The variance implied conditional correlation

Andres Algaba[a,*], Kris Boudt[a,b,c] and Steven Vanduffel[a]

[a]Solvay Business School, Vrije Universiteit Brussel, Belgium
[b]School of Business and Economics, Vrije Universiteit Amsterdam, The Netherlands
[c]Department of Economics, Ghent University, Belgium

ARTICLE HISTORY
Compiled August 28, 2019

ABSTRACT
We apply univariate GARCH models to construct a computationally simple filter for estimating the conditional correlation matrix of asset returns. The proposed Variance Implied Conditional Correlation (VICC) exploits the polarization result that links the correlation between two standardized variables with the variances of linear combinations thereof. In a Monte Carlo study, we show that the VICC yields accurate correlation estimates for common choices of the correlation dynamics. We also provide an empirical application to cross hedging that confirms the effectiveness of the VICC.

KEYWORDS
conditional correlation; cross hedging; Dynamic Conditional Correlation (DCC); GARCH; hedge ratio; regularization.

JEL CLASSIFICATION
C10, G11

"Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration."

- George E. P. Box, 1976.

1. Introduction

Financial markets are inherently multidimensional. In this context, portfolio risk involves not only the volatility of asset returns, but also the correlations among them. The latter can be used for managing diversification and hedging purposes, and are thus an important element of financial planning for any investor. Due to the dynamic nature of the comovement between assets, the main difficulty lies in obtaining timely conditional estimates of the correlation between the asset returns.

Typically, conditional correlations are jointly estimated via a multivariate modelling approach, such as Multivariate Generalized AutoRegressive Conditional Heteroskedasticity (MGARCH) models. This approach usually requires the optimization of a multivariate likelihood function, which can be numerically challenging and can lead to

* Corresponding author. Andres Algaba, Pleinlaan 2, 1050 Brussel, Belgium. Email: andres.algaba@vub.be.

parameter instability in the case of a general parametrization (Bauwens, Laurent, and Rombouts 2006). Moreover, most types of MGARCH models suffer from the so-called curse of dimensionality which makes their use in practice often infeasible. More constrained versions of the MGARCH models, like the Dynamic Conditional Correlation (DCC) model of Engle (2002) or the Generalized Orthogonal GARCH (GO-GARCH) model of van der Weide (2002), are computationally more convenient, but they may be too restrictive in terms of the conditional correlation dynamics that they can accommodate (see e.g., Caporin and McAleer, 2013).

In this paper, we present a simple and flexible filter for estimating the pairwise conditional correlations among variables. The proposed Variance Implied Conditional Correlation (VICC) exploits the polarization result that links the correlation between two standardized variables with the variances of linear combinations thereof. The VICC thus only requires the estimation of *univariate* variance models. There is a rich variety of well-studied univariate variance models available, namely the class of univariate GARCH models, which offer a lot of flexibility in modelling the univariate variance process, while only requiring the optimization of a univariate likelihood function. Furthermore, by using a pairwise estimation method, the VICC avoids the curse of dimensionality which most types of MGARCH models suffer from.

The VICC is referred to as a filter since we do not ambition efficiency under a particular model specification, but rather aim for reliability and accuracy in terms of correlation estimates under a wide set of possible models for the conditional correlation dynamics. The use of a filter rather than a fully specified MGARCH model is consistent with the statement of George E. P. Box (1976) that "Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration." It is also consistent with the view of Caporin and McAleer (2013) that, as the exact specification of the conditional correlation is unknown, conditional correlation models should be considered as filters for obtaining reliable estimates of the conditional correlations, even if they arise through possible model misspecification.

We expect that, since the VICC only requires the optimization of univariate likelihood functions, it will lead to more stable and reliable conditional correlation estimates when comparing with the more complex MGARCH models. To assess the reliability of the conditional correlation estimates we perform a Monte Carlo study in which the VICC is misspecified for all the considered correlation processes. However, we indeed find that the VICC yields accurate correlation estimates for common choices of the correlation dynamics.

We study the usefulness of the VICC to determine the portfolio allocation needed when cross hedging the weekly S&P 500 return using weekly returns of futures on the interest rate, the exchange rate between the US Dollar and the Euro, and the VIX. We conclude that VICC-based cross hedging performs at least as good as cross hedging using the DCC and GO-GARCH model in terms of variance reduction and achieved decorrelation, while generating a lower turnover and being more simple to compute. Cross heding by using the VICC also outperforms the the approach of cross hedging using an unconditional hedge ratio and an Exponentially Weighted Moving Average (EWMA) covariance model in terms of variance reduction and achieved decorrelation.

The remainder of the paper is organised as follows. Section 2 introduces the Variance Implied Conditional Correlation (VICC) filter and discusses its properties. Section 3 presents a Monte Carlo study to evaluate the performance of the VICC compared to some benchmark models. In Section 4, we assess the usefulness of the VICC in an empirical cross hedging application. Section 5 concludes.

2. Variance implied conditional correlation

We first present the well-known polarization result for the unconditional correlation estimator and extend this result to the conditional case. Next, we use this result to construct the Variance Implied Conditional Correlation (VICC) pairwise correlation filter. We then elaborate on some further properties of the VICC correlation matrix filter. Finally, we use news impact surfaces to analyze the responsiveness of the VICC-based correlation to the standardized return innovations.

2.1. *Definitions*

The polarization result connects the variances of a sum of random variables (possibly standardized) and their difference to their covariance. The result is often used in robust statistics and high-frequency financial econometrics. For example, Gnanadesikan and Kettenring (1972) use it to obtain robust estimates of the unconditional correlation, while Ma and Genton (2002) exploit the polarization result to robustly estimate the autocovariance function. More recently, Aït-Sahalia, Fan, and Xiu (2010) estimate a high-frequency data based realized covariance via the polarization result. We work with standardized data, henceforth we focus on variables with zero mean and unit variance. Our first result pertains to the iid case.

Property 1. *Let Z_i and Z_j be bivariate standard normally distributed with correlation coefficient ρ_{ij}. Assume there are T observations of Z_i and Z_j collected in samples z_i and z_j, i.e.,*

$$z_i = (z_{i,1}, ..., z_{i,T}) \quad \text{and} \quad z_j = (z_{j,1}, ..., z_{j,T}).$$

Denote the sample variances of $Z_i + Z_j$ and $Z_i - Z_j$ by $\widehat{h}_{i+j}$ and $\widehat{h}_{i-j}$, respectively. We then have that $\widehat{\rho}_{ij}$ defined as:

$$\widehat{\rho}_{ij} = \frac{\widehat{h}_{i+j} - \widehat{h}_{i-j}}{\widehat{h}_{i+j} + \widehat{h}_{i-j}}$$

is a consistent estimator for ρ_{ij}. Moreover, it has the lowest asymptotic variance (i.e., it is the most efficient) among the following class of estimators,

$$\widehat{\rho}_{ij}(\gamma) = \frac{\widehat{h}_{i+j}(\gamma) - \widehat{h}_{i-j}(\gamma)}{\widehat{h}_{i+j}(\gamma) + \widehat{h}_{i-j}(\gamma)},$$

where $\widehat{h}_{i+j}(\gamma)$ and $\widehat{h}_{i-j}(\gamma)$ denote the sample variances of $\gamma Z_i + (1-\gamma)Z_j$ and $\gamma Z_i - (1-\gamma)Z_j$, respectively, with $\gamma \in [0, 1]$.

The consistency of $\widehat{\rho}_{ij}$ follows directly from the law of large numbers. We prove the asymptotic efficiency result in Appendix A.

For most asset returns, there is overwhelming evidence of time-variation in their comovement. To exploit this feature, we use time series models in which we denote two time series processes by $\{Z_{i,s}\}$ and $\{Z_{j,s}\}$, whereby the relevant information set to predict their future comovement changes on each date. The following property can then be used to obtain a conditional correlation estimate.

Property 2. *Let* $\{Z_{i,s}\}$ *and* $\{Z_{j,s}\}$ *be two stochastic time series processes with* $s \leq t-1$*. Assume that conditionally on the information set* $\mathcal{F}_{t-1}$*,* $Z_{i,t}$ *and* $Z_{j,t}$ *have mean zero, unit variance, and correlation equal to* $\rho_{ij,t}$*. We then have that*

$$\rho_{ij,t} = \frac{\mathbb{E}[(Z_{i,t} + Z_{j,t})^2|\mathcal{F}_{t-1}] - \mathbb{E}[(Z_{i,t} - Z_{j,t})^2|\mathcal{F}_{t-1}]}{\mathbb{E}[(Z_{i,t} + Z_{j,t})^2|\mathcal{F}_{t-1}] + \mathbb{E}[(Z_{i,t} - Z_{j,t})^2|\mathcal{F}_{t-1}]},$$

where $\mathbb{E}[\cdot|\mathcal{F}_{t-1}]$ *denotes the conditional expectation operator.*

It thus follows that the conditional variances of the sum and difference of standardized variables can be exploited to obtain the pairwise conditional correlation.

2.2. *Implementation*

We now use Property 2 to construct a filter for the conditional correlation matrix $\mathbf{R}_t$ of the asset return vector $\mathbf{r}_t = (r_{1,t}, ..., r_{N,t})'$, conditional on the information available up until time $t-1$. The procedure requires to first standardize the returns.

We suppose to have a filter for the conditional mean $\widehat{\mu}_{i,t}$ and variance $\widehat{h}_{i,t}$ of each series of asset returns $r_{i,t}$ for all N assets:

$$\begin{aligned} \widehat{\mu}_{i,t} &= \widehat{\mu}\left(r_{i,1}, ..., r_{i,t-1}\right), \\ \widehat{h}_{i,t} &= \widehat{h}\left(r_{i,1}, ..., r_{i,t-1}\right). \end{aligned} \tag{1}$$

Traditional filters for the conditional mean are often based on (extensions of) Autoregressive Moving Average (ARMA) models, while traditional filters for the conditional variance are often based on (extensions of) GARCH models. There is a rich variety of well-studied ARMA and GARCH models which offer a lot of flexibility in modelling the conditional mean and variance process. Without loss of generality, we use the standard GARCH(1,1) setting of Bollerslev (1986) as a conditional variance filter throughout this paper. We summarize this model in Appendix B. Extensions to other GARCH-type models or the inclusion of ARMA terms in the mean filter are straightforward.

Using the mean and variance filters in Equation (1), we can compute the standardized returns $z_{i,t}$ as follows:

$$z_{i,t} = \frac{r_{i,t} - \widehat{\mu}_{i,t}}{\sqrt{\widehat{h}_{i,t}}}. \tag{2}$$

Next, we also suppose to have a filter to compute the conditional variances of the sum and difference of these standardized returns:

$$\begin{aligned} \widehat{h}_{i+j,t} &= \widehat{h}_{i+j}(z_{i,1} + z_{j,1}, ..., z_{i,t-1} + z_{j,t-1}), \\ \widehat{h}_{i-j,t} &= \widehat{h}_{i-j}(z_{i,1} - z_{j,1}, ..., z_{i,t-1} - z_{j,t-1}). \end{aligned} \tag{3}$$

Motivated by Property 2, we can now construct the Variance Implied Conditional Correlation (VICC) which proxies the conditional correlation by using only the univariate conditional variances of Equation (3):

$$\widehat{\rho}_{ij,t} = \frac{\widehat{h}_{i+j,t} - \widehat{h}_{i-j,t}}{\widehat{h}_{i+j,t} + \widehat{h}_{i-j,t}} \quad \text{for} \quad i, j \in \{1, 2, \ldots, N\}, \tag{4}$$

where $\widehat{\rho}_{ij,t}$ denotes the bivariate VICC correlation filter.

