JOB MARKET PAPER

(Un)Reliable Realized Risk Measures: Portfolio Compositions and Conditional Betas

This Version: December 1, 2015

Rogier Quaedvlieg[a,*]

[a] *Department of Finance, Maastricht University, Netherlands*

Abstract

The estimation and forecasting of dynamically varying covariance matrices plays a crucial role in a host of different economic and financial decision making processes. This paper proposes multivariate volatility models that use the asymptotic theory of high-frequency covariance estimates to endogenously deal with time-variation in the precision of the estimates, by allowing the degree of measurement error attenuation to vary over time. The models allow for increased responsiveness when recent estimates are precise, and limit the impact of noisy high-frequency estimates on forecasts. I illustrate the working of the models in two financial applications. First, I show in a portfolio allocation setting that the forecasts from the new models result in more efficient portfolios and significantly reduced turnover for which a risk-averse investor would be willing to pay up to 180 basis points per year. Second, realized beta forecasts from the dynamic attenuation models lead to improved hedging decisions.

Keywords: Forecasting, Measurement Error, Asset Allocation, Conditional Betas
JEL: C32, C58, G11, G32

☆This research was financially supported by a grant of the Netherlands Organization for Scientific Research (NWO). Part of this research was carried out while I was a visiting scholar at Duke University. I would like to thank Tim Bollerslev, Sébastien Laurent, Andrew Patton and Peter Schotman for comments, and Bingzhi Zhao for providing me with the cleaned data.

*Corresponding author: Department of Finance, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.

Email address: r.quaedvlieg@maastrichtuniversity.nl (Rogier Quaedvlieg)

1. Introduction

The covariation among asset returns is of critical importance for many economic and financial applications, such as portfolio management and hedging decisions. In practice, the covariance matrix has to be estimated and forecasted from historical data, and the out-of-sample performance can be significantly impacted by parameter uncertainty and estimation errors.

A large component of the estimation error is the result of time-variation in covariances. This results in a trade-off between using many observations to accurately estimate the conditional covariance matrix, and only using recent history to better capture the short-term dynamics, likely increasing estimation error. The wide availability of high-frequency data has ameliorated this trade-off as it allows for more accurate estimation and forecasting of covariance matrices, even on a daily basis (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003; Fleming, Kirby, and Ostdiek, 2003).

Although high-frequency measures lead to better estimation of both (co)variances and derived measures of risk such as betas, they are still estimates of latent population quantities and therefore subject to estimation error. The use of these estimates in dynamic models leads to the classical errors-in-variables problem, where parameters are attenuated towards zero. The attenuation intensity is increasing in the variance of the measurement error. If measurement errors are heteroskedastic, the degree of attenuation is proportional to the average measurement error variance. This means that all observations are treated equal, while some estimates are likely to be close to the true values and others are not. An observation with below average measurement error provides a stronger signal and the autoregressive parameter for this observation would ideally be attenuated less and vice versa. Hence, for the purpose of forecasting a constant parameter is suboptimal as the degree of attenuation is never in line with the degree of measurement error.

In this paper I propose dynamic models for realized covariances which explicitly take into account their measurement error. The models use the asymptotic distribution of the high-frequency estimates to allow for time-varying attenuation of the autoregressive parameters. The parameters are high on days where the covariance is estimated precisely and they are attenuated more heavily on days when the measurement error is large and the signal is weak. The models parsimoniously allow for different dynamics for each individual element of the covariance matrix. The concept was used in the simpler univariate setting by Bollerslev, Patton, and Quaedvlieg (2015), who found promising results. By studying the multivariate problem, I can consider the impact of dynamic attenuation in a number of economically relevant applications that are not possible in the univariate case.

I evaluate the models both statistically and economically, using a variety of out-of-sample methods. Firstly, I consider a standard statistical evaluation of the models, where I find that the models provide significantly better forecasts compared to their constant attenuation benchmarks. I show that the models work well because the autoregressive parameters are tailored for the information content in past estimates. They allow more responsive short-term dynamics when the signal is strong, and limit the propagation of potentially large estimation errors when the signal is weak. Using various loss functions I document improved forecast accuracy, and find that improvements are made both after precise and after noisy covariance estimates, demonstrating that dynamic attenuation offers benefits in both directions.

Since measurement error has a large impact on economic decision making, the dynamic attenuation models have widespread practical applicability. I provide evidence of their empirical qualities by using them for several financial applications. In the first application I use the models for portfolio allocation. The problem of measurement error has received significant attention in this literature, as it often leads to extreme positions. The extreme portfolio weights are rarely the optimal allocation based on population values, but the result of inaccurate and poorly conditioned, estimated covariance matrices. The spurious nature of the positions leads to portfolios that are under-diversified, due to large weights in single stocks, and unstable, as the spurious positions have to be unwound in the next period (e.g. Li, 2015). This leads to poor out-of-sample performance because of bad allocations and excessive turnover.

Two of the most popular techniques to reduce the impact of estimation error on portfolio allocation are to either 'shrink' the covariance matrix or to impose constraints on the portfolio weights. For shrinkage, the portfolio allocation is based on a linear combination of the covariance estimate and some target matrix that implies stable portfolios, such as the identity or an equicorrelation matrix (Ledoit and Wolf, 2003, 2004a). Another strategy is to constrain the portfolio weights directly, as the extreme positions are typically the result of estimation error, not of the population values. For instance, Jagannathan and Ma (2003) study no short-sale global minimum variance portfolios, and DeMiguel et al. (2009) and Brodie et al. (2009) propose L-norm constrained portfolios.

The time-varying parameter models proposed in this paper endogenously provide a dynamic alternative to the shrinkage estimators. Covariance forecasts are shrunk, with time-varying intensity, from the conditional to the unconditional covariance matrix, which leads to less extreme and more stable portfolio allocations. I find that the dynamic adjustment of the parameter leads to more accurate (i.e. lower ex-post portfolio volatility) and more stable forecasts, which reduces turnover by about 30% compared to their respective benchmark models. This result is economically significant. Using the Fleming et al. (2003) framework I find that a risk averse investor would be willing to sacrifice up to 150 basis points annually to switch to a dynamic attenuation model and, using realistic values, typically gains an additional 30 basis points due to reduced transaction costs. The improvements persist after applying other methods to counter estimation error, such as imposing weight constraints. Moreover, I find that despite the higher level of transactions costs and increased noise in the estimates, portfolios that are rebalanced daily using the models proposed in this paper outperform portfolios which are rebalanced weekly or monthly, which is not the case for the base models.

Similar to how Realized (Co)variances allow effective exploitation of the information inherent in high-frequency data, they also allow for direct reduced-form modeling of the otherwise latent quantities. Indeed, in multivariate applications the covariance is seldom the object of interest and an important derived quantity is the realized beta (Andersen et al., 2006; Hansen et al., 2014). In the second application of the models I consider the forecasting of risk exposures. The ex-post estimates of beta are similarly impacted by heteroskedastic measurement error. To obtain beta forecasts, I innovate by comparing forecasts of beta implied by the covariance matrix to beta forecasts obtained by directly modeling realized beta. I find that by taking into account the heteroskedasticity of the measurement error, betas are more persistent than previously thought. I consider various hedging applications to evaluate the forecasts. I find that modeling beta directly using realized betas outperforms those implied by covariance forecasts and more importantly, that explicitly taking into account the time variation in estimation uncertainty of realized beta offers additional improvements in the forecasts.

The remainder of the paper is organized as follows. Section 2 introduces notation and the various models. In Section 3 I report the result of statistical evaluation of covariance forecasts. Sections 4 and 5 report the results of simulations and empirical application on the performance of the models proposed in this paper for the purpose of portfolio selection and beta forecasting respectively. Section 6 concludes.

2. Dynamic Modeling of Realized Covariance with Heteroskedastic Measurement Error

In this section I introduce the models for forecasting the conditional covariance matrix of daily returns. I build on popular models in the literature and extend the models to take into account the measurement error in realized covariances. I propose to model the dynamics of the covariance matrix as autoregressive processes, where the coefficients are allowed to vary throughout time as a function of the degree of measurement error, of which we can obtain estimates using the asymptotic theory of the high-frequency estimators. The responsiveness to recent observations is reduced when their measurement error is large and vice-versa.

The intuition for the models follows from the errors-in-variables literature. The time-series of the estimated (co)variances are the sum of two components, the latent quantity and measurement error. Since the latter is just noise, the series appear to be less persistent than they truly are, and the autoregressive parameters are attenuated towards zero. When measurement error is homoskedastic the degree of attenuation is proportional to the measurement errors variance. However, if the measurement error is heteroskedastic, estimates will be attenuated based on the *average* measurement error variance. This results in attenuation which is never optimal. It is too strong when the past observation has little measurement error and too weak when it is estimated imprecisely. The dynamics of the estimates series can be more accurately modeled by taking into account the heteroskedastic nature of measurement error.

We know from the asymptotic theory of the realized covariance that its measurement error is in fact heteroskedastic, and moreover, we can feasibly estimate its magnitude every single day. I develop models that use these estimates to obtain daily varying parameters such that attenuation can be tailored specifically to each individual estimate. On some days there is sufficient information to allow for less attenuation, while on other days one might have to attenuate more dramatically than a constant parameter model would allow. By directly incorporating estimates of the asymptotic distribution, dynamic attenuation can be obtained in a robust, parsimonious, and easy to estimate manner.

In the following, I first develop some notation and provide intuition for the modeling choices I make, followed by the definition of the various models. As base models I consider three popular models in the literature, the vech-HAR (Chiriac and Voev, 2010), the Exponentially Weighted Moving Average (EWMA) (used in, amongst others, Fleming et al., 2003) and HEAVY (Noureldin, Shephard, and Sheppard, 2012) models. The former can easily be estimated using OLS and is potentially the most popular model in the literature. The latter two models are estimated using maximum likelihood techniques.[1]

The purpose of using multiple models is not to run a horse-race between the various models, but to evaluate the relative performance of the models with dynamically attenuated parameters versus the ones with constant parameters. It also serves to highlight the fact that the principle is widely applicable. The same ideas could be applied to any other model using high-frequency estimates or where estimates of the measurement error variance are available.

2.1. Notation and Definitions

I assume the following log-price process

$$P(s) = \int_0^s \mu(u)du + \int_0^s \sigma(u)dW(u), \tag{1}$$

where $W(u)$ is an N-dimensional vector of independent Brownian motions, with σ the volatility process such that $\Sigma(u) = \sigma(u)\sigma(u)'$ is the spot covariance matrix.[2] The objective is to model the ex-post covariation for day t, called the Integrated Covariance, which for price-processes of this type equals

$$\Sigma_t = \int_{t-1}^t \Sigma(u)du. \tag{2}$$

[1] Pakel et al. (2014) demonstrate that parameter estimates in large-dimensional covariance models may be severely biased with MLE estimation, they propose a technique known as composite likelihood, that significantly reduces the bias, and moreover greatly reduces computation time. For this paper I use their 2MSCLE technique. More details are provided in Appendix C.

[2] For simplicity I only consider a data generating process without jumps. The general principle behind the models in this paper readily extend to a price process that includes a jump-component.

This quantity is not directly observable, but can be consistently estimated with the use of high-frequency returns. Define $r_{i,t} = P_{t-1+i\Delta} - P_{t-1+(i-1)\Delta}$ as the i-th return vector of day t. Returns for individual series $n = 1, ..., N$, are denoted by $r_{i,t}^{(n)}$. We observe $M = 1/\Delta$ uniformly spaced, intra-daily returns for each series. The vector of daily returns is obtained as $r_t = \sum_{i=1}^{M} r_{i,t}$.

The daily integrated covariance matrix, Σ_t, can be consistently estimated, denoted S_t, using a wide class of estimators. Although realized variances can be very accurately estimated at the 5-minute frequency, realized covariances suffer from the Epps (1979) effect, where correlation estimates tend to decrease when sampling at higher frequencies due to asynchronous trading and observation of the underlying price process. A number of alternatives to realized covariance exist that are robust to this effect and I use the Multivariate Kernel (MKernel) of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) applied to 5-minute returns.[3,4]

Throughout the paper, I will vectorize the covariance matrix, and consider the $N^* = N(N+1)/2$ dimensional vector $\varsigma_t = \text{vech } \Sigma_t$. The MKernel is an estimator of Σ_t and is therefore subject to estimation error, $s_t = \varsigma_t + \eta_t$, with $s_t = \text{vech } S_t$. The asymptotic theory of the MKernel provides us with the distribution of the measurement error η_t, which is mixed-normal with $N^* \times N^*$ covariance matrix Π_t. The measurement error variance Π_t is proportional to a quantity called the Integrated Quarticity (IQ), defined as

$$IQ_t = \int_{t-1}^{t} \Sigma(u) \otimes \Sigma(u) du, \tag{3}$$

where $\otimes$ denotes the kronecker product. The asymptotic variance of the variance, or IQ, can similarly be consistently estimated based on high-frequency data. Appendix A provides estimation details of the MKernel and its asymptotic distribution.

2.2. Vector Autoregressive Models with Measurement Error

The effect of measurement error on autoregressive models is extensively studied in the univariate case (e.g. Staudenmayer and Buonaccorsi, 2005). In the multivariate autoregressive setting the problem is less well-developed and typically looks at the effect of measurement error on identification and hypothesis testing (e.g Holtz-Eakin, Newey, and Rosen, 1988; Komunjer and Ng, 2014). It is however well-known that the estimates of autoregressive parameters are attenuated towards zero, both in uni- and multivariate case. In this section I review some of the results relevant for the forecasting of realized covariances.

In order to provide intuition for my modeling choices in the next sections, I assume a simple model for the integrated covariance and analyze the effect of measurement error. Specifically, I

[3]One disadvantage of using the MKernel is that it is theoretically slightly biased. In practice, especially at this sampling frequency, this term is negligible, and I disregard it throughout the rest of the paper.

