Testing for Vector Autoregressive Dynamics under Heteroskedasticity

Christian M. Hafner* Helmut Herwartz †

September 2002

Abstract

In this paper we introduce a bootstrap procedure to test parameter restrictions in vector autoregressive models which is robust in cases of conditionally heteroskedastic error terms. The adopted wild bootstrap method does not require any parametric specification of the volatility process and takes contemporaneous error correlation implicitly into account. Via a Monte Carlo investigation empirical size and power properties of the new method are illustrated. We compare the bootstrap approach with standard procedures either ignoring heteroskedasticity or adopting a heteroskedasticity consistent estimation of the relevant covariance matrices in the spirit of the White correction. In terms of empirical size the proposed method clearly outperforms competing approaches without paying any price in terms of size adjusted power. We apply the alternative tests to investigate the potential of causal relationships linking daily prices of natural gas and crude oil. Unlike standard inference ignoring time varying error variances, heteroskedasticity consistent test procedures do not deliver any evidence in favor of short run causality between the two series.

Keywords: vector autoregression, hypothesis testing, heteroskedasticity, bootstrap, causality, energy markets

JEL Classification: C12, C32

*Econometric Institute, Erasmus University Rotterdam, P.O.B. 1738, 3000 DR Rotterdam, Netherlands, chafner@few.eur.nl

†Institut für Statistik und Ökonometrie, Humboldt–Universität zu Berlin, Spandauer Str. 1, D–10178 Berlin, helmut@wiwi.hu-berlin.de. The authors would like to thank Michael H. Neumann for helpful comments. Financial support by the Deutsche Forschungsgemeinschaft, SFB 373 "Simulation and Quantification of Economic Processes" at Humboldt University, is gratefully acknowledged.

1 Introduction

The magnitude of price variations at speculative markets typically exhibits positive autocorrelation and cross correlation among a set of assets, goods, stock market indices, exchange rates, etc. The observation that periods of higher and lower volatility alternate has generated a huge body of econometric literature after the seminal contributions by Engle (1982), Bollerslev (1986), and Taylor (1986) introducing the (generalized) autoregressive conditionally heteroskedastic ((G)ARCH) process and the stochastic volatility model, respectively. Numerous proposals of parametric models characterizing multivariate volatility dynamics are now available in the literature, see e.g. Bollerslev (1990), Bollerslev and Engle (1993), Braun, Sunier and Nelson (1995), Danielsson (1998), and Engle and Kroner (1995).

In the multivariate case, estimation of volatility dynamics typically requires highly specialized optimization algorithms which are employed to maximize some (quasi) log likelihood function. In the multivariate framework results on the asymptotic properties of the (Q)ML-estimator have been derived only recently, e.g. by Jeantheau (1998) Comte and Lieberman (2000), and Bollerslev and Wooldridge (1992). Nevertheless, most of the practical analyses are numerical in nature and thus highly dependent on data and problem specific features. The reliability of QML-procedures may suffer from large parameter spaces necessary for the joint modelling of variances and covariances. In addition, given large samples of empirical data, the implicit assumption of structural invariance of the volatility process may also be criticized. Applying standard tests on structural invariance of GARCH-type error processes as introduced by Chu (1995), the assumption of dynamic homogeneity of empirical volatility processes is often (strongly) rejected. Thus, in practice, QML-methods could lack robustness especially if the assumed volatility model, GARCH say, amounts to misspecification or structural variation of the volatility process.

When heteroskedastic error terms generate a vector autoregressive (VAR) model, the analyst might be interested in inference on significance of (specific) parameter estimates. First and second order moments of VAR-estimates obtained from QML-procedures are typically dependent on the particular specification of the volatility dynamics. Thus, the conclusions to be made when testing e.g. non-causality in the VAR framework might not be robust with respect to the a-priori assumed underlying volatility dynamics. Least squares based approaches to testing significance of VAR parameter estimates are more robust with respect to the specification of volatility dynamics. Moreover, least squares procedures provide unique numerical results and, thus, might be preferred to QML-estimates when the interest of the analyst is focused on estimating the conditional mean of a VAR-process. However, inference along standard lines ignoring potential heteroskedasticity involves invalid empirical levels of tests derived under iid assumptions. To correct for these size distortions, White (1980) introduced a correction of standard $t-$ratios in a univariate framework under heteroskedasticity which is easily implemented. Furthermore, bootstrap procedures designed for heteroskedastic innovations may be seen as a reasonable framework to retain the convenience of least squares procedures, as shown by Hafner and Herwartz (2000).

In this paper we introduce a wild bootstrap method to test parameter restrictions in vector autoregressive models that is robust under conditional heteroskedasticity of unspecified (unknown) form. We show consistency of the new approach and compare it with standard Wald-tests. Implementing the latter, we evaluate the relevant covariance matrices under the assumption of underlying iid error terms and, alternatively, allowing for conditional heteroskedasticity. Via Monte Carlo analysis we show that in terms of its empirical size the bootstrap approach outperforms competing test strategies. Differences in power estimates mirror the size properties and, thus, we summarize that the bootstrap is to be recommended for empirical work.

In an empirical application we show that the statistical decision for or against short term causality from crude oil to natural gas prices hinges on the way the asymptotic covariance matrix of the parameter estimates is calculated: standard Granger causality tests indicate causality, whereas tests correcting for heteroskedasticity do not. Both series exhibit strong heteroskedasticity, so that the decision of no causality based on heteroskedasticity consistent procedures is more reliable.

The remainder of the paper is organized as follows. In the next section we formalize the considered testing problem and discuss alternative testing procedures. Empirical properties of competing test approaches are investigated in Section 3. An empirical analysis of causal relationships linking daily prices for crude oil and natural gas follows in Section 4. A brief summary concludes the paper. To improve readability of the paper proofs of the propositions are given in the appendix.

2 Testing for VAR(1) dynamics in case of heteroskedastic error terms

Let us consider the vector autoregressive process of order one, VAR(1), given by

$$y_t = Ay_{t-1} + u_t, \tag{1}$$

where y_t contains K components, A is a $K \times K$ parameter matrix, and u_t is a mean zero error term. One of our objectives is to keep the notation simple, so that we do not include an intercept in (1), nor do we consider VAR models of higher order. For VAR(p) models with intercept, analogous results are easily obtained.

For a given sample of T observations, $y_1, \ldots, y_T$, we can collect the variables in the $K \times T$ matrices $Y = (y_1, \ldots, y_T)$, $Z = (y_0, \ldots, y_{T-1})$, and $U = (u_1, \ldots, u_T)$. The model then reads compactly as

$$Y = AZ + U. \tag{2}$$

Denote $\mathbf{a} = \text{vec}A$, $\mathbf{y} = \text{vec}Y$ and $\mathbf{u} = \text{vec}U$. The OLS-estimator is given by

$$\hat{\mathbf{a}} = \left\{(ZZ')^{-1}Z \otimes I_K\right\}\mathbf{y}.$$

For deriving the properties of $\hat{\mathbf{a}}$ one often assumes that the vector u_t is an iid white noise vector with finite covariance matrix. In the following we abandon this strong assumption and allow for dependence of u_t by making the following assumptions.

(A1) All eigenvalues of A have modulus smaller than one.

(A2) u_t is a mixing process.

(A3) $\mathrm{E}[u_t \mid \mathcal{F}_{t-1}] = 0$ with $\mathcal{F}_t$ denoting the information set up to time t.

(A4) $E[|u_t|^{2r}] \leq B < \infty$, for some $r > 2$ and for all t.

(A5) $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}[(y_{t-1}y_{t-1}') \otimes (u_t u_t')] = W$ with some finite, positive definite matrix W.

(A1) ensures that y_t is stable, (A2) says that the temporal dependence of u_t decays in a certain sense, (A3) says that u_t is a martingale difference, and (A4) ensures that all fourth moments of u_t exist. (A5) is a kind of asymptotic stationarity assumption. Finiteness of the expectation $\mathrm{E}[(y_{t-1}y_{t-1}') \otimes (u_t u_t')]$ follows already by (A1) and (A4), as shown in the proof of Proposition 2, so (A5) merely assumes that the averages of these expectations converge to a fixed matrix W. We can now state the first proposition.