Next, we stack the various bivariate VICC correlation filters in a $N \times N$ matrix $\widehat{\mathbf{R}}_t$ with ones on the diagonal and with the (i,j)–th element equal to $\widehat{\rho}_{ij,t}$, for $i \neq j$. In the bivariate case, it is clear from the determinant value $1 - \widehat{\rho}^2_{12,t}$ that the resulting correlation matrix filter is guaranteed to be positive-definite. However, when $N > 2$, $\widehat{\mathbf{R}}_t$ is no longer guaranteed to be positive-definite. Therefore, we define the VICC correlation matrix as a regularized version of $\widehat{\mathbf{R}}_t$ based on the regularization approach of Boudt et al. (2018). Specifically, the VICC correlation filter $\widehat{\mathbf{R}}_t^{\text{VICC}}$ is a convex combination between $\widehat{\mathbf{R}}_t$ and the lagged VICC correlation matrix $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}}$:

$$\widehat{\mathbf{R}}_t^{\text{VICC}} = (1 - \kappa_t)\, \widehat{\mathbf{R}}_t + \kappa_t\, \widehat{\mathbf{R}}_{t-1}^{\text{VICC}}, \tag{5}$$

where $\kappa_t \in [0, 1]$ denotes the regularization intensity, which is only different from zero when the smallest eigenvalue of $\widehat{\mathbf{R}}_t^{\text{VICC}}$ is smaller than or equal to zero. It is given by:

$$\kappa_t = \max\left\{\frac{\psi_{\min} - \lambda_{\min,t}}{1 - \lambda_{\min,t}}, 0\right\}, \tag{6}$$

where $\lambda_{\min,t}$ is the smallest eigenvalue of $\mathbf{G}_{t-1}^{-1}\widehat{\mathbf{R}}_t\mathbf{G}_{t-1}^{-\intercal}$, with $\mathbf{G}_{t-1}$ being the square root of $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}}$ obtained by using the Cholesky factorization such that $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}} = \mathbf{G}_{t-1}\mathbf{G}_{t-1}^{\intercal}$. Furthermore, $\psi_{\min}$ is a positive, near-zero tuning parameter, which we set at 10^{-6} in the application, and $\max\{\cdot, \cdot\}$ is the maximum operator. In Appendix C, we show the explicit derivation of κ_t assuming a positive-definite initialization of the VICC correlation filtering process.

The VICC-based correlation matrix has the typical shrinkage notation as in Ledoit and Wolf (2004), but uses the lagged correlation matrix $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}}$ as target matrix instead of the usual target matrices based on the identity matrix, equicorrelation matrix or single factor model. As such, it is aligned with the time-variation in the correlation filter. As in Boudt et al. (2018), we find that the time-varying regularization parameter κ_t is typically a small number and the regularization is only applied when needed. Finally, note in Equations (5)–(6) that $\widehat{\mathbf{R}}_t^{\text{VICC}} = \widehat{\mathbf{R}}_t$ when $\lambda_{\min,t} > \psi_{\min}$. The regularization should thus not be confused with exponential smoothing as used in Pozzi, Di Matteo, and Aste (2012), where the weight on the lagged correlation prediction is typically high and constant over time. The dynamics of the VICC filter are driven by the dynamics in the $\widehat{h}_{i+j,t}$ and $\widehat{h}_{i-j,t}$ estimates defining the $\widehat{\rho}_{ij,t}$ elements in $\widehat{\mathbf{R}}_t$.

2.3. *Further properties*

A key property of the VICC filter is that its implementation only requires univariate GARCH estimations. It thus avoids the curse of dimensionality which affects the (quasi-)Maximum likelihood estimation of many multivariate GARCH models (see e.g., Boudt et al. (2019) for a recent survey and Pakel et al. (2019) for a recent discussion in the context of DCC models). In contrast with the (quasi-)Maximum likelihood estimation of the DCC model, the VICC parameter estimation is embarrassingly parallel and therefore computationally scalable. A further property is that the VICC filter is designed to yield a well-defined correlation matrix, irrespective of the mean and GARCH variance specification used. As such, it is flexible and it can accomodate the many existing GARCH model specifications (see, e.g., Bollerslev (2008) for an overview). Alternatively, model averaging can be used by setting:

$$\widehat{\rho}_{ij,t}^{avg} = \sum_{k=1}^{K} w^{(k)} \frac{\widehat{h}_{i+j,t}^{(k)} - \widehat{h}_{i-j,t}^{(k)}}{\widehat{h}_{i+j,t}^{(k)} + \widehat{h}_{i+j,t}^{(k)}} \quad \text{for} \quad i,j \in \{1,2,\ldots,N\}, \tag{7}$$

where K is the number of GARCH model implementations considered and $w^{(1)},\ldots,w^{(k)}$ are the weights assigned to each implementation, with $\sum_{k=1}^{K} w^{(k)} = 1$.

We further have that the VICC filter directly leads to dynamic covariance matrix filters, which are useful for portfolio optimization (see e.g., Boudt, Daníelsson, and Laurent (2013)), dynamic beta estimation (see e.g., Engle (2016)) and multivariate hedging, as we document in Section 4. In fact, let $\widehat{\mathbf{D}}_t$ be the $N \times N$ diagonal matrix with element (i,i) equal to $\sqrt{\widehat{h}_{i,t}}$, as defined in Equation (1). Then the VICC covariance matrix filter is given by:

$$\widehat{\mathbf{H}}_t^{\text{VICC}} = \widehat{\mathbf{D}}_t \widehat{\mathbf{R}}_t^{\text{VICC}} \widehat{\mathbf{D}}_t, \tag{8}$$

where the values of $\widehat{\mathbf{D}}_t$ and $\widehat{\mathbf{R}}_t^{\text{VICC}}$ follow from the application of the mean, variance and correlation filters in Equations (1)–(7). Those equations thus provide a flexible dynamic filtering setup. They do not describe a model for the return generating process, for the same reasons as mentioned by Caporin and McAleer (2013) in case of the DCC model. As compared to the DCC filter, the VICC filter uses a different correlation specification and has the advantage of using univariate estimations. For these reasons, it can thus be seen as an alternative method.

2.4. *VICC news impact surface*

The VICC correlation filter $\widehat{\rho}_{12,t}$ in Equation (4) inherits its dynamic properties from the dynamics in the estimated conditional variances of the sum and difference of the standardized return innovations $z_{1,t}$ and $z_{2,t}$. Due to the non-linear transformation of those input series, it is analytically cumbersome to derive the dynamic properties of the VICC correlation filter. Instead, we recommend to use the so-called news impact surface for visualizing the impact of the standardized return innovation on the VICC correlation filter.

The news impact surface is the multivariate extension of the news impact curve which was proposed by Pagan and Schwert (1990) and Engle and Ng (1993) to analyze the impact of innovations on the GARCH variance. Cappiello, Engle, and Sheppard (2006) used news impact surfaces to analyze the impact on correlation dynamics. The VICC news impact surface depends on the variance processes chosen in (3). Suppose for instance that the VICC in Equation (4) is constructed using a GARCH(1,1) specification in $h_{1+2,t}$ and $h_{1-2,t}$ with parameters $\theta = (\omega_{1+2}, \alpha_{1+2}, \beta_{1+2}, \omega_{1-2}, \alpha_{1-2}, \beta_{1-2})'$ (see Appendix B for more details). We can then compute the news impact surface of $\widehat{\rho}_{ij,t}$ by setting $h_{1+2,t-1}$ and $h_{1-2,t-1}$ to their long term mean value (namely, $\overline{h}_{1+2} = \omega_{1+2}/[1-\alpha_{1+2}-\beta_{1+2}]$ and $\overline{h}_{1-2} = \omega_{1-2}/[1-\alpha_{1-2}-\beta_{1-2}]$, respectively).

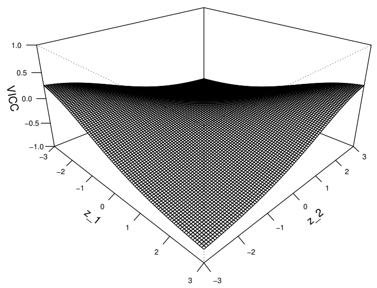
Under these assumptions we then have that:

$$\widehat{\rho}^{\text{VICC}}_{\text{GARCH}}(z_1, z_2) = \frac{c + \alpha_{1+2}(z_1 + z_2)^2 - \alpha_{1-2}(z_1 - z_2)^2}{d + \alpha_{1+2}(z_1 + z_2)^2 + \alpha_{1-2}(z_1 - z_2)^2}, \tag{9}$$

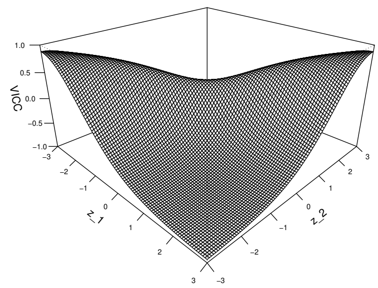
with $c = \omega_{1+2} - \omega_{1-2} + \beta_{1+2}\overline{h}_{1+2} - \beta_{1-2}\overline{h}_{1-2}$ and $d = \omega_{1+2} + \omega_{1-2} + \beta_{1+2}\overline{h}_{1+2} + \beta_{1-2}\overline{h}_{1-2}$.

Figure (1) shows $\widehat{\rho}^{\text{VICC}}_{ij,t}$ as a function of z_1 and z_2 with values ranging from -3 to 3. We consider four different news impact surfaces for the VICC-based correlation. The left panels (a) and (c) show persistent VICC correlation processes, while the right panels (b) and (d) show reactive VICC correlation processes. The top plots correspond to a positive value for c while c is negative for the lower news impact surfaces. Panel (a) has the parameter values $\theta = (0.04, 0.05, 0.85, 0.02, 0.01, 0.90)'$. Note that the GARCH processes used for the sum and difference of the innovations are persistent, leading to a persistent VICC correlation filter with a rather flat news impact surface. Panel (b) corresponds to the parameter vector $\theta = (0.04, 0.05, 0.25, 0.02, 0.01, 0.20)'$. The lower persistence of the GARCH(1,1) models lead to a more reactive VICC correlation. A large degree of co-movement in the standardized return innovations has a bigger impact on the resulting VICC-based correlation estimate. Note that a larger value for either α_{1+2} or α_{1-2} would result in an even more reactive VICC correlation. Panels (c) and (d) have the parameter values $\theta = (0.02, 0.05, 0.85, 0.04, 0.01, 0.90)'$ and $\theta = (0.02, 0.05, 0.25, 0.04, 0.01, 0.20)'$, respectively. Note that only the ω_{1-2} and ω_{1-2} parameters are switched compared to the previous news impact surfaces. In this case when there are no return innovations the VICC correlations in panels (c) and (d) are negative, i.e. -0.69 and -0.31, respectively.

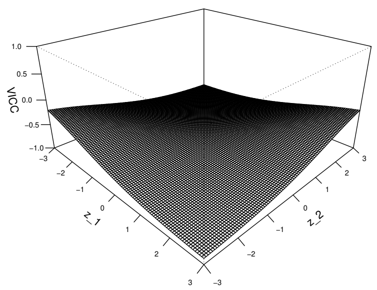
Figure 1. VICC news impact surfaces in case a symmetric GARCH(1,1) specification is used for the variances of the sum and difference of asset returns.



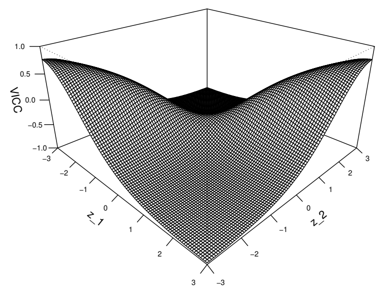
(a) Persistent VICC correlation with $\overline{h}_{1+2} > \overline{h}_{1-2}$.



(b) Reactive VICC correlation with $\overline{h}_{1+2} > \overline{h}_{1-2}$.



(c) Persistent VICC correlation with $\overline{h}_{1+2} < \overline{h}_{1-2}$.



(d) Reactive VICC correlation with $\overline{h}_{1+2} < \overline{h}_{1-2}$.

