Testing for nonlinearity in (co)variances

Bilel Sanhaji

Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS

21 February 2014

Abstract

In this paper, we propose two Lagrange multiplier tests for nonlinearity in conditional covariances. The null hypothesis is the scalar BEKK model in which covolatilities of time series are driven by a linear function of its own lags and lagged squared innovations. The alternative hypothesis is an extension of that model in which covolatilities are modelled by a nonlinear function in the lagged squared innovations. The nonlinearity is represented by an exponential and a logistic transition function. We define the asymptotic properties of the scalar BEKK and determine the moment conditions of the test statistics. We investigate the size and the power of these tests through Monte Carlo experiments, and we show empirical illustrations.

Keywords: Lagrange multiplier test, nonlinearity test, smooth transition function, scalar BEKK, multivariate GARCH.

JEL classification: C12, C32, C58.
1 Introduction

Since the last quarter of century an important amount of multivariate model handling the conditional second moments of financial time series have been developed. While the quantity of financial data has hugely increased, practitioners who manage large data sets need to use accurate tools.

Caporin and McAleer (2008, 2012) ask about the pertinence to discriminate among the two most used multivariate GARCH models (MGARCH): the conditional covariances BEKK model (Engle and Kroner, 1995) and the conditional correlations DCC model (Engle, 2002). They shed light on the existence of asymptotic theory for the BEKK model which is a key determinant for any econometricians who want to explore the asymptotic limits.

Whereas the number of MGARCH models increases, only few diagnostic tests are available (Bauwens, Laurent and Rombouts, 2006; Silvennoinen and Teräsvirta, 2009, for large surveys). The literature covering the nonlinearity tests in the multivariate framework remains relatively low and based on threshold models (Tsay, 1998; Kwan, Li and Ng, 2010), or for testing the constancy of conditional correlations (Péguin-Feissolle and Sanhaji, 2013). Note to capture asymmetric effects of negative and positive shocks in time series, the smooth transition GARCH model allow intermediate regimes by passing from one to the other smoothly (see van Dijk, Teräsvirta and Franses, 2002, for a survey).

The scalar BEKK defined by Ding and Engle (2001) is a high dimensional model which solves large-scale problem. It permits to deal with a very large amount of (financial) time series. This model can face to the curse of dimensionality. For 10 assets, the number of parameters to estimate is equal to 57, where for the diagonal BEKK it is equal to 75 and 255 for a full BEKK (see Table 1 in Caporin and McAleer, 2012). One can demonstrate straightforwardly that the scalar BEKK model is a special case of the BEKK model for which Comte and Lieberman (2003) have proved strong consistency of the quasi-maximum likelihood estimator (QMLE) using Jeantheau (1998), and asymptotic normality under eight-order moments. Hafner and Preminger (2009) derived the general multivariate GARCH model which nests, among numerous special cases, the BEKK model (Engle and Kroner, 1995, Proposition 2.4).

We develop two tests based on an exponential and a logistic smooth transition functions in order to test the nonlinearity in variances and covariances. Under the null hypothesis, we define the asymptotic properties and determine the moment conditions. Finite-sample properties of the two new tests are examined using Monte Carlo methods. We show that the tests perform well in our small-sample simulations. Empirical illustrations using real data, point
out that these tests can be useful to reject the linearity hypothesis in multivariate modelling of conditional heteroscedasticity. All computations have been performed using Matlab 8.1, and the tests are available for Matlab and R upon request.

The paper is organized as follows. We begin in next section to introduce the model basis. In Section 3, we first present the strong consistency and thereafter show the asymptotic normality of the QML estimator under the null hypothesis. Section 4 gives the tests statistic. We provide Monte Carlo experiments in Section 5 and present empirical illustration in Section 6. Section 7 concludes. Section 8 gives the appendix.

2 The model

Let the observations \( \{y_t\} \) be multivariate time series of \( \mathbb{R}^N \) defined by

\[
y_t = \mu_t + \varepsilon_t, \quad \text{for} \quad t = 1, \ldots, T
\]

(1)

where the conditional mean \( \mu_t = \mathbb{E}(y_t|\mathcal{F}_{t-1}) = 0 \) to simplify the discussion, \( \mathcal{F}_{t-1} \) is the \( \sigma \)-field generated by all the information until time \( t - 1 \).

Consider \( \varepsilon_t \) is an \( N \times 1 \) disturbance vector with conditional mean

\[
\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0
\]

and

\[
\mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = H_t;
\]

\( H_t \) is the conditional covariance matrix defined by \( H_t = H_t(\theta) \) where \( \theta \) is a parameter vector.

We define a multivariate GARCH process such that

\[
\varepsilon_t = H_t^{1/2} z_t
\]

(2)

where \( z_t \sim iid(0, I_N) \), is a martingale difference sequence with respect to \( \mathcal{F}_{t-1} \), and \( I_N \) is the \( N \times N \) identity matrix. The conditional covariance is defined, for the first order model (for only one lag), by

\[
H_{t+1} = CC' + (\alpha + \varphi G(s_t; \gamma, c)) \varepsilon_t\varepsilon_t' + \beta H_t
\]

(3)

where \( CC' \) is a positive-definite matrix of parameters composed by two non-singular lower triangular matrices \( C \) which ensures positive-definiteness by construction, \( \alpha, \varphi \) and \( \beta \) are scalars. The sufficient conditions to ensure the positivity of the model are \( \beta \geq 0, \alpha > 0 \) and \( \alpha + \varphi \geq 0 \) (or \( \alpha \geq 0 \) and \( \alpha + \varphi > 0 \)). The sufficient conditions to be second-order stationary if and
only if \( s_t \) and \( \varepsilon_t \) are independent are given by \( \alpha + \beta + \varphi < 1; \varphi \) can be negative.

The transition function is given by \( G(s_t; \gamma, c) \). This function depends on a transition variable \( s_t \), a transition parameter \( \gamma \), and a location parameter \( c \). It can adopt an exponential form:

\[
G(s_t; \gamma, c) = 1 - \exp \left\{ -\gamma (s_t - c)^2 \right\}, \quad \gamma > 0
\]

and bounded by \([0, 1]\). When \( s_t \to -\infty \) or \( \infty \) : \( G(s_t; \gamma, c) \to 1 \), when \( s_t \to c \) : \( G(s_t; \gamma, c) \to 0 \). According to van Dijk, Ter"asvirta and Franses (2002), the transition variable \( s_t \) can be an exogenous variable, or a function of lagged endogenous variables, or a function of time \( t \), corresponding to a frequency \( t/T : ]0, 1] \). The transition variable \( s_t \) is assumed to be stationary and have a continuous distribution. Moreover, the transition function can adopt a logistic form:

\[
G(s_t; \gamma, c) = \left[ 1 + \exp \left\{ -\gamma (s_t - c) \right\} \right]^{-1}, \quad \gamma > 0.
\]  

The vectorization vec of (3) is

\[
\text{vec}(H_{t+1}) = \text{vec}(CC') + (\alpha + \varphi G(s_t; \gamma, c)) \text{vec}(\varepsilon_t \varepsilon_t') + \beta \text{vec}(H_t).
\]  

