SPECIFICATION TESTING OF GARCH REGRESSION MODELS

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Humanities

Year of Submission
2011

WASEL BIN SHADAT

School of Social Science
Economics

Contents

Final Word Count: 36372

List of Tables

Abstract

This thesis analyses, derives and evaluates specification tests of Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) regression models, both univariate and multivariate. Of particular interest, in the first half of the thesis, is the derivation of robust test procedures designed to assess the Constant Conditional Correlation (CCC) assumption often employed in multivariate GARCH (MGARCH) models. New asymptotically valid conditional moment tests are proposed which are simple to construct, easily implementable following the full or partial Quasi Maximum Likelihood (QML) estimation and which are robust to non-normality. In doing so, a non-normality robust version of the Tse's (2000) LM test is provided. In addition, a new and easily programmable expressions of the expected Hessian matrix associated with the QMLE is obtained. The finite sample performances of these tests are investigated in an extensive Monte Carlo study, programmed in GAUSS.

In the second half of the thesis, attention is devoted to nonparametric testing of GARCH regression models. First simultaneous consistent nonparametric tests of the conditional mean and conditional variance structure of univariate GARCH models are considered. The approach is developed from the Integrated Generalized Spectral (IGS) and Projected Integrated Conditional Moment (PICM) procedures proposed recently by Escanciano (2008 and 2009, respectively) for time series models. Extending Escanciano (2008), a new and simple wild bootstrap procedure is proposed to implement these tests. A Monte Carlo study compares the performance of these nonparametric tests and four parametric tests of nonlinearity and/or asymmetry under a wide range of alternatives. Although the proposed bootstrap scheme does not strictly satisfy the asymptotic requirements, the simulation results demonstrate its ability to control the size extremely well and therefore the power comparison seems justified. Furthermore, this suggests there may exist weaker conditions under which the tests are implementable. The simulation exercise also presents the new evidence of the effect of conditional mean misspecification on various parametric tests of conditional variance. The testing procedures are also illustrated with the help of the S&P 500 data. Finally the PICM and IGS approaches are extended to the MGARCH case. The procedure is illustrated with the help of a bivariate CCC-GARCH model, but can be generalized to other MGARCH specifications. Simulation exercise shows that these tests have satisfactory size and are robust to non-normality. The marginal mean and variance tests have excellent power; however the covariance marginal tests lack power for some alternatives.

Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

To the T4 family (Tuba, Taniba and Tazima).

Acknowledgement

It gives me immense pleasure to express my deep gratitude and appreciation to the members of my supervisory team, Professor Chris Orme, Professor Alastair Hall and Dr Len Gill, who accepted to supervise this PhD research. Their continuing academic support and encouragement, useful comments and suggestions, and friendly way to discuss the problems and issues, throughout the writing of this thesis, were invaluable. A special mention must be given to Chris Orme to whom I am deeply indebted for introducing me to the beauty of the econometrics theory whilst a master student. Without his patient guidance, wisdom, enthusiasm and friendship it is doubtful whether this work could have taken its current form.

I would also like to thank Professor Denies Osborn, Ralf Becker, Simon Peters, Andreea Halunga and Christos Savva for their valuable inputs on my thesis.

This PhD research was supported by a Commonwealth Scholarship and Fellowship Plan, a Manchester School Award and a Charles Wallace Bangladesh Trust Doctoral Bursary Award; all of which are gratefully acknowledged. I would also like to thank the British Council for administering the Commonwealth Scholarship and Fellowship Plan for my PhD degree.

Many thanks go to my colleagues and friends in Manchester for their generosity and friendship. I gratefully acknowledge the continuous support, encouragements and love I have received from my friends and family back in Bangladesh.

Finally, a very special thanks to my wife Tuba and my two angels Taniba and Tazima, for their constant support, patience and love. In addition to taking care of our daughters and thereby freeing up time for me to do this research, my wife has played an invaluable role to complete this research by providing encouragement and assistance during the difficult time when nothing was going right.

Chapter 1

Introduction

The univariate Auto-Regressive Conditional Heteroskedasticity (ARCH) model, introduced by Engle (1982), and its univariate extension the Generalized ARCH (GARCH) model proposed by Bollerslev (1986) have enjoyed remarkable success in various fields, particularly modelling the volatility of economic and financial time series data. For more than twenty years many extensions of the original ARCH and GARCH models have been proposed in the literature. The existence of a gigantic body of literature, which uses these processes in modelling conditional volatility, demonstrates the popularity of various (G)ARCH models. Recognizing the importance of the interdependence among economic and financial variables, numerous multivariate extensions of the models (multivariate GARCH, or MGARCH) have been proposed in the literature.[1]

However, an investigation of the literature reveals that despite the escalating attention in designing various GARCH and MGARCH specifications, little interest has been paid in the applied literatures to testing the adequacy of the models employed. Although a number of studies deals with the parametric diagnostic tests of GARCH models (see for example Halunga and Orme, 2009 and the references therein), testing (both parametric and non-

[1]For a survey of univariate (G)ARCH models, see, e.g., Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994), Shephard (1996) among others. For surveys on MGARCH models, see, e.g., Laurent, Bauwens and Rombouts (2006) and Silvennoinen and Terasvirta (2008).

parametric) of the adequacy of MGARCH models and nonparametric testing of GARCH models are less explored by the literature. The two recent surveys on MGARCH models, one by Laurent, Bauwens and Rombouts (2006) and the other by Silvennoinen and Teräsvirta (2009b), put emphasis on the paucity of diagnostic tests for the MGARCH models and identified developing specification tests as one of the most intriguing research topics in this field. Acknowledging this inadequacy, the overarching theme of this thesis is the specification testing of GARCH regression models.

We begin in Chapter 2 with a review of the ever-growing literature on MGARCH models with particular attention on the conditional correlation approach. The correlation approach is very popular and parsimonious way of modelling MGARCH models which, instead of modelling the variance-covariance matrix directly, uses the classical decomposition of the variance-covariance matrix; and model first the individual conditional volatility through some GARCH model and then specify the correlation separately. In the MGARCH literature both the Constant Conditional Correlation (CCC) (e.g., Bollerslev, 1990) and time varying (or dynamic) conditional correlation (e.g., Engle, 2002) approaches are suggested.

In Chapter 3 an asymptotically valid Conditional Moment (CM) testing procedure of the CCC assumption of MGARCH models is proposed considering both the Full Quasi Maximum Likelihood Estimation (FQMLE) and Partial or two-stage QMLE (PQMLE) framework. A "new" and easily programmable expression for the expected Hessian is provided for the FQMLE. The Outer Product of Gradients (OPG) and robust to non-normality versions of the test statistics are derived. The Tse (2000) OPG-type Lagrange Multiplier (LM) test of the CCC assumption is analyzed within our CM framework. In addition a new robust version of this test is proposed. An extensive Monte Carlo investigation demonstrates good size and power properties. The OPG versions suffer from size distortion under non-normality whereas robust versions perform much better. All simulation experiments in this thesis are programmed in GAUSS.

Chapter 4 deals with simultaneous nonparametric (consistent) testing of the conditional mean and conditional variance of the GARCH regression model. It has been widely noted that a misspecified mean may induce bias in the parametric diagnostic tests for variance specification which inherently assume a correct specification for the mean, and thereby may lead to misleading conclusions.[2] Therefore it is important to test the conditional mean and variance simultaneously. We consider the Projected Integrated Conditional Moment (PICM) (Escanciano, 2009) and Integrated Generalized Spectral (IGS) (Escanciano, 2008) testing framework for this purpose. The integrated approach of consistent testing requires an approximation to the limit null distribution of the test statistic by some bootstrap technique and the wild bootstrapping is the most popular choice. The full parametric model based bootstrap of GARCH models is an operationally problematic issue. Escanciano (2009) proposed a Fixed Design Wild Bootstrap (FDWB) algorithm for joint and marginal testing problem for certain conditionally heteroskedastic models. This thesis argued that his FDWB is not applicable in the case of null GARCH regression model. To address this, a new and easily implementable FDWB procedure for GARCH regression models is proposed. Although the asymptotic analysis shows that the proposed bootstrap scheme does not satisfy the sufficient conditions established in the literature (Escanciano, 2007b), it does not necessarily mean the procedure is asymptotically invalid. The simulation study suggests that it controls the size extremely well, even in small sample, and thereby justifying the power comparison.[3] Furthermore, this suggests there may exist weaker conditions under which the tests are implementable. An extensive Monte Carlo study is conducted to compare the performance of these two nonparametric tests and four parametric tests of nonlinearity and/or asymmetry considered in Halunga and Orme (2009), Engle and Ng (1993) and Lundbergh and Teräsvirta (2002). The simulation exercise also provides us with the opportunity to investigate the effect of conditional mean misspecification on various parametric LM and CM tests of the conditional variance in the regression context which is not

[2] See, e.g., Halunga and Orme (2009), McAleer (2005), Lumsdaine and Ng (1999) among others.

[3] There are many instances in diagnostic checking literature that asymptotically valid tests display appalling size property which subsequently lead to inconsequential power comparison.

done in the literature before. The testing procedures are also illustrated with the help of the S&P 500 data.

In Chapter 5 we extend the work of Chapter 4 to test the moments of MGARCH models simultaneously. To be precise, this Chapter investigates whether the extensions of the univariate PICM and IGS tests are applicable to assess the conditional mean and variance-covariance specification simultaneously. Unlike the parametric tests of the correlation assumption which consider a particular null correlation model in advance (e.g., CM tests proposed in Chapter 3, Tse's (2000) LM test), these nonparametric tests are applicable to any MGARCH models. As a leading case of interest the CCC-GARCH model is used to illustrate the testing procedures but can be generalized to other MGARCH specifications. To implement the IGS tests we need to extend the wild bootstrap technique developed in Chapter 4 to accommodate the conditional covariance structure. The finite sample performances of these tests are investigated via a Monte Carlo study.

Finally, Chapter 6 concludes the thesis.

1.1 Notation

In this thesis throughout we will use the following notation:

- $\mathrm{E}(.)$ is the expectation operator, $\mathrm{E}_0(.)$ and $\mathrm{E}_{t-1}(.)$ denote the expectation with respect to true parameter value and conditional on previous history up to $t-1$, respectively.

- $\mathrm{tr}(A)$, $|A|$ and $\|A\| = (\mathrm{tr}(A'A))^{1/2}$ denote the trace, determinant and Euclidean norm of A, respectively.

- $\otimes$ and $\odot$ denote the Kronecker and Hadarmard product, respectively.

- $\mathrm{vec}(.)$, $\mathrm{vech}(.)$ and $\mathrm{vecl}(.)$ denote the operator that stacks the columns of $(N \times N)$ matrix as a $(N^2 \times 1)$ vector, the lower triangular portion of a $(N \times N)$ matrix as a $\left(\frac{N(N+1)}{2} \times 1\right)$ vector and the strictly lower triangular portion of a $(N \times N)$ matrix as a $\left(\frac{N(N-1)}{2} \times 1\right)$ vector, respectively.

- $I_N = \{\delta_{ik}\}$ is the identity matrix of order N; with δ_{ik} denoting the kronecker delta, $\iota'_K = (1, 1, ..., 1)$ is $(1 \times K)$ vector of ones, and $\mathcal{J}_K = \iota_K \iota'_K$ is the $(K \times K)$ matrix of ones.

Chapter 2

A Review of Conditional Correlation MGARCH Model

2.1 Introduction

Since the seminal work of Engle (1982) which introduced the ARCH model, many researchers have been working to model the time-varying volatility of economic data both in the univariate and, more recently, multivariate contexts. The empirical success of the univariate ARCH and its extension the "Generalized ARCH" (GARCH) models, proposed by Bollerslev (1986), have inspired and motivated a new generation of models to capture the dynamic conditional volatility in financial time series data. A number of specifications have been proposed in the literature to make the original ARCH and GARCH models more flexible so that they can capture the stylized facts and features demonstrated by the economic data. This stream of research has resulted in several variants and extensions of the GARCH including, for example, the Glosten, Jagannathan, and Runkle (1993) asymmetric (or threshold) GARCH (GJR) model, the Nelson (1991) Exponential GARCH (EGARCH) model, the Ding, Granger and Engle (1993) Asymmetric Power GARCH (APGARCH) model, the Threshold ARCH (TARCH) model of Zakoian (1994). For a survey of univariate ARCH and GARCH type models, see Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994), Pagan (1996), Palm (1996), and Shephard (1996) among others.

Although these univariate models are useful in modelling volatility of returns, they cannot capture the interdependence or spillover effects in volatility across different markets or assets. Estimating and forecasting the conditional covariance structure of returns has critical practical importance in various areas of finance such as portfolio construction, asset pricing and risk management. For example, portfolio construction and risk management depend on finding and updating optimal hedging positions, asset pricing depends on the covariance of the assets in a portfolio. To accommodate spillovers in conditional volatility, the search for adequate multivariate models of a large number of assets has continued for more than twenty years and several versions of MGARCH models have been proposed in the literature. Two excellent surveys on MGARCH models have been conducted recently by Laurent, Bauwens and Rombouts (2006) and Silvennoinen and Teräsvirta (2009b) and one less recently by Bollerslev, Engle and Nelson (1994).

In their survey, Laurent, Bauwens and Rombouts (2006) broadly categorize MGARCH models into three non mutually exclusive classes based on the modelling approaches. The first is the early generation MGARCH models which attempted to model conditional covariances by generalizing directly the univariate GARCH model; for example, the VEC and Diagonal-VEC (DVEC) model of Bollerslev, Engle and Wooldridge (1988), the Baba, Engle, Kraft and Kroner (BEKK) MGARCH models of Engle and Kroner (1995). Instead of modelling directly the conditional covariance matrix and motivated by parsimony, another group of models utilizes a few number of factors to model conditional covariance matrix, such as the Factor GARCH (FGARCH) model of Engle, Ng and Rothschild (1990), the Full Factor MGARCH (FFMGARCH) model of Vrontos, Dellaportas and Politis (2003). However, it can be seen that these factor models are a special case of the BEKK model.

The second approach is based on the linear combination of univariate GARCH models such as the Orthogonal GARCH (OGARCH) model of Kariya (1988) and Alexander and Chibumba (1997) and the Generalized Orthogonal GARCH (GOGARCH) of van der Weide (2002). The univariate models that are used to derive the OGARCH or GOGARCH models are not necessarily the standard GARCH models. They could be the EGARCH model of Nelson (1991), the APARCH model of Ding, Granger and Engle (1993), Quadratic ARCH model of Sentana (1995), the fractionally integrated GARCH of Baillie, Bollerslev and Mikkelsen (1996), or the contemporaneous asymmetric GARCH model of El Babsiri and Zakoian (2001). These orthogonal models are particular cases of the FGARCH models and therefore are nested in the BEKK model and their properties follow from those of the BEKK model.

The third category is the collection of models which can be seen as nonlinear combinations of univariate GARCH models. These models allow a decomposition of the conditional covariance matrix by specifying the conditional variance and conditional correlation matrix (or another measure of dependence between individual series) separately. The first subgroup of this category contains the family of models that assumes the conditional correlation to be constant. For example, the constant conditional correlation (CCC) MGARCH model of Bollerslev (1990), the Ling and McAleer (2003) Vector ARMA-GARCH (VARMA-GARCH) model and the McAleer, Hoti and Chan (2009) VARMA-Asymmetric GARCH (VARMA-AGARCH). Next there is a group of models which relaxes the CCC assumption and model the conditional correlations and conditional covariances jointly; for example, the Engle (2002) Dynamic Conditional Correlation (DCC) model, the Tse and Tsui (2002) Varying Conditional Correlation (VCC) model, the Asymmetric Generalized DCC (AGDCC) GARCH model of Cappiello, Engle, and Sheppard (2006), the General Dynamic Covariance (GDC) model of Kroner and Ng (1998), the Smooth Transition CC (STCC) GARCH model of Silvennoinen and Teräsvirta (2005), the Double Smooth Transition CC (DSTCC) GARCH model of Silvennoinen and Teräsvirta (2009a), the Regime Switching Dynamic Correlation (RSDC) GARCH model of Pelletier (2006), the McAleer, Chan, Hoti and Lieberman (2008) Generalized Auto-Regressive Conditional Correlation (GARCC) model. The copula MGARCH models of Patton (2006) and Jondeau and Rockinger (2006) also fall under this category of nonlinear combinations of univariate GARCH models. Apart from these parametric specifications, a number of studies attempted to model the conditional variance and conditional correlation non-parametrically and semi-parametrically; see, for example, Hafner and Rombouts (2007), Hafner, Dijk and Franses (2005), Long and Ullah (2005). More recently Noureldin, Shephard and Sheppard (2011) introduced a new class of multivariate High-frEquency-bAsed VolatilitY (HEAVY) models that utilize the realized measure of volatility and are capable of producing better multi-step forecasts of the conditional covariance matrix.

However, as the literature related to MGARCH models is vast and ever increasing, we need to place some limits on our discussion in this Chapter. For this reason, models associated with factors, orthogonalization, copula GARCH and HEAVY will not be discussed here. Of all MGARCH specifications, the class of conditional correlation models, particularly the dynamic one, is the most popular among the empirical researchers. In modelling financial data and risk, correlations play a vital role. Recognizing that the economy is an interconnected set of economic agents, it is hardly surprising that the movement in prices are correlated. If they were independent, then it would be possible to form a portfolio with negligible volatility. In reality, this is clearly not the case. For example, stock returns in a particular stock exchange market are correlated, stocks in one country or market are correlated with stocks in another, bond returns on one firm or country or maturity are generally correlated with returns on others.[1] Correlation structure is instrumental to determine the risk of a portfolio. The risk of a portfolio is greater if all the assets are highly correlated. It may go down (or up) further, if they all move together. Thus the correlation structure across assets is a key feature of the portfolio choice problem. Moreover, it is now widely agreed that the return volatilities and correlations are time varying with persistent dynamics which makes the problem more difficult. Therefore, in optimal risk management, portfolio selection and hedging, one has to "anticipate correlations" with a forward looking correlation estimator. The field of financial econometrics devotes considerable attention to modelling dynamic volatility and correlation. The existing methods of estimating and forecasting correlations

[1] However, stock and bond returns sometimes appear uncorrelated.

include historical correlation (which use a rolling window of N observations for both covariances and variances), exponential smoothing and MGARCH models. Among these, MGARCH models are the most recent, sophisticated and popular among researchers. This motivates us to study the MGARCH models with conditional correlation specification.

In the next section we will discuss various MGARCH specifications, starting from the basic MGARCH framework and moving towards to the focus of our interest; i.e., the conditional correlation models. This will be followed by the discussion on parametric and nonparametric estimation techniques emerged in the literature. For the sake of completeness the univariate GARCH model is also presented in the Appendix.

2.2 Basic MGARCH Model

Consider a stochastic vector process $\{y_t\} = (y_{1t}, \cdots, y_{Nt})'$ with dimension $N \times 1$. We write the regression function as:

$$y_{it} = m\left(w_{it}; \varphi_i\right) + \varepsilon_{it}, \ i = 1, ..., N \ \ t = 1, ..., T, \tag{2.1}$$

where $\varphi_i \in \Psi \subset \Re^K$. In vector/matrix form (at the observational, t, level), this can be expressed as:

$$y_t = m\left(W_t; \varphi\right) + \varepsilon_t, \ \ t = 1, ..., T, \tag{2.2}$$

where $y_t = (y_{1t}, \cdots, y_{Nt})'$and $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})'$ are $N \times 1$ vectors, $\varphi = (\varphi_1', .., \varphi_N')'$ is a $(NK \times 1)$ vector of conditional mean parameters and W_t' is the $(N \times NK)$ data matrix of the t-th observation with the form:

$$\underset{N\times NK}{W_t'} = \begin{bmatrix} w_{1t}' & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_{Nt}' \end{bmatrix}. \tag{2.3}$$

Note that $m\left(W_t; \varphi\right)$ can possibly be nonlinear and we implicitly assume that W_t contains current and lagged exogenous variables, and lagged dependent (predetermined) variables $y_{t-1}, y_{t-2}, ...$ However, for simplicity of exposition we assume a linear specification for the conditional mean function; i.e., $m(W_t;\varphi) = W_t'\varphi$ so that $y_t = W_t'\varphi + \varepsilon_t \quad t = 1, ..., T$. Define $\mathcal{F}_{t-1}$ as the $\sigma-$field generated by the past information up to and including time $t-1$.

In stacked form we can write the above as:

$$\begin{aligned}
y &= W\varphi + \varepsilon, \qquad (2.4)\\
y &= [y_1', y_2', ..., y_T']', (NT \times 1),\\
W &= \begin{bmatrix} W_1' \\ W_2' \\ \vdots \\ W_T' \end{bmatrix}, (NT \times NK),\\
\varepsilon &= [\varepsilon_1', \varepsilon_2', ..., \varepsilon_T']', (NT \times 1).
\end{aligned}$$

We assume that, the error $\{\varepsilon_t, \mathcal{F}_t\}$ is a Martingale Difference Sequence (MDS) and given by:

$$\varepsilon_t = H_t^{1/2}(\theta)\xi_t, \qquad (2.5)$$

where $H_t^{1/2}$ is a $N \times N$ positive definite matrix and θ is the vector of unknown parameters which includes conditional mean parameters defined in (2.1) or (2.4) as well (for notational convenience, we drop θ in $H_t^{1/2}(\theta)$).

We further assume, following the literature, that the $N \times 1$ random vector ξ_t is an independent and identically distributed (i.i.d.) process with the following first two moments:

$$\begin{aligned}
\mathrm{E}(\xi_t) &= 0,\\
Var(\xi_t) &= \mathrm{E}(\xi_t\xi_t') = I_N, \qquad (2.6)
\end{aligned}$$

where I_N is the identity matrix of order N. Again note that,

$$\begin{aligned} Var(y_t|\mathcal{F}_{t-1}) &\equiv Var_{t-1}(y_t) \\ &= Var_{t-1}(\varepsilon_t) \\ &= H_t^{1/2} Var_{t-1}(\xi_t') H_t'^{1/2} \\ &= H_t. \end{aligned} \tag{2.7}$$

Hence $H_t^{1/2}$is any $N \times N$ positive definite matrix such that $H_t = [h_{ijt}]$ is the $N \times N$ (positive definite and symmetric) conditional ($\mathcal{F}_{t-1}$ measurable) variance-covariance matrix of y_t; for example, $H_t^{1/2}$ may be obtained from Cholesky factorization. Given this specification, the conditional mean, conditional variance and conditional correlation of the error ε_t are given by,

$$\begin{aligned} \mathrm{E}\left[\varepsilon_{it}|\mathcal{F}_{t-1}\right] &= 0, \\ \mathrm{E}\left[\varepsilon_{it}^2|\mathcal{F}_{t-1}\right] &= h_{it} \equiv h_{iit}, \\ corr\left[\varepsilon_{it}, \varepsilon_{jt}|\mathcal{F}_{t-1}\right] &= \frac{h_{ijt}}{\sqrt{h_{it}}\sqrt{h_{jt}}}, \end{aligned}$$

where $i, j = 1, ..., N, \quad t = 1, ..., T$. Note that $E\left[\varepsilon_t \varepsilon_{t-j}'|\mathcal{F}_{t-1}\right] = E\left[\varepsilon_t|\mathcal{F}_{t-1}\right] \varepsilon_{t-j}' =$ 0, almost surely, for all $j \geq 1$.

Now we need to specify the matrix process H_t. In Multivariate GARCH (MGARCH) models H_t is specified parametrically by generalizing the univariate GARCH process.[2] The generic form of the MGARCH(p, q) is given by:

$$H_t = C + \sum_{k=1}^{q} A_k^* \left(\varepsilon_{t-k}\varepsilon_{t-k}'\right) A_k^{*\prime} + \sum_{m=1}^{p} B_m^* H_{t-j} B_m^{*\prime}. \tag{2.8}$$

where C, A_k^* and B_m^* are the parameter matrices. Both the conditional mean and conditional variance depend on unknown parameter vector θ, which can generally be split in two disjoint parts one for conditional mean and one for conditional variance, H_t.[3] However although the GARCH parameters do not

[2] For a detailed discussion on the univariate GARCH, see Appendix 2.A and references therein.

[3] One example where the conditional mean and conditional parameters can not be separated is the GARCH-in-mean models, where conditional mean is functionally dependent

affect the conditional mean, the conditional mean parameters generally enter the conditional variance specification through the lagged squared errors.

The problem is now to achieve a parsimonious parameterization of this structure.

2.2.1 Critical Issues Related to MGARCH Modelling

Before turning to the various specifications for H_t, few general points regarding the MGARCH models should be emphasized:

1. One needs to ensure the positive definiteness of H_t, which is a numerically difficult problem in large systems.

2. The number of parameters can increase very rapidly as the dimension of ε_t increases creating enormous difficulties in estimation of the models. So one of the important objectives for MGARCH specifications is to make them reasonably parsimonious while maintaining flexibility.

Although univariate GARCH models are applied extensively by empirical researchers, MGARCH models have not yet enjoyed the same degree of popularity. McAleer, Chan, Hoti, and Lieberman (2008) identified four reasons behind this:

1. lack of a theoretical foundation of some MGARCH models,

2. the interpretations of parameters are not always straightforward,

3. the rapid and significant increase of number of parameters with number of assets and/or markets,

4. computational problem for large system.

Next, we will briefly discuss a few selected MGARCH specifications before we move to the our main interest; i.e., the conditional correlation models.

on H_t.

2.3 Modelling H_t Directly

2.3.1 The VEC-GARCH Model

First, consider the VEC-GARCH model of Bollerslev, Engle and Wooldridge (1988) which is a straightforward generalization of the univariate GARCH model. In this case, each element of H_t is a linear function of all lagged conditional variances and covariances (i.e., lagged values of the elements of H_t), lagged squared errors and cross product of errors. The VEC-GARCH(p, q) model may be written as:

$$\text{vech}(H_t) = c + \sum_{k=1}^{q} A_k \,\text{vech}(\varepsilon_{t-k}\varepsilon'_{t-k}) + \sum_{m=1}^{p} B_m \,\text{vech}(H_{t-m}), \tag{2.9}$$

where c is an $\frac{N(N+1)}{2} \times 1$ vector and A_k and B_m are square parameter matrices of order $\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}$. For the VEC-GARCH$(p, q)$ model we have, in fact, for $r \leq s = 1, ..., N$,

$$h_{rs,t} = c_{rs} + \sum_{k=1}^{q}\sum_{i=1}^{N}\sum_{j=i}^{N} \alpha_{ijk}\varepsilon_{i,t-k}\varepsilon_{j,t-k} + \sum_{m=1}^{p}\sum_{i=1}^{N}\sum_{j=i}^{N} \beta_{ijm}h_{ij,t-m}, \tag{2.10}$$

whereby *all* conditional variances and covariances, from all series, affect $h_{rs,t}$.

This model is very general and highly flexible on the one hand, while it is quite restrictive in two counts on the other. First, there exists only sufficient, which is indeed quite restrictive, conditions for H_t to be positive definite for all t (Gouriéroux, 1997, Chapter 6). Secondly, the number of parameters is $(p + q)(N(N + 1)/2)^2 + N(N + 1)/2$, which is quite large unless N is small. For example, for the VEC-GARCH$(1, 1)$ model with $N = 3$, the number of parameters is 78 implying that in practice this model could only be used in the bivariate case.

The DVEC Model

Bollerslev, Engle and Wooldridge (1988) proposed the Diagonal VEC (DVEC) to simplify the VEC-GARCH model where both the matrices A_k and B_m in (2.9) are assumed to be diagonal. The DVEC model implies that each element of H_t depends only on its own lags and lagged values of ε_t^2. This yields the DVEC(p, q) model specification as follows:

$$h_{ij,t} = c_{ij} + \sum_{k=1}^{q} \alpha_{ijk}\varepsilon_{i,t-k}\varepsilon_{j,t-k} + \sum_{m=1}^{p} \beta_{ijm} h_{ij,t-m}, \quad j \leq i = 1, ..., N. \quad (2.11)$$

For example, the bivariate DVEC-GARCH$(1, 1)$ is (with appropriate redefinition of the c's, a's and b's)

$$\begin{aligned} h_{1t} &= c_{11} + a_{11}\varepsilon_{1,t-1}^2 + b_{11}h_{11,t-1}, \\ h_{12t} &= c_{12} + a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{12}h_{12,t-1}, \\ h_{2t} &= c_{22} + a_{22}\varepsilon_{2,t-1}^2 + b_{22}h_{22,t-1}. \end{aligned}$$

In this case, it is possible to derive conditions for H_t to be positive definite for all t.[4] This diagonal restriction reduces the number of parameters to $(p + q + 1)N(N + 1)/2$.

However, due to this restriction the model loses its flexibility since no interaction between conditional variances and covariances is allowed. Under standard regularity conditions the ML estimators for θ (both in VEC and DVEC models) is asymptotically normal and traditional inference procedures can readily be applied.[5] It can be seen from (2.55) that the conditional covariance matrix H_t has to be inverted for every t in each iteration, which may be problematic and tedious when both T and N are large.

[4] See Bollerslev, Engle and Nelson (1994).

[5] Following Bollerslev and Wooldridge (1992) and Weiss (1986), if the first two conditional moments are specified correctly by the model, but the conditional normality assumption is violated, under suitable regularity conditions the quasi-maximum likelihood estimators (QMLE) obtained from conditional log-likelihood function (under conditional normality assumption) will still be consistent and asymptotically normal, but the usual standard errors have to be modified.

2.3.2 The BEKK Model

As an alternative to the VEC-GARCH model, Engle and Kroner (1995) proposed the BEKK (Baba, Engle, Kraft and Kroner) model in which the conditional covariance matrix H_t is positive definite by construction. This model can be viewed as a restricted version of the general VEC model, with the BEKK(p, q, K) is defined as:

$$H_t = CC' + \sum_{i=1}^{q}\sum_{l=1}^{K} A'_{li}\varepsilon_{t-i}\varepsilon'_{t-i}A_{li} + \sum_{m=1}^{p}\sum_{l=1}^{K} B'_{lm}H_{t-m}B_{lm}, \tag{2.12}$$

where A_{li}, B_{lm} and C are $N\times N$ parameter matrices and C is lower triangular. The number of parameters of this model is $(p+q)KN^2+N(N+1)/2$ which is again very large. A simplified (and restrictive as well) version is the diagonal BEKK model which is again a restricted version of the DVEC model. Here the number of parameters reduces to $(p + q)KN + N(N + 1)/2$ which still is quite large. Although the model ensure positive definiteness, due to the non-linearity in parameters in (2.12) numerical optimization is a problematic issue for this model with large system. In practice, therefore, $p = q = K = 1$ is assumed. For the BEKK$(p, q, 1)$, (2.12) becomes

$$H_t = CC' + \sum_{i=1}^{q} A'_i\left(\varepsilon_{t-i}\varepsilon'_{t-i}\right)A_i + \sum_{m=1}^{p} B'_m H_{t-m}B_m. \tag{2.13}$$

Remark 2.1 *As noted in Section 2.3.1, this has a VEC-GARCH representation. That is, (2.13) a restricted version of (2.9); see Stelzer (2008).*

Remark 2.2 *If C is lower triangular and $H_0, H_{-1}, ..., H_{1-p}$, the starting values for H, are positive semidefinite (psd), then H_t is positive definite (pd) for all $t = 1, ..., T$. This is intuitively obvious and can be proved by induction. Consider just the BEKK$(1, 1, 1)$ model, $H_t = CC' + A\left(\varepsilon_{t-1}\varepsilon'_{t-1}\right)A' + BH_{t-1}B'$. Suppose H_τ is pd, $\tau \geq 1$, then*

$$H_{\tau+1} = CC' + A\left(\varepsilon_\tau\varepsilon'_\tau\right)A' + BH_\tau B',$$

which must be pd, because (a) C has full rank $\Longrightarrow$ CC' is pd, (b) $A\left(\varepsilon_\tau\varepsilon_\tau'\right)A'$ is always psd, and (c) $BH_\tau B'$ is psd (it is pd iff B has full rank). Thus, if H_τ is pd, $\tau \geq 1$, then $H_{\tau+1}$ must be pd. Now $H_1 = CC' + A\left(\varepsilon_0\varepsilon_0'\right)A' + BH_0B'$, must be pd. Therefore, by induction, H_t is pd for all $t = 1, ..., T$.

It should be noted that the interpretation of parameters in (2.12) is not easy, unlike the VEC model, they do not represent directly the impact of different lagged terms on the elements of H_t. In addition, whenever $K > 1$, an identification problem arises as several parameterization could yield the same representation of the model. Engle and Kroner (1995) provide sufficient conditions for eliminating redundant, observationally equivalent representations.

Remark 2.3 *Another option to guarantee the positivity of H_t in the VEC representation is to use the Cholesky decomposition of H_t (e.g., define $H_t = L_tL_t'$ as proposed by Kawakatsu, 2003). In this case H_t is always positive definite without any restriction on the parameters, but identification restrictions are needed and the interpretation of parameters is difficult.*

2.3.3 A Few Other Models

The following specifications also model the conditional covariance matrix H_t directly, however these models will be not discussed later in this study. Another model which ensures positive definiteness of H_t is the Kawakatsu's (2006) matrix exponential GARCH model which is a generalization of the univariate EGARCH model of Nelson (1991) and is defined as follows:

$$\mathrm{vech}(\ln H_t - C) = \sum_{i=1}^{q} A_i\xi_{t-i} + \sum_{i=1}^{q} F_i(|\xi_{t-i}| - \mathrm{E}\,|\xi_{t-i}|) + \sum_{j=1}^{q} B_j\,\mathrm{vech}(\ln H_{t-j} - C), \tag{2.14}$$

where ξ_{t-i} is defined in (2.6), A_i and F_i are parameter matrices of sizes $N(N+1)/2 \times N$, B_j is parameter matrix of size $N(N+1)/2 \times N(N+1)/2$ and C is a symmetric $N \times N$ parameter matrix.

The Factor GARCH (FGARCH) models parameterize H_t using the idea that comovements of returns are determined by a small number of common underlying variables, which are called factors. For instance, in the arbitrage pricing theory of Ross (1976), returns are generated by a number of unobserved components. Assuming H_t is generated by $K(< N)$ factors which may be correlated, Engle, Ng and Rothschild (1990) introduced the FGARCH model as follows:

$$H_t = \Omega + \sum_{k=1}^{K} w_k w_k' f_{k,t}, \tag{2.15}$$

where Ω is an $N \times N$ positive semi-definite matrix, $w_k, k = 1, ..., K$ are linearly independent $N \times 1$ vectors of factor weights, and the $f_{k,t}$'s are the K factors. It is assumed that these factors have a first-order GARCH structure:

$$f_{kt} = w_k + \alpha_k(\gamma_k' \varepsilon_{t-1})^2 + \beta_k f_{k,t-1},$$

where w_k, α_k and β_k are scalars and γ_k is an $N \times 1$ vector of weights. Engle, Ng and Rothschild (1990) discussed a consistent, but not efficient, two-step maximum likelihood estimation strategy. Several other versions of the factor MGARCH model are proposed in the literature; for example, the FFMGARCH model (Vrontos, Dellaportas and Politis, 2003), the OGARCH (Kariya, 1988 and Alexander and Chibumba, 1997), the GOGARCH (van der Weide, 2002).

2.4 Models of Conditional Variances and Correlations

Early MGARCH models, some of which were briefly discussed above, attempted to specify the dynamics of all elements of the covariance matrix using a large number of parameters in a more flexible way. As mentioned earlier, the problem with this approach that the models suffer from the "curse of dimensionality". The conditional correlation models are developed to be more parsimonious compared to these models. The correlation models are based on the decomposition of the conditional covariance matrix H_t into conditional standard deviations and correlations in the following way

$$H_t = D_t \Gamma_t D_t, \tag{2.16}$$

$$D_t = diag(h_{1t}^{1/2}, ..., h_{Nt}^{1/2}), \tag{2.17}$$

where $\Gamma_t = [\rho_{ijt}]$ is a symmetric positive definite matrix with $\rho_{iit} = 1$, $i = 1, ..., N$; i.e., Γ_t is the conditional correlation matrix with the structure

$$\Gamma_t = \begin{bmatrix} 1 & \rho_{12t} & \cdots & \rho_{1Nt} \\ \rho_{21t} & 1 & \rho_{23t} & \cdots \\ & & \cdots & \\ \rho_{N,1t} & \cdots & \rho_{N,N-1t} & 1 \end{bmatrix},$$

and h_{it} can be defined as any univariate GARCH model, e.g., a GARCH(p, q) specification $h_{it} = \eta_i' s_{i,t-1}$ where $s_{i,t-1} = \left(1, \varepsilon_{i,t-1}^2, ..., \varepsilon_{i,t-q}^2, h_{i,t-1}, .., h_{i,t-p}\right)'$ and $\eta_i = \left(\alpha_{i0}, \alpha_{i1}, ..., \alpha_{iq}, \beta_{i1}, .., \beta_{ip}\right)'$. H_t is positive definite if and only if the conditional correlation matrix Γ_t is positive definite and all the N conditional variances are positive. (2.16) implies that the diagonal elements of the conditional covariance matrix are simply the conditional variances while the off-diagonal elements are defined as follows:

$$h_{ijt} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ijt}, \quad i \neq j,$$

where $1 \leq i, j \leq N$. The models under this class, in general, estimate the conditional variance matrix in a hierarchical way which we shall term as the *Partial Quasi Maximum Likelihood (PQML)* approach. First, one chooses and estimates a univariate GARCH type model for each conditional variance. Second, the conditional correlation matrix is estimated based on the estimated conditional variances at the first stage (see Section 2.5 for more details). Here we will discuss first various Constant Conditional Correlation (CCC) models. Next we focus on the Dynamic Conditional Correlation (DCC) models.

2.4.1 The Constant Correlation Models

A natural scale invariant measure of coherence between y_{it} and y_{jt} evaluated at time $t-1$ is given by the conditional correlation $\rho_{ijt} = h_{ijt}/\sqrt{(h_{it}h_{jt})}$, where $-1 \leq \rho_{ijt} \leq 1$, a.s., for all t. In general, ρ_{ijt} is time varying or dynamic as H_t is dynamic. However, Bollerslev (1990) argued that in some applications the conditional covariances (i.e., the off-diagonal elements of H_t denoted by $h_{ijt}, \quad i \neq j$) might be considered as proportional to the square root of the product of the corresponding two conditional variances (namely, h_{it} and h_{jt}) leaving the conditional correlations constant through time; i.e.,

$$h_{ijt} = \rho_{ij}\sqrt{h_{it}h_{jt}} \quad j = 1, ..., N, \quad i = j+1, ..., N. \tag{2.18}$$

The CCC models use the same decomposition given in (2.16) but the conditional correlation matrix is constant (Γ); i.e., time invariant. However, this model allows time varying conditional variances and covariances. Indeed, the fact that assumption (2.18) should be tested is an issue which will be discussed in more details later (see Chapter 3).

Hence for the CCC model, we can write

$$H_t = D_t \Gamma D_t, \tag{2.19}$$

where $D_t = diag\left(h_{it}^{1/2}\right)$, $\Gamma = \left\{\rho_{ij}\right\}$, and pd with $\rho_{ii} \equiv 1$. There are, therefore, $\frac{1}{2}N(N-1)$ free parameters in Γ. Because of this structure, we can write $\Gamma = \Lambda\Lambda'$, via Cholesky Decomposition,where $\Lambda = \{\lambda_{ij}\}$ is a lower triangular $(N \times N)$ matrix with $\lambda_{11} = 1$, and $\Lambda^{-1} = \left\{\lambda^{ij}\right\}$ is lower triangular with $\lambda^{11} = 1$, but also having $\frac{1}{2}N(N-1)$ free parameters.

Example 2.1 *For example, in the bivariate case* $\Gamma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, *then*

$$\Lambda = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}, \ \Lambda^{-1} = \begin{bmatrix} 1 & 0 \\ -\dfrac{\rho}{\sqrt{1-\rho^2}} & \dfrac{1}{\sqrt{1-\rho^2}} \end{bmatrix},$$

and $\rho_{12} \equiv \rho$. *In general, note that since* $\Lambda\Lambda^{-1} = I_N$, $\sum_{j=1}^{N} \lambda_{ij}\lambda^{ji} = 1$ *but because* $\lambda_{ij} = 0$, $j \geq i$, *and* $\lambda^{ji} = 0$, $i \geq j$, *this implies that* $\lambda_{ii}\lambda^{ii} = 1$, $i = 1, ..., N$, *with* $\lambda_{11} = \lambda^{11} = 1$.

