Identification of structural multivariate GARCH models

Hafner, Christian M.

2022-03

Hafner , C M , Herwartz , H & Maxand , S 2022 , ' Identification of structural multivariate GARCH models ' , Journal of Econometrics , vol. 227 , no. 1 , pp. 212-227 . https://doi.org/10.1016/j.jeconom.2020.0

http://hdl.handle.net/10138/339818
https://doi.org/10.1016/j.jeconom.2020.07.019

Journal of Econometrics 227 (2022) 212–227

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Identification of structural multivariate GARCH models

Christian M. Hafner[a,b,*], Helmut Herwartz[c], Simone Maxand[d]

[a] Louvain Institute for Data Analysis and Modelling in Economics and Statistics (LIDAM), Université catholique de Louvain, Belgium
[b] ISBA, Université catholique de Louvain, Belgium
[c] Department of Economics, University of Goettingen, Germany
[d] Department of Political and Economic Studies, University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 26 October 2018
Received in revised form 7 June 2020
Accepted 1 July 2020
Available online 29 July 2020

JEL classification:
C32
G15

Keywords:
Structural innovations
Identifying assumptions
MGARCH
Portfolio risk
Volatility transmission

ABSTRACT

The class of multivariate GARCH models is widely used to quantify and monitor volatility and correlation dynamics of financial time series. While many specifications have been proposed in the literature, these models are typically silent about the system inherent transmission of implied orthogonalized shocks to vector returns. In a framework of non-Gaussian independent structural shocks, this paper proposes a loss statistic, based on higher order co-moments, to discriminate in a data-driven way between alternative structural assumptions about the transmission scheme, and hence identify the structural model. Consistency of identification is shown theoretically and via a simulation study. In its structural form, a four dimensional system comprising US and Latin American stock market returns points to a substantial volatility transmission from the US to the Latin American markets. The identified structural model improves the estimation of classical measures of portfolio risk, as well as corresponding variations.

1. Introduction

In the wake of a growing dependence across markets and countries, evidenced during the financial and economic crisis, 2007–2009, and the European debt crisis, concern has risen among central bankers, regulators, policy makers and portfolio managers about understanding the volatility linkages between countries, markets, and asset classes from a fundamental perspective. In this context, multivariate GARCH models (MGARCH) have been used extensively to analyse contagion and transmission of risks, either via variance impulse response functions, tests of specific parameter restrictions, tests for causality in variance, or forecast error variance decompositions, see Bauwens et al. (2006) for a review. The latter approach has been initially proposed by Diebold and Yilmaz (2009) to model contemporaneous variance transmissions in vector autoregressions (VAR) of realized variances.

While conveying insightful information about the volatility and correlation dynamics, both realized volatility VAR and MGARCH models are limited in the sense that in most studies the underlying model of shock transmissions lacks identification in a strictly structural sense. For instance, Diebold and Yilmaz (2009) use the order-dependent Cholesky factorization, and similar ad-hoc choices are often encountered in MGARCH specifications.

Identification of structural MGARCH models has been addressed previously in the framework of the orthogonal GARCH model of van der Weide (2002) and Rigobon (2003), as well as the structural conditional correlation model of Weber (2010). Both approaches come at the cost of imposing reduced form dynamic profiles that are more restrictive than

* Corresponding author at: ISBA, Université catholique de Louvain, Belgium.
E-mail addresses: christian.hafner@uclouvain.be (C.M. Hafner), hherwartz@uni-goettingen.de (H. Herwartz), simone.maxand@helsinki.fi (S. Maxand).

https://doi.org/10.1016/j.jeconom.2020.07.019

a general MGARCH model such as the BEKK model of Engle and Kroner (1995). Alternatively, one may use external economic information for identification as in Herwartz and Roestel (2018). On the high frequency scale of financial markets, however, such external information is generally scarce and often not consensual.

In this paper, retaining the full flexibility of unrestricted BEKK models, we build upon recent advances in data based identification of SVARs to develop a new structural MGARCH model. In particular, we exploit the uniqueness of non-Gaussian structural shocks for MGARCH identification, see e.g. Lanne et al. (2017), Gouriéroux et al. (2017) and Lütkepohl and Kilian (2017). The starting point is the definition of structural shocks as stochastically independent innovations as in Hafner and Herwartz (2006). This definition exploits the full information of the joint distribution of shocks, and not only their second order moment structure. The idea is that under our assumptions, higher order conditional moments, as opposed to second order moments, are not invariant with respect to rotations of a decomposed conditional covariance matrix. For the identification of independent shocks we suggest a feasible loss (test) statistic that is a weighted sum of squared differences between empirical third and fourth order moments and their theoretical counterparts under the independence assumption. We show theoretically and via a simulation study that the minimization of the proposed test statistic consistently identifies the structural MGARCH model that has been estimated by quasi maximum likelihood (QML).

The identified structural model is particularly helpful in modelling the higher order moment structure of portfolio returns which is an important ingredient for modern risk management and regulatory purposes. In a non-Gaussian framework, risk measures such as Value-at-Risk (VaR) and expected shortfall (ES) directly depend on the *a priori* unknown structural model. Furthermore, the structural model is also helpful in quantifying uncertainties inherent in conditional VaR and ES statistics. To further improve the modelling of tail events, we introduce a higher order risk measure by considering the difference between squared returns and conditional variances. While the conditional mean of this variable is zero, its conditional variance is directly linked to the conditional kurtosis of portfolio returns. Analysing the variation and tail risk of this variable provides important insights for risk management with respect to variations of classical risk measures over time.

The merits of our approach are discussed in detail based on an application to a four dimensional system of weekly stock returns of US and Latin American markets. Unlike an ad-hoc symmetric covariance decomposition, the estimation results suggest an active role of US markets in transmitting volatilities to Latin American markets. We show via simulations that VaR and ES measures are well approximated using the estimated structural model, as opposed to a symmetric model. We also show that this holds for quantities measuring the higher order type risk.

The next section introduces the structural MGARCH model and discusses the moment based identification criterion. A simulation study sheds light on the discriminatory strength of the identification criterion in finite samples. In Section 3 we highlight how structural information can be beneficial for both types of risk analysis. Section 4 provides an empirical illustration of identified volatility transmission patterns, and shows how structural information improves practical matters of monitoring portfolio risks. Section 5 concludes.

2. Structural MGARCH and identification

Suppose that an N-dimensional vector of financial returns, denoted r_t, is decomposed into a conditional mean and an error component,

$$r_t = \mu_t + e_t, \tag{1}$$

where $\mu_t = E[r_t|\mathcal{F}_{t-1}]$, $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$, e_t is an error term with conditional mean equal to zero. We suppose that e_t is conditionally heteroskedastic, $H_t := \mathrm{Var}[e_t|\mathcal{F}_{t-1}]$, where the $N \times N$ matrix process $\{H_t\}$ is symmetric, positive definite, and $\mathcal{F}_{t-1}$-measurable.

Numerous MGARCH specifications for H_t have been proposed, see e.g. Bauwens et al. (2006) for a review. In our applications we will focus on the popular BEKK specification of Engle and Kroner (1995), which in its simplest form can be written as

$$H_t = CC' + A'e_{t-1}e_{t-1}'A + B'H_{t-1}B, \tag{2}$$

where C is a lower triangular matrix and A and B are $N \times N$ parameter matrices. The BEKK model has the convenient feature to ensure positive definite covariance matrices under mild regularity and initial conditions. Conditions for ergodicity and stationarity have been derived by Engle and Kroner (1995) and Boussama et al. (2011). Higher order terms can be added but are rarely used in practice. The parameters of the model are stacked into a vector $\theta = (\ \mathrm{vech}(C)',\ \mathrm{vec}(A)',\ \mathrm{vec}(B)')'$, and we can emphasize the dependence of H_t on θ by writing $H_t(\theta)$.

We note also that there is some discussion about the caveats of a BEKK representation with full parameter matrices A and B, as opposed to a diagonal BEKK model, in terms of the existence of an underlying stochastic process, regularity conditions, invertibility and asymptotic properties, see in particular McAleer (2019). Moreover, in higher dimensions, further restrictions may be necessary to avoid the curse of dimensionality, i.e., vastly increasing numbers of parameters, for example by specifying A and B as scalar parameters, as advocated by Caporin and McAleer (2012). The diagonal form of a BEKK model has been advocated e.g. by Chang et al. (2017) to test for (co-)volatility spillover, as this model has a well-defined underlying stochastic process and asymptotic theory.

To fully specify the dynamics of e_t and not only its conditional second moment structure, one often uses ad-hoc decompositions of H_t such as, for example, the symmetric eigenvalue decomposition or Cholesky factors. Denoting by $H_t^{1/2}$ the matrix square root of H_t obtained by eigenvalue decomposition, the MGARCH process reads as

$$e_t = H_t^{1/2}\xi_t,\ \ \xi_t \sim iid(0, I_N). \tag{3}$$

Viewing the expression in (3) as a structural scheme, the jth column of $H_t^{1/2}$ formalizes how single orthogonalized shocks ξ_{jt} in ξ_t affect the returns (or their reduced form residuals) collected in r_t (e_t). Similarly, the ith row of $H_t^{1/2}$ unravels the contribution of each shock in ξ_t to uncertainty/volatility received by a single market r_{it} (e_{it}). Importantly, the eigenvalue decomposition implies for each market a symmetry of cross market volatility reception and transmission.

On a priori grounds one might argue that the implication of symmetry lacks economic justification in many contexts. For instance, it appears intuitive to expect a wedge between patterns of market specific volatility transmission and reception if the considered markets differ considerably in terms of economic importance or market valuation. Alternatively, a Cholesky decomposition of H_t is conditional on the presumed ordering of variables. However, an analyst might warrant a more flexible model framework which could also nest as special cases particular a-priori schemes.

Generalizing the exposition in (3), the identification problem in MGARCH models can be made more explicit in terms of a structural transmission scheme as

$$e_t = W_t\xi_t, \tag{4}$$

where W_t is an $N \times N$, $\mathcal{F}_{t-1}$-measurable matrix such that $\mathrm{Var}[e_t|\mathcal{F}_{t-1}] = W_tW_t' = H_t$. Note that this decomposition is not unique without further assumptions. For instance, under the assumption of conditional normality all possible decompositions of the form $H_t = W_tW_t'$ are observationally equivalent, and one would have to rely on external non-data based information to identify a structural model.