[4]Other robust choices include estimator of Hayashi and Yoshida (2005) and the pre-averaging estimator of Christensen, Kinnebrock, and Podolskij (2010), but I choose the MKernel as its asymptotic variance can easily be estimated.

assume that ς_t follows a VAR(1) model. Under this assumption the integrated covariance follows

$$\varsigma_t = \Phi_0 + \Phi_1 \varsigma_{t-1} + \epsilon_t, \tag{4}$$

with Φ_0 of dimension N^* and where Φ_1 is an $N^* \times N^*$ matrix. In practice one would have to estimate the model on estimated quantities $s_t = \varsigma_t + \eta_t$,[5]

$$s_t = \Theta_0 + \Theta_1 s_{t-1} + \epsilon_t. \tag{5}$$

Under the assumption that ϵ_t and η_t are i.i.d. such that measurement error is homoskedastic and uncorrelated over time, it is straightforward to derive that the OLS estimate relates to the population value Φ_1 in the following way

$$\begin{aligned} \Theta_1 &= (\mathbf{s}'\mathbf{s})^{-1}(\boldsymbol{\varsigma}'\boldsymbol{\varsigma})\Phi_1 \\ &= (\boldsymbol{\varsigma}'\boldsymbol{\varsigma} + \boldsymbol{\eta}'\boldsymbol{\eta})^{-1}(\boldsymbol{\varsigma}'\boldsymbol{\varsigma})\Phi_1, \end{aligned} \tag{6}$$

where $\mathbf{s} = (s_1, ..., s_{T-1})$ is the $T-1 \times N^*$ matrix of lagged covariances, with $\boldsymbol{\varsigma}$ and $\boldsymbol{\eta}$ defined analogously.[6] The estimated parameter is the population parameter times the ratio of variation of the latent process over the variation of the estimated process, which is often called the reliability ratio. This result shows that in the general VAR(1) case, the attenuation in the multivariate case is more subtle than the univariate case. Specifically, depending on the covariance structure of η, certain parameters may be biased towards zero, while other parameters may be biased upwards. In contrast, in the univariate setting, parameters are always biased towards zero. Regardless of the direction, the degree of bias of the parameter is proportional to the degree of measurement error.

In practice, since the dimension of covariances, and therefore the parameter matrix, increases rapidly with the number of assets, it is customary in the covariance modeling literature to consider scalar models where the autoregressive parameter is restricted to be a scalar. One typically estimates the full dimension of the constant or uses covariance targeting for further parameter reduction. The scalar model removes the subtleties of the measurement error and its parameter's attenuation is proportional to the average of all measurement errors.[7]

The bias as quantified by (6), could be removed using estimates of the measurement error covariance matrix $\eta'\eta$, by simply reverse engineering the equation to retrieve Φ_1. A similar result

[5] In the univariate case the population parameters of an AR(p) model with homoskedastic measurement error can be identified by estimating an ARMA(p,p) model. In the multivariate case the composite error term $\epsilon_t - \eta_t + \Theta_1 \eta_{t-1}$ typically fails to be a finite order moving average process, and a VARMA will not identify the population parameter.

[6] In the univariate case this relationship simplifies to the easier to interpret $\Theta_1 = \Phi_1(1 + Var(\eta_t)/Var(\varsigma_t))^{-1}$, where the estimated parameter is the population parameter times a noise-to-signal ratio.

[7] To see this, note that to estimate the scalar model using OLS amounts to stacking the columns of the $\mathbf{s}$ matrix, i.e. $\mathbf{s}^* = \text{vec}\,\mathbf{s}$. Hence, the scalar parameter has the same bias formula (6), where s, ς and η are replaced with their stacked counterparts, and $\boldsymbol{\eta}^{*\prime}\boldsymbol{\eta}^*$ is simply the average measurement error variance.

can be produced for the scalar model. Although this identifies the population parameter, which may often be of interest, this is not optimal for forecasting purposes, as attenuation serves a purpose by limiting the impact of noisy estimates.

When the measurement error is heteroskedastic, attenuation is proportional to the average measurement error variance. The attenuation serves the same purpose, but constant attenuation is not optimal. Equation (6) could similarly be used to obtain an optimally attenuated coefficient at each point in time by replacing the unconditional noise covariance matrix with an estimate of the conditional one. However, due to the difficulty of estimating the measurement error variance and the required inversion of the matrix, this will typically result in extremely poor and unstable parameter estimates.

In the following sections I propose parsimonious dynamic specifications for the high-frequency covariance estimates that combine the features of the general VAR and restricted scalar models to deal with heteroskedastic measurement error in a robust manner. The autoregressive parameters are linear in the measurement error covariance matrix of the past estimates. This enables the model to allow for complicated and general dynamics with a small amount of parameters, where each element of the covariance matrix has different dynamic autoregressive parameters based on the degree of measurement error of individual series on each individual day.

2.3. HAR Models

The widely popular Heterogeneous Autoregressive (HAR) model of Corsi (2009) was first extended to the multivariate setting in Chiriac and Voev (2010) who apply the HAR approach to the vech of the covariance matrix, defined as

$$s_t = \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t, \tag{7}$$

where $s_{t-h|t-1} = \frac{1}{h}\sum_{i=1}^{h} s_{t-i}$. θ_0 is a N^* dimensional vector and the other θs are scalar. One could allow for more general specifications, by letting the different θ coefficients be (possibly restricted) $N^* \times N^*$ matrices. However, the more general specifications greatly complicate the model, and moreover, their forecasts are not guaranteed to be positive-definite. By considering scalar coefficients, the forecasts are sums of positive definite matrices and as such positive definite.

One way to allow for a more general specification whilst ensuring positive definite forecasts is to model the variances and correlations separately, as proposed in Oh and Patton (2015). I refer to this version as the HAR-DRD, where DRD refers to the decomposition of S into variances and correlations

$$S_t = D_t R_t D_t, \tag{8}$$

with D_t a diagonal matrix of standard deviations and R_t the correlation matrix. Each variance can be modeled using a univariate HAR, and the correlation matrix $r_t = \text{vech } R_t$ can be modeled using the scalar HAR of Equation (7). I then recombine the forecasted correlation matrix with the forecasted variances to obtain the covariance forecasts. These forecasts are positive-definite as long as the forecasted variances are positive.

In order to exploit the heteroskedastic measurement error for forecasting purposes, Bollerslev, Patton, and Quaedvlieg (2015) propose a simple modification to the univariate HAR, which is Equation (7) with scalar s_t, by replacing the parameter for the daily lag by

$$\theta_{1,t} = (\theta_1 + \theta_{1Q} \Pi_t^{1/2}), \tag{9}$$

where Π_t is an estimate of twice the Integrated Quarticity, the asymptotic variance of the univariate Realized Variance. The specification can be seen as a linear approximation to the inverse of (6) in the univariate setting, which is more robust to the estimation error in Π_t. They call this model the HARQ, where the Q suffix refers to the Integrated Quarticity. One could also consider time-varying parameters for the weekly and monthly lag, but Bollerslev, Patton, and Quaedvlieg (2015) find that, due to the difficulty of estimating Π_t, these tend to add more noise than dynamic improvements, as the measurement error and its heteroskedasticity are significantly reduced when averaging over five or twenty-two days.

There are many potential ways of defining a multivariate analogue. Although measurement error in s_t is expected to be of greater importance in the multivariate setting, it is even more difficult to accurately estimate its distribution Π_t, which restricts the choice of specification significantly. It turns out that estimates of the full Π_t matrix are too noisy to successfully be incorporated in the models. Its estimation error leads to attenuation in the dynamics of the autoregressive parameter itself, which then essentially remains constant. Improvements can be made by restricting the way Π_t is incorporated in the dynamic specification.

One way too extract useful information from Π_t is to only use a subset of its elements. Similar to the covariance matrix s_t itself, the variance elements of Π_t are much more accurately estimated than the covariance elements. Hence, for robustness purposes, I choose to define $\pi_t = \sqrt{diag(\Pi_t)}$, i.e. the matrix that has the asymptotic standard deviation of each individual element of the covariance matrix on its diagonal.[8] I define the vech HARQ as

$$\begin{aligned} s_t &= \theta_0 + \theta_{1,t} \circ s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t \\ \theta_{1,t} &= \theta_1 \iota + \theta_{1Q} \pi_{t-1}, \end{aligned} \tag{10}$$

where $\circ$ denotes the Hadamard product and ι an N^* dimensional vector of ones. In the HARQ, as in the HAR, θ_1 and θ_{1Q} are scalar. However, as π_t is a vector, $\theta_{1,t}$ is a vector as well, and the different elements of s_t obtain different degrees of persistence. The $\theta_{1,t}$ parameter vector thus allows

[8]In practice I demean π_t to make the θ_1 coefficient interpretable as the value at average measurement error.

for different dynamics for each individual element of the covariance matrix based on their relative measurement error variance using only two estimated parameters, θ_1 and θ_{1Q}. One could easily allow for different and possibly more general dynamics by incorporating Π_t in a less restrictive way, but I found this specification of π_t to work best empirically, as it balances the information content on the heteroskedasticity with the estimation error in Π_t.

For the HARQ-DRD, I simply model the variances using the univariate HARQ of Bollerslev, Patton, and Quaedvlieg (2015) as described in Equation (9). I use the normal scalar HAR for the correlations. One could also consider a HARQ specification on the correlations, but I refrain from a dynamic specification for the parameter as the heteroskedasticity in the measurement error of the correlations is limited.[9,10] The potential benefits of the marginal variability in the parameters would be outweighed by the estimation noise in π_t.

The HAR(Q) and HAR(Q)-DRD have the main advantage that their estimation is very straightforward as the models can be easily estimated via OLS. For the HAR-DRD this is a 2-step estimation where the univariate models and the correlation model are separated.

2.4. EWMA Filters

Another popular method for forecasting covariance matrices which incorporates realized covariances is the Exponentially Weighted Moving Average (EWMA) used by, amongst others, Fleming et al. (2003). Contrary to the HAR model that directly models the realized covariance, the EWMA, and the HEAVY model discussed in the next subsection, forecast the covariance matrix of daily returns, by incorporating realized information.

They consider an exponentially weighted moving average representation for the daily covariance matrix $V_t = E(r_t r_t' \mid \mathcal{F}_{t-1})$, with $\mathcal{F}_{t-1}$ the information set available at time $t-1$, and incorporate Realized Covariance in their rolling estimator to improve forecasts. The specification they propose is

$$v_t = \exp(-\alpha) v_{t-1} + \alpha \exp(-\alpha) s_{t-1}, \tag{11}$$

with $v_t = \text{vech } V_t$. Here α captures the decay rate of the lags. When α is low the process is very persistent and when α is high more of the new information coming from s_{t-1} is incorporated in the fitted values for v_t. The model guarantees positive-definite forecasts, which is essential for many applications. I estimate the optimal decay rate under the assumption that daily returns are conditionally normal using composite maximum likelihood as described in Appendix C.

[9] Using Proposition 3 in Barndorff-Nielsen and Shephard (2004), and assuming constant spot volatility, the correlation coefficient between two assets has asymptotic standard deviation of approximately $1-\rho_t^2$. The heteroskedasticity of measurement error therefore only depends on the variability of the square of the correlation which is by definition limited compared to (co)variances, since $|\rho_t| \leq 1$.

[10] The asymptotic variance of the correlation matrix can be derived more generally using the delta method, i.e. $\tilde{\Pi}_t = \nabla \Pi_t \nabla'$, where $\nabla = \frac{\partial \text{vech} D^{-1} \Sigma D^{-1}}{\partial \text{vech} \Sigma}$.

Similar to the HARQ, I propose to let the parameter α vary with estimates of the uncertainty in s_t. Specifically, the EWMAQ is defined as

$$\begin{aligned} v_t &= \exp(-\alpha_t) \circ v_{t-1} + \alpha_t \circ \exp(-\alpha_t) \circ s_{t-1} \\ \alpha_t &= (\alpha\iota + \alpha_Q \pi_{t-1}). \end{aligned} \tag{12}$$

Typically, the α parameter in the EWMAQ will be higher than in the EWMA, allowing for more short term dynamics stemming from s_{t-1}, with α_Q negative such that the impact of highly uncertain estimates of the covariance are not incorporated in v_t. Instead, if s_t is estimated imprecisely, the model shifts weight towards the weighted average of past covariances v_{t-1}.

2.5. HEAVY Models

The HEAVY model was introduced in Noureldin et al. (2012). The HEAVY model is similar to the EWMA in that it models v_t by incorporating high-frequency covariance estimates, but is more richly parametrized. Similarly, the HEAVY model allows for short response times to changes in volatility by incorporating new information from the high-frequency estimates. Here I introduce the HEAVYQ, which again differentiates the response to the high-frequency estimate by incorporating the strength of the signal of s_t as measured by its asymptotic variance.

In this paper I consider the most parsimonious HEAVY specification, which is the scalar version with covariance targeting, but the principle can be applied to any HEAVY specification.

$$v_t = (I_{d^*} - B - A\kappa)\lambda_V + Bv_{t-1} + As_{t-1}, \tag{13}$$

where A and B are scalar. Denote by Λ_Σ the unconditional expectation of the high-frequency covariance matrix and by Λ_V the unconditional expectation of the low-frequency daily covariance matrix. $\kappa = L_N(\bar{\kappa} \otimes \bar{\kappa})D_N$, with $\bar{\kappa} = \Lambda_\Sigma^{1/2}\Lambda_V^{-1/2}$ and L_N and D_N the elimination and duplication matrices respectively (Magnus and Neudecker, 1980). The κ term is an adjustment to match the expectation of the high-frequency covariance with the daily covariance. Finally, $\lambda_V = \text{vech}\, \Lambda_V$. Again, the models are estimated using composite likelihood.

Like the previous two models, the HEAVYQ model simply allows the impact of s_{t-1} to vary over time with the degree of measurement error and is defined as

$$\begin{aligned} v_t &= (I_{d^*} - B - A_t\kappa)\lambda_V + Bv_{t-1} + A_t \circ s_{t-1} \\ A_t &= (A\iota + A_Q \pi_{t-1}), \end{aligned} \tag{14}$$

where again A_Q is scalar.

3. Forecasting the Covariance Matrix

In this section I evaluate the empirical qualities of the Q models using statistical loss functions. Section 3.2 provides in-sample estimates and fit, and Section 3.3 presents the results of an out-of-sample model evaluation.

3.1. Data

The empirical investigation focuses on the 27 Dow Jones constituents as of September 20, 2013 that traded continuously from the start to the end of the sample. I use 5-minute returns retrieved from the TAQ database over the period February 1993 until December 2013, for a total of 5,267 daily observations. For the purpose of constructing betas, I supplement the individual stocks with high-frequency counterparts of the four risk factors of Fama and French (1993) and Carhart (1997). The market is proxied by the S&P500 ETF, the SPY. The other three factors, Small-minus-Big (SMB), High-minus-Low (HML) and Momentum (MOM), are constructed as the difference between the returns of the highest and lowest decile portfolios of the full TAQ database.[11] These risk factors are used in the two empirical applications presented in Sections 4 and 5. In the former I use them to obtain portfolios to track the market index and in the latter I forecast the exposures to the four risk factors.

The analysis focuses on open-to-close covariance matrices, where the noisy overnight returns are not included.[12] This is often done in the high-frequency covariance literature, even in the context of portfolio allocation (Lunde, Shephard, and Sheppard, 2015; Hautsch, Kyj, and Malec, 2015; De Lira Salvatierra and Patton, 2015). Moreover, Liu (2009) finds that including the overnight return in covariance estimation leads to greater out-of-sample portfolio turnover. Omitting the overnight return is in line with Andersen, Bollerslev, Frederiksen, and Nielsen (2010) who treat overnight returns as deterministic jumps, which have a transitory effect in the dynamics of variances. Similarly, I measure the vector of daily returns using open-to-close returns.