In conditional correlation MGARCH models standardized errors play a crucial role both in estimation and post-estimation diagnostic testing of the models. Apart from the i.i.d. error ξ_t appeared in (2.5) which we will term as "fully" standardized errors, two other types of standardized errors have emerged in the MGARCH literature: one is standardized error (ζ_t) and the other is the Tse's transformed version of standardized errors (ε_t^*). The formal definitions of these are given below:

Definition 2.1 *The* $N \times 1$ *vector of the standardized errors are defined as*

$$\zeta_t = (\zeta_{1t}, ..., \zeta_{Nt})' = D_t^{-1}\varepsilon_t, \tag{2.20}$$

satisfying $\mathrm{E}[\zeta_t|\mathcal{F}_{t-1}] = 0$, $\mathrm{E}[\zeta_t\zeta_t'|\mathcal{F}_{t-1}] = \Gamma_t$ *(in general) and* $\mathrm{E}[\zeta_t\zeta_t'|\mathcal{F}_{t-1}] = \Gamma$ *(in case of CCC model). In the finance literature this is also known as volatility adjusted return.*

Definition 2.2 *The "fully" standardized errors are defined as the* $N \times 1$ *vector*

$$\xi_t = \Lambda^{-1}\zeta_t = \Lambda^{-1}D_t^{-1}\varepsilon_t = H_t^{-1/2}\varepsilon_t, \tag{2.21}$$

so that $\mathrm{E}[\xi_t\xi_t'|\mathcal{F}_{t-1}] = I_N$.

Definition 2.3 *The (Tse's) transformed standardized errors are defined as*

$$\varepsilon_t^* = (\varepsilon_{1t}^*, ..., \varepsilon_{Nt}^*)' = \Gamma^{-1}\zeta_t = \Gamma^{-1}D_t^{-1}\varepsilon_t, \tag{2.22}$$

satisfying $\mathrm{E}[\varepsilon_t^*|\mathcal{F}_{t-1}] = 0$, $\mathrm{E}[\varepsilon_t^*\varepsilon_t^{*\prime}|\mathcal{F}_{t-1}] = \Gamma_t^{-1}$ *(in general) and* $\mathrm{E}[\varepsilon_t^*\varepsilon_t^{*\prime}|\mathcal{F}_{t-1}] = \Gamma^{-1}$ *(in case of CCC model). Note that* $\varepsilon_{it}^* = \left\{\sum_{m=1}^{N} \rho^{im}\zeta_{mt}\right\}$, *where* $\Gamma^{-1} = \{\rho^{ij}\}$ $i, j = 1, \cdots, N$.

Ling and McAleer (2003) provided an explanation of the CCC hypothesis in the following way. Suppose that h_{it} captures completely the past information, with $\mathrm{E}(h_{it}) = E(\varepsilon_{it}^2)$. Then the standardized residuals $\zeta_{it} = h_{it}^{-1/2}\varepsilon_{it}$

will be independent of the past information and for each i, $\{\zeta_{it},\ t = 0, \pm 1, \cdots\}$ will be a sequence of i.i.d. random variables, with $\mathrm{E}(\zeta_{it}) = 0$ and $Var(\zeta_{it}) = \mathrm{E}(\zeta_{it}^2) = 1$. In general ζ_{it} and ζ_{jt} are correlated for $i \neq j$. Hence, it is natural to assume that $\zeta_t = (\zeta_{it}, \cdots \zeta_{it})'$ is a sequence of i.i.d. random vectors, with zero mean and covariance equals to the correlation matrix, Γ.

The CCC models are computationally very attractive since estimation and inference procedure is much simpler due to the assumption (2.18) which allows the decomposition (2.19). Although the log-likelihood function is still highly nonlinear in parameters, as always the case with MGARCH models, while optimizing the log-likelihood function through iterative methods one has to invert the conditional correlation matrix Γ only once in each iteration. In addition to the univariate GARCH model applied for modelling individual conditional variance, the number of parameters in the CCC GARCH models equals $N(N-1)/2$ which is considerably lesser compared to the VEC and BEKK type models.

Besides, the correlation matrix can be concentrated out from the log-likelihood function, resulting in a reduction in the number of parameters to be optimized. Moreover, it is relatively easy to control the parameters of the conditional variance to ensure the positivity of h_{it} during the optimization whereas controlling a matrix of parameters to be positive definite during optimization is quite difficult. A detailed discussion on the QMLE framework for the CCC models will follow later in Section 2.5.3.

The Bollerslev's CCC Model

The CCC GARCH model proposed by Bollerslev (1990) is the simplest and earliest version of the multivariate correlation model which is nested in the other more general conditional correlation model. The original CCC model has a GARCH(1,1) specification for each conditional variance in D_t which is the most common in the literature; i.e.,

$$h_{it} = \alpha_{i0} + \alpha_{i1}\varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}, \quad i = 1, ..., N. \tag{2.23}$$

However, one can use the GARCH(p, q) specification. Note that conditional variances in (2.23) is positive if all elements of η_i are positive. However, positivity of η_i is not a necessary condition for Γ to be positive definite unless $p = q = 1$.[6] In a multivariate regression framework, this model can be viewed as an extension of Seemingly Unrelated Regressions (SUR) allowing for conditional and/or unconditional heteroskedasticity. Bollerslev (1990) reparameterize the conditional variance which is only unique up to scale in the following way,

$$h_{it} = \alpha_{i0}\sigma_{it}^2, \ \ i = 1, \cdots, N, \tag{2.24}$$

where α_{i0} is a positive time invariant scalar and $\sigma_{it}^2 > 0$ a.s., for all t. In this model, given (2.18) and (2.24) the conditional covariance matrix, H_t is decomposed as:

$$H_t = D_t^* \Gamma^* D_t^*,$$

where $D_t^* = diag(\sigma_{1t}, ..., \sigma_{Nt})$ and $\Gamma^* = \{\gamma_{ij}\} = \{\rho_{ij}\sqrt{\alpha_{i0}\alpha_{j0}}\}$ $i, j = 1, ..., N$, is an $N \times N$ time invariant matrix. That is,

$$H_t = D_t A^{-1}(A\Gamma A)A^{-1}D_t,$$

where $A = diag\left(\alpha_{i0}^{1/2}\right)$, $D_t^* = D_t A^{-1}$, and $\Gamma^* = A\Gamma A$. The subsequent implication of this is that, in terms of the new parameters in Γ^*, the diagonal elements $\rho_{ii}^* = |\alpha_{i0}| = \alpha_{i0} > 0$ rather than 1.

To illustrate Bollerslev's reparameterization with a GARCH(p, q) specification for individual variance equation, note that:

$$\begin{aligned} h_{it} &= \alpha_{i0} + \sum_{k=1}^{q} \alpha_{ik}\varepsilon_{i,t-k}^2 + \sum_{m=1}^{p} \beta_{im} h_{i,t-m} \\ &= \alpha_{i0}\left(1 + \sum_{k=1}^{q} \alpha_{ik}^*\varepsilon_{i,t-k}^2 + \sum_{m=1}^{p} \beta_{im}\sigma_{i,t-m}^2\right) \\ &= \alpha_{i0}\sigma_{it}^2, \quad i = 1, ..., N, \end{aligned} \tag{2.25}$$

where $\alpha_{ik}^* = \alpha_{ik}/\alpha_{i0}$ and $\sigma_{it}^2 = 1 + \sum_{k=1}^{q} \alpha_{ik}^*\varepsilon_{i,t-k}^2 + \sum_{m=1}^{p} \beta_{im}\sigma_{i,t-m}^2$.

[6] See Nelson and Cao (1992) for a discussion of the positivity condition for h_{it} in univariate GARCH(p, q) models.

The Extended CCC (ECCC) Model

Note that in case of the CCC-GARCH(p, q) specification the conditional variances can be written in a vector form

$$\begin{aligned} h_t &= (h_{1t}, \cdots, h_{Nt})' \\ &= \alpha + \sum_{k=1}^{q} A_k \varepsilon_{t-k} \odot \varepsilon_{t-k} + \sum_{m=1}^{p} B_m h_{t-m}, \end{aligned}$$

where α is $N \times 1$ vector, A_k and B_k are diagonal $N \times N$ matrices. The diagonality of A_k and B_m implies that this specification does not allow feedback from other series while estimating individual conditional correlation. Jeantheau (1998) proposed an extension of the CCC model by relaxing this diagonality assumption by allowing the past squared errors and variances of all series to enter the individual conditional variance equation. For example, the ith variance equation of the first order ECCC GARCH model is:

$$h_{it} = a_{i0} + a_{11}\varepsilon_{1,t-1}^2 + ... + a_{1N}\varepsilon_{N,t-1}^2 + b_{11}h_{1,t-1} + ... + b_{1N}h_{N,t-1}, \ i = 1, ..., N. \tag{2.26}$$

The i-th conditional variances for the (p, q) order ECCC model is thus given by

$$h_{it} = \sum_{j=1}^{N} \eta'_{ji} s_{ji,t-1}, \tag{2.27}$$

where $s_{i,t-1}$ and η_i is defined before. This model accommodates a richer autocorrelation structure than the standard CCC GARCH model. He and Teräsvirta (2004) also used a VEC-type formulation for $h_t = (h_{1t}, ..., h_{Nt})'$ to allow for interactions between the conditional variances.

The VARMA-GARCH and VARMA-AGARCH Model

To accommodate interdependence among conditional variances h_{it}, Ling and McAleer (2003) proposed VARMA-GARCH model such that h_{it} contains some past information not only from ε_{it} but also from ε_{jt}. More specifically, they proposed a vector autoregressive moving average (VARMA) specification for the mean function and the following specification for conditional variance:

$$h_t = C + \sum_{i=1}^{r} D_i \overrightarrow{\varepsilon}_{t-i} + \sum_{j=1}^{s} E_j h_{t-j}, \tag{2.28}$$

where $h_t = (h_{1t}, \cdots, h_{Nt})'$, $\overrightarrow{\varepsilon}_t = (\varepsilon_{1t}^2, \cdots, \varepsilon_{Nt}^2)'$, and C, D_i for $i = 1, \cdots, r$ and E_j for $j = 1, \cdots, s$ are $(N \times N)$ parameter matrices. It is to be mentioned that this model still assume a constant conditional correlation matrix Γ and utilize the decomposition (2.19) in the likelihood function.

However, this VARMA GARCH does not allow asymmetric effect; i.e., assumes negative and positive shocks have identical impacts on volatility. To incorporate this asymmetric effect McAleer, Hoti and Chan (2009) proposed the extension VARMA-AGARCH as follows:

$$h_t = C + \sum_{i=1}^{r} D_i \overrightarrow{\varepsilon}_{t-i} + \sum_{j=1}^{s} E_j h_{t-j} + \sum_{i=1}^{r} F_i I_{t-i} \overrightarrow{\varepsilon}_{t-i}, \tag{2.29}$$

where F_i for $i = 1, \cdots, r$ are $(N \times N)$ parameter matrices and $I_t = diag\,(I_{it})$, where

$$I_{it} = \begin{cases} 0, & \varepsilon_{it} > 0 \\ 1, & \varepsilon_{it} < 0 \end{cases}.$$

Note that if $N = 1$, (2.29) reduces to the GJR (or asymmetric) model and with $F_i = 0$ and with D_i and E_j are diagonal for all i, j then the VARMA-AGARCH collapses to the CCC-GARCH model.

2.4.2 The Dynamic (or Time Varying) Conditional Correlation Models

Although the CCC-GARCH model is very attractive with a comparatively simple parameterization, the assumption of constant correlations is not always supported by the data and empirical studies often find that this CCC assumption is quite restrictive (see, for example, Engle and Sheppard, 2001; Tse, 2000; Tsui and Yu, 1999). Therefore, by relaxing the CCC model a number of specifications have been proposed in the literature that make the conditional correlation matrix time dependent. This class of models is known as the DCC model and apply the decomposition as given in (2.16). Almost all DCC models use standardized residuals (obtained from first step estimation) as input to estimate the correlation parameter (in the second stage). Clearly, the correlation matrix is same as the covariance matrix of the standardized residuals ζ_t; i.e., $Var_{t-1}(\zeta_t) = Var_{t-1}\left(D_t^{-1}\varepsilon_t\right) = \Gamma_t$. The difficulty of these specifications is that the dynamic conditional correlation matrix has to be positive definite for all t. Hence the search for DCC models focuses on ensuring the positive definiteness of the conditional correlation matrix (hence that of H_t) with simple and less restrictive conditions on parameters.

The Christodoulakis and Satchell's Specification

Christodoulakis and Satchell (2002) proposed a DCC model for bivariate system which uses the Fisher transformation of the correlation coefficient. This model use the transformation:

$$\rho_{12,t} = \frac{\exp(2r_t) - 1}{\exp(2r_t) + 1}, \tag{2.30}$$

where r_t can be estimated by using any univariate GARCH model using $\varepsilon_{1t}\varepsilon_{2t}/\sqrt{h_{1t}h_{2t}}$. This transformation ensures the positive definiteness of the correlation matrix; however, this is not a genuine multivariate model. The more general multivariate DCC models were first proposed by Engle (2002) and Tse and Tsui (2002) to model high-dimensional dataset. Next we will discuss these specifications.

The VC-GARCH Model of Tse and Tsui

Due to the intuitive interpretation of correlations, a number of specifications have been proposed in the literature. The varying correlation GARCH (VC-GARCH) model of Tse and Tsui (2002) imposes GARCH type dynamics on the conditional correlations. The conditional correlations in their model are functions of the conditional correlation of the previous period and a set of estimated correlations:

$$\Gamma_t = (1-a-b)\Gamma + a\Gamma_{t-1} + b\Psi_{t-1}, \tag{2.31}$$

where a and b are nonnegative scalar parameters satisfying $a+b \leq 1$, Γ is constant (time invariant) $N \times N$ symmetric positive definite matrix with ones on the diagonal. Ψ_{t-1} is the $N \times N$ sample correlation matrix of past M standardized residuals ζ_τ defined in (2.20), for $\tau = t-M, t-M+1, ..., t-1$ and its i,j-th element is given by:

$$\psi_{ij,t-1} = \frac{\sum_{m=1}^{M} \zeta_{i,t-m}\zeta_{j,t-m}}{\left(\sum_{m=1}^{M} \zeta_{i,t-m}^2\right)^{1/2} \left(\sum_{m=1}^{M} \zeta_{j,t-m}^2\right)^{1/2}}. \tag{2.32}$$

The matrix Ψ_{t-1} can be expressed as:

$$\Psi_{t-1} = K_{t-1}^{-1} L_{t-1} L_{t-1}' K_{t-1}^{-1}, \tag{2.33}$$

where $K_{t-1} = diag(\sqrt{\sum_{m=1}^{M} \zeta_{1,t-m}^2}, ..., \sqrt{\sum_{m=1}^{M} \zeta_{N,t-m}^2})$ is a $N \times N$ diagonal matrix and $L_{t-1} = (\zeta_{t-1}, ..., \zeta_{t-M})$ is a $N \times M$ matrix with $\zeta_t = (\zeta_{1t}, ..., \zeta_{Nt})'$.

The positive definiteness of Γ_t is ensured by construction if Γ_0 and Ψ_{t-1} are positive definite. A necessary condition for the latter to hold is $M \geq N$. Note that when $M = 1$, Ψ_{t-1} is equal to a matrix of ones. When positive definiteness of $\Psi_{t-1}$is ensured, then it is a well defined correlation matrix. Tse and Tsui (2002) employed $M = N$ in their paper for computational purpose. It is to be noted here that the definition of the intercept, $(1-a-b)$ in (2.31), corresponds to the idea of "variance targeting" of Engle and Mezrich (1996).

The Engle's DCC Model

An alternative specification proposed by Engle (2002) is known as DCC model. In this model, the correlation matrix Γ_t is specified as:

$$\begin{aligned}\Gamma_t &= (I \odot Q_t)^{-1/2} Q_t (I \odot Q_t)^{-1/2} \\ &= diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \\ &= diag(q_{11,t}^{-1/2} ... q_{NN,t}^{-1/2}) Q_t diag(q_{11,t}^{-1/2} ... q_{NN,t}^{-1/2}),\end{aligned} \tag{2.34}$$

where $Q_t = (q_{ij,t})$ is a $N \times N$ positive definite matrix which is defined either by the integrated model

$$Q_t = (1-\lambda)\zeta_{t-1}\zeta_{t-1}' + \lambda Q_{t-1}, \tag{2.35}$$

or by the mean reverting model

$$Q_t = \widetilde{\Omega} + \alpha\zeta_{t-1}\zeta_{t-1}' + \beta Q_{t-1}, \tag{2.36}$$

where $\zeta_{t-1}$is the standardized residuals given by (2.20).

Each specification has its own advantages and disadvantages. Both (2.35) and (2.36) generate positive definite Q matrices, and therefore positive definite Γ_t, if the initial condition is positive definite and the matrix intercept $\widetilde{\Omega}$ in (2.36) is positive definite. The integrated model assumes all changes in the correlations to be permanent which may be a reasonable assumption in some cases. However, it implies that asymptotically correlations go to $+1$ or -1 which can be verified by simulating (2.20), (2.34) and (2.35). This may not be a satisfactory representation of the data (Engle, 2009). Nevertheless, it may be a good filter in the sense of Nelson and Foster (1994). On the other hand, the mean reverting model assumes that all changes in correlations are transitory. It is to be, however, noted that if $(\alpha + \beta)$ is close to one in (2.36), the changes in correlations can last for quite a long time. The downside of mean reverting specification is very large number of parameters, i.e., $(N(N-1)/2+2)$, as compared to only one parameter of the integrated specification. The number of parameters increase with the dimension of system.

To cope with this problem, another set of estimating equations is introduced. These equations are moment conditions that can be used with the First Order Condition (FOC) of the likelihood function. Letting the sample (unconditional) correlation of standardized residuals (ζ_t) be $\widetilde{\Gamma}$:

$$\widetilde{\Gamma} = \frac{1}{T}\sum_{t=1}^{T} \zeta_t \zeta_t', \tag{2.37}$$

another relationship among the parameters can be found:

$$\overline{Q} = \frac{1}{T}\sum_{t=1}^{T} Q_t \cong \widetilde{\Omega} + \alpha\widetilde{\Gamma} + \beta\overline{Q}. \tag{2.38}$$

Finally, with the additional assumption that on average, $\overline{Q} = \widetilde{\Gamma}$, the intercept can be written as:

$$\widetilde{\Omega} = (1 - \alpha - \beta)\widetilde{\Gamma}, \tag{2.39}$$

and the equation (2.36) can be expressed as:

$$\begin{aligned} Q_t &= \widetilde{\Gamma} + \alpha(\zeta_{t-1}\zeta_{t-1}' - \widetilde{\Gamma}) + \beta(Q_{t-1} - \widetilde{\Gamma}) \\ &= (1 - \alpha - \beta)\widetilde{\Gamma} + \alpha\zeta_{t-1}\zeta_{t-1}' + \beta Q_{t-1}, \end{aligned} \tag{2.40}$$

where α and β are nonnegative scalar parameters and $\alpha + \beta < 1$. Note that, if $\alpha + \beta = 1$, (2.40) collapses to (2.35). Hence, equation (2.40) has only two parameters irrespective of the size of the system and Q_t is mean reverting to the average correlation as long as $\alpha + \beta < 1$. The assumption (2.39), a generalization of the idea of "variance targeting" of Engle and Mezrich (1996), is known as "correlation targeting" which restricts the unconditional correlation implied by the model to be equal to the unconditional sample correlation of standardized residuals. The estimator of the parameter $\widetilde{\Omega}$ is different from the MLE and asymptotically inefficient although it may be relatively robust to some form of misspecification.

Remark 2.4 *The primary structural difference between the VC-GARCH and DCC-GARCH is that the VC-GARCH use recursion (2.32) to obtain time varying conditional correlation matrix while the DCC-GARCH standardized Q_t applying (2.34) to obtain the dynamic conditional correlation matrix. To illustrate the difference between Tse and Tsui (2002)'s VC-GARCH model and Engle's (2002) DCC-GARCH model, we write the expression of correlation coefficient in the bivariate case. For the VC-GARCH model,*

$$\rho_{12,t} = (1-a-b)\rho_{12} + a\rho_{12,t-1} + b\frac{\sum_{m=1}^{M}\zeta_{1,t-m}\zeta_{2,t-m}}{\left(\sum_{m=1}^{M}\zeta_{1,t-m}^2\right)^{1/2}\left(\sum_{m=1}^{M}\zeta_{2,t-m}^2\right)^{1/2}}, \quad (2.41)$$

whereas for the DCC-GARCH model it can be written as:

$$\rho_{12,t} = \frac{(1-\alpha-\beta)\overline{q}_{12} + \alpha\zeta_{1,t-1}\zeta_{2,t-1} + \beta q_{12,t-1}}{\sqrt{\left((1-\alpha-\beta)\overline{q}_{11} + \alpha\zeta_{1,t-1}^2 + \beta q_{11,t-1}\right)\left((1-\alpha-\beta)\overline{q}_{22} + \alpha\zeta_{2,t-1}^2 + \beta q_{22,t-1}\right)}}. \quad (2.42)$$

One can see that the VC-GARCH model formulates the conditional correlation as a weighted sum of past correlations, whereas in the DCC-GARCH model the matrix Q_t is expressed as a GARCH $(1,1)$-type equation and then transformed to a correlation matrix.[7] *However, for both models, the restrictions $a = b = 0$ or $\alpha = \beta = 0$ can be tested to examine the empirical validity of imposing the CCC assumption.*

The DECO DCC Model

A new covariance matrix model, dynamic equicorrelation (DECO) is proposed by Engle and Kelly (2008) under the assumption that at every time period all pairwise correlations are equal. This assumption, which is prag-

[7] It is possible to use other univariate GARCH or even complex multivariate GARCH specification to model Q_t. For example, Engle (2002) proposed another extension using the multivariate GARCH family of Ding and Engle (2001) which can be expressed in first order form as:

$$Q_t = \widetilde{\Gamma} \odot (\iota_N \iota_N' - A - B) + A \odot \overline{\varepsilon}_{t-1}\overline{\varepsilon}_{t-1}' + B \odot Q_{t-1},$$

where ι_N' is $(1 \times N)$ vector of ones, and A and B are $N \times N$ matrices of parameter. This, however, increases the number of parameters considerably.

matically applied in various areas of finance, makes it possible to estimate arbitrarily large covariance matrices with ease. The DECO is a special case of the CCC and DCC models which involve first adjusting for individual volatilities and then estimating the correlations. Following Engle and Kelly (2008), a matrix $\overleftrightarrow{\Gamma}_t$ is an equicorrelation matrix of an $N \times\ 1$ vector of random variables if it is positive definite and takes the form

$$\overleftrightarrow{\Gamma}_t = (1 - \rho_t)\, I_N + \rho_t \mathcal{J}_N, \tag{2.43}$$

where ρ_t is the equicorrelation. The inverse and determinant of $\overleftrightarrow{\Gamma}_t$ are given by:

$$\overleftrightarrow{\Gamma}_t^{-1} = \frac{1}{(1-\rho_t)} \left[I_N - \frac{\rho_t}{1+(N-1)\rho_t} \mathcal{J}_N \right], \tag{2.44}$$

and

$$\left| \overleftrightarrow{\Gamma}_t \right| = (1-\rho_t)^{N-1} [1 + (N-1)\rho_t]. \tag{2.45}$$

Now, a $N \times 1$ vector $\{y_t\}$ satisfies a DECO model if $Var_{t-1}(y_t) = D_t \overleftrightarrow{\Gamma}_t D_t$, where D_t is a diagonal matrix with conditional standard deviations on the diagonal and $\overleftrightarrow{\Gamma}_t$ is given by Equation (2.43) for all t. The equicorrelation coefficient is a general, potentially time-varying, function. The DECO is adopted to individual applications by specifying a ρ_t process and specifying a conditional volatility model. Engle and Kelly (2008) considered the DCC model of Engle (2002) as the basic ρ_t specification and subsequently referred as the DECO-DCC. They specify the equicorrelation ρ_t as in the DECO-DCC model as the pairwise average of off-diagonal elements of Γ_t as defined in the Engle's (2002) DCC model in (2.34):

$$\rho_t = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}, \tag{2.46}$$

where $q_{ij,t}$ is the i, j-th element of Q_t defined either by (2.40)or (2.35). If Q_t is mean reverting (integrated) then the eqicorrelation will also be mean reverting (integrated). Also the correlation matrices generated by every realization of a DECO process are positive definite and invertible. See Engle (2009b), for a detailed review and discussion on various DCC model specifications and estimation strategies.

A Few Other DCC Models

A new era of flexible modeling in the variance part has begun with the DCC models and numerous studies, both theoretical and empirical, have emerged in the literature. A general drawback of the DCC models is that the parameters (a and b in case of the VC-GARCH or α and β in case of the DCC-GARCH) are scalar, therefore all the conditional correlations obey the same dynamics. There may be many situations where this restriction may be violated. For example, correlations between stocks from the same industry may behave rather differently from correlations between stocks from different industries. Another drawback is the symmetry imposed by these models in the sense that a pair of positive standardized residuals ζ_{it} and ζ_{jt} has the same effect on the conditional correlation as a pair of negative standardized residuals of same magnitude. Empirical studies found evidence of the presence of asymmetries in correlations (see, for example, Ang and Chen, 2002). To avoid these limitations, various generalizations of the DCC model have been proposed.

The Quadratic Flexible DCC (QFDCC) Model: For example, Billio and Caporin (2009) suggest quadratic flexible DCC GARCH model by imposing a BEKK structure on the conditional correlations where the matrix process Q_t is defined as:

$$Q_t = C'\widetilde{\Gamma}C + A'\zeta_{t-1}\zeta'_{t-1}A + B'Q_{t-1}B, \tag{2.47}$$

where the parameter matrices A, B and C are symmetric, $\widetilde{\Gamma}$ is the unconditional covariance matrix of standardized residuals defined in (2.37). To obtain stationarity, $C'\widetilde{\Gamma}C$ has to be positive definite and the eigenvalues of $A+B$ has to be less than one in modulus. However, the number of parameters governing the correlation is $3N(N+1)/2$ which is unfeasible in large system. Several special cases proposed by the authors; for example, restricting the coefficient matrices to be diagonal with possible partition, group the variables according to some criterion and the restricting the coefficient matrices to be block diagonal.

The Semi Generalized DCC (SGDCC) Model: Hafner and Franses (2003) suggested the Semi Generalized DCC (SGDCC) model that allows for asset specific news impact parameters. Here the matrix process Q_t is expressed as:

$$Q_t = (1 - \overline{\alpha}^2 - \beta)\widetilde{\Gamma} + \alpha\alpha' \odot \zeta_{t-1}\zeta'_{t-1} + \beta Q_{t-1}, \tag{2.48}$$

where $\alpha = (\alpha_1, \ldots, \alpha_N)$ is now $N \times 1$ parameter vector and $\overline{\alpha} = N^{-1}\sum_{i=1}^{N}\alpha_i$ is the average news impact parameter. The memory parameter β can be asset specific. Again note the number of parameters in (2.48) increases linearly with N.

The Asymmetric Generalized DCC (AGDCC) Model: Asymmetric correlation models were introduced by Cappoello, Engle and Sheppard (2007) where the correlations become larger when two returns are both negative than they are equally positive and all other factors remain constant. In their Asymmetric Generalized DCC (AGDCC) model, the Q_t has been formulated as:

$$Q_t = (\widetilde{\Gamma} - A'\widetilde{\Gamma}A - B'\widetilde{\Gamma}B - C'\overrightarrow{\Gamma}C) + A'\zeta_{t-1}\zeta'_{t-1}A + B'Q_{t-1}B + C'\overrightarrow{\varepsilon}_{t-1}\overrightarrow{\varepsilon}'_{t-1}C), \tag{2.49}$$

where A, B and C are $N \times N$ parameter matrices, $\overrightarrow{\varepsilon}_t = \tilde{I}_{\{\zeta_t<0\}} \odot \zeta_t$, where $\tilde{I}$ is an indicator function and $\widetilde{\Gamma}$ and $\overrightarrow{\Gamma}$ are the unconditional covariance matrix of ζ_t and $\overrightarrow{\varepsilon}_t$. The number of parameters again increases rapidly with the dimension and various restrictions (e.g., diagonal, scalar and symmetric versions), were proposed.

2.4.3 The Regime Switching and Smooth Transition Models

All of the above DCC specifications use standardized residuals ζ_t as input to model the dynamics of conditional correlations. There is another class of models that allows the dynamic structure of conditional correlations to be controlled by exogenous variable which may be either a single observable, or combination of several observable variables or a latent variable that represents factors that are difficult to quantify. For example, in the Smooth Transition CC (STCC) GARCH model of Silvennoinen and Teräsvirta (2005) Γ_t varies smoothly between two extreme states according to a transition variable:

$$\Gamma_t = (1 - G(s_t))\Gamma_{(1)} + G(s_t)\Gamma_{(2)}, \tag{2.50}$$

where $\Gamma_{(1)}$, $\Gamma_{(2)}$ are positive definite correlation matrices describing the two extreme states of correlation and $\Gamma_{(1)} \neq \Gamma_{(2)}$, and $G(.) \to (0, 1)$ is a monotonic function of an observable transition variable $s_t \in F_{t-1}$. Silvennoinen and Teräsvirta (2005) proposed $G(.)$ as a logistic function:

$$G(s_t) = \frac{1}{1 + \exp(-\gamma(s_t - c))}, \quad \gamma > 0, \tag{2.51}$$

where the parameter γ determines the velocity and c is the location parameter. In addition to univariate variance equations, this model has $N(N-1)+2$ additional parameters. The positive definiteness of $\Gamma_{(1)}$ and $\Gamma_{(2)}$ ensure the positive definiteness of Γ_t as the latter is a convex combination of $\Gamma_{(1)}$ and $\Gamma_{(2)}$. An important component of the STCC model is the transition variable s_t. Berben and Jansen (2005) take $s_t = t$ which effectively model a gradual structural change in correlation. Hafner, Dijk and Franses (2005) choose transition variable as lagged market return and volatility motivated by the "correlation breakdown" effect and by the presence of factor structure, respectively.[8]

[8] Literature on correlation breakdown focused on the fact that typically asset correlations increase in bear markets, but are not affected (or to a much lesser extent) in bull markets. See, for example, Longin and Solink (2001), Ang and Chen (2002), Butler and Joaquin (2002) among others.

Silvennoinen and Teräsvirta (2009a) extended their STCC GARCH model by allowing for another transition around the first one which they call Double Smooth transition CC (DSTCC) GARCH model. In this model, Γ_t is specified as:

$$\begin{aligned}\Gamma_t &= (1 - G_2(s_{2t}))\{(1 - G_1(s_{1t}))\Gamma_{(11)} + G_1(s_{1t})\Gamma_{(21)}\} \\ &\quad + G_2(s_{2t})\{(1 - G_1(s_{1t}))\Gamma_{(12)} + G_1(s_{1t})\Gamma_{(22)}\}. \end{aligned} \tag{2.52}$$

Pelletier (2006) proposed the Regime Switching Dynamic Correlation (RSDC) GARCH model which falls somewhere between the CCC GARCH model and the models for which correlation changes continuously at every period. This model imposes constant correlation within a regime while the dynamics enter through switching regimes:

$$\Gamma_t = \sum_{r=1}^{R} \tilde{I}_{\{\triangle_t = r\}} \Gamma_{(r)}, \tag{2.53}$$

where $\triangle_t$ is usually a first order Markov Chain independent of ξ_t, that can take R possible values and governed by a transition probability matrix Π, $\tilde{I}$ is the indicator function and $\Gamma_{(r)},\ r = 1, \cdots, R$ are positive definite regime-specific correlation matrices and $\Gamma_{(r)} \neq \Gamma_{(r')}$ for $r \neq r'$. Hence the correlation component of this specification has $RN(N-1)/2 - R(R-1)$ parameters. To reduce the parameters, a version is proposed by restricting R possible states of correlations to be linear combination of a state of zero correlation and that of possibly high correlation:

$$\Gamma_t = (1 - \lambda(\Delta_t))I_N + \lambda(\Delta_t)\Gamma, \tag{2.54}$$

where I_N is the identity matrix implying no correlation, Γ is a correlation matrix representing the state of possibly high correlation and $\lambda(.) : \{1, \cdots, R\} \rightarrow [0, 1]$ is a monotonic function of Δ_t. The correlation matrix at time t is a weighted average of two extreme states of the world. In one state, the returns are uncorrelated ($\lambda(\Delta_t) = 0$) and in the other the returns are (highly) correlated ($\lambda(\Delta_t) = 1$). We then have regimes of generally higher or lower correlations and the changes across correlations in a given regime are proportional. The number of regime R is not a parameter to be estimated. One advantage of the RSDC-GARCH model is that the conditional correlation matrices are positive definite at each point in time by construction both in restricted and unrestricted version of the model. Pelletier (2006) proposed two stage estimation: in the first stage the parameters in the GARCH equation are estimated while in the second step, conditioning on these estimates of the first stage, the correlations and switching probabilities are estimated using the Expectation Maximization (EM) algorithm of Dempster, Laird and Rubin (1977).

2.5 Parametric Estimation of MGARCH Models

2.5.1 A General Set-up and the QMLE

Without further restrictions, and following the path similar to that taken by Bollerslev and Wooldridge (1992), along with the linear conditional mean specification in (2.4), we assume the following parameterization of H_t

$$H_t = H_t(W_t; \varphi, \eta) = H_t(\theta),$$

where $\theta' = (\varphi', \eta')$, $\varphi' = (\varphi_1', ..., \varphi_N')$, $\eta' = (\eta_1', ..., \eta_N')$. For example, the GARCH$(p, q)$ specification with $N = 1$ yields $h_t = \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 + \sum_{m=1}^{p} \beta_j h_{t-p}$, so that $\eta' = \left(\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p\right)$ and φ enters through the ε_{t-k}^2 terms.

Define $\varepsilon_t \equiv \varepsilon_t(\varphi) = y_t - W_t'\varphi$ and $H_t \equiv H_t(\theta)$. The quasi-conditional log-likelihood, for observation t, is (ignoring constants)

$$l_t(\theta) = -\frac{1}{2} \ln |H_t| - \frac{1}{2} \varepsilon_t{}' H_t{}^{-1} \varepsilon_t = \frac{1}{2} \ln |H_t| - \frac{1}{2} \operatorname{tr}\left(H_t^{-1} \varepsilon_t \varepsilon_t'\right), \tag{2.55}$$

where $|A| = \det A$; i.e., estimation proceeds "as if" the errors are conditionally Gaussian. Taking differentials (Magnus and Neudecker, 1999, Chps. 5

and 8), where dH_t and $d\varepsilon_t$ are the differentials of H_t and ε_t, respectively, we obtain[9]

$$\begin{aligned}
\mathrm{d}\, l_t(\theta) &= -\frac{1}{2}\operatorname{tr}\left(H_t^{-1}\,\mathrm{d}\, H_t\right) - \varepsilon_t' H_t^{-1}\,\mathrm{d}\,\varepsilon_t + \frac{1}{2}\operatorname{tr}\left(H_t^{-1}\,\mathrm{d}\, H_t H_t^{-1}\varepsilon_t\varepsilon_t'\right) \\
&= -\frac{1}{2}\operatorname{tr}\left(H_t^{-1}\,\mathrm{d}\, H_t\right) - \varepsilon_t' H_t^{-1}\,\mathrm{d}\,\varepsilon_t + \frac{1}{2}\operatorname{tr}\left(H_t^{-1}\varepsilon_t\varepsilon_t' H_t^{-1}\,\mathrm{d}\, H_t\right) \\
&= \frac{1}{2}\left\{\operatorname{vec}\left(H_t^{-1}\varepsilon_t\varepsilon_t' H_t^{-1}\right) - \operatorname{vec}\left(H_t^{-1}\right)\right\}'\mathrm{d}\operatorname{vec} H_t - \varepsilon_t' H_t^{-1}\,\mathrm{d}\,\varepsilon_t \\
&= \frac{1}{2}\operatorname{vec}\left(H_t^{-1/2}\left\{\xi_t\xi_t' - I_N\right\}H_t^{-1/2}\right)'\mathrm{d}\operatorname{vec} H_t \\
&\quad -\xi_t' H_t^{-1/2}\,\mathrm{d}\,\varepsilon_t, \qquad (2.56)
\end{aligned}$$

where $\xi_t = H_t^{-1/2}\varepsilon_t$ is the "fully" standardized residuals and we are assuming H_t is positive definite.

Using (2.56), we can differentiate (2.55), with respect to φ and then η, to obtain the conditional score vectors

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \varphi'} &= \frac{1}{2}\operatorname{vec}\left(H_t^{-1/2}\left\{\xi_t\xi_t' - I_N\right\}H_t^{-1/2}\right)'\frac{\partial \operatorname{vec} H_t}{\partial \varphi'} - \xi_t' H_t^{-1/2}\frac{\partial \varepsilon_t}{\partial \varphi'}. \\
\frac{\partial l_t(\theta)}{\partial \eta'} &= \frac{1}{2}\operatorname{vec}\left(H_t^{-1/2}\left\{\xi_t\xi_t' - I_N\right\}H_t^{-1/2}\right)'\frac{\partial \operatorname{vec} H_t}{\partial \eta'}.
\end{aligned}$$

The above expressions of scores make use of the $\operatorname{vec}(.)$ operator and do not take into account the fact of the symmetry of H_t. Making use of the following Remark we obtain alternative expressions for scores which use $\operatorname{vech}()$ operator.

Remark 2.5 *For non-singular matrices, A and B, $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$ and for any conformable matrices, A, B, C, $\operatorname{vec}(ABC) = (C'\otimes A)\operatorname{vec}(B)$; see Magnus and Neudecker (1999, p. 28 and 30). For a $(n\times n)$ symmetric matrix, A, there exists a unique $\left(n^2\times \frac{1}{2}n\,(n+1)\right)$ duplication matrix, D, with full column rank, such that $D\operatorname{vech}(A) = \operatorname{vec}(A)$, or $\operatorname{vech}(A) = D^{+}\operatorname{vec}(A)$, $A = A'$, where $D^{+} = (D'D)^{-1}D'$, $\left(\frac{1}{2}n\,(n+1)\times n^2\right)$ is the Moore-Penrose inverse of D; see Magnus and Neudecker (1999, p. 49).*

[9] Using $d\ln|A| = tr\left(A^{-1}dA\right)$, $dA^{-1} = -A^{-1}(dA)A^{-1}$ and $tr(A'B) = vec(A)'vec(B)$.

