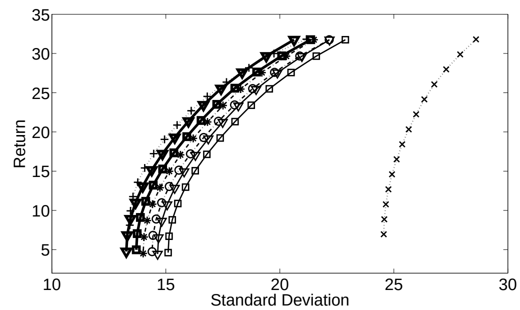
(b) RCOV 5 min vs. RCOV 1 min



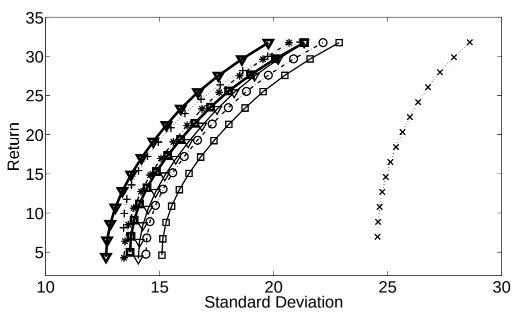
(c) RCOV 5 min vs. RCOV SS 5 min



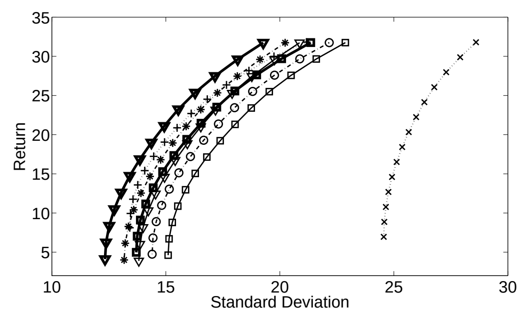
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

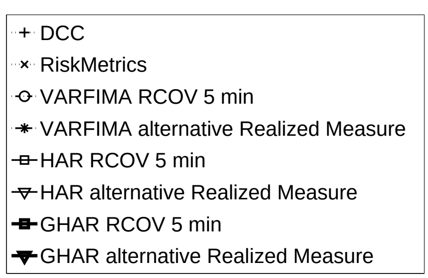
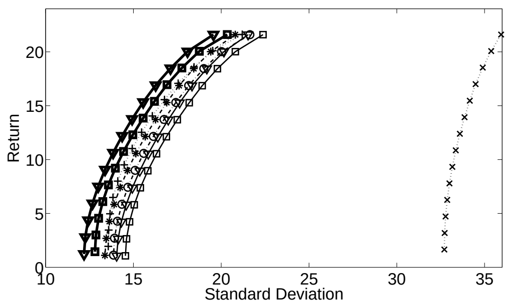
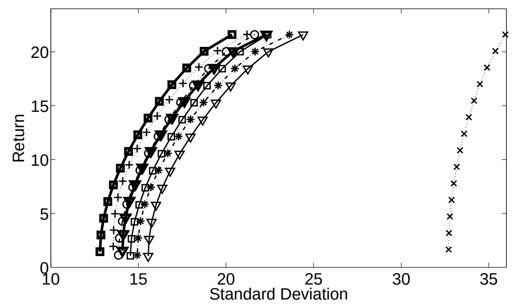