To avoid redundancy we can consider only one side of the off-diagonal elements. As seen in (3), the matrices involved are symmetric and can be defined following a vech representation

\[
\text{vech}(H_{t+1}) = \text{vech}(CC') + (\alpha + \varphi G(s_t; \gamma, c)) \text{vech}(\varepsilon_t \varepsilon_t') + \beta \text{vech}(H_t),
\]  

where vech, the vector-half operator, stacks the lower triangular half into a single vector of length \( N(N+1)/2 \). The configurations given by (6) and (7) are the vector representations of (3) without any modification in the estimates. The structure of the model implies that the model is driven by the same parameters \( \alpha, \varphi, \beta, \gamma, c \), and the \( C \) matrix, which is a strong restriction. The number of parameters to estimate is \( K = N(N+1)/2 + 3 + 2 \) where \( N(N+1)/2 \) are the intercepts, 3 corresponds to the scalars \( \alpha, \varphi, \beta \) and 2 is the number of parameters \( \gamma, c \) in the transition function (4) and (5). This model is actually the ‘Scalar-Diagonal or Two-parameters Model’ commonly named the ‘Scalar BEKK’ of Ding and Engle (2001), itself derived from the general BEKK model, and augmented by a smooth transition function. The test developed in this paper is \( H_0 : \varphi = 0 \).
3 Moment conditions

Strong consistency and asymptotic normality of the quasi maximum likelihood estimator (QMLE) in multivariate heteroscedastic frameworks have been studied by Ling and McAleer (2003) for the CCC-GARCH model, Comte and Lieberman (2003) for the BEKK model and Hafner and Preminger (2009) for the VEC model (see Francq and Zakoïan (2010), Chapter 11, for a review). In this section, we give the moment conditions under the null hypothesis, when there is no smooth transition in the (co)variances (see Section 4). The model (1 – 3) defined in Section 2 is, under the null, the so-called scalar BEKK of Ding and Engle (2001):

\[ H_{t+1} = CC' + \alpha \varepsilon_t \varepsilon_t' + \beta H_t. \] (8)

Since the aim of the paper is to build a Lagrange multiplier test, only the restricted model have to be estimated. We give the sufficient conditions for consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE). The quasi-loglikelihood function for observation \( t \) is given by:

\[ L_T(\theta) = -\frac{1}{2T} \sum_{t=1}^{T} \left( \log |H_t(\theta)| + \varepsilon_t' H_t(\theta)^{-1} \varepsilon_t \right) \]

where \( \theta = (\text{vech}(C)', \alpha, \beta)' \) is the parameter vector and assume that \( \theta \in \Theta \subset \mathbb{R}^p, p = \frac{N(N+1)}{2} + 2, \) and \( \theta_0 \) denotes the true parameter vector. The estimator of the QML is defined as \( \hat{\theta}_T = \arg \max_{\theta \in \Theta} L_T(\theta) \).

In the case of all orders are set to one, for easiness of notation and without any loss of generality, we can state that the scalar BEKK is a particular case of the BEKK model since (8) is equivalent to

\[ H_{t+1} = CC' + A \varepsilon_t \varepsilon_t' A' + BH_t B' \] (9)

where \( A = \sqrt{\alpha} I_N \) and \( B = \sqrt{\beta} I_N \).

Strong consistency

The paper of Comte and Lieberman (2003) (Theorem 2), based on Jeantheau (1998) (Theorem 2.1), proves strong consistency of the QMLE when the process is defined by (9). We make the following assumptions for the model defined by (8).

**Assumption 3.1.** The parameter space \( \Theta \) is compact and \( \alpha + \beta < 1 \)

**Assumption 3.2.** The rescaled errors \( z_t \) admit a density absolutely continuous with respect to the Lebesgue measure and positive in a neighbourhood of the origin.
Assumption 3.3. The model is identifiable: $\alpha > 0$

Assumption 3.4. $E||z_t||^2 < \infty$, $\text{var}(z_t) = I_N$

Theorem 1. Under Assumptions 3.1-3.4, $\hat{\theta}_T \xrightarrow{T \to \infty} a.s. \theta_0$.

Asymptotic normality

The asymptotic normality of the QMLE has been established by the theorem 3 of Comte and Lieberman (2003) (Appendix A) for the BEKK model under the eighth moment, whereas Hafner and Preminger (2009) need the sixth moment of the innovations for the VEC model.

Assumption 3.5. $\theta_0$ is an interior point of $\Theta$

Assumption 3.6. $E||\varepsilon_t||^6 < \infty$

Theorem 2. Under Assumptions 3.1-3.6, the asymptotic distribution of the QML estimators is given by

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} \frac{T \to \infty}{N(0, J(\theta_0)^{-1} J(\theta_0) J(\theta_0)^{-1})}.$$ 

Theorem 2 shows that whether the innovations $\varepsilon_t$ are conditionally Gaussian, the ML estimation provides the most efficient estimators and the sandwich matrix becomes equal to $J(\theta_0)^{-1}$:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{D} \frac{T \to \infty}{N(0, J(\theta_0)^{-1})}.$$ 

Note that the matrices $J(\theta_0)$ and $I(\theta_0)$ are consistently estimated by

$$\hat{J}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left. \frac{\partial l_t(\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}} \left. \frac{\partial l_t(\theta)}{\partial \theta'} \right|_{\theta = \hat{\theta}}$$

(10)

and

$$\hat{I}(\hat{\theta}) = -\frac{1}{T} \sum_{t=1}^{T} \left. \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right|_{\theta = \hat{\theta}}.$$ 

(11)
Unconditional fourth moment

The unconditional fourth moment is computed to prove its existence. Its finiteness will be demonstrated to complete the necessary and sufficient moment conditions for the asymptotic normality. Following He and Teräsvirta (1999a,b) and Ling and McAleer (2002a,b) in GARCH, and Hafner (2003) in MGARCH, we have not considered moments of higher order since the fourth moments are probably more interesting in practice than any higher-order ones.

We consider the matrix reduced-form of the model defined by (8) to estimate under the null hypothesis of linear covariances such that

$$H_t = CC' + \Gamma_{t-1} H_{t-1}$$

(12)

where $\Gamma_t = \alpha z_t z'_t + \beta I_N$.

**Theorem 3.** Let $E||z_t||^2 < \infty$. The sufficient condition for the model (12) to be second-order stationary is given by $\Gamma_1 < 1$:

$$E[\varepsilon_t \varepsilon'_t] = (I_N - \Gamma_1)^{-1} CC'$$

(13)

where $\Gamma_1 = E\Gamma_t = \alpha \nu_2 + \beta I_N$ and $\nu_2 = E[z_t z'_t]$.

The unconditional mean is obtained by applying the unconditional expectation operator on both side of the equation (12).