3. Monte Carlo study

It is clear that by construction the estimation of the parameters of the variance processes determining the dynamics in the VICC correlation matrix filter in Equations (4)–(5) is computationally convenient, as it only requires univariate GARCH model estimations and is thus embarassingly parallel. In this section, we use numerical simulations to evaluate the performance of the VICC in capturing the dynamics in the conditional correlation for a broad range of conditional correlation processes. For all the considered processes, the VICC is misspecified. The study shows that, despite the misspecification, the VICC still yields reliable estimates of the conditional correlation, whatever the underlying conditional correlation process is. We first discuss the benchmark models and then present the bivariate and multivariate Monte Carlo study, respectively.

3.1. *Benchmark models*

We compare the performance of the VICC-based correlation filter against four benchmark estimators by comparing the Mean Absolute Error (MAE) and Mean Squared Error (MSE) in estimating the in-sample correlation. As a first benchmark, we opt for the EWMA covariance model which requires no estimation (see e.g., Engle (2009)). We describe the model in Appendix D. The second benchmark model is the standard workhorse in conditional correlation estimation, namely the DCC specification of Engle (2002). Its estimation still requires the optimization of a multivariate likelihood function. We review the DCC model in Appendix E. The third benchmark model is the GO-GARCH model of van der Weide (2002). By using an unobserved independent factor framework, this model is another feasible MGARCH model for the estimation of larger systems. We provide details of the GO-GARCH model in Appendix F. The last benchmark is from the related work of Harris and Shen (2003) (HS hereafter) in which they imply the conditional covariance via the polarization result applied to variances of returns rather than to variances of standardized returns. We expect the VICC to be a more reliable estimator, as the correlation embedded in the HS method is an unbounded function of the univariate variance estimates which may lead to correlation estimates that violate the absolute bounds for a correlation, i.e., $|\widehat{\rho}_t^{\mathrm{HS}}| > 1$. If this violation occurs we truncate the HS correlation estimate at -1 or 1 as appropriate (see e.g., Zhang (2011)). Moreover, the HS method does not distinguish between the dynamics of the conditional correlation and the conditional variance. Since the seminal paper of Engle (2002) introducing the DCC model, it has become the standard to use separate equations for modelling the conditional variances and correlations. Details of the HS method are provided in Appendix G.

3.2. *Bivariate simulation*

We consider a similar bivariate Monte Carlo setup as in Engle (2002) and Creal, Koopman, and Lucas (2011) in which the true correlation structure can be different from the one assumed by the econometrician. The return series with conditional variances $h_{1,t}$ and $h_{2,t}$ are constructed as follows:

$$\begin{aligned} r_{1,t} &= \sqrt{h_{1,t}} z_{1,t}, \\ r_{2,t} &= \sqrt{h_{2,t}} z_{2,t}, \end{aligned} \tag{10}$$

where $z_{1,t}$ and $z_{2,t}$ are bivariate standard normally distributed with a (possibly time-varying) correlation coefficient ρ_t. In a similar way as in Engle (2002), the data generating process further consists of the following two standard GARCH(1,1) models:

$$\begin{aligned} h_{1,t} &= 0.01 + 0.05 r_{1,t-1}^2 + 0.94 h_{1,t-1}, \\ h_{2,t} &= 0.5 + 0.2 r_{2,t-1}^2 + 0.5 h_{2,t-1}. \end{aligned} \tag{11}$$

We examine how the correlation filters perform under different forms of misspecification by considering seven possible true correlation processes. First, as in Engle (2002) and Creal, Koopman, and Lucas (2011), we consider the constant, sine, fast sine, step and ramp correlation processes:

- **constant** $\rho_t = 0.9$,
- **sine** $\rho_t = 0.5 + 0.4\ \cos(2\pi t/200)$,
- **fast sine** $\rho_t = 0.5 + 0.4\ \cos(2\pi t/20)$,
- **step** $\rho_t = 0.9 - 0.5(t > 500)$,
- **ramp** $\rho_t = \mod (t/200)$.

The sine correlation process exhibits gradual changes, while the fast sine correlation process contains fast changes. The step correlation process appears to be constant, but exhibits an abrupt change. These abrupt changes are also found in the ramp correlation process.

In addition, we also assume a mean-reverting Dynamic Conditional Correlation (DCC) process:

- **DCC** $\rho_t = \frac{\boldsymbol{Q}_{12,t}}{\sqrt{\boldsymbol{Q}_{11,t}\boldsymbol{Q}_{22,t}}}$,
  with $\boldsymbol{Q}_t = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix} + 0.05 \begin{bmatrix} z_{1,t-1}^2 & z_{1,t-1}z_{2,t-1} \\ z_{1,t-1}z_{2,t-1} & z_{2,t-1}^2 \end{bmatrix} + 0.85\, \boldsymbol{Q}_{t-1}$.

The seventh process is a conditional correlation process which is associated to the Diagonal BEKK (DBEKK) model (Engle and Kroner 1995), namely:

- **DBEKK** $\rho_t = \frac{h_{12,t}}{\sqrt{h_{1,t}h_{2,t}}}$ with $h_{12,t} = 0.14 + 0.1 r_{1,t-1} r_{2,t-1} + 0.69 h_{12,t-1}$,

with $h_{1,t}$ and $h_{2,t}$ as defined in Equation (11). In terms of the number of replications and time series length, we follow Engle (2002) by simulating 200 series of 1000 observations.

Table 1 shows the MAE and MSE of the VICC, DCC, EWMA, GO-GARCH and HS method for each of the seven considered correlation processes, and their corresponding HAC standard errors between parentheses. The MAE of the VICC ranges in between 0.0060 for the constant correlation process, and 0.2267 for the fast sine correlation process. In all the cases where the true correlation process is not a DCC process, the performance of the VICC is either better than (sine, fast sine, step, ramp and DBEKK) or similar to (constant) the DCC model. Note from the HAC standard errors that for the sine, step and ramp correlation process the outperformance is statistically significant for both the MAE and MSE. For the fast sine correlation process the outperformance is only statistically significant for the MSE. In case the econometrician knows that the true process is a DCC process, it is clearly optimal to use a DCC. However, in practice, as mentioned by Caporin and McAleer (2013), the DCC process is not a realistic data generating process. In addition to its better or equal performance performance, the VICC is also easier to compute. Finally, for all the considered correlations processes, the EWMA, GO-GARCH and HS method perform (drastically) worse than both the VICC and the DCC correlation filters.

3.3. *Multivariate simulation*

Table 2 and 3 present the MAE and MSE results, respectively, for the N-dimensional returns simulated using a two-block Dynamic Equicorrelation (DECO) specification for the conditional correlation matrix given by:

$$\mathbf{R}_t = \begin{bmatrix} (1-\rho_{1,t}^2)\mathcal{I}_{N_1} & 0 \\ 0 & (1-\rho_{2,t}^2)\mathcal{I}_{N_2} \end{bmatrix} + \begin{pmatrix} \rho_{1,t}\mathcal{L}_{N_1} \\ \rho_{2,t}\mathcal{L}_{N_2} \end{pmatrix} \begin{pmatrix} \rho_{1,t}\mathcal{L}'_{N_1} & \rho_{2,t}\mathcal{L}'_{N_2} \end{pmatrix}, \quad (12)$$

where $\rho_{1,t}$ and $\rho_{2,t}$ denote the dynamic equicorrelation parameters in the diagonal blocks, $\mathcal{I}_N$ is the N-dimensional identity matrix, $\mathcal{L}_N$ is a $N \times 1$ vector of ones, and N_1 and N_2 are the number of assets in each block with $N_1 + N_2 = N$. Note that our setup allows for a direct relation between $\rho_{1,t}$ and $\rho_{2,t}$ and the dynamic cross-equicorrelation parameter in the off-diagonal blocks. The same process was considered in Lucas, Schwaab, and Zhang (2017). $\mathbf{R}_t$ is a positive definite correlation matrix if

Table 1. Mean absolute error and mean squared error of the correlation estimates obtained using the VICC, DCC, EWMA, GO-GARCH and HS method for the constant, sine, fast sine, step, ramp, DCC and DBEKK correlation processes.

	VICC	DCC	EWMA	GO-GARCH	HS
Mean absolute error					
constant	0.0060	**0.0059**	0.0276	0.0206	0.0701
	(0.0001)	(<0.0001)	(0.0002)	(0.0002)	(0.0008)
sine	**0.1312**	0.1390	0.1501	0.1688	0.1623
	(0.0008)	(0.0009)	(0.0011)	(0.0013)	(0.0010)
fast sine	0.2267	**0.2266**	0.2603	0.2449	0.2370
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0004)
step	**0.0665**	0.0712	0.0791	0.0999	0.1248
	(0.0011)	(0.0013)	(0.0014)	(0.0013)	(0.0016)
ramp	**0.1504**	0.1569	0.1551	0.1777	0.1785
	(0.0012)	(0.0011)	(0.0011)	(0.0015)	(0.0010)
DCC	0.0479	**0.0324**	0.0741	0.0852	0.0806
	(0.0004)	(0.0004)	(0.0006)	(0.0007)	(0.0008)
DBEKK	**0.0807**	0.0819	0.1049	0.0987	0.0821
	(0.0004)	(0.0004)	(0.0008)	(0.0006)	(0.0004)
Mean squared error					
constant	0.0001	**0.0001**	0.0013	0.0006	0.0076
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.0003)
sine	**0.0264**	0.0306	0.0347	0.0428	0.0402
	(0.0003)	(0.0004)	(0.0005)	(0.0006)	(0.0006)
fast sine	**0.0669**	0.0679	0.0926	0.0795	0.0782
	(0.0002)	(0.0002)	(0.0004)	(0.0002)	(0.0004)
step	**0.0093**	0.0112	0.0137	0.0172	0.0249
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0008)
ramp	**0.0402**	0.0439	0.0473	0.0512	0.0520
	(0.0006)	(0.0006)	(0.0007)	(0.0008)	(0.0010)
DCC	0.0037	**0.0019**	0.0089	0.0120	0.0109
	(<0.0001)	(<0.0001)	(0.0002)	(0.0002)	(0.0004)
DBEKK	**0.0106**	0.0109	0.0181	0.0160	0.0119
	(<0.0001)	(<0.0001)	(0.0003)	(0.0002)	(0.0001)

Note: This table presents the mean absolute error and mean squared error for each of the seven considered correlation processes and for each considered correlation estimator. Between parentheses are the Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors. For each correlation process, the result of the best performing estimator is put in bold.

and only if $\rho_{i,t} \in (0, 1)$ for $i = 1, 2$. We set both $\rho_{1,t}$ and $\rho_{2,t}$ according to one of the seven bivariate correlation processes which we have defined in Section 3.2 (the i.e. constant $\rho_{1,t} = 0.9$) and consider all 28 possible combinations for a 10×10 correlation matrix with $N_1 = N_2 = 5$.

We use the VICC, DCC, EWMA and GO-GARCH method to estimate the 10×10 two-block DECO correlation matrix and show the resulting MAE and MSE for the lower diagonal elements in Table 2 and 3, respectively.[1] The MAE of the VICC ranges in between 0.0108 for the constant-constant correlation process, and 0.1894 for the combination between between the ramp and fast sine correlation processes.

For all the cross combinations between the considered correlation processes (i.e. the constant-sine correlation process), the VICC outperforms all the other benchmark models. The MAE indicates that the outperformance in the multivariate simulation is more substantial compared to the bivariate simulation setup. For example, the MAE difference between the VICC and DCC for the cross combination between the constant and sine correlation processes is equal to 0.04. When the same correlation processes are combined in the two-block DECO correlation matrix (i.e. the constant-constant correlation process), the DCC method outperforms the VICC method twice, although not substantially, and the GO-GARCH method outperforms the VICC three times. Overall, the EWMA seems to be the worst performing model in the multivariate simulation setup. The bottom line is that in most cases the VICC outperforms all the other benchmark models and that the outperformance becomes more substantial in higher dimensional correlation matrices.

