Supplementary Internet Appendix for

Are Correlations Constant? Empirical and Theoretical Results on Popular Correlation Models in Finance*

Zeno Adams, Roland Füss, Thorsten Glück

This Internet Appendix contains the following supplementary content:

Appendix A: Empirical Results for Other Popular MGARCH Models

Appendix B: The Scalar $\hat{D}$ for the Test Statistic of Wied, Krämer and Dehling (2012)

Appendix C: List of Assets

Appendix D: Approximation Accuracy of $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$

Appendix E. Proof of Proposition 1

Appendix F. Correlations and Non-Elliptical Return Distributions
Appendix A: Empirical Results for Other Popular MGARCH Models (referring to page 3)

In this paper we focus on Engle’s DCC model which appears to be the most popular and most widely used MGARCH model in empirical research. In the following, we will show that our findings are not confined to the DCC specification but also extend to other more complex MGARCH models. In particular, we find that correlation breaks have similar implications for the coefficient estimates of the diagonal VECH model (Bollerslev, Engle, and Wooldridge, 1988), the BEKK model (Engle and Kroner, 1995), and the corrected DCC model (Aielli, 2013). While differences in model specifications prevent a direct comparison between models, we can observe a distinct change in coefficient estimates across all models.

The diagonal VECH model of Bollerslev, Engle, and Wooldridge (1988) is a restricted version of the more general VECH model and is expressed as

\[
H_t = \Omega + A \otimes \epsilon_{t-1} \epsilon_{t-1}' + B \otimes H_{t-1}, \tag{A.1}
\]

where the parameter matrices \(A\) and \(B\) are indefinite matrices, i.e. the parameters can vary without any restrictions. While this specification does not ensure that the conditional covariance matrix is positive semidefinite, it is also the most general way of writing the diagonal VECH model.

The diagonal BEKK model of Engle and Kroner (1995) is defined as

\[
H_t = \Omega \Omega + A \epsilon_{t-1} \epsilon_{t-1}' A' + B H_{t-1} B'. \tag{A.2}
\]

The general form of the BEKK model in which \(A\) and \(B\) are unrestricted contains many parameters and is computationally expensive. We therefore use the more common diagonal form in which \(A\) and \(B\) are restricted to be diagonal matrices.

Aielli (2013) shows that the constant \(\theta_{i,j}\) in \(\hat{q}_{i,j,t} = (1-a-b) \theta_{i,j} + a \epsilon_{t-1} \epsilon_{t-1} + b \hat{q}_{i,j,t-1}\) can be inconsistent. The constant \(\theta_{i,j}\) is thought of as the second moment of \(\epsilon_t\) or \(\theta = \mathbb{E}(\epsilon_t \epsilon_t')\). For certain parameter values and large systems containing many assets, this may not be the case.
which can cause the DCC estimator to be biased. However, Aielli also shows that for parameter values that are common for financial applications the bias is negligible. The correction proposed by Aielli (2013) is

\[
\hat{q}_{i,j} = (1-a-b)q_{i,j} + a \cdot \hat{q}_{i,j} \cdot e_{i,j-1} e_{j,j-1} + b \hat{q}_{i,j,j-1}.
\] (A.3)

We estimate the three models for all assets in our data set. The dynamic correlations produced by the models are very similar. For instance, Figure A1 shows the daily dynamic correlations between the S&P 500 and the NASDAQ composite index. The upper graph is based on the DCC model and is the same as Figure 1 shown in the introduction. The lower graph shows the deviations of the DCC model from the corrected DCC (cDCC), the diagonal Vech, and the diagonal BEKK model. As can be seen, the differences are quite small, so that we should expect the findings in this paper to hold also for other autoregressive MGARCH specifications.
This figure shows typical correlation dynamics generated by a DCC model as well as the deviations from these DCC correlations. The deviations of other popular correlation models are quite small indicating that the results in this paper also extend to other frequently used multivariate GARCH models.
Table A1 repeats the correlation coefficient analysis discussed earlier in Panel A of Figure 6, this time comparing coefficients across different MGARCH models. The first column shows how the news and decay parameters change from full samples that contain at least one correlation break to subsamples that occur between breaks. While the news parameter increases slightly, the decay parameter decreases by 0.134 to 0.836. As we have shown in the paper, this reduction is sufficiently large to eliminate most of the variation in $\hat{\rho}$. Columns 2 to 4 show a similar decrease in the persistence parameter for the other three models. The reduction is not as large as in the case of the DCC model but it moves the sum of both parameters away from the 0.99 threshold at which correlation dynamics become very pronounced (Aielli, 2013). From the findings in Table A1 we conclude that typical autoregressive MGARCH specifications respond in a similar way to correlations breaks so that our results based on the DCC model can be also extended to other MGARCH specifications.