**Proof.** can be directly obtained from Comte and Lieberman (2003) (Proposition A.1.) for the BEKK, or from He and Teräsvirta (2004) (Theorem 1.) for the ECCC-GARCH, also used in Nakatani and Teräsvirta (2009), (see also Francq and Zakoïan, 2010, p. 430 (11.7)) using $\rho(\Gamma) < 1$ which is equivalent and yields to the sufficient condition $\Gamma_1 < 1$. $\square$

**Theorem 4.** Let $E||z_t^4|| < \infty$ and condition given in Theorem 3 holds. Then, the fourth-order moment of $\varepsilon_t$ exists:

$$E(\varepsilon_t \varepsilon'_t \otimes \varepsilon_t \varepsilon'_t) = \nu_4[(CC' \otimes CC') + 2(CC' \otimes \Gamma_1 H)](I_{N^2} - \Gamma_2)^{-1},$$

(14)

if and only if $\Gamma_2 < 1$ in modulus, where $\Gamma_1 = \alpha \nu_2 + \beta I_N$, $\Gamma_2 = \alpha^2 \nu_4 + \beta^2 I_{N^2} + 2(\alpha \nu_2 \otimes \beta I_N)$, $H = E[\varepsilon_t \varepsilon'_t]$ and $\nu_4 = E(z_t z'_t \otimes z_t z'_t)$.

---

1To compute the kurtosis and the autocorrelation functions of residuals and squared residuals.
4 Nonlinearity test

4.1 LM test statistic

Assuming the model with one lag in the innovations and the conditional covariances, we construct a Lagrange Multiplier (LM) test for smooth transition in the conditional covariances. Under the null hypothesis the model is a scalar BEKK, under the alternative the model is a scalar BEKK with smooth transition in the conditional covariances.

The loglikelihood function is

\[ L_T(\theta) = -\frac{1}{2T} \sum_{t=1}^{T} l_t(\theta) \]

with

\[ l_t(\theta) = \log |H_t(\theta)| + \epsilon_t' H_t(\theta)^{-1} \epsilon_t. \]

Under mild regularity conditions, the LM test statistic is given by

\[ LM = T^{-1} S(\hat{\theta})' \hat{I}(\hat{\theta})^{-1} S(\hat{\theta}). \]  \hspace{1cm} (15)

Note \( \hat{\theta} \) the maximum likelihood estimator of \( \theta_0 \), the true parameter vector under the null hypothesis. The score evaluated at \( \hat{\theta} \) equals

\[ S(\hat{\theta}) = \sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}}. \]

The population information matrix \( \hat{I}(\hat{\theta}) \) is the consistent estimator using the score’s outer product (10) or using the negative of the Hessian matrix (11). We apply the rules of matrix partition to the inverse of (11) and extract the relevant part.

Once the LM statistic is computed analytically, we estimate the restricted model under the null hypothesis at the quasi maximum likelihood estimator. Then, we compute the LM statistic (15) by replacing \( \theta_0 \), the true unknown parameter vector, by the consistent estimator \( \hat{\theta} \), under the null hypothesis.

4.2 Approximation of the null hypothesis

Under the null hypothesis of linear conditional covariances \( H_0 : \varphi = 0 \), some parameters in the functions (4) and (5) remain indeterminate: \( \gamma \) and \( c \). Following Luukkonen, Saikkonen and Teräsvirta (1988), we approximate the transition functions by a Taylor expansion around \( \gamma = 0 \).
The function defined by (4) is approximated by a first-order Taylor expansion:

\[ G(s_t; \gamma, c) \approx \gamma (s_t - c)^2. \]  
(16)

Reparametrizing (3) with (16) yields to

\[ H_{t+1} = CC' + (\varphi_1 + \varphi_2 s_t + \varphi_3 s_t^2) \varepsilon_t \varepsilon_t' + \beta H_t \]  
(17)

where \( \varphi_1 = \alpha + \gamma c^2 \), \( \varphi_2 = -2\gamma c \) and \( \varphi_3 = \gamma \).

The auxiliary null hypothesis is as follows:

\[ H_{01} : \varphi_2 = \varphi_3 = 0. \]  
(18)

The function defined by (5) is approximated by a second-order Taylor expansion:

\[ G(s_t; \gamma, c) \approx \frac{1}{2} + \frac{1}{4}(s_t - c)\gamma. \]  
(19)

Reparametrizing (3) with (19) yields to

\[ H_{t+1} = CC' + (\phi_1 + \phi_2 s_t) \varepsilon_t \varepsilon_t' + \beta H_t \]  
(20)

where \( \phi_1 = \alpha + 1/2 - 1/4\gamma c \) and \( \phi_2 = 1/4\gamma \).

The auxiliary null hypothesis is as follows:

\[ H_{02} : \phi_2 = 0. \]  
(21)

The LM statistic for testing \( H_0 \) is asymptotically distributed as a \( \chi^2 \) with 2 or 1 degrees of freedom, say \( \chi^2_2 \) or \( \chi^2_1 \), when the unrestricted model contains an exponential transition function or a logistic transition function, respectively.

5 Monte Carlo experiments

5.1 Size simulations

The empirical size is investigated under a bivariate first-order scalar BEKK model with normal errors. The true data generating process follows (2). Following Silvennoinen and Teräsvirta (2005), the transition parameter is generated from an exogenous GARCH(1,1) process, such that \( s_t = h_t^{1/2} z_t \), where \( z_t \sim N(0,1) \). The generation parameters are as follows:

\[ H_t = CC' + 0.06 \varepsilon_{t-1} \varepsilon_{t-1}' + 0.84 H_{t-1} \]
\[ h_t = 0.02 + 0.03 s_{t-1}^2 + 0.94 h_{t-1} \]
The sample sizes are 100, 250, 500 and 1000. Before generating the actual observations, we remove the first 1000 observations from the series in order to eliminate initialization effects. The number of replications is $S = 5000$. The results are reported in Table 1. The size of the tests is already very close to the nominal size for the sample of 500.

### Table 1: Small sample sizes

<table>
<thead>
<tr>
<th></th>
<th>Logistic test</th>
<th></th>
<th>Exponential test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>100</td>
<td>0.0170</td>
<td>0.0570</td>
<td>0.0960</td>
<td>0.0468</td>
</tr>
<tr>
<td>250</td>
<td>0.0108</td>
<td>0.0440</td>
<td>0.0848</td>
<td>0.0188</td>
</tr>
<tr>
<td>500</td>
<td>0.0116</td>
<td>0.0510</td>
<td>0.0972</td>
<td>0.0110</td>
</tr>
<tr>
<td>1000</td>
<td>0.0108</td>
<td>0.0462</td>
<td>0.0902</td>
<td>0.0118</td>
</tr>
</tbody>
</table>

### 5.2 Power simulations

The empirical power is investigated under a bivariate and a five-variate first-order smooth transition scalar BEKK. The true data generating process follows (2). The sample sizes are 100, 250, 500 and 1000 and the number of replications is $S = 5000$, excepted for $N = 5$, $S = 1000$. The power is directly related to the transition parameter $s_t$. In line with Silvennoinen and Teräsvirta (2005), a natural choice is to use a function of squared lagged returns containing relative information of covariances. The generation parameters are as follows:

$$H_t = CC' + [0.02 + 0.05G(s_{t-1}; 2, 0)] \varepsilon_{t-1}\varepsilon_{t-1}' + 0.91H_{t-1}$$

where

$$C = \begin{bmatrix} 0.07 \\ 0.03 \\ 0.09 \end{bmatrix}$$

for $N = 2$, and

$$H_t = CC' + [0.1 + 0.2G(s_{t-1}; 5, 0)] \varepsilon_{t-1}\varepsilon_{t-1}' + 0.65H_{t-1}$$
where

$$\mathbf{C} = \begin{bmatrix}
0.5 \\
0.9 \\
0.1 \\
0.2 \\
0.6 \\
0.9
\end{bmatrix}$$

for $N = 5$.

