

Hafner, Christian M.; Herwartz, Helmut

**Working Paper**

## Testing for Causality in Variance using Multivariate GARCH Models

Economics Working Paper, No. 2004-03

**Provided in Cooperation with:**

Christian-Albrechts-University of Kiel, Department of Economics

*Suggested Citation:* Hafner, Christian M.; Herwartz, Helmut (2004) : Testing for Causality in Variance using Multivariate GARCH Models, Economics Working Paper, No. 2004-03, Kiel University, Department of Economics, Kiel,  
<http://nbn-resolving.de/urn:nbn:de:101:1-200911033770>

This Version is available at:

<http://hdl.handle.net/10419/21980>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Testing for Causality in Variance using Multivariate GARCH Models

by Christian M. Hafner, Helmut Herwartz

C | A | U

Christian-Albrechts-Universität Kiel

Department of Economics

*Economics Working Paper*

*No 2004-03*



# Testing for Causality in Variance using Multivariate GARCH Models

Christian M. Hafner\*      Helmut Herwartz †

March 17, 2004

## Abstract

Tests of causality in variance in multiple time series have been proposed recently, based on residuals of estimated univariate models. Although such tests are applied frequently little is known about their power properties. In this paper we show that a convenient alternative to residual based testing is to specify a multivariate volatility model, such as multivariate GARCH (or BEKK), and construct a Wald test on noncausality in variance. We compare both approaches to testing causality in variance in terms of asymptotic and finite sample properties. The Wald test is shown to have superior power properties under a sequence of local alternatives. Furthermore, we show by simulation that the Wald test is quite robust to misspecification of the order of the BEKK model, but that empirical power decreases substantially when asymmetries in volatility are ignored.

Keywords: causality, multivariate volatility, local power

JEL Classification: C22, C52

---

\*Econometric Institute, Erasmus University Rotterdam, P.O.B 1738, 3000 DR Rotterdam, The Netherlands [chafner@few.eur.nl](mailto:chafner@few.eur.nl) (corresponding author)

†Institut für Statistik und Ökonometrie, Christian Albrechts Universität zu Kiel, Ohlshausenstr. 40-60, D-24118 Kiel, Germany, [Herwartz@stat-econ.uni-kiel.de](mailto:Herwartz@stat-econ.uni-kiel.de)

# 1 Introduction

Causal relationships in systems of economic time series variables are often defined according to forecasting principles exploiting the idea that a cause must precede its effect in time. Tests of Granger causality (Granger (1969), Granger (1980), Granger (1988)) have become a standard step when analyzing linear systems of time series. In light of a still growing interest in dynamics of financial data recent work on causality also addresses the issue of second order causality and/or causality in variance (Granger, Robins and Engle (1986), Cheung and Ng (1996), Comte and Liebermann (2000)).

For testing the hypothesis of noncausality in variance two approaches have been followed in the literature. On the one hand two step methodologies have been introduced which concentrate on the cross correlation function (CCF) of univariate residual estimates. Building upon tests on causality in mean (Haugh (1976), Pierce and Haugh (1977)) Cheung and Ng (1996) follow these lines to infer on cross sectional dependence of squared GARCH innovations. Kanas and Kouretas (2002) employ the CCF-test introduced by Cheung and Ng (1996) to detect volatility spillovers between official and black currency markets.

Alternatively, causality in variance is often diagnosed by means of (Quasi) Maximum-Likelihood ((Q)ML) methods which utilize a parametric specification of volatility dynamics of systems of financial data. In particular, the BEKK form of the multivariate GARCH model (Engle and Kroner, 1995) allows to establish a one-to-one relation between non-causality in variance and particular testable zero restrictions imposed on the parametric model. One of the arguments of Cheung and Ng (1996) against using specifications of multivariate GARCH models was that a rigorous proof of asymptotic QML theory in the multivariate GARCH framework was still missing. This argument, however, is no longer valid after recent progress on the theoretical side, e.g., by Comte and Lieberman (2003).

Comte and Lieberman (2000) provide a unified treatment of first and second order causality in the framework of VARMA models with multivariate GARCH error terms. The BEKK model is widely used to test causal relationships between financial time series (see e.g. Hafner and Herwartz (1998), Herwartz and Lütkepohl (2000), Caporale, Spittis and Spagnolo (2002)). It is worthwhile to mention that both approaches to inference on causality, the two step and the ((Q)ML) methodology, have not yet been compared in terms of their empirical properties.

In this paper we pick up the definitions of causality in variance and linear causality in variance given in Comte and Lieberman (2000) for the multivariate GARCH model and relate it to the notions of strong, semi-strong and weak GARCH processes going back to Drost and Nijman (1993). We give sufficient and necessary conditions for (linear) noncausality in variance and derive testable parametric restrictions covering these restrictions. We provide a local power analysis and a Monte Carlo study to investigate the relative performance of the CCF-test and Wald-type tests derived from asymptotic (Q)ML theory. The robustness of the latter methodology under misspecification of the (quasi) log-likelihood function is also addressed.

The remainder of the paper is organized as follows: Section 2 provides the method-

ological framework for our analysis, the multivariate GARCH model and its VARMA representation. In Section 3 causality in variance and linear causality in variance are defined and competing tests of the null hypothesis of higher order noncausality are motivated. A local power investigation compares two approaches to causality testing which are frequently used, the CCF and the Wald test. A Monte Carlo study in Section 4 is provided to assess the finite sample properties of alternative tests on noncausality in variance. Section 5 briefly summarizes our main results and underscores their scope for empirical multivariate volatility modelling. Proofs and details of implementing the Wald test are given in Appendices A and B.

## 2 Multivariate GARCH Models

Let us first introduce the terminology of weak, semi-strong and strong multivariate GARCH models, analogously to the univariate case. Later we will define causality in variance concepts that naturally apply these alternative notions of GARCH models.

**Definition 1 (Multivariate GARCH)** *Let  $\varepsilon_t$  denote a stochastic vector process with  $K$  components and  $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ . Now define a positive definite and symmetric matrix  $H_t$  such that  $h_t = \text{vech}(H_t)$  has the representation*

$$h_t = \omega + \sum_{i=1}^q A_i \eta_{t-i} + \sum_{j=1}^p B_j h_{t-j}, \quad (1)$$

where  $\omega = \text{vech}(\Omega)$ ,  $\eta_t = \text{vech}(\varepsilon_t \varepsilon_t')$  and  $\Omega, A_i, B_j$ , are  $K^* \times K^*$  parameter matrices with  $K^* = K(K+1)/2$ . Then we say that  $\varepsilon_t$  is a

1. *strong multivariate GARCH( $p, q$ ) process, if  $\xi_t = H_t^{-1/2} \varepsilon_t$  is an i.i.d. process with mean zero and variance the identity matrix,*
2. *semi-strong multivariate GARCH( $p, q$ ) process, if  $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = H_t$ ,*
3. *weak multivariate GARCH( $p, q$ ) process, if  $h_t$  is the best linear predictor of  $\eta_t$  in terms of a constant and lagged values of  $\eta_t$ , that is*

$$h_t = P(\eta_t | \mathcal{H}_{t-1}) = [P(\eta_{t,1} | \mathcal{H}_{t-1}), \dots, P(\eta_{t,K^*} | \mathcal{H}_{t-1})]'$$

where  $\mathcal{H}_t$  denotes the Hilbert space spanned by a constant and  $\eta_{t-\tau,1}, \dots, \eta_{t-\tau,K^*}$ ,  $\tau \geq 0$ .

A strong multivariate GARCH( $p, q$ ) process is also semi-strong, and a semi-strong multivariate GARCH( $p, q$ ) process is also weak, which justifies the terminology.

To establish the analogy to VARMA models, rewrite the process (1) as

$$\eta_t = \omega + \sum_{i=1}^{\max(p,q)} (A_i + B_i) \eta_{t-i} - \sum_{j=1}^p B_j u_{t-j} + u_t, \quad (2)$$

with  $u_t = \eta_t - h_t$  and where we set  $A_{q+1} = \dots = A_p = 0$  if  $p > q$  and  $B_{p+1} = \dots = B_q = 0$  if  $q > p$ . It now depends on the properties of  $u_t$  if we can consider (2) as a VARMA process.

**Assumption 1** *The covariance matrix of  $u_t$ ,  $\Sigma_u = E[u_t u_t']$ , is assumed to be finite and positive definite.*

For the case of a strong multivariate GARCH( $p, q$ ) model with spherical distribution of  $\xi_t$ , necessary and sufficient conditions for Assumption 1 to hold are given in Hafner (2003). We now have the following result.