We now introduce the following parameterization of W_t:

$$W_t = H_t^{1/2}R_\delta, \tag{5}$$

where R_δ is a rotation matrix, parameterized by a vector δ, such that $R_\delta R_\delta' = I_N$. A special case is $R_\delta = I_N$, which corresponds to the symmetric model in (3). Another special case is the Cholesky factorization, which can be recovered by choosing the rotation angles δ such that W_t becomes triangular. To be more specific, R_δ is parameterized as the product of distinct forms of Givens rotation matrices where the elements of δ, denoted δ_i, $0 \le \delta_i < \pi$, are rotation angles. For a model of dimension N, δ comprises $N(N-1)/2$ rotation angles. For instance, in the case of $N = 3$,

$$R_\delta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\delta_1) & -\sin(\delta_1) \\ 0 & \sin(\delta_1) & \cos(\delta_1) \end{bmatrix} \times \begin{bmatrix} \cos(\delta_2) & 0 & -\sin(\delta_2) \\ 0 & 1 & 0 \\ \sin(\delta_2) & 0 & \cos(\delta_2) \end{bmatrix} \times \begin{bmatrix} \cos(\delta_3) & -\sin(\delta_3) & 0 \\ \sin(\delta_3) & \cos(\delta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

Obviously, the matrices W_t depend on both the reduced form MGARCH parameters $\theta \in \Theta$ and the rotation angles δ, i.e., $W_t(\theta, \delta) = H_t(\theta)^{1/2}R_\delta$. Consequently, model implied structural shocks read as $\xi_t(\theta, \delta) = W_t^{-1}(\theta, \delta)e_t$. While the vector of coefficients θ can be uniquely determined by means of QML estimation, δ lacks identification without additional information. Henceforth, the specification of W_t focuses on the most appropriate choice of R_δ for given (estimates of) θ and $\{H_t\}$.

The recent literature on identification in structural VAR models has shown that data-based identification of structural relations offers unique solutions if structural shocks are non-Gaussian and independently distributed (Lanne et al., 2017; Gouriéroux et al., 2017). In empirical applications of volatility models the supposition of Gaussian innovations is regularly confirmed as overly restrictive and models incorporating skewed or leptokurtic innovations have been put forward. This motivates our approach to exploit the particular structure of third and fourth order co-moments that is implied by independently distributed non-Gaussian model innovations.

2.1. Moment based identification

Under non-normality, the core idea to identify the structural MGARCH model is to start with an a-priori decomposition $H_t(\theta)^{1/2}$, obtaining standardized reduced form errors $e_t^{std} := H_t(\theta)^{-1/2}e_t$. In a second step, this factor is subjected to systematic rotations R_δ to obtain a specific (rotated) matrix $W_t(\theta, \delta) = H_t(\theta)^{1/2}R_\delta$ that implies innovations which are best in matching a set of co-moment conditions that apply to independent shocks.

Before characterizing the identification technique, we need to impose some assumptions on the vector of innovations $\xi_t(\theta, \delta) = W_t^{-1}(\theta, \delta)e_t$. In the following, we distinguish between the vector of independent structural error terms which is related to rotation angle δ_0 by $\xi_t := \xi_t^{\delta_0} = \xi_t(\theta, \delta_0)$ and the general vector of structural shocks $\xi_t^{\delta} = \xi_t(\theta, \delta)$ which depends on rotation angle δ and is not necessarily independent. The following assumptions are imposed on ξ_t.

Assumption 1. ξ_t, $t = 1, \ldots, T$, is an N-dimensional vector with the following properties:

(i) At most one of the components of the random vector ξ_t has a normal distribution.
(ii) The components ξ_{it} are mutually independent with $E[\xi_{it}] = 0$ and $\mathrm{Var}[\xi_t] = I_N$.
(iii) For some $\epsilon > 0$, $\mathrm{E}|\xi_{kt}|^{6+\epsilon} < \infty$, $k = 1, \ldots, N$.

In the following, we discuss moment properties of independent shocks in order to propose a diagnostic statistic in Section 2.2 to identify them based on higher order moments. We first introduce some notation and define respectively the marginal skewness and kurtosis of innovations, $m_i^{(3)} := E[\xi_{it}^3]$ and $m_i^{(4)} := E[\xi_{it}^4]$, $i = 1, \dots, N$. Furthermore, let $m^{(g)} := (m_1^{(g)}, \dots, m_N^{(g)})'$ denote N-dimensional moment vectors for $g \in \mathbb{N}$. In particular, the skewness vector corresponds to $m^{(3)}$, the kurtosis vector to $m^{(4)}$, and the second order moment $m^{(2)}$ is unity by definition. The operator $\operatorname{diag}(a)$ stacks an n-dimensional vector a into a diagonal $n \times n$ matrix.

Let us now define the matrices of third and fourth order cross-products of ξ_t as

$$\Phi_t := (\xi_t \otimes \xi_t)\xi_t' \quad \text{and} \quad \Psi_t := \xi_t\xi_t' \otimes \xi_t\xi_t'$$

and corresponding expectations,

$$\Phi := E[\Phi_t] = \operatorname{diag}(m^{(3)})L_N \tag{7}$$

$$\Psi := E[\Psi_t] = 2D_N D_N^+ + \operatorname{vec}(I_N)\operatorname{vec}(I_N)' + \operatorname{diag}(\operatorname{vec}(\operatorname{diag}(m^{(4)} - 3\iota_N))), \tag{8}$$

where L_N is the unique $N^2 \times N$ matrix defined by the property $\operatorname{diag}(A) = L_N' \operatorname{vec}(A)$ for any $N \times N$ matrix A, D_N is the duplication matrix, D_N^+ its generalized inverse, and $\iota_N := (1, 1, \dots, 1)'$ an N-dimensional vector of ones. A derivation of these expressions is straightforward following the lines of Proposition 5.3 of Hafner and Rombouts (2007). An elementwise characterization of the matrix Ψ was derived in Fengler and Herwartz (2018).

Furthermore, define $\boldsymbol{\phi}_t$ as the vector containing the non-redundant elements of $\operatorname{vec}(\Phi_t)$, except for the terms ξ_{it}^3, $i = 1, \dots, N$. That is, $\boldsymbol{\phi}_t$ contains all cross-products of the type $\xi_{it}^2\xi_{jt}$ and $\xi_{it}\xi_{jt}\xi_{kt}$, so that $\boldsymbol{\phi}_t$ is of dimension $q_\phi := N(N-1)(N+4)/6$. Similarly, define the vector $\boldsymbol{\psi}_t$ as the vector containing the unique elements of $\operatorname{vec}(\Psi_t)$, except for the terms ξ_{it}^4, $i = 1, \dots, N$. That is, $\boldsymbol{\psi}_t$ contains all cross-products of the type $\xi_{it}^3\xi_{jt}$, $\xi_{it}^2\xi_{jt}\xi_{kt}$, $\xi_{it}^2\xi_{jt}^2$ and $\xi_{it}\xi_{jt}\xi_{kt}\xi_{lt}$, and it can easily be checked that $\boldsymbol{\psi}_t$ is of dimension $q_\psi := N(N-1)(N^2+7N+18)/24$. For example, if $N = 4$ as in our application, then $q_\phi = 16$ and $q_\psi = 31$. Note that $q_\phi = O(N^3)$ and $q_\psi = O(N^4)$. Finally, let $\boldsymbol{\phi} := \mathrm{E}[\boldsymbol{\phi}_t]$ and $\boldsymbol{\psi} := \mathrm{E}[\boldsymbol{\psi}_t]$ be the vectors of expectations of third and fourth order cross-products of innovations.

The variance–covariance matrix of $\boldsymbol{\phi}_t$ can be obtained as

$$\operatorname{vech}\left(\operatorname{Var}(\boldsymbol{\phi}_t)\right) = C_1 \begin{pmatrix} 0 \\ 1 \\ \operatorname{vecl}(m^{(3)}m^{(3)'}) \\ m^{(4)} \end{pmatrix}$$

where $\operatorname{vecl}(\cdot)$ denotes the operator that stacks the lower triangular part of a matrix, excluding the diagonal, into a column vector, and C_1 is a $q_\phi(q_\phi+1)/2 \times N(N+1)/2+2$-dimensional binary selection matrix, that is, each row contains exactly one entry of unity, and zeros elsewhere. For example, in the bivariate case ($N = 2$), we have $\boldsymbol{\phi}_t = (\xi_{1t}^2\xi_{2t}, \xi_{1t}\xi_{2t}^2)'$ and

$$\operatorname{Var}(\boldsymbol{\phi}_t) = \begin{pmatrix} m_1^{(4)} & m_1^{(3)}m_2^{(3)} \\ m_1^{(3)}m_2^{(3)} & m_2^{(4)} \end{pmatrix}$$

such that C_1 has dimension 3×5 and is given by

$$C_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Similarly, the variance–covariance matrix of $\boldsymbol{\psi}_t$ is given by

$$\operatorname{vech}\left(\operatorname{Var}(\boldsymbol{\psi}_t)\right) = C_2 \begin{pmatrix} 0 \\ 1 \\ m^{(4)} \\ \operatorname{vecl}(m^{(4)}m^{(4)'}) - \iota_{N(N-1)/2} \\ m^{(6)} \end{pmatrix}$$

where C_2 is a $q_\psi(q_\psi+1)/2 \times (2(N+1) + N(N-1)/2)$-dimensional binary selection matrix.

To construct the covariance-matrix of the vector $(\boldsymbol{\phi}_t', \boldsymbol{\psi}_t')'$, it remains to calculate the covariance between $\boldsymbol{\phi}_t$ and $\boldsymbol{\psi}_t$, which is obtained as

$$\operatorname{vec}\left(\operatorname{Cov}(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t)\right) = C_3 \begin{pmatrix} 0 \\ m^{(3)} \\ m^{(5)} \\ \operatorname{vec}(m^{(3)}m^{(4)'}) \end{pmatrix},$$

where C_3 is a $q_\phi q_\psi \times (N+1)^2$ binary selection matrix.

We now stack both vectors into the q-dimensional vector $S_t = \left(\boldsymbol{\phi}_t', \boldsymbol{\psi}_t'\right)'$ (e.g. $q = 47$ for $N = 4$), and $q = q_\phi + q_\psi$. All variances and covariances of $\boldsymbol{\phi}_t$ and $\boldsymbol{\psi}_t$ are then used to construct $\Sigma := \operatorname{Var}((\boldsymbol{\phi}_t', \boldsymbol{\psi}_t')')$. Furthermore, $\hat{\boldsymbol{\phi}}_t$ and $\hat{\boldsymbol{\psi}}_t$ ($\hat{S}_t$)

define the respective estimators obtained from $\hat{\xi}_t = \xi_t(\hat{\theta}, \delta_0)$. We discuss consistency of these estimators in the following Section 2.2.

2.2. A diagnostic and its asymptotic properties

The co-moment structure detailed above holds for the case of independent structural shocks ξ_t. Alternative choices of δ in (5) yield shocks ξ_t^δ with non-trivial third and fourth order moments. We suggest to select the vector of rotation angles in a way such that implied empirical third and fourth order co-moments of ξ_t^δ are closest to the moments of independent shocks ξ_t.