Figure 5: Efficient frontiers - portfolio of 10 stocks

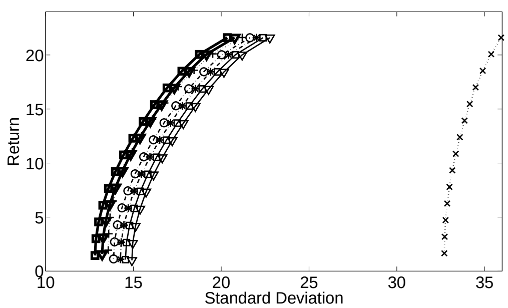
(a) RCOV 5min vs. MRK



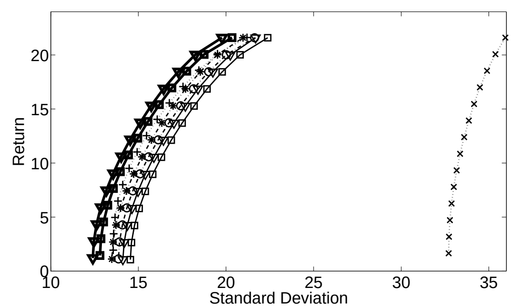
(b) RCOV 5 min vs. RCOV 1 min



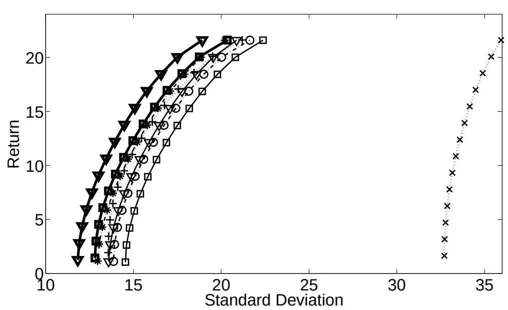
(c) RCOV 5 min vs. RCOV SS 5 min



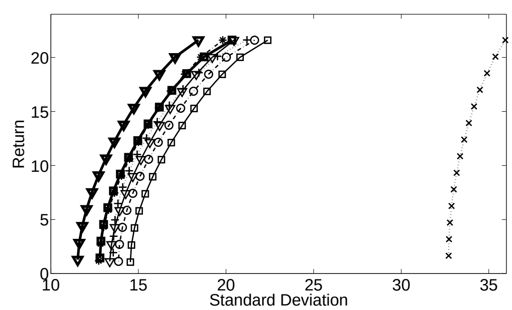
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

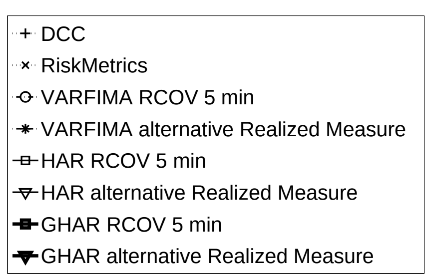


Figure 6: Efficient frontiers - portfolio of 15 stocks

Appendix C: 10 step ahead forecasts

Table 14: GMVP - portfolio of 5 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	30.50	30.50	30.50	30.50	30.50	30.50	30.50
RiskMetrics	40.58	40.58	40.58	40.58	40.58	40.58	40.58
VARFIMA	30.30	33.75	31.80	32.20	30.53	29.41	28.88
GHAR	30.35	33.66	31.94	32.21	30.58	29.40	28.81
HAR	31.16	34.40	32.68	33.01	31.30	30.24	29.74
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	17,39	17,39	17,39	17,39	17,39	17,39	17,39
RiskMetrics	23,22	23,22	23,22	23,22	23,22	23,22	23,22
VARFIMA	17,07	18,82	17,84	18,02	17,15	16,54	16,26
GHAR	17,11	18,74	17,86	18,04	17,15	16,49	16,16
HAR	17,72	19,32	18,49	18,62	17,73	17,13	16,86

Note: Model with the lowest risk for given frequency is highlighted; values are scaled by forecasting horizon

Table 15: RMSE – portfolio of 5 stocks

	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	1.208	1.294	1.375	1.291	1.173	1.107	1.101
RiskMetrics	1.389	1.401	1.431	1.404	1.388	1.384	1.380
VARFIMA	1.153	1.147	1.266	1.173	1.106	1.072	1.078
GHAR	1.287	1.256	1.409	1.307	1.237	1.197	1.205
HAR	1.138	1.133	1.242	1.163	1.091	1.058	1.067

Note: Values are scaled by 10^{-3} and by forecasting horizon; highlighted cells belongs to 5% MCS