**Lemma 1** *Under Assumption 1, if  $\{\varepsilon_t\}$  is*

1. *strong or semi-strong multivariate GARCH( $p, q$ ), then  $\{u_t\}$  in (2) is a martingale difference process.*
2. *weak multivariate GARCH( $p, q$ ), then  $\{u_t\}$  is weak white noise in the sense that  $E[u_t] = 0$ ,  $E[u_t u_s'] = 0$ ,  $\forall t \neq s$ , and  $E[u_t u_t'] = \Sigma_u < \infty$ .*

As a result of this lemma,  $\{\eta_t\}$  in (2) follows a VARMA( $\max(p, q), p$ ) process under Assumption 1. In the case of strong and semi-strong multivariate GARCH, this VARMA process characterizes the conditional mean of  $\eta_t$ , whereas in the case of weak multivariate GARCH, it characterizes the best linear predictor of  $\eta_t$  in terms of lagged values of  $\eta_t$ . Thus, when defining causality concepts we will have to distinguish between concepts that are based on the conditional mean and concepts based on best linear predictors.

In the following we will assume covariance stationarity of the process  $\varepsilon_t$ .

**Assumption 2** *All eigenvalues of the matrix  $\sum_{i=1}^{\max(p, q)} (A_i + B_i)$  have modulus smaller than one.*

The multivariate GARCH( $p, q$ ) process  $\varepsilon_t$  is covariance stationary if and only if Assumption 2 holds, see e.g. Engle and Kroner (1995). In that case, the components of the unconditional covariance matrix  $\Sigma = \text{Var}(\varepsilon_t)$  are given by

$$\sigma = \text{vech}(\Sigma) = \left( I_{K^*} - \sum_{i=1}^{\max(p, q)} (A_i + B_i) \right)^{-1} \omega. \quad (3)$$

Under Assumption 2, one obtains the VMA( $\infty$ ) representation from the VARMA representation (2),

$$\eta_t = \sigma + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \quad (4)$$

where the  $K^* \times K^*$  matrices  $\Phi_i$  can be determined recursively by  $\Phi_0 = I_{K^*}$ ,

$$\Phi_i = -B_i + \sum_{j=1}^i (A_j + B_j) \Phi_{i-j}, \quad i = 1, 2, \dots, \quad (5)$$

see Lütkepohl (1993, pp. 220).

In practice it is often easier to work with the so-called BEKK model of Engle and Kroner (1995), which is a special case of the vec model. The BEKK model involves less parameters to be estimated and ensures positive definiteness of  $H_t$  under weak conditions. In its general form, the BEKK( $p, q, S$ ) model can be written as

$$H_t = CC' + \sum_{s=1}^S \sum_{i=1}^q A_{si}^* \varepsilon_{t-i} \varepsilon_{t-i}' A_{si}^{*'} + \sum_{s=1}^S \sum_{i=1}^p B_{si}^* H_{t-i} B_{si}^{*'}, \quad (6)$$

where  $C$  is a lower triangular matrix and  $A_{si}^*$  and  $B_{si}^*$  are  $K \times K$  parameter matrices. For illustrative purposes, we will only consider the case  $p = q = S = 1$ , which is also the mostly applied model order. Thus, the model simplifies to

$$H_t = CC' + A^* \varepsilon_{t-1} \varepsilon_{t-1}' A^{*'} + B^* H_{t-1} B^{*'}, \quad (7)$$

with

$$\vartheta = (\text{vech}(C)', \text{vec}(A^*)', \text{vec}(B^*)')' \quad (8)$$

being a  $K(5K + 1)/2$ -dimensional parameter vector. Note that each BEKK model has a corresponding unique vec representation, but not vice versa, see Engle and Kroner (1995).

### 3 Causality Tests

In order to define causality concepts for the variance of a vector process  $\varepsilon_t$ , we make in the following two assumptions that simplify the presentation and allow us to focus on the issue of inference for causality. The first assumption is that the conditional mean is zero, i.e.,  $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ . Without this assumption, there would be a difference between a concept that corrects for the mean using all available information and one that corrects for the mean using only the information of the variable that is to be caused by the others. The first notion was introduced by Granger, Robins and Engle (1986), and the second one by Comte and Lieberman (2000). Under the assumption  $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ , however, both notions are equivalent.

The second simplifying assumption concerns the number of sub-groups of the vector  $\varepsilon_t$ . We assume that there are only two sub-groups, and we investigate concepts of causality between these two groups. As is well known, e.g. from Dufour and Renault (1998), in such setups it suffices to investigate the causality horizon of one period. If there is noncausality at horizon one, then there is noncausality at every horizon. If there were more sub-groups of the vector  $\varepsilon_t$ , and we were investigating the causality between the first two sub-groups, say, then there could be causality at larger horizons even though there may not be causality at horizon one. The intuitive reason is that there may be a causality chain going from the causing sub-group to a third sub-group, and then in a later period from this third sub-group to the sub-group to be caused. Thus, our restriction to only two sub-groups means that we can restrict our attention to a horizon of one period, which is notationally convenient. About all results of Dufour and Renault (1998) apply in our

setting as well, so that we only discuss the simple case here and all extensions follow by analogy.

First, define the index sets  $\mathcal{I} = (i_1, \dots, i_k)$  and  $\mathcal{J} = (j_1, \dots, j_{K-k})$ , where  $\mathcal{I} \cup \mathcal{J} = (1, \dots, K)$  and  $\mathcal{I} \cap \mathcal{J} = \emptyset$ . We will investigate the issue whether the variables indexed by  $\mathcal{J}$  cause the variables indexed by  $\mathcal{I}$ . We define the sub-vectors of  $\varepsilon_t$  by  $\varepsilon_t^{\mathcal{I}} = (\varepsilon_{t,i_1}, \dots, \varepsilon_{t,i_k})'$  and  $\varepsilon_t^{\mathcal{J}} = (\varepsilon_{t,j_1}, \dots, \varepsilon_{t,j_{K-k}})'$ , and let  $\eta_t^{\mathcal{I}} = \text{vech}(\varepsilon_t^{\mathcal{I}} \varepsilon_t^{\mathcal{I}'})$ , which is a vector of length  $k^* = k(k+1)/2$ . The  $\sigma$ -algebras generated by  $\varepsilon_s^{\mathcal{I}}$  and  $\varepsilon_s^{\mathcal{J}}$ ,  $s \leq t$ , are denoted by  $\mathcal{F}_t^{\mathcal{I}}$  and  $\mathcal{F}_t^{\mathcal{J}}$ , respectively. Moreover, denote by  $\mathcal{H}_t^{\mathcal{I}}$  the Hilbert space spanned by the variables  $\eta_{s,1}^{\mathcal{I}}, \dots, \eta_{s,k^*}^{\mathcal{I}}$ ,  $s \leq t$ .

Now, similar to Comte and Lieberman (2000), we define causality in variance and linear causality in variance.

**Definition 2** *We say that*

- $\varepsilon_t^{\mathcal{J}}$  does not cause  $\varepsilon_t^{\mathcal{I}}$  in variance, denoted by  $\varepsilon_t^{\mathcal{J}} \not\stackrel{V}{\rightarrow} \varepsilon_t^{\mathcal{I}}$ , if

$$\text{Var}(\varepsilon_t^{\mathcal{I}} | \mathcal{F}_{t-1}) = \text{Var}(\varepsilon_t^{\mathcal{I}} | \mathcal{F}_{t-1}^{\mathcal{I}})$$

- $\varepsilon_t^{\mathcal{J}}$  does not cause  $\varepsilon_t^{\mathcal{I}}$  linearly in variance, denoted by  $\varepsilon_t^{\mathcal{J}} \not\stackrel{LV}{\rightarrow} \varepsilon_t^{\mathcal{I}}$ , if

$$P(\eta_t^{\mathcal{I}} | \mathcal{H}_{t-1}) = P(\eta_t^{\mathcal{I}} | \mathcal{H}_{t-1}^{\mathcal{I}})$$

### 3.1 The CCF test

Based on squared residuals  $\hat{\xi}_{i,t}^2 = \varepsilon_{i,t}^2 / \hat{\sigma}_{i,t}^2$ , where  $\hat{\sigma}_{i,t}^2$  is the estimated conditional variance of  $\varepsilon_{i,t}$  using univariate GARCH, Cheung and Ng (1996) introduce a portmanteau statistic to test the null hypothesis of noncausality in variance,

$$H_0 : \varepsilon_{j,t} \not\stackrel{V}{\rightarrow} \varepsilon_{i,t}, \quad \forall i \in \mathcal{I}, \forall j \in \mathcal{J}.$$

The test statistic builds upon sample cross correlations and reads as

$$P_m = T \sum_{l=1}^m r_{ij,l}^2, \quad i \in \mathcal{I}, j \in \mathcal{J}, \quad (9)$$

where

$$r_{ij,l} = \frac{c_{ij,l}}{\sqrt{c_{ii,0} c_{jj,0}}}, \quad c_{ij,l} = \frac{1}{T} \sum_{t=1}^T (\xi_{i,t}^2 - \bar{\xi}_i^2)(\xi_{j,t-l}^2 - \bar{\xi}_j^2),$$

and where  $\bar{\xi}_i^2 = T^{-1} \sum_{t=1}^T \xi_{i,t}^2$ .