For the vector $E[S_t^\delta]$ of stacked theoretical third and fourth order moments of ξ_t^δ, let $\bar{S}_T^\delta = \frac{1}{T}\sum_{t=1}^T S_t^\delta = \left(\bar{\boldsymbol{\phi}}_T^{\delta'}, \bar{\boldsymbol{\psi}}_T^{\delta'}\right)'$ denote the empirical counterpart. To discriminate between distinct choices of the rotation angle δ we consider the statistic

$$\lambda_\delta^T = T(\bar{S}_T^\delta - E[S_t])'\Sigma_\delta^{-1}(\bar{S}_T^\delta - E[S_t]), \tag{9}$$

where the covariance matrix Σ_δ collects the second order (co-)moments of the elements in $S_t^\delta = \left(\boldsymbol{\phi}_t^{\delta'}, \boldsymbol{\psi}_t^{\delta'}\right)'$. The entries in Σ_δ are calculated under the assumption of independence with respect to rotation angle δ. Similarly, the expectation vector $E[S_t]$ comprises $E[\boldsymbol{\phi}_t]$ and $E[\boldsymbol{\psi}_t]$ defined under independence.[1]

A feasible evaluation of (9) relies on estimated rather than true innovation estimates $\hat{\xi}_t^\delta = W_t(\hat{\theta}, \delta)^{-1}e_t$, and the corresponding statistic will be denoted

$$\hat{\lambda}_\delta^T = T(\bar{\hat{S}}_T - E[S_t])'\widehat{\Sigma}^{-1}(\bar{\hat{S}}_T - E[S_t]). \tag{10}$$

If independence is violated, and under non-sphericity, the third and fourth order moment conditions provided in Section 2.1 are (partly) violated such that $\hat{\lambda}_\delta^T$ in (10) is expected to diverge for increasing T. Hence, minimizing $\hat{\lambda}_\delta^T$ provides a particular decomposition of H_t closest to independence in a third and fourth order moment sense. Using the statistic in (10), the identified structural model reads as

$$\widehat{W}_t = W_t(\hat{\theta}, \hat{\delta}), \text{ with } \hat{\delta} = \text{argmin}_\delta\{\hat{\lambda}_\delta^T \,|\, \xi_t = W_t(\hat{\theta}, \delta)^{-1}e_t\}. \tag{11}$$

Remark 1. Simplifications are possible by assuming that the marginal distributions of the shocks ξ_{it} are symmetric, as the deviations of the Gaussian distribution are often due to excess kurtosis rather than skewness. Only a few higher order moments would have to be estimated. Moreover, cross products of the form $\xi_{it}^3\xi_{jt}$ are removed from $\boldsymbol{\psi}_t$, S_t and $\bar{S}_T$. Accordingly, the asymptotics of (9) can be established with the weaker assumption of finite moments of order $4+\epsilon$, $\epsilon > 0$.

Remark 2. As an alternative identification scheme, one might consider the minimization of a non-parametric independence diagnostic such as, for example, the Cramér-von-Mises distance of Genest et al. (2007). However, such rank-based statistics typically suffer from low power compared to a moment-based criterion as λ_δ^T in (9). This is confirmed in our empirical analysis of Section 4.1, where we provide some details on the identification based on a Cramér-von-Mises distance.

For the next theorem we will use an additional assumption.

Assumption 2. The estimator of the reduced form parameter θ is such that $\hat{\theta} \xrightarrow{p} \theta$ and $\sqrt{T}(\hat{\theta} - \theta)$ has an asymptotic normal distribution.

For the BEKK model, extensive results are available for the quasi maximum likelihood estimator (QMLE), giving sufficient conditions under which Assumption 2 holds, starting with Comte and Lieberman (2003). For example, Hafner and Preminger (2009) show consistency under the condition of finite second order moments of the (non-Gaussian) innovations ξ_t, and asymptotic normality given sixth order moments of the MGARCH process. These assumptions, however, are not easily testable, and more difficult to substantiate than those given in Ling and McAleer (2003). We now have the following theorem.

Theorem 1. *Under Assumptions 1 and 2, as $T \to \infty$, $\hat{\lambda}_\delta^T \xrightarrow{d} \chi^2(q)$ and $\hat{\delta} \xrightarrow{p} \delta_0$.*

Proof. See Appendix.

There is a connection of our approach with classical independent component analysis (ICA). The statistic λ_δ^T minimizes the sum of squared standardized elements of the matrices of third and fourth order co-moments, except for the marginal

[1] For the class of spherical distributions, of which the standard normal is a special case, the above moment conditions hold as well and they are invariant with respect to orthogonal rotations. However, the standard normal is the only member of this class with independent components, so that spherical distributions are excluded by Assumption 1(i)–(ii). In practice, observing a statistic λ_δ^T that is a non-trivial function of δ is an indicator of non-sphericity.

skewness and kurtosis. Alternatively, one could maximize the information contained in the marginal skewness and kurtosis. This technique is used in a classical ICA algorithm, the so-called joint approximate diagonalization of eigenmatrices (JADE). Following, for instance, Miettinen et al. (2015), the described identification technique can be formulated as the counterpart to standard ICA methodologies. While we are interested in minimizing the standardized cross-moments, JADE maximizes the standardized marginal moments. As the overall entropy is constant, both approaches yield the same results. The construction of the third and fourth moments criterion is similar to that of entropy and negentropy (cf. equation (14.2) and (14.3) of Hyvärinen et al. (2001)). In this sense, it serves as a measure of mutual information and non-Gaussianity, respectively. Minimizing/maximizing the respective criteria equivalently yields independent components under given conditions. More details are provided in Hyvärinen et al. (2001).

2.3. Simulation study

The purpose of the simulation study is to uncover the scope of distance measures as defined in (9) for identifying the structural parameters in W_t. Owing to consistency of QML estimation, we (mostly) discard QML estimation steps in the simulations, and evaluate model selection outcomes under the assumption that $\{H_t\}_{t=1}^T$ is known to the analyst. Throughout, we set $\mu_t = 0$. We investigate model performance by means of trivariate DGPs ($N = 3$, benchmark experiments) and, alternatively, set model dimensions to $N = 2$ and $N = 4$. Typical reduced form parameters of the BEKK model in (2) are denoted $C^{(N)}, A^{(N)}$ and $B^{(N)}$ to account for model dimensionality. With regard to the loss statistic λ, we consider mostly vectors collecting fourth order co-moments (i.e. $\bar{\boldsymbol{\psi}}$). Robustness analyses focus on the informational content of using third order cross moments (i.e. $\bar{\boldsymbol{\phi}}$) and third and fourth order cross moments jointly $((\bar{\boldsymbol{\phi}}', \bar{\boldsymbol{\psi}}')')$.

For benchmarking purposes we use trivariate DGPs, such that $C^{(3)} = C, A^{(3)} = A$ and $B^{(3)} = B$, where

$$C = \begin{pmatrix} 4.00 & 0.00 & 0.00 \\ 14.5 & 2.00 & 0.00 \\ 25.0 & -8.50 & 2.50 \end{pmatrix} /1000,\ A = \begin{pmatrix} .14 & .05 & .05 \\ -.05 & .14 & .05 \\ -.03 & .05 & .14 \end{pmatrix},\ B = \begin{pmatrix} .96 & -.06 & .02 \\ .04 & .96 & .02 \\ .04 & .02 & .96 \end{pmatrix}.$$

Moreover, we set for the case of bivariate systems the parameters equal to the upper left block of the benchmark parameters, i.e. $C^{(2)} = C[1:2, 1:2], A^{(3)} = A[1:2, 1:2]$ and $B^{(3)} = B[1:2, 1:2]$. For bivariate models we also evaluate the performance of the suggested identification scheme if applied to model residuals determined subsequent to QML-estimation of the BEKK model. For the case of four-dimensional models we employ the parameterization

$$C^{(4)} = \begin{pmatrix} C & 0 \\ c_{21} & 0.0025 \end{pmatrix}, A^{(4)} = \begin{pmatrix} A & a_{12} \\ a_{21} & 0.14 \end{pmatrix} \text{ and } B^{(4)} = \begin{pmatrix} B & b_{12} \\ b_{21} & 0.96 \end{pmatrix},$$

where $c_{21} = (10, 4, -5)/1000$, $a_{12} = (0.05, 0.05, 0.03)'$, $a_{21} = (0.03, 0.05, -0.03)$, $b_{12} = (0.02, 0.02, 0.02)'$ and $b_{21} = (-0.02, -0.04, 0.02)$.

To generate (excess) returns e_t from the structural model, the transmission matrix W_t is determined by means of a rotation of $H_t^{1/2}$ as $W_t = H_t^{1/2} R_\delta$, where R_δ is of the form in (6) and rotation angles are $\delta^{(3)} = \delta, \delta = (\delta_1, \delta_2, \delta_3)' = (.10,\ .25,\ .40)\pi$. Furthermore we set $\delta^{(2)} = 0.25\pi$ in the bivariate case and $\delta^{(4)} = (\delta', \delta')'$ in case of $N = 4$. The stochastic model components are drawn from standardized Student-t innovations ξ_t with $\nu = 5,\ 10,\ 15,\ 30$ and 100 degrees of freedom. We also provide results for the unidentified Gaussian model. Sample sizes are $T + 100 = 1100, 2100, 4100$ and 8100, before discarding the first 100 observations to immunize the analysis from initialization effects. Each experiment is replicated 10,000 times.

The simulated data $\{e_t, H_t\}_{t=1}^T$ are subjected to a structural analysis presuming rival specifications of the decomposition in (5). On the one hand, a candidate decomposition matrix is chosen in accordance with the true model, $W_t^{(0)} = W_t$, for which one would expect a minimal loss statistic λ_δ^T. In addition, seven alternative (and false) covariance decomposition schemes are considered, namely

$$\begin{aligned} W_t^{(0)} &= H_t^{1/2} R_\delta, \ \text{ true model, clockwise rotation of } H_t^{1/2}, \\ W_t^{(1)} &= H_t^{1/2}, \ \text{unrotated model, eigenvalue decomposition}, \\ W_t^{(\overline{q})} &= H_t^{1/2} R_{\overline{q}\delta}, \overline{q} = 1.010,\ 1.025,\ 1.050 \text{ 'excess' clockwise rotations}, \\ W_t^{(\underline{q})} &= H_t^{1/2} R_{\underline{q}\delta}, \underline{q} = (\overline{q})^{-1} \text{ 'insufficient' clockwise rotations}. \end{aligned}$$

Table 1 reports results for benchmark simulation experiments ($N = 3$, $\bar{\boldsymbol{\psi}}$). Contrasting loss measures from the true decomposition against counterparts obtained from the symmetric decomposition is supportive for the true model with selection frequencies close to unity under various scenarios with either sizeable deviations from the Gaussian model ($\nu = 5$, $T = 1000$) or sufficiently rich sample information. For a standardized Student-t distribution with 30 degrees of freedom, samples as large as $T = 8000$ are informative to rule out the symmetric model against the true structural model in 92% of all Monte Carlo replications. In the Gaussian case (i.e. $\nu \to \infty$), the outcomes of model selection are purely random, with selection frequencies of 50% for both alternatives.

Comparing the loss statistic for the true model with those derived from over- and under-rotations shows that minimizing the loss statistic yields a local minimum in the neighbourhood of the true structural model. With $T = 1000$

Table 1
Benchmark simulation results ($N = 3$, $\bar{\psi}$). The table documents estimated frequencies of particular identification results obtained from 10,000 Monte Carlo replications. Columns entitled '$F(W^{(0)})$' show the frequency of a smaller diagnostic λ_δ^T in (9) obtained for the true rotated model $W^{(0)}$ in comparison with the unrotated model $W^{(1)}$. Columns entitled with entries for 'q' document how often the diagnostic λ_δ^T is smaller than both $\lambda_{\delta q}^T$ and $\lambda_{\delta/q}^T$. Distinct choices of q indicate varying degrees of over- or under-rotations of the symmetric model. Sample sizes and degrees of freedom of standardized Student-t distributed shocks are T and ν, respectively, with $\nu = \infty$ indicating the Gaussian case.