The transition variable is defined such that

$$\bar{s}_t = \begin{bmatrix} 0.2, 0.2, 0.2, 0.2 \end{bmatrix}^\prime \begin{bmatrix} \bar{\varepsilon}^{(2)}_{t-1}, \bar{\varepsilon}^{(2)}_{t-2}, \bar{\varepsilon}^{(2)}_{t-3}, \bar{\varepsilon}^{(2)}_{t-4}, \bar{\varepsilon}^{(2)}_{t-5} \end{bmatrix},$$

where $\bar{\varepsilon}^{(2)}_t$ is the squared elementwise of the mean of $\varepsilon_t$ over $N$.

Figure 1 - 2 report the results of our small sample power simulations for $N = 2$ and $N = 5$, respectively. We remark that the power of the two tests decreases when dimension raises. Actually, the scalar BEKK is known to be ill-conditioned when the dimension reaches the observations. The test based on an exponential function (Figure 1) is likely more powerful, whatever we simulate a smooth transition scalar BEKK model with a logistic or an exponential transition function.
Figure 1: Small sample powers for the nonlinear exponential test when the models simulated are a smooth transition scalar BEKK with a logistic function (left) and an exponential function (right), $N = 2$
Figure 2: Small sample powers for the nonlinear logistic test when the models simulated are a smooth transition scalar BEKK with a logistic function (left) and an exponential function (right), $N = 2$

Table 2: Small sample powers for linear conditional covariance test with exponential transition function when $N = 5$

<table>
<thead>
<tr>
<th>Model:</th>
<th>$T=100$</th>
<th>$T=250$</th>
<th>$T=500$</th>
<th>$T=1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.461</td>
<td>0.588</td>
<td>0.682</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.620</td>
<td>0.772</td>
<td>0.832</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.889</td>
<td>0.955</td>
<td>0.973</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Small sample powers for linear conditional covariance test with logistic transition function when $N = 5$

<table>
<thead>
<tr>
<th>Model: Exponential</th>
<th>T</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>0.322</td>
<td>0.472</td>
<td>0.566</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.662</td>
<td>0.817</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.918</td>
<td>0.971</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.998</td>
<td>0.998</td>
<td>0.999</td>
</tr>
</tbody>
</table>

6 Empirical illustration

We illustrate the performance of the two tests in an empirical study. We consider the S&P500 (SPX) index, eight assets: Apple Inc. (AAPL), Bank of America Corp. (BAC), The Coca-Cola Company (KO), Ford Motor (F), International Business Machines Corp. (IBM), JPMorgan and Chase & Co. (JPM), McDonald’s Corp. (MCD), Exxon Mobil Corp. (XOM), three exchange rates: the Australian Dollar (AUD), the British Pound (GBP) and the New-Zealand Dollar (NZD), and the West Texas Intermediate (WTI) oil price, over the period which extends from May 29, 1986 to January 3, 2014. Daily transactions are turned into returns: $100 \times \log(P_t/P_{t-1})$, where $P_t$ represents the daily closing price at time $t$, for $t = 1, \ldots, 6960$.

Table 4 presents the summary statistics of the asset returns. In order to identify nonlinearity without outliers, the extreme values are removed. Outliers are likely to come from market structure. We remove the few peaks by cutting them off such that each return is in the range $[-20, +20]$. The cutoff is mentioned by the superscript $(co)$. The number of truncations is mentioned (Trunc.) in the last column of Table 4.

The returns exhibit a positive excess of kurtosis and a negative excess of skewness. When the cutoff is applied, the positive excess of kurtosis and the negative excess of skewness hugely decrease.

The test is designed to be easy to implement. The relevant transition variable has been chosen among different candidates as discussed in Silvennoinen and Teräsvirta (2005). For the two tests, we set it as the absolute value of the two days lagged S&P500 index returns for $N = 2, 4, 8$ and $9$. To save space, the results for the absolute S&P500 index returns and for the absolute seven days lagged S&P500 index returns as transition variable are not reported here. For $N = 3$, with exchange rate returns, we set the squared
Table 4: Summary statistics

<table>
<thead>
<tr>
<th>Asset</th>
<th>Mean (co)</th>
<th>Min.</th>
<th>Std. (co)</th>
<th>Skew (co)</th>
<th>Kurt (co)</th>
<th>Trunc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPX</td>
<td>0.0287</td>
<td>0.0291</td>
<td>22.899</td>
<td>1.1882</td>
<td>1.0712</td>
<td>30.505</td>
</tr>
<tr>
<td>AAPL</td>
<td>0.0385</td>
<td>0.0661</td>
<td>28.679</td>
<td>3.9165</td>
<td>2.9195</td>
<td>101.10</td>
</tr>
<tr>
<td>BAC</td>
<td>-0.0173</td>
<td>0.0072</td>
<td>30.209</td>
<td>-0.8164</td>
<td>3.0664</td>
<td>129.05</td>
</tr>
<tr>
<td>KO</td>
<td>-0.0152</td>
<td>0.0265</td>
<td>17.958</td>
<td>-107.69</td>
<td>2.6193</td>
<td>132.70</td>
</tr>
<tr>
<td>F</td>
<td>-0.0230</td>
<td>0.0092</td>
<td>25.865</td>
<td>2.6186</td>
<td>5.6297</td>
<td>132.70</td>
</tr>
<tr>
<td>IBM</td>
<td>-0.0028</td>
<td>0.0182</td>
<td>23.363</td>
<td>2.6186</td>
<td>5.6297</td>
<td>132.70</td>
</tr>
<tr>
<td>JPM</td>
<td>0.0013</td>
<td>0.0205</td>
<td>23.363</td>
<td>2.6186</td>
<td>5.6297</td>
<td>132.70</td>
</tr>
<tr>
<td>XOM</td>
<td>0.00075</td>
<td>0.0292</td>
<td>23.363</td>
<td>2.6186</td>
<td>5.6297</td>
<td>132.70</td>
</tr>
<tr>
<td>AUD</td>
<td>0.00012</td>
<td>-8.2120</td>
<td>4.4349</td>
<td>4.9692</td>
<td>0.6079</td>
<td>6.8727</td>
</tr>
<tr>
<td>GBP</td>
<td>0.0057</td>
<td>5.9322</td>
<td>6.1782</td>
<td>0.7670</td>
<td>-0.2988</td>
<td>6.6726</td>
</tr>
<tr>
<td>NZD</td>
<td>0.0200</td>
<td>19.150</td>
<td>8.1069</td>
<td>0.7670</td>
<td>-0.2988</td>
<td>6.6726</td>
</tr>
<tr>
<td>WTI</td>
<td>-0.0200</td>
<td>19.150</td>
<td>8.1069</td>
<td>0.7670</td>
<td>-0.2988</td>
<td>6.6726</td>
</tr>
</tbody>
</table>