**Table A1: The Impact of Correlation Breaks across Different MGARCH Models**

<table>
<thead>
<tr>
<th></th>
<th>DCC</th>
<th>Corrected DCC</th>
<th>VECH</th>
<th>BEKK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News Par.</td>
<td>0.021</td>
<td>0.021</td>
<td>0.025</td>
<td>0.049</td>
</tr>
<tr>
<td>Decay Par.</td>
<td>0.970</td>
<td>0.970</td>
<td>0.941</td>
<td>0.939</td>
</tr>
<tr>
<td><strong>Subsample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News Par.</td>
<td>0.027 (0.006)</td>
<td>0.025 (0.004)</td>
<td>0.026 (0.001)</td>
<td>0.045 (-0.004)</td>
</tr>
<tr>
<td>Decay Par.</td>
<td>0.836 (-0.134)</td>
<td>0.892 (-0.078)</td>
<td>0.894 (-0.047)</td>
<td>0.909 (-0.03)</td>
</tr>
</tbody>
</table>

This table shows that the presence of correlation breaks affects the persistence parameter and hence the overall correlation dynamics of all four popular MGARCH models in a similar way.
Appendix B: The Scalar \( \hat{D} \) for the Test Statistic of Wied, Krämer and Dehling (2012) (referring to page 13)

We briefly describe the construction of the scalar \( \hat{D} \) which is part of the expression in Equation (3). For a general and in-depth treatment we refer to Wied, Krämer, and Dehling (2012, Appendix A.1). Let \( \{(x_{t,i}, x_{t,j})\} \) be the bivariate time-series with \( E[(x_{t,i}, x_{t,j})] = 0 \).

Given is a sample of size \( T \). For \( i = 1, 2 \), denote \( \bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{t,i} \), \( \bar{x}_i^2 = T^{-1} \sum_{t=1}^{T} x_{t,i}^2 \) and \( \hat{\sigma}_x = \sqrt{\bar{x}_i^2 - \bar{x}_i^2} \). Further, denote \( x_1 x_2 = T^{-1} \sum_{t=1}^{T} x_{t,1} x_{t,2} \) and \( \hat{\sigma}_{x_1 x_2} = \bar{x}_1 \bar{x}_2 - \bar{x}_1 \bar{x}_2 \). Let \( k(\cdot) \) be the Bartlett kernel function. The scalar \( \hat{D} \) is then given by

\[
\hat{D} = \sqrt{\hat{D}_1 \hat{D}_2 \hat{D}_3 \hat{D}_4}, \tag{B.1}
\]

where

\[
\hat{D}_1 = \sum_{i=1}^{T} \sum_{u=1}^{T} k\left( \frac{t-u}{\log T} \right) V_t V_u',
\]

with \( V_t = T^{-1/2} \left( x_{t,1}^2 - x_{1,1}^2, x_{t,2}^2 - x_{1,2}^2, x_{t,1} - x_{1,1}, x_{t,2} - x_{1,2}, x_{t,1} x_{t,2} - x_{1,1} x_{1,2} \right)' \),

\[
\hat{D}_2 = \begin{pmatrix}
1 & 0 & -2x_1 & 0 & 0 \\
0 & 1 & 0 & -2x_1 & 0 \\
0 & 0 & -x_2 & -x_1 & 1
\end{pmatrix}, \tag{B.3}
\]

and

\[
\hat{D}_3 = \begin{pmatrix}
\frac{1}{2} \hat{\sigma}_{x_1 x_2} \hat{\sigma}_x^{-3} & -\frac{1}{2} \hat{\sigma}_{x_1 x_2} \hat{\sigma}_x^{-3} & \frac{1}{\hat{\sigma}_{x_1 x_2}}
\end{pmatrix}.
\]

The purpose of the scalar \( \hat{D} \) is to appropriately rescale the cumulated sum of empirical correlation coefficients in such a way that convergence of \( Q_t \) to the asymptotic null distribution is achieved.
## Appendix C: List of Assets (referring to page 15)