ν	$F(W^{(0)})$	$q = 1.010$	1.025	1.05	$W^{(0)}$ vs. $W^{(1)}$	$q = 1.010$	1.025	1.05
	$T = 1000$				$T = 2000$			
5	0.998	.103	.256	.480	1.000	.146	.347	.625
10	0.975	.060	.145	.284	0.998	.089	.214	.412
15	0.903	.041	.100	.196	0.976	.059	.148	.291
30	0.721	.025	.059	.109	0.817	.030	.078	.152
100	0.560	.015	.039	.071	0.576	.015	.038	.076
∞	0.505	.012	.032	.065	0.497	.012	.033	.065
	$T = 4000$				$T = 8000$			
5	1.000	.194	.453	.770	1.000	.260	.597	.888
10	1.000	.122	.293	.549	1.000	.179	.414	.721
15	0.999	.083	.207	.403	1.000	.128	.304	.557
30	0.921	.046	.107	.211	0.982	.064	.157	.308
100	0.612	.019	.045	.086	0.663	.019	.052	.099
∞	0.504	.013	.033	.067	0.495	.011	.031	.065

Table 2
Further simulation results. Simulation scenarios are indicated with regard to the dimension of the DGP (N) and the co-moments ('cmom') employed for identification ($\bar{\psi}$ or $\bar{\phi}$ or both). The column labelled 'QML' indicates if simulation results have been obtained from known ('No') or estimated ('Yes') covariances. For further notes see Table 1.

N	QML	ν	cmom	$F(W^{(0)})$	$q = 1.010$	1.025	1.05	$F(W^{(0)})$	$q = 1.010$	1.025	1.05
				$T = 1000$				$T = 2000$			
3	No	10	$(\bar{\phi}', \bar{\psi}')'$	0.970	.052	.141	.272	0.998	.085	.205	.391
		30		0.711	.020	.054	.108	0.793	.029	.073	.144
3	No	10	$\bar{\phi}$	0.740	.020	.047	.094	0.758	.022	.048	.096
		30		0.570	.014	.031	.057	0.575	.011	.028	.057
2	No	10	$\bar{\psi}$	0.951	.029	.072	.145	0.995	.040	.100	.194
		30		0.663	.012	.027	.053	0.759	.015	.035	.070
4	No	10	$\bar{\psi}$	0.997	.100	.234	.438	1.000	.139	.337	.608
		30		0.813	.037	.089	.174	0.902	.048	.117	.230
2	Yes	10	$\bar{\psi}$	0.918	.026	.066	.128	0.987	.041	.094	.183
		30		0.439	.006	.020	.039	0.524	.010	.025	.051
				$T = 4000$				$T = 8000$			
3	No	10	$(\bar{\phi}', \bar{\psi}')'$	1.000	.116	.284	.536	1.000	.169	.412	.710
		30		0.898	.044	.103	.197	0.973	.062	.156	.297
3	No	10	$\bar{\phi}$	0.765	.022	.053	.103	0.769	.021	.052	.101
		30		0.576	.013	.030	.059	0.578	.011	.027	.057
2	No	10	$\bar{\psi}$	1.000	.056	.145	.281	1.000	.086	.208	.395
		30		0.867	.022	.055	.103	0.960	.032	.078	.151
4	No	10	$\bar{\psi}$	1.000	.202	.471	.788	1.000	.286	.638	.924
		30		0.971	.071	.171	.327	0.998	.106	.263	.476
2	Yes	10	$\bar{\psi}$	1.000	.056	.141	.271	1.000	.086	.203	.387
		30		0.670	.017	.039	.078	0.823	.025	.061	.116

observations and Student-t shocks with 15 degrees of freedom, the true decomposition leads to a loss measure smaller than those from both bounds of false rotations with $q = 1.050$ in about 20% of all Monte Carlo experiments. Pointing at the consistency of the loss minimization, this proportion increases to almost 30%, 40% and 60% in larger samples of sizes $T = 2000$, 4000 and $T = 8000$, respectively.

Table 2 documents results obtained from further simulation experiments. For space considerations we document only outcomes for DGPs generated from independent standardized Student-t distributed shocks where the degrees of freedom parameter are either $\nu = 10$ or $\nu = 30$. Complementing benchmark outcomes, the robustness analysis obtains the following results:

First, for the case of trivariate DGPs ($N = 3$) the identification outcomes are very similar if model selection is based jointly on third and fourth order co-moments, i.e. on $(\bar{\phi}', \bar{\psi}')'$. Although the true underlying innovations do not show any co-skewness, subjecting an extended vector of co-moments to identification involves only a negligible power loss.

Second, if one focuses model identification exclusively on third order moments, i.e. on $\bar{\boldsymbol{\phi}}$, then model selection suffers from weak power in comparison with benchmark outcomes for $\bar{\boldsymbol{\psi}}$. However, given that the true model innovations do not exhibit any co-skewness, it is interesting to see that the identification scheme is not completely uninformative. For instance, the true model is preferred over the symmetric covariance decomposition in significantly more than 50% of all experiments, and unsurprisingly this effect is stronger for $\nu = 10$ degrees of freedom in comparison with $\nu = 30$. Although $E[\boldsymbol{\phi}] = 0$ holds for all model rotations, the diagnostic in (10) also depends on covariance assessments. These involve even order marginal moments of ξ_t, which are underestimated in falsely rotated models. In this case, the $\hat{\lambda}_\delta^T$ statistic does not diverge but converges to a distribution different from a χ^2. As the sample size increases, the success frequencies of the true model then converge to the proportion of statistics using this distribution that are larger than the theoretical χ^2 distribution.

Third, the dimensionality of the process under scrutiny seems to be an important factor for the performance of the criterion in (10) for structural model selection. Other things equal, i.e. using $\bar{\boldsymbol{\psi}}$ for identification and setting $\nu = 10$ or $\nu = 30$, the structural model selection is more (less) successful in systems of dimension $N = 4$ ($N = 2$) in comparison with systems with benchmark dimension of $N = 3$.

For computational reasons almost all simulation results have been obtained under knowledge of the true covariances $\{H_t\}_{t=1}^T$. A fourth set of robustness results obtains from simulating identification outcomes with estimated BEKK models. In larger samples ($T > 2000$) or under marked deviations from the Gaussian model ($\nu = 10$) the performance statistics documented in Table 2 are quite similar if the identification step relies on QML estimates of the BEKK model and hence on estimated covariances $\{\widehat{H}_t\}_{t=1}^T$.

3. Portfolio risk analysis

MGARCH models have become a widespread tool for risk analysis and management of investment portfolios. In a non-Gaussian framework, alternative structural MGARCH models exhibit distinct higher order characteristics. The importance of higher order moments for risk measures has been recognized by regulatory authorities. For example, the European supervisory authorities EBA, EIOPA and ESMA have established regulatory technical standards for packaged retail and insurance-based investment products (see ESA (2016)) that include the use of skewness and kurtosis via Cornish–Fisher expansions to evaluate market risk measures.

In the following, we first show that the choice of the structural model has non-trivial consequences for the conditional kurtosis of portfolio returns. We then discuss two implications for risk management: First, in the context of approximations of tail risk measures such as Value-at-Risk and expected shortfall, and second, for measures of variability of squared portfolio returns around their conditional expectation.

Henceforth, let a_t denote an $\mathcal{F}_{t-1}$-measurable, N-dimensional vector of portfolio weights and define portfolio returns as $\tau_t := a_t' e_t$, where for notational simplicity we set the conditional mean to zero. The conditional portfolio variance is given by $\sigma_t^2 = \text{Var}[\tau_t|\mathcal{F}_{t-1}] = a_t' H_t a_t$. In the univariate case, the conditional kurtosis of a GARCH process with i.i.d. innovations is a constant, given by the kurtosis of the innovations. In the multivariate case however, as we will see in the following, the conditional kurtosis of portfolio returns is in general no longer constant and, in particular, is not invariant with respect to orthogonal rotations of the innovation vector.

Theorem 2. *Under Assumption 1, the conditional kurtosis of portfolio returns is a non-trivial function of δ and given by*

$$\varsigma_t(\delta) := E[\tau_t^4|\mathcal{F}_{t-1}] = vec(a_t a_t')'(W_t \otimes W_t)\Psi(W_t' \otimes W_t')vec(a_t a_t'). \tag{12}$$

where $W_t = W_t(\theta, \delta)$, and Ψ is defined in (8).

Proof. See Appendix.

The conditional kurtosis coefficient is defined as $\kappa_t(\delta) = \varsigma_t(\delta)/\sigma_t^4$, which can be compared with the corresponding coefficient retrieved from the symmetric decomposition $H_t^{1/2}$, i.e. $\kappa_t(0)$. We provide a numerical example.

Example 1. Consider the special case $N = 2$, equally weighted portfolio $a_t = (1/2, 1/2)'$, unit conditional variances, conditional correlation given by ρ_t, $|\rho_t| < 1$, and symmetric marginal distributions of ξ_{1t} and ξ_{2t}, having identical marginal kurtosis coefficient k. Then, straightforward calculations show that $\kappa_t(\delta) = k + (3 - k)\cos^2(2\delta)/2$, which is a non-trivial function of δ provided that $k \neq 3$, while $\kappa_t(0) = (k + 3)/2$. Note that, in this special case, $\kappa_t(\delta)$ is constant as it does not depend on ρ_t. Note also that, if $k = 3$ as for the case of a normal distribution, then $\kappa_t(\delta) = 3$, independent of δ, which confirms the impossibility to identify δ based on kurtosis measures in the Gaussian case.

Therefore, identifying structural innovations using the λ_δ statistic of the previous section will have a non-negligible impact on the conditional kurtosis of portfolio returns. As we will show, this affects estimation uncertainty of common risk measures. Moreover, it could be used for kurtosis diversification of portfolios that are sensitive to higher order risk, see e.g. Lassance et al. (2019).

3.1. First order risk

The main measures for evaluating market risk, recommended by the Basel committee of Banking Supervision, see e.g. Chapter 2 of McNeil et al. (2nd edition, 2016), are the Value-at-Risk (VaR) and the expected shortfall (ES). In a non-Gaussian context VaR and ES may depend on higher order moments such that the choice between a structural or symmetric model for the decomposition of H_t is not innocuous for first order risk assessment.

Our approach consists of simulating risk measures for alternative models. At each point in time t, conditional on the information set $\mathcal{F}_{t-1}$, we simulate a very large number (i.e. 10^6) of portfolio returns τ_t using independent bootstrap draws from model implied innovations ξ_t, and then obtain the α-quantile of their empirical distribution, which will give an approximation of the conditional VaR at time t.[2] Formally,

$$\mathrm{VaR}_{\alpha t}(\delta) = -F_{\alpha t}^{-1}(\delta), \tag{13}$$

where $F_{\cdot t}(\delta)$ is the conditional distribution function of portfolio returns as implied by the structural model parameter δ. The nominal probability level α is set alternatively to $\alpha = .010, .025, .050, .100, .250$.