15
GBP returns as transition variable.

It is interesting to see what the variances and covariances between two or more asset returns are. Table 5 shows the two tests in bivariate cases. The both tests applied on returns are in favour of nonlinearity in conditional covariances in most of bivariate relationships. For some of them, one or both tests fail to reject the null hypothesis of linearity. This is the case for AAPL–F for the both tests. In general, the exponential transition function-based test rejects more often than the logistic transition function-based test, as shown in Section 5.2. Actually, 10 of the 29 pairs fail to reject the null hypothesis of linearity with the logistic-based test and only 1 for the exponential-based test, at the 10% level of significance.

Table 6 shows the results for the both tests in multivariate analyses. The tests fail to reject the hypothesis of linearity on exchange rate returns. For \( N = 4, 8 \) and 9, the exponential-based test rejects the null at the nominal size of 1%, and the logistic-based test fails to reject the null for AAPL-BAC-KO-F.

7 Conclusion

We propose two tests for the linearity of variances and covariances in the multivariate GARCH models. The first one is based on an exponential transition function, and the second one on a logistic transition function. Under the null hypothesis, we define the asymptotic properties and determine the moment conditions. The main practical findings in this paper are that these two new tests perform very well in our small sample Monte Carlo experiments. An empirical illustration shows that the two tests detect nonlinearity in covariances.

8 Appendix

8.1 Derivatives from the dynamic specification of the model

We compute the derivatives of the approximate loglikelihood of \( H_t \) with respect to the parameters in \( \theta \). All the rules of matrix calculus can be found in the Handbook of Matrices of Lütkepohl (1996) and in The Matrix Cookbook of Petersen and Pedersen (2008). Moreover, Comte and Lieberman (2003) found similar results.

It is more direct, convenient and easier to calculate the score and the
Table 5: Linear conditional covariance tests for $N = 2$, the cutoff is applied and the transition variable is the absolute value of the two days lagged S&P500 returns.

<table>
<thead>
<tr>
<th>Transition function:</th>
<th>Logistic $p$-value</th>
<th>Exponential $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPX-WTI</td>
<td>0.0424 0.8369</td>
<td>8.8682 0.0119</td>
</tr>
<tr>
<td>AAPL-BAC</td>
<td>4.1798 0.0409</td>
<td>39.714 0.0000</td>
</tr>
<tr>
<td>AAPL-KO</td>
<td>14.892 0.0001</td>
<td>17.874 0.0001</td>
</tr>
<tr>
<td>AAPL-F</td>
<td>0.0023 0.9619</td>
<td>1.2840 0.5262</td>
</tr>
<tr>
<td>AAPL-IBM</td>
<td>17.570 0.0000</td>
<td>19.551 0.0001</td>
</tr>
<tr>
<td>AAPL-JPM</td>
<td>3.3776 0.0661</td>
<td>3.3787 0.1846</td>
</tr>
<tr>
<td>AAPL-MCD</td>
<td>16.125 0.0001</td>
<td>17.992 0.0001</td>
</tr>
<tr>
<td>AAPL-XOM</td>
<td>28.237 0.0000</td>
<td>30.580 0.0000</td>
</tr>
<tr>
<td>BAC-KO</td>
<td>10.037 0.0015</td>
<td>72.443 0.0000</td>
</tr>
<tr>
<td>BAC-F</td>
<td>44.067 0.0000</td>
<td>136.80 0.0000</td>
</tr>
<tr>
<td>BAC-IBM</td>
<td>11.556 0.0007</td>
<td>85.012 0.0000</td>
</tr>
<tr>
<td>BAC-JPM</td>
<td>7.0902 0.0078</td>
<td>30.577 0.0000</td>
</tr>
<tr>
<td>BAC-MCD</td>
<td>5.1405 0.0234</td>
<td>54.303 0.0000</td>
</tr>
<tr>
<td>BAC-XOM</td>
<td>22.046 0.0000</td>
<td>108.79 0.0000</td>
</tr>
<tr>
<td>KO-F</td>
<td>5.9015 0.0151</td>
<td>24.189 0.0000</td>
</tr>
<tr>
<td>KO-IBM</td>
<td>0.2634 0.6078</td>
<td>13.795 0.0010</td>
</tr>
<tr>
<td>KO-JPM</td>
<td>0.1272 0.7214</td>
<td>8.1833 0.167</td>
</tr>
<tr>
<td>KO-MCD</td>
<td>2.1329 0.1442</td>
<td>12.265 0.0022</td>
</tr>
<tr>
<td>KO-XOM</td>
<td>2.8932 0.0890</td>
<td>22.710 0.0000</td>
</tr>
<tr>
<td>F-IBM</td>
<td>1.5854 0.2080</td>
<td>28.757 0.0000</td>
</tr>
<tr>
<td>F-JPM</td>
<td>7.9544 0.0048</td>
<td>14.194 0.0008</td>
</tr>
<tr>
<td>F-MCD</td>
<td>3.3461 0.0674</td>
<td>25.898 0.0000</td>
</tr>
<tr>
<td>F-XOM</td>
<td>1.5972 0.2063</td>
<td>18.001 0.0001</td>
</tr>
<tr>
<td>IBM-JPM</td>
<td>0.3578 0.5497</td>
<td>19.275 0.0001</td>
</tr>
<tr>
<td>IBM-MCD</td>
<td>4.8987 0.0269</td>
<td>12.463 0.0020</td>
</tr>
<tr>
<td>IBM-XOM</td>
<td>1.9868 0.1587</td>
<td>18.377 0.0001</td>
</tr>
<tr>
<td>JPM-MCD</td>
<td>1.5093 0.2192</td>
<td>4.7999 0.0907</td>
</tr>
<tr>
<td>JPM-XOM</td>
<td>3.0398 0.0812</td>
<td>8.4500 0.0146</td>
</tr>
<tr>
<td>XOM-MCD</td>
<td>8.9608 0.0028</td>
<td>15.697 0.0004</td>
</tr>
</tbody>
</table>
Table 6: Linear conditional covariance tests for $N > 2$, the cutoff is applied and the transition variable is the absolute value of the two days lagged S&P500 returns.

