Temporal Aggregation of Multivariate GARCH Processes

Christian M. Hafner[1]

Econometric Institute Report EI 2004–29

Abstract

This paper derives results for the temporal aggregation of multivariate GARCH processes in the general vector specification. It is shown that the class of weak multivariate GARCH processes is closed under temporal aggregation. Fourth moment characteristics turn out to be crucial for the low frequency dynamics for both stock and flow variables. It is shown that spurious instantaneous causality in variance will only appear in degenerated cases, but that spurious Granger causality will be more common. Forecasting volatility, it is generally advisable to aggregate forecasts of the disaggregate series rather than forecasting the aggregated series directly, and unlike for VARMA processes the advantage does not diminish for large forecast horizons. Results are derived for the distribution of multivariate realized volatility if the high frequency process follows multivariate GARCH. Finally, the estimation problem is discussed. A numerical example illustrates some of the results.

Keywords: multivariate GARCH, temporal aggregation, causality in variance, volatility forecasts, realized volatility

JEL Classification: C22

[1]Econometric Institute, Erasmus University Rotterdam, P.O.B. 1738, 3000 DR Rotterdam, The Netherlands. e-mail `chafner@few.eur.nl`.

1 Introduction

Financial time series such as stock prices or exchange rates usually are available on very high frequencies such as minute by minute. Typically, however, the econometrician uses highly aggregated data such as daily or weekly returns. This poses the question how the low frequency dynamics depend on the characteristics of the high frequency process. It is an important general topic in econometrics whenever the sample frequency does not correspond to the 'natural' frequency, where the natural frequency of financial time series is so high that the series is often represented by continuous time stochastic processes.

For financial time series in discrete time, the GARCH modelling class has proved to be successful to describe the volatility. Drost and Nijman (1993) have derived the low frequency parameters if the high frequency dynamics follows univariate GARCH. However, they also show that only a weak version of GARCH is closed under temporal aggregation, that is, GARCH does not explain the conditional variance but rather the best linear prediction in terms of lagged returns and lagged squared returns. Meddahi and Renault (2004) extend the weak GARCH model to a class of autoregressive stochastic volatility models that is closed under temporal aggregation. Their model is characterized by multi-period conditional moment conditions that allow for estimation and inference by the generalized method of moments. Also, it is less restrictive in terms of moment conditions. However, due to their simplicity GARCH models remain the principal volatility model used in econometric practice, and its widespread implementation guarantees a need for thorough understanding of its theoretical properties. This is even more so in the multivariate case, since multivariate GARCH models also start to become a standard in statistical and econometric programming packages. Other multivariate volatility models such as multivariate stochastic volatility quickly become intractable in empirical work. Throughout the paper I will use the so-called vec form of multivariate GARCH, as introduced by Bollerslev, Engle, and Wooldridge (1988). It nests the so-called BEKK model of Engle and Kroner (1995) that has been introduced mainly to overcome some practical disadvantages of the vec model. It also nests the factor ARCH models introduced by Diebold and Nerlove (1989) and Engle, Ng, and Rothschild (1990), as well as the orthogonal GARCH model of Alexander (2001) and its generalization by van der Weide (2002). However, it does not nest the constant conditional correlation (CCC) model of Bollerslev (1990) or its extension, the dynamic conditional correlation (DCC) model of Engle (2002).

Due to their nonlinear character, it will be difficult to derive aggregation results for both of these models. For a recent review of the various multivariate GARCH specifications, see Bauwens, Laurent and Rombouts (2003).

This paper extends the results of Drost and Nijman (1993) to the multivariate case. Mainly, I show that the class of weak multivariate GARCH processes is closed under temporal aggregation and provide formulae how to to obtain the low frequency dynamics for a given high frequency process. I make use of some well known aggregation results of VARMA models. However, there are important differences that occur in multivariate GARCH models compared to VARMA models. This is mainly due to the fact that in GARCH models it is not the second order process, i.e. the squared returns, that is aggregated but the returns themselves. This creates cross-products and therefore additional noise in the aggregated series. The variance and auto-covariance of this additional noise affects the dynamics of the aggregated series. Distinguishing between stock and flow variables, there appears a major difference between univariate and multivariate GARCH processes: Whereas in the univariate case only the aggregated flow variable process depends on the fourth moment characteristics, so does also the aggregated stock variable process in the multivariate case.

Further to the derivation of the low frequency dynamics, I discuss some issues related to causality in volatility. In VARMA processes, Breitung and Swanson (2002) investigate the phenomenon of *spurious instantaneous causality*, that is, instantaneous causality of the low frequency process that is solely induced by temporal aggregation without any causal relationship at the high frequency. For multivariate GARCH processes, I show that such misleading causality can be ruled out whenever there is a nonzero conditional correlation between the series, or if the dimension is not larger than two. Spurious Granger causality, i.e. uni- or bi-directional causality, is of more practical relevance, since if the parameter matrices of the high frequency process are diagonal (i.e. no Granger causality), those of the low frequency will in general not be diagonal. However, as measures for causality suggest, this spurious Granger causality is typically much smaller than the instantaneous causality. All Granger causality in volatility disappears as the series is more and more aggregated. Moreover, the normalized series converges to a multivariate Gaussian white noise series with increasing aggregation level.

For the prediction of volatility, it is no surprise that the method that predicts the disaggregate process and then aggregates the forecasts has a smaller mean square prediction error than the method that directly predicts the aggregated series. In the VARMA framework this has been demonstrated e.g. by Lütkepohl (1987). However, whereas in VARMA models the two methods become identical when the prediction horizon increases, this is not the case for multivariate GARCH processes. The reason is the additional noise terms, referred to above, in the aggregated series which are absent in the aggregation of VARMA processes.

Finally, I try to build a link to the increasing literature on so-called realized volatilities, that is, aggregation of the high-frequency (typically intra-day) second order process to obtain a measure rather than a model for the low frequency volatility, see e.g. Andersen et al. (2003). Based on results of Breitung and Swanson (2002), it can be shown that if the high frequency process follows multivariate GARCH, then the multivariate realized volatility process for finite but large aggregations can be approximated by a VMA(1) process.

The paper is organized as follows. Section 2 introduces the notation, some definitions and preliminary results such as the fourth moment structure of multivariate GARCH processes. Section 3 derives the main results of the paper, where I distinguish between the cases of stock and flow variables. Section 4 discusses the causality in volatility and Section 5 the prediction of volatility. Section 6 derives results for realized volatility. Finally, Section 7 discusses the estimation problem, and Section 8 concludes. Throughout the paper I use a numerical example to illustrate the results. Proofs of the theorems are given in the appendix.

2 Preliminaries

To begin with, the notion of vector white noise is at the core of most multivariate stochastic processes, but it is often defined in three alternative ways. In the context of modelling the conditional mean the exact notion of white noise has not been of much interest and importance. For the study of temporal aggregation of multivariate GARCH processes, however, the distinction of these definitions will turn out to be crucial.

Definition 1 (White Noise) *Let* $\{u_t, t \in \mathbb{Z}\}$ *denote a stochastic vector process of dimension* K. *We say that* u_t *is*

1. *strong white noise, if it is i.i.d. with* $E[u_t] = 0$ *and* $E[u_t u_t'] = \Sigma_u < \infty$,

2. *semi-strong white noise, if* $E[u_t \mid \mathcal{F}_{t-1}] = 0$ *and* $E[u_t u_t'] = \Sigma_u < \infty$*, where* $\mathcal{F}_t = \sigma(u_s, -\infty < s \leq t)$*,*

3. *weak white noise, if* $E[u_t] = 0$*,* $E[u_t u_s'] = 0$*,* $\forall t \neq s$*, and* $E[u_t u_t'] = \Sigma_u < \infty$*.*

A semi-strong white noise process can be characterized as a martingale difference. Processes that build on martingale differences are not closed under temporal aggregation, see e.g. Meddahi and Renault (2004), and it is therefore important to consider the weak white noise process. Before turning to GARCH processes it is convenient to define three versions of vector autoregressive moving average (VARMA) processes based on the above white noise notions.

Definition 2 (VARMA) *Let* $\{y_t, t \in \mathbb{Z}\}$ *be a stochastic process given by*

$$y_t = \nu + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{j=0}^{q} \Theta_j u_{t-j},$$

where u_t *is a white noise vector process,* ν *is a* K *dimensional parameter vector,* Φ_i *and* Θ_j *are square parameter matrices of order* K*, and where we set* $\Theta_0 = I_K$*. Then* y_t *is called a*

1. *strong VARMA*(p,q) *process if* u_t *is strong white noise,*

2. *semi-strong VARMA*(p,q) *if* u_t *is semi-strong white noise, and*

3. *weak VARMA*(p,q) *if* u_t *is weak white noise.*

VARMA processes are widely known to be closed under temporal aggregation, but in fact this holds only for weak VARMA processes, see the monograph by Lütkepohl (1987). Analogous to the above definitions we now consider three versions of multivariate GARCH processes. [1]

[1]Throughout the paper, vec denotes the operator that stacks all columns of a matrix into a vector, and vech denotes the operator that stacks only the lower triangular part including the diagonal of a symmetric matrix into a vector.

Definition 3 (Multivariate GARCH) *Let ε_t denote a stochastic vector process with K components and $E[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$. Now define a positive definite and symmetric matrix H_t such that vech$(H_t) = h_t$ and where the stochastic vector process h_t has the representation*

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

where $\omega = \text{vech}(\Omega)$, $\eta_t = \text{vech}(\varepsilon_t \varepsilon_t')$ and $N \times N$ parameter matrices Ω, A_i, B_j, with $N = K(K+1)/2$. Then we say that ε_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 Var$(\varepsilon_t \mid \mathcal{F}_{t-1}) = H_t$, where $\mathcal{F}_t = \sigma(\varepsilon_s, -\infty < s \leq t)$,*

3. *weak multivariate GARCH(p, q) process, if h_t is the best linear predictor of η_t in terms of a constant and lagged values of η_t, that is*

$$h_t = P(\eta_t \mid \mathcal{H}_{t-1}) = [P(\eta_{t,1} \mid \mathcal{H}_{t-1}), \ldots, P(\eta_{t,N} \mid \mathcal{H}_{t-1})]'$$

where $\mathcal{H}_t = sp\{1, \eta_{t-\tau,1}, \ldots, \eta_{t-\tau,N}, \tau \geq 0\}$ denotes the infinite dimensional Hilbert space spanned by all linear combinations of a constant and $\eta_{t-\tau,1}, \ldots, \eta_{t-\tau,N}$.

Note that 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, consider the process

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

where $Q_i = A_i + B_i$, $u_t = \eta_t - h_t$ and where we set $A_{q+1} = \ldots = A_p = 0$ if $p > q$ and $B_{p+1} = \ldots = B_q = 0$ if $q > p$. Roughly speaking, (2) is a VARMA process if u_t is white noise with finite covariance matrix, which we assume in the following. The type of VARMA process of η_t depends on the type of GARCH process of ε_t and will be made more precise in Proposition 2.

Assumption 1 *$E[|\varepsilon_t|^{4+\delta}] \leq b < \infty$ for some $\delta > 0$ and for all $t \in \mathbb{Z}$, where $|\cdot|$ denotes the Euclidean norm.*

Proposition 1 *Under Assumption 1, the following matrices exist:*

$$\Sigma = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[\varepsilon_t \varepsilon_t'], \tag{3}$$

$$\Sigma_\eta = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[\eta_t \eta_t'], \tag{4}$$

$$\Sigma_h = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[h_t h_t'], \tag{5}$$

$$\Sigma_u = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[u_t u_t']. \tag{6}$$

Note that Σ_η, Σ_h and Σ_u are positive semi-definite. To ensure that they are strictly positive definite we make the following assumption.

Assumption 2 *The matrices Σ, Σ_η, Σ_h and Σ_u have full rank.*

If ε_t is semi-strong multivariate GARCH, then $\Sigma_u = \Sigma_\eta - \Sigma_h$. This follows directly by writing out the expectations and applying the law of iterated expectations.

In semi-strong and strong GARCH(p, q) processes, Σ exists if and only if ε_t is covariance stationary. This is the case if and only if all eigenvalues of the matrix $\sum_{i=1}^{\max(p,q)} Q_i$ have modulus smaller than one, see Engle and Kroner (1995). The unconditional covariance matrix $\Sigma = \mathrm{Var}(\varepsilon_t)$ would then be given by

$$\sigma = \mathrm{vech}(\Sigma) = \left(I_N - \sum_{i=1}^{\max(p,q)} Q_i \right)^{-1} \omega, \tag{7}$$

where the $(N \times 1)$ vector σ contains the K unconditional variances and the $K(K-1)/2$ unconditional covariances of ε_t.

We now have the following result.

Proposition 2 *Under Assumption 1, if $\{\varepsilon_t\}$ is*

1. *strong or semi-strong multivariate GARCH(p, q), then $\{u_t\}$ is semi-strong white noise, which means that $\{\eta_t\}$ in (2) follows a semi-strong VARMA$(\max(p, q), p)$ process.*

2. *weak multivariate GARCH(p, q), then $\{u_t\}$ is weak white noise, which means that $\{\eta_t\}$ in (2) follows a weak VARMA($\max(p, q), p$) process.*

It should be emphasized that a strong multivariate GARCH process only permits a semi-strong VARMA representation for η_t given by (2). The same holds for a semi-strong multivariate GARCH process, whereas for a weak multivariate GARCH(p, q) process, (2) is only weak VARMA, and H_t is not necessarily the conditional variance matrix of ε_t.