<table>
<thead>
<tr>
<th>Name</th>
<th>DS Mnemonic</th>
<th>Name</th>
<th>DS Mnemonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500</td>
<td>S&amp;PCOMP</td>
<td>US 10 yr. Gov.</td>
<td>BMUS10Y</td>
</tr>
<tr>
<td>TSX</td>
<td>TTOCOMP</td>
<td>CA 10 yr. Gov.</td>
<td>BMCN10Y</td>
</tr>
<tr>
<td>MIB</td>
<td>FTSEMIB</td>
<td>IT 10 yr. Gov.</td>
<td>BMIT10Y</td>
</tr>
<tr>
<td>OMX Stockholm</td>
<td>SWEDOMX</td>
<td>SE 10 yr. Gov.</td>
<td>BMSD10Y</td>
</tr>
<tr>
<td>DAX 30</td>
<td>DAXINDZ</td>
<td>GER 10 yr. Gov.</td>
<td>ABDGVG4</td>
</tr>
<tr>
<td>CAC 40</td>
<td>FRCAC40</td>
<td>FR 10 yr. Gov.</td>
<td>BMFR10Y</td>
</tr>
<tr>
<td>AEX</td>
<td>AMSTEOE</td>
<td>NL 10 yr. Gov.</td>
<td>BMNL10Y</td>
</tr>
<tr>
<td>NIKKEI 225</td>
<td>JAPDOWA</td>
<td>JP 10 yr. Gov.</td>
<td>BMJP10Y</td>
</tr>
<tr>
<td>KOSPI</td>
<td>KORCOMP</td>
<td>UK 10 yr. Gov.</td>
<td>BMUK10Y</td>
</tr>
<tr>
<td>FTSE100</td>
<td>FTSE100</td>
<td>CH 10 yr. Gov.</td>
<td>BMSW10Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>DS Mnemonic</th>
<th>Name</th>
<th>DS Mnemonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper</td>
<td>GSICTOT</td>
<td>USD–EUR</td>
<td>USEURSP</td>
</tr>
<tr>
<td>Corn</td>
<td>GSCNTOT</td>
<td>USD–JPY</td>
<td>JAPYNUS</td>
</tr>
<tr>
<td>Crude Oil</td>
<td>SGCRTOT</td>
<td>USD–CAD</td>
<td>CDNDLUS</td>
</tr>
<tr>
<td>Heating Oil</td>
<td>GSHOTOT</td>
<td>USD–KKW</td>
<td>SKORWUS</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>GSGNTOT</td>
<td>USD–SEK</td>
<td>SWEDKUS</td>
</tr>
<tr>
<td>Gold</td>
<td>GSGCTOT</td>
<td>USD–CHF</td>
<td>SWISFUS</td>
</tr>
<tr>
<td>Aluminum</td>
<td>GSIATOT</td>
<td>USD–MXN</td>
<td>MEXPFUS</td>
</tr>
<tr>
<td>Sugar</td>
<td>GSSBTOT</td>
<td>USD–GBP</td>
<td>BRITPUS</td>
</tr>
<tr>
<td>Cotton</td>
<td>GSCTTOT</td>
<td>USD–NOK</td>
<td>NORGKUS</td>
</tr>
<tr>
<td>Cattle</td>
<td>GSLCTOT</td>
<td>USD–BRL</td>
<td>BRAZLUS</td>
</tr>
</tbody>
</table>

This table lists the 40 assets that are used in the paper. All time series are from Thomson Reuters DataStream. The commodities are S&P GSCI Total Return Indices.
Appendix D: Approximation Accuracy of $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$ (referring to page 21)

To assess whether $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$ is a good approximation for $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$, we simulate a series of 500,000 innovations following the DGP as defined in Equation (10). We can then generate a series of conditional correlations $\hat{\rho}_t$ from the simulated sample according to Equation (1), where $\hat{\mathbf{q}}_t$ is defined by a rolling window estimator as in Equation (7), an EWMA model as in (8), or by the DCC specification of Equation (9). We can measure the approximation accuracy using the percentage deviation of the approximation $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$ from the sample value of $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$.