Proposition 1 *Under (A1) to (A5),*

1. $plim \frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t' = 0$
2. $plim \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-1}' = \Gamma$ *exists and is nonsingular*
3. $plim \hat{\mathbf{a}} = \mathbf{a}$
4. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1} \otimes u_t \xrightarrow{d} N(0, W)$
5. $\sqrt{T}(\hat{\mathbf{a}} - \mathbf{a}) \xrightarrow{d} N(0, V^{-1}WV^{-1})$ *with* $V = \Gamma \otimes I_K$.

Note that replacing (A2) by the stronger assumption of conditional homoskedasticity, i.e., $\mathrm{E}[u_t u_t' \mid \mathcal{F}_{t-1}] = \Sigma_u$ with some positive definite matrix Σ_u, W would be given by

$$W = \lim \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}[(y_{t-1}y_{t-1}') \otimes \Sigma_u] = \Gamma \otimes \Sigma_u, \tag{3}$$

and the asymptotic covariance matrix would simplify to

$$\begin{aligned} V^{-1}WV^{-1} &= (\Gamma \otimes I_K)^{-1}(\Gamma \otimes \Sigma_u)(\Gamma \otimes I_K)^{-1} \\ &= \Gamma^{-1} \otimes \Sigma_u = C, \quad \text{say.} \end{aligned} \tag{4}$$

In general, however, the matrix W is unknown. In a univariate framework, White (1980) has shown that a consistent estimate of W can be easily obtained by using the residuals $\hat{u}_t$ of the least squares regression. We adopt this idea in our multivariate framework and propose the following estimate of W.

$$W_T = \frac{1}{T}\sum_{t=1}^{T}(y_{t-1}y_{t-1}') \otimes (\hat{u}_t \hat{u}_t'). \tag{5}$$

The following proposition states the consistency of W_T.

Proposition 2 *Under Assumptions (A1) to (A5),* $\text{plim}\, W_T = W$.

One can now easily construct a consistent estimate of the asymptotic covariance matrix $V^{-1}WV^{-1}$ by defining $V_T = \Gamma_T \otimes I_K$ and $\Gamma_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-1}'$. By Proposition 1, Γ_T is consistent for Γ and hence V_T is consistent for V. Thus, making use of Slutsky's Theorem, the estimator $V_T^{-1}W_T V_T^{-1}$ consistently estimates the asymptotic covariance matrix $V^{-1}WV^{-1}$.

We now want to test the hypothesis $H_0 : R\mathbf{a} = r$ against $H_1 : R\mathbf{a} \neq r$ where R is an $(N \times K^2)$ matrix of rank N and r is an $(N \times 1)$ vector. Examples of specific hypotheses are the presence of autoregressive dynamics ($R = I_{K^2}, r = 0$) and the absence of Granger causality which imposes zero restrictions on some elements of $\mathbf{a}$ that are collected with an appropriate restriction matrix R.

2.1 The standard Wald statistic

The common Wald test statistic for testing H_0 under the assumption of homoskedastic white noise errors reads

$$\lambda_T = T(R\hat{\mathbf{a}} - r)'(RC_T R')^{-1}(R\hat{\mathbf{a}} - r), \tag{6}$$

where $C_T = \Gamma_T^{-1} \otimes \widehat{\Sigma}_u$, and $\widehat{\Sigma}_u$ is a consistent estimate of Σ_u, e.g., $\widehat{\Sigma}_u = \sum_{t=1}^{T} \hat{u}_t\hat{u}_t'/T$.

In terms of the underlying vector error terms u_t the test statistic in (6) has the following structure:

$$\lambda_T = \frac{1}{T}\mathbf{u}'B_T'Q_T B_T\mathbf{u}, \tag{7}$$

where

$$B_T = (\Gamma_T \otimes \widehat{\Sigma}_u)^{-1/2}(Z \otimes I_K) \quad \text{and} \quad Q_T = C_T^{1/2}R'(RC_T R')^{-1}RC_T^{1/2}.$$

Since each of the factors in C_T and Q_T is consistent, one can use Slutsky's Theorem to obtain $C_T \xrightarrow{p} C$ and $Q_T \xrightarrow{p} Q$, where C is given in (4) and $Q = C^{1/2}R'(RCR')^{-1}RC^{1/2}$. Note that Q is an idempotent matrix of dimension $K^2 \times K^2$, i.e. $QQ = Q$. This allows us to prove the following proposition.

Proposition 3 *Under assumptions (A1) to (A5) and* H_0,

$$\lambda_T \xrightarrow{d} \beta' X, \tag{8}$$

where X is an N-dimensional vector of independent $\chi^2(1)$ random variables and β is an N-dimensional vector containing the eigenvalues of Ω given by

$$\Omega = W^{1/2}V^{-1}R'(RCR')^{-1}RV^{-1}W^{1/2}. \tag{9}$$

So, the test statistic λ_T is asymptotically distributed as a weighted mixture of independent $\chi^2(1)$-random variables where the weights are the eigenvalues of the matrix Ω.

Corollary 1 *If one replaces (A2) by the stronger assumption of conditional homoskedasticity,* $E[u_t u_t' \mid \mathcal{F}_{t-1}] = \Sigma_u$*, then* $W = \Gamma \otimes \Sigma_u$*, see (3), and* Ω *reduces to* $\Omega = Q$*, which is idempotent and therefore its eigenvalues are 0 and 1, where the number of eigenvalues equal to 1 is equal to the rank* N *of* Q*. Thus,* λ_T *has an asymptotic* χ^2_N*-distribution.*

A direct proof of this corollary is given e.g. in Proposition 3.5 of Lütkepohl (1993).

In general, however, using the standard Wald statistic with critical values obtained from the χ^2_N distribution leads to inconsistency. In the following section, we use the results of Proposition 1 to propose a test statistic that modifies the covariance matrix of the standard Wald statistic to obtain a standard χ^2_N distribution.

2.2 A modified Wald statistic

Taking into account the modification of the asymptotic covariance matrix for the case of heteroskedasticity, we suggest the following modified Wald test statistic:

$$\phi_T = T(R\hat{\mathbf{a}} - r)' \left[R V_T^{-1} W_T V_T^{-1} R' \right]^{-1} (R\hat{\mathbf{a}} - r), \tag{10}$$

Proposition 4 *Under Assumptions (A1) to (A5) and* H_0*,* ϕ_T *has an asymptotic* χ^2_N*-distribution, i.e.*

$$\phi_T \xrightarrow{d} \chi^2_N. \tag{11}$$

Thus, in the general case allowing for heteroskedasticity, it is preferable to use ϕ_T rather than λ_T in combination with critical values of the χ^2_N distribution. Note, however, that ϕ_T is more complex because it involves estimation of W, which is essentially a matrix of fourth moments, whereas λ_T only needs estimation of V, a matrix of second moments. Thus, ϕ_T may be more affected by high variation in small samples than λ_T does. The empirical performance of both test statistics will be investigated in our simulation study.

2.3 Consistency of the wild bootstrap

As derived in Section 2.1, the test statistic λ_T has a nonstandard limit distribution in case of conditional heteroskedasticity. A first order asymptotic approximation for the distribution of λ_T is hardly available in practice although the relevant nuisance parameters, namely the eigenvalues of Ω in (9) can be consistently estimated. In this case bootstrap methods become a convenient means to estimate the distribution of λ_T if the resampling scheme allows for conditional heteroskedasticity. Wu (1986) introduced the wild bootstrap coping with heteroskedastic error distributions. This procedure is advocated by Mammen (1993) to estimate the distribution of F–type statistics in parametric regression models with random explanatory variables under heteroskedasticity. Adopting a nonparametric framework, Neumann and Kreiss (1998) show that the validity of regression type bootstrap procedures is maintained for autoregressive models if the error term u_t follows a martingale difference sequence, as assumed in (A3).

Bootstrap error terms can be obtained as $u_t^* = \hat{u}_t \eta_t$, where the $\hat{u}_t$ are estimated residuals obtained from the VAR(1) model in (1) and $\eta_t \sim \text{iid}(0,1)$ and independent of y_{t-1} and $\hat{u}_t$. By construction, the first two moments of estimated residuals and bootstrap errors is identical, i.e., $E[u_{it}^*] = 0$ and $E[u_{it}^* u_{jt}^*] = \hat{u}_{it}\hat{u}_{jt}$ for all $i, j = 1, \ldots K$. To implement the bootstrap approximation of λ_T (ϕ_T) denoted as λ_T^* (ϕ_T^*) we draw $R = 500$ replications of λ_T^* (ϕ_T^*). A particular null hypothesis is rejected with significance level α if λ_T (ϕ_T) exceeds the $(1-\alpha)$-quantile of the bootstrap distribution.

