Forecasting crude oil and refined products volatilities and correlations: New evidence from fractionally integrated multivariate GARCH models

Malvina Marchese[a], Ioannis Kyriakou[a], Michael Tamvakis[a,*], Francesca Di Iorio[b]

[a] *Cass Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK*
[b] *Dipartimento di Scienze Politiche, Università Degli Studi di Napoli Federico II, Via Rodinò 22 a / Largo San Marcellino, Italy*

Abstract

The relationship between the prices of crude oil and its refined products is at the heart of the oil industry. Crude oil and refined products volatilities and correlations have been modelled extensively using short-memory multivariate GARCH models. This paper investigates the potential benefits from using fractionally integrated multivariate GARCH models from a forecasting and a risk management perspective. Several models for the spot returns on three major oil-related markets are compared. In-sample results show significant evidence of long-memory decay and leverage effects in volatilities and of time-varying autocorrelations. The forecasting performance of the models is assessed by means of three approaches: the Superior Predictive Ability test, the Model Confidence Set and the Value-at-Risk. The results indicate that the multivariate models incorporating long-memory outperform the short-memory benchmarks in forecasting the conditional covariance matrix and associated risk magnitudes. The paper makes an innovative contribution to the analysis of the relationship between crude oil and its refined products providing refiners, physical oil traders, non-commercial oil traders and other energy markets agents with significant insights for hedging and risk management operations.

Keywords: multivariate GARCH, long memory, Superior Predictive Ability test, Model Confidence Set, Value-at-Risk
JEL: C32, C51, C52, Q40

*Corresponding author.
Email addresses: malvina.marchese@city.ac.uk (Malvina Marchese), ioannis.kyriakou@city.ac.uk (Ioannis Kyriakou), m.tamvakis@city.ac.uk (Michael Tamvakis), fdiiorio@unina.it (Francesca Di Iorio)

1. Introduction

Oil prices, especially that of crude oil, are central to global economic activity, be it the physical handling and trading of the commodity and its products, or the effect of these prices on current and future economic prosperity. Crude oil, as a commodity, is of limited direct usage as a fuel. It is the range of products yielded by refining which are consumed either directly (e.g., gasoline and diesel for motor vehicles) or indirectly (e.g., fuel oil to generate electricity, or naphtha as petrochemical feedstock). Because of the need to transform crude oil into refined products, the interaction between upstream producers and downstream consumers is not direct. Prices for refined products can be linked back to those of crude oil through the netback mechanism. Refined product prices should theoretically be linked to the cost of acquiring crude oil (of various qualities and provenances), transporting it (via pipelines or tankers, often from abroad) to the transformation point, storing it, refining it, storing the refined products and distributing these products to a myriad of consumption points, which may be located abroad as well.

Such calculations might be feasible if all the relevant information were publicly available and easily accessible. As this is not usually the case, researchers investigate the relationship using data for the most commonly traded crude oils and refined products. In fact, the volatility and correlations of oil and refined products prices are key inputs to anything from macroeconomic models, option pricing models, investment portfolio construction, hedging and risk management operations. The latter is of particular significance to the refining industry, which forms the nexus between crude oil production and final consumption and which is exposed to risks from the supply and demand sides. Several authors have studied the relationship between crude oil and refined products. Among them, Borenstein et al. (1997) look at the asymmetrical response of gasoline to crude oil prices; Kaufmann and Laskowski (2005) do a similar study for crude oil, gasoline and heating oil; Lee and Zyren (2007) look at the volatility, rather than the price, of crude oil and products; Ji and Fan (2012) employ several GARCH-type models to devise a dynamic hedging strategy for an oil market portfolio; Suenaga and Smith (2011) study the volatility dynamics and seasonality of crude oil, gasoline and heating oil futures contacts traded on NYMEX; Vacha and Barunik (2012) use wavelet coherence analysis to test the co-movements of crude oil, gasoline and heating oil; Tong et al. (2013) perform a similar analysis but with the use of copulas. Block et al. (2015) investigate the dynamic conditional correlation among crude oil, refined products and natural gas and the role of structural breaks with a Copula multivariate GARCH model.

There is a strong consensus in the current literature on the effectiveness of multivariate GARCH (MGARCH) models in exploring and forecasting volatility spillovers and co-movements between crude oil and refined products (Wang and Wu, 2012, Chang et al., 2010, 2011). However, all the MGARCH models used in the literature implicitly impose a short-memory decay rate on crude oil and refined products volatilities. Such assumption is, in fact, over-restrictive: a large number of empirical studies suggest that crude oil and refined products price volatilities display a strong degree of persistence, consistent with the notion of long memory rather than with the exponential decay rate implied by the short-memory assumption. Several univariate long-memory models, including the fractionally integrated auto-regressive (ARFIMA) and the fractionally integrated GARCH (FI-GARCH) model, have been successfully used to forecast crude oil and refined products price series individually (Brunetti and Gilbert, 2000, Tabak and Cajueiro, 2007, Kang et al., 2009, Chang et al., 2010) but, to the best of our knowledge, no attempt to include such feature in multivariate models has yet been made. In practice, failure to account for this decay rate in the volatility of crude oil and refined products prices will result in model misspecification and potentially incorrect conclusions about the response of refined products volatility to crude oil price shocks, and, further, to incorrect volatility forecasts and unreliable risk management evaluations. This paper addresses such lack in the literature by assessing whether, in the investigation of co-movements between crude oil and refined products, the use of multivariate long-memory GARCH models with asymmetries and dynamic correlations significantly improves the models' in-sample and forecasting performance as well as their attractiveness in terms of risk monitoring.

The purpose of the paper is three-fold. First, we analyze the co-movements between crude oil (West Texas Intermediate-Cushing) and two refined products price series, conventional gasoline (New York Harbor) and heating oil (New York Harbor), by means of different MGARCH models, including the fractionally integrated DCC models, and assess the gains from using long-memory specifications by comparing the models' in-sample performance. The choice of US, where all three commodities are traded, is justified not only by the depth and breadth of spot, forward and futures markets for these energy commodities, but also by the position of US as the world's largest producer of crude oil, largest consumer of crude oil and refined products, second largest importer of crude oil, largest refiner (and refining capacity holder) and largest exporter of refined products. The empirical analysis is carried out for 30 different specifications of MGARCH models deriving from the combination of several univariate volatility processes with different multivariate structures under different distributional assumptions (Amendola and Candila, 2016).

Second, we evaluate the forecasting accuracy gains by means of two statistical approaches: the Superior Predictive Ability (SPA) test of Hansen (2005) and the Model Confidence Set (MCS) method of Hansen et al. (2011). The SPA test focuses on the predictive ability of a predefined benchmark model with respect to several alternatives: we employ it to assess if specific assumptions for the multivariate structure, such as constant correlations, and for the dynamics of individual volatilities, such as short memory, can be rejected. With the MCS method, we identify from the initial set of competing models those which display equal predictive ability and outperform the others at a given confidence level. Both tests are executed using several symmetric and asymmetric matrix loss functions, which are robust to the choice of the volatility proxy (Laurent et al., 2012, 2013, Patton, 2011). Thus, we further contribute to the energy literature by extending the existing oil prices forecasting framework and providing a comprehensive conditional variance matrix forecasts' comparison assessing simultaneously volatility and correlation forecasting. We explore the sensitivity of the models' forecasting accuracy with respect to different forecasting horizons (1, 5 and 20 days ahead) and forecasting sample periods. We consider three different periods with homogeneous volatility dynamics (calm, turbulent and fairly volatile markets) and find that, while during calm periods symmetry and constant correlations cannot be rejected, during turbulent periods the set of superior models includes only specifications with long memory and dynamic correlations.

Finally, when comparing different competing models, the evaluation of the best performance in an economically meaningful way is very relevant. Managing and assessing risk in oil markets is a key issue for practitioners and the Value-at-Risk (VaR) is a widespread method of quantifying it (Agnolucci, 2009). The benefits of univariate GARCH models with long memory in forecasting VaR have been investigated by several authors (Chkili et al., 2014, Aloui and Mabrouk, 2010). Recent findings in financial econometrics (Giot and Laurent, 2003) suggest that MGARCH models outrun their univariate counterparts on an out-of-sample basis in VaR prediction. Our last application in the forecasting exercise explores the efficiency gains from using the fractionally integrated DCC models in one-step ahead VaR prediction for short and long positions.

The remainder of the paper is organized as follows. Section 2 provides a brief outline of the multivariate volatility models considered in the paper. Section 3 describes the data and analyzes the in-sample performance of the models. Section 4 presents the forecasting exercise. Section 5 offers some robustness checks. Concluding remarks and directions for future research are given in Section 6. Overall, the results demonstrate the benefits from using MGARCH models with long memory, from both an in-sample and an out-of-sample perspective.

2. Multivariate conditional volatility models

This section presents the models which we estimate and compare in Sections 3 and 4, namely the Baba–Engle–Kraft–Kroner (BEKK) model of Engle and Kroner (1995), its asymmetric extension (ABEKK) by Grier et al. (2004), the vector asymmetric GARCH (AGARCH) model of McAleer et al. (2009), the Constant Conditional Correlation (CCC) model of Bollerslev (1990), the Dynamic Conditional Correlation (DCC) model of Engle (2002), and its long-memory extensions, namely the fractionally integrated symmetric and asymmetric DCC (FIGARC-DCC and FIEGARCH-DCC) models.