The parameter of the EWMA model is set to $\lambda = 0.94$, which is the daily parameter value proposed by RiskMetrics. The choice of the DCC model parameters is based on the estimated values derived from Engle and Sheppard (2001), Engle (2002), and Engle and Colacito (2006). Table D1 gives an overview of the ML estimates reported in these studies. The parameter $a$ ranges from 0.01 and 0.07, $b$ ranges from 0.91 to 0.98, and on average $1 - a - b$ is about 0.01. We therefore set $a$ equal to 0.02, 0.04, or 0.06, while $b$ is defined as $0.99 - a$. Finally, the window size $n$ of the rolling window estimator is set to 60, 120, and 240.

The simulation results are shown in Table D.1. For the EWMA model, $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$ tends to overstate $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$ for $\rho = 0.0$ and 0.3, and understate it for $\rho = 0.6$ and 0.9. For the DCC model, $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$ tends to overstate $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$, while the opposite appears to be true for the rolling window. Overall, the deviations range from -10.26% to 5.96%. 

### Table D1: Simulated Values $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$ versus the Approximation $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>EWMA</th>
<th>DCC with $b = 0.99 - \alpha$</th>
<th>Rolling Window</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda = 0.94$</td>
<td>$a = 0.02$</td>
<td>$a = 0.04$</td>
</tr>
</tbody>
</table>

**Panel A: lag $s = 0$:** $(\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_t) / \text{Cov}(\hat{\rho}_t, \hat{\rho}_t) - 1) \cdot 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>4.9943</td>
<td>1.4060</td>
<td>5.5689</td>
<td>5.9551</td>
<td>-1.1186</td>
<td>-0.2574</td>
<td>-1.2697</td>
<td>-1.2697</td>
</tr>
<tr>
<td>0.3</td>
<td>1.5494</td>
<td>1.4909</td>
<td>3.7652</td>
<td>3.8787</td>
<td>-2.6126</td>
<td>-0.5663</td>
<td>-0.5497</td>
<td>-0.5497</td>
</tr>
<tr>
<td>0.6</td>
<td>-1.3444</td>
<td>0.8873</td>
<td>0.5631</td>
<td>0.2382</td>
<td>-5.1407</td>
<td>-1.8598</td>
<td>-1.9055</td>
<td>-1.9055</td>
</tr>
<tr>
<td>0.9</td>
<td>-10.2614</td>
<td>0.0674</td>
<td>-2.4617</td>
<td>-6.0443</td>
<td>-9.1979</td>
<td>-5.4237</td>
<td>-0.6458</td>
<td>-0.6458</td>
</tr>
</tbody>
</table>

**Panel B: lag $s = 5$:** $(\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-5}) / \text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-5}) - 1) \cdot 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>4.2139</td>
<td>0.9175</td>
<td>5.0035</td>
<td>4.4564</td>
<td>-0.7880</td>
<td>-0.1607</td>
<td>-1.2761</td>
<td>-1.2761</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1484</td>
<td>1.1537</td>
<td>3.0081</td>
<td>2.0319</td>
<td>-2.3207</td>
<td>-0.4636</td>
<td>-0.5311</td>
<td>-0.5311</td>
</tr>
<tr>
<td>0.6</td>
<td>-2.0026</td>
<td>0.6319</td>
<td>0.1008</td>
<td>-0.8287</td>
<td>-4.8108</td>
<td>-1.7221</td>
<td>-1.8935</td>
<td>-1.8935</td>
</tr>
<tr>
<td>0.9</td>
<td>-10.1639</td>
<td>0.2974</td>
<td>-2.3910</td>
<td>-6.0815</td>
<td>-8.6375</td>
<td>-5.3226</td>
<td>-0.5925</td>
<td>-0.5925</td>
</tr>
</tbody>
</table>

**Panel C: lag $s = 10$:** $(\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-10}) / \text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-10}) - 1) \cdot 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
<th>$\hat{\rho}_t$</th>
<th>$\hat{\rho}_t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>4.0909</td>
<td>0.4415</td>
<td>4.5275</td>
<td>3.7571</td>
<td>-0.4778</td>
<td>-0.0782</td>
<td>-1.2807</td>
<td>-1.2807</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.9402</td>
<td>0.9954</td>
<td>2.3103</td>
<td>0.1669</td>
<td>-2.0155</td>
<td>-0.3886</td>
<td>-0.5222</td>
<td>-0.5222</td>
</tr>
<tr>
<td>0.6</td>
<td>-2.6713</td>
<td>0.3867</td>
<td>-0.0721</td>
<td>-1.6956</td>
<td>-4.5631</td>
<td>-1.5754</td>
<td>-1.8715</td>
<td>-1.8715</td>
</tr>
<tr>
<td>0.9</td>
<td>-10.1972</td>
<td>0.1999</td>
<td>-2.5478</td>
<td>-6.4674</td>
<td>-8.0247</td>
<td>-5.2235</td>
<td>-0.5493</td>
<td>-0.5493</td>
</tr>
</tbody>
</table>