I provide some descriptives of the data in Appendix B. In the remainder of this Section, as well as in Section 4, I analyze a subset of ten of the most liquid stocks to keep estimation feasible. In order to effectively model covariance matrices of higher dimension more structure needs to be

[11] As the construction of the high-frequency risk-factors slightly different from the standard method, their daily returns do not perfectly match the factors typically used in empirical work. However, the correlation between the high-frequency factors and the daily factors obtained from Kenneth French's website is between 70% and 90%, which is mainly due to the exclusion of the overnight return. When included the correlation increases to values between 90% and 95%.

[12] The single overnight return covers seventeen and a half hours compared to the five-minute returns during trading hours. As such, the estimate of the close-to-open variation is very noisy. Moreover, the estimation of the asymptotic variance for this part of the daily variation is equally poor. In order to incorporate the overnight return, one could use an auxiliary specification to predict the overnight variation and open-to-close variation separately. For the overnight variation, a Q model is unlikely to provide great benefits due to the noise in its asymptotic variance estimates. Another possibility to deal with the overnight return is to scale up the intraday variation to reflect the variation for the whole day, similar to the κ adjustment term in the HEAVY model.

applied to the model, which I leave for future research. The subset consists of stocks with the following tickers: AXP, BA, CVX, DD, GE, IBM, JPM, KO, MSFT and XOM. In Section 5, I separately model the covariance matrix of each of the 27 stocks with the four risk-factors, whose limited dimension remains feasible.

3.2. In-Sample Results

First consider the in-sample estimation results for the ten stock sub-sample. Parameters with robust standard errors are reported in Table 1 for each of the models. For the HAR(Q)-DRD I report the average of the parameters and standard errors over each of the individual variance specifications to save space. Finally, in the bottom two rows I report the Frobenius distance and QLIKE loss of the fitted values H_t with respect to the ex-post covariance S_t. The Frobenius distance is defined as

$$L_t^{Frobenius} = \sqrt{Tr(H_t - S_t)}, \tag{15}$$

and the quasi-likelihood (QLIKE) as

$$L_t^{QLIKE} = \log |H_t| + Tr(H_t^{-1} S_t), \tag{16}$$

which is the negative of the log-likelihood of a multivariate normal density. Both functions measure loss, and hence lower values are preferable.[13]

As expected, all the Q coefficients are negative and significant, which is consistent with the intuition that as the measurement error variance π_t increases, the informativeness of the past covariance estimate decreases, and the model endogenously increases the degree of attenuation. To put the magnitude of the parameters in perspective, the cross-sectional average of the composite parameter $\theta_{1,t}$ varies between roughly 0.21 and 0.87 for the HARQ model.

As in Bollerslev, Patton, and Quaedvlieg (2015) for the HAR there is a distinct redistribution of weight from the monthly lag to the daily lag. This effect is more pronounced in the HAR-DRD model, due to its increased flexibility in the time-variation of parameters. For the HARQ-DRD all individual variances have separate dynamics, and they can allow for individual sensitivities to the degree of measurement error. As such the average increase in θ_1 compared to the standard models is greater for the HARQ-DRD than for the HARQ, and similarly the θ_{1Q} parameter is greater in absolute value. This in itself is an attenuation story, as the measurement error in π_t aggregates in the cross-section for the HARQ, making the dynamic parameter less effective. Relative to the EWMA, the EWMAQ has a higher value of α which shows it allows for a greater impact of the MKernel estimates on the sample path of v_t. However, given the negative value of α_Q, when estimation error becomes greater the weight is shifted towards the long-run weighted average away from the

[13] Both loss functions offer consistent rankings of models, despite the use of an ex-post estimate as a proxy for the true covariance matrix when evaluating the forecasts (Patton, 2011; Laurent, Rombouts, and Violante, 2013).

Table 1: In-Sample Estimates

	HAR	HARQ		HAR-DRD	HARQ-DRD		EWMA	EWMAQ		HEAVY	HEAVYQ
				Average Variances							
θ_1	0.247	0.541	θ_1	0.260	0.660	α	0.098	0.118	A	0.106	0.148
	(0.040)	(0.040)		(0.064)	(0.065)		(0.005)	(0.005)		(0.009)	(0.009)
θ_2	0.410	0.333	θ_2	0.395	0.223	α_Q		-0.007	B	0.876	0.825
	(0.038)	(0.038)		(0.016)	(0.022)			(0.003)		(0.004)	(0.004)
θ_3	0.244	0.113	θ_3	0.240	0.101				A_Q		-0.026
	(0.038)	(0.038)		(0.028)	(0.028)						(0.012)
θ_{1Q}		-0.043	θ_{1Q}		-0.067						
		(0.018)			(0.026)						
				Correlation							
			θ_1	0.049	0.049						
				(0.007)	(0.007)						
			θ_2	0.159	0.159						
				(0.003)	(0.003)						
			θ_3	0.560	0.560						
				(0.003)	(0.003)						
Frob.	12.282	12.190		11.987	11.952		12.295	12.242		12.539	12.532
QLIKE	15.484	15.392		15.289	15.205		15.439	15.441		15.517	15.505

Note: The table provides in-sample parameter estimates and measures of fit for the various benchmark and Q models. For the HAR-DRD the reported parameter estimates and standard errors for the univariate variance specifications are the average across the ten assets. The bottom panel shows the in-sample fit of the various models as measured by the Frobenius distance and QLIKE.

noisy estimate. The table displays a similar picture for the HEAVYQ, where the A parameter of the HEAVYQ is higher than that of the HEAVY model, and high estimation uncertainty in s_{t-1} decreases the responsiveness to the high-frequency estimate in favor of the long term weighted average v_{t-1}.

3.3. Out-of-Sample Results

In this section I consider out-of-sample forecasts of the covariance matrix derived from the various models. I focus on one-day-ahead forecasts of the ten-dimensional covariance matrix. For the forecasts, the parameters of the different models are re-estimated each day using a rolling window sample of 1,000 days.[14] The forecasts are evaluated based on the Frobenius distance and QLIKE defined in Equations (15) and (16) respectively, where the fitted value, H_t, is replaced with the forecasts based on the $t-1$ information set, denoted $H_{t|t-1}$.

Table 2 presents the forecasting results for the different models. The Q models produce improvements on their standard models for both loss functions. Unreported results show that both

[14]Similar to the univariate case in Bollerslev, Patton, and Quaedvlieg (2015), due to the sometimes erratic behavior of the difficult to estimate π_t, the HARQ and HARQ-DRD may sometimes produce negative definite forecasts. To ensure positive-definite forecasts, I apply the same insanity filter as Bollerslev, Patton, and Quaedvlieg (2015) and replace the entire covariance forecast with the time-series average of the estimation sample if the forecast is non positive-semidefinite. This happens for one or two forecasts per series throughout the empirical section.

Table 2: Out-of-Sample Forecast Results

	Full Sample		Lower 95% $\|\|\pi_t\|\|$		Upper 5% $\|\|\pi_t\|\|$	
	Frob.	QLIKE	Frob.	QLIKE	Frob.	QLIKE
HAR	12.305	14.382	9.274	13.275	**69.723**	32.651
HARQ	**12.107***	14.159*	9.161*	13.206*	**69.703**	**32.190***
HAR-DRD	**12.134**	14.140	**8.940**	**13.022**	70.467	32.603
HARQ-DRD	**11.976***	**13.896***	**8.888***	**12.990***	**70.056***	**31.050***
EWMA	12.379	14.105	9.303	13.109	71.104	32.686
EWMAQ	**12.178***	**14.091***	9.172*	13.122	70.633*	32.643*
HEAVY	12.473	**14.051**	9.299	13.041	71.789	32.263
HEAYVQ	**12.161***	**14.004***	**9.093***	13.050	**70.170***	32.258

Note: This table provides the loss of out-of-sample forecasts for the various models. The full-sample results are split up in days where the measurement error variance was low, where $||\pi_t||$ was in the lower 95%, and when estimation error was severe, the upper 5%. Entries in bold depict the models that are part of the 90% Model Confidence Set for each column. Q models that significantly improve relative to their non-Q benchmark are indicated by an asterisk.

the standard and Q models forecast the right covariance matrix on average, with average bias of each element close to zero, but the Q models make great improvements in terms of the variance of the forecast error, which is significantly reduced.

To better understand whether most of the improvements are coming from the increased short-term dynamics when the estimates are precise or the increased attenuation when the estimates are poorly estimated, I split the results up in the 95% of days with lowest measurement error variance and the 5% highest. Since the asymptotic variance is a large-dimensional matrix, I summarize the total uncertainty in s_t by means of the Frobenius norm of π_t. For all models, the forecasts are more accurate both when the high-frequency estimates are precise and when they are imprecise. However, the largest percentage improvements appear to stem from the increased responsiveness when the estimates are precise.

To test whether the quality of the forecasts differ significantly across models, I apply the Model Confidence Set (MCS) of Hansen et al. (2011). This procedure determines the (sub)set of models that contains the best forecasting model with 90% confidence. For both the full series, as well as the two uncertainty-split samples, I compare each Q model with its basemodel separately (significance indicated by *), as well as all eight forecasts jointly (members of the set indicated in bold). For the former, the MCS procedure finds that each of the Q models significantly outperforms their respective benchmark in the full sample based on either loss function. For the two sub-sample split based on the degree of uncertainty, the Q models fail to significantly improve for small set of models. Overall, this provides significant evidence that the dynamic attenuation models produce better forecasts.

The results of the MCS applied to all eight models simultaneously are reported in Table 2. The models that are part of the MCS are highlighted in bold. The only model that is consistently in the confidence set is the HARQ-DRD. In the full sample, the EWMAQ and HEAVYQ are also in the

MCS based on either loss-function. The HARQ, on the other hand, appears to work particularly well when there is high uncertainty in covariance estimates. Overall, the results of the eight-model MCS show that the Q models jointly outperform the non-Q models from a statistical point of view. In the next sections I evaluate their relative merit from a practical point of view by using them in a dynamic portfolio allocation setting and for beta forecasting.

4. Global Minimum Variance Portfolios

In this section the empirical qualities of the various models are evaluated by using their forecasts to obtain Global Minimum Variance (GMV) portfolios. The GMV portfolio is a popular tool to evaluate the accuracy of covariance forecasts as its optimal weights are a function of the covariance matrix only. Any other portfolio application would also depend on the estimation error of the mean. Moreover, Jagannathan and Ma (2003) and DeMiguel et al. (2009) demonstrate that mean-variance efficient portfolios do not perform as well as the global minimum variance portfolios in terms of out-of-sample Sharpe ratios, exactly because of the noisiness of the mean estimates. There is a large literature based on the performance of various methods for obtaining the covariance matrix for portfolio allocation using daily data (Chan et al., 1999; Jagannathan and Ma, 2003; Fan et al., 2008; Brodie et al., 2009), and more recently the benefits of using high-frequency data for portfolio allocations have been extensively researched (Fleming et al., 2003; Bandi et al., 2008; Pooter et al., 2008; Liu, 2009; Varneskov and Voev, 2013; Hautsch et al., 2015). Without exception, the papers find that integrating high-frequency data leads to superior out-of-sample portfolio performance. However, these papers typically focus on estimation considerations, such as which sampling frequency or estimator to use. Instead, in this application I take the estimate as given, and compare different forecasting methods.

Implementation of the GMV portfolio remains difficult despite all the econometric advances and data availability. Covariances vary over time, and are subject to significant forecasting and measurement error. The former leads to the optimal portfolio allocation shifting over time, and the latter often leads to extreme allocations. Indeed, turnover in portfolios arises from two sources. Part of turnover is fundamental and induced by the time-variation in covariances, and part of this turnover is spurious, induced by forecast and measurement error.

To counter the effect of estimation error, a technique called *shrinkage* is often applied. Optimal allocations are based on a convex combination of the estimated, or forecasted, covariance matrix, and some target matrix. The target is often taken to be an equicorrelation matrix or some factor-model based matrix (Ledoit and Wolf, 2003, 2004b). The degree to which they are shrunk is typically based on some optimality result that shows that the shrinkage should be stronger when the estimation uncertainty is greater. The Q models directly incorporate this same intuition and endogenously offer an empirically more realistic *conditional* shrinkage, making them ideally suited towards the goal of portfolio allocation.

Indeed, by allowing for the time-variation in parameters as a function of measurement error, the models produce better forecasts when the signal is strong, and endogenously shrink the forecasts towards a long-term average when the signal is weak. As such, not only do they reduce the spurious turnover, but they also more accurately model the fundamental turnover when possible. To illustrate this I conduct a small simulation where I can distinguish between fundamental and spurious turnover, and an empirical application where I show the significant improvements in out-of-sample performance of the Q models.

4.1. Setup

I consider a risk-averse investor who allocates funds into N risky assets and uses conditional mean-variance analysis to make daily allocation decisions. Since the models described in this paper mainly improve on their basemodels by modeling the short-term dynamics, the merits of the Q model are better highlighted on the daily horizon than on the slower weekly or monthly horizons. Moreover, daily rebalancing is an often studied and important problem, and it is the ideal frequency for volatility-timing strategies (e.g. Fleming et al., 2003; Fan et al., 2012).

The portfolio is rebalanced using forecasts of the daily covariance matrix. To minimize conditional volatility the investor solves the global minimum variance portfolio:

$$\begin{aligned} w_t &= \arg\min\, w_t' H_{t|t-1} w_t \\ s.t. &\quad w_t'\iota = 1, \end{aligned} \tag{17}$$

where ι is a vector of ones of appropriate size. The solution to this problem is the weight vector

$$w_t = \frac{H_{t|t-1}^{-1}\iota}{\iota' H_{t|t-1}^{-1}\iota}. \tag{18}$$

whose elements are denoted by $w_t^{(n)}$, which is the allocation to asset n.

One feature of the Q models is that they should lead to more stable forecasts. Hence, of particular interest for the forecasting models developed in this paper is the distinction between fundamental and spurious turnover, and the effect on portfolio performance net of transaction costs. To evaluate the impact of the turnover implied by the different models I assume a fixed transaction cost c on each traded dollar for any stock. Total transaction costs are proportional to the turnover rates of the portfolios.

I use a standard set-up for transaction costs due to turnover, used in amongst others Han (2006), Liu (2009) and DeMiguel et al. (2014). At day t, the investor has portfolio weights $w_t^{(n)}$ in stock n. Before re-balancing, the return on stock n has pushed the size of the investment in stock n to $w_t^{(n)}(1 + r_t^{(n)})$, and the total value of the portfolio has increased by a factor $1 + w_t' r_t$. As such, the actual weight of stock n at the end of the day equals $w_t^{(n)} \frac{1+r_t^{(n)}}{1+w_t' r_t}$.