Table 16: GMVP - portfolio of 10 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	22.05	22.05	22.05	22.05	22.05	22.05	22.05
RiskMetrics	42.08	42.08	42.08	42.08	42.08	42.08	42.08
VARFIMA	22.89	26.91	24.17	24.97	23.31	22.09	21.45
GHAR	22.16	26.23	23.55	24.40	22.61	21.33	20.66
HAR	24.15	27.94	25.42	26.14	24.45	23.25	22.66
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	13.03	13.03	13.03	13.03	13.03	13.03	13.03
RiskMetrics	24.40	24.40	24.40	24.40	24.40	24.40	24.40
VARFIMA	13.16	15.13	13.80	14.20	13.32	12.67	12.33
GHAR	12.56	14.67	13.28	13.75	12.76	12.06	11.69
HAR	14.01	15.85	14.67	15.01	14.10	13.45	13.14

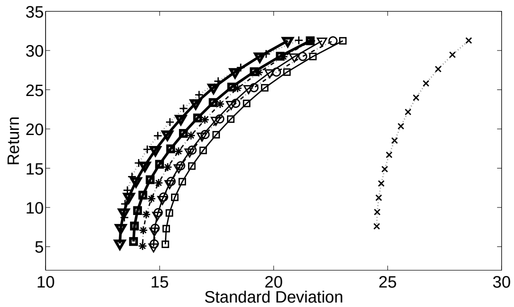
Note: Model with the lowest risk for given frequency is highlighted; values are scaled by forecasting horizon

Table 17: RMSE – portfolio of 10 stocks

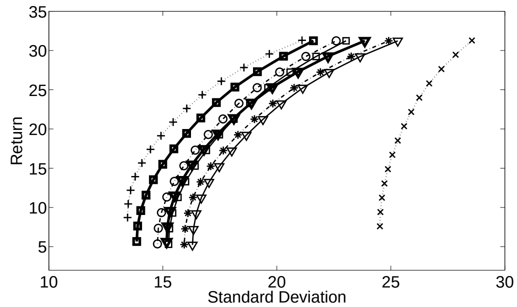
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	2.437	2.609	2.687	2.610	2.362	2.260	2.271
RiskMetrics	3.445	3.461	3.455	3.448	3.458	3.487	3.481
VARFIMA	2.139	2.165	2.327	2.208	2.057	2.011	2.041
GHAR	2.605	2.514	2.729	2.607	2.494	2.449	2.491
HAR	2.114	2.110	2.276	2.174	2.026	1.986	2.024

Note: Values are scaled by 10^{-3} and by forecasting horizon; highlighted cells belongs to 5% MCS