In the same vein, we can simulate the conditional ES, i.e.,

$$\mathrm{ES}_{\alpha t}(\delta) = -\mathbb{E}_\delta\left[\tau_t \mid (\tau_t < -\mathrm{VaR}_{\alpha t}(\delta)), \mathcal{F}_{t-1}\right], \tag{14}$$

where the expectation on the right hand side is taken with respect to the simulated distribution of portfolio returns. To highlight the role of tail events for the determination of ES, one might also consider the expected excess shortfall, i.e.

$$\mathrm{EES}_{\alpha t}(\delta) = \mathrm{ES}_{\alpha t}(\delta) - \mathrm{VaR}_{\alpha t}(\delta). \tag{15}$$

In analysing VaR exceedances, however, we do not only consider their average but also address the extent to which the structural MGARCH models are useful in managing their distribution. For this purpose, we extract from the simulated return distributions the interquartile range of VaR exceedances which provides a statistical tool that can be contrasted against empirical patterns of exceedances. Specifically, one would expect that, on average, about 50% of all VaR exceedances fall within the model implied interquartile range. Formally, the interquartile range is given by the interval $\left[\underline{\mathrm{ES}}_{\alpha t}(\delta), \overline{\mathrm{ES}}_{\alpha t}(\delta)\right]$, where the bounds are implicitly defined by the two equations $\mathbb{P}_t\left[\tau_t > \underline{\mathrm{ES}}_{\alpha t}(\delta) | \tau_t < -\mathrm{VaR}_{\alpha t}(\delta)\right] = 0.25$ and $\mathbb{P}_t\left[\tau_t < \overline{\mathrm{ES}}_{\alpha t}(\delta) | \tau_t < -\mathrm{VaR}_{\alpha t}(\delta)\right] = 0.75$, and $\mathbb{P}_t(\cdot)$ denotes probability conditional on $\mathcal{F}_{t-1}$.

3.2. Kurtosis risk

While measures of volatility, VaR and ES convey important information for risk management, they are subject to inherent estimation uncertainty. In particular, the variation of squared returns around the conditional variances contributes to estimation uncertainty of conditional measures of VaR and ES. These fluctuations are directly related to the conditional kurtosis of portfolio returns. Accordingly, we refer to such patterns of higher order risk as kurtosis risk.

Let us define the difference between squared portfolio returns and conditional variances as

$$m_t := \tau_t^2 - \sigma_t^2, \tag{16}$$

which is a martingale difference sequence that describes surprises to portfolio risk. Such surprises can be expected to be small (large) if $\mathrm{Var}[m_t|\mathcal{F}_{t-1}]$ is small (large). A standardized measure for the conditional variance of m_t is given by

$$v_t(\delta) := \frac{\mathrm{Var}[m_t|\mathcal{F}_{t-1}]}{\sigma_t^4} = \frac{E[\tau_t^4|\mathcal{F}_{t-1}] - \sigma_t^4}{\sigma_t^4} = \kappa_t(\delta) - 1, \tag{17}$$

which is directly linked to the conditional kurtosis $\kappa_t(\delta)$. Unlike the portfolio variance σ_t^2, the risk statistic in (17) depends on the specification of the structural model, parameterized by δ. Hence, it is interesting to contrast rival structural specifications with regard to their scope in quantifying time varying patterns of kurtosis risk. Accordingly, the empirical counterpart of $v_t(\delta)$ is given by

$$\widetilde{m}_t^2 := \frac{m_t^2}{\sigma_t^4}, \tag{18}$$

which is conditionally observable.

Apart from modelling mean profiles of $\widetilde{m}_t^2$, we evaluate model accuracies in capturing the (tail) event that $\widetilde{m}_t^2$ exceeds a pre-specified critical quantile of its (simulated) distribution. Formally, we define the conditional 'kurtosis-at-risk' (KaR) as

$$\mathrm{KaR}_{\gamma t}(\delta) = G_{\gamma t}^{-1}(\delta), \tag{19}$$

where $G_{\cdot t}(\delta)$ is the conditional distribution function of $\widetilde{m}_t^2$ as implied by the structural model parameter δ, and γ is a nominal probability level.

[2] We provide a more detailed description of the resampling scheme in Section 4.1.

Similar to expected shortfall analysis of first order risk, we also consider the expected violation of the threshold given that an event of threshold violation has occurred. Formally, we define the conditional 'expected kurtosis shortfall' (EKS) as

$$\text{EKS}_{\gamma t}(\delta) = \mathbb{E}_\delta\left[\widetilde{m}_t^2 | \left(\widetilde{m}_t^2 > \text{KaR}_{\gamma t}(\delta)\right), \mathcal{F}_{t-1}\right]. \tag{20}$$

Furthermore, we also consider the distribution of risk exceedances, and in particular the interquartile range, analogous to our analysis of first order risk. For kurtosis risk, the interquartile range is defined as the interval $\left[\underline{\text{EKS}}_{\gamma t}(\delta), \overline{\text{EKS}}_{\gamma t}(\delta)\right]$, where the bounds are implicitly determined by the two equations $\mathbb{P}_t\left[\widetilde{m}_t^2 > \underline{\text{EKS}}_{\gamma t}(\delta) \mid \widetilde{m}_t^2 > \text{KaR}_{\gamma t}(\delta)\right] = 0.25$ and $\mathbb{P}_t\left[\widetilde{m}_t^2 < \overline{\text{EKS}}_{\gamma t}(\delta) \mid \widetilde{m}_t^2 > \text{KaR}_{\gamma t}(\delta)\right] = 0.75$.

4. Shock transmission in American stock markets

We illustrate the merits of the structural MGARCH approach by analysing a four dimensional system of the US and Latin American stock markets. Regarding US stock markets as important issuers of information, profiles of volatility transmission that are implied by a symmetric covariance decomposition might be critical for such a system. We study weekly real returns of four stock markets in the US (r_{1t}), Argentina (r_{2t}), Brazil (r_{3t}) and Chile (r_{4t}) for the period January 1992 to November 2007.[3] We assume time invariant conditional return expectations and subject centred real returns to QML MGARCH estimation of a BEKK model. The sample size is $T = 829$. Estimates of the BEKK parameters (with QML t-ratios in parentheses) are given as follows[4]:

$$\widehat{A} = \begin{pmatrix} 0.328 & 0.028 & -0.035 & -0.020 \\ (3.655) & (0.122) & (-0.125) & (-0.309) \\ -0.019 & 0.216 & 0.054 & 0.058 \\ (-0.565) & (3.721) & (0.941) & (3.322) \\ -0.042 & 0.023 & 0.281 & -0.033 \\ (-1.394) & (0.375) & (2.205) & (-1.074) \\ -0.033 & 0.082 & 0.042 & 0.179 \\ (-0.451) & (0.635) & (0.306) & (4.884) \end{pmatrix}, \ \widehat{B} = \begin{pmatrix} 0.914 & 0.012 & 0.015 & -0.002 \\ (9.748) & (0.086) & (0.061) & (-0.050) \\ -0.019 & 0.958 & -0.025 & -0.015 \\ (-2.005) & (46.80) & (-1.055) & (-1.095) \\ 0.033 & 0.021 & 0.949 & 0.019 \\ (1.314) & (0.742) & (16.03) & (1.081) \\ 0.034 & -0.034 & -0.060 & 0.969 \\ (0.502) & (-0.766) & (-0.596) & (89.49) \end{pmatrix};$$

Testing the joint significance of the off-diagonal parameters in $\widehat{A}$ and $\widehat{B}$ obtains a Wald-statistic of 121.75 with a respective p-value which is close to zero according to an asymptotic $\chi^2(24)$-distribution. Subsequent to QML estimation of the BEKK model we use the estimated covariance paths to extract model innovations. Minimization of squared deviations of third and fourth order cross moments of rotated shocks from their theoretical counterparts obtains estimated structural MGARCH matrices $W_t = H_t^{1/2} R_{\hat{\delta}}$, where

$$R_{\hat{\delta}} = \begin{pmatrix} 0.863 & -0.134 & -0.324 & -0.364 \\ 0.348 & 0.805 & 0.466 & 0.113 \\ 0.142 & -0.508 & 0.823 & -0.210 \\ 0.338 & -0.274 & 0.002 & 0.900 \end{pmatrix}.$$

The independence diagnostics of the symmetric decomposition ($\delta = 0,\ R_\delta = I_N$) and the rotated model are $\lambda_{\delta=0} = 126.6$ and $\lambda_{\hat{\delta}} = 39.73$, respectively. While the former is clearly significant at any conventional level, the latter implies an associated p-value of 0.76 from a supposed χ^2-distribution with 47 degrees of freedom.

In the following, we illustrate the plausibility of the identified model. Time varying elements of estimated covariance decompositions are displayed in Fig. 1. Confirming a-priori intuition, the estimated structural model $\{\widehat{W}_t\}$ implies that all Latin American markets are more affected by innovations in the US markets than under model symmetry ($\{\widehat{H}_t^{1/2}\}$). Likewise, the US market is less affected by Latin American markets under the estimated structural model. In the time dimension, volatility spillovers operating from the US to the Latin American markets are of particular strength during the period 10/03/1997 to 09/03/1999, which roughly corresponds to the great economic recession in Argentina (1998–2002), and to the Brazilian financial/currency crisis of 1998 and 1999 which culminated in the 35% devaluation of the Real (Samba effect). Reflecting the traditionally close trade relationships between Argentina and Brazil, the Argentinian stock markets received substantial volatility spillovers from the Brazilian markets in 1998 and 1999. While volatility transmission from the Brazilian markets to Chile is of similar shape during the crisis in 1998 and 1999, the structural model indicates that this is of less importance (i.e., of smaller magnitude) in comparison with transmission to the Argentinian stock market.

[3] The data are taken from Diebold and Yilmaz (2009), https://estima.com/procedures/dieboldyilmaz_ej2009.zip. Weekly returns are changes in log prices, Friday-to-Friday (or Thursday when Friday is not available) subsequently converted to real quotes by means of consumer price indices from the IMF's International Financial Statistics.

[4] Estimates of the intercept matrix C are less important and available upon request.