[1]We omit the results of the HS method as the use of an improper pairwise correlation estimator causes some severe outliers in the HS estimates of the 10×10 two-block DECO correlation matrix (i.e. $|\widehat{\rho}_t^{\text{HS}}| > 1$). After truncation of the correlations at 1 or -1 as appropriate, we try two methods to obtain a positive-definite correlation matrix, namely via our proposed regularization method, and via a brute force method which extracts the negative eigenvalues via an eigendecomposition and sets these equal to a small positive number (i.e. 10^{-1}) before reconstructing the correlation matrix. While these methods improve the results, they remain drastically worse compared to all the other models. In practice, additional truncations are needed, but this is beyond the scope of our paper. The results with and without truncation are available from the authors upon request.

Table 2. Mean absolute error for the the lower diagonal elements of the simulated 10×10 two-block DECO correlation matrix using the VICC, DCC, EWMA and GO-GARCH method.

	$\rho_{1,t}$ \ $\rho_{2,t}$	**constant**	**sine**	**fast sine**	**step**	**ramp**	**DCC**	**DBEKK**
VICC	**constant**	0.0108						
	sine	**0.1113**	0.1546					
	fast sine	**0.1760**	**0.1832**	0.2452				
	step	**0.0615**	**0.1250**	**0.1728**	**0.0877**			
	ramp	**0.1235**	**0.1663**	**0.1894**	**0.1329**	0.1732		
	DCC	**0.0658**	**0.1191**	**0.1483**	**0.0915**	**0.1245**	0.0767	
	DBEKK	**0.0777**	**0.1287**	**0.1617**	**0.0994**	**0.1361**	**0.0861**	**0.0972**
DCC	**constant**	**0.0086**						
	sine	0.1522	0.1921					
	fast sine	0.1848	0.2016	0.2546				
	step	0.0760	0.1476	0.1861	0.1011			
	ramp	0.1574	0.1838	0.2043	0.1524	0.2061		
	DCC	0.0667	0.1319	0.1501	0.0977	0.1348	**0.0761**	
	DBEKK	0.0805	0.1426	0.1644	0.1057	0.1472	0.0865	0.0990
EWMA	**constant**	0.0487						
	sine	0.1276	0.1646					
	fast sine	0.2039	0.2009	0.2693				
	step	0.0831	0.1383	0.1934	0.1006			
	ramp	0.1344	0.1742	0.2075	0.1438	0.1761		
	DCC	0.1103	0.1466	0.1877	0.1254	0.1515	0.1371	
	DBEKK	0.1124	0.1498	0.1945	0.1279	0.1556	0.1392	0.1417
GO-GARCH	**constant**	0.0310						
	sine	0.1382	**0.1344**					
	fast sine	0.1848	0.1919	**0.2394**				
	step	0.0996	0.1561	0.1911	0.0965			
	ramp	0.1453	0.1666	0.1946	0.1580	**0.1509**		
	DCC	0.0775	0.1308	0.1519	0.1114	0.1332	0.0830	
	DBEKK	0.0893	0.1427	0.1663	0.1223	0.1454	0.0933	0.1045

Note: This table presents the mean absolute error for the the lower diagonal elements of the simulated 10×10 two-block DECO correlation matrix using the VICC, DCC, EWMA and GO-GARCH method. The 28 considered correlation processes are combinations of the constant, sine, fast sine, step, ramp, DCC and DBEKK correlation processes described in Section 3.2. For each correlation process, the result of the best performing estimator is put in bold.

Table 3. Mean squared error for the the lower diagonal elements of the simulated 10 × 10 two-block DECO correlation matrix using the VICC, DCC, EWMA and GO-GARCH method.

	$\rho_{1,t}$ \ $\rho_{2,t}$	**constant**	**sine**	**fast sine**	**step**	**ramp**	**DCC**	**DBEKK**
VICC	**constant**	0.0002						
	sine	**0.0223**	0.0344					
	fast sine	**0.0498**	**0.0485**	0.0767				
	step	**0.0087**	**0.0250**	**0.0468**	**0.0156**			
	ramp	**0.0312**	**0.0419**	**0.0522**	**0.0304**	0.0468		
	DCC	**0.0082**	**0.0212**	**0.0333**	**0.0136**	**0.0249**	0.0093	
	DBEKK	**0.0117**	**0.0243**	**0.0381**	**0.0161**	**0.0290**	**0.0117**	**0.0148**
DCC	**constant**	**0.0001**						
	sine	0.0364	0.0485					
	fast sine	0.0536	0.0564	0.0807				
	step	0.0122	0.0320	0.0512	0.0198			
	ramp	0.0435	0.0500	0.0594	0.0375	0.0608		
	DCC	0.0085	0.0264	0.0340	0.0156	0.0297	**0.0091**	
	DBEKK	0.0126	0.0297	0.0391	0.0180	0.0338	0.0118	0.0153
EWMA	**constant**	0.0040						
	sine	0.0282	0.0412					
	fast sine	0.0682	0.0621	0.1018				
	step	0.0138	0.0314	0.0604	0.0201			
	ramp	0.0373	0.0496	0.0670	0.0377	0.0539		
	DCC	0.0212	0.0334	0.0552	0.0259	0.0379	0.0295	
	DBEKK	0.0223	0.0345	0.0586	0.0271	0.0401	0.0304	0.0316
GO-GARCH	**constant**	0.0015						
	sine	0.0308	**0.0276**					
	fast sine	0.0533	0.0540	**0.0757**				
	step	0.0157	0.0365	0.0530	0.0159			
	ramp	0.0369	0.0435	0.0562	0.0392	**0.0380**		
	DCC	0.0106	0.0265	0.0355	0.0190	0.0289	0.0109	
	DBEKK	0.0143	0.0307	0.0411	0.0226	0.0334	0.0138	0.0172

Note: This table presents the mean squared error for the the lower diagonal elements of the simulated 10×10 two-block DECO correlation matrix using the VICC, DCC, EWMA and GO-GARCH method. The 28 considered correlation processes are combinations of the constant, sine, fast sine, step, ramp, DCC and DBEKK correlation processes described in Section 3.2. For each correlation process, the result of the best performing estimator is put in bold.

4. Application

The prediction of conditional correlations is of great practical importance in a lot of financial applications, such as portfolio optimization, hedging and risk management. In this section, we examine whether the VICC can be a useful tool in the estimation of conditional hedge ratios. We first introduce the concept of futures cross hedging. Next, we present the data. Then, we explain how the cross hedging performance of the various hedge ratios can be measured and finally we show the bivariate and multivariate cross hedging results.

4.1. *Cross hedging*

Futures cross hedging is a hedging strategy where futures contracts of a (correlated) asset that differs from the underlying asset are used.[2] Often, an investor cross hedges when he wants to avoid a certain type of exposure in his portfolio. More specifically, we consider three cross hedging applications where the investor seeks to protect his S&P 500 index portfolio value against changes in (either) the interest rate, the exchange rate between the US Dollar and the Euro, or the VIX index. He aims to achieve this by cross hedging his fixed long spot position in the S&P 500 index using the CBOT 10-y US T-Note (TY), the CME Euro FX (EC) and the CBOE VIX (VX) futures, respectively. We consider bivariate as well as multivariate cross hedging. In the former the investor only wants to eliminate one type of exposure, while in the latter all three types of exposure are eliminated simultaneously. In a similar way as in Wang, Wu, and Yang (2015), we assume that the investor rebalances his hedging position at the close of each trading week.

Let us consider an investor with a one-period hedging horizon who wants to hedge a fixed long spot position with a one period return $r_{s,t}$ by taking a position in N (correlated) futures contracts with $\mathbf{r}_{f,t}$ representing the vector of one period futures returns. To simplify the notation we assume that the investor has a long spot position of one unit. The hedged portfolio return $r_{p,t}$ is:

$$r_{p,t} = r_{s,t} - \mathbf{b}_t^\intercal \mathbf{r}_{f,t}, \tag{13}$$

where $\mathbf{b}_t$ denotes a vector containing the futures positions' units. We follow Kroner and Sultan (1993), among others, by determining the optimal values for the single-period hedge ratios $\mathbf{b}_t$ as those that minimize the conditional variance of the hedged portfolio return.[3] This leads to the following optimization problem:

$$\underset{\mathbf{b}_t}{\arg\min} \quad \mathrm{Var}\left(r_{p,t}|\mathcal{F}_{t-1}\right), \tag{14}$$

where $\mathrm{Var}\left(\cdot|\mathcal{F}_{t-1}\right)$ denotes the conditional variance operator with $\mathcal{F}_{t-1}$ being the avail-

[2]Another typical application is direct hedging, which is a hedging strategy in which the futures contract is based on the same underlying as the spot position. We do not consider this type of hedging in this paper since Wang, Wu, and Yang (2015) find that, due to model misspecification and estimation errors, an estimated direct hedge ratio often does not lead to substantial gains with respect to a naive hedging strategy of using a unit hedge ratio.

[3]Optimal hedge ratios can also be studied under the expected-utility maximization paradigm. In this regard, note that in the application a weekly hedging horizon is used and for such short horizon the expected return contribution to the mean-variance utility function is negligible with respect to the magnitude of the portfolio variance. The optimal hedge ratio obtained under the expected-utility maximization paradigm is then similar to the one obtained under the minimum variance criterion (see e.g., Chen, Lee, and Shrestha (2003)).

able information set up until time $t-1$. It follows from the first-order condition that:

$$\mathbf{b}_t^* = \mathbf{H}_{ff,t}^{-1}\mathbf{H}_{sf,t}, \quad (15)$$

where $\mathbf{b}_t^*$ is a vector containing the optimal hedge ratios at time t conditional on the available information at time $t-1$, $\mathbf{H}_{ff,t}$ is the conditional covariance matrix of the futures returns, and $\mathbf{H}_{sf,t}$ is a vector containing the conditional covariances between the spot returns and each of the futures returns. In the simple case of bivariate cross hedging where $N=1$, the optimal hedge ratio is equal to the ratio of the conditional covariance between spot and futures returns to the conditional variance of futures returns.

We propose to estimate the hedge ratios in Equation (15) by using the VICC covariance matrix filter as defined in Equation (8). We compare its performance with an unconditional hedge ratio, which is the slope of the Ordinary Least Squares (OLS) regression of spot returns on futures returns (see e.g., Kavussanos and Visvikis (2008)). The DCC, EWMA, GO-GARCH and HS method are the additional conditional benchmark estimators.

4.2. *Data*

The dataset contains daily prices on the S&P 500 index, the CBOT 10-y US T-Note (TY), the CME Euro FX (EC), and the CBOE VIX (VX) futures from January 1, 2008 until December 31, 2018. We use Friday closing prices to compute weekly simple percentage returns. If Friday is a holiday, Thursday closing prices are used. We use the nearest futures contract to delivery and roll over to the next nearest contract when the current contract reaches the first day of the delivery month or its expiry date to avoid thin trading and expiration effects (see e.g., Lypny and Powalla (1998)). The spot data is obtained via Wharton Research Data Services (WRDS) and the futures data via Quandl (Stevens Continuous Futures).

Table 4 shows the sample means and standard deviations (in %) of the weekly spot and futures returns on the S&P 500 index, and the CBOT 10-y US T-Note, the CME Euro FX and the CBOE VIX futures from January 2008 until December 2018. We further report the relative volatilities and the sample correlation coefficients between the corresponding spot and futures returns, and the unconditional bivariate OLS hedge ratios. Besides the full-sample summary statistics, we also provide sub-sample summary statistics by dividing the sample into five sub-samples.