This table shows the percentage deviation of the $\text{Cov}(\hat{\rho}_t^*, \hat{\rho}_{t-s}^*)$ approximation from the sample value $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$, with values for $s$ changing from 0 in Panel A to 5 in Panel B, and finally 10 in Panel C. $\text{Cov}(\hat{\rho}_t, \hat{\rho}_{t-s})$ is estimated from conditional correlations generated by applying an EWMA model, a DCC model, and a rolling window estimator to a series of simulated innovations as defined in Equation (10). The true underlying correlation $\rho$ changes over typical values: 0, 0.3, 0.6, and 0.9.
Appendix E. Proof of Proposition 1 (referring to page 21)

We begin with the expressions \( \hat{\mathbf{q}}_t, \hat{\mathbf{q}}_{t-s} \) for the DCC and the EWMA model (labeled \( (i) \) and \( (ii) \) in Proposition 1). First, let \( \mathbf{g}_t = \begin{pmatrix} e_{1,t}^2 & e_{2,t}^2 & e_{1,t} e_{2,t} \end{pmatrix}' \) so that \( \mathbb{E}(\mathbf{g}_t) = \varphi = (1, 1, \rho)' \) and

\[
\hat{\mathbf{q}}_t = (1-a-b)\varphi + a\mathbf{g}_{t-1} + b\hat{\mathbf{q}}_{t-1},
\]
with \( 0 < a < 1, 0 < b < 1 \) and \( a + b < 1 \). Defining \( \mathbf{u}_t = a(\mathbf{g}_{t-1} - \varphi) \) such that \( \mathbb{E}(\mathbf{u}_t) = 0 \) and \( \mathbf{B} = b\mathbf{I} \), we can rewrite (E.1) as a VAR(1) process:

\[
\hat{\mathbf{q}}_t = (1-b)\varphi + \mathbf{B}\hat{\mathbf{q}}_{t-1} + \mathbf{u}_t,
\]

\[
= (1-b)\varphi \lim_{t \to \infty} \left( \mathbf{I} + \mathbf{B} + \cdots + \mathbf{B}^t \right) + \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{u}_{t-s}, \tag{E.2}
\]

\[
= \varphi + \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{u}_{t-s},
\]

where we use the fact that \( \lim_{t \to \infty} \left( \mathbf{I} + \mathbf{B} + \cdots + \mathbf{B}^t \right) = \frac{1}{1-b} \mathbf{I} \). Equation (E.2) is stationary if all eigenvalues of \( \mathbf{B} \) are less than 1 in absolute values, which is satisfied here as we assume \( 0 < b < 1 \) (this is a common assumption, see e.g. Lütkepohl, 2006). Because \( \hat{\mathbf{q}}_t \) is stationary, \( \mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = 0 \) and \( \mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \mathbf{u}_t \) for all \( t \). Also, \( \mathbf{u}_t \) is invertible for \( |\rho| < 1 \). It follows that

\[
\mathbb{Cov}(\hat{\mathbf{q}}_t, \hat{\mathbf{q}}_{t-s}) = \sum_{i=0}^{\infty} \mathbf{B}^{i+1} \mathbf{u}_t (\mathbf{B}^i)' .
\]

Hence, all that is needed to get \( \mathbb{Cov}(\hat{\mathbf{q}}_t, \hat{\mathbf{q}}_{t-s}) \) is \( \mathbb{E}(\mathbf{u}_t \mathbf{u}_t') \). Denoting \( \varphi \varphi' = \Lambda \) and

\[
\mathbb{E}(\mathbf{g}_t \mathbf{g}_t') = \Gamma, \text{ we can write } \mathbf{u}_t = a^2(\Gamma - \Lambda) \text{ and therefore }
\]

\[
\mathbb{Cov}(\hat{\mathbf{q}}_t, \hat{\mathbf{q}}_{t-s}) = \sum_{i=0}^{\infty} \mathbf{B}^{i+1} a^2 (\Gamma - \Lambda) (\mathbf{B}^i)' = a^2(\Gamma - \Lambda) \frac{b^s}{1-b^s}. \tag{E.4}
\]

Because \( \Omega = \Gamma - \Lambda \), Equation (E.4) implies the DCC case \( (i) \) for \( (a+b) < 1 \) and the EWMA case \( (ii) \) for \( 0 < \lambda = b = (1-a) < 1 \).