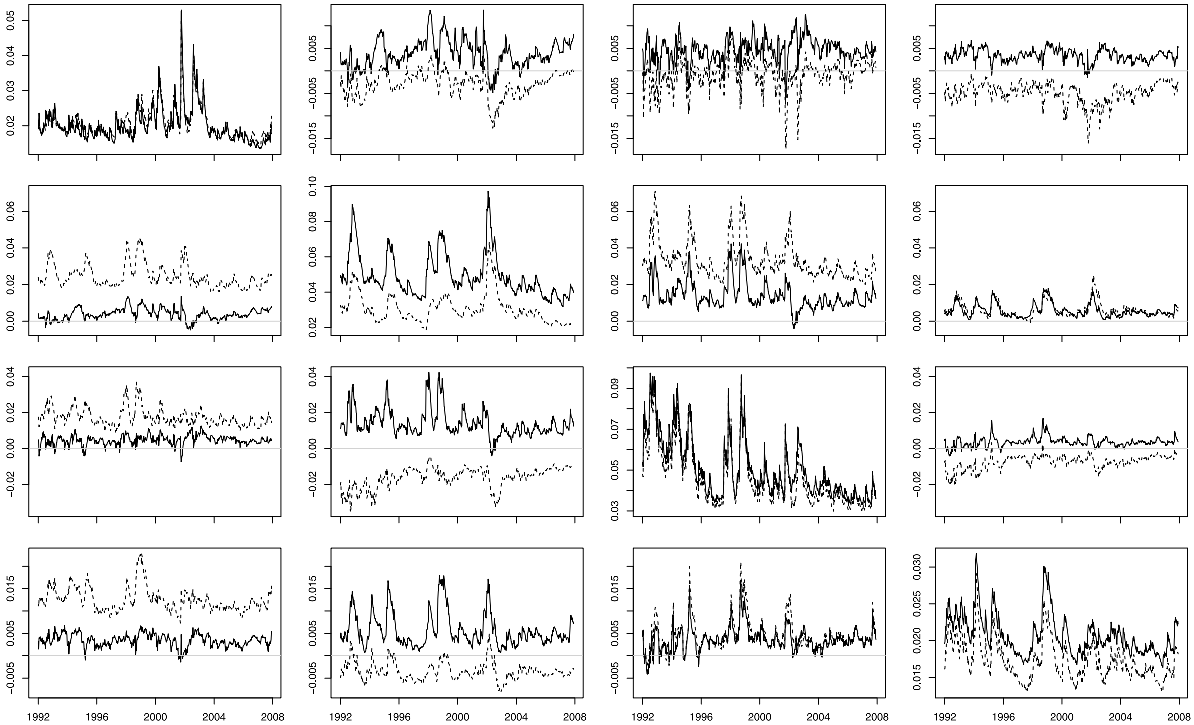

Fig. 1. Elements of alternative transmission matrices $H_t^{1/2}$ (symmetric model, solid lines) and W_t (dashed lines). The ordering of markets (rows) is US, Argentina, Brazil, Chile. Columns (rows) show volatility transmission of a given shock (reception from distinct markets).

4.1. Empirical risk measurement

To further highlight the merits of the identified model, we now discuss whether the distinction of alternative MGARCH model structures is useful for an active management of portfolio risks. For both types of risk analysis described in Section 3 we choose five alternative nominal coverage levels ranging from $\alpha = 0.01$ to 0.25 for first order risk modelling, and $\gamma = 0.75$ to 0.99 for kurtosis risk modelling.

We consider the following six stylized portfolios: Equally weighted portfolio, P_2: minimum variance portfolio, i.e. $a_t = H_t^{-1}\iota/(\iota' H_t^{-1}\iota)$ where ι is a four dimensional vector of ones, P_3: $a_{1t} = 0.5, a_{jt} = 0.5/3, j \neq 1$; P_4: $a_{2t} = 0.5, a_{jt} = 0.5/3, j \neq 2$; P_5: $a_{3t} = 0.5, a_{jt} = 0.5/3, j \neq 3$; and P_6: $a_{4t} = 0.5, a_{jt} = 0.5/3, j \neq 4$.[5] Throughout, the simulation of portfolio returns and their higher order properties is conditional on the processes of estimated covariances H_t and their alternative decompositions $\{W_t\}$ and $\{H_t^{1/2}\}$. Loss measures are obtained from averaging over samples of empirical portfolio returns $\{\tau_t\}_{t=1}^T$, whereas the conditional return features are determined at each time instance t over the number of generated bootstrap samples.

All simulations rely on $B = 10^6$ bootstrap samples. Bootstrap samples $\{\xi_t^*\}_{t=1}^T$ are drawn from estimated innovations $\{\hat{\xi}_t\}_{t=1}^T$ originating alternatively from the symmetric ($\hat{\xi}_t = \widehat{H}_t^{-1/2} e_t$) or the asymmetric model ($\hat{\xi}_t = \widehat{W}_t^{-1} e_t$). Assuming cross equation independence, we compose bootstrap vectors ξ_t^* by drawing its elements ξ_{it}^* with replacement from the marginal distributions, i.e. from $\{\hat{\xi}_{it}\}_{t=1}^T$. After their generation, bootstrap vectors $\{\xi_t^*\}_{t=1}^T$ are used to obtain samples of bootstrap returns based alternatively on the symmetric ($e_t^* = H_t^{1/2}\xi_t^*$) or the asymmetric model ($e_t^* = W_t \xi_t^*$). Let $b,\ b = 1, 2, \ldots, B$, be an index to distinguish single bootstrap draws. Then, bootstrap samples of portfolio returns read explicitly as

$$\{\{\tau_{t,b}^* = a_t' e_{t,b}^*\}_{t=1}^T\}_{b=1}^B,\ B = 10^6. \tag{21}$$

In the following we omit the replication index b for notational convenience. At each time instance, the samples in (21) are used to determine the risk statistics that have been introduced in Section 3. The results are discussed in the following.

4.2. First order risk analysis

To rank alternative model specifications we consider the following loss functions, where $I\{\}$ denotes an indicator function:

[5] In an alternative set of experiments we draw randomly 1000 vectors of positive portfolio weights. Results from these exercises are fully in line with average outcomes for the stylized portfolios and available from the authors upon request.

1. $\mathcal{L}_1 = \frac{1}{T}\sum_{t=1}^{T} I\left\{\tau_t < -\text{VaR}^{\bullet}_{\alpha t}\right\}$ (Empirical coverage of conditional VaR estimates)
2. $\mathcal{L}_2 = \frac{1}{T}\sum_{t=1}^{T}\left(|\tau_t| - \text{VaR}^{\bullet}_{\alpha t}\right) I\left\{\tau_t < -\text{VaR}^{\bullet}_{\alpha t}\right\}$ (Mean excess shortfall)
3. $\mathcal{L}_3 = \frac{1}{T}\sum_{t=1}^{T} I\left\{\tau_t \in \left[\underline{\text{ES}}^{\bullet}_{\alpha t}, \overline{\text{ES}}^{\bullet}_{\alpha t}\right]\right\}$ (Empirical coverage of interquartile ranges)

While the first loss function is common for VaR assessments, $\mathcal{L}_2$ provides an empirical counterpart of EES as defined in (15). Empirical outcomes of the interquartile range coverage $\mathcal{L}_3$ are the more favourable the closer they are to a nominal coverage of 50%. Before looking at the distinct model performances in risk modelling, it is worth to highlight that the common loss measures $\mathcal{L}_1$ and $\mathcal{L}_2$ largely depend on the conditional VaR estimate. This estimate in turn depends on the conditional portfolio variance and an unconditional quantile of the distribution of standardized returns. While the former quantity (σ_t^2) does not depend on the structural model representation at all, one might expect only small differences of alternative structural representation with regard to the latter quantity. Hence, using $H_t^{1/2}$ or W_t to determine the conditional VaR will likely obtain similar results and, accordingly, similar outcomes for $\mathcal{L}_1$ and $\mathcal{L}_2$ losses. Given the dependence of the tail properties of portfolio returns on the transmission $e_t = W_t\xi_t$, however, we expect the choice of the structural model as particularly beneficial for understanding the distributional features of returns in the tails of the distribution. In particular, we expect the choice of the structural model to be essential for $\mathcal{L}_3$ loss which summarizes risk model performance conditional on tail events.

Fig. 2 displays empirical equal weight portfolio returns (first line) joint with VaR estimates at levels $\alpha = 0.025$ (second line) and $\alpha = 0.01$ (third line). Corresponding estimates of expected excess shortfall (EES) are shown in the fourth ($\alpha = 0.025$) and fifth panel ($\alpha = 0.01$) of Fig. 2. For both nominal levels $\alpha = 0.025$ and $\alpha = 0.01$ the VaRs obtained from the asymmetric model are slightly more conservative, i.e. larger in absolute value. As it becomes apparent from the comparison of model implied EES statistics, the symmetric and the asymmetric MGARCH specification differ in particular with respect to the assignment of probabilities to tail events of (very) small portfolio returns. The EES statistics issued from the asymmetric model are throughout more conservative, i.e., larger in absolute value.

Detailed statistics for modelling portfolio risks are documented in the top panels of Table 3. Confirming a-priori expectations, in terms of the loss statistics $\mathcal{L}_1$ (coverage) and $\mathcal{L}_2$ (mean excess shortfall) and taking a joint perspective over all considered nominal coverage levels, the performance of both alternative MGARCH variants is similar. For 4 out of 5 nominal coverage levels, however, the asymmetric model leads to more favourable outcomes for $\mathcal{L}_3$. For instance, in case of the nominal level of $\alpha = 0.01$ ($\alpha = 0.025$) the empirical coverage of interquartile ranges for shortfall returns ($\mathcal{L}_3$) are 0.125 and 0.40 (0.682 and 0.476) for the symmetric and asymmetric model, respectively.

Summarizing the results for six stylized portfolios (see also Table 3), we find that the asymmetric model overall improves accuracy of risk assessments. The coverage of interquartile ranges of shortfall returns ($\mathcal{L}_3$) provides strongest support for the asymmetric model. For instance, with nominal level of $\alpha = 0.01$ and over six portfolios the total counts of shortfall returns are 49 and 34 for the symmetric and the asymmetric model, respectively. Out of these, 13 and 19 shortfall returns are covered by the respective interquartile ranges obtaining empirical coverage frequencies of 26.7% and 55.8%. While the latter cannot be distinguished statistically from the nominal 50% coverage, the former violates this level with 5% significance. On average and conditional on the six stylized portfolios, the empirical coverage of interquartile ranges determined by means of the asymmetric model is closer to the nominal 50% coverage for each choice of α except for $\alpha = 0.05$.

4.3. Empirical analysis of kurtosis risk

To rank alternative model specifications in kurtosis risk assessment we consider the following two loss functions:

1. $\mathcal{M}_1 = \frac{1}{T}\sum_{t=1}^{T} I\left\{\widetilde{m}_t^2 > \text{KaR}_{\gamma t}\right\}$ (Empirical frequency of risk excess variations)
2. $\mathcal{M}_2 = \frac{1}{T}\sum_{t=1}^{T} I\left\{\widetilde{m}_t^2 \in \left[\underline{\text{EKS}}_{\gamma t}, \overline{\text{EKS}}_{\gamma t}\right]\right\}$ (Coverage of interquartile ranges)

As it turns out, the empirical profiles of $\widetilde{m}_t^2$ exhibit a couple of strong outliers which coincide with sizeable new information entering the conditional variance processes. For instance, the strong dispersion of the empirical distribution of $\widetilde{m}_t^2$ is reflected in the fact that, on average, 8.08%, 6.03% and 3.25% of all observations characterizing equal weight portfolios are above thresholds of 3, 5 and 10, respectively. Throughout, model implied kurtosis (minus 1) statistics are markedly larger for the asymmetric model ($\upsilon_t(\hat{\delta}) \approx 3.2$ on average (with standard deviation of 0.07)) in comparison with statistics obtained from the symmetric model ($\upsilon_t(0) \approx 2.5$ (0.06)). Hence, seeing large and frequent outlying observations for $\widetilde{m}_t^2$, we may expect from the kurtosis differential that the asymmetric model has some lead in managing tail events.