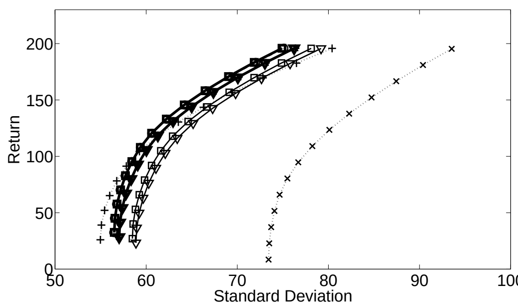
(a) RCOV 5min vs. MRK



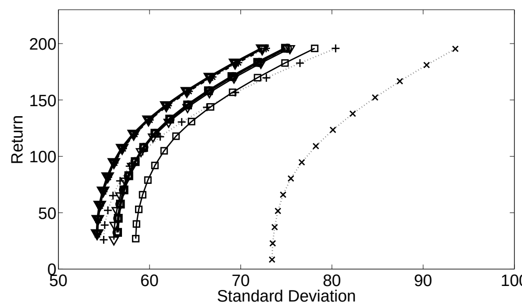
(b) RCOV 5 min vs. RCOV 1 min



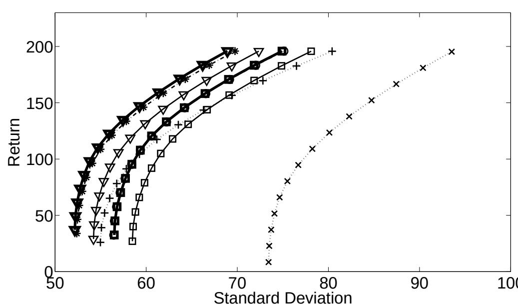
(c) RCOV 5 min vs. RCOV SS 5 min



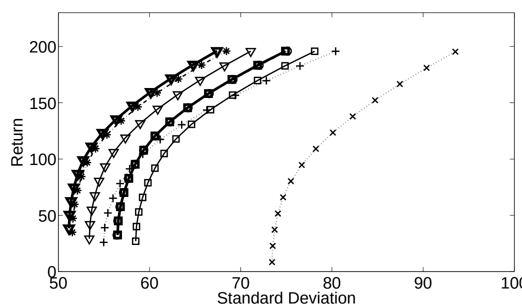
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

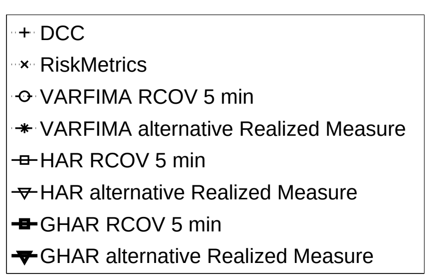


Figure 7: Efficient frontiers - portfolio of 5 stocks

Table 18: GMVP - portfolio of 15 stocks

	Cumulative						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	20.67	20.67	20.67	20.67	20.67	20.67	20.67
RiskMetrics	56.60	56.60	56.60	56.60	56.60	56.60	56.60
VARFIMA	21.08	25.00	22.33	23.21	21.56	20.36	19.72
GHAR	20.21	24.13	21.54	22.30	20.72	19.53	18.83
HAR	22.31	26.00	23.46	24.32	22.68	21.49	20.88
	Annualized						
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	12,56	12,56	12,56	12,56	12,56	12,56	12,56
RiskMetrics	32,32	32,32	32,32	32,32	32,32	32,32	32,32
VARFIMA	12,32	14,21	12,95	13,38	12,50	11,86	11,53
GHAR	11,44	13,55	12,16	12,59	11,70	11,04	10,66
HAR	13,19	14,95	13,82	14,19	13,31	12,68	12,37

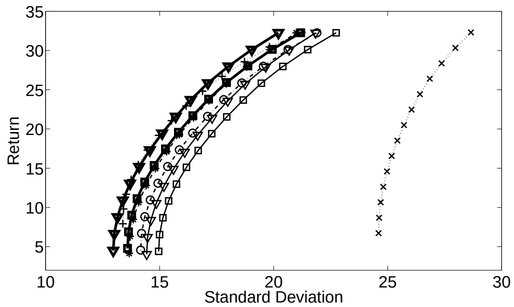
Note: Model with the lowest risk for given frequency is highlighted; values are scaled by forecasting horizon

Table 19: RMSE – portfolio of 15 stocks

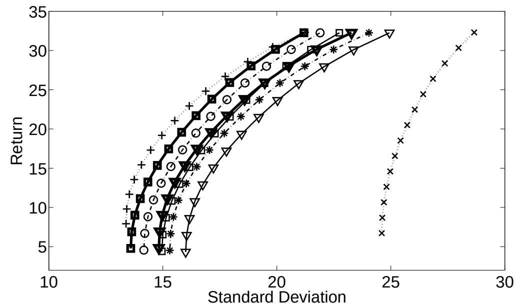
	MRK	RCOV		Sub-Sampled RCOV			
	1 min	1 min	5 min	5 min	10 min	15 min	20 min
DCC	4.141	4.258	4.385	4.323	4.054	3.989	4.010
RiskMetrics	11.806	11.735	11.720	11.719	11.884	11.981	11.961
VARFIMA	3.690	3.542	3.821	3.680	3.496	3.439	3.509
GHAR	4.807	4.514	4.859	4.746	4.571	4.508	4.613
HAR	3.666	3.471	3.767	3.635	3.468	3.424	3.512