The next assumption will be useful for proving asymptotic normality of the aggregated process. It restricts the type of temporal dependence of (ε_t).

Assumption 3 *The process $(\varepsilon_t, t \in \mathbb{Z})$ is α-mixing.*

In the univariate context, Drost and Nijman (1993) define weak GARCH models as h_t being the projection on a constant and lagged η_t, but also on lagged ε_t. However, the orthogonality of the projection error u_t w.r.t. lagged ε_t is not a necessary requirement to obtain a GARCH model that is closed under temporal aggregation. It is true that, without further assumption, the weak GARCH model as defined in Definition 3 is not closed under temporal aggregation of flow variables. As it will become clear in the next section, what is needed is the following assumption on the structure of fourth moments of (ε_t).

Assumption 4

$$E[vec(\varepsilon_t \varepsilon_{t-i}')vec(\varepsilon_t \varepsilon_{t-j}')] = 0, \quad \forall i, j \geq 0, \quad i \neq j \tag{8}$$

A sufficient condition for (8) to hold is that all conditional skewness and co-skewness measures are zero, i.e., $\mathrm{E}[\eta_t \varepsilon_t' \mid \mathcal{F}_{t-1}] = 0$, and that there is no leverage effect, that is, the conditional variance of ε_t is conditionally uncorrelated to all lagged ε_t, $\mathrm{E}[\eta_t \varepsilon_{t-i}' \mid \mathcal{F}_{t-i-1}] = 0, \forall i \geq 1$.

To derive the autocovariance structure of η_t it is convenient to work with the pure vector moving average (VMA(∞)) representation of η_t. From the VARMA representation (2) we obtain

$$\eta_t = \sigma + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \tag{9}$$

where the $N \times N$ matrices Φ_i can be determined recursively by $\Phi_0 = I_N$,

$$\Phi_i = -B_i + \sum_{j=1}^{i} Q_j \Phi_{i-j}, \quad i = 1, 2, \ldots, \tag{10}$$

see Lütkepohl (1993, pp. 220). From (9) we see directly that $\mathrm{E}[\eta_t] = \sigma$ and $\mathrm{Var}(\eta_t) = \sum_{i=0}^{\infty} \Phi_i \Sigma_u \Phi_i'$, whereas the autocovariance matrix is given by

$$\begin{aligned} \Gamma(\tau) &= \mathrm{E}\left[(\eta_t - \sigma)(\eta_{t-\tau} - \sigma)'\right] \\ &= \sum_{i=0}^{\infty} \Phi_{\tau+i} \Sigma_u \Phi_i'. \end{aligned} \tag{11}$$

Using the notation $\Sigma_\eta = \mathrm{E}[\eta_t \eta_t']$ we can also write $\Gamma(0) = \Sigma_\eta - \sigma\sigma'$ for the unconditional variance matrix of η_t. In Section 3 we will also need the following structure of fourth moments,

$$\widetilde{\Gamma}(\tau) = \mathrm{E}[D_K^+ \mathrm{vec}(\varepsilon_t \varepsilon_{t-\tau}) \mathrm{vec}(\varepsilon_t \varepsilon_{t-\tau})' D_K^{+,\prime}] \tag{12}$$

which using Lemma 2 in the appendix is linked to $\Gamma(\tau)$ by

$$\mathrm{vec}(\widetilde{\Gamma}(\tau)) = \mathcal{G}_K \mathrm{vec}(\Gamma(\tau) + \sigma\sigma'), \tag{13}$$

where the matrix $\mathcal{G}_K$ is square of order N^2 and given by

$$\mathcal{G}_K = (D_K^+ \otimes D_K^+)(I_K \otimes C_{KK} \otimes I_K)(D_K \otimes D_K), \tag{14}$$

with D_m and C_{mn} denoting the duplication and commutation matrices, respectively, and where $D_m^+ = (D_m' D_m)^{-1} D_m'$.

Assumption 1 implies finiteness of Σ_u. However, to determine Σ_u numerically one has to specify further how u_t is generated. For all numerical calculations in this paper I assume that the disaggregate process is strong multivariate GARCH with innovations $\xi_t = H_t^{-1/2} \varepsilon_t$ that belong to the spherical class of distributions. This is to obtain numerical values for Σ_u and is not necessary for the validity of the temporal aggregation results. If other ways are found how to determine Σ_u for other distributions or even for not strong multivariate GARCH processes, these could be used here equally well. Thus, to calculate Σ_u I assume that the disaggregated process ε_t is strong multivariate GARCH with innovations ξ_t whose distribution belongs to the class of spherical distributions with finite fourth moments. Spherical distributions include the multinormal and multivariate t distributions as special cases. They are characterized by the fact that the density is a function of ξ_t only through $\xi_t' \xi_t$. See Fang, Kotz and Ng (1989) for a monograph on spherical distributions. All moments of spherical distributions containing odd orders are zero and the marginal distributions (which are all the same) have fourth moments $\mathrm{E}[\xi_{t,i}^4]$

that are linked to the co-kurtosis $c = \mathrm{E}[\xi_{t,i}^2 \xi_{t,j}^2], i \neq j$ by $\mathrm{E}[\xi_{t,i}^4] = 3c$. This follows by Lemma 2. For example, for a multinormal distribution $c = 1$, and for a multivariate t distribution with ν degrees of freedom $c = (\nu - 2)/(\nu - 4)$ if $\nu > 4$. It can be argued that if the disaggregated process is sampled on a sufficiently high frequency, then it could well approximate a diffusion process with Wiener innovations (whose distribution over discrete time intervals is multi-normal). Another implication is that Assumption 4 is satisfied.

Proposition 3 *If* (ε_t) *is strong multivariate GARCH with spherical innovations, then (8) holds.*

Since strong GARCH with spherical innovations is a quite strong assumption, we only use it when the calculation of Σ_u is of interest, but the weaker Assumption 4 if the temporal aggregation result is of interest for a given Σ_u.

Finiteness of fourth moments of ξ_t is necessary for a finite covariance matrix of u_t, Σ_u, but it is not sufficient. Recall that for semi-strong multivariate GARCH, $\Sigma_u = \Sigma_\eta - \Sigma_h$, so that Σ_u exists if and only if Σ_η and Σ_h exist. The following simple relationship between Σ_η and Σ_h holds under sphericity of ξ_t,

$$\mathrm{vec}(\Sigma_\eta) = c(2\mathcal{G}_K + I_{N^2})\mathrm{vec}(\Sigma_h), \tag{15}$$

where $\mathcal{G}_K$ is given by (14) and $c = \mathrm{E}[\xi_{t,1}^4]/3$, by Theorem 1 of Hafner (2003). Thus, it suffices to consider the condition for a finite Σ_η. Theorem 2 of Hafner (2003) states that under spherical innovations, Σ_η is finite if and only if all eigenvalues of the matrix $\sum_{i=1}^\infty (\Phi_i \otimes \Phi_i)\{2c\mathcal{G}_K + (c-1)I_{N^2}\}$ have modulus smaller than one. In that case, the vectorized matrix of fourth moments of ε_t is given by

$$\mathrm{vec}(\Sigma_\eta) = c(2\mathcal{G}_K + I_{N^2}) \left(I_{N^2} - \sum_{i=1}^\infty (\Phi_i \otimes \Phi_i)\{2c\mathcal{G}_K + (c-1)I_{N^2}\} \right)^{-1} \mathrm{vec}(\sigma\sigma'). \tag{16}$$

Consequently, we obtain for Σ_u

$$\mathrm{vec}(\Sigma_u) = \{2c\mathcal{G}_K + (c-1)I_{N^2}\} \left(I_{N^2} - \sum_{i=1}^\infty (\Phi_i \otimes \Phi_i)\{2c\mathcal{G}_K + (c-1)I_{N^2}\} \right)^{-1} \mathrm{vec}(\sigma\sigma'). \tag{17}$$

Simpler expressions for the often used GARCH(1,1) model are readily available. It should be emphasized that a correct understanding of the fourth moment structure will turn out to be essential for the study of temporal aggregation.

Example 1 *To illustrate the results we will use the following bivariate example process throughout the paper.*

$$\begin{aligned} \varepsilon_t &= H_t^{1/2}\xi_t, \quad \xi_t \sim i.i.d.N(0, I_2), \qquad (18) \\ vech(H_t) = h_t &= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.16 & 0.08 & 0.01 \\ 0 & 0.12 & 0.03 \\ 0 & 0 & 0.09 \end{bmatrix} \eta_{t-1} + \begin{bmatrix} 0.64 & 0 & 0 \\ 0 & 0.72 & 0 \\ 0 & 0 & 0.81 \end{bmatrix} h_{t-1} \end{aligned}$$

This process is stationary with maximum eigenvalue of Q equal to 0.9. Fourth moments are finite as the maximum eigenvalue of the matrix $\sum_{i=1}^{\infty}(\Phi_i \otimes \Phi_i)\{2c\mathcal{G}_K + (c-1)I_{N^2}\}$ is 0.8262. The unconditional covariance matrix is $\sigma = (6.25, 1.875, 10)'$, so that $\rho = 0.237$. The unconditional kurtosis of $\varepsilon_{t,1}$ is 4.17, that of $\varepsilon_{t,2}$ is 3.28, and the unconditional co-kurtosis is 1.4. The normal kurtosis and co-kurtosis is 3 and $1+2\rho^2 = 1.1125$, respectively, so there is excess kurtosis and excess co-kurtosis. One issue to be investigated is how kurtosis and co-kurtosis change when the series is temporally aggregated. Note that for this example process the conditional variance of the second component of ε_t is only affected by its own squared lagged values, and therefore one can speak of absence of causality from the first to the second component in volatility. Section 4 formalizes this and discusses the impact of temporal aggregation on causality.

3 Temporal aggregation

In order to keep the notation simple I will only discuss temporal aggregation of multivariate GARCH(1,1) models. Most empirical applications use models of this order and it is in the tradition of Drost and Nijman (1993). Thus, in the following I consider the multivariate GARCH$(1,1)$ model, where the best linear predictor of η_t in terms of a constant and $\eta_{t-1}, \eta_{t-2}, \ldots$, is given by

$$h_t = P(\eta_t \mid \mathcal{H}_{t-1}) = \omega + A\eta_{t-1} + Bh_{t-1}.$$

Recall from (2) that η_t has the VARMA(1,1) representation

$$\eta_t = \omega + Q\eta_{t-1} - Bu_{t-1} + u_t, \qquad (19)$$

where $Q = A + B$ and $u_t = \eta_t - h_t$.

We will look at two types of aggregation that are typically used in the case of stock and flow variables. By far more relevant is the case of flow variables, e.g. when financial returns are under study, whereas stock variables are easier to analyze. Denote the process ε_t that is aggregated over m periods by $\{\varepsilon_{mt}^{(m)}, t \in \mathbb{Z}\}$ which is then given by

$$\varepsilon_{mt}^{(m)} = \begin{cases} \varepsilon_{mt}, & \text{stock variables} \\ \varepsilon_{mt} + \varepsilon_{mt-1} + \ldots + \varepsilon_{mt-m+1}, & \text{flow variables.} \end{cases}$$

Since ε_t is a white noise process, it follows immediately that the unconditional variances of the aggregated process $\varepsilon_{mt}^{(m)}$ are Σ in the case of stock variables and $m\Sigma$ in the case of flow variables, where $\mathrm{vech}(\Sigma)$ is given in (7). This implies that in both cases the unconditional correlation matrix remains unchanged under temporal aggregation.

Now denote by $\eta_{mt}^{(m)} = \mathrm{vech}(\varepsilon_{mt}^{(m)} \varepsilon_{mt}^{(m)\prime})$ the vector process that contains the squares and cross-products of the aggregated process $\varepsilon_{mt}^{(m)}$. Since for arbitrary vectors a and b of dimension K, $\mathrm{vech}(ab') + \mathrm{vech}(ba') = 2D_K^+ \mathrm{vec}(ab')$, we have

$$\eta_{mt}^{(m)} = \begin{cases} \eta_{mt}, & \text{stock variables} \\ \eta_{mt} + \eta_{mt-1} + \ldots + \eta_{mt-m+1} + w_{mt}^{(m)}, & \text{flow variables.} \end{cases} \tag{20}$$

where, using the lag operator $L^k x_t = x_{t-k}$,

$$w_{mt}^{(m)} = 2D_K^+ \left\{ \sum_{i=0}^{m-2} L^i \mathrm{vec}(\varepsilon_{mt} \varepsilon_{mt-1}') + \sum_{i=0}^{m-3} L^i \mathrm{vec}(\varepsilon_{mt} \varepsilon_{mt-2}') + \cdots + \mathrm{vec}(\varepsilon_{mt} \varepsilon_{mt-m+1}')\} \right\} \tag{21}$$

For example, if $m = 2$ then $w_{2t}^{(2)} = 2D_K^+ \mathrm{vec}(\varepsilon_{2t} \varepsilon_{2t-1})$. Each term of $w_{mt}^{(m)}$ has expectation zero and due to Assumption 4 it is uncorrelated with every other term. Thus, it acts as a noise term that is added to the sum of the high frequency second order process η_t. It turns out that this noise complicates the analysis of temporal aggregation when compared with VARMA processes where this term is missing. See however Section 6 for the approach of realized volatility that suppresses this term and thus aims at aggregating not the returns but rather volatility directly. For later reference and recalling equation (12), we can calculate the variance matrix of $w_{mt}^{(m)}$, $\Sigma_w^{(m)}$ say, as

$$\Sigma_w^{(m)} = 4 \sum_{i=1}^{m-1} (m-i) \widetilde{\Gamma}(i), \tag{22}$$

where $\widetilde{\Gamma}(i)$ is given by (13).