In practice, the choice of  $m$  should allow to cover the highest potential lag of causality in variance. Cheung and Ng (1996) prove that under consistent estimation of the univariate GARCH parameters,  $P_m$  follows asymptotically a  $\chi^2(m)$ -distribution under the



null hypothesis. Analogous statistics can be defined for testing the hypothesis of bidirectional causality. Also, in small samples one can use modified portmanteau statistics in the standard manner.

The CCF test has the appealing feature to be easily computable. A drawback, however, is that the order  $m$  has to be determined. If  $m$  is chosen too small, one may miss causalities at higher lags, if it is chosen too large, the degrees of freedom increase and, hence, the power of the test decreases. We will show later that the CCF test has very poor power properties if the alternative is multivariate GARCH, irrespective of the choice of  $m$ .

### 3.2 A pseudo likelihood ratio statistic

The CCF test estimates univariate GARCH models and then tests for cross-correlations between standardized (squared) residuals. If there are only two series, it can be thought of estimating the model under the null hypothesis of no bi-directional causality, so that it is in the spirit of Lagrange multiplier statistics. In the next section we are going to discuss Wald type statistics. It is also possible to consider statistics in the spirit of likelihood ratio statistics, where both univariate and multivariate models are estimated. Likelihood ratio tests of causality in linear VARMA type models have been introduced by Geweke (1982) and extended to multivariate GARCH models by Hafner (2003).

To define the test statistic, consider a bivariate GARCH process and its VARMA representation (2), where the error term  $u_t$  has, by Assumption 1, finite covariance matrix  $\Sigma_u$ . Having a sample of  $T$  observations and considering causality in variance from  $\varepsilon_{2t}$  to  $\varepsilon_{1t}$ , we may alternatively estimate a univariate GARCH model of appropriate order for  $\varepsilon_{1t}$ , obtain its ARMA representation and the corresponding residual variance,  $\sigma_v^2$  say. Now the statistic is given by

$$LR = T \log \frac{\sigma_v^2}{\Sigma_{u,11}}. \quad (10)$$

In VARMA models with Gaussian errors, (10) is the usual likelihood ratio statistic. The problem in multivariate GARCH models is that the errors of the VARMA representation (2) are not Gaussian and typically highly skewed. Thus, (10) is not the true LR statistic and will be biased if compared with a  $\chi^2$  distribution. Nevertheless, it might be useful for descriptive purposes. We have used bootstrapped versions of the statistic (10) to correct for the size, but the power turned out to be equally poor as for the CCF test. Therefore, we do not report these results in this paper to economize on space.

### 3.3 Tests based on multivariate GARCH models

Noncausality in variance amounts to certain zero restrictions of the matrices  $A_i$  and  $B_j$  in (1). To find these restrictions, we first define the index

$$k_{ij}^K = i + (j - 1) \left( K - \frac{j}{2} \right) \quad (11)$$

for  $i, j \in \mathcal{I} \cup \mathcal{J}$  and  $i \geq j$ , which is the position of the  $(i, j)$ -th element of a  $(K \times K)$  symmetric matrix  $M$  in the vector  $\text{vech}(M)$ . Recall that  $\text{vech}(M)$  contains  $K^* = K(K + 1)/2$  elements. Moreover, we define the index sets

$$\mathcal{I}^* = \{k_{ij}^K \mid i, j \in \mathcal{I}\} \quad (12)$$

and

$$\mathcal{J}^* = \{1, \dots, K^*\} \setminus \mathcal{I}^*. \quad (13)$$

This notation at hand, we can now give conditions for noncausality in variance. Let us consider the following two conditions,

$$[\Phi_n]_{ij} = 0, \quad \forall n \geq 1, \quad \forall i \in \mathcal{I}^*, \quad \forall j \in \mathcal{J}^*, \quad (14)$$

and

$$[A_a]_{ij} = 0, \quad a = 1, \dots, q, \quad [B_b]_{ij} = 0, \quad b = 1, \dots, p, \quad \forall i \in \mathcal{I}^*, \quad \forall j \in \mathcal{J}^*. \quad (15)$$

**Theorem 1** *If  $\varepsilon_t$  is a weak multivariate GARCH process, then condition (14) is necessary and sufficient and condition (15) is sufficient for  $\varepsilon_t^{\mathcal{J}} \xrightarrow{LV} \varepsilon_t^{\mathcal{I}}$ . If  $\varepsilon_t$  is a semi-strong multivariate GARCH process, then each one of conditions (14) and (15) is sufficient for  $\varepsilon_t^{\mathcal{J}} \xrightarrow{V} \varepsilon_t^{\mathcal{I}}$ .*

**Proof:** see Appendix.

The first part of Theorem 1 is well known, see e.g. Lütkepohl (1993, p. 236f.). Unlike in the first part, no equivalence between Condition (14) and variance noncausality can be established in the second part. The reason for this is that, if  $\varepsilon_t$  is a semi-strong multivariate GARCH process, then a subprocess is only weak GARCH, see Nijman and Sentana (1996). Thus, the conditional expectation of a subprocess of  $\eta_t$  may not be a linear function of lagged  $\eta_t$ , and the restrictions that variance noncausality implies become impossible to express in terms of the moving average coefficients  $\Phi_n$ .

Turning to the testing problem, note that Condition (14) is infeasible to test due to the large number of coefficient matrices to be tested. We therefore focus attention on testing Condition (15), but according to Theorem 1 it is only a sufficient condition for variance noncausality, even in the linear causality sense. Fortunately, for the often used multivariate GARCH(1,1) model,

$$h_t = \omega + A\eta_{t-1} + Bh_{t-1}, \quad (16)$$

where Condition (15) simplifies to

$$[A]_{ij} = 0, \quad [B]_{ij} = 0, \quad \forall i \in \mathcal{I}^*, \quad \forall j \in \mathcal{J}^*, \quad (17)$$

it turns out that both conditions are equivalent under an additional assumption.

**Lemma 2** *In the multivariate GARCH(1,1) model (16), if  $A$  is invertible, then conditions (14) and (17) are equivalent.*

**Proof:** see Appendix.

As a consequence of this lemma, if  $A$  is invertible, then testing (17) is equivalent to testing variance noncausality and linear variance noncausality, respectively. Note the difference to VARMA(1,1) models where (17) is only a sufficient condition for noncausality, see e.g. Lütkepohl (1993, p. 236f.). The reason for this difference is the particular parameter structure of GARCH models, that is, the sum of the autoregressive parameter matrix  $(A+B)$  and the moving average matrix  $(-B)$  is just equal to  $A$ , giving  $\Phi_1 = A$ , so that  $A$  directly inherits the properties of  $\Phi_1$ . In the VARMA(1,1) model,  $\Phi_1$  is equal to the sum of the autoregressive and moving average parameter matrices, so that a restriction of  $\Phi_1$  does not necessarily convey to each one of these.

Since Condition (17) in weak GARCH(1,1) models is not only sufficient but also necessary for linear noncausality in variance, testing this condition under the null hypothesis of no linear causality in variance should provide correct Type I and Type II errors. It is also likely to have more power than alternative tests that are based on only necessary conditions for noncausality, such as the CCF test of Cheung and Ng (1996).

In the following we will only discuss testing in GARCH(1,1) models, so that we will only consider tests of Condition (17). It is now straightforward to define a test statistic that tests the zero restrictions on  $A$  and  $B$ . Let us first define the following restriction matrix.

**Definition 3** Let  $\tilde{R}$  be a matrix of dimension  $k^*(K^* - k^*) \times (K^*)^2$ , of rank  $k^*(K^* - k^*)$ . The  $(r, \tau)$  element of  $\tilde{R}$  is defined by

$$\tilde{R}_{r,\tau} = \begin{cases} 1, & \tau = s_{mn} \\ 0, & \tau \neq s_{mn} \end{cases}$$

where  $r = m + (n - 1)k^*$ ,  $s_{mn} = i_m + (j_n - 1)K^*$ ,  $i_m \in \mathcal{I}^*$ ,  $j_n \in \mathcal{J}^*$ , and  $m = 1, \dots, k^*$ ,  $n = 1, \dots, K^* - k^*$ .

