



Large dynamic covariance matrices: Enhancements based on intraday data

Gianluca De Nard^{a,b}, Robert F. Engle^c, Olivier Ledoit^{d,e}, Michael Wolf^{d,*}

^a Department of Banking and Finance University of Zurich Zurich CH-8032, Switzerland

^b NYU Stern Volatility and Risk Institute, New York, NY 10012, USA

^c Department of Finance New York University New York, NY 10012, USA

^d Department of Economics University of Zurich Zurich CH-8032, Switzerland

^e AlphaCrest Capital Management, New York, NY 10036, USA

ARTICLE INFO

Article history:

Received 24 July 2020

Accepted 26 January 2022

Available online 1 February 2022

JEL classification:

C13

C58

G11

Keywords:

Dynamic conditional correlations

Intraday data

Markowitz portfolio selection

Multivariate GARCH

Nonlinear shrinkage

ABSTRACT

Multivariate GARCH models do not perform well in large dimensions due to the so-called curse of dimensionality. The recent DCC-NL model of Engle et al. (2019) is able to overcome this curse via nonlinear shrinkage estimation of the unconditional correlation matrix. In this paper, we show how performance can be increased further by using open/high/low/close (OHLC) price data instead of simply using daily returns. A key innovation, for the improved modeling of not only dynamic variances but also of dynamic correlations, is the concept of a *regularized return*, obtained from a volatility proxy in conjunction with a smoothed sign of the observed return.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Modeling and forecasting dynamic (or time-varying) covariance matrices of a vector of asset returns has many important applications in finance, such as Markowitz portfolio selection and risk management. A popular tool to this end are multivariate GARCH models. Historically, such models did not perform well in large dimensions due to the so-called curse of dimensionality. The recent DCC-NL model of Engle et al. (2019) is able to overcome this curse via *nonlinear shrinkage* (NL) estimation of the unconditional correlation matrix.

Just as the original *dynamic conditional correlations* (DCC) model of Engle (2002), also the DCC-NL model is based on daily returns for the modeling of both (univariate) dynamic variances and the dynamic correlation matrix; those quantities are then combined for the modeling of the dynamic covariance matrix. It is known that the use of intraday price data can lead to improved modeling of dynamic variances: Even if one uses only the four pieces of

information given by open/high/low/close (OHLC) prices, the improvements can be considerable. The intuition is that using an improved volatility proxy as the innovation of a GARCH-type model works better than using the squared daily return, which is the original proposal of Engle (1982) and Bollerslev (1986).

Therefore, it is natural to use such an approach in the first step of a DCC(-NL) model, the modeling of dynamic variances. Doing so has already been suggested before; for example, see Molnár (2016) and Fiszeder et al. (2019). On the other hand, we are the first to extend this approach to the second step of a DCC(-NL) model, the modeling of the dynamic correlation matrix. The crux of our proposal is as follows. The intraday counterpart of the daily squared return is the (improved) volatility proxy. Hence, we suggest to use as the intraday counterpart of the daily return itself the square root of the volatility proxy in conjunction with a (smoothed) sign of the return, a construct we call *regularized return*. Then these regularized returns are used in the second step of the DCC(-NL) model instead of the daily returns.

Importantly, the resulting models remain computationally feasible also for large dimensions of $N \geq 1000$ assets. In contrast, the HEAVY-DCC model, for example, which is also based on intraday data, can be applied only to small investment universes; for ex-

* Corresponding author at: Zürichbergstrasse 14, CH-8032 Zurich, Switzerland.

E-mail addresses: gianluca.denard@bf.uzh.ch (G. De Nard), rengle@stern.nyu.edu (R.F. Engle), olivier.ledoit@econ.uzh.ch (O. Ledoit), michael.wolf@econ.uzh.ch (M. Wolf).

ample, see the empirical analyses with $N \leq 10$ of Xu (2019) and Noureldin et al. (2012), respectively.

The remainder of the paper is organized as follows. Section 2 reviews existing DCC(-NL) models. Section 3 details our new models based on intraday data. Section 4 gives a brief description of existing estimators of dynamic variances that we deem the most useful for our purpose. Section 5 describes the empirical methodology and presents the results of out-of-sample backtest exercises based on real-life stock return data. Section 6 concludes. An appendix contains all figures and tables.

2. Large dynamic covariance matrices

2.1. Notation

In what follows, the subscript i indexes assets and covers the range of integers from 1 to N , where N denotes the dimension of the investment universe; the subscript t indexes dates and covers the range of integers from 1 to T , where T denotes the sample size. $\text{Cor}(\cdot)$ denotes the correlation matrix of a random vector, $\text{Cov}(\cdot)$ denotes the covariance matrix of a random vector, and $\text{Diag}(\cdot)$ denotes the function that sets to zero all the off-diagonal elements of a square matrix. Furthermore, we use the following notations:

- $o_{i,t}$: observed opening price (“open”) for asset i at date t , stacked into $o_t := (o_{1,t}, \dots, o_{N,t})'$
- $h_{i,t}$: observed highest price (“high”) transacted for asset i at date t , stacked into $h_t := (h_{1,t}, \dots, h_{N,t})'$
- $l_{i,t}$: observed lowest price (“low”) transacted for asset i at date t , stacked into $l_t := (l_{1,t}, \dots, l_{N,t})'$
- $c_{i,t}$: observed closing price (“close”) for asset i at date t , stacked into $c_t := (c_{1,t}, \dots, c_{N,t})'$
- $r_{i,t}$: observed return for asset i at date t , stacked into $r_t := (r_{1,t}, \dots, r_{N,t})'$
- $\tilde{r}_{i,t}$: regularized return for asset i at date t , stacked into $\tilde{r}_t := (\tilde{r}_{1,t}, \dots, \tilde{r}_{N,t})'$
- $x_{i,t}$: underlying time series for covariance matrix estimation; thus $x_{i,t} \in \{r_{i,t}, \tilde{r}_{i,t}\}$
- $d_{i,t}^2 := \text{Var}(x_{i,t} | \mathcal{F}_{t-1})$: conditional variance of the i th asset at t
- $s_{i,t} := x_{i,t}/d_{i,t}$: devolatilized series, stacked into $s_t := (s_{1,t}, \dots, s_{N,t})'$
- D_t : the N -dimensional diagonal matrix whose i th diagonal element is $d_{i,t}$
- $R_t := \text{Cor}(x_t | \mathcal{F}_{t-1}) = \text{Cov}(s_t | \mathcal{F}_{t-1})$: conditional correlation matrix at date t
- $\Sigma_t := \text{Cov}(x_t | \mathcal{F}_{t-1})$: conditional covariance matrix at date t ; thus $\text{Diag}(\Sigma_t) = D_t^2$
- $C := \mathbb{E}(R_t) = \text{Cor}(x_t) = \text{Cov}(s_t)$: unconditional correlation matrix

Here, the symbol $:=$ denotes a definition sign where the left-hand side is defined to be equal to the right-hand side, whereas the symbol $=$ (to be used below) denotes a definition sign where the right-hand side is defined to be equal to the left-hand side.

Remark 2.1 (Terminology). In this paper, the terms “dynamic” and “conditional” are used interchangeably. As an example, a dynamic covariance matrix is equal to a conditional covariance matrix, such as the covariance matrix Σ_t defined above; analogously for a correlation matrix and, necessarily then, also for any entries of such matrices, such as a variance, a covariance, or a correlation.

2.2. Averaged forecasting of dynamic covariance matrices

In our empirical analysis, as is common in the literature, we use (intra-)daily data to forecast dynamic covariance matrices but then hold the portfolio for an entire ‘month’ (that is, for a period of 21

subsequent trading days) before updating it again. Doing so creates a certain ‘mismatch’ for dynamic models that assume that the (conditional) covariance matrix changes at the forecast frequency, that is, at the daily level: Why use a covariance matrix forecasted only for the next day to construct a portfolio that will then be held for an entire month?

To address this mismatch, we use an ‘averaged-forecasting’ approach for all dynamic models; this approach was first suggested by De Nard et al. (2021): At portfolio construction date k , forecast the covariance matrix for all days of the upcoming month, that is, for $t = k, k + 1, \dots, k + 20$; then average those 21 forecasts and use this ‘averaged forecasts’ to construct the portfolio at date k .

To model conditional variances, we use a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + \delta_{1,i} x_{i,t-1}^2 + \delta_{2,i} d_{i,t-1}^2 \tag{2.1}$$

where $(\omega_i, \delta_{1,i}, \delta_{2,i})$ are the variable-specific GARCH(1,1) parameters. We assume that the evolution of the conditional correlation matrix over time is governed as in the DCC-NL model of Engle et al. (2019):

$$Q_t = (1 - \delta_1 - \delta_2)C + \delta_{1,t} s_{t-1} s_{t-1}' + \delta_2 Q_{t-1} \tag{2.2}$$

where (δ_1, δ_2) are the DCC-NL parameters analogous to $(\delta_{1,i}, \delta_{2,i})$. The matrix Q_t can be interpreted as a conditional pseudo-correlation matrix, or as a conditional covariance matrix of devolatilized residuals. It cannot be used directly because its diagonal elements, although close to one, are typically not exactly equal to one. Therefore, we obtain the conditional correlation matrix, and from it the conditional covariance matrix, as

$$R_t := \text{Diag}(Q_t)^{-1/2} Q_t \text{Diag}(Q_t)^{-1/2} \tag{2.3}$$

$$\Sigma_t := D_t R_t D_t. \tag{2.4}$$

Finally, the data-generating process is driven by the multivariate normal law

$$x_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma_t). \tag{2.5}$$

Hence, to determine the average of the L forecasts of the conditional covariance matrices $\Sigma_{k+l} = D_{k+l} R_{k+l} D_{k+l}$, for $l = 0, 1, \dots, L - 1$, we suggest a three-step approach where D_{k+l} and R_{k+l} can be forecasted separately.

2.2.1. Step one: Forecasting conditional univariate volatilities

According to Baillie and Bollerslev (1992), the multi-step ahead forecasts of the $i = 1, \dots, N$ GARCH(1,1) volatilities can be written as

$$\mathbb{E}[d_{i,k+l}^2 | \mathcal{F}_{k-1}] = \sum_{j=0}^{l-1} \omega_i (\delta_{1,i} + \delta_{2,i})^j + (\delta_{1,i} + \delta_{2,i})^l \mathbb{E}[d_{i,k}^2 | \mathcal{F}_{k-1}] \tag{2.6}$$

where $\mathbb{E}[d_{i,k}^2 | \mathcal{F}_{k-1}] = \omega_i + \delta_{1,i} x_{i,k-1}^2 + \delta_{2,i} d_{i,k-1}^2$. Therefore, we compute the forecasts of the N -dimensional diagonal matrix D_{k+l} as

$$\mathbb{E}[D_{k+l} | \mathcal{F}_{k-1}] = \text{Diag}\left(\sqrt{\mathbb{E}[d_{1,k+l}^2 | \mathcal{F}_{k-1}]}, \dots, \sqrt{\mathbb{E}[d_{N,k+l}^2 | \mathcal{F}_{k-1}]}\right). \tag{2.7}$$

2.2.2. Step two: Forecasting conditional correlation matrices

For the multivariate case we consider the approach of Engle and Sheppard (2001) where the multi-step ahead forecasts of the conditional correlation matrices are computed as

$$\mathbb{E}[R_{k+l} | \mathcal{F}_{k-1}] = \sum_{j=0}^{l-1} (1 - \delta_1 - \delta_2) C (\delta_1 + \delta_2)^j + (\delta_1 + \delta_2)^l \mathbb{E}[R_k | \mathcal{F}_{k-1}] \tag{2.8}$$

using the approximation $\mathbb{E}[R_k|\mathcal{F}_{k-1}] \approx \mathbb{E}[Q_k|\mathcal{F}_{k-1}]$. In practice, the diagonal elements of the matrix C tend to deviate from one slightly, in spite of the fact that devolatilized returns are used as inputs. Therefore, every column and every row has to be divided by the square root of the corresponding diagonal entry, so as to produce a proper correlation matrix.