<table>
<thead>
<tr>
<th></th>
<th>Transition function:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logistic</td>
</tr>
<tr>
<td></td>
<td>$LM_{LCC}$</td>
</tr>
<tr>
<td>AUD-NZD-GBP</td>
<td>0.3246</td>
</tr>
<tr>
<td>AAPL-BAC-KO-F</td>
<td>1.3356</td>
</tr>
<tr>
<td>IBM-JPM-MCD-XOM</td>
<td>11.284</td>
</tr>
<tr>
<td>AAPL-JPM-KO-XOM</td>
<td>27.102</td>
</tr>
<tr>
<td>IBM-BAC-MCD-F</td>
<td>17.640</td>
</tr>
<tr>
<td>AAPL-JPM-KO-XOM</td>
<td></td>
</tr>
<tr>
<td>IBM-BAC-MCD-F</td>
<td>20.502</td>
</tr>
<tr>
<td>SPX-AAPL-JPM-KO-XOM-IBM-BAC-MCD-F</td>
<td>28.742</td>
</tr>
</tbody>
</table>

Hessian from the dynamic specification of the model. The approximate log-likelihood is

$$l_t(\theta) = \log |H_t(\theta)| + \varepsilon_t'H_t(\theta)^{-1}\varepsilon_t, \text{ for } t = 1, \ldots, T.$$ 

For easiness of notation, $H_t$ denotes $H_t(\theta)$. The first derivative of the log-likelihood with respect to $\theta_i$, the elements of the parameter vector $\theta$, is

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \left( \log |H_t| + \varepsilon_t'H_t^{-1}\varepsilon_t \right)$$

where

$$\frac{\partial}{\partial \theta_i} \left( \log |H_t| \right) = \text{Tr} \left( \left( \frac{\partial \log |H_t|}{\partial H_t} \right)' \frac{\partial H_t}{\partial \theta_i} \right)$$

$$= \text{Tr} \left( H_t^{-1} \frac{\partial H_t}{\partial \theta_i} \right)$$

and

$$\frac{\partial}{\partial \theta_i} \left( \varepsilon_t'H_t^{-1}\varepsilon_t \right) = \text{Tr} \left( \frac{\partial \varepsilon_t'H_t^{-1}\varepsilon_t}{\partial H_t} \frac{\partial H_t}{\partial \theta_i} \right)$$

$$= \text{Tr} \left( -H_t^{-1}\varepsilon_t\varepsilon_t'H_t^{-1} \frac{\partial H_t}{\partial \theta_i} \right).$$
Note 5. The matrix calculus rules applied here are chain rule involving a function such as \( \frac{\partial g(U)}{\partial X} \) involving a function such as \( \frac{\partial g(U)}{\partial U} \) and \( \frac{\partial \log |X|}{\partial X} = (X^{-1})' \) and \( \frac{\partial aX^{-1}b}{\partial X} = -(X^{-1})'ab'(X^{-1}) \).

Thus, we obtain the \( i \)-th element of the score

\[
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta_i} = \sum_{t=1}^{T} \text{Tr} \left[ \frac{\partial H_t}{\partial \theta_i} H_t^{-1} - \epsilon_t \varepsilon_t' H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \right]
\]

or

\[
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} = \sum_{t=1}^{T} \left[ \left( \frac{\partial \text{vec} H_t}{\partial \theta} \right)' \text{vec} H_t^{-1} - \left( \frac{\partial \text{vec} H_t}{\partial \theta} \right)' (H_t^{-1} \otimes H_t^{-1}) \text{vec}(\epsilon_t \epsilon_t') \right]
\]

for \( \theta = (\theta_1, \ldots, \theta_m)' \).

Note 6. Using \( \text{Tr}(AB) = \text{vec}(A)' \text{vec}(B) \), \( \text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B) \) and \( \text{Tr}(ABC) = \text{vec}(A)'(D' \otimes B) \text{vec}(C) \).

Remark 7. In line with Comte and Lieberman (2003), when the conditional expectation operator is applied on the \( i \)-th element of the score when evaluated at \( \theta_0 \),

\[
\mathbb{E}_{t-1} \left[ \text{Tr} \left( \frac{\partial H_t}{\partial \theta_i} H_t^{-1} - \epsilon_t \varepsilon_t' H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \right) \right]_{\theta = \theta_0} = 0 \quad \text{a.s.}
\]

which is a martingale difference sequence where \( H_t \) denotes \( H_t(\theta_0) \).

The \( ij \)-th element of the Hessian is given by the first derivative of (22) w.r.t. \( \theta_j \), the elements of the parameter vector \( \theta \).

\[
\frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} = \text{Tr} \left[ \frac{\partial^2 H_t}{\partial \theta_i \partial \theta_j} H_t^{-1} - \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} H_t^{-1} + \epsilon_t \varepsilon_t' H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} H_t^{-1} \right. \\
- \left. \epsilon_t \varepsilon_t' H_t^{-1} \frac{\partial^2 H_t}{\partial \theta_i \partial \theta_j} H_t^{-1} + \epsilon_t \varepsilon_t' H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \right].
\]

Note 8. The matrix calculus rule applied here is \( \frac{\partial X^{-1}}{\partial x} = -X^{-1} \frac{\partial X}{\partial x} X^{-1} \) (Petersen and Pedersen, 2008, p. 8, 2.2 (53)).
Since $E[\varepsilon_t \varepsilon_t'|\mathcal{F}_{t-1}] = E_{t-1}[\varepsilon_t \varepsilon_t'] = H_t(\theta_0)$, under the expectations operator, conditioned on the information set $\mathcal{F}_{t-1}$, the expression above simplifies to
\[
E_{t-1} \frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} \bigg|_{\theta = \theta_0} = \text{Tr} \left[ \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} H_t^{-1} \right]
\]
when evaluated at the true unknown value of $\theta = \theta_0$. Then, the consistent estimator is given by
\[
\hat{I}(\hat{\theta}) = -\frac{1}{T} \sum_{t=1}^{T} \text{Tr} \left[ \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} H_t^{-1} \right]_{\theta = \hat{\theta}}
\]
for $\theta = (\theta_1, \ldots, \theta_m; \theta_1, \ldots, \theta_n)'$.

### 8.2 Derivatives of $H_t$ with exponential transition function

As $H_t$ and $\varepsilon_t \varepsilon_t'$ are symmetric, we use the vech representation of (17)
\[\text{vech}H_t = \text{vech}(CC') + (\varphi_1 + \varphi_2 s_{t-1} + \varphi_3 s^2_{t-1}) \text{vech}(\varepsilon_{t-1} \varepsilon_{t-1}') + \beta \text{vech}H_{t-1}.\]

We derive $\text{vech}H_t$ with respect to the parameters in $\theta = (\text{vech}(C)', \varphi_1, \varphi_2, \varphi_3, \beta)'$, the vector of all the parameters of the model.

\[
\frac{\partial \text{vech}H_t}{\partial \text{vech}(C)'} = \frac{2D_N}{N(N+1)} (C \otimes I_N) \frac{L_{N}'}{N^2} + \beta \frac{\partial \text{vech}H_{t-1}}{\partial \text{vech}(C)'}
\]

\[
\frac{\partial \text{vech}H_t}{\partial \varphi_1} = \text{vech}(\varepsilon_{t-1} \varepsilon_{t-1}') + \beta \frac{\partial \text{vech}H_{t-1}}{\partial \varphi_1}
\]

\[
\frac{\partial \text{vech}H_t}{\partial \varphi_2} = \text{vech}(\varepsilon_{t-1} \varepsilon_{t-1}') s_{t-1} + \beta \frac{\partial \text{vech}H_{t-1}}{\partial \varphi_2}
\]

\[
\frac{\partial \text{vech}H_t}{\partial \varphi_3} = \text{vech}(\varepsilon_{t-1} \varepsilon_{t-1}') s^2_{t-1} + \beta \frac{\partial \text{vech}H_{t-1}}{\partial \varphi_3}
\]

\[
\frac{\partial \text{vech}H_t}{\partial \beta} = \text{vech}(H_{t-1}) + \beta \frac{\partial \text{vech}H_{t-1}}{\partial \beta}
\]
These derivatives can be implemented in the score (23) and the information matrix (26).