The sub-sample summary statistics from 2008–2010 show the substantial impact of the Financial Crisis of 2008 on the hedge ratio dynamics as the high volatility on the S&P 500 index leads to the highest relative volatility between spot and futures returns for the Euro FX and VIX futures, and for the second highest relative volatility between spot and futures returns for the 10-y US T-Note futures, compared to the other sub-samples. This means that, at constant correlation, the value of the unconditional hedge ratio should be decreasing in the consecutive sub-samples. However, this is often not the case, as the effect of the decreasing relative volatility is partially offset by an increase in the absolute value of the correlation. Furthermore, the time-varying correlation also has a notable impact on the hedge ratio of the Euro FX futures, where the sign of the correlation even changes from positive to negative, and back to positive in the last few years of the sample. These sub-sample summary statistics show the need for a time-varying hedge ratio and the potential benefits in using separate equations for estimating the conditional variances and correlations as they show different time-series

Table 4. Full-sample and sub-sample summary statistics of the weekly spot and futures returns over the period January 1, 2008 until December 31, 2018.

	$\widehat{\mu}$	$\sqrt{\widehat{h}}$	$\sqrt{\widehat{h}_s/\widehat{h}_f}$	$\widehat{\rho}_{sf}$	$\widehat{b}$	$\widehat{\mu}$	$\sqrt{\widehat{h}}$	$\sqrt{\widehat{h}_s/\widehat{h}_f}$	$\widehat{\rho}_{sf}$	$\widehat{b}$
			2008–2018					*2013–2014*		
S&P 500	0.13	2.52				0.34	1.43			
CBOT 10-y US T-Note	0.01	0.84	3.01	-0.35	-1.04	-0.03	0.74	1.92	-0.22	-0.43
CME Euro FX	-0.03	1.38	1.82	0.24	0.43	-0.07	0.92	1.55	-0.10	-0.15
CBOE VIX	0.40	9.08	0.28	-0.67	-0.18	0.33	7.24	0.20	-0.80	-0.16
			2008–2010					*2015–2016*		
S&P 500	-0.01	3.68				0.10	1.80			
CBOT 10-y US T-Note	0.04	1.12	3.29	-0.31	-1.01	-0.02	0.77	2.35	-0.31	-0.73
CME Euro FX	-0.05	1.74	2.11	0.36	0.75	-0.11	1.46	1.23	-0.16	-0.20
CBOE VIX	0.23	9.07	0.41	-0.62	-0.25	0.26	9.56	0.19	-0.80	-0.15
			2011–2012					*2017–2018*		
S&P 500	0.14	2.41				0.10	1.88			
CBOT 10-y US T-Note	0.09	0.78	3.11	-0.63	-1.97	-0.02	0.49	3.85	-0.26	-1.02
CME Euro FX	0.03	1.43	1.68	0.49	0.82	0.09	0.91	2.07	0.05	0.10
CBOE VIX	0.37	10.51	0.23	-0.74	-0.17	0.89	8.80	0.21	-0.73	-0.16

Note: This table presents the sample means ($\widehat{\mu}$) and standard deviations ($\sqrt{\widehat{h}}$) of the weekly returns (in %) of the S&P 500 index, and of the CBOT 10-y US T-Note, CME Euro FX and CBOE VIX futures over the period January 1, 2008 until December 31, 2018. The relative volatility ($\sqrt{\widehat{h}_s/\widehat{h}_f}$) and the sample correlation coefficients ($\widehat{\rho}_{sf}$) between the corresponding spot and futures returns, and the unconditional bivariate OLS hedge ratios ($\widehat{b}$) are also reported.

behaviour.

4.3. *Performance measures*

We estimate weekly out-of-sample hedge ratios using the VICC, DCC, EWMA, GO-GARCH, HS and OLS method on a rolling estimation window for cross hedging the S&P 500 index using the CBOT 10-y US T-Note, the CME Euro FX and the CBOE VIX futures from January 2008 until December 2018. In a similar way as in Wang, Wu, and Yang (2015), we split our sample into two subsamples, namely an estimation window and an evaluation window, by splitting it in half. We use the estimation window to compute the one-step-ahead weekly forecasts of the OHR and the corresponding out-of-sample hedged portfolio returns for all the considered models. Then we reestimate the model parameters each time by rolling the estimation window one week forward and dropping the first observation.[4] Finally, we evaluate the models' cross hedging effectiveness by comparing the out-of-sample variance reduction, the achieved decorrelation and the stability of the hedge ratio.

We quantify the first measure of hedging effectiveness with the Variance Reduction Ratio (VRR) which compares the variance of the out-of-sample hedged portfolio returns to the variance of the unhedged portfolio returns as follows:

$$\text{VRR} = 100 \left(\frac{\widehat{h}_s - \widehat{h}_p}{\widehat{h}_s} \right), \tag{16}$$

where $\widehat{h}_p$ and $\widehat{h}_s$ are the sample variances of the weekly hedged and unhedged portfolio returns, respectively. Following Wang, Wu, and Yang (2015), we use the Diebold-Mariano test (DM-test) to account for the estimation uncertainty when evaluating the difference in out-of-sample performance (Diebold and Mariano 2002). We compare the cross hedging performance of the best performing hedge ratio with each of the competing hedge ratios. We use the squared weekly out-of-sample returns and a Null hypothesis of equal or worse performance by the best performing hedge ratio. We refer to the original paper of Diebold and Mariano (2002) for a more detailed discussion.

The second measure of hedging effectiveness is the achieved decorrelation between the out-of-sample hedged portfolio returns and the corresponding futures returns. If a hedge ratio is effective, then the resulting hedged portfolio returns should be uncorrelated with the futures returns. We compute the correlation coefficient between the out-of-sample hedged portfolio returns and the futures returns, and test whether it is significant with a Null hypothesis of no correlation.

A last performance measure is the stability of the hedge ratio. A timely hedge ratio should immediately react to changes in the underlying correlation, and relative volatility, between spot and futures returns. However, if two separate hedge ratios lead to an equal performance in terms of variance reduction and achieved decorrelation, the more stable hedge ratio is preferred as it generates less transaction costs. We define a more stable hedge ratio, as a hedge ratio that generates a lower portfolio turnover. We define the portfolio turnover over the sample with T observations as follows:

$$\text{Turnover} = 100 \, \frac{1}{T-1} \sum_{t=2}^{T} |\widehat{b}_t - \widehat{b}_{t-1}|, \tag{17}$$

where $\widehat{b}_t$ denotes an estimated hedge ratio. In the case of multiple hedge ratios, we average over all the resulting turnovers.

4.4. *Bivariate results*

Let us first consider the actual time series of estimated bivariate hedge ratios for the various methods. We show these in Figure 2. We observe clear differences in time series properties. The most stable estimate is obtained by the OLS hedge ratio. The other methods are more reactive to recent data. The VICC and DCC hedge ratio show a large degree of co-movement, while the HS method yields the most volatile predictions

[4] We also considered an expanding estimation window but found that, for all methods considered, a rolling estimation window tend to lead to a better hedging performance. The expanding estimation window results are available from the authors upon request.

of the optimal hedge ratio. Despite the similar dynamics, the GO-GARCH and EWMA hedge ratio are easy to distinguish from both the VICC and DCC hedge ratio.

In Figure 3, we plot the correlation prediction that is associated with each of the hedge ratios. Note that the correlation and hedge ratio have of course the same sign. They differ by the scaling factor, which is equal to the ratio of the underlying relative volatility between spot and futures returns. The VICC, DCC, EWMA, GO-GARCH and OLS yield by construction conditional correlation estimates that are in between -1 and 1. In contrast, we observe that the HS correlation effectively violates this condition and is sometimes truncated at -1 or 1 as appropriate. Its interpretation is therefore problematic.

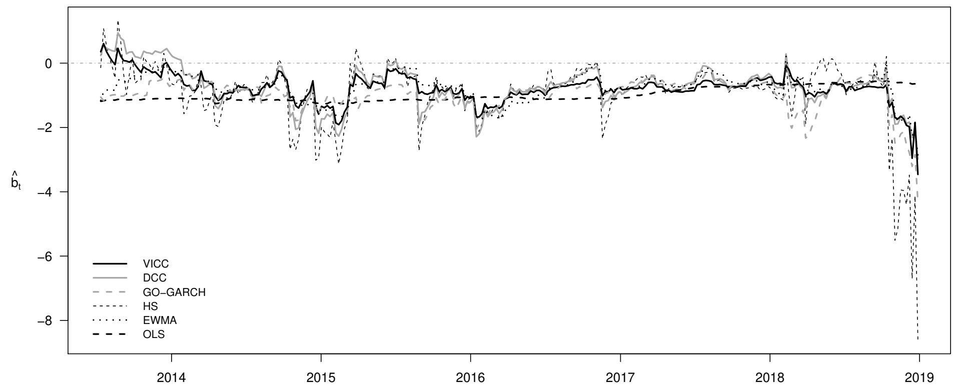
Table 5 shows the VRR, the correlation between the out-of-sample hedged portfolio returns and the futures returns, and the turnover of the VICC, DCC, EWMA, GO-GARCH, HS and OLS hedge ratios for the bivariate cross hedging applications. We observe that the VICC hedge ratio is the top performer in terms of variance reduction for the 10-y US T-Note and Euro FX futures cross hedging applications with a VRR of 9.34 and 0.47 percentage points, respectively. It is also the second best performing hedge ratio for the VIX futures where the DCC hedge ratio reduces the most portfolio return variance with a VRR of 65.18 percentage points. The DM-test indicates that the VRR of the VICC and DCC hedge ratios are only significantly different in the Euro FX futures cross hedging application in favor of the VICC. The HS and OLS hedge ratios are significantly outperformed by either the VICC or the DCC hedge ratio in all three applications, while the EWMA hedge ratio is significantly outperformed twice. The GO-GARCH hedge ratio performs quite well in the 10-y US T-Note and VIX applications, but performs drastically worse for the Euro FX. Besides the statistical significance, the VRR numbers indicate that there are some substantial economic gains in using the VICC. For example, for the Euro FX futures, the VICC hedge ratio has a VRR that is 3.75, 1.42, 12.75, 6.42 and 12.83 percentage points larger than the VRR of the DCC, EWMA, GO-GARCH, HS and OLS hedge ratio, respectively.

The correlation coefficients of the out-of-sample hedged portfolio returns with the futures returns indicate that all the hedge ratios, except for the GO-GARCH and OLS hedge ratios, lead to hedged portfolio returns that are not significantly correlated with the futures returns for each cross hedging application. This means that the hedge ratios succeed in eliminating the exposure to the interest rate, the exchange rate between the US Dollar and the Euro, and the VIX. The GO-GARCH hedge ratio fails to eliminate the exchange rate exposure, as the hedged portfolio returns are still significantly correlated with the Euro FX futures returns, and the OLS hedge ratio only manages to eliminate interest rate exposure.

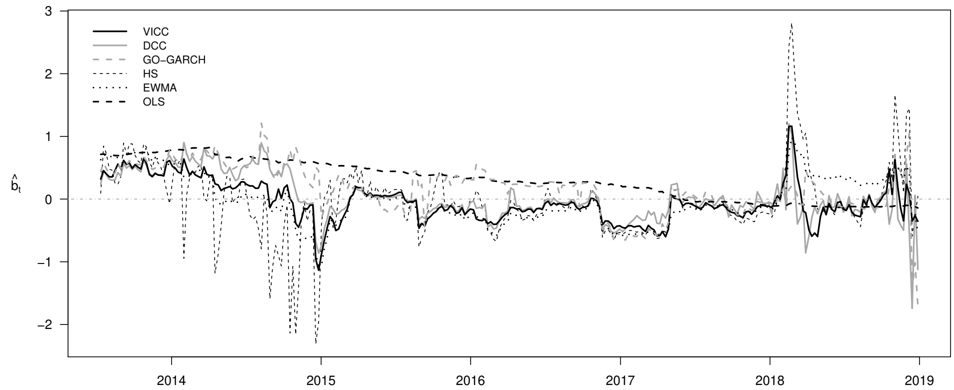
Consistent with the stable time series line of the OLS hedge ratio in Figure 2, we find that the portfolio turnover indicates that the OLS hedge ratio is always by far the most stable hedge ratio. However, as the VRR and achieved decorrelation indicate, this stability comes at a cost of a less (or even a non-)effective hedging strategy. One should evaluate whether the higher hedging effectiveness is worth the extra turnover that the conditional models generate for each particular case. The EWMA hedge ratio is on average the second most stable hedge ratio but again at the cost of a less effective hedging strategy, while the VICC hedge ratio comes third as a substantially more stable hedge ratio, on average, compared to the DCC and GO-GARCH hedge ratio (e.g., about a 4.83 and 2.04 percentage points lower turnover for the 10-y US T-Note, respectively). Lastly, the HS method does not only seem to lead to a poor performance in terms of variance reduction, but also to an unstable hedge ratio.