2.2.3. Step three: Averaging forecasted conditional covariance matrices

By using the notation $\hat{\Sigma}_{k+l} := \mathbb{E}[\Sigma_{k+l}|\mathcal{F}_{k-1}]$, $\hat{R}_{k+l} := \mathbb{E}[R_{k+l}|\mathcal{F}_{k-1}]$ and $\hat{D}_{k+l} := \mathbb{E}[D_{k+l}|\mathcal{F}_{k-1}]$ we finally calculate $\hat{\Sigma}_{k+l} := \hat{D}_{k+l}\hat{R}_{k+l}\hat{D}_{k+l}$, for $l = 0, 1, \dots, L-1$. Therefore, to get the estimated covariance matrix on portfolio construction day k we average over the L forecasts:

$$\hat{\Sigma}_k := \frac{1}{L} \sum_{l=0}^{L-1} \hat{\Sigma}_{k+l}. \quad (2.9)$$

2.3. Estimation of parameters

In practice, both the GARCH parameters in step one and the DCC(-NL) parameters in step two need to be estimated. In doing so, we mainly follow the suggestions of Engle et al. (2019, Section 3).

In step one, the GARCH parameters of Eq. (2.1) are estimated using (pseudo) maximum likelihood assuming normality. This results in estimators $(\hat{w}_i, \hat{\delta}_{1,i}, \hat{\delta}_{2,i})$ that are used for devolatilizing returns and are also used for forecasting conditional variances via Eq. (2.6).

In step two, the correlation-targeting matrix C of Eq. (2.2) is estimated in one of two ways. For DCC, we use the sample covariance matrix of the devolatilized returns $\{s_t\}$, whereas for DCC-NL we use nonlinear shrinkage applied to the $\{s_t\}$, with post-processing analogous to (2.3) to enforce a proper correlation matrix.¹ Having an estimator \hat{C} , obtained either way, we then estimate the DCC parameters (δ_1, δ_2) via the (pseudo) composite likelihood method of (Pakel et al., 2021) assuming normality.² In this way, $(\hat{w}_i, \hat{\delta}_{1,i}, \hat{\delta}_{2,i}, \hat{\delta}_1, \hat{\delta}_2)$ are used for forecasting conditional correlation matrices via Eq. (2.8).

Combining forecasts of conditional variances with forecasts of conditional correlation matrices yields forecasts of conditional covariance matrices in the usual fashion.

3. Models based on OHLC data

One might ask why not use monthly data instead of daily data for the estimation of the various models given that the investment horizon is one month? Part of the justification is that at the monthly frequency, there generally are not enough data to estimate a multivariate GARCH model. Another part is that using daily data for the estimation tends to lead to better results even if the investment period is one month; for example, compare Tables 1 and 10 of Ledoit and Wolf (2017). So given that using daily data leads to improved performance when investing at the monthly frequency, it natural to go a step further and use intraday data, which hopefully will lead to another round of improvement.

In GARCH-type models operating at the daily frequency, the volatility of today is expressed as a function of the volatility of yesterday and a volatility innovation, typically taken to be the

¹ To speed up the computations, we use the analytical nonlinear shrinkage method of Ledoit and Wolf (2020); in contrast, Engle et al. (2019, Section 3) use the numerical method of Ledoit and Wolf (2015).

² As Engle et al. (2019, Section 3) do, we using neighboring pairs of assets to build up a (pseudo) composite likelihood.

squared (possibly demeaned) return of yesterday. From this angle, the squared return serves as a proxy for the return variance of yesterday. Although this proxy is unbiased, it is rather crude. Using a more accurate proxy based on intraday data instead should improve the performance of GARCH-type models, in terms of both in-sample fit and out-of-sample forecasting ability. In following this route, we will restrict attention to volatility proxies that only require easily obtained intraday data in the form of open/high/low/close (OHLC) prices.

3.1. Volatility proxy

Replacing the squared return $r_{i,t-1}^2$ of Eq. (2.1) with an improved volatility proxy, denoted $\hat{v}_{i,t-1}$, yields the following GARCH(1,1) model:

$$d_{i,t}^2 = \omega_i + \delta_{1,i}\hat{v}_{i,t-1} + \delta_{2,i}d_{i,t-1}^2. \quad (3.1)$$

If one only modifies the first step of the DCC(-NL) model in this way, then the returns that get devolatilized based on (3.1) for use in the second step are still the observed daily returns. We call the resulting models ID-DCC(-NL), where "ID" stands for 'volatility proxy based on IntraDay data'. Such models have been suggested before in the literature; for example, see Molnár (2016) and Fiszeder et al. (2019).

But one can forge ahead by using improved volatility proxies also in the second step of DCC(-NL) models. Especially in large dimensions, there are many more conditional covariances/correlations $((N^2 - N)/2)$ to estimate than conditional variances (N). Hence, the idea is to use intraday volatility proxies not only to improve upon the diagonal of a covariance matrix estimator but also upon its off-diagonal. To this end, we introduce the new concept of a *regularized return*.

3.2. Regularized returns

At this point, we know that the intraday approach brings substantial improvements to the diagonal of the covariance matrix. However, the diagonal is not the prize because it has only N entries. The off-diagonal is the *real* prize because it has $N(N-1)/2$ entries – one order of magnitude higher – and because it carries the biggest hazard to adversely impact the inversion of the covariance matrix for the purpose of portfolio selection, namely through ill-conditioning or even singularity.

How can one extend the intraday approach to the off-diagonal? The off-diagonal is based on cross-products of returns, not on squared returns. Hence, we need an intraday counterpart to returns themselves, which we call *regularized returns*. For a given stock, the intraday counterpart of the squared return is the volatility proxy, which then suggests to use as the intraday counterpart of the return itself the 'signed' square root of the volatility proxy. But some care is needed. On any given trading day, there are many stocks that go up and many stocks that go down. This is often easy to identify: If Apple goes up 1% and MacDonald's goes down 1%, say, then their signs are +1 and -1, respectively. However, what if Apple goes up 1 bp and MacDonald's goes down 1 bp? Is this still a straight +1/-1 signing? It could very well be that the closing price of Apple ended up on the asking price, due to short-term trading pressures, and the closing price of MacDonald's ended up on the bid price, due to other trading pressures that are uncontrollable.

At a higher level, we can refer to Max Planck's concept of the *quantum* from theoretical physics to reason that certain things are not knowable as continuously as we would wish, but only up to a certain (discrete) level of precision; and, as a matter of fact, stock prices move in quota known as ticks rather than continuously. At a more down-to-earth level, we can refer to the well-documented

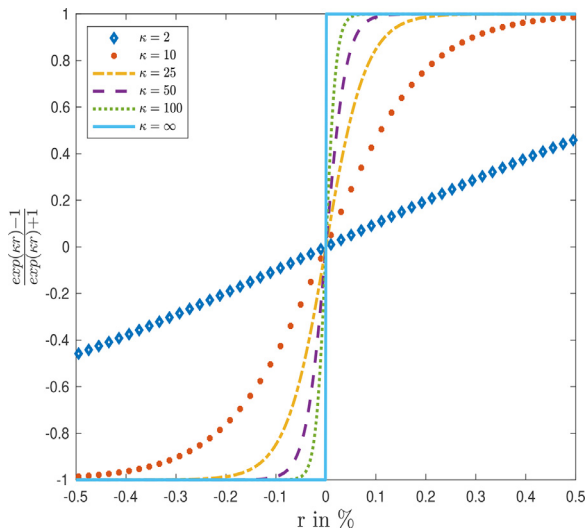


Fig. 1. The scaled hyperbolic tangent function: some examples for various values of κ .

bid-ask-bounce effect going back to at least Roll (1984). In mathematical terms, the discontinuity of the sign function is uncomfortable and may propagate deleterious effects into the statistical analysis. The point is that there is such a thing between “today Apple went up” and “today Apple was flat”, which is: “today we think Apple may sort of have gone up but we are not completely sure”. In such an intermediate case, it seems appropriate to assign Apple a middle-of-the-road signing factor in the open interval (0,1). As for MacDonalds, if it goes down 1 bp, then we would assign it a middle-of-the-road signing factor in the open interval (−1, 0).

Therefore, when ‘signing’ the root of the volatility proxy to obtain the regularized return, we recommend replacing the ordinary sign with a similar function but smoothed down by some controllable parameter that represents the amount of uncertainty about whether an observed uptick or downtick was real, and actionable, or maybe due to last-minute randomness in order-placing near the close. For tractability, we choose the ‘scaled’ hyperbolic tangent function to this end:

$$\text{stanh}(r, \kappa) := \frac{e^{\kappa r} - 1}{e^{\kappa r} + 1} \tag{3.2}$$

which is graphically displayed in Fig. 1. The regularized return is then defined as

$$\tilde{r}_{i,t-1} := \text{stanh}(r_{i,t-1}, \kappa) \cdot \sqrt{\hat{v}_{i,t-1}}. \tag{3.3}$$

The function (3.2) can be considered a smoothed sign function, where the amount of smoothing is determined by the scaling factor κ : the smaller κ , the larger the amount of smoothing. In particular, for $\kappa = 2$, one obtains the ‘standard’ hyperbolic tangent function

$$\text{stanh}(r, 2) = \frac{e^{2r} - 1}{e^{2r} + 1} = \frac{e^r - e^{-r}}{e^r + e^{-r}} = \frac{\sinh(r)}{\cosh(r)} =: \tanh(r) \tag{3.4}$$

whereas for $\kappa \rightarrow \infty$, one recovers the ordinary sign function in the limit:

$$\text{stanh}(r, \infty) := \lim_{\kappa \rightarrow \infty} \text{stanh}(r, \kappa) = \begin{cases} 1, & \text{for } r > 0 \\ 0, & \text{for } r = 0 \\ -1, & \text{for } r < 0 \end{cases} =: \text{sign}(r). \tag{3.5}$$

The reason why we use the more general scaled hyperbolic tangent function definition (3.3) instead of the sign function is, as reasoned above, that the sign function is intuitively not ‘quite right’ when the observed return $r_{i,t-1}$ is close to zero and that the scaled

hyperbolic tangent function ‘corrects’ for that: It believes the big moves but tends to disbelieve the small ones, as illustrated by Fig. 2. For observed returns close to zero, it is desirable to enforce the regularized return to have a smaller magnitude than the root of the volatility proxy, and this is exactly what the scaled hyperbolic tangent function does: shrinking the root of the volatility proxy to zero while keeping the sign of the observed return.

Of course, the relevant question is what is to be considered “close to zero” in practice? That is, for observed returns in which range should (noticeable) shrinkage of the root of the volatility proxy occur? Our proposal, going back to Roll (1984), is the bid-ask bounce, since a return in that range can be considered ‘noise’ whereas a return outside that range can be considered a true ‘signal’. The bid-ask bounce varies (somewhat) across stocks, but for simplicity we use 5 bps as a ‘typical’ common number. Armed with this range, we can now consult Fig. 1 to determine an appropriate scaling factor κ for the scaled hyperbolic tangent function. It can be seen that the choice $\kappa = 100$ basically limits shrinkage to the range [- 5 bps, 5 bps] and is therefore the natural suggestion; but any $\kappa \in [50, 200]$ would also be a reasonable choice. Some robustness checks in the empirical analysis of Section 5 will come back to this issue; see Remark 5.3.

Remark 3.1 (Returns in percent vs. raw returns). The discussion above applies to returns in percent. If raw returns are used instead, the value of κ needs to be multiplied by 100 to achieve the same amount of shrinkage; see Eq. (3.2) and Fig. 1. For example, the choice $\kappa = 100$ for returns in percent corresponds to the choice $\kappa = 10,000$ for raw returns.

Finally, for internal consistency, when we use regularized returns in the second step to model the conditional correlation matrix, we also use (squared) regularized returns in the first step to model conditional variances; that is, we upgrade (3.1) to

$$d_{i,t}^2 = \omega_i + \delta_{1,i} \tilde{r}_{i,t-1}^2 + \delta_{2,i} d_{i,t-1}^2. \tag{3.6}$$

We call the resulting models IDR-DCC(-NL), where “R” stands for “regularized returns”. Appendix A provides a detailed description of these models.

Fig. 2 display a representative example of a time series of absolute observed returns juxtaposed with the square root of the volatility proxies. on the one hand, and the absolute regularized returns, on the other hand. One can see that the absolute regularized returns are somewhat more stable than the absolute observed returns (but not as much ‘bounded away from zero’ as the square root of the volatility proxies). Hence, one could say that regularized returns place more weight on the sign of the observed return relative to its magnitude.. Arguably, this is a desirable feature, as the magnitude of daily stock returns is generally regarded as unpredictable, whereas their sign is regarded as predictable to some extent; for example, see Welch and Goyal (2008), Henriksson and Merton (1981), Pesaran and Timmermann (1995), and Christoffersen and Diebold (2006).