Taking into account the shifts in portfolio composition on the relative portfolio weights, the total turnover of the portfolio is equal to

$$TO_t = \sum_{n}^{N} \left| w_{t+1}^{(n)} - w_t^{(n)} \frac{1 + r_t^{(n)}}{1 + w_t' r_t} \right|. \tag{19}$$

Total transactions are assumed to be proportional to the turnover, and defined as cTO_t, where c is a transaction cost on each traded dollar. Portfolio excess returns net of transaction cost are given by $r_{pt} = w_t' r_t - cTO_t$. I consider values for c ranging from 0 to 2% chosen to cover reasonable transaction cost levels for the recent sample, in line with amongst others Fleming et al. (2003) and Brown and Smith (2011).

The other main quality of the models is that they do not propagate highly uncertain covariances which likely have large estimation error. Since estimation error leads to extreme portfolio allocations, particularly if the matrix is ill-conditioned, I also report details on the composition of the portfolios. To summarize these features, I report the portfolio concentration

$$PC_t = \left(\sum_n w_t^{(n)2} \right)^{1/2} \tag{20}$$

and the total portfolio short position

$$PS_t = \sum_n w_t^{(n)} \mathbb{I}_{\{w_t^{(n)} < 0\}}. \tag{21}$$

Finally, in order to evaluate the economic significance of the forecasting models, I use the framework of Fleming et al. (2001, 2003), who consider an investor with quadratic utility, who places a fixed amount of wealth in the set of assets each day. The realized daily utility generated by the portfolio for model k,

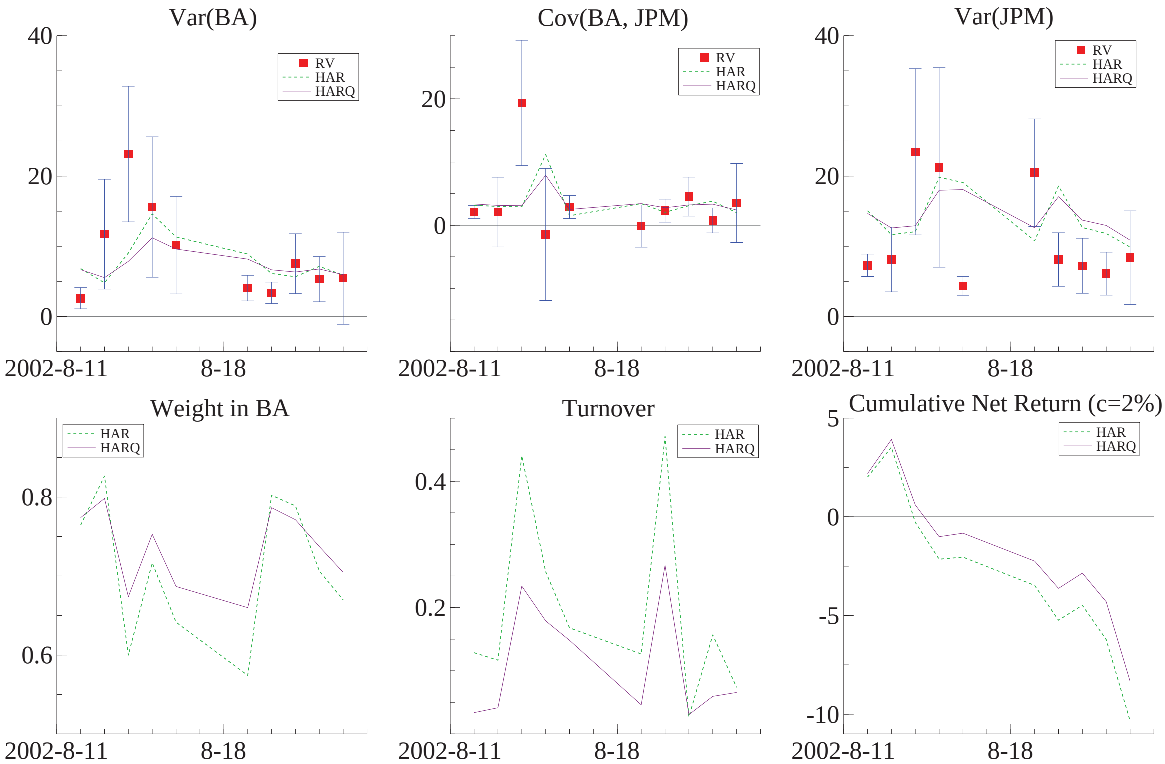
$$U(r_{pt}^k, \gamma) = (1 + r_{pt}^k) - \frac{\gamma}{2(1+\gamma)} (1 + r_{pt}^k)^2.$$

To determine the economic value of switching from one model to another I solve for Δ_γ in

$$\sum_{t=1}^{T} U(r_{pt}^k, \gamma) = \sum_{t=1}^{T} U(r_{pt}^l - \Delta_\gamma, \gamma), \tag{22}$$

where Δ_γ can be interpreted as the return an investor with risk aversion γ would be willing to sacrifice to switch from model k to model l.

Figure 1: Dynamic Shrinkage



Note: The top row of panels shows the estimated (co)variance for BA and JPM, along with the forecasted values of the HAR and HARQ model. The bottom left panel shows the weights in the first asset for the HAR and HARQ model, the middle panel shows the turnover, and the right panel the cumulative portfolio return net of transaction costs over the two weeks.

4.2. HARQ for portfolio allocation

First I provide some intuition on how the models work by means of an illustrative example using the HAR and HARQ. I consider a simple bivariate example (using BA and JPM) to better visualise the dynamics of the models. I plot the (co)variance and portfolio dynamics of both models over a two week period during the dot-com bubble in Figure 1. The top row shows ex-post estimated (co)variances along with 95% confidence bounds. In the same graphs, I show the ex-ante forecasts for the same dates based on the $t-1$ information set. What is evident from the top row is that, in line with intuition, the HARQ is less responsive to new information when estimation uncertainty in those estimates is large. As a result the forecasts are more stable as large uncertainty is often accompanied by transitory spikes in volatility. This leads to a more stable portfolio allocation, as evidenced by two of the bottom graphs which display the portfolio weight in one of the assets, and the implied turnover. The allocations of the HAR and HARQ do not differ substantially, as both models on average have about a 70% allocation to BA, but the variability around this weight is much lower for the HARQ, which leads to uniformly lower turnover. The reduction in transaction costs is significant, as shown by the bottom right graph, which displays the cumulative net return over the two weeks based on $c = 2\%$ transaction costs. The daily return on the HARQ portfolio does not differ much from those of the HAR, as the portfolio composition is similar, but the reduced turnover ensures that the daily net return on the HARQ portfolio is higher every single day. Over the two week period this adds up to a significantly higher return on the HARQ portfolio. While the daily improvements may appear to be small, the empirical analyses will show that they amount to up to 70 basis points annually.

4.3. Simulation Results: Fundamental and Spurious Turnover

In the previous sections I argued that the Q models are ideally suited to the purpose of portfolio management. The better forecasts of the HARQ should lead to portfolio allocations closer to the fundamental weights implied by the true covariance matrix and hence lower ex-post portfolio volatility. Second, the more stable forecasts that the Q models obtain by reducing the impact of noisy estimates should lead to a more stable portfolio allocation and a reduction in turnover.

To analyze these predictions, this section presents the results form a small simulation study. I consider an out-of-sample conditional portfolio allocation, where I make one-day-ahead predictions of the covariance matrix using the HAR and HARQ. I then compare the estimated GMV portfolio with the fundamental one based on the population value in terms of the portfolio composition, ex-post variance and degree of turnover.

In order to simulate realistic sample paths of the covariance, for each iteration I sample a random block of 2,000 consecutive estimated covariance matrices from five random series of the dataset outlined in the previous section. On each day I simulate returns at one-second frequency with a diurnal volatility pattern (Andersen, Dobrev, and Schaumburg, 2012), such that the Integrate Covariance of the simulated data matches that of the empirical estimate.[15] I then aggregate the paths to 1, 5 and 15-minute returns ($M = 390, 78, 26$), and estimate the MKernel and its asymptotic distribution. The different sampling frequencies allow me to attribute the gains to the HARQ's ability to better cope with the heteroskedastic measurement error, as the overall level of measurement error is decreasing in sampling frequency. To set a base-level of performance, I also consider HAR forecasts using the true population covariances ($M = \infty$), which I will denote HAR_∞. In this case there is no measurement error, only forecast error. By comparing the results based on estimated quantities we can directly infer the effect of estimation error on portfolio allocation.

As in the previous section I use a rolling window of 1,000 observations to obtain one-step-ahead forecasts. Each day, I obtain the GMV portfolio based on the forecasts and analyze their

[15] Although the simulated process is clearly oversimplified (i.e. no stochastic volatility), the diurnal pattern ensures that the asymptotic variance of the estimates is not simply their square, as would be the case if volatility were constant.

properties over the out-of-sample period. I consider the portfolio turnover as in (19) and the average portfolio standard deviation based on the population covariance $\sqrt{\hat{w}_t \Sigma_t \hat{w}_t'}$. To compare the estimated portfolio composition with the fundamental one, I report the average distance of the portfolio weights with respect to those obtained from the population value $\sqrt{\sum_n \left(\hat{w}_t^{(n)} - w_t^{(n)}\right)^2}$. Finally, to demonstrate that the forecasts of the HARQ are truly shrunk based on the degree of measurement error, I also show the distance of the portfolio weights to those implied by the HAR_∞. Results are based on 1,000 simulations.

The results of the simulation are reported in Table 3. First consider the results for the HAR model. The column marked ∞ directly models the latent population covariances, while the other columns use estimates based on M intraday observations. Note that even if the true covariance is observed, there will still be forecast errors, but it sets a baseline for the HAR's performance with estimated covariances. As expected, the estimation error in the MKernel has large consequences on the quality of the forecasts. When estimates are based on just 26 observations, the portfolio performance is significantly reduced compared to the HAR_∞ as evidenced by the higher portfolio standard deviation and distance to the fundamental weights. Importantly, the estimation error also induces much more variability in the portfolio allocation, as turnover increases by over 60%. When using more observations, the estimates become more precise and the portfolio allocations converge to the fundamental weights and the ones of HAR using population quantities.

In contrast, the HARQ obtains much better performance using small estimation samples. Compared to the HAR, the HARQ reduces the spurious turnover induced by estimation error by over 50% when $M = 26$, and even more than 60% when $M = 78$ or $M = 390$. The greater improvements relative to the HAR for larger M are due to the fact that the degree of measurement error also needs to be estimated, and its estimate similarly becomes better when M increases. In addition to the reduced excess turnover, the HARQ portfolio obtains lower ex-post portfolio variance than the HAR, and portfolio weights that are significantly closer to the fundamental and HAR_∞ weights. Moreover, as the final row demonstrates, the difference in HARQ and HAR covariance forecasts is truly driven by the measurement error, as the optimal weights are shrunk in the direction of HAR forecasts based on population quantities, and not some other more accurate forecast.

4.4. Empirical Results: Minimum Variance and Tracking Error Portfolios

In this section I evaluate the out-of-sample performance of the various models in the context of a portfolio selection problem. Of particular interest is whether the Q models improve allocation with lower portfolio variance and reduced turnover. I consider the standard GMV problem, as well as a tracking error portfolio, where the SPY is tracked. Constructing the minimum tracking-error variance portfolio is equivalent to constructing the GMV portfolio using returns in excess of the SPY. The minimum tracking portfolio is of interest for at least two reasons. First, it is empirically relevant as many institutional investors are evaluated based on their performance relative to a

Table 3: Simulation Results: Fundamental and Spurious Turnover

	HAR				HARQ		
M	∞	26	78	390	26	78	390
Turnover	0.1878	0.2999	0.2830	0.2379	0.2376	0.2257	0.2071
St. Dev	0.1829	0.1920	0.1879	0.1844	0.1887	0.1853	0.1837
Distance to population	0.0446	0.0601	0.0531	0.0471	0.0527	0.0488	0.0457
Distance to HAR_∞		0.0231	0.0198	0.0104	0.0141	0.0102	0.0062

Note: The table provides the results of a simulation study of out-of-sample portfolio allocation. Forecasts are based on MKernel estimates based on different numbers of observations M using the HAR and HARQ model. The $M = \infty$ column provides the result of the portfolio allocation using the HAR model applied to population values of the covariance. Turnover is as in (19), St. Dev is the portfolio standard deviation based on the population covariance, and Distance reports the distance between the portfolio allocations $\sqrt{\sum_n \left(\hat{w}_t^{(n)} - w_t^{(n)}\right)^2}$.

benchmark and full replication of the benchmark is often not desired or practical. Second, the covariance matrices of returns, net of a benchmark, are typically more dispersed as assets correlate to a different degree to the benchmark. This makes the minimum tracking error portfolio more challenging and prone to extreme weights.

As in Section 3, I obtain one-step-ahead forecasts using a rolling window of observations, re-estimating the parameters every day. The portfolios are re-balanced daily and evaluated according to their out-of-sample performance. To statistically compare the different models I use the MCS procedure of Hansen et al. (2011) to obtain the set of models with minimum ex-post portfolio variance. Additionally, I use the utility based framework of Fleming et al. (2003) to evaluate the economic gains of switching from the standard to the Q model and the economic impact of turnover through transaction costs.

4.4.1. Portfolio Allocation Results

Table 4 compares the out-of-sample performance of the GMV and minimum tracking error portfolio for the various models. Along with the dynamic models I consider a standard random walk forecast and the 1/N portfolio. The latter has been shown to be empirically hard to beat out-of-sample in terms of Sharpe ratio (DeMiguel et al., 2009).

Panel A shows the results of the GMV portfolios. Amongst all models considered the HEAVYQ has the lowest out-of-sample portfolio variance. However, based on the MCS results in bold, its portfolio variance is statistically indistinguishable from that of all other Q models and the standard HAR. Moreover, the out-of-sample Sharpe Ratio (without transaction cost) of all models, including the non-Q models, is higher than that of the 1/N portfolio. For most models, this result remains when accounting for differences in turnover with all reasonable values of c. This finding is at odds with DeMiguel et al. (2009), who report that strategies based on covariance forecasts cannot consistently outperform a naive diversification strategy. However, they consider unconditional Sharpe ratios, whereas I focus on conditional portfolio volatility.