Instead of obtaining $\hat{u}_t$ from the unrestricted VAR-model, residual terms can alternatively be estimated under the null hypothesis. Under the latter approach the analyst runs the risk of losing power due to the fact that under the alternative hypothesis the obtained error terms may be far away from their true values. Sampling u_t^* from restricted OLS-residuals, however, promises a close approximation of the data generating process if the null hypothesis is actually true. Therefore more accurate size properties could be expected for estimating $\hat{u}_t$ under the null hypothesis. Note that the risk of sampling from biased error estimates when the alternative hypothesis is actually true is more pronounced if the null hypothesis and the unrestricted model are far apart from each other. For the implementation of the bootstrap in our Monte Carlo study we will shed light on the latter issue by drawing u_t^* from both error estimates obtained under the null and alternative hypothesis.

In the following we will argue that the distributions of λ_T and its bootstrap counterpart λ_T^* coincide asymptotically. Since this distribution is nonstandard it turns out to be cumbersome to show the asymptotic validity of the bootstrap method under heteroskedasticity directly. In deriving the desired result indirectly we make use of the fact that the idempotent matrix Q_T converges in probability to a limit which only depends on the unconditional expectations of $y_{t-1}y_{t-1}'$ and $u_t u_t'$ and, thus, is identical under the actual sampling scheme and the recommended wild bootstrap procedure. Therefore our arguments will exploit the fact that the asymptotic distribution of λ_T defined in (6) coincides with that of λ_T^* defined by

$$\lambda_T^* = \frac{1}{T}\mathbf{u}^{*\prime} B_T' Q_T B_T \mathbf{u}^*, \tag{12}$$

if $\mathbf{b}_T = \frac{1}{\sqrt{T}} B_T \mathbf{u}$ and $\mathbf{b}_T^* = \frac{1}{\sqrt{T}} B_T \mathbf{u}^*$ share the same asymptotic distribution. It was shown in the proof of Proposition 3 that $\mathbf{b}_T$ is asymptotically normally distributed with mean zero and finite covariance matrix $(\Gamma \otimes \Sigma_u)^{-1/2} W (\Gamma \otimes \Sigma_u)^{-1/2}$, or explicitly,

$$(\Gamma \otimes \Sigma_u)^{-1/2} \lim_{T\to\infty} \left[\frac{1}{T}\sum_{t=1}^{T} E[y_{t-1}y_{t-1}' \otimes u_t u_t']\right] (\Gamma \otimes \Sigma_u)^{-1/2}. \tag{13}$$

To derive the asymptotic distribution of the bootstrap random vector $\mathbf{b}_T^*$, the $\hat{u}_t$ are treated as if they were non-stochastic fixed variables. Then the bootstrap counterpart of $\mathbf{b}_T$, i.e.

$$\mathbf{b}_T^* = \frac{1}{\sqrt{T}} B_T \mathbf{u}^* = \frac{1}{\sqrt{T}} (\Gamma_T \otimes \widehat{\Sigma}_u)^{-1/2} \sum_{t=1}^{T} y_{t-1} \otimes \hat{u}_t \eta_t$$

is also asymptotically normally distributed with mean zero and covariance matrix

$$(\Gamma \otimes \Sigma_u)^{-1/2} \lim_{T\to\infty} \left[\frac{1}{T} \sum_{t=1}^{T} E[y_{t-1}y'_{t-1} \otimes \hat{u}_t\hat{u}'_t\eta_t^2] \right] (\Gamma \otimes \Sigma_u)^{-1/2}. \tag{14}$$

Since η_t is independent from y_{t-1} and $E[\eta_t^2] = 1$, we have by construction the equivalence of $E[y_{t-1}y'_{t-1} \otimes \hat{u}_t\hat{u}'_t\eta_t^2]$ and $E[y_{t-1}y'_{t-1} \otimes \hat{u}_t\hat{u}'_t]$. Therefore, $\mathbf{b}_T$ and $\mathbf{b}_T^*$ share the same asymptotic normal distribution. Given that Q is unaffected under the bootstrap procedure, the latter argument implies also that $\lambda_T^* = \mathbf{b}_T^{*\prime} Q_T \mathbf{b}_T^*$ and $\lambda_T = \mathbf{b}_T' Q_T \mathbf{b}_T$ share the same asymptotic distribution. So we have proved the following proposition.

Proposition 5 *Under assumptions (A1) to (A5) and H_0,*

$$\lambda_T^* \xrightarrow{d} \beta' X, \tag{15}$$

where X is an N-dimensional vector of independent $\chi^2(1)$ random variables and β is an N-dimensional vector containing the eigenvalues of Ω given in (9).

The asymptotic distribution of the statistics λ_T and λ_T^* is not asymptotically pivotal, which means that it depends on nuisance parameters. However, it is well known that with respect to interval estimation the bootstrap procedure is particularly fruitful if the simulated statistic is (asymptotically) pivotal (Hall 1992). As was shown in Proposition 4, the statistic ϕ_T is asymptotically pivotal and, following similar arguments as above, we can show that ϕ_T^*, the bootstrap version of ϕ_T, shares the same property. In the simulation experiment we will investigate all competing test procedures $(\lambda_T, \lambda_T^*, \phi_T, \phi_T^*)$ and compare their empirical performance.

3 Monte Carlo Investigation

To shed light on the empirical properties of the test statistics proposed in the previous section we conduct a Monte Carlo analysis.

3.1 The simulation design

We generate bivariate autoregressive processes $(K = 2)$ of lag order $p = 1$:

$$y_t = Ay_{t-1} + u_t. \tag{16}$$

The hypothesis to be tested is $H_0 : \text{vec}A = 0$ implying that y_t is free of serial correlation. To investigate the empirical size properties of the competing test procedures we set $A = 0$ and draw u_t alternatively from a Gaussian distribution and according to a multivariate GARCH(1,1) process,

$$u_t = \Sigma_t^{1/2}\xi_t, \tag{17}$$

$$\Sigma_t = D'D + F'u_{t-1}u'_{t-1}F + G'\Sigma_{t-1}G. \tag{18}$$

In (17) ξ_t is a two dimensional standard Gaussian random vector. In (18) F and G are $K \times K$ parameter matrices and D is an upper triangular matrix. The parametric model in (18) has become popular as the so-called BEKK-model which provides a rich dynamic volatility structure including cross equation dependencies. It is discussed in detail e.g. by Engle and Kroner (1995). To be specific we draw heteroskedastic error terms using the following choices of parameter matrices in (18):

$$D = 10^{-3} \begin{pmatrix} 1.15 & .31 \\ 0 & .76 \end{pmatrix}, F = \begin{pmatrix} .282 & -.050 \\ -.057 & .293 \end{pmatrix}, G = \begin{pmatrix} .939 & .028 \\ .025 & .939 \end{pmatrix}. \tag{19}$$

The particular model detailed in (18) and (19) is found in Fengler and Herwartz (2002) to characterize joint volatility dynamics of daily quotes of the DEM and GBP measured against the USD over the period December 31, 1979 to April 1, 1994.

To examine the empirical power properties we let u_t follow the GARCH(1,1) specification given above and use the following choices for the parameter matrix A:

$$A^{(1)} = \begin{pmatrix} 0.10 & 0.00 \\ 0.00 & 0.10 \end{pmatrix}, A^{(2)} = \begin{pmatrix} 0.10 & 0.05 \\ 0.05 & 0.10 \end{pmatrix}, A^{(3)} = \begin{pmatrix} 0.00 & 0.10 \\ 0.10 & 0.00 \end{pmatrix}.$$

Each data generating process is generated with 5000 replications. The relevant sample sizes are alternatively $T = 25, 50, 100, 500, 1000$. As the nominal test level we mostly consider $\alpha = 0.05$.