Intuitively, squares of the regularized returns are improved innovations for modeling dynamic variances compared to squares of observed returns (first step of DCC-type models); and similarly, cross-products of regularized returns are improved innovations for modeling dynamic covariances compared to cross-products of observed returns (second step of DCC-type models). Relating to the latter point, Fig. 3 displays a representative example of a time series of absolute cross-products of observed returns juxtaposed with the cross-products of regularized returns; again, it can be seen that using regularized returns results in somewhat more stable values.

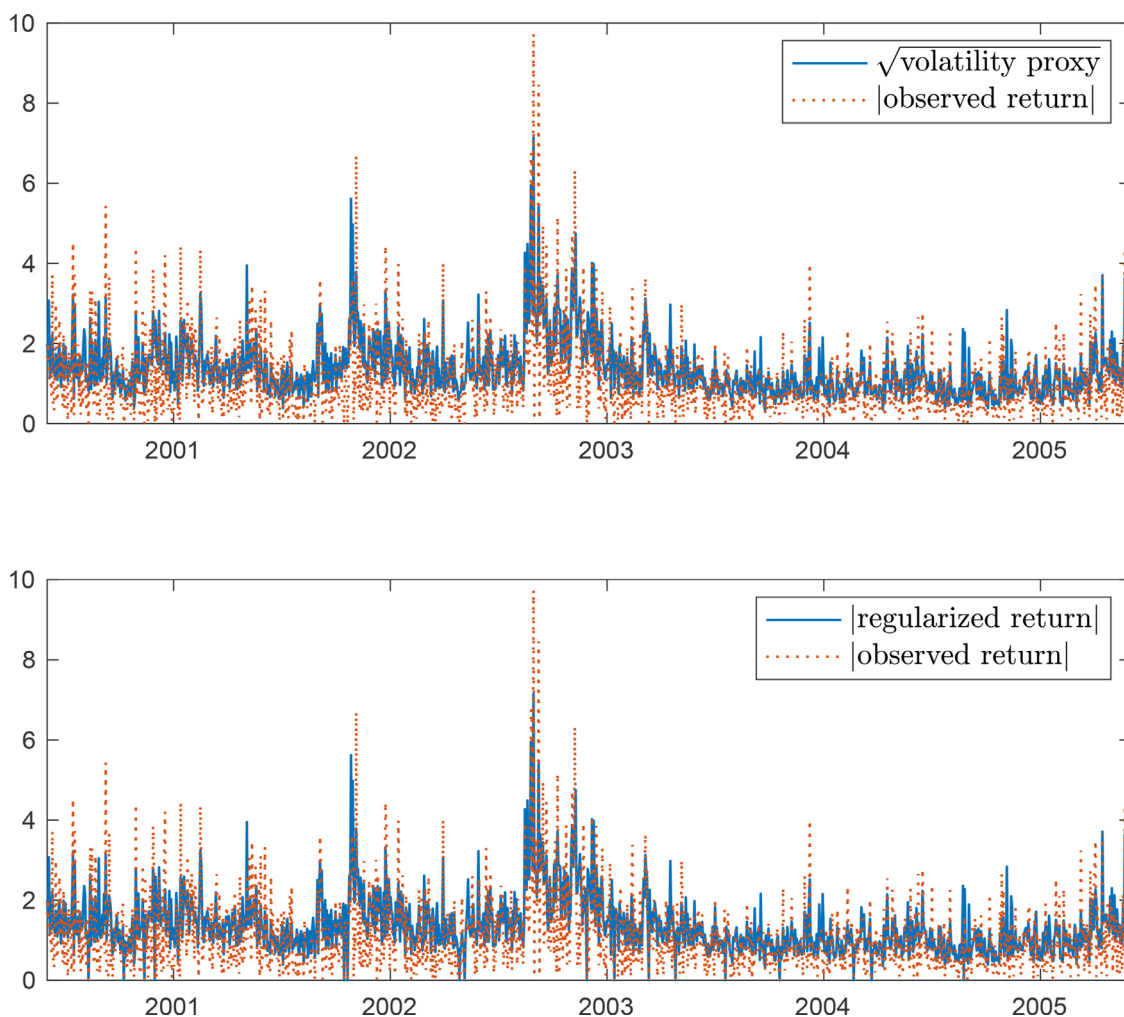


Fig. 2. Absolute observed returns vs. the square root of the volatility proxies (upper panel) and the absolute regularized returns (lower panel) for a representative stock (Exxon Mobil Corporation) and a representative period (4/19/2000–4/26/2005), in percent. The volatility proxies, which also form the basis for the regularized returns are computed using formula (4.7).

3.3. Intraday data vs. high-frequency data

Based on what we have promoted so far, it might be tempting to go even further and use high-frequency data such as 5-minute returns, or even tick-by-tick data, instead of ‘only’ intraday data in the form of OHLC prices. However, doing so would give rise to a number of difficulties. First, high-frequency data is not easily available to everybody and, when it is, it tends to be expensive, especially if one wants a large universe and a long time series. Second, even if such data is available, it requires expert preprocessing to be put in usable form, which is tedious and time-consuming; for example, see [Barndorff-Nielsen et al. \(2009\)](#). Third, using high-frequency data stretches computing resources because the data takes a lot of space, and running multiple simulations or backtest exercises can be very slow. Fourth, if the investment universe is large, there will be some smaller stocks included for which the realized variance will be downward biased due to liquidity issues (that is, due to many near-zero values for 5-minute returns). For all these reasons, we stick to intraday data in the form of OHLC prices in this paper. Nevertheless, if someone wants to go down the high-frequency route, our methodology, including regularized returns, can be adapted easily.³

³ For example, one could use realized variances as volatility proxies in [Eqs. \(3.1\) and \(3.3\)](#).

There exists already a multivariate realized GARCH model; see [Archakov et al. \(2020\)](#). It is similar to our model in that it separately models conditional variances and conditional correlations. It is more flexible than our model in that it models the evolution of the conditional correlation matrix in a more flexible way. However, doing so comes at the expense of the dimension N that can be handled in practice. It is not clear whether this model can handle dimensions $N \geq 100$; in particular, the empirical application in the paper considers $N = 9$ assets only.

A different approach based on realized measures is proposed by [Brito et al. \(2018\)](#). They forecast future realized covariances as a function of current and past realized covariance matrices in the spirit of a vector autoregressive (VAR) model. (Note that a covariance matrix can be equivalently represented as a vector by means of the vech operator.) In order to keep the number of parameters manageable when the dimension N is large, the authors make use of a covariance decomposition based on a factor model and estimate the VAR-type coefficients by LASSO or adaLASSO instead of OLS. Their empirical application considers $N = 430$ assets; see [Remark 5.4](#).

4. Volatility proxies/estimators of conditional variance

In this section, we review the existing volatility proxies (or conditional-variance estimators) that we deem the most useful for our purposes. Note that returns are not demeaned, as is common

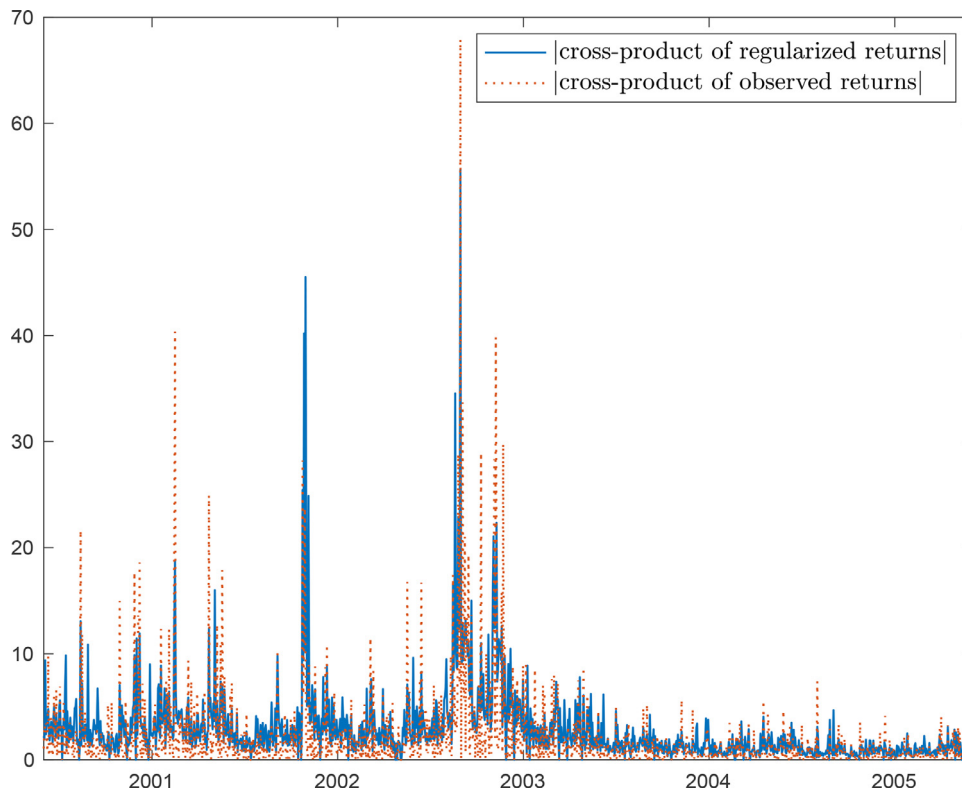


Fig. 3. Absolute cross-products of the observed returns respectively the regularized returns for two representative stocks (General Electric Corporation and Exxon Mobil Corporation) and a representative period (4/19/2000–4/26/2005), in percent. The volatility proxies, which form the basis for the regularized returns, are computed using formula (4.7).

practice in the literature when working with daily returns; for example, Hansen and Lunde (2005) demonstrate that modeling expected returns, whether conditionally or unconditionally, makes virtually no difference when forecasting conditional variances. As stated before, all of the proxies considered make only use of OHLC price data.

4.1. Close/Close

To start out, it is customary to back out a synthetic closing price on day $t - 2$ from the closing price on day $t - 1$ together with the return on day $t - 1$ as

$$\tilde{c}_{i,t-2} := \frac{c_{i,t-1}}{1 + r_{i,t-1}}. \tag{4.1}$$

Most of the time, $\tilde{c}_{i,t-2}$ is simply equal to $c_{i,t-2}$. But there can be exceptions, for example when there is an overnight dividend payment, stock split, or other corporate action. In such exceptional cases, $\tilde{c}_{i,t-2}$ can be interpreted as a suitable rescaling of $c_{i,t-2}$ that is economically more suitable to be used in formulas alongside $c_{i,t-1}$ (or any other price data recorded on day $t - 1$).

The first, and most obvious, estimator of the conditional variance on day $t - 1$ is then given as

$$r_{i,t-1}^2 = \left(\frac{c_{i,t-1}}{\tilde{c}_{i,t-2}} - 1 \right)^2. \tag{4.2}$$

$r_{i,t-1}^2$ is the usual ‘ingredient’ in the standard ARCH/GARCH models of Engle (1982) and Bollerslev (1986).

As an alternative, one can take the square of the log return instead:

$$\hat{v}_{i,t-1}^{CC} := \left[\log(1 + r_{i,t-1}) \right]^2, \tag{4.3}$$

where $\log(\cdot)$ denotes the natural logarithm.

Taking logarithms at the daily return frequency matters little in numerical terms and thus there is little difference between $r_{i,t-1}^2$ and $\hat{v}_{i,t-1}^{CC}$. The value of introducing the latter proxy is that, in the relevant literature, it is the natural stepping stone toward improved estimators, which traditionally have been couched in terms of continuously compounded returns because of mathematical grounding in the random-walk model favored by the (Black and Scholes, 1973) option-pricing formula.

4.2. Introducing the open

One of the first contributions of Garman and Klass (1980) is to realize that decomposing the close-to-close log-return as

$$\log(1 + r_{i,t-1}) = \log\left(\frac{o_{i,t-1}}{\tilde{c}_{i,t-2}}\right) + \log\left(\frac{c_{i,t-1}}{o_{i,t-1}}\right) \tag{4.4}$$

opens the door to a family of improved estimators. An important issue is scaling: Both overnight and open-market variances are on a different scale than daily, so they need to be adjusted appropriately. To this end, Garman and Klass (1980, Section III) introduce a factor $f \in (0, 1)$ that represents the proportion of variance realized when the market is closed. In consequence, they propose an improved estimator as

$$\hat{v}_{i,t-1}^{OC} := \frac{1}{2f} \left[\log\left(\frac{o_{i,t-1}}{\tilde{c}_{i,t-2}}\right) \right]^2 + \frac{1}{2(1-f)} \left[\log\left(\frac{c_{i,t-1}}{o_{i,t-1}}\right) \right]^2. \tag{4.5}$$

Yang and Zhang (2000, p. 485), based on an empirical study of U.S. equity data at the daily frequency, recommend the choice $f = 0.25$.