Overall, we find that, in terms of variance reduction, the VICC hedge ratio signif-

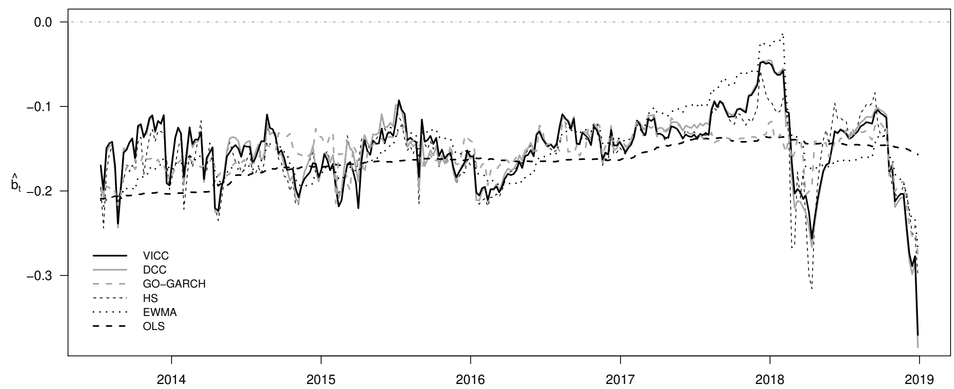
Figure 2. Out-of-sample estimates of the VICC, DCC, EWMA, GO-GARCH, HS and OLS bivariate hedge ratios for the S&P 500 index using either the 10-y US T-Note, Euro FX and VIX futures as cross hedging asset.



(a) 10-y US T-Note.

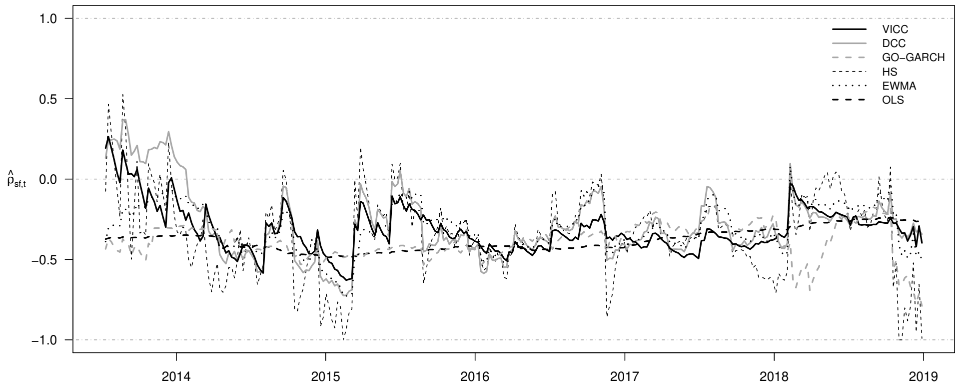


(b) Euro FX.

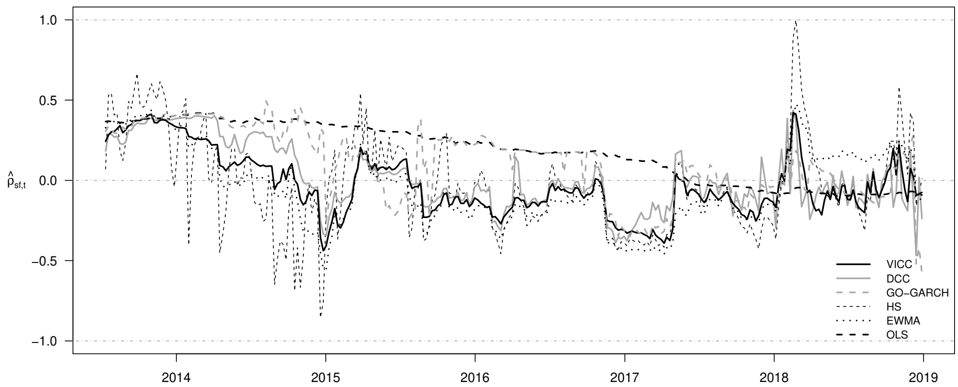


(c) VIX.

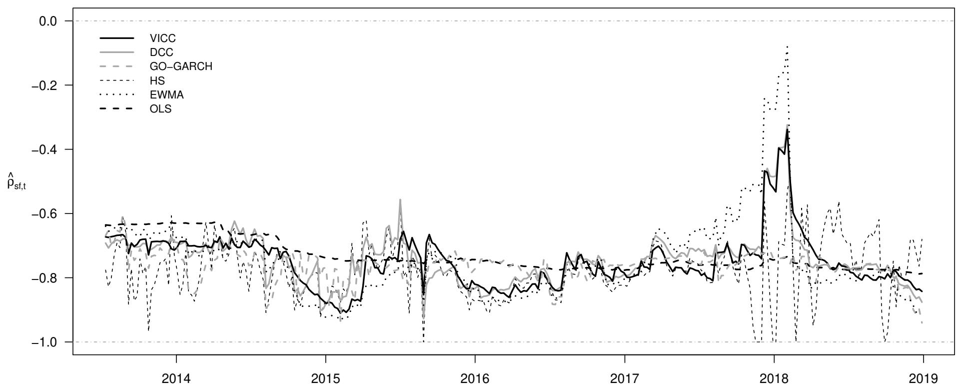
Figure 3. Out-of-sample estimates of the correlation between the weekly S&P 500 index return and the returns on the 10-y US T-Note, Euro FX and VIX futures obtained via the VICC, DCC, EWMA, GO-GARCH, HS and OLS models.



(a) 10-y US T-Note.



(b) Euro FX.



(c) VIX.

Table 5. Bivariate hedging performance of the VICC, DCC, EWMA, GO-GARCH, HS and OLS hedge ratio for cross hedging the S&P 500 index using the 10-y US T-Note, Euro FX and VIX futures.

	VICC	DCC	EWMA	GO GARCH	HS	OLS
10-y US T-Note						
VRR	**9.34**	8.71	4.69**	8.72	3.20*	4.38*
$\widehat{\rho}_{r_p,\text{10-y US T-Note}}$	0.02	-0.01	0.02	0.09	0.07	0.13**
Turnover	11.60	16.43	9.58	13.64	35.36	1.13
Euro FX						
VRR	**0.47**	-2.81*	-0.95	-12.28***	-5.95*	-12.36***
$\widehat{\rho}_{r_p,\text{Euro FX}}$	-0.04	-0.07	-0.03	-0.17***	-0.08	-0.29***
Turnover	8.58	13.13	5.61	10.03	22.01	0.75
VIX						
VRR	64.60	**65.18**	62.48*	63.82	63.52*	60.34**
$\widehat{\rho}_{r_p,\text{VIX}}$	-0.04	-0.04	-0.09	-0.03	-0.03	0.03
Turnover	1.25	1.31	0.49	0.79	1.47	0.07

Note: This table shows the VRR, correlation between the out-of-sample hedged portfolio returns and the futures returns, and portfolio turnover of the VICC, DCC, EWMA, GO-GARCH, HS and OLS hedge ratio for bivariate cross hedging the S&P 500 index using the CBOT 10-y US T-Note, the CME Euro FX and the CBOE VIX futures. The VRR is defined in Equation (16) and the portfolio turnover is defined in Equation (17). Numbers in bold indicate that it is the highest VRR in the set. The DM-test is used to determine whether the outperformance in VRR is significant with a Null hypothesis of equal or worse performance by the best performing hedge ratio, and the significance of the correlation is tested with a Null hypothesis of no correlation. The significance at the 10%, 5%, and 1% levels are denoted as *, **, and ***, respectively.

icantly outperforms all the competing models in at least one of the three bivariate cross hedging applications. Importantly, these benefits are not eliminated by a large turnover as the VICC has a lower turnover, on average, compared with the second and third best model, namely the DCC and GO-GARCH hedge ratios. This makes the VICC more appealing to use in practice. A last important finding is that the use of the unconditional OLS hedge ratio significantly underperforms in all the applications and fails to eliminate the exposure to the interest rate and the exchange rate between the US Dollar and the Euro. This finding stresses the importance of conditional hedging.

4.5. *Multivariate results*

Table 6 shows the VRR, the correlation between the out-of-sample hedged portfolio returns and the futures returns, and the turnover of the VICC, DCC, EWMA,

Table 6. Multivariate hedging performance of the VICC, DCC, EWMA, GO-GARCH and OLS hedge ratio for cross hedging the S&P 500 index using the 10-y US T-Note, Euro FX and VIX futures.

	VICC	DCC	EWMA	GO GARCH	OLS
VRR	**65.30**	62.05**	62.31*	39.37***	60.87*
$\widehat{\rho}_{r_p,\text{10-y US T-Note}}$	0.02	-0.03	0.03	0.01	0.05
$\widehat{\rho}_{r_p,\text{Euro FX}}$	0.04	0.06	0.00	-0.06	-0.16***
$\widehat{\rho}_{r_p,\text{VIX}}$	-0.03	-0.04	-0.10*	-0.04	-0.07
Turnover	6.21	7.80	2.97	14.18	0.45

Note: This table shows the VRR, correlation between the out-of-sample hedged portfolio returns and the futures returns, and portfolio turnover of the VICC, DCC, EWMA, GO-GARCH and OLS hedge ratio for bivariate cross hedging the S&P 500 index using the CBOT 10-y US T-Note, the CME Euro FX and the CBOE VIX futures. The VRR is defined in Equation (16) and the portfolio turnover is defined in Equation (17). Numbers in bold indicate that it is the highest VRR in the set. The DM-test is used to determine whether the outperformance in VRR is significant with a Null hypothesis of equal or worse performance by the best performing hedge ratio, and the significance of the correlation is tested with a Null hypothesis of no correlation. The significance at the 10%, 5%, and 1% levels are denoted as *, **, and ***, respectively.

GO-GARCH and OLS hedge ratios for the multivariate cross hedging application.[5] In terms of VRR, we find that the VICC hedge ratio is the top performer with a VRR of 65.30 percentage points, and that it significantly outperforms all of the other considered models. The economic gains vary between 3 to 5 percentage points compared to the DCC, EWMA and OLS hedge ratios. Surprisingly, the GO-GARCH hedge ratio performs drastically worse with a VRR of only 39.37 percentage points.

The correlation coefficients of the out-of-sample hedged portfolio returns with the futures returns indicate that the VICC, DCC and GO-GARCH hedge ratio succeed in simultaneously eliminating the interest rate, the exchange rate between the US Dollar and the Euro, and the VIX index exposure from the S&P 500 portfolio. The EWMA hedge ratio fails to do so for the VIX exposure with a significant correlation of -0.10, and the unconditional OLS hedge ratio for the exchange rate exposure with a significant correlation of -0.16. Again the OLS hedge ratio generates the lowest turnover at a cost of a less (and even non-)effective hedging strategy, while the VICC hedge ratio is considerably stabler than the DCC and GO-GARCH hedge ratio. Finally, note that the regularization weight (κ_t) is never different from zero, meaning that the resulting VICC matrix is always positive-definite by itself.