Table 4: Unconstrained Portfolio Results

		HAR	HARQ	HAR-DRD	HARQ-DRD	EWMA	EWMAQ	HEAVY	HEAVYQ	RW	1/N
					Panel A: Global Minimum Variance						
TO		0.522	0.385	0.391	0.339	0.135	0.096	0.172	0.122	2.506	0.009
CO		0.513	0.517	0.487	0.497	0.505	0.506	0.497	0.497	0.868	0.316
SP		-0.105	-0.109	-0.070	-0.082	-0.095	-0.096	-0.089	-0.089	-0.547	0.000
Mean Ret		3.176	3.274	3.726	4.270	3.565	4.105	3.969	4.354	3.741	1.039
Std Ret		**15.147**	**15.020**	15.253	**15.060**	15.208	**14.911**	15.157	**14.883**	19.158	18.871
$c = 0\%$	Sharpe	0.210	0.218	0.244	0.283	0.234	0.275	0.262	0.293	0.195	0.055
	$\varDelta_1$		0.113		1.225		0.567		0.413		
	$\varDelta_{10}$		0.280		1.481		0.958		0.772		
$c = 1\%$	Sharpe	0.120	0.151	0.178	0.291	0.211	0.259	0.232	0.271	-0.145	0.054
	$\varDelta_1$		0.459		1.355		0.664		0.540		
	$\varDelta_{10}$		0.626		1.611		1.055		0.899		
$c = 2\%$	Sharpe	0.031	0.085	0.111	0.233	0.188	0.242	0.203	0.250	-0.485	0.053
	$\varDelta_1$		0.804		1.485		0.762		0.667		
	$\varDelta_{10}$		0.972		1.742		1.153		1.026		
					Panel B: Minimum Tracking Error Variance						
TO		0.173	0.102	0.134	0.114	0.063	0.045	0.080	0.057	1.128	0.009
CO		0.339	0.338	0.339	0.342	0.340	0.340	0.338	0.338	0.461	0.316
SP		-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.087	0.000
Mean TE		1.123	1.273	1.069	1.215	0.990	1.010	0.799	0.811	0.560	0.949
Std TE		7.119	**6.665**	7.086	**6.738**	7.107	**6.735**	7.122	**6.706**	9.268	7.616
$c = 0\%$	$\varDelta_1$		0.880		0.276		0.797		0.829		
	$\varDelta_{10}$		1.153		0.485		1.022		1.079		
$c = 1\%$	$\varDelta_1$		1.057		0.325		0.843		0.889		
	$\varDelta_{10}$		1.330		0.535		1.068		1.139		
$c = 2\%$	$\varDelta_1$		1.234		0.375		0.889		0.949		
	$\varDelta_{10}$		1.508		0.585		1.113		1.199		

Note: The top panel shows the portfolio results for the Global Minimum Variance Portfolio and the bottom panel shows the Minimum Tracking Error results under the constraint of no short sales. Each panel shows turnover (TO), portfolio concentration (CO), the average short position (SP), as well as the average annualized return (tracking error) and its volatility. Standard deviations in bold depict the models that belong to the 90% Model Confidence Set of lowest ex-post daily volatility. For both portfolios I show the economic gains of switching from the standard model to the Q-model in annual percentage points, $\varDelta_\gamma$, for various transaction cost levels c and risk aversion coefficients γ, as outlined in Equation (22).

The variability between turnover of the different base models is large. Interestingly, the increased turnover of, for instance, the HAR does not consistently lead to better Sharpe ratios or lower portfolio variance than the models that imply less turnover. On the other hand, all Q models obtain a reduction in variance, an increase in average returns and in Sharpe ratio in conjunction with a decrease in turnover relative to their respective benchmarks.[16] Turnover is reduced by a minimum of 14% for the HARQ-DRD up to almost 30% for the HEAVYQ. On the downside, and contrary to expectation, the Q portfolios appear to be slightly more concentrated and ask for

[16]This finding is in line with the low volatility anomaly documented in amongst others, Chan et al. (1999), Jagannathan and Ma (2003) and Baker et al. (2011).

greater short-positions, although the effect is economically small.

The bottom of Panel A shows the economic gains of switching to the Q models compared to their standard counterparts using the framework of Equation (22). When $c = 0\%$, there are no transaction costs, and the gains of switching are limited from about 0.1% up to 1.5% annually depending on the model. However, as all models have less turnover, the higher the transaction costs the greater the economic gains of switching. Depending on the specification, going from $c = 0\%$ to $c = 2\%$, increases the economic gains between 0.3 and 0.7% annually, purely due to decreased turnover. Under different assumptions on c, the economic gains can simply be extrapolated.

Panel B of Table 4 contains the results of the minimum tracking error portfolios. The results are similar to those of the Global Minimum Variance portfolio. The standard deviation of the tracking error is significantly reduced by the Q models, which jointly form the model confidence set of minimum variance models. Again, turnover is significantly reduced by the Q models with similar magnitude as in the GMV portfolios. The mean tracking error is slightly further from zero for the Q models than for the non-Q models. However, as the average return is positive this is not of concern, and actually shows that the models obtain higher returns in excess of the benchmark while obtaining lower volatility. As such, the models similarly offer economic gains as reported in the bottom part of Panel B. The gains are consistently positive and amount to between 0.3% and 1.5% annually.

4.4.2. No Shortselling

As a robustness check, I consider a no short-sale portfolio (e.g. Frost and Savarino, 1986). This is an empirically relevant restriction, as unlimited short positions are practically infeasible. Moreover, as shown by Jagannathan and Ma (2003), imposing short-selling constrains may enhance portfolio performance by limiting the impact of estimation error. In fact, portfolio restrictions have been shown to be equivalent to shrinkage of the covariance (Jagannathan and Ma, 2003). Given that imposing this constraint also reduces the impact of forecast and estimation error it is expected to reduce the gains of the Q models.

For the no-short sell minimum variance portfolio, there is the additional restriction that each position $w_t^{(n)}$ be non-negative.

$$\begin{aligned} w_t = \arg\min\ & w_t' H_{t|t-1} w_t \\ s.t.\quad & w_t'\iota = 1, \\ & w_t^{(n)} \geq 0, \forall n, \end{aligned} \tag{23}$$

which can be solved numerically using standard quadratic programming tools.

The results for the GMV and minimum tracking error portfolios under a no short-sale constraint are reported in Table 5. They show that the Q models still improve on the non-Q models, even after doing one of the standard adjustments to deal with measurement error. Indeed, the gains of the Q models with respect to their standard counterparts are reduced compared to the unconstrained

Table 5: No Short-sale Portfolio Results

		HAR	HARQ	HAR-DRD	HARQ-DRD	EWMA	EWMAQ	HEAVY	HEAVYQ	RW	1/N
						Panel A: Global Minimum Variance					
TO		0.390	0.284	0.322	0.280	0.101	0.071	0.131	0.093	1.139	0.009
CO		0.477	0.479	0.466	0.471	0.475	0.475	0.469	0.469	0.572	0.316
Mean Ret		3.874	4.210	4.046	4.271	4.050	4.632	4.363	4.967	2.800	1.039
Std Ret		15.607	15.274	15.537	15.260	15.500	**15.108**	15.489	**15.075**	16.837	18.871
$c = 0\%$	Sharpe	0.248	0.276	0.260	0.279	0.261	0.307	0.282	0.330	0.166	0.055
	Δ_1		0.375		0.98		0.623		0.647		
	Δ_{10}		0.825		1.301		1.147		1.200		
$c = 1\%$	Sharpe	0.183	0.227	0.207	0.298	0.244	0.294	0.260	0.314	-0.010	0.054
	Δ_1		0.642		1.035		0.697		0.744		
	Δ_{10}		1.091		1.407		1.221		1.297		
$c = 2\%$	Sharpe	0.118	0.179	0.153	0.250	0.227	0.282	0.238	0.298	-0.185	0.053
	Δ_1		0.908		1.141		0.771		0.840		
	Δ_{10}		1.358		1.514		1.295		1.393		
						Panel B: Minimum Tracking Error Variance					
TO		0.172	0.102	0.134	0.114	0.063	0.045	0.080	0.056	0.884	0.009
CO		0.339	0.337	0.339	0.342	0.340	0.340	0.338	0.338	0.416	0.316
Mean TE		1.118	1.269	1.068	1.214	0.990	1.008	0.794	0.810	0.542	0.949
Std TE		7.120	**6.667**	7.087	**6.742**	7.110	**6.735**	7.123	**6.708**	9.311	7.616
$c = 0\%$	Δ_1		0.879		0.272		0.792		0.826		
	Δ_{10}		1.150		0.482		1.021		1.077		
$c = 1\%$	Δ_1		1.052		0.320		0.842		0.886		
	Δ_{10}		1.325		0.533		1.067		1.134		
$c = 2\%$	Δ_1		1.230		0.372		0.887		0.945		
	Δ_{10}		1.507		0.584		1.110		1.195		

Note: The top panel shows the portfolio results for the Global Minimum Variance Portfolio and the bottom panel shows the Minimum Tracking Error results under the constraint of no short sales. Each panel shows turnover (TO), portfolio concentration (CO), as well as the average annualized return (tracking error) and its volatility. Standard deviations in bold depict the models that belong to the 90% Model Confidence Set of lowest ex-post daily volatility. For both portfolios I show the economic gains of switching from the standard model to the Q-model in annual percentage points, Δ_γ, for various transaction cost levels c and risk aversion coefficients γ, as outlined in Equation (22).

portfolios. For the GMV portfolio only the EWMAQ and HEAVYQ are in the set of lowest ex-post portfolio volatility while for the minimum tracking error volatility, all four Q models obtain roughly equal performance.

4.4.3. Weekly and Monthly Rebalancing

Up until now I have focused on daily re-balancing of the portfolios. While this is an empirically relevant problem, in practice not all investors would be willing to re-balance their portfolios on a daily basis. Moreover, at longer horizons more data is available to accurately estimate the covariance matrix, which is of course an alternative method to deal with measurement error.

In this section, I investigate the performance of the Q models for weekly and monthly forecasts. Specifically, the models are used to predict next week's total volatility $\Sigma_{t|t+4}$ on each Friday, and

Table 6: Longer Horizon Portfolio Results

		Weekly				Monthly			
		HAR-DRD	HARQ-DRD	RW	1/N	HAR-DRD	HARQ-DRD	RW	1/N
		Panel A: Global Minimum Variance							
TO		0.005	0.005	0.006	0.005	0.003	0.002	0.003	0.003
CO		0.458	0.477	0.545	0.324	0.444	0.456	0.494	0.340
SP		-0.058	-0.063	-0.148	0.000	-0.046	-0.047	-0.089	0.000
Mean Ret		1.364	1.741	1.801	1.213	1.966	1.992	1.087	1.328
Std Ret		16.565	**16.030**	16.968	19.278	16.667	**16.451**	16.991	20.019
c=0%	Sharpe	0.082	0.109	0.106	0.063	0.118	0.121	0.064	0.066
	Δ_1		0.450				0.060		
	Δ_{10}		1.212				0.371		
c=1%	Sharpe	0.082	0.108	0.105	0.063	0.118	0.121	0.064	0.066
	Δ_1		0.451				0.060		
	Δ_{10}		1.213				0.372		
c=2%	Sharpe	0.081	0.107	0.104	0.062	0.117	0.120	0.063	0.066
	Δ_1		0.452				0.061		
	Δ_{10}		1.214				0.372		
		Panel B: Minimum Tracking Error Variance							
TO		0.004	0.003	0.004	0.004	0.002	0.002	0.002	0.002
CO		0.337	0.342	0.358	0.304	0.334	0.338	0.343	0.328
SP		0.000	0.000	-0.003	0.000	0.000	0.000	0.000	0.000
Mean TE		0.796	0.801	0.799	1.020	0.804	0.885	0.783	0.998
Std TE		**7.667**	**7.603**	7.981	8.178	**7.689**	**7.692**	7.778	8.253
c=0%	Δ_1		0.239				0.058		
	Δ_{10}		0.337				0.067		
c=1%	Δ_1		0.238				0.057		
	Δ_{10}		0.336				0.066		
c=2%	Δ_1		0.236				0.056		
	Δ_{10}		0.335				0.065		

Note: The top panel shows the portfolio results for the Global Minimum Variance Portfolio and the bottom panel shows the Minimum Tracking Error results where portfolios are re-balanced weekly or monthly based on forecasts of the weekly and monthly covariance matrix. Each panel shows turnover (TO), portfolio concentration (CO), as well as the average annualized return (tracking error) and its volatility. Standard deviations in bold depict the models that belong to the 90% Model Confidence Set of lowest ex-post daily volatility. For both portfolios and horizons I show the economic gains of switching from the standard HAR-DRD to the HARQ-DRD in annual percentage points, Δ_γ, for various transaction cost levels c and risk aversion coefficients γ, as outlined in Equation (22).

next month's volatility $\Sigma_{t|t+21}$ on each last trading day of the month. The manager re-balances the portfolio to the weights implied by the forecasts and holds that portfolio for the next week or month, where the process is repeated.

I focus on the HAR(Q)-DRD as an illustration, as it is straightforward to extend the model to longer horizons. The simplest way to obtain longer horizon forecasts is by means of direct projection, where instead of Equation (7), for $h-$step ahead forecasts we use

$$s_{t|t+h} = \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t. \quad (24)$$

Motivated by the long-horizon forecast results in Bollerslev, Patton, and Quaedvlieg (2015), I adapt the HARQ to the forecast horizon. While for the daily forecasts the daily lag is of greatest importance, when forecasting the weekly or monthly covariance matrix, the weekly and monthly lag increase in importance. Instead of time-varying attenuation in the daily lag, where θ_1 is made time-varying, I allow for time-variation in θ_2 for the weekly forecasts, and in θ_3 for the monthly forecasts. Specifically, the weekly and monthly specifications for the univariate HARQ are

$$\begin{aligned} s_{t|t+4} &= \theta_0 + \theta_1 s_{t-1} + \theta_{2,t} s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t \\ \theta_{2,t} &= (\theta_2 \iota + \theta_{2Q} \pi_{t-1|t-5}), \end{aligned} \tag{25}$$

and

$$\begin{aligned} s_{t|t+21} &= \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_{3,t} s_{t-22|t-1} + \epsilon_t \\ \theta_{3,t} &= (\theta_3 \iota + \theta_{3Q} \pi_{t-1|t-22}), \end{aligned} \tag{26}$$

where $\pi_{t-1|t-h} = \left(\frac{1}{h}\sum_{i=1}^{h} \text{diag}(\Pi_{t-i})\right)^{1/2}$. Again I also consider the 1/N portfolio and portfolios based on random walk forecasts. In line with the out-of-sample horizon, the random walk forecast is based on the average covariance of the past week or month.