The following test procedures are studied: The conventional Wald statistic (λ_T) with the covariance matrix of vec$\hat{A}$ estimated under the assumption of homoskedastic error terms. The evaluation of the latter covariance matrix under the assumption of heteroskedasticity as given in Proposition 4 delivers the test statistic ϕ_T. The heteroskedasticity consistent bootstrap counterparts of the latter two statistics are denoted as λ_T^* and ϕ_T^*, respectively. To provide a comparison with commonly used test procedures we also test the null hypothesis of interest by means of portmanteau statistics. Since the simulations cover the small sample cases with $T = 25$ and $T = 50$ we decide to employ the modified portmanteau test statistic (see e.g. Lütkepohl 1993)

$$P_h = T^2 \sum_{i=1}^{h} (T-i)^{-1} \mathrm{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \quad \hat{C}_i = \frac{1}{T} \sum_{t=i+1}^{T} \hat{u}_t \hat{u}_{t-i}', \tag{20}$$

where $\hat{u}_t$ are residual estimates obtained from estimating a VAR model under the null hypothesis. We consider alternative test orders $h = 1$ and $h = 10$. Under the null hypothesis of no serial correlation, P_h has an asymptotic χ^2_q distribution with $q = hK^2$ degrees of freedom. So, in the bivariate case, the asymptotic distributions of P_1 and P_{10} are χ^2_4 and χ^2_{40}, respectively.

3.2 Simulation results

Rejection frequencies obtained from competing approaches to test the hypothesis $H_0 : R\mathbf{a} = 0, R = I_{K^2}$ in the two dimensional VAR(1)-model are shown in Table 1. To facilitate the comparison of competing testing strategies size estimates violating the respective

	$\alpha = 0.05$					$\alpha = 0.10$		
T	25	50	100	500	1000	25	100	500
	Size 1: u_t ∼iid							
λ_T	**.1206**	**.0824**	**.0638**	.0524	.0508	**.1916**	**.1180**	.1092
ϕ_T	**.2742**	**.1550**	**.0952**	**.0586**	.0530	**.3596**	**.1616**	**.1208**
λ_T^*	**.0904**	**.0706**	**.0596**	.0536	.0498	**.1550**	**.1136**	.1084
ϕ_T^*	**.1402**	**.0986**	**.0724**	.0514	.0512	**.2182**	**.1236**	.1102
$\tilde{\lambda}_T^*$	.0560	.0510	.0508	.0526	.0488	**.1132**	.1028	.1058
P_1	**.0400**	.0480	.0460	.0496	.0496	.0932	.0958	.1054
P_{10}	.0562	.0500	.0508	.0450	.0454	.1018	.0956	.0944
	Size 2: u_t follows multivariate GARCH(1,1)							
λ_T	**.1282**	**.0940**	**.0810**	**.0954**	**.1150**	**.2030**	**.1384**	**.1630**
ϕ_T	**.2748**	**.1574**	**.0920**	**.0648**	**.0600**	**.3622**	**.1578**	**.1182**
λ_T^*	**.0886**	**.0734**	**.0610**	.0562	.0566	**.1530**	.1080	.1052
ϕ_T^*	**.1398**	**.1024**	**.0682**	.0576	.0560	**.2102**	**.1218**	.1060
$\tilde{\lambda}_T^*$	**.0582**	.0566	.0508	.0522	.0550	**.1134**	.0944	.1010
P_1	**.0418**	.0558	.0578	**.0912**	**.1130**	.0938	**.1142**	**.1586**
P_{10}	**.0606**	**.0606**	**.0860**	**.1702**	**.2188**	.1062	**.1470**	**.2570**
	Power 1 under GARCH: $A = A^{(1)}$							
λ_T	.1316	.1372	.1886	.6884	.9246	.2082	.2800	.7740
ϕ_T	.2806	.2006	.1974	.5428	.8382	.3654	.2906	.6676
λ_T^*	.0890	.1138	.1504	.5852	.8692	.1604	.2378	.7024
ϕ_T^*	.1398	.1324	.1536	.5232	.8292	.2178	.2392	.6448
$\tilde{\lambda}_T^*$	.0602	.0884	.1310	.5768	.8684	.1202	.2188	.6984
P_1	.0420	.0832	.1554	.6804	.9232	.0980	.2476	.7692
P_{10}	.0690	.0798	.1248	.4492	.7466	.1200	.1984	.5690
	Power 2 under GARCH: $A = A^{(2)}$							
λ_T	.1370	.1526	.2230	.7910	.9760	.2166	.3200	.8602
ϕ_T	.2824	.2238	.2434	.7022	.9386	.3786	.3402	.7964
λ_T^*	.0928	.1266	.1812	.7036	.9450	.1656	.2766	.8002
ϕ_T^*	.1478	.1520	.1928	.6850	.9324	.2230	.2866	.7798
$\tilde{\lambda}_T^*$	.0598	.0982	.1612	.6986	.9440	.1256	.2594	.7954
P_1	.0456	.0924	.1848	.7842	.9756	.1024	.2840	.8560
P_{10}	.0684	.0824	.1338	.5216	.8398	.1220	.2132	.6384
	Power 3 under GARCH: $A = A^{(3)}$							
λ_T	.1592	.1640	.2110	.7220	.9554	.2410	.3174	.8184
ϕ_T	.3124	.2346	.2378	.6514	.9196	.4040	.3390	.7592
λ_T^*	.1114	.1336	.1708	.6132	.9056	.1830	.2684	.7382
ϕ_T^*	.1596	.1594	.1874	.6354	.9132	.2430	.2822	.7396
$\tilde{\lambda}_T^*$	.0724	.1040	.1538	.6070	.9020	.1408	.2488	.7344
P_1	.0538	.1032	.1716	.7134	.9540	.1214	.2832	.8130
P_{10}	.0658	.0754	.1106	.4386	.7592	.1112	.1896	.5596

Table 1: *Size and power estimates for competing procedures testing $H_0 : vecA = 0$. λ_T and ϕ_T are the standard Wald test and its heteroskedasticity consistent counterpart. λ_T^* and ϕ_T^* denote corresponding bootstrap approximations. $\tilde{\lambda}_T^*$ is analogous to λ_T^* except that bootstrap errors are drawn from restricted OLS residuals. P_h denotes the multivariate portmanteau statistic of order h.*

nominal levels with 1% significance are indicated with bold entries. In case of underlying homoskedastic innovations we obtain empirical size properties of the standard Wald test (λ_T) and portmanteau statistics ($P_{(.)}$) which are in line with the results of other contributions to the topic. In small samples ($T = 25$) serious size distortions are obtained for the Wald test with the empirical significance level exceeding twice its nominal counterpart. In one particular case, $T = 25$, the portmanteau statistic P_1 turns out to be significantly conservative, i.e. the actual empirical level of the test is far below the nominal one. For the remaining scenarios both portmanteau statistics deliver empirical significance levels that cannot be distinguished from their nominal counterparts. Estimating the covariance matrix of $\hat{\mathbf{a}} = \text{vec}\hat{A}$ under the assumption of heteroskedastic innovations (ϕ_T) delivers serious size distortions which are significant at the 1% level up to sample size $T = 500$. In the worst case ($T = 25$) the latter device shows an empirical size of 27.42% when the nominal level is $\alpha = 0.05$. Bootstrap procedures (λ_T^*, ϕ_T^*) improve the empirical size properties of the corresponding Wald tests (λ_T, ϕ_T) in all cases. In small samples ($T = 25$), for instance, the empirical size estimates obtained from λ_T^* and ϕ_T^* are 9.04% and 14.02% respectively, being considerably closer to the nominal level of $\alpha = 0.05$ as if the test is performed by means of first order asymptotic approximations. Further improvements of the empirical size estimates are obtained when the bootstrap distribution is simulated with error estimates $\hat{u}_t$ according to the restricted model ($\tilde{\lambda}_T^*$). At the nominal 5% significance level this test strategy is the only procedure that delivers for all sample sizes T empirical rejection frequencies under the null hypothesis which cannot be distinguished from 0.05.

If y_t is driven by heteroskedastic error terms both portmanteau tests (P_1, P_{10}) and the standard Wald statistic (λ_T) show huge size distortions which do not vanish even if the sample size gets quite large. For the case $T = 1000$ the latter test device shows an empirical size which is at least 11.3% thus exceeding the nominal level of 5% by far. Employing the heteroskedasticity consistent covariance estimate of $\hat{\mathbf{a}}$ (ϕ_T) improves the latter size properties considerably in larger samples ($T = 500$, $T = 1000$). For the case $T = 1000$, however, the empirical significance level of the ϕ_T^* statistic is 6.0% and, thus, still exceeds significantly the nominal level. In small samples ($T = 25, 50, 100$) it appears that there is almost no gain from evaluating the covariance matrix of $\hat{\mathbf{a}}$ under the assumption of conditional heteroskedasticity. The bootstrap procedures in general and the λ_T^* and $\tilde{\lambda}_T^*$ versions in particular deliver superior size estimates. Over all considered sample sizes the empirical rejection frequencies obtained from λ_T^* ($\tilde{\lambda}_T^*$) vary between 8.86% and 5.66% (5.82% and 5.50%). For sample sizes $T = 500$ and $T = 1000$ bootstrap procedures are the only approaches delivering size estimates which cannot be distinguished from the nominal level with 1% significance. Drawing bootstrap error terms from restricted residual estimates ($\tilde{\lambda}_T^*$) significant size distortions vanish already for samples of size $T = 50$ and larger.