In its lower panels Table 3 documents the performance of alternative approaches to quantify kurtosis risk of equal weight portfolio returns. Except for the most conservative nominal level, the empirical frequencies of excessive statistics $\widetilde{m}_t^2$ ($\mathcal{M}_1$) are throughout closer to the nominal counterparts when conditioning the analysis on the asymmetric model. Using the simulated interquartile ranges for conditional interval 'prediction' ($\mathcal{M}_2$), we find that the quantities determined by means of the asymmetric model are more reliable for several nominal levels γ. In particular, for the most conservative nominal level ($\gamma = 0.99$) both model variants yield five violations of the respective quantile. None of these violations is covered by the interquartile range determined under model symmetry, while the asymmetric specification obtains intervals capturing three out of five critical events.

Table 3
Loss statistics for risk modelling. Alternative stylized portfolios include the equal weight (P_1) and minimum variance portfolio (P_2) and portfolios assigning weights of $a_i = 0.5, a_j = 0.5/3, j \neq i$ to single markets $i = 1, 2, 3, 4$ (P_3 to P_6). Columns entitled '$H_t^{1/2}$' and 'W_t' refer to the symmetric and asymmetric structural MGARCH model, respectively. Rows labelled 'av.' document average outcomes for P1 to P6. Alternative nominal coverage levels are indicated with α (first order risk) and γ (kurtosis risk).

α	1.0%		2.5%		5.0%		10.0%		25.0%	
	$H_t^{1/2}$	W_t	$H_t^{1/2}$	W_t	$H_t^{1/2}$	W_t	$H_t^{1/2}$	W_t	$H_t^{1/2}$	W_t
	Empirical coverage ($\mathcal{L}_1$)									
P_1	0.96	0.60	2.65	2.53	4.95	4.83	9.41	9.89	21.7	23.5
P_2	1.09	0.60	2.77	2.41	3.98	3.86	8.08	7.48	23.6	24.5
P_3	1.21	0.96	2.65	2.29	5.07	4.83	9.89	9.89	22.4	23.4
P_4	0.84	0.48	2.77	2.53	5.43	5.43	9.53	9.65	21.5	22.2
P_5	1.09	0.60	2.77	2.65	4.34	4.22	9.17	9.41	23.0	23.8
P_6	1.21	0.72	2.41	2.17	4.70	4.46	9.05	9.53	21.5	22.7
av.	1.067	0.660	2.670	2.430	4.745	4.605	9.188	9.308	22.28	23.35
	Mean excess shortfall ($\mathcal{L}_2$)									
P_1	1.84	2.28	1.57	1.44	1.63	1.64	1.69	1.67	2.01	1.96
P_2	0.94	1.33	0.85	0.85	1.10	1.07	1.01	1.05	1.02	1.03
P_3	1.45	1.39	1.42	1.48	1.37	1.39	1.36	1.37	1.65	1.65
P_4	1.95	2.67	1.44	1.42	1.69	1.70	2.06	2.09	2.41	2.42
P_5	2.37	3.51	2.14	2.03	2.49	2.53	2.28	2.30	2.47	2.49
P_6	1.24	1.35	1.41	1.35	1.38	1.39	1.41	1.37	1.64	1.63
	IQR coverage ($\mathcal{L}_3$)									
P_1	.125	.400	.682	.476	.488	.550	.436	.451	.539	.523
P_2	.444	.600	.522	.600	.545	.531	.448	.484	.520	.522
P_3	.300	.500	.500	.526	.429	.500	.415	.451	.511	.500
P_4	.000	.750	.565	.333	.511	.578	.456	.475	.449	.446
P_5	.333	.600	.652	.455	.528	.514	.500	.513	.503	.518
P_6	.400	.500	.500	.611	.513	.568	.480	.468	.506	.511
av.	.267	.558	.570	.500	.502	.540	.456	.474	.505	.503
	Empirical coverage ($\mathcal{M}_1$)									
P_1	25.9	26.5	8.44	8.56	3.74	4.10	1.81	1.45	.603	.603
P_2	22.4	24.7	7.24	8.69	3.86	4.34	1.81	1.67	.603	.724
P_3	22.7	23.2	8.56	7.24	3.74	3.62	1.81	1.93	.965	.844
P_4	26.2	26.4	8.69	8.69	4.70	4.46	1.93	1.57	.603	.603
P_5	22.8	25.0	7.36	9.41	3.98	4.58	2.05	1.93	.844	.724
P_6	23.3	24.0	8.93	7.84	4.34	4.10	2.05	2.05	.844	.603
av.	23.9	25.0	8.20	8.41	4.06	4.20	1.91	1.77	0.74	0.68
	IQR coverage ($\mathcal{M}_2$)									
P_1	.447	.473	.529	.521	.613	.559	.333	.333	.000	.600
P_2	.446	.551	.450	.514	.500	.528	.667	.571	.600	.500
P_3	.468	.458	.521	.483	.484	.567	.400	.500	.125	.286
P_4	.456	.475	.556	.569	.513	.541	.313	.462	.600	.600
P_5	.481	.536	.475	.538	.606	.526	.588	.500	.429	.333
P_6	.466	.442	.514	.477	.472	.559	.353	.529	.429	.600
av.	.474	.476	.505	.524	.541	.535	.471	.458	.333	.486

Summary loss statistics for all stylized portfolios are also displayed in the lower panel of Table 3. Empirical frequencies of kurtosis risks in excess of the specified thresholds are closer to the nominal counterparts for the asymmetric models. Moreover, the coverage of the interquartile range of simulated excessive risks is generally closer to the nominal 50% coverage for the asymmetric specification. While extreme events characterized by $\gamma = 0.99$ are rare for a given portfolio, aggregating the number of respective threshold violations over all six portfolios obtains that empirical statistics $\widetilde{m}_t^2$ exceed the thresholds implied by the asymmetric and symmetric model in 35 and 36 cases, respectively. From these instances of quite strong surprises to risk (in total) 48.6% and 33.3% are covered by the model specific interquartile ranges. While the former statistic is in line with the nominal coverage, the latter differs from its nominal reference with 5% significance. Similar to outcomes for the empirical modelling of first order risk, summarizing the alternative model performance for 1000 portfolios with random compositions yields largely similar conclusions as the set of stylized portfolios.

5. Conclusions

We provide a new methodology to identify structural MGARCH models, which is shown to be consistent if the structural analysis follows consistent QML estimation and innovations are not normally distributed. Simulation based evidence confirms our theoretical results and evaluates the sensitivity to deviations from the unidentified Gaussian model.

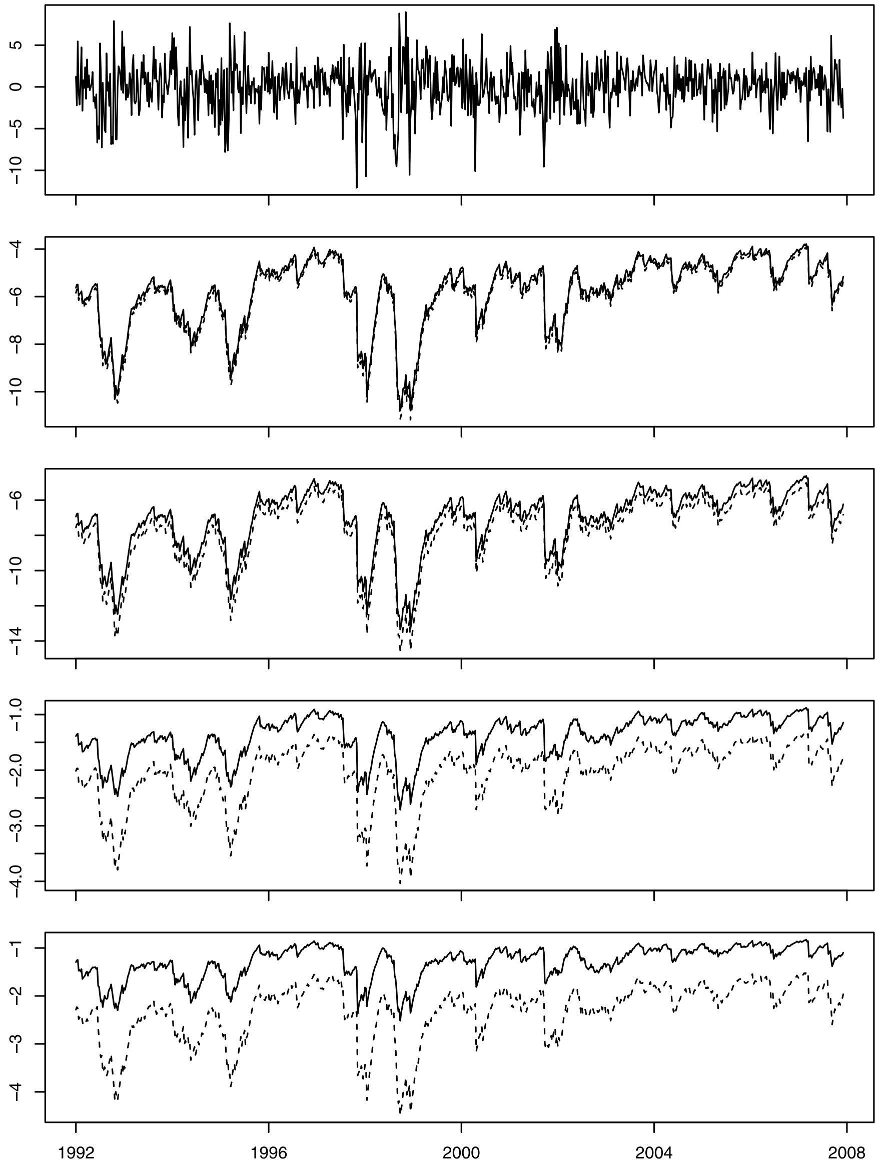

Fig. 2. First order risk modelling. First row: Equal weight portfolio returns. Second and third row: Conditional VaR estimates with 2.5% and 1% nominal coverage. Fourth and fifth row: Excess ES statistics for nominal coverage levels 2.5% and 1%, respectively. Time varying estimates for the symmetric (asymmetric) model are shown in solid (dashed) curves. To facilitate the comparison with empirical returns VaR and ES estimates are multiplied with minus unity.

Our empirical analysis provides structural insights into volatility transmission and reception characterizing a four dimensional system of US and Latin American stock markets (Argentina, Brazil, Chile). The devised structural model obtains volatility transmission patterns which are better justified in economic terms in comparison with corresponding profiles retrieved from a symmetric ad-hoc covariance decomposition. Moreover, the identified structural model turns out to be preferable in terms of an active management of first order (conditional VaR, expected shortfall) and higher order (conditional kurtosis) risk patterns inherent in portfolio returns.

There are risks we have not addressed in this paper, for example the risks of dynamic hedging strategies using optimal hedge ratios (see, e.g., Chang et al., 2017). The optimal hedge ratio only depends on the conditional second order moment structure and is, therefore, invariant with respect to orthogonal rotations. However, in the non-Gaussian case, the distribution of the hedging error is not invariant with respect to orthogonal rotations, and higher order risk measures would depend on the particular structural model. This is a promising line for future research.