Noting the above *Remark*, the fact that $\dfrac{\partial \varepsilon_t}{\partial \varphi'} = -W_t'$, H_t is symmetric and writing $Z_t' = H_t^{-1/2} W_t'$, $\nabla_\varphi H_t = \left\{ \dfrac{\partial \operatorname{vech} H_t}{\partial \varphi'} \right\}'$, and $\nabla_\eta H_t = \left\{ \dfrac{\partial \operatorname{vech} H_t}{\partial \eta'} \right\}'$ yields

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \varphi'} &= \frac{1}{2} \operatorname{vec}\left(\xi_t \xi_t' - I_N\right)' \left(H_t \otimes H_t\right)^{-1/2} \left(\nabla_\varphi H_t\right)' + \xi_t' Z_t' \\
&= \frac{1}{2} \operatorname{vech}\left(\xi_t \xi_t' - I_N\right)' \left\{ D' \left(H_t \otimes H_t\right)^{-1/2} D \right\} \left(\nabla_\varphi H_t\right)' + \xi_t' Z_t'. \\
\frac{\partial l_t(\theta)}{\partial \eta'} &= \frac{1}{2} vech\left(\xi_t \xi_t' - I_N\right)' \left\{ D' \left(H_t \otimes H_t\right)^{-1/2} D \right\} \left(\nabla_\eta H_t\right)'. \qquad (2.57)
\end{aligned}$$

Special Case: The VEC GARCH Model

In (2.9), $c = \operatorname{vech}(C) = D^+ \operatorname{vec}(C)$, which is $\left(\frac{1}{2} N (N+1) \times 1\right)$ vector and

$$\begin{aligned}
\operatorname{vech}\left(A_k^* \left(\varepsilon_{t-k} \varepsilon_{t-k}'\right) A_k^{*\prime}\right) &= D^+ \left(A_k^* \otimes A_k^*\right) \operatorname{vec}\left(\varepsilon_{t-k} \varepsilon_{t-k}'\right) \\
&= D^+ \left(A_k^* \otimes A_k^*\right) D \operatorname{vech}\left(\varepsilon_{t-k} \varepsilon_{t-k}'\right),
\end{aligned}$$

so that $A_k = D^+ \left(A_k^* \otimes A_k^*\right) D$ and, similarly, $B_m = D^+ \left(B_m^* \otimes B_m^*\right) D$, both $\left(\frac{1}{2} N (N+1) \times \frac{1}{2} N (N+1)\right)$. In this case

$$\frac{\partial \operatorname{vech} H_t}{\partial \varphi'} = \sum_{k=1}^{q} A_k \frac{\partial \operatorname{vech}\left(\varepsilon_{t-k} \varepsilon_{t-k}'\right)}{\partial \varphi'} + \sum_{m=1}^{p} B_m \frac{\partial \operatorname{vech}(H_{t-m})}{\partial \varphi'}.$$

Taking differentials, $\mathrm{d}\left(\varepsilon_{t-k} \varepsilon_{t-k}'\right) = \left(\mathrm{d}\, \varepsilon_{t-k}\right) \varepsilon_{t-k}' + \varepsilon_{t-k} \left(\mathrm{d}\, \varepsilon_{t-k}'\right)$, so $\mathrm{d} \operatorname{vec}\left(\varepsilon_{t-k} \varepsilon_{t-k}'\right) = \left(\varepsilon_{t-k} \otimes I_N\right) \mathrm{d}\, \varepsilon_{t-k} + \left(I_N \otimes \varepsilon_{t-k}\right) \mathrm{d}\, \varepsilon_{t-k}$. Thus,

$$\begin{aligned}
\frac{\partial \operatorname{vech}\left(\varepsilon_{t-k} \varepsilon_{t-k}'\right)}{\partial \varphi'} &= D^+ \left\{ \left(\varepsilon_{t-k} \otimes I_N\right) + \left(I_N \otimes \varepsilon_{t-k}\right) \right\} \frac{\partial \varepsilon_{t-k}}{\partial \varphi'} \\
&= -D^+ \left\{ \left(\varepsilon_{t-k} \otimes I_N\right) + \left(I_N \otimes \varepsilon_{t-k}\right) \right\} W_{t-k}'.
\end{aligned}$$

and from (2.10)

$$\frac{\partial h_{rs,t}}{\partial \varphi_i'} = -2 \sum_{k=1}^{q} \alpha_{iik} \varepsilon_{i,t-k} w_{i,t-k}' - \sum_{k=1}^{q} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \alpha_{ijk} \varepsilon_{j,t-k} w_{i,t-k}' + \sum_{m=1}^{p} \sum_{l=1}^{N} \sum_{j=l}^{N} \beta_{ljm} \frac{\partial h_{lj,t-m}}{\partial \varphi_i}.$$

In the case of DVEC GARCH model the estimation becomes much easier as each equation can be treated separately with

$$\begin{aligned}
\frac{\partial h_{it}}{\partial \varphi_i'} &= -2\sum_{k=1}^{q}\alpha_{iik}\varepsilon_{i,t-k}w_{i,t-k}' + \sum_{m=1}^{p}\beta_{iim}\frac{\partial h_{ii,t-m}}{\partial \varphi_i'}, \quad i=1,...,N, \\
\frac{\partial h_{ijt}}{\partial \varphi_i'} &= -\sum_{k=1}^{q}\alpha_{ijk}\varepsilon_{j,t-k}w_{i,t-k}' + \sum_{m=1}^{p}\beta_{ijm}\frac{\partial h_{ij,t-m}}{\partial \varphi_i'}, \quad j\le i=1,...,N, \\
\frac{\partial h_{ijt}}{\partial \varphi_j'} &= -\sum_{k=1}^{q}\alpha_{ijk}\varepsilon_{i,t-k}w_{j,t-k}' + \sum_{m=1}^{p}\beta_{ijm}\frac{\partial h_{ij,t-m}}{\partial \varphi_j'}, \quad j\le i=1,...,N.
\end{aligned}$$

2.5.2 The Conditional Correlation Model: Two stage Estimation

Here we provide the two-stage likelihood estimation framework, which we will term as the Partial QMLE (PQMLE), of the dynamic conditional correlation models. To start with, the statistical specification of a general DCC model can be given as:

$$\begin{aligned}
& \varepsilon_t|\mathcal{F}_{t-1} \sim N(0,H_t) \quad or \quad \varepsilon_t|\mathcal{F}_{t-1} \sim N(0,D_t\Gamma_t D_t), \\
\zeta_t &= {D_t}^{-1}\varepsilon_t, \\
D_t &= diag(h_{1t}^{1/2},.....,h_{Nt}^{1/2}) \text{ so that each } h_{it} \text{ follows a GARCH process}, \\
\Gamma_t &= [\rho_{ij,t}] \sim \text{symmetric PD with } \rho_{ii,t}=1,\ i=1,...,N.
\end{aligned} \tag{2.58}$$

Let $\theta' = (\theta_1', \theta_2')$ where θ_1 and θ_2 denote the parameters involved in D_t and Γ_t, respectively. Assuming conditional normality (without this assumption, the estimator still have the QML interpretation), the corresponding log-likelihood function can be written as (ignoring constants):

$$\begin{aligned}
\sum_{t=1}^{T} l_t(\theta) &= -\frac{1}{2}\sum_{t=1}^{T} \ln|D_t\Gamma_t D_t| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t'(D_t\Gamma_t D_t)^{-1}\varepsilon_t \\
&= -\frac{1}{2}\sum_{t=1}^{T} \ln|\Gamma_t| - \sum_{t=1}^{T} \ln|D_t| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t' D_t^{-1}\Gamma_t^{-1} D_t{}^{-1}\varepsilon_t \\
&= -\frac{1}{2}\left(\sum_{t=1}^{T} \ln|\Gamma_t| - \sum_{t=1}^{T}\sum_{i=1}^{N} \ln|h_{it}| - \sum_{t=1}^{T} \zeta_t'\Gamma_t^{-1}\zeta_t\right) \\
&= -(1/2)\sum_{t=1}^{T}\left(\begin{array}{c} \sum_{i=1}^{N} \ln|h_{it}| + \varepsilon_t' D_t^{-1} D_t{}^{-1}\varepsilon_t \\ -\zeta_t'\zeta_t + \ln|\Gamma_t| + \sum_{t=1}^{T} \zeta_t'\Gamma_t^{-1}\zeta_t \end{array}\right). \qquad (2.59)
\end{aligned}$$

This log likelihood must be maximized over all parameters involved in (2.59). However, it could be very difficult for large systems. Therefore a two stage estimation procedure has been proposed by Engle (2002). The resulting estimators are inefficient but consistent. Following the arguments of Newey and McFadden (1994), sufficient conditions for consistency and asymptotic normality of these estimators are also provided in Engle (2002).[10] Then (2.59) can be written as the sum of a volatility part and a correlation part:

$$L(\theta_1, \theta_2) = L_V(\theta_1) + L_C(\theta_1, \theta_2), \qquad (2.60)$$

where

$$L_V(\theta_1) = -(1/2)\sum_{t=1}^{T}\left(\sum_{i=1}^{N} \ln|h_{it}| + \varepsilon_t' D_t^{-1} D_t{}^{-1}\varepsilon_t\right) \qquad (2.61)$$

and

$$L_C(\theta_1, \theta_2) = -\,(1/2)\sum_{t=1}^{T}\left(-\zeta_t'\zeta_t + \ln|\Gamma_t| + \sum_{t=1}^{T} \zeta_t'\Gamma_t^{-1}\zeta_t\right). \qquad (2.62)$$

Note that (2.61) is the sum of individual GARCH likelihoods which is jointly maximized by separately maximizing each equation. For example, if univariate GARCH (1,1) models are used to estimate conditional volatilities

[10] See Engle and Sheppard (2001) for more details of the argument.

h_{it} then θ_1 is of dimension $3N \times 1$. For estimating the correlation parameters θ_2, (2.62) is used and since the squared standardized residuals do not depend on θ_2 it can be ignored while optimization. Therefore, the two step approach of maximizing (2.59) involves finding the consistent estimates of θ_1 in first step

$$\widehat{\theta}_1 = \arg\max_{\theta_1}[L_V(\theta_1)], \tag{2.63}$$

then employing this estimate as given in second stage and using the estimated standardized residuals $(\hat{\zeta}_t)$ as input:

$$\widehat{\theta}_2 = \arg\max_{\theta_2}[L_C(\widehat{\theta}_1, \theta_2)]. \tag{2.64}$$

To reiterate, the decomposition in (2.60) is inefficient but still yields valid martingale difference based moment constraints. The above two stage estimation provides an attractive and simple solution of estimating correlation matrices in the DCC framework. However, there are still some problems. For example, inversion of the $N \times N$ correlation matrix is required for evaluation of (2.62) for each observation. For large N, it could be a daunting task due to numerical issues, e.g., problematic as convergence is not guaranteed and is sensitive to starting values. Secondly, Engle and Sheppard (2001) showed by simulations that in the correctly specified Engle's (2002) DCC models, there is a downward bias in the estimation of α when N is large. In other words, the correlations are estimated to be smoother and less variable when a large number of assets are considered than when a small number of assets are considered.

2.5.3 The Full QMLE of the CCC Model

The CCC model can indeed be estimated by the above two-stage estimation algorithm. In practice applied researchers mostly use this method, see, for example, Engle and Sheppard (2008), Hafner, Dijk and Franses (2005), Billio, Caporin and Gobbo (2006) among others. However, the CCC hypothesis affords a considerable simplification of the model specification and one can estimate the model by maximizing the full likelihood function in one step (we shall refer this as the Full QMLE (FQMLE)). We first discuss the FQMLE framework based on the usual decomposition (2.19). This will be followed by the discussion in the context of Bollerslev's reparameterization $H_t = D_t^* \Gamma^* D_t^*$ (as discussed in Section 2.4.1).

With $H_t = D_t \Gamma D_t$, it is assumed that:

1. $h_{ij,t} = E\left[\varepsilon_{it}\varepsilon_{jt}|F_{t-1}\right] = \rho_{ij}\sqrt{h_{ii,t}}\sqrt{h_{jj,t}}$, so that $corr\left[\varepsilon_{it}, \varepsilon_{jt}\right] = \rho_{ij}$, a constant with $\rho_{ii} \equiv 1$.

2. $h_{ii,t} = \alpha_{i0} + \sum_{k=1}^{q} \alpha_{ik}\varepsilon_{i,t-k}^2 + \sum_{m=1}^{p} \beta_{im} h_{ii,t-m}$.

3. $\Gamma = \left\{\rho_{ij}\right\}$, is symmetric and pd with $\rho_{ii} \equiv 1$. There are, therefore, $\frac{1}{2}N\left(N-1\right)$ free parameters in Γ. Hence we can write $\Gamma = \Lambda\Lambda'$, via Cholesky Decomposition,where $\Lambda = \left\{\lambda_{ij}\right\}$ is a lower triangular $(N \times N)$ matrix with $\lambda_{11} = 1$, and $\Lambda^{-1} = \left\{\lambda^{ij}\right\}$ is lower triangular with $\lambda^{11} = 1$, but also having $\frac{1}{2}N\left(N-1\right)$ free parameters.

Express the parameter vector $\varphi' = \left(\varphi_1', ..., \varphi_N'\right)$, $\eta' = \left(\eta_1', ..., \eta_N'\right)$ and the distinct elements of $\Gamma = \left\{\rho_{ij}\right\}$, denoted by ρ. It may also be useful to define $\theta' = \left(\theta_1', ..., \theta_N'\right)$, with $\theta_i' = \left(\varphi_i', \eta_i'\right)$ and $\varpi = \left(\theta', \rho'\right)'$. Under the assumption of conditional normality, the quasi-conditional log-likelihood per observation, t, (ignoring any constant terms) is, via (2.55),

$$
\begin{aligned}
l_t^* \equiv l_t^*(\varpi) &= -\frac{1}{2}\ln\left|H_t\right| - \frac{1}{2}\varepsilon_t' H_t^{-1}\varepsilon_t = -\frac{1}{2}\ln\left|H_t\right| - \frac{1}{2}tr\left(H_t^{-1}\varepsilon_t\varepsilon_t'\right) \\
&= -\frac{1}{2}\ln\left|\Gamma\right| - \frac{1}{2}\sum_{j=1}^{N}\ln h_{jt} - \frac{1}{2}\zeta_t'\Gamma^{-1}\zeta_t \\
&= -\frac{1}{2}\ln\left|\Lambda\Lambda'\right| - \frac{1}{2}\sum_{j=1}^{N}\ln h_{jt} - \frac{1}{2}\zeta_t'\Lambda'^{-1}\Lambda^{-1}\zeta_t \\
&= \ln\left|\Lambda^{-1}\right| - \frac{1}{2}\sum_{j=1}^{N}\ln h_{jt} - \frac{1}{2}\sum_{j=1}^{N}\xi_{jt}^2,
\end{aligned}
\tag{2.65}
$$

where, now, $H_t \equiv H_t\left(\varpi\right)$, $D_t \equiv D_t\left(\theta\right)$.

Remark 2.6 *Here we have made use of* $|AB| = |A|\,|B|$, $|A'| = |A|$ *and* $|A|^{-1} = |A^{-1}|$, *so* $-\dfrac{1}{2}\ln|\Lambda\Lambda'| = -\dfrac{1}{2}\ln|\Lambda|^2 = -\ln|\Lambda| = \ln|\Lambda|^{-1}$.

Scores of the CCC Model Based on $H_t = D_t\Gamma D_t$

Let the average score of the CCC model is $G_T^*(\varpi) = T^{-1}\sum_{t=1}^{T} g_t^*(\varpi)$ where $g_t^*(\varpi) = \left(\dfrac{\partial l_t^*}{\partial\theta'}, \dfrac{\partial l_t^*}{\partial\rho'}\right)' = \left(\dfrac{\partial l_t^*}{\partial\varphi'}, \dfrac{\partial l_t^*}{\partial\eta'}, \dfrac{\partial l_t^*}{\partial\rho'}\right)'$. For the i^{th} variable, define $\underset{(T\times K)}{F_i}$, $\underset{(T\times K)}{C_i}$ and $\underset{(T\times K')}{X_i}$ with rows $f'_{it} = \dfrac{w'_{it}}{\sqrt{h_{it}}}$; $c'_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial\varphi'_i}$ and $x'_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial\eta'_i}$, respectively. Then define

$$
\begin{aligned}
\underset{(NT\times NK)}{F} &= diag\,(F_i)\,,\\
\underset{(N\times NK)}{F'_t} &= diag\,(f'_{it}) \text{ for } t = 1,\cdots,T.
\end{aligned}
$$

In a similar way, define C, X, C'_t and X'_t. For example, C'_t has the following form:

$$
C'_t = \begin{bmatrix} c'_{1t} & 0 & 0 & 0 \\ 0 & c'_{2t} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & c'_{Nt} \end{bmatrix}.
$$

Also, define $\underset{(N\times N)}{E_t} = diag\,(\zeta_{it})$; and the $(N\times T)$ matrices $E = \{\zeta_{it}\}$ and $E^* = \{\varepsilon_{it}^*\} = \Gamma^{-1}E$ having columns ζ_t and ε_t^*, respectively.

Lemma 2.1 *The score vector for observation* t *of (2.65),* $g_t^*(\varpi) = \left(\dfrac{\partial l_t^*}{\partial\theta'}, \dfrac{\partial l_t^*}{\partial\rho'}\right)' = \left(\dfrac{\partial l_t^*}{\partial\varphi'}, \dfrac{\partial l_t^*}{\partial\eta'}, \dfrac{\partial l_t^*}{\partial\rho'}\right)'$ *can be expressed as:*

$$
\begin{aligned}
\frac{\partial l_t^*}{\partial\varphi} &= F_t\Gamma^{-1}\zeta_t + \frac{1}{2}C_t\left\{E_t\Gamma^{-1}\zeta_t - \iota_N\right\} = F_t\varepsilon_t^* + \frac{1}{2}C_t\left\{E_t\varepsilon_t^* - \iota_N\right\},\\
\frac{\partial l_t^*}{\partial\eta} &= \frac{1}{2}X_t\left\{E_t\Gamma^{-1}\zeta_t - \iota_N\right\} = \frac{1}{2}X_t\left\{E_t\varepsilon_t^* - \iota_N\right\},\\
\frac{\partial l_t^*}{\partial\rho_{ij}} &= vecl\,(M_t) = m_{ij,t},\;\; j < i = 2,...,N \;\; (with\ \rho_{ii} \equiv 1), \qquad (2.66)
\end{aligned}
$$

where , E_t, F_t, C_t and X_t defined earlier and $M_t = \{m_{ij,t}\} = \Gamma^{-1}(\zeta_t\zeta_t' - \Gamma)\Gamma^{-1}$.

Remark 2.7 *Recall that* $\hat{\varepsilon}^*_{it} = \sum_{j=1}^{N} \widehat{\rho}^{ij}\widehat{\zeta}_{ijt}$, *then the corresponding score equations (first order conditions) are:*

$$\begin{aligned}
\sum_{t=1}^{T} \hat{f}_{it}\hat{\varepsilon}^*_{it} + \frac{1}{2}\sum_{t=1}^{T} \hat{c}_{it}\left\{\hat{\zeta}_{it}\hat{\varepsilon}^*_{it} - 1\right\} &= 0, i = 1, ..., N. \\
\frac{1}{2}\sum_{t=1}^{T} \hat{x}_{it}\left\{\hat{\zeta}_{it}\hat{\varepsilon}^*_{it} - 1\right\} &= 0, i = 1, ..., N. \\
\sum_{t=1}^{T} \left(\hat{\varepsilon}^*_{it}\hat{\varepsilon}^*_{jt} - \hat{\rho}^{ij}\right) &= 0, \ .j < i = 2, ..., N.
\end{aligned}$$

In matrix notation the average score function $G^*_T(\varpi)$ *can be written as:*

$$\begin{aligned}
G^*_T(\varphi) &= T^{-1}\left[F'\left(\Gamma^{-1} \otimes I_T\right)\zeta + \frac{1}{2}C'\left(\widetilde{e} - \iota_{NT}\right)\right]. \\
G^*_T(\eta) &= \frac{1}{2}T^{-1}X'\left(\widetilde{e} - \iota_{NT}\right). \\
G^*_T(\rho) &= T^{-1}vecl\left(E^*E^{*\prime} - T\Gamma^{-1}\right). \qquad (2.67)
\end{aligned}$$

where $\underset{(NT\times 1)}{\widetilde{e}} = vec(\widetilde{E}')$, $\widetilde{E} = E^* \odot E = \{\varepsilon^*_{it}\zeta_{it}\}$, *and* F, C, X, E *and* E^* *are defined earlier.*

Remark 2.8 *Now, because* $\rho_{ii} = 1$, *the first order condition (for* ρ*) then requires that* $\sum_{t=1}^{T} \hat{M}^L_t = 0$ *where* M^L_t *signifies the strictly lower triangle elements of* M_t *That is to say,* $\sum_{t=1}^{T}\left(\hat{\varepsilon}^*_{it}\hat{\varepsilon}^*_{jt} - \hat{\rho}^{ij}\right) = 0$, $j < i$. *But this is not the same as* $\sum_{t=1}^{T}\left(\hat{\zeta}_{it}\hat{\zeta}_{jt} - \hat{\rho}_{ij}\right) = 0$. *For example, consider the bivariate case*

$$\varepsilon^*_t = \begin{bmatrix} \varepsilon^*_{1t} \\ \varepsilon^*_{2t} \end{bmatrix} = \Gamma^{-1}\zeta_t = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} \zeta_{1t} - \rho\zeta_{2t} \\ \zeta_{2t} - \rho\zeta_{1t} \end{bmatrix},$$

so that we require

$$\sum_{t=1}^{T}\left(\hat{\varepsilon}^*_{1t}\hat{\varepsilon}^*_{2t} + \frac{\hat{\rho}}{1-\hat{\rho}^2}\right) = 0.$$

Note, however, that:

$$\begin{aligned}\varepsilon_{1t}^{*}\varepsilon_{2t}^{*}+\frac{\rho}{1-\rho^{2}} &= \frac{1}{\left(1-\rho^{2}\right)^{2}}\left[\left\{\zeta_{1t}-\rho\zeta_{2t}\right\}\left\{\zeta_{2t}-\rho\zeta_{1t}\right\}+\hat{\rho}\left(1-\hat{\rho}^{2}\right)\right] \\ &= \frac{1}{\left(1-\rho^{2}\right)^{2}}\left[\left(1+\rho^{2}\right)\left(\zeta_{1t}\zeta_{2t}-\rho\right)-\rho\left(\zeta_{1t}^{2}-1\right)-\rho\left(\zeta_{2t}^{2}-1\right)\right].\end{aligned}$$

So that $\sum_{t=1}^{T}\left(\hat{\zeta}_{1t}\hat{\zeta}_{2t}-\hat{\rho}\right)=0$ *does not guarantee that* $\sum_{t=1}^{T}\left(\hat{\varepsilon}_{1t}^{*}\hat{\varepsilon}_{2t}^{*}+\frac{\hat{\rho}}{1-\hat{\rho}^{2}}\right)=0$, *since* $\sum_{t=1}^{T}\left(\hat{\zeta}_{it}^{2}-1\right)=0$ *is not guaranteed in finite samples, by the first order conditions, although* $\text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\left(\hat{\zeta}_{it}^{2}-1\right)=0$, *under correct model specification.*

A way around this, through a reparameterization, is discussed next.

Scores with Bollerslev's Reparameterization (Based on $H_t = D_t^{*}\Gamma^{*}D_t^{*}$)

As discussed in section 2.4, Bollerlsev (1990) notes that, given $\alpha_{i0} > 0$, we can write the separate GARCH(p,q) specifications as (2.25); i.e., $h_{it} = \alpha_{i0}+\sum_{k=1}^{q}\alpha_{ik}\varepsilon_{i,t-k}^{2}+\sum_{m=1}^{p}\beta_{im}h_{i,t-m} = \alpha_{i0}\sigma_{it}^{2}$. Given this and the decomposition discussed earlier, the quasi-conditional log-likelihood per observation, t, (ignoring any constant terms) is

$$\begin{aligned}l_t^{*} &\equiv l_t^{*}(\theta^{*},\rho^{*}) = -\frac{1}{2}\ln|H_t|-\frac{1}{2}\varepsilon_t' H_t^{-1}\varepsilon_t \\ &= -\frac{1}{2}\ln|G_t^{*}\Gamma^{*}G_t^{*}|-\frac{1}{2}\varepsilon_t' G_t^{*-1}\Gamma^{*-1}G_t^{*-1}\varepsilon_t \\ &= -\frac{1}{2}\ln|\Gamma^{*}|-\frac{1}{2}\sum_{j=1}^{N}\ln\sigma_{jt}^{2}-\frac{1}{2}\zeta_t^{*\prime}\Gamma^{*-1}\zeta_t^{*},\end{aligned}$$

where $\zeta_t^{*} = (\zeta_{1t}^{*},...,\zeta_{Nt}^{*})' = D_t^{*^{-1}}\varepsilon_t = \left\{\frac{\varepsilon_{it}}{\sqrt{\sigma_{it}^{2}}}\right\}$ is a $N\times 1$ vector which is different from the standardized residuals defined in (2.20) or "fully" standardized residuals defined in (2.21) and $\sigma_{it}^{2} = 1+\sum_{k=1}^{q}\alpha_{ik}^{*}\varepsilon_{i,t-k}^{2}+\sum_{m=1}^{p}\beta_{im}\sigma_{i,t-m}^{2}$, $\alpha_{ik}^{*} = \frac{\alpha_{ik}}{\alpha_{i0}}$ and the new parameter is $\theta_i^{*\prime} = \left(\varphi_i', \eta_i^{*\prime}\right)$, $i = 1,...,N$, is as before but with α_{i0} removed from η_i^{*} with the latter's remaining parameters being made up of α_{ik}^{*} and β_{im}, $k = 1,...,q$, $m = 1,...,p$. The estimate of the parameter α_{i0} will be retrieved from estimating $\rho_{ii}^{*} = \alpha_{i0}$.

Corollary 2.1 *The score for φ, η^{*} are given by*

$$\begin{aligned}\frac{\partial l_t}{\partial \varphi} &= W_t D_t^{*} \Gamma^{*-1} \zeta_t^{*} + \frac{1}{2} C_t \left\{ E_t^{*} \Gamma^{*-1} \zeta_t^{*} - \iota_N \right\}, \\ \frac{\partial l_t}{\partial \eta^{*}} &= \frac{1}{2} X_t^{*} \left\{ E_t^{*} \Gamma^{*-1} \zeta_t^{*} - \iota_N \right\},\end{aligned}$$

where $E_t^{*} = diag\left(\zeta_{it}^{*}\right)$ is a $(N \times N)$ matrix, $C_t^{*\prime} = \begin{bmatrix} c_{1t}^{*\prime} & 0 & 0 & 0 \\ 0 & c_{2t}^{*\prime} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & c_{Nt}^{*\prime} \end{bmatrix}$,and $X_t^{*\prime} = \begin{bmatrix} x_{1t}^{*\prime} & 0 & 0 & 0 \\ 0 & x_{2t}^{*\prime} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & x_{Nt}^{*\prime} \end{bmatrix}$ are both $(N \times NK)$ matrix, constructed in a similar way to W_t' in (2.3) with $c_{it}^{*} = \dfrac{1}{\sigma_{it}^{2}} \dfrac{\partial \sigma_{it}^{2}}{\partial \varphi_i'}$ and $x_{it}^{*} = \dfrac{1}{\sigma_{it}^{2}} \dfrac{\partial \sigma_{it}^{2}}{\partial \eta_i^{*}}$.

Remark 2.9 *Defining $\varepsilon_t^{**} = \Gamma^{*-1} \zeta_t^{*} = \Gamma^{*-1} G_t^{*-1} \varepsilon_t = \left\{ \varepsilon_{it}^{**} \right\}$, and $z_{it}^{*} = w_{it} / \sqrt{\sigma_{it}^{2}}$ these scores can be written alternatively as:*

$$\frac{\partial l_t}{\partial \varphi_i} = z_{it}^{*} \varepsilon_{it}^{**} + \frac{1}{2} c_{it}^{*} \left\{ \zeta_{it}^{*} \varepsilon_{it}^{*} - 1 \right\}, \quad i = 1, ..., N. \tag{2.68}$$

The corresponding score equation (first order condition) is

$$\sum_{t=1}^{T} \hat{z}_{it}^{*} \hat{\varepsilon}_{it}^{*} + \frac{1}{2} \sum_{t=1}^{T} \hat{c}_{it} \left\{ \hat{\zeta}_{it} \hat{\varepsilon}_{it}^{*} - 1 \right\} = 0, \ i = 1, ..., N.$$

Similarly the score for η_i^{*} and corresponding FOC is therefore

$$\frac{\partial l_t}{\partial \eta_i^{*}} = \frac{1}{2} x_{it}^{*} \left\{ \zeta_{it}^{*} \varepsilon_{it}^{**} - 1 \right\}, \quad i = 1, ..., N. \tag{2.69}$$

The simplicity of this reparameterization is derived, however, from estimating Γ^{*}, which is now unconstrained (it is, however, still symmetric).

Taking differentials of the likelihood function, we obtain

$$dl_t = \frac{1}{2} tr \left\{ \Gamma^{*-1} \left(\zeta_t^* \zeta_t^{*\prime} - \Gamma^* \right) \Gamma^{*-1} d\Gamma^* \right\},$$

yielding (since the elements ρ_{ij}^*, $i \leq j$, are now unconstrained) the score for ρ_{ij}^* as

$$\frac{\partial l_t}{\partial \rho_{ij}^*} = m_{ij,t}^*, \tag{2.70}$$

where $M_t^* = \Gamma^{*-1} \left(\zeta_t^* \zeta_t^{*\prime} - \Gamma^* \right) \Gamma^{*-1} = \left\{ m_{ij,t}^* \right\} = \left\{ \varepsilon_{it}^{**} \varepsilon_{jt}^{**\prime} - \rho^{*ij} \right\}$with $\Gamma^{*-1} = \left\{ \rho^{*ij} \right\}$.

This implies that the FOC $\sum_t \hat{m}_{ij,t}^* = 0$ yields $\hat{\Gamma}^* = \frac{1}{T} \sum_{t=1}^{t} \hat{\zeta}_t^* \hat{\zeta}_t^{*\prime}$. Notice that $\widehat{\rho}_{ii}^* = \frac{1}{T} \sum_{t=1}^{T} \hat{\zeta}_{it}^{*2}$, and this provides the estimate of α_{i0}, and $\widehat{\rho}_{ij}^* = \frac{1}{T} \sum_{t=1}^{T} \hat{\zeta}_{it}^* \hat{\zeta}_{jt}^*$ $i \neq j$. Since $\rho_{ij} = \rho_{ij}^* / \sqrt{\alpha_{i0}} \sqrt{\alpha_{j0}}$, with $\rho_{ii}^* = \alpha_{i0}$, we can estimate $\hat{\rho}_{ij}$, by the invariance property of MLE, as

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} \hat{\zeta}_{it}^* \hat{\zeta}_{jt}^*}{\sqrt{\sum_{t=1}^{T} \hat{\zeta}_{it}^{*2}} \sqrt{\sum_{t=1}^{T} \hat{\zeta}_{jt}^{*2}}}.$$

From these parameter estimates we can construct

$$\begin{aligned} \hat{h}_{iit} &= \hat{\rho}_{ii}^* \hat{\sigma}_{it}^2 = \hat{\rho}_{ii}^* \left\{ 1 + \sum_{k=1}^{q} \hat{\alpha}_{ik}^* \hat{\varepsilon}_{i,t-k}^2 + \sum_{m=1}^{p} \hat{\beta}_{im} \hat{\sigma}_{i,t-m}^2 \right\}, \\ \hat{\zeta}_{it}^* &= \hat{\varepsilon}_{it} / \sqrt{\hat{\sigma}_{it}^2} = \hat{\rho}_{ii}^{*-1/2} \frac{\hat{\varepsilon}_{it}}{\sqrt{\hat{h}_{it}}} = \hat{\rho}_{ii}^{*-1/2} \hat{\zeta}_{it}, \end{aligned}$$

where ζ_{it} is defined as before. Thus, with this parameterization, since $\hat{\rho}_{ii}^* = \frac{1}{T} \sum_{t=1}^{T} \hat{\zeta}_{it}^{*2}$ this enforces $\frac{1}{T} \sum_{t=1}^{T} \hat{\zeta}_{it}^2 = 1$, an equality which is not enforced in under the original parameterization.

2.5.4 The MacGyver Estimator of the DCC Model

In response to the criticisms associated with two-stage estimation of DCC models, Engle (2009a) introduced a new estimation method which he termed as "MacGyver" method.[11] The MacGyver method is based on bivariate estimation of correlations. It assumes that the selected DCC model is correctly specified between every pair of assets i and j. For example, for the mean reverting DCC model of Engle (2002) as given in (2.40) the correlation process is simply

$$\begin{aligned} \rho_{ij,t} &= \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}}, \\ q_{ij,t} &= (1-\alpha-\beta)\widetilde{\Gamma}_{ij} + \alpha\zeta_{i,t-1}\zeta'_{j,t-1} + \beta q_{ij,t-1}, \end{aligned} \tag{2.71}$$

and the log likelihood function for this pair of assets is extracted from (2.62). It is given by

$$L_{C,ij}(\theta_2) = -(1/2)\left(\ln\left[1-\rho^2_{ij,t}\right] + \frac{\zeta^2_{i,t}+\zeta^2_{j,t}-2\rho_{ij,t}\zeta_{i,t}\zeta_{j,t}}{\left[1-\rho^2_{ij,t}\right]}\right). \tag{2.72}$$

If the high dimension model is correctly specified, so is the bivariate model. The MLE of α, β should be consistent using only data on one pair of assets. However, information that is ignored at the estimation stage could yield more efficient estimates. Engle (2009a) suggested combining parameter estimates from these bivariate models for improving estimation. The combined parameters are then used with equation (2.71) to calculate the correlations.

An analytical solution to optimal combination of bivariate parameter estimates seems extremely difficult as the data are dependent from one pair to another and the dependence is a function of the parameters. Knowing the dependence of the data does not lead easily to measures of the dependence of the parameter estimates. Engle used Monte Carlo experiment to find the optimal functions and found that median has the lowest RMS biases and errors compared to mean and trimmed mean.

In addition to the computational simplification and bias reduction, Engle (2009a) mentioned three other advantages of MacGyver method of estimating

[11] He named this technique after the popular old TV show "MacGyver" which showed MacGyver using whatever was at hand to cleverly solve his problem.

a DCC model. There are $\binom{N}{2}$ pairs of bivariate pairs of correlations. For example, for 50 assets, there are 1225 bivariate pairs; while for 100 assets, there are 4950 asset pairs. Hence the number of bivariate estimations increases as well. However, since only the median of all these estimations is needed as found in his experiment, there is little loss of efficiency if some are not run. This opens the possibility of estimating a subset of the bivariate pairs, as it is clear that there is little advantage to doing all of them. When new assets are added to the collection, it may not be necessary to reestimate at all if the investigator is confident that the specification is adequate. However, it is not clear how to select a good subset. The second advantage is that the data sets for each bivariate pair need not be of the same length. Thus, an asset with only a short history can be added to the system without requiring the shortening of all other series. This is particularly important when examining large asset classes and cross country correlations as there are many assets which are newly issued, merged or otherwise associated with short time histories. A potential third advantage which has not been explored by Engle (2009a) but is mentioned in his paper, is that there may be evidence in these bivariate parameter estimates that the selected DCC model is not correctly specified. Presumably, the bivariate models would show less dispersion if the model is correctly specified than if it is incorrect.

Since the MacGyver is a very new technique, several issues are unsolved and yet to be answered. Some of these are mentioned earlier (e.g., the optimal blend of bivariate estimators, selection of subset of pairs for large system). The construction of standard errors of MacGyver estimators is one such issue which needs further research.

2.5.5 The DECO DCC Estimator

Another recent development in this field of estimating correlation matrix is the DECO DCC estimator of Engle and Kelly (2008). It may often the case that the equicorrelation assumption fails so that there is cross sectional variation in pairwise correlations, as in the DCC. Engle and Kelly (2008) showed that in this case the DECO model remains a powerful tool and the maximum likelihood estimate of DECO is consistent for the DCC parameters when DCC is the true model. In other words, the DECO can continue to give consistent parameter estimates when the equicorrelation assumption is violated. This model applied the two-stage estimation procedure and the volatility-correlation decomposition of the likelihood function as shown in (2.60). In addition this model has a simpler likelihood function than the DCC so that the model is estimable for large cross sections while remaining a consistent (QMLE) estimator of the original DCC model. In the second step, to obtain estimates of the parameters for the ρ_t process, the log-likelihood function in (2.62) becomes (using (2.44) and (2.45)):

$$\begin{aligned} L_C(\theta_2) &= -(1/2)\sum_{t=1}^{T}\left(\ln|\overleftrightarrow{\Gamma}_t| + \sum_{t=1}^{T}\zeta_t'\overleftrightarrow{\Gamma}_t^{-1}\zeta_t\right) \\ &= -(1/2)\left(\begin{array}{c} \ln\left[(1-\rho_t)^{N-1}[1+(N-1)\rho_t]\right] \\ +\frac{1}{(1-\rho_t)}\left[I_N - \frac{\rho_t}{1+(N-1)\rho_t}\mathcal{J}_N\right] \end{array}\right), \end{aligned} \tag{2.73}$$

where ρ_t is governed by (2.46). The advantage from the equicorrelation assumption is the computational ease. Note that only the scalar equicorrelation parameter for each t is recorded, and the compact analytical forms for the determinant and inverse of the covariance matrix under the assumption of equicorrelation make the computational demands for solving the likelihood optimization problem manageable for large cross sections. Whereas in the DCC models, the conditional correlation matrices must be recorded and inverted for all t. Further, these T inversions and determinant calculations are repeated for each of the many iterations required in a numeric optimization program. This is costly for small cross sections and potentially infeasible for very large ones.

2.5.6 The Semiparametric Estimation Approach

All of the above estimation frameworks are parametric, an alternative is the non- and semi-parametric estimation techniques which do not impose any particular structure (possibly misspecified) on the data. There are few advantages of parametric specifications: for example; they offer an interpretation of the dynamic structure of the conditional covariance or correlation matrices, the consistency and asymptotic normality of the QMLE (even under non-normality). However, there may be considerable efficiency losses in finite samples if the returns are not normally distributed which is often the case. Semiparametric models combine the advantages of a parametric model by retaining consistency and interpretability, and those of a nonparametric model which is robust against distributional misspecification.

Hafner, van Dijk, and Franses (2005) put forward a Semi-Parametric Conditional Correlation (SPCC–) GARCH model where the conditional variances are modelled parametrically by any choice of univariate GARCH model, where $\hat{\zeta}_t = \hat{D}_t^{-1}\hat{\varepsilon}_t$ is the vector of the standardized residuals. The conditional correlations Γ_t are then estimated using a transformed Nadaraya-Watson estimator:

$$\Gamma_t = (I \odot Q_t)^{-1/2} \, Q_t \, (I \odot Q_t)^{-1/2} \,,$$

where

$$Q_t = \frac{\sum_{\tau=1}^{T} \hat{\zeta}_\tau \hat{\zeta}_\tau' K_h(x_\tau - x_t)}{\sum_{\tau=1}^{T} K_h(x_\tau - x_t)}, \tag{2.74}$$

and $x_\tau \in F_{t-1}$ is a conditioning (observed) variable, $K_h(.) = \frac{K(./h)}{h}$, $K(\cdot)$ is a kernel function, and h is the bandwidth parameter. It is to be mentioned here that the choice of the kernel function is not important and it could be any probability density function, whereas the choice of the bandwidth parameter h is crucial, see for instance Sections 2.4.2 and 2.7 of Pagan and Ullah (1999). Hafner, van Dijk, and Franses (2005) discuss a way of choosing a dynamic bandwidth parameter such that the bandwidth is larger in the tails of the marginal distribution of the conditioning variable x_τ than it is in the mid-region of the distribution.

2.6 Concluding Remarks

In this Chapter we review some important MGARCH model specifications with specific focus on conditional correlation models. We also discuss the various estimation framework for MGARCH models. The QMLE framework of conditional correlation models, particularly of the CCC model, has been discussed at length. We also provide an explicit expression for the score function of CCC-GARCH regression model which will be useful in the next chapter.

Appendices

2.A Appendix A: GARCH(p,q) model

We consider the GARCH(p,q) model in the regression context. That is, the regression model for the variable of interest, y_t, is defined as:

$$y_t = m\left(w_t;\varphi_0\right) + \varepsilon_{0t}, \qquad t = 1,...,T \tag{2.75}$$

where $w_t = (y'_{t-1}, z'_t)$, $y_{t-1} = (1, y_{t-1,}..., y_{t-l})' \in \Re^{l+1}$, $z_t = (z_{t1,}..., z_{tk})' \in \Re^k$ are exogenous variables, $\varphi_0 = (\varphi_{01}, .., \varphi_{0r})'$ is a vector of unknown parameters and the conditional mean function, $m\left(w_t;\varphi_0\right)$, is possibly non-linear but at least twice continuously differentiable in φ.[12] The error $\{\varepsilon_{0t}, \mathcal{F}_t\}$, where $\mathcal{F}_t$ is a σ-field generated by the current and past errors, is a martingale difference sequence (MDS) given by

$$\varepsilon_{0t} = \xi_t h_{0t}^{1/2}, \tag{2.76}$$

where the standardized error process, ξ_t, is an independent and identically distributed ($i.i.d.$) sequence with mean zero and variance one. Since the ξ_t are $i.i.d.$, they are also conditionally homokurtic, yielding $\mathrm{E}\left[\left(\dfrac{\varepsilon_t^2}{h_t} - 1\right)^2 |\mathcal{F}_{t-1}\right] = k_c - 1$, where k_c is a finite constant. Also define the conditional third moment of ξ_t as $\mathrm{E}\left[\left(\dfrac{\varepsilon_t^3}{h_t^{3/2}}\right)|\mathcal{F}_{t-1}\right] = v_c$, where v_c is a finite constant. The conditional variance follows the GARCH (p,q) process and is specified as

$$\begin{aligned} h_{0t} &= \eta'_0 s_{0,t-1} \qquad (2.77) \\ &= \alpha_{00} + A_0(L)\varepsilon_{0t}^2 + B_0(L)h_{0t} \\ &= \alpha_{00} + \sum_{k=1}^{q} \alpha_{0k}\varepsilon_{0,t-k}^2 + \sum_{j=1}^{p} \beta_{0j} h_{0,t-j}, \end{aligned}$$

[12] For example, Lundbergh and Teräsvirta (1999) proposed the STAR-GARCH model and the statistical properties of this model were investigated by Chan and McAleer (2002).

where $s_{t-1} = \left(1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, h_{t-1}, .., h_{t-p}\right)'$, $\eta = \left(\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, .., \beta_p\right)'$, $A_0(L) = \alpha_{01}L + ... + \alpha_{0q}L^q$, and $B_0(L) = \beta_{01}L + ... + \beta_{0p}L^p$. Note that α_k and β_j represent the ARCH effects (or the short-run persistence) and the GARCH effects (or the contribution of shocks to long-run persistence), respectively.

The preceding process is defined for the true parameter $\theta_0 = (\varphi_0', \eta_0')'$ and correspondingly the model for the unknown parameter vector $\theta = (\varphi', \eta')'$ is defined as

$$\begin{aligned} y_t &= m(w_t; \varphi) + \varepsilon_t, \qquad t = 1, ..., T \\ h_t &= \eta' s_{t-1} \\ &= \alpha_0 + A(L)\varepsilon_t^2 + B(L)h_t \\ &= \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}. \end{aligned} \tag{2.78}$$

The following assumptions are made to ensure the identifiability, stationarity, and ergodicity of the above process.