Let $\boldsymbol{r}_t$ be the vector of log-returns of n oil prices and θ a finite vector of parameters. The general form of a multivariate GARCH model is

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t(\theta) + \boldsymbol{\epsilon}_t, \tag{1}$$

$$\boldsymbol{\epsilon}_t = \boldsymbol{H}_t^{1/2}(\theta)\,\boldsymbol{z}_t, \tag{2}$$

where $\boldsymbol{z}_t$ is a zero-mean i.i.d. random vector with $\mathrm{Var}(\boldsymbol{z}_t) = \boldsymbol{I}_n$, the vector $\boldsymbol{\mu}_t$ is the conditional mean of the process, and the positive definite matrix $\boldsymbol{H}_t$ is its conditional variance. In what follows, we specify, without loss of generality, the conditional mean equation as a vector autoregressive process. Different MGARCH specifications in the literature are based on different parameterizations of $\boldsymbol{H}_t$ (comprehensive reviews of MGARCH models can be found in Bauwens et al., 2006 and Silvennoinen and Teräsvirta, 2009). Popular in the econometric analysis of energy markets is the BEKK(k, p, q) model of Engle and Kroner (1995). In practice, empirical studies fix $k = q = p = 1$ and estimate the BEKK$(1, 1, 1)$ model with conditional variance matrix

$$\boldsymbol{H}_t = \boldsymbol{C}_0'\boldsymbol{C}_0 + \boldsymbol{A}_{11}'\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\boldsymbol{A}_{11} + \boldsymbol{G}_{11}'\boldsymbol{H}_{t-1}\boldsymbol{G}_{11}, \tag{3}$$

where $\boldsymbol{C}_0$, $\boldsymbol{A}_{11}$ and $\boldsymbol{G}_{11}$ are $n \times n$ parameter matrices with $\boldsymbol{C}_0$ upper triangular. Identification of the BEKK(1, 1, 1) is achieved under simple and straightforward conditions, which can be imposed during the estimation relatively easily (Engle and Kroner, 1995). The model has $2n^2+n(n+1)/2$ parameters to be estimated. To reduce the computational burden, we consider also a diagonal BEKK(1, 1, 1) imposing diagonality of A_{1k} and G_{1k}. Model (3) does not allow for asymmetric impact of positive and negative shocks on the conditional variance. Grier et al. (2004) propose the asymmetric extension

$$\boldsymbol{H}_t = \boldsymbol{C}'\boldsymbol{C} + \boldsymbol{A}_{11}^{'}\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}'\boldsymbol{A}_{11} + \boldsymbol{G}_{11}^{'}\boldsymbol{H}_{t-1}\boldsymbol{G}_{11} + \boldsymbol{D}_{11}'\boldsymbol{\xi}_{t-1}\boldsymbol{\xi}_{t-1}'\boldsymbol{D}_{11}, \tag{4}$$

where the vector $\boldsymbol{\xi}_t$ has elements given by $\min(0, \boldsymbol{\epsilon}_{it})$ and $\boldsymbol{D}_{11}$ is a $n \times n$ parameter matrix. This specification nests the full and diagonal BEKK models and, if the elements of $\boldsymbol{D}_{11}$ are significantly different from zero, it detects asymmetries; if they are negative, it detects leverage.

The CCC and DCC models are based on the attractive decomposition of the conditional variance matrix into conditional standard deviations and conditional correlations matrices

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t, \tag{5}$$

where $\boldsymbol{D}_t$ is the diagonal matrix of conditional standard deviations and $\boldsymbol{R}_t$ is a symmetric, positive definite correlation matrix. In the CCC model, $\boldsymbol{R}_t$ is assumed to be constant over time, i.e., $\boldsymbol{R}_t = \boldsymbol{R}$, and the overall stationarity is ensured by the stationarity of the individual GARCH series. In the DCC model,

$$\mathbf{R}_t = (\boldsymbol{I} \odot \mathbf{Q}_t)^{-\frac{1}{2}}\mathbf{Q}_t(\mathbf{I} \odot \mathbf{Q}_t)^{-\frac{1}{2}}, \tag{6}$$

$$\boldsymbol{Q}_t = (1 - \lambda_1 - \lambda_2)\,\bar{\boldsymbol{Q}} + \lambda_1 e_t e_t' + \lambda_2\boldsymbol{Q}_{t-1}, \tag{7}$$

with λ_1 and λ_2 nonnegative scalar parameters satisfying $\lambda_1 + \lambda_2 < 1$, $e_t = diag(\boldsymbol{Q}_t)^{1/2}\epsilon_t$ and $\bar{\boldsymbol{Q}}$ set equal to the unconditional correlation matrix of the standardized residuals. In both the CCC and DCC multivariate specifications, the diagonal elements of $\boldsymbol{H}_t$, i.e., the individual series volatilities, can evolve according to different univariate GARCH processes. In our application, we consider 6 different univariate specifications in the CCC and DCC framework respectively, including the short-memory GARCH(1, 1), EGARCH(1, 1), IGARCH(1, 1, 1), GJR-GARCH(1, 1), and the long-memory FI-GARCH(1, d, 1) and FI-EGARCH(1, d, 1). The functional forms of the competing univariate models are reported in Table 1. Among the long-memory specifications, the fractionally integrated exponential GARCH models log-volatilities rather than volatilities:

$$\log \sigma_{it} = \omega_i + \frac{a_i\left(L\right)}{b_i\left(L\right)}\left(1 - L\right)^{-d} g_i\left(z_{it-1}\right), \tag{8}$$

where $\boldsymbol{z}_t$ is a vector of zero-mean i.i.d. shocks with variance Σ_z, $a_i\left(L\right)$ and $b_i\left(L\right)$ are univariate polynomials in the lag operator of known degree with no common zeros, $(1 - L)^{-d}$ is the univariate fractional operator[1], and

$$g_i\left(z_{it-1}\right) = \theta_i + \delta_i\left(\left|z_{it}\right| - \mu_{|z_i|}\right). \tag{9}$$

Depending on the significance and the sign of θ_i, volatilities may display leverage and asymmetries, and exhibit hyperbolic decay if the memory parameter d_i is found significantly different from zero. For benchmarking purposes, we estimate a restricted version of this model, that is, the FI-GARCH$(1, d, 1)$ process:

$$h_{it} = \omega + \beta_1 h_{it-1} + [1 - (1-\beta_1 L)^{-1}(1-\phi_1 L)\,(1-L)^d]\epsilon_{it}^2. \tag{10}$$

A limitation of the CCC/DCC class is its inability to capture spillover effects. To overcome this limitation, Ling and McAleer (2003) and McAleer et al. (2009) propose the AGARCH model which includes cross-volatility and cross-innovation spillovers. The model assumes equation (5) with constant conditional correlation matrix $\boldsymbol{R}$ and sets

$$\boldsymbol{h}_t = diag(h_{it}, \ldots, h_{nt}) = \boldsymbol{w} + \sum_{i=1}^{r} \boldsymbol{A}_i \boldsymbol{\epsilon}_{t-i} + \sum_{j=1}^{s} \boldsymbol{B}_j \boldsymbol{h}_{t-j} + \sum_{l=1}^{q} \boldsymbol{C}_l \boldsymbol{I}_{t-l} \boldsymbol{\epsilon}_{t-i}, \tag{11}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{1t}^2, \ldots, \epsilon_{nt}^2)$, $\boldsymbol{A}_i$, $\boldsymbol{C}_l$ and $\boldsymbol{B}_j$ are $n \times n$ parameter matrices, $\boldsymbol{w}$ is a $n \times 1$ vector, and $\boldsymbol{I_t} = diag(I_{1t}, \ldots, I_{nt})$ is the matrix indicator function taking value 1 if $\epsilon_{it} \leq 0$ and 0 otherwise. Sufficient conditions for stationarity and ergodicity of the AGARCH models are derived in McAleer et al. (2009). We follow Chang et al. (2010) and fit the AGARCH$(1, 1, 1)$ model to our data. As with the CCC model, the main limitation of the AGARCH model is that it imposes constant correlations across time, which might be a stringent assumption for most returns in energy markets (Rahman and Serletis, 2012, Chevallier, 2012, Chang et al., 2010, 2011).

We consider 15 different model specifications and, for each model, two different distributions for the innovations: the normal and the skewed-t distribution. The latter is motivated by the need to account for potential heavy tails and large skewness in the distribution of oil returns consistently with empirical evidence (Vo, 2011, Gronwald, 2012). In total, we have 30 different model specifications. For convenience, we summarize in Table 2 the MGARCH models estimated in the paper together with their main characteristics.

3. Data

We estimate the models of Section 2 for three series of spot price returns: crude oil (CO), conventional gasoline (CG) and heating oil (HO). We use daily observations from 1 June 1993 to 1 June 2018 from the Energy Information Administration (EIA) of the US Department of Energy; we have 6,421 valid observations.

[1] The operator has binomial expansion

$$(1-L)^{-d} = \sum_{k=0}^{\infty} \Gamma\left(k-d\right) \Gamma\left(k+1\right)^{-1} \Gamma\left(-d\right)^{-1} L^k, \quad d < 1/2,$$

where Γ is the gamma function. For the estimation, we truncate at $k = 1,000$.

We define the return r_t as the first-order difference of the logarithmic closing price. Table 3 reports descriptive statistics for the three returns series. The average daily returns are very small compared to their sample standard deviation. The returns display some evidence of skewness and excess kurtosis: the p-values of the Jarque–Bera test statistic suggest rejection of the hypothesis of normality. The Ljung–Box Q statistic for serial correlation shows that the null hypothesis of no autocorrelation up to lag 10 is rejected at the 10% level of significance, implying that some autocorrelation might exist in the conditional mean of the returns. On the other hand, the correlogram and the Ljung–Box Q statistic for serial correlation of the squared returns suggests an extremely strong degree of persistence in all volatility series, consistently with a long-memory decay rate. To assess the memory properties of the returns, we estimate semi-parametrically for each series the long-memory parameter d using the local Whittle estimator of Robinson (1995) with bandwidth $m = 100$ and no trimming. The asymptotic standard error is equal to $\left(2m^{1/2}\right)^{-1} = 0.050$. We find significant persistence in the CO and CG series with estimated memory parameters d of 0.35 and 0.33, respectively. All models are estimated via quasi-maximum likelihood methods in one step to ensure comparability of the in-sample information criteria. Estimation and forecasting are conducted on Matlab 2019a using Kevin Sheppard's Oxford Matlab MFE Toolbox.