Generating y_t under the alternative of (weak) serial correlation all test procedures show some power which becomes apparent at least for samples of size $T = 50$ and larger. Differences in power estimates mirror the different size properties especially of the Wald tests (λ_T, ϕ_T) and their bootstrap counterparts (λ_T^*, ϕ_T^*). As outlined before the risk of simulating the bootstrap distribution by means of restricted OLS-residuals is to reduce the power of the test. For all processes simulated under the alternative, however, it turns out that this risk might be negligible under weak serial correlation patterns since the power losses involved when applying $\tilde{\lambda}_T^*$, in comparison to λ_T^* say, can be entirely addressed to the better size properties of the former statistic. In most cases considered the high order Portmanteau test (P_{10}) delivers the weakest power estimates which is obviously related to choosing too high a test order.

For the purpose of illustration Table 1 also provides some simulation results for the nominal level of $\alpha = 0.10$. As can be seen almost all results obtained for the higher significance level are analogous to the case $\alpha = 0.05$.

4 Short-term causality from crude oil to natural gas spot prices

Multivariate economic time series can often be categorized into series that are integrated without common trends, series that are cointegrated or have common stochastic trends, and series that are mean reverting or trend stationary. A very rough ordering with respect to the type of series would be financial markets, macroeconomic and commodity markets series being typically described by the first, second and third category, respectively. For financial time series, the argument of the efficient markets hypothesis says that prices should contain all available and relevant information at any time, excluding any predictability beyond a given economic model. As noted in many papers, however, one often observes predictability of financial series that is stronger than what established economic theories explain. In commodity markets, where mean reversion is the rule rather than the exception, it is not surprising that prices tend to be autocorrelated around a given trend. More interesting here is the question of causality among alternative goods. For example, in energy markets, the issue of alternative fuels such as products linked to natural gas and others linked to oil is long-debated. The economic argument for a causality is the substitution effect. If prices shoot up in oil-linked products, people will try to move to alternative products such as gas, thereby increasing the price of gas. However, contracts are usually not flexible in the short run, so that a physical effect of an oil-price shock on gas, say, is to be expected only after a few weeks or months. Most border prices of natural gas, for example, are linked to a lagged six month average of a reference oil product. But still, there may be a reaction in the short run, from day to day, if people think that the oil price shock is persistent. In this paper, we will investigate this issue by comparing the standard Granger causality test using the assumption of homoskedasticity, with tests correcting for heteroskedasticity. Indeed, the degree of heteroskedasticity in the oil price series seems to be very high. We use a reference crude oil series (north sea Brent) and the gas spot price series of the British National Balance Point (NBP) for the period March 1996 to April 2002 (1421 daily prices).

For the discussed sample period, a unit root cannot be rejected to be contained in the (log) Brent series. The ADF statistic with four lags is -1.26 with a ten percent critical value of -2.57. However, if we take a longer sample starting on May 16, 1983, the ADF statistic with four lags (-2.69) rejects at the ten percent level and the PP statistic with truncation lag 9 (-3.21) rejects at the five percent level. This supports the view that oil

Figure 1: *Solid line: Price in \$/bbl of Brent crude oil, dashed line: Price in p/thm of NBP natural gas.*

prices are mean reverting with a very slow mean reversion rate, and that our sample of 6 years is not long enough to detect the mean reversion. As for log NBP, both ADF(4) with -3.89 and PP(7) with -4.36 reject the null hypothesis of a unit root at the one percent level. To summarize, we will consider both log series as stationary and will fit a bivariate VAR model to $y_t = (\log \mathrm{Brent}_t, \log \mathrm{NBP}_t)'$. To check robustness of our results we will also investigate a bivariate VAR model specified for Δy_t.

To illustrate conditional heteroskedasticity of both log price series we apply an ARCH-LM test of order 1 (Engle 1982) to the residuals of the ADF-regressions discussed before. Testing residuals of the ADF regression for crude oil prices and natural gas prices the obtained LM-statistics are 10.30 and 251.67, respectively. Comparing both statistics with a χ_1^2 distribution we reject the assumption of homoskedasticity which, in light of the preceding sections, should alert us against using the standard causality test mentioned above. Rather, we calculate the heteroskedasticity consistent Wald test and a corresponding bootstrap version.

In modelling gas prices, one often observes seasonality that arises because of heating during the winter months. To take seasonality into account, we consider the exogenous variable x_t which is defined as $x_t = \max(H - \mathrm{Temp}_t, 0)$, where is H is a parameter and Temp_t is the relevant temperature. We take the mean of daily minimum and maximum temperatures in London as explanatory variable for all heating-based gas demand.

The empirical models specify vector autoregressive dynamics alternatively for the level series y_t and first differences Δy_t as follows:

1. Level representation
$$Y = AZ + B_l W + U_l, \tag{21}$$
where similar to (2) $Y = (y_1, \ldots, y_T)$ $A = [A_1 : \cdots : A_p]$, $Z = (z_0, \ldots, z_{T-1})$, $z_t = (y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p})'$ $U_l = (u_{l1}, \ldots, u_{lT})$. In addition $X = (w_1, \ldots, w_T)$, $w_t = (1, x_t, x_{t-1})'$ and B_l contains parameters governing the impact of the exogenous variables on y_t.

2. Growth rates
$$\Delta Y = D\Delta Z + B_d X + U_d, \tag{22}$$
where apart from obvious definitions $D = [D_1 : \cdots : D_{p-1}]$.

As indicated in (21) and (22) we determine the autoregressive order of our empirical model merely for the level representation by statistical criteria. Instead of doing specification and causality tests for the models in (21) and (22) directly we first condition the entire analysis on the variables in W for convenience, i.e. estimate vector autoregressive dynamics for oil and gas prices by means of the following models:

$$\begin{aligned} YM &= AZM + U_l \text{ (levels)}, & (23) \\ \Delta YM &= D\Delta ZM + U_d \text{ (growth rates)}, & (24) \end{aligned}$$

where $M = I - X'(XX')^{-1}X$.

To determine the lag order p in the level representation (21) we use the SIC criterion which yields $p = 3$, and H is selected by means of a grid search over all integer values between -1 and 20 degrees. Along these lines we find that $H = 1$ provides the best fit in terms of the R^2.

For both empirical models, multivariate portmanteau tests on serial residual autocorrelation are highly significant throughout. In particular, we obtain $P_{10} = 53.64$ and $P_{20} = 104.33$ for the levels representation and $P_{10} = 57.53$ and $P_{20} = 104.80$ for the growth rates model, respectively. Given that portmanteau tests are highly oversized under conditional heteroskedasticity we do not take these results to indicate misspecification of both models. Since for all specification and causality tests which are to be performed portmanteau statistics are expected to become even larger we do not further report empirical results for this test. Covariance estimates for the residuals of both models are rather close. Therefore one may conclude that switching from the levels representation to growth rates does not involve overdifferencing of the data. We obtain the following covariance estimators

$$\hat{\Sigma}_{u_l} = \begin{pmatrix} 0.717 & -0.075 \\ -0.075 & 6.093 \end{pmatrix} 10^{-3} \text{ and } \hat{\Sigma}_{u_d} \begin{pmatrix} 0.724 & -0.064 \\ -0.064 & 6.471 \end{pmatrix} 10^{-3},$$

respectively.