Next, we turn to the expression \( \hat{\mathbf{q}}_t, \hat{\mathbf{q}}_{t-s} \) for the rolling window estimator (labeled \( (iii) \) in Proposition 1). Let again \( \mathbf{g}_t = \begin{pmatrix} e_{1,t}^2 & e_{2,t}^2 & e_{1,t} e_{2,t} \end{pmatrix}' \), \( \mathbb{E}(\mathbf{g}_t) = \varphi = (1, 1, \rho)' \), \( \varphi \varphi' = \Lambda \) and \( \mathbb{E}(\mathbf{g}_t \mathbf{g}_t') = \Gamma \). Because \( \mathbb{Cov}(\hat{\mathbf{q}}_t, \hat{\mathbf{q}}_{t-s}) = \mathbb{E}(\hat{\mathbf{q}}_t \hat{\mathbf{q}}_{t-s}) - \Lambda \) we have to derive \( \mathbb{E}(\hat{\mathbf{q}}_t \hat{\mathbf{q}}_{t-s}) \).
for all $s \geq 0$. For the rolling window estimator, we rewrite $\hat{q}_t$ as $\hat{q}_t = \frac{1}{n} \sum_{t=1}^{n} g_{t-t}$. First, let $s = 0$. In this case

$$E(\hat{q}_t \hat{q}'_t) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} E(g_{t-l} g'_{t-k}),$$

$$= \frac{1}{n^2} (n\Gamma + n(n-1)\Lambda),$$

$$= \frac{n}{n^2} (\Gamma - \Lambda) + \Lambda. \quad \text{(E.5)}$$

Next consider $0 < s = n$. Note that we can write $\hat{q}_{t-s} = \hat{q}_{t-(s-1)} - \frac{1}{n} (g_{t-s} g'_{t-s-n})$ such that

$$E(\hat{q}_t \hat{q}'_{t-s}) = E(\hat{q}_t \hat{q}'_{t-(s-1)}) - \frac{1}{n} \left( E(\hat{q}_t g'_{t-s}) - E(\hat{q}_t g'_{t-s-n}) \right),$$

$$= E(\hat{q}_t \hat{q}'_{t-(s-1)}) - \frac{1}{n} \left( \frac{1}{n} \Gamma + \frac{n-1}{n} \Lambda - \Lambda \right),$$

$$= E(\hat{q}_t \hat{q}'_{t-(s-1)}) - \frac{1}{n^2} (\Gamma - \Lambda). \quad \text{(E.6)}$$

Substituting $\frac{n-(s-1)}{n^2} (\Gamma - \Lambda) + \Lambda$ for $E(\hat{q}_t \hat{q}'_{t-(s-1)})$ in (E.6) yields $E(\hat{q}_t \hat{q}'_{t-s}) - \frac{n-s}{n^2} (\Gamma - \Lambda) + \Lambda$. Recalling $\text{Cov}(\hat{q}_t, \hat{q}_{t-s}) = E(\hat{q}_t \hat{q}'_{t-s}) - \Lambda$ and $\Omega = \Gamma - \Lambda$, (E.5) and (E.6) thus imply $\text{Cov}(\hat{q}_t, \hat{q}_{t-s}) = \frac{n-s}{n^2} \Omega$ for $0 \leq s < n$. Obviously, for all $s \geq n$ we have $E(\hat{q}_t \hat{q}'_{t-s}) = \Lambda$ so that $\text{Cov}(\hat{q}_t, \hat{q}_{t-s})$ is the null matrix. Hence, the proof is complete.
Appendix F. Correlations and Non-Elliptical Return Distributions (referring to page 24)