Assumption 2.A.1 *The parameter space, Θ, is compact, and θ_0 lies in the interior of Θ.*

Assumption 2.A.2 *The elements (y_t, z_t') are strictly stationary and ergodic, and $m(w_t;\varphi)$ is continuous and $\mathcal{F}_{t-1}$ measurable for all $\varphi \in \Theta$.*

Assumption 2.A.3 *All the roots of $1 - A(z) - B(z) = 0$ lie outside the unit circle;*

Assumption 2.A.4 *The parameter space is constrained such that $0 < \lambda \leq \min\{\eta_l\} \leq \max\{\eta_l\} < \lambda^*$, $l = 1, \cdots, p + q + 1$, where λ and λ^* are independent of θ;*

Assumption 2.A.5 *The polynomials $A(z)$ and $1 - B(z)$ are coprimes.*

An alternative representation of the above GARCH(p, q) specification can be given by writing (2.78) as

$$
\begin{aligned}
h_t &= \alpha_0 \left\{ 1 + \frac{1}{\alpha_0} \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 + \frac{1}{\alpha_0} \sum_{j=1}^{p} \beta_j h_{t-j} \right\} \\
&= \alpha_0 \left\{ 1 + \sum_{k=1}^{q} \alpha_k^* \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j \frac{h_{t-j}}{\alpha_0} \right\} \\
&= \alpha_0 \left\{ 1 + \sum_{k=1}^{q} \alpha_k^* \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \right\} \\
&= \alpha_0 \sigma_t^2, \ (say) \qquad (2.79)
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma_t^2 &= 1 + \sum_{k=1}^{q} \alpha_k^* \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 = 1 + \eta^{*\prime} s_{t-1}^*, \\
\eta^{*\prime} &= \left(\alpha_{1,}^* ..., \alpha_q^*, \beta_1, .., \beta_p \right) = \left(\frac{\alpha_{1,}}{\alpha_0} ..., \frac{\alpha_q}{\alpha_0}, \beta_1, .., \beta_p \right), \\
s_{t-1}^{*\prime} &= \left(\varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, \sigma_{t-1}^2, .., \sigma_{t-p}^2 \right). \qquad (2.80)
\end{aligned}
$$

Hence, the new parameter vector is now $\theta^B = (\varphi', \alpha_0, \eta^{*\prime})'$.

2.A.1 GARCH QML Estimation Framework

The (average) quasi log-likelihood (ignoring constants), conditional on available presample values, is

$$
L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = -\frac{1}{2} \left[\ln(h_t) + \frac{\varepsilon_t^2}{h_t} \right]. \qquad (2.81)
$$

Note that the log-likelihood function is not only conditional on available presample values $(y_0, \cdots, y_{1-l})'$, from which $\varepsilon_t, \ t = 1, ..., T$, can be constructed, but also on $\bar{\varepsilon}_0 = (\varepsilon_0, \cdots, \varepsilon_{1-q}, h_0, \cdots, h_{1-p})'$, from which h_t can be constructed using (2.78). However, ε_t and the process h_t, $t \leq 0$ are unobserved. Hence, in practice, for standard inferential procedures a constant value is chosen for $\bar{\varepsilon}_0$ in order to generate $h_t, \ t = 1, \cdots T$:

The Score

Assuming $L_T(\theta)$ is twice continuously differentiable in $\theta = (\varphi', \eta')'$, denote the average score $G_T(\theta) = T^{-1}\sum_{t=1}^{T} g_t(\theta)$ where $g_t(\theta) = \dfrac{\partial l_t}{\partial \theta}$ is given by

$$\begin{aligned} g_t(\theta) &= \frac{\partial l_t}{\partial \theta} = \left(\frac{\partial l_t}{\partial \varphi'}, \frac{\partial l_t}{\partial \eta'}\right)' \\ &= \begin{pmatrix} \frac{\varepsilon_t}{\sqrt{h_t}} f_t + \dfrac{1}{2}\left(\dfrac{\varepsilon_t^2}{h_t} - 1\right) c_t \\ \dfrac{1}{2}\left(\dfrac{\varepsilon_t^2}{h_t} - 1\right) x_t \end{pmatrix}, \end{aligned} \tag{2.82}$$

where $f_t = \frac{1}{\sqrt{h_t}} \dfrac{\partial m(w_t;\varphi)}{\partial \varphi}$ (for linear conditional mean specification $\dfrac{\partial m(w_t;\varphi)}{\partial \varphi} = w_t$), $c_t = \dfrac{1}{h_t}\dfrac{\partial h_t}{\partial \varphi}$, $x_t = \dfrac{1}{h_t}\dfrac{\partial h_t}{\partial \eta}$.

Remark 2.10 *For the reparameterized GARCH(p, q) model as given in (2.79), the log-likelihood becomes*

$$l_t\left(\theta^B\right) = -\frac{1}{2}\left[\ln\left(\alpha_0\sigma_t^2\right) + \frac{\varepsilon_t^2}{\alpha_0\sigma_t^2}\right] = -\frac{1}{2}\left[\ln\left(\alpha_0\right) + \ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\alpha_0\sigma_t^2}\right],$$

and the score for α_0 is given by $\sum \dfrac{\partial l_t\left(\theta^B\right)}{\partial \alpha_0} = \dfrac{1}{2}\alpha_0^{-1} T^{-1}\sum_{t=1}^{T}\left(\dfrac{\varepsilon_t^2}{\alpha_0\sigma_t^2} - 1\right)$.
Notice that the first order condition (FOC) $\sum_{t=1}^{T} \dfrac{\partial l_t\left(\hat{\theta}^B\right)}{\partial \alpha_0} = 0$ *now enforces* $T^{-1}\sum_{t=1}^{T}\left(\dfrac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1\right) = 0$, *where* $\hat{h}_t = \hat{\alpha}_0\hat{\sigma}_t^2$ *and this is not guaranteed by the original parameterization.*

The consistency and asymptotic normality of the QMLE estimator $\hat{\theta} = \arg\max_\theta L_T(\theta)$ are presented in the following sections (for proof see Halunga and Orme (2009, section 2.1)). In addition to the earlier set of assumptions, following Weiss (1986), the existence of moments is assumed when required as follows, where $\| \,.\, \|$denotes the Euclidean norm.

Assumption 2.A.6 $\mathrm{E}\left\|\varepsilon_{0t}\right\|^{4(1+b)} < \infty$ *for some $b > 0$ and for all t.*

Assumption 2.A.7 $\mathrm{E}\left\|m(w_t;\varphi)-m(w_t;\varphi_0)\right\|^2>0$ *for all* $\varphi\neq\varphi_0$.

Assumption 2.A.8 $m(w_t;\varphi_0)$ *is at least twice continuously differentiable in* φ *for all* t.

The Consistency and limit Distribution of the QMLE

Given the above assumptions, $\hat{\theta}\xrightarrow{d}\theta_0$ and $\sqrt{T}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}N\left(0,J_{\theta\theta}^{-1}\Sigma_{\theta\theta}J_{\theta\theta}^{-1}\right)$ where $J_{\theta\theta}=-\mathrm{E}\left[\frac{\partial g_t(\theta_0)}{\partial\theta'}\right]$ and $\Sigma_{\theta\theta}=\mathrm{E}\left[g_t(\theta_0)\,g_t(\theta_0)'\right]$ are both finite and positive definite with

$$J_{\theta\theta}=\begin{bmatrix}J_{\varphi\varphi} & J_{\eta\varphi}'\\ J_{\eta\varphi} & J_{\eta\eta}\end{bmatrix}=\frac{1}{2}\mathrm{E}\begin{bmatrix}c_tc_t' & c_tx_t'\\ x_tc_t' & x_tx_t'\end{bmatrix}_{\theta=\theta_0}+\mathrm{E}\begin{bmatrix}f_tf_t' & 0\\ 0 & 0\end{bmatrix}_{\theta=\theta_0},$$

and

$$\begin{aligned}\Sigma_{\theta\theta}&=\begin{bmatrix}\Sigma_{\varphi\varphi} & \Sigma_{\eta\varphi}'\\ \Sigma_{\eta\varphi} & \Sigma_{\eta\eta}\end{bmatrix}\\ &=\frac{(k_c-1)}{4}\mathrm{E}\begin{bmatrix}c_tc_t' & c_tx_t'\\ x_tc_t' & x_tx_t'\end{bmatrix}_{\theta=\theta_0}+\frac{v_c}{2}\mathrm{E}\begin{bmatrix}f_tc_t' & f_tx_t'\\ x_tf_t' & 0\end{bmatrix}_{\theta=\theta_0}\\ &\quad+\mathrm{E}\begin{bmatrix}f_tf_t' & 0\\ 0 & 0\end{bmatrix}_{\theta=\theta_0}.\end{aligned}$$

Consistent Standard Errors of the QMLE

Under above assumptions, $\hat{\Sigma}_{\theta\theta}-\Sigma_{\theta\theta}=o_p(1)$, where

$$\hat{\Sigma}_{\theta\theta}=\frac{\left(\hat{k}_c-1\right)}{4T}\begin{bmatrix}\hat{C}'\hat{C} & \hat{C}'\hat{X}\\ \hat{X}'\hat{C} & \hat{X}'\hat{X}\end{bmatrix}+\frac{\hat{v}_c}{2T}\begin{bmatrix}\hat{F}'\hat{C} & \hat{F}'\hat{X}\\ \hat{X}'\hat{F} & 0\end{bmatrix}+\frac{1}{T}\begin{bmatrix}\hat{F}'\hat{F} & 0\\ 0 & 0\end{bmatrix},$$

where C,X and F are matrices with rows c_t', x_t' and f_t', respectively evaluated at $\hat{\theta}$, $\hat{k}_c-1=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t}-1\right)$ and $\hat{v}_c=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^3}{\hat{h}_t^{3/2}}\right)$.

Also under above assumptions, $\hat{J}_{\theta\theta} - J_{\theta\theta} = o_p(1)$, where

$$\hat{J}_{\theta\theta} = \frac{1}{2T}\begin{bmatrix} \hat{C}'\hat{C} & \hat{C}'\hat{X} \\ \hat{X}'\hat{C} & \hat{X}'\hat{X} \end{bmatrix} + \frac{1}{T}\begin{bmatrix} \hat{F}'\hat{F} & 0 \\ 0 & 0 \end{bmatrix}.$$

The i^{th} equation, $i = 1, ..., N$ of our model is:

$$y_{it} = \varphi_i' w_{it} + \varepsilon_{it}, \quad t = 1, ..., T$$

where φ_i is $(K \times 1)$, $\mathrm{E}[\varepsilon_{it}|F_{t-1}] = 0$ and

$$\begin{aligned} \mathrm{E}\left[\varepsilon_{it}^2|F_{t-1}\right] &\equiv h_{it} = \eta_i' s_{i,t-1}, \\ s_{i,t-1}' &= \left(1, \varepsilon_{i,t-1}^2, \varepsilon_{i,t-2}^2, ..., \varepsilon_{i,t-q}^2, h_{i,t-1}, h_{i,t-2}, ..., h_{i,t-p}\right). \end{aligned}$$

2.B Appendix B: Proofs

For the CCC we have,

$$corr\left[\varepsilon_{it}, \varepsilon_{jt}|F_{t-1}\right] = \mathrm{E}\left[\left.\frac{\varepsilon_{it}\varepsilon_{jt}}{\sqrt{h_{it}}\sqrt{h_{jt}}}\right| F_{t-1}\right] = \rho_{ij}.$$

We have the following definitions:

1. $\zeta_{it} = \varepsilon_{it}/\sqrt{h_{it}}$ is i.i.d. $(0, 1)$, for $t = 1, ..., T$, with $\mathrm{E}\left[\zeta_{it}\zeta_{jt}|F_{t-1}\right] = \rho_{ij}$; or $\mathrm{E}\left[\zeta_t\zeta_t'|F_{t-1}\right] = \Gamma = \left\{\rho_{ij}\right\}$, $(N \times N)$, where $\zeta_t = \{\zeta_{it}\}$, $(N \times 1)$, and $\rho_{ii} \equiv 1$, $\Gamma = \Gamma'$ is symmetric and positive definite.
2. Let $\Gamma^{-1} = \{\rho^{ij}\}$, so that $\sum_{m=1}^{N} \rho^{im}\rho_{mj} = \delta_{ij}$, the Kronecker Delta; i.e., $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$.
3. $\varepsilon_{it}^* = \sum_{m=1}^{N} \rho^{im}\zeta_{mt}$, so that $\varepsilon_t^* = \{\varepsilon_{it}^*\} = \Gamma^{-1}\zeta_t$.
4. $f_{it} = w_{it}/\sqrt{h_{it}}$, $c_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \varphi_i}$, $x_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \eta_i}$.

Then, in addition to the properties of ζ_{it} listed in (1), we have the following:

$$\begin{aligned}
\mathrm{E}\left[\varepsilon_{it}^{*}|F_{t-1}\right] &= 0, \\
\mathrm{E}\left[\varepsilon_{t}^{*}\varepsilon_{t}^{*\prime}|F_{t-1}\right] &= \Gamma^{-1} = \left\{\rho^{ij}\right\} = \left\{E\left[\varepsilon_{it}^{*}\varepsilon_{jt}^{*}|F_{t-1}\right]\right\}, \\
\mathrm{E}\left[\varepsilon_{t}^{*}\zeta_{t}^{\prime}|F_{t-1}\right] &= I_N = \left\{E\left[\varepsilon_{it}^{*}\zeta_{jt}|F_{t-1}\right]\right\},
\end{aligned}$$

so that, in particular, $\mathrm{E}\left[\varepsilon_{it}^{*}\zeta_{jt}|F_{t-1}\right] = \delta_{ij}$.

2.B.1 Proof of Lemma 2.1

Proof. Since ζ_t and h_{it} are functionally independent of Γ, we can differentiate (2.65) with respect to φ and η, $i = 1, ..., N$, to obtain the respective scores. we have the following results:

(a) $\dfrac{\partial \varepsilon_t}{\partial \varphi^{\prime}} = \dfrac{\partial}{\partial \varphi^{\prime}}\left(y_t - W_t^{\prime}\varphi\right) = -W_t^{\prime}$,

(b) $\dfrac{\partial h_{it}^{-1/2}}{\partial \varphi_i^{\prime}} = -\dfrac{1}{2}h_{it}^{-3/2}\dfrac{\partial h_{it}}{\partial \varphi_i^{\prime}} = -\dfrac{1}{2}h_{it}^{-1/2}c_{it}^{\prime}$, where $c_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \varphi_i^{\prime}}$, see 2.A.1, and

(c) $\varepsilon_{it}\dfrac{\partial h_{it}^{-1/2}}{\partial \varphi_i^{\prime}} = -\dfrac{1}{2}\zeta_{it}c_{it}^{\prime}$.

Differentiating (2.65) with respect to φ we get

$$\frac{\partial l_t^*}{\partial \varphi^{\prime}} = -\zeta_t^{\prime}\Gamma^{-1}\frac{\partial \zeta_t}{\partial \varphi^{\prime}} - \frac{1}{2}\sum_{j=1}^{N}\frac{1}{h_{jt}}\frac{\partial h_{jt}}{\partial \varphi^{\prime}}. \tag{2.83}$$

Now, because $\zeta_t = D_t^{-1}\varepsilon_t = \left\{h_{it}^{-1/2}\varepsilon_{it}\right\}$ and (in the CCC model) ε_{it} and h_{it} are functionally independent of φ_j, $j \neq i$, using the above results we have: $\underset{(N\times NK)}{\dfrac{\partial \zeta_t}{\partial \varphi^{\prime}}} = -F_t^{\prime} - \dfrac{1}{2}E_tC_t^{\prime}$, (say), where $\underset{(N\times NK)}{F_t^{\prime}} = D_t^{-1}W_t^{\prime} = diag\left(f_{it}^{\prime}\right)$, $\underset{(N\times N)}{E_t} = diag\left(\zeta_{it}\right)$ and $\underset{(N\times NK)}{C_t^{\prime}} = diag\left(c_{1t}^{\prime}, \cdots, c_{Nt}^{\prime}\right)$. Thus, from (2.83) the score for φ is

$$\frac{\partial l_t^*}{\partial \varphi^{\prime}} = \zeta_t^{\prime}\Gamma^{-1}F_t^{\prime} + \frac{1}{2}\zeta_t^{\prime}\Gamma^{-1}E_tC_t^{\prime} - \frac{1}{2}\iota_N^{\prime}C_t^{\prime},$$

or, as a column vector,

$$\frac{\partial l_t^*}{\partial \varphi} = F_t \Gamma^{-1} \zeta_t + \frac{1}{2} C_t \left\{ E_t \Gamma^{-1} \zeta_t - \iota_N \right\}. \qquad (2.84)$$

The score with respect to η, by analogy, is

$$\begin{aligned} \frac{\partial l_t^*}{\partial \eta} &= -\left(\frac{\partial \zeta_t}{\partial \eta'}\right)' \Gamma^{-1\prime} \zeta_t - \frac{1}{2} \sum_{j=1}^{N} \frac{1}{h_{jt}} \frac{\partial h_{jt}}{\partial \eta} \\ &= \frac{1}{2} X_t \left\{ E_t \Gamma^{-1} \zeta_t - \iota_N \right\}, \qquad (2.85) \end{aligned}$$

where $\underset{(N \times NK')}{X_t} = diag\,(x_{it})\,,\; x_{it} = \frac{1}{h_{it}} \frac{\partial h_{it}}{\partial \eta_i}$, see Appendix 2.A.1.

We now turn to the score for ρ, the distinct elements of Γ (with $\rho_{ii} \equiv 1$). Note that when we take differentials, with respect to elements of Γ, $d \ln |\Gamma| = tr\,(\Gamma^{-1} d\Gamma)\,,\; dtr\,(\Gamma^{-1} \zeta_t \zeta_t') = tr\,(d\Gamma^{-1} \zeta_t \zeta_t') = -tr\,(\Gamma^{-1}\; d\Gamma\; \Gamma^{-1} \zeta_t \zeta_t')\,.$ Thus,

$$\begin{aligned} \mathrm{d}\, l_t^* &= -\frac{1}{2} \mathrm{d} \ln |\Gamma| - \frac{1}{2} \mathrm{tr}\left(\mathrm{d}\, \Gamma^{-1} \zeta_t \zeta_t'\right) \\ &= -\frac{1}{2} \mathrm{tr}\left(\Gamma^{-1}\, \mathrm{d}\, \Gamma\right) + \frac{1}{2} \mathrm{tr}\left(\Gamma^{-1} \left(\mathrm{d}\, \Gamma\right) \Gamma^{-1} \zeta_t \zeta_t'\right) \\ &= \frac{1}{2} \mathrm{tr}\left(\Gamma^{-1} \zeta_t \zeta_t' \Gamma^{-1} \left(\mathrm{d}\, \Gamma\right)\right) - \frac{1}{2} \mathrm{tr}\left(\Gamma^{-1}\, \mathrm{d}\, \Gamma\right) \\ &= \frac{1}{2} \mathrm{tr}\left\{\Gamma^{-1} \left(\zeta_t \zeta_t' - \Gamma\right) \Gamma^{-1}\, \mathrm{d}\, \Gamma\right\}. \end{aligned}$$

Let $M_t = \Gamma^{-1} \left(\zeta_t \zeta_t' - \Gamma\right) \Gamma^{-1} = \left\{m_{ij,t}\right\}$, so that (with $\rho_{ii} \equiv 1$)

$$\begin{aligned} \mathrm{tr}\left(M_t \left(\mathrm{d}\, \Gamma\right)\right) &= \sum_{i=1}^{N} \sum_{j=1}^{N} \left(1 - \delta_{ij}\right) m_{ij,t}\, \mathrm{d}\, \rho_{ji} \\ &= 2 \sum_{i=2}^{N} \sum_{j=1}^{i-1} m_{ij,t}\, \mathrm{d}\, \rho_{ij}, \end{aligned}$$

where δ_{ij} is the Kronecker delta: $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$, and we have used symmetry of M_t and Γ. This gives the score for ρ_{ij} :

$$\frac{\partial l_t^*}{\partial \rho_{ij}} = m_{ij,t}, \quad j < i = 2, ..., N. \tag{2.86}$$

■

2.B.2 Proof of Corollary 2.1

Proof. By analogy with the previous analysis. ■

Chapter 3

An Investigation of Parametric Tests of the Constant Conditional Correlation Assumption

3.1 Introduction

Applied researchers have increasingly been using the conditional correlation approach to model multivariate volatility through a multivariate GARCH (MGARCH) model. Although the Dynamic Conditional Correlation (DCC) model is by far the most popular specification among applied researchers, a number of empirical researches have applied the Constant Conditional Correlation (CCC) model; see for example, Bollerslev (1990), Kroner and Claessens (1991), Kroner and Sultan (1991, 1993), Park and Switzer(1995) and Lien and Tse (1998). Due to the simplicity and computational advantages of the CCC model compared to that of the DCC model, on the one hand, but the

*The paper version of this Chapter, jointly authored with my supervisor Chris Orme, has been accepted and will be presented in the contributed sessions of the 65th European Meeting (2011) and 2011 Asian Meeting of the Econometric Society in August 2011. Our thanks to the participants of the Doctoral Conference 2009, University of Manchester, where an earlier version of this Chapter was presented.

restrictiveness of the CCC assumption on the other, testing the adequacy of the CCC-GARCH model is very important both from practical and theoretical point of view. The validity of the CCC assumption remains an empirical question. However, only a few tests of this assumption have been proposed in the literature.

To test the CCC assumption, Bollerslev (1990) suggested some diagnostics applying Ljung-Box portmanteau test statistics based on the crossproducts of the standardized residuals obtained from the CCC-GARCH model. The idea is that if the CCC assumption is valid, then these crossproducts should also be serially uncorrelated. He found that the standardized residuals are uncorrelated in case of five European countries' monthly exchange rate and suggested that this provided evidence of constancy of the correlations. However, serially uncorrelated standardized residuals implies they are linearly independent over time and does not guarantee that the conditional correlations are constant over time. Further, critical values for this test procedure were based on a χ^2 distribution whereas Li and Mak (1994) pointed out that the portmanteau statistic is not asymptotically χ^2 and the use of a χ^2 approximation is inappropriate. Bollerslev (1990) used another diagnostic based on an artificial regression involving the products of the standardized residuals. However, the optimality of portmanteau and residual based tests is not established. Therefore, there remains the question of how powerful these tests are against the dynamic conditional correlation.

Longin and Solnik (1995) suggested another test by taking pairs of variables at a time, explicitly specifying the conditional correlation as a function of potential sources of deviation from constant correlation and then testing the significance of the associated parameters.[1] However, their alternative correlation specification is not guaranteed to be bounded by -1 and 1 (i.e., $|\rho| \leq 1$). This would appear to be a crucial defect. In their empirical applica-

[1]For example, they specify the conditional covariance between two assets as follows:

$$h_{12t} = (\rho_0 + \rho_1 x_{1t} + \cdots + \rho_r x_{rt}) \sqrt{h_{1t} h_{2t}}$$

where h_{it} is the conditional variance of i^{th} asset $i = 1, 2$, x_{it}'s are possible sources of deviation. The CCC assumption corresponds to the null $H_0 : \rho_1 = \cdots = \rho_r = 0$.

tion with monthly excess returns of stock markets of seven major countries from 1960 to 1990, they considered three sources of deviation: a time trend, the presence of threshold and influence of related economic variables (dividend yields and interest rates) and found that the correlation was increased over time and related to dividend yields and interest rates implying the rejection of the CCC hypothesis.

Bera and Kim (2002) developed a test of a bivariate CCC-GARCH model against the alternative that the correlation coefficient is random (over time). This test is an Information Matrix (IM) test (White, 1982) in the form of a LM or score test of random variation in correlation parameter ρ, see e.g., Chesher (1984) and Cox (1983). The null hypothesis of this score test is that the variance of the parameter of interest is zero and the test checks the local behavior of the log-likelihood function is close to the null of no parameter variation. It does not check the CCC assumption directly. Secondly, this test is not robust to non-normality. Thirdly, this test is derived for bivariate case only, limiting its applicability in high dimensional cases. Finally, the IM test assesses several features of the model. Bera and Kim (2002, p. 182) also recognized the fact that "ability of the IM test principle to check various features of the underlying model might be viewed as a drawback rather than an advantage".

However, all the above-mentioned tests are not specifically designed for testing the CCC assumption and in practice they may not be very helpful to address this issue. Tse (2000) proposed a LM test of the CCC assumption. This is a multivariate test which can be applied to high-dimensional data and, among applied workers, the most widely used test of the CCC assumption (see, for example, Tse, 2000; Lien, Tse and Tsui, 2002; Andreou and Ghysels, 2003; Lee, 2006; Aslanidis, Osborn and Sensier, 2008; among others). This test involves the Full QMLE (FQMLE) approach; i.e., simultaneous estimation of the volatility and correlation parameters under the null of CCC. Therefore it might not be robust to the GARCH misspecifications in individual volatility equations. Moreover, Tse uses the OPG version of the LM test which is based on the normality assumption, therefore it may demonstrate relatively poor finite sample properties and may not be robust under non-normality (see, for example, Davidson and MacKinnon, 1983; Bera and McKenzie, 1986; Chesher and Spady, 1991). Finally the time varying alternative specification of correlation matrix as presented by Tse is not necessarily a positive definite matrix for all t (see Section 3.4). For this reason Silvennoinen and Teräsvirta (2009b) interpreted this test as a general misspecification test. In a recent paper, Nakatani and Teräsvirta (2009) proposed a LM test for volatility interaction where the null model is CCC GARCH model against the alternative of Extended CCC (ECCC) GARCH model.

3.1.1 Contributions and Structure of the Chapter

Nevertheless it is evident that the topic of testing the CCC assumption is relatively under-developed compared to other aspects of the MGARCH literature. The aim of this study is to put forward some alternative asymptotically valid testing strategies of the CCC assumption. Firstly, we present and review a conditional moment (CM) testing framework based on the FQMLE of the null CCC model. However, in practice, while estimating a MGARCH model adopting the conditional correlation approach (both constant and dynamic, but particularly for the dynamic one), most researchers use a two-step or PQMLE approach. For the PQMLE, in the first stage the volatility parameters are estimated using a univariate GARCH specification for individual variables and the correlation parameters are estimated using the volatility parameter estimates obtained in the first stage (see Engle and Sheppard, 2008; Hafner, Dijk and Franses, 2005; Billio, Caporin and Gobbo, 2006; among others). There appears to be no testing approach of the CCC assumption available in the literature which allows for the partial estimation. The implication of this is that one has to first estimate the FQMLE of the null CCC model in order to test the null CCC assumption; and if the null is rejected the researcher needs to use the DCC specification which generally use the two-step estimation procedure. Again, there is a well-developed literature which deals with the specification testing for the GARCH models and their asymptotic properties.[2] These two facts motivate us to develop asymptotically valid CM tests of the CCC assumption based on two-step estimation and utilizing the GARCH results. The second contribution of this research is to devise a simple test after the PQML estimation. Thirdly, both the OPG and robust versions of the tests are developed. The proposed tests (both the FQMLE and PQMLE) are easy to implement and demonstrate satisfactory size and good power properties in the simulation experiments. Fourthly, we derive a "new" expression for the average Hessian of the CCC GARCH regression model which is easy to programme. Finally, we have analyzed the Tse's LM test within our CM testing framework and suggested a robust version of this which demonstrate superior size properties under non-normality.

The rest of this Chapter is organized in the following way. The specification and estimation framework of the CCC-MGARCH model are presented in Section 3.2. In Section 3.3, a class of parametric tests with their asymptotic properties is described. An analysis of the Tse's LM test is presented in the next Section. Section 3.5 provides some Monte Carlo evidence and Section 3.6 concludes. The proof of lemmas, propositions and theorems are relegated to Appendix.

3.2 The Null Constant Conditional Correlation Model

Here we consider the CCC-MGARCH model discussed in Chapter 2. Suppose we are interested in the $(N \times 1)$ time-series vector $\{y_t\} = (y_{1t}, \cdots, y_{Nt})'$ and

[2] For example, Lundbergh and Terasvirta (2002) proposed a parametric Lagrange multiplier (LM) type tests of no ARCH effect in standardized errors, linearity, and parameter constancy. Testing for leverage effect developed by Engle and Ng (1993) is widely used in empirical finance. Bollerslev (1986) presented another LM-type test for testing a GARCH model against a higher order GARCH model. One important work in this field is of Halunga and Orme's (2009) unifying parametric testing framework based on the CM principal which takes into account the asymptotically non-negligible estimation effect from the conditional mean parameters. This is the major point of departure of the Halunga and Orme's (2009) test with that of the abovementioned tests. They demonstrated that these tests are asymptotically invalid in the regression context and may have low power. A Monte Carlo study also showed better empirical power properties of their proposed test than those of Engle and Ng (1993) and Lundbergh and Teräsvirta (2002).

$\mathcal{F}_{t-1} = \sigma\left(W_t', W_{t-1}', \cdots\right)$ is the $\sigma-$field generated by the past information up to and including time $t-1$. We consider the following CCC-GARCH specification to model this series:

$$\begin{aligned} y_t &= m\left(W_t;\varphi\right)+\varepsilon_t, \quad t=1,...,T \\ \varepsilon_t &= H_t^{1/2}\left(\varpi\right)\xi_t, \\ H_t &= D_t \Gamma D_t, \\ D_t &= diag(h_{11t}^{1/2},.....,h_{NNt}^{1/2}), \end{aligned} \tag{3.1}$$

where $\varphi' = (\varphi_1', \cdots, \varphi_N')$, $\varphi_i \in \Psi \subset \Re^K$ is a $(NK \times 1)$ vector of conditional mean parameters and W_t' is the $(N \times NK)$ data matrix of the t-th observation, $H_t^{1/2}(\varpi)$ is a $(N \times N)$ positive definite matrix such that $H_t = Var\left(\varepsilon_t|\mathcal{F}_{t-1}\right)$ and ϖ is the vector of unknown parameters which includes conditional mean parameter φ as well (for notational convenience, we drop ϖ in $H_t^{1/2}(\varpi)$), D_t is a $(N \times N)$ diagonal matrix of conditional standard deviation, and $\Gamma = [\rho_{ij}]$ is a time invariant symmetric positive definite conditional correlation matrix with $\rho_{ii} = 1$, $i = 1, ..., N$. The mean function $m\left(W_t;\varphi\right)$ can possibly be nonlinear and W_t contains current and lagged exogenous variables, and lagged dependent variables. However, for simplicity of exposition, we assume a linear specification for the conditional mean function, i.e., $m\left(W_t;\varphi\right) = W_t'\varphi$, so that the conditional mean function becomes $y_t = W_t'\varphi + \varepsilon_t \;\; t = 1, ..., T$. The stochastic sequence $\{\xi_t\}$ is an i.i.d. process with $\mathrm{E}(\xi_t) = 0$ and $Var(\xi_t) = \mathrm{E}(\xi_t\xi_t') = I_N$. We further assume that the error $\{\varepsilon_t, \mathcal{F}_t\}$ is a MDS.

With these assumptions,

$$\mathrm{E}\left[\varepsilon_t|\mathcal{F}_{t-1}\right] = 0, \text{ and } \mathrm{E}\left[\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right] = H_t = \begin{cases} h_{it} & i = j \\ h_{it}^{1/2}h_{jt}^{1/2}\rho_{ij} & i \neq j \end{cases}.$$

Also note that $corr\left[\varepsilon_{it}, \varepsilon_{jt}|\mathcal{F}_{t-1}\right] = \rho_{ij} = \dfrac{h_{ij,t}}{\sqrt{h_{it}}\sqrt{h_{jt}}}$ and $\mathrm{E}\left[\varepsilon_t\varepsilon_{t-j}'|\mathcal{F}_{t-1}\right] = \mathrm{E}\left[\varepsilon_t|\mathcal{F}_{t-1}\right]\varepsilon_{t-j}' = 0$, almost surely (a.s.), for all $j \geq 1$. The CCC models uses the following decomposition of H_t:

$$H_t = D_t \Gamma D_t, \tag{3.2}$$

where $\Gamma = [\rho_{ij}]$ is a symmetric positive definite matrix with $\rho_{ii} = 1,\ i = 1, ..., N$. (3.2) implies that the diagonal elements of the conditional covariance matrix denote the conditional variances, while the off-diagonal elements are $h_{ijt} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij},\ i \neq j,\ 1 \leq i, j \leq N$.

Here we assume that each $h_{it},\ i = 1, ..., N$ has a GARCH (p, q) specification

$$h_{it} = \eta_i' s_{i,t-1} = \alpha_{i0} + \sum_{k=1}^{q} \alpha_{ik} \varepsilon_{i,t-k}^2 + \sum_{j=1}^{p} \beta_{i,j} h_{i,t-j}. \tag{3.3}$$

Denoting $h_t = (h_{1t}, \cdots, h_{Nt})'$, we can write

$$h_t = a_0 + \sum_{k=1}^{q} \widetilde{A}_k \overrightarrow{\varepsilon}_{t-k} + \sum_{j=1}^{p} \widetilde{B}_j h_{t-j},$$

where $\widetilde{A}_k$ and $\widetilde{B}_j$ are both $(N \times N)$ diagonal matrix and a_0 and $\overrightarrow{\varepsilon}_t = (\varepsilon_{1t}^2, \cdots, \varepsilon_{Nt}^2)'$ are $(N \times 1)$ vector.

In the conditional correlation MGARCH models standardized errors play a crucial role. The three types of standardized errors that will appear in subsequent analysis are (see Chapter 2, Definition 2.1-2.3):

$$\zeta_t = D_t^{-1} \varepsilon_t,\ \mathrm{E}[\zeta_t | \mathcal{F}_{t-1}] = 0,\ \mathrm{E}[\zeta_t \zeta_t' | \mathcal{F}_{t-1}] = \Gamma. \tag{3.4}$$

$$\xi_t = H_t^{-1/2} \varepsilon_t,\ \mathrm{E}[\xi_t | \mathcal{F}_{t-1}] = 0,\ \mathrm{E}[\xi_t \xi_t' | \mathcal{F}_{t-1}] = I_N. \tag{3.5}$$

$$\varepsilon_t^* = \Gamma^{-1} \zeta_t = \Gamma^{-1} D_t^{-1} \varepsilon_t,\ \mathrm{E}[\varepsilon_t^* | \mathcal{F}_{t-1}] = 0,\ \mathrm{E}[\varepsilon_t^* \varepsilon_t^{*\prime} | \mathcal{F}_{t-1}] = \Gamma^{-1}. \tag{3.6}$$

3.2.1 The FQMLE Estimation Framework

We will start by defining and introducing some notations which will be useful when deriving the expressions for the scores and expected Hessian. Without loss of generality we assume that each variable correspond to a parameter vector of same dimension; i.e., for the $i-$th variable, $i = 1, \cdots, N$, define $\theta_i = (\varphi_i', \eta_i')' \subset \Re^{K+K'}$ with $\varphi_i \subset \Re^K$ (corresponding to conditional mean function) and $\eta_i \subset \Re^{K'}$ (corresponding to volatility function). Hence, $\theta = (\theta_1', ..., \theta_N')' \subset \Re^{N(K+K')}$ is the parameter vector consisting of conditional mean and volatility parameters for N variables and $\rho \subset \Re^{\frac{N(N-1)}{2}}$ is the vector of distinct correlation parameters. Then define the collection of all parameters $\varpi = (\theta', \rho')' \in \Theta \subset \Re^{N'}$ where $N' = N(K+K') + \frac{N(N-1)}{2}$.[3]

For the i^{th} variable, define $\underset{(T\times K)}{F_i}$, $\underset{(T\times K)}{C_i}$ and $\underset{(T\times K')}{X_i}$ with rows $f_{it}' = \frac{w_{it}'}{\sqrt{h_{it}}}$, $c_{it}' = \frac{1}{h_{it}} \frac{\partial h_{it}}{\partial \varphi_i'}$ and $x_{it}' = \frac{1}{h_{it}} \frac{\partial h_{it}}{\partial \eta_i'}$, respectively. Then define

$$\begin{aligned} \underset{(NT\times NK)}{F} &= diag(F_i), \\ \underset{(N\times NK)}{F_t'} &= diag(f_{it}') \text{ for } t = 1, \cdots, T. \end{aligned}$$

In a similar way, define C, X, C_t' and X_t' matrices. It will be useful to define $\underset{(N\times N)}{E_t} = diag(\zeta_{it})$, $\Gamma_A = I_N + (\Gamma^{-1} \odot \Gamma)$ and suppose Γ^{-1} has a typical element ρ^{ij}. Finally, let ρ^k be the k^{th} column of Γ^{-1}, define $\Gamma^k = \Gamma^{-1} diag(\tau_k)$, where $\tau_k = \{\delta_{ik}\}$, $(N \times 1)$, $i = 1, ..., N$; i.e., Γ^k be the $(N \times N)$ matrix of zeros, except for column k which is ρ^k. Define the following two $(N \times N)$ symmetric matrices:

$$\begin{aligned} P_k &= \Gamma^k + \left(\Gamma^k\right)', \\ \Gamma_{km} &= \rho^k (\rho^m)' + \rho^m \left(\rho^k\right)'. \end{aligned}$$

The Score, Hessian and Limit Distribution of the FQMLE

Define the average log-likelihood function as $L_T^*(\varpi) = \frac{1}{T} \sum l_t^*(\theta, \rho)$, where $l_t^* \equiv l_t^*(\theta, \rho)$ is the quasi-conditional log-likelihood per observation, t, (ignoring any constant terms) which can be written as,

$$l_t^* = -\frac{1}{2} \ln |\Gamma| - \frac{1}{2} \sum_{j=1}^{N} \ln h_{jt} - \frac{1}{2} \zeta_t' \Gamma^{-1} \zeta_t, \tag{3.7}$$

[3]For example, for a AR(1)-bivariate CCC specification with GARCH (1,1) model for individual volatility, we have $N = 2$, $K = 2$, $K' = 3$ and $N' = 11$.

where, $H_t \equiv H_t(\varpi)$, $D_t \equiv D_t(\theta)$.[4] The parameter estimates can be obtained by the FQML method; i.e.,

$$\hat{\varpi} = \arg\max_{\varpi} \sum_{t=1}^{T} l_t^*.$$

Assuming $L_T^*(\varpi) = T^{-1}\sum_{t=1}^{T} l_t^*(\theta,\rho)$ is at least twice continuously differentiable, define the average score for the CCC model $G_T^*(\varpi) = T^{-1}\sum_{t=1}^{T} g_t^*(\varpi)$, where $g_t^*(\varpi) = \left(\frac{\partial l_t^*}{\partial \theta'}, \frac{\partial l_t^*}{\partial \rho'}\right)' = \left(\frac{\partial l_t^*}{\partial \varphi'}, \frac{\partial l_t^*}{\partial \eta'}, \frac{\partial l_t^*}{\partial \rho'}\right)'$, and S^* as a $(T \times N')$ matrix with rows $g_t^{*\prime}(\varpi)$. Using the similar notation, define the Hessian of the log-likelihood function for observation t as $\mathcal{H}_t^*(\varpi) = \frac{\partial^2 l_t^*}{\partial \varpi \partial \varpi'} = \frac{\partial g_t^*(\varpi)}{\partial \varpi'}$. The expression for $g_t^*(\varpi)$ and $G_T^*(\varpi)$ are provided in Lemma 2.1 and Remark 2.7.

The FQMLE $\hat{\varpi}' = \left(\hat{\theta}', \hat{\rho}'\right)$ satisfies $G_T^*(\hat{\varpi}) = 0$. Bollerslev and Wooldridge (1992) showed that under regularity conditions the conditional heteroskedasticity FQML estimators are consistent and asymptotically normal. However, they did not verify whether the regularity conditions hold for specific MGARCH model. Jeantheau (1998) gave conditions for strong consistency of the FQMLE for MGARCH and verified the conditions for extended CCC (ECCC) model. Comte and Lieberman (2003) proved the strong consistency and asymptotic normality of the FQMLE (both when initial state is stationary or fixed) for the BEKK MGARCH specification which requires the finiteness of the moments of the non-Gaussian process ε_t up to order 8; i.e., $\mathrm{E}\left[\varepsilon_{i,t}^8\right] < \infty$, $i = 1, \cdots, , N$. Ling and McAleer (2003) presented a theoretical framework for a class of vector ARMA-GARCH models with the ECCC specification for the conditional heteroskedasticity and their conditions require $\mathrm{E}\left[\varepsilon_{i,t}^6\right] < \infty$, $i = 1, \cdots, , N$. Since the CCC model is nested within this class, we can make use of the following results. Following Ling and McAleer (2003) and Nakatani and Teräsvirta (2009), to ensure the asymptotic normality of QMLE $\hat{\varpi}$ we assume that the followings to hold:

[4] Note that we make use of asterisk (*) to differentiate joint log-likelihood from the univariate GARCH log-likelihood.

Assumption 3.2.1 *The elements* (y_{it}, W'_{it}) *are strictly stationary and ergodic for all* $i = 1, \cdots, N$*; and* $m(W_{it}; \varphi_i)$ *is continuous and* $\mathcal{F}_{t-1}$*-measurable for all* $\varphi_i \in \Psi \subset \Re^K$.

Assumption 3.2.2 *The spectral radius* $\varsigma(\Gamma)$ *has a positive lower bound over the parameter space* Θ *which is a compact subset of the Euclidean space such that* ϖ_0 *lie in the interior of* Θ*. In addition each element of* a_0 *has a positive lower and upper bounds over* Θ.