Parameter estimates $\hat{A}_k, k = 1, 2, 3$ and $\hat{D}_k, k = 1, 2$ are given in Table 2. Apart from the point estimates p-values obtained from competing significance tests are shown. In

$\hat{A}_1$		$\hat{A}_2$		$\hat{A}_3$	
1.023	0.006	−0.062	0.004	0.035	−0.010
(.000, .000)	(.477, .502)	(.101, .152)	(.628, .651)	(.188, .231)	(.217, .276)
[.000, .000]	[.472, .486]	[.144, .146]	[.622, .632]	[.232, .250]	[.278, .242]
0.176	0.887	−0.201	−0.050	0.046	0.132
(.022, .033)	(.000, .000)	(.068, .094)	(.132, .589)	(.547, .591)	(.000, .048)
[.034, .042]	[.000, .000]	[.098, .102]	[.608, .624]	[.580, .586]	[.058, .044]
$\hat{D}_1$		$\hat{D}_2$			
0.025	0.004	−0.034	0.010		
(.341, .433)	(.723, .733)	(.184, .220)	(.233, .293)		
[.426, .426]	[.736, .728]	[.218, .216]	[.300, .270]		
0.145	−0.111	−0.013	−0.137		
(.064, .069)	(.000, .253)	(.870, .881)	(.000, .045)		
[.072, .068]	[.244, .254]	[.878, .876]	[.030, .018]		

Table 2: *Coefficient estimates for the level ($\hat{A}_k$) and growth rates ($\hat{D}_k$) VAR-model and p-values obtained from competing testing strategies. Results for the standard and the heteroskedasticity consistent Wald-test are shown in parentheses ($p(\lambda_T)$, $p(\phi_T)$). p-values from bootstrap approximations are in square brackets [$p(\lambda_T^*)$, $p(\tilde{\lambda}_T^*)$].*

parentheses directly underneath the estimates p-values obtained from the standard Wald test (λ_T) and the corresponding statistic ϕ_T are given the latter of which is consistent under conditional heteroskedasticity. In square brackets p-values obtained from bootstrapping λ_T from unrestricted (λ_T^*) and restricted OLS-residuals ($\tilde{\lambda}_T^*$) are shown. Taking the 5% level to decide on significance of single parameter estimates $a_{i,j}^{(k)}, k = 1, 2, 3, i, j = 1, 2$, it turns out that most procedures find 4 estimates with p-values smaller than 0.05. Bootstrapping λ_T by means of unrestricted residuals delivers only three significant parameters. Since under this scheme the coefficient $\hat{a}_{22}^{(3)}$ is not significant it might be sensible to reduce the autoregressive order of the model. All inference procedures find the diagonal elements of the matrix $\hat{A}_1$ to be highly significant. For the dynamic system of growth rates the statistic λ_T indicates significance of the estimate $\hat{d}_{22}^{(1)}$. Whereas for this coefficient the p-value obtained by this test statistic is close to zero all heteroskedasticity consistent approaches provide p-values of at least 24%.

Further results from tests on joint parameter restrictions are shown in Table 3. Joint parameter restrictions are mostly not rejected by heteroskedasticity consistent test procedures (ϕ, λ_T^*, ϕ_T^*) whereas the standard Wald test (λ_T) (falsely) supports rejection of the respective null hypotheses. For instance, applying the heteroskedasticity consistent procedures to test the hypothesis that all parameters in the D_1 matrix are zero delivers p-values of at least 24% whereas the corresponding p-value obtained from λ_T is close to zero. Conclusions to be drawn from testing the hypotheses $H_0 : A_3 = 0$ or $H_0 : D_2 = 0$ via bootstrap procedures differ largely with respect to the residuals from which bootstrap replications are drawn. Estimating these residuals under the null hypothesis, the corresponding test statistics turn out to be insignificant up to the 20% level, whereas from

H_0	λ_T	$p(\lambda_T)$	ϕ_T	$p(\phi_T)$	$p(\lambda_T^*)$	$p(\phi_T^*)$	$p(\tilde{\lambda}_T^*)$	$p(\tilde{\phi}_T^*)$
	VAR(3) for log price levels							
$a_{i,j}^{(\cdot)} = 0, i,j = 1,2$	208257.1	.000	172218.8	.000	.000	.000	.000	.000
$a_{i,j}^{(\cdot)} = 0, i \neq j$	13.990	.030	12.488	.052	.064	.068	.054	.068
$a_{i,j}^{(\cdot)} = 0, j > i$	1.921	.589	1.460	.692	.628	.712	.608	.700
$a_{i,j}^{(\cdot)} = 0, i > j$	12.053	.007	11.273	.010	.018	.008	.020	.012
$a_{i,j}^{(1)} = 0, i,j = 1,2$	2756.4	.000	1200.14	.000	.000	.000	.000	.000
$a_{i,j}^{(2)} = 0, i,j = 1,2$	8.378	.079	5.659	.226	.508	.280	.508	.288
$a_{i,j}^{(3)} = 0, i,j = 1,2$	31.647	.000	6.166	.187	.064	.232	.046	.204
	VAR(2) for log price changes							
$d_{i,j}^{(\cdot)} = 0, i,j = 1,2$	53.552	.000	8.914	.350	.112	.482	.106	.432
$d_{i,j}^{(\cdot)} = 0, i \neq j$	4.993	.288	4.708	.319	.374	.390	.354	.388
$d_{i,j}^{(\cdot)} = 0, j > i$	1.562	.458	1.196	.550	.516	.590	.504	.580
$d_{i,j}^{(\cdot)} = 0, i > j$	3.431	.180	3.319	.190	.220	.208	.218	.208
$d_{i,j}^{(1)} = 0,$	24.595	.000	4.747	.314	.240	.366	.242	.360
$d_{i,j}^{(2)} = 0,$	33.448	.000	6.059	.195	.040	.252	.022	.232

Table 3: *Test statistics and p-values obtained when testing diverse joint parameter restrictions in the VAR-model for energy prices specified in levels and growth rates. See also Table 2.*

T^*	λ_T	$p(\lambda_T)$	ϕ_T	$p(\phi_T)$	$p(\lambda_T^*)$	$p(\phi_T^*)$	$p(\tilde{\lambda}_T^*)$	$p(\tilde{\phi}_T^*)$
	H_0 : Gas prices do not Granger cause crude oil prices							
50	0.769	.681	0.693	.707	.702	.774	.698	.786
100	6.367	.041	4.049	.132	.104	.202	.118	.200
200	12.07	.002	8.735	.013	.004	.016	.010	.018
300	6.898	.032	4.987	.083	.062	.082	.060	.082
400	4.312	.116	3.305	.192	.190	.202	.196	.188
500	3.478	.176	3.105	.212	.268	.250	.270	.242
1000	0.714	.700	0.568	.753	.742	.794	.706	.774
	H_0 : Crude oil prices do not Granger cause Gas prices							
50	0.442	.802	0.516	.773	.838	.862	.834	.868
100	0.077	.962	0.068	.967	.982	.982	.982	.984
200	4.060	.131	3.683	.159	.320	.206	.380	.212
300	1.574	.455	1.818	.403	.614	.468	.636	.468
400	4.638	.098	5.757	.056	.180	.074	.168	.064
500	8.045	.018	7.688	.021	.028	.020	.040	.020
1000	3.471	.176	3.650	.161	.208	.178	.200	.174

Table 4: *Sequential causality tests with increasing time windows of length T^* at the actual end of the sample. See also Table 2 and Table 3.*

unrestricted residuals bootstrap p-values are .064 or less. Thus, for this particular hypotheses one might diagnose some loss of power when resampling from restricted residual estimates.

With respect to testing for causality we find that the hypotheses $H_0 : a_{ij}^{(.)} = 0, j > i$ and $H_0 : d_{ij}^{(.)} = 0, j > i$ (i.e. gas prices do not Granger cause crude oil prices) deliver higher p-values in comparison to the opposite hypotheses $H_0 : a_{ij}^{(.)} = 0, j > i$ and $H_0 : d_{ij}^{(.)} = 0, j < i$ (i.e. gas prices are not Granger caused by prices of crude oil). Moreover, in contrast to the former hypothesis the latter is rejected by all test procedures with 5% significance when the level representation is employed for the test.

Given that for both empirical models more than 1400 observations are available and taking the results from the Monte Carlo exercises into account it should not be too surprising that all heteroskedasticity consistent test statistics deliver almost unique results when testing at a particular nominal level, 5% say. For practical purposes it is interesting to investigate the robustness of the conclusions on causality drawn before. Therefore we perform the causality test at the actual end of the sample for time windows of varying sizes. Doing so we also get some more insight into the performance of the competing test procedures when less sample information is available. In Table 4, test results for a sequence of tests are given where the employed sample covers the last $T^* = 50, 100, 200, 300, 500, 1000$ available observations. All tests are performed for the growth rate specification. It turns out that the hypothesis that gas prices are not Granger caused by crude oil prices cannot be rejected for most sample sizes under consideration. Using merely the last 500 available observations, however, the latter hypothesis is rejected by all test statistics. On the one hand one may address this finding to the Type I error of statistical inference. On the other hand one may also call the stability of the underlying data generating dynamics (conditional mean or volatility) into question. Both considerations motivate an interest in procedures which work quite well even in small samples.