[5]Obtaining the multivariate hedge ratios requires the inverse of the 3×3 covariance matrix of the futures returns ($\mathbf{H}_{ff,t}$). However, the use of an improper correlation estimator may cause some severe outliers in the HS hedge ratios (i.e. $|\widehat{\rho}_t^{\text{HS}}| > 1$). After truncation of the correlations at 1 or -1 as appropriate, we try two methods to obtain a positive-definite correlation matrix, namely via our proposed regularization method, and via a brute force method which extracts the negative eigenvalues via an eigendecomposition and sets these equal to a small positive number (i.e. 10^{-1}) before reconstructing the correlation matrix. While all these methods lead to an invertible covariance matrix, the resulting multivariate hedge ratios often take extreme values. As an example, for the 10-y US T-Note futures, the HS hedge ratio ranges between -16.23 and 21.07, while the VICC hedge ratio ranges between -1.85 and 0.83. The multivariate HS hedge ratios are therefore not usable in practice and lead to severe underperformance when cross hedging. In practice, additional truncations are needed, but this is beyond the scope of our paper. The results with and without truncation are available from the authors upon request.

Overall, we find that, in terms of variance reduction, the VICC hedge ratio significantly outperforms all the competing models in the multivariate hedging application, and that the outperformance entails some substantial economic gains. Moreover, the VICC succeeds in eliminating all the relevant exposures with a manageable turnover. Again, we find that the unconditional OLS hedge ratio fails to eliminate the exposure to the exchange rate between the US Dollar and the Euro, hence confirming the importance of dynamic hedging.

5. Conclusion

Accurate estimates of the conditional correlations between asset returns are of great practical importance in a lot of financial applications, such as portfolio optimization, hedging and risk management. Traditionally, they are jointly estimated via MGARCH models which require the optimization of a multivariate likelihood function. This can be numerically challenging and can lead to parameter instability in the case of a general parametrization. Moreover, most MGARCH models suffer from the so-called curse of dimensionality.

To avoid these issues we present a computationally simple and flexible filter to predict the time-varying correlation matrix of asset returns. The proposed Variance Implied Conditional Correlation (VICC) filter exploits the polarization result that links the correlation between two standardized variables with the variances of linear combinations of their standardized values. By using flexible univariate GARCH models the VICC only requires the estimation of univariate likelihood functions and the curse of dimensionality is avoided.

We assess the reliability of the VICC estimates by performing a Monte Carlo study and find that the VICC yields accurate correlation estimates for common choices of correlation dynamics. We also study the cross hedging of the S&P 500 against changes in (either) the interest rate, the exchange rate between the US Dollar and the Euro, and the VIX. We conclude that the VICC either leads to a better, or to an equal, hedging performance compared to standard benchmark models. Moreover, in case of an equal performance it has the benefit of providing substantially more stable hedge ratio estimates.

The proposed VICC is a flexible framework for dynamic correlation filtering. In this paper, we have focussed on the implementation with a symmetric GARCH(1,1) variance model in the VICC filter and the application to cross hedging. It would be interesting to investigate the added value of asymmetric news impact surfaces when using the VICC filter in practice and studying the usefulness in other applications, such as high-dimensional dynamic portfolio allocation.

Acknowledgements

A preliminary version of this paper previously circulated as “The variance implied hedge ratio”. We thank David Ardia, Filip Bekarovski, Samuel Borms, Nabil Bouamara, Prabesh Luitel, Pieter-Jan Pauwelyn, Stefan Straetmans, Marjan Wauters, and participants at the workshop in honor of Peter Rousseeuw’s 60th birthday in Leuven, the 2018 CFE conference in London, the 8th international MAF conference in Madrid, the 16th meeting of the BFRF in Brussels, and the 4th edition of EAJ Conference in Leuven for their valuable comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The views expressed in this paper are the sole responsibility of the authors. Any remaining errors or shortcomings are those of the authors.

References

Aït-Sahalia, Yacine, Jianqing Fan, and Dacheng Xiu. 2010. "High-frequency covariance estimates with noisy and asynchronous financial data." *Journal of the American Statistical Association* 105 (492): 1504–1517.

Bauwens, Luc, Sébastien Laurent, and Jeroen V.K. Rombouts. 2006. "Multivariate GARCH models: A survey." *Journal of Applied Econometrics* 21 (1): 79–109.

Bollerslev, Tim. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 31 (3): 307–327.

Bollerslev, Tim. 2008. "Glossary to ARCH (GARCH)." In T. Bollerslev, J. R. Russell & M. Watson (Eds.), Festschrift in Honor of Robert F. Engle.

Boudt, Kris, Jón Daníelsson, and Sébastien Laurent. 2013. "Robust forecasting of dynamic conditional correlation GARCH models." *International Journal of Forecasting* 29 (2): 244 –257.

Boudt, Kris, Alexios Ghalanos, Scott Payseur, and Eric Zivot. 2019. "Multivariate GARCH models for large-scale applications: A survey." In *H.D. Vinod and C.R. Rao (Ed.) Handbook of Statistics, Volume 41*, forthcoming.

Boudt, Kris, Peter J. Rousseeuw, Steven Vanduffel, and Tim Verdonck. 2018. "The Minimum Regularized Covariance Determinant estimator." *Working paper* .

Box, George E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71 (356): 791–799.

Broda, Simon A., and Marc S. Paolella. 2009. "CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation." *Journal of Financial Econometrics* 7 (4): 412–436.

Caporin, Massimiliano, and Michael McAleer. 2013. "Ten Things You Should Know about the Dynamic Conditional Correlation Representation." *Econometrics* 1 (1): 115–126.

Cappiello, Lorenzo, Robert F. Engle, and Kevin Sheppard. 2006. "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns." *Journal of Financial Econometrics* 4 (4): 537–572.

Chen, Sheng-Syan, Cheng-Few Lee, and Keshab Shrestha. 2003. "Futures hedge ratios: A review." *The Quarterly Review of Economics and Finance* 43 (3): 433–465.

Creal, Drew, Siem Jan Koopman, and André Lucas. 2011. "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations." *Journal of Business & Economic Statistics* 29 (4): 552–563.

Daníelsson, Jón. 2011. *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*. John Wiley & Sons.

Diebold, Francis X., and Robert S. Mariano. 2002. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 20 (1): 134–144.

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

Engle, Robert F. 2009. *Anticipating correlations: A new paradigm for risk management*. Princeton University Press.

Engle, Robert F. 2016. "Dynamic Conditional Beta." *Journal of Financial Econometrics* 14 (4): 643–667.

Engle, Robert F., and Kenneth F. Kroner. 1995. "Multivariate simultaneous generalized ARCH." *Econometric theory* 11 (1): 122–150.

Engle, Robert F., and Victor K. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *The Journal of Finance* 48 (5): 1749–1778.

Ghalanos, Alexios. 2016. *rmgarch: Multivariate GARCH models.* R package version 1.3-0, https://cran.r-project.org/package=rmgarch.
Ghalanos, Alexios. 2018. *rugarch: Univariate GARCH Models.* R package version 1.4-0, https://cran.r-project.org/package=rugarch.
Gnanadesikan, R., and J.R. Kettenring. 1972. "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data." *Biometrics* 28 (1): 81–124.
Harris, Richard D.F., and Jian Shen. 2003. "Robust estimation of the optimal hedge ratio." *Journal of Futures Markets* 23 (8): 799–816.
Iwashita, Toshiya, and Minoru Siotani. 1994. "Asymptotic distributions of functions of a sample covariance matrix under the elliptical distribution." *Canadian Journal of Statistics* 22 (2): 273–283.
Kavussanos, Manolis G., and Ilias D. Visvikis. 2008. "Hedging effectiveness of the Athens stock index futures contracts." *The European Journal of Finance* 14 (3): 243–270.
Kroner, Kenneth F., and Jahangir Sultan. 1993. "Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures." *The Journal of Financial and Quantitative Analysis* 28 (4): 535–551.
Ledoit, Olivier, and Michael Wolf. 2004. "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis* 88 (2): 365–411.
Lucas, André, Bernd Schwaab, and Xin Zhang. 2017. "Modeling Financial Sector Joint Tail Risk in the Euro Area." *Journal of Applied Econometrics* 32 (1): 171–191.
Lypny, G., and M. Powalla. 1998. "The hedging effectiveness of DAX futures." *The European Journal of Finance* 4 (4): 345–355.
Ma, Yanyuan, and Marc G. Genton. 2002. "Highly Robust Estimation of the Autocovariance Function." *Journal of Time Series Analysis* 21 (6): 663–684.
Pagan, Adrian R., and G.William Schwert. 1990. "Alternative models for conditional stock volatility." *Journal of Econometrics* 45 (1): 267–290.
Pakel, Cavit, Neil Shephard, Kevin Sheppard, and Robert F. Engle. 2019. "Fitting vast dimensional time-varying covariance models." *Working paper* .
Pozzi, F., T. Di Matteo, and T. Aste. 2012. "Exponential smoothing weighted correlations." *The European Physical Journal B* 85 (6): 175.
van der Weide, Roy. 2002. "GO-GARCH: A multivariate generalized orthogonal GARCH model." *Journal of Applied Econometrics* 17 (5): 549–564.
Wang, Yudong, Chongfeng Wu, and Li Yang. 2015. "Hedging with Futures: Does Anything Beat the Naïve Hedging Strategy?" *Management Science* 61 (12): 2870–2889.
Zhang, Lan. 2011. "Estimating covariation: Epps effect, microstructure noise." *Journal of Econometrics* 160 (1): 33–47.

Appendix A. Optimal γ in the Gaussian case

Under the assumptions of Property 2, we have that Z_i and Z_j are bivariate standard normally distributed with correlation coefficient ρ_{ij}. We maximize estimation efficiency by minimizing the scaled asymptotic variance of the sample variance of the weighted sum between Z_i and Z_j. This leads to the following objective function:

$$S_+(\gamma) = \text{Avar}\left(\sqrt{T-1}\ \widehat{h}_{i+j}(\gamma)\right), \tag{A1}$$

where Avar $(\cdot)$ is the asymptotic variance operator and $\widehat{h}_{i+j}(\gamma)$ is the sample variance estimator of the weighted sum of Z_i and Z_j. We now assume there are T observations of Z_i and Z_j collected in samples z_i and z_j. We then have:

$$\begin{aligned}
\widehat{h}_{i+j}(\gamma) &= \frac{1}{T-1}\sum_{t=1}^{T}\left(\gamma z_{i,t} + (1-\gamma)z_{j,t} - \gamma\bar{z}_i - (1-\gamma)\bar{z}_j\right)^2 \\
&= \gamma^2 \underbrace{\frac{1}{T-1}\sum_{t=1}^{T}\left(z_{i,t}-\bar{z}_i\right)^2}_{\widehat{h}_i} + (1-\gamma)^2 \underbrace{\frac{1}{T-1}\sum_{t=1}^{T}\left(z_{j,t}-\bar{z}_j\right)^2}_{\widehat{h}_j} \\
&+ 2\gamma(1-\gamma) \underbrace{\frac{1}{T-1}\sum_{t=1}^{T}\left(z_{i,t}-\bar{z}_i\right)\left(z_{j,t}-\bar{z}_j\right)}_{\widehat{h}_{ij}},
\end{aligned} \tag{A2}$$

where $\bar{z}_i$ and $\bar{z}_j$ are the sample means, $\widehat{h}_i$ and $\widehat{h}_j$ are the sample variances, and $\widehat{h}_{ij}$ is the sample covariance of $z_{i,t}$ and $z_{j,t}$. Under the above assumptions we have that, for $T \to \infty$:

$$\mathrm{Cov}\left(\sqrt{T-1}\begin{pmatrix} \widehat{h}_i - 1 \\ \widehat{h}_{ij} - \rho_{ij} \\ \widehat{h}_j - 1 \end{pmatrix}\right) = \begin{pmatrix} 2 & 2\rho_{ij} & 2\rho_{ij}^2 \\ 2\rho_{ij} & 1+\rho_{ij}^2 & 2\rho_{ij} \\ 2\rho_{ij}^2 & 2\rho_{ij} & 2 \end{pmatrix}, \tag{A3}$$

see e.g., Equation (2.1) in Iwashita and Siotani (1994). We can then write $S_+(\gamma)$ as:

$$8\gamma^4\left(\rho_{ij}^2 - 2\rho_{ij} + 1\right) - 16\gamma^3\left(\rho_{ij}^2 - 2\rho_{ij} + 1\right) + 8\gamma^2\left(\rho_{ij}^2 - 3\rho_{ij} + 2\right) + 8\gamma\left(\rho_{ij} - 1\right) + 2. \tag{A4}$$