The correlation coefficient measures the linear dependence between two random variables. If the joint return distribution of two assets is the bivariate normal distribution then the correlation coefficient measures the entire dependence structure. In fact, the interpretation of the correlation coefficient is unproblematic as long as the joint distribution belongs to the class of elliptical distributions (Embrechts, McNeil, and Straumann, 1999). A distribution is elliptical if the points of equal density form an ellipse. The normal distribution is therefore a special case of the larger family of elliptical distributions. Although correlation does not assume the asset returns to be elliptical, the correlation coefficient no longer has the same informational content if the joint distribution is non-elliptical. For instance, zero correlation between two assets no longer implies independence. More generally, the maximum attainable range of correlations no longer lies between -1 and +1. In particular, fat-tailed distributions and asymmetries can significantly reduce this attainable range. As a consequence, low correlation coefficients are wrongly interpreted as implying low dependence between assets. To illustrate this point, Figure F1 simulates the maximum and minimum attainable correlations for three different cases. In Panel A, both assets are normally distributed. In this case, the attainable range for the correlation coefficient always lies between -1 and +1 independent of the variance of the assets. The situation changes in Panel B where one asset has the same standard normal distribution as before, but the other asset has a fat-tailed $t$-distribution. Here, the attainable correlation range responds strongly to the presence of fat tails. A lower degrees-of-freedom parameter corresponds to more pronounced tails and hence more extreme outliers that can affect the correlation coefficient. Whereas correlations cover the full range for moderate tails generated by a distribution with 5 degrees of freedom, the attainable range quickly declines for more pronounced tails and is close to zero for a $t$-distribution with one degree of freedom.
Figure F1: Attainable Range of Correlations Could be Reduced in Practice

Panel A: Both Variables are Normally Distributed

Panel B: One Variable has Fat Tails

Panel C: Both Variables are Skewed and have Fat Tails

This figure shows the simulated maximum and minimum attainable range for correlations estimated on 1 million random draws from two marginal distributions. Panel A shows that the full range between -1 and +1 can be obtained for the sample correlation if the two random variables each have Gaussian marginal distributions. This holds irrespective of the relative size of the variances of the two variables. Panel B illustrates the case when one variable has a t-distribution with varying degrees of freedom (dof). Low dof values generate fat tails. The more pronounced the tails of this marginal distribution, the lower is the...
attainable range for the sample correlation coefficient. For instance, if the dof parameter is one, the maximum and minimum attainable range are close to zero even when the actual dependence is very high. In Panel C, both variables have a lognormal distribution which exhibits both, a high skewness and heavy tails. The attainable range is again reduced and converges towards zero with increasing standard deviation of one variable.

In Panel C we show the effects when asset returns exhibit both fat tails and asymmetries. Like before, we hold one distribution fixed while increasing the scale parameter of the other. An increase in the standard deviation of the second asset now generates both, more pronounced tails and positive skewness. In this case, the attainable range of correlation coefficients again converges towards zero but the convergence is asymmetric with the minimum correlation value being more strongly affected. In summary, our results indicate that moderate deviation from non-normality may be harmless but that extreme tails can generate outcomes that render the correlation coefficient as a measure of the overall dependence structure useless.

An indication for the degree of non-normality in our actual return data can be obtained by comparing the time-varying correlations generated by the standard DCC model with those of the dynamic copula GARCH model (Demarta and McNeil, 2005; Lee and Long, 2009; Ghalanos, 2015). In this model, the standardized returns are estimated using univariate GARCH models on each return series with the remaining dependence structure being modeled by a copula. We use a multivariate student-$t$ (MVT) copula where the shape parameter $\eta$ is determined by the data. If our standardized return vector is characterized by a high degree of non-normality we would expect the dynamic correlations generated by the copula GARCH model to deviate significantly from the benchmark DCC model that is used in our paper.\footnote{Although the copula is multivariate-$t$, the marginal distributions are taken from the empirical univariate distributions which are not $t$-distributed. As a consequence, the resulting bivariate distribution is not a $t$-distribution.} Panel A of Figure F2 shows the theoretical normal distribution with the mean vector and covariance matrix being equal to the average values of our standardized return data. In comparison, the right graph in Panel A shows the empirical distribution that is obtained from the bivariate kernel density plot over all asset pairs. The visual impression from the empirical distribution is that the actual standardized returns have more density in the tails and are not elliptical.\footnote{The Henze-Zirkler (1990) test rejects the null hypothesis of a bivariate normal distribution for all of our 780 asset pairs.} However, as we have seen in Figure F1 above, moderate fat tails do not reduce the attainable range of the correlation coefficient. It is therefore unclear whether this deviation is sufficient to generate distorting effects in our paper.
This figure compares the dynamic correlations generated by the DCC model of Engle (2002) with those generated by a multivariate Student $t$-Copula GARCH model (Demarta and McNeil, 2005). For estimating the marginal distributions for both series, we take the semiparametric empirical approach of Genest et al. (1995). The graphs show that the standard DCC model generates time-varying correlations that are very similar to the more complex $t$-Copula specification. The application of a copula GARCH model is motivated in Panel A, which compares the theoretical bivariate returns with the actual standardized returns from the DCC model.
Panel B of Figure F2 compares the DCC and the MVT copula GARCH correlation models for the time series that were used in Figure 2 in the introduction of the paper. The dashed line shows the DCC correlation between the returns of the S&P 500 and the NSDAQ index. The solid line shows the dynamic correlations generated by the copula GARCH model. Although the two correlation estimates can deviate slightly when correlations spike in either direction, the comovement dynamics between both time series are very similar. We can confirm this observation in Panel C where both models are compared for the correlation between the returns of the S&P 500 and crude oil. Like before, there is little evidence that the dynamic correlations estimated by a copula GARCH model deviate in any systematic way from our benchmark DCC model. Our findings from Figure F2 are at odds with the notion that non-normality or non-elliptical return behavior is substantially biasing our correlation estimates.