5 Conclusions

We have shown that standard Wald type statistics can perform very poorly if heteroskedasticity of error terms is ignored. A modified Wald statistic that is heteroskedasticity consistent is shown to behave better in large samples but even worse in small samples. The reason for this was explained to lie in the high variability of involved moment estimates in small samples. We proposed to use bootstapped versions of these statistics that were shown to be consistent and have a superior empirical performance, in particular with respect to the size of the test. Applied to a bivariate series of daily prices of natural gas and crude oil which both exhibit strong heteroskedasticity, we found that standard tests would reject the hypothesis of no causality from oil to gas, whereas this was not the case for the heteroskedasticity consistent statistics. One could investigate this issue more deeply by looking at aggregated series, monthly say, since then economically there should be a causality from crude oil to natural gas prices, since delivery contracts of the latter are often linked directly to some moving average of oil prices. This may be a topic for future research, but longer series would be needed. Our tests have shown that in the daily trading activity of gas and oil, causality between them is doubtful due to the high degree of heteroskedasticity in both series.

Appendix

Denote by $|\cdot|$ the Euclidean norm and by $\|\cdot\|_p$ the L_p norm $\mathrm{E}[|\cdot|^p]^{1/p}$. The eigenvalue of A with maximum modulus is denoted by λ_{max} with $|\lambda_{max}| < 1$ by Assumption (A1).

Proof of Proposition 1

1. Follows the arguments of the proof of Theorem 6.5.1. of Davidson (2000) by noting that the vector $y_{t-1} \otimes u_t$ is a martingale difference. It therefore suffices to show that $\mathrm{E}[|y_{t-1} \otimes u_t|^{1+\delta}] < \infty$ for some $\delta > 0$. Noting that $y_t = \sum_{i=0}^{\infty} A^i u_{t-i}$,

$$\|y_{t-1} \otimes u_t\|_{1+\delta} \leq \sum_{i=0}^{\infty} |\lambda_{max}|^i \|u_{t-i-1} \otimes u_t\|_{1+\delta} \quad (25)$$

$$\leq \frac{\max_{j\geq 0} \|u_{t-j-1} u_t\|_{1+\delta}}{1 - |\lambda_{max}|}. \quad (26)$$

The Cauchy-Schwarz inequality gives

$$\mathrm{E}[|u_{t-i-1} \otimes u_t|^{1+\delta}] \leq \left(\mathrm{E}[|u_{t-i-1}|^{2+2\delta}]\mathrm{E}[|u_t|^{2+2\delta}]\right)^{1/2} \leq B < \infty$$

by Assumption (A4). Therefore, $\|y_{t-1} \otimes u_t\|_{1+\delta} < \infty$, which completes the proof of the first part.

2. Use the decomposition

$$\frac{1}{T}\sum_{t=1}^{T} y_t y_t' = \frac{1}{T}\sum_{t=1}^{T} A y_{t-1} y_{t-1}' A' + \frac{1}{T}\sum_{t=1}^{T} u_t y_{t-1}' A' + \frac{1}{T}\sum_{t=1}^{T} A y_{t-1} u_t' + \frac{1}{T}\sum_{t=1}^{T} u_t u_t'.$$

The probability limits of the second and third terms are zero by Proposition 1.1. The difference between $\frac{1}{T}\sum_{t=1}^{T} y_t y_t'$ and $\frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}'$ is bounded in probability. Therefore, taking vecs, and using the stability assumption (A1),

$$\text{plim}\frac{1}{T}\sum_{t=1}^{T} \text{vec}(y_t y_t') = (I_{K^2} - A \otimes A)^{-1}\text{plim}\frac{1}{T}\sum_{t=1}^{T} \text{vec}(u_t u_t') + O_p(1)$$

To see that $\text{plim}\frac{1}{T}\sum_{t=1}^{T} \text{vec}(u_t u_t')$ is bounded note that $\text{E}[u_t u_t'] = \Sigma_t < \infty$ for all t by Assumption (A4) and therefore $\lim \frac{1}{T}\sum_{t=1}^{T} \Sigma_t < \infty$. Since u_t is mixing by (A2), assumptions of Theorem 6.4.4 of Davidson (2000) hold and $\text{plim}\frac{1}{T}\sum_{t=1}^{T}(\text{vec}(u_t u_t') - \Sigma_t) = 0$.

3. For the consistency of $\hat{\mathbf{a}}$, note that $\hat{\mathbf{a}} = \mathbf{a} + (\frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}')^{-1}\frac{1}{T}\sum_{t=1}^{T} \text{vec}(u_t y_{t-1}')$. By Slutsky's Theorem,

$$\text{plim}\hat{\mathbf{a}} - \mathbf{a} = (\text{plim}\frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}')^{-1}\text{plim}\frac{1}{T}\sum_{t=1}^{T} \text{vec}(u_t y_{t-1}') = 0$$

since $\text{plim}\frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}'$ is bounded by Proposition 1.2 and $\text{plim}\frac{1}{T}\sum_{t=1}^{T} \text{vec}(u_t y_{t-1}') = 0$ by Proposition 1.1.

4. Because $v_t = y_{t-1} \otimes u_t$ is a martingale difference w.r.t. $\mathcal{F}_t$, 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 condition $\sup_T T^{-1}\sum_{t=1}^{T} \text{E}[|v_t|^{2+\delta}] < \infty$ for some $\delta > 0$ is fulfilled by noting that for every t, $\text{E}[|v_t|^{2+\delta}] \leq B < \infty$ by Assumption (A4). The condition of part (b) of their Theorem 10.1 is $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \text{E}[v_t v_t'] = W < \infty$, which is just our Assumption (A5).

5. The asymptotic distribution of $\sqrt{T}(\hat{\mathbf{a}} - \mathbf{a})$ now follows by the consistency of $\frac{1}{T}\sum_{t=1}^{T} y_t y_t'$ in Proposition 1.2, by the asymptotic distribution of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1} \otimes u_t$ in Proposition 1.4 and Cramér's Theorem.

Proof of Proposition 2

Since $\hat{u}_t = u_t - (\hat{A} - A)y_{t-1}$ we can decompose the estimate of W as

$$W_T = \frac{1}{T}\sum_{t=1}^{T} \text{vec}(y_{t-1} y_{t-1}')\text{vec}(\hat{u}_t \hat{u}_t')' \tag{27}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} \text{vec}(y_{t-1}y'_{t-1})\text{vec}(u_t u'_t)' && (28)\\
&- \frac{1}{T}\sum_{t=1}^{T} \text{vec}(y_{t-1}y'_{t-1})\text{vec}(u_t y'_{t-1})'((\hat{A}-A)' \otimes I_K) && (29)\\
&- \frac{1}{T}\sum_{t=1}^{T} \text{vec}(y_{t-1}y'_{t-1})\text{vec}(y_{t-1}u'_t)'(I_K \otimes (\hat{A}-A)') && (30)\\
&+ \frac{1}{T}\sum_{t=1}^{T} \text{vec}(y_{t-1}y'_{t-1})\text{vec}(y_{t-1}y'_{t-1})'((\hat{A}-A) \otimes (\hat{A}-A))' && (31)\\
&= T_1 + T_2 + T_3 + T_4 && (32)
\end{aligned}
$$

First we show the convergence in probability of T_1 to W. The expectation of T_1 is bounded:

$$
\text{E}[y_{t-1}y'_{t-1} \otimes u_t u'_t] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(A^i \otimes A^j)\text{E}[\text{vec}(u_{t-j}u'_{t-i})\text{vec}(u_t u'_t)']. \qquad (33)
$$