The proof of Theorem 1 in the appendix shows that the aggregated process $\eta_{mt}^{(m)}$ has the following VARMA representation,

$$(I_N - Q^m L)\eta_{mt}^{(m)} = \omega^{(m)} + v_{mt}^{(m)}, \tag{23}$$

where

$$\omega^{(m)} = \begin{cases} (I_N + Q + \ldots + Q^{m-1})\omega, & \text{for stock variables} \\ m(I_N + Q + \ldots + Q^{m-1})\omega, & \text{for flow variables} \end{cases} \tag{24}$$

and $v_{mt}^{(m)}$ is a vector moving average process of order one, that is, it has expectation zero, finite covariance matrix $\Sigma_v^{(m)} = \mathrm{E}[v_{mt}^{(m)} v_{mt}^{(m)\prime}]$, first order autocovariance matrix $\Gamma_v^{(m)} = \mathrm{E}[v_{mt}^{(m)} v_{m(t-1)}^{(m)\prime}]$, and higher order autocovariances equal to zero. By convention, the lag operator in (23) that operates on an aggregated process lags it on the low frequency scale, that is, $L\eta_{mt}^{(m)} = \eta_{m(t-1)}^{(m)}$. [2] The coefficient matrix of the autoregressive part is given by Q^m. Assumption 1 implies that all eigenvalues of Q have modulus smaller than one, so that Q^m converges to the zero matrix exponentially fast. However, if the largest eigenvalue of Q^m is very close to unity, then it may require a large aggregation level m for the autoregressive part to become negligible.

The moving average part is more difficult to obtain and depends on the particular type of aggregation. For the case of stock variables it takes the form

$$v_{mt}^{(m)} = \sum_{i=0}^{m} J_i^s u_{mt-i} \tag{25}$$

where

$$\begin{aligned} J_0^s &= I_N \\ J_i^s &= Q^{i-1}A, \quad i = 1, \ldots, m-1 \\ J_m^s &= -Q^{m-1}B. \end{aligned}$$

From (25) we obtain immediately the form of the variance and autocovariances of $v_{mt}^{(m)}$,

$$\Sigma_v^{(m)} = \sum_{i=0}^{m} J_i^s \Sigma_u J_i^{s\prime} \tag{26}$$

$$\Gamma_v^{(m)} = J_m^s \Sigma_u. \tag{27}$$

[2] Alternatively, one could define $L\eta_{mt}^{(m)} = \eta_{mt-1}^{(m)}$ and replace L in (23) by L^m.

For the case of flow variables the moving average term takes the form

$$v_{mt}^{(m)} = \sum_{i=0}^{2m-1} J_i^f u_{mt-i} + w_{mt}^{(m)} - Q^m w_{m(t-1)}^{(m)}. \tag{28}$$

The J_i^f matrices are determined as follows:

$$\begin{aligned}
J_0^f &= I_N \\
J_i^f &= I_N + A + QA + \cdots + Q^{i-1}A, \quad i = 1, \ldots, m-1 \\
J_m^f &= \{I_N + Q + \cdots + Q^{m-2}\}A - Q^{m-1}B \\
J_i^f &= \{Q^{i-m} + Q^{i-m+1} + \cdots + Q^{m-2}\}A - Q^{m-1}B, \quad i = m+1, \ldots, 2m-2 \\
J_{2m-1}^f &= -Q^{m-1}B
\end{aligned}$$

Note that J_i^f can also be calculated recursively as $J_i^f = J_{i-1}^f + Q^{i-1}A$ for $i = 1, \ldots, m-1$, and as $J_i^f = J_{i-1}^f - Q^{i-m-1}A$ for $i = m+1, \ldots, 2m-1$.

From equation (28) we obtain the variance and first order auto-covariance of $v_{mt}^{(m)}$ as

$$\Sigma_v^{(m)} = \sum_{i=0}^{2m-1} J_i^f \Sigma_u J_i^{f\prime} + \Sigma_w^{(m)} + Q^m \Sigma_w^{(m)} (Q')^m \tag{29}$$

$$\Gamma_v^{(m)} = \sum_{i=0}^{m-1} J_{i+m}^f \Sigma_u J_i^{f\prime} - Q^m \Sigma_w^{(m)} \tag{30}$$

where $\Sigma_w^{(m)}$ is the variance matrix of $w_{mt}^{(m)}$ given in (22).

The following theorem summarizes the main result.

Theorem 1 *Under Assumptions 1, 2 and 4, the class of weak multivariate GARCH(1,1) processes is closed under temporal aggregation. By Definition 3, this means that for the aggregated process* $\varepsilon_{mt}^{(m)}$, $E[\varepsilon_{mt}^{(m)} \mid \mathcal{F}_{m(t-1)}^{(m)}] = 0$, *where* $\mathcal{F}_{mt}^{(m)} = \sigma(\varepsilon_{ms}^{(m)}, -\infty < s \leq t)$. *Moreover,* $h_{mt}^{(m)} = P(\eta_{mt}^{(m)} \mid \mathcal{H}_{m(t-1)}^{(m)})$, *with* $\mathcal{H}_{mt}^{(m)} = sp(1, \eta_{m(t-\tau),1}^{(m)}, \ldots, \eta_{m(t-\tau),N}^{(m)}, \tau \geq 0)$, *and where*

$$h_{mt}^{(m)} = \omega^{(m)} + A^{(m)} \eta_{m(t-1)}^{(m)} + B^{(m)} h_{m(t-1)}^{(m)}, \tag{31}$$

where $\omega^{(m)}$ *is given by (24),* $B^{(m)}$ *is given by the solution to the system of quadratic equations*

$$B^{(m)} \Gamma_v^{(m)} B^{(m)\prime} + B^{(m)} \Sigma_v^{(m)} + \Gamma_v^{(m)} = 0, \tag{32}$$

where the matrices $\Sigma_v^{(m)}$ *and* $\Gamma_v^{(m)}$ *are given by (26) and (27) for the case of stock variables and by (29) and (30) for the case of flow variables,* $A^{(m)}$ *is given by*

$$A^{(m)} = Q^m - B^{(m)}, \tag{33}$$

and where the projection error $\{u_{mt}^{(m)}, t \in \mathbb{Z}\}$, $u_{mt}^{(m)} = \eta_{mt}^{(m)} - h_{mt}^{(m)}$, *is a weak white noise vector process with covariance matrix* $\Sigma_u^{(m)}$ *with*

$$vec(\Sigma_u^{(m)}) = (I_{N^2} + B^{(m)} \otimes B^{(m)})^{-1} vec(\Sigma_v^{(m)}). \tag{34}$$

Introducing the notation $Q^{(m)} = A^{(m)} + B^{(m)}$, it follows from (33) that $Q^{(m)} = Q^m$. Using Proposition 2, we immediately obtain the following corollary.

Corollary 1 *The aggregated process* $\eta_{mt}^{(m)}$ *follows a weak VARMA(1,1) process that can be written as*

$$\eta_{mt}^{(m)} = \omega^{(m)} + Q^{(m)} \eta_{m(t-1)}^{(m)} - B^{(m)} u_{m(t-1)}^{(m)} + u_{mt}^{(m)}, \tag{35}$$

Theorem 1 shows how the parameter matrices of the aggregated process can be obtained from the high frequency process. The matrices $\Sigma_v^{(m)}$ and $\Gamma_v^{(m)}$ given by (26) and (27) and by (29) and (30), respectively, are functions of the matrices A, B, and Σ_u and thus can be calculated if the high frequency process is known. As for $B^{(m)}$, (32) is a system of nonlinear equations that can not be solved explicitly. The analysis of existence and uniqueness of solutions for (32) goes beyond the scope of the present paper, but is certainly important for future research. In practice any numerical search algorithm will work well. In all investigated situations with stationary high frequency processes, I found that convergence to a solution is very fast if the disaggregate process is not too close to the stationarity boundary and not too close to a white noise process. Also, the solutions were unique under the restriction of invertibility, that is, all eigenvalues of $B^{(m)}$ smaller than one in modulus. [3]

Note that equation (32) can be directly compared to equation (10) of Drost and Nijman (1993) for the univariate case. We can vectorize equation (32) and write for the case of stock variables

$$\left[(B^{(m)} \otimes B^{(m)} + I_{N^2})(I_N \otimes Q^{m-1} B) + (I_N \otimes B^{(m)}) \sum_{i=0}^{m} J_i^s \otimes J_i^s \right] \mathrm{vec}(\Sigma_u) = 0 \tag{36}$$

[3] A computer program is available for download at `http://www.few.eur.nl/few/people/chafner/`.

In the univariate case, Σ_u (which is linked to the fourth moment structure) is a positive scalar so that it can be dropped from (36). A solution then just solves the term in squared brackets being zero. In the multivariate case, however, (36) may hold even if the term in squared brackets is not zero. The implication of this is that, in general, the low frequency parameters depend on the fourth moment characteristics even in the case of stock variables. This is different from the univariate case, where this dependence occurred only for flow variables.

In the following let us look at the case of flow variables, the practically more relevant one. One interesting aspect of the aggregated series is its fourth moment structure, in particular the kurtosis of each marginal series. We expect these kurtosis measures to decline eventually towards 3 as m increases. But it will turn out later that the kurtosis can actually increase for small values of m, before it decreases. The matrix of fourth moments of the aggregated process is given by

$$\Sigma_\eta^{(m)} = \mathrm{E}[\eta_{mt}^{(m)}\eta_{mt}^{(m)\prime}] = m\Sigma_\eta + \Sigma_w^{(m)} + \sum_{i=1}^{m-1}(m-i)\left\{\Gamma(i) + \Gamma(i)' + 2\sigma\sigma'\right\} \tag{37}$$

The first two terms on the right hand side of (37) are the sum of the variances of each individual term of $\eta_{mt}^{(m)}$, whereas the third term arises because of the non-zero covariance between η_t and $\eta_{t-\tau}$ for $\tau \neq 0$ given in (11). This allows to compute the kurtosis and co-kurtosis of the aggregated series. The following theorem states that excess kurtosis and co-kurtosis disappear under temporal aggregation, a fact that in the univariate case has already been shown by Diebold (1988).

Theorem 2 *Under Assumptions 1 to 4, conditional heteroskedasticity, excess kurtosis and excess co-kurtosis of the aggregated process* $\varepsilon_{mt}^{(m)} = \varepsilon_{mt} + \varepsilon_{mt-1} + \cdots + \varepsilon_{mt-m+1}$ *disappear asymptotically as* $m \longrightarrow \infty$. *Moreover,*

$$m^{-1/2}\varepsilon_{mt}^{(m)} \xrightarrow{\mathcal{D}} N\left(0, \Sigma\right)$$

Figure 1 shows the kurtosis and co-kurtosis of the example process (18) as a function of the aggregation level m. Both kurtosis coefficients converge to 3, whereas the co-kurtosis converges to $1 + 2\rho^2 = 1.1125$. Note however the slow rate of convergence with still substantial excess kurtosis and excess co-kurtosis at $m = 50$. Moreover, it is remarkable that both kurtosis and co-kurtosis increase for small m. Thus, a series may become even more leptokurtic under temporal aggregation, if the aggregation level is small.

From the weak VARMA representation (35) one obtains the weak VMA(∞) representation

$$\eta_{mt}^{(m)} = \sigma^{(m)} + \sum_{i=0}^{\infty} \Phi_i^{(m)} u_{m(t-i)}^{(m)}, \tag{38}$$

where $\sigma^{(m)} = \left(I_N - Q^{(m)}\right)^{-1} \omega^{(m)}$, and where the $N \times N$ matrices $\Phi_i^{(m)}$ are given by $\Phi_0 = I_N$ and

$$\Phi_i^{(m)} = (Q^{(m)})^{i-1} A^{(m)}, \quad i = 1, 2, \ldots, \tag{39}$$

4 Causality

There is a substantial literature on the effects of temporal aggregation for causality between time series, see e.g. Marcellino (1999) for a recent overview and references. The general difficulty in empirical work is that only data of the temporally aggregated series is available, for which one typically observes contemporaneous correlation between the series. The question for the investigator is whether this correlation stems from a true causal relation of the high frequency series or whether it is a mere artefact of temporal aggregation. We will address this issue here in the volatility context and show that, again, there are important differences to the VARMA case.

As is common in econometrics, we use the term causality in the sense of 'Granger causality', which for volatility has been defined by Granger, Robins, and Engle (1984). However, there are at least three alternative versions of Granger causality, one based on the entire distribution of a variable to be forecast, another on the conditional expecation, and yet another on optimal linear forecasts. Knowing from Section 3 that temporally aggregated multivariate GARCH processes are only weak multivariate GARCH, we have to be careful in defining causality in variance, because notions based on conditional expectations or conditional variances become difficult to check for the aggregated series. Rather, one has to weaken the concept and use the notion of best linear predictors, but this stands in the tradition of, for example, Boudjellaba et al. (1992) and Comte and Lieberman (2000). Also, I use the term 'Granger causality' for the case of a causal lag greater than zero (sometimes this is also called 'directional causality'), whereas I use 'instantaneous causality' for the causal lag being actually zero.