4.3. High-Low range

Parkinson (1980) proposes an estimator for the conditional variance during market hours on day $t - 1$ based on the range, deter-

mined by high and low, as

$$\hat{v}_{i,t-1}^{\text{HL}} := \frac{1}{4 \log 2} \left[\log \left(\frac{h_{i,t-1}}{l_{i,t-1}} \right) \right]^2 \quad (4.6)$$

where the normalizing coefficient $4 \log 2$ is derived from random-walk mathematics. In reality, however, overnight jumps matter too. Hence, Garman and Klass (1980, Section IV) propose to amend the range-based estimator as follows:

$$\hat{v}_{i,t-1}^{\text{OHL}} := \frac{a_3}{f} \left[\log \left(\frac{o_{i,t-1}}{\bar{c}_{i,t-2}} \right) \right]^2 + \frac{1-a_3}{1-f} \frac{1}{4 \log 2} \left[\log \left(\frac{h_{i,t-1}}{l_{i,t-1}} \right) \right]^2 \quad (4.7)$$

and claim that for the optimal choice of parameter $a_3 = 0.17$, $\hat{v}_{i,t-1}^{\text{OHL}}$ is around 6.2 times more efficient than the naïve close-to-close estimator $\hat{v}_{i,t-1}^{\text{CC}}$.

Remark 4.1 (Choice of tuning parameters). The estimator (4.5) involves a single tuning parameter, f , whereas the estimator (4.7) involves two tuning parameters, f and a_3 . We simply follow recommended choices from the literature and refer to the specified papers for justification. It is possible that alternative choices might result in even better performance for our purposes but we do not want to open the door to such experiments and possible ensuing suspicions of data mining.

5. Empirical analysis

5.1. Aim of the analysis

We carry out backtest exercises where various estimators of the covariance matrix are used to estimate Markowitz portfolios. To compare (pseudo) out-of-sample performance, as opposed to in-sample performance, we always estimate the portfolios based on a fixed window of ‘past’ observations and then compute the resulting returns over a fixed stretch into the ‘future’. These exercises are primarily designed to compare the accuracy of estimators of the covariance matrix; they are not designed to compare the performance of real-life investment strategies; for example, see Remark 5.1.

5.2. Data and general portfolio-construction rules

We download daily stock return data from the Center for Research in Security Prices (CRSP) starting on 01/01/1994 and ending on 12/31/2018. We restrict attention to stocks from the NYSE, AMEX, and NASDAQ stock exchanges. We also download daily OHLC price data (in dollars per share).

For simplicity, and in line with literature, we adopt the common convention that 21 consecutive trading days constitute one (trading) ‘month’. The out-of-sample period ranges from 12/18/1998 through 12/31/2018, resulting in a total of 240 months (or 5,040 days). All portfolios are updated monthly.⁴ We denote the investment dates by $k = 1, \dots, 240$. At any investment date k , a covariance matrix is estimated based on the most recent 1260 daily returns, which roughly corresponds to using five years of past data.

We consider the portfolio sizes $N \in \{100, 500, 1000\}$. For a given combination (k, N) , the investment universe is obtained as follows. We first determine the set of stocks that have an almost complete return history over the most recent $T = 1,260$ days as well as a

⁴ Monthly updating is common practice to avoid an unreasonable amount of turnover and thus transaction costs. During a month, from one day to the next, we hold number of shares fixed rather than portfolio weights; in this way, there are no transactions at all during a month.

complete return ‘future’ over the next 21 days.⁵ Additionally, we require every stock in the universe to have all four OHLC price data available at least 90% of the time and either high/low or open available at least 95% of the time. We next look for possible pairs of highly correlated stocks, that is, pairs of stocks that have returns with a sample correlation exceeding 0.95 over the past 1260 days. In such pairs, if any should exist, we remove the stock with the lower market capitalization of the two on investment date k .⁶ Of the remaining set of stocks, we then pick the largest N stocks (as measured by their market capitalization on investment date k) as our investment universe. In this way, the investment universe changes relatively slowly from one investment date to the next.

There is a great advantage in having a well-defined rule that does not involve drawing stocks at random, as such a scheme would have to be replicated many times and averaged over to give stable results. As far as rules go, the one we have chosen seems the most reasonable because it avoids so-called “penny stocks” whose behavior is often erratic; also, high-market-cap stocks tend to have the lowest bid-ask spreads and the highest depth in the order book, which allows large investment funds to invest in them without breaching standard safety guidelines.

In the best-case scenario where all the price data are available for a given stock on day $t - 1$, we use (4.7) as the volatility proxy $\hat{v}_{i,t-1}$. If high/low are available but open is missing, we use (4.6). If high/low are missing but open is available, we use (4.5). Finally, if neither high/low nor open are available, we go back to the traditional setting and use (4.3).

5.3. Competing models

The following models are included in our empirical analysis:

- **DCC**: the multivariate GARCH model of Engle (2002).
- **ID-DCC**: a model based on DCC with intraday-based volatility proxy in the first step; see formula (3.1).
- **IDR-DCC**: as ID-DCC but, additionally, with regularized returns as underlying time series for estimating DCC dynamics in the second step; see formula (3.3) and Appendix A.
- **DCC-NL**: the multivariate GARCH model of Engle et al. (2019) where the unconditional correlation matrix C is estimated via nonlinear shrinkage.
- **ID-DCC-NL**: a model based on DCC-NL with intraday-based volatility proxy in the first step; see formula (3.1).
- **IDR-DCC-NL**: as ID-DCC-NL but, additionally, with regularized returns as underlying time series for estimating DCC dynamics in the second step; see formula (3.3) and Appendix A.

Note that in the acronyms of the new models proposed, “ID” stands for “volatility proxy based on IntraDay data” and “R” stands for “regularized returns”.

5.4. Global minimum variance portfolio

We consider the problem of estimating the global minimum variance (GMV) portfolio in the absence of short-sales constraints. The problem is formulated as

$$\min_w w' \Sigma_t w \quad (5.1)$$

⁵ The first restriction allows for up to 2.5% of missing returns over the most recent 1260 days, and replaces missing values by zero. The latter, ‘forward-looking’ restriction is not feasible in practice but is commonly used in the literature. Although it might induce a (very minor) bias in absolute performance due to survivorship bias, it does not systematically affect relative performance of various portfolio-selection strategies.

⁶ The reason is that we do not want to include highly similar stocks. In the early years, there are no such pairs; in the most recent years, there are never more than three such pairs.

$$\text{subject to } w' \mathbb{1} = 1 \quad (5.2)$$

where $\mathbb{1}$ denotes a vector of ones of dimension $N \times 1$. It has the analytical solution

$$w = \frac{\Sigma_t^{-1} \mathbb{1}}{\mathbb{1}' \Sigma_t^{-1} \mathbb{1}}. \quad (5.3)$$

The natural strategy in practice is to replace the unknown covariance matrix Σ_t by an estimator $\hat{\Sigma}_t$ in formula (5.3), yielding a feasible portfolio

$$\hat{w} := \frac{\hat{\Sigma}_t^{-1} \mathbb{1}}{\mathbb{1}' \hat{\Sigma}_t^{-1} \mathbb{1}}. \quad (5.4)$$

Estimating the GMV portfolio is a 'clean' problem in terms of evaluating the quality of a covariance matrix estimator, as it abstracts from having to estimate the vector of expected returns at the same time. In addition, researchers have established that estimated GMV portfolios have desirable out-of-sample properties not only in terms of risk but also in terms of reward-to-risk, that is, in terms of the information ratio; for example, see [Haugen and Baker \(1991\)](#), [Jagannathan and Ma \(2003\)](#), and [Nielsen and Aylur-subramanian \(2008\)](#).

Remark 5.1 (Academia vs. real life). The portfolio (5.4) must be considered an 'academic' portfolio, since the only constraint it imposes is that the weights sum up to one (that is, that the portfolio be fully invested). In real life, portfolio managers face many potential additional constraints, such as gross-exposure constraints, leverage constraints, maximum-position constraints, factor-exposure constraints, and style constraints. Furthermore, it is customary to (explicitly) limit turnover, respectively trading costs, when rebalancing. Although the portfolio (5.4) is unrealistic for these reasons, it is still useful to compare the accuracy of estimators of the covariance matrix. The point is that real-life portfolio managers who operate in a Markowitz framework will always benefit from using a more accurate estimator of the covariance matrix regardless of the number and the nature of the constraints they face. In this sense, there certainly is real(-life) value to our analysis.

In addition to Markowitz portfolios based on formula (5.4), we also include as a simple-minded benchmark the equal-weighted portfolio promoted by [DeMiguel et al. \(2009\)](#), among others, as it has been claimed to be difficult to outperform. We call this portfolio **EW**. In addition, we include the value-weighted portfolio, called **VW**, which is sometimes favored over the equal-weighted portfolio, for example because it incurs lower turnover and controls for small-cap biases.

We report the following three out-of-sample performance measures for each scenario. (All of them are annualized and in percent for ease of interpretation.)

- **AV**: We compute the average of the 5,040 out-of-sample returns and then multiply by 252 to annualize.
- **SD**: We compute the standard deviation of the 5,040 out-of-sample returns and then multiply by $\sqrt{252}$ to annualize.
- **IR**: We compute the (annualized) information ratio as the ratio AV/SD ; so implicitly the benchmark is zero here.
- **SR**: We compute the (annualized) Sharpe ratio, which is defined as the information ratio with the riskfree rate as the benchmark.⁷

Our stance is that in the context of the GMV portfolio, the most important performance measure is the out-of-sample standard deviation (SD). The true (but unfeasible) GMV portfolio is

given by (5.3). It is designed to minimize the variance (and thus the standard deviation) rather than to maximize the expected return or the information ratio. Therefore, any portfolio that implements the GMV portfolio should be primarily evaluated by how successfully it achieves this goal. A high out-of-sample average return (AV) and a high out-of-sample information ratio (IR) are naturally also desirable, but should be considered of secondary importance when evaluating the quality of a covariance matrix estimator.

We also study whether one model delivers a lower out-of-sample standard deviation than another model with statistical significance. As we include 7 models, there are 21 pairwise comparisons. To avoid a multiple testing problem and since a major goal of this paper is to show that using higher-frequency (intraday) data improves the estimation of large-dimensional covariances matrices, we restrict attention to two comparisons: (i) DCC with IDR-DCC and (ii) DCC-NL with IDR-DCC-NL. For a given universe size N , a two-sided p -value for the null hypothesis of equal standard deviations is obtained by the prewhitened HAC_{PW} method described in [Ledoit and Wolf \(2011, Section 3.1\)](#).⁸

The results are presented in [Table 1](#) and can be summarized as follows; unless stated otherwise, the findings are with respect to SD as the performance measure.

- All models outperform EW and VW by a wide margin.
- Each DCC-NL model outperforms its DCC version.
- Each intraday model, IDR-DCC(-NL) and ID-DCC(-NL), outperforms its traditional base model, DCC(-NL). Additionally, IDR-DCC(-NL) outperforms its ID-DCC(-NL) version.
- Moreover, the outperformance of IDR-DCC-NL over DCC-NL, respectively IDR-DCC over DCC, is always statistically significant and also economically meaningful.
- There is a consistent ranking across all portfolio sizes N (from best to worst):
IDR-DCC-NL, ID-DCC-NL, IDR-DCC, ID-DCC, DCC-NL, DCC, VW, EW.⁹

To sum up, models using intraday data such as IDR-DCC(-NL) and ID-DCC(-NL) outperform the traditional DCC(-NL) models using daily data only. Furthermore, IDR models, which use regularized returns in the second step, outperform their ID counterparts. As a visual addition to these tabulated results, [Fig. 4](#) displays boxplots of the various out-of-sample portfolio returns.