Note: Values are scaled by 10^{-3} and by forecasting horizon; highlighted cells belongs to 5% MCS

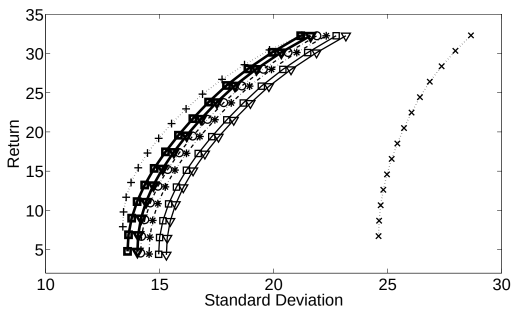
(a) RCOV 5min vs. MRK



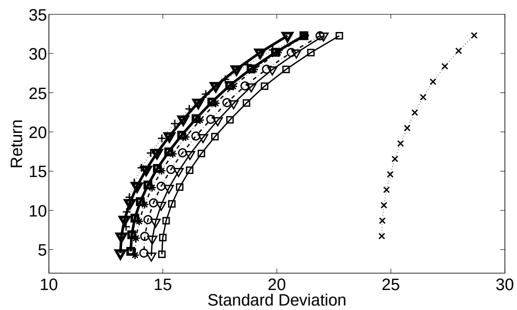
(b) RCOV 5 min vs. RCOV 1 min



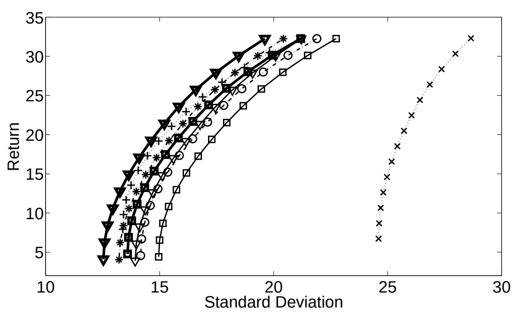
(c) RCOV 5 min vs. RCOV SS 5 min



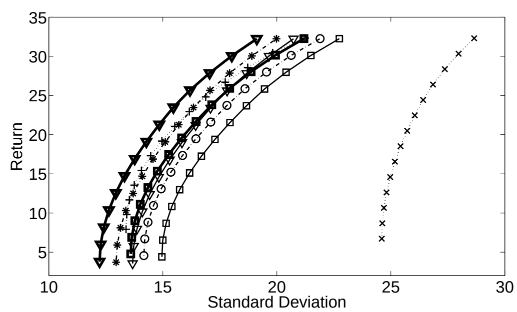
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

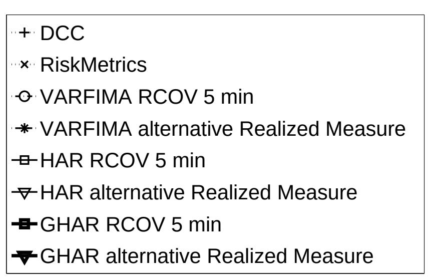
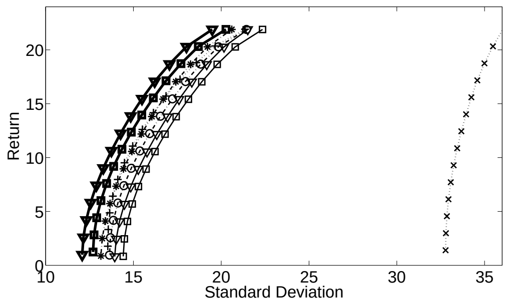
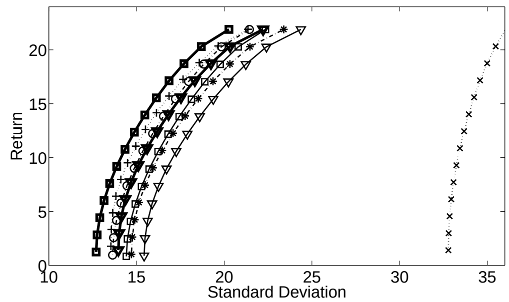