Suppose we are interested in the causality in variance between the first two elements of ε_t, $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$. Let us introduce the following notation. Denote the σ-algebra generated by $\varepsilon_{s,i},\ s \leq t,\ i,j = 1,3,4,\ldots,K$ by $\mathcal{F}_t^{(-2)}$. Moreover, denote by $\mathcal{H}_t$ the set of all linear combinations of a constant and $\varepsilon_{s,i}\varepsilon_{s,j},\ s \leq t,\ i,j = 1,\ldots,K$ (as before in Definition 3), by $\mathcal{H}_t^{(-2)}$ the set of all linear combinations of a constant and $\varepsilon_{s,i}\varepsilon_{s,j},\ s \leq t,\ i,j = 1,3,4,\ldots,K$, and by $\mathcal{H}_t^{(+2)}$ the set of all linear combinations of a constant, $\varepsilon_{s,i}\varepsilon_{s,j},\ s \leq t,\ i,j = 1,\ldots,K$, and $\varepsilon_{t+1,2}\varepsilon_{t+1,i},\ i = 2,\ldots,K$.

Definition 4 1. *We say that $\varepsilon_{t,2}$ Granger causes $\varepsilon_{t,1}$ in variance (GCV), denoted by $\varepsilon_{t,2} \overset{GCV}{\to} \varepsilon_{t,1}$ if, for some $h \geq 1$,*

$$Var(\varepsilon_{t+h,1} \mid \mathcal{F}_t) \neq Var(\varepsilon_{t+h,1} \mid \mathcal{F}_t^{(-2)}), \tag{40}$$

2. *There is said to be instantaneous causality in variance (ICV) between $\varepsilon_{t,2}$ and $\varepsilon_{t,1}$, denoted by $\varepsilon_{t,1} \overset{ICV}{\leftrightarrow} \varepsilon_{t,2}$ if*

$$Var(\varepsilon_{t+1,1} \mid \mathcal{F}_t) \neq Var(\varepsilon_{t+1,1} \mid \mathcal{F}_t \vee \sigma(\varepsilon_{t+1,2})) \tag{41}$$

where $\mathcal{F}_t \vee \sigma(\varepsilon_{t+1,2})$ denotes the augmentation of $\mathcal{F}_{t-1}$ by the information contained in $\varepsilon_{t,2}$.

3. *We say that $\varepsilon_{t,2}$ linearly Granger causes $\varepsilon_{t,1}$ in variance (LGCV), denoted by $\varepsilon_{t,2} \overset{LGCV}{\to} \varepsilon_{t,1}$ if, for some $h \geq 1$,*

$$P(\varepsilon_{t+h,1}^2 \mid \mathcal{H}_t) \neq P(\varepsilon_{t+h,1}^2 \mid \mathcal{H}_t^{(-2)}), \tag{42}$$

4. *There is said to be linear instantaneous causality in variance (LICV) between $\varepsilon_{t,2}$ and $\varepsilon_{t,1}$, denoted by $\varepsilon_{t,1} \overset{LICV}{\leftrightarrow} \varepsilon_{t,2}$ if*

$$P(\varepsilon_{t+1,1}^2 \mid \mathcal{H}_t) \neq P(\varepsilon_{t+1,1}^2 \mid \mathcal{H}_t^{(+2)}) \tag{43}$$

For weak multivariate GARCH processes it is only possible to investigate linear causality since the conditional variances are not specified or not known. On the other hand, for semi-strong multivariate GARCH processes it is well possible to investigate causality, but that would only be relevant for the high-frequency process. Absence of either of these causality concepts now amounts to zero restrictions on the parameter matrices. Hafner and Herwartz (2004) give necessary and sufficient conditions for absence of GCV and LGCV.

In temporally aggregated VARMA models, Breitung and Swanson (2002) have investigated the effect of so-called *spurious instantaneous causality*, as first investigated by Renault and Szafarz (1991) and Renault, Sekkat and Szafarz (1998). This occurs if there is no causality between the disaggregated time series, but instantaneous causality between the aggregated time series. We adapt this definition to the volatility case. If there is no causality in volatility (instantaneous or directional) between the series $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$, we denote this by $\varepsilon_{t,1} \overset{CV}{\not\leftrightarrow} \varepsilon_{t,2}$, and correspondingly we write $\varepsilon_{t,1} \overset{LCV}{\not\leftrightarrow} \varepsilon_{t,2}$ if there is no linear causality in volatility (instantaneous or directional) between the series.

Definition 5 1. *There is said to be spurious ICV, if* $\varepsilon_{t,1} \overset{CV}{\not\leftrightarrow} \varepsilon_{t,2}$, *but* $\varepsilon^{(m)}_{mt,1} \overset{ICV}{\leftrightarrow} \varepsilon^{(m)}_{mt,2}$ *for some* $m \geq 2$ *and some* $t \in \mathbb{Z}$.

2. *There is said to be spurious LICV, if* $\varepsilon_{t,1} \overset{LCV}{\not\leftrightarrow} \varepsilon_{t,2}$, *but* $\varepsilon^{(m)}_{mt,1} \overset{LICV}{\leftrightarrow} \varepsilon^{(m)}_{mt,2}$ *for some* $m \geq 2$ *and some* $t \in \mathbb{Z}$.

It has sometimes been argued that spurious instantaneous causality can be problematic in empirical work, since if two aggregated time series are found to show instantaneous causality, it may be because there is causality between the disaggregated series or because it is induced by temporal aggregation. Breitung and Swanson (2003) give sufficient conditions to exclude spurious instantaneous causality in VARMA models. In the volatility case, the following theorem gives a necessary condition for spurious instantaneous causality.

Theorem 3 *If the high frequency process follows strong multivariate GARCH with Gaussian innovations, then a necessary condition for spurious LICV between* $(\varepsilon_{t,1})$ *and* $(\varepsilon_{t,2})$ *is*

$$h_{t,2} = 0 \quad \text{and} \quad K \geq 3,$$

for all t, *where* $h_{t,2}$ *is the second component of* h_t, *i.e. the conditional covariance of* $\varepsilon_{t,1}$ *and* $\varepsilon_{t,2}$.

In the following let us be a bit more loose in terminology and only refer to GCV and ICV when it could also mean LGCV or LICV. Theorem 3 implies that in empirical work spurious ICV is of much less relevance than spurious instantaneous causality in the conditional mean, because the two series will in most cases show some non-zero conditional covariance, be it constant or not. Financial series such as stock returns, for example, tend to be positively correlated at high frequencies. So, ICV will be the rule rather than the exception if high frequency financial series are investigated.

Rather than ICV, it is far more interesting to see whether there is GCV. It turns out that there may be absence of GCV between the disaggregate series, but presence of GCV between the aggregated series. This might be called *spurious Granger causality in volatility*. A sufficient condition for absence of GCV is that the parameter matrices A and B of the multivariate GARCH model are diagonal. Many empirical studies have shown that diagonal GARCH models may give good descriptions of the DGP at many frequencies. This can be due to the fact that even though there may be GCV induced by temporal aggregation, it is possibly much less important numerically than ICV. To see whether this is the case for a given multivariate GARCH model, we need measures for the alternative causalities, which we will look at in the following.

Measures for the causality in variance have been considered by Hafner (2003) based on well known measures for causality in VARMA models introduced by Geweke (1982). For simplicity, I only consider the bivariate case in the following, but extensions to causality measures conditional on other variables follow in analogy to Geweke (1984). Let $x_t = \varepsilon_{t,1}^2$ and $y_t = \varepsilon_{t,2}^2$. By the results of Nijman and Sentana (1996), the marginal process $\varepsilon_{t,1}$ follows a weak univariate GARCH process and therefore x_t has a weak ARMA(q^*, p^*) representation such as

$$x_t = \omega^x + \sum_{i=1}^{q^*} (\alpha_i^x + \beta_i^x) x_{t-i} - \sum_{j=1}^{p^*} \beta_j^x w_{t-j} + w_t, \tag{44}$$

where $w_t = x_t - P(\varepsilon_{t,1}^2 \mid \mathcal{H}_{t-1}^{(-2)})$, ω^x, α_i^x and β_j^x are parameters. Upper bounds for the AR and MA orders are given by $q^* \leq 3$ and $p^* \leq 3$, respectively, by Corollary 4.2.2 of Lütkepohl (1987) or Nijman and Sentana (1996). The process w_t is univariate weak white noise with variance σ_w^2, say. A measure for GCV from y_t to x_t is given by

$$\mathcal{GCV}_{y \to x} = \log \frac{\sigma_w^2}{[\Sigma_u]_{11}}, \tag{45}$$

By symmetry, one obtains a causality measure for the reverse causality direction, $\mathcal{GCV}_{x \to y}$. Summing up these unidirectional causality measures, we can define a measure for bidirectional causality as

$$\mathcal{GCV}_{y \leftrightarrow x} = \mathcal{GCV}_{y \to x} + \mathcal{GCV}_{x \to y} \tag{46}$$

A measure for ICV between x_t and y_t is given by

$$\mathcal{ICV}_{x\leftrightarrow y} = \log \frac{\Sigma_{u,11}\Sigma_{u,33}}{\Sigma_{u,11}\Sigma_{u,33} - \Sigma_{u,13}^2}, \tag{47}$$

Finally, the measure for linear dependence between x_t and y_t is denoted by $\mathcal{CV}_{x,y}$. This measure can be decomposed into the three causality measures:

$$\mathcal{CV}_{x,y} = \mathcal{GCV}_{x\to y} + \mathcal{GCV}_{y\to x} + \mathcal{ICV}_{x\leftrightarrow y} = \mathcal{GCV}_{y\leftrightarrow x} + \mathcal{ICV}_{x\leftrightarrow y}. \tag{48}$$

Now suppose one is mainly interested in the bidirectional GCV measure, $\mathcal{GCV}_{y\leftrightarrow x}$, because, for example, one wants to see how important spurious GCV can become. For example, the hypothesis of a diagonal GARCH model amounts to testing whether this bidirectional measure is zero. For a given multivariate GARCH process there is no obvious way to find the unidirectional measures $\mathcal{GCV}_{y\to x}$ and $\mathcal{GCV}_{x\to y}$, other than via determining the univariate GARCH models for the marginal processes, which is straightforward but tedious, see Nijman and Sentana (1996). However, there is a simple way to find the bidirectional measure $\mathcal{GCV}_{y\leftrightarrow x}$, as we will see immediately. The measure for linear dependence can be decomposed in the frequency domain as

$$\mathcal{CV}_{x,y} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{f_{11}(\lambda)f_{33}(\lambda)}{f_{11}(\lambda)f_{33}(\lambda) - |f_{13}(\lambda)|^2} d\lambda,$$

see e.g. Geweke (1982), where $f(\lambda)$ denotes the spectral density matrix of $\eta_t = \mathrm{vech}(\varepsilon_t\varepsilon_t')$ which is given by

$$f(\lambda) = \left(\sum_{j=0}^{\infty} \Phi_j e^{ij\lambda}\right) \Sigma_u \left(\sum_{j=0}^{\infty} \Phi_j e^{ij\lambda}\right)'. \tag{49}$$

The bidirectional measure $\mathcal{GCV}_{y\leftrightarrow x}$ can now easily be obtained as a residual of equation (48), i.e., by the difference between $\mathcal{CV}_{x,y}$ and $\mathcal{ICV}_{x\leftrightarrow y}$. The advantage of this approach is that $f(\lambda)$ and therefore the bidirectional measure can be calculated directly using the representation of the joint process ε_t. The alternative way of summing up the two unidirectional measures requires the determination of the marginal processes $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$, which is somewhat more involved, see Section 3 of Nijman and Sentana (1996).

The above causality measures can now also be obtained for the aggregated series $\eta_{mt}^{(m)}$ by replacing Σ_u in (49) and (47) by $\Sigma_u^{(m)}$ given in (34) and replacing Φ_i in (49) by $\Phi_i^{(m)}$ given in (39). This gives us a measure of bidirectional causality in volatility for the aggregated series, defined as

$$\mathcal{GCV}_{y\leftrightarrow x}^{(m)} = \mathcal{CV}_{x,y}^{(m)} - \mathcal{ICV}_{x\leftrightarrow y}^{(m)}. \tag{50}$$

Since $\forall i \geq 1$, $\Phi_i^{(m)} \to 0$ as $m \to \infty$, the spectral density matrix of the series $m^{-1}\eta_{mt}^{(m)}$ converges to the limit of $m^{-2}\Sigma_u^{(m)}$, U say. For example, by the results of Section 2, this would be given by $\text{vec}(U) = (cG_K - I_{N^2})\text{vec}(\sigma\sigma')$ under the assumption of spherical innovations. Thus, $\mathcal{CV}_{x,y}^{(m)}$ and $\mathcal{ICV}_{x\leftrightarrow y}^{(m)}$ converge to the same limit given by

$$\lim_{m\to\infty} \mathcal{CV}_{x,y}^{(m)} = \lim_{m\to\infty} \mathcal{ICV}_{x\leftrightarrow y}^{(m)} = \log \frac{U_{11}U_{33}}{U_{11}U_{33} - U_{13}^2}$$

Using (50), this implies that $\lim_{m\to\infty} \mathcal{GCV}_{y\leftrightarrow x}^{(m)} = 0$, meaning that all directional Granger causality in variance disappears eventually as the series is aggregated. This is of course no surprise as it corresponds to the aggregation results in VARMA processes.