3.1. In-sample results

To account for serial correlation in the data, we fit a VAR(p) model to the returns vector. Lag selection criteria and the LR test, reported in Table A.1, suggest that a VAR(1) parameterization accounts well for the conditional mean dynamics of the series. The model, the estimated parameters and their robust t-statistics are reported in Table A.2 along with the diagnostic tests. Only 3 out of the 12 estimated parameters are significant at the 10% level. Only CG displays time-dependence in the mean equation which might be arising from the seasonal patterns related to the driving season in the US. There is no evidence of spillover effects between the means series. Post-estimation diagnostic tests for the residuals of the estimated VAR(1) model, reported in Table A.3, confirm the presence of strong GARCH effects, non-normality and no serial correlation.

We fit the MGARCH specifications discussed in the previous section to the residuals. Estimation results are reported in the appendix. Tables A.4–A.6 report, respectively, estimation results for the diagonal, full and asymmetric BEKK$(1, 1, 1)$ models. In all specifications, the main diagonal parameters of A and G are highly significant confirming the presence of strong ARCH and GARCH effects which capture own past shocks and volatility effects in the residual series: the highest ARCH estimate is 0.105 and the GARCH estimates range from 0.818 to 0.902. In the full and asymmetric BEKK, α_{21}, α_{23}, g_{21}, and g_{23} are significant at the 10% level implying existence of volatility spillovers between CO and CG, CO and HO. All the series display leverage effects, with d_{11} and d_{22} significant at the 5% level and d_{33} significant at the 10% and negative. The degree of long-run persistence, $\alpha_{ii} + g_{ii}$, is very close to 1 for the CO and CG series insinuating long-range dependence, which however cannot be modelled explicitly in the BEKK framework.

Estimation results for the AGARCH$(1, 1, 1)$ model, reported in Table A.7, confirm the presence of significant ARCH and GARCH effects. The ARCH estimates are generally small, while the GARCH are high and close to 1. Estimated persistence for the CO and CG volatilities is, respectively, at 0.998 and 0.999, and for the HO series at 0.850 suggesting a short-memory decay rate. Volatility spillovers between the series are significant: we find cross-innovation and cross-volatility spillovers from CO to CG and HO significant at the 5% level, implying that the volatility of CG and HO are affected by the previous long-run shocks in the CO market. The constant conditional correlation estimates between the series are all significant, with highest correlation of 0.683 between CO and HO returns insinuating their positive co-movement. There are significant and negative asymmetry effects for CO and CG. The off-diagonal elements of the asymmetry parameters matrix are not significant, confirming the absence of cross-leverage effects.

Results from the CCC models, shown in Table A.8, are consistent with the empirical evidence from the previous specifications. The estimated conditional correlations are significant at the 1% level and of higher magnitude than in the AGARCH model, ranging from 0.794 to 0.901. This discrepancy, however, might be due to the inability of the CCC model to account explicitly for spillover effects as the AGARCH.

Tables A.9–A.11 report the DCC models' estimates. All estimates are significant at the 5% level. The strong significance of λ_1 and λ_2 confirms that the hypothesis of constant conditional correlations is inadequate in the analysis of the co-movements between crude oil and refined products markets. The long-run persistence of shocks to the conditional correlations is quite high, estimated at 0.891 ($= 0.328 + 0.563$). The estimates for the fractionally integrated DCC, based on univariate FI-GARCH$(1, 1)$ or FI-EGARCH$(1, 1)$ processes, are significant at the 1% level for the first two series, suggesting the adequateness of a hyperbolic date rate for the volatilities of CO and CG. The estimated value of θ in the FI-EGARCH models and in the GJR model are significant and negative for all the series, corroborating the existence of leverage effects.

Post-estimation diagnostic tests, available upon request, confirm that all MGARCH specifications successfully capture the volatility clustering, skewness and excess kurtosis found in the residuals of the VAR(1) model. Table 4 reports the maximized log-likelihood and information criteria for all the fitted models. Boldface values correspond to the best-performing models according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Under either criterion, constant conditional correlation specifications are surpassed by their dynamic counterparts and symmetric specifications by those including leverage effects. This suggests that the correlation between crude oil and refined products evolves dynamically. The strong evidence of, respectively, long and short-memory decay in CO and CG and in HO implies that convergence to the long-run equilibrium after shocks is slower in the CO and CG than in the HO spot returns.

4. The forecasting exercise

Comparison of MGARCH models in terms of variance matrix forecasting accuracy has only recently been addressed in the literature. Volatility forecasting is particularly challenging as volatility itself is latent and thus unobservable even ex-post. In general, to compare model-based forecasts with ex-post realizations, the researcher must choose either a statistical or an economic loss function, as well as a proxy for the true unobservable conditional variance matrix.

As pointed out by Andersen et al. (2005) and Laurent et al. (2012), the use of a proxy might lead to a different ordering of competing models which would be obtained if the true volatility were observed. This issue is particularly relevant when only noisy proxies, such as daily or weekly data, are available. To avoid a distorted outcome, the choice of an appropriate loss function is crucial. It turns out that a number of popular loss functions (MAE, SD-MAE, HMSE) are not robust to noisy volatility proxies and, for this reason, their use has frequently led to conflicting rankings of volatility forecasts (Nomikos and Pouliasis, 2011). Laurent et al. (2012) derive conditions for the functional form of the loss function ensuring consistency of the proxy-based ranking, providing a parametric expression for the entire class of consistent loss functions. We follow Bauwens et al. (2016) and use several loss functions which are robust to noisy proxies, i.e., are expected to provide the same forecast ranking using the true conditional covariance or a conditionally unbiased proxy; we define these in Table 5. As a proxy at day t, we use the matrix of the outer products of the daily mean forecast errors, $e_{T+1}e'_{T+1}$, which is a conditionally unbiased proxy (Patton and Sheppard, 2009 and Becker et al., 2015). The Frobenius, Euclidean and Mean Squared Forecasting Error (MSFE) functions are quadratic functions based on the forecast error and symmetric with respect to over/under-predictions. The Euclidean distance considers the unique elements of the covariance matrix, the Frobenius double-counts the loss associated with the conditional covariances. The Stein loss function is based on the standardized forecast error and is asymmetric with respect to over/under-prediction, heavily penalizing under-predictions. The von Neumann divergence (VDN), on the other hand, penalizes over-predictions.

The forecasting ability of the set of proposed models is evaluated over a series of 630 out-of-sample predictions. We compare the one-day ahead conditional variance matrix forecasts based on the models estimated in Section 3. To carry out the forecasting exercise, we divide the full data set into two periods:

- Period I is the in-sample period from 1 June 1993 to 24 December 2015 (i.e., 5,791 observations) and is reserved for the models' initial estimation
- Period II is the out-of-sample set comprising the remaining 630 observations from 28 December 2015 to the end of the sample period, and is used for forecasting evaluation.

Forecasts are constructed using a fixed rolling window scheme: the estimation period is rolled forward by adding one new daily observation and dropping the most distant observation. Model parameters are re-estimated each day to obtain tomorrow's volatility forecasts and the sample size used for the estimation is fixed. Any dependence on the mean dynamics has been accounted for by fitting a VAR(1), so the mean forecasts do not depend on the models. This scheme satisfies the assumptions required by the MCS method of Hansen et al. (2011) and the SPA test of Hansen (2005) and allows a unified treatment of nested and unnested models. For each statistical loss function, we evaluate the significance of the differences by means of the SPA test and the MCS methodology.

4.1. Assessing the benchmarks: the SPA test

In this section, we study the forecasting performance of a pre-specified benchmark model with respect to alternative models using the SPA test. As benchmarks, we choose the most parsimonious models taking into account the different assumptions for the multivariate structure and the individual volatility dynamics. We use the CCC specification with GARCH$(1,1)$ volatilities to test the hypotheses of constant correlation, short memory and symmetry; we use the DCC specification with GARCH$(1,1)$ volatilities as a flexible and parsimonious benchmark to assess whether relaxing the constant-correlation assumption improves the predictive ability. The CCC-FIGARCH$(1,d,1)$, CCC-FIEGARCH$(1,d,1)$, DCC-FIGARCH$(1,1)$ and DCC-FIEGARCH$(1,1)$ allow us to test whether including long memory and asymmetries in the individual dynamics improves the forecasting accuracy.

For a given loss function, the test is based on the loss differential between the benchmark model, indexed by 0, and an alternative model $k = 1,\dots,m$. Each alternative leads to a sequence of losses during the evaluation period, $t = 1,\dots,T$, and for each period and each model we compute

$$d_{j,k,t} = L_{j,0,t} - L_{j,k,t}, \qquad k = 1,2,\dots m \;\text{ and }\; t = 1,2,\dots,T,$$

where $L_{j,0,t}$ denotes the jth loss function at time t for the benchmark model and $L_{j,k,t}$ the corresponding value of the loss function for the competitor k. The null hypothesis of the test is that the benchmark model is as good as any of the competitors in terms of expected loss:

$$H_0\text{: } \lambda_{j,k} = E(d_{j,k,t}) \leq 0 \text{ for all } k = 1,\dots,m.$$

Note that $\lambda_{j,k} > 0$ corresponds to the case of the competitor k outperforming the benchmark model. For the jth loss function the test statistic is

$$T_{SPA} = \max\left(\sqrt{T}\max_{k=1,\dots,m}\frac{\bar{d}_{j,k}}{\hat{\omega}_k}, 0\right),$$

where $\bar{d}_{j,k} = \frac{1}{T}\sum_{t=1}^{T} d_{j,k,t}$ is the sample loss differential between the benchmark and the competing model k and $\hat{\omega}_k^2$ is a consistent estimator of $\omega_k = \lim_{T\to\infty}(\sqrt{T}\text{variance}(\bar{d}_{j,k}))$. Under α-mixing conditions, a central limit theorem holds and $\sqrt{T}(\bar{\boldsymbol{d}}_j - \boldsymbol{\lambda}_j) \xrightarrow{d} N_m(0,\Omega)$, where $\bar{\boldsymbol{d}}_j$ is the vector of the sample differentials for the jth loss function. To compute the test statistic, only the diagonal elements of Ω are required. While this greatly simplifies the estimation when m is large, it also implies that some elements of the covariance matrix are unknown under the null hypothesis and the asymptotic distribution of the test statistic depends on nuisance parameters. To avoid this, we follow Laurent et al. (2012) and obtain the p-values of the test by bootstrap. There is an extensive literature on the use of bootstrap methods for weakly dependent, i.e., short-memory processes. For example, Politis and Romano (1994) propose an automatic block length selection procedure and Patton et al. (2009) establish a data-dependent method which successfully provides the optimal block length in the case of short-memory data. It is unclear, however, whether such methods are still valid in the case of long-range dependence processes. Lahiri (1999) shows that the block bootstrap is in general not valid even when large block lengths are used and the residual-based bootstrap, known as sieve bootstrap, is asymptotically valid for stationary fractionally integrated processes. We implement the sieve bootstrap for long-memory processes of Kapetanios et al. (2019) with 10,000 bootstrap samples, pre-filtering with the local Whittle estimator.