Assumption 3.2.3 *All the roots of* $\det\left(I_N - \sum_{k=1}^{q} \widetilde{A}_k x^k - \sum_{j=1}^{p} \widetilde{B}_j x^j\right) = 0$ *lie outside the unit circle.*

Assumption 3.2.4 *The following identifiability conditions presented in Jeantheau (1998) are satisfied:*

$$\forall \varpi \in \Theta,\ \forall \varpi_0 \in \Theta, \quad \left.\begin{array}{c} m_{t,\varpi} = m_{t,\varpi_0}\ P_{\varpi_0} -\ a.s. \\ and \\ H_{t,\varpi} = H_{t,\varpi_0}\ P_{\varpi_0} -\ a.s. \end{array}\right\} \Longrightarrow \varpi = \varpi_0.$$

Assumption 3.2.5 $\mathrm{E}\left[\varepsilon_{i,t}^6\right] < \infty$, $i = 1, \cdots, N$.

Assumption 3.2.6 $\operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{H}_t^*(\varpi)$ *exists and finite for all* $\varpi \in \Theta$ *such that the* $N \times N$ *nonrandom matrix*

$$J_{\varpi\varpi}^* = -\mathrm{E}_0\left[\mathcal{H}_t^*(\varpi_0)\right] = \operatorname*{plim}_{T\to\infty} -\frac{1}{T}\sum_{t=1}^{T} \mathcal{H}_t^*(\varpi_0).$$

Theorem 3.1 *Given these assumptions,* $\hat{\varpi} \xrightarrow{p} \varpi_0$ *and*

$$\sqrt{T}\left(\hat{\varpi} - \varpi_0\right) \xrightarrow{d} N\left(0, J_{\varpi\varpi}^{*-1}\Sigma_{GG}^* J_{\varpi\varpi}^{*-1}\right),$$

where $J_{\varpi\varpi}^* = -\mathrm{E}_0\left[\mathcal{H}_t^*(\varpi_0)\right]$ *and* $\Sigma_{GG}^* = \mathrm{E}_0\left[g_{0t}^* g_{0t}^{*\prime}\right]$ *are both finite and positive definite and* $\mathrm{E}_0[.]$ *denotes expectation evaluated at the true parameter values* $\varpi_0 = (\theta_0', \ \rho_0')'$.

The matrix $J_{\varpi\varpi}^*$ is the negative of the expected Hessian while Σ_{GG}^* is the expectation of the outer product of the score vector both evaluated at ϖ_0 and the later is the population information matrix. Moreover, if $\xi_t \sim N(0, I_N)$, then $\Sigma^*_{GG} = J^*_{\varpi\varpi}$ and the asymptotic covariance matrix reaches to the Cramer-Rao lower bound; i.e., Σ^{*-1}_{GG}. By the consistency of the QMLE $\hat{\varpi}$, $J^*_{\varpi\varpi}$ can be consistently estimated by $\hat{J}^*_{\varpi\varpi T} = -\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}^*_t(\hat{\varpi}) = -\frac{1}{T}\sum_{t=1}^{T}\left.\frac{\partial^2 l_t}{\partial\varpi\partial\varpi'}\right|_{\varpi=\hat{\varpi}}$. Note that by definition $H_t(\varpi_0) = \mathrm{E}_0(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1})$ which implies that it would be computationally easier to work with $\widetilde{\mathcal{H}}^*_t(\varpi_0) = \mathrm{E}_0[\mathcal{H}^*_t(\varpi_0)|\,\mathcal{F}_{t-1}]$ (say), as under conditional expectation operator a number of terms in $\frac{\partial^2 l_t}{\partial\varpi\partial\varpi'}$ cancel when evaluated at ϖ_0. Further by the law of iterated expectation we have $J^*_{\varpi\varpi} = -\mathrm{E}_0[\mathrm{E}_0[\mathcal{H}^*_t(\varpi_0)|\,\mathcal{F}_{t-1}]]$ and a simpler estimate of $J^*_{\varpi\varpi}$ is obtained as $\hat{J}^*_{\varpi\varpi T} = -\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathcal{H}}^*_t(\hat{\varpi})$.[5] The Hessian can be derived with reference to Nakatani and Teräsvirta (2009, 2008) who provided the general expression of $\widetilde{\mathcal{H}}^*_t(\varpi)$ for the ECCC-GARCH model. However, these authors derived the expression assuming a known or zero conditional mean. In addition, they did not specify any particular D_t.

Since our focus is on the CCC-GARCH regression model, Lemma 3.1 provides a *new and simple* expression for $\widetilde{\mathcal{H}}^*_t(\varpi_0)$ in the regression context which considers the conditional mean function and GARCH (p, q) specification for individual conditional variances in D_t.

Lemma 3.1 $\widetilde{\mathcal{H}}^*_t(\varpi_0) = \mathrm{E}_0[\mathcal{H}^*_t(\varpi_0)|\,\mathcal{F}_{t-1}]$ *where*

$$\widetilde{\mathcal{H}}^*_t(\varpi) = \begin{bmatrix} \widetilde{\mathcal{H}}^*_{\varphi\varphi} & \widetilde{\mathcal{H}}^*_{\varphi\eta} & \widetilde{\mathcal{H}}^*_{\varphi\rho} \\ \widetilde{\mathcal{H}}^{*\prime}_{\varphi\eta} & \widetilde{\mathcal{H}}^*_{\eta\eta} & \widetilde{\mathcal{H}}^*_{\eta\rho} \\ \widetilde{\mathcal{H}}^{*\prime}_{\varphi\rho} & \widetilde{\mathcal{H}}^{*\prime}_{\eta\rho} & \widetilde{\mathcal{H}}^*_{\rho\rho} \end{bmatrix},$$

and the typical (i, j)*-th block of* $\widetilde{\mathcal{H}}^*_{\varphi\varphi}$, $\widetilde{\mathcal{H}}^*_{\varphi\eta}$, $\widetilde{\mathcal{H}}^*_{\varphi\rho}$, $\widetilde{\mathcal{H}}^*_{\eta\eta}$, $\widetilde{\mathcal{H}}^*_{\eta\rho}$ *and* $\widetilde{\mathcal{H}}^*_{\rho\rho}$, $i, j =$

[5] Although both $\hat{J}^*_{\varpi\varpi T} = -\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}^*_t(\hat{\varpi})$ and $\hat{J}^*_{\varpi\varpi T} = -\frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathcal{H}}^*_t(\hat{\varpi})$ are asymptotically equivalent, in finite sample their performance may vary (see Hafner and Herwartz, 2008).

$1, \cdots, N$ *are given as, respectively:*

$$
\begin{aligned}
\widetilde{\mathcal{H}}^{*}_{\varphi_i\varphi_j} &= -\rho^{ij} f_{it} f'_{jt} - \frac{1}{4}\left(\delta_{ij} + \rho^{ij}\rho_{ij}\right) c_{it} c'_{jt}, \\
\widetilde{\mathcal{H}}^{*}_{\varphi_i\eta_j} &= -\frac{1}{4}\left(\delta_{ij} + \rho^{ij}\rho_{ij}\right) c_{it} x'_{jt}, \\
\widetilde{\mathcal{H}}^{*}_{\varphi_i\rho_{ij}} &= -\frac{1}{2}\delta_{jk}\rho^{ik} c_{kt} - \frac{1}{2}\delta_{ik}\rho^{jk} c_{kt}, \ i > j \\
\widetilde{\mathcal{H}}^{*}_{\eta_i\eta_j} &= -\frac{1}{4}\left(\delta_{ij} + \rho^{ij}\rho_{ij}\right) x_{it} x'_{jt}, \\
\widetilde{\mathcal{H}}^{*}_{\eta_i\rho_{ij}} &= -\frac{1}{2}\delta_{jk}\rho^{ik} x_{kt} - \frac{1}{2}\delta_{ik}\rho^{jk} x_{kt}, \ i > j \\
\widetilde{\mathcal{H}}^{*}_{\rho_{ij}\rho_{km}} &= -\rho^{ik}\rho^{jm} - \rho^{im}\rho^{jk}, \ i > j.
\end{aligned}
$$

Using the above results, Lemma 3.2 provides the expression for $\hat{J}^{*}_{\varpi\varpi T}$ that will be required to construct our tests discussed in Section 3.3.

Lemma 3.2 *For QMLE* $\hat{\varpi}$, $J^{*}_{\varpi\varpi} - \hat{J}^{*}_{\varpi\varpi T} = o_p(1)$, *and* $\hat{J}^{*}_{\varpi\varpi T} = -\frac{1}{T}\sum_{t=1}^{T} \widetilde{\mathcal{H}}^{*}_t(\hat{\varpi})$ *has the form*

$$
\hat{J}^{*}_{\varpi\varpi T} = \begin{bmatrix} \hat{J}^{*}_{\varphi\varphi T} & \hat{J}^{*}_{\varphi\eta T} & \hat{J}^{*}_{\varphi\rho T} \\ \hat{J}^{*\prime}_{\varphi\eta T} & \hat{J}^{*}_{\eta\eta T} & \hat{J}^{*}_{\eta\rho T} \\ \hat{J}^{*\prime}_{\varphi\rho} & \hat{J}^{*\prime}_{\eta\rho T} & \hat{J}^{*}_{\rho\rho T} \end{bmatrix} = T^{-1}\sum_{t=1}^{T} \left[\begin{array}{ccc|c} \frac{\partial^2 l^{*}_t}{\partial\varphi\partial\varphi'} & \frac{\partial^2 l^{*}_t}{\partial\varphi\partial\eta'} & \frac{\partial^2 l^{*}_t}{\partial\varphi\partial\rho'} & \\ \frac{\partial^2 l^{*}_t}{\partial\eta\partial\varphi'} & \frac{\partial^2 l^{*}_t}{\partial\eta\partial\eta'} & \frac{\partial^2 l^{*}_t}{\partial\eta\partial\rho'} & \mathcal{F}_{t-1} \\ \frac{\partial^2 l^{*}_t}{\partial\rho\partial\varphi'} & \frac{\partial^2 l^{*}_t}{\partial\rho\partial\eta'} & \frac{\partial^2 l^{*}_t}{\partial\rho\partial\rho'} & \end{array} \right]_{\varpi=\hat{\varpi}},
$$

where

$$
\begin{aligned}
\hat{J}^{*}_{\varphi\varphi T} &= \frac{1}{T}\left[\hat{F}'\left(\hat{\Gamma}^{-1}\otimes I_T\right)\hat{F} + \frac{1}{4}\hat{C}'\left(\hat{\Gamma}_A\otimes I_T\right)\hat{C}\right] \longrightarrow J^{*}_{\varphi\varphi}, \\
\hat{J}^{*}_{\varphi\eta T} &= \frac{1}{4T}\hat{C}'\left(\hat{\Gamma}_A\otimes I_T\right)\hat{X} \longrightarrow J^{*}_{\varphi\eta}, \\
\hat{J}^{*}_{\varphi\rho T} &= \frac{1}{2T}\hat{C}'\left(I_N\otimes\iota'_T\right)\hat{P} \longrightarrow J^{*}_{\varphi\rho}, \\
\hat{J}^{*}_{\eta\eta T} &= \frac{1}{4T}\hat{X}'\left(\hat{\Gamma}_A\otimes I_T\right)\hat{X} \longrightarrow J^{*}_{\eta\eta}, \\
\hat{J}^{*}_{\eta\rho T} &= \frac{1}{2T}\hat{X}'\left(I_N\otimes\iota'_T\right)\hat{P} \longrightarrow J^{*}_{\eta\rho}, \\
\hat{J}^{*}_{\rho\rho T} &= \hat{\widetilde{P}} \longrightarrow J^{*}_{\varphi\varphi},
\end{aligned}
$$

where $\underset{N\times\frac{N(N-1)}{2}}{P}$ *has rows* $p_k' = vecl(P_k)'$, $k = 1, ..., N$ *and* $\underset{\frac{N(N-1)}{2}\times\frac{N(N-1)}{2}}{\widetilde{P}}$ *has columns* $\widetilde{p}_{km} = vecl\left(\Gamma_{km}\right)$, $m = 1, ..., N-1$, $k = m+1, ..., N$ *(*k *changes more quickly than* m*) while* F', C', X', Γ_A, P_k *and* Γ_{km} *are defined at the onset of this section.*

The above expression of $\hat{J}^*_{\varpi\varpi T}$ is required to obtain the consistent robust estimator of the asymptotic variance covariance matrix of the test indicator based on FQMLE (see Lemma 3.3).

3.2.2 The PQMLE (or Two-step) Estimation

Because of the structure of log-likelihood of the conditional correlation model, a simplified two step estimation procedure can be implemented as suggested by Engle (2002), which involves (at the first step) separate estimation of the N univariate GARCH models to get the volatility estimates, and then using these obtain the correlation parameter estimates. Such a procedure is consistent, but asymptotically inefficient relative to the FQMLE procedure. This partial estimation technique is mostly useful for the DCC models due to the complexity of the estimation procedure, but can be used for the CCC model.

Note that (3.7) can be expressed as the sum of two components, $l_t^*(\theta,\rho) = \sum_{j=1}^N l_t^V(\theta_j) + l_t^C(\theta,\rho)$ where $\sum_{j=1}^N l_t^V(\theta_j) = -\frac{1}{2}\sum_{j=1}^N \left\{\ln h_{jt} + h_{jt}^{-1}\varepsilon_{jt}^2\right\}$ represents conditional log-likelihood contributions for N separate GARCH(p,q) models which is functionally independent of ρ, and $l_t^C(\theta,\rho) = -\frac{1}{2}\ln|\Gamma| - \frac{1}{2}\zeta_t'\Gamma^{-1}\zeta_t + \frac{1}{2}\zeta_t'\zeta_t$ contains the correlation structure. Two step estimation is then pursued as follows:

1. Obtain $\hat{\theta}_j = \arg\max_{\theta_j}\sum_{t=1}^T l_t^v(\theta_j)$, $\quad j = 1, ..., N$ by the QML applying to the GARCH(p,q) specification for individual variables.[6] Then construct standardized residuals as $\hat{\zeta}_{jt} = \hat{h}_{jt}^{-1/2}\hat{\varepsilon}_{jt}$, and $l_t^C(\hat{\theta},\rho) = -\frac{1}{2}\ln|\Gamma| - \frac{1}{2}\hat{\zeta}_t'\Gamma^{-1}\hat{\zeta}_t + \frac{1}{2}\hat{\zeta}_t'\hat{\zeta}_t = k_t - \frac{1}{2}\ln|\Gamma| - \frac{1}{2}\hat{\zeta}_t'\Gamma^{-1}\hat{\zeta}_t$ where k_t is a constant as far as ρ is concerned.

[6] For the scores, see e.g., Appendix 2.A or Halunga and Orme (2009).

2. Obtain $\hat{\rho} = \arg\max_{\rho} \sum_{t=1}^{T} l_t^C(\hat{\theta}, \rho)$, which satisfies the score equations $\sum_{t=1}^{T} \left(\hat{\varepsilon}_{it}^* \hat{\varepsilon}_{jt}^* - \hat{\rho}^{ij} \right) = 0$, $j < i$, with $\hat{\varepsilon}_t^* = \left\{ \hat{\varepsilon}_{jt}^* \right\} = \hat{\Gamma}^{-1} \hat{\zeta}_t$.

Hence the PQMLE $\hat{\varpi} = \left(\hat{\theta}', \hat{\rho}' \right)' = \left(\hat{\theta}_1', \cdots \hat{\theta}_N', \hat{\rho}' \right)'$ can be obtained from the above two steps. Note that to avoid notational complexity we use "hat" to denote both the FQMLE and PQMLE which should not make any confusion, as later while deriving the test statistics notational differences will clearly distinguish the estimation procedure employed. Hafner and Herwartz (2008) provided an analytical expression for the variance of the PQMLE, for both the CCC and DCC models.

As noted by Engle (2002), the correlation matrix Γ_t is also the conditional covariance matrix of standardized errors; i.e., $\mathrm{E}\left[\zeta_t \zeta_t' | F_{t-1} \right] = \Gamma$. Although the score for ρ obtained in second step is not equal to $\sum_{t=1}^{T} \left(\hat{\zeta}_{it} \hat{\zeta}_{jt} - \hat{\rho}_{ij} \right) = 0, j < i$, Bollerslev's (1990) pointed out that a suitable reparameterization ensures that $T^{-1} \sum_{t=1}^{T} \left(\frac{\hat{\varepsilon}_t^2}{\hat{h}_t} - 1 \right) = 0$ so that $\sum_{t=1}^{T} \left(\hat{\zeta}_{it} \hat{\zeta}_{jt} - \hat{\rho}_{ij} \right) = \sum_{t=1}^{T} \left(\hat{\varepsilon}_{it}^* \hat{\varepsilon}_{jt}^* - \hat{\rho}^{ij} \right) = 0$, $j < i$. Therefore, we can use $\hat{\rho}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \hat{\zeta}_{it} \hat{\zeta}_{jt}$, $j < i$ as a consistent estimator for ρ_{ij}. However, noting that the covariance matrix of standardized residuals will never be a correlation matrix in finite sample, as the diagonal will not be exactly equal (though very close) to 1, another option is to use the usual correlation estimator; i.e.,

$$\hat{\rho}_{ij}^* = \frac{\sum_{t=1}^{T} \hat{\zeta}_{it} \hat{\zeta}_{jt}}{\sqrt{\sum_{t=1}^{T} \hat{\zeta}_{it}^2 \sum_{t=1}^{T} \hat{\zeta}_{jt}^2}}, \; i, j = 1, \cdots, N,$$

which is a linear (one-to-one) transformation of $\hat{\rho}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \hat{\zeta}_{it} \hat{\zeta}_{jt}$, $j < i$. In the literature both versions are used to estimate the correlation parameters. For testing of the CCC assumption we only need $\hat{\rho}_{ij}$, $i \neq j$, and score tests are invariant to linear transformation of parameter space (see Dagenais and Dufour, 1991). Hence in this Chapter when developing the asymptotic theory we will use $\hat{\rho} = \frac{1}{T} \hat{\zeta}_t \hat{\zeta}_t'$ as the PQML estimator of ρ.

3.3 A Class of Asymptotically Valid CM Test Procedures

In this section, we develop a class of asymptotically valid parametric testing procedures, along with the first order asymptotic distribution results, of the CCC assumption that are derived from the conditional moment (CM) principle. If both individual GARCH specifications and the CCC assumption are correct, then the definition of standardized residuals, given in (3.4), provides the moment condition corresponding to the CCC assumption; i.e., $\mathrm{E}\left[\zeta_t\zeta_t' - \Gamma|\mathcal{F}_{t-1}\right] = 0$. Note that, the diagonal elements of $(\zeta_t\zeta_t' - \Gamma)$ correspond to the individual GARCH (or volatility) specifications, *whereas the off-diagonal elements correspond to the CCC assumption.* Also due to the symmetry of $(\zeta_t\zeta_t' - \Gamma)$, there are $\frac{N(N+1)}{2}$ independent restrictions in this moment condition, hence we can write these distinct moment restrictions as

$$\mathrm{E}\left[\mathrm{vech}\left(\zeta_t\zeta_t' - \Gamma|\mathcal{F}_{t-1}\right)\right] = 0. \tag{3.8}$$

If we are interested in testing simply the CCC assumption leaving the individual GARCH specifications aside, then we need to consider the strictly lower triangular portion of $(\zeta_t\zeta_t' - \Gamma)$; i.e.,

$$\mathrm{E}\left[\mathrm{vecl}\left(\zeta_t\zeta_t' - \Gamma|\mathcal{F}_{t-1}\right)\right] = 0. \tag{3.9}$$

The parametric misspecification tests of the conditional correlation models can be constructed by considering either (3.8) or (3.9). If the test is based on (3.8), which will be referred as the *Full CM (FCM) test*, it can be treated as a joint misspecification test of the complete MGARCH specification as this would also pick any misspecification in individual volatility specifications with that of the correlation specification. On the other hand, if the underlying moment restriction of the test is (3.9), we will refer this as the *CCC CM (CCM) test.*

Therefore, a joint parametric misspecification test of the CCC and individual volatility assumptions may be constructed as test of the following null moment restriction :

$$\mathrm{E}\left[\operatorname{vech}\left(\zeta_t\zeta_t' - \Gamma\right) \otimes r_t\left(\varpi_0\right)\right] = 0, \tag{3.10}$$

where $r_t(\varpi_0)$ be a $\mathcal{F}_{t-1}$ measurable test variables. To test this null, the generic CM test indicator is constructed as

$$\hat{M}_T^j = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{v}_t \otimes \hat{r}_t\right) = \frac{1}{T}\sum_{t=1}^{T}\hat{m}_t^j, \tag{3.11}$$

where the superscript j denotes joint testing of the CCC and of the individual volatility specifications and $\hat{v}_t = \operatorname{vech}\left(\hat{\zeta}_t\hat{\zeta}_t' - \hat{\Gamma}\right)$ with "hat" denoting that everything is evaluated at a consistent null parameter estimator (either the FQMLE or PQMLE), $\hat{\varpi} = \left(\hat{\theta}', \hat{\rho}'\right)'$. Similarly, a misspecification test of the CCC assumption, *only*, can be conducted by testing the moment restriction:

$$\mathrm{E}\left[\operatorname{vecl}\left(\zeta_t\zeta_t' - \Gamma\right) \otimes r_t\left(\varpi_0\right)\right] = 0. \tag{3.12}$$

The corresponding CM test indicator would have the following form

$$\hat{M}_T^c = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{v}_t^* \otimes \hat{r}_t\right) = \frac{1}{T}\sum_{t=1}^{T}\hat{m}_t^c, \tag{3.13}$$

where $\hat{v}_t^* = \operatorname{vecl}\left(\hat{\zeta}_t\hat{\zeta}_t' - \hat{\Gamma}\right)$ and superscript c denote testing of the only CCC assumption. It is to be noted here that (3.13) is simply a subset of (3.11).

Example 3.1 *For example, in the bivariate case,*

$$\mathrm{E}\left[\zeta_t\zeta_t' - \Gamma|\mathcal{F}_{t-1}\right] = \mathrm{E}\left[\left.\begin{pmatrix}\zeta_{1t}^2 & \zeta_{1t}\zeta_{2t}\\ \zeta_{1t}\zeta_{2t} & \zeta_{2t}^2\end{pmatrix} - \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right|\mathcal{F}_{t-1}\right] = 0.$$

Then (3.8) becomes

$$\mathrm{E}\left[\left.\left(\zeta_{1t}^2 - 1,\ \zeta_{1t}\zeta_{2t} - \rho,\ \zeta_{2t}^2 - 1\right)'\right|\mathcal{F}_{t-1}\right] = 0.$$

The first and third components refer to the individual GARCH equations *while second one corresponds to the CCC assumption. Therefore, in this case (3.9) is simply* $\mathrm{E}\left[\zeta_{1t}\zeta_{2t} - \rho|\mathcal{F}_{t-1}\right] = 0$. *Subsequently for a* $\mathcal{F}_{t-1}$ *measurable test variable* $r_t\left(\varpi_0\right)$ *(3.10) becomes*

$$\mathrm{E}\begin{bmatrix}\zeta_{1t}^2 - 1 \\ \zeta_{1t}\zeta_{2t} - \rho \\ \zeta_{2t}^2 - 1\end{bmatrix} \otimes r_t\left(\varpi_0\right) = 0, \tag{3.14}$$

and (3.12) becomes $\mathrm{E}\left[\left(\zeta_{1t}\zeta_{2t} - \rho\right) \otimes r_t\left(\varpi_0\right)\right] = 0$.

To develop asymptotically valid tests of the CCC hypothesis we need to establish the limit distributions of the test indicator vector $\sqrt{T}\hat{M}_T$. Both the FQMLE and PQMLE approaches are considered while deriving the test statistics and their asymptotic distributions. We illustrate the procedure of constructing the test statistics considering both Gaussian and non-Gaussian distributions of the fully standardized error process, ξ_t. In case of non-normally distributed ξ_t we develop a *non-normality robust* procedure in the similar spirit of Wooldridge (1990).[7] When ξ_t follows a normal distribution the generalized IM inequality holds (see, e.g., Newey, 1985) and the OPG covariance matrix estimator can be employed in deriving the test statistics.

3.3.1 Case 1: Tests Based on the FQMLE

The test indicator under consideration is $\hat{M}_{FT} \equiv M_{FT}(\hat{\varpi}) = T^{-1}\sum_{t=1}^{T}\hat{m}_{Ft}$, where subscript F represents the FQMLE case, with $\mathrm{E}_0\left[m_{Ft}\right] = 0$. Define the $(T \times r)$ matrix R with rows m'_{Ft}. Hereafter we will use the notation $G_{0T}^* = G_T^*\left(\varpi_0\right)$, $M_{0FT} \equiv M_{FT}(\varpi_0)$, etc. where ϖ_0 denotes the true parameter values and $\hat{M}_{FT}$, $\hat{J}_{\varpi\varpi T}^{*-1}$ etc. to denote evaluation at $\hat{\varpi}$. We assume that the following central limit theorem holds:

Proposition 3.1 $\sqrt{T}\begin{bmatrix} M_{0FT} \\ G_{0T}^* \end{bmatrix} = T^{-1/2}\begin{bmatrix} \sum_{t=1}^{T} m_{0Ft} \\ \sum_{t=1}^{T} g_{0t}^* \end{bmatrix} \xrightarrow{d} N\left(0, \Sigma^*\right),$

[7] Similar approach was employed by Halunga and Orme (2009).

where $\Sigma^* = \begin{bmatrix} \Sigma_{MM} & \Sigma^*_{MG} \\ \Sigma^*_{GM} & \Sigma^*_{GG} \end{bmatrix}$,
$\Sigma_{MM} = \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} m_{0Ft} m'_{0Ft}$,
$\Sigma^*_{GG} = \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} g^*_{0t} g^{*\prime}_{0t}$, and
$\Sigma^*_{MG} = \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} m_{0Ft} g^{*\prime}_{0t}$.

Remark 3.1 *Assuming a suitable ULLN, Σ_{MM} might be consistently estimated, for example, by $T^{-1}\sum_{t=1}^{T} \hat{m}_{Ft}\hat{m}'_{Ft}$ (see, Halunga and Orme, 2009).*

We then have the following result.

Theorem 3.2 *Given $\hat{\varpi} \xrightarrow{p} \varpi_0$, the CLT stated in Proposition 3.1 and a suitable ULLN,*

$$\sqrt{T}\hat{M}_{FT} \xrightarrow{d} N(0, V),$$

where
$V = A^* \Sigma^* A^{*\prime}$, *and*
$A^* = \left[I_r : -J^*_{M\varpi} J^{*-1}_{\varpi\varpi} \right]$, *with* $J^*_{\varpi\varpi} = -\underset{T\longrightarrow\infty}{\text{plim}} \frac{1}{T}\left[\sum_{t=1}^{T} \mathcal{H}^*_t(\varpi_0)\right]$, $J^*_{M\varpi} = -\underset{T\longrightarrow\infty}{\text{plim}} \left[\frac{\partial M_{0Ft}}{\partial \varpi'}\right]$, *and I_r is the identity matrix of rank $r = rank(\Sigma_{MM})$.*

Remark 3.2 *Note that the variance-covariance matrix V can be written as*

$$V = \Sigma_{MM} - \Sigma^*_{MG} J^{*-1}_{\varpi\varpi} J^{*\prime}_{M\varpi} - J^*_{M\varpi} J^{*-1}_{\varpi\varpi} \Sigma^*_{GM} + J^*_{M\varpi} J^{*-1}_{\varpi\varpi} \Sigma^*_{GG} J^{*-1}_{\varpi\varpi} J^{*\prime}_{M\varpi}. \quad (3.15)$$

From the preceding result, the general form of the CCC misspecification test statistic, based on the FQMLE, is the quadratic form

$$T_F = T\hat{M}'_{FT}\hat{V}_T^{-1}\hat{M}_{FT}, \quad (3.16)$$

under the null which has a χ^2_r limiting distribution, where $\hat{V}_T$ is any consistent estimator for V; i.e., $\hat{V}_T = V + o_p(1)$.

Case 1a: The Robust FQMLE Test

To construct a robust (to non-normality) test statistics we need a consistent estimator $\hat{V}^r_T = \hat{A}^*\hat{\Sigma}^*\hat{A}^{*\prime}$ where the superscript r signifies the robust estimator, for which we require $\hat{J}^*_{M\varpi T} = -T^{-1}\sum \frac{\partial m_{Ft}}{\partial \hat{\varpi}'}$, $\hat{J}^*_{\varpi\varpi T}$ and $\hat{\Sigma}^*$. For $\hat{\varpi}$, we can construct $\hat{J}^*_{M\varpi T}$ using the results provided in Proposition 3.2.

Proposition 3.2 *It can be shown that for $i = 1, \cdots N$,*

$$\mathrm{E}_0\left[\frac{\partial\left(\zeta_{it}^2 - 1\right)}{\partial\theta'} r_t\right] = -\operatorname*{plim}_{T\longrightarrow\infty}\frac{1}{T}\left[R^{*\prime}(0,\ldots,Z_i,\cdots,0)\right],$$

and for $i \neq j,\ j < i = 2, \cdots, N$,

$$\begin{aligned}
\mathrm{E}_0\left[\frac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial\theta'} r_t\right] &= -\frac{1}{2}\rho_0 \operatorname*{plim}_{T\longrightarrow\infty}\frac{1}{T}\left\{R^{*\prime}(0,\cdots,Z_i,\cdots,Z_j,\cdots 0)\right\}, \\
\mathrm{E}_0\left[\frac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial\rho'} r_t\right] &= -\operatorname*{plim}_{T\longrightarrow\infty}\frac{1}{T}(0,\ldots,1',\cdots,0)R^*,
\end{aligned}$$

where Z_i is a $(T \times k_i)$ matrix having rows $z_{it}' = (c_{it}', x_{it}')$ and R^* having rows r_t', if r_t is a vector of test variables, or R^* is a vector with typical element r_t if the latter is a scalar.

Example 3.2 *Again consider $N = 2$, then the full moment condition is given in (3.14). Hence we have*

$$\hat{J}^*_{M\varpi T} = \frac{1}{T}\begin{bmatrix} \hat{R}^{*\prime}(\hat{Z}_1, 0, 0) \\ \frac{1}{2}\hat{\rho}\hat{R}^{*\prime}(\hat{Z}_1, \hat{Z}_2), \iota_T'\hat{R}^* \\ \hat{R}^{*\prime}(0, \hat{Z}_2, 0) \end{bmatrix}.$$

Clearly, for the only CCC moment condition $\hat{J}^*_{M\varpi T} = \left\{\frac{1}{2T}\hat{\rho}\hat{R}^{*\prime}(\hat{Z}_1, \hat{Z}_2), \iota_T'\hat{R}^*\right\}$.

The next Lemma shows how to obtain a consistent robust estimator $\hat{V}_T^r$.

Lemma 3.3 *Under suitable assumptions, $\hat{\Sigma}_T^* - \Sigma^* = o_p(1)$ and $\hat{A}^* - A^* = o_p(1)$, where*

$$\begin{aligned}
\hat{\Sigma}_T^* &= \begin{bmatrix} \hat{\Sigma}_{MM} & \hat{\Sigma}^*_{MG} \\ \hat{\Sigma}^*_{GM} & \hat{\Sigma}^*_{GG} \end{bmatrix} = \frac{1}{T}\begin{bmatrix} \hat{R}'\hat{R} & \hat{S}^{*\prime}\hat{R} \\ \hat{R}'\hat{S}^* & \hat{S}^{*\prime}\hat{S}^* \end{bmatrix}, \\
\hat{A}^* &= \left[I_r : -\hat{J}^*_{M\varpi T}\hat{J}^{*-1}_{\varpi\varpi T}\right],
\end{aligned}$$

with R and S^* denoting matrices with rows m'_{Ft} and $g_t^{*\prime}(\varpi)$, respectively, $\hat{J}^*_{\varpi\varpi T}$ and $\hat{J}^*_{M\varpi T}$ are constructed using Lemma 3.2 and Proposition 3.2, respectively, all evaluated at $\hat{\varpi}$. Therefore we have,

$$\hat{V}_T^r = \hat{\Sigma}_{MM} - \hat{\Sigma}^*_{MG}\hat{J}^{*-1}_{\varpi\varpi T}\hat{J}^{*\prime}_{M\varpi T} - \hat{J}^*_{M\varpi T}\hat{J}^{*-1}_{\varpi\varpi T}\hat{\Sigma}^*_{GM} + \hat{J}^*_{M\varpi T}\hat{J}^{*-1}_{\varpi\varpi T}\hat{\Sigma}^*_{GG}\hat{J}^{*-1}_{\varpi\varpi T}\hat{J}^{*\prime}_{M\varpi T}.$$

Tests based on this estimator $\hat{V}_T^r$ will be referred as the *robust FQMLE test* and will be denoted as $T_F^{(r)}$.

Case 1(b) The OPG FQMLE Test

If $\xi_t \sim N(0, I_N)$, then $\Sigma^*_{GG} = J^*_{\varpi\varpi}$ and (3.15) reduces to $V = \Sigma_{MM} - \Sigma^*_{MG}\Sigma^{*-1}_{GG}\Sigma^*_{GM}$. The following Lemma provides an expression for the consistent estimator of V when the normality assumption holds.

Lemma 3.4 *Under suitable assumptions and $\xi_t \sim N(0, I_N)$, V can be consistently estimated by*

$$\hat{V}_T = \hat{\Sigma}_{MM} - \hat{\Sigma}^*_{MG}\hat{\Sigma}^{*-1}_{GG}\hat{\Sigma}^*_{GM} = \frac{1}{T}\hat{W}^{*\prime}\hat{W}^*,$$

where

$$\begin{aligned}
\hat{\Sigma}_{MM} &= T^{-1}\sum_{t=1}^{T}\hat{m}_{Ft}\hat{m}'_{Ft} = T^{-1}\hat{R}'\hat{R}, \\
\hat{\Sigma}^*_{GG} &= T^{-1}\sum_{t=1}^{T}\hat{g}^*_t\hat{g}^{*\prime}_t = T^{-1}\hat{S}^{*\prime}\hat{S}^*, \\
\hat{\Sigma}^*_{MG} &= T^{-1}\sum_{t=1}^{T}\hat{m}_{Ft}\hat{g}^{*\prime}_t = T^{-1}\hat{S}^{*\prime}\hat{R}, \\
\hat{W}^* &= \hat{B}^*\hat{A}^{*\prime}, \\
\hat{A}^* &= \left[I_r : -\hat{\Sigma}^*_{MG}\hat{\Sigma}^{*-1}_{GG}\right], \text{ and} \\
B^* &= [R,\ S^*],
\end{aligned}$$

with R and S^* evaluated at $\hat{\varpi}$.

In this case, the test statistic (3.16) has a convenient OPG form. To see this, note that

$$\hat{W}^* = \hat{B}^* \hat{A}^{*\prime} = \hat{R} - \hat{S}^* \left(\hat{S}^{*\prime}\hat{S}^*\right)^{-1} \hat{S}^{*\prime}\hat{R}.$$

Exploiting the FOC that $\hat{S}^{*\prime}\iota_T \equiv 0$ we have $\hat{W}^{*\prime}\iota_T \equiv \hat{R}^{\prime}\iota_T$. Hence an alternative form of the test statistic under normality is given by:

$$T_F = \iota_T^{\prime}\hat{R}\left(\hat{W}^{*\prime}\hat{W}^*\right)^{-1}\hat{R}^{\prime}\iota_T = \iota_T^{\prime}\hat{W}^*\left(\hat{W}^{*\prime}\hat{W}^*\right)^{-1}\hat{W}^{*\prime}\iota_T, \tag{3.17}$$

where ι_T is the $(T \times 1)$ column vector of ones. (3.17) can be interpreted as $T - RSS$ where RSS is the residual sum of squares from the regression of ι_T on $\hat{W}$. Note that, this test can be constructed easily by defining $\hat{U}^* = \left(\hat{R}, \hat{S}^*\right)$, then

$$T_F = \iota_T^{\prime}\hat{U}^*\left(\hat{U}^{*\prime}\hat{U}^*\right)^{-1}\hat{U}^{*\prime}\iota_T, \tag{3.18}$$

which can be obtained as $T - RSS$ from a regression of ι_T on $\hat{U}^*$.

Summary: The FQMLE Tests

From the above results, for each of the FCM and CCM test statistics, based on the FQMLE $\hat{\varpi}$, we have two versions namely, the robust and the OPG; i.e.,

1. Robust (to non-normality) FCM test

$$T_F^{j(r)} = T\hat{M}_{FT}^{j\prime}\left(\hat{V}_T^{j(r)}\right)^{-1}\hat{M}_{FT}^{j}. \tag{3.19}$$

2. OPG FCM test

$$T_F^{j} = T\hat{M}_{FT}^{j\prime}\left(\hat{V}_T^{j}\right)^{-1}\hat{M}_{FT}^{j}. \tag{3.20}$$

3. Robust (to non-normality) CCM test

$$T_F^{c(r)} = T\hat{M}_{FT}^{c\prime}\left(\hat{V}_T^{c(r)}\right)^{-1}\hat{M}_{FT}^{c}. \tag{3.21}$$

4. OPG CCM test

$$T_F^c = T\hat{M}_{FT}^{c\prime}\left(\hat{V}_T^c\right)^{-1}\hat{M}_{FT}^c. \tag{3.22}$$

The robust variance estimator $\hat{V}_T^{j(r)}$ and $\hat{V}_T^{c(r)}$ can be obtained using Lemma 3.3 while the OPG test statistics are constructed using the artificial regression as given in (3.18) with appropriate test indicators.

3.3.2 Case 2: Tests Based on the PQMLE

Define the test indicator under investigation as

$$\hat{M}_{PT} \equiv M_{PT}(\hat{\theta}, \hat{\rho}) = \frac{1}{T}\sum_{t=1}^{T} m_{Pt}(\hat{\theta}, \hat{\rho}),$$

with $\mathrm{E}_0\left[m_{Pt}\right] = 0$, where subscript P represents the PQMLE case; i.e., the correlation parameters are estimated by $\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{\zeta}_{it}\hat{\zeta}_{jt}$, $j < i$. We will establish the results considering the bivariate case for ease of exposition, so that $\theta' = (\theta_1', \theta_2')$. It is straightforward to generalize the results to higher dimensional cases.

Let $L_{iT}(\theta_i) = \dfrac{1}{T}\sum_{t=1}^{T} l_{it}(\theta_i)$ be the average log-likelihood of the univariate GARCH models for the i-th variable where $l_{it}(\theta_i) = -\frac{1}{2}\left[\ln\left(h_{it}\right) + \dfrac{\varepsilon_{it}^2}{h_{it}}\right]$ (ignoring constants). Define $G_i(\theta_i) = \dfrac{1}{T}\sum_{t=1}^{T} g_{it}(\theta_i) = \dfrac{1}{T}\sum_{t=1}^{T}\dfrac{\partial l_{it}(\theta_i)}{\partial\theta_i}$, and

$$J_{\theta\theta} = \begin{bmatrix} J_1(\theta_1) & 0 \\ 0 & J_2(\theta_2) \end{bmatrix},$$

where $J_i\left(\theta_i\right) \equiv J_i = -\mathrm{E}_0\left[\frac{\partial^2 l_{it}(\theta_i)}{\partial\theta_i\partial\theta_i'}\right] = -\operatorname{plim}_{T\to\infty}\dfrac{1}{T}\sum_{t=1}^{T}\dfrac{\partial^2 l_{it}(\theta_i)}{\partial\theta_i\partial\theta_i'}$. The block diagonal structure of $J_{\theta\theta}$ follows from the fact, that in the PQMLE framework, the univariate GARCH estimating equations are functionally independent. Also define $Q = (Q_1, Q_2)$ and $S = (S_1, S_2)$, where Q_i and S_i are both $(T \times k_i)$ matrix with $k_i = K + K'$, and rows $g_{it}'(\theta_i) = \dfrac{\partial l_{it}(\theta_i)}{\partial\theta_i'}$ and

$g_t^{*\prime}(\theta_i) = \dfrac{\partial l_t^*(\theta, \rho)}{\partial \theta_i'}$, $i = 1, 2$, respectively.[8] As before, R is a $(T \times r)$ matrix, but now with rows m_{Pt}'.