Figure 2 shows the alternative causality measures for the example process (18). Clearly, the bidirectional GCV measure is much smaller here than the ICV measure and also dissipates to zero very quickly. Note that the bidirectional GCV measure of the disaggregate process (18) is equal to the unidirectional GCV measure from $\varepsilon_{t,2}$ to $\varepsilon_{t,1}$, since the matrices A and B are upper triangular, so that there is no GCV from $\varepsilon_{t,1}$ to $\varepsilon_{t,2}$. However, the bidirectional GCV measure of the aggregated process incorporates some causality from $\varepsilon_{t,1}$ to $\varepsilon_{t,2}$, although smaller than from $\varepsilon_{t,2}$ to $\varepsilon_{t,1}$. But this is not shown in the figure.

Finally, the discussed causality measures could be used for testing causality for a given empirical time series. If the errors u_t of the VARMA representation of multivariate GARCH models were Gaussian, then an estimate of the $\mathcal{GCV}$ measure, multiplied by the sample size T, would be the usual likelihood ratio statistic, having an asymptotic χ^2 distribution, see Geweke (1982). Now u_t is not Gaussian but skewed and conditionally heteroskedastic. Thus, $T\widehat{\mathcal{GCV}}$ could be called pseudo likelihood ratio statistic with a nonstandard asymptotic distribution. To obtain valid critical values one can use the bootstrap as in Hafner and Herwartz (2004). They find that this statistic has similar size and power properties as the so-called CCF test of Cheung and Ng (1996). The CCF test estimates univariate GARCH models and computes cross-correlations of standardized residuals. It is therefore in the spirit of Lagrange Multiplier statistics. A third way to approach the testing problem is to use Wald type statistics, for example based on QML estimation and inference of the multivariate model. Hafner and Herwartz (2004) show that the Wald test has more power under local alternatives than both the CCF and the pseudo likelihood ratio test. However, their framework is a semi-strong multivariate GARCH model and it is as yet unknown whether this carries over to weak GARCH models. It is related to the problem of estimating weak GARCH models, briefly discussed in Section 7.

5 Forecasting

Suppose one is interested in the prediction of multivariate volatility of the aggregated series h periods ahead. That is, given information at time mt one wants to predict the volatility of $\varepsilon^{(m)}_{m(t+h)}$. Let us only consider the flow variable case here, so that $\varepsilon^{(m)}_{m(t+h)} = \varepsilon_{m(t+h)} + \varepsilon_{m(t+h)-1} + \ldots + \varepsilon_{m(t+h-1)+1}$. Prediction of the volatility of $\varepsilon^{(m)}_{m(t+h)}$ is the same as prediction of $\eta^{(m)}_{m(t+h)}$. One can now build a forecast based on the VMA(∞) representation of the aggregated series in (38). It is given by

$$\eta^{(m)}_{mt}(h) = \sigma^{(m)} + \sum_{i=0}^{\infty} \Phi^{(m)}_{h+i} u^{(m)}_{m(t-i)}$$

The mean square error of this forecast is given by the matrix

$$\Sigma_a(h) = \sum_{i=0}^{h-1} \Phi^{(m)}_i \Sigma^{(m)}_u \Phi^{(m)\prime}_i.$$

Another possibility is to predict the disaggregated series and then aggregate the forecasts. Based on the VMA(∞) representation of the disaggregated series in (9), the optimal r-step forecast in a mean square error sense is given by

$$\eta_t(r) = \sigma + \sum_{i=0}^{\infty} \Phi_{r+i} u_{t-i}$$

The forecast for $\eta^{(m)}_{m(t+h)}$ is then given by $\eta_{mt}(mh) + \eta_{mt}(mh-1) + \ldots + \eta_{mt}(m(h-1)+1)$. The mean square error of this forecast is given by

$$\Sigma_d(h) = F\Sigma_{dm}(h)F',$$

where $F = (I_N, \ldots, I_N)$ is an $(N \times mN)$ aggregation matrix, and $\Sigma_{dm}(h)$ is a symmetric, positive definite $(mN \times mN)$ matrix given by

$$\Sigma_{dm}(h) = \begin{bmatrix} \sum_{i=0}^{m(h-1)} \Phi_i \Sigma_u \Phi'_i & \sum_{i=0}^{m(h-1)} \Phi_i \Sigma_u \Phi'_{i+1} & \cdots & \sum_{i=0}^{m(h-1)} \Phi_i \Sigma_u \Phi'_{i+m-1} \\ \sum_{i=0}^{m(h-1)} \Phi_{i+1} \Sigma_u \Phi'_i & \sum_{i=0}^{m(h-1)+1} \Phi_i \Sigma_u \Phi'_i & \cdots & \sum_{i=0}^{m(h-1)+1} \Phi_i \Sigma_u \Phi'_{i+m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{m(h-1)} \Phi_{i+m-1} \Sigma_u \Phi'_i & \sum_{i=0}^{m(h-1)+1} \Phi_{i+m-2} \Sigma_u \Phi'_i & \cdots & \sum_{i=0}^{mh-1} \Phi_i \Sigma_u \Phi'_i \end{bmatrix},$$

see e.g. chapter 8 of Lütkepohl (1987). There it is also shown that for VARMA models in general $\Sigma_d(h) \leq \Sigma_a(h)$ in the sense that the matrix $\Sigma_a(h) - \Sigma_d(h)$ is positive semi-definite, and that equality only holds in special cases such as periodicity with period equal to the aggregation level. An implication of this result is that the forecasts based on the disaggregated series are superior to the forecasts based on the aggregated series in terms of forecast precision. On the other hand, both forecasts become equivalent as the forecast horizon increases, as both mean square error matrices approach the same unconditional covariance matrix.

For the aggregation of multivariate GARCH processes, however, the difference between both forecasts turns out to be stronger than for VARMA processes and not dissipating for increasing horizons. The reason is the additional noise term in the aggregated series, $w_{mt}^{(m)}$. The expectation of this term is zero, but it has a positive definite covariance matrix $\Sigma_w^{(m)}$ given by (22). Therefore, the unconditional variance of $\eta_{mt}^{(m)}$ is larger than that of $\eta_{mt} + \eta_{mt-1} + \ldots + \eta_{m(t-1)+1}$, and the forecast mean square error matrices converge to two different levels with increasing horizon. Thus, we have a strict inequality, $\Sigma_d(h) < \Sigma_a(h)$ for all $h > 0$. Asymptotically, the difference is given by

$$\lim_{h \to \infty} \Sigma_a(h) - \Sigma_d(h) = \Sigma_w^{(m)}, \tag{51}$$

where $\Sigma_w^{(m)}$ is given by (22). As the difference between the two forecasting methods is negligible in VARMA models for sufficiently large horizons, it turns out to be substantial in multivariate GARCH models. Equation (51) says that in the limit this difference is just given by the variance matrix of the noise term $w_{mt}^{(m)}$ in (21) that was added to the sum of the indivual η_{mt} in constructing the aggregate $\eta_{mt}^{(m)}$. It should be emphasized that this noise term is missing in the aggregation of VARMA processes. The implication of (51) is that forecasting weekly volatility, for example, by aggregating daily volatility forecasts will always be better than forecasting the weekly series directly, no matter how large the forecasting horizon. This is also the reason why in forecasting volatility one should use the highest frequency for which data is available, provided that there are no biases coming from microstructure effects, for example. Recent empirical research has shown that predicting daily volatility of a financial time series using intra-day returns can substantially improve the precision of forecasts using the daily series only, see e.g. Andersen et al. (2003). See also Section 6, where this so-called realized volatility is investigated in the context of multivariate GARCH models.

Figure 3 shows the mean square prediction errors of the two forecasting methods for the example process (18) with $m = 2$. In this example, the mean square prediction error can be reduced by almost 50 % for all forecasting horizons by doubling the sampling frequency and using the high frequency data for prediction.

6 Multivariate realized volatility

There is a growing literature on so-called realized volatilities, see, e.g., Andersen et al. (2003) for an overview. Realized volatilities are estimates of low-frequency volatilities using high frequency data. For example, the volatility of a daily return series could be estimated by the sum of squared intra-day returns. When the sampling frequency goes to infinity, realized volatilities converge to the actual volatility and are therefore consistent, unbiased estimates of daily volatility. In the multivariate context, the same idea applies to the vector of squares and cross-products, $\eta_t = \mathrm{vech}(\varepsilon_t \varepsilon_t')$. The aggregation scheme is no longer $\varepsilon_{mt}^{(m)} = \varepsilon_{mt} + \varepsilon_{mt-1} + \ldots + \varepsilon_{mt-m+1}$ but $\bar{\eta}_{mt} = \eta_{mt} + \eta_{mt-1} + \ldots + \eta_{mt-m+1}$. Thus, all the cross-terms that appeared in our previous aggregation scheme $\eta_{mt}^{(m)} = \mathrm{vech}(\varepsilon_{mt}^{(m)} \varepsilon_{mt}^{(m)\prime})$ are absent here.

First, it is clear that for any finite m, $\bar{\eta}_{mt}$ is an unbiased estimate of the unobservable daily volatility. It is more efficient than the noisy $\eta_{mt}^{(m)} = \mathrm{vech}(\varepsilon_{mt}^{(m)} \varepsilon_{mt}^{(m)\prime})$ but, for every finite m it is inefficient compared to $\bar{h}_{mt} = h_{mt} + h_{mt-1} + \ldots + h_{mt-m+1}$. The practical advantage of using $\bar{\eta}_{mt}$ is, of course, that no parametric model of volatility needs to be specified, but a drawback is given by the restriction that m can not be chosen arbitrarily large. In other words, the time interval between observations can not be arbitrarily small due to market microstructure effects. If the true volatility process follows multivariate GARCH, we quantify below the loss of efficiency of $\bar{\eta}_{mt}$ compared with $\bar{h}_{mt}$.

To calculate the variance of $\bar{\eta}_{mt}$, note that this is just the sum of the variances of the individual terms η_{mt}, each one equal to $\Sigma_\eta - \sigma\sigma'$, plus the sum of all covariances. This is given by

$$\mathrm{Var}(\bar{\eta}_{mt}) = m(\Sigma_\eta - \sigma\sigma') + \sum_{i=1}^{m-1} (m-i)\left(\Gamma(i) + \Gamma(i)'\right).$$

Similarly, we obtain for the variance of $\bar{h}_{mt}$

$$\mathrm{Var}(\bar{h}_{mt}) = m(\Sigma_h - \sigma\sigma') + \sum_{i=1}^{m-1} (m-i)\left(\Gamma(i) + \Gamma(i)'\right), \qquad (52)$$

so that the difference is given by

$$\mathrm{Var}(\bar{\eta}_{mt}) - \mathrm{Var}(\bar{h}_{mt}) = m(\Sigma_\eta - \Sigma_h), \tag{53}$$

which is positive semi-definite. Note that (52) is $O(m^2)$ and (53) is $O(m)$, so that the relative difference between the two variances is $O(m^{-1})$. In other words, the loss of efficiency of realized volatilities w.r.t. the model (supposing that this is correctly specified) is diminishing with rate $O(m^{-1})$. In practice, m can not increase without bounds, so that the relative efficiency for a given m depends on features such as the volatility persistence and the correlation. Let us define the relative efficiency of the i-th component of realized volatility w.r.t. the model as the i-th diagonal element of $\mathrm{Var}(\bar{\eta}_{mt})$ divided by the corresponding diagonal element of $\mathrm{Var}(\bar{h}_{mt})$, that is,

$$RE_i(m) = \frac{[\mathrm{Var}(\bar{\eta}_{mt})]_{ii}}{\left[\mathrm{Var}(\bar{h}_{mt})\right]_{ii}}. \tag{54}$$

Note that $RE_i(m) = 1 + O(m^{-1})$ so that for m sufficiently large the efficiency loss is negligible. However, if m can not be chosen arbitrarily large in practice, the efficiency loss may be substantial. For our example process (18), Table 1 lists the values of $RE_i(m)$ for selected levels m. Obviously, even at $m = 50$ the variance of the realized volatility estimator is still 29% higher than that of the optimal one for the first component of $\eta_{mt}^{(m)}$. For the other two components the loss is even higher. For their exchange rate example, Andersen et al. (2003) use a value of $m = 48$, having half-hourly data for a 24 hours per day market. They can not choose m much larger because of the problems with interfering microstructure effects such as bid-ask bounces. The values of $RE_i(m)$ in Table 1 therefore appear relevant if our example process can be considered as a typical high frequency process. In such a situation the practitioner has to weigh the risk of mis-specifying a parametric volatility model for the high frequency process against the efficiency loss of the nonparametric estimation using realized volatilities.