Each row of  $\tilde{R}$  contains a one at the  $i + (j - 1)K^*$ -th position, where  $i \in \mathcal{I}^*$  and  $j \in \mathcal{J}^*$ , and zeros elsewhere. Appendix B gives some examples how to find the restriction matrix  $\tilde{R}$ .

The null hypothesis of no causality can be written as

$$H_0 : R\theta = 0 \tag{18}$$

with  $\theta = (\omega', \text{vec}(A)', \text{vec}(B)')'$ , and

$$R = [0_{(k^*(K^*-k^*) \times K^*)}, \tilde{R}, \tilde{R}]$$

If the BEKK(1,1,1) model (7) model is used, then equivalent conditions can be found that involve less equations. To formalize this idea, let us define the restriction matrix  $\tilde{Q}$  associated with the BEKK model.

**Definition 4** Let  $\tilde{Q}$  be a matrix of dimension  $k(K - k) \times (K)^2$ , of rank  $k(K - k)$ . The  $(r, \tau)$  element of  $\tilde{Q}$  is defined by

$$\tilde{Q}_{r,\tau} = \begin{cases} 1, & \tau = s_{mn} \\ 0, & \tau \neq s_{mn} \end{cases}$$

where  $r = m + (n - 1)k$ ,  $s_{mn} = i_m + (j_n - 1)K$ ,  $i_m \in \mathcal{I}$ ,  $j_n \in \mathcal{J}$ , and  $m = 1, \dots, k$ ,  $n = 1, \dots, K - k$ .

The null hypothesis of no causality in the BEKK(1,1,1) model can now be written as

$$H_0 : Q\vartheta = 0 \tag{19}$$

with  $\vartheta$  given by (8), and where

$$Q = [0_{(k(K-k) \times K)}, \tilde{Q}, \tilde{Q}].$$

The following theorem states the equivalence of the conditions (18) and (19).

**Theorem 2** *If the multivariate GARCH model has the BEKK(1,1,1) representation (7), then the noncausality conditions (18) and (19) are equivalent.*

**Proof:** see Appendix.

Note that, although equivalent, condition (19) only involves  $k(K - k)$  equations as opposed to the  $(k^*(K^* - k^*))$  equations of condition (18). In standard likelihood based tests, the reduction in degrees of freedom may therefore result in more power if a BEKK model is used instead of a vec model, provided the data are well described by a BEKK model. Essentially, many conditions of (18) are redundant and just implied by (19). For example, if  $K = 2$  and  $k = 1$ , then (18) has two equations and (19) only one.

Suppose now that we have  $T$  observations,  $\varepsilon_1, \dots, \varepsilon_T$ . We assume in the following that the true process is known to belong to the BEKK class, for which asymptotic theory of estimation and inference is well developed, see Comte and Lieberman (2003). Denote a consistent estimator of the true parameter vector  $\vartheta_0$  by  $\hat{\vartheta}$  and assume that its asymptotic distribution is given by

$$\sqrt{T}(\hat{\vartheta} - \vartheta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_\vartheta), \tag{20}$$

with some positive definite and symmetric matrix  $\Sigma_\vartheta$ . Assume also that a consistent estimate for  $\Sigma_\vartheta$  is given by  $\hat{\Sigma}_\vartheta$ . For example, if QML estimation is used, then (20) holds under regularity conditions listed by Comte and Lieberman (2003), and  $\Sigma_\vartheta$  is given by

$$\Sigma_\vartheta = \mathcal{S}^{-1} \mathcal{D} \mathcal{S}^{-1},$$

where

$$\mathcal{D} = \text{E} \left[ \frac{\partial l_t(\vartheta)}{\partial \vartheta} \frac{\partial l_t(\vartheta)}{\partial \vartheta'} \bigg|_{\vartheta_0} \right], \quad \mathcal{S} = -\text{E} \left[ \frac{\partial^2 l_t(\vartheta)}{\partial \vartheta \partial \vartheta'} \bigg|_{\vartheta_0} \right],$$

with

$$l_t(\vartheta) = -\frac{K}{2} \ln(2\pi) - \frac{1}{2} \ln |H_t(\vartheta)| - \frac{1}{2} \varepsilon_t' H_t^{-1}(\vartheta) \varepsilon_t. \quad (21)$$

Hafner and Herwartz (2003) provide expressions for  $\mathcal{D}$  and  $\mathcal{S}$  and of their estimates. For significance testing Hafner and Herwartz (2003) show that making use of analytical expressions for  $\Sigma_\vartheta$  is by far superior to using numerical derivatives in terms of empirical size and power estimates.

We propose the following standard Wald statistic for testing the hypothesis (19),

$$W_T = T(Q\hat{\vartheta})'(Q\hat{\Sigma}_\vartheta Q')^{-1}(Q\hat{\vartheta}). \quad (22)$$

Using (20) and Proposition C.4 of Lütkepohl (1993), the asymptotic distribution of the Wald statistic is given by

$$W_T \xrightarrow{\mathcal{L}} \chi_{k(K-k)}^2.$$

An analogous statistic can be defined for the vec model based on the null hypothesis (18), provided that conditions for asymptotic normality of estimators hold, which is as yet unknown. Note that the degrees of freedom of the Wald statistic for the vec model would be  $k^*(K^* - k^*)$ .

### 3.4 A power comparison

We now investigate the asymptotic power of the Wald and the CCF tests under a sequence of local alternatives. To ensure that all asymptotic results are valid we still assume that the model is known to belong to the BEKK class. Consider the parameter vector

$$\vartheta_{0T} = \vartheta_0 + \delta/\sqrt{T}$$

where  $\vartheta_0$  is the parameter vector under the null hypothesis and  $\delta$  is a fixed vector of the same length as  $\vartheta_0$ . Under local alternatives, we have the asymptotic distribution

$$\sqrt{T}(\hat{\vartheta} - \vartheta_{0T}) = \sqrt{T}(\hat{\vartheta} - \vartheta_0) + \delta \xrightarrow{\mathcal{L}} N(\delta, \Sigma_\vartheta). \quad (23)$$

As a consequence, the Wald statistic in (22) has the following asymptotic distribution under local alternatives,

$$W_T \xrightarrow{\mathcal{L}} \chi_{k(K-k)}^2(\lambda),$$

i.e., a noncentral  $\chi^2$  distribution with  $k(K - k)$  degrees of freedom and noncentrality parameter  $\lambda$  given by

$$\lambda = (\delta' Q')(Q \Sigma_\vartheta Q')^{-1} Q \delta.$$

One can now derive the asymptotic power of  $W_T$  as a function of  $\delta$ . If only one element of  $\delta$  is different from zero, then one can plot the asymptotic power as a function of this element, as done in Figure 1 for the process defined in the next section.

Next, we can derive the asymptotic power of the CCF test under a sequence of local alternatives  $\vartheta_{0T}$  that characterizes a multivariate GARCH alternative. We obtain, using a Taylor expansion around  $\vartheta_0$ ,

$$\sqrt{T}\hat{r}_{ij,l}(\vartheta_{0T}) = \sqrt{T}\hat{r}_{ij,l}(\vartheta_0) + \left. \frac{\partial \hat{r}_{ij,l}}{\partial \vartheta'} \right|_{\vartheta_0} \delta + O_p(T^{-1/2}).$$

Assuming consistency of correlation estimates, i.e.,  $\text{plim } \hat{r}_{ij,l}(\vartheta_0) = r_{ij,l}(\vartheta_0)$ , we obtain for the vector of correlations up to lag  $m$ ,

$$\sqrt{T}\hat{r}_{ij}^m(\vartheta_{0T}) \xrightarrow{\mathcal{L}} N \left( \left. \frac{\partial r_{ij}^m}{\partial \vartheta'} \right|_{\vartheta_0} \delta, I_m \right).$$

where  $r_{ij}^m = (r_{ij,1}, \dots, r_{ij,m})'$ . Hence, the CCF statistic

$$P_m = T \sum_{l=1}^m r_{ij,l}^2 \xrightarrow{\mathcal{L}} \chi_m^2(\lambda)$$

has, asymptotically, a noncentral  $\chi^2$  distribution with  $m$  degrees of freedom and noncentrality parameter  $\lambda$  given by

$$\lambda = \delta' \left. \frac{\partial r_{ij}^m}{\partial \vartheta} \right|_{\vartheta_0} \left. \frac{\partial r_{ij}^m}{\partial \vartheta'} \right|_{\vartheta_0} \delta.$$

The derivative can be calculated numerically. We generate 500 bivariate diagonal BEKK processes with  $T = 10000$ , estimate univariate GARCH processes, obtain residuals  $\hat{\xi}_{1,t}$  and  $\hat{\xi}_{2,t}$  and calculate correlations  $\hat{r}_{ij}^m$ . The same is done for a bivariate BEKK process with lower left element of the  $A^*$  matrix changed to -0.01 and 0.01, and the mean of the corresponding derivatives of  $\hat{r}_{ij}^m$  is a good approximation of  $\partial r_{ij}^m / \partial \vartheta'$ . For the BEKK(1,1,1) process (7) with parameters specified in Section 4, the asymptotic power function of the CCF test is depicted in Figure 1 together with the corresponding function for the Wald test, assuming Gaussian innovations. Clearly, the Wald test has uniformly higher power in a neighborhood of  $\vartheta_0$ . We get a very similar picture in Figure 2 when assuming multivariate  $t_8$  distributed innovations, where the power drops slightly for both the Wald and the CCF tests.