Results for the different benchmarks are reported in Table 6 where p_C, p_L and p_U are, respectively, the consistent p-values and their lower and upper bounds. Boldface consistent p-values indicate non-rejection of the null at the 10% significance level. We find that the hypothesis of constant correlation (benchmarks 1–4) is always rejected, as well as the hypothesis of short memory. The hypothesis of symmetry in the volatility dynamics is rejected for most benchmarks. Allowing for dynamic correlations significantly improves the models' forecasting accuracy. Overall, it appears that the most valid specification in this application is the fractionally integrated exponential DCC model (benchmark 8). For this benchmark, the null hypothesis is rejected under the Euclidean and Frobenius loss functions but not under the Stein loss function, indicating that FI-EDCC possibly tends to overestimate the variance-covariance matrix.

4.2. The MCS

The Model Confidence Set (MCS) identifies a set of models with equivalent predictive ability which outperform all the other competing models at a given confidence level with respect to a loss function. This method does not require pre-specifying a preferred benchmark model; in fact, it is a statistical test of equivalence with respect to a particular loss function. Let M^0 be the initial set of models for which we compute the series of one-step ahead conditional covariance forecasts for period t, denoted by H_{it}, where i denotes the ith model. The initial assumption is that all the models in M^0 have equal forecasting performance according to the loss function L. By sequentially trimming M^0, the MCS removes those models which are found statistically inferior and determines the set of models M^* which have the best forecasting performance for a given confidence level. The trimming is achieved via a sequence of equal predictive ability (EPA) tests. At each step, the hypothesis

$$H_0\colon\ E(d_{ij,t}) = 0 \text{ for all } i, j \in M$$

is tested for a set of models $M \in M^0$, with $d_{ij,t} = L_{i,t} - L_{j,t}$ denoting the sequence of loss function differentials between forecast i and j. If H_0 is rejected, the worst performing model is eliminated from M and the trimming ends when the first non-rejection occurs. The test statistic is computed as

$$T_M = \max_{i \in M} t_i,$$

where

$$t_i = \frac{\bar{d}_i}{\sqrt{\hat{V}(\bar{d}_i)}},$$

$\bar{d}_i = M^{-1} \sum_{j=1}^{M} \bar{d}_{ij}$ is the relative sample loss statistic of forecast i with respect to all the other forecasts, $\bar{d}_{ij} = T^{-1} \sum_{t=1}^{T} d_{ij,t}$ is the sample loss statistic between forecasts i and j, and $\hat{V}$ is a consistent estimator for the variance of $\bar{d}_i$. Under regularity conditions on $d_{ij,t}$, the asymptotic distribution of T_M depends on the asymptotic correlation matrix of the vector $\left(\bar{d}_1, \ldots, \bar{d}_m\right)'$ (Hansen et al., 2011). For a large number of competing models, to avoid estimation of the high-dimensional correlation matrix, the quantiles of the asymptotic distribution of the test statistic can be obtained by bootstrap method consistently. To this end, we implement a the sieve bootstrap for long-memory processes of Kapetanios et al. (2019) with 10,000 bootstrap samples and pre-filtering by local Whittle estimator. If the null hypothesis is rejected, an elimination rule is needed. We adopt the rule $e_{\max,M} = \arg\max_{i \in M} t_i$, which removes the model contributing more to the test statistic. We repeat this process until non-rejection of the null occurs and a $(1-\alpha)$ confidence set containing the set of models with the best forecasting performance is obtained.

Tables 7–8 report the MCS results at, respectively, the 90% and 75% confidence levels. The last column of each table displays a measure of model performance given by the percentage of inclusions in the MCS across the six loss functions. At the 90% level, the highest number of models (eight) is included for the Euclidean and Frobenious loss functions. At the 75% level, no benchmark model is included in the MCS for any loss function. The asymmetric BEKK and the DCC-FIEGARCH are included in the MCS resulting from the Euclidean, Frobenius, MSFE, and VDN loss functions, whereas the DCC-FIGARCH is included in the MCS from the Euclidean, Frobenius and MSFE loss functions only. The most striking result is the inclusion of the DCC-FIEGARCH model in the MCS of four loss functions, supporting the hypothesis that the inclusion of long memory, asymmetries and time-varying correlations significantly improve the forecasting accuracy of crude oil and refined products volatilities and correlations.

4.3. Portfolio VaR forecasting

This section shifts the focus from a statistical to a decision-theoretical framework for model evaluation. More specifically, we examine the possible efficiency gains from using long-memory asymmetric MGARCH models over short-memory benchmarks for one-step ahead VaR forecasting. To this end, we focus on the models' ability to predict the tail behavior of the returns rather than obtaining the 'best' volatility model. We forecast the one-day ahead VaR for each of the models compared at the 5%, 2.5% and 1% levels, and assess their accuracy using statistical back-testing. We are concerned with both the long and short positions' VaR. In the first case, the risk originates from a price drop, whereas in the second from a price increase. So, we focus, respectively, on the left and right tail of the forecasted distribution of returns and assess the models' joint ability to deliver accurate VaR forecasts for both tails. For each model, the portfolio VaR at level α on day t, conditional on the information available at time $t-1$ and assuming no misspecification, is computed as

$$\mathrm{VaR}_t(\alpha) = z_\alpha \sqrt{w' H_t w},$$

where w is a 3×1 vector of weights, H_t the forecasted conditional covariance matrix for model k, $k = 1, \ldots, m$, and z_α the right or left quantile of the standard normal distribution. For simplicity, we consider only equally weighted portfolios. To carry on with our analysis, we initially estimate the models using the 5,791 observations of the in-sample period. We then compare the predicted one-day ahead VaR for both the long and short positions with the observed returns and record the results. In the second iteration, the models are re-estimated by adding one more day to the estimation sample and the VaRs are forecasted again and compared with the observed returns. We repeat until the in-sample period comprises all the observations minus one. For each model, we then compute the failure rates and VaR exceptions by comparing the long and short forecasted VaR_{t+1} with the observed returns over the whole forecasting period. The percentage of negative (positive) returns which are smaller (larger) than the forecasted one-step ahead VaR_{t+1} for long (short) positions is denoted as $\hat{\pi}_L$ ($\hat{\pi}_S$). To assess the accuracy of the VaRs corresponding to the different models, we test whether the failure rate implied by each model is statistically equal to the expected one. A popular back-testing procedure in the literature is based on the unconditional coverage Kupiec test (e.g., see Giot and Laurent, 2003). The test is a likelihood ratio test built on the assumption that VaR violations are independent. In particular, we rely on the conditional coverage test of Christoffersen (1998) which jointly tests if the total number of failures is equal to the expected number and if the failure process is independently distributed across time. The test statistic is

$$LR_{CC} = -2\log\left[(1-\alpha)^{T-E}\alpha^{E}\right] + 2\log\left[(1-\hat{\pi}_i)^{T-E}\,\hat{\pi}_i^{E}\right],$$

where T is the number of observations in the forecasting period, E the total number of exceptions in the forecasting period, and $i = S$ or L indicates testing for short or long positions. Under the null, the test statistic is distributed as a χ^2 distribution with two degrees of freedom.

The p-values for Christoffersen's test are reported in Table 9. Boldface p-values correspond to rejection of the null hypothesis at the 5% significance level. Results for the short-memory constant-correlation models are homogenous for short and long VaRs leading in all the cases to rejection of the null hypothesis, independently of the model structure. Models with dynamic conditional correlations do better, passing all the tests with occasional rejection in the most extreme quantiles. Models with dynamic conditional correlations and long memory adequately forecast VaR at all levels. In summary, for equally weighted portfolios, reliable VaR forecasts can be obtained under the assumption of conditionally normally standardized portfolio returns using DCC-type models with long-range dependence and asymmetries in the individual volatilities.

5. Robustness checks

In this section, we investigate the sensitivity of the models' forecasting performance with respect to the choice of the forecasting sample and the forecasting horizon.

5.1. Robustness to sub-samples

The overall sample period considered is quite long and characterized by dramatic changes in the volatility dynamics. As pointed out by Hansen et al. (2003), the MCS is specific to the set of candidate models and the sample period. Here, we investigate the sensitivity of the models' forecasting performance with respect to the forecast evaluation sample based on three sub-samples which are homogeneous in their volatility dynamics. The choice of periods reflects the dynamics of crude oil prices.

The first sub-sample, from June 1993 to December 1997, corresponds to a relatively calm period for the market as opposed to later periods in our sample. Crude oil price oscillated around \$20 as the world came to terms with the collapse of the Soviet Union.

Our second sub-sample, from January 1998 to December 2010, represents probably the most turbulent period in the history of the oil industry. The period starts at the beginning of 1998 with the collapse of the oil price to almost \$10 in the aftermath of the SE Asian financial crisis. From Q1 1999 the market started recovering and climbed all the way to \$35, only to retreat to \$20 after the DotCom bubble burst. The price started climbing again from Q1 2002 which coincided with the beginning of the Chinese economic rally of the early 2000s. In Q2 2004, the oil price crossed the \$40 mark, an important psychological barrier which was only reached during the first Gulf War in the summer of 1990. Between Q3 2004 and Q3 2007, the oil price rallied almost continuously, boosted by Asian economic growth as well as haphazard events, such as the aftermath of hurricane Katrina in the US refined products market in 2005. The only exception was the second half of 2006 when prices retreated, only to bounce back and cross the new psychological barrier of \$80 in Q4 2007. At this point this sub-sample includes the most tempestuous period in oil price history to date, with the price climbing to \$145 in July 2008, only to collapse to \$30 in December of the same year. The price bounced again above \$60 and stayed between \$60-80 for most of the time until the end of 2010.

The third sub-sample covers the period from 2011 to 2018. This is a fairly volatile period, but not to the extent witnessed in the previous ones. Between 2011 and 2014, the main characteristic was the oscillation of prices between \$80-110, with only a few drops to \$75 and an average price of around \$100. From early 2011, the US market started receiving substantial amounts of shale oil (following the shale gas boom of 2005), which at the time could not be exported to the international market. As the world adjusted to the rapid increase of US shale oil supply, it became evident that conventional oil producers, such as OPEC members, had to cope with fresh competition to their supply. After 2014, the price crossed \$80 downwards and moved rapidly below \$60. That was when Saudi Arabia signaled its determination to fight for market share in the hope that shale oil producers would find it difficult to survive. Prices dived to near \$25 in Q1 2016, before recovering again above \$40 and remaining between \$40-60 until Q1 2018. The remaining of this sub-sample saw prices trying to find an equilibrium between \$60-70.

Clearly, the volatility dynamics and its scale vary widely between sub-periods. As expected, there are differences with the MCS obtained for the full sample, however our findings support the benefits of the long-memory DCC specifications. The results for the three sub-samples are reported respectively in Tables 10–12. In periods of relatively calm markets, the data show weaker evidence of dynamics in the correlation process and asymmetry. These periods are characterized by a relatively smaller and slow-moving volatility, therefore the result is not surprising and, as expected, most of the MGARCH models exhibit a good fit. Looking at the composition of the MCS, we can draw the following conclusions. First, the AGARCH and IGARCH specifications are excluded from the MCS under all loss functions. Second, the MCS contains CCC and DCC specifications, with GARCH conditional variances, confirming that the hypotheses of constant conditional correlation and symmetry cannot be rejected in calm markets. Finally, all the long-memory specifications are still included in the MCS which also includes two asymmetric specifications, i.e., DCC-GJR(1, 1) and DCC-EGARCH(1, 1), both characterized by weaker sample performance within the MCS. In periods of high turbulence, modelling directly the conditional correlation and accounting for the leverage effect in the conditional variances becomes more important than in the full sample. Table 11 shows that DCC-type models with FI-EGARCH conditional variances dominate the MCS and have the smallest losses. Among these, we find one CCC specification, with FI-EGARCH dynamics for the conditional variances, which suggests that adequately modelling long memory and asymmetry in the conditional variances can in some cases compensate for the restrictive assumption of no dynamics in the conditional correlation over a shorter period of time. Furthermore, the exclusion of other specifications which account for asymmetry in the variance, e.g., DCC with GJR dynamics, underlines the importance of the EGARCH parameterization of volatilities. Results for the last sub-sample, reported in Table 12, are in line with those obtained for the full sample. The MCS is dominated by specifications in the DCC family and only those including long memory are included under all loss functions. In this sub-sample, the non-rejection of the full BEKK specification is somehow surprising and may be suggesting that modelling spillover effects can in some cases, over short horizons, compensate the loss of accuracy induced by the restrictive short-memory assumption.

5.2. Robustness to the forecasts horizon

As a second robustness check, we test our findings with respect to longer forecast horizons. The MCS for the multi-step (5 and 20-day) forecast evaluation over the full sample are reported in Tables 12 and 13. As expected, the average loss increases with the forecast horizon, irrespectively of the evaluation period or the choice of the loss function. For longer horizons, the performance of models with similar properties and structure tend to cluster since they converge to the same long-run variance matrix, but differences between clusters increase since different specifications can imply different levels for the long-run variance. The composition of the MCS is in line with the one-step ahead case. For longer horizons, the MCS reduces in size making it easier to separate between superior and inferior models. For both horizons, the MCS includes only models with dynamic correlations and long memory supporting strongly the need to account for fractional integration in the volatility decay rates.

6. Concluding discussion and remarks

Several multivariate GARCH models have been used in the energy literature to explore the volatilities and correlations of oil and oil-related product prices. However, no specification including long memory has been tested yet at multivariate level. In practice, such investigation is important to avoid misspecification of volatilities decay rate which may lead to inaccurate forecasting and unreliable risk assessments.

This paper advances research on the co-movements of crude oil and refined products by looking into the forecasting accuracy gains from using multivariate GARCH models with long memory over the short-memory benchmarks commonly used in the energy literature. The empirical analysis considers spot price returns for three major oil-related markets. We compare 30 multivariate GARCH models with different characteristics. All models are estimated in one step using pseudo-maximum likelihood methods to simplify the computation of robust standard errors and to avoid discrepancies in the forecasting performance arising from different estimation methods. In-sample results, based on asymptotic standard errors, show strong evidence of GARCH-type dynamics, long-range dependence and leverage effects in the individual volatilities. In terms of the multivariate structure, the data strongly support the hypothesis of dynamic conditional correlations.

Using a fixed rolling window scheme, we assess the one, five and twenty-day ahead forecasting accuracy of the models with two statistical approaches: the MCS method and the SPA test. We employ several matrix loss functions, which are robust to the choice of the volatility proxy. We then study the models' forecasting performances in an economically meaningful way by predicting the Value-at-Risk for short and long positions. Our results suggest that models with long-memory decay rate surpass the short-memory counterparts from a statistical as well as an economic perspective and their use can significantly improve oil markets risk assessments. The sensitivity of the results with respect to the forecasting sample is tested by considering, in addition to the full sample, three sub-samples with homogenous volatility dynamics. Our findings indicate that over calm markets, constant conditional correlation specifications cannot be rejected. However, in the full sample and turbulent market periods, the short-memory constant-correlation models are always rejected in favour of long-memory dynamic-correlation models. Finally, for longer forecasting horizons, we find that the set of superior models includes only long-memory specifications suggesting that such feature is indeed essential for successful prediction of risk in oil markets.

Our results are important for agents trading in any of the three commodities and particularly so to those who trade in crack spreads. Such agents include refiners, who are by nature exposed to both crude oil and refined products, as well as oil trading companies who tend to have risk exposures to both the crude and refined sides of the market. Risk managers in such companies seek better ways to improve their VaR forecasts and we find strong evidence of superior performance of models with fractional integration, dynamic correlations and EGARCH-type asymmetries.

This paper considers only forecasts based on MGARCH models. It would be interesting to investigate the forecasting performance of other types of multivariate volatility models with long memory and asymmetries, such as the factor multivariate stochastic volatility model with long memory of Asai and McAleer (2015) and the long-memory regime switching model in Diebold and Inoue (2001), and Bayesian network (e.g., see Cuestas and Ordóñez, 2018, Li et al., 2016, Peraza and Halliday, 2010).

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to thank Sebastien Laurent and Giuseppe Storti and the participants of the International Conference MAF 2018, Madrid, Spain, for thoughtful comments and suggestions. Our thanks are extended to the Editor-in-Chief, Professor Richard S. J. Tol, and two anonymous reviewers whose valuable comments helped us improve our paper. All remaining errors are, of course, our own. Declarations of interest: none.

Agnolucci, P., 2009. Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. Energy Economics 31 (2), 316–321.

Aloui, C., Mabrouk, S., 2010. Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. Energy Policy 38 (5), 2326–2339.

Amendola, A., Candila, V., 2016. Evaluation of volatility predictions in a VaR framework. Quantitative Finance 16 (5), 695–709.

Andersen, T. G., Bollerslev, T., Meddahi, N., 2005. Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities. Econometrica 73 (1), 279–296.

Asai, M., McAleer, M., 2015. Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance. Journal of Econometrics 189 (2), 251–262.

Bauwens, L., Braione, M., Storti, G., 2016. Forecasting comparison of long term component dynamic models for realized covariance matrices. Annals of Economics and Statistics 123/124, 103–134.

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

Becker, R., Clements, A. E., Doolan, M. B., Hurn, A. S., 2015. Selecting volatility forecasting models for portfolio allocation purposes. International Journal of Forecasting 31 (3), 849–861.

Block, A. S., Righi, M. B., Schlender, S. G., Coronel, D. A., 2015. Investigating dynamic conditional correlation between crude oil and fuels in non-linear framework: The financial and economic role of structural breaks. Energy Economics 49, 23–32.

Bollerslev, T., 1990. Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model. The Review of Economics and Statistics 72 (3), 498–505.

Borenstein, S., Cameron, A. C., Gilbert, R., 1997. Do gasoline prices respond asymmetrically to crude oil price changes? The Quarterly Journal of Economics 112 (1), 305–339.

Brunetti, C., Gilbert, L. C., 2000. Bivariate FIGARCH and fractional cointegration. Journal of Empirical Finance 7 (5), 509–530.

Chang, C.-L., McAleer, M., Tansuchat, R., 2010. Analyzing and forecasting volatility spillovers, asymmetries and hedging in major oil markets. Energy Economics 32 (6), 1445–1455.

Chang, C.-L., McAleer, M., Tansuchat, R., 2011. Crude oil hedging strategies using dynamic multivariate GARCH. Energy Economics 33 (5), 912–923.

Chevallier, J., 2012. Time-varying correlations in oil, gas and CO_2 prices: an application using BEKK, CCC, and DCC-MGARCH models. Applied Economics 44 (32), 4257–4274.

Chkili, W., Hammoudeh, S., Nguyen, D. K., 2014. Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Economics 41, 1–18.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International Economic Review 39 (4), 841–862.

Cuestas, J. C., Ordóñez, J., 2018. Oil prices and unemployment in the UK before and after the crisis: A Bayesian VAR approach. A note. Physica A: Statistical Mechanics and its Applications 510, 200–207.

Diebold, F., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105 (1), 131–159.

Engle, R. 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, R. F., Kroner, K. F., 1995. Multivariate Simultaneous Generalized ARCH. Econometric Theory 11 (1), 122–150.

Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. Journal of Applied Econometrics 18 (6), 641–663.

Grier, K. B., Henry, O. T., Olekalns, N., Shields, K., 2004. The Asymmetric Effects of Uncertainty on Inflation and Output Growth. Journal of Applied Econometrics 19 (5), 551–565.

Gronwald, M., 2012. A characterization of oil price behavior – Evidence from jump models. Energy Economics 34 (5), 1310–1317.

Hansen, P. R., 2005. A Test for Superior Predictive Ability. Journal of Business & Economic Statistics 23 (4), 365–380.

Hansen, P. R., Lunde, A., Nason, J. M., 2003. Choosing the Best Volatility Models: The Model Confidence Set Approach. Oxford Bulletin of Economics and Statistics 65, 839–861.

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

Ji, Q., Fan, Y., 2012. How does oil price volatility affect non-energy commodity markets? Applied Energy 89 (1), 273–280.

Kang, S. H., Kang, S. M., Yoon, S. M., 2009. Forecasting volatility of crude oil markets. Energy Economics 31 (2), 119–125.

Kapetanios, G., Papailias, F., Taylor, A. R., 2019. A generalised fractional differencing bootstrap for long memory processes. Journal of Time Series Analysis 40 (4), 467–492.

Kaufmann, R. K., Laskowski, C., 2005. Causes for an asymmetric relation between the price of crude oil and refined petroleum products. Energy Policy 33 (12), 1587–1596.

Lahiri, S. N., 1999. Theoretical comparisons of block bootstrap methods. The Annals of Statistics 27 (1), 386–404.

Laurent, S., Rombouts, J. V. K., Violante, F., 2012. On the forecasting accuracy of multivariate GARCH models. Journal of Applied Econometrics 27 (6), 934–955.

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

Lee, T. K., Zyren, J., 2007. Volatility spillover versus contagion: Empirical evidence from crude oil and gasoline markets. In: Developing & Delivering Affordable Energy in the 21st Century. 27th USAEE/IAEE North American Conference.

Li, X., Chen, G., Zhu, H., 2016. Quantitative risk analysis on leakage failure of submarine oil and gas pipelines using Bayesian network. Process Safety and Environmental Protection 103, 163–173.

Ling, S., McAleer, M., 2003. Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19 (2), 280–310.

McAleer, M., Hoti, S., Chan, F., 2009. Structure and asymptotic theory for multivariate asymmetric conditional volatility. Econometric Reviews 28 (5), 422–440.

Nomikos, N. K., Pouliasis, P. K., 2011. Forecasting petroleum futures markets volatility: The role of regimes and market conditions. Energy Economics 33 (2), 321–337.

Patton, A., Politis, D. N., White, H., 2009. Correction to "Automatic block-length selection for the dependent bootstrap". Econometric Reviews 28 (4), 372–375.

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

Patton, A. J., Sheppard, K., 2009. Evaluating volatility and correlation forecasts. In: Mikosch, T., Kreiß, J.-P., Davis, R. A., Andersen, T. G. (Eds.), Handbook of Financial Time Series. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 801–838.

Peraza, L., Halliday, D. M., 2010. A fast dynamic Bayesian network algorithm for structure learning using a time-lagged partial correlation matrix. In: ICSES 2010 International Conference on Signals and Electronic Circuits. pp. 25–28.

Politis, D. N., Romano, J. P., 1994. The Stationary Bootstrap. Journal of the American Statistical Association 89 (428), 1303–1313.

Rahman, S., Serletis, A., 2012. Oil price uncertainty and the canadian economy: Evidence from a VARMA, GARCH-in-Mean, asymmetric BEKK model. Energy Economics 34 (2), 603–610.

Robinson, P. M., 1995. Gaussian Semiparametric Estimation of Long Range Dependence. The Annals of Statistics 23 (5), 1630–1661.

Silvennoinen, A., Teräsvirta, T., 2009. Multivariate GARCH models. In: Mikosch, T., Kreiß, J.-P., Davis, R. A., Andersen, T. G. (Eds.), Handbook of Financial Time Series. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 201–229.

Suenaga, H., Smith, A., 2011. Volatility dynamics and seasonality in energy prices: implications for crack-spread price risk. The Energy Journal 32 (3), 27–58.

Tabak, B. M., Cajueiro, D. O., 2007. Are the crude oil markets becoming weakly efficient over time? A test for time-varying long-range dependence in prices and volatility. Energy Economics 29 (1), 28–36.

Tong, B., Wu, C., Zhou, C., 2013. Modeling the co-movements between crude oil and refined petroleum markets. Energy Economics 40, 882–897.

Vacha, L., Barunik, J., 2012. Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. Energy Economics 34 (1), 241–247.

Vo, M., 2011. Oil and stock market volatility: A multivariate stochastic volatility perspective. Energy Economics 33 (5), 956–965.

Wang, Y., Wu, C., 2012. Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? Energy Economics 34 (6), 2167–2181.

Model	Equation	Parameters
GARCH(1,1)	$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-j}$	$\{\omega, \alpha_1, \beta_1\}$
IGARCH(1,1)	$h_t = \omega + \alpha \epsilon_{t-1}^2 + (1-\alpha) h_{t-j}$	$\{\omega, \alpha\}$
EGARCH(1,1)	$\log h_t = \omega + \alpha \left\lvert \frac{\epsilon_{t-1}}{\sqrt{h_{t-1}}} \right\rvert + \gamma \frac{\epsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta \log h_{t-1}$	$\{\omega, \alpha, \gamma, \beta\}$
GJR-GARCH(1,1)	$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-1}^2 I_{\epsilon_{t-1}<0} + \beta_1 h_{t-j}$	$\{\omega, \alpha_1, \alpha_2, \beta_1\}$
FIGARCH(1,d,1)	$h_{it} = \omega + \beta_1 h_{it-1} + [1 - (1-\beta_1 L)^{-1} (1-\phi_1 L)(1-L)^d] \epsilon_{it}^2$	$\{\omega, \phi_1, \beta_1, d\}$
FIEGARCH(1,d,1)	$\log h_t = \omega + \frac{a(L)}{b(L)} (1-L)^{-d} g(\epsilon_{t-1})$	$\{\omega, a, b, d, \theta, \delta\}$

Table 1: Univariate volatility processes for the CCC/DCC class

Model	Dynamic correlation	Asymmetries	Volatility long memory decay	Spillover effects
DBEKK (3)	✓			
BEKK (3)	✓			✓
ABEKK (4)	✓	✓		✓
AGARCH (5), (11)		✓		✓
CCC-GARCH (5)				
CCC-IGARCH (5)				
CCC-EGARCH (5)		✓		
CCC-GJR (5)		✓		
CCC-FIGARCH (5)			✓	
CCC-FIEGARCH (5)		✓	✓	
DCC-GARCH (5)–(7)	✓			
DCC-IGARCH (5)–(7)	✓			
DCC-EGARCH (5)–(7)	✓	✓		
DCC-GJR (5)–(7)	✓	✓		
DCC-FIGARCH (5)–(7), (10)	✓		✓	
DCC-FIEGARCH (5), (6)–(8)	✓	✓	✓	

Table 2: MGARCH models and characteristics. *Notes:* For univariate specifications for the CCC/DCC class, refer to Table 1.

	Mean	Max	Min	Standard deviation	Skewness coefficient	Kurtosis	JB	$Q(10)$	$Q^2(10)$
CO	0.0004	0.159	-0.181	0.025	-0.057	5.75	**370.6**	**58.12**	**352.1**
CG	0.0003	0.137	-0.145	0.027	-0.074	6.08	**432.2**	**52.70**	**654.3**
HO	0.0003	0.164	-0.205	0.025	-0.055	6.56	**521.3**	**57.34**	**743.3**

Table 3: Descriptive statistics of energy price returns. *Notes:* JB is the Jarque–Bera test statistic; $Q(10)$ and $Q^2(10)$ are the Ljung–Box statistics, respectively, for the returns and the squared returns for correlation up to lag 10. Boldface entries are significant at the 10% significance level.

Model	N_p	LLk	AIC	BIC
DBEKK	12	-18321	36666	36737
BEKK	24	-18218	36484	36615
ABEKK	33	-17989	36044	36240
AGARCH	33	-18011	36088	36284
CCC-GARCH	12	-18202	36428	36499
CCC-IGARCH	9	-18301	36428	36499
CCC-EGARCH	15	-18065	36428	36499
CCC-GJR	15	-18063	36428	36499
CCC-FIGARCH	16	-18011	36428	36499
CCC-FIEGARCH	22	-18002	36428	36499
DCC-GARCH	11	-17695	35414	35515
DCC-IGARCH	8	-17832	35976	35665
DCC-EGARCH	14	-17611	35391	35509
DCC-GJR	14	-17607	35402	35517
DCC-FIGARCH	17	-17684	**35402**	**35503**
DCC-FIEGARCH	23	-17661	**35368**	**35501**

Table 4: Information criteria. *Notes:* N_p is the number of estimated parameters of each model. LLk is the log-likelihood of the models: these values are not directly comparable across models due to the varying number of parameters. The AIC and BIC information criteria are computed respectively as $-2\text{LLk} + 2N_p$ and $-2\text{LLk} + N_p \ln n$, where n is the total number of observations in the sample. Boldface values correspond to the best-performing models.

Loss function	Type
Frobenius $\text{tr}\left[\left(\hat{\Sigma}_t - H_{it}\right)'\left(\hat{\Sigma}_t - H_{it}\right)\right]$	Symmetric
Euclidean $\text{vech}\left(\hat{\Sigma}_t - H_{it}\right)'\text{vech}\left(\hat{\Sigma}_t - H_{it}\right)$	Symmetric
MSFE $\frac{1}{T}\text{vec}\left(\hat{\Sigma}_t - H_{it}\right)'\text{vec}\left(\hat{\Sigma}_t - H_{it}\right)'$	Symmetric
QLIKE $\log\lvert H_t\rvert + \text{vec}\left(H_{it}^{-1}\hat{\Sigma}_t\right)'\iota$	Symmetric
Stein $\text{tr}\left(H_{it}^{-1}\hat{\Sigma}_t\right) - \log\left\lvert H_{it}^{-1}\hat{\Sigma}_t\right\rvert - n$	Asymmetric
VDN $\text{tr}(\hat{\Sigma}_t \log \hat{\Sigma}_t - \hat{\Sigma}_t \log H_{it} - \hat{\Sigma}_t + H_{it})$	Asymmetric

Table 5: Loss functions. *Notes:* H_{it} denotes the predicted covariance matrix for day t, $\hat{\Sigma}_t$ the conditional covariance matrix proxy, ι a vector of ones, T the out-of-sample length, and n the sample size. Operators vec and vech stack, respectively, the columns and the lower triangular portion of a matrix into a vector; tr denotes the trace of a matrix.

Benchmark	L_E			L_F			L_S		
	p_L	p_C	p_U	p_L	p_C	p_U	p_L	p_C	p_U
CCC-GARCH(1,1)	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02
CCC-EGARCH(1,1)	0.01	0.02	0.02	0.00	0.00	0.00	0.02	0.03	0.03
CCC-FIGARCH(1,1)	0.00	0.00	0.00	0.02	0.02	0.02	0.05	0.05	0.08
CCC-FIEGARCH(1,1)	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.02	0.02
DCC-GARCH(1,1)	0.05	0.04	0.04	0.09	**0.10**	0.14	0.02	0.02	0.02
DCC-EGARCH(1,1)	0.10	**0.11**	0.10	0.08	0.08	0.09	0.08	0.09	0.13
DCC-FIGARCH(1,1)	0.09	0.02	0.02	0.16	**0.23**	0.46	0.04	0.04	0.05
DCC-FIEGARCH(1,1)	0.11	**0.11**	0.17	0.32	**0.82**	0.98	0.05	0.05	0.05

Table 6: The SPA test. *Notes:* L_E, L_F and L_S denote, respectively, the Euclidean, Frobenius and Stein loss functions; p_C, p_L, p_U are, respectively, the consistent p-values, their lower and upper bounds. Boldface consistent p-values indicate non-rejection of the null at the significance level 10% (see Hansen, 2005 for the details). The number of sieve bootstrap samples used to obtain the distribution under the null is 10,000.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK	✓	✓					33
DBEKK	✓	✓					33
ABEKK	✓	✓	✓			✓	67
AGARCH	✓	✓					33
CCC-GARCH	✓	✓					33
CCC-IGARCH	✓						17
CCC-EGARCH	✓	✓					33
CCC-GJR	✓	✓					33
CCC-FIGARCH	✓	✓					33
CCC-FIEGARCH	✓	✓					40
DCC-GARCH	✓	✓					50
DCC-IGARCH	✓						25
DCC-EGARCH	✓	✓					54
DCC-GJR	✓	✓					52
DCC-FIGARCH	✓	✓				✓	70
DCC-FIEGARCH	✓	✓	✓			✓	72

Table 7: Full Sample Model Confidence Set at the 90% level. *Notes:* Ec and Fr denote, respectively, the Euclidean and Frobenius loss functions, MSFE is the Mean Squared Forecast Error, VDN the von Neumann distance. Perf is the percentage of inclusion of each model in the MCS across the six loss functions.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK							0
DBEKK							0
ABEKK	✓	✓				✓	50
AGARCH							0
CCC-GARCH							0
CCC-IGARCH							0
CCC-EGARCH							0
CCC-GJR							0
CCC-FIGARCH							0
CCC-FIEGARCH		✓					16
DCC-GARCH						✓	17
DCC-IGARCH							0
DCC-EGARCH			✓			✓	20
DCC-GJR							0
DCC-FIGARCH	✓	✓				✓	55
DCC-FIEGARCH	✓	✓	✓			✓	69

Table 8: Full Sample Model Confidence Set at the 75% level.

Model	Long positions			Short positions		
	5%	2.5%	1%	5%	2.5%	1%
DBEKK	**0.00**	**0.00**	**0.00**	**0.01**	**0.01**	**0.02**
BEKK	0.11	**0.02**	0.18	0.12	0.17	**0.01**
ABEKK	0.24	0.22	**0.04**	0.27	0.28	0.21
AGARCH	**0.04**	**0.04**	**0.04**	**0.02**	**0.02**	**0.04**
CCC-GARCH	**0.00**	**0.02**	**0.04**	**0.03**	**0.03**	**0.03**
CCC-IGARCG	**0.00**	**0.02**	**0.04**	**0.03**	**0.03**	**0.03**
CCC-EGARCH	**0.00**	**0.02**	**0.04**	**0.03**	**0.03**	**0.03**
CCC-GJR	**0.00**	**0.02**	**0.04**	**0.03**	**0.03**	**0.03**
CCC-FIGARCH	**0.00**	**0.02**	**0.04**	**0.03**	**0.03**	**0.03**
CCC-FIEGARCH	**0.00**	**0.02**	**0.04**	**0.03**	**0.03**	**0.03**
DCC-GARCH	0.14	0.12	0.12	0.11	0.06	**0.04**
DCC-IGARCH	0.14	0.12	0.12	0.11	0.06	**0.04**
DCC-EGARCH	0.14	0.12	0.12	0.11	0.06	**0.04**
DCC-GJR	0.14	0.12	0.12	0.11	0.06	**0.04**
DCC-FIGARCH	0.40	0.27	**0.04**	0.49	0.13	0.36
DCC-FIEGARCH	0.59	0.40	0.40	0.49	0.76	0.47

Table 9: Likelihood Ratio (LR) test results. *Notes:* p-values of the LR Conditional Coverage Test for short and long positions for the equal weighted portfolios' Value-at-Risk. Boldface p-values correspond to rejection of the null hypothesis at the 5% significance level.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK	✓	✓	✓				55
DBEKK	✓						16
ABEKK	✓						17
AGARCH							0
CCC-GARCH	✓	✓					33
CCC-IGARCH							0
CCC-EGARCH	✓	✓	✓	✓			60
CCC-GJR	✓						17
CCC-FIGARCH	✓	✓	✓				55
CCC-FIEGARCH	✓	✓	✓				55
DCC-GARCH	✓	✓		✓			51
DCC-IGARCH							0
DCC-EGARCH	✓	✓					50
DCC-GJR	✓	✓					50
DCC-FIGARCH	✓	✓				✓	70
DCC-FIEGARCH	✓	✓	✓				70

Table 10: Calm market (sub-sample 1) Model Confidence Set at the 90% level.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK	✓						17
DBEKK	✓	✓					0
ABEKK	✓	✓					33
AGARCH							0
CCC-GARCH	✓						17
CCC-IGARCH	✓	✓					33
CCC-EGARCH	✓	✓					17
CCC-GJR							0
CCC-FIGARCH	✓						17
CCC-FIEGARCH	✓	✓	✓				40
DCC-GARCH	✓	✓					17
DCC-IGARCH	✓	✓					33
DCC-EGARCH	✓	✓	✓				54
DCC-GJR	✓						33
DCC-FIGARCH	✓	✓		✓		✓	80
DCC-FIEGARCH	✓	✓	✓		✓	✓	85

Table 11: Turbulent market (sub-sample 2) Model Confidence Set at the 90% level.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK	✓						17
DBEKK							0
ABEKK	✓	✓					33
AGARCH							0
CCC-GARCH	✓						17
CCC-IGARCH	✓						17
CCC-EGARCH	✓	✓					33
CCC-GJR	✓	✓					33
CCC-FIGARCH	✓	✓			✓		50
CCC-FIEGARCH	✓	✓		✓			50
DCC-GARCH	✓	✓					33
DCC-IGARCH	✓	✓					33
DCC-EGARCH	✓	✓	✓				54
DCC-GJR	✓	✓					52
DCC-FIGARCH	✓	✓	✓	✓		✓	76
DCC-FIEGARCH	✓	✓	✓			✓	77

Table 12: Fairly volatile market (sub-sample 3) Model Confidence Set at the 90% level.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK	✓						17
DBEKK							0
ABEKK	✓						17
AGARCH							0
CCC-GARCH							0
CCC-IGARCH	✓						17
CCC-EGARCH	✓						17
CCC-GJR							0
CCC-FIGARCH	✓	✓	✓				33
CCC-FIEGARCH	✓	✓					33
DCC-GARCH	✓						17
DCC-IGARCH	✓						17
DCC-EGARCH	✓	✓					45
DCC-GJR	✓	✓					33
DCC-FIGARCH	✓	✓				✓	75
DCC-FIEGARCH	✓	✓	✓			✓	78

Table 13: Five-day ahead Model Confidence Set at the 90% level.

Model	Ec	Fr	MSFE	QLIKE	STEIN	VDN	Perf
BEKK							0
DBEKK							0
ABEKK						0	
AGARCH							0
CCC-GARCH							0
CCC-IGARCH							0
CCC-EGARCH	✓						17
CCC-GJR	✓						33
CCC-FIGARCH	✓	✓					33
CCC-FIEGARCH	✓	✓					33
DCC-GARCH	✓	✓					33
DCC-IGARCH							0
DCC-EGARCH	✓	✓					47
DCC-GJR	✓						17
DCC-FIGARCH	✓	✓				✓	70
DCC-FIEGARCH	✓	✓	✓			✓	72

Table 14: Twenty-day ahead Model Confidence Set at the 90% level.

Appendix A. In-sample estimates

Criterion \ Lag l	1	2	3	4	5	6	7
BIC(l)	**11.1762**	11.1231	11.1001	10.9761	10.8651	10.8322	10.7844
AIC(l)	**11.3374**	11.3342	11.3301	11.2201	11.2109	11.2051	11.2001
HQ(l)	**11.2733**	11.2721	11.2705	11.2001	11.1987	11.1562	11.1438

Table A.1: Selection criteria for optimal lag length determination. *Notes:* BIC(l) denotes the Bayesian information criterion; AIC(l) the Akaike information criterion; HQ(l) the Hannan information criterion. Boldface reports correspond to the optimal lag length for each criterion.

	μ	Φ		
CO	0.0012	-0.20111	0.00312	0.01612
	(0.008)	(-2.998)	(0.132)	(0.072)
CG	0.0004	**0.00134**	**-0.10000**	0.00742
	(0.032)	(0.00145)	(-1.997)	(0.0154)
HO	**0.0012**	0.00240	0.00126	**-0.18131**
	(1.762)	(0.0347)	(0.141)	(-4.356)

Table A.2: Estimated coefficients of VAR mean equation. *Notes:* Estimation is carried out via maximum likelihood methods with robust standard errors. All the eigenvalues of the companion matrix are smaller than 1 in absolute value. Boldface entries indicate significance at the 10% level. The log-likelihood is -18.364 and the AIC is 36716.

Diagnostic tests	$Q(5)$	$Q(10)$	JB	ARCH_{10}
$\hat{\epsilon}_{CO}$	112.11	96.14	424.71	675.44
	(0.076)	(0.094)	(0.0011)	(0.000)
$\hat{\epsilon}_{CG}$	132.11	102.01	122.11	522.03
	(0.087)	(0.102)	(0.076)	(0.001)
$\hat{\epsilon}_{HO}$	132.31	89.55	122.11	601.18
	(0.076)	(0.108)	(0.076)	(0.000)

Table A.3: Post-estimation diagnostic tests. *Notes:* Diagnostic post-estimation tests are conducted on the residuals of the series, $\hat{\epsilon}$. Q(l) denotes the Ljung–Box–Pierce portmanteau test statistic with maximal lag equal to l for the residuals, JB is the Jarque–Bera test for normality, ARCH_{10} the ARCH tests for a maximal lag of order 10. p-values are reported in parentheses.

	C			A			G		
CO	**0.017**** (2.113)	0.043 (1.032)	**0.021*** (1.987)	**0.103**** (5.041)			**0.887**** (4.021)		
CG		**0.003**** (2.042)	0.019 (0.879)		**0.096**** (4.276)			**0.902**** (5.321)	
HO			**0.014*** (1.995)			**0.062**** (3.553)			**0.874**** (3.873)

Table A.4: Diagonal BEKK$(1, 1, 1)$: model (3) with off-diagonal elements of A and G set equal to 0. *Notes:* The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	C			A			G		
CO	**0.013**** (3.538)	**0.054** (1.785)	**0.033*** (2.033)	**0.084**** (2.984)	-0.014 (-1.223)	0.005 (0.098)	**0.915*** (7.675)	- 0.037 (-0.238)	-0.032 (-1.378)
CG		**0.018*** (2.421)	**0.026** (0.879)	**-0.034** (-1.754)	**0.100**** (3.548)	**0.042** (1.711)	**0.041** (1.888)	**0.897**** (7.090)	**0.153** (1.703)
HO			**0.005*** (2.067)	**0.063** (1.688)	0.075 (1.501)	**0.034**** (3.774)	**-0.061** (-1.811)	**0.058** (1.685)	**0.854**** (8.113)

Table A.5: Full BEKK$(1, 1, 1)$: model (4). *Notes:* The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	C			A		
CO	**0.001*** (2.081)	**0.037** (1.785)	**0.053** (1.703)	**0.098*** (2.561)	-0.028 (-1.183)	0.013 (1.792)
CG		**0.004**** (4.982)	0.032 (0.003)	**-0.188** (-5.627)	**0.102**** (3.548)	**0.048** (1.651)
HO			**0.007**** (3.981)	**0.012** (1.682)	0.073 (1.986)	**0.054**** (7.381)
	G			D		
CO	**0.900**** (5.441)	- 0.028 (-1.011)	-0.023 (-1.386)	**-0.327*** (-2.031)	0.041 (0.056)	0.012 (1.061)
CG	**0.022** (1.663)	**0.896**** (4.936)	**0.147** (1.865)	0.016 (0.037)	**-0.545**** (-4.561)	0.025 (0.012)
HO	**-0.063** (-1.777)	0.027 (1.030)	**0.818**** (6.463)	0.033 (0.065)	0.018 (0.197)	**-0.009** (-1.765)

Table A.6: Asymmetric BEKK$(1, 1, 1)$: model (5). *Notes:* The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	w	A			B		
CO	**0.005*** (2.531)	**0.101**** (3.652)	**-0.016** (-1.178)	0.011 (0.096)	**0.897**** (4.356)	-0.014 (1.002)	**-0.008** (-1.765)
CG	**0.003**** (3.456)	**-0.072** (-1.753)	**0.098**** (7.324)	0.011 (1.012)	**0.012** (0.189)	**0.901**** (3.882)	**0.017** (1.673)
HO	**0.008*** (1.999)	**0.021** (1.674)	0.015 (0.037)	**0.061**** (2.891)	**-0.037** (-1.765)	0.013 (1.001)	**0.793**** (5.683)
	C			R			
CO	**-0.356*** (-0.198)	0.002 (0.073)	0.035 (1.002)	1	**0.576*** (2.061)	**0.683**** (3.875)	
CG	0.018 (0.045)	**-0.312*** (-2.031)	0.101 (0.389)	**0.576*** (2.061)	1	**0.451**** (4.564)	
HO	0.029 (0.851)	0.017 (0.101)	**-0.057** (1.688)	**0.683**** (3.875)	**0.451**** (4.564)	1	

Table A.7: AGARCH$(1, 1, 1)$: model (11). *Notes:* The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	ω	α_1	β_1		$\mathbf{R}$	
CO	**0.001***	**0.098****	**0.901****	1	**0.458****	**0.591****
	(2.031)	(5.456)	(4.229)		(4.136)	(5.432)
CG	**0.012***	**0.021****	**0.977****	**0.458****	1	**0.653****
	(2.157)	(4.374)	(4.108)	(4.136)		(3.882)
HO	**0.006***	**0.033****	**0.744****	**0.591****	**0.653****	1
	(1.999)	(2.987)	(3.554)	(5.432)	(3.882)	

Table A.8: CCC model. *Notes:* The individual volatilities follow GARCH$(1,1)$ processes as in equation (6) with $p = q = 1$. The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	ω	α_1	β_1	λ_1	λ_2
CO	**0.014***	**0.084****	**0.912****	**0.338****	**0.593****
	(2.051)	(5.456)	(3.514)	(5.987)	(8.378)
CG	**0.023***	**0.062****	**0.935****		
	(2.103)	(7.881)	(2.873)		
HO	**0.018***	**0.049****	**0.751****		
	(2.010)	(6.987)	(5.467)		

Table A.9: DCC model with the conditional covariance matrix evolving according to (7) and (8). *Notes:* The individual volatilities follow GARCH$(1,1)$ processes as in equation (6) with $p = q = 1$. The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	ω	β_1	ϕ_1	d	λ_1	λ_2
CO	**0.021**	**0.054****	**0.932****	**0.43****	**0.354****	**0.601****
	(1.691)	(5.456)	(3.514)	(3.588)	(6.238)	(4.982)
CG	**0.023**	**0.042****	**0.911****	**0.33****		
	(1.700)	(7.881)	(2.873)	(4.716)		
HO	**0.018***	**0.049****	**0.921****	0.011		
	(2.010)	(6.987)	(5.467)	(1.352)		

Table A.10: DCC-FIGARCH model with R_t evolving according to (7) and (8). *Notes:* The individual volatilities follow FI-GARCH$(1, d, 1)$ processes. The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.

	ω	a	b	δ	θ	d	λ_1	λ_2
CO	**0.021**	**0.354****	**0.932****	**0.071****	**-0.044***	**0.40****	**0.386****	**0.623****
	(1.691)	(5.456)	(3.514)	(4.110)	(2.031)	(7.451)	(5.882)	(4.023)
CG	**0.023**	**0.362****	**0.934****	**1.034****	**-0.076***	**0.38****		
	(1.700)	(7.881)	(2.873)	(6.187)	(2.011)	(6.291)		
HO	**0.018**	**0.349****	**0.916****	**1.001****	**-0.008***	**0.003**		
	(2.010)	(6.987)	(5.467)	(5.151)	(1.1998)	(1.971)		

Table A.11: DCC-FIEGARCH model with D_t as in (10) and R_t evolving according to (7) and (8). *Notes:* The individual volatilities follow FI-EGARCH$(1, d, 1)$ processes as in equations (8) and (9). The two entries for each parameter are their respective estimated value and robust t-ratio. Boldface entries indicate significance at the 10% level; asterisked boldface entries significance at the 5% level; double-asterisked boldface entries significance at the 1% level.