**Note 9.** The derivative (27) of \( vechH_t \) w.r.t. the lower triangular matrix \( C \) has dimension \( \frac{N(N+1)}{2} \times \frac{N(N+1)}{2} \) and the others (28–31) are of dimension \( \frac{N(N+1)}{2} \times 1 \). \( L_N \) is the elimination matrix such that \( vech(X) = L_N vec(X) \). Since \( C \) is a lower triangular matrix, the derivative of (27) is obtained from Lütkepohl (1996) (p. 196, 10.5.4)

\[
\frac{\partial vechH_t}{\partial vech(C)} = 2D_N^+ (C \otimes I_N) L'_N
\]

where \( D_N^+ \) is the \( \left( \frac{N(N+1)}{2} \times N^2 \right) \) Moore-Penrose inverse such that \( D_N^+ = (D'_N D_N)^{-1} D'_N \), and \( D_N \) is the \( \left( N^2 \times \frac{N(N+1)}{2} \right) \) duplication matrix such that \( vec(X) = D_N vec(X) \).

### 8.3 Derivatives of \( H_t \) with logistic transition function

As \( H_t \) and \( \varepsilon_t \varepsilon'_t \) are symmetric, we use the vech representation of (20)

\[
vechH_t = vech(CC') + (\phi_1 + \phi_2 s_{t-1}) vech(\varepsilon_{t-1} \varepsilon'_{t-1}) + \beta vechH_{t-1}.
\]

We derive \( vechH_t \) with respect to the parameters in \( \theta = (vech(C)', \phi_1, \phi_2, \beta)' \), the vector of all the parameters of the model.

\[
\frac{\partial vechH_t}{\partial vech(C)'} = 2D_N^+ \left( \frac{N(N+1)}{2} \times N^2 \right) \left( \frac{N(N+1)}{2} \times N^2 \right) + \beta \frac{\partial vechH_{t-1}}{\partial vech(C)'}
\]

\[
\frac{\partial vechH_t}{\partial \phi_1} = vech(\varepsilon_{t-1} \varepsilon'_{t-1}) + \beta \frac{\partial vechH_{t-1}}{\partial \phi_1}
\]

\[
\frac{\partial vechH_t}{\partial \phi_2} = vech(\varepsilon_{t-1} \varepsilon'_{t-1}) s_{t-1} + \beta \frac{\partial vechH_{t-1}}{\partial \phi_2}
\]

\[
\frac{\partial vechH_t}{\partial \beta} = vech(H_{t-1}) + \beta \frac{\partial vechH_{t-1}}{\partial \beta}
\]

These derivatives can be implemented in the score (23) and the information matrix (26).
Note 10. The derivative (27) of $\vech H_t$ w.r.t. the lower triangular matrix $C$ has dimension $\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}$ and the others (28–31) are of dimension $\frac{N(N+1)}{2} \times 1$. $L_N$ is the elimination matrix such that $\vech(X) = L_N \text{vec}(X)$.

Since $C$ is a lower triangular matrix, the derivative of (27) is obtained from Lütkepohl (1996) (p. 196, 10.5.4)

$$
\frac{\partial \vech H_t}{\partial \vech(C)'} = 2D_N^+(C \otimes I_N)L_N'
$$

where $D_N^+$ is the $\left(\frac{N(N+1)}{2} \times N^2\right)$ Moore-Penrose inverse such that $D_N^+ = (D_N' D_N)^{-1} D_N$, and $D_N$ is the $\left(N^2 \times \frac{N(N+1)}{2}\right)$ duplication matrix such that $\text{vec}(X) = D_N \vech(X)$.

8.4 The LM-statistic: exponential transition function

First, consider the score

$$
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} = \begin{bmatrix}
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} \\
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \varphi_1} \\
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \beta} \\
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \varphi_2} \\
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \varphi_3}
\end{bmatrix}
\begin{bmatrix}
\frac{N(N+1)}{2} \times 1 \\
1 \times 1 \\
1 \times 1 \\
1 \times 1
\end{bmatrix}
= \begin{bmatrix}
V_C \\
V_{\varphi_1} \\
V_{\beta} \\
V_{\varphi_2} \\
V_{\varphi_3}
\end{bmatrix}.
$$

Equation (23) can be written as

$$
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} = -\frac{1}{2} \sum_{t=1}^{T} \left[ \left(D_N ^{\frac{N(N+1)}{2}} \frac{\partial \vech H_t}{\partial \theta} \right)' D_N \vech H_t^{-1} - \left(D_N ^{\frac{N(N+1)}{2}} \frac{\partial \vech H_t}{\partial \theta} \right)' (H_t^{-1} \otimes H_t^{-1}) D_N \vech(\varepsilon_t \varepsilon_t') \right].
$$
Second, consider the information matrix

$$\mathcal{I}(\theta) = \frac{1}{2T} \sum_{t=1}^{T} \left( \frac{\partial \text{vech} H_t}{\partial \theta} \right)' D_N (H_t^{-1} \otimes H_t^{-1}) D_N \frac{\partial \text{vech} H_t}{\partial \theta}.$$ 

Third, consider the analytical computation of the LM test

The LM test of No Exponential Smooth Transition, denoted $LM_{NEST}$, can be expressed as

$$LM_{NEST} = \begin{pmatrix} V_C \\ V_{\varphi_1} \\ V_{\beta} \\ V_{\varphi_2} \\ V_{\varphi_3} \end{pmatrix} \left[ \begin{array}{ccc} M_{CC} & M'_{\varphi_1 C} & M'_{\varphi_2 C} \\ M_{\varphi_1 C} & M_{\varphi_1 \varphi_1} & M_{\varphi_2 \varphi_1} \\ M_{\beta C} & M_{\beta \varphi_1} & M_{\beta \varphi_2} \\ M_{\varphi_2 C} & M_{\varphi_2 \varphi_1} & M_{\varphi_2 \varphi_2} \\ M_{\varphi_3 C} & M_{\varphi_3 \varphi_1} & M_{\varphi_3 \varphi_2} \end{array} \right]^{-1} \begin{pmatrix} V_C \\ V_{\varphi_1} \\ V_{\beta} \\ V_{\varphi_2} \\ V_{\varphi_3} \end{pmatrix}$$

where $M_{..}$ is the information matrix w.r.t its parameters.