The separate limit distributions of $\sqrt{T}\left(\hat{\theta}_i - \theta_{i0}\right) = J_i(\theta_{i0})^{-1}\sqrt{T}G_i(\theta_{i0}) + o_p(1)$, for true parameter values θ_{i0}, $(k_i \times 1)$, $i = 1, 2$, are essentially given in Halunga and Orme (1990, Theorem 1). We have, $\sqrt{T}G(\theta_0) \xrightarrow{d} N(0, \Sigma_{GG})$, where, $G_{0T} = (G_1(\theta_{10})', G_2(\theta_{20})')'$, and

$$\Sigma_{GG} = \operatorname*{plim}_{T\to\infty} \frac{1}{T} Q'Q = \operatorname*{plim}_{T\to\infty} \frac{1}{T} \begin{bmatrix} Q_1'Q_1 & Q_1'Q_2 \\ Q_2'Q_1 & Q_2'Q_2 \end{bmatrix}.$$

Now to test the significance of the test indicator, $\hat{M}_{PT}$, the limit joint distribution needs to take account of the estimation effect from correlation parameters, $\hat{\rho}$. We can ignore this estimation effect from ρ, which will eventually lead to relatively simple to construct asymptotically valid tests, if we can impose the following condition:

Condition 3.1 $\sqrt{T}M_T(\hat{\theta}, \hat{\rho}) = \sqrt{T}M_T(\hat{\theta}, \rho_0) + o_p(1)$.

This implies that the effect of estimating ρ using the first step estimator $\hat{\theta}' = \left(\hat{\theta}_1', \hat{\theta}_2'\right)$ can be ignored (asymptotically). Although it seems a very restricted condition, in our case, this condition can easily be met by using a centered (i.e., demeaned) test variable $\left(\hat{r}_t - \widehat{\bar{r}}\right)$ and thereby making the test indicator $\hat{M}_{PT}$ functionally independent of ρ.

Example 3.3 *For example, in the bivariate context, consider the only correlation test indicator given by*

$$\begin{aligned} \hat{M}_{PT}^c &= \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\zeta}_{1t}\hat{\zeta}_{2t} - \hat{\rho}\right]\hat{r}_t \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{\zeta}_{1t}\hat{\zeta}_{2t}\left(\hat{r}_t - \widehat{\bar{r}}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\zeta}_{1t}\hat{\zeta}_{2t} - \hat{\rho}\right)\left(\hat{r}_t - \widehat{\bar{r}}\right), \quad (3.23) \end{aligned}$$

[8] That is, Q_i is the matrix having rows univariate GARCH scores $g_{it}'(\theta_i)$ while the rows of S_i contains the FQMLE scores $g_t^{*\prime}(\theta_i)$, given in Lemma 2.1, corresponding to the conditional mean and volatility parameters θ_i barring correlation parameter ρ.

where $\widehat{\overline{r}} = \dfrac{1}{T}\sum_{t=1}^{T}\hat{r}_t$ *and* $\hat{\rho} = \dfrac{1}{T}\sum_{t=1}^{T}\hat{\zeta}_{1t}\hat{\zeta}_{2t}$.

This simple demeaning trick produces algebraically equivalent test indicators that do not involve $\hat{\rho}$. In other words, $\hat{M}_{PT}$ is simply a function of $\left(\hat{\theta}_1', \hat{\theta}_2'\right)$ and does not involve $\hat{\rho}$ and hence allows us to deduce the limit distribution under Condition 3.1.

Remark 3.3 *Note that, since* $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\zeta}_{it}^2 - 1\right) \neq 0$, *we have*

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\zeta}_{it}^2 - 1\right)\hat{r}_t \neq \frac{1}{T}\sum_{t=1}^{T}\hat{\zeta}_{it}^2\left(\hat{r}_t - \widehat{\overline{r}}\right).$$

However, Condition 3.1 does apply to the full CM test indicator M_{PT}^j *since* $\dfrac{1}{T}\sum_{t=1}^{T}\left(\hat{\zeta}_{it}^2 - 1\right)\hat{r}_t,\ i = 1, 2$ *does not involve* $\hat{\rho}$. *Hence, it is not necessary to demean the test variable for the elements of test indicator* $\hat{M}_{PT}^j$ *that do not involve* $\hat{\rho}$. *However, hereafter for simplicity (see Remark 3.5 later), we will consider demeaned test variables for all elements of* $\hat{M}_{PT}^j$; *i.e.,*

$$\hat{M}_{PT}^j = \frac{1}{T}\sum_{t=1}^{T}\left(\begin{array}{c}\left[\hat{\zeta}_{1t}^2 - 1\right]\left(\hat{r}_t - \widehat{\overline{r}}\right)\\ \left(\hat{\zeta}_{1t}\hat{\zeta}_{2t} - \hat{\rho}\right)\left(\hat{r}_t - \widehat{\overline{r}}\right)\\ \left[\hat{\zeta}_{2t}^2 - 1\right]\left(\hat{r}_t - \widehat{\overline{r}}\right)\end{array}\right) = \frac{1}{T}\sum_{t=1}^{T}\left\{\hat{\upsilon}_t \otimes \left(\hat{r}_t - \widehat{\overline{r}}\right)\right\}. \quad (3.24)$$

To derive the asymptotic distribution, we assume the following central limit theorem to hold:

Proposition 3.3 *Under suitable regularity conditions,*

$$\sqrt{T}\left[\begin{array}{c}M_{0PT}\\ G_{0T}\end{array}\right] = \frac{1}{\sqrt{T}}\left[\begin{array}{c}\sum_{t=1}^{T} m_{0Pt}\\ \sum_{t=1}^{T} g_{0t}\end{array}\right] \xrightarrow{d} N\left(0, \Sigma\right),$$

where $\Sigma = \begin{bmatrix} \Sigma_{MM} & \Sigma_{MG} \\ \Sigma_{GM} & \Sigma_{GG} \end{bmatrix}$ *with*

$$\begin{aligned}
\Sigma_{MM} &= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} m_{0Pt} m_{0Pt}' = \operatorname*{plim}_{T\to\infty} \frac{1}{T} R'R, \\
\Sigma_{GG} &= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} g_{0t} g_{0t}', \ \textit{and} \\
\Sigma_{MG} &= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} m_{0t} g_{0t}' = \operatorname*{plim}_{T\to\infty} \frac{1}{T} R'Q.
\end{aligned}$$

The above arguments enable us to construct asymptotically valid test from the first stage estimates of $\hat{\theta}$ only. The following theorem gives the limit distribution of $\sqrt{T}\hat{M}_{PT}$ under condition (3.1).

Theorem 3.3 *Given* $\hat{\varpi} \overset{p}{\longrightarrow} \varpi_0$, *the CLT stated in proposition (3.3) and a suitable ULLN,* $\sqrt{T}\hat{M}_{PT} \overset{d}{\longrightarrow} N(0, V_1)$ *where*
$V_1 = A_1 \Sigma A_1'$,
$A_1 = \left[I_r, \ -J_{M\theta} \times J_{\theta\theta}^{-1} \right]$, *with*
$J_{\theta\theta} = \begin{bmatrix} J_1(\theta_{10}) & 0 \\ 0 & J_2(\theta_{20}) \end{bmatrix}$, $J_{M\theta} = -\operatorname*{plim}_{T\longrightarrow\infty} \left[\frac{\partial M_{0Pt}}{\partial \theta'} \right]$ *and* I_r *is the identity matrix of rank* $r = rank\,(\Sigma_{MM})$.

From this result, the general form of the CCC misspecification test using the PQMLE has the quadratic form

$$T_P = T\hat{M}_{PT}' \hat{V}_{1T}^{-1} \hat{M}_{PT}, \tag{3.25}$$

under the null which has a χ_r^2 limiting distribution, where $\hat{V}_{1T}$ is any consistent estimator for V_1; i.e., $\hat{V}_{1T} = V_1 + o_p(1)$. We want to stress here that to get asymptotically valid test statistic, one has to use $\hat{\upsilon}_t \otimes \left(\hat{r}_t - \widehat{\bar{r}}\right)$ for the FCM and $\left(\hat{\zeta}_{1t}\hat{\zeta}_{2t} - \hat{\rho}\right)(\hat{r}_t - \bar{r})$ for the CCM rather than $\zeta_{1t}\zeta_{2t}(r_t - \bar{r})$ when constructing the test indicator $\hat{M}_{PT}$. This has no effect on the numerator of the test statistic, as $\sum_{t=1}^{T} \left[\hat{\zeta}_{1t}\hat{\zeta}_{2t} - \hat{\rho}\right] \hat{r}_t = \sum_{t=1}^{T} \left(\hat{\zeta}_{1t}\hat{\zeta}_{2t} - \hat{\rho}\right)\left(\hat{r}_t - \widehat{\bar{r}}\right) = \sum_{t=1}^{T} \hat{\zeta}_{1t}\hat{\zeta}_{2t}\left(\hat{r}_t - \widehat{\bar{r}}\right)$, but gives us the right expression for the asymptotic variance estimate.

To construct asymptotically valid test statistics we need a consistent expression of V_1. Similar to the FQMLE case, we will consider both the robust and OPG versions in the following.

Case 2a: The Robust-PQMLE Test

To construct a robust (to non-normality) test of (3.25), first note that $\hat{J}_{M\theta T}$ can be obtained using the results of Proposition 3.2, but replacing r_t by $\left(\hat{r}_t - \widehat{\overline{r}}\right)$, the demeaned test variables. Let us define $\overline{R}^*$ as a matrix having rows $(r_t - \overline{r}_t)'$ if r_t is a vector of test variables, or as a vector with typical element $(r_t - \overline{r}_t)$ in case of scalar r_t.

Example 3.4 *In the bivariate case with full moment condition, we have:*

$$\hat{J}^*_{M\varpi T} = \frac{1}{T}\begin{bmatrix} \widehat{\overline{R}}^{*\prime}(\hat{Z}_1, 0) \\ \frac{1}{2}\hat{\rho}\widehat{\overline{R}}^{*\prime}(\hat{Z}_1, \hat{Z}_2) \\ \widehat{\overline{R}}^{*\prime}(0, \hat{Z}_2) \end{bmatrix}, \tag{3.26}$$

and, for the only correlation moment condition, it becomes

$$\hat{J}^*_{M\varpi T} = \frac{1}{2T}\hat{\rho}\widehat{\overline{R}}^{*\prime}(\hat{Z}_1, \hat{Z}_2).$$

In addition, from Halunga and Orme (1990, Lemma 1), for i^{th} variable we have:

$$\hat{J}_i = \frac{1}{2T}\begin{bmatrix} \hat{C}_i'\hat{C}_i & \hat{C}_i'\hat{X}_i \\ \hat{X}_i'\hat{C}_i & \hat{X}_i'\hat{X}_i \end{bmatrix} + \frac{1}{T}\begin{bmatrix} \hat{F}_i'\hat{F}_i & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.27}$$

where F_i, C_i and X_i have rows $f_{it}' = \dfrac{w_{it}'}{\sqrt{h_{it}}}$, $c_{it}' = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \varphi_i}$ and $x_{it}' = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \eta_i}$, all evaluated at the PQMLE $\hat{\theta}_i' = (\hat{\varphi}_i', \hat{\eta}_i')$.

Combining the above two results, the next lemma provides an expression for the robust consistent variance estimator $\hat{V}_{1T}^r$ for the bivariate case which can be generalized, for $N > 2$, in a straightforward way.

Lemma 3.5 *Suppose $\hat{M}^j_{PT}$ is the joint PQMLE-CM test indicator for $N=2$, a robust (to non-normality) consistent estimator of V_1 is given by*

$$\hat{V}^r_{1T} = \frac{1}{T}\hat{W}^{r\prime}\hat{W}^r = \frac{1}{T}\hat{A}^r_1\hat{B}'\hat{B}A^{r\prime}_1,$$

where

$W^r = BA^{r\prime}_1$, $\hat{B} = \left[\hat{R},\ \hat{Q}_1,\ \hat{Q}_2\right]$*; i.e., $\hat{B}$ has rows* $\left(\hat{m}'_{Pt},\ \hat{g}'_{1t},\ \hat{g}'_{2t}\right)$,

$\hat{A}^r_1 = \left[I_r : -\hat{J}^*_{M\varpi T}\hat{J}^{-1}_{\theta\theta}\right]$, *with* $\hat{J}_{\theta\theta} = \begin{bmatrix} \hat{J}_1 & 0 \\ 0 & \hat{J}_2 \end{bmatrix}$, *$\hat{J}_i$ is obtained from (3.27), $\hat{J}^*_{M\varpi T}$ is given in (3.26) and I_r is the identity matrix of rank $r = rank\left(\Sigma_{MM}\right)$.*

Remark 3.4 *For $N=2$, we can write $\hat{A}^r_1$ and $\hat{W}^r$ as*

$$\begin{aligned}
\hat{A}^r_1 &= \left[I_r : -\frac{1}{T}\begin{bmatrix} \hat{\overline{R}}^{*\prime}(\hat{Z}_1, 0) \\ \hat{\rho}\hat{\overline{R}}^{*\prime}(\hat{Z}_1, \hat{Z}_2) \\ \hat{\overline{R}}^{*\prime}(0, \hat{Z}_2) \end{bmatrix}\begin{bmatrix} \hat{J}^{-1}_1 & 0 \\ 0 & \hat{J}^{-1}_2 \end{bmatrix}\right] \\
&= \left[I_r : -\frac{1}{T}\begin{bmatrix} \hat{\overline{R}}^{*\prime}\hat{Z}_1\hat{J}^{-1}_1 & 0 \\ \hat{\rho}\hat{\overline{R}}^{*\prime}\hat{Z}_1\hat{J}^{-1}_1 & \hat{\rho}\hat{\overline{R}}^{*\prime}\hat{Z}_2\hat{J}^{-1}_2 \\ 0 & \hat{\overline{R}}^{*\prime}\hat{Z}_2\hat{J}^{-1}_2 \end{bmatrix}\right],
\end{aligned}$$

and

$$\hat{W}^r = \hat{R} - \frac{1}{T}\left(\hat{Q}_1\hat{J}^{-1}_1\hat{Z}'_1\hat{\overline{R}}^*, \frac{1}{2}\hat{\rho}\left(\hat{Q}_1\hat{J}^{-1}_1\hat{Z}'_1 + \hat{Q}_2\hat{J}^{-1}_2\hat{Z}'_2\right)\hat{\overline{R}}^*, \hat{Q}_2\hat{J}^{-1}_2\hat{Z}'_2\hat{\overline{R}}^*\right),$$

respectively.

Tests based on this estimator $\hat{V}^r_{1T}$ will be referred as the *robust PQMLE test* and will be denoted as $T^{(r)}_P$. Note that exploiting the FOC $\hat{Q}'_i\iota_T \equiv 0$ for $\hat{\theta}_i$, $i = 1, 2$, we have $\hat{W}^{r\prime}\iota_T \equiv \hat{R}'\iota_T$, where ι_T is the $(T \times 1)$ column vector of ones, implying that

$$T^{(r)}_p = T\hat{M}'_{PT}\hat{V}^{r-1}_{1T}\hat{M}_{PT} = \iota'_T\hat{W}^r\left(\hat{W}^{r\prime}\hat{W}^r\right)^{-1}\hat{W}^{r\prime}\iota_T. \qquad (3.28)$$

This corresponds to $T - RSS$ from a regression of ι_T on $\hat{W}^r$.

Case 2(b): The OPG-PQMLE Test

The following Lemma provides an expression for a consistent estimator of the variance-covariance matrix V_1 under normality.

Lemma 3.6 *Assuming that the specification of the log-likelihood for the joint estimation of parameters in Section 3.3.1 is correct (i.e., $\xi_t \sim N(0, I_N)$), $\hat{A}_1\hat{\Sigma}\hat{A}_1' - V_1 = o_p(1)$ where*

$$\hat{A}_1 = \left[I_r : -\tfrac{1}{T}\hat{R}'\hat{S} \begin{pmatrix} \left(\tfrac{1}{T}\hat{Q}_1'\hat{S}_1\right)^{-1} & 0 \\ 0 & \left(\tfrac{1}{T}\hat{Q}_2'\hat{S}_2\right)^{-1} \end{pmatrix} \right],$$

$\hat{\Sigma} = \frac{1}{T}\hat{B}'\hat{B}$, *and*

$\hat{B} = \left[\hat{R},\ \hat{Q}_1,\ \hat{Q}_2\right]$; *i.e., $\hat{B}$ has rows $\left(\hat{m}_{Pt}',\ \hat{g}_{1t}',\ \hat{g}_{2t}'\right)$.*

Hence, V_1 can be consistently estimated by $\hat{V}_{1T} = \dfrac{1}{T}\hat{W}'\hat{W}$ where

$\hat{W} = \hat{B}\hat{A}_1' = \hat{R} - \hat{Q}_1\left(\hat{S}_1'\hat{Q}_1\right)^{-1}\hat{S}_1'\hat{R} - \hat{Q}_2\left(\hat{S}_2'\hat{Q}_2\right)^{-1}\hat{S}_2'\hat{R}.$

Again, the FOC $\hat{Q}_i'\iota_T \equiv 0$ implies $\hat{W}'\iota_T \equiv \hat{R}'\iota_T$ so that

$$T_p = T\hat{M}_{PT}'\hat{V}_{1T}^{-1}\hat{M}_{PT} = \iota_T'\hat{W}\left(\hat{W}'\hat{W}\right)^{-1}\hat{W}'\iota_T, \tag{3.29}$$

which can be obtained as $T - RSS$ from a regression of ι_T on $\hat{W}$.

Summary: The PQMLE tests

Hence, in case of the PQMLE, we again consider the following four test statistics:

5. Robust (to non-normality) FCM test

$$T_p^{j(r)} = T\hat{M}_{pT}^{j(r)\prime}\left(\hat{V}_{1T}^{j(r)}\right)^{-1}\hat{M}_{pT}^{j(r)}. \tag{3.30}$$

6. OPG FCM test

$$T_p^{j} = T\hat{M}_{pT}^{j\prime}\left(\hat{V}_{1T}^{j}\right)^{-1}\hat{M}_{pT}^{j}. \tag{3.31}$$

7. Robust (to non-normality) CCM test

$$T_p^{c(r)} = T\hat{M}_{pT}^{c(r)\prime}\left(\hat{V}_{1T}^{c(r)}\right)^{-1}\hat{M}_{pT}^{c(r)}. \tag{3.32}$$

8. OPG CCM test

$$T_p^{c} = T\hat{M}_{pT}^{c\prime}\left(\hat{V}_{1T}^{c}\right)^{-1}\hat{M}_{pT}^{c}. \tag{3.33}$$

The robust and OPG variance estimator can be obtained using Lemma 3.5 and Lemma 3.6, respectively.

3.4 Analysis of the Tse's LM Test

Tse (2000) proposed a LM test for the multivariate CCC-GARCH model against the alternative that the correlations are changing as functions of the previous standardized residuals, having the form

$$\rho_{ijt} = \rho_{ij} + \tau_{ij}\zeta_{i,t-1}\zeta_{j,t-1} \text{ or } \Gamma_t = \Gamma + \Delta \odot \zeta_{t-1}\zeta_{t-1}', \tag{3.34}$$

where Δ is a symmetric parameter matrix with the leading diagonal elements equal to zero. Note that (3.34) does not define a particular alternative to the CCC as Γ_t is not necessarily a positive definite matrix for all t. Therefore, Silvennoinen and Teräsvirta (2009b) interpreted this as a general misspecification test. Here τ_{ij}, $1 \leq i < j \leq N$ are $\frac{N(N-1)}{2}$ additional parameters in the extended model and the null hypothesis under consideration is $H_0 : \Delta = 0$ or $H_0 : \text{vecl}(\Delta) = 0$. Under this setting Tse proposed the following statistic:

$$\begin{aligned} LM_T &= \widehat{\widetilde{s}}'\left(\widehat{\widetilde{S}}'\widehat{\widetilde{S}}\right)^{-1}\widehat{\widetilde{s}} &(3.35)\\ &= \iota_T'\widehat{\widetilde{S}}\left(\widehat{\widetilde{S}}'\widehat{\widetilde{S}}\right)^{-1}\widehat{\widetilde{S}}'\iota_T, &(3.36)\end{aligned}$$

where $\widehat{\widetilde{s}}$ is the $\left(\left(N' + \frac{N(N-1)}{2}\right) \times 1\right)$ score vector, $\widehat{\widetilde{S}}$ is $\left(T \times N' + \frac{N(N-1)}{2}\right)$ matrix, with rows of partial derivatives of the log likelihood function and ι_T

is the $(T \times 1)$ column vector of ones.[9] Note that (3.36) can be interpreted as $T - RSS$, where RSS is the residual sum of squares from the regression of ι_T on $\widehat{\widetilde{S}}$. Under the usual regularity conditions LM_T is asymptotically distributed as $\chi^2_{\frac{N(N-1)}{2}}$.

It is informative to note that this LM_T can be interpreted as a test of moment condition $\mathrm{E}\left[\mathrm{vecl}\left(\varepsilon_t^* \varepsilon_t^{*\prime} - \Gamma^{-1} | \mathcal{F}_{t-1}\right)\right] = 0$, where ε_t^* is the transformed standardized errors as given in (3.6). This test is based on the FQMLE approach and cannot be implemented directly within the PQMLE framework. We can, however, modify the test indicator in such a way so that the testing procedure based on the PQMLE developed in previous section can be employed (see Section 3.4.2). Again the procedure will be demonstrated in the bivariate context.

In the bivariate case, $\varepsilon_t^* = (\varepsilon_{1t}^*, \varepsilon_{2t}^*)' = \Gamma^{-1}\zeta_t = \dfrac{1}{1-\rho^2}\begin{bmatrix} \zeta_{1t} - \rho\zeta_{2t} \\ \zeta_{2t} - \rho\zeta_{1t} \end{bmatrix}$, and the implicit null of the CCC is $\mathrm{E}\left[\varepsilon_{1t}^*\varepsilon_{2t}^* + \dfrac{\rho}{1-\rho^2} | \mathcal{F}_{t-1}\right] = 0$. Note that

$$
\begin{aligned}
\varepsilon_{1t}^*\varepsilon_{2t}^* + \frac{\rho}{1-\rho^2} &= \frac{1}{(1-\rho^2)^2}\left[\{\zeta_{1t} - \rho\zeta_{2t}\}\{\zeta_{2t} - \rho\zeta_{1t}\} + \rho\left(1-\rho^2\right)\right] \\
&= \frac{1}{(1-\rho^2)^2}\begin{bmatrix} -\rho\left(\zeta_{1t}^2 - 1\right) + (1+\rho^2)\left(\zeta_{1t}\zeta_{2t} - \rho\right) \\ -\rho\left(\zeta_{2t}^2 - 1\right) \end{bmatrix} \\
&= \frac{1}{(1-\rho^2)^2}\pi' v_t
\end{aligned}
$$

where $\pi'(\rho) \equiv \pi' = \left(-\rho, \left(1+\rho^2\right), -\rho\right)$, $v_t' = \left(\zeta_{1t}^2 - 1,\ \zeta_{1t}\zeta_{2t} - \rho,\ \zeta_{2t}^2 - 1\right)$.

3.4.1 The FQMLE Case

Assuming that r_t is a scalar and ignoring the irrelevant factor of proportionality, $1/\left(1-\rho^2\right)^2$, define Tse's "modified" test indicator as

$$
\hat{M}_{FT}^t = \frac{1}{T}\sum_{t=1}^{T} \hat{m}_{Ft}^t\left(\hat{\varpi}\right) = \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}'\hat{v}_t\hat{r}_t = \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}'\hat{m}_{Ft}^j\left(\hat{\varpi}\right), \tag{3.37}
$$

[9] For the full expressions of the first partial derivatives of the likelihood function with respect to the model parameters, see Tse (2000).

where the superscript t represents Tse's indicator and $\hat{m}^{j}_{Ft}(\hat{\varpi})$ is the contribution of t^{th} observation to the test indicator for the FCM test $\hat{M}^{j}_{FT}$, all evaluated at the FQMLE $\hat{\varpi}$.

Corollary 3.1 *From Theorem (3.2),* $\sqrt{T}\hat{M}^{j}_{FT} \xrightarrow{d} N\left(0, V^{j}\right)$, *hence*

$$\sqrt{T}\hat{M}^{t}_{FT} \xrightarrow{d} N\left(0, V^{t}\right),$$

where $V^{t} = \pi' V^{j} \pi$.

Then, an equivalent procedure to the Tse's LM test can be obtained applying the CM testing framework developed in Section 3.3.1 by using $\hat{M}^{t}_{FT}$ and constructing the OPG version of the test as:

$$T^{t}_{F} = T\hat{M}^{t\prime}_{FT}\left(\hat{V}^{t}_{T}\right)^{-1}\hat{M}^{t}_{FT} = T\hat{M}^{t\prime}_{FT}\left(\hat{\pi}'\hat{V}^{j}\hat{\pi}\right)^{-1}\hat{M}^{t}_{FT}, \qquad (3.38)$$

where $\hat{V}^{j} = \hat{\Sigma}_{MM} - \hat{\Sigma}^{*}_{MG}\hat{\Sigma}^{*-1}_{GG}\hat{\Sigma}^{*}_{GM}$ comes from the OPG FCM test given in (3.20).

As noted earlier that, Tse assumed the generalized IM equality to hold while developing his OPG version of the LM test, which may not be robust under non-normality. Using the robust variance estimator of Lemma 3.3, we can now robustify this LM test; i.e.,

$$T^{t(r)}_{F} = T\hat{M}^{t\prime}_{FT}\left(\hat{V}^{t(r)}_{T}\right)^{-1}\hat{M}^{t}_{FT} = T\hat{M}^{t\prime}_{FT}\left(\hat{\pi}\hat{V}^{j(r)}_{T}\hat{\pi}'\right)^{-1}\hat{M}^{t}_{FT}, \qquad (3.39)$$

where $\hat{V}^{t(r)}_{T} = \hat{\pi}\hat{V}^{j(r)}_{T}\hat{\pi}'$ and $\hat{V}^{j(r)}_{T}$ is the variance-covariance matrix defined in (3.19).

3.4.2 The PQMLE Case

For the PQMLE, to obtain asymptotically valid test statistic, we apply the demeaning technique so that the estimation effect from ρ asymptotically negligible (i.e., condition (3.1) holds) in the following way:

$$
\begin{aligned}
\hat{M}_{PT}^{t} &= \frac{1}{T}\sum_{t=1}^{T}\hat{m}_{Pt}^{t}\left(\hat{\theta},\hat{\rho}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{1t}^{*}\hat{\varepsilon}_{2t}^{*} + \frac{\hat{\rho}}{1-\hat{\rho}^{2}}\right)\left(\hat{r}_{t} - \widehat{\overline{r}}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}'\hat{v}_{t}\left(\hat{r}_{t} - \widehat{\overline{r}}\right) = \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}'\hat{m}_{Pt}^{j}\left(\hat{\theta},\hat{\rho}\right), \qquad (3.40)
\end{aligned}
$$

where the last equality follows from (3.24).

Also noting that $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{1t}^{*}\hat{\varepsilon}_{2t}^{*} + \frac{\hat{\rho}}{1-\hat{\rho}^{2}}\right) \neq 0$ unless the FQMLE is employed, we have $\hat{M}_{PT}^{t} \neq \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{1t}^{*}\hat{\varepsilon}_{2t}^{*} + \frac{\hat{\rho}}{1-\hat{\rho}^{2}}\right)\hat{r}_{t}$. Now using the following Corollary and the results of preceding section, we can construct asymptotically valid tests of the CCC assumption employing Tse's modified indicator $\hat{M}_{PT}^{t}$ based on the PQMLE.

Corollary 3.2 *From Theorem (3.3),* $\sqrt{T}\hat{M}_{PT}^{j} \xrightarrow{d} N\left(0, V_{1}^{j}\right)$; *hence*

$$
\sqrt{T}\hat{M}_{PT}^{t} \xrightarrow{d} N\left(0, V_{1}^{t}\right),
$$

where $V_{1}^{t} = \pi' V_{1}^{j}\pi$.

In particular, the OPG and robust test statistics with $\hat{M}_{PT}^{t}$ can be constructed easily by:

$$
T_{P}^{t} = T\hat{M}_{PT}^{t\prime}\left(\hat{V}_{1T}^{t}\right)^{-1}\hat{M}_{PT}^{t} = T\hat{M}_{PT}^{t\prime}\left(\hat{\pi}'\hat{V}_{1T}^{j}\hat{\pi}\right)^{-1}\hat{M}_{PT}^{t}, \qquad (3.41)
$$

and

$$
T_{p}^{t(r)} = T\hat{M}_{PT}^{t\prime}\left(\hat{V}_{1T}^{t(r)}\right)^{-1}\hat{M}_{PT}^{t} = T\hat{M}_{PT}^{t\prime}\left(\hat{\pi}'\hat{V}_{1T}^{j(r)}\hat{\pi}\right)^{-1}\hat{M}_{PT}^{t}, \qquad (3.42)
$$

where $\hat{V}_{1T}^{j(r)}$ and $\hat{V}_{1T}^{j}$ are given in (3.30) and (3.31), respectively.

Alternatively, $\hat{V}_{1T}^{t}$ can be constructed as $\hat{V}_{1T}^{t} = \frac{1}{T}\hat{W}^{t\prime}\hat{W}^{t}$ where $\hat{W}^{t} = \hat{B}\hat{A}_{1}'\hat{\pi}$, and $\hat{B}$ and $\hat{A}_{1}$ are defined as before. Similarly, $\hat{V}_{1T}^{t(r)} = \frac{1}{T}\hat{W}^{t(r)\prime}\hat{W}^{t(r)}$ where $\hat{W}^{t(r)} = \hat{B}\hat{A}_{1}^{r\prime}\hat{\pi}$. And then the test statistics are obtained as $T - RSS$ from a regression of ι_{T} on $\hat{W}^{t}$ or $\hat{W}^{t(r)}$.

Remark 3.5 *Notice that if we demean the test indicator* $\hat{M}^{j}_{PT}$ *only for the components which involve* $\hat{\rho}$ *then, as noted before (Remark 3.3), the OPG-FCM and the robust FCM procedure will be asymptotically valid. However (3.41) or (3.42) cannot be employed. In other words, it is necessary to demean all elements of the moment condition to obtain valid test statistics based on* $\hat{M}^{t}_{PT}$.

3.5 Monte Carlo Evidence

In this section, we present Monte Carlo evidence on finite sample size and power performance of the 12 tests defined in (3.19)-(3.22), (3.30)-(3.33), (3.38), (3.39), (3.41) and (3.42). To recapitulate, we consider three test indicators, namely FCM $\left(\hat{M}^{j}_{T}\right)$, CCM $\left(\hat{M}^{c}_{T}\right)$ and Tse's "modified" indicator $\left(\hat{M}^{t}_{T}\right)$, each having four versions (FQMLE OPG, FQMLE robust, PQMLE OPG and PQMLE robust). Table 3.1 displays various test indicators and associated test statistics under consideration.

Table 3.1: Various test indicators and tests considered in the simulation

Test indicator	Tests
$\hat{M}^{j}_{T}$	T^{j}_{F}, $T^{j(r)}_{F}$, T^{j}_{P} and $T^{j(r)}_{p}$
$\hat{M}^{c}_{T}$	T^{c}_{F}, $T^{c(r)}_{F}$, T^{c}_{P} and $T^{c(r)}_{p}$
$\hat{M}^{t}_{T}$	T^{t}_{F}, $T^{t(r)}_{F}$, T^{t}_{P} and $T^{t(r)}_{p}$

The parameter values for the null and alternative Data Generating Process (DGPs) are taken from the existing literature (e.g., Engle and Ng, 1993; Tse, 2000; Lundbergh and Teräsvirta, 2002; Halunga and Orme, 2009). For each experiment, three series of 1200, 900 and 700 data realizations were generated with the first 200 observations being discarded to avoid initialization effects, yielding sample sizes of $T = 1000$, 700 and 300, respectively. Each model is replicated and estimated, 10,000 times (for size experiments) and 2000 times (for robustness to non-normality and power experiments), both by FQMLE and PQMLE. Next, the above mentioned 12 test statistics are calculated. For this simulation study, we consider the product of the 1-period lagged standardized residuals as the scalar test variable, i.e., $\hat{r}_t = \zeta_{1,t-1}\zeta_{2,t-1}$, to calculate all 12 test statistics. All simulation experiments are conducted in GAUSS programming language.

3.5.1 Size

To assess the size properties of the tests we consider a bivariate AR(1)-CCC-GARCH (1,1) DGP as our null model;[10] viz.,

$$\begin{aligned} y_{it} &= \varphi_{i0} + \varphi_{i1} y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, \\ Var\left(\varepsilon_t | \mathcal{F}_{t-1}\right) &= H_t \Rightarrow \mathrm{E}\left[\varepsilon_{it}^2 | \mathcal{F}_{t-1}\right] = h_{it},\ \varepsilon_t = H_t^{1/2}\left(\varpi\right) \xi_t,\ \xi_t \sim N(0, I), \\ h_{ii,t} &= \alpha_{i0} + \alpha_{i1}\varepsilon_{i,t-1}^2 + \beta_{i1} h_{i,t-1}, \\ H_t &= D_t \Gamma D_t,\ D_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 \\ 0 & \sqrt{h_{2t}} \end{bmatrix} \text{ and } \Gamma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \end{aligned} \tag{3.43}$$

Four experiments are considered and the corresponding true parameter vectors are presented in Table 3.2. E1 and E2 represent models with relatively high persistence ($\alpha + \beta = 0.95$ for E1 and $\alpha + \beta = 0.90$ for E2) while E3 and E4 correspond to relatively low persistence model ($\alpha + \beta = 0.70$ for both E3 and E4). On the other hand, E1 and E3 represent high correlation models and E2 and E4 represent low correlation models. Hence E1, E2, E3 and E4 represent high-persistent-high-correlation, high-persistent-low-correlation, low-persistent-high-correlation and low-persistent-low-correlation specification, respectively.

Table 3.3 reports the actual rejection frequencies when the null of the CCC is true and $\xi_t \sim N(0, I)$. The results are reported for a nominal size of 5%. It can be seen that for low correlation DGPs (E2 and E4), the empirical sizes for all test statistics, except the OPG FCM (T_F^j and T_P^j), are very close to the nominal size of 5%. The OPG version based on the FQMLE of other two tests (T_F^t and T_F^c) slightly over-rejects when $T = 500$, though size property improves as T increases. Interestingly, the PQMLE and robust versions

[10]We also consider a CCC-GARCH (1,1) DGP (i.e., assuming a zero or known conditional mean) to evaluate size property. However, to save space and due to the qualitative similarity of the findings we will discuss only the AR(1)-CCC GARCH (1,1) results.

of all these tests demonstrate better performance even in small sample. In case of experiments with high correlation, particularly with high persistence volatility (E1), all FQMLE-OPG tests (T_F^t, T_F^c and T_F^j) are slightly oversized, robust version of these statistics, however, corrects this size distortion. Tests based on $\hat{M}_T^t$ perform comparatively better for high correlation case.

Our finding that size performance depends on correlation, but volatility persistence does not have much impact on rejection frequencies, is in line with that of Tse (2000). He reports "correlations seem to play a role in determining the rate of convergence to the nominal size. Models with low correlations are less subject to over-rejection in small samples....the persistence of the conditional variance does not have much effect.. " (Tse, 2000 p. 115).

In summary, tests with Tse's modified indicator based on the PQMLE (i.e., T_P^t and $T_P^{t(r)}$) provide the most reliable size property, the robust versions perform better than the OPG (in general), the OPG-FCM tests are slightly oversized and all test statistics perform better in low correlation experiments.

3.5.2 Effect of Non-normality

Table 3.4 presents the actual rejection frequencies when the null of CCC is true and $\xi_t \sim t(6)$, $\xi_t \sim t(8)$ and $\xi_t \sim t(10)$. The inclusion of $t(6)$ offers some evidence on the robustness of the procedure to violations of the underlying moment assumptions (cf. Assumption 3.2.5). First thing to observe that all OPG-FQMLE tests (T_F^t, T_F^c and T_F^j) over-rejects the null under both high and low correlation structure, but more severe in high correlation models. Particularly, note that the Tse's original LM test (T_F^t) is sensitive to the departure from normality assumption. Interestingly, the OPG-PQMLE tests, T_P^t and T_P^c, demonstrate robust size performance under non-normality. The FQMLE-robust version of tests reduce the over-rejection rate considerably and in fact $T_F^{c(r)}$ and $T_F^{c(r)}$ are slightly undersized for low persistent-low correlation model with $t(8)$ and $t(10)$ errors. The empirical size of the robust tests based on Tse's indicator (particularly $T_P^{t(r)}$) and the CCM tests, in general, are close to nominal level of 5% while all versions of the FCM tests show unreliable size property (in general, they are oversized).

Table 3.2: True parameter values for size simulation

	E1	E2	E3	E4
$\varphi_1' = (\varphi_{10}, \varphi_{11})$	$(1.00, 0.10)$			
$\eta_1' = (\alpha_{10}, \alpha_{11}, \beta_{11})$	$(0.01, 0.15, 0.80)$	$(0.01, 0.15, 0.80)$	$(0.40, 0.40, 0.30)$	$(0.40, 0.40, 0.30)$
$\varphi_2' = (\varphi_{20}, \varphi_{21})$	$(1.00, 0.50)$			
$\eta_2' = (\alpha_{20}, \alpha_{21}, \beta_{21})$	$(0.05, 0.20, 0.70)$	$(0.05, 0.20, 0.70)$	$(0.20, 0.50, 0.20)$	$(0.20, 0.50, 0.20)$
ρ	0.80	0.20	0.80	0.20

Note: The parameter vectors $\varphi_1, \eta_1, \varphi_2, \eta_2$ and ρ refer to the AR(1)-CCC-GARCH(1,1) model in (3.43).

Table 3.3: Empirical size at 5 per cent nominal level with the Normal errors

	DGP:E1			DGP:E2			DGP:E3			DGP:E4		
	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500
T_F^t	6.49	6.84	7.42	5.54	5.51	6.31	6.41	6.83	7.00	5.63	5.54	6.65
$T_F^{t(r)}$	5.61	5.37	5.51	4.75	4.33	4.70	5.28	5.47	5.18	4.70	4.46	4.88
T_P^t	5.25	5.02	5.13	5.07	4.65	5.30	4.99	4.80	4.59	5.08	4.89	5.57
$T_P^{t(r)}$	5.01	4.79	4.83	4.66	4.32	4.68	5.18	5.04	4.69	4.70	4.39	4.91
T_F^c	6.43	6.99	6.96	5.76	5.63	6.37	6.37	6.99	7.08	5.74	5.51	6.47
$T_F^{c(r)}$	5.64	5.90	5.58	4.91	4.52	4.68	5.59	5.78	5.71	4.86	4.54	4.78
T_P^c	5.60	6.10	6.05	5.29	4.99	5.67	4.51	4.56	4.13	5.32	4.99	5.89
$T_P^{c(r)}$	5.70	6.37	6.12	4.90	4.50	4.66	5.85	6.49	6.23	4.84	4.52	4.73
T_F^j	7.80	9.02	8.99	6.66	6.81	7.53	7.57	8.75	8.78	6.80	6.83	7.82
$T_F^{j(r)}$	5.85	5.96	5.12	4.92	4.42	4.15	5.29	5.74	5.05	4.76	4.32	4.01
T_P^j	7.07	6.96	6.75	5.71	5.71	6.21	5.73	6.23	5.56	6.21	6.06	6.69
$T_P^{j(r)}$	5.44	5.54	4.96	4.86	4.47	4.16	5.26	5.89	5.08	4.77	4.26	4.03

Notes: 1. Data generated according to the AR(1)-CCC-GARCH(1,1) DGP in (3.43) with parameter values as detailed in Table 3.2. Values in the Table represent the empirical rejection frequencies, against 5% nominal level, of the null of CCC assumption.

2. The first, second and third block correspond to the Tse, CCM and FCM tests respectively (as denoted by the superscript t, c and j). Within each block th first two and last two represent the FQMLE and PQMLE tests (as denoted by subscript F and P), respectively, while the presence of superscript r indicate robust tests.