Remark 5.2 (Subperiod analysis). [Table 1](#) presents 'single' results over the entire out-of-sample period 12/18/1998–12/31/2018. It is natural to ask whether the relative performance of the various models is stable during that period or whether it changes during certain subperiods, such as periods of 'boom' or 'bust' compared to normal periods. To address this question, we carry out a rolling-window analysis based on shorter out-of-sample periods: one month (21 days), one year (252 days), and five years (1,260 days). The results are displayed in [Fig. 5](#), where any given number represents the out-of-sample standard deviation (SD) over the corresponding period ending on that day; the universe size is $N = 1000$ always. It can be seen that the relative performance is remarkably stable over time and that, in particular, IDR-DCC-NL generally performs best during all subperiods. Interestingly, the benefit of using intraday data instead of daily data, namely the reduction in SD from DCC(-NL) to IDR-DCC(-NL), is pretty much

⁸ As the out-of-sample size is very large at 5,040, there is no need to use the computationally more involved bootstrap method described in [Ledoit and Wolf \(2011, Section 3.2\)](#), which is preferred for small sample sizes.

⁹ With the single exception of $N = 1000$ where DCC-NL outperforms ID-DCC. Thus, for 'medium'-sized investment universes, $N = 100, 500$, even the proposed DCC models using intraday data outperform the traditional DCC-NL model. However, for larger dimensions, $N = 1000$, the benefit of nonlinear shrinkage is too important to neglect; see [Fig. 4](#).

⁷ The data for the riskfree rate are obtained from the website of Kenneth French.

Table 1

Annualized performance measures (in percent) for various estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; IR stands for information ratio; and SR stands for Sharpe ratio. All measures are based on 5,040 daily out-of-sample returns. In the rows labeled SD, the lowest number appears in **bold face**. In the columns labeled IDR-DCC respectively IDR-DCC-NL, significant outperformance over DCC respectively DCC-NL in terms of SD is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

Period: 12/18/1998–12/31/2018								
	EW	VW	DCC	ID-DCC	IDR-DCC	DCC-NL	ID-DCC-NL	IDR-DCC-NL
N=100								
AV	7.54	6.77	5.57	6.63	7.66	5.52	6.43	7.29
SD	18.94	18.70	12.45	12.24	12.18**	12.33	12.14	12.07**
IR	0.40	0.36	0.45	0.54	0.63	0.45	0.53	0.60
SR	0.31	0.27	0.31	0.40	0.49	0.31	0.39	0.46
N=500								
AV	10.05	7.81	6.94	5.88	6.62	7.75	6.91	7.04
SD	19.50	18.80	9.67	8.81	7.94***	9.01	8.19	7.44***
IR	0.52	0.42	0.72	0.67	0.83	0.86	0.84	0.95
SR	0.43	0.32	0.54	0.47	0.61	0.67	0.63	0.71
N=1000								
AV	10.91	8.07	8.33	6.20	6.08	8.32	6.76	6.29
SD	20.26	18.89	10.05	8.76	7.72***	7.88	6.85	5.90***
IR	0.54	0.43	0.83	0.71	0.79	1.06	0.99	1.07
SR	0.45	0.33	0.65	0.51	0.56	0.83	0.73	0.77

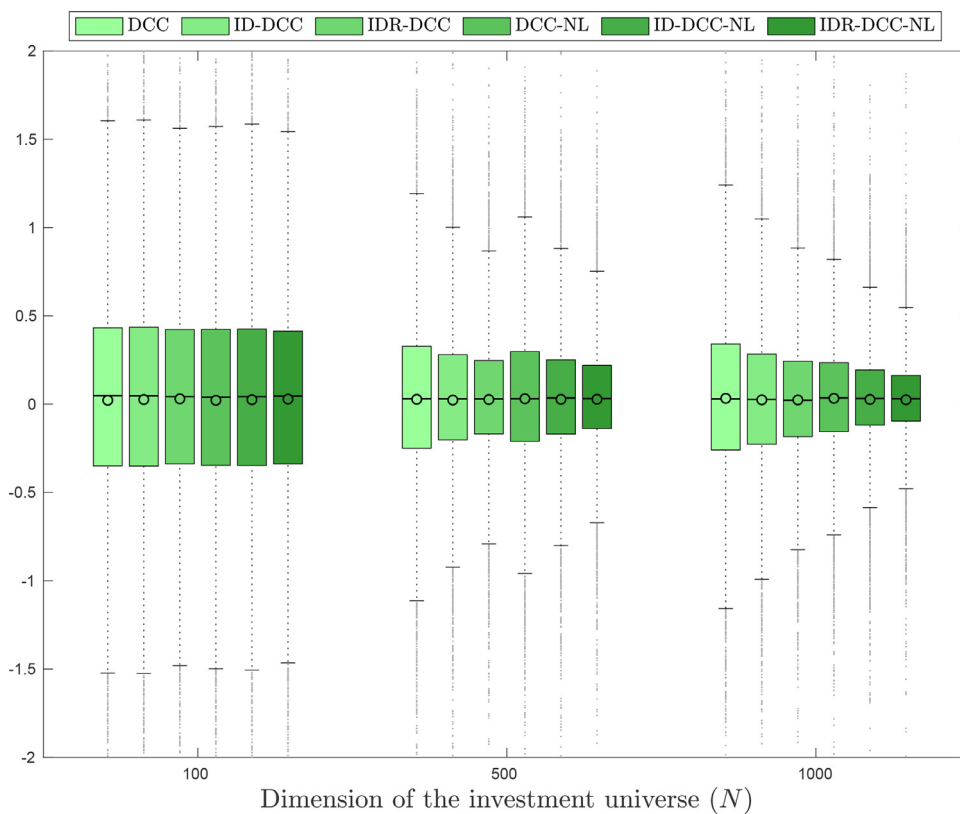


Fig. 4. Boxplots of the 5,040 daily out-of-sample returns (in percent) for various estimators of the GMV portfolio; the period is 12/18/1998–12/31/2018. The relative benefit of using intraday data (ID respectively IDR) and nonlinear shrinkage (NL) gets more pronounced for larger dimensions.

time-invariant: We do not find that the outperformance of intraday models is systematically different during periods of ‘boom’ or ‘bust’ compared to normal periods, which is a reassuring finding.

Inspired by Leippold and Svaton (2019), who discuss the measurement of idiosyncratic volatility, we also analyze whether events like the trade decimalization in April 2001, meaning how prices are reported on different exchanges, have a visible effect on the estimation of dynamic covariance matrices, based on daily and intradaily data. As our out-of-sample period starts in December 1998, we only have less than two years of data before the mentioned trade decimalization. Based on this short pre-event period at least, we do not observe any noticeable differences in the rel-

ative performance of the various models (before the event compared to after the event).

Remark 5.3 (Choice of κ). The two IDR models use regularized returns, which are based on shrinking the root of the volatility proxy to zero when the observed return is small. In practice, the amount of shrinkage is determined by the choice of the parameter κ of the scaled hyperbolic tangent function (3.2). Based on theoretical arguments, we had considered the range $\kappa \in [50, 200]$ to be reasonable and, specifically, had suggested the use of $\kappa = 100$. Robustness checks in Tables 3 and 4 provide empirical support for this suggestion. In particular, for the IDR-DCC-NL model, which is

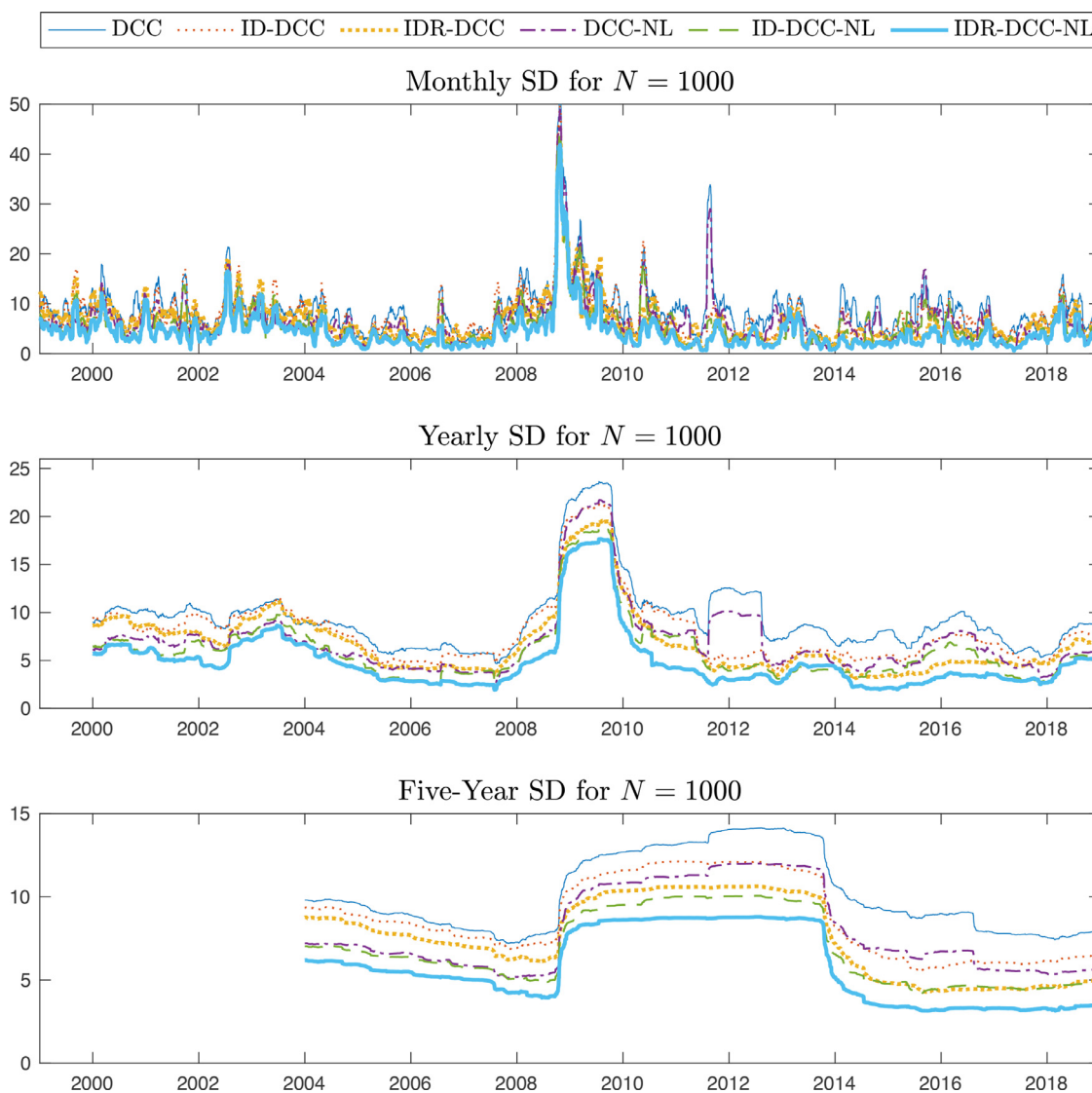


Fig. 5. Rolling-window out-of-sample standard deviations (SD) for various models and $N = 1000$. The lengths of the out-of-sample period are one month, one year, and five years, respectively. Any given number represents the out-of-sample standard deviation (SD) over the corresponding subperiod ending on that day.

our ultimate proposal, $\kappa = 100$ is uniformly the best of the three choices $\kappa \in \{50, 100, 200\}$, in terms of both SD and IR as the performance measure. But the other two choices, $\kappa \in \{50, 200\}$, also result in good performance and improve upon the naïve choice $\kappa = \infty$, that is, upon the standard sign function.

Remark 5.4 (Alternative models). Another way to improve upon the DCC-NL model, while only using daily returns, is to combine it with an approximate factor model. In particular, De Nard et al. (2021) suggest the AFM1-DCC-NL model which is based on an approximate single-factor model using the market factor. Based on unreported results, we find that the performance of this model (in terms of SD) is generally somewhere between the performances of DCC-NL and ID-DCC, and always below the performance of IDR-DCC-NL. The approximate factor model uses DCC-NL to model the dynamic covariance matrix of the residuals from a factor model. Although tempting at first sight, it is therefore not possible to ‘marry’ approximate factor models with our new proposal, since doing so would require intraday data for the residuals instead of the actual stocks, and such data is not available. Our recommendation is thus the following: If one only has access to daily

data, use AFM1-DCC-NL; but if one also has access to OHLC price data, use IDR-DCC-NL instead.