Acknowledgements

We thank three anonymous reviewers and the editor Torben G. Andersen for helpful suggestions on a former version of this work. Much of the research documented in this paper has been done while the second author was visiting the Department of Economics and the Center of Environmental and Natural Economics (CENRE) of the University of Chile, Santiago de Chile. The hospitality and productive research environment of both institutions are gratefully acknowledged. We are furthermore grateful for financial support by the Belgian government (ARC 18/23-089), the Deutsche Forschungsgemeinschaft, Germany (HE2188/14-1) and the Academy of Finland (308628).

Appendix. Proofs

Proof of Theorem 1. The proof proceeds in three steps. First, keeping the rotation angles δ_0 fixed, we derive consistency of the estimates $\widehat{S}_t = (\hat{\boldsymbol{\psi}}_t', \hat{\boldsymbol{\phi}}_t')'$. Second, we show the asymptotic properties of $\lambda^T_{\delta_0}$ and $\hat{\lambda}^T_{\delta_0}$. Third, we establish consistency of $\hat{\delta}$.

Consistency of the QML estimator for θ, and application of a continuous mapping theorem imply that $W_t(\hat{\theta}, \delta_0) - W_t(\theta, \delta_0) \xrightarrow{p} 0$. Thus,

$$\hat{\xi}_t - \xi_t = \left(W_t(\hat{\theta}, \delta_0)^{-1} - W_t(\theta, \delta_0)^{-1}\right) e_t \xrightarrow{p} 0 \quad \text{and} \quad \hat{\xi}_t^2 - \xi_t^2 \xrightarrow{p} 0. \tag{22}$$

In turn, this implies $\hat{\boldsymbol{\psi}}_t - \boldsymbol{\psi}_t \xrightarrow{p} 0$ and $\hat{\boldsymbol{\phi}}_t - \boldsymbol{\phi}_t \xrightarrow{p} 0$. By a weak law of large numbers, $\bar{\boldsymbol{\phi}} \xrightarrow{p} E(\boldsymbol{\phi})$ and $\bar{\boldsymbol{\psi}} \xrightarrow{p} E(\boldsymbol{\psi})$. For example, $\frac{1}{T}\sum_{t=1}^T \xi_{ti}^2\xi_{tj}^2 \xrightarrow{p} \mathrm{E}(\xi_i^2\xi_j^2) = 1$.

Assumption 2 implies that the structural model determination by means of the statistic in (9) also applies asymptotically to estimated vectors of orthogonalized shocks $\hat{\xi}_t$, and hence $\bar{\hat{\boldsymbol{\phi}}} \xrightarrow{p} E(\boldsymbol{\phi})$ and $\bar{\hat{\boldsymbol{\psi}}} \xrightarrow{p} E(\boldsymbol{\psi})$. While we have derived convergence under the assumption of $\delta = \delta_0$, the results hold for any rotation angle $\delta \in [0, 2\pi]$. Consequently, for the statistic in (9) it follows that for any rotation angle δ, $|\lambda_\delta^T - \lambda_\delta| \xrightarrow{p} 0$. The limit λ_δ measures the distance between the expected third and fourth moments under rotation angle δ and δ_0.

Under independence, i.e. $S_T = S_T^{\delta_0}$, $\sqrt{T}(\bar{S}_T - E[S_t]) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ by the Lindeberg–Levy central limit theorem. Consequently,

$$T(\bar{S}_T - E[S_t])'\Sigma^{-1}(\bar{S}_T - E[S_t]) \xrightarrow{d} \chi^2(q). \tag{23}$$

By the consistency of $\widehat{\bar{S}}$ and Slutsky's lemma, $T(\widehat{\bar{S}}_T - E[S_t])'\widehat{\Sigma}^{-1}(\widehat{\bar{S}}_T - E[S_t]) \xrightarrow{d} \chi^2(q)$, which shows the first part of the theorem.

We next show the consistency of $\hat{\delta}$. The decomposition $H_t = W_tW_t'$ is unique for independent non-Gaussian components $\xi_{k,t}$ (Comon, 1994; Lancaster, 1954). Applying a result of Miettinen et al. (2015), the minimum of λ_{δ_0} is unique up to permutation, sign-changes, and scaling. Thus, for independent ξ_t the criterion λ_δ is minimal, i.e. the minimum of λ_δ is obtained at rotation angle δ_0. We have $\lambda_{\hat{\delta}} \geq \lambda_{\delta_0}$ and $\lambda^T_{\delta_0} \geq \lambda^T_{\hat{\delta}}$. Therefore, $\lambda^T_{\delta_0} - \lambda_{\delta_0} \geq \lambda^T_{\hat{\delta}} - \lambda_{\delta_0} \geq \lambda^T_{\hat{\delta}} - \lambda_{\hat{\delta}}$. This implies

$$\begin{aligned} |\lambda^T_{\hat{\delta}} - \lambda_{\delta_0}| &\leq \max(|\lambda^T_{\delta_0} - \lambda_{\delta_0}|, |\lambda^T_{\hat{\delta}} - \lambda_{\hat{\delta}}|) \\ &\leq \sup_\delta |\lambda^T_\delta - \lambda_\delta| \xrightarrow{p} 0. \end{aligned}$$

By continuity it follows that $R_{\hat{\delta}} \xrightarrow{p} R_{\delta_0}$ and $\hat{\delta} \xrightarrow{p} \delta_0$. □

Proof of Theorem 2. Using result 11, p.98 of Lütkepohl (1996), the conditional fourth order moment of portfolio returns is given by

$$\begin{aligned} E[\tau_t^4|\mathcal{F}_{t-1}] &= E[\mathrm{tr}[a_ta_t'(e_te_t')a_ta_t'(e_te_t')]|\mathcal{F}_{t-1}] \\ &= E[\mathrm{vec}(a_ta_t')'(W_t\xi_t\xi_t'W_t' \otimes W_t\xi_t\xi_t'W_t')\mathrm{vec}(a_ta_t')], \end{aligned} \tag{24}$$

where e_t has been replaced by its structural representation which is known conditional on $\mathcal{F}_{t-1}$. Noticing that the Kronecker product in (24) involves identical matrices and using the results 7, 7.2(6) and 8(a) (Lütkepohl, 1996, p.97), one obtains

$$\begin{aligned} E[\tau_t^4|\mathcal{F}_{t-1}] &= E[\mathrm{vec}(a_ta_t')'(W_t \otimes W_t)\mathrm{vec}(\xi_t\xi_t')\mathrm{vec}(\xi_t\xi_t')'(W_t' \otimes W_t')\mathrm{vec}(a_ta_t')] \\ &= \mathrm{vec}(a_ta_t')'(W_t \otimes W_t)E[\mathrm{vec}(\xi_t\xi_t')\mathrm{vec}(\xi_t\xi_t')'(W_t' \otimes W_t')\mathrm{vec}(a_ta_t')] \\ &= \mathrm{vec}(a_ta_t')'(W_t \otimes W_t)\Psi(W_t' \otimes W_t')\mathrm{vec}(a_ta_t') =: \varsigma_t. \end{aligned} \tag{25}$$

where Ψ is defined in (8). This proves the stated result. □

References

Bauwens, L., Laurent, S., Rombouts, J.V.K., 2006. Multivariate GARCH models: A survey. J. Appl. Econometrics 31, 79–109.
Boussama, F., Fuchs, F., Stelzer, R., 2011. Stationarity and geometric ergodicity of BEKK multivariate GARCH models. Stochastic Process. Appl. 121, 2331–2360.
Caporin, M., McAleer, M., 2012. Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. J. Econ. Surv. 26 (4), 736–751.
Chang, C.-L., McAleer, M., Zuo, G., 2017. Volatility spillovers and causality of carbon emissions, oil and coal spot and futures for the EU and USA. Sustainability 9 (10), 1789.
Comon, P., 1994. Independent component analysis, A new concept?. Signal Process. 36 (3), 287–314.
Comte, F., Lieberman, O., 2003. Asymptotic theory for multivariate GARCH processes. J. Multivariate Anal. 84, 61–84.
Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Econom. J. 119, 158–171.
Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.
ESA, 2016. Final Draft Regulatory Technical Standards. Working paper, Joint Committee of the European Supervisory Authorities ESMA, EBA, and EIOPA.
Fengler, M.R., Herwartz, H., 2018. Measuring spot variance spillovers when (co)variances are time-varying – the case of multivariate GARCH models. Oxf. Bull. Econ. Stat. 80, 135–159.
Genest, C., Quessy, J.-F., Rémillard, B., 2007. Asymptotic local efficiency of Cramér von Mises tests for multivariate independence. Ann. Statist. 35, 166–191.
Gouriéroux, C., Monfort, A., Renne, J., 2017. Statistical inference for independent component analysis: Application to structural VAR models. J. Econometrics 196, 111–126.
Hafner, C.M., Herwartz, H., 2006. Volatility impulse responses for multivariate GARCH models: An exchange rate illustration. J. Int. Money Finance 25, 719–740.
Hafner, C.M., Preminger, A., 2009. On asymptotic theory for multivariate GARCH models. J. Multivariate Anal. 100 (9), 2044–2054.
Hafner, C.M., Rombouts, J.V., 2007. Semiparametric multivariate volatility models. Econometric Theory 23 (2), 251–280.
Herwartz, H., Roestel, J., 2018. A Structural Approach to Identify Financial Transmission in Scenarios of Crises. Research paper, Kiel Economics Working Papers 2018-8.
Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons.
Lancaster, H.O., 1954. Traces and cumulants of quadratic forms in normal variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 16 (2), 247–254.
Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. J. Econometrics 196, 288–304.
Lassance, N., Demiguel, V., Vrins, F., 2019. Optimal Portfolio Diversification Via Independent Component Analysis. Working paper, London Business School.
Ling, S., McAleer, M., 2003. Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19 (02), 280–310.
Lütkepohl, H., 1996. Handbook of Matrices. John Wiley & Sons, Ltd., Chichester, p. xvi+304.
Lütkepohl, H., Kilian, L., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press, Cambridge.
McAleer, M., 2019. What they did not tell you about algebraic (non-)existence, mathematical (ir-)regularity and (non-)asymptotic properties of the full BEKK dynamic conditional covariance model. J. Risk Financ. Manage. 12 (2), 66.
McNeil, A., Frey, R., Embrechts, P., 2016. Quantitative Risk Management, second ed. Princeton University Press, Princeton, NJ.
Miettinen, J., Taskinen, S., Nordhausen, K., Oja, H., 2015. Fourth moments and independent component analysis. Statist. Sci. 30 (3), 372–390.
Rigobon, R., 2003. Identification through heteroskedasticity. Rev. Econ. Stat. 85 (4), 777–792.
Weber, E., 2010. Structural conditional correlation. J. Financ. Econom. 8 (3), 392–407.
van der Weide, R., 2002. GO-GARCH: A multivariate generalized orthogonal GARCH model. J. Appl. Econometrics 17, 549–564.