## 4 Finite sample performance

The following Monte Carlo investigation is thought to shed light on the empirical performance of two strategies for inference on noncausality in variance. We compare the empirical properties of the Wald statistic in (22) on the one hand and of the CCF test introduced by Cheung and Ng (1996) on the other hand.

## 4.1 The Monte Carlo design - Wald vs. CCF

To illustrate the empirical size properties of competing approaches to test noncausality in variance we simulate bivariate GARCH-processes of the BEKK-form ( $S = p = q = 1$ ) according to the following choice of parameter matrices:

$$C = \begin{pmatrix} 1.10 & 0.00 \\ 0.30 & 0.90 \end{pmatrix}, A^* = \begin{pmatrix} 0.25 & 0.00 \\ 0.00 & 0.25 \end{pmatrix}, B^* = \begin{pmatrix} 0.90 & 0.00 \\ 0.00 & 0.90 \end{pmatrix}. \quad (24)$$

We test three null hypotheses. The first null hypothesis states that there is no causality in variance in the system at all. The second and third null hypothesis formalize that  $\varepsilon_{1t}$  does not cause  $\varepsilon_{2t}$  in variance and vice versa. In summary we test the following null hypotheses:

$$\begin{aligned} H_0^{(1)} &: \varepsilon_{1t} \not\overset{V}{\rightarrow} \varepsilon_{2t}, \varepsilon_{2t} \not\overset{V}{\rightarrow} \varepsilon_{1t}, \\ H_0^{(2)} &: \varepsilon_{1t} \overset{V}{\rightarrow} \varepsilon_{2t}, \\ H_0^{(3)} &: \varepsilon_{2t} \overset{V}{\rightarrow} \varepsilon_{1t}. \end{aligned}$$

We also provide empirical power estimates for the cases when testing  $H_0^{(1)}$  or  $H_0^{(2)}$ . Under the alternative hypotheses we choose the parameter matrices  $A^*$  and  $B^*$  as

$$A^* = \begin{pmatrix} .250 & .000 \\ .025 & .250 \end{pmatrix}, B^* = \begin{pmatrix} .900 & .000 \\ .025 & .900 \end{pmatrix}. \quad (25)$$

To indicate the relative performance of exact ML inference on the one hand and the QML methodology on the other hand we draw underlying innovations alternatively from a bivariate Gaussian distribution or as standardized and independent innovations from a  $t$ -distribution with 8 degrees of freedom. Note that under Gaussian innovations estimating the asymptotic covariance matrix as  $\hat{\Sigma}_\vartheta = \hat{\mathcal{S}}^{-1} \hat{\mathcal{D}} \hat{\mathcal{S}}^{-1}$ , could be inefficient in small samples (Hafner and Herwartz (2003)). Under conditional leptokurtosis making use of a covariance estimator  $\hat{\Sigma}_\vartheta = \hat{\mathcal{D}}^{-1}$  will result in size distortions. We consider sample sizes  $T = 1000, 2000, 4000, 8000$ . The nominal significance level for all performed tests is  $\alpha = 0.05$ . Each process is generated 2000 times.

## 4.2 Simulation results

Table 1 shows empirical rejection frequencies for the Wald statistics implemented alternatively with covariance estimators  $\hat{\mathcal{D}}^{-1}$  (W1) and  $\hat{\mathcal{S}}^{-1} \hat{\mathcal{D}} \hat{\mathcal{S}}^{-1}$  (W2). Overall the empirical size estimates are larger than their nominal counterparts, and often exceed the latter significantly at the 5% level even in large samples ( $T = 8000$ ). Empirical size estimates are in almost any case somewhat larger when testing on overall noncausality ( $H_0^{(1)}$ ) in comparison to testing on unidirectional noncausality ( $H_0^{(2)}, H_0^{(3)}$ ). Under leptokurtic (standardized  $t$ -distributed) innovations  $\xi_t$ , W1 shows huge size distortions whereas the empirical size estimates obtained for W2 come close to the nominal level as the sample

size increases ( $T = 8000$ ). Moreover, under conditional normality estimating the asymptotic covariance matrix as  $\hat{\Sigma}_\vartheta = \hat{\mathcal{S}}^{-1} \hat{\mathcal{D}} \hat{\mathcal{S}}^{-1}$  yields higher size estimates in comparison with W1. For instance, testing  $H_0^{(1)}$  under conditional normality in samples of size  $T = 1000$  obtains empirical rejection frequencies of  $\hat{\alpha} = 0.077$  and  $\hat{\alpha} = 0.121$  for W1 and W2, respectively. Relative to the nominal level of  $\alpha = 0.05$  it is evident, that the choice of the robust covariance estimator may go at the cost of size distortions.

With respect to power properties testing overall noncausality ( $H_0^{(1)}$ ) turns out to be less effective than unidirectional testing ( $H_0^{(2)}$ ) when actually  $\varepsilon_{1t}$  is causing  $\varepsilon_{2t}$  and  $\varepsilon_{2,t} \overset{V}{\nrightarrow} \varepsilon_{1,t}$ . Under conditional normality (conditional leptokurtosis), for instance, W1 (W2) delivers empirical rejection frequencies which are up to 9% higher when testing  $H_0^{(2)}$  instead of  $H_0^{(1)}$ .

Table 2 displays selected simulation results for the CCF-test, namely size estimates for the case  $T = 1000$  and power estimates for samples of size  $T = 8000$  under both, conditional normality and leptokurtosis. Results are shown for alternative test orders ( $m$ ). Apparently the size properties obtained from CCF are close to the nominal level and therefore superior relative to the performance of the Wald statistics. In terms of power, however, the CCF approach performs rather poor. For example, testing under Gaussian innovations in samples of size  $T = 8000$  the most favorable rejection frequency obtained for the CCF-test is 15.75% which is by far inferior to the Wald test delivering empirical power estimates of about 80.0%.

Given that W1 and W2 tend to reject under the null hypothesis more often than CCF it is sensible to compare size adjusted power estimates. For this purpose Table 3 displays rejection frequencies obtained when testing  $H_0^{(1)}$  and  $H_0^{(2)}$  for the Gaussian model with  $\varepsilon_{1t}$  causing  $\varepsilon_{2t}$  in variance. Size adjustment is here achieved by tuning the nominal level of the CCF test such that under the null hypothesis both test procedures, the Wald and the CCF, give identical empirical size estimates. Apparently, W1 clearly outperforms the CCF-test after size adjustment. In samples with  $T = 4000$  or  $T = 8000$  size adjusted power estimates of W1 are up to five times larger than the corresponding estimates obtained for the CCF-test. Note that the latter result is particularly important for practical purposes. Adopting the CCF approach to test for causality in variance will often fail to uncover causal relationships and will thereby tend to preclude multivariate volatility models allowing cross equation dynamics, as e.g. the BEKK-model.

### 4.3 The Wald test under misspecification of the DGP

As outlined before the Wald statistic is obtained from QML-estimation of the multivariate GARCH process. To indicate the impact of misspecification of the underlying DGP and, thus, of the (quasi) log-likelihood function we follow two lines. First, we estimate BEKK models of order  $S = p = q = 1$  when the true DGP has a higher BEKK order, namely  $S = 2$ . In this case we use the following parameter choices to evaluate size and power properties, respectively:



- Size:

$$A_{11}^* = \begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.15 \end{pmatrix}, A_{21}^* = \begin{pmatrix} 0.15 & 0.00 \\ 0.00 & 0.20 \end{pmatrix}, B_{11}^* = B_{21}^* = \begin{pmatrix} 0.90 & 0.00 \\ 0.00 & 0.90 \end{pmatrix}. \quad (26)$$

- Power

$$A_{11}^* = \begin{pmatrix} .200 & .000 \\ .025 & .150 \end{pmatrix}, A_{21}^* = \begin{pmatrix} .150 & .000 \\ .000 & .200 \end{pmatrix}, B_{11}^* = B_{21}^* = \begin{pmatrix} .900 & .000 \\ .025 & .900 \end{pmatrix}. \quad (27)$$

Second, we estimate symmetric multivariate GARCH models in case the true underlying DGP exhibits an asymmetric impact of current (co)variances to lagged innovations. For this purpose we use a bivariate threshold GARCH specification as in Hafner and Herwartz (1998) or Herwartz and Lütkepohl (2000) generalizing the univariate process introduced by Glosten, Jaganathan and Runkle (1993). Formally the asymmetric DGP is specified by means of a state dependent parameter matrix  $\tilde{A}_{11}^*$  replacing the corresponding parameters in (6). The latter is chosen as

$$\tilde{A}_{11}^* = A_{11}^* I(\varepsilon_{1,t-1} < 0) + A_{21}^* I(\varepsilon_{2,t-1} < 0), \quad (28)$$

where  $I(\cdot)$  is an indicator function and  $A_{11}^*$  and  $A_{21}^*$  are given in (26) and (27) for assessing size and power properties, respectively. With respect to the choice of the matrix  $B_{11}^*$  the asymmetric process is identical to the symmetric specification with parameters given in (24) and (25).