The portfolio allocation results are reported in Table 6. The table shows that also at weekly and monthly frequencies accurate forecasts of the covariance matrix lead to better portfolio characteristics out-of-sample compared to the robust 1/N strategy. However, the relative gains of the HARQ-DRD compared to the standard HAR-DRD decrease when the forecasting horizon increases. This again highlights the fact that, due to the measurement error, the standard models particularly fail to capture the short-term dynamics, while the long-term trend is modeled accurately. Regardless, also on the weekly and monthly horizon, the HARQ-DRD improves on the standard HAR-DRD for both the GMV and minimum tracking error portfolios.

Another interesting finding is that the Sharpe Ratio of the portfolios based on HARQ-DRD forecasts with daily re-balancing (as reported in Table 4), is higher than those based on weekly and monthly re-balancing, even after accounting for transaction costs. This is not uniformly true for the standard HAR-DRD model, which is more heavily affected by transaction costs. This shows that the degree of noise in the short-term dynamics, if appropriately taken into account, does not prohibit the implementation of daily managed portfolios, even with high transaction costs.

5. Beta Forecasting

The realized beta is an important derived quantity of realized covariances and plays a central role in many asset-pricing and risk management applications. For instance, Bollerslev and Zhang (2003) and Boguth et al. (2011) use realized betas to assess multi-factor asset pricing models more accurately, Corradi et al. (2013) use them to back out conditional alphas, and Patton and Verardo (2012) use high-frequency betas to obtain market-neutral portfolios.

If the objective is to forecast the exposure to risk factors, modeling the full covariance matrix can be overly complicated. Instead, it is much simpler to directly model the realized betas themselves. In this section I investigate the merits of direct reduced form modeling of realized beta by comparing it to betas derived from covariance matrix forecasts. Similar to (co)variances, the realized beta also suffers from heteroskedastic measurement error, and reduced form modeling based on the ex-post beta estimates requires the same considerations.

Let $r_{i,t} = \{r_{i,t}^{(n)}, r_{i,t}^{(MKT)}, r_{i,t}^{(HML)}, r_{i,t}^{(SMB)}, r_{i,t}^{(MOM)}\}$, i.e. the returns of asset n, followed by the four risk factors. Its covariance matrix can be partitioned as follows

$$S_t = \begin{bmatrix} S_{nn,t} & S_{fn,t} \\ S_{nf,t} & S_{ff,t} \end{bmatrix}, \tag{27}$$

with $S_{nn,t}$ the variance of asset n, $S_{fn,t} = S'_{nf,t}$ the covariance of asset n with each of the factors, and $S_{ff,t}$ the covariance matrix of the factors. Then, the realized beta is defined as

$$b_t^{(n)} = S_{fn,t} S_{ff,t}^{-1}. \tag{28}$$

Modeling b_t rather than the covariance matrix s_t, has the benefit that the problem's dimension is significantly reduced. With just the market beta, the dimension reduction is limited from three to one, but for the four factor model, the reduction is significant from fifteen to just four.

In this section I use the HAR and HARQ models to directly forecast realized beta and compare its forecasting accuracy in a number of empirical applications with betas derived from forecasted covariance matrices. The vech HAR format directly carries over to the four dimensional beta vector. In order to obtain a HARQ specification, the asymptotic distribution of the beta estimate is needed. As beta is a function of the covariance estimate whose distribution is known, this is a straightforward application of the delta-method. Here I extend the result of Barndorff-Nielsen and Shephard (2004) to the case when there is more than just one factor.

Let $\Psi_t = D_N \Pi_t D'_N$, with D_N the duplication matrix as before. Then

$$M^{1/k}(b_t - \beta_t) \rightarrow_{Ls} MN(0, \nabla_t \Psi_t \nabla'_t), \tag{29}$$

where k is dependent on the convergence rate of the estimator used and

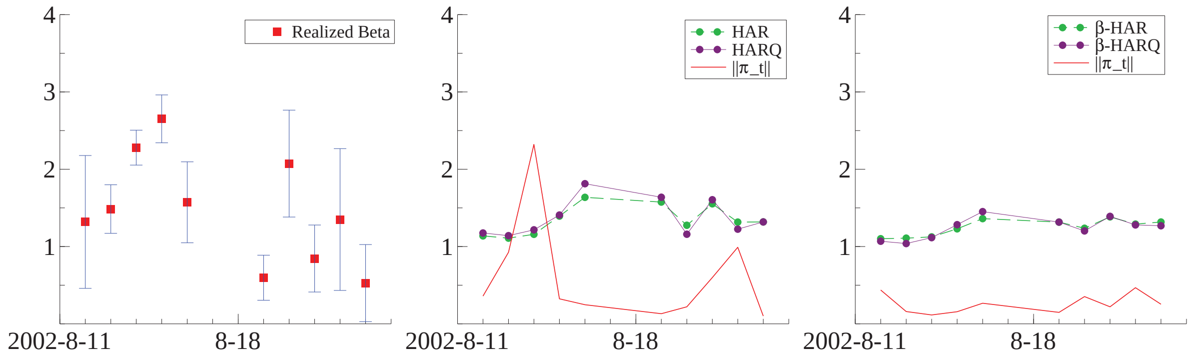
$$\nabla_t = \frac{dvec(\beta_t)}{dvec(\Sigma_t)'} = ((A\Sigma_t A')^{-1} A) \otimes (B - B\Sigma_t A'(A\Sigma_t A')^{-1} A) \tag{30}$$

with $A = \begin{bmatrix} 0_{N-1} & I_{N-1} \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0'_{N-1} \end{bmatrix}$. The result is derived in Appendix D.[17]

In order to model β directly, we can apply the HAR to the realized beta vector. I define the

[17] Note that in the bivariate case this result reduces to the standard result for scalar realized beta. Indeed, in the bi-

Figure 2: Beta Estimates and Forecasts for General Electric



Note: The graph on the left displays the realized beta estimates with 95% confidence bounds for a two week period during the dot-com crisis. The middle graph provides HAR and HARQ forecasts along with the norm of the measurement error π_t which governs the dynamics of HARQ. The right graph similarly plots the forecasts from the β-HAR and β-HARQ along the degree of measurement error which governs the dynamics of β-HARQ.

β-HAR as

$$b_t = \theta_0 + \theta_1 b_{t-1} + \theta_2 b_{t-5|t-1} + \theta_3 b_{t-22|t-1} + \epsilon_t. \tag{31}$$

Using the previous result, the $\beta-$HARQ can be defined by replacing θ_1 with $\theta_{1,t} = \theta_1 \iota + \theta_{1Q}\pi_t$, with $\pi_t = \sqrt{diag(\nabla_t \Psi_t \nabla_t')}$, analogous to the HARQ for covariances. Of course, forecasts for beta could be improved by allowing more general specifications and potentially even by modeling them all separately, but for now I consider models that are similarly parsimoniously parametrized to those of the covariance matrix to keep the comparison fair.

The distinction between modeling beta and the covariances is even more important for the HARQ and $\beta-$HARQ. Due to the non-linear nature of the beta transformation, the overall degree of measurement error of the covariance and beta can significantly diverge. Where the measurement error of the covariance may be above average, requiring greater attenuation, the beta estimate may be estimated with above average precision, allowing for strong short-term dynamics. This is especially important given the fact that small changes in covariances can have a large impact on betas.

5.1. HARQ for Beta Forecasting

The workings of the $(\beta-)$HAR(Q) are illustrated in Figure 2, where I plot General Electric's (GE) market beta estimates along with forecasts from the various models. The left-most graph

---

variate case, $\nabla = \begin{bmatrix} 0 & 0 & \Sigma_{22}^{-1} & \Sigma_{12}\Sigma_{22}^{-2} \end{bmatrix}$ which reduces to $M^{1/k}(b_t - \beta_t) \to N\left[0, (\Sigma_{22,t})^{-2}\begin{bmatrix}1 & -\beta_t\end{bmatrix}\Psi_t\begin{bmatrix}1 & -\beta_t\end{bmatrix}'\right]$ with $\beta_t = \Sigma_{12,t}\Sigma_{22,t}^{-1}$.

plots the realized beta estimates along with 95% confidence bounds. The middle and right graph show the forecasts from the covariance and beta models respectively, along with the norm of the measurement error variance, $||\pi_t||$ to proxy the total degree of measurement error which governs the dynamic attenuation of the Q models.

The figure illustrates that the measurement error variance of the covariance matrix and the beta estimates is largely uncorrelated, and as a result the daily variation in responsiveness of beta differs across the two models. First, the overall responsiveness of the Q models is greater than of the non-Q models. An interesting data point to analyze is the forecast for the 15th of August, the fourth observation in the panels. This forecast follows large uncertainty in the covariance matrix, but a precisely estimated beta. Compared to the HAR, the HARQ is less responsive to this observation, and beta only rises the day after, when there is a precise covariance estimate. In contrast, compared to the β-HAR, the $\beta-$HARQ is immediately responsive to the high beta estimate of the previous day. Comparing the covariance with the beta models, I find that overall there is less short-term variability in the beta forecasts derived from the $\beta-$HAR models. This is due to the simple fact that small changes in the covariance can lead to large changes in betas.

In the remainder I investigate which model captures the dynamics of beta best. First I consider a small simulation study in which I evaluate the relative merits of the different models for forecasting beta, and investigate the effect of dimension reduction by considering a range of one to four betas, where the vech of the covariance matrix has dimension between three and fifteen. Next I report the in- and out-of-sample performance of the HAR(Q) and $\beta-$HAR(Q) for the purpose of modeling beta and finally, I present the results of a number of empirical applications using the beta forecasts.

5.2. Simulation Results: The merits of dimension reduction

In the previous section I highlighted two reasons why modeling beta directly rather than inferring it from covariance forecasts might be advantageous. First, there is a significant dimension reduction of the vector to forecast, and second, the dynamics of the attenuation intensity of beta models are distinctly different from covariance models. Hence the covariance models might attenuate too heavily on the wrong days for the purpose of beta forecasts.

To analyze these model characteristics, this section reports the results of a small simulation study. I consider the out-of-sample performance of one-step-ahead beta forecasts from the various models. I forecast beta vectors of dimension one to four, randomly selecting one of all possible permutations of the Fama and French (1993) and Carhart (1997) factors. When forecasting a single beta, the dimension reduction is limited from three to one, while for all four betas, the dimension is reduced from fifteen to four by directly modeling the betas.

Similar to the previous simulation, in order to simulate realistic sample paths of the covariance, for each iteration, I sample a random block of 2,000 consecutive estimated covariance matrices from a randomly selected Dow Jones constituent. Within this block of data I take the covariance matrix

Table 7: Simulation Results: Beta Forecasts

Dimension Beta		1		2		3		4	
	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
HAR	∞	-0.004	0.473	0.009	0.777	-0.008	0.849	-0.003	0.912
	26	-0.006	0.489	0.005	0.806	-0.008	0.892	-0.003	0.952
	78	-0.005	0.488	0.006	0.801	-0.007	0.885	-0.002	0.951
	390	-0.004	0.483	0.006	0.800	-0.007	0.886	-0.002	0.951
HARQ	26	-0.004	0.479	0.004	0.783	-0.008	0.856	-0.001	0.920
	78	0.000	0.476	0.005	0.779	-0.006	0.852	-0.001	0.914
	390	0.000	0.474	0.005	0.778	-0.008	0.849	-0.010	0.913
β-HAR	∞	-0.008	0.469	-0.007	0.760	-0.008	0.827	-0.004	0.879
	26	-0.010	0.478	-0.008	0.769	-0.010	0.842	-0.004	0.899
	78	-0.009	0.476	-0.007	0.766	-0.009	0.839	-0.004	0.895
	390	-0.008	0.474	-0.007	0.764	-0.008	0.834	-0.004	0.889
β-HARQ	26	-0.010	0.474	-0.007	0.764	-0.008	0.830	-0.004	0.882
	78	-0.009	0.471	-0.006	0.762	-0.008	0.828	-0.004	0.880
	390	-0.007	0.470	-0.006	0.761	-0.008	0.827	-0.004	0.879

Note: The table provides the results of a simulation study of out-of-sample beta forecasting. Forecasts are based on MKernel estimates of either the realized covariance or realized beta, based on different numbers of observations M using the HAR and HARQ models. The $M = \infty$ row provides the result of the HAR models applied to population values of the covariance and beta.

of the stock with a randomly selected subset of between one and all four of the risk factors. On each day I simulate returns at one-second frequency with a diurnal volatility pattern (Andersen et al., 2012), such that the Integrate Covariance of the simulated data matches that of the empirical estimate. I then aggregate the paths to 1, 5 and 15-minute returns ($M = 390, 78, 26$), and estimate the MKernel, the implied realized beta, and their asymptotic distributions.

Using a rolling window of 1,000 observations I make one-step-ahead forecasts of the covariances and betas and compare the forecasted betas with the population beta. Again, the different sampling frequencies allow us to attribute the gains to the Q models' ability to better cope with the heteroskedastic measurement error, while the different dimensions allow us to attribute gains to the degree of dimension reduction. To set a base-level of performance, I also consider HAR forecasts using the true population covariances and betas ($M = \infty$).

The results of the simulation are reported in Table 7. First, the models' forecasts become less accurate as the dimension of the beta vector increases, with RMSE increasing by a factor two between the one- and four-dimensional betas. For both the covariance and beta models, we see improvements for the HARQ models compared to the HAR models. The HARQ models converge more quickly to the minimum obtainable bias and RMSE of the models applied to population values. Comparing the beta and covariance models, the simulations show that the beta models have slightly greater bias, but typically lower RMSE. Moreover, the percentage reduction in RMSE of the beta models increases with the dimension of beta showing that the dimension reduction plays a large role in the improvements of the models.

Table 8: In-Sample Parameter Estimates Betas

	HAR		HARQ		$\beta-$HAR		$\beta-$HARQ	
	Average	StDev	Average	StDev	Average	StDev	Average	StDev
θ_1	0.2355	0.0445	0.3481	0.0318	0.0231	0.0112	0.1629	0.0412
θ_2	0.3609	0.0478	0.3464	0.0482	0.1071	0.0594	0.0870	0.0595
θ_3	0.3001	0.0318	0.2526	0.0322	0.6197	0.0969	0.5586	0.0876
θ_{1Q}			-0.0394	0.0182			-0.0122	0.0120

Note: The table shows the average and standard deviation of the parameter estimates across the 27 DJIA assets.

5.3. In-Sample Results

I consider the 27 assets described in Section 3.1 along with the four high-frequency risk factors. For each of the assets I estimate the 5-dimensional covariance matrix and compute realized betas. In this section I present in-sample parameter estimates for the HAR(Q) and $\beta-$HAR(Q) models. The parameter estimates are reported in Table 8. For each of the models I report the average and standard deviation of the parameter estimates across the 27 assets. Interestingly, the magnitude of the θ_{1Q} coefficients relative to θ_1 is greater for these models than in the previous section where I modeled ten-dimensional matrices. This is in line with expectations as the measurement error in π_t itself aggregates across the cross-section, leading to attenuation of the Q parameter.