Last, consider the LM test

Under the null hypothesis, we maximize the quasi-likelihood function with respect to $\text{vech}(C)'$, $\varphi_1$ and $\beta$. Then, the first derivatives of the quasi-likelihood function with respect to these three parameters are set to zero:

$$\sum_{t=1}^{T} \left. \frac{\partial l_t(\theta)}{\partial \theta} \right|_{\theta = \hat{\theta}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and

$$LM_{NEST} = \begin{pmatrix} V_{\varphi_2} \\ V_{\varphi_3} \end{pmatrix}' \left[ \begin{array}{ccc} M_{\varphi_2 \varphi_2} & M'_{\varphi_2 \varphi_2} & M'_{\varphi_3 \varphi_2} \\ M_{\varphi_2 \varphi_2} & M_{\varphi_2 \varphi_2} & M_{\varphi_3 \varphi_2} \\ M_{\varphi_3 \varphi_3} & M_{\varphi_3 \varphi_3} & M_{\varphi_3 \varphi_3} \end{array} \right]^{-1} \begin{pmatrix} M'_{\varphi_2 C} & M'_{\varphi_2 \varphi_1} & M'_{\varphi_2 \varphi_2} \\ M'_{\varphi_2 \varphi_1} & M'_{\varphi_2 \varphi_1} & M'_{\varphi_2 \varphi_2} \\ M'_{\varphi_3 C} & M'_{\varphi_3 \varphi_1} & M'_{\varphi_3 \varphi_2} \end{pmatrix}^{-1} \begin{pmatrix} V_{\varphi_2} \\ V_{\varphi_3} \end{pmatrix}$$
using the matrix partition (Lütkepohl, 1996, p.30) and extracting the relevant part of the inverse of the information matrix corresponding to the bottom right block, for which \( \varphi_2 \) and \( \varphi_3 \) are set to zero under the null hypothesis.

Thus, the test statistic is given by

\[
LM_{NEST} = T^{-1} \left( \sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta})}{\partial \varphi_2, \varphi_3} \right) ' \left[ \hat{I}(\hat{\theta}) \right]^{-1} \left( \sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta})}{\partial \varphi_2, \varphi_3} \right),
\]

where \( LM_{NEST} \xrightarrow{D} \chi^2_2 \) under \( H_0 : \varphi_2 = \varphi_3 = 0 \).
8.5 The LM-statistic: logistic transition function

First, consider the score

\[
T \sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} = \begin{bmatrix}
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \phi_1} \\
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \beta} \\
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \phi_2}
\end{bmatrix}
\times \begin{bmatrix}
\frac{N(N+1)}{2} \times 1
\end{bmatrix} = \begin{bmatrix}
Dimension \n\end{bmatrix}
\times \begin{bmatrix}
V_C \\
V_{\phi_1} \\
V_\beta \\
V_{\phi_2}
\end{bmatrix}.
\]

Equation (23) can be written as

\[
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} = -\frac{1}{2} \sum_{t=1}^{T} \left[ \left( \frac{\partial \text{vech}H_t}{\partial \theta} \right) \right]^t \left( \frac{\partial \text{vech}H_t}{\partial \theta} \right)' D_N \text{vech}H_t^{-1}
\]

\[\quad - \left( \frac{\partial \text{vech}H_t}{\partial \theta} \right)' (H_t^{-1} \otimes H_t^{-1}) D_N \text{vech}(\varepsilon_t \varepsilon_t') \right].
\]

Second, consider the information matrix

\[
I(\theta) = -T^{-1} T \sum_{t=1}^{T} \left( \frac{\partial \text{vech}H_t}{\partial \theta} \right) \left( \frac{\partial \text{vech}H_t}{\partial \theta} \right)'

\]

\[\quad - \left( \frac{\partial \text{vech}H_t}{\partial \theta} \right)' (H_t^{-1} \otimes H_t^{-1}) D_N \frac{\partial \text{vech}H_t}{\partial \theta}.
\]

Third, consider the inversion of the matrix \(I(\theta)\) and the analytical computation of the LM test

The LM test of No Logistic Smooth Transition, denoted \(LM_{NLST}\), can be expressed as

\[
LM_{NLST} = \begin{bmatrix}
V_C \\
V_{\phi_1} \\
V_\beta \\
V_{\phi_2}
\end{bmatrix}
\]

\[\begin{bmatrix}
M_{CC} & M_{\phi_1C} & M_{\beta C} & M_{\phi_2C} \\
M_{\phi_1C} & M_{\phi_1\phi_1} & M_{\beta\phi_1} & M_{\phi_2\phi_1} \\
M_{\beta C} & M_{\beta\phi_1} & M_{\beta\beta} & M_{\phi_2\beta} \\
M_{\phi_2C} & M_{\phi_2\phi_1} & M_{\phi_2\beta} & M_{\phi_2\phi_2}
\end{bmatrix}^{-1}
\]

\[\begin{bmatrix}
V_C \\
V_{\phi_1} \\
V_\beta \\
V_{\phi_2}
\end{bmatrix}
\]

where \(M_{..}\) is the information matrix w.r.t its parameters.
Last, consider the LM test

Under the null hypothesis, we maximize the quasi-likelihood function with respect to $\text{vech}(C)'$, $\varphi_1$ and $\beta$. Then, the first derivatives of the quasi-likelihood function with respect to these three parameters are set to zero:

$$
\sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} = \begin{bmatrix}
0 \\
0 \\
0 \\
\end{bmatrix}
$$

and

$$
LM_{NLST}^{1 \times 1} = V'_{\phi_2} \begin{bmatrix} M_{\phi_2 \phi_2} - (M_{\phi_2 C} M_{\phi_2 \phi_1} M_{\phi_2 \beta} ) \\
M_{\phi_1 C} M_{\phi_1 \phi_1} M_{\beta C} \\
M_{\beta C} M_{\beta \phi_1} M_{\beta \beta} \\
\end{bmatrix}^{-1} \begin{bmatrix} M_{\phi_2 C} M_{\phi_2 \phi_1} M_{\phi_2 \beta} \\
M_{\phi_2 C} M_{\phi_2 \phi_1} M_{\phi_2 \beta} \\
M_{\phi_2 C} M_{\phi_2 \phi_1} M_{\phi_2 \beta} \\
\end{bmatrix} V_{\phi_2}
$$

using the matrix partition (Lütkepohl, 1996, p.30) and extracting the relevant part of the inverse of the information matrix corresponding to the bottom right block, for which $\phi_2$ is set to zero under the null hypothesis.

Thus, the test statistic is given by

$$
LM_{NLST} = T^{-1} \left( \sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta})}{\partial \phi_2} \right)' \left[ \tilde{I}(\hat{\theta}) \right]_{(\phi_2,\phi_2)}^{-1} \left( \sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta})}{\partial \phi_2} \right),
$$

where $LM_{NLST} \xrightarrow{D} \chi_1^2$ under $H_0: \phi_2 = 0$. 

26
References


Ling S. and M. McAleer (2002a). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) model, *Econometric Theory*, 18, 722–729.