There is a second issue concerning standardized residuals using realized volatilities which turns out to be intimately related to the relative efficiency issue. Standardized residuals are typically obtained by $\bar{H}_{mt}^{-1/2}\varepsilon_{mt}^{(m)}$, where $\bar{H}_{mt}$ is the de-vectorized $\bar{h}_{mt}$, for the given multivariate GARCH model. Alternatively, without an assumption on the underlying process, one can define standardized residuals by $\Upsilon_{mt}^{-1/2}\varepsilon_{mt}^{(m)}$, where Υ_{mt} is the de-vectorized $\bar{\eta}_{mt}$. Due to the higher variance of $\bar{\eta}_{mt}$ compared to $\bar{h}_{mt}$, the kurtosis of the

m	$RE_1(m)$	$RE_2(m)$	$RE_3(m)$
2	3.2264	4.3470	6.8479
3	2.3839	3.0344	4.4386
4	2.0356	2.5008	3.4832
5	1.8460	2.2121	2.9713
10	1.5075	1.6985	2.0663
20	1.3632	1.4774	1.6745
30	1.3217	1.4127	1.5585
40	1.3030	1.3832	1.5056
50	1.2925	1.3668	1.4763

Table 1: *Relative efficiencies according to definition (54) of realized volatilities with respect to the optimal estimates when the high frequency process is known to be the process given in (18).* RE_1 *is the measure for the conditional variance of* $\varepsilon^{(m)}_{mt,1}$, RE_2 *is the measure for the conditional covariance of* $\varepsilon^{(m)}_{mt,1}$ *and* $\varepsilon^{(m)}_{mt,2}$, *and* RE_3 *is the measure for the conditional variance of* $\varepsilon^{(m)}_{mt,2}$.

residuals standardized by realized volatilities $\bar{\eta}_{mt}$ will be smaller than that of residuals standardized by $\bar{h}_{mt}$. In particular, if the innovation distribution is Gaussian, the kurtosis of the residuals standardized by realized volatilities is smaller than three, which is also apparent in the empirical results of Andersen et al. (2003), Table 1. They claim that standardized residuals are close to being Gaussian, but for their sample of ten years of daily returns on the DM/Dollar exchange rate a value of 2.57 for the kurtosis of standardized residuals is likely to violate the normality assumption.[4] It can also be shown that, using first order expansions, the negative bias of the kurtosis estimate is directly related to the efficiency loss expressed by $RE_i(m)$.

Recently, interest has focused on the distribution of realized volatilities. If the true underlying DGP is multivariate GARCH and m is sufficiently large, this may be approximated by the asymptotic distribution of the centered and normalized realized volatilities,

[4] This can be seen by noting that for an i.i.d. Gaussian white noise, the standard error of the kurtosis estimator is $(24/n)^{1/2}$, where n is the sample size. If $n = 2500$, which roughly corresponds to ten years of daily data, the standard error takes the value 0.098, so that with an estimate of 2.57 one would reject the null hypothesis of Gaussian white noise at the 95% significance level.

which is given in the following theorem.

Theorem 4 *Under Assumptions 1 to 3, the asymptotic distribution of realized volatilities for $m \to \infty$ is given by*

$$m^{-1/2}(\bar{\eta}_{mt} - m\sigma) \xrightarrow{\mathcal{D}} N\left(0, 2\pi f(0)\right)$$

where $f(\lambda)$ is the spectral density matrix of η_t at frequency λ given in (49). Moreover,

$$\lim_{m\to\infty} Cov(\bar{\eta}_{mt}, \bar{\eta}_{m(t+\tau)}) = \begin{cases} \sum_{j=1}^{\infty} \left(\sum_{i=0}^{j} \Phi_i\right) \Sigma_u \left(\sum_{i=j+1}^{\infty} \Phi_i'\right), & \tau = 1 \\ 0, & \tau \geq 2 \end{cases}$$

where Σ_u is given in (17).

An implication of this theorem is that, for m sufficiently large, the centered and normalized realized volatilities may be approximated by a multinormal distribution. However, due to the asymmetric nature of the distribution of volatilities, typically being strongly skewed to the right, it may require very large values of m before the normality result of Theorem 4 applies. In fact, Andersen et al. (2003) find that for moderately large m the distribution of foreign exchange realized volatilities can be well approximated by a log-normal distribution. Further empirical evidence is required to assess how these results depend on the aggregation level m. Also, one may do Monte Carlo simulations to find the distribution of $m^{-1/2}(\bar{\eta}_{mt} - m\sigma)$ for finite m and a known high frequency process such as (18). This is beyond the scope of this paper but interesting for future research.

The second result of Theorem 4 implies that the aggregated process $\bar{\eta}_{mt}$ for large but finite aggregation levels m can be approximated by a VMA(1) process. This is because $\mathrm{Cov}(\bar{\eta}_{mt}, \bar{\eta}_{m(t+\tau)})$ is $O(m)$ for $\tau = 0$, $O(1)$ for $\tau = 1$ and $o(1)$ for $\tau \geq 2$. That is, for $m \to \infty$ the process converges to white noise since the autocorrelations tend to zero, but for finite m the first order autocorrelation will be much larger than higher order autocorrelations. In other words, the vector of realized volatilities can be approximated by a VMA(1) process for large but finite values of m if the underlying DGP is multivariate GARCH. Hence, in practice one may directly specify a VMA(1) model for the realized volatilities for finite but large aggregation level m. Alternatively, one may even use standard model selection procedures to specify a VARMA(p, q) model for the realized volatilities.

7 Estimation

Suppose one has a sample $(\varepsilon_t, t = 1 \ldots, T)$ of observations which are sampled at a low frequency, and that one has correctly specified a weak multivariate GARCH model. Denote the finite dimensional parameter vector characterizing the dynamics of h_t by $\theta \subset \Theta$, where Θ is a compact set.

The quasi maximum likelihood estimator is defined by

$$\hat{\theta}_{QML} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{T} \ln|H_t(\theta)| + \varepsilon_t' H_t(\theta)^{-1}\varepsilon_t$$

conditional on some starting value for H_0. The consistency of QMLE in conditionally heteroskedastic models requires that the first two moments of the process are correctly specified, see e.g. Bollerslev and Wooldridge (1992). For semi-strong multivariate GARCH models, precise conditions for consistency and asymptotic normality of QMLE have been provided by Jeantheau (1998) and Comte and Lieberman (2003), although their models are restricted versions of the vec representation (1). However, aggregated GARCH processes are only weak GARCH, so that the conditional second moment is not correctly specified. There are as yet no results on the theoretical properties of QML estimators for weak GARCH models. In the univariate case, Drost and Nijman (1993) find that the bias is not big. On the other hand, Meddahi and Renault (2004) find it to be more important under high persistence and large aggregation levels.

An alternative to QMLE is nonlinear least squares, which has been proved to be consistent for weak GARCH models by Francq and Zakoian (2000). It is to be conjectured that this carries over to the multivariate case. The nonlinear least squares estimator is defined by

$$\hat{\theta}_{NLS} = \arg\min_{\theta\in\Theta} \sum_{t=1}^{T} u_t(\theta)' u_t(\theta)$$

where for example in the GARCH(1,1) case,

$$u_t(\theta) = \eta_t - \omega - Q\eta_{t-1} + Bu_{t-1}(\theta)$$

conditional on some starting values u_0 and η_0.

Supposing that QML is asymptotically biased and NLS is not, there will be a sample size above which it will always be preferable to use NLS rather than QML, if one takes the mean square error criterion. However, in a simulation experiment Hafner and Rombouts (2004) find that this 'critical sample size' may be much larger than sample sizes typically encountered in practice. In most practical situations one would therefore prefer to use QML rather than NLS. The reason is that NLS has a much higher variance, which for moderately large sample sizes outweighs by far the advantage in terms of bias.

8 Conclusions and Outlook

The main conclusion of this paper is that the class of weak multivariate GARCH processes is closed under temporal aggregation and that the dynamics of the aggregated process can be obtained in a straightforward manner. Although there are many similar results for VARMA processes and univariate GARCH processes, there are also many differences. To recall just two examples, the aggregated process of a stock variable does not depend on the kurtosis in the univariate case, but it depends on the fourth moment structure in the multivariate case. Secondly, the forecasting performance of the method that directly predicts the aggregated process does not become identical to the optimal procedure for increasing horizons. Thus, there is a substantial difference between forecasting a VARMA process and the volatility of a multivariate GARCH process. Concerning realized volatility, it will be important to shed more empirical light on the multivariate distribution of realized volatilities, for which this paper derives an asymptotic result if the high frequency process is multivariate GARCH.

Finally, it will be important to bridge the gap to continuous time processes, as was done in the univariate case by Nelson (1990) and Drost and Werker (1996). This is also left to future research.

Appendix

Lemma 1 *Let* $\Gamma(\tau) = E[(\eta_t - \sigma)(\eta_{t-\tau} - \sigma)']$ *and* $\widetilde{\Gamma}(\tau) = E[D_K^+ vec(\varepsilon_t \varepsilon_{t-\tau}) vec(\varepsilon_t \varepsilon_{t-\tau})' D_K^{+,\prime}]$. *Then*

$$vec(\widetilde{\Gamma}(\tau)) = \mathcal{G}_K vec(\Gamma(\tau) + \sigma\sigma'), \tag{55}$$

where the matrix $\mathcal{G}_K$ *is square of order* N^2 *and given by*

$$\mathcal{G}_K = (D_K^+ \otimes D_K^+)(I_K \otimes C_{KK} \otimes I_K)(D_K \otimes D_K), \tag{56}$$

with D_m *and* C_{mn} *denoting the duplication and commutation matrices, respectively, and where* $D_m^+ = (D_m' D_m)^{-1} D_m'$.

Proof: Follows by making use of the following results for some square matrices A, B, C, D of order N: $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, $\text{vec}(A \otimes B) = (I_N \otimes C_{NN} \otimes I_N)(\text{vec}A \otimes \text{vec}B)$, and $(AC) \otimes (BD) = (A \otimes B)(C \otimes D)$. Thus,

$$\begin{aligned}
\text{vec}(\widetilde{\Gamma}(\tau)) &= (D_K^+ \otimes D_K^+) \text{ vec E}\left[\text{vec}(\varepsilon_t \varepsilon_{t-\tau}')\text{vec}(\varepsilon_t \varepsilon_{t-\tau}')'\right] \\
&= (D_K^+ \otimes D_K^+) \text{ vec E}\left[(\varepsilon_{t-\tau} \otimes \varepsilon_t)(\varepsilon_{t-\tau} \otimes \varepsilon_t)'\right] \\
&= (D_K^+ \otimes D_K^+) \text{ vec E}\left[(\varepsilon_{t-\tau}\varepsilon_{t-\tau}') \otimes (\varepsilon_t \varepsilon_t')\right] \\
&= (D_K^+ \otimes D_K^+)(I_K \otimes C_{KK} \otimes I_K)\text{E}\left[\text{vec}(\varepsilon_{t-\tau}\varepsilon_{t-\tau}') \otimes \text{vec}(\varepsilon_t \varepsilon_t')\right] \\
&= (D_K^+ \otimes D_K^+)(I_K \otimes C_{KK} \otimes I_K)(D_K \otimes D_K)\text{E}\left[\eta_{t-\tau} \otimes \eta_t\right] \\
&= \mathcal{G}_K(\Gamma(\tau) + \sigma\sigma'),
\end{aligned}$$

where $\mathcal{G}_K = (D_K^+ \otimes D_K^+)(I_K \otimes C_{KK} \otimes I_K)(D_K \otimes D_K)$, which proves the lemma. $\square$

Lemma 2 *For any spherical distribution,*

$$E\left[\prod_{j=1}^N X_j^{\alpha_j}\right] = \begin{cases} 0 & \textit{if one (or more) } \alpha_j \textit{ is odd} \\ K_\alpha \prod_{j=1}^N \frac{\alpha_j!}{(\alpha_j/2)!} & \textit{if all } \alpha_j \textit{ are even} \end{cases}$$

where $\alpha = \sum_{j=1}^N \alpha_j$ *and* K_α *depends on* α *only.*

Proof: see Box and Hunter (1957).

Proof of Proposition 1

First, by Jensen's inequality, $\text{E}[|\varepsilon_t|^2] \leq \text{E}[|\varepsilon_t|^4]^{1/2} \leq \sqrt{b} < \infty$. This holds for all t, so that Σ, the limit of time averages of second moment matrices, exists. Second, Assumption

1 implies that all fourth order moments $\mathrm{E}[\varepsilon_{t,i}^4]$, $i = 1, \ldots, K$, exist. A double application of the Cauchy-Schwartz inequality leads to $\mathrm{E}[|\varepsilon_{t,i}\varepsilon_{t,j}\varepsilon_{t,k}\varepsilon_{t,l}|] \leq (\mathrm{E}[\varepsilon_{t,i}^4]\mathrm{E}[\varepsilon_{t,j}^4]\mathrm{E}[\varepsilon_{t,k}^4]\mathrm{E}[\varepsilon_{t,l}^4])^{1/4} \leq b < \infty$, $\forall i, j, k, l = 1, \ldots, K$. Thus $\mathrm{E}[\eta_t \eta_t']$ exists for all t and Σ_η must be finite. Third, h_t is a linear combination of lagged η_t where the weights are absolutely summable as implied by Assumption 1. Thus, $\mathrm{E}[h_t h_t']$ is finite if $\mathrm{E}[\eta_t \eta_{t-i}']$ is finite for all i. This follows again by Cauchy-Schwartz, i.e., for $j, k = 1, \ldots, N$, we have $\mathrm{E}[|\eta_{t,j}\eta_{t-i,k}|] \leq (\mathrm{E}[\eta_{t,j}^2]\mathrm{E}[\eta_{t-i,k}^2])^{1/2} \leq b < \infty$. As this holds for all t, Σ_h must be finite. The same argument applies to $\mathrm{E}[u_t u_t']$, writing out the expectation with $u_t = \eta_t - h_t$, so that Σ_u is finite. □

Proof of Proposition 2

If ε_t is semi-strong multivariate GARCH, then by definition $\mathrm{E}[\eta \mid \mathcal{F}_{t-1}] = h_t$ and, thus, $\mathrm{E}[u_t \mid \mathcal{F}_{t-1}] = \mathrm{E}[\eta_t - h_t \mid \mathcal{F}_{t-1}] = 0$. Thus, u_t is a martingale difference which finite fourth moments by Assumption 1. This is semi-strong white noise according to Definition 1. This proves the first part of the statement.