#### 4.4 QML under misspecification - Simulation results

Table 4 shows empirical size and power estimates for QML inference under conditional normality. Empirical size estimates of both Wald tests, W1 and W2, are almost unaffected when modelling a process parameterized with (26) under misspecification of the order parameter  $S$ . When testing  $H_0^{(1)}$  or  $H_0^{(2)}$  under the alternative of causal relations, however, underestimating the BEKK order involves losses in terms of power. For instance, in case  $T = 8000$  testing  $H_0^{(2)}$  in presence of causal relationships yields empirical rejection frequencies for W1 of .851 and .716 if the BEKK order of the underlying model is  $S = 1$  and  $S = 2$ , respectively. Under the null hypothesis W1 shows an empirical size of 5.09% for both (true) BEKK orders underlying the DGP. In comparison with misspecifying the BEKK order, ignoring the potential of an asymmetric response of volatility with respect to the sign of lagged error terms  $\varepsilon_{t-1}$  involves slightly higher empirical size distortions under the null hypothesis but huge power losses under the alternative. For example, employing the symmetric BEKK model to test on overall noncausality ( $H_0^{(1)}$ ) by means of W1 and in case  $T = 8000$  yields empirical size and power estimates of 6.35% and 77.75% (7.9% and 19.05%) if the true DGP exhibits a symmetric (asymmetric) impact of  $\varepsilon_{t-1}$  on volatility.

## 5 Conclusions

We formalize the concepts of strong, semi strong, and weak multivariate GARCH. Using the general vec representation of this model and building on Comte and Lieberman (2000) we prove sufficiency or necessity of particular parameter restrictions for noncausality in variance (linear causality in variance). Two approaches to testing for causality in variance, namely the CCF test introduced by Cheung and Ng (1996) and a Wald test based on (Q)ML theory are discussed. Evaluating the asymptotic local power properties we find that the CCF test is inferior.

A Monte Carlo investigation indicates that the CCF test has more favorable empirical size properties in comparison with the Wald test. The former test, however, is also characterized by a severe shortfall in terms of empirical power. As a particular drawback of the (Q)ML based approach one may regard the necessity of a (potentially misspecified) parametric model to formalize the log likelihood function. Our results indicate that ignoring an asymmetric impact of volatility on lagged innovations will involve significant power losses in causality testing whereas underestimating the so-called BEKK order of a particular DGP appears to have less severe implications for the power of Wald type tests.

For practical aspects of (co)variance modelling our results imply that using the CCF test will in general mitigate the evidence in favor of volatility spillovers. Moreover, specification tests on asymmetric impacts of lagged innovations on current volatility should be applied before formalizing higher dimensional parametric volatility models.

## Appendix A

**Definition 5** Let  $\mathcal{K} = \{1, \dots, K\}$ ,  $K \geq 2$ , and  $\mathcal{I} \subset \mathcal{K}$ . Let  $X$  be a square matrix of order  $K$ . Then  $X$  is said to be  $0_{\mathcal{I}}$  if

$$X_{ij} = 0 \quad \forall i \in \mathcal{I}, \quad \forall j \notin \mathcal{I}.$$

**Lemma 3** Let  $X$  and  $Y$  be some square matrices of order  $K \geq 2$ , and  $\mathcal{I} \subset \{1, \dots, K\}$ . If both  $X$  and  $Y$  are  $0_{\mathcal{I}}$ , then the matrix product  $XY$  is also  $0_{\mathcal{I}}$ .

**Proof:** By definition of the Cayley matrix product,

$$[XY]_{ij} = \sum_{k \in \mathcal{I}} X_{ik} Y_{kj} + \sum_{k \notin \mathcal{I}} X_{ik} Y_{kj}. \quad (29)$$

Now  $\forall i \in \mathcal{I}^*$  and  $\forall j \in \mathcal{J}^*$  the first term on the right hand side of (29) is zero because  $Y_{kj} = 0$ , and the second term is zero because  $X_{ik} = 0$ . Thus,  $XY$  is  $0_{\mathcal{I}}$ . Q.E.D.

**Lemma 4** If  $A(L) = I_K + \sum_{n=1}^{\infty} A_n L^n$  is an invertible linear filter and  $A_n$  is  $0_{\mathcal{I}} \forall n \geq 1$ , then  $A(L)^{-1} = \Pi(L) = I_K - \sum_{n=1}^{\infty} \Pi_n L^n$  is such that  $\Pi_n$  is  $0_{\mathcal{I}} \forall n \geq 1$ .

**Proof:** The inverse filter is obtained recursively by  $\Pi_1 = A_1$  and  $\Pi_n = A_n - \sum_{m=1}^{n-1} \Pi_{n-m} A_m$ , see Lütkepohl (1993, p.219). Applying Lemma 3 recursively yields the statement. Q.E.D.

**Lemma 5** *Condition (15) implies Condition (14).*

**Proof:** Follows immediately by applying Lemma 3 recursively to the matrices  $\Phi_n$  defined by (5). Q.E.D.

**Proof of Theorem 1:** The first part follows by the fact that weak multivariate GARCH allows for the VARMA representation (2). The definition of linear causality in variance corresponds to the definition of causality employed by Lütkepohl (1993). Hence, the equivalence of (14) and linear noncausality in variance follows by Proposition 2.2 of Lütkepohl (1993), and the sufficiency of (15) follows by Proposition 6.3 of Lütkepohl (1993).

In the second part of the theorem,  $\varepsilon_t$  is a semi-strong multivariate GARCH process. Due to Lemma 5, we only need to show that Condition (14) implies variance noncausality. Note first that

$$\begin{aligned}\text{Var}(\varepsilon_t^{\mathcal{I}} | \mathcal{F}_{t-1}^{\mathcal{I}}) &= \text{E}[\eta_t^{\mathcal{I}} | \mathcal{F}_{t-1}^{\mathcal{I}}] \\ &= \text{E}[\text{E}[\eta_t^{\mathcal{I}} | \mathcal{F}_{t-1}] | \mathcal{F}_{t-1}^{\mathcal{I}}] \\ &= \text{E}[h_t^{\mathcal{I}} | \mathcal{F}_{t-1}^{\mathcal{I}}],\end{aligned}$$

which follows because of  $\mathcal{F}_{t-1}^{\mathcal{I}} \subset \mathcal{F}_{t-1}$ . Thus, using Definition 2 and the measurability of  $h_t$  with respect to  $\mathcal{F}_{t-1}$ , we have that  $\varepsilon_t^{\mathcal{I}} \overset{V}{\rightsquigarrow} \varepsilon_t^{\mathcal{I}}$  is equivalent to

$$\text{E}[h_{t,i} | \mathcal{F}_{t-1}^{\mathcal{I}}] = \text{E}[h_{t,i} | \mathcal{F}_{t-1}] = h_{t,i}, \quad \forall i \in \mathcal{I}^*. \quad (30)$$

Under Assumption 2, the process  $h_t$  can be written as  $h_t = \Phi(1)^{-1}\sigma + (I_{K^*} - \Phi(L)^{-1})\eta_t$ . Denoting  $\Phi(L)^{-1} = \Pi(L) = I_K - \sum_{n=1}^{\infty} \Pi_n L^n$ , we obtain  $h_t = \Phi(1)^{-1}\sigma + \sum_{n=1}^{\infty} \Pi_n \eta_{t-n}$ . Thus, we have  $\forall i \in \mathcal{I}^*$ ,

$$\text{E}[h_{t,i} | \mathcal{F}_{t-1}^{\mathcal{I}}] = [\Phi(1)^{-1}\sigma]_i + \sum_{i' \in \mathcal{I}^*} [\Pi(L)]_{ii'} \eta_{t,i'} + \sum_{j \in \mathcal{J}^*} [\Pi(L)]_{ij} \text{E}[\eta_{t,j} | \mathcal{F}_{t-1}^{\mathcal{I}}]. \quad (31)$$

Under Condition (14), all  $\Phi_n$  are  $0_{\mathcal{I}^*}$  and by Lemma 4 all  $\Pi_n$  are also  $0_{\mathcal{I}^*}$ . Thus, the third term on the right hand side of (31) is zero and (30) holds, which proves the result. Q.E.D.