There are more complicated multivariate volatility models based on high-frequency data. For example, one can use all 5- or 10-minute returns during the day to compute realized covariance matrices and then use a sophisticated methodology, involving factor structure and regularization, to make corresponding forecasts. Using such an approach, Brito et al. (2018) obtain a reduction of 22.1% in SD (from 10.65 to a typical number of 8.3) compared to DCC-NL for a (constant) universe of $N = 430$ stocks; see their Table 5. Note that with our simpler IDR-DCC-NL model, we get a not much lower reduction in SD for a (time-varying) universe of $N = 500$ stocks, namely a reduction of 17.4% (from 9.01 to 7.44).¹⁰

Some further remarks are in order.

DeMiguel et al. (2009) claim that it is difficult to outperform EW in terms of the out-of-sample Sharpe ratio (SR) with ‘sophisti-

¹⁰ It makes sense to compare reduction in percentage points rather than nominal reduction, as the actual SD numbers of Brito et al. (2018) are not one-to-one comparable to ours because of different universes, different out-of-sample periods, and different portfolio-rebalancing frequency.

Table 2

Additional performance measures for various estimators of the GMV portfolio. TO stands for average turnover; GL stands for average gross leverage; PL stands for average proportion of leverage; \tilde{IR} stands for the information ratio net of transaction costs (5 bps); \tilde{SR} stands for the Sharpe ratio net of transaction costs (5 bps); and MDD stands for maximum drawdown (in percent). All measures are based on 240 monthly weight vectors. For the DCC models and the rows labeled TO, GL, PL, and MDD, the lowest (and thus best) number appears in **bold face**.

Period: 12/18/1998–12/31/2018								
	EW	VW	DCC	ID-DCC	IDR-DCC	DCC-NL	ID-DCC-NL	IDR-DCC-NL
N=100								
TO	0.11	0.03	2.84	3.06	2.44	2.64	2.87	2.26
GL	1.00	1.00	3.20	3.19	2.54	3.00	3.00	2.37
PL	0.00	0.00	0.47	0.46	0.45	0.46	0.46	0.45
\tilde{IR}	0.39	0.36	0.31	0.39	0.51	0.32	0.39	0.49
\tilde{SR}	0.30	0.27	0.17	0.25	0.37	0.18	0.24	0.35
MDD	55.47	52.70	37.47	36.64	35.34	36.79	36.33	35.08
N=500								
TO	0.10	0.02	4.56	4.47	3.83	3.05	3.01	2.57
GL	1.00	1.00	4.99	4.41	3.83	3.46	3.12	2.62
PL	0.00	0.00	0.49	0.49	0.49	0.51	0.50	0.51
\tilde{IR}	0.51	0.41	0.43	0.36	0.54	0.66	0.62	0.74
\tilde{SR}	0.42	0.32	0.25	0.17	0.32	0.46	0.40	0.50
MDD	56.51	53.34	32.28	28.59	29.59	25.21	23.93	23.68
N=1000								
TO	0.10	0.01	7.72	7.20	6.20	2.71	2.87	2.56
GL	1.00	1.00	7.90	6.80	6.07	3.07	2.83	2.54
PL	0.00	0.00	0.50	0.49	0.49	0.51	0.51	0.51
\tilde{IR}	0.54	0.43	0.37	0.21	0.22	0.85	0.73	0.77
\tilde{SR}	0.45	0.33	0.19	0.17	0.01	0.63	0.48	0.48
MDD	57.15	53.58	34.00	32.45	25.69	28.69	25.25	23.45

Table 3

Annualized performance measures (in percent) for various IDR-DCC estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; IR stands for information ratio; and SR stands for Sharpe ratio. All measures are based on 5,040 daily out-of-sample returns. In the rows labeled SD, the lowest number appears in **bold face**. In the column labeled $\kappa = 100$, significant outperformance over $\kappa = \infty$ in terms of SD is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

Period: 12/18/1998–12/31/2018					
IDR-DCC with					
	$\kappa = 2$	$\kappa = 50$	$\kappa = 100$	$\kappa = 200$	$\kappa = \infty$
N=100					
AV	5.46	7.06	7.66	7.00	6.94
SD	12.59	12.18	12.18	12.24	12.17
IR	0.43	0.58	0.63	0.57	0.57
SR	0.29	0.44	0.49	0.43	0.43
N=500					
AV	6.35	5.99	6.62	5.88	6.84
SD	9.45	8.13	7.94***	8.08	8.60
IR	0.67	0.74	0.83	0.73	0.79
SR	0.49	0.52	0.61	0.51	0.59
N=1000					
AV	5.98	5.51	6.08	6.26	7.23
SD	8.80	7.88	7.72***	8.12	8.73
IR	0.68	0.70	0.79	0.77	0.82
SR	0.48	0.48	0.56	0.56	0.63

cated' portfolios (that is, with Markowitz portfolios that estimate input parameters). However, contradicting this claim, it can be seen that all DCC models outperform EW (and also VW) in terms of SR.

Additionally, we report results on average turnover, gross leverage, and proportion of leverage, defined as follows.

- **TO:** We compute average (monthly) turnover as $\frac{1}{239} \sum_{k=1}^{239} \|\hat{w}_{k+1} - \hat{w}_k^{\text{hold}}\|_1$, where $\|\cdot\|_1$ denotes the L^1 norm

Table 4

Annualized performance measures (in percent) for various IDR-DCC-NL estimators of the GMV portfolio. AV stands for average; SD stands for standard deviation; IR stands for information ratio; and SR stands for Sharpe ratio. All measures are based on 5,040 daily out-of-sample returns. In the rows labeled SD, the lowest number appears in **bold face**. In the column labeled $\kappa = 100$, significant outperformance over $\kappa = \infty$ in terms of SD is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

Period: 12/18/1998–12/31/2018					
IDR-DCC-NL with					
	$\kappa = 2$	$\kappa = 50$	$\kappa = 100$	$\kappa = 200$	$\kappa = \infty$
N=100					
AV	5.35	6.78	7.29	6.69	6.64
SD	12.47	12.07	12.07	12.13	12.07
IR	0.43	0.56	0.60	0.55	0.55
SR	0.29	0.42	0.46	0.41	0.41
N=500					
AV	6.80	6.75	7.04	6.65	7.62
SD	9.03	7.69	7.44***	7.50	8.14
IR	0.75	0.88	0.95	0.89	0.94
SR	0.56	0.65	0.71	0.53	0.71
N=1000					
AV	6.25	6.11	6.29	6.65	7.15
SD	7.72	6.13	5.90***	6.18	6.84
IR	0.81	1.00	1.07	1.08	1.04
SR	0.58	0.71	0.77	0.77	0.77

and \hat{w}_k^{hold} denotes the vector of the 'hold' portfolio weights at the end of month k .¹¹

- **GL:** We compute average (monthly) gross leverage as $\frac{1}{240} \sum_{k=1}^{240} \|\hat{w}_k\|_1$.
- **PL:** We compute average (monthly) proportion of leverage as $\frac{1}{240 \times N} \sum_{k=1}^{240} \sum_{i=1}^N \mathbb{1}_{\{\hat{w}_{i,k} < 0\}}$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

¹¹ The vector \hat{w}_k^{hold} is determined by the initial vector of portfolio weights, \hat{w}_k , together with the evolution of the various prices of the N stocks in the portfolio during month k .

Table 5

Annualized performance measures (in percent) for various estimators of the constrained GMV portfolio, allowing for a gross leverage of at most 2.0. AV stands for average; SD stands for standard deviation; IR stands for information ratio; and SR stands for Sharpe ratio. All measures are based on 5,040 daily out-of-sample returns. In the rows labeled SD, the lowest number appears in **bold face**. In the columns labeled IDR-DCC respectively IDR-DCC-NL, significant outperformance over DCC respectively DCC-NL in terms of SD is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

Period: 12/18/1998–12/31/2018						
	DCC	ID-DCC	IDR-DCC	DCC-NL	ID-DCC-NL	IDR-DCC-NL
N=1000						
AV	7.64	6.43	5.81	8.07	6.69	6.24
SD	7.88	6.84	5.80***	7.77	6.67	5.73***
IR	0.97	0.94	1.00	1.04	1.00	1.09
SR	0.75	0.68	0.70	0.81	0.74	0.78

The results are presented in Table 2 and can be summarized as follows.

- IDR-DCC-NL leads to markedly lower turnover and gross leverage than all other models.
- IDR-DCC leads to the lowest proportion of leverage, but the differences across all DCC models are minor.
- Using regularized returns instead of observed returns in the estimation of the DCC(-NL) models (that is, using models with an IDR part instead of models with an ID part) reduces turnover and gross leverage, but has no noticeable effect on percentage of leverage.

As we have stated before, the point of the backtest exercise is not to evaluate realistic portfolio strategies but to compare the accuracy of estimators of the covariance matrix. Still, based on previous experience, there are always readers interested in additional measures of portfolio performance. The remainder of this section is dedicated to such readers.

First, we investigate the impact of gross-leverage (GL) constraints which are often used by quantitative investors. Specifically, we consider so-called 150-50 portfolios that allow for a total short position of at most 50% and thus a gross leverage of at most 2.0; for example, such portfolios are also considered by Zhao et al. (2022). Table 5 presents the corresponding results for N = 1000. The fact that imposing the GL constraint actually reduces SD for models without a NL part quite a bit whereas there is almost no difference for models with a NL part, is consistent with the findings of Zhao et al. (2022, Table 2). IDR-DCC-NL remains the best model in the end, though now its performance is basically matched by IDR-DCC.

Second, it may be of interest to see whether the outperformance of DCC models over EW and VW in terms of the information ratio survives transactions costs. To this end, we compute the information ratio net of transaction costs, called \tilde{IR} , where we assume a constant transaction cost of 5 bps for all stocks.¹² Arguably, such a comparison is not overly meaningful, since any real-life portfolio managers would take transaction costs into account at the portfolio-selection stage, which is expected to result in better ‘net’ performance compared to ignoring transaction costs at the selection stage and just paying them afterwards. Nevertheless, Table 2 demonstrates that all DCC portfolios with a NL part still outperform EW and VW in terms of \tilde{IR} for N = 500, 1000. On the other hand, the DCC portfolios without a NL part, which incur higher turnover compared to their NL counterparts, often underperform EW and VW.

¹² Many academic studies consider higher transaction costs, up to 50 bps, but 5 bps is a typical number for large stocks this day and age.

Third, we compute the maximum drawdown, called **MDD**. Compared to EW and VW, the DCC models achieve more favorable (that is, smaller) numbers and IDR-DCC-NL is again best. For example, for N = 1000, the maximum drawdown for EW is 57.15% whereas for IDR-DCC-NL it is only 23.45%, less than half.

As a visual addition to these tabulated results, Fig. 6 displays the trajectories of the portfolios EW, DCC-NL, and IDR-DCC-NL for N = 1000; we restrict attention to these three portfolios in order to not over-crowd the figure. EW achieves the highest value at the end but suffers from periods of extreme turmoil and a maximum drawdown of nearly 60%. DCC-NL achieves the second-highest value at the end but is also more volatile compared to IDR-DCC-NL.

5.5. Markowitz portfolio with momentum signal

We next turn attention to a ‘full’ Markowitz portfolio with a signal, where the term ‘signal’ stands for an estimate of

$$\mu_t := \mathbb{E}(r_t | \mathcal{F}_{t-1}).$$

By now a large number of variables have been documented that can be used to construct a signal in practice. For simplicity and reproducibility, we use the well-known momentum factor (or simply momentum for short) of Jegadeesh and Titman (1993). For a given investment period k and a given stock, the momentum is defined as the geometric average of the previous 252 returns on the stock but excluding the most recent 21 returns; in other words, one uses the geometric average over the previous ‘year’ but excluding the previous ‘month’. Collecting the individual momentums of all the N stocks contained in the portfolio universe yields the return-predictive signal m_t .