If ε_t is weak multivariate GARCH, then by definition, the projection error u_t is orthogonal to all $\eta_{t-\tau}$, $\tau \geq 1$. Since $u_t = \eta_t - h_t$ is a linear combination of current and lagged η_t, this implies that u_t is also orthogonal to all $u_{t-\tau}$, $\tau \geq 1$. This corresponds to our definition of weak white noise in Definition 1. If u_t is weak white noise, then η_t has a weak VARMA representation according to Definition 2. This completes the proof of the second part. □

Proof of Proposition 3

By the law of iterated expecations, (8) holds if $\forall i \geq 1$, $\mathrm{E}[\eta_t \varepsilon_{t-i}'] = 0$. As strong GARCH models are also semi-strong GARCH, we have $h_t = \mathrm{E}[\eta_t \mid \mathcal{F}_{t-1}]$ and, thus, $\mathrm{E}[\eta_t \varepsilon_{t-i}'] = \mathrm{E}[h_t \varepsilon_{t-i}']$. Considering first the case $i = 1$ and GARCH(1,1), this can be written as $\mathrm{E}[h_t \varepsilon_{t-1}'] = \mathrm{E}[\omega \varepsilon_{t-1}'] + A\mathrm{E}[\eta_{t-1}\varepsilon_{t-1}'] + B\mathrm{E}[h_{t-1}\varepsilon_{t-1}']$. the first and third of these expectations are zero due to the martingale difference property of ε_t. The second expectation contains terms of the form $\mathrm{E}[\varepsilon_{tj}\varepsilon_{tk}\varepsilon_{tl}]$, $j, k, l = 1, \ldots, K$. If, as assumed, innovations ξ_t are i.i.d. spherical, then the conditional distribution of ε_t is elliptic, whose odd-order moments are all zero, see e.g. Fang et al. (1989). Applying this argument recursively for $i \geq 2$ then proves that $\mathrm{E}[h_t \varepsilon_{t-i}'] = 0$ for all i. The same argument applies to GARCH models of higher order. □

Proof of Theorem 1

First, the aggregated series $\varepsilon_{mt}^{(m)}$ is a martingale difference w.r.t the information set $\mathcal{F}_{m(t-1)}^{(m)}$, because $\mathcal{F}_{m(t-1)}^{(m)} \subset \mathcal{F}_{m(t-1)}$ and, by the law of iterated expectations, $\mathrm{E}[\varepsilon_{mt-j} \mid$

$\mathcal{F}_{m(t-1)}] = 0$ for $j = 1, \ldots, m-1$, and therefore also $\mathrm{E}[\varepsilon_{mt}^{(m)} \mid \mathcal{F}_{m(t-1)}] = 0$.

To prove the volatility part, I will use so-called macro processes, based on the discussion for VARMA models in Lütkepohl (1987, Chapter 6). The advantage of this approach is that it allows for considering temporal and contemporaneous aggregation in a joint framework. Recalling the notation $Q = A + B$, the VARMA representation in (2) can be rewritten as the macro process

$$A_0 \tilde{\eta}_{mt} = \tilde{\omega} + A_1 \tilde{\eta}_{m(t-1)} + M_0 \tilde{u}_{mt} + M_1 \tilde{u}_{m(t-1)} \tag{57}$$

with the $(mN \times mN)$ matrices

$$A_0 = \begin{bmatrix} I_N & 0 & 0 & \cdots & 0 \\ -Q & I_N & 0 & \cdots & 0 \\ 0 & -Q & I_N & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_N \end{bmatrix} \quad A_1 = \begin{bmatrix} 0 & 0 & \cdots & Q \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

$$M_0 = \begin{bmatrix} I_N & 0 & 0 & \cdots & 0 \\ -B & I_N & 0 & \cdots & 0 \\ 0 & -B & I_N & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_N \end{bmatrix} \quad M_1 = \begin{bmatrix} 0 & 0 & \cdots & -B \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

and with the $(mN \times 1)$ vectors

$$\tilde{\omega} = \begin{bmatrix} \omega \\ \vdots \\ \omega \end{bmatrix}, \quad \tilde{\eta}_{mt} = \begin{bmatrix} \eta_{m(t-1)+1} \\ \eta_{m(t-1)+2} \\ \vdots \\ \eta_{mt} \end{bmatrix}, \quad \tilde{u}_{mt} = \begin{bmatrix} u_{m(t-1)+1} \\ u_{m(t-1)+2} \\ \vdots \\ u_{mt} \end{bmatrix}.$$

After multiplying both sides of (57) from the left by the inverse of A_0 one obtains

$$A(L)\tilde{\eta}_{mt} = A_0^{-1}\tilde{\omega} + M(L)\tilde{u}_{mt} \tag{58}$$

with

$$A(L) = \begin{bmatrix} I_N & 0 & \cdots & -QL \\ 0 & I_N & \cdots & -Q^2 L \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_N - Q^m L \end{bmatrix}$$

$$M(L) = \begin{bmatrix} I_N & 0 & \cdots & -BL \\ A & I_N & \cdots & -QBL \\ QA & A & \cdots & -Q^2BL \\ \vdots & \vdots & \ddots & \vdots \\ Q^{m-2}A & Q^{m-3}A & \cdots & I_N - Q^{m-1}BL \end{bmatrix},$$

Now, denoting the block-adjoint of $A(L)$ by $A(L)^*$, which is given by

$$A(L)^* = \begin{bmatrix} I_N - Q^mL & 0 & \cdots & QL \\ 0 & I_N - Q^mL & \cdots & Q^2L \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_N \end{bmatrix},$$

we can write

$$A(L)^*A(L) = \mathrm{diag}(I_N - Q^mL),$$

where $\mathrm{diag}(X)$ denotes a block diagonal matrix with matrices X on the diagonal. Multiplying both sides of (58) from the left by $A(L)^*$, we obtain

$$\mathrm{diag}(I_N - Q^mL)\tilde{\eta}_{mt} = A(L)^*A_0^{-1}\tilde{\omega} + Z(L)\tilde{u}_{mt}, \tag{59}$$

with $Z(L) = A(L)^*M(L)$. First, note that the constant is given by

$$A(L)^*A_0^{-1}\tilde{\omega} = \begin{bmatrix} (I_N + Q + \ldots + Q^{m-1})\omega \\ \vdots \\ (I_N + Q + \ldots + Q^{m-1})\omega \end{bmatrix}.$$

Next, the matrix $Z(L)$ determines the moving average term and is given by

$$Z(L) = \begin{bmatrix} I_N - Q^{m-1}BL & Q^{m-2}AL & Q^{m-3}AL & \cdots & AL \\ A & I_N - Q^{m-1}BL & Q^{m-2}AL & \cdots & QAL \\ QA & A & I_N - Q^{m-1}BL & \cdots & Q^2AL \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q^{m-2}A & Q^{m-3}A & Q^{m-4}A & \cdots & I_N - Q^{m-1}BL \end{bmatrix}. \tag{60}$$

The matrices on the block diagonal of $Z(L)$ are all $I_N - Q^{m-1}BL$. The matrices on the j-th sub-diagonal are all $Q^{j-1}A$, and the matrices on the j-th super-diagonal are all $Q^{m-j-1}AL$, $j = 1, \ldots, m-1$.

The idea is now to represent temporal aggregation of the process ε_t as a linear transformation of the macro process $\tilde{\eta}_t$. Let us define the $(N \times mN)$ transformation matrix F by

$$F = \begin{cases} (0, 0, \cdots, I_N), & \text{for stock variables} \\ (I_N, I_N, \cdots, I_N), & \text{for flow variables} \end{cases} \tag{61}$$

Then, by definition of $\eta_{mt}^{(m)}$ in (20), $\eta_{mt}^{(m)} = F\tilde{\eta}_t$ for stock variables and $\eta_{mt}^{(m)} = F\tilde{\eta}_t + w_{mt}^{(m)}$ for flow variables. Now, multiplying both sides of (59) from the left by F we obtain the following VARMA representation for the aggregated process,

$$(I_N - Q^m L)\eta_{mt}^{(m)} = \omega^{(m)} + v_{mt}^{(m)}, \tag{62}$$

where

$$\omega^{(m)} = \begin{cases} (I_N + Q + \ldots + Q^{m-1})\omega, & \text{for stock variables} \\ m(I_N + Q + \ldots + Q^{m-1})\omega, & \text{for flow variables} \end{cases}$$

and

$$v_{mt}^{(m)} = \begin{cases} FZ(L)\tilde{u}_{mt}, & \text{for stock variables} \\ FZ(L)\tilde{u}_{mt} + (I_N - Q^m L)w_{mt}^{(m)}, & \text{for flow variables} \end{cases} \tag{63}$$

First, from (62) the coefficient matrix of the autoregressive part is given by Q^m. To see what the moving average part is, one has to determine the matrix $FZ(L)$. In the case of stock variables this is just the last block-row of $Z(L)$, whereas for flow variables one needs to construct the column-wise sums of $Z(L)$. From the structure of $Z(L)$ given in (60) one easily finds the expressions given in (25) and (28), respectively. From these expressions it follows that on the low frequency time scale $v_{mt}^{(m)}$ has a VMA(1) representation since the first order autocorrelation is different from zero whereas all higher order autocorrelations are zero. Thus, we can write

$$v_{mt}^{(m)} = -B^{(m)}u_{m(t-1)}^{(m)} + u_{mt}^{(m)}$$

where $u_{mt}^{(m)}$ is a weak white noise vector process with variance matrix $\Sigma_u^{(m)}$, say. The variance and autocovariance matrices of the VMA(1) process $v_{mt}^{(m)}$ are given by

$$\Sigma_v^{(m)} = \Sigma_u^{(m)} + B^{(m)}\Sigma_u^{(m)}B^{(m)\prime} \tag{64}$$

$$\Gamma_v^{(m)} = -B^{(m)}\Sigma_u^{(m)} \tag{65}$$

which can be reduced to $B^{(m)}\Gamma_v^{(m)}B^{(m)\prime} + B^{(m)}\Sigma_v^{(m)} + \Gamma_v^{(m)} = 0$, which proves (32). The equation for $A^{(m)}$ in (33) follows by noting that the coefficient matrix of the autoregressive part, $A^{(m)} + B^{(m)}$, in the representation (35) has to be equal to the corresponding matrix of the representation (62), Q^m. Finally, (34) follows directly by vectorizing (64). This completes the proof. $\square$

Proof of Theorem 2

For all $k > 0$, $\lim_{m\to\infty} J_{m+k} = 0$, and therefore $\lim_{m\to\infty} \Gamma_v^{(m)} = 0$, so that the system of equations (32) reduces asymptotically to $B^{(\infty)}\Sigma_v^{(\infty)} = 0$, where $B^{(\infty)} = \lim_{m\to\infty} B^{(m)}$ and $\Sigma_v^{(\infty)} = \lim_{m\to\infty} \Sigma_v^{(m)}$. Since $\Sigma_v^{(\infty)}$ is positive definite, this can only hold if $B^{(\infty)} = 0$. It follows that $A^{(\infty)} = \lim_{m\to\infty} A^{(m)} = 0$. This shows that asymptotically, conditional heteroskedasticity disappears.

Vectorizing (37) and using (13),

$$m^{-2}\mathrm{vec}\left(\Sigma_\eta^{(m)}\right) = m^{-1}\mathrm{vec}(\Sigma_\eta)+2\sum_{i=1}^{m-1}\frac{m-i}{m^2}\left\{[2\mathcal{G}_K + D_K]\mathrm{vec}(\Gamma(i)) + (2\mathcal{G}_K + I_{N^2})\mathrm{vec}(\sigma\sigma')\right\}$$

Notice that the term $(2\mathcal{G}_K + D_K)\sum_{i=1}^{m-1}\frac{m-i}{m^2}\mathrm{vec}(\Gamma(i))$ is $O(m^{-1})$ and $\sum_{i=1}^{m-1}\frac{m-i}{m^2} = 1/2 + O(m^{-1})$. Thus, $m^{-2}\mathrm{vec}\left(\Sigma_\eta^{(m)}\right) = (2\mathcal{G}_K + I_{N^2})\mathrm{vec}(\sigma\sigma') + O(m^{-1})$. Now the structure of $\mathcal{G}_K$ is such that $k_{ii}(\bar{\varepsilon}_{mt}) = 3 + O(m^{-1})$ and $k_{ij}(\bar{\varepsilon}_{mt}) = 1 + 2\rho_{ij}^2 + O(m^{-1})$, see also Hafner (2003). Thus, asymptotically the kurtosis and co-kurtosis are given by 3 and $1+2\rho_{ij}^2$, the values of a normal distribution with correlation ρ_{ij}.