Figure 8: Efficient frontiers - portfolio of 10 stocks

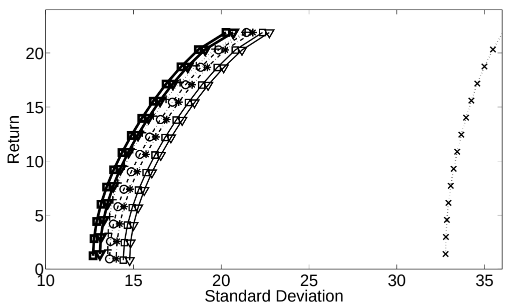
(a) RCOV 5min vs. MRK



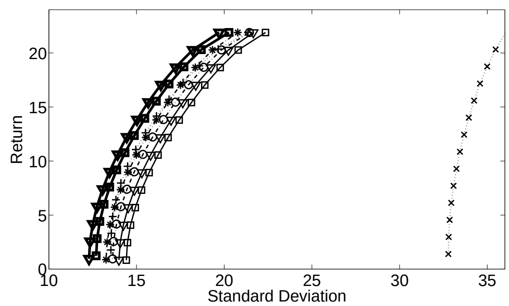
(b) RCOV 5 min vs. RCOV 1 min



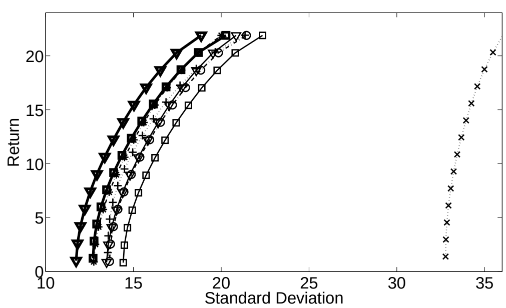
(c) RCOV 5 min vs. RCOV SS 5 min



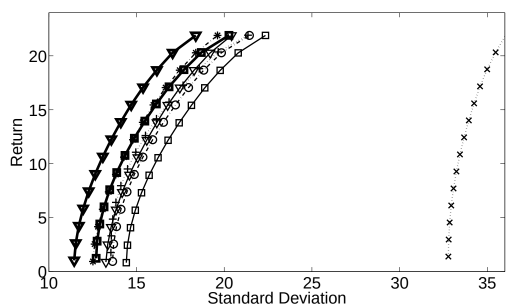
(d) RCOV 5 min vs. RCOV SS 10 min



(e) RCOV 5 min vs. RCOV SS 15 min



(f) RCOV 5 min vs. RCOV SS 20 min



(g) legend

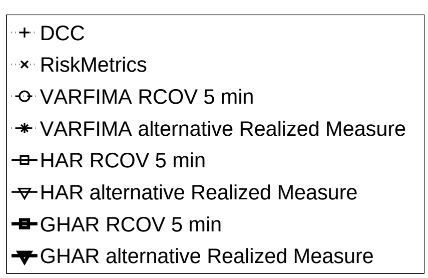


Figure 9: Efficient frontiers - portfolio of 15 stocks

Appendix D

Table 20: Descriptive statistics of returns over the period 01.07.2005 – 03.01.2012