Andersen et al. (2006) report that the realized beta is less persistent than the covariances which is confirmed in my analysis. However, as the parameter estimates of the $\beta-$HARQ model show, part of this is due to the measurement error of b, as θ_1 increases significantly compared to the standard $\beta-$HAR. The model allows for far stronger short term dynamics when the beta estimates are precise, whereas the standard model is mostly driven by the long-term average.

5.4. Out-of-Sample Results

In this section I compare the four models in terms of their ability to forecast the beta vector. I use the same forecasting set-up as in the previous setting, re-estimating parameters daily using a rolling window of 1,000 observations. The forecasts are evaluated based on their bias and RMSE relative to the ex-post estimates for each beta separately, even though they are modeled jointly.

The forecast results are reported in Table 9. First, there is great variability in the forecasting accuracy across betas. The momentum factor loading is easier to forecast, while the SMB loading has the worst results. Overall we see that typically the $\beta-$HAR(Q) models have lower bias and also lower RMSE than the the beta forecasts implied by covariance forecasts. Additionally, the Q models reduce the RMSE of the forecasts compared to the non-Q models without exception, while the bias is only slightly reduced, similar to the covariance results reported in Section 3.

To statistically determine whether the $\beta-$HAR and Q-models significantly improve the forecasts I again use the 90% MCS to determine per risk factor, which model forecasts the most accurately. The models that make up the MCS are highlighted in bold in Table 9. Although the improvements

Table 9: Out-of-Sample Forecast Results

	HAR		HARQ		$\beta-$HAR		$\beta-$HARQ	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MKT	-0.0153	0.0289	-0.0086	**0.0263**	-0.0108	0.0281	-0.0109	**0.0261**
HML	0.0036	0.0254	0.0001	0.0253	-0.0168	**0.0247**	-0.0162	**0.0243**
SMB	-0.0231	0.0279	-0.0217	0.0274	-0.0200	0.0267	-0.0187	**0.0245**
MOM	-0.0037	0.0172	-0.0068	**0.0163**	0.0014	0.0168	0.0011	**0.0163**

Note: The table gives the average bias and RMSE of beta forecasts across time and the 27 DJIA assets. RMSE in bold depict the models that belong to the 90% Model Confidence Set of best forecasting models for that particular beta.

in RMSE may seem small, they are consistent enough to reject the null of equality of models. The $\beta-$HARQ is part of the set of best models for all risk factors signifying the merits of modeling realized beta directly.

5.5. Economic Evaluation of Beta Forecasts

In this section I compare the beta forecasts produced by the HAR(Q) and $\beta-$HAR(Q) using two methods. First, I evaluate the different models by comparing their adequacy for hedging factor risk, which is captured by their out-of-sample residual variance. I then analyze these results more deeply by comparing them directly and jointly in a regression-based comparison.

5.5.1. Beta Hedging

In order to economically evaluate the beta forecasts, I use them for beta hedging. The factor structure should most adequately capture the returns for the model with the most accurate beta forecasts. To compare the different models based on this premise, for each model and each asset, I compute the time series

$$Z_t^{(n)} = b_{MKT,t}^{(n)} r_{MKT,t} + b_{t,HML}^{(n)} r_{HML,t} + b_{t,SMB}^{(n)} r_{SMB,t} + b_{t,MOM}^{(n)} r_{MOM,t}, \tag{32}$$

where each model has its own $Z_t^{(n)}$ implied by their respective beta forecasts. I then compute the implied factor tracking error

$$r_t^{(n)} - Z_t^{(n)}. \tag{33}$$

The best model should have the lowest tracking error variance. I compute the sample variance of their time-series, and use the Model Confidence Set of Hansen et al. (2011) to identify the beta forecasting specification whose variance is significantly smaller than that of the competitors. A similar evaluation is performed in Hansen et al. (2014) and Boudt et al. (2015).

The differences in tracking error variances between the models are large. The average residual variance across assets of the HAR is 1.40, compared to 1.31 for the HARQ. This large reduction is a result of the better conditioned forecasts produced by the HARQ, where the inversion needed for the beta forecasts is less affected by the measurement error. Both β models significantly reduce the tracking error variance relative to the covariance models, with the β-HAR obtaining just 1.14 and the $\beta-$HARQ modeling the factor structure slightly better at 1.13. To statistically evaluate the magnitude of these differences I construct the model confidence such that it contains the model with the smallest tracking error with a probability of no less than 90%. I compute the set for each of the 27 DJIA stocks and find that the HAR is included just once, while the HARQ is included for six of the assets. The $\beta-$HAR models lead to much lower tracking error, with the normal $\beta-$HAR being included in the set seventeen times, and the $\beta-$HARQ 25 times.

5.5.2. Regression Based Comparison

In the previous section I considered the tracking errors based on the different models separately. In this section I run the following regression to allow for a direct competition between the different models

$$r_t^{(n)} = \alpha^{(n)} + \delta_{HAR}^{(n)} Z_{t,HAR}^{(n)} + \delta_{HARQ}^{(n)} Z_{t,HARQ}^{(n)} + \delta_{\beta-HAR}^{(n)} Z_{t,\beta-HAR}^{(n)} + \delta_{\beta-HARQ}^{(n)} Z_{t,\beta-HARQ}^{(n)} + \epsilon_t^{(n)}. \quad (34)$$

$Z_t^{(n)}$ is as before in Equation (32), with the subscript denoting from which model the beta forecasts are derived. Perfect forecasts of a model would correspond to its δ coefficient being equal to one, while all other models' δ should be equal to 0. Engle (2015) uses this approach to find support for time-variation in beta by comparing betas obtained from a DCC model with static betas. I use it to investigate whether directly modeling beta offers advantages over modeling it indirectly through the covariance matrix, and whether the Q models lead to better forecasts compared to the non-Q models.

First, consider the full Equation (34), where we can compare all models at once. To test whether any specification dominates all others I test the following four hypotheses:

$$\begin{aligned}
H_{HAR} \quad & : \delta_{HAR} = 1, \delta_{HARQ} = 0, \delta_{\beta-HAR} = 0, \delta_{\beta-HARQ} = 0 \\
H_{HARQ} \quad & : \delta_{HAR} = 0, \delta_{HARQ} = 1, \delta_{\beta-HAR} = 0, \delta_{\beta-HARQ} = 0 \\
H_{\beta-HAR} \quad & : \delta_{HAR} = 0, \delta_{HARQ} = 0, \delta_{\beta-HAR} = 1, \delta_{\beta-HARQ} = 0 \\
H_{\beta-HARQ} \quad & : \delta_{HAR} = 0, \delta_{HARQ} = 0, \delta_{\beta-HAR} = 0, \delta_{\beta-HARQ} = 1,
\end{aligned}$$

using robust inference (White, 1980). These hypotheses are very strong, especially considering the strong correlation between the forecasts, and are expected to be rejected for almost every DJIA stock.

In order to better highlight the relative merits of the different models I split the analysis up in parts. First, in order to directly compare the Q models with their standard counterparts I estimate Equation (34) with either just the HAR and HARQ, or the $\beta-$HAR and $\beta-$HARQ. I then test the following hypotheses in the first regression

$$H'_{HARQ} : \delta_i^{HAR} = 0 \quad \text{and} \quad H'_{HAR} : \delta_i^{HARQ} = 0$$

and

$$H'_{\beta-HARQ} : \delta_i^{\beta-HAR} = 0 \quad \text{and} \quad H'_{\beta-HAR} : \delta_i^{\beta-HARQ} = 0$$

in the second regression. Note that for the hypothesis H'_{HARQ} I test whether the HAR model is still significant within the regression. Hence, a non-rejection is in favor of the HARQ, similar to the first set of hypotheses.

Next, in order to directly test the usefulness of the direct modeling of beta compared to using the forecasts for the covariance matrix, I estimate Equation 34 with just the HAR and β-HAR, and similarly with the HARQ and β-HARQ. Based on these two regressions we can test the following set of hypotheses:

$$H''_{\beta-HAR} : \delta_i^{HAR} = 0 \quad \text{and} \quad H''_{HAR} : \delta_i^{\beta-HAR} = 0$$
$$H''_{\beta-HARQ} : \delta_i^{HARQ} = 0 \quad \text{and} \quad H''_{HARQ} : \delta_i^{\beta-HARQ} = 0$$

in the first and second regressions, respectively. I estimate the five different regressions and test all twelve hypotheses for each of the 27 assets. A summary of the parameter estimates, as well as the rejection frequencies, are reported in Table 10.

Panel A reports the point estimates of the various models. The top half of the panel shows descriptives of the parameter estimates across the assets for the full specification of Equation (34), while the bottom half shows the parameter estimates when one Z is included at a time. First, due to the high collinearity of the various Z variables, large variation can be seen in the point estimates. On average, the HARQ and $\beta-$HARQ obtain the point estimates closest to one. When only considering one model at a time, all models appear able to capture the betas relatively well, with average point estimates close to one, and again the $\beta-$HAR(Q) significantly closer than the covariance models.

Panel B shows the rejection rates for the various hypotheses. First consider the results on the full specification in the top row. The hypothesis that the HAR forecasts the ideal time series of beta is rejected for every asset, leaving no doubt that the HAR is not the best model. The HARQ is not rejected for a small set of series, with even fewer rejections for the $\beta-$HAR and $\beta-$HARQ specification. The $\beta-$HARQ survives this stringent test the best, as the model is not rejected as the single best model for seven of the series.

Next consider the results of the direct comparison of Q models with non-Q models in the second row of Panel B. The HAR model is almost uniformly rejected, implying that the HARQ almost always obtains non-zero weight in the direct comparison between HAR and HARQ. On the other

Table 10: Regression Based Comparison of Betas

Panel A: Parameter Estimates				
	δ_i^{HAR}	δ_i^{HARQ}	$\delta_i^{\beta-HAR}$	$\delta_i^{\beta-HARQ}$
		Full Model		
Mean	-0.183	0.618	-0.086	0.648
StDev	0.307	0.361	1.017	1.070
25% Quartile	-0.279	0.383	-0.681	-0.119
Median	-0.152	0.682	-0.143	0.840
75% Quartile	0.041	0.804	0.648	1.271
		Individual		
Mean	0.853	0.917	0.991	0.992
StDev	0.074	0.060	0.095	0.095
25% Quartile	0.800	0.884	0.923	0.926
Median	0.873	0.917	0.995	0.997
75% Quartile	0.901	0.961	1.059	1.060
Panel B: Rejection Frequencies Hypothesis Testing				
Full Model				
	H_{HAR}	H_{HARQ}	$H_{\beta-HAR}$	$H_{\beta-HARQ}$
	27	25	24	20
Q vs. Non-Q				
	H'_{HAR}	H'_{HARQ}	$H'_{\beta-HAR}$	$H'_{\beta-HARQ}$
	26	0	15	5
Beta vs. Covariance				
	H''_{HAR}	H''_{HARQ}	$H''_{\beta-HAR}$	$H''_{\beta-HARQ}$
	25	23	3	5

Note: This table reports the results of regression comparison based on Equation (34). Panel A shows parameter estimates of the full model and of the regression models which have one Z at a time. Panel B shows rejection frequencies of the null-hypotheses that each model best fits the data, hence lower numbers are better. The number reported is the frequency of rejections out of the possible 27 for each stock. The first row directly compares all models at once, and tests whether the models δ parameter equals one and all others equal zero jointly. The second row compares the Q with the non-Q models, i.e. $r_t = \alpha + \delta_{HAR} Z_{HAR,t} + \delta_{HARQ} Z_{HARQ,t}$ and analogous for the $\beta - HAR$ models. Here the hypotheses are $H'_{HAR} : \delta_{HARQ} = 0$ and vice versa, such that rejection is in favor of the model in the hypothesis subscript. The final row compares beta forecasts with betas derived from covariance forecasts using the regressions $r_t = \alpha + \delta_{HAR} Z_{HAR,t} + \delta_{\beta-HAR} Z_{\beta-HAR,t}$, and analogous for the Q-models. Here, $H"_{HAR} : \delta_{\beta-HAR} = 0$ and vice versa.

hand, the HARQ hypothesis is never rejected, implying that there is no significant loading on HAR. We see a similar, though less extreme, picture for the $\beta-$HAR(Q) models, where the forecasts from the Q model also better describe the data.

Finally consider the bottom row in Panel B, where I test whether modeling beta directly leads to better results than modeling the covariance. Here, the HAR and $\beta-$HAR are directly compared. Both for the Q and non-Q models, the β specification dominates the covariance specification, with far greater rejection frequencies for the covariance models. This confirms the overall conclusion of this section that modeling beta directly rather than indirectly leads to improved forecasts, with a possibility of additional improvements through dynamic attenuation models.

6. Conclusion

This paper proposes a simple extension to multivariate volatility models that incorporate high-frequency covariance estimates. The models improve on their benchmarks by allowing the responsiveness to the high-frequency estimates to vary over time with the heteroskedasticity of the measurement error. The models use the asymptotic theory of the high-frequency estimator to parsimoniously allow for dynamic attenuation of the autoregressive parameters. The new high-frequency information is influential when the estimate is precise, and the impact is significantly reduced when the estimate is noisy. This allows the models to incorporate the information of high-frequency returns much more efficiently, compared to standard models, that have a fixed degree of attenuation. I illustrate the concept on three of the most popular multivariate volatility models, the vech HAR, EWMA and HEAVY. The models with time-varying parameters are dubbed HARQ, EWMAQ and HEAVYQ, as the adjustment is based on the Integrated Quarticity of high-frequency returns. The simple extension leads to significantly improved out-of-sample performance.

I evaluate the performance of the models in various financial applications. First, I consider a dynamic portfolio allocation for which the Q models are perfectly suited. I use Global Minimum Variance portfolios such that the performance of the various models is not contaminated by estimation error in the mean. Here the models can be interpreted as dynamic shrinkage models, where the covariance forecasts are shrunk towards the unconditional covariance matrix when estimation uncertainty is large. The more accurate forecasts of the Q models lead to better portfolio allocation, with reduced ex-post variance and higher Sharpe ratios. Moreover, by limiting the influence of noisy covariance estimates, the forecasts are more stable, which leads to reduced turnover. As a result, an investor implementing this volatility-timing strategy would be willing to pay up to 180 basis points per year to switch to the Q models from their respective benchmark.

In addition, I use the models to forecast betas. Here, I evaluate the forecasting accuracy of dynamically modeling realized betas, versus betas implied by forecasts from a covariance model. I similarly derive a Q version for the dynamics of beta, incorporating the asymptotic distribution of the realized beta estimate. The main advantage of directly modeling beta compared to indirectly through covariances is the significant reduction in the dimension of the problem. Moreover, the attenuation dynamics of the Q models can differ significantly between the beta and covariance models. I find that out-of-sample forecasts of dynamic models of beta are more accurate from a statistical point of view, and also perform better in a number of risk management applications, with the single best model the Q version of the dynamic model of beta.