Due to the law of iterated expectations, the expectation on the right hand side of (33) is zero for all $i \neq j$. Applying the Cauchy-Schwartz inequality, we have for $k, l, m, n = 1 \ldots, K$

$$
\text{E}[|u_{t-i,k}u_{t-i,l}u_{t,m}u_{t,n}|] \leq \left(\text{E}u^4_{t-i,k}\text{E}u^4_{t-i,l}\text{E}u^4_{t,m}\text{E}u^4_{t,n}\right)^{1/4} \leq B < \infty
$$

by Assumption(A4). Thus, the expectation on the right hand side of (33) is bounded by B and, hence,

$$
\text{E}[y_{t-1}y'_{t-1} \otimes u_t u'_t] \leq B\sum_{i=0}^{\infty}(A \otimes A)^i = B(I_{K^2} - A \otimes A)^{-1} < \infty
$$

by Assumption (A1). Now, let $x_t = \text{vec}(y_{t-1}y'_{t-1} \otimes u_t u'_t)$ and note that

$$
\begin{aligned}
|x_t| &= \left|\text{vec}\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(A^i \otimes A^j)((u_{t-j-1}u'_{t-i-1}) \otimes (u_t u'_t))\right)\right| && (34)\\
&\leq \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|\lambda_{max}|^{i+j}|\text{vec}((u_{t-j-1}u'_{t-i-1}) \otimes (u_t u'_t))| && (35)
\end{aligned}
$$

almost surely. Thus,

$$
\begin{aligned}
\|x_t - \text{E}[x_t]\|_{1+\delta} &\leq \|x_t\|_{1+\delta} + \|\text{E}[x_t]\|_{1+\delta} && (36)\\
&\leq 2\|x_t\|_{1+\delta} && (37)\\
&\leq 2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|\lambda_{max}|^{i+j}\|\text{vec}((u_{t-j-1}u'_{t-i-1}) \otimes (u_t u'_t))\|_{1+\delta} && (38)\\
&\leq \frac{2B}{(1-|\lambda_{max}|)^2} < \infty && (39)
\end{aligned}
$$

This, together with x_t being $L_1 - NED$ on the mixing process u_t (see Davidson, 2000) ensures that a weak law of large numbers applies to x_t, i.e., $\text{plim}\frac{1}{T}\sum_{t=1}^{T}(x_t - \text{E}[x_t]) = 0$.

Since $\hat{A} - A = O_p(T^{-1/2})$, the terms T_2, T_3 and T_4 converge to zero in probability provided the means of $\text{vec}(y_t y_t')\text{vec}(y_t y_t')'$ and $\text{vec}(y_{t-1} y_{t-1}')\text{vec}(y_{t-1} u_t')'$ converge in probability to finite limits. To see that this is the case, one can argue just as above. For example, let $z_t = \text{vec}((y_t y_t') \otimes (y_t y_t')$ and note that $\text{E}[z_t] < \infty$ as implied by (A1) and (A4). Then,

$$\|z_t - \text{E}[z_t]\|_{1+\delta} \leq 2\|z_t\|_{1+\delta} \leq \frac{2B}{(1 - |\lambda_{max}|)^4} < \infty$$

Proof of Proposition 3

Based on Proposition 1.4, $\mathbf{b}_T = \frac{1}{\sqrt{T}} B_T \mathbf{u}$ is asymptotically normally distributed with asymptotic covariance matrix given by $(\Gamma \otimes \Sigma_u)^{-1/2} W (\Gamma \otimes \Sigma_u)^{-1/2}$, by using (A3), $\text{plim}\Gamma_T = \Gamma$, $\text{plim}\widehat{\Sigma_u} = \Sigma_u$, and Proposition C.4 (1) of Lütkepohl (1993). Defining $\tilde{\mathbf{b}}_T = W_T^{-1/2}(\Gamma_T \otimes \widehat{\Sigma}_u)^{1/2} b_T$ and using the same arguments as before, one obtains

$$\tilde{\mathbf{b}}_T \xrightarrow{d} N(0, I_{K^2}) \tag{40}$$

We can now rewrite the test statistic as $\lambda_T = \tilde{\mathbf{b}}_T' \Omega_T \tilde{\mathbf{b}}_T$ with

$$\Omega_T = W_T^{1/2}(\Gamma_T^{-1} \otimes I_K) R' (R C_T R')^{-1} R (\Gamma_T^{-1} \otimes I_K) W_T^{1/2}.$$

Again using Slutsky's Theorem and Proposition C.4 (1) of Lütkepohl (1993), Ω_T converges in probability to a finite positive semi-definite matrix given by $\Omega = W^{1/2}(\Gamma^{-1} \otimes I_K) R' (R C R')^{-1} R (\Gamma^{-1} \otimes I_K) W^{1/2}$. Decompose Ω_T as $\Omega_T = \Theta_T \Lambda_T \Theta_T'$, where Θ_T contains the eigenvectors and Λ_T is diagonal with the eigenvalues $\Lambda_T(1), \ldots, \Lambda_T(K^2)$ of Ω_T on its diagonal. The test statistic can now be written as

$$\lambda_T = \tilde{\mathbf{b}}_T \Theta_T \Lambda_T \Theta_T' \tilde{\mathbf{b}}_T' = \sum_{i=1}^{K^2} \Lambda_T(i) (\Theta_T' \tilde{\mathbf{b}}_T)_i^2$$

Using (40) and the orthogonality of Θ_T, $\Theta_T' \tilde{\mathbf{b}}_T \xrightarrow{d} N(0, I_{K^2})$, which shows that λ_T is asymptotically distributed as a weighted mixture of independent $\chi^2(1)$-random variables where the weights are the eigenvalues $\Lambda(i)$ of the matrix Ω. Q.E.D.

Proof of Proposition 4

As in the proof of Proposition 3, one can write the test statistic as $\phi_T = \tilde{\mathbf{b}}_T' \Omega_T \tilde{\mathbf{b}}_T$ where $\tilde{\mathbf{b}}_T \xrightarrow{d} N(0, I_{K^2})$, but where Ω_T is now given by

$$\Omega_T = W_T^{1/2} V_T^{-1} R' (R V_T^{-1} W V_T^{-1} R')^{-1} R V_T^{-1} W_T^{1/2}.$$

Note that Ω_T is now idempotent and, as before, converges in probability to a finite positive semi-definite matrix given by $\Omega = W^{1/2} V^{-1} R' (R V^{-1} W V^{-1} R')^{-1} R V^{-1} W^{1/2}$. Since Ω is idempotent, its eigenvalues are 0 and 1, where the number of non-zero eigenvalues is equal to the rank N of Ω. Thus, ϕ_T has an asymptotic χ_N^2 distribution. Q.E.D.

References

Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, 31, 307–327.

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

Bollerslev, T., Engle, R.F. (1993), “Common Persistence in Conditional Variances”, *Econometrica*, 61, 167–186.

Bollerslev, T., Engle, R.F., Wooldridge, J.M. (1988), “A Capital Asset Pricing Model with Time Varying Covariances”, *Journal of Political Economy*, 96, 116–131.

Bollerslev, T., Wooldridge, J.M. (1992), “Quasi–Maximum Likelihood Estimation and Inference in Dynamic Models with Time–varying Covariances”, *Econometric Reviews*, 11, 143–172.

Braun, P.A., Nelson D.B., Sunier A.M. (1995) Good news, bad news, volatility, and betas. *Journal of Finance* 50, 1575–1603

Chu, C.-S.J. (1995). Detecting Parameter Shift in GARCH Models. *Econometric Reviews*, 14, 241-266.

Comte, F., Lieberman, O. (2000), “Asymptotic Theory for Multivariate GARCH Processes”, *Manuscript, Universities Paris 6 and Paris 7.*

Danielsson, J., (1998), “Multivariate Stochastic Volatility Models: Estimation and a Comparison with VGARCH Models”, *Journal of Empirical Finance*, 5, 155–174.

Davidson, J. (2000), *Econometric Theory*, Blackwell.

Engle, R. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation”, *Econometrica*, 50, 987–1008.

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

Fengler, M., Herwartz, H. (2002) Multivariate volatility models, in: Härdle, W., T. Kleinow, G. Stahl (Eds.): Applied Quantitative Finance, Berlin, Springer, 221–236.

Hafner, C.M., Herwartz, H. (2000), “Testing for Linear Autoregressive Dynamics Under Heteroskedasticity”, *The Econometrics Journal*, 3, 177–197.

Hall, P. (1992), The Bootstrap and Edgeworth Expansion, New York, Springer.

Harvey, A.C., Ruiz, E., Shephard, N. (1994), “Multivariate Stochastic Variance Models”, *Review of Economic Studies*, 61, 247–264.

Jeantheau, T. (1998) "Strong Consistency of Estimators for Multivariate ARCH Models", *Econometric Theory*, 14, 70-86.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.

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

White, H. (1980), "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica* 48, 817–838.