As a final robustness check, we would like to show that the theoretical hump-shaped relationship between the level of the true but unobserved correlation and the volatility of its estimate can also be confirmed empirically if we use the distribution free rank correlation measure Kendall’s $\tau$. The relationship is discussed in subsection 4.3 and is formally derived in Appendix E. For the empirical verification of this phenomenon, we assumed that the estimated DCC correlation fluctuates around its unconditional true value $\rho$ so that the average DCC correlation was our estimate of the true correlation level. The relationship is shown in Panel A of Figure F3 which repeats Figure 7 from the paper. If the bivariate return distribution changes to a non-elliptical type over time, it may impact the attainable range of the underlying correlation and hence distort our empirical estimate of the relationship between $\rho$ and $\hat{\rho}_p$.\textsuperscript{3} If distribution changes to non-elliptical forms are distorting our results, we would expect to see different empirical results when repeating the left graph in Panel A of Figure F3 with the distribution free rank correlation measure Kendall’s $\tau$. The right graph shows the results when we replace the average DCC correlation in estimating the

\textsuperscript{3} A typical cause of such a change would be a single major economic or financial event such as the burst of the Dot-Com bubble or the bankruptcy of Lehman Brothers. However, these types of distribution changes are already accounted for in our analysis since we look at between-break subsamples. Hence, only the remaining distribution changes can have distorting effects on our analysis which reduces the number of instances where this issue matters.
underlying level of $\rho$ by the sample estimate of Kendall’s $\tau$ applied to the asset returns. Correlation estimates from Kendall’s $\tau$ are very similar to their DCC counterparts and the location and shape of the curves are almost identical. There is only a slight deviation between both estimates for very high levels of correlation, where the relationship based on Kendall’s $\tau$ indicates that the volatility in $\hat{\rho}$, $\sigma_\rho$, is actually closer to zero than what is suggested by the DCC model. We can confirm the similarities between both correlation measures in Panel B where we regress average DCC correlations against Kendall’s $\tau$ estimates. Panel B shows that the relationship is very strong and only deviates for high levels of correlation where high DCC correlation estimates are associated with slightly lower values for Kendall’s $\tau$. The findings in Figure F3 suggest that deviations from non-normality could have a minor impact for high levels of correlation but are generally unlikely to have major distorting effects. Overall, our findings in this appendix indicate that non-normal marginal and bivariate return distributions are a common feature of financial data but that the density in the tails of the distributions may not be sufficiently large to have serious distorting effects on the attainable correlation range. Although the topic remains an important problem in theory, it appears to be a much smaller issue in practice.
This figure shows that our empirical findings do not change if we use a distribution free correlation measure (Kendalls $\tau$). Panel A compares the hump-shaped relationship between the level and the volatility of the correlation coefficient. The left graph repeats Figure 7 from the paper while the right graph shows a similar relationship based on Kendalls $\tau$ correlation estimates. This measure is closer to the theoretical functional form at the borders when the correlation is estimated to be very high, but in general, the graph confirms our previous findings indicating that the distortion of our DCC correlations due to non-normal bivariate returns is likely to be small. This is again confirmed in Panel B which shows the scatterplot of the time-averaged DCC correlation and Kendalls $\tau$ correlation.
References