	AAPL	CVX	GE	GOOG	IBM	JNJ	JPM	KO	MSFT	PFE	PG	T	WFC	WMT	XOM
	1 min														
Mean	-0.105	0.086	-0.218	-0.204	0.215	-0.035	-0.036	0.011	-0.040	-0.150	0.133	-0.079	-0.049	-0.018	0.099
Max	0.046	0.043	0.050	0.022	0.030	0.044	0.032	0.020	0.019	0.032	0.032	0.039	0.042	0.030	0.040
Min	-0.037	-0.027	-0.032	-0.041	-0.020	-0.033	-0.060	-0.038	-0.025	-0.028	-0.028	-0.038	-0.049	-0.021	-0.034
SD	1.046	0.891	1.088	0.936	0.749	0.559	1.336	0.619	0.833	0.825	0.616	0.834	1.464	0.685	0.827
Skewness	0.037	0.262	0.181	-0.369	0.077	0.394	-0.144	-0.482	-0.063	0.103	-0.066	-0.122	0.068	0.365	-0.188
Kurtosis	37.100	44.353	43.467	34.940	40.526	126.707	41.663	70.490	18.848	30.256	71.747	52.025	40.587	39.587	50.690
	5 min														
Mean	-0.629	0.423	-0.994	-1.037	1.177	-0.132	-0.165	0.113	-0.138	-0.745	0.741	-0.353	-0.252	-0.057	0.526
Max	0.065	0.061	0.052	0.046	0.053	0.032	0.069	0.028	0.030	0.030	0.050	0.034	0.066	0.048	0.053
Min	-0.048	-0.068	-0.046	-0.069	-0.036	-0.040	-0.068	-0.037	-0.028	-0.038	-0.062	-0.073	-0.077	-0.042	-0.059
SD	2.258	1.916	2.280	2.023	1.580	1.174	2.871	1.297	1.756	1.694	1.316	1.779	3.151	1.478	1.779
Skewness	-0.008	-0.062	0.268	-0.509	0.121	-0.127	0.059	-0.300	-0.091	0.109	-0.460	-0.518	0.095	0.432	-0.098
Kurtosis	28.079	37.935	31.858	39.555	37.413	44.133	35.720	33.284	17.395	18.627	86.196	46.840	34.793	42.507	40.512
	10 min														
Mean	-0.960	1.129	-1.740	-1.820	2.856	0.118	-0.611	0.387	0.350	-1.416	1.832	-0.444	-0.827	0.289	1.410
Max	0.050	0.039	0.052	0.043	0.029	0.030	0.067	0.031	0.029	0.029	0.024	0.038	0.069	0.053	0.051
Min	-0.079	-0.034	-0.058	-0.073	-0.036	-0.025	-0.102	-0.040	-0.038	-0.027	-0.031	-0.043	-0.092	-0.035	-0.067
SD	3.150	2.630	3.168	2.776	2.169	1.592	3.937	1.793	2.399	2.297	1.765	2.412	4.372	1.999	2.423
Skewness	-0.301	0.229	0.256	-0.373	-0.161	0.310	0.035	-0.321	-0.018	0.244	-0.042	-0.114	0.082	0.435	-0.098
Kurtosis	24.897	15.055	29.864	25.773	20.025	21.991	31.039	27.228	15.157	12.988	20.916	21.321	30.512	23.805	27.665
	15 min														
Mean	-1.415	2.229	-2.493	-2.598	4.830	0.250	-0.769	0.643	0.765	-1.986	2.917	-0.370	-1.003	0.712	3.193
Max	0.058	0.046	0.071	0.049	0.038	0.025	0.113	0.030	0.032	0.039	0.028	0.046	0.099	0.051	0.047
Min	-0.053	-0.037	-0.070	-0.068	-0.053	-0.024	-0.086	-0.041	-0.042	-0.029	-0.034	-0.053	-0.075	-0.035	-0.037
SD	3.801	3.186	3.877	3.350	2.630	1.946	4.896	2.175	2.921	2.794	2.138	2.951	5.335	2.445	2.925
Skewness	-0.012	0.237	0.161	-0.242	-0.159	0.332	0.314	-0.319	-0.050	0.264	0.086	-0.052	0.421	0.562	0.263
Kurtosis	16.529	15.547	31.899	22.586	21.422	19.781	35.685	22.721	14.785	13.138	21.422	21.883	30.411	20.882	19.090
	20 min														
Mean	-1.950	2.227	-4.353	-3.743	5.912	-0.276	-1.903	0.412	0.445	-2.604	3.494	-0.812	-2.259	0.542	3.371
Max	0.050	0.059	0.062	0.043	0.036	0.034	0.074	0.035	0.034	0.041	0.026	0.053	0.086	0.053	0.069
Min	-0.048	-0.037	-0.068	-0.118	-0.040	-0.021	-0.102	-0.040	-0.038	-0.029	-0.029	-0.049	-0.080	-0.024	-0.067
SD	4.245	3.608	4.350	3.775	2.939	2.159	5.381	2.446	3.259	3.131	2.380	3.300	6.064	2.741	3.325
Skewness	-0.075	0.252	0.130	-0.867	-0.034	0.420	-0.051	-0.148	-0.044	0.256	0.087	-0.086	0.161	0.510	0.177
Kurtosis	13.207	14.377	26.975	42.831	17.258	17.410	23.170	19.581	12.493	11.697	15.784	19.932	25.774	17.183	25.829