7. References

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

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

Andersen, T. G., Bollerslev, T., Frederiksen, P., Nielsen, M. Ø., 2010. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. Journal of Applied Econometrics 25 (2), 233–261.

Andersen, T. G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics 169 (1), 75–93.

Baker, M., Bradley, B., Wurgler, J., 2011. Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. Financial Analysts Journal 67 (1), 40–54.

Bandi, F. M., Russell, J. R., Zhu, Y., 2008. Using high-frequency data in dynamic portfolio choice. Econometric Reviews 27 (1-3), 163–198.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76 (6), 1481–1536.

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

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

Boguth, O., Carlson, M., Fisher, A., Simutin, M., 2011. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. Journal of Financial Economics 102 (2), 363–389.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2015. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics (Forthcoming).

Bollerslev, T., Zhang, B. Y., 2003. Measuring and modeling systematic risk in factor pricing models using high-frequency data. Journal of Empirical Finance 10 (5), 533–558.

Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., Orimar, S., 2015. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. Working Paper.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., Loris, I., 2009. Sparse and stable markowitz portfolios. Proceedings of the National Academy of Sciences 106 (30), 12267–12272.

Brown, D. B., Smith, J. E., 2011. Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds. Management Science 57 (10), 1752–1770.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52 (1), 57–82.

Chan, L. K., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. Review of Financial Studies 12 (5), 937–974.

Chiriac, R., Voev, V., 2010. Modelling and forecasting multivariate realized volatility. Journal of Applied Econometrics 26 (6), 922–947.

Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. Journal of Econometrics 159 (1), 116–133.

Corradi, V., Distaso, W., Fernandes, M., 2013. Conditional alphas and realized betas. Working Paper.

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

De Lira Salvatierra, I., Patton, A. J., 2015. Dynamic copula models and high frequency data. Journal of Empirical Finance 30, 120–135.

DeMiguel, V., Garlappi, L., Nogales, F. J., Uppal, R., 2009. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science 55 (5), 798–812.

DeMiguel, V., Nogales, F. J., Uppal, R., 2014. Stock return serial dependence and out-of-sample portfolio performance. Review of Financial Studies 27 (4), 1031–1073.

Doornik, J., 2009. Object-oriented Matrix Programming Using Ox. Timberlake Consultants Press.

Engle, R. F., 2015. Dynamic conditional beta. Working Paper.

Epps, T. W., 1979. Comovements in stock prices in the very short run. Journal of the American Statistical Association 74 (366a), 291–298.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. Journal of Econometrics 147 (1), 186–197.

Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. Journal of the American Statistical Association 107 (497), 412–428.

Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. The Journal of Finance 56 (1), 329–352.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using realized volatility. Journal of Financial Economics 67 (3), 473–509.

Frost, P. A., Savarino, J. E., 1986. An empirical bayes approach to efficient portfolio selection. Journal of Financial and Quantitative Analysis 21 (3), 293–305.

Han, Y., 2006. Asset allocation with a high dimensional latent factor stochastic volatility model.

Review of Financial Studies 19 (1), 237–271.

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

Hansen, P. R., Lunde, A., Voev, V., 2014. Realized beta garch: a multivariate garch model with realized measures of volatility. Journal of Applied Econometrics 29 (5), 774–799.

Hautsch, N., Kyj, L. M., Malec, P., 2015. Do high-frequency data improve high-dimensional portfolio allocations? Journal of Applied Econometrics 30 (2), 263–290.

Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. Bernoulli 11 (2), 359–379.

Holtz-Eakin, D., Newey, W., Rosen, H. S., 1988. Estimating vector autoregressions with panel data. Econometrica 56 (6), 1371–1395.

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

Komunjer, I., Ng, S., 2014. Measurement errors in dynamic models. Econometric Theory 30 (1), 150–175.

Laurent, S., Rombouts, J. V., Violante, F., 2013. On loss functions and ranking forecasting performances of multivariate volatility models. Journal of Econometrics 173 (1), 1–10.

Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10 (5), 603–621.

Ledoit, O., Wolf, M., 2004a. Honey, I shrunk the sample covariance matrix. Journal of Portfolio Management 30 (4), 110–119.

Ledoit, O., Wolf, M., 2004b. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88 (2), 365–411.

Li, J., 2015. Sparse and stable portfolio selection with parameter uncertainty. Journal of Business & Economic Statistics 33 (3), 381–392.

Liu, Q., 2009. On portfolio optimization: How and when do we benefit from high-frequency data? Journal of Applied Econometrics 24 (4), 560–582.

Lunde, A., Shephard, N., Sheppard, K., 2015. Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. Journal of Business & Economic Statistics (Forthcoming).

Magnus, J. R., Neudecker, H., 1980. The elimination matrix: some lemmas and applications. SIAM Journal on Algebraic Discrete Methods 1 (4), 422–449.

Noureldin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics 27 (6), 907–933.

Oh, D. H., Patton, A. J., 2015. Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. Working Paper.

Pakel, C., Shephard, N., Sheppard, K., Engle, R. F., 2014. Fitting vast dimensional time-varying covariance models. Working Paper.

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

Patton, A. J., Verardo, M., 2012. Does beta move with news? Firm-specific information flows and learning about profitability. Review of Financial Studies 25 (9), 2789–2839.

Pooter, M. d., Martens, M., Dijk, D. v., 2008. Predicting the daily covariance matrix for S&P 100 stocks using intraday databut which frequency to use? Econometric Reviews 27 (1-3), 199–229.

Staudenmayer, J., Buonaccorsi, J. P., 2005. Measurement error in linear autoregressive models. Journal of the American Statistical Association 100 (471), 841–852.

Varneskov, R., Voev, V., 2013. The role of realized ex-post covariance measures and dynamic model choice on the quality of covariance forecasts. Journal of Empirical Finance 20, 83–95.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838.

Appendix A. MKernel Theory and Estimation

The Multivariate Kernel is defined as

$$\text{MKernel}_t = \sum_{h=-M}^{M} k(h/H)\Gamma_h, \tag{A.1}$$

where $\Gamma_h = \sum_{i=h+1}^{M} r_{i,t} r'_{i-h,t}$, $\Gamma_h = \Gamma'_{-h}$ and $k(.)$ is an appropriate kernel function, where I use the parzen kernel. The asymptotic theory of the MKernel requires 'jittering', i.e. averaging out the first few and last few observations, which I do implement but leave out for notational convenience.

According to Theorem 2 of Barndorff-Nielsen et al. (2011),

$$M^{1/5}(\text{vech MKernel}_t - \text{vech } \Sigma_t) \rightarrow_{Ls} MN(\omega_t, \Pi_t) \tag{A.2}$$

The bias term, ω_t, is negligibly small and ignored in this paper, and $\Pi_t = 3.777 IQ_t$, where IQ_t is the Integrated Quarticity. Following the suggestions in Barndorff-Nielsen et al. (2011), I estimate Π_t using the estimator of Barndorff-Nielsen and Shephard (2004), but on pre-averaged data. Defining $x_{i,t} = \text{vech}(\bar{r}_{i,t}\bar{r}'_{i,t})$, where $\bar{r}_{i,t}$ is the pre-averaged data using bandwidth H, the IQ_t can be estimated as

$$\widehat{IQ_t} = M \sum_{i=1}^{M-H} x_{i,t}x'_{i,t} - \frac{M}{2} \sum_{j=1}^{M-H-1} (x_{i,t}x'_{i+1,t} + x_{i+1,t}, x'_{i,t}). \tag{A.3}$$

Appendix B. Descriptive Statistics

Table B.1 provides a set of summary statistics for the daily variance estimates. The estimates in this table are obtained using the Realized Kernel estimator of Barndorff-Nielsen et al. (2008), the univariate counterpart of the MKernel.

Table B.1: Realized Kernel Summary Statistics

Name		Min	Mean	Median	Max	Skewness	Kurtosis
				Factors			
Market	MKT	0.025	0.954	0.483	62.824	11.465	221.171
High-minus-Low	HML	0.022	0.563	0.265	36.710	13.226	315.023
Small-minus-Big	SMB	0.034	0.831	0.435	44.386	10.219	173.225
Momentum	MOM	0.041	1.865	0.608	92.309	9.003	121.663
				Individual Stocks			
American Express	AXP	0.035	3.906	1.876	167.268	7.853	107.127
Boeing	BA	0.084	2.782	1.641	82.516	7.042	86.481
Caterpillar	CAT	0.086	3.329	1.961	107.856	6.681	84.979
Cisco Systems	CSCO	0.086	5.284	2.920	313.587	13.314	348.012
Chevron	CVX	0.069	1.975	1.249	118.871	14.147	319.889
DuPont	DD	0.023	2.723	1.579	97.924	7.072	102.556
Walt Disney	DIS	0.085	2.835	1.584	82.146	5.926	60.945
General Electric	GE	0.037	2.709	1.268	147.268	9.763	151.835
The Home Depot	HD	0.053	3.215	1.837	121.143	7.411	107.310
IBM	IBM	0.054	2.250	1.262	62.827	5.983	66.064
Intel	INTC	0.098	4.147	2.448	88.254	4.910	45.995
Johnson & Johson	JNJ	0.028	1.582	0.956	54.406	7.455	106.876
JPMorgan Chase	JPM	0.046	4.329	2.021	200.125	9.030	128.000
Coca-Cola	KO	0.021	1.740	1.014	49.874	7.128	89.965
McDonald's	MCD	0.027	2.122	1.309	88.773	8.871	167.448
3M	MMM	0.050	1.929	1.121	113.378	13.476	363.764
Merck	MRK	0.041	2.346	1.431	108.603	10.357	201.485
Microsoft	MSFT	0.041	2.897	1.781	54.623	4.735	42.371
Nike	NKE	0.087	3.202	1.731	97.949	6.024	68.289
Pfizer	PFE	0.088	2.350	1.483	74.954	6.888	90.296
Procter & Gamble	PG	0.040	1.687	0.993	75.425	9.049	160.917
Travelers	TRV	0.073	2.805	1.294	134.942	9.020	135.165
UnitedHealth Group	UNH	0.070	4.097	2.231	156.822	8.046	106.611
United Technologies	UTX	0.066	2.436	1.354	136.111	11.758	288.459
Verizon	VZ	0.084	2.238	1.279	146.248	14.234	457.643
Wal-Mart	WMT	0.067	2.483	1.467	88.586	6.698	94.866
ExxonMobil	XOM	0.065	1.789	1.066	147.160	18.078	563.450

Note: The table provides summary statistics for the Realized Kernel for each of the 27 DJIA stocks and the high-frequency risk factors. The sample covers the period from February 1993 through December 2013.

Appendix C. Composite Likelihood Estimators

Estimation of large-dimensional models often becomes very time-consuming due to the need to invert the large dimensional conditional covariance matrix.

Pakel et al. (2014) propose a composite likelihood theory. The method is based on summing up quasi-likelihoods of subsets of assets. Each subset has a valid quasi-likelihood, but has only limited information on the parameters. By summing over many subsets the estimator has the desirable property that we do not have to invert large-dimensional matrices.

The quantity of interest is $V_t = E(r_t r_t')$, where V_t is modeled parametrically, by means of e.g. the EWMA or HEAVY model, with parameter vector ψ. Assuming normality, the quasi-likelihood is given by

$$\log L(\psi; r) = \sum_{t=1}^{T} \ell(\psi; r_t) \text{ with } \ell(\psi; r_t) = -\frac{1}{2}\log|V_t| - \frac{1}{2} r_t' V_t^{-1} r_t. \tag{C.1}$$

The main problem with this is that V_t is potentially of large dimension and it has to be inverted often with numerical optimization.

Composite likelihoods sidestep this issue by instead approximating the likelihood with a large number of lower dimensional marginal densities. The dimension of the problem is reduced from N to 2. While different options are available, I will consider contiguous pairs, $Y_{1t} = (r_{1,t}, r_{2,t})$, $Y_{2t} = (r_{2,t}, r_{3,t})$,..., $Y_{d-1,t} = (r_{d-1,t}, r_{d,t})$, with composite likelihood

$$CL(\psi) = \frac{1}{T(N-1)} \sum_{j=1}^{N-1} \sum_{t=1}^{T} \ell_{jt}(\psi; Y_{jt}) \tag{C.2}$$

The estimator is consistent and asymptotically normal. Distributions of the estimates based on this theory are available in Pakel et al. (2014).

Appendix D. Distribution of $\boldsymbol{\beta}$

Assuming returns are ordered as $r_t = \{r_t^{(i)}, r_t^{(MKT)}, r_t^{(HML)}, r_t^{(SMB)}, r_t^{(MOM)}\}$, i.e. the individual stock first and the risk factors afterwards with Σ its $N \times N$ covariance matrix. Then, realized beta as defined in Equation (28), can be written as

$$\beta = (B\Sigma A')(A\Sigma A')^{-1},$$

with $A = \begin{bmatrix} 0_{N-1} & I_{N-1} \end{bmatrix}$, and $B = \begin{bmatrix} 1 & 0'_{N-1} \end{bmatrix}$. Here, I_N is an identity matrix of size N and 0_N is a vector of zeros. To apply the delta method the derivative of β with respect to elements of Σ is needed.

Note that

$$d(A\Sigma A')^{-1} = -(A\Sigma A')^{-1} A d\Sigma A' (A\Sigma A')^{-1}$$

and thus

$$\begin{aligned} d(B\Sigma A')(A\Sigma A')^{-1} &= Bd\Sigma A'(A\Sigma A')^{-1} + B\Sigma A' d(A\Sigma A')^{-1} \\ &= Bd\Sigma A'(A\Sigma A')^{-1} - B\Sigma A'(A\Sigma A')^{-1} A d\Sigma A'(A\Sigma A')^{-1} \\ &= (B - B\Sigma A'(A\Sigma A')^{-1} A) d\Sigma A'(A\Sigma A')^{-1} \end{aligned}$$

Using $vec(XYZ) = (Z' \otimes X)vec(Y)$,

$$\nabla = \frac{dvec(\beta)}{dvec(\Sigma)'} = ((A\Sigma A')^{-1} A \otimes (B - B\Sigma A'(A\Sigma A')^{-1} A)$$

Let $\Psi_t = D_N \Pi_t D'_N$ is the distribution of vec S_t. Using the delta method we find that the MKernel estimate of β, denoted b has distribution

$$M^{1/5}(b - \beta) \to_{Ls} MN(0, \nabla \Psi \nabla'). \qquad \text{(D.1)}$$