In the absence of short-sales constraints, the investment problem is formulated as

$$\min_w w' \Sigma_t w \tag{5.5}$$

$$\text{subject to } w' m_t = b \text{ and} \tag{5.6}$$

$$w' \mathbb{1} = 1 \tag{5.7}$$

where b is a selected target expected return. The problem has the analytical solution

$$w = c_1 \Sigma_t^{-1} \mathbb{1} + c_2 \Sigma_t^{-1} m_t \tag{5.8}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \text{ and } c_2 := \frac{bA - B}{AC - B^2} \tag{5.9}$$

$$\text{with } A := \mathbb{1}' \Sigma_t^{-1} \mathbb{1} \quad B := \mathbb{1}' \Sigma_t^{-1} m_t \quad \text{and } C := m_t' \Sigma_t^{-1} m_t. \tag{5.10}$$

The natural strategy in practice is to replace the unknown covariance matrix Σ_t by an estimator $\hat{\Sigma}_t$ in formulas (5.8)–(5.10), yielding the feasible portfolio

$$\hat{w} := c_1 \hat{\Sigma}_t^{-1} \mathbb{1} + c_2 \hat{\Sigma}_t^{-1} m_t \tag{5.11}$$

$$\text{where } c_1 := \frac{C - bB}{AC - B^2} \text{ and } c_2 := \frac{bA - B}{AC - B^2} \tag{5.12}$$

$$\text{with } A := \mathbb{1}' \hat{\Sigma}_t^{-1} \mathbb{1} \quad B := \mathbb{1}' \hat{\Sigma}_t^{-1} m_t \quad \text{and } C := m_t' \hat{\Sigma}_t^{-1} m_t. \tag{5.13}$$

In addition to Markowitz portfolios based on formulas (5.11)–(5.13), we also include as a simple-minded benchmark the equal-weighted portfolio of the top-quintile stocks. This portfolio is obtained by sorting the stocks, from lowest to highest, according to momentum and then putting equal weight on all the stocks in the top 20%, that is, in the top quintile. We call this portfolio **EW-TQ**.

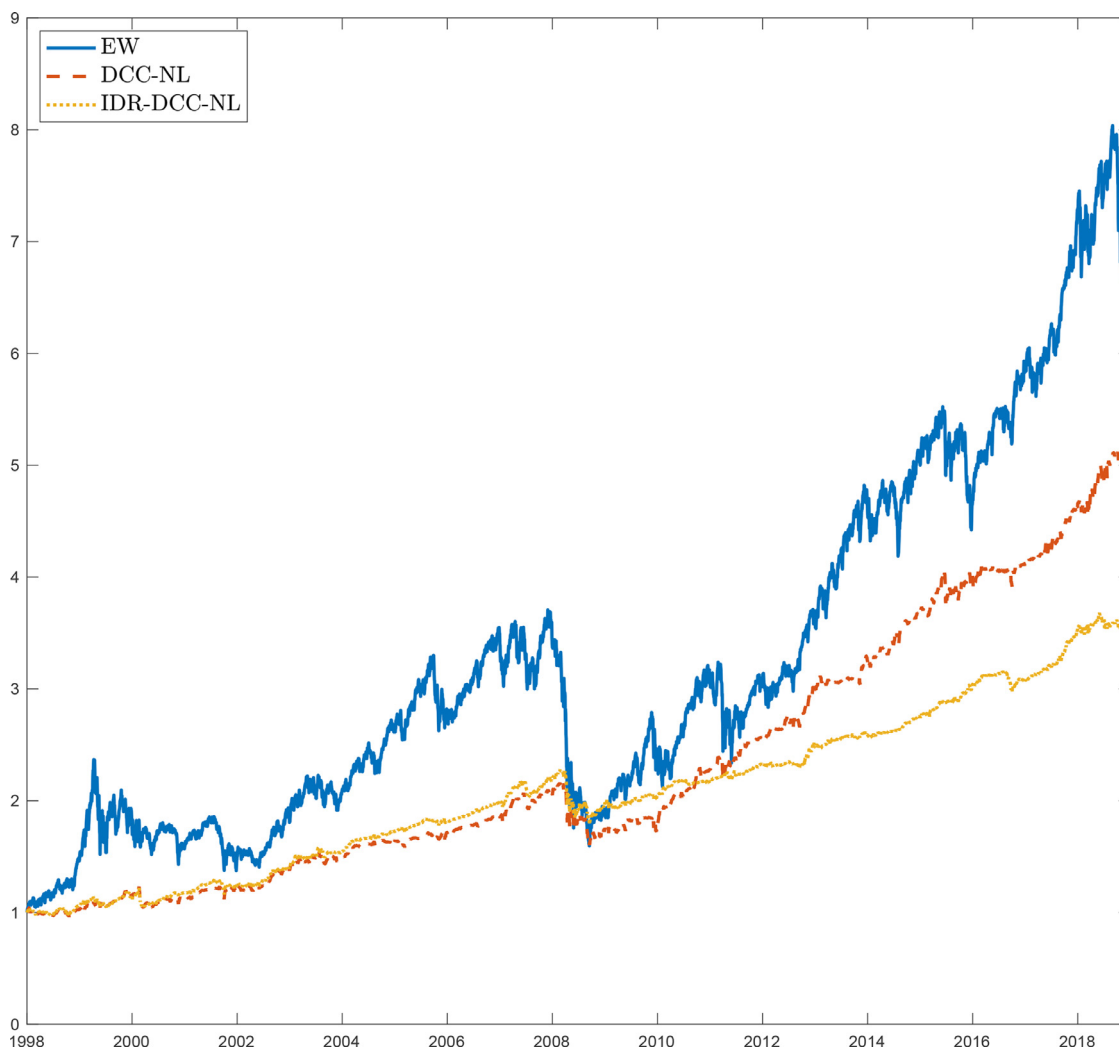


Fig. 6. Trajectories of three estimated GMV portfolio for $N = 1000$.

A related strategy, called **VW-TQ**, uses value-weighting instead of equal-weighting for the stocks in the top 20%.

Our stance is that in the context of a ‘full’ Markowitz portfolio, the most important performance measure is the out-of-sample information ratio (IR). In the investment problem (5.5)–(5.7), minimizing the variance (for a fixed target expected return b) is equivalent to maximizing the information ratio if the true expected return is equal to the target. In practice, because of the estimation error in m_t relative to μ_t , different strategies do not necessarily have the same expected return. As a consequence, rankings based on the standard deviation may not coincide with rankings based on the information ratio.

Remark 5.5 (Information ratio vs. Sharpe ratio). For completeness, the tables also report the out-of-sample Sharpe ratio (SR) in addition to the out-of-sample information ratio (IR). But as the performance criterion, we focus on the information ratio for two reasons. First, the feasible portfolio (5.11) is based on actual returns instead of excess returns, since we use actual returns to compute the two inputs $\hat{\Sigma}_t$ and m_t , as is customary with real-life portfolio managers. It would then appear inconsistent to base the performance measure on excess returns instead of actual returns (when dividing the average return by the standard deviation of the returns); after all, our empirical analysis has to be a competition about who has the ‘strongest muscle’ (that is, the most accurate estimator of the covariance matrix) on a level playing field. Second, if one advocates

that out-of-sample returns should be benchmarked, then a suitable benchmark for a fully invested portfolio (that is, a portfolio for which \hat{w}_t sums up to one) is, arguably, a stock market index, such as the S&P 500, rather than the riskfree rate.

We also study whether IDR-DCC(-NL) delivers a higher out-of-sample information ratio than DCC(-NL) with statistical significance; so there are two comparisons again just as in Section 5.4. For a given universe size, a two-sided p -value for the null hypothesis of equal information ratios is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2008, Section 3.1).¹³

The results are presented in Table 6 and can be summarized as follows; unless stated otherwise, the findings are with respect to IR as performance measure.

- All models outperform EW-TQ and VW-TQ by a wide margin.
- Each DCC-NL model outperforms its DCC counterpart.
- IDR-DCC-NL outperforms all other models.
- Having said that, we do not find statistical significance (for the two comparisons considered).

¹³ As the out-of-sample size is very large at 5,040, there is no need to use the computationally more expensive bootstrap method described in Ledoit and Wolf (2008, Section 3.2), which is preferred for small sample sizes.

Table 6

Annualized performance measures (in percent) for various estimators of the Markowitz portfolio with momentum signal. AV stands for average; SD stands for standard deviation; IR stands for information ratio; and SR stands for Sharpe ratio. All measures are based on 5,040 daily out-of-sample returns. In the rows labeled IR, the largest number appears in **bold face**. In the columns labeled IDR-DCC respectively IDR-DCC-NL, significant outperformance over DCC respectively DCC-NL in terms of SD and IR (separately) is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

Period: 12/18/1998–12/31/2018								
	EW-TQ	VW-TQ	DCC	ID-DCC	IDR-DCC	DCC-NL	ID-DCC-NL	IDR-DCC-NL
N=100								
AV	9.75	9.42	9.41	10.23	10.76	9.63	10.38	10.69
SD	23.49	23.17	14.89	14.71	14.72	14.77	14.61	14.65
IR	0.42	0.41	0.63	0.70	0.73	0.65	0.71	0.73
SR	0.34	0.33	0.51	0.58	0.61	0.53	0.59	0.61
N=500								
AV	11.82	10.21	7.77	6.28	7.03	8.80	7.56	7.80
SD	22.97	23.04	10.81	9.96	9.12***	10.13	9.29	8.58***
IR	0.51	0.44	0.72	0.63	0.77	0.87	0.81	0.91
SR	0.44	0.37	0.56	0.46	0.58	0.70	0.63	0.71
N=1000								
AV	12.02	10.33	8.36	6.34	6.03	8.51	6.74	6.54
SD	23.66	23.23	11.32	10.02	8.73***	8.64	7.52	6.54***
IR	0.51	0.44	0.74	0.63	0.69	0.98	0.90	1.00
SR	0.43	0.37	0.58	0.46	0.49	0.78	0.66	0.73

- Moreover, in terms of SD as performance measure, we find statistical significance (for the two comparisons considered) for $N = 500, 1000$, with IDR-DCC-NL again being the best model.

Some further remarks are in order.

DeMiguel et al. (2009) claim that it is difficult to outperform EW in terms of the out-of-sample Sharpe ratio (SR) with 'sophisticated' portfolios (that is, with Markowitz portfolios that estimate input parameters). However, comparing with Table 1, it can be seen that all DCC strategies using the (simple-minded) momentum signal outperform EW. Even though momentum is not a very powerful return-predictive signal, the differences compared to EW can be substantial. For example, for $N = 1000$, the Sharpe ratio of EW is only 0.45 whereas the Sharpe ratio of IDR-DCC-NL is 0.73, more than 60% larger.

Engle and Colacito (2006) argue for the use of the out-of-sample standard deviation, SD, as a performance measure also in the context of a 'full' Markowitz portfolio. It can be seen that

according to this performance all models outperform EW-TQ and VW-TQ by a wide margin as well, with IDR-DCC-NL being always best again.

Additionally, we report results on average turnover, gross leverage, and percentage of leverage. The results are presented in Table 7 and can be summarized as follows. (The summary is *verbatim* identical to the one of Section 5.4.)

- IDR-DCC-NL leads to markedly lower turnover and gross leverage than all other models.
- IDR-DCC leads to the lowest proportion of leverage, although the differences across all DCC models are minor.
- Using regularized returns instead of observed returns in the estimation of the DCC(-NL) models (that is, using models with an IDR part instead of models with an ID part) reduces turnover and gross leverage, but has no noticeable effect on percentage of leverage.

Table 7

Additional performance measures for various estimators of the Markowitz portfolio with momentum signal. TO stands for average turnover; GL stands for average gross leverage; PL stands for average proportion of leverage; \tilde{IR} stands for the information ratio net of transaction costs (5 bps); \tilde{SR} stands for the Sharpe ratio net of transaction costs (5 bps); and MDD stands for maximum drawdown (in percent). All measures are based on 240 monthly weight vectors. For the DCC models and the rows labeled TO, GL, PL, and MDD, the lowest (and thus best) number appears in **bold face**, whereas for the row labeled \tilde{IR} , the highest (and thus best) number appears in **bold face**.