3. T is the sample size and results are based on 10,000 simulations.

Table 3.4: Empirical size at 5 per cent nominal level under non-normality (with the Student's t errors)

	DGP:E1			DGP:E2			DGP:E3			DGP:E4		
	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500
						$t(6)$						
T_F^t	11.05	12.15	12.65	7.70	7.35	7.40	11.40	10.65	11.80	7.30	8.50	9.10
$T_F^{t(r)}$	7.10	6.70	6.90	5.70	4.65	4.10	7.50	6.25	6.40	4.85	5.00	5.20
T_P^t	5.75	5.70	5.75	6.45	5.70	5.25	5.60	4.90	5.60	5.15	6.15	6.65
$T_P^{t(r)}$	5.40	5.50	5.20	5.65	4.45	4.25	6.05	5.10	5.25	4.85	4.95	5.40
T_F^c	11.60	12.30	10.75	7.80	6.90	7.15	10.05	10.65	12.45	7.00	8.45	9.05
$T_F^{c(r)}$	7.80	7.30	5.70	5.20	4.30	4.15	6.35	6.95	6.85	4.80	5.40	5.55
T_P^c	8.05	7.75	6.80	6.45	5.50	5.75	5.70	5.55	4.65	5.80	6.75	7.50
$T_P^{c(r)}$	8.05	7.55	6.45	5.25	4.30	4.20	7.45	7.55	7.05	4.70	5.45	5.70
T_F^j	16.50	16.55	17.05	11.40	12.15	11.85	15.85	15.10	17.55	9.95	11.05	12.55
$T_F^{j(r)}$	8.05	6.10	6.05	4.65	3.55	3.85	6.45	5.10	5.25	3.45	3.65	3.45
T_P^j	10.45	10.85	9.55	8.80	8.75	8.90	11.05	9.55	8.95	8.35	8.80	10.00
$T_P^{j(r)}$	6.10	4.95	4.35	4.65	3.75	4.20	6.50	4.90	3.90	3.75	3.70	3.45
						$t(8)$						
T_F^t	8.80	9.25	9.90	6.10	6.10	7.20	8.70	9.55	11.90	7.05	7.95	10.05
$T_F^{t(r)}$	6.15	5.80	6.25	4.80	4.15	4.30	6.65	5.40	7.50	5.10	4.95	5.70
T_P^t	5.30	5.65	4.40	5.35	4.75	4.95	5.20	5.05	5.75	5.65	6.30	7.65

Table 3.4: (continued)

	DGP:E1			DGP:E2			DGP:E3			DGP:E4		
	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500
$T_P^{t(r)}$	5.30	5.40	4.05	4.65	4.05	4.25	5.45	4.70	5.90	5.25	4.90	5.65
T_F^c	10.05	9.90	10.25	5.80	6.75	6.70	7.80	9.20	10.05	6.15	7.65	8.80
$T_F^{c(r)}$	7.10	7.05	7.00	4.30	4.85	4.05	6.05	6.50	6.25	4.60	5.40	6.20
T_P^c	7.55	6.80	7.40	4.85	5.70	5.20	4.00	5.30	5.05	5.30	6.60	7.85
$T_P^{c(r)}$	7.65	7.25	7.85	4.20	4.80	4.00	5.80	7.35	7.50	4.45	5.20	6.00
T_F^j	12.60	12.75	14.55	8.70	8.90	10.05	12.95	13.55	15.55	7.80	9.75	11.10
$T_F^{j(r)}$	7.50	6.30	5.55	5.20	3.95	3.80	6.15	5.95	7.20	3.75	4.15	4.05
T_P^j	9.45	8.95	8.00	7.15	7.15	7.20	8.65	8.95	8.40	6.45	7.85	9.40
$T_P^{j(r)}$	5.85	5.25	4.05	5.25	4.10	4.15	4.95	5.60	6.00	3.60	4.05	4.15
$t(10)$												
T_F^t	8.95	8.45	9.30	5.65	5.50	6.85	8.25	8.45	9.40	7.15	7.50	7.20
$T_F^{t(r)}$	6.70	5.55	5.80	4.35	3.95	4.15	6.15	6.05	6.25	5.90	5.55	4.55
T_P^t	6.20	5.00	5.20	4.65	4.55	4.65	5.75	4.90	5.80	6.15	6.35	5.55
$T_P^{t(r)}$	6.45	5.05	5.50	4.45	4.05	4.05	5.25	5.05	5.35	5.90	5.50	4.50
T_F^c	8.30	8.45	10.30	5.55	6.25	5.95	7.80	7.30	8.85	6.35	7.30	6.55
$T_F^{c(r)}$	6.05	5.75	7.05	3.85	4.50	3.95	6.35	5.00	5.65	4.90	5.05	4.40
T_P^c	7.45	6.35	7.75	4.60	5.55	4.75	4.95	4.55	4.65	5.55	6.60	5.55

Table 3.4: (continued)

	DGP:E1			DGP:E2			DGP:E3			DGP:E4		
	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500
$T_P^{c(r)}$	6.80	6.20	7.20	3.85	4.60	3.95	7.05	6.25	6.05	4.85	5.20	4.20
T_F^{j}	11.95	11.90	14.25	7.70	7.60	9.55	11.10	11.20	13.35	8.35	9.70	9.40
$T_F^{j(r)}$	7.50	6.40	6.00	4.40	3.00	4.15	6.25	5.95	5.50	4.10	3.80	3.70
T_P^{j}	8.05	8.35	8.30	6.00	5.70	7.40	8.70	8.00	8.75	7.25	7.95	7.60
$T_P^{j(r)}$	6.35	5.40	5.25	4.35	3.10	4.10	5.95	4.80	5.30	4.10	4.00	3.55

Notes: See the notes 1-3 of Table 3.3, but results are based on 2,000 simulations.

3.5.3 Impact of Misspecified Univariate Volatility

We consider four experiments (M1, M2, M3 and M4) in the regression context to investigate the effect of misspecification in the univariate GARCH model when the true correlation structure is constant. M1 and M3 has low correlation ($\rho = 0.20$) and M2 and M4 follow high correlation ($\rho = 0.80$) structure. The conditional mean parameters are the same as in the size experiments. For M1 and M2, the univariate volatility specification for first variable is given by high persistence GARCH (1,1) model; i.e., $h_{1t} = 0.01 + 0.15\varepsilon_{1,t-1}^2 + 0.80h_{1,t-1}$, while the second variable follows the EGARCH(1,1) model of Nelson (1991) with parameter values considered by Engle and Ng (1993) and Halunga and Orme (2009) in their simulation study: $\log(h_{2,t}) = -0.23 + 0.9\log(h_{2,t-1}) + 0.25\left[\left|\xi_{t-1}\right| - 0.3\xi_{t-1}\right]$. On the other hand, in the experiment M3 and M4, we assume that both variables are subject to volatility spillover (i.e., ECCC model) in the following way:

$$h_{it} = \alpha_{i0} + \alpha_{i1}\varepsilon_{i,t-1}^2 + \beta_{ii}h_{i,t-1} + \beta_{ij}h_{j,t-1},\ i = 1, 2 \text{ and } i \neq j,$$

with

$$\begin{aligned}(\alpha_{10}, \alpha_{11}, \beta_{11}, \beta_{12}) &= (0.01, 0.15, 0.80, 0.02)\,, \text{ and} \\ (\alpha_{20}, \alpha_{21}, \beta_{22}, \beta_{21}) &= (0.05, 0.20, 0.70, 0.03)\,.\end{aligned}$$

Table 3.5 reports the results of the simulation study based on 2000 replications where the data is generated with normal errors and the nominal level of significance is set to 5%. It can be observed that the tests are fairly robust to the volatility spillover case (i.e., M3 and M4). On the other hand these tests seem to be sensitive to the GARCH-EGARCH-High correlation specification (M2) where all tests over-reject the null of CCC. It is to be noted here that due to the fact that the FCM test indicator involve the volatility moment condition, these tests, as expected, display the power to pick the misspecification. For M1 (GARCH-EGARCH-low correlation), the tests, except the FCM, are less sensitive to univariate conditional variance misspecification.

Table 3.5: The effect of univariate volatility misspecification against 5 per cent nominal level

	DGP:M1			DGP:M2			DGP:M3			DGP:M4		
	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500
T_F^t	7.75	8.60	9.21	19.50	20.55	21.36	5.60	6.45	5.65	6.25	7.10	7.70
$T_F^{t(r)}$	4.70	5.05	5.21	14.95	12.95	11.57	4.70	5.00	4.50	5.25	5.90	5.55
T_P^t	5.60	5.95	6.32	10.55	9.65	9.05	5.10	5.65	5.00	4.50	5.15	4.60
$T_P^{t(r)}$	4.80	5.15	5.54	9.25	8.55	8.33	4.70	5.10	4.45	4.95	5.20	4.30
T_F^c	6.45	6.60	6.87	14.85	15.15	15.35	5.70	6.25	6.45	6.90	7.70	7.00
$T_F^{c(r)}$	5.65	4.85	4.41	13.05	12.45	12.05	4.90	5.00	4.70	6.05	6.80	5.65
T_P^c	6.45	6.00	5.70	12.60	12.85	13.67	5.30	5.50	5.75	5.65	6.55	5.90
$T_P^{c(r)}$	5.75	4.90	4.77	14.50	14.20	14.20	4.85	4.90	4.80	5.90	6.60	5.90
T_F^j	12.70	11.95	11.45	61.70	51.80	45.42	6.60	7.05	6.70	9.05	9.40	10.55
$T_F^{j(r)}$	9.05	7.25	6.15	52.50	38.40	29.30	4.70	4.25	4.00	7.00	6.15	7.25
T_P^j	12.35	10.85	9.85	41.65	34.00	28.90	5.30	6.10	5.80	7.50	7.35	7.45
$T_P^{j(r)}$	9.15	7.20	6.10	33.85	27.00	22.45	4.70	4.25	4.20	5.55	6.20	5.55

Notes: 1. The DGP M1 represents AR(1)-GARCH-EGARCH-CCC (low correlation), M2 represents AR(1)-GARCH-EGARCH-CCC (high correlation), M3 represents AR(1)-ECCC (low correlation)-GARCH and M4 represents AR(1)-ECCC (low correlation)-GARCH model. Values in the Table represent the empirical rejection frequencies, against 5% nominal level, of the null of CCC assumption.

2. The first, second and third block correspond to the Tse, CCM and FCM tests respectively (as denoted by the superscript t, c and j). Within each block th first two and last two represent the FQMLE and PQMLE tests (as denoted by subscript F and P), respectively, while the presence of superscript r indicate robust tests.

3. T is the sample size and results are based on 2,000 simulations.

3.5.4 Power Simulation

To examine power, we consider two types of MGARCH models with time varying correlations. The AR(1) specification for the conditional mean function introduced for size simulation is retained. First we assume that the true DGP for conditional variance matrix H_t has the following Engle's (2002) DCC-GARCH(1,1) model:

$$\begin{aligned}
y_{it} &= \varphi_{i0} + \varphi_{i1} y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, \\
Var\left(\varepsilon_t | \mathcal{F}_{t-1}\right) &= H_t \Rightarrow \mathrm{E}\left[\varepsilon_{it}^2 | \mathcal{F}_{t-1}\right] = h_{it}, \\
h_{it} &= \alpha_{i0} + \alpha_{i1} \varepsilon_{i,t-1}^2 + \beta_{i1} h_{i,t-1}, \\
H_t &= D_t \Gamma_t D_t, \; D_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 \\ 0 & \sqrt{h_{2t}} \end{bmatrix}, \\
\Gamma_t &= \left(I \odot Q_t\right)^{-1/2} Q_t \left(I \odot Q_t\right)^{-1/2} = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2}, \\
Q_t &= (1 - \widetilde{\alpha} - \widetilde{\beta}) \overline{Q} + \widetilde{\alpha} \zeta_{t-1} \zeta_{t-1}' + \widetilde{\beta} Q_{t-1}. \qquad (3.44)
\end{aligned}$$

Secondly, we consider the BEKK model of Engle and Kroner (1995) as the true DGP for conditional variance matrix H_t

$$\begin{aligned}
y_{it} &= \varphi_{i0} + \varphi_{i1} y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, \\
Var\left(\varepsilon_t | \mathcal{F}_{t-1}\right) &= H_t, \\
H_t &= C_B + A_B' \left(\varepsilon_{t-1} \varepsilon_{t-1}'\right) A_B + B_B' H_{t-1} B_B. \qquad (3.45)
\end{aligned}$$

Five experiments are considered: P1, P2 and P3 with the DCC DGP and remaining two (P4 and P5) follow the BEKK DGP. The true parameter values for the conditional mean functions of size simulation experiment are maintained for all DGPs. In addition, for the DCC DGPs high persistence individual volatility specification for both variables, as given in E1 and E2, is retained (i.e., $\eta_1' = (0.01, 0.15, 0.80)$ and $\eta_2' = (0.05, 0.20, 0.70)$). The remaining true parameter vectors are given in Table 3.6. The parameter values for the BEKK models are taken from Tse (2000).

Table 3.7 and Table 3.8 present the power results with 2000 replications for the DCC and BEKK DGPs, respectively, where the nominal size is again

Table 3.6: True parameter values for power simulation

	P1	P2	P3	P4	P5
$\widetilde{\alpha}$	0.05	0.10	0.15	-	-
$\widetilde{\beta}$	0.90	0.85	0.80	-	-
$\overline{Q}$		$\begin{bmatrix} 1.00 & 0.60 \\ 0.60 & 1.00 \end{bmatrix}$		-	-
A_B	-	-	-	$\begin{bmatrix} 0.30 & 0.10 \\ 0.10 & 0.30 \end{bmatrix}$	$\begin{bmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{bmatrix}$
B_B	-	-	-	$\begin{bmatrix} 0.60 & 0.20 \\ 0.20 & 0.60 \end{bmatrix}$	$\begin{bmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{bmatrix}$
C_B	-	-	-	$\begin{bmatrix} 0.20 & 0.10 \\ 0.10 & 0.20 \end{bmatrix}$	$\begin{bmatrix} 0.20 & 0.04 \\ 0.04 & 0.20 \end{bmatrix}$

Note: The parameters $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\overline{Q}$ refer to the DCC-GARCH(1,1) model in (3.44). The parameter matrices A_B, B_B and C_B refer to the DEKK(1,1) model in (3.45).

5%. The data is generated assuming normality. As a measure of the variability of the conditional correlation coefficients, we calculate the range (i.e., maximum - minimum) of the conditional correlation coefficients in each simulated sample of T observations. In the last panel we report the maximum and minimum ranges of the true conditional correlation coefficients in the Monte Carlo samples of 2000. We also report the average of estimated correlation parameter over all Monte Carlo samples.

It can be seen when the true DGP is the DCC, P3 has the largest variability in correlations followed by P2 and P1; i.e., variability increases as $\widetilde{\alpha}$ increases and $\widetilde{\beta}$ decreases. In general, the tests based on the Tse's indicator is found to have higher power in all three DCC experiments. However, as the variability in correlation decreases power decreases. The FCM tests also show nice power property. It is to be noted that even with $T = 500$, both the Tse and FCM tests show high power especially in P2 and P3. But the CCM tests lack power considerably, particularly for P1. The OPG-FQMLE tests show higher power, however using the robust and PQMLE versions do not cost much power. In case of the BEKK DGP the conclusions are quite similar to the DCC models. P5 has larger variability in correlation than P4 and the tests also oblige the fact. All tests show excellent power for P5. However, the CCM tests, particularly the PQMLE based CCM tests, lack power for P4.

Table 3.7: Empirical power against 5 per cent nominal level with the DCC DGP

	DGP: P1			DGP: P2			DGP: P3		
	T=1000	T=700	T=500	T=1000	T=700	T=500	T=1000	T=700	T=500
T_F^t	51.15	39.45	31.90	96.35	88.40	79.85	99.90	99.50	96.50
$T_F^{t(r)}$	48.45	36.00	27.70	95.45	86.30	75.25	99.85	99.10	95.50
T_P^t	47.60	35.65	27.40	94.85	84.25	71.90	99.70	98.70	93.30
$T_P^{t(r)}$	47.10	35.10	26.50	95.05	85.15	73.50	99.85	99.10	94.95
T_F^c	16.65	12.20	9.85	60.35	45.75	33.75	92.05	80.00	69.10
$T_F^{c(r)}$	14.70	9.65	6.95	57.15	42.10	29.15	91.15	76.65	63.25
T_P^c	15.60	10.85	7.70	55.70	41.00	30.05	90.55	76.10	62.65
$T_P^{c(r)}$	13.80	9.55	6.55	54.55	38.85	28.00	90.10	75.05	61.65
T_F^j	40.35	30.90	27.25	92.20	81.40	70.15	99.60	97.95	94.20
$T_F^{j(r)}$	34.40	24.20	18.85	89.30	74.70	60.10	99.55	96.90	89.60
T_P^j	36.05	26.65	20.15	88.60	73.95	57.50	99.35	95.65	87.25
$T_P^{j(r)}$	33.60	24.25	18.20	88.80	73.10	57.15	99.60	96.80	89.00
Estimated Correlation (Average over all simulation)									
FQMLE	0.593	0.595	0.596	0.575	0.580	0.666	0.552	0.558	0.557
PQMLE	0.590	0.596	0.591	0.570	0.573	0.571	0.545	0.550	0.547
Range of true correlation in simulated sample									
Average	0.568	0.534	0.503	1.056	1.001	0.947	1.400	1.344	1.286
Max	0.961	0.930	0.876	1.457	1.506	1.455	1.711	1.727	1.728
Min	0.369	0.305	0.279	0.679	0.548	0.462	0.937	0.727	0.579

Note: Data generated according to the AR(1)-DCC-GARCH(1,1) DGP in (3.44) with parameter values as detailed in Table 3.6. Also see notes 2 and 3 of Table 3.5.

Table 3.8: Empirical power against 5 per cent nominal level with the BEKK DGP

	DGP: P4			DGP: P5		
	T=1000	T=700	T=500	T=1000	T=700	T=500
T_F^t	85.50	73.95	64.80	99.85	98.45	95.30
$T_F^{t(r)}$	83.10	69.30	57.00	99.80	97.55	93.35
T_P^t	91.10	76.95	62.80	99.95	99.10	96.05
$T_P^{t(r)}$	89.10	73.75	59.40	99.80	98.75	94.10
T_F^c	72.45	56.75	42.55	100.00	99.50	97.80
$T_F^{c(r)}$	68.35	49.90	34.10	99.95	99.15	95.35
T_P^c	25.35	16.20	10.10	100.00	98.45	94.75
$T_P^{c(r)}$	22.85	15.25	9.65	99.95	97.80	92.25
T_F^j	81.25	66.95	55.10	99.95	99.20	95.75
$T_F^{j(r)}$	74.75	56.85	42.25	99.80	97.60	88.80
T_P^j	81.25	62.40	46.15	100.00	98.95	94.70
$T_P^{j(r)}$	74.80	53.55	38.35	99.85	97.55	88.55
Estimated Correlation (Average over all simulation)						
FQMLE	0.839	0.839	0.840	0.616	0.617	0.617
PQMLE	0.831	0.831	0.831	0.595	0.596	0.596
Range of true correlation in simulated sample						
Average	0.247	0.239	0.231	0.668	0.653	0.637
Max	0.349	0.347	0.321	0.829	0.824	0.802
Min	0.191	0.189	0.178	0.567	0.557	0.537

Note: Data generated according to the AR(1)-BEKK(1,1) DGP in (3.45) with parameter values as detailed in Table 3.6. Also see notes 2 and 3 of Table 3.5.

3.6 Concluding Remarks

In this Chapter, we propose a set of asymptotically valid CM tests of testing the CCC hypothesis for MGARCH model. We develop these tests considering both the FQMLE and PQMLE framework for the CCC model. We want to underscore the point that tests with the PQMLE is nonexistent in the literature. Moreover, the robust and OPG versions of these tests are developed. These tests are very easy to implement. We also analyze and accommodate the Tse's (2000) LM test, which is a FQMLE-OPG type test, within our CM testing framework. Further we provide the PQMLE and robust versions of the Tse's LM test. We examine the finite sample performance of these asymptotically valid tests.

Monte Carlo experiments indicate that in general all tests have desirable size property and the robust version perform better than the OPG version. It is found that the correlation parameter has a significant impact on empirical size of these tests (low correlation is associated with better size property). The size is, however, not affected by the degree of univariate volatility persistence. In case of the departure from normality assumption of true error, the robust versions demonstrate better size as compared to the OPG tests. In particular, all OPG-FQMLE tests are oversized. Interestingly, the PQMLE based tests exhibit more robustness compared to the FQMLE tests. Besides, when the assumption of the null model is violated by assuming misspecified univariate volatility structure but maintaining the CCC assumption, size of these tests are not affected by volatility spillover effect. However when one equation is misspecified and true correlation is high all tests over-reject the null of the CCC assumption. The rejection rate is higher in case of the FCM tests as expected, since by construction these tests consider the individual volatility moment conditions as well.

The power of these tests depends on the variability of the true correlation parameter and it is found that the tests based on Tse's modified indicator and FCM show excellent power, even in models with less dispersed correlations. The CCM tests, in general, show comparatively lower power and particularly in models with less dispersed correlations have limited power. In terms of power there is a very little to choose between the OPG and robust, and between the FQMLE and PQMLE.

To sum up, testing correlation constancy depends on the true correlation parameter and no significant difference is observed whether one use the FQMLE and PQMLE approach. The robust versions manifest better size under non-normality. The FCM tests check the individual volatility along with CCC assumption, hence can be treated as a general diagnostic test. The CCM test has desirable size properties, but lacks power under certain DGPs. The Tse's LM test has good size and power properties but is sensitive to the departure from normality while its OPG-PQMLE display impressive robustness maintaining the high power performance. The robust version of the Tse's test, both the FQMLE and PQMLE, has excellent power and size property.

The tests here are derived for to check the CCC assumption which in many situations may not be a realistic or reasonable one. It is therefore is of interest to devise test of time varying correlation. In practice, a two-stage estimation approach is almost always applied to estimate time varying correlation models indicating the need to develop a testing framework based on the PQMLE approach. However, in this case correlations are not scalar and the simple demeaning technique, that we have used to derive the PQMLE tests in this Chapter, is not possible. In other words we need to consider the estimation effect emerging from correlation parameters to derive the limit distribution of the test statistics. Such extensions, however, left for future research.

Appendix

3.A Appendix A: Proofs

For the CCC we have,

$$corr\left[\varepsilon_{it}, \varepsilon_{jt} | \mathcal{F}_{t-1}\right] = \mathrm{E}\left[\left. \frac{\varepsilon_{it}\varepsilon_{jt}}{\sqrt{h_{it}}\sqrt{h_{jt}}} \right| \mathcal{F}_{t-1} \right] = \rho_{ij}.$$

We have the following definitions:

1. $\zeta_{it} = \varepsilon_{it}/\sqrt{h_{it}}$ is i.i.d. $(0,1)$, for $t = 1, ..., T$, with $\mathrm{E}\left[\zeta_{it}\zeta_{jt}|\mathcal{F}_{t-1}\right] = \rho_{ij}$, or $\mathrm{E}\left[\zeta_t\zeta_t'|\mathcal{F}_{t-1}\right] = \Gamma = \left\{\rho_{ij}\right\}, (N \times N)$, where $\zeta_t = \{\zeta_{it}\}, (N \times 1)$, and $\rho_{ii} \equiv 1$, $\Gamma = \Gamma'$ is symmetric and positive definite.
2. Let $\Gamma^{-1} = \{\rho^{ij}\}$, so that $\sum_{m=1}^{N} \rho^{im}\rho_{mj} = \delta_{ij}$, the Kronecker Delta; i.e., $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$.
3. $\varepsilon_{it}^* = \sum_{m=1}^{N} \rho^{im}\zeta_{mt}$, so that $\varepsilon_t^* = \{\varepsilon_{it}^*\} = \Gamma^{-1}\zeta_t$.
4. $f_{it} = w_{it}/\sqrt{h_{it}}$, $c_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \varphi_i}$, $x_{it} = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \eta_i}$.

Then, in addition to the properties of ζ_{it} listed in (1), we have the following:

$$\begin{aligned}
\mathrm{E}\left[\varepsilon_{it}^*|\mathcal{F}_{t-1}\right] &= 0, \\
\mathrm{E}\left[\varepsilon_t^*\varepsilon_t^{*\prime}|\mathcal{F}_{t-1}\right] &= \Gamma^{-1} = \left\{\rho^{ij}\right\} = \left\{\mathrm{E}\left[\varepsilon_{it}^*\varepsilon_{jt}^*|\mathcal{F}_{t-1}\right]\right\}, \\
\mathrm{E}\left[\varepsilon_t^*\zeta_t'|\mathcal{F}_{t-1}\right] &= I_N = \left\{\mathrm{E}\left[\varepsilon_{it}^*\zeta_{jt}|\mathcal{F}_{t-1}\right]\right\},
\end{aligned}$$

so that, in particular, $\mathrm{E}\left[\varepsilon_{it}^*\zeta_{jt}|\mathcal{F}_{t-1}\right] = \delta_{ij}$.

3.A.1 Proof of Lemma 3.1

To construct the expected Hessian matrix conditional on $\mathcal{F}_{t-1}$, we first obtain the second partial derivatives given $\mathcal{F}_{t-1}$. Since $\varepsilon_t' H_t^{-1}\varepsilon_t = \zeta_t'\Gamma^{-1}\zeta_t = \zeta_t'\varepsilon_t^*$ we can write the likelihood function (2.65) as:

$$
\begin{aligned}
l_t^* &= -\frac{1}{2}\ln|\Gamma| - \frac{1}{2}\sum_{j=1}^{N}\ln h_{jt} - \frac{1}{2}\zeta_t'\Gamma^{-1}\zeta_t \\
&= -\frac{1}{2}\ln|\Gamma| - \frac{1}{2}\sum_{j=1}^{N}\ln h_{jt} - \frac{1}{2}\sum_{j=1}^{N}\zeta_{jt}\varepsilon_{jt}^*.
\end{aligned}
$$

Using the results of Lemma 2.1, and noting that $\varepsilon_{it}^* = \sum_{j=1}^{N}\rho^{ij}\zeta_{jt}$ and $\rho^{ij} = \rho^{ji}$, we have $\dfrac{\partial\varepsilon_{jt}^*}{\partial\varphi_i} = -\rho^{ij}\left(f_{it} + \dfrac{1}{2}\zeta_{it}c_{it}\right)$, and $\dfrac{\partial\varepsilon_{jt}^*}{\partial\eta_i} = -\dfrac{1}{2}\rho^{ij}\zeta_{it}x_{it}$. Then we have:

$$
\begin{aligned}
\frac{\partial l_t^*}{\partial\varphi_i} &= -\frac{1}{2}c_{it} + \frac{1}{2}\left(f_{it} + \frac{1}{2}\zeta_{it}c_{it}\right)\varepsilon_{it}^* - \frac{1}{2}\sum_{j=1}^{N}\zeta_{jt}\frac{\partial\varepsilon_{jt}^*}{\partial\varphi_i} \\
&= f_{it}\varepsilon_{it}^* + \frac{1}{2}c_{it}\left(\zeta_{it}\varepsilon_{it}^* - 1\right). \qquad (3.46)
\end{aligned}
$$

Similarly,

$$
\frac{\partial l_t^*}{\partial\eta_i} = \frac{1}{2}\left(\zeta_{it}\varepsilon_{it}^* - 1\right)x_{it}. \qquad (3.47)
$$

Finally, as in Tse (2000, p. 113) we have:

$$
\frac{\partial l_t^*}{\partial\rho_{ij}} = \varepsilon_{it}^*\varepsilon_{jt}^* - \rho^{ij}, \quad i > j. \qquad (3.48)
$$

To see how (3.48) is derived, note from Magnus and Neudecker (p. 178) we have:

$$
\begin{aligned}
\mathrm{d}\ln|\Gamma| &= |\Gamma|^{-1}\,\mathrm{d}\,|\Gamma| = tr\left(\Gamma^{-1}\,\mathrm{d}\,\Gamma\right) \\
&= \sum_{k=1}^{N}\sum_{m=1}^{N}\rho^{ij}\left(\mathrm{d}\,\rho_{ji}\right) = \sum_{k=1}^{N}\sum_{m=1}^{N}\rho^{ij}\left(\mathrm{d}\,\rho_{ij}\right) \\
&= \sum_{k=1}^{N}\sum_{m\neq 1}^{N}\rho^{ij}\left(\mathrm{d}\,\rho_{ij}\right) = 2\sum_{k=2}^{N}\sum_{m=1}^{k-1}\rho^{ij}\left(\mathrm{d}\,\rho_{ij}\right),
\end{aligned}
$$

where $\Gamma = \left\{\rho_{ij}\right\}$, $\Gamma^{-1} = \left\{\rho^{ij}\right\}$, and where (in line 2) we have used the fact that $\rho_{kk} \equiv 1$ and that $\rho_{km} = \rho_{mk}$. Thus

$$\frac{\partial \ln |\Gamma|}{\partial \rho_{ij}} = 2\rho^{ij}, \quad i > j.$$

Furthermore, since $\varepsilon^*_{kt} = \sum_{m=1}^{N} \rho^{km} \zeta_{mt}$ we can write

$$\frac{\partial \varepsilon^*_{kt}}{\partial \rho_{ij}} = \sum_{m=1}^{N} \frac{\partial \rho^{km}}{\partial \rho_{ij}} \zeta_{mt}.$$

Note from Magnus and Neudecker (p. 183) we have (the differential) $\mathrm{d}\,\Gamma^{-1} = -\Gamma^{-1} (\mathrm{d}\,\Gamma)\,\Gamma^{-1}$. Then looking at the elements we see (with $\Gamma = \{\rho_{ij}\}$, $\Gamma^{-1} = \{\rho^{ij}\}$)

$$\begin{aligned}
\mathrm{d}\,\rho^{km} &= -\sum_{r=1}^{N}\sum_{s=1}^{N} \rho^{kr} (\mathrm{d}\,\rho_{rs})\,\rho^{sm} \\
&= -\sum_{r=1}^{N}\sum_{s\neq r}^{N} \rho^{kr} (\mathrm{d}\,\rho_{rs})\,\rho^{sm} \\
&= -\sum_{r=2}^{N}\sum_{s=1}^{r-1} \rho^{kr} (\mathrm{d}\,\rho_{rs})\,\rho^{sm} - \sum_{s=2}^{N}\sum_{r=1}^{s-1} \rho^{kr} (\mathrm{d}\,\rho_{rs})\,\rho^{sm} \\
&= -\sum_{r=2}^{N}\sum_{s=1}^{r-1} \rho^{kr} (\mathrm{d}\,\rho_{rs})\,\rho^{sm} - \sum_{r=2}^{N}\sum_{s=1}^{s-1} \rho^{ks} (\mathrm{d}\,\rho_{rs})\,\rho^{rm},
\end{aligned}$$

where (in line 2) we have used the fact that $\rho_{rr} \equiv 1$ and (in line 4) that $\rho_{rs} = \rho_{sr}$. Thus,

$$\begin{aligned}
\frac{\partial \rho^{km}}{\partial \rho_{ij}} &= -\rho^{ki}\rho^{jm} - \rho^{kj}\rho^{im} \\
&= -\rho^{ki}\rho^{mj} - \rho^{kj}\rho^{mi}.
\end{aligned}$$

Then,

$$
\begin{aligned}
\frac{\partial \varepsilon_{kt}^{*}}{\partial \rho_{ij}} &= \sum_{m=1}^{N} \frac{\partial \rho^{km}}{\partial \rho_{ij}} \zeta_{mt} \\
&= -\sum_{m=1}^{N} \left(\rho^{ki} \rho^{jm} + \rho^{kj} \rho^{im} \right) \zeta_{mt} \\
&= -\rho^{ki} \varepsilon_{jt}^{*} - \rho^{kj} \varepsilon_{it}^{*}.
\end{aligned}
$$

Now we have,

$$
\begin{aligned}
\frac{\partial l_t^{*}}{\partial \rho_{ij}} &= -\rho^{ij} - \frac{1}{2} \sum_{k=1}^{N} \zeta_{kt} \frac{\partial \varepsilon_{kt}^{*}}{\partial \rho_{ij}} \\
&= \frac{1}{2} \sum_{k=1}^{N} \zeta_{kt} \left(\rho^{ki} \varepsilon_{jt}^{*} + \rho^{kj} \varepsilon_{it}^{*} \right) - \rho^{ij} \\
&= \frac{1}{2} \left(\varepsilon_{jt}^{*} \sum_{k=1}^{N} \rho^{ik} \zeta_{kt} + \varepsilon_{it}^{*} \sum_{k=1}^{N} \rho^{jk} \zeta_{kt} \right) - \rho^{ij} \\
&= \frac{1}{2} \left(\varepsilon_{jt}^{*} \varepsilon_{it}^{*} + \varepsilon_{it}^{*} \varepsilon_{jt}^{*} \right) - \rho^{ij} \\
&= \varepsilon_{it}^{*} \varepsilon_{jt}^{*} - \rho^{ij}.
\end{aligned}
$$

Recall that $\mathrm{E}\left[\varepsilon_{it}^{*} | \mathcal{F}_{t-1}\right] = 0$, $\mathrm{E}\left[\zeta_{it} \varepsilon_{it}^{*} | \mathcal{F}_{t-1}\right] = 1$ and $\mathrm{E}\left[\varepsilon_{it}^{*} \varepsilon_{jt}^{*} | \mathcal{F}_{t-1}\right] = \rho^{ij}$, so that each of the above scores has zero mean. Note that $\partial f_{it}/\partial \varphi_i$, $\partial f_{it}/\partial \eta_i$, $\partial c_{it}/\partial \varphi_i$, $\partial c_{it}/\partial \eta_i$, $\partial x_{it}/\partial \varphi_i$, and $\partial x_{it}/\partial \eta_i$ are all $\mathcal{F}_{t-1}$ measurable. Also, $\Gamma = \left\{\rho_{ij}\right\}$ and $\Gamma^{-1} = \left\{\rho^{ij}\right\}$, so that $\sum_{m=1}^{N} \rho^{im} \rho_{mj} = \delta_{ij}$,where $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$. And $\varepsilon_t^{*} = \{\varepsilon_{it}^{*}\} = \Gamma^{-1} \zeta_t$; hence $\varepsilon_{it}^{*} = \sum_{m=1}^{N} \rho^{im} \zeta_{mt}$.

1. Differentiating (3.46) with respect to φ_j, we obtain

$$
\begin{aligned}
\frac{\partial^2 l_t^{*}}{\partial \varphi_i \partial \varphi_j'} &= \delta_{ij} \frac{\partial f_{it}}{\partial \varphi_j'} \varepsilon_{it}^{*} + f_{it} \frac{\partial \varepsilon_{it}^{*}}{\partial \varphi_j'} + \frac{1}{2} \delta_{ij} \frac{\partial c_{it}}{\partial \varphi_j'} \left(\zeta_{it} \varepsilon_{it}^{*} - 1 \right) \\
&\quad + \frac{1}{2} \delta_{ij} c_{it} \varepsilon_{it}^{*} \frac{\partial \zeta_{it}}{\partial \varphi_j'} + \frac{1}{2} c_{it} \zeta_{it} \frac{\partial \varepsilon_{it}^{*}}{\partial \varphi_j'} \\
&= f_{it} \frac{\partial \varepsilon_{it}^{*}}{\partial \varphi_j'} + \frac{1}{2} \delta_{ij} c_{it} \varepsilon_{it}^{*} \frac{\partial \zeta_{it}}{\partial \varphi_j'} + \frac{1}{2} c_{it} \zeta_{it} \frac{\partial \varepsilon_{it}^{*}}{\partial \varphi_j'} + \varkappa_{1t},
\end{aligned}
$$

where $\varkappa_{1t} = \delta_{ij} \frac{\partial f_{it}}{\partial \varphi_j} \varepsilon_{it}^* + \frac{1}{2} \delta_{ij} \frac{\partial c_{it}}{\partial \varphi_j'} \left(\zeta_{it} \varepsilon_{it}^* - 1 \right)$ so that $\mathrm{E}_0 \left[\varkappa_t | \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0} = 0$. Hence,

$$\begin{aligned} \frac{\partial^2 l_t^*}{\partial \varphi_i \partial \varphi_j'} &= -\rho^{ij} f_{it} \left(f_{jt}' + \frac{1}{2} \zeta_{jt} c_{jt}' \right) - \frac{1}{2} \delta_{ij} c_{it} \varepsilon_{it}^* \left(f_{jt}' + \frac{1}{2} \zeta_{jt} c_{jt}' \right) \\ &\quad - \frac{1}{2} \rho^{ij} c_{it} \zeta_{it} \left(f_{jt}' + \frac{1}{2} \zeta_{jt} c_{jt}' \right) + \varkappa_{1t} \\ &= -\rho^{ij} f_{it} f_{jt}' - \frac{1}{4} \delta_{ij} c_{it} c_{jt}' - \frac{1}{4} \rho^{ij} \rho_{ij} c_{it} c_{jt}' + \varkappa_{1t}^* \text{ (say)}, \end{aligned}$$

where $\mathrm{E}_0 \left[\varkappa_{1t}^* | \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0} = 0$. Thus we can write

$$\widetilde{\mathcal{H}}_{\varphi_i \varphi_j}^* = -\rho^{ij} f_{it} f_{jt}' - \frac{1}{4} \left(\delta_{ij} + \rho^{ij} \rho_{ij} \right) c_{it} c_{jt}', \tag{3.49}$$

such that $\mathrm{E}_0 \left[\widetilde{\mathcal{H}}_{\varphi_i \varphi_j}^* (\varpi_0) \right] = \mathrm{E}_0 \left[\left. \frac{\partial^2 l_t^*}{\partial \varphi_i \partial \varphi_j'} \right| \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0}$.