**Proof of Lemma 2:** For the multivariate GARCH(1,1) model,  $\Phi_n = (A + B)^{n-1}A$ , and Condition (14) becomes  $(A + B)^{n-1}A$  is  $0_{\mathcal{I}^*}$ ,  $\forall n \geq 1$ . The proof then follows the same line of argument as the proof of Lemma 2 in Comte and Lieberman (2000). Q.E.D.

**Proof of Theorem 2:** We first show that (18) implies (19). From the BEKK representation (7), the equivalent vec representation (1) can be obtained by setting  $\omega = \text{vech}(CC')$ ,  $A = D_K^+(A^* \otimes A^*)D_K$  and  $B = D_K^+(B^* \otimes B^*)D_K$ , where  $D_K$  is the duplication matrix and  $D_K^+$  its generalized inverse. Thus, the condition  $\tilde{R}\text{vec}(A) = 0$  is equivalent to  $\tilde{R}(D_K' \otimes D_K^+)\text{vec}(A^* \otimes A^*) = 0$ . Now  $\text{vec}(A^* \otimes A^*)$  can be written as  $(I_K \otimes C_{KK} \otimes I_K)\text{vec}(a^* a^{*'})$ , where  $a^* = \text{vec}(A^*)$  and  $C_{KK}$  is the commutation matrix. (For a definition of  $D_K$  and  $C_{KK}$  see, e.g., Lütkepohl, 1996). Thus,  $\tilde{R}\text{vec}(A) = 0$  is equivalent to  $\tilde{R}Z_K\text{vec}(a^* a^{*'}) = 0$ , with  $Z_K = (D_K' \otimes D_K^+)(I_K \otimes C_{KK} \otimes I_K)$ . The matrix  $\tilde{R}Z_K$  contains in its  $u$ th row a 1 at the  $v$ -th position and zeros elsewhere, where

$u = (j - 1)k^*(K - k) + i + (i - 1)(k - i/2)$  and  $v = \{i + (j - 1)K\}(K^2 + 1) - K^2$ , for  $i \in \mathcal{I}$  and  $j \in \mathcal{J}$ . However,  $v$  is just the index of the  $i + (j - 1)K$ th diagonal element of  $a^*a'$  in the vector  $\text{vec}(a^*a')$ . Thus, the corresponding equation reads  $(A_{ij}^*)^2 = 0$  or, equivalently,  $A_{ij}^* = 0$ . This condition is equivalent to the  $m + (n - 1)k$ th row of condition (19), where  $m$  is the index of  $i$  in  $\mathcal{I}$  and  $n$  is the index of  $j$  in  $\mathcal{J}$ . This holds for all  $i \in \mathcal{I}$  and  $j \in \mathcal{J}$ , which proves that (18) implies (19).

Conversely, assume that  $A_{ij}^* = 0$  for all  $i \in \mathcal{I}$  and  $j \in \mathcal{J}$ . Then,  $A = D_K^+(A^* \otimes A^*)D_K$  is such that  $A_{i^*,j^*} = 0$  for all  $i^* \in \mathcal{I}^*$  and  $j^* \in \mathcal{J}^*$ . But this is equivalent to condition (18). Hence, (19) implies (18) and we have established the equivalence of conditions (19) and (18). Q.E.D.

## Appendix B

In the following, let us give some examples how to find the restriction matrices  $\tilde{R}$  and  $\tilde{Q}$ . Suppose we are interested in the conditions for  $\varepsilon_t^{\mathcal{J}} \rightarrow \varepsilon_t^{\mathcal{I}}$  ( $V$  or  $LV$ ), where only the composition of  $\mathcal{I}$  and  $\mathcal{J}$  change.

1.  $K = 2$ : Let  $\mathcal{I} = \{1\}$  and  $\mathcal{J} = \{2\}$ . Then,  $\mathcal{I}^* = \{1\}$  and  $\mathcal{J}^* = \{2, 3\}$ . The matrix  $\tilde{R}$  is of dimension  $2 \times 9$  and given by

$$\tilde{R} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0, \end{bmatrix}$$

whereas  $\tilde{Q}$  is of dimension  $1 \times 4$ , given by

$$\tilde{Q} = [ 0 \quad 0 \quad 1 \quad 0 ].$$

2.  $K = 3$

- (a) First let  $\mathcal{I} = \{1, 2\}$  and  $\mathcal{J} = \{3\}$ . Then,  $\mathcal{I}^* = \{1, 2, 4\}$  and  $\mathcal{J}^* = \{3, 5, 6\}$ . The matrix  $\tilde{R}$  is of dimension  $9 \times 36$ . The positions of the 1 in the respective rows are given by 13,14,16, 25,26,28,31,32,34.

The matrix  $\tilde{Q}$  is of dimension  $2 \times 9$  and given by

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0, \end{bmatrix}$$

- (b) Now consider the reverse causality direction,  $\mathcal{I} = \{3\}$  and  $\mathcal{J} = \{1, 2\}$ . Then  $\mathcal{I}^* = \{6\}$  and  $\mathcal{J}^* = \{1, 2, 3, 4, 5\}$ . The matrix  $R$  is of dimension  $5 \times 36$ , with a 1 in the respective rows at the positions 6,12,18,24,30.

The matrix  $\tilde{Q}$  is now given by

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0, \end{bmatrix}$$

3.  $K = 4$

- (a) First let  $\mathcal{I} = \{1, 2\}$  and  $\mathcal{J} = \{3, 4\}$ . Then,  $\mathcal{I}^* = \{1, 2, 5\}$  and  $\mathcal{J}^* = \{3, 4, 6, 7, 8, 9, 10\}$ . The matrix  $\tilde{R}$  is of dimension  $21 \times 100$ . The positions of the 1 in the respective rows are given by 21,22,25,31,32,35,51,52,55,61,62,65,71,72,75,81,82,85,91,92,95. The matrix  $\tilde{Q}$  is of dimension  $4 \times 16$  and given by

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

- (b) Now let  $\mathcal{I} = \{1\}$  and  $\mathcal{J} = \{2, 3, 4\}$ . Then  $\mathcal{I}^* = \{1\}$  and  $\mathcal{J}^* = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ . The matrix  $\tilde{R}$  is of dimension  $9 \times 100$ , with positions of the 1 in the respective rows given by 11,21,31,41,51,61,71,81,91.  $\tilde{Q}$  is now  $(3 \times 16)$  and given by

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

## References

- Caporale, G.M., N. Spittis, and N. Spagnolo (2002), Testing for causality-in-variance: An application to the East Asian markets, *International Journal of Finance and Economics*, 7, 235–245.
- Cheung, Y.W. and L.K. Ng (1996), A causality in variance test and its application to financial market prices, *Journal of Econometrics*, 72, 33–48.
- Comte, F. and O. Lieberman (2000), Second order noncausality in multivariate GARCH processes, *Journal of Time Series Analysis* 21, 535–557.
- Comte, F. and O. Lieberman (2003), Asymptotic theory for multivariate GARCH processes, *Journal of Multivariate Analysis* 84, 61–84.
- Drost, F.C. and T.E. Nijman (1993), Temporal aggregation of GARCH processes, *Econometrica*, 61, 909–927.
- Dufour, J.M. and E. Renault (1998), Short run and long run causality in time series: theory, *Econometrica* 66, 1099–1126.
- Engle, R.F. and K.F. Kroner (1995), Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Glosten, L., R. Jagannathan and D. Runkle (1993), Relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

- Granger, C.W.J. (1969), Investigating causal relations by econometric models and cross spectral methods, *Econometrica*, 37, 424-438.
- Granger, C.W.J. (1980), Testing for causality: A personal viewpoint, *Journal of Economic Dynamics and Control*, 2, 329-352.
- Granger, C.W.J. (1988), Some recent developments in a concept of causality, *Journal of Econometrics*, 39, 199-211.
- Granger, C.W.J., R.P. Robins and R.F. Engle (1986). Wholesale and retail prices: Bivariate time series modelling with forecastable error variances, in: D. Belsley and E. Kuh, eds., *Model Reliability*, MIT Press, Cambridge, MA, 1-17.
- Hafner, C.M. (2003), Fourth moment structure of multivariate GARCH models, *Journal of Financial Econometrics* 1, 26-54.
- Hafner, C.M. and H. Herwartz (2003), Analytical quasi maximum likelihood inference in multivariate volatility models, Econometric Institute Report 2003-21, Erasmus University Rotterdam.
- Hafner, C.M. and H. Herwartz (1998). Structural analysis of portfolio risk using beta impulse response functions. *Statistica Neerlandica* 52, 336-355.
- Herwartz, H. and H. Lütkepohl (2000), Multivariate volatility analysis of VW stock prices. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 9, 35-54.
- Haugh, L.D. (1976), Checking the independence of two covariance-stationary time series: A univariate residual correlation approach. *Journal of the American Statistical Association*, 71, 378-385.
- Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, 2nd ed., Berlin, New York: Springer Verlag.
- Lütkepohl, H. (1996), *Handbook of Matrices*, New York: Wiley.
- Kanas, A. and G.P.Kouretas (2002), Mean and variance causality between official and parallel currency markets: Evidence from four Latin American Countries. *The Financial Review*, 37, 137-163.
- Nijman, T. and E. Sentana (1996), Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71, 71-87.
- Pierce, D.A. and L.D.Haugh (1977), Causality in temporal systems. *Journal of Econometrics*, 5, 265-293.