Period: 12/18/1998–12/31/2018								
	EW-TQ	VW-TQ	DCC	ID-DCC	IDR-DCC	DCC-NL	ID-DCC-NL	IDR-DCC-NL
N=100								
TO	0.58	0.58	3.40	3.54	2.86	3.17	3.34	2.66
GL	1.00	1.00	3.90	3.87	3.14	3.67	3.65	2.95
PL	0.00	0.00	0.47	0.47	0.46	0.47	0.47	0.46
\tilde{IR}	0.40	0.39	0.50	0.55	0.61	0.52	0.57	0.62
\tilde{SR}	0.33	0.32	0.38	0.43	0.50	0.41	0.45	0.50
MDD	56.85	58.29	37.16	35.99	32.86	36.05	34.49	32.63
N=500								
TO	0.56	0.59	5.04	4.89	4.19	3.41	3.43	2.84
GL	1.00	1.00	5.68	5.04	4.35	3.99	3.59	3.02
PL	0.00	0.00	0.50	0.50	0.50	0.50	0.50	0.51
\tilde{IR}	0.50	0.43	0.53	0.42	0.50	0.68	0.60	0.71
\tilde{SR}	0.42	0.35	0.37	0.24	0.30	0.51	0.42	0.51
MDD	57.24	51.46	32.38	29.42	27.61	23.17	21.15	21.01
N=1000								
TO	0.56	0.60	8.99	8.33	6.95	3.05	3.24	2.83
GL	1.00	1.00	9.58	8.20	6.98	3.59	3.33	2.93
PL	0.00	0.00	0.50	0.50	0.50	0.51	0.51	0.51
\tilde{IR}	0.49	0.43	0.56	0.42	0.21	0.75	0.63	0.76
\tilde{SR}	0.42	0.35	0.40	0.25	0.02	0.56	0.42	0.47
MDD	56.97	51.37	29.31	28.26	23.69	25.94	22.54	21.04

As further demonstrated in Table 7, all DCC models with a NL part outperform EW-TQ and VW-TQ in terms of the information ratio net of transaction costs ($\tilde{\mathbf{IR}}$), where again a constant transaction cost of 5 bps is assumed for all stocks.

Last but not least, compared to EW-TQ and VW-TQ, all DCC models are able to achieve more favorable (that is, smaller) numbers in terms of maximum drawdown and IDR-DCC-NL is again best. For example, for $N = 1000$, the maximum drawdown for EW-TQ is 56.97% whereas for IDR-DCC-NL it is only 21.04%, down to nearly a third.

6. Conclusion

In this paper we have shown that using intraday data, in the form of open/high/low/close (OHLC) prices, leads to improved forecasts of dynamic covariance matrices via multivariate GARCH models, where our focus has been on the original DCC model of Engle (2002) and its recent extension, the DCC-NL model of Engle et al. (2019).

The first step of a DCC(-NL) model consists of modeling dynamic (univariate) variances via a GARCH(1,1) model where the innovation is the daily squared return. Hence, the first idea is to use an improved volatility proxy based on OHLC data, instead of the squared return, in this step. We call the resulting multivariate GARCH models ID-DCC(-NL), where “ID” stands for “volatility proxy based on IntraDay data”.

The second step of a DCC(-NL) model consists of modeling the dynamic correlation matrix where the innovation now is the outer product of the vector of daily (devolitized) returns. Hence, the second idea is to use an improved innovation that is based on volatility proxies instead of returns. The counterpart of a volatility proxy itself is the squared return. Thus, it is natural to take the sign of the return together with the square root of the volatility proxy as the counterpart of the return itself. This already works well, but a ‘smoothed’ sign of the return together with the root of the volatility proxy works even better, which is the new concept of a *regularized return*. Using these regularized returns in the second step, and their squares in the first step (for reasons of internal consistency), gives rise to multivariate GARCH models that we call IDR-DCC(-NL) models, where “R” stands for “regularized returns”.

An important feature of our newly proposed models is that they remain computationally feasible for universes of $N \geq 1000$ assets, unlike most other multivariate GARCH-type models (especially when based on intraday data).

Empirical backtest exercises demonstrate that ID-DCC(-NL) models already deliver a substantial improvement over DCC(-NL) models, and that IDR-DCC(-NL) models deliver another improvement of roughly the same magnitude. In particular, the IDR-DCC-NL model is the clear winner and its performance for large investment universes is rather impressive; for example, the out-of-sample standard deviation of an estimated global-minimum-variance (GMV) portfolio of $N = 1000$ stocks improves from 7.88% to 5.90% (on an annual basis) when one upgrades from DCC-NL to IDR-DCC-NL. A further advantage of the IDR-DCC-NL model is that it leads to reduced turnover, reduced gross leverage, and reduced maximum drawdown compared to the DCC-NL model.

Appendix A. Detailed description of IDR-DCC(-NL) models

First, compute the regularized returns for a suitable scaling factor κ ¹⁴

$$\tilde{r}_{i,t-1} := \text{stanh}(r_{i,t-1}, \kappa) \sqrt{\hat{v}_{i,t-1}} \quad (\text{A.1})$$

¹⁴ We suggest the choice $\kappa = 100$ for returns in percent, which corresponds to the choice $\kappa = 10,000$ for raw returns.

where $\text{stanh}(\cdot, \cdot)$ is the ‘scaled’ hyperbolic tangent function defined in Eq. (3.2) and $\hat{v}_{i,t-1}$ denotes a generic volatility proxy based on OHLC price data; see Section 4 for specific proposals.

Second, for modeling conditional covariances, use a GARCH(1,1) model with squared regularized returns, instead of squared returns, as innovations:

$$d_{i,t}^2 = \omega_i + \delta_{1,i} \tilde{r}_{i,t-1}^2 + \delta_{2,i} d_{i,t-1}^2 \quad (\text{A.2})$$

where $(\omega_i, \delta_{1,i}, \delta_{2,i})$ are the asset-specific GARCH(1,1) parameters. Then use the conditional variances to devolitize the regularized returns: $s_{i,t} := \tilde{r}_{i,t}/d_{i,t}$.

Third, for modeling the conditional (pseudo) correlation matrix use the DCC model with correlation targeting:

$$Q_t = (1 - \delta_1 - \delta_2)C + \delta_1 s_{t-1} s'_{t-1} + \delta_2 Q_{t-1} \quad (\text{A.3})$$

where (δ_1, δ_2) are the DCC parameters and $C := \text{Cor}(\tilde{r}_t) = \text{Cov}(s_t)$ denotes the unconditional correlation matrix of the regularized returns. Note that Q_t cannot be used directly because its diagonal elements, although close to one, are typically not exactly equal to one. Therefore, the conditional correlation matrix, and from it the conditional covariance matrix, are obtained as follows:

$$R_t := \text{Diag}(Q_t)^{-1/2} Q_t \text{Diag}(Q_t)^{-1/2} \quad (\text{A.4})$$

$$\Sigma_t := D_t R_t D_t \quad (\text{A.5})$$

Finally, if the portfolio is held for more than one day use the ‘averaged-forecasting’ approach of De Nard et al. (2021): At portfolio-formation date k , forecast the covariance matrix for all L days of the holding period, that is, for $t = k, k + 1, \dots, k + (L - 1)$; then average those L forecasts and use this ‘averaged’ forecast to construct the portfolio at date k ; see Section 2.2:

$$\hat{\Sigma}_k := \frac{1}{L} \sum_{l=0}^{L-1} \mathbb{E}[D_{k+l} | \mathcal{F}_{k-1}] \mathbb{E}[R_{k+l} | \mathcal{F}_{k-1}] \mathbb{E}[D_{k+l} | \mathcal{F}_{k-1}]. \quad (\text{A.6})$$

In practice, the parameters in Eqs. (A.2) and (A.3) need to be estimated. To this end, follow the same prescriptions as in Section 2.3, just with regularized returns \tilde{r}_t in place of returns r_t . In particular, if the correlation-targeting matrix C of Eq. (A.2) is estimated using the sample covariance matrix of the devolitized regularized-returns $\{s_t\}$, the IDR-DCC model obtains; if instead C is estimated by applying nonlinear shrinkage to the $\{s_t\}$, with post-processing analogous to (A.4), the IDR-DCC-NL model obtains.

CRedit authorship contribution statement

Gianluca De Nard: Conceptualization, Methodology, Investigation, Software, Writing – original draft. **Robert F. Engle:** Conceptualization, Methodology. **Olivier Ledoit:** Conceptualization, Methodology, Investigation, Software, Writing – original draft. **Michael Wolf:** Conceptualization, Methodology, Investigation, Software, Writing – original draft.

References

- Archakov, I., Hansen, P.R., Lunde, A., 2020. A Multivariate Realized GARCH Model. Technical Report.
- Baillie, R.T., Bollerslev, T., 1992. Prediction in dynamic models with time-dependent conditional variances. *J. Econom.* 52, 91–113.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: trades and quotes. *Econom. J.* 119, C1–C32.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *J. Polit. Economy* 81, 637–659.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* 31, 307–327.
- Brito, D., Medeiros, M.C., Ribeiro, R., 2018. Forecasting Large Realized Covariance Matrices: The Benefits of Factor Models and Shrinkage. Technical Report. Department of Economics, PUC-Rio. Available at <https://ssrn.com/abstract=3163668>.

- Christoffersen, P., Diebold, F., 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Manage. Sci.* 52 (8), 1273–1287.
- De Nard, G., Ledoit, O., Wolf, M., 2021. Factor models for portfolio selection in large dimensions: the good, the better and the ugly. *J. Financ. Econom.* 19 (2), 236–257.
- DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Rev. Financ. Stud.* 22, 1915–1953.
- Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Engle, R.F., 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.* 20 (3), 339–350.
- Engle, R.F., Colacito, R., 2006. Testing and valuing dynamic correlations for asset allocation. *J. Bus. Econ. Stat.* 24 (2), 238–253.
- Engle, R.F., Ledoit, O., Wolf, M., 2019. Large dynamic covariance matrices. *J. Bus. Econ. Stat.* 37 (2), 363–375.
- Engle, R.F., Sheppard, K., 2001. Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH. Working Paper 8554. National Bureau of Economic Research.
- Fiszeder, P., Faldziński, M., Molnár, P., 2019. Range-based DCC models for covariance and value-at-risk forecasting. *J. Empir. Finance* 54, 58–76.
- Garman, M.B., Klass, M.J., 1980. On the estimation of security price volatilities from historical data. *J. Bus.* 53 (1), 67–78.
- Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *J. Appl. Econom.* 20, 873–889.
- Haugen, R.A., Baker, N.L., 1991. The efficient market inefficiency of capitalization-weighted stock portfolios. *J. Portf. Manage.* 17 (3), 35–40.
- Henriksson, R., Merton, R., 1981. On market timing and investment performance. *J. Bus.* 54 (4), 513–533.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. *J. Finance* 54 (4), 1651–1684.
- Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *J. Finance* 48 (1), 65–91.
- Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. *J. Empir. Finance* 15, 850–859.
- Ledoit, O., Wolf, M., 2011. Robust performance hypothesis testing with the variance. *Wilmott Mag.* September, 86–89.
- Ledoit, O., Wolf, M., 2015. Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivar. Anal.* 139 (2), 360–384.
- Ledoit, O., Wolf, M., 2017. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* 30 (12), 4349–4388.
- Ledoit, O., Wolf, M., 2020. Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Ann. Stat.* 40 (5), 3043–3065.
- Leippold, M., Svaton, M., 2019. Trend and Reversal of Idiosyncratic Volatility Revisited. Technical Report. Available at <https://ssrn.com/abstract=3321678>.
- Molnár, P., 2016. High-low range in GARCH models of stock return volatility. *Appl. Econ.* 48 (51), 4977–4991.
- Nielsen, F., Aylursubramanian, R., 2008. Far From the Madding Crowd – Volatility Efficient Indices. Research Insights. MSCI Barra.
- Noureddin, D., Shephard, N., Sheppard, K., 2012. Multivariate high-frequency-based volatility (HEAVY) models. *J. Appl. Econom.* 27, 907–933.
- Pakel, C., Shephard, N., Sheppard, K., Engle, R.F., 2021. Fitting vast dimensional time-varying covariance models. *J. Bus. Econ. Stat.* 39 (3), 652–668.
- Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. *J. Bus.* 53 (1), 61–65.
- Pesaran, H., Timmermann, A., 1995. Predictability of stock returns: robustness and economic significance. *J. Finance* 50 (4), 1201–1228.
- Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *J. Finance* 39 (4), 1127–1139.
- Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financ. Stud.* 21 (4), 1455–1508.
- Xu, Y., 2019. DCC-HEAVY: A Multivariate GARCH Model with Realized Measures of Variance and Correlation. Cardiff Economics Working Paper. Cardiff Business School.
- Yang, D., Zhang, Q., 2000. Drift-independent volatility estimation based on high, low, open, and close prices. *J. Bus.* 73 (3), 477–492.
- Zhao, Z., Ledoit, O., Jiang, H., 2022. Risk reduction and efficiency increase in large portfolios: Leverage and shrinkage. *J. Financ. Econom.*. Forthcoming.