Note: The means are scaled by 10^5, the standard deviations are scaled by 10^3

IES Working Paper Series

2014

1. Diana Zigraiova, Petr Jakubik: *Systemic Event Prediction by Early Warning System*
2. Pavla Blahova, Karel Janda, Ladislav Kristoufek: *The Perspectives for Genetically Modified Cellulosic Ethanol in the Czech Republic*
3. Oxana Babecká Kucharčuková: *The impact of exchange rate volatility on trade: Evidence for the Czech Republic*
4. Oliver Polyák: *The Impact of Euro Adoption on Export Performance: Comparison of the Czech Republic and Slovakia*
5. Lucie Davidová, Vladimír Benáček: *Determinants of Austrian International Trade: Analysis Based on the Gravity Model*
6. Julie Chytilová, Václav Korbel: *Individual and Group Cheating Behavior: A Field Experiment with Adolescents*
7. Petr Janský: *Consumer Demand System Estimation and Value Added Tax Reforms in the Czech Republic*
8. Pavel Hait, Petr Janský: *Inflation Differentials among Czech Households*
9. Hana Džmuráňová, Petr Teplý: *Risk management of savings accounts*
10. Tomáš Fiala, Tomáš Havránek: *Ailing Mothers, Healthy Daughters? Contagion in the Central European Banking Sector*
11. Petra Andrlíková: *Is Barrier version of Merton model more realistic? Evidence from Europe*
12. Martin Pospíšil, Jiří Schwarz: *Bankruptcy, Investment, and Financial Constraints: Evidence from a Post-Transition Economy*
13. Matt Hunt, Goran Serdarević, Tom Ovington, Clive Kenny: *Evidence for a ladder of investment in Central and Eastern European countries*
14. Petra Andrlíková: *Bayesian default probability models*
15. Petra Buzková: *Has the Relationship Between Market and Model CDS Price Changed during the EMU Debt Crisis?*
16. Jakub Černý, Jiří Witzany: *Interest Rate Swap Credit Valuation Adjustment*
17. Martin Gregor: *Receiver's access fee for a single sender*
18. Pavel Ryska: *Deflation and Economic Growth in Long-Term Perspective*
19. Michal Paulus, Eva Michalíková, Vladimír Benáček: *German International Trade: Interpreting Export Flows According to the Gravity Model*
20. Michal Bauer, Nathan Fiala, Ian Levely: *Trusting Former Rebels: An Experimental Approach to Understanding Reintegration after Civil War*
21. Michal Skořepa: *Family Matters: Concurrent Capital Buffers in a Banking Group*
22. Martin Gregor: *Access Fees for Competing Lobbies*
23. Jozef Baruník, František Čech: *On the Modelling and Forecasting Multivariate Realized Volatility: Generalized Heterogeneous Autoregressivee (GHAR) Model*

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz