DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Models with Multiplicative Decomposition of Conditional Variances and Correlations

Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta

CREATES Research Paper 2018-14

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Models with Multiplicative Decomposition of Conditional Variances and Correlations

Cristina Amado
University of Minho and NIPE
CREATES and Aarhus University
camado@eeg.uminho.pt

Annastiina Silvennoinen
School of Economics and Finance
Queensland University of Technology
silvennoinen@qut.edu.au

Timo Teräsvirta
CREATES and Aarhus University
C.A.S.E., Humboldt-Universität zu Berlin
tterasvirta@econ.au.dk

March 23, 2018

Abstract

Univariate and multivariate GARCH type models with multiplicative decomposition of the variance to short and long run components are surveyed. The latter component can be either deterministic or stochastic. Examples of both types are studied.

JEL Classification Codes: C12; C32; C51; C52.

Keywords: Conditional heteroskedasticity; Deterministically varying correlations; Multiplicative decomposition; Nonstationary volatility.

Acknowledgements. This research has been supported by Center for Research in Econometric Analysis of Time Series (CREATES), funded by the Danish National Research Foundation, Grant No. DNFR 78. The first author also acknowledges funding from COMPETE (Ref. No. POCI-01-0145-FEDER-006683), with the FCT/MEC's (Fundação para a Ciência e a Tecnologia, I.P.) financial support through national funding and by the ERDF through the Operational Programme on "Competitiveness and Internationalization" - COMPETE 2020 under the PT2020 Partnership Agreement. We wish to thank Christian Conrad for useful remarks. Any errors and shortcomings in this work are the authors' responsibility.

1 Introduction

Many daily or weekly volatility series appear nonstationary. In the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) framework this nonstationarity has been explicitly modelled by Integrated GARCH models (Engle and Bollerslev, 1986) or using a more general version, the Fractionally Integrated GARCH model (Baillie, Bollerslev and Mikkelsen, 1996). Another strand of literature, see for example Lamoureux and Lastrapes (1990) or Mikosch and Stărică (2004), builds on the assumption that nonstationarity is due to structural changes in the volatility process. One way of adjusting the GARCH model to the latter type of nonstationarity is to modify the observations (typically but not exclusively daily returns of financial assets) to fit a stationary GARCH model. This is done by augmenting the GARCH model multiplicatively by a positive-valued component. van Bellegem and von Sachs (2004) and Feng (2004) are the first examples of this approach.

A GARCH model may be multiplicatively augmented also because of the desire to explain and predict variations in volatility by economic variables. GARCH-MIDAS models, pioneered by Engle, Ghysels and Sohn (2013), are an example of this. In this review models of both types of multiplicative decomposition are considered. The plan of the review is as follows. The decomposition is described in Section 2. Section 3 concerns models with a deterministic multiplicative component and in Section 4 this component is stochastic. Multivariate generalisations are discussed in Section 5. Section 6 contains final remarks.

2 Multiplicative decomposition of variance

Most univariate or single-equation models to be considered in this article are of the following form

$$y_t - \mu_t = \varepsilon_t = z_t h_t^{1/2} g_t^{1/2} \tag{1}$$

where z_t is iid(0,1). In this review it is assumed that μ_t is known so that ε_t is observable (y_t is assumed observable). The conditional variance component is assumed to have a GARCH representation (Bollerslev, 1986, and Taylor, 1986). Thus, when $g_t \equiv 1$,

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2}$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$, $j = 1, ..., q-1$, $\alpha_q > 0$, $\beta_j \geq 0$, $j = 1, ..., p$. The process is weakly stationary if and only if $\sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j < 1$.

The positive-valued function g_t is either deterministic or stochastic and represents the slowly moving component of ε_t^2. It follows that $\sigma_t^2 = \mathsf{E}h_t g_t$ is the total time-varying variance of ε_t at time t. The role of h_t is to chacterise clustering of volatility present for instance in asset or index return series of sufficiently high frequency, as already observed by Mandelbrot (1963) and originally parameterised by Engle (1982). When g_t is deterministic and nonconstant, ε_t is nonstationary. One of the important purposes of g_t is to render $\phi_t = \varepsilon_t / g_t^{1/2}$ weakly stationary in situations where the return series appears nonstationary. A standard GARCH model (2) fitted to the data would in that case be inadequate. From (1) it is seen that ϕ_t would then follow a GARCH process. There exist many variants of (2), but in this article the standard GARCH structure with $p = q = 1$ or its asymmetric counterpart, the GJR-GARCH by Glosten, Jagannathan and Runkle (1993), is mostly sufficient for our purposes. When $p = q = 1$, the subscript will be omitted from α_1, β_1 and κ_1 (the coefficient of the asymmetry term in the GJR-GARCH model).

3 Models with a deterministic long-run component

3.1 Nonparametric deterministic component

One of the first examples of the use of a deterministic component in modelling returns is the time-modulated (tm) process proposed by van Bellegem and von Sachs (2004) and further discussed in van Bellegem (2012). The observable return (the conditional mean is abstracted away) equals $\varepsilon_t = \zeta_t g(t/T)$, where $g(t/T)$ is deterministic positive-valued function of rescaled time and T is the number of observations. The error term ζ_t is assumed to be either white noise with zero mean and unit variance, an ARMA process, or a GARCH process. In the latter case, $\zeta_t = z_t h_t^{1/2}$, where $z_t \sim \text{iid}(0, 1)$, so that ε_t is defined as in (1) and h_t by (2). When $h_t \equiv 1$, the white noise assumption is sufficient for $\mathsf{E}\varepsilon_t \varepsilon_{t-j} = 0$ for $j \neq 0$, which is a property for many financial time series. The lack of a GARCH component is (partly) compensated by the assumption that the error ζ_t in $\varepsilon_t = \zeta_t g(t/T)$ has a leptokurtic density.

As already seen, in the tm-process time is rescaled between zero and one. It may be assumed that $g(t/T)$ is a smooth function of its argument. More specifically, $g(t/T)$ is assumed Lipschitz continuous: $|g(r_1) - g(r_0)| < C|r_1 - r_0|$ for all $0 < r_0, r_1 < 1$. The reason for rescaling time is that $g(t/T)$ is estimated nonparametrically. The observations ε_t, $t = 1, ..., T$, are squared, and the sequence $\{\varepsilon_t^2\}$ is smoothed using kernel estimation. The smoothed values are the estimated values of $g(t/T)$. Using the terminology of Dahlhaus (1997), the tm-variance process is locally stationary if z_t is stationary.

van Bellegem and von Sachs (2004) also consider the situation in which the process contains breaks but is piecewise Lipschitz continuous. The variance for the continuous segments is estimated by kernel estimation, but finding the break-points then becomes an essential part of the modelling process. To choose the segments or break-points, the authors suggest using two tests for detecting breaks in the unconditional variance based on the cumulative sum of squares. The first one is the post-sample prediction test which depends on knowing the time point where the time series is split. The second one is the CUSUM test which does not require splitting the returns into two subsamples but instead controls changes in the unconditional variance at each time point.

When $\zeta_t = z_t h_t^{1/2}$ (the GARCH case), van Bellegem and von Sachs (2004) explain that the purpose of the deterministic component is to transform the potentially nonstationary sequence $\{\varepsilon_t\}$ into a weakly stationary one $\{\phi_t\}$ by the normalisation $\phi_t = \varepsilon_t / g^{1/2}(t/T)$. They point out that once this has been done, 'standard econometric techniques' can be used to build models for ϕ_t.

Feng (2004) also considers the multiplicative decomposition of variance (1) such that h_t is a weakly stationary GARCH(p,q) process. The deterministic component $g(t/T)$ is assumed at least twice continuously differentiable on $[0,1]$ and the errors are iid$\mathcal{N}(0,1)$. The ensuing model is called the Semiparametric GARCH (SEMIGARCH) model since $g(t/T)$ is estimated nonparametrically as in the tm-model of van Bellegem and von Sachs. Even here, the observations ε_t are squared, and the estimates of $\widehat{g}(t/T)$, $t = 1, ..., T$, are obtained as smoothed values of the sequence $\{\varepsilon_t^2\}$ by the Nadaraya-Watson kernel estimator. The GARCH model is fitted into the normalised observations $\widehat{\phi}_t = \varepsilon_t / \widehat{g}^{1/2}(t/T)$.

Feng (2004) derives the bias and variance of $\widehat{g}(t/T)$ (or a transformation of it). Both are functions of the squared bandwidth, and the asymptotic bias is the same as in the nonparametric regression of iid variables. The results also include the asymptotic distribution of $\widehat{g}(t/T)$. As expected, the rate of convergence is a function of the bandwidth.

To investigate consequences of two-step estimation, Feng uses the normal log-likelihood for ϕ_t when the model is GARCH(1,1), which he calls approximate because it is conditional on $\widehat{g}(t/T)$. Under regularity conditions and denoting the GARCH parameter vector by $\boldsymbol{\theta}$, its maximum likelihood estimator by $\widehat{\boldsymbol{\theta}}$ and the true parameter vector by $\boldsymbol{\theta}_0$, the following result emerges:

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\beta}_\theta - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_{\theta_0}^{-1})$$

where $\boldsymbol{\beta}_\theta$ is the asymptotic bias and $\boldsymbol{\Sigma}_{\theta_0}$ the expected Hessian evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The bias term is of the order $O(b^2 + (Tb)^{-1})$ and negligible when the bandwidth b in the estimation of $g(t/T)$ is sufficiently small, that is, $O(T^{-1/2}) < b < O(T^{-1/4})$. In that case one can confidently base the statistical inference on the standard asymptotic theory for the stationary GARCH(1,1) model.

Properties of the SEMIGARCH model are studied by simulation. The purpose of the experiments is twofold: study both the choice of bandwidth and the behaviour of the GARCH parameter estimates. One of the simulation experiments is highlighted here, the focus being on GARCH parameter estimates. The GARCH component equals

$$h_t = 0.15 + 0.1\varepsilon_{t-1}^2 + 0.75h_{t-1}$$

so the total variance when $g(t/T) = 1$ equals one. The true $g(t/T)$ is a linear combination of a linear trend, a cosine function, and a hyperbolic tangent function which is close to a logistic function. The linear trend means that the amplitude of the clusters in the data increases over time. The other two components add extra movements to these changes.

The average estimated GARCH equation based on 2000 realisations and ignoring $g(t/T)$ becomes

$$\widehat{h}_t = 0.0363 + 0.0540\varepsilon_{t-1}^2 + 0.9432\widehat{h}_{t-1}$$

which yields the total variance 12.96. As can be expected, $\widehat{\alpha} + \widehat{\beta} = 0.9972$ is very close to one. When ε_t is rescaled and the GARCH model fitted to $\widehat{\phi}_t$, the equation has the following average form

$$\widehat{h}_t = 0.2052 + 0.0937\widehat{\phi}_{t-1}^2 + 0.6965\widehat{h}_{t-1}.$$

The coefficient of h_{t-1} is slightly underestimated while the average estimate of the total variance of ϕ_t equals 0.978, which is not far off the mark. It seems that rescaling is very important and that it works quite well. It is expected to do so even when there are several amplitude changes in the data at irregular intervals.

There are two empirical examples in Feng (2004). The SEMIGARCH model is fitted to daily returns of the New York S&P 500 index and to the Frankfurt DAX 100 index for the period from 3 January 1994 to 23 August 2000. It may be mentioned that the aforementioned simulation experiment was fashioned after the observed behaviour of DAX 100. For S&P 500 returns, the estimated GARCH(1,1) model is

$$\widehat{h}_t = 5.684 \times 10^{-7} + 0.0674\varepsilon_{t-1}^2 + 0.9302\widehat{h}_{t-1} \tag{3}$$

so $\widehat{\alpha} + \widehat{\beta} = 0.9976$, and the estimated total variance equals 2.4×10^{-4}, which is unrealistically low. The SEMIGARCH model yields

$$\widehat{h}_t = 0.0649 + 0.0686\widehat{\phi}_{t-1}^2 + 0.8676\widehat{h}_{t-1}. \tag{4}$$

In (4), $\widehat{\alpha} + \widehat{\beta} = 0.9362$ and the total variance estimate is 1.018. In comparing (3) and (4) it is seen that $\widehat{\beta}$ decreases, whereas $\widehat{\alpha}$ does not change much. This is in fact typical for many applications in which rescaling is employed. The weight of the lagged conditional variance in the GARCH model diminishes when the scale change is properly modelled. This is important in forecasting, because the decay rate of the conditional variance in the GARCH(1,1) model equals $\alpha + \beta$. A rate close to one leads to forecasts in which volatility remains too high for too long when the starting-value, the most recent estimated conditional variance, is high. Results for DAX 100 are quite similar to the ones for S&P500: $\widehat{\alpha} + \widehat{\beta} = 0.9849$ for GARCH equals 0.9354 for SEMIGARCH. For details, see Feng (2004).

Zhang, Feng and Peitz (2017) suggest estimating the scale function from $|\varepsilon_t|^\lambda$ instead of ε_t^2. This is motivated by moment requirements on financial return series in choosing the bandwidth. The relationship between the resulting scale function $g_\lambda(t/T)$ and the one based on smoothing ε_t^2 is demonstrated. The resulting model is called the Box-Cox SEMIGARCH model. It may be mentioned that in the paper the rescaled observations are modelled using several GARCH models, including the Exponential GARCH model by Nelson (1991). The application is to daily S&P 500 and DAX index returns.

When high-frequency (intradaily) returns r_t are being modelled, one also has to consider diurnal variation patterns due to investors' average behaviour over the day. Feng and McNeil (2008) extend SEMIGARCH to this situation. The decomposition (1) is augmented by a periodic component $s_{k(t)}$, $k = 1, ..., K$, where K is the length of the period. For instance, if ε_t is a five-minute return, K is the number of five-minute returns included in the 'day', a subset of the time the exchange is open for trading. The periodic SEMIGARCH has the form

$$r_t = z_t V_0^{1/2} h_t^{1/2} g^{1/2}(t/T) s_{k(t)} \tag{5}$$

where $t = 1, ..., T$, $z_t \sim \text{iid}\mathcal{N}(0,1)$, $V_0^{1/2} > 0$, $(1/K)\sum_{k=1}^{K} s_k = 1$, and $\int_0^1 g(u)\mathrm{d}u = 1$. The positive constant V_0 is a consequence of variance targeting: $\mathsf{E}h_t = 1$ because the intercept is defined as

$$\alpha_0 = 1 - \sum_{j=1}^{q} \alpha_j - \sum_{j=1}^{p} \beta_j \tag{6}$$

in (2). Rescaling leads to $\phi_t = r_t/(V_0^{1/2} g^{1/2}(t/T) s_{k(t)})$ such that ϕ_t has a stationary GARCH representation with $\mathsf{E}\{\phi_t|\mathcal{F}_{t-1}\} = 1$, where $\mathcal{F}_{t-1}$ contains the conditioning information. Feng and McNeil (2008) discuss nonparametric estimation and asymptotic properties of the estimators. The authors also suggest a test of the null hypothesis $g(t/T) = 1$, whose empirical null distribution is obtained by simulation. The model is fitted to 20-minute returns of four German stocks from 28 November 1997 to 30

December 1999.

3.2 Deterministic splines

Engle and Rangel (2008) introduce another multiplicative decomposition which is based on exponential quadratic splines. The aim of the authors is to examine links between return volatility and macroeconomics. To do this they develop the spline-GARCH model 'to allow the high frequency financial data to be linked with the low-frequency macro data.' In this review these links are not considered, and the focus is instead on the model. The scaling function of the spline-GARCH model is defined as

$$g_t = c\exp[w_0 t + \sum_{i=1}^{k} w_i\{(t-t_i)_+\}^2 + \gamma x_t] \qquad (7)$$

where $(t-t_i)_+ > 0$ for $t-t_i > 0$ and zero otherwise, and x_t is a weakly exogenous random variable. The exponential form (7) is used to make sure that $g_t > 0$ for all t. Here the decomposition is classified as nonparametric because g_t is a spline function, but g_t also contains the parameter vector $\mathbf{w} = (c, w_0, w_1, ..., w_k, \gamma)'$. To facilitate estimation, the knots are assumed equidistant, and their number in the spline function, k, is determined by the data. This is done by estimating the spline-GARCH model with $1, 2, ..., K$ splines and choosing the final model as the one for which $k \in \{1, ..., K\}$ minimises BIC of Rissanen (1978) and Schwarz (1978). Selecting another model selection criterion such as AIC (Akaike, 1974) which is more generous than BIC in selecting knots may sometimes lead to substantially larger number of them, see, for example, Amado, Silvennoinen and Teräsvirta (2017).

Using model selection criteria may cause identification problems. Let the true $k = k_0$, but suppose that the largest k to be considered exceeds k_0. When this is the case, the search leads to estimating unidentified models. Also, since variance forecasts from rescaled GARCH models depend on the end-point of the deterministic component (and also on how this component is extrapolated), the choice of the model selection criterion may have a large effect on forecasts. Furthermore, the larger the number of knots and splines, the smaller the sum $\widehat{\alpha} + \widehat{\beta}$ in the GARCH(1,1) equation. This has an effect on persistence of a shock, which in turn affects forecasts for several periods ahead.

Unlike the SEMIGARCH model, there is no asymptotic theory available for the maximum likelihood or other estimators of the parameters of the spline-GARCH model. Obviously, the reported standard deviation estimates are based on the assumption that the estimators are consistent and asymptotically normal. Engle and Rangel (2008) fit the spline-GARCH model to daily return series for 48 stock indexes from stock exchanges around the world. Variance targeting is used in estimation of spline-GARCH models, which means setting $\alpha_0 = 1-\alpha-\beta$ in (2). The same appears not to be true for the estimated GARCH models because in some cases $\widehat{\alpha}+\widehat{\beta} > 1$. The spline-GARCH results show the same pattern as in the SEMIGARCH applications. First, the sum $\widehat{\alpha}+\widehat{\beta}$ from the spline-GARCH model is generally clearly lower than what is obtained with the GARCH model. Second, the estimate $\widehat{\beta}$ decreases, sometimes quite strongly, when one moves from GARCH to spline-GARCH, whereas changes in $\widehat{\alpha}$ remain relatively minor. The number of knots varies from one to 15. The extreme case is Russia with 14 knots, one knot for only 167 observations. The estimated models for this dataset do not contain macroeconomic or other weakly exogenous variables, i.e., $\gamma = 0$ in (7).

Brownlees and Gallo (2010) define g_t with a different spline function. They consider the Multiplicative Error Model (MEM) that is similar to GARCH but used for realised variances (a daily realised variance is denoted as RV_t; there are several definitions for it), but this function can also be used for GARCH specifications. A first-order MEM has the following form:

$$h_t = \alpha_0 + \alpha_1 RV_{t-1} + \beta_1 h_{t-1} \tag{8}$$

where $RV_{t-1} > 0$. Brownlees and Gallo (2010) include an asymmetry term as in the GJR-GARCH model, but for notational simplicity this extension is omitted here. The spline function is a modification of the so-called B-spline. In the exponential case

$$g_t = c\exp\{\sum_{i=1}^{k} w_i B_i(t)\} \tag{9}$$

where $B_i(t)$ consists of pieces of polynomials. If $B_i(t)$ is of order q, it means that it consists of $q+1$ polynomial pieces of degree q that join at inner knots. The total number of knots spanning $B_i(t)$ equals $q+2$, and outside the outer knots the spline equals zero. The sum $\sum_{i=1}^{k} w_i B_i(t)$ is a B-spline. For more properties and information, see Eilers and Marx (1996). One of the advantages of B-splines is that they are easy to compute.

In practice, B-splines may not be used as such. An approach recommended by Eilers and Marx (1996) and followed by Brownlees and Gallo (2010) is to first select a large number of (equidistant) knots and reduce the dimension of the problem imposing a roughness penalty (Good and Gaskins, 1971) on the log-likelihood. There are many ways of doing that, one of them being to assume the penalty to be a function of the jth differences of the adjacent spline coefficients w_i. If the log-likelihood for T observations is denoted as $L_T(\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Theta$ contains the parameters in (8) and (9), their estimates are obtained as

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta}\{L_T(\boldsymbol{\theta}) - \frac{\lambda}{2}\sum_{i=j+1}^{k}(\Delta^j w_i)^2\}. \tag{10}$$

The resulting splines are called penalised B-splines, or P-splines for short. Useful properties of P-splines are listed in Eilers and Marx (1996). The idea is that while a B-spline with a large number of knots is not very smooth, the penalty of type (10) smooths the spline. Brownlees and Gallo (2010) point out, among other things, that quadratic splines have poor numerical properties compared to P-splines and that choosing the knots using a model selection criterion is not an appealing procedure. As already suggested, it may lead to numerical difficulties if a quadratic spline with 'too many' knots is estimated. But then λ, the size of the penalty in the P-spline, is determined by a model selection criterion. Eilers and Marx (1996) prefer AIC; note, however, the way they define the penalty.

The purpose of Brownlees and Gallo (2010) is to forecast the Value at Risk, and they are also interested in the performance of various estimators of realised variance. Since the main interest in this review lies in how well rescaling handles nonstationarity in the original return or realised variance series, their empirical results are bypassed here.

3.3 Flexible Fourier Form

Parameterising the scaling function using the Flexible Fourier Form (FFF) by Gallant (1981, 1984) constitutes an alternative to splines. Mazur and Pipień (2012) introduce a model consisting of (1) and (2) in which ε_t is replaced by $\phi_t = \varepsilon_t / g^{1/2}(t/T)$ such that

$$g(t/T) = \exp\{\sum_{i=1}^{k}\{w_i^c \cos(\frac{2\pi i}{T}t) + w_i^s \sin(\frac{2\pi i}{T}t)\} \tag{11}$$

where w_i^c and w_i^s are parameters, k is in practice small and the terms in the exponent of $g(t/T)$ are the lowest frequencies of the Fourier decomposition of the unknown scaling function. This model is called an almost periodically correlated (APC-) GARCH model. The inference is Bayesian, and the authors consider both normal and t-distributed errors for z_t. The application is to the daily returns of the S&P500 returns from 18 January 1950 to 7 February 2012. First-order APC-GARCH models with $k = 0, 1, 2, 3, 4$ are fitted to the data. The results show that standard GARCH ($k = 0$) fits clearly less well than the APC-GARCH models, and this outcome does not depend on the error process. Interestingly, but not unexpectedly, when the error are assumed normal estimated posterior probabilities are largest for models with $k \geq 2$, whereas $k = 1$ or 2 is favoured for models with t-distributed errors. This shows how an error distribution with thicker tails than the standard normal is able to absorb some of the movements in the series.

A look at the GARCH coefficients shows that the sum $\widehat{\alpha} + \widehat{\beta}$ or the corresponding sum for the GJR-GARCH model hardly changes when one moves from GARCH to APC-

GARCH. This is different from both SEMIGARCH and spline-GARCH and somewhat surprising as the purpose of rescaling is to handle long-run movements so that the GARCH coefficients would only reflect short-run movements (volatility clustering) in the data. The joint prior distribution for the GARCH parameters α and β is uniform $[0,1]^2$, which means that nonstationarity is not excluded even under scaling.

Multiplicative decomposition with FFF as the scaling function is also used in modelling realised variance. The purpose of the deterministic component is to describe the diurnal variation in high-frequency returns as in Feng and McNeil (2008). Andersen and Bollerslev (1998) apply this idea as follows. Let $R_{t,n}$ be an intradaily, in their case a 5-minute, return and $\mathsf{E}R_{t,n}$ the expected return, and set $r_{t,n} = R_{t,n} - \overline{R}$, where $\mathsf{E}R_{t,n}$ is approximated by the sample mean $\overline{R}$. The decomposition is

$$r_{t,n} = z_{t,n} h_{t,n}^{1/2} g_{t,n}^{1/2} \tag{12}$$

where $z_{t,n} \sim \text{iid}(0,1)$, $h_{t,n}$ has a GARCH or stochastic volatility structure, and $g_{t,n}$ represents diurnal variation and intradaily announcement ('news') effects. This is different from the approach of Feng and McNeil (2008) as the focus is solely on the periodic variation. The authors square (12) and take logarithms, so

$$x_{t,n} = \ln r_{t,n}^2 - \ln h_{t,n} = \mathsf{E}\ln z_{t,n}^2 + \ln g_{t,n} + u_{t,n}$$

where the iid error term $u_{t,n} = \ln z_{t,n}^2 - \mathsf{E}\ln z_{t,n}^2$. This is made operational by estimating $h_{t,n}$ from the data and assuming that while $\ln g_{t,n}$ is stochastic and mean-stationary, $\mathsf{E}\ln g_{t,n}$ has an FFF augmented by M news dummies $I_m(t,n)$:

$$\mathsf{E}\ln g_{t,n} = c_0 + \sum_{m=1}^{M} \lambda_m I_m(t,n) + \sum_{i=1}^{k} \{w_i^c \cos(\frac{2\pi i}{N} n) + w_i^s \sin(\frac{2\pi i}{N} n)\} \tag{13}$$

where N is the number of intervals within a day. Apart from the news dummies, (11) is similar to (13).

Assume now that $\widehat{h}_{t,n}$ is an appropriate estimator of $h_{t,n}$ and write

$$x_{t,n} = \ln r_{t,n}^2 - \ln \widehat{h}_{t,n} = \widehat{c} + \mathsf{E}\ln g_{t,n} + u_{t,n}^* \tag{14}$$

where $\widehat{c}$ is an intercept and $u_{t,n}^*$ a stationary error ($u_{t,n}$ is iid). The model is fitted to 5-minute Deutsche Mark-US Dollar logarithmic bid-ask spot price quotes from 1 October 1992 to 29 September 1993. For details of (14) and empirical results, see Andersen and Bollerslev (1998).

It may be mentioned that multiplicative decomposition is also used in modelling diurnal variation in duration series when the idea is to model the dynamic behaviour of the length of the interval between adjacent trades (duration). Since in this review is about volatility models and modelling, multiplicative decomposition of durations in models of autoregressive conditional duration, see Engle and Russell (1998), is not discussed here.

3.4 Parametric deterministic component

The models in which g_t is defined using splines or FFF already contain parameters, although they are in this review classified as nonparametric ones. In this section the focus is on models in which the decomposition is as in (1) but time is rescaled, and the smooth long-run component $g_t = g(t/T)$ is now fully parametric. The basic model of this kind is called the Multiplicative Time-Varying GARCH (MTV-GARCH) model. It is introduced in Amado and Teräsvirta (2008); see also Amado and Teräsvirta (2013, 2017). The purpose of the MTV-GARCH model is the same as before: jointly describe the short- and long-run movements in nonstationary return series. In (1), $\{z_t\} \sim \mathrm{iid}(0,1)$ with $\mathrm{E}z_t^3 = 0$ and $\mathrm{E}|z_t^2|^{2+\phi} < \infty$, $\phi > 0$. The conditional variance component is a GARCH process, although the GJR-GARCH is also used, in particular when modelling stock returns and indexes. The positive-valued long-run component $g_t = g(t/T)$ is defined as follows:

$$g(t/T) = 1 + \sum_{l=1}^{r} \delta_l G_l(t/T; \gamma_l, \mathbf{c}_l) \tag{15}$$

where $\delta_l, l = 1, \ldots, r$, are parameters and $G_l(t/T; \gamma_l, \mathbf{c}_l)$ is the generalised logistic transition function:

$$G_l(t/T; \gamma_l, \mathbf{c}_l) = (1 + \exp\{-\gamma_l \prod_{j=1}^{k_l} (t/T - c_{lj})\}^{-1} \tag{16}$$

with $\gamma_l > 0$ and $c_{l1} \leq c_{l2} \leq \ldots \leq c_{lk}$. Positivity imposes restrictions on δ_l, $l = 1, ..., r$. The intercept in (15) is set to equal one for identification reasons: the multiplicative decomposition can only have one free intercept. When variance targeting is not used, the intercept in (15) has to be fixed to a known positive value. In parameter estimation some choices are for numerical reasons better than some others. Alternatively, one can use variance targeting, see (6), and have a free intercept in (15).

When $k = 1$, (16) is a monotonic function of rescaled time, whereas it is nonmonotonic and symmetric around $(c_{l1} + c_{l2})/2$. When $k = 1$, the parameter γ_l controls the slope of the transition function, i.e., the speed of the transition. When $k = 1$ and $\gamma_l \to \infty$, (16) becomes a step function. Breaks in returns may not be straightforward to characterise by splines or FFF but are not difficult to model in the MTV-GARCH framework. In applications, almost invariably $k = 1$ or $k = 2$. Depending on the number of transitions, $g(t/T)$ can be a very flexible function of its argument.

In model specification there are issues similar to choosing the number of knots in spline-GARCH. If the number of transitions in (15) is too large, $g(t/T)$ is not identified. In order to avoid estimating unidentified models, r is determined by sequential testing. Constancy of (15) is tested first. If it is rejected, the TV-GARCH model is estimated and tested against a model with two transitions. Testing and estimation continues until the first non-rejection. The identification problem is circumvented by approximating the alternative model following Luukkonen, Saikkonen and Teräsvirta (1988). Details of the specification technique can be found in Amado and Teräsvirta (2017) and Amado et al. (2017).

In the SEMIGARCH model, the length of the series does not seem to affect modelling in any way. This is not quite the case in the MTV-GARCH framework. If the time series under consideration is very long and the number of transitions potentially large, the specification strategy outlined in Amado and Teräsvirta (2017) cannot be expected to work well. The solution is to split the series into subseries and identify the transitions in them before parameters of the model for the whole series are estimated. See Amado and Teräsvirta (2014b) and Amado et al. (2017) for examples of how this can be successfully done.

Maximum likelihood estimates of the parameters are obtained by dividing the maximisation problem into two parts; see Song, Fan and Kalbfleisch (2005). Amado and Teräsvirta (2013) show that under regularity conditions and using the results in Song et al. (2005), maximum likelihood estimators of the parameters of the MTV-GARCH (or MTV-GJR-GARCH) model are consistent and asymptotically normal. After the model has been estimated, its adequacy is examined using misspecification tests; see Amado and Teräsvirta (2017).

The MTV-GARCH model or its GJR version has been applied to several daily stock and stock index returns, exchange rate returns and commodity price returns. As an example, Amado and Teräsvirta (2017) describe the US Dollar/Singapore Dollar exchange rate returns using the model. The well-specified model has one transition with $k = 2$ in (16). The standard GARCH(1,1) model fitted to the series has the following form:

$$\widehat{h}_t = \underset{(0.001)}{0.001} + \underset{(0.016)}{0.056}\varepsilon_{t-1}^2 + \underset{(0.019)}{0.938}\widehat{h}_{t-1}$$

so $\widehat{\alpha} + \widehat{\beta} = 0.994$. The figures in parentheses are estimated standard errors. The corresponding component from the MTV-GARCH model is

$$\widehat{h}_t = \underset{(0.005)}{0.011} + \underset{(0.020)}{0.065}\widehat{\phi}_{t-1}^2 + \underset{(0.042)}{0.868}\widehat{h}_{t-1}$$

where $\widehat{\alpha} + \widehat{\beta} = 0.933$. As expected from Feng (2004), $\widehat{\beta}$ has changed much more than $\widehat{\alpha}$. The total variance estimates are 0.167 for GARCH and 0.164 for MTV-GARCH.

The MTV-GARCH model may be modified by assuming that the argument of $g(\cdot)$ is a random variable. This is discussed in Section 4.2. Other models with a random long-run component are considered in Section 4.

4 Stochastic multiplicative decomposition

4.1 Nonparametric stochastic component

As discussed in Section 3.1, Feng (2004) constructs a semiparametric GARCH model with a deterministic nonparametric long-run component. He shows how this helps remove the 'IGARCH effect', that is, $\widehat{\alpha} + \widehat{\beta} \approx 1$, often found in the applications of the first-order GARCH. Han and Kristensen (2017) develop another semiparametric GARCH model which they call Semiparametric Multiplicative GARCH-X, or SEMIX for short. It differs from the SEMIGARCH model in that in SEMIX the long-run component is stochastic. The stated goal of the authors is the same as that of Feng (2004): to develop a model in which the long-run component alleviates the IGARCH effect. The decomposition employed by Han and Kristensen (2017) can be written as $\varepsilon_t = z_t h_t^{1/2} g^{1/2}(x_{t-1})$, where the random variable x_t is quite persistent and strongly exogenous. It is defined as

$$x_t = (1 - \frac{c}{T})x_{t-1} + v_t \tag{17}$$

where $c \geq 0$. The error term v_t has mean zero and is independent of the GARCH error term z_t.

Under the assumption that $\{v_t\}$ is an iid sequence and some moment conditions it is found that the autocorrelation function of ε_t^2 of the SEMIX model converges to a positive random variable. In other words, the autocorrelation function then displays the 'long memory property' in that the decay of the autocorrelations as a function of the lag is slower than exponential. Interestingly, Mikosch and Stărică (2004) have shown that if there is a break in the unconditional variance of the GARCH model, ignoring it when computing the autocorrelations of ε_t^2 the sequence of autocorrelations also converges to a positive value when the lag length approaches infinity. A break can be viewed as a nonsmooth version of change in the long-run component.

The asymptotic theory becomes nonstandard when x_t is nonstationary as in (17). For details, some of them still open at this moment, the reader is referred to Han and Kristensen (2017). For the case where x_t is stationary, the limiting distribution of the (quasi) maximum likelihood estimator of the GARCH parameter vector is shown to be mixed normal.

Han and Kristensen (2017) fit the SEMIX model to three European daily index return series. The observation period stretches from 2 January 2004 to 30 December 2013. The indexes are FTSE (London), CAC (Paris) and DAX (Frankfurt). The random variable is the Chicago Board Options Exchange volatility index VIX. It represents implied volatility calculated from the options of the S&P 500 index. Given that the US stock market is likely to influence European markets, VIX may be used as the random variable for SEMIX. Furthermore, the index is close to being nonstationary and its dynamic behaviour agrees with the definition (17). The GARCH model of the authors is GJR-GARCH(1,1), and in all three cases the sum $\widehat{\alpha} + \widehat{\kappa}/2 + \widehat{\beta}$ is slightly below but very close to one. The corresponding sums for GJR-SEMIX are 0.931 (FTSE), 0.937 (CAC) and 0.947 (DAX), so the model works as intended. These results are quite similar to those of Feng (2004) who applies his SEMIGARCH model to S&P 500 and DAX returns, as discussed Section 3.1.

Han and Kristensen (2017) also consider the case where $g(\mathbf{x}'_{t-1}\boldsymbol{\delta})$, that is, the argument of $g(\cdot)$ is a linear combination of more than one random variable. In addition, $||\boldsymbol{\delta}|| = 1$, which makes it possible to compare coefficient estimates with each other. A GJR-SEMIX model with three variables, VIX, the country's industrial production index, and the price of crude oil is fitted to the three return series. The industrial production index is interpolated from a monthly to the daily level. The results show that VIX dominates the linear combination, and the oil price appears the least important variable as it has the smallest (in absolute value) coefficient estimate.

4.2 Time-varying ambiguity GARCH

Amado and Laakkonen (2013) introduce another model with a stochastic decomposition called the Time-Varying Ambiguity (TVA-)GARCH model. The word 'ambiguity' derives from Knight (1921) and is motivated by the application; for a detailed explanation see Amado and Laakkonen (2013). The decomposition is the same as in Han and Kristensen (2017), but the model is a stochastic variant of the TV-GARCH model studied in Section 3.4. It is obtained simply by replacing t/T in (15) and (16) by an exogenous random variable x_{t-1}. The authors discuss the modelling strategy consisting of specification, estimation and evaluation of the TVA model. At the specification stage GARCH is tested against TVA-GARCH. Performing the test described in the paper requires x_t to be weakly stationary and possess a sufficient amount of higher moments. As is the case with the TV-GARCH model, maximum likelihood estimation of parameters in TVA is carried out by estimating the parameters jointly but splitting each iteration into two parts. Note, however, that the asymptotic theory for maximum likelihood estimators derived for TV-GARCH has not yet been extended to TVA-GARCH. The estimated model is evaluated by misspecification tests.

The TVA model is applied to three daily bond return series. The bonds are US, German and French 10-year bonds, and the time series cover the period from 3 January 2000 to 30 December 2011. VIX and some of its transformations (ΔVIX_t, $\Delta|VIX_t|$ and $(\Delta VIX_t)^2$ are used as the stochastic variable in the model. The null hypothesis of GARCH is rejected for all transformations, and the strongest rejections (the p-values being really minimal) are obtained for the last two of them. The TVA model with $x_t = \Delta|VIX_t|$ is fitted to the three series. The estimated transition functions of type (16) are quite smooth, so the effect of the transformed VIX on volatility is gradual. Compared to Han and Kristensen (2017), an interesting observation is that the sum of the GARCH parameters remains very close to one even after accounting for the effect of VIX. A possible reason for this is there may be other factors which influence the conditional variance and which Amado and Laakkonen (2013) have not been able to consider because their model only allows a single stochastic variable in $g(x_{t-1})$.

4.3 Stochastic splines

Audrino and Bühlmann (2009) consider a model with stochastic B-splines. In fact, their model can be more general than a GARCH model, but the version discussed in the paper has the GARCH(1,1) model as the starting-point of iterations. The general task is to fit a nonparametric model of conditional heteroskedasticity to the data. The authors also include a conditional mean in their model, but for simplicity it is ignored here. This means that ε_t is observable, has mean zero, and $\varepsilon_t = z_t\sigma_t$, where $z_t \sim \text{iid}\mathcal{N}(0,1)$. The conditional variance σ_t^2 has the following form:

$$\sigma_t^2 = \mathsf{E}\{\varepsilon_t^2|\mathcal{F}_{t-1}\} = f(\varepsilon_{t-1}^2, \sigma_{t-1}^2)$$

where the unknown $f(\varepsilon_{t-1}^2, \sigma_{t-1}^2)$ can be quite general and does not have to be smooth. To indicate that σ_t^2 contains parameters, denote $\sigma_t^2 = \sigma_t^2(\boldsymbol{\theta})$. The model equals

$$\begin{aligned} \ln \sigma_t^2(\boldsymbol{\theta}) &= \ln f(\varepsilon_{t-1}^2, \sigma_{t-1}^2) = u_0(\varepsilon_{t-1}^2, \sigma_{t-1}^2) \\ &\quad + \sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} \beta_{k_1,k_2} B_{k_1,k_2}(\varepsilon_{t-1}^2, \sigma_{t-1}^2) \end{aligned} \tag{18}$$

where $B_{k_1,k_2}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)$ is a two-dimensional B-spline, and β_{k_1,k_2} its weight. The spline function $B_{k_1,k_2}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)$ is defined as follows:

$$B_{k_1,k_2}(\varepsilon_{t-1}^2, \sigma_{t-1}^2) = \sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} B_{k_1}(\varepsilon_{t-1}^2) B_{k_2}(\sigma_{t-1}^2). \tag{19}$$

Furthermore,

$$u_0(\varepsilon_{t-1}^2, \sigma_{t-1}^2) = \ln(\alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}) = \ln h_t \tag{20}$$

that is, the logarithm of the (weakly stationary) GARCH(1,1) process, and $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \beta_{k_1,k_2}, k_1 = 1, ..., K_1; k_2 = 1, ..., K_2)$. It is assumed that the polynomials in $B_{k_1}(\varepsilon_{t-1}^2)$ are quadratic, and in $B_{k_2}(\sigma_{t-1}^2)$ they are linear. For illustration, graphs of B-splines of these dimensions can be found in Eilers and Marx (1996, p. 91). Before taking logarithms, (18) is a multiplicative decomposition (1) with h_t defined in (20) and the positive-valued g_t equalling

$$g_t = \exp\{\sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} \beta_{k_1,k_2} B_{k_1,k_2}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)\}.$$

The knots in (19) are determined as quantiles of the arguments ε_{t-1}^2 and σ_{t-1}^2.

The final model is obtained by iteratively updating (18). After the mth iteration,

$$u_{mt}(\varepsilon_{t-1}^2, \sigma_{t-1}^2) = u_{m,t-1}(\varepsilon_{t-1}^2, \sigma_{t-1}^2) + \beta_m B_m(\varepsilon_{t-1}^2, \exp\{u_{m,t-1}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)\})$$

where β_m is a scalar weight function (note the change of notation from (k_1, k_2) to m), and $B_m(\varepsilon_{t-1}^2, \exp\{u_{m,t-1}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)\})$ is the spline determined for this iteration. For space reasons is not possible to describe details of how the spline is constructed and the weight β_m obtained. They can be found in the paper. Assuming the estimation is terminated after M iterations, the final model can be expressed as a function of the starting-value as follows:

$$u_{Mt}(\varepsilon_{t-1}^2, \sigma_{t-1}^2) = u_0(\varepsilon_{t-1}^2, \sigma_{t-1}^2) + \sum_{m=1}^{M} \beta_m B_m(\varepsilon_{t-1}^2, \exp\{u_{m-1,t-1}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)\}).$$

The stopping rule is determined by out-of-sample considerations. The time series is divided into two parts. The first 70% of observations are used to estimate the model and the remaining 30% are for determining M. The fit of the estimated model in the out-of-sample part determines M. In the multiplicative form (1), the final model has h_t defined as GARCH(1,1), and

$$g_t = \exp\{\sum_{m=1}^{M} \beta_m B_m(\varepsilon_{t-1}^2, \exp\{u_{m-1,t-1}(\varepsilon_{t-1}^2, \sigma_{t-1}^2)\}).$$

Audrino and Bühlmann (2009) remark that in practice it would be desirable to shrink β_m towards zero for every iteration. This would mean using $\beta_m(\kappa) = \kappa\beta_m$, $0 < \kappa < 1$, instead of β_m. The authors report that $0.1 < \kappa < 0.2$ works well in practice.

The paper contains two empirical examples. The first series is the daily annualised log-return series of the S&P 500 index and the second is the similarly defined returns of the 30-year US Treasury bill. Both series run from January 1990 to October 2003, 3376 observations in total. The first 2212 observations are used for estimation and the remaining 1164 for out-of-sample forecasting. The performance of the spline-GARCH model is compared with that of GARCH(1,1) and an earlier nonparametric model the authors have developed (Audrino and Bühlmann, 2003). In order to make comparisons possible, the true volatility is proxied by realised volatility. The results indicate that the spline-GARCH model performs better in terms of the mean square errors and mean absolute deviations both in-sample and out of sample than its two competitors. The Diebold and Mariano (1995) test shows that many but not all of the out-of-sample improvements are significant at the 0.1 level.

4.4 GARCH-MIDAS

The SEMIX model in Section 4.1 and the TVA-GARCH model in Section 4.2 are designed to allow random variables in the long-run component of the model. Engle et al. (2013) introduce another way of incorporating macroeconomic variables into GARCH equations. Their aim is to find out how well these variables explain stock market volatility. The model they develop can be seen as another variant of multiplicative decomposition such that the 'long-run component' is stochastic. To emphasise the different time scales (after abstracting away the conditional mean) the decomposition of the return ε_t is written, analogously to (1), as $\varepsilon_{i,t} = z_{i,t} h_{i,t}^{1/2} g_t^{1/2}$. Now i is the short scale (for example, $\varepsilon_{i,t}$ is a daily return), and t is the long or aggregated scale (the unit is one month or one quarter). The short-run component $h_{i,t}$ is measured in days, whereas the long-run g_t is available on a monthly or maybe a quarterly basis. The short-run component (conditional variance) is defined as before:

$$h_{i,t} = (1 - \alpha - \beta) + \alpha \frac{\varepsilon_{i-1,t}^2}{g_t} + \beta h_{i-1,t} \tag{21}$$

where $\alpha + \beta < 1$. This implies that $\mathsf{E} h_{i,t} = 1$. As already discussed, fixing this expectation (to one) is necessary for identification reasons, but in fact any *known* positive value would do.

The first example of a model based on this decomposition is one in which g_t is defined as follows:

$$g_t = m + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) RV_{t-k}. \tag{22}$$

The realized variance RV_t is the sum of daily squared returns,

$$RV_t = \sum_{i=1}^{M_t} \varepsilon_{i,t}^2 \tag{23}$$

In (23), M_t is the number of days in the month t of trading days. The purpose of the nonnegative function $\varphi_k(\omega_1, \omega_2)$ is to provide a parsimonious representation of a lag structure when K may be large. The beta-function

$$\varphi_k(\omega_1, \omega_2) = \frac{(\frac{k}{K})^{\omega_1 - 1}(1 - \frac{k}{K})^{\omega_2 - 1}}{\sum_{j=1}^{K}(\frac{j}{K})^{\omega_1 - 1}(1 - \frac{j}{K})^{\omega_2 - 1}} \tag{24}$$

is a popular choice. Engle et al. (2013) also consider the case in which the realized variance is defined by a rolling window:

$$RV_i = \sum_{j=1}^{M'} \varepsilon_{i-j}^2 \tag{25}$$

where the realized measure consists of M' terms from period i backwards in time.

The purpose of the lag structure is to smooth out erratic fluctuations in RV_t, meaning that (22) is preferred to $g_t = m' + \theta' RV_{t-1}$. Equations (21), (22) and (23) jointly with the definition of $\varepsilon_{i,t}$ form the GARCH-MIDAS-RV model.

If a (monthly) macroeconomic variable x_t replaces RV_t, to guarantee positivity g_t appears in the exponential form:

$$g_t = c \exp\{\theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) x_{t-k}\} \tag{26}$$

where $\mathsf{E}\varepsilon_t x_{t-k} = 0$, $k \geq 1$. The resulting model may be called the GARCH-MIDAS-X model. Engle et al. (2013) present (26) in the logarithmic form, but to stress similarity between (7) and (26) the exponential form is preferred here. Parameters of h_t and g_t are estimated jointly after the number of equidistant knots has been determined. Asymptotic properties of the maximum likelihood estimators of these parameters are not discussed in the paper. However, Wang and Ghysels (2015) prove consistency and asymptotic normality of maximum likelihood estimators of parameters in the GARCH-MIDAS-RV model in the rolling window (25) case.

In the application of Engle et al. (2013), the MIDAS function is exponential:

$$\varphi_k(\omega) = \frac{\omega^k}{\sum_{j=1}^{K} \omega^j}.$$

Furthermore, x_t is a quarterly series, and the variables are the growth rates of industrial production and the production price index, respectively. These are called 'level' series, and their error variance is named 'volatility'. The latter is computed by fitting an autoregressive model to the level series, taking the residuals, squaring them, and forming a sum of the squared residuals as in (23). One can of course consider a model in which only one of the four macro variables is present, but it is also possible to build more complicated models in which two or all four of them appear simultaneously. The authors also consider a spline-GARCH model in which g_t is defined by (7) but the longer time scale is retained.

The series to be modelled is long and consists of daily U.S. stock returns over the period from 16 February 1885 to 31 December 2010. The two macroeconomic series also start in 1884. They are originally monthly but are temporally aggregated to the quarterly level. In addition to the complete period, GARCH-MIDAS models are also fitted to subperiods. The parameter estimates from the logarithmic g_t equation are strongly significant. Estimated GARCH equations are not reported, so it is not possible to assess the effect of g_t on the estimated sum of the GARCH(1,1) parameters α and β. The results indicate that for GARCH-MIDAS-RV models this sum is generally well below one, the only notable expectation being the subperiod 1953–1984. For the GARCH-MIDAS-X models this sum is well below one for the subperiod 1890–1919 and close to one, generally greater than 0.98 for all models independent of the level/volatility and the variable or variables included in the model. Results of fitting spline-GARCH models show that this sum is slightly lower than in corresponding MIDAS models. One may conclude that including macro variables in the GARCH model does not affect the amplitude of the volatility clusters in the way it does when g_t is deterministic. This outcome is quite different from what Han and Kristensen (2017) obtain with their SEMIX model, but the economic variable (VIX) in SEMIX is different from the four variables used by Engle et al. (2013).

As to the contribution of the macro variables, the authors find that quite a significant fraction of variation in expected volatility can be ascribed to economic sources. Variables with the strongest contribution are not the same throughout the whole period, but in general terms they seem useful in explaining stock market volatility. It is not possible here to describe forecasting experiments Engle et al. (2013) conduct with GARCH-MIDAS-X models but they seem to support this conclusion.

The work of Engle et al. (2013) has generated plenty of interest and applications. Girardin and Joyeux (2013) employ the GARCH-MIDAS-RV model to estimate the long-run component g_t for daily returns of indexes of A and B stocks in the Shanghai Stock Exchange. They then study connections between g_t and economic fundamentals as in Engle and Rangel (2008). The long-run time scale is monthly. Ashgarian, Hou and Javed (2013) combine the realized variance, the 'level' and the 'volatility' into a single model. A noteworthy detail is that they do not use the exponential form but rather extend the RV specification (22). Their long-run component has the form

$$\begin{aligned} g_t &= m + \theta_{RV} \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) RV_{t-k} + \theta_L \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) x_{L,t-k} \\ &\quad + \theta_V \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) x_{V,t-k} \end{aligned} \tag{27}$$

where $x_{L,t}$ is a 'level' economic variable and $x_{V,t}$ is the corresponding 'volatility' variable. Principal components of economic variables are also used as an economic variable. This approach requires the level variable to remain positive during the observation period. In the application the long time scale t is monthly. Unlike Engle et al. (2013), these authors use squared first differences of economic variables as the measure of their volatility.

The series to be modelled and forecast is the S&P 500 daily return series from January 1991 to June 2008. The out-of-sample forecasting period begins in January 2004. Results show that adding the first principal component as the economic variable to the GARCH-MIDAS-RV model as in (27) improves forecasting performance compared to that of the GARCH-MIDAS-RV model.

Conrad and Loch (2015) employ both one- and two-sided MIDAS filters. The latter are employed to study lead and lag relationships between stock market volatility and macroeconomic variables and were already considered by Engle et al. (2013). The difference between these two is that the two-sided filter of Engle et al. (2013) contains future values that by definition are not available at $t-1$. Conrad and Loch (2015) replace these unknown observations by forecasts available at $t-1$, so their two-sided MIDAS filter can be used for forecasting. The returns are again continuously compounded daily returns of the S&P 500 index, and the observation period extends from 2 January 1969 to 30 December 2011. The long-run time scale is quarterly. Furthermore, the short-term GARCH component is a GJR-GARCH one. GJR-GARCH-MIDAS-X equations are estimated for 11 economic variables. The authors also fit a GJR-GARCH-MIDAS-RV model with a quarterly RV component to the data and finally, combine these two into a GJR-GARCH-MIDAS-RV-X model, in which, denoting the relevant macro variable by x_t,

$$g_t = \exp\{m + \theta_{RV} \sum_{k=1}^{K} \varphi_k(1, \omega_2^{RV}) RV_{t-k} + \theta_x \sum_{k=1}^{K} \varphi_k(\omega_1^x, \omega_2^x) x_{t-k}\}. \tag{28}$$

The exponential form is used to make sure that g_t remains positive. Forecasting with GARCH-MIDAS is discussed, and forecasting with models with two-sided filters receives special attention. The reader is referred to the paper for a rich set of results concerning both fitted one-and two-sided filter models and forecasts from them.

4.5 Misspecification testing

When a volatility model such as a GARCH model (2) has been estimated, its adequacy should be tested before using it for forecasting or other purposes. Multiplicative decomposition may be used for constructing alternative hypotheses to the estimated specification. As an example, Lundbergh and Teräsvirta (2002) derive a test statistic for testing the adequacy of the GARCH model. Under H_0, $\varepsilon_t = z_t h_t^{1/2}$ where $z_t \sim \mathrm{iid}\mathcal{N}(0,1)$ and h_t follows a GARCH process, whereas under the alternative when the alternative $z_t = \zeta_t g_t^{1/2}$, where $\zeta_t \sim \mathrm{iid}\mathcal{N}(0,1)$ and

$$g_t = 1 + \sum_{j=1}^{r} \delta_j z_{t-j}^2. \tag{29}$$

The null hypothesis is $\delta_i = 0$, $i = 1, ..., r$. The idea is that under the alternative there is 'ARCH nested in GARCH'. Under regularity conditions, the resulting statistic follows a χ^2-distribution with r degrees of freedom.

As already indicated, another variant of (29) is one in which GARCH is tested against (15) with (16). The testing situation is, however, nonstandard because the alternative model is not identified when the null hypothesis is valid. More discussion about deriving a test statistic in that situation can be found in Amado and Laakkonen (2013) and Amado and Teräsvirta (2017).

Conrad and Schienle (2017) construct a Lagrange multiplier misspecification test for testing GARCH against GARCH-MIDAS, see (26). The null hypothesis is that the long-run component is constant. This implies $\theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) = 0$. The authors show that their test statistic has an asymptotic χ^2-distribution with K degrees of freedom. They discuss the choice of K and find in simulations that $K = 1$ often already suffices to reject the null hypothesis when the alternative holds.

The multiplicative decomposition may also be used for improving a misspecified model or, in other words, dividing estimation of the true (unknown) model into two parts. Following Mishra, Su and Ullah (2010), one first fits a GARCH model to the series (or residuals if a conditional mean has already been estimated) under consideration to parametrically remove some of the variation in ε_t^2. Then one makes use of the identity

$$\mathsf{E}\{\varepsilon_t^2 | \mathcal{F}_{t-1}\} = h_t \mathsf{E}\{\frac{\varepsilon_t^2}{h_t} | \mathcal{F}_{t-1}\} = h_t g_t.$$

If h_t is correctly specified, $\mathsf{E}\varepsilon_t^2 / h_t = 1$. If it is not, there may be structure left in g_t, and it is assumed that it can be estimated nonparametrically. This looks a bit like nonparametric 'ARCH nested in GARCH'. The resulting model is called the Semiparametric GARCH (SPGARCH) model. Mishra et al. (2010) assume that g_t is locally linear, describe the subsequent estimation problem in detail and under regularity conditions prove consistency and asymptotic normality of the resulting maximum likelihood estimator.

To demonstrate the modelling strategy, the authors use the S&P500 daily returns from 3 January 2002 through 3 January 2007, a total of 1,258 observations. They fit an ARCH(1), the GARCH(1,1) and the GJR-GARCH(1,1) model to this series. They report the ARCH or GARCH parameter estimates and the amount of variation explained by the corresponding parametric model. The results are surprising in that this ratio is largest in ARCH (88.2%) and lowest in the GJR-GARCH model (81.1%). The likely explanation is, however, that g_t is a function of the variables in h_t. Since GARCH and GJR-GARCH have one variable more than ARCH, this gives the latter model better possibilities to improve the fit than it does to the former two. Misspecification checks suggest that the SPARCH model is misspecified, whereas the two GARCH models seem adequate. The persistence for h_t in the GARCH model is high: $\widehat{\alpha}_1+\widehat{\beta}_1 = 0.982$, and the corresponding number for GJR-GARCH equals 0.983. It is very low for ARCH: $\widehat{\alpha}_1 = 0.27$. From the plots in the paper Mishra et al. (2010) conclude that the nonparametric component is higher for negative than for positive shocks, so it contributes to explaining the asymmetry in the series.

5 Multivariate models

5.1 Stochastic discount factor model

Multiplicative decomposition of variance can be generalised to decomposing the covariance matrix. Section 5 is in its entirety devoted to this generalisation. In this case, correlations between the return variables or functions of them come into play. To fix notation, let $\boldsymbol{\varepsilon}_t$ be an $N \times 1$ log-return vector, $\mathsf{E}\boldsymbol{\varepsilon}_t = \mathbf{0}$, and $\mathsf{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \boldsymbol{\Sigma}_t$. There are several models in which $\boldsymbol{\Sigma}_t$ is multiplicatively decomposed to short and long run components. To begin with, Osiewalski (2009) and Osiewalski and Pajor (2009) consider the following simple multiplicative decomposition:

$$\boldsymbol{\varepsilon}_t = \mathbf{z}_t g_t^{1/2} \tag{30}$$

where $\mathbf{z}_t \sim \mathrm{iid}\mathcal{N}(\mathbf{0}, \mathbf{C})$, and $\mathbf{C}$ is positive definite covariance matrix. Furthermore, g_t is a positive-valued stochastic random variable such that

$$\ln g_t = \phi \ln g_{t-1} + \sigma_g \eta_t \tag{31}$$

where $\sigma_g > 0$, $\eta_t \sim \text{iid}\mathcal{N}(0,1)$ and η_t and $\mathbf{z}_t$ are mutually independent. The ensuing model is called the Stochastic Discount Factor (SDF) model. The conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ equals

$$\boldsymbol{\Omega}_t = \mathsf{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\} = g_t\mathbf{C} = g_t[c_{ij}] \tag{32}$$

where $\mathbf{C}$ is a positive definite matrix. Since $\mathbf{C}$ is not a correlation matrix, it has to be transformed into one. This is done by defining $\mathbf{P} = (\mathbf{I}_N \odot \mathbf{C})^{-1/2}\mathbf{C}(\mathbf{I}_N \odot \mathbf{C})^{-1/2}$, where $\mathbf{P}$ is now the constant conditional correlation matrix. Time-variation in (32) is due to a scalar stochastic component g_t. It follows that the SDF model has two independent sources of noise, which makes estimation numerically demanding. The authors therefore adopt a Bayesian approach with priors on ϕ, $\mathbf{C}^{-1}$ and σ_g^{-2}. The model (even one with a VAR(1) conditional mean) can be analysed using the Gibbs sampler. Osiewalski (2009), however, points out that the SDF model is too simple for practical purposes because the time-variation in covariances is controlled by a single stochastic variable.

Consequently, Osiewalski (2009) and Osiewalski and Pajor (2009) study a generalisation of the SDF model. It consists of making $\mathbf{C}$ in (32) time-varying. The ensuing models are called Hybrid SDF-Scalar BEKK (SDF-SBEKK) models because the time-varying covariance matrix has a BEKK structure; for the BEKK-GARCH model, see Engle and Kroner (1995). The first-order scalar BEKK with covariance targeting has the following form

$$\mathbf{H}_t = (1-\alpha-\beta)\mathbf{C} + \alpha\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \beta\mathbf{H}_{t-1} \tag{33}$$

where α and β are positive scalars, $\alpha+\beta < 1$, $\mathbf{H}_t$ is a conditional covariance matrix, and $\mathbf{C}$ is a symmetric positive definite matrix. The hybrid model combines scalar BEKK and stochastic volatility. The multiplicative decomposition of $\boldsymbol{\varepsilon}_t$ has the following form:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Omega}_t^{1/2}\mathbf{z}_t g_t^{1/2} \tag{34}$$

where g_t is defined as in (31), $\boldsymbol{\Omega}_t^{1/2} \in \mathcal{F}_{t-1}$ is a time-varying positive definite matrix, $\eta_t \sim \text{iid}\mathcal{N}(0,1)$, $\mathbf{z}_t \sim \text{iid}\mathcal{N}(\mathbf{0},\mathbf{I}_N)$, and $\mathbf{z}_t$ and η_t are mutually independent. The decomposition (34) implies $\mathsf{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\} = g_t\boldsymbol{\Omega}_t$. Two hybrid models emerge. The first one, called the type I hybrid, has $\boldsymbol{\Omega}_t = \mathbf{H}_t$, which means that $\boldsymbol{\Omega}_t$ is defined by (33) and does not depend on g_t. In type II model it is assumed that $\boldsymbol{\varepsilon}_t$ is rescaled: $\boldsymbol{\phi}_t = \boldsymbol{\varepsilon}_t/g_t^{1/2}$, and that $\boldsymbol{\phi}_t$ has the scalar BEKK structure. This is analogous to the rescaling of ε_t in the univariate GARCH case in Section 2, where the rescaled returns are assumed to follow a GARCH process. In the type II model, $\boldsymbol{\Omega}_t = \mathbf{H}_t^*$, where

$$\mathbf{H}_t^* = (1-\alpha-\beta)\mathbf{C} + \alpha\boldsymbol{\phi}_{t-1}\boldsymbol{\phi}_{t-1}' + \beta\mathbf{H}_{t-1}^*. \tag{35}$$

When $N = 1$, the rescaled BEKK (35) collapses into the rescaled GARCH in (1), where now g_t is stochastic and follows a stochastic volatility process (31). Up to now, this 'hybrid SDF-GARCH' may not have been used in applied work.

The Bayesian analysis (estimation) of these two hybrid models by Gibbs sampling is described in detail in Osiewalski (2009). To alleviate numerical problems, $\mathbf{C}$ is estimated by the sample covariance matrix of $\boldsymbol{\phi}_t$. This is analogous to how the 'intercept matrix' is estimated in the DCC-GARCH model of Engle (2002). As discussed in Osiewalski and Pajor (2009), $\boldsymbol{\Omega}_t$ in (34) need not be of BEKK type. The DCC-GARCH is deemed as a promising alternative, although it would be computationally more complicated to handle than the relatively simple scalar BEKK parameterisation. More information about this can be found in Osiewalski and Pajor (2007).

Osiewalski and Pajor (2009) apply the SDF-GARCH and a few other models to two bivariate datasets, of which the first one consists of two daily exchange rate log-returns for the period 1 February 1996 - 31 December 2001 and the second one of daily log-returns of WIG, the Warsaw stock exchange index, and the S&P 500 from 8 January 1999 to 1 December 2006. The estimated models are ranked according to their Bayes factors. In both cases, the SDF-BEKK fares better than the alternatives that include, among other things, BEKK, scalar BEKK, and DCC, all with t-distributed errors. Interestingly, type I SDF-BEKK is ranked above the computationally more demanding type II. SDF-DCC is also ranked ahead of the models without the SDF component. It may be noted from Tables 5 and 6 in the paper that the persistence calculated from the posterior means is very high in both bivariate SDF-BEKK models, so the SDF extension to BEKK does not have much impact on the standard BEKK in this respect.

The authors also present a high-dimensional application in which the dataset consists of daily log-returns of 23 stocks from the mWIG40 index and another 11 from WIG20. The period runs from 30 January 2003 to 29 August 2007. (Amado et al. (2017) have recently modelled daily log-retuns of the latter index using the MTV-GARCH model discussed in Section 3.4.) In this case the persistence, estimated from both the type I and type II model, remains below 0.9. Whether or not the persistence generally decreases when the dimension of the return vector increases is an open (empirical) question, but further study of this possibility could be interesting. Besides, the estimate (posterior mean) of ϕ in (31) is also fairly small, about 0.5 for both the type I and type II model. Anyway, the analysis shows that is quite possible to describe large (or medium-sized) return vectors with the SDF-BEKK model.

5.2 Local dynamic conditional correlation model

Feng (2006) presents a multivariate GARCH model in which both the variance and the correlation component are deterministically time-varying. He calls the model the

Local Dynamic Conditional Correlation (LDCC-)GARCH model. It is a generalisation of the SEMIGARCH model considered in Section 3.1. In this review, the fact that Feng even estimates the conditional mean (nonparametrically) as a part of the modelling process is ignored, and only modelling and estimating the variance and correlations are considered. The return or error vector $\boldsymbol{\varepsilon}_t$ is multiplicatively decomposed as follows: $\boldsymbol{\varepsilon}_t = \mathbf{S}(t/T)\mathbf{D}_t\mathbf{z}_t$, where $\mathbf{D}_t = \text{diag}(h_{1t}^{1/2}, ..., h_{Nt}^{1/2})$ contains the short-run GARCH components and $\mathbf{S}(t/T) = \text{diag}(g_1^{1/2}(t/T), ..., g_N^{1/2}(t/T))$ the deterministic elements. Furthermore, $\mathbf{z}_t \sim \text{indep}(\mathbf{0}, \mathbf{P}(t/T))$, where $\mathbf{P}(t/T) = \mathsf{E}\mathbf{z}_t\mathbf{z}_t'$ contains the deterministically time-varying correlations. This implies the following decomposition for the correlation matrix $\boldsymbol{\Sigma}(t/T) = \mathsf{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\}$:

$$\boldsymbol{\Sigma}(t/T) = \mathbf{S}(t/T)\mathbf{D}_t\mathbf{P}(t/T)\mathbf{D}_t\mathbf{S}(t/T) \tag{36}$$

The covariance matrix (36) is estimated as follows. First estimate the deterministic components $g_i(t/T)$, $i = 1, ..., N$, nonparametrically as in Feng (2004). This yields the residuals $\widehat{\boldsymbol{\phi}}_t = \widehat{\mathbf{S}}^{-1}(t/T)\boldsymbol{\varepsilon}_t$. The elements of $\widehat{\boldsymbol{\phi}}_t$ are assumed to follow a GARCH process, so α_i and β_i are estimated from

$$h_{it} = (1 - \alpha_i - \beta_i) + \alpha_i\widehat{\phi}_{t-1}^2 + \beta_i h_{i,t-1}$$

$i = 1, ..., N$, where $\alpha_i + \beta_i < 1$ when the standard first-order GARCH process is used. For identification reasons, $\mathsf{E}h_{it} = 1$, $i = 1, ..., N$. This operation gives the residuals $\widehat{\mathbf{z}}_t = \mathbf{D}_t^{-1}\widehat{\boldsymbol{\phi}}_t$. The correlations will be functions of $\mathbf{z}_{t-j}$, $j = 1, ..., p$, where the lag length p is determined by the user. The correlation matrix is estimated nonparametrically. Curse of dimensionality when N is large is dealt with in the following way. Consider the random vector $\mathbf{y}_t = (y_{1t}, ..., y_{pt})'$, where $y_{jt} = (\mathbf{1}'\mathbf{z}_{t-j})^2$. Define two kernels, a univariate one, K_0, for t/T and a multivariate spheric kernel K for $\mathbf{y}_t$. The estimator for the correlation matrix becomes

$$\widehat{\mathbf{Q}}(\tau, \mathbf{y}) = \sum_{t=n_1}^{n_2} w_t\widehat{\mathbf{z}}_t\widehat{\mathbf{z}}_t' \tag{37}$$

where

$$w_t = \frac{K_0(\frac{t/T-\tau}{b_0})K(\frac{y_{1t}-y_1}{b}, ..., \frac{y_{pt}-y_p}{b})}{\sum_{t=n_1}^{n_2} K_0(\frac{t/T-\tau}{b_0})K(\frac{y_{1t}-y_1}{b}, ..., \frac{y_{pt}-y_p}{b})}.$$

The two bandwidths b_0 and b are different from each other because t/T and y_{jt} are of different magnitude. How to select the bandwidths and n_1 and n_2 in (37) is discussed in the paper. Theoretical results obtained for biases and variances of the elements of $\widehat{\mathbf{Q}}(\tau, \mathbf{y})$ are derived under the assumption that even a nonparametric mean is estimated. Note that $\mathbf{Q}(\tau, \mathbf{y})$ does not automatically become a correlation matrix, so the by now familiar adjustment

$$\widehat{\mathbf{P}}(\tau, \mathbf{y}) = (\mathbf{I}_N \odot \widehat{\mathbf{Q}}(\tau, \mathbf{y}))^{-1/2} \widehat{\mathbf{Q}}(\tau, \mathbf{y}) (\mathbf{I}_N \odot \widehat{\mathbf{Q}}(\tau, \mathbf{y}))^{-1/2} \tag{38}$$

is required. Results on optimal bandwidths are building on the same premises. As already mentioned, the mean has been ignored in this exposition.

Feng (2006) fits the LDCC-GARCH model to the daily foreign exchange rate series of the British Pound, Euro, Japanese Yen and Canadian Dollar vis-à-vis the US Dollar from 4 January 1999 to 30 December 2005. It may be noted that conditional variance of the Euro returns is modelled as a GARCH(2,2) process, whereas for the other three series GARCH(1,1) is deemed adequate. The six estimated correlations show a tendency to increase towards the end of the period. The deterministic component is clearly useful, and the short-run fluctuations around it are relatively minor in comparison. The short-run component in correlations is due to $\mathbf{D}(t/T)$. $\mathbf{Q}(\tau, \mathbf{y})$ itself does not contain a short-run component, but Feng (2006) mentions that having one could be worth investigating. His suggestion has been followed up later, see Section 5.7.

5.3 Local BEKK model

Hafner and Linton (2010) use the following multiplicative decomposition for $\boldsymbol{\varepsilon}_t$:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}^{1/2}(t/T) \mathbf{H}_t^{1/2} \mathbf{z}_t \tag{39}$$

where $\mathbf{z}_t$ is strictly stationary martingale difference sequence with $\mathsf{E}\{\mathbf{z}_t | \mathcal{F}_{t-1}\} = \mathbf{0}$ and $\mathsf{E}\{\mathbf{z}_t \mathbf{z}_t' | \mathcal{F}_{t-1}\} = \mathbf{I}_N$. The matrix $\boldsymbol{\Sigma}(t/T)$ is a nonparametric function of rescaled time, positive definite and at least twice continuously differentiable. The stochastic $\mathbf{H}_t \in \mathcal{F}_{t-1}$ is also a positive definite matrix but with a parametric representation. Decomposition (39) is another generalisation of SEMIGARCH to the multivariate case. Writing $\boldsymbol{\phi}_t = \boldsymbol{\Sigma}^{-1/2}(t/T)\boldsymbol{\varepsilon}_t$ one obtains $\boldsymbol{\phi}_t = \mathbf{H}_t^{1/2}\mathbf{z}_t$. As an example of $\mathbf{H}_t$ the authors use the 'full' first-order BEKK-GARCH:

$$\mathbf{H}_t = \mathbf{I}_N - \mathbf{A}\mathbf{A}' - \mathbf{B}\mathbf{B}' + \mathbf{A}\boldsymbol{\phi}_{t-1}\boldsymbol{\phi}_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}'$$

where $\mathsf{E}\mathbf{H}_t = \mathbf{I}_N$. This condition is needed for identification of the model. The model of Hafner and Linton (2010) could then, following Feng (2006), be called the Local-BEKK (LBEKK) model.

Estimation of LBEKK can be carried out in stages by first estimating $\boldsymbol{\Sigma}(t/T)$ non-parametrically from $\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}^{1/2}(t/T)\mathbf{z}_t$ using kernel estimation as in Feng (2004). How this is done when $\boldsymbol{\Sigma}(t/T)$ is a matrix is discussed in Rodríguez-Poo and Linton (2001).

Analogously to Feng (2006), one then forms rescaled returns $\widetilde{\boldsymbol{\phi}}_t = \widetilde{\boldsymbol{\Sigma}}^{-1/2}(t/T)\boldsymbol{\varepsilon}_t$, where $\widetilde{\boldsymbol{\Sigma}}(t/T)$ is a nonparametric estimate of $\boldsymbol{\Sigma}(t/T)$ and estimates the BEKK parameters from

$$\mathbf{H}_t^* = \mathbf{I}_N - \mathbf{A}\mathbf{A}' - \mathbf{B}\mathbf{B}' + \mathbf{A}\widetilde{\boldsymbol{\phi}}_{t-1}\widetilde{\boldsymbol{\phi}}_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}^*\mathbf{B}'. \tag{40}$$

For a scalar BEKK variant of this, see (35). Since $\widetilde{\boldsymbol{\Sigma}}(t/T)$ is not a function of any of the BEKK parameters, after assuming $\mathbf{z}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ maximum likelihood estimation of $\mathbf{A}$ and $\mathbf{B}$ in (40) proceeds as in the standard BEKK case. Hafner and Linton (2010) write that the estimator $\widetilde{\gamma}$ of $\gamma = (\text{vec}(\mathbf{A})', \text{vec}(\mathbf{B})')'$ is 'expected to be consistent and asymptotically normal but not efficient.' In the univariate case Feng (2004) found an asymptotic bias which, however, was negligible in large samples. Whether or not the situation is similar here is not known. Efficiency may nevertheless be improved by re-estimating $\boldsymbol{\Sigma}(t/T)$ from $\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}^{1/2}(t/T)\widetilde{\mathbf{u}}_t$, where $\widetilde{\mathbf{u}}_t = (\widetilde{\mathbf{H}}_t^*)^{1/2}\mathbf{z}_t$ and then re-estimating the BEKK parameters. Computational details of the estimation procedure can be found in Hafner and Linton (2010).

The LBEKK model is fitted to the bivariate series of daily Dow Jones and NASDAQ index returns decomposition from 2 January 1990 to 7 January 2009, a total of 4795 observations. The autocorrelation of returns is first removed by fitting a VAR(1) model to the returns. In order to model the residuals from this model when the returns are stock index returns the BEKK model is augmented by an asymmetry term analogous to that in the univariate GJR-GARCH model. The results of modelling the residuals show that persistence of the BEKK-GARCH component declines when the time-varying component $\boldsymbol{\Sigma}(t/T)$ is included in the model. They also strongly support inclusion of the asymmetry component. The time-varying unconditional correlation between the two return series generally exceeds 0.5 and lies close to unity at the end of the sample. As expected, the conditional correlations fluctuate around the unconditional 'trend'. The paper does not contain higher-dimensional examples, so comparisons with the work of Osiewalski and Pajor (2009) are not possible.

5.4 Multiplicative DCC model

In modelling dynamics of electricity futures, Bauwens, Hafner and Pierret (2013) apply the decomposition (36). They call their model a multiplicative DCC (mDCC) model. The difference between mDCC and LDCC is that the correlation matrix is estimated nonparametrically in LDCC and parametrically in mDCC. (Incidentally, 'DCC' means different things in these two acronyms.) Estimation of parameters proceeds as in Feng (2006) and Hafner and Linton (2010). First, estimate the deterministic structure $\boldsymbol{\Sigma}(t/T)$ nonparametrically from $\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}^{1/2}(t/T)\boldsymbol{\phi}_t$, assuming that the unconditional expectation $\mathsf{E}\boldsymbol{\phi}_t\boldsymbol{\phi}_t' = \mathbf{I}_N$. The vector $\boldsymbol{\phi}_t = \boldsymbol{\Sigma}^{-1/2}(t/T)\boldsymbol{\varepsilon}_t$ is now free of slowly moving variation in the variance of $\boldsymbol{\varepsilon}_t$ characterised by the deterministic component $\boldsymbol{\Sigma}(t/T)$. Next, estimate the conditional variances in $\mathbf{h}_t = (h_{1t}, ..., h_{Nt})'$, represented by the diagonal matrix $\mathbf{D}_t$ in (36). Bauwens et al. (2013) assume that they follow the GJR-GARCH model:

$$h_{it} = \alpha_{i0} + \alpha_{i1}\phi_{i,t-1}^2 + \kappa_i I(\phi_{t-1} < 0)\phi_{i,t-1}^2 + h_{i,t-1}$$

This yields the (estimated) error vector $\widehat{\mathbf{z}}_t = \mathbf{D}_t^{-1}\widehat{\boldsymbol{\phi}}_t$. Since $\mathsf{E}\boldsymbol{\phi}_t\boldsymbol{\phi}_t' = \mathbf{I}_N$, the unconditional variance $\mathsf{E}\mathbf{z}_t\mathbf{z}_t' = \mathbf{I}_N$ as well, while $\mathsf{E}\{\mathbf{z}_t\mathbf{z}_t'|\mathcal{F}_{t-1}\} = \mathbf{P}_t$, where $\mathbf{P}_t$ follows a DCC process. Estimating its parameters completes the estimation of the model. A noteworthy detail of the DCC equation

$$\mathbf{Q}_t = (1 - a - b)\mathbf{I}_N + a\widehat{\mathbf{z}}_{t-1}\widehat{\mathbf{z}}_{t-1}' + b\mathbf{Q}_{t-1}$$

is that the intercept matrix is an identity matrix. This is the case because, as already noted, $\mathsf{E}\mathbf{z}_t\mathbf{z}_t' = \mathbf{I}_N$.

The application consists of jointly modelling dynamics of volatilities and correlations of three electricity futures contracts written on the index of the European Energy Exchange (EEX). They correspond to monthly, quarterly, and yearly maturities. In this case $\boldsymbol{\varepsilon}_t$ is not a daily return vector but a residual vector from an estimated vector error correction model, see Bauwens et al. (2013) for details. The GJR-GARCH equations for the conditional variances of ϕ_{it} are in fact GJR-GARCH-X equations because the GJR-GARCH component is additively completed by a number of exogenous variables. The sum of the DCC coefficient estimates $\widehat{a}$ and $\widehat{b}$ is remarkably low compared to typical results, only 0.879. (In many applications this sum exceeds 0.99.) This demonstrates the importance of the deterministic component $\boldsymbol{\Sigma}(t/T)$ in this application. Other, slightly more general, DCC specifications are considered as well, but discussing them would be outside the scope of this review.

5.5 Multivariate spline-GARCH models

In the multivariate GARCH model by Rangel and Engle (2013) spline-GARCH equations are in a central role. The paper differs from the previous ones in that the starting point is the CAPM model by Sharpe (1964), and the interest lies in deriving correlations between excess returns of assets. Let r_{it} denote the excess return of asset i, and let r_{mt} be the market excess return. Then for asset i,

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it} \qquad (41)$$

In order to work out the correlation between r_{it} and r_{jt} it is assumed that $\varepsilon_{it} = z_{it}h_{it}^{1/2}g_{it}^{1/2}$, where $z_{it} \sim \text{iid}(0,1)$. Furthermore, $r_{mt} = \alpha_m + \varepsilon_{mt}$ where $\varepsilon_{mt} = z_{mt}h_{mt}^{1/2}g_{mt}^{1/2}$. The GJR-GARCH type conditional variance of ε_{it} has the following representation:

$$h_{it} = 1 - \alpha_i - \kappa_i/2 - \beta_i + \alpha_i \frac{\varepsilon_{i,t-1}^2}{g_{i,t-1}} + \kappa_i \frac{\varepsilon_{i,t-1}^2}{g_{i,t-1}} I(r_{i,t-1} < 0) + \beta_i h_{it-1} \tag{42}$$

$i = 1, ..., N$, where it is assumed that $\alpha_i + \kappa_i/2 + \beta_i < 1$ and that g_{it} is as in (7) but with $\gamma = 0$ (the exogenous variable x_t is excluded). Also note that the argument in the indicator variable is $r_{i,t-1}$ and not $\varepsilon_{i,t-1}$ as would be the case in the standard GJR-GARCH model. This implies that $\mathsf{E}h_{it} \neq 1$. The conditional variance for r_{mt} equals

$$h_{mt} = 1 - \alpha_m - \kappa_m/2 - \beta_m + \alpha_m \frac{\varepsilon_{m,t-1}^2}{g_{m,t-1}} + \kappa_i \frac{\varepsilon_{m,t-1}^2}{g_{m,t-1}} I(r_{m,t-1} < 0) + \beta_m h_{m,t-1}.$$

Analogously to (42), the argument of $I(\cdot)$ is $r_{m,t-1} < 0$, not $\varepsilon_{t,m-1} < 0$, so $\mathsf{E}h_{mt} \neq 1$, unless $\alpha_m = 0$. The error vector $\mathbf{z}_t = (z_{mt}, z_{1t}, ..., z_{Nt})'$ has a DCC structure. Consequently, the (conditional) correlation between z_{it} and z_{jt} has a time-varying structure that involves the market long-run component g_{mt} as well as g_{it} and g_{jt}. The conditional correlation between the two excess returns r_{it} and r_{jt} has the following form:

$$\begin{aligned}
\rho_{ijt} \;=\; & \{\beta_i\beta_j h_{it}g_{it} + \beta_i h_{mt}^{1/2}g_{mt}^{1/2}h_{jt}^{1/2}g_{jt}^{1/2}\rho_{mjt} + \beta_j h_{mt}^{1/2}g_{mt}^{1/2}h_{it}^{1/2}g_{it}^{1/2}\rho_{mit} \\
& + h_{it}^{1/2}g_{it}^{1/2}h_{jt}^{1/2}g_{jt}^{1/2}\rho_{ijt}\} \\
& \times \{\beta_i^2 h_{mt}g_{mt} + h_{it}g_{it} + 2\beta_i h_{mt}^{1/2}g_{mt}^{1/2}h_{it}^{1/2}g_{it}^{1/2}\rho_{mit}\}^{-1/2} \\
& \times \{\beta_i^2 h_{mt}g_{mt} + h_{jt}g_{jt} + 2\beta_j h_{mt}^{1/2}g_{mt}^{1/2}h_{jt}^{1/2}g_{jt}^{1/2}\rho_{mjt}\}^{-1/2}
\end{aligned}$$

where ρ_{mit} is the correlation between the market error $z_{mt} = \varepsilon_{mt}/(h_{mt}^{1/2}g_{mt}^{1/2})$ and $z_{it} = \varepsilon_{it}/(h_{it}^{1/2}g_{it}^{1/2})$, and ρ_{ijt} is the conditional correlation between z_{it} and $z_{jt} = \varepsilon_{jt}/(h_{jt}^{1/2}g_{jt}^{1/2})$. The deterministic components g_{it} and g_{jt} contribute to z_{it} and z_{jt}, respectively, and this way, together with g_{mt}, to ρ_{ijt}.

Rangel and Engle (2013) write that the quasi-maximum likelihood estimators of the parameters in (41) and the GJR-GARCH type equations are consistent under mild regularity conditions. Estimating these parameters forms the first stage of estimation. The second stage consists of estimating the correlation parameters conditionally on the first stage parameter estimates. The authors point out that misspecification of the number of knots may lead to inconsistent estimates in stage one. They suggest minimising the consequences of this by using t-distributed errors instead of the normally distributed ones. They also draw attention to biases in the estimates of the correlation (DCC) component when N is large and suggest alternatives to maximum likelihood.

The authors consider daily returns on Dow Jones Industrial Average (DJIA) stocks from December 1988 to December 2006 and have a sample of 33 stocks. Daily returns on the S&P 500 are used as a market factor, and the one-month T-bill rate functions as the time-varying risk-free rate. The market excess return r_m is the difference between these two returns. A set of models with various degrees of complexity are estimated. It is found that the models with the spline-GARCH deterministic component dominate models in which this component is constant (GARCH).

The performance of the SFG-DCC model is compared with that of a selection of competing approaches including a pure DCC-GARCH model and alternatives based on estimating large covariance matrices without any GARCH structure. The forecasts are generated by first estimating the models from December 1988 to June 1995 and forecasting from 1 to 126 days ahead. These 126 days are then added to the sample, the models are re-estimated and another set of forecasts up to 126 days are generated. This produces 22 sets of non-overlapping forecasts. Describing the whole experiment here would take too much space, but the main conclusion, based on long-run forecasts (from 87 to 126 days) and various measures is that the SFG-DCC model on the average performs better than the alternatives.

The research problem of Opschoor, van Dijk and van der Wel (2014) is to assess the impact of financial conditions on volatility and correlations of returns of bank equities. They study it in the spline-GARCH framework. In their multivariate model, the decomposition (50) is defined such that $\mathbf{S}_t$ has a spline-GARCH formulation: $\mathbf{S}_t = \mathrm{diag}(g_{1t}^{1/2}, ..., g_{N,t}^{1/2})$ with $g_{it} = \exp\{\kappa_{i0} + \kappa_{i1}x_{t-1}\}$, $i = 1, ..., N$, where x_t is a stochastic random variable. In the application it is a financial conditions index. In $\mathbf{D}_t = \mathrm{diag}(h_{1t}^{1/2}, ..., h_{Nt}^{1/2})$, the diagonal elements h_{it} have a GJR-GARCH representation. Following Connor and Suurlaht (2013), the authors define

$$\mathbf{P}_t = \overline{\mathbf{P}} + m_{t-1}(\mathbf{11}' - \overline{\mathbf{P}}) = (1 - m_{t-1})\overline{\mathbf{P}} + m_{t-1}\mathbf{11}' \tag{43}$$

where the N-vector $\mathbf{1} = (1, ...,1)'$, $\overline{\mathbf{P}}$ is the (sample) correlation matrix of $\mathbf{z}_t$, and

$$m_t = \frac{\exp\{x_t\beta\} - 1}{\exp\{x_t\beta\} + 1}.$$

This is called the DC-X correlation model. The expression (43) resembles the one in the STCC-GARCH model, see Section 5.7, but the resemblance is superficial. The time-varying correlation matrix $\mathbf{P}_t$ is not a convex combination of $\overline{\mathbf{P}}$ and $\mathbf{11}'$ because m_t fluctuates between -1 and 1, not between 0 and 1. Besides, $\overline{\mathbf{P}}$ and $\mathbf{11}'$ are known matrices, the former after the spline-GARCH equations have been estimated. When $m_t \to -1$, $\mathbf{P}_t \to 2\overline{\mathbf{P}} - \mathbf{11}'$, and when $m_t \to +1$, $\mathbf{P}_t \to \mathbf{11}'$. The latter limit may appear a bit strange because $\mathbf{11}'$ is no longer a positive definite correlation matrix but rather a matrix of rank one, suggesting that all errors z_{it} are perfectly linearly correlated with each other. Some conditions on the elements of $\overline{\mathbf{P}}$ are required for $\mathbf{P}_t$ to remain positive definite for $m_t < 1$.

This situation will change if the matrix $\mathbf{11}'$ in (43) is replaced by a positive definite correlation matrix $\mathbf{P}$ and m_t by a function bounded between zero and one. Setting

$$\mathbf{P}_t = (1 - G_t)\overline{\mathbf{P}} + G_t\mathbf{P} \tag{44}$$

where G_t is defined as in (16) with x_{t-1} as the transition variable, $\mathbf{P}_t$ is, as a convex combination of two positive definite correlation matrices, itself a positive definite correlation matrix. Compared to (43), definition (44) implies $N(N-1)/2$ additional parameters to be estimated. If, however, one wants both to save the 'spirit' of converging to $\mathbf{11}'$ and to save parameters, one could assume that $\mathbf{P}$ is an equicorrelation matrix, see Engle and Kelly (2012). But then, in this parameterisation the correlations would no longer fluctuate around $\overline{\mathbf{P}}$.

The results for Morgan Stanley and Citigroup (for space reasons results for other pairs are not reported) show that the estimated coefficient of x_{t-1} is significantly different from zero in both spline-GARCH models. Figure 3 indicates, however, that nonstationarity visible in both return series does not diminish by the introduction of g_t in the GARCH equations. The correlations estimated by DC-X and DCC, respectively, look different. The former are more stable than the latter. The paper also contains a portfolio Value at Risk analysis, but discussing it here would be beyond the scope of this review.

5.6 Multivariate GARCH-MIDAS

The MIDAS approach discussed in Section 4.4 can be generalised into multivariate GARCH models. Colacito, Engle and Ghysels (2011) consider this possibility. The GARCH equations contain a multiplicative decomposition as in the GARCH-MIDAS-RV model of Engle et al. (2013), see (22). Following the notation in (24), the weight function is defined as $\varphi_k(1, \omega_2)$, $k = 1, ..., K$. In Engle and Rangel (2008), the correlations follow a DCC structure, but Colacito et al. (2011) define them as having a MIDAS-type representation. This implies modifying the intercept (or sample correlation) matrix in the DCC-GARCH model. This is done in two ways. First, the matrix is estimated only through a rolling window. Let $v_{it} = (\sum_{j=t-n_c}^{t} z_{ij}^2)$, $i = 1, ..., N$, $\mathbf{V}_t = \text{diag}(v_{1t}, ..., v_{Nt})$ and $\mathbf{Z}_t = \sum_{j=t-n_c}^{t} \mathbf{z}_j\mathbf{z}_j'$. Then the rolling matrix $\mathbf{C}_t = \mathbf{V}_t^{-1/2}\mathbf{Z}_t\mathbf{V}_t^{-1/2}$.

Second, the matrices are smoothed using a MIDAS type smoother. In the simplest case the beta function is the same for all elements of $\mathbf{C}_t$, and the correlation matrix becomes

$$\mathbf{P}_t(\omega) = \sum_{k=1}^{K} \varphi_k(1,\omega)\mathbf{C}_{t-k}. \tag{45}$$

but more complicated situations are possible as well. Incorporating (45) into the standard DCC model yields the following short-run dynamic structure:

$$\mathbf{Q}_t = (1-a-b)\mathbf{P}_t(\omega) + a\mathbf{z}_{t-1}\mathbf{z}'_{t-1} + b\mathbf{Q}_{t-1}.$$

Colacito et al. (2011) consider more general DCC structures in which a and b are no longer scalars, but they are not discussed here.

Parameters are estimated in two stages as is the case with all DCC models. The GARCH equations are estimated first, which gives the estimates of $\mathbf{z}_t$. From these one obtains the rolling matrices $\mathbf{C}_t$, which allows one to compute an estimate of $\mathbf{P}_t(\omega)$. Given $\widehat{\mathbf{P}}_t(\omega)$, $\mathbf{Q}_t$, a and b can be estimated. Asymptotic properties of the maximum likelihood estimators of these parameters are not known.

The paper contains several examples, of which only one is touched upon here. The main object of interest is correlations between industry portfolios and a 10-year bond. The observation period runs from 15 July 1971 to 30 June 2006. As an example, consider the combination of energy and hi-tech portfolios and the bond. The inclusion of the MIDAS-RV component in the GARCH has a strong impact on the GARCH(1,1) coefficient estimates. The sum $\widehat{\alpha}+\widehat{\beta}$ is now low, for the energy equation even below 0.9. As to DCC, adding the MIDAS component has a negligible impact on estimates of a and b. The sum $\widehat{a}+\widehat{b}$ exceeds 0.995 in both cases. The situation does not change when the DCC-MIDAS component is more richly parameterised, see Table 2 in Colacito et al. (2011) for details.

There exist variants of the DCC-MIDAS model. Conrad, Loch and Rittler (2014) construct a bivariate DCC-MIDAS-X model ('X' will be explained later) for investigating the oil–US stock market relationship. Their purpose is to consider macroeconomic determinants of the long-term correlation between the daily US stock market and crude oil price returns. The spline-GARCH daily short-run component (conditional variance) is defined as before:

$$h_{ji,t} = 1 - \alpha_j - \beta_j + \alpha_j \frac{\varepsilon^2_{j,i-1,t}}{g_{jt}} + \beta_j h_{j,i-1,t}$$

where i indicates the day, t the month, and $\alpha_j + \beta_j < 1$, $j = 1,2$. The positive-valued monthly macroeconomic component g_{jt} is defined as follows:

$$g_{jt} = \exp\{m_j + \theta_j \sum_{k=1}^{K_j} \varphi_k(1, \omega)x_{t-k}\}$$

where x_t is a macro variable.

Correlations are defined as follows. Analogously to LDCC and other models, the conditional covariance matrix has the decomposition $\mathbf{\Sigma}_{jt} = \mathbf{S}_t\mathbf{D}_{jt}\mathbf{P}_t\mathbf{D}_{jt}\mathbf{S}_t$ where $\mathbf{D}_{jt} = \text{diag}(h_{1i,t}^{1/2}, h_{2i,t}^{1/2})$ and $\mathbf{S}_t = \text{diag}(g_{1t}^{1/2}, g_{2t}^{1/2})$. The correlation matrix $\mathbf{P}_t$ is a two-dimensional long-term or macroeconomic matrix, and its only time-varying correlation is defined as

$$\rho_{12t} = \frac{\exp\{2u_{12t}\} - 1}{\exp\{2u_{12t}\} + 1} \tag{46}$$

where u_{12t} has a spline-GARCH-X structure (hence 'X' in DCC-MIDAS-X)

$$u_{12t} = m_{12} + \theta_{12} \sum_{k=1}^{K_{12}} \varphi_k(1, \omega)x_{t-k}.$$

The expression (46) guarantees that ρ_{12t} qualifies as a correlation coefficient because it fluctuates between -1 and $+1$. As a result, the short-run DCC correlation fluctuates around the long-run one driven by lags of x_t and defined by (46). Unlike g_{jt}, u_{12t} does not have to be positive. The parameters of the model are estimated by (quasi) maximum likelihood. Asymptotic properties of the estimators are still unknown.

The stock market is represented through daily returns on the CRSP value-weighted portfolio. The oil price returns are constructed from the daily spot price for West Texas Intermediate (WTI) crude oil for delivery in Cushing, Oklahoma. The observation period extends from January 1993 to November 2011. Five candidate variables are considered in the empirical part of the model, and a DCC-MIDAS-X model is estimated for each of them. Estimates of the pure GARCH model suggest nonstationarity or near-nonstationarity in stock returns, as $\widehat{\alpha} + \widehat{\beta} = 0.992$. Adding the MIDAS component does have some effect on this sum for the five equations, and the MIDAS parameter estimates are significant, however. The estimated GARCH equation for the oil returns has $\widehat{\alpha} + \widehat{\beta} = 1.004$, and in this case rescaling by MIDAS pulls this sum safely below one.

A large number of models with macro variables and correlations are estimated. If the best model is for variances is selected by BIC, it is GARCH(1,1) for the stock returns and GARCH-MIDAS-X based on the 'Leading Index' (LI) in for the oil. For the definition of LI, see Conrad et al. (2014). As to conditional correlations, there is little difference between DCC and DCC-MIDAS for each type of GARCH residuals.

It appears that it is more important to specify GARCH equations correctly than to extend the DCC model. At least in the case of oil returns this is understandable, because there the GARCH equation is mildly explosive. Measured by AIC and BIC, the best models have GARCH-MIDAS-X residuals with LI, but after that there is little to choose between DCC and DCC-MIDAS-X. As to $\widehat{a}+\widehat{b}$, the sum of the estimated DCC coefficients, it equals 0.997 in the former and is marginally lower, equal to 0.989 in the latter.

Connor and Suurlaht (2013) apply the GARCH-MIDAS-RV model (21), (22) and (23) to analysing correlations between daily returns of 11 European stock market indices. They are calculated from daily closing prices from 31 December 1991 to 31 December 31 2010. In the GARCH-MIDAS equations the choice of MIDAS weights is $\varphi(\omega_i, 1)$, $i = 1, ..., 11$, and the correlations are defined using (43). In the application, the sum of the GARCH parameters is close to one for the 11 GARCH models, and including the MIDAS component does not change this outcome. There is a discussion of how to define the variable m_t in (43), given a number of economic variables deemed useful for the purpose, but for space reasons details cannot be considered here.

Chen, Choudhry and Wu (2013) build their analysis on the MEM model of the observed range of returns $r_{it} = \mu_i + \varepsilon_{it}$, where $\mathsf{E}\varepsilon_{it} = 0$, which is the difference between the maximum and minimum logarithmic price within a time interval t. This range, OR_{it} for asset i, is decomposed as

$$OR_{it} = z_{it} h_{it} g_{it} \tag{47}$$

where $z_{it} \sim \text{iid}Gamma(\varphi_i, 1/\varphi_i)$,

$$\begin{aligned} h_{it} &= 1 - \alpha_i - \kappa_i/2 - \beta_i \\ &+ \alpha_i \frac{OR_{i,t-1}}{g_{i,t-1}} + \kappa_i I(r_{i,t-1} < 0) \frac{OR_{i,t-1}}{g_{i,t-1}} + \beta_i h_{i,t-1} \end{aligned} \tag{48}$$

and the MIDAS component g_{it} equals

$$g_{it} = m + \theta \sum_{k=1}^{K} \varphi_k(\omega_{i1}, \omega_{i2}) RVM_{i,t-k} \tag{49}$$

with the Rolling Window Volatility Measure $RVM_{i,t} = \sum_{j=0}^{N'-1} OR_{i,t-j}$. Note that $\mathsf{E}h_{it} \neq 1$, because the indicator variable in (48) is $r_{i,t-1}$ and not $\varepsilon_{i,t-1} = r_{i,t-1} - \mu_i$. The equations (47), (48) and (49) define the Conditional Autoregressive Range MIDAS (CARR-MIDAS) model.

The application concerns the relationship between the oil price and US dollar returns as those series are expected to move together. The original series are US dollar index futures and West Texas Intermediate oil futures prices. As the CARR model indicates, the observed range of returns is the variable of interest. The dependence between returns of them is modelled by a copula. For details, see Chen et al. (2013).

5.7 Multivariate Time-Varying GARCH model

The MTV-GARCH model discussed in Section 3.4 can also be generalised to the multivariate case. The resulting model, the multivariate MTV-GARCH model, is proposed and studied by Amado and Teräsvirta (2014a). It is a Conditional Correlation (CC-) GARCH model with the difference that the GARCH equations are augmented by a multiplicative component as in the TV-GARCH model. More formally, set $\boldsymbol{\varepsilon}_t = \mathbf{S}_t\mathbf{D}_t\mathbf{z}_t$, where $\mathbf{z}_t \sim \text{indep}(\mathbf{0}, \mathbf{P}_t)$, so the time-varying $N \times N$ conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ becomes, as in (36),

$$\mathsf{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\} = \boldsymbol{\Sigma}_t = \mathbf{S}(t/T)\mathbf{D}_t\mathbf{P}_t\mathbf{D}_t\mathbf{S}(t/T) \tag{50}$$

where $\mathbf{D}_t = \text{diag}(h_{1t}^{1/2}, ..., h_{Nt}^{1/2})$, $\mathbf{S}(t/T) = \text{diag}(g_1^{1/2}(t/T), ..., g_N^{1/2}(t/T))$ and $\mathbf{P}_t$ is a positive definite time-varying correlation matrix. The positive-valued functions $g_i(t/T)$, $i = 1, ..., N$ are defined as in (15) and (16). Analogously to the univariate case, $\boldsymbol{\phi}_t = \mathbf{S}_t^{-1}\boldsymbol{\varepsilon}_t$ follows a standard conditional correlation GARCH model. Setting $\mathbf{P}_t \equiv \mathbf{P}$ yields the Constant Conditional Correlation (CCC-) GARCH model by Bollerslev (1990). If $\mathbf{P}_t \equiv \mathbf{P}$, $\mathbf{D}_t = \mathbf{I}_N$ and $\mathbf{S}_t = g(t/T)\mathbf{I}_N$, one obtains an analogue to the model (32) that Osiewalski and Pajor (2009) considered but discarded as too simple.

Defining the recursion

$$\mathbf{Q}_t = (1 - a - b)\overline{\mathbf{P}} + a\sum_{j=1}^{n}\mathbf{z}_{t-j}\mathbf{z}_{t-j}' + b\mathbf{Q}_{t-1} \tag{51}$$

where $n \geq N$, $a+b < 1$, and $\overline{\mathbf{P}}$ is the sample correlation matrix of $\mathbf{z}_t$, gives the Varying-Correlation (VC-) GARCH model by Tse and Tsui (2002). Setting $n = 1$ in (51) leads to the Dynamic Conditional Correlation (DCC-) GARCH model of Engle (2002). As in LDCC, since recursions in $\mathbf{Q}_t$ do not automatically generate correlation matrices, the adjustment (38) has to be made to obtain a proper correlation matrix for time t.

The multivariate MTV-GARCH or MTV-CC-GARCH model thus consists of N univariate TV-GARCH components and a (possibly) time-varying correlation matrix $\mathbf{P}_t$. Maximum likelihood estimation of parameters can be carried out in two stages: by first jointly estimating $\mathbf{S}(t/T)$ and $\mathbf{D}_t$ and then, given the estimates $\widehat{\mathbf{z}}_t$, the correlations. From Amado and Teräsvirta (2013) it follows that maximum likelihood estimators of the parameters $\mathbf{S}(t/T)$ and $\mathbf{D}_t$ are consistent and asymptotically normal under the assumption $\mathbf{P}_t = \mathbf{I}_N$. No asymptotic theory is available for the model when $\mathbf{P}_t$ (or $\mathbf{Q}_t$)

is time-varying as in (51).

The conditional correlations are defined as the ones in $\mathbf{P}_t$, so they are correlations between the elements of $\mathbf{z}_t = \mathbf{S}^{-1}(t/T)\mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t$. Similarly to the LDCC, it follows that the deterministic components (diagonal elements of $\mathbf{S}(t/T)$) affect the correlations. This may be useful in cases where the correlations are changing systematically over time and do not fluctuate around a constant level as in, say, the VC- or DCC-GARCH models.

The functions $g_i(t/T)$, $i = 1, ..., N$, may also be defined using other than TV-GARCH specifications. Asymptotic properties of maximum likelihood estimators of GARCH parameters in SEMIGARCH are known from Feng (2004), whereas the exponential quadratic spline- or P-spline GARCH models they are unknown.

Amado and Teräsvirta (2014a) apply the MTV-CC-GARCH model to daily returns of seven frequently traded stocks belonging to the S&P 500 index. The modelling period extends from 29 September 1998 to 7 October 2008. The observations from 8 October 2008 to 31 December 2009 are saved for forecasting. Because the returns are stock returns, GJR-GARCH equations are used instead of standard GARCH. The aforementioned asymptotic results obtained for the GARCH equations remain valid in the GJR-GARCH case; see Amado and Teräsvirta (2013) for discussion.

Persistence estimates from the seven TV-GJR-GARCH models vary from 0.97 to 0.78. The estimated $g_i(t/T)$ have a rather similar shape: values of the function decrease around 2002-2004 and increase again around 2008. The authors fit a model with a CCC, VC and DCC structure to the estimated residuals $\widehat{\mathbf{z}}_t$. Judging from the maximised log-likelihood, MTV-GJR-VC-GARCH fits the data best and, as may be expected, CCC-GARCH has the lowest maximum. For comparison, the performance of CC-spline-GJR-GARCH is studied as well. It has the best fit overall, whereas, again as expected, GJR-CCC-GARCH with $\mathbf{S}(t/T) = \mathbf{I}_N$ is worst in this respect.

The covariance matrix $\boldsymbol{\Sigma}_t$ is being forecast one day ahead. Since the true $\boldsymbol{\Sigma}_t$ is unknown, following Andersen, Bollerslev, Diebold and Labys (2003) it was proxied by the realised covariance matrix based on 5-minute returns. The main forecast error loss function is the Frobenius distance

$$L_{F,T+i} = (1/N^2)\mathrm{vec}(\boldsymbol{\Sigma}_{T+i} - \widehat{\boldsymbol{\Sigma}}_{T+i})'\mathrm{vec}(\boldsymbol{\Sigma}_{T+i} - \widehat{\boldsymbol{\Sigma}}_{T+i}) \tag{52}$$

where $\widehat{\boldsymbol{\Sigma}}_{T+i}$ is the one-step-ahead forecast of $\boldsymbol{\Sigma}_{T+i}$. The Root Mean Square Error (RMSE) over the forecasting period is computed from the Frobenius distance (52). The mean Absolute Deviation and Median Squared Error were also used but not discussed here. Since the estimated splines of the spline-GARCH model by Engle and Rangel (2008) all point strongly upwards at the end of the estimation period, there is the question of how to extrapolate them when forecasting. This is not an important issue if one is only forecasting one day ahead but becomes one when the forecast horizon is long. Amado and Teräsvirta (2014a) assume that when the starting-point for forecasting is T, $g_i(1 + 1/T) = g(1)$ for $i = 1, ..., N$..

The results indicate that the models with MTV-GJR-GARCH equations generate better forecasts than the ones with Spline-GJR-GARCH or GJR-GARCH. This is true for both CCC, VC and DCC specifications. The best model overall is the MTV-GJR-VC-GARCH model. The Model Confidence Set obtained from this set of models, see Hansen, Lunde and Nason (2011), consists of the three models (CCC, VC and DCC) in which the GARCH equations are MTV-GJR-GARCH equations of Amado and Teräsvirta (2013). It seems that at least in this seven-dimensional example getting the levels, i.e., g_{iT}, $i = 1, ..., 7$, 'right' is more important than choosing between the three correlation structures.

Due to the correlation structure (51) of VC and DCC, the correlations are restricted to fluctuate around $\overline{\mathbf{P}}$. Besides, the structure is aimed at capturing 'correlation clustering'. If the correlations in reality for example move monotonically in one direction over time, this translates into $\widehat{a} + \widehat{b}$ being very close to one (because of correlation targeting, the restriction $a + b < 1$ has to be maintained). This problem is avoided by applying the multiplicative Time-Varying Smooth Transition Conditional Correlation GARCH (TVC) model by Silvennoinen and Teräsvirta (2017). In this model, the correlation matrix $\mathbf{P}_t = \mathbf{P}(t/T)$ is changing deterministically:

$$\mathbf{P}(t/T) = G_{\text{corr}}(t/T, \gamma, c)\mathbf{P}_{(1)} + \{1 - G_{\text{corr}}(t/T, \gamma, c)\}\mathbf{P}_{(2)} \tag{53}$$

where $\mathbf{P}_{(1)}$ and $\mathbf{P}_{(2)}$, $\mathbf{P}_{(1)} \neq \mathbf{P}_{(2)}$, are positive definite correlation matrices, and

$$G_{\text{corr}}(t/T, \gamma, \mathbf{c}) = (1 + \exp\{-\gamma \prod_{k=1}^{K}(t/T - c_k)\})^{-1}, \ \gamma > 0 \tag{54}$$

and $c_1 \leq ... \leq c_K$. As a convex combination of $\mathbf{P}_{(1)}$ and $\mathbf{P}_{(2)}$, $\mathbf{P}(t/T)$ is positive definite. The definition of $\mathbf{P}(t/T)$ in (53) is analogous to that in the Smooth Transition Conditional Correlation (STCC-) GARCH, see Silvennoinen and Teräsvirta (2005, 2015). The difference is that the transition variable in (53) is deterministic, not stochastic.

Silvennoinen and Teräsvirta (2017) show that maximum likelihood estimators of the parameters of the TVC model, parameters in the correlation component included, are consistent and asymptotically normal. This paves way for misspecification testing which should be an essential part of any model building exercise.

With the exception of Silvennoinen and Teräsvirta (2017), correlations become deterministically time-varying through deterministic components in GARCH equations, see (50). Preceding these developments Silvennoinen and Teräsvirta (2009a), assuming that $\mathbf{S}_t = \mathbf{I}_N$ in (50), have proposed the following correlation matrix:

$$\begin{aligned} \mathbf{P}_t &= (1 - G_{2t})\{(1 - G_{1t})\mathbf{P}_{(11)} + G_{1t}\mathbf{P}_{(21)}\} \\ &\quad + G_{2t}\{(1 - G_{1t})\mathbf{P}_{(21)} + G_{1t}\mathbf{P}_{(22)}\} \end{aligned} \tag{55}$$

where G_{1t} and G_{2t} are logistic transition functions defined in (16) for $k_i = 1$, $i = 1, 2$, and $\mathbf{P}_{(ij)}$, $i, j = 1, 2$, are positive definite correlation matrices. The ensuing model is called the Double Smooth Transition Conditional Correlation (DSTCC-)GARCH model. Suppose that the argument of G_{2t} is rescaled time t/T whereas G_{1t} controls rapid movements in correlations. Then (55) allows for long-run changes in the dynamics of correlations. The correlations may gradually as a function of time move to fluctuate around a new level. In the standard Smooth Transition Conditional Correlation model (Silvennoinen and Teräsvirta 2005, 2015) where $G_{2t} = 0$, the correlations fluctuate between two positive definite correlation matrices and do not experience systematic changes unless of course G_{1t} is a function of t/T. This special case has been first studied in the bivariate setting by Berben and Jansen (2005).

As already mentioned, Silvennoinen and Teräsvirta (2017) reintroduce $\mathbf{S}(t/T)$ with $\mathbf{P}_t = \mathbf{P}(t/T)$ in $\mathbf{\Sigma}(t/T)$ and prove consistency and asymptotic normality of maximum likelihood estimators for all parameters of the model, provided G_{1t} is a function of t/T, while still assuming $G_{2t} = 0$. The TVC model can be generalised to the DSTCC case, but the asymptotic properties of maximum likelihood estimators for parameters of that model are not known. (The parameters are estimated jointly, unless G_{2t} is a function of lagged z_{it}.) However, given the available asymptotic theory, TVC can be tested against DSTCC. Silvennoinen and Teräsvirta (2009a) have developed an LM test for this purpose with just assuming that the maximum likelihood estimators of the parameters of the null model are consistent. Despite the lack of asymptotic theory, it may argued that DSTCC offers an alternative to the LDCC model by Feng (2006) and the FSG-DCC model by Rangel and Engle (2013) when the number of assets in the model remains reasonably small.

The paper contains a bivariate application to daily log-returns of the S&P 500 index and the 30-year US Treasury bill from 3 January 2000 to 6 July 2015, 4046 observations. The amplitude of clusters in both return series varies strongly over time, and after testing and rejecting constancy of $g_i(t/T)$, $i = 1, 2$, TV-GARCH equations are specified for both series. Constancy of the correlation between z_t^S (S for stock) and z_t^B (B for bond) is also tested and rejected. (The relevant test will be discussed in a forthcoming working paper.) Joint modelling of GARCH parameters and time-varying correlations, that is, a complete TVC model, gives a correlation function with two sharp changes, one from about -0.3 to close to zero around 2004 and another back to -0.3

around 2007.

5.8 Misspecification testing

There do not seem to be many misspecification tests available for testing the adequacy of multivariate GARCH models with multiplicative decomposition of the (conditional) covariance matrix. Catani, Teräsvirta and Yin (2017) develop a Lagrange multiplier test of the model

$$\boldsymbol{\varepsilon}_t = \mathbf{D}_t \mathbf{z}_t$$

where $\mathbf{D}_t = (h_{1t}^{1/2}, ..., h_{Nt}^{1/2})$ contains the GARCH equations, so $\mathbf{z}_t = \mathbf{D}_t^{-1}\boldsymbol{\varepsilon}_t$. Under the alternative it is assumed that $\mathbf{z}_t$ is not iid but contains dynamic structure such that $\mathbf{z}_t = \mathbf{G}_t \boldsymbol{\zeta}_t$, where $\boldsymbol{\zeta}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \mathbf{P})$. Under H_0, $\mathbf{G}_t = \mathbf{I}_N$, so the null model is the CCC-GARCH model. The diagonal matrix $\mathbf{G}_t = (g_{1t}^{1/2}, ..., g_{Nt}^{1/2})$ is a multivariate generalisation of (29) in that $g_{it} = 1 + \sum_{j=1}^{r} \delta_{ij} z_{i,t-j}^2$, $i = 1, ..., N$. Letting $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', ..., \boldsymbol{\delta}_N')'$ be an Nr-vector whose $r \times 1$ vector blocks are $\boldsymbol{\delta}_i = (\delta_{i1}, ..., \delta_{ir})'$, the null hypothesis can be expressed as $\boldsymbol{\delta} = \mathbf{0}$. The authors develop a Lagrange multiplier test for testing this hypothesis and show that the resulting test statistic has an asymptotic χ^2-distribution with Nr degrees of freedom. The Lagrange multiplier statistic by Lin and Li (1997) turns out to be a parsimonious special case of the statistic by Catani et al. (2017).

Generalising the test to the situation in which the null model is a DCC or VC model, say, is hindered by the fact that the asymptotic properties of the maximum likelihood estimators of parameters in these models are not known. For an illuminating discussion, see Engle and Kelly (2012). For the same reason, testing for example DCC against DCC-MIDAS does not seem possible.

6 Final remarks

This survey only concerns a small subset of GARCH models. It consists of models in which the usual conditional variance component is multiplicatively augmented by another time-varying component that can be either deterministic or stochastic. For readers who want more background information about GARCH there is a modern overview by Francq and Zakoïan (2010) that also covers the statistical inference, and a more compact exposition by Gouriéroux (1997). Many econometrics or financial econometrics texbooks contain chapters on ARCH and GARCH; see for example Tsay (2010), Teräsvirta, Tjøstheim and Granger (2010) or Box, Jenkins, Reinsel and Ljung (2015). Engle (1995) contains the most important early contributions reprinted in a single volume. Surveys of univariate GARCH models published over the years include Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994), Palm (1996), Teräsvirta

(2009) and Zivot (2009). Li, Zhang, Zhu and Ling (2018) develop a nonstationary GARCH model (Zero-Drift GARCH). It describes volatility series that generate clusters with varying amplitudes but does it without multiplicative decomposition. Multivariate GARCH models are surveyed by Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009b). de Almeida, Hotta and Ruiz (2018) provide a useful discussion of popular BEKK- and DCC-GARCH type of models and compare their performance in forecasting in cases where the data are generated by a more general model than either of the two. Finally, in addition to articles already mentioned, the two volumes, Andersen, Davis, Kreiss and Mikosch (2009) and Bauwens, Hafner and Laurent (2012), contain several useful articles on various aspects of GARCH models.

References

Akaike, H.: 1974, A new look at statistical model identification, *IEEE Transactions on Automatic Control* **AC–19**, 716–723.

Amado, C. and Laakkonen, H.: 2013, Modeling time-varying volatility in financial returns: Evidence from the bond markets, *in* N. Haldrup, M. Meitz and P. Saikkonen (eds), *Essays in nonlinear time series econometrics*, Oxford University Press, Oxford, pp. 139–160.

Amado, C., Silvennoinen, A. and Teräsvirta, T.: 2017, Modelling and forecasting WIG20 daily returns, *Central European Journal of Economic Modelling and Econometrics* **9**, 173–200.

Amado, C. and Teräsvirta, T.: 2008, Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure, *SSE/EFI Working Paper Series in Economics and Finance 691*, Stockholm School of Economics.

Amado, C. and Teräsvirta, T.: 2013, Modelling volatility by variance decomposition, *Journal of Econometrics* **175**, 153–165.

Amado, C. and Teräsvirta, T.: 2014a, Conditional correlation models of autoregressive conditional heteroscedasticity with nonstationary GARCH equations, *Journal of Business and Economic Statistics* **32**, 69–87.

Amado, C. and Teräsvirta, T.: 2014b, Modelling changes in the unconditional variance of long stock return series, *Journal of Empirical Finance* **25**, 15–35.

Amado, C. and Teräsvirta, T.: 2017, Specification and testing of multiplicative time-varying GARCH models with applications, *Econometric Reviews* **36**, 421–446.

Andersen, T. G. and Bollerslev, T.: 1998, Deutsche Mark-Dollar volatility intraday activity patters, macroeconomic announcements, and longer run dependencies, *Journal of Finance* **53**, 219–265.

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

Andersen, T. G., Davis, R. A., Kreiss, J.-P. and Mikosch, T. (eds): 2009, *Handbook of Financial Time Series*, Springer, Berlin.

Ashgarian, H., Hou, A. J. and Javed, F.: 2013, The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach, *Journal of Forecasting* **32**, 600–612.

Audrino, F. and Bühlmann, P.: 2003, Volatility estimation with functional gradient descent for very high-dimensional financial time series, *Journal of Computational Finance* **6**, 65–89.

Audrino, F. and Bühlmann, P.: 2009, Splines for financial volatility, *Journal of the Royal Statistical Society, Series B* **71**, 655–670.

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O.: 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **74**, 3–30.

Bauwens, L., Hafner, C. and Laurent, S. (eds): 2012, *Handbook of volatility models and their applications*, Wiley, New York.

Bauwens, L., Hafner, C. M. and Pierret, D.: 2013, Multivariate volatility modeling of electricity futures, *Journal of Applied Econometrics* **28**, 743–761.

Bauwens, L., Laurent, S. and Rombouts, J. V. K.: 2006, Multivariate GARCH models: A survey, *Journal of Applied Econometrics* **21**, 79–109.

Berben, R.-P. and Jansen, W. J.: 2005, Comovement in international equity markets: A sectoral view, *Journal of International Money and Finance* **24**, 832–857.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Bollerslev, T.: 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72**, 498–505.

Bollerslev, T., Chou, R. Y. and Kroner, K. F.: 1992, ARCH modeling in finance. A review of the theory and empirical evidence, *Journal of Econometrics* **52**, 5–59.

Bollerslev, T., Engle, R. F. and Nelson, D. B.: 1994, ARCH models, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, North-Holland, Amsterdam, pp. 2959–3038.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M.: 2015, *Time Series Analysis: Forecasting and Control*, fifth edn, Wiley, New York.

Brownlees, C. T. and Gallo, G. M.: 2010, Comparison of volatility measures: A risk management perspective, *Journal of Financial Econometrics* **8**, 29–56.

Catani, P., Teräsvirta, T. and Yin, M.: 2017, A Lagrange multiplier test for testing the adequacy of constant conditional correlation GARCH model, *Econometric Reviews* **36**, 599–621.

Chen, W.-P., Choudhry, T. and Wu, C.-C.: 2013, The extreme value in crude oil and US dollar markets, *Journal of International Money and Finance* **36**, 191–210.

Colacito, R., Engle, R. F. and Ghysels, E.: 2011, A component model for dynamic correlations, *Journal of Econometrics* **164**, 45–59.

Connor, G. and Suurlaht, A.: 2013, Dynamic stock market covariances in the Eurozone, *Journal of International Money and Finance* **37**, 353–370.

Conrad, C. and Loch, K.: 2015, Anticipating long-term stock market volatility, *Journal of Applied Econometrics* **30**, 1090–1114.

Conrad, C., Loch, K. and Rittler, D.: 2014, On the macroeconomic determinants of long-term volatilities and correlations in U.S. stock and crude oil markets, *Jounal of Empirical Finance* **29**, 26–40.

Conrad, C. and Schienle, M.: 2017, Testing for an omitted multiplicative long-term component in GARCH models, *Working paper*, Heidelberg University, Available at SSRN: http://dx.doi.org/10.2139/ssrn.2631976.

Dahlhaus, R.: 1997, Fitting time series models to nonstationary processes, *The Annals of Statistics* **25**, 1–37.

de Almeida, D., Hotta, L. K. and Ruiz, E.: 2018, MGARCH models: Trade-off between feasibility and flexibility, *International Journal of Forecasting* **34**, 45–63.

Diebold, F. X. and Mariano, R. S.: 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**, 253–263.

Eilers, P. H. C. and Marx, B. D.: 1996, Flexible smoothing with B-splines, *Statistical Science* **11**, 89–121.

Engle, R. F.: 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1007.

Engle, R. F.: 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models, *Journal of Business and Economic Statistics* **20**, 339–350.

Engle, R. F. and Bollerslev, T.: 1986, Modelling the persistence of conditional variances, *Econometric Reviews* **5**, 1–50.

Engle, R. F. (ed.): 1995, *ARCH Selected readings*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Engle, R. F., Ghysels, E. and Sohn, B.: 2013, Stock market volatility and macroeconomic fundamentals, *Review of Economics and Statistics* **95**, 776–797.

Engle, R. F. and Kelly, B.: 2012, Dynamic equicorrelation, *Journal of Business and Economic Statistics* **30**, 212–228.

Engle, R. F. and Kroner, K. F.: 1995, Multivariate simultaneous generalized ARCH, *Econometric Theory* **11**, 122–150.

Engle, R. F. and Rangel, J. G.: 2008, The spline-GARCH model for low-frequency volatility and its global macroeconomic causes, *Review of Financial Studies* **21**, 1187–1222.

Engle, R. F. and Russell, J. R.: 1998, Autoregressive Conditional Duration: A new model for irregularly spaced transaction data, *Econometrica* **66**, 1127–1162.

Feng, Y.: 2004, Simultaneously modeling conditional heteroskedasticity and scale change, *Econometric Theory* **20**, 563–596.

Feng, Y.: 2006, A local dynamic conditional correlation model, *MPRA Paper 1592*, http://mpra.ub.uni-muenchen.de/1592.

Feng, Y. and McNeil, A. J.: 2008, Modelling of scale change, periodicity and conditional heteroskedasticity in return volatility, *Economic Modelling* **25**, 850–867.

Francq, C. and Zakoïan, J.-M.: 2010, *GARCH Models. Structure, Statistical Inference and Financial Applications*, Wiley, Chichester.

Gallant, A. R.: 1981, On the bias in flexible functional forms and an essentially unbiased form. The Fourier flexible form, *Jounal of Econometrics* **15**, 211–245.

Gallant, A. R.: 1984, The Fourier flexible form, *American Journal of Agricultural Economics* **66**, 204–208.

Girardin, E. and Joyeux, R.: 2013, Macro fundamentals as a source of stock market volatility in China: A GARCH-MIDAS approach, *Economic Modelling* **34**, 59–68.

Glosten, L. W., Jagannathan, R. and Runkle, D. E.: 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**, 1779–1801.

Good, I. J. and Gaskins, R. A.: 1971, Nonparametric roughness penalties for probability densities, *Biometrika* **58**, 255–277.

Gouriéroux, C.: 1997, *ARCH Models and Financial Applications*, Springer, Berlin.

Hafner, C. M. and Linton, O.: 2010, Efficient estimation of a multivariate multiplicative volatility model, *Journal of Econometrics* **159**, 55–73.

Han, H. and Kristensen, D.: 2017, Semiparametric multiplicative GARCH-X model: Adopting economic variables to explain volatility, *Unpublished paper*.

Hansen, P. R., Lunde, A. and Nason, J. M.: 2011, The model confidence set, *Econometrica* **79**, 453–497.

Knight, F.: 1921, *Risk, uncertainty, and profit*, Houghton Mifflin, Boston.

Lamoureux, C. G. and Lastrapes, W. G.: 1990, Persistence in variance, structural change and the GARCH model, *Journal of Business and Economic Statistics* **8**, 225–234.

Li, D., Zhang, X., Zhu, K. and Ling, S.: 2018, The ZD-GARCH model: A new way to study heteroscedasticity, *Journal of Econometrics* **202**, 1–17.

Lin, S. and Li, W. K.: 1997, Diagnostic checking of multivariate nonlinear time series with multivariate ARCH errors, *Journal of Time Series Analysis* **18**, 447–464.

Lundbergh, S. and Teräsvirta, T.: 2002, Evaluating GARCH models, *Journal of Econometrics* **110**, 417–435.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T.: 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Mandelbrot, B.: 1963, The variation of certain speculative prices, *Journal of Business* **36**, 394–419.

Mazur, B. and Pipień, M.: 2012, On the empirical importance of periodicity in the volatility of financial returns - time varying GARCH as a second order APC(2) process, *Central European Journal of Economic Modelling and Econometrics* **4**, 95–116.

Mikosch, T. and Stărică, C.: 2004, Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects, *Review of Economics and Statistics* **86**, 378–390.

Mishra, S., Su, L. and Ullah, A.: 2010, Semiparametric estimator of time series conditional variance, *Journal of Business and Economic Statistics* **28**, 256–274.

Nelson, D. B.: 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**, 347–370.

Opschoor, A., van Dijk, D. and van der Wel, M.: 2014, Predicting volatility and correlations with financial conditions indexes, *Jounal of Empirical Finance* **29**, 435–447.

Osiewalski, J.: 2009, New hybrid models of multivariate volatility (a Bayesian perspective), *Przeglad Statystycny* **56**, 15–22.

Osiewalski, J. and Pajor, A.: 2007, Flexibility and parsimony in multivariate financial modelling: a hybrid bivariate DCC-SV model, *in* W. Milo and P. Wdowiński (eds), *Financial Markets. Principles of Modeling, Forecasting and Decision-Making*, Vol. 3 of *FindEcon Monograph Series*, Ldź University Press, Ldź, pp. 11–26.

Osiewalski, J. and Pajor, A.: 2009, Bayesian analysis for Hybrid MSF-SBEKK models of multivariate volatility, *Central European Journal of Economic Modelling and Econometrics* **1**, 179–202.

Palm, F. C.: 1996, GARCH models of volatility, *in* G. Maddala and C. Rao (eds), *Statistical Methods in Finance*, Vol. 14 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 209–240.

Rangel, J. G. and Engle, R. F.: 2013, The Factor-Spline-GARCH model for high and low frequency correlations, *Journal of Business & Economic Statistics* **30**, 109–124.

Rissanen, J.: 1978, Modeling by shortest data description, *Automatica* **14**, 465–471.

Rodríguez-Poo, J. M. and Linton, O.: 2001, Nonparametric factor analysis of residual time series, *TEST* **10**, 161–182.

Schwarz, G.: 1978, Estimating the dimension of a model, *Annals of Statistics* **6**, 461–464.

Sharpe, W. F.: 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* **19**, 425–442.

Silvennoinen, A. and Teräsvirta, T.: 2005, Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations, *SSE/EFI Working Papers in Economics and Finance 577*, Stockholm School of Economics.

Silvennoinen, A. and Teräsvirta, T.: 2009a, Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model, *Journal of Financial Econometrics* **7**, 373–411.

Silvennoinen, A. and Teräsvirta, T.: 2009b, Multivariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 201–229.

Silvennoinen, A. and Teräsvirta, T.: 2015, Modeling conditional correlations of asset returns: A smooth transition approach, *Econometric Reviews* **34**, 174–197.

Silvennoinen, A. and Teräsvirta, T.: 2017, Consistency and asymptotic normality of maximum likelihood estimators of a multiplicative time-varying smooth transition correlation GARCH model, *Research Paper 2017-28*, CREATES, Aarhus University, Aarhus.

Song, P. X., Fan, Y. and Kalbfleisch, J. D.: 2005, Maximization by parts in likelihood inference, *Journal of the American Statistical Association* **100**, 1145–1158.

Taylor, S. J.: 1986, *Modelling Financial Time Series*, Wiley, Chichester.

Teräsvirta, T.: 2009, An introduction to univariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, New York, pp. 17–42.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J.: 2010, *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

Tsay, R.: 2010, *Analysis of financial time series*, third edn, Wiley, Hoboken, NJ.

Tse, Y. K. and Tsui, K. C.: 2002, A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business and Economic Statistics* **20**, 351–362.

van Bellegem, S.: 2012, Locally stationary volatility models, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Wiley Handbook in Financial Engineering and Econometrics: Volatility Models and Their Applications*, Wiley, New York, pp. 249–268.

van Bellegem, S. and von Sachs, R.: 2004, Forecasting economic time series with unconditional time-varying variance, *International Journal of Forecasting* **20**, 611–627.

Wang, F. and Ghysels, E.: 2015, Econometric analysis of volatility component models, *Econometric Theory* **31**, 362–393.

Zhang, X., Feng, Y. and Peitz, C.: 2017, A general class of SemiGARCH models based on the Box-Cox transformation, *Working Paper 2017-06*, Center for International Economics, Paderborn University.

Zivot, E.: 2009, Practical issues in the analysis of univariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer, Berlin, pp. 113–155.

Research Papers 2018

CREATES

Center for Research in Econometric Analysis of Time Series