$T$		Size				Power			
		1000	2000	4000	8000	1000	2000	4000	8000
		Gaussian model							
$H_0^{(1)}$	W1: $\mathcal{D}^{-1}$	.0765	.0795	.0640	.0635	.1285	.2135	.4370	.7775
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.1205	.0970	.0730	.0740	.1870	.2530	.4740	.7895
$H_0^{(2)}$	W1: $\mathcal{D}^{-1}$	.0570	.0665	.0605	.0590	.1430	.2825	.5505	.8510
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0925	.0790	.0655	.0620	.1970	.3180	.5630	.8595
$H_0^{(3)}$	W1: $\mathcal{D}^{-1}$	.0555	.0720	.0645	.0560	.0690	.0700	.0580	.0610
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0800	.0835	.0695	.0625	.0990	.0885	.0685	.0610
		$t$ -model							
$H_0^{(1)}$	W1: $\mathcal{D}^{-1}$	.2625	.2950	.2885	.2450	.3505	.5220	.7310	.9250
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.1315	.1110	.0825	.0735	.1905	.2535	.4545	.7370
$H_0^{(2)}$	W1: $\mathcal{D}^{-1}$	.2110	.2250	.2110	.2005	.3335	.5290	.7580	.9405
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0950	.0915	.0810	.0685	.2005	.3310	.5490	.8305
$H_0^{(3)}$	W1: $\mathcal{D}^{-1}$	.1930	.2210	.2030	.1820	.1965	.2140	.1945	.1665
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0885	.0825	.0690	.0595	.0910	.0905	.0755	.0595

Table 1: Size and power estimates for the Wald statistics (W1 and W2).

	Size $T = 1000$			Power $T = 8000$			
	Gaussian model						
	$m = 1$	5	10	1	5	10	df
$H_0^{(1)}$	.0475	.0590	.0545	.0575	.0820	.1105	$2m$
$H_0^{(2)}$	.0500	.0590	.0585	.0625	.1115	.1575	$m$
$H_0^{(3)}$	.0510	.0500	.0480	.0485	.0475	.0460	$m$
	$t$ -model						
$H_0^{(1)}$	.0555	.0665	.0780	.0620	.0915	.1280	$2m$
$H_0^{(2)}$	.0460	.0600	.0650	.0650	.1085	.1565	$m$
$H_0^{(3)}$	.0460	.0625	.0795	.0420	.0465	.0560	$m$

Table 2: Selected size and power estimates for CCF tests of alternative orders  $m$ .  $df$  denotes the degrees of freedom under the null hypothesis.

	Wald				CCF			
$T$	1000	2000	4000	8000	1000	2000	4000	8000
$H_0^{(1)}$	.1285	.2135	.4370	.7775	.0929	.0982	.1088	.1490
$H_0^{(2)}$	.1430	.2825	.5505	.8510	.0767	.1118	.1182	.1868

Table 3: Size adjusted power estimates for the Wald test using  $\hat{\mathcal{D}}^{-1}$  as covariance estimator (W1) and the CCF test ( $m = 10$ ) under conditional normality of the DGP.



$T$		Size estimates				Power estimates			
		1000	2000	4000	8000	1000	2000	4000	8000
		True BEKK order $S = 2$							
$H_0^{(1)}$ :	W1: $\mathcal{D}^{-1}$	.0820	.0815	.0625	.0645	.1070	.1690	.3260	.6220
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.1210	.0995	.0735	.0765	.1715	.2020	.3475	.6335
$H_0^{(2)}$ :	W1: $\mathcal{D}^{-1}$	.0605	.0670	.0620	.0590	.1135	.2230	.4165	.7155
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0950	.0800	.0675	.0610	.1540	.2485	.4450	.7300
$H_0^{(3)}$ :	W1: $\mathcal{D}^{-1}$	.0585	.0730	.0655	.0575	.0655	.0710	.0585	.0605
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0830	.0850	.0710	.0630	.0940	.0855	.0700	.0615
		Bivariate threshold GARCH							
$H_0^{(1)}$ :	W1: $\mathcal{D}^{-1}$	.0590	.0640	.0645	.0790	.0540	.0725	.0975	.1905
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0840	.0880	.0740	.0815	.0710	.0875	.1000	.1880
$H_0^{(2)}$ :	W1: $\mathcal{D}^{-1}$	.0280	.0455	.0555	.0700	.0335	.0630	.1230	.2525
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0560	.0570	.0615	.0720	.0610	.0775	.1355	.2510
$H_0^{(3)}$ :	W1: $\mathcal{D}^{-1}$	.0280	.0435	.0475	.0655	.0275	.0405	.0500	.0675
	W2: $\mathcal{S}^{-1}\mathcal{D}\mathcal{S}^{-1}$	.0475	.0460	.0580	.0650	.0405	.0520	.0565	.0675

Table 4: Size and power estimates for the Wald statistics (W1 and W2) under misspecification of the (quasi) log-likelihood function. Upper block: Underestimation of the BEKK order  $S$ . Lower block: Asymmetry of the underlying volatility process.

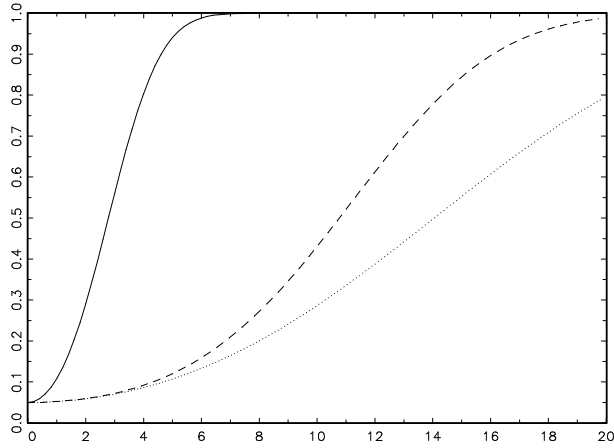


Figure 1: *Asymptotic power functions of the Wald test (solid), the CCF test with  $m = 1$  (dotted) and the CCF test with  $m = 10$  (dashed) assuming Gaussian innovations. The abscissa is the lower left element of the BEKK matrix  $A^*$  multiplied by  $\sqrt{T}$ .*

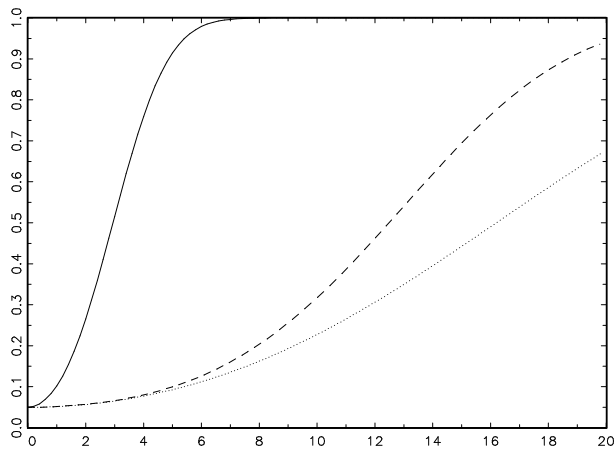


Figure 2: *Asymptotic power functions of the Wald test (solid), the CCF test with  $m = 1$  (dotted) and the CCF test with  $m = 10$  (dashed) assuming  $t_8$  distributed innovations. The abscissa is the lower left element of the BEKK matrix  $A^*$  multiplied by  $\sqrt{T}$ .*