Its first and second order derivatives are:

$$\begin{aligned}
S_+'(\gamma) &= 32\gamma^3\left(\rho_{ij}^2 - 2\rho_{ij} + 1\right) - 48\gamma^2\left(\rho_{ij}^2 - 2\rho_{ij} + 1\right) + 16\gamma\left(\rho_{ij}^2 - 3\rho_{ij} + 2\right) + 8\left(\rho_{ij} - 1\right), \\
S_+''(\gamma) &= 96\gamma^2\left(\rho_{ij}^2 - 2\rho_{ij} + 1\right) - 96\gamma\left(\rho_{ij}^2 - 2\rho_{ij} + 1\right) + 16\left(\rho_{ij}^2 - 3\rho_{ij} + 2\right).
\end{aligned} \tag{A5}$$

From the first and second order conditions, it is trivial to see that $\gamma = 0.5$ minimizes $S_+(\gamma)$ for $\gamma \in [0,1]$ and $|\rho_{ij}| < 1$. The same holds for $\widehat{h}_{i-j}(\gamma)$ for which the objective function is:

$$S_-(\gamma) = \mathrm{Avar}\left(\sqrt{T-1}\,\widehat{h}_{i-j}(\gamma)\right), \tag{A6}$$

where $\widehat{h}_{i-j}(\gamma)$ is the sample variance estimator constructed using $\gamma z_{i,1} - (1-\gamma)z_{j,1}, \ldots, \gamma z_{i,T} - (1-\gamma)z_{j,T}$. Its first and second order derivatives are:

$$\begin{aligned}
S_-'(\gamma) &= 32\gamma^3\left(\rho_{ij}^2 + 2\rho_{ij} + 1\right) - 48\gamma^2\left(\rho_{ij}^2 + 2\rho_{ij} + 1\right) + 16\gamma\left(\rho_{ij}^2 + 3\rho_{ij} + 2\right) - 8\left(\rho_{ij} + 1\right), \\
S_-''(\gamma) &= 96\gamma^2\left(\rho_{ij}^2 + 2\rho_{ij} + 1\right) - 96\gamma\left(\rho_{ij}^2 + 2\rho_{ij} + 1\right) + 16\left(\rho_{ij}^2 + 3\rho_{ij} + 2\right).
\end{aligned} \tag{A7}$$

We find that $\gamma = 0.5$ satisfies the first order conditions for a minimum.

Appendix B. Univariate GARCH model

Denote the squared innovation in the return by $\varepsilon_{i,t}^2 = (r_{i,t} - \mu_{i,t})^2$, with $\mu_{i,t}$ the predicted return, conditional on the information available at time $t-1$. The standard GARCH(1,1) specification of Bollerslev (1986) models the conditional variance as a linear combination of the lagged squared error term $\varepsilon_{i,t-1}^2$ and the lagged conditional variance $h_{i,t-1}$. This leads to the following equation:

$$h_{i,t} = \omega_i + \alpha_i \ \varepsilon_{i,t-1}^2 + \beta_i \ h_{i,t-1}. \tag{B1}$$

The coefficients ω_i, α_i and β_i should all be positive to ensure a positive variance and the sum of α_i and β_i should be lower than one for the conditional variance process to be covariance stationary. The model is estimated via Gaussian (Quasi) Maximum Likelihood estimation using the rugarch package of Ghalanos (2018). We refer to the original paper of Bollerslev (1986) for a more detailed discussion.

Appendix C. Calibration of the regularization parameter κ_t

Given that $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}}$ is positive-definite, let $\mathbf{G}_{t-1}$ be the square root of $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}}$ obtained by using the Cholesky factorization such that $\widehat{\mathbf{R}}_{t-1}^{\text{VICC}} = \mathbf{G}_{t-1}\mathbf{G}_{t-1}^{\intercal}$. We can rewrite Equation (5) as follows:

$$\begin{aligned} \widehat{\mathbf{R}}_t^{\text{VICC}} &= (1-\kappa_t)\ \widehat{\mathbf{R}}_t + \kappa_t\ \mathbf{G}_{t-1}\ \mathbf{G}_{t-1}^{\intercal}, \\ &= \mathbf{G}_{t-1} \left[(1-\kappa_t)\ \mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal} + \kappa_t\ \mathcal{I}_N \right]\ \mathbf{G}_{t-1}^{\intercal}. \end{aligned} \tag{C1}$$

Since $\mathbf{G}_{t-1}$ is positive-definite, $\widehat{\mathbf{R}}_t^{\text{VICC}}$ is positive-definite if $(1-\kappa_t)\ \mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal} + \kappa_t\ \mathcal{I}_N$ is positive-definite. We ensure this by setting κ_t such that the lowest eigenvalue of $(1-\kappa_t)\ \mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal} + \kappa_t\ \mathcal{I}_N$ equals a small positive number $\psi_{\min}$. This is the case for the regularized VICC correlation matrix in Equation (6). To see this, note first that the eigenvalues of $(1-\kappa_t)\ \mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal} + \kappa_t\ \mathcal{I}_N$ are linearly related to those of $\mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal}$. Recall that $\lambda_{\min,t}$ is defined as the smallest eigenvalue of $\mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal}$ and note that, if the vector of eigenvalues of $\mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal}$ is $\boldsymbol{\lambda}_t$, then the eigenvalues of $(1-\kappa_t)\ \mathbf{G}_{t-1}^{-1}\ \widehat{\mathbf{R}}_t\ \mathbf{G}_{t-1}^{-\intercal} + \kappa_t\ \mathcal{I}_N$ are $(1-\kappa_t)\boldsymbol{\lambda}_t + \kappa_t$. It then follows that:

$$\begin{aligned} (1-\kappa_t)\lambda_{\min,t} + \kappa_t &= \psi_{\min}, \\ (1-\lambda_{\min,t})\kappa_t &= \psi_{\min} - \lambda_{\min,t}, \\ \kappa_t &= \max\left\{ \frac{\psi_{\min} - \lambda_{\min,t}}{1-\lambda_{\min,t}}, 0 \right\}. \end{aligned} \tag{C2}$$

Appendix D. EWMA model

Let the $N \times 1$ innovation vector in the returns $\mathbf{r}_t$ be $\boldsymbol{\epsilon}_t = \mathbf{r}_t - \widehat{\boldsymbol{\mu}}_t$, where $\widehat{\boldsymbol{\mu}}_t$ is a $N \times 1$ vector with the sample averages of the returns from inception till time $t-1$. The

EWMA prediction is then as follows:

$$\mathbf{H}_t = (1-\delta)\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\intercal} + \delta\mathbf{H}_{t-1}, \tag{D1}$$

where $\mathbf{H}_t$ is the conditional covariance matrix and δ is the EWMA smoothing parameter. As usual (see e.g., Daníelsson (2011) and Engle (2009)), we set $\delta = 0.94$ and initialize the process with the unconditional sample covariance matrix. The EWMA correlation matrix can then be easily obtained as follows:

$$\mathbf{R}_t = diag\{dg(\mathbf{H}_t)\}^{-1/2}\ \mathbf{H}_t\ diag\{dg(\mathbf{H}_t)\}^{-1/2}, \tag{D2}$$

where the matrix operator $dg(\cdot)$ returns a vector that contains the elements of the main diagonal and $diag\{\cdot\}$ creates a diagonal matrix.

Appendix E. DCC model

Let $\mathbf{r}_t$ be a $N \times 1$ return vector with the following joint dynamics:

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \mathbf{D}_t\boldsymbol{z}_t, \tag{E1}$$

where $\boldsymbol{\mu}_t$ is a vector with the conditional means, $\mathbf{D}_t$ is a diagonal matrix with the conditional standard deviations of the returns on the main diagonal and $\boldsymbol{z}_t$ is a vector with the standardized error terms. The conditional covariance matrix $\mathbf{H}_t$ can be expressed as:

$$\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t, \tag{E2}$$

in which $\mathbf{R}_t$ is the conditional correlation matrix. The diagonal elements of $\mathbf{D}_t$ are obtained via the standard GARCH(1,1) model. The DCC model then specifies the dynamics in $\mathbf{R}_t$ using the standardized values $\boldsymbol{z}$ as follows:

$$\begin{aligned} \boldsymbol{Q}_t &= (1-\omega_1-\omega_2)\ \overline{\boldsymbol{Q}} + \omega_1\ \boldsymbol{z}_t\boldsymbol{z}_t^{\intercal} + \omega_2\ \boldsymbol{Q}_{t-1}, \\ \mathbf{R}_t &= diag\{dg(\boldsymbol{Q}_t)\}^{-1/2}\ \boldsymbol{Q}_t\ diag\{dg(\boldsymbol{Q}_t)\}^{-1/2}, \end{aligned} \tag{E3}$$

where $\boldsymbol{Q}_t$ is a symmetric positive definite matrix and $\overline{\boldsymbol{Q}}$ is the unconditional covariance matrix of $\boldsymbol{z}_t$. The scalar parameters ω_1 and ω_2 are strictly positive and their sum is strictly lower than one. The DCC model is estimated using the standard two-step Gaussian (Quasi) Maximum Likelihood estimation as implemented in the rmgarch package of Ghalanos (2016). We refer to the original paper of Engle (2002) for a more detailed discussion.

Appendix F. GO-GARCH model

Following the GO-GARCH model of van der Weide (2002) the innovation vector $\boldsymbol{\epsilon}_t$ is modelled as a linear combination of N unobserved factors $\mathbf{f}_t$:

$$\boldsymbol{\epsilon}_t = \mathbf{A}\mathbf{f}_t, \tag{F1}$$

where $\mathbf{A}$ is a time-invariant, invertible mixing matrix. The factors are specified as follows:

$$\mathbf{f}_t = \mathbf{\Sigma}_t^{1/2}\mathbf{v}_t, \tag{F2}$$

where $\mathbf{\Sigma}_t^{1/2}$ is a diagonal matrix with the conditional variances of the factors, and $v_{i,t}$ is an iid random variable with zero mean and unit variance. The variances of the factors are in our case modelled via the standard GARCH(1,1) model. The conditional covariance matrix $\mathbf{H}_t$ is $\mathbf{A}\mathbf{\Sigma}_t\mathbf{A}^\intercal$. To estimate the GO-GARCH model we follow Broda and Paolella (2009) and use the implementation in the rmgarch package of Ghalanos (2016).

Appendix G. Harris and Shen (2003)

Harris and Shen (2003) propose to estimate the conditional covariance $h_{ij,t}$ as follows:

$$\widehat{h}_{ij,t}^{\text{HS}} = \frac{\widehat{h}_{+,t} - \widehat{h}_{-,t}}{4}, \tag{G1}$$

where $\widehat{h}_{+,t}$ is the estimated conditional variance of the sum of $r_{i,t}$ and $r_{j,t}$ and $\widehat{h}_{-,t}$ that of the difference between them. Both are obtained by using the standard GARCH(1,1) model. Note that the correlation embedded in the HS method is an unbounded function of the conditional variance estimates $\widehat{h}_{i,t}$ and $\widehat{h}_{j,t}$ and may therefore be sensitive to extreme observations in the data. This may even lead to $|\widehat{h}_{ij,t}^{\text{HS}}/\sqrt{\widehat{h}_{i,t}\widehat{h}_{j,t}}| > 1$. We refer to Harris and Shen (2003) for a more detailed discussion.