Because ε_t is a martingale difference sequence, we invoke a central limit theorem for square integrable martingale difference sequences. In the multivariate case this is given e.g. by Theorem 10.1 of Pötscher and Prucha (1997). Their first condition is that ε_t is L_0-approximable by some α-mixing process. This holds trivially as ε_t is itself α-mixing by Assumption 3. Their second condition $\sup_T T^{-1}\sum_{t=1}^{T}\mathrm{E}[|\varepsilon_t|^{2+\delta}] < \infty$ for some $\delta > 0$ is fulfilled by noting that for every t, $\mathrm{E}[|\varepsilon_t|^{2+\delta}] \le \sqrt{b} < \infty$ by Assumption 1. This proves asymptotic normality of the aggregated process. $\square$

Proof of Theorem 3:

First, $\varepsilon_{t,1} \overset{LICV}{\not\leftrightarrow} \varepsilon_{t,2}$ is equivalent to $[\Sigma_u]_{13} = 0$, by Proposition 2.3 of Lütkepohl (1993, p.40). Now $[\Sigma_u]_{13} = \mathrm{E}[\varepsilon_{t,1}^2\varepsilon_{t,2}^2] - \mathrm{E}[h_{t,1}h_{t,3}]$. Under the assumption of conditional normality, the first term is given by $\mathrm{E}[\varepsilon_{t,1}^2\varepsilon_{t,2}^2] = \mathrm{E}[h_{t,1}h_{t,3} + 2h_{t,2}^2]$ by Theorem 1 of Hafner (2003). Thus, $[\Sigma_u]_{13} = 0$ is equivalent to $h_{t,2} = 0$. But if $h_{t,2} = 0$, $\varepsilon_{t,1} \overset{GCV}{\not\leftrightarrow} \varepsilon_{t,2}$, and $K = 2$, then the diagonality of the matrices A, B, and Σ_u implies also diagonality of the matrices $\Sigma_v^{(m)}$ and $\Gamma_v^{(m)}$, and therefore $A^{(m)}$, $B^{(m)}$, and $\Sigma_u^{(m)}$. Thus, if $\varepsilon_{t,1} \overset{LCV}{\not\leftrightarrow} \varepsilon_{t,2}$ and $K = 2$, then we

also have $\varepsilon^{(m)}_{mt,1} \overset{LCV}{\not\leftrightarrow} \varepsilon^{(m)}_{mt,2}$. Hence, spurious LICV can only appear if $h_{t,2} = 0$ and $K \geq 3$.
$\square$

Proof of Theorem 4:

The aggregated process $\bar{\eta}_{mt}$ has a weak finite order VARMA representation that is stationary and invertible. Thus, it also has a linear VMA(∞) representation, for which Breitung and Swanson (2003) have shown the asymptotic results for $m^{-1}\text{Var}(\bar{\eta}_{mt})$ and $\text{Cov}(\bar{\eta}_{mt}, \bar{\eta}_{m(t+\tau)})$, $\tau \geq 1$. The asymptotic normality follows similar to Proposition 3.3 of Lütkepohl (1993). The formulae for $f(\lambda)$ and for Σ_u have been derived by Hafner (2003).
$\square$

Acknowledgments

The author would like to thank Jörg Breitung, Rob Engle, Nour Meddahi, and Jeroen Rombouts for helpful discussions. The paper has been presented at the Quantitative Finance seminar of Humboldt University Berlin, January 2003, the Workshop on Econometric Time Series Analysis - Methods and Applications, September 2003, Linz, the North American Winter Meeting of the Econometric Society, San Diego, January 2004, and the European Meeting of the Econometric Society, Madrid, August 2004. Valuable comments of seminar participants are gratefully acknowledged. Of course, the author assumes full scientific responsibility.

References

Alexander, C.O. (2001). *Orthogonal GARCH*, in *Mastering Risk*, Volume II, edited by C.O. Alexander, pp. 21–38, Prentice Hall.

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

Bauwens, L., S. Laurent, and J.V.K. Rombouts (2003). Multivariate GARCH models: A survey, CORE Discussion paper 31/2003, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

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

Bollerslev, T., R.F. Engle and J.M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96, 116–131.

Bollerslev, T., J.M. Wooldridge (1992). Quasi–maximum likelihood estimation and inference in dynamic models with time–varying covariances, *Econometric Reviews* 11, 143–172.

Boudjellaba, H., J.-M. Dufour and R. Roy (1992). Testing causality between two vectors in multivariate autoregressive moving average models, *Journal of the American Statistical Association* 87, 1082–1090.

Box, G.E.P. and J.S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces, *Annals of Mathematical Statistics*, 28, 195–241.

Breitung, J. and N.R. Swanson (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis* 23, 651 - 665.

Cheung, Y.W., 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.

Diebold, F.X. (1988). *Empirical Modeling of Exchange Rates*, Springer Verlag, New York.

Diebold, F.X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics* 4, 1–21.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.

Drost, F. C. and B. J. M. Werker (1996). Closing the GARCH gap: continuous time GARCH modeling. *Journal of Econometrics* 74, 31–57.

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

Engle, R.F., V.M. Ng and M. Rothschild (1990). Asset pricing with a factor-ARCH structure: Empirical estimates for treasury bills, *Journal of Econometrics* 45, 213–237.

Engle, R.F. (2002). Dynamic conditional correlation - A simple class of multivariate GARCH models, *Journal of Business and Economic Statistics* 20, 339–350.

Fang, K.-T., S. Kotz and K.W. Ng (1989). *Symmetric Multivariate and Related Distributions*, Monograph on Statistics and Applied Probability 36, Chapman and Hall.

Francq, R.F. and J.-M. Zakoian (2000). Estimating weak GARCH representations, *Econometric Theory* 16, 692–728.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* 77, 304–313.

Geweke, J. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* 79, 907–915.

Granger, C.W.J., R.P. Robins, and R.F. Engle (1984). 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.

Hafner, C.M. (2003). Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics* 1, 26–54.

Hafner, C.M. and H. Herwartz (2004). Testing for causality in variance using multivariate GARCH models. Econometric Institute Report 20, Erasmus University Rotterdam.

Hafner, C.M. and J.V.K. Rombouts (2004). Estimation of temporally aggregated multivariate GARCH models, Econometric Institute Report 30, Erasmus University Rotterdam.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14, 70–86.

Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Lecture Notes in Economics and Mathematical Systems, Berlin, New York: Springer Verlag.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2nd ed., Berlin, New York: Springer Verlag.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis, *Journal of Business and Economic Statistics* 17, 129–136.

Meddahi, N. and E. Renault (2004). Temporal aggregation of volatility models. *Journal of Econometrics*, forthcoming.

Nelson, D.B. (1990) ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–39.

Nijman, T. and E. Sentana (1996). Marginalization and contemporaneous aggregation in multivariate GARCH processes, *Journal of Econometrics* 71, 71–87.

Pötscher, B.M., and I.R. Prucha (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*, Berlin, Springer Verlag.

Renault, E. and A. Szafarz (1991). True versus spurious instantaneous causality, Discussion paper 9103, CEME, Brussels.

Renault, E., K. Sekkat and A. Szafarz (1998). Testing for spurious causality in exchange rates, *Journal of Empirical Finance* 5, 47–66.

van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* 17, 549–564.

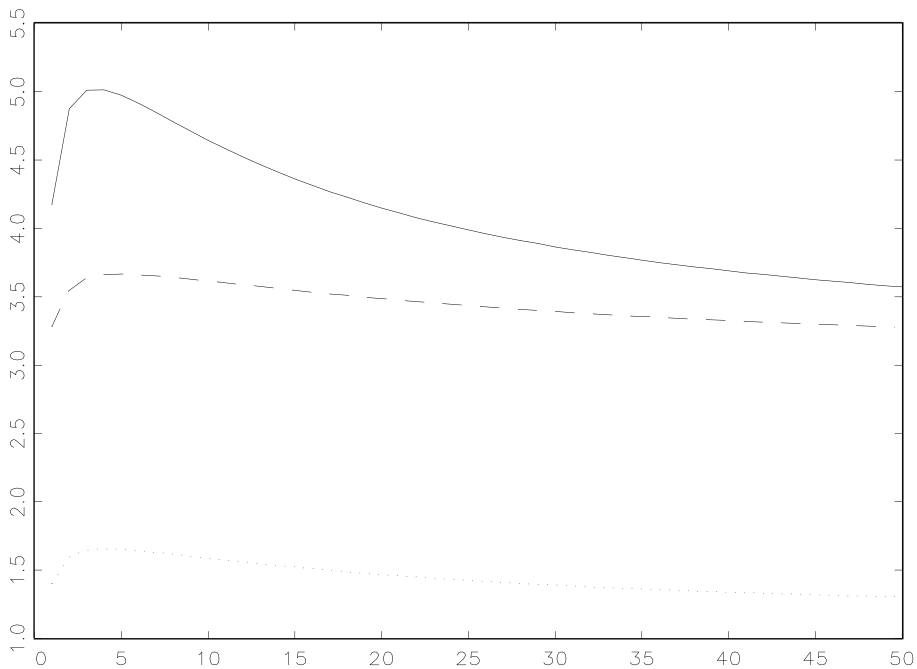

Figure 1: *Kurtosis and co-kurtosis of the example process (18) as a function of the aggregation level m. Solid line: kurtosis of* $\varepsilon^{(m)}_{mt,1}$*, dashed line: kurtosis of* $\varepsilon^{(m)}_{mt,2}$*, dotted line: co-kurtosis of* $\varepsilon^{(m)}_{mt,1}$ *and* $\varepsilon^{(m)}_{mt,2}$*.*

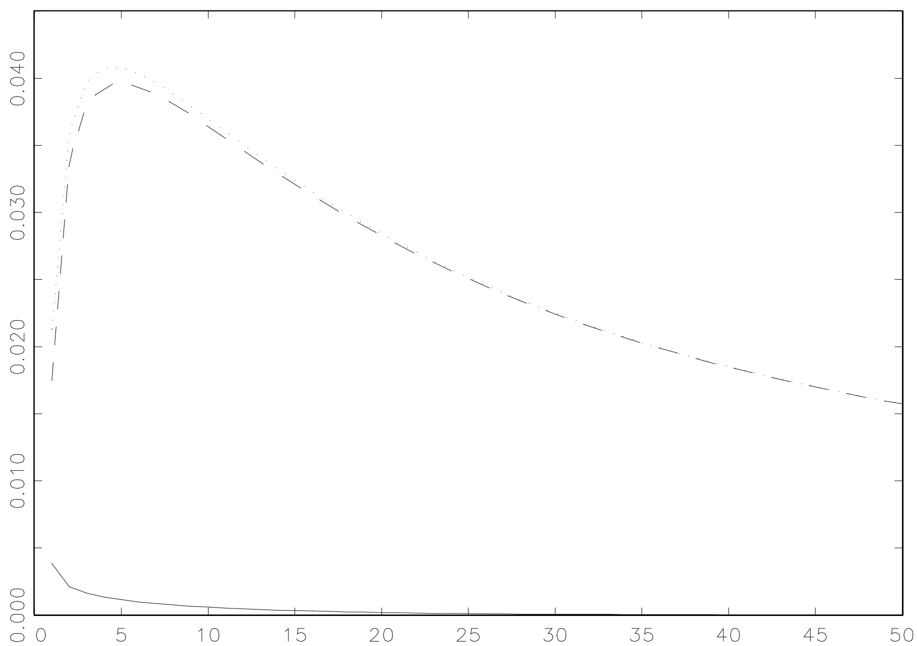

Figure 2: *Causality measures for the example process (18) as a function of the aggregation level m. Dashed line: the instantaneous causality measure* $\mathcal{ICV}^{(m)}_{y\leftrightarrow x}$*, dotted line: the linear dependence measure* $\mathcal{CV}^{(m)}_{y\leftrightarrow x}$*, solid line: the bi-directional Granger causality measure* $\mathcal{GCV}^{(m)}_{y\leftrightarrow x}$*, where* $x = \varepsilon^{(m)}_{mt,1}$ *and* $y = \varepsilon^{(m)}_{mt,2}$

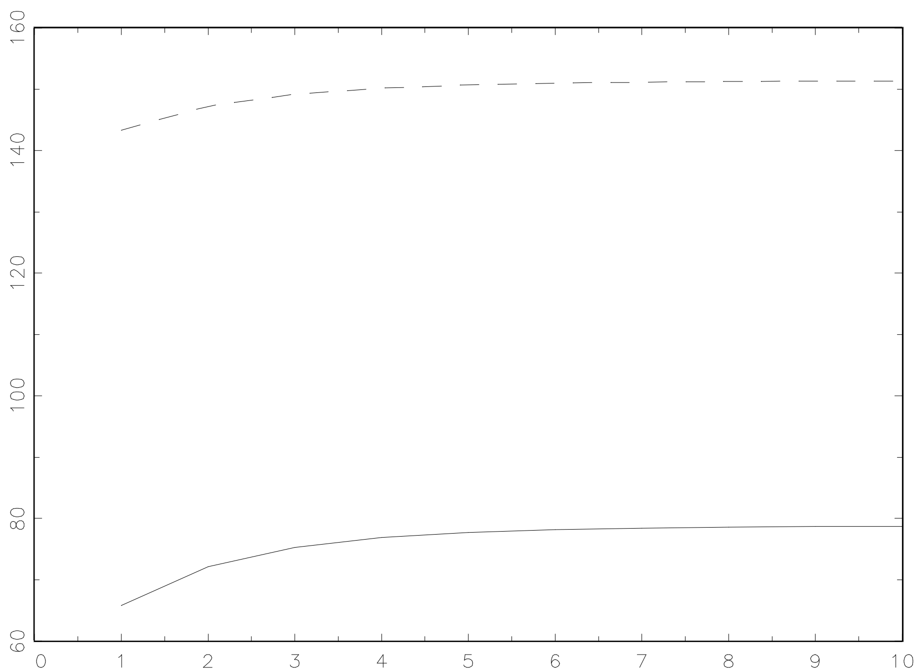

Figure 3: *Mean square prediction error of forecasting the volatility of* $\varepsilon^{(m)}_{mt,1}$ *for the example process (18) with* $m = 2$ *as a function of the forecast horizon* h*. Solid line: Prediction using the disaggregated process and then aggregating the forecasts. Dashed line: Prediction of the aggregated process. The values are scaled by the factor* m^{-2}.