2. Differentiating (3.47) with respect to η_j, we obtain

$$\begin{aligned} \frac{\partial^2 l_t^*}{\partial \varphi_i \partial \eta_j'} &= \delta_{ij} \frac{\partial f_{it}}{\partial \eta_j'} \varepsilon_{it}^* + f_{it} \frac{\partial \varepsilon_{it}^*}{\partial \eta_j'} + \frac{1}{2} \delta_{ij} \frac{\partial c_{it}}{\partial \eta_j'} \left(\zeta_{it} \varepsilon_{it}^* - 1 \right) \\ &\quad + \frac{1}{2} \delta_{ij} c_{it} \varepsilon_{it}^* \frac{\partial \zeta_{it}}{\partial \eta_j'} + \frac{1}{2} c_{it} \zeta_{it} \frac{\partial \varepsilon_{it}^*}{\partial \eta_j'} \\ &= f_{it} \frac{\partial \varepsilon_{it}^*}{\partial \eta_j'} + \frac{1}{2} \delta_{ij} c_{it} \varepsilon_{it}^* \frac{\partial \zeta_{it}}{\partial \eta_j'} + \frac{1}{2} c_{it} \zeta_{it} \frac{\partial \varepsilon_{it}^*}{\partial \eta_j'} + \varkappa_{2t}, \end{aligned}$$

where $\varkappa_{2t} = \delta_{ij} \frac{\partial f_{it}}{\partial \eta_j} \varepsilon_{it}^* + \frac{1}{2} \delta_{ij} \frac{\partial c_{it}}{\partial \eta_j'} \left(\zeta_{it} \varepsilon_{it}^* - 1 \right)$ so that $\mathrm{E}_0 \left[\varkappa_{2t} | \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0} = 0$. Therefore we have,

$$\begin{aligned} \frac{\partial^2 l_t^*}{\partial \varphi_i \partial \eta_j'} &= -\frac{1}{2} \rho^{ij} f_{it} \zeta_{jt} x_{jt}' - \frac{1}{4} \delta_{ij} c_{it} \varepsilon_{it}^* \zeta_{jt} x_{jt}' - \frac{1}{4} \rho^{ij} c_{it} \zeta_{it} \zeta_{jt} x_{jt}' + \varkappa_{2t} \\ &= -\frac{1}{4} \delta_{ij} c_{it} x_{jt}' - \frac{1}{4} \rho^{ij} \rho_{ij} c_{it} x_{jt}' + \varkappa_{2t}^* \text{ (say)}, \end{aligned}$$

where $\mathrm{E}_0 \left[\varkappa_{2t}^* | \mathcal{F}_{t-1} \right]_{\varpi=\varpi_0} = 0$. Thus, similarly

$$\widetilde{\mathcal{H}}^*_{\varphi_i \eta_j} = -\frac{1}{4} \left(\delta_{ij} + \rho^{ij} \rho_{ij} \right) c_{it} x'_{jt}. \tag{3.50}$$

3. Differentiating (3.48) with respect to φ_k yields

$$\begin{aligned} \frac{\partial^2 l_t^*}{\partial \rho_{ij} \partial \varphi'_k} &= \varepsilon_{jt}^* \frac{\partial \varepsilon_{it}^*}{\partial \varphi'_k} + \varepsilon_{it}^* \frac{\partial \varepsilon_{jt}^*}{\partial \varphi'_k} \\ &= -\rho^{ik} \varepsilon_{jt}^* \left(f_{kt} + \frac{1}{2} \zeta_{kt} c_{kt} \right) - \rho^{jk} \varepsilon_{it}^* \left(f_{kt} + \frac{1}{2} \zeta_{kt} c_{kt} \right) \\ &= -\frac{1}{2} \delta_{jk} \rho^{ik} c_{kt} - \frac{1}{2} \delta_{ik} \rho^{jk} c_{kt} + \varkappa_{3t}, \end{aligned} \tag{3.51}$$

where $\mathrm{E}_0 \left[\varkappa_{3t} | \mathcal{F}_{t-1} \right]_{\varpi=\varpi_0} = 0$ and we have

$$\widetilde{\mathcal{H}}^*_{\varphi_i \rho_{ij}} = -\frac{1}{2} \delta_{jk} \rho^{ik} c_{kt} - \frac{1}{2} \delta_{ik} \rho^{jk} c_{kt}.$$

Note that $i > j$, here, so that

$$\widetilde{\mathcal{H}}^*_{\varphi_i \rho_{ij}} = \begin{cases} -\frac{1}{2} \rho^{ji} c_{it} = -\frac{1}{2} \rho^{ij} c_{it}, & k = i > j \\ -\frac{1}{2} \rho^{ij} c_{jt}, & k = j < i \\ 0, & k \neq i, k \neq j. \end{cases}$$

4. Differentiating (3.47) with respect to η_j, we obtain

$$\begin{aligned} \frac{\partial^2 l_t^*}{\partial \eta_i \partial \eta'_j} &= \frac{1}{2} \delta_{ij} \frac{\partial x_{it}}{\partial \eta'_j} \left(\zeta_{it} \varepsilon_{it}^* - 1 \right) + \frac{1}{2} \delta_{ij} x_{it} \varepsilon_{it}^* \frac{\partial \zeta_{it}}{\partial \eta'_j} + \frac{1}{2} x_{it} \zeta_{it} \frac{\partial \varepsilon_{it}^*}{\partial \eta'_j} \\ &= \frac{1}{2} \delta_{ij} x_{it} \varepsilon_{it}^* \frac{\partial \zeta_{it}}{\partial \eta'_j} + \frac{1}{2} x_{it} \zeta_{it} \frac{\partial \varepsilon_{it}^*}{\partial \eta'_j} + \varkappa_{4t} \\ &= -\frac{1}{4} \delta_{ij} x_{it} \varepsilon_{it}^* \zeta_{jt} x'_{jt} - \frac{1}{4} \rho^{ij} x_{it} \zeta_{it} \zeta_{jt} x'_{jt} + \varkappa_{4t} \\ &= -\frac{1}{4} \delta_{ij} x_{it} x'_{jt} - \frac{1}{4} \rho^{ij} \rho_{ij} x_{it} x'_{jt} + \varkappa_{4t}, \end{aligned}$$

where $\varkappa_{4t} = \frac{1}{2}\delta_{ij}\frac{\partial x_{it}}{\partial \eta_j'}\left(\zeta_{it}\varepsilon_{it}^* - 1\right)$ so that $\mathrm{E}_0\left[\varkappa_{4t}|\mathcal{F}_{t-1}\right]_{\varpi=\varpi_0} = 0$. Thus

$$\widetilde{\mathcal{H}}_{\eta_i\eta_j}^* = -\frac{1}{4}\left(\delta_{ij} + \rho^{ij}\rho_{ij}\right)x_{it}x_{jt}'. \tag{3.52}$$

5. Differentiating (3.48) with respect to η_k yields

$$\begin{aligned}\frac{\partial^2 l_t^*}{\partial\rho_{ij}\partial\eta_k'} &= \varepsilon_{jt}^*\frac{\partial\varepsilon_{it}^*}{\partial\eta_k'} + \varepsilon_{it}^*\frac{\partial\varepsilon_{jt}^*}{\partial\eta_k'} \\ &= -\frac{1}{2}\rho^{ik}\varepsilon_{jt}^*\zeta_{kt}x_{kt} - \frac{1}{2}\rho^{jk}\varepsilon_{it}^*\zeta_{kt}x_{kt} \\ &= -\frac{1}{2}\delta_{jk}\rho^{ik}x_{kt} - \frac{1}{2}\delta_{ik}\rho^{jk}x_{kt}.\end{aligned} \tag{3.53}$$

Note that $i > j$, here, so that

$$\widetilde{\mathcal{H}}_{\eta_i\rho_{ij}}^* = \begin{cases} -\frac{1}{2}\rho^{ji}x_{it} = -\frac{1}{2}\rho^{ij}x_{it}, & k = i > j, \\ -\frac{1}{2}\rho^{ij}x_{jt}, & k = j < i, \\ 0, & k \neq i, k \neq j. \end{cases}$$

6. Differentiating (3.48), $\frac{\partial l_t^*}{\partial\rho_{ij}} = \varepsilon_{it}^*\varepsilon_{jt}^* - \rho^{ij}$, $i > j$, with respect to ρ_{km} yields

$$\begin{aligned}\frac{\partial^2 l_t^*}{\partial\rho_{ij}\partial\rho_{km}} &= \frac{\partial\varepsilon_{it}^*}{\partial\rho_{km}}\varepsilon_{jt}^* + \varepsilon_{it}^*\frac{\partial\varepsilon_{jt}^*}{\partial\rho_{km}} - \frac{\partial\rho^{ij}}{\partial\rho_{km}} \\ &= -\rho^{ik}\varepsilon_{mt}^*\varepsilon_{jt}^* - \rho^{im}\varepsilon_{kt}^*\varepsilon_{jt}^* - \rho^{jk}\varepsilon_{mt}^*\varepsilon_{it}^* - \rho^{jm}\varepsilon_{kt}^*\varepsilon_{it}^* \\ &\quad +\rho^{ik}\rho^{jm} + \rho^{im}\rho^{jk},\end{aligned}$$

where we have used the previous results: $\frac{\partial\rho^{ij}}{\partial\rho_{km}} = -\rho^{ik}\rho^{jm} - \rho^{im}\rho^{jk}$ and

$\dfrac{\partial \varepsilon_{it}^*}{\partial \rho_{km}} = -\rho^{ik}\varepsilon_{mt}^* - \rho^{im}\varepsilon_{kt}^*$. We thus obtain, using symmetry,

$$\begin{aligned}
\widetilde{\mathcal{H}}^*_{\rho_{ij}\rho_{km}} &= -\rho^{ik}\rho^{mj} - \rho^{im}\rho^{kj} - \rho^{jk}\rho^{mi} \\
&\quad -\rho^{jm}\rho^{ki} + \rho^{ik}\rho^{jm} + \rho^{im}\rho^{jk} \\
&= -\rho^{jk}\rho^{mi} - \rho^{jm}\rho^{ki} \\
&= -\rho^{ik}\rho^{jm} - \rho^{im}\rho^{jk} = \frac{\partial \rho^{ij}}{\partial \rho_{km}}. \qquad (3.54)
\end{aligned}$$

3.A.2 Proof of Lemma 3.2

First we define the following:

$\rho = \text{vecl}(\Gamma) = \{\rho_{ij}\}$, $j = 1, ..., N-1$, $i = j+1, ..., N$ (i.e., the i subscript changes more quickly than the j subscript).

For the i^{th} variable, define $\underset{(T\times K)}{F_i}$, $\underset{(T\times K)}{C_i}$ and $\underset{(T\times K')}{X_i}$ with rows $f_{it}' = \dfrac{w_{it}'}{\sqrt{h_{it}}}$, $c_{it}' = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \varphi_i'}$ and $x_{it}' = \dfrac{1}{h_{it}}\dfrac{\partial h_{it}}{\partial \eta_i'}$, respectively. Then define

$$\begin{aligned}
\underset{(NT\times NK)}{F} &= diag\left(F_i\right), \\
\underset{(T\times NK)}{\tilde{F}} &= [F_1, F_2, ..., F_N] \\
\underset{(N\times NK)}{F_t'} &= diag\left(f_{it}'\right) fort = 1, \cdots, T.
\end{aligned}$$

In a similar way, define C, X, $\tilde{C}$, $\tilde{X}$. C_t' and X_t'. Also define $\underset{(N\times N)}{E_t} = diag\left(\zeta_{it}\right)$, and the $(N \times T)$ matrices $E = \{\zeta_{it}\}$ and $E^* = \{\varepsilon_{it}^*\} = \Gamma^{-1}E$ having columns ζ_t and ε_t^*, respectively. It will be useful to define $\Gamma_A = I_N + (\Gamma^{-1} \odot \Gamma)$.

Let ρ^k be the k^{th} column of Γ^{-1}, define $\Gamma^k = \Gamma^{-1} diag\left(\tau_k\right)$, where $\tau_k = \{\delta_{ik}\}$, $(N \times 1)$, $i = 1, ..., N$; i.e., Γ^k be the $(N \times N)$ matrix of zeros, except for column k which is ρ^k. Define the following two $(N \times N)$ symmetric matrices:

$$\begin{aligned}
P_k &= \Gamma^k + \left(\Gamma^k\right)', \\
\Gamma_{km} &= \rho^k \left(\rho^m\right)' + \rho^m \left(\rho^k\right)'.
\end{aligned}$$

Note that $\tilde{F} = (\iota_N' \otimes I_T) F$, so that $\tilde{F}'\tilde{F} = F'(\iota_N \otimes I_T)(\iota_N' \otimes I_T)\ F = F'(\mathcal{J}_N \otimes I_T) F$. Now

1.

$$\mathrm{E}_0\left[\sum_{t=1}^{T}\frac{\partial^2 l_t}{\partial\varphi_i\partial\varphi_j'}\right|\left.\mathcal{F}_{t-1}\right]_{\varpi=\varpi_0} = -\rho^{ij}\sum_{t=1}^{T} f_{it}f_{jt}' - \frac{1}{4}\left(\delta_{ij}+\rho^{ij}\rho_{ij}\right)\sum_{t=1}^{T} c_{it}c_{jt}'$$
$$= -\rho^{ij}F_i'F_j - \frac{1}{4}\left(\delta_{ij}+\rho^{ij}\rho_{ij}\right)C_i'C_j.$$

Hence,

$$\mathrm{E}_0\left[\sum_{t=1}^{T}\frac{\partial^2 l_t}{\partial\varphi\partial\varphi'}\right|\left.\mathcal{F}_{t-1}\right]_{\varpi=\varpi_0} = -\left(\Gamma^{-1}\otimes J_K\right)\odot\tilde{F}'\tilde{F} - \frac{1}{4}\left(\Gamma_A\otimes J\right)\odot\tilde{C}'\tilde{C}$$
$$= -\left(\Gamma^{-1}\otimes J_K\right)\odot F'\left(\mathcal{J}_N\otimes I_T\right)F$$
$$-\frac{1}{4}\left(\Gamma_A\otimes\mathcal{J}_K\right)\odot C'\left(J_N\otimes I_T\right)C$$
$$= -F'\left(\Gamma^{-1}\otimes I_T\right)F - \frac{1}{4}C'\left(\Gamma_A\otimes I_T\right)C.$$

2. Similarly,

$$\mathrm{E}_0\left[\sum_{t=1}^{T}\frac{\partial^2 l_t}{\partial\varphi\partial\eta'}\right|\left.\mathcal{F}_{t-1}\right]_{\varpi=\varpi_0} = -\frac{1}{4}\left(\Gamma_A\otimes J_K\right)\odot\tilde{C}'\tilde{X}$$
$$= -\frac{1}{4}\left(\Gamma_A\otimes J_K\right)\odot C'\left(J_N\otimes I_T\right)X$$
$$= -\frac{1}{4}C'\left(\Gamma_A\otimes I_T\right)X.$$

3.

$$\mathrm{E}_0\left[\sum_{t=1}^{T}\frac{\partial^2 l_t}{\partial\eta\partial\eta'}\right|\left.\mathcal{F}_{t-1}\right]_{\varpi=\varpi_0} = -\frac{1}{4}\left(A\otimes J_K\right)\odot\tilde{X}'\tilde{X}$$
$$= -\frac{1}{4}X'\left(A\otimes I_T\right)X.$$

4. For $i > j$,

$$\mathrm{E}_0 \left[\sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \rho_{ij} \partial \varphi_k'} \middle| \mathcal{F}_{t-1} \right]_{\varpi=\varpi_0} = -\frac{1}{2} \left(\delta_{ik}\rho^{jk} + \delta_{jk}\rho^{ik} \right) \iota_T' C_k$$
$$= -\frac{1}{2} \left(\delta_{ik}\rho^{kj} + \rho^{ik}\delta_{kj} \right) \iota_T' C_k.$$

Then, the matrix with typical element $\rho^{ik}\delta_{kj}$ is $\rho^k \tau_k'$, where $\tau_k = \{\delta_{ik}\}$, $i = 1, ..., N$. $(N \times N)$, ρ^k is the k^{th} column of Γ^{-1}, and e_k is the k^{th} column of I_N, $k = 1, ..., N$. Similarly, $\left(\rho^k e_k'\right)' = \left\{\delta_{ik}\rho^{kj}\right\}$. Alternatively, let Γ^k be the $(N \times N)$ matrix of zeros, except for column k which is ρ^k, the k^{th} column of Γ^{-1}; i.e., $\Gamma^k = \Gamma^{-1} diag(\tau_k)$. Define the symmetric matrix $P_k = \Gamma^k + \left(\Gamma^k\right)'$, $p_k = vecl(P_k)$, and $R_k = \iota_T p_k'$. Then, since $i > j$,

$$\mathrm{E}_0 \left[\sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \rho \partial \varphi_k'} \middle| \mathcal{F}_{t-1} \right]_{\varpi=\varpi_0} = -\frac{1}{2} \mathrm{vecl}\left(P_k\right) \iota_T' C_k$$
$$= -\frac{1}{2} p_k \iota_T' C_k.$$

Collecting the k blocks together we get

$$\mathrm{E}_0 \left[\sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \rho \partial \varphi'} \middle| \mathcal{F}_{t-1} \right]_{\varpi=\varpi_0} = -\frac{1}{2} P' \left(I_N \otimes \iota_T'\right) C$$

where $\underset{N \times \frac{N(N-1)}{2}}{P}$ has rows $p_k' = \mathrm{vecl}(P_k)'$, $k = 1, ..., N$.

5. In a similar way to the previous result,

$$\mathrm{E}_0 \left[\sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \rho \partial' \eta} \middle| \mathcal{F}_{t-1} \right]_{\varpi=\varpi_0} = -\frac{1}{2} P' \left(I_N \otimes \iota_T'\right) X.$$

6.

$$\mathrm{E}_0 \left[\left. \frac{\partial^2 l_t}{\partial \rho_{ij} \partial \rho_{km}} \right| \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0} = -\rho^{ik}\rho^{jm} - \rho^{im}\rho^{jk}$$
$$= -\rho^{ik}\rho^{mj} - \rho^{im}\rho^{kj}.$$

Now, the $(N \times N)$ matrix with typical $(i,j)^{th}$ element equal to $\rho^{ik}\rho^{mj}$ is $\rho^k (\rho^m)'$ and that with typical $(i,j)^{th}$ element equal to $\rho^{im}\rho^{kj}$ is $\rho^m \left(\rho^k\right)'$. Let $\Gamma_{km} = \rho^k (\rho^m)' + \rho^m \left(\rho^k\right)'$, and let $\widetilde{p}_{km} = \mathrm{vecl}\,(\Gamma_{km})$. Then

$$\mathrm{E}_0 \left[\left. \frac{\partial^2 l_t}{\partial \rho \partial \rho_{km}} \right| \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0} = -\widetilde{p}_{km}$$

or

$$\mathrm{E}_0 \left[\left. \frac{\partial^2 l_t}{\partial \rho \partial \rho'} \right| F_{t-1} \right]_{\varpi = \varpi_0} = -\widetilde{P}$$
$$\mathrm{E}_0 \left[\sum_{t=1}^{T} \left. \frac{\partial^2 l_t}{\partial \rho \partial \rho'} \right| \mathcal{F}_{t-1} \right]_{\varpi = \varpi_0} = -T\widetilde{P}$$

where $\widetilde{P} = [\widetilde{p}_{21}, \widetilde{p}_{31}, ..., \widetilde{p}_{N,N-1}]$, a matrix with columns $\widetilde{p}_{km} = \mathrm{vecl}\,(\Gamma_{km})$, $m = 1, ..., N-1$, $k = m+1, ..., N$ (k changes more quickly than m).

3.A.3 Proof of Theorem 3.2

Proof. The test indicator under consideration is $\hat{M}_{FT} \equiv T^{-1} \sum_{t=1}^{T} \hat{m}_{Ft}$. By the consistency of $\hat{\varpi}$ we have, $\sqrt{T}(\hat{\varpi} - \varpi_0) = J_{\varpi\varpi}^{*-1} \sqrt{T} G_{0T}^{*} + o_p(1)$ where $J_{\varpi\varpi}^{*} = -\mathrm{E}_0 [\mathcal{H}_t^{*}(\varpi_0)] = \mathrm{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \frac{\partial g_{0t}^{*}}{\partial \varpi'}$ and assuming (by a ULLN) $J_{M\varpi}^{*} = \mathrm{plim}_{T\to\infty} J_{M\varpi T}^{*}$ where $J_{M\varpi T}^{*} = -T^{-1} \sum_{t=1}^{T} \frac{\partial m_{0Ft}}{\partial \varpi}$. Taking a mean value expansion of $\hat{M}_{FT}$ about $\varpi_0 = (\theta_0', \rho_0')'$,

$$\sqrt{T}\hat{M}_{FT} = \sqrt{T} M_{0FT} - \overline{J}_{M\varpi T}^{*} \sqrt{T}(\hat{\varpi} - \varpi_0)$$
$$= \sqrt{T} M_{0FT} - J_{M\varpi}^{*} J_{\varpi\varpi}^{*-1} \sqrt{T} G_{0T}^{*} + o_p(1)$$
$$= A^{*} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} m_{0Ft} \\ g_{0t}^{*} \end{pmatrix} + o_p(1),$$

where $\overline{J}^{*}_{M\varpi T} = -T^{-1}\sum_{t=1}^{T} \left.\dfrac{\partial m_{Ft}}{\partial \varpi}\right|_{\overline{\varpi}}$ and $\overline{\varpi}$ is the usual "mean value" satisfying $\overline{\varpi} = \varpi_0 + o_p(1) \Rightarrow \overline{J}^{*}_{M\varpi T} = J^{*}_{M\varpi} + o_p(1)$, and $A^{*} = [I_r : -J^{*}_{M\varpi} J^{*-1}_{\varpi\varpi}]$. Now using Proposition 3.1, we conclude that

$$\sqrt{T}\hat{M}_{FT} \xrightarrow{d} N(0, V),$$

where $V = A^{*}\Sigma^{*}A^{*\prime}$ with $\Sigma^{*} = \begin{bmatrix} \Sigma_{MM} & \Sigma^{*}_{MG} \\ \Sigma^{*}_{GM} & \Sigma^{*}_{GG} \end{bmatrix}$. ■

3.A.4 Proof of Proposition 3.2

Proof. The test indicator under consideration is $\hat{M}_{FT} = \dfrac{1}{T}\sum_{t=1}^{T}(\hat{\upsilon}_t \otimes \hat{r}_t) = \dfrac{1}{T}\sum_{t=1}^{T}\hat{m}_{Ft}$ where $\hat{\upsilon}_t = \text{vech}\left(\hat{\zeta}_t\hat{\zeta}_t^{\prime} - \hat{\Gamma}\right)$. Define $J^{*}_{M\varpi} = \underset{T\longrightarrow\infty}{\text{plim}}\ J^{*}_{M\varpi T}$ where $J^{*}_{M\varpi T} = -\dfrac{1}{T}\sum_{t=1}^{T}\dfrac{\partial m_{Ft}}{\partial \varpi^{\prime}} = -\left[\dfrac{1}{T}\sum_{t=1}^{T}\dfrac{\partial m_{Ft}}{\partial \theta^{\prime}}, \dfrac{1}{T}\sum_{t=1}^{T}\dfrac{\partial m_{Ft}}{\partial \rho^{\prime}}\right]$, $Z = (Z_1, \cdots, Z_N)$ where Z_i is $(T \times k_i)$ matrix having rows $z^{\prime}_{it} = (c^{\prime}_{it}, x^{\prime}_{it})$ for $i = 1, \cdots N$. Also define R^{*} having rows $r^{\prime}_t$, if r_t is a vector of test variables, or R^{*} is a vector with typical element r_t if r_t is a scalar.

Note that $\upsilon_t = \left\{\zeta^2_{it} - 1\right\}$ if $i = j$ and $\upsilon_t = \left\{\zeta_{it}\zeta_{jt} - \rho_{ij}\right\}$ if $i \neq j$, $j < i = 2, \cdots, N$. Since ζ_{jt} are functionally independent of both φ_i and η_i, $i \neq j$, and ρ does not enter in θ. Now, we have $\dfrac{\partial \zeta_{it}}{\partial \varphi_i} = -f_{it} - \dfrac{1}{2}\zeta_{it}c_{it}$, $\dfrac{\partial \zeta_{it}}{\partial \eta_i} = -\dfrac{1}{2}\zeta_{it}x_{it}$, hence $\dfrac{\partial\left(\zeta^2_{it} - 1\right)}{\partial \varphi_i} = -2\zeta_{it}\left(f_{it} + \dfrac{1}{2}\zeta_{it}c_{it}\right)$ and $\dfrac{\partial\left(\zeta^2_{it} - 1\right)}{\partial \eta_i} = -2\zeta^2_{it}x_{it}$ and for $i \neq j, j < i = 2, \cdots, N$, $\dfrac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial \varphi_i} = -\zeta_{jt}\left(f_{it} + \dfrac{1}{2}\zeta_{it}c_{it}\right)$, $\dfrac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial \eta_i} = -\zeta_{jt}\left(\dfrac{1}{2}\zeta_{it}x_{it}\right)$, $\dfrac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial \rho_{ij}} = -1$. Hence,

$$\begin{aligned} \frac{\partial\left(\zeta^2_{it} - 1\right)}{\partial \theta^{\prime}_i} &= \left(-2\zeta_{it}\left(f^{\prime}_{it} + \frac{1}{2}\zeta_{it}c^{\prime}_{it}\right), -2\zeta^2_{it}x^{\prime}_{it}\right), \\ \frac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial \theta^{\prime}_i} &= -\zeta_{jt}\left(f^{\prime}_{it} + \frac{1}{2}\zeta_{it}c^{\prime}_{it}, \frac{1}{2}\zeta_{it}x^{\prime}_{it}\right). \end{aligned}$$

Now, note that $\mathrm{E}\left[\zeta^2_{jt}|\mathcal{F}_{t-1}\right] = 1$, $\mathrm{E}\left[f_{it}\zeta_{jt}|\mathcal{F}_{t-1}\right] = \mathrm{E}\left[f_{it}\zeta_{it}|\mathcal{F}_{t-1}\right] = 0$, since f_{it} is $\mathcal{F}_{t-1}$ measurable and $\mathrm{E}_0\left[\zeta_{1t}\zeta_{2t}|\mathcal{F}_{t-1}\right] = \rho_0$. Therefore

$$\mathrm{E}_0\left[\frac{\partial\left(\zeta_{it}^2 - 1\right)}{\partial\theta'} r_t|\mathcal{F}_{t-1}\right] = \underset{T\longrightarrow\infty}{\text{plim}} \frac{1}{T} R^{*\prime}(0,\ldots,Z_i,\cdots,0), \tag{3.55}$$

and for $i \neq j$, $j < i = 2,\cdots,N$,

$$\mathrm{E}_0\left[\sum_{t=1}^{T}\frac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial\theta'} r_t|\mathcal{F}_{t-1}\right] = \frac{1}{2}\rho_0 \underset{T\longrightarrow\infty}{\text{plim}} \frac{1}{T}\left\{R^{*\prime}(0,\ldots,Z_i,\cdots,Z_j,\cdots 0)\right\}. \tag{3.56}$$

Finally,

$$\mathrm{E}_0\left[\frac{\partial\left(\zeta_{it}^2 - 1\right)}{\partial\rho'} r_t|\mathcal{F}_{t-1}\right] = 0, \tag{3.57}$$

and

$$\mathrm{E}_0\left[\frac{\partial(\zeta_{it}\zeta_{jt} - \rho_{ij})}{\partial\rho'} r_t|\mathcal{F}_{t-1}\right] = \underset{T\longrightarrow\infty}{\text{plim}} \frac{1}{T}(0,\ldots,1',\cdots,0)R^{*}. \tag{3.58}$$

■

3.A.5 Proof of Lemma 3.4

Proof. Under normality, from the generalized IM inequality (e.g., Newey, 1985), we have $J^{*}_{M\varpi} = \Sigma^{*}_{MG}$ and $J^{*}_{\varpi\varpi} = \Sigma^{*}_{GG}$ and the result follows. Further (by ULLN) the consistent estimator of Σ^{*}_{MG}, Σ^{*}_{GG}.and Σ^{*}_{MM} are given by

$$\begin{aligned}
\hat{\Sigma}^{*}_{MG} &= T^{-1}\sum_{t=1}^{T}\hat{m}_{Ft}\hat{g}^{*\prime}_t = T^{-1}\hat{S}^{*\prime}\hat{R},\\
\hat{\Sigma}^{*}_{GG} &= T^{-1}\sum_{t=1}^{T}\hat{g}^{*}_t\hat{g}^{*\prime}_t = T^{-1}\hat{S}^{*\prime}\hat{S}^{*},\\
\hat{\Sigma}_{MM} &= T^{-1}\sum_{t=1}^{T}\hat{m}_{Ft}\hat{m}'_{Ft} = T^{-1}\hat{R}'\hat{R}.
\end{aligned}$$

Hence, $\hat{A}^{*} = \left[I_r : -\hat{\Sigma}^{*}_{MG}\hat{\Sigma}^{*-1}_{GG}\right]$. Now Define $\hat{B}^{*} = \left[\hat{R},\ \hat{S}^{*}\right]$, and $\hat{W}^{*} = \hat{B}^{*}\hat{A}^{*\prime}$, where R and $\hat{S}^{*}$are $(T \times r)$ and $(T \times N')$ matrices having rows $\hat{m}'_{Ft}$

and $\dfrac{\partial l_t^*}{\partial \varpi'}$, evaluated at $\hat{\varpi}$. Then V can be consistently estimated by $\hat{V}_T = \dfrac{1}{T}\hat{W}^{*\prime}\hat{W}^* = \hat{\Sigma}_{MM} - \hat{\Sigma}^*_{MG}\hat{\Sigma}^{*-1}_{GG}\hat{\Sigma}^*_{GM}$. ■

3.A.6 Proof of Theorem 3.3

Proof. Define $\hat{M}_{PT} \equiv \hat{M}$, $J_i(\theta_0) = -\operatorname{plim}_{T\to\infty} \frac{1}{T}\left[\sum_{t=1}^{T} \frac{\partial^2 l_{it}(\theta_i)}{\partial\theta_i \partial\theta_i'}\right]_{\theta=\theta_0}$, $\nabla M_i \equiv \nabla M_{PTi}(\theta_1, \theta_2) = \operatorname{plim}_{T\to\infty} \dfrac{\partial M_{0PT}}{\partial \theta_i}$, $i = 1, 2$. Now taking a mean value expansion of $\sqrt{T}\hat{M}$ about $\hat{\theta} = \theta_0$

$$\begin{aligned}
\sqrt{T}\hat{M} &= \sqrt{T}M(\theta_0) + \nabla M_1(\theta_{10}, \theta_{20}) \times J_1(\theta_{10})^{-1}\sqrt{T}G_1(\theta_{10}) \\
&\quad + \nabla M_2(\theta_{10}, \theta_{20}) \times J_2(\theta_{20})^{-1}\sqrt{T}G_2(\theta_{20}) + o_p(1) \\
&= \sqrt{T}M(\theta_0) - J_{M\theta} \times J_{\theta\theta}^{-1}\sqrt{T}G(\theta_0) + o_p(1) \\
&= A_1 \begin{bmatrix} \sqrt{T}M(\theta_0) \\ \sqrt{T}G(\theta_0) \end{bmatrix} + o_p(1),
\end{aligned} \tag{3.59}$$

where $J_{M\theta} = -\begin{bmatrix} \nabla M_1 & \nabla M_2 \end{bmatrix}$, $G(\theta_0) = \begin{bmatrix} G_1(\theta_{10}) \\ G_2(\theta_{20}) \end{bmatrix}$, $J_{\theta\theta} = \begin{bmatrix} J_1(\theta_{10}) & 0 \\ 0 & J_2(\theta_{20}) \end{bmatrix}$ and $A_1 = [I_r, \ -J_{M\theta} \times J_{\theta\theta}]$. Thus, when the proposition (3.3) holds we can write that

$$\sqrt{T}\hat{M}_T \xrightarrow{d} N(0, V_1),$$

where $V_1 = A_1 \Sigma A_1'$. ■

3.A.7 Proof of Lemma 3.6

Proof. Assuming that the specification of the log-likelihood for the FQML estimation of parameters is correct, we can use a generalized (conditional) IM equality which says that

$$\begin{aligned}
\mathrm{E}_0\left[\frac{\partial}{\partial\theta_i}\left(\frac{\partial l_{it}(\theta_i)}{\partial\theta_i'}\right)|\mathcal{F}_{t-1}\right] &= -\mathrm{E}_0\left[\left(\frac{\partial l_{it}(\theta_i)}{\partial\theta_i}\right)\left(\frac{\partial l_t^*(\theta,\rho)}{\partial\theta_i'}\right)\middle|\mathcal{F}_{t-1}\right], \\
\mathrm{E}_0\left[\frac{\partial m_t(\theta)}{\partial\theta_i'}\middle|\mathcal{F}_{t-1}\right] &= -\mathrm{E}_0\left[m_t(\theta)\frac{\partial l_t^*(\theta,\rho)}{\partial\theta_i'}\middle|\mathcal{F}_{t-1}\right],
\end{aligned}$$

where $\dfrac{\partial l_t^*(\theta,\rho)}{\partial \theta_i'}$ is the score for θ_i, $i=1,2$, from the FQMLE log-likelihood.

Then $J_i = -\operatorname{plim}_{T\to\infty} \frac{1}{T}\left[\sum_{t=1}^{T} \frac{\partial^2 l_{it}(\theta_i)}{\partial\theta_i \partial\theta_i'}\right] = \operatorname{plim}_{T\to\infty} \dfrac{1}{T} Q_i' S_i$ and $J_{M\theta} = -\operatorname{plim}_{T\to\infty} \dfrac{1}{T} R' S_i$. Substituting these into (3.59) yields

$$
\begin{aligned}
\sqrt{T}\hat{M} &= \sqrt{T}M(\theta_0) - \frac{1}{T}R'S \begin{bmatrix} \frac{1}{T}Q_1'S_1 & 0 \\ 0 & \frac{1}{T}Q_2'S_2 \end{bmatrix}^{-1} \sqrt{T}G(\theta_0) + o_p(1) \\
&= A_1 \begin{bmatrix} \sqrt{T}M(\theta_0) \\ \sqrt{T}G(\theta_0) \end{bmatrix} + o_p(1),
\end{aligned}
$$

where $A_1 = \left[I_r : \ -R'S \begin{pmatrix} (Q_1'S_1)^{-1} & 0 \\ 0 & (Q_2'S_2)^{-1} \end{pmatrix} \right]$ and $S = [S_1,\ S_2]$. Now defining $B = [R,\ Q_1,\ Q_2]$, $\Sigma = \operatorname{plim}_{T\to\infty} B'B$. Hence, the variance-covariance matrix V_1 can be written as $V_1 = \operatorname{plim}_{T\to\infty} \dfrac{1}{T} A_1 B' B A_1' = \operatorname{plim}_{T\to\infty} \dfrac{1}{T} W'W$ where $W = BA_1' = R - Q_1 (S_1'Q_1)^{-1} S_1'R - Q_2 (S_2'Q_2)^{-1} S_2'R$.

Now V_1 can be consistently estimated by $\hat{V}_{1T} = \dfrac{1}{T}\hat{W}'\hat{W}$, where hats denote θ_0 replaced by the individual GARCH estimators, $\hat{\theta}$, and ρ_0 replaced by the estimator $\hat{\rho} = \dfrac{1}{T}\sum_{t=1}^{T} \hat{\zeta}_{1t}\hat{\zeta}_{2t}$. ■

Chapter 4

On the Nonparametric Tests of Univariate GARCH Regression Models

4.1 Introduction

In many economic and financial time series applications, such as portfolio selection and asset pricing, deciding whether the dynamics are determined by the conditional mean and/or conditional variance has significant implications. The widespread application of GARCH type models in financial econometrics is a clear indication of the popularity and success of these models. Specification testing procedures of these volatility models, particularly parametric ones, have rightfully received considerable attention in the literature and almost all of them implicitly assume a correct specification for the conditional mean. A separate stream of literature deals with the problem of diagnostic testing of the mean function in the presence of heteroskedascity. On the other hand, only a few studies consider simultaneously testing both the mean and variance function.

*Earlier versions of this Chapter were presented in the Doctoral conference 2010, University of Manchester and Econometrics and Applied Economics Workshops 2010/11, University of Manchester. Our thanks to workshop participants for their insightful comments and discussions.

To test for the presence of the ARCH effects Engle (1982) provided a LM type testing procedure. There is a singularity problem of the block of information matrix required to construct a LM test for the GARCH disturbances against white noise disturbances (Bollerslev, 1986). Lee (1991) bypassed this problem by showing that for the linear regression model, LM tests for the GARCH and ARCH disturbances are the same. A number of diagnostic testing procedures have also been proposed in the literature; for example, a portmanteau type test to test the null that the squared standardized error process is serially uncorrelated (Li and Mak, 1994), score-type tests for testing a GARCH specification against a higher order GARCH (Bollerslev, 1986), for asymmetry (Engle and Ng, 1993) and for (a) no remaining ARCH effects in standardized errors, (b) linearity or symmetry against a smooth transition GARCH, and (c) parameter constancy against smoothly changing parameters (Lundbergh and Teräsvirta, 2002). Halunga and Orme (2009) provided a unifying framework based on a Conditional Moment (CM) principle and proposed "new" tests for asymmetry and non-linearity.

The parametric CM tests, however, are not necessarily consistent against all possible alternatives (see, for example, Bierens, 1982; Holly, 1982; Newey, 1985; Tauchen, 1985) as they only employ a finite number of moment restrictions implied by the model; e.g.,

$$\mathrm{E}\left[e\left(\theta_0\right)|Z\right] = 0 \text{ a.s. for some } \theta_0 \in \Theta \subset \Re^p, \tag{4.1}$$

where $e\left(\theta_0\right)$ is the regression error and Z is the conditioning set. Further we want to reemphasize that all above mentioned tests are developed assuming a correct mean specification. It is worthwhile to quote Engle's (1982, pp.990) third interpretation of the ARCH regression model which says "an approximation to a more complex regression which has non-ARCH disturbances. The ARCH specification might then be picking up the effect of variables omitted from the estimated model. The existence of an ARCH effect would be interpreted as evidence of misspecification, either by omitted variables or through structural change. If this is the case, ARCH may be a better approximation to reality than making standard assumptions about the disturbances, but trying to find the omitted variable or determine the nature of the structural change would be even better". Noting the possible spill-over effect from neglected misspecification of conditional mean on testing the conditional variance and the inconsistency of the parametric CM tests, this Chapter aims to test both joint (mean and variance) specification of a GARCH regression model and its marginal components simultaneously applying a consistent nonparametric approach.

If the conditional mean is misspecified, existing specification tests of the conditional variance perform poorly by over-rejecting the correct conditional variance specification. In fact, such tests can be interpreted as joint tests of conditional mean and conditional variance. A number of studies examined the effect of misspecified mean (e.g., omitted variables, structural change parameter instability, noisy chaotic function) on the diagnostic tests of variance specification. For example, Lumsdaine and Ng (1999) found that, under mean misspecification, the ARCH-LM test, in general, over-rejects the null of conditional homoskedasticity and suggested using recursive residuals to robustify the ARCH-LM test. For a noisy chaotic (highly non-linear) conditional mean model with homoskedastic errors, Kyrtsou (2008) showed that the ARCH-LM (Engle, 1982) and the McLeod and Li (1983) tests for non-linearity in the second moment may exhibit spurious heteroskedasticity due to inappropriate filtering of neglected non-linearity. Blake and Kapetanios (2007) also found that misspecification of the conditional mean may lead to the spurious rejection of the null of no ARCH and suggested a new testing procedure based on an artificial neural network (ANN) which is robust to the presence of neglected non-linearity. On the other hand, there is also the problem of testing the conditional mean specification in the presence of (G)ARCH error. To address this problem a number of techniques have been suggested in the literature; e.g., using a heteroskedasticity-consistent covariance matrix estimator or correctly specifying ARCH process (Lee et al., 1993), using a heteroskedastic consistent auxiliary regression together with the wild bootstrap (Becker and Hurn, 2009).

The above evidence suggests that it is desirable to test the conditional mean and variance specification simultaneously. There are some suggestions of sequential testing; i.e., the conditional mean first and then testing the conditional variance with correctly specified mean function. If sole interest lies in the conditional variance specification, another possibility is to estimate the conditional mean by some nonparametric method and then use the residuals in the variance diagnostic test (cf. Blake and Kapetanios, 2007). Although the problem of testing many conditional moment restrictions has been considered in the literature (see, for example, Chen and Fan, 1999 for mixing data or Delgado, Dominguez and Lavergne, 2006 for independent data), the literature on simultaneous diagnostic tests of conditional mean and variance is very limited. Ngatchou-Wandji (2005) suggested a Wald-type test based on χ^2-discrepancy measures which is, however, not consistent against a wide class of alternatives to the correct specification. The aim of this Chapter is to develop simultaneous tests of both specifications by employing consistent testing procedures.

Bierens (1982) first proposed a consistent testing procedure, an Integrated CM (ICM) test, for non-linear parametric regression models involving i.i.d. data. Since then a vast amount of literature has addressed the issue of consistent testing in both the i.i.d. and time series contexts. For time series, a few prominent examples are Bierens (1984), de Jong (1996), Bierens and Polberger (1997), Koul and Stute (1999), Hong and Lee (2005), and Escanciano (2006a, 2006b, 2007a, 2008). These can be broadly categorized into two classes of tests: namely tests based on a local approach and tests based on an integrated approach. The first approach uses nonparametric smoothing estimators of a local measure of dependence $\mathrm{E}\left[e\left(\theta_0\right)|Z\right]$. The local approach requires smoothing of the data which leads to a less precise fit and for high (or even moderate) dimension of Z. This approach thus suffers from the "curse of dimensionality"; i.e., considerable bias, even for large sample (see for example, Section 7.1 of Fan and Gijbels, 1996). But these tests have standard asymptotic null distributions, though finite-sample distributions depend on the choice of the bandwidth and on the nonparametric estimator. On the other hand, ICM tests use integrated (or cumulative) measures of dependence and avoid the smoothing by converting conditional orthogonality conditions of (4.1) to uncountably many unconditional (parametric) orthogonality moment restrictions; i.e.,

$$\mathrm{E}\left[e\left(\theta_0\right)|Z\right]=0 \text{ a.s.} \Leftrightarrow \mathrm{E}\left[e\left(\theta_0\right)w\left(Z,x\right)\right]=0, \text{ a.e. in } \Pi\subset\Re^q, \tag{4.2}$$

where the parametric family $\{w(.,x): x\in\Pi\}$ is such that (4.2) holds and $\Pi\subset\Re^q$ is a properly chosen space. More details are provided in Section 4.3, also see e.g., Bierens and Polberger (1997), Stinchcombe and White (1998), and Escanciano (2006a). However, in this case the asymptotic null distribution depends on the Data Generating Process (DGP) and null specification. Hence critical values cannot be tabulated for general cases and a bootstrap procedure is required to implement the test. In terms of power the ICM tests have higher power than local tests for low frequency alternatives and have lower power compared to local tests for high frequency alternatives (Fan and Li, 2000).

We can, in principle, use the classical ICM tests in our case. However, for GARCH regression model bootstrapping is a complex and problematic operational issue. One solution is to use a *feasible projected* version of the classical ICM test (Escanciano, 2009), which we shall term as the Projected ICM (PICM) test, to address our problem of simultaneous testing of joint and marginal hypothesis. Although this procedure also requires bootstrap, it does not need to estimate the parameters in the bootstrap world, therefore making it much simpler to implement. To the best of our knowledge there is no study which examines this approach in detail in the context of GARCH regression models.

Escanciano (2008) proposed a consistent joint and marginal testing procedure, an Integrated Generalized Spectral (IGS) test, for conditional mean and conditional variance for a general conditionally heteroskedastic model. This test is based on a pair-wise generalized spectral approach. To calculate the critical values of the IGS tests, the author suggested and theoretically justified a Fixed Design Wild Bootstrap (FDWB) procedure which requires estimating the parameters in each bootstrap replication. The null model considered by Escanciano indeed includes GARCH regression models, however the null DGPs in his Monte Carlo experiments do not consider GARCH models with non-zero and potentially time varying conditional mean (see Remark 4.2). However, this Chapter argues that Escanciano's FDWB procedure is not directly applicable in this case (see Section 4.5 for more on this topic).

4.1.1 Contributions and Structure of the Chapter

The first major contribution of this Chapter is to identify a problem of Escanciano's (2008) Fixed Design Wild Bootstrap (FDWB) in the context of a general GARCH regression model, implying that a modification of this procedure is required to accommodate this model. Since a full parametric GARCH bootstrap procedure for IGS tests using the GARCH QMLE is operationally problematic, a simple alternative bootstrap procedure for IGS tests, using least squares estimation, has been proposed. Asymptotic analysis suggests that it does not strictly satisfy the sufficient conditions identified in the previous literature (Escanciano 2007b) which however does not necessarily mean that our procedure is asymptotically invalid. Nonetheless, due to the simplicity of the procedure a Monte Carlo study is conducted to evaluate its ability to control for the size of the IGS tests. Our simulation study demonstrates excellent size properties for the proposed bootstrap procedure. This suggests the possibility that the IGS tests can be implemented under a weaker set of conditions. Moreover the ability of our procedure to control the size very well makes power comparisons justified. The specification testing literature reveals that asymptotically valid tests often display poor size properties leading to inconsequential power comparisons.

Secondly, we illustrate in detail how to execute the PICM tests for this model. Thirdly, an extensive Monte Carlo study is conducted to compare the performance of these two nonparametric tests and four parametric CM tests of nonlinearity and/or asymmetry considered in Halunga and Orme (2009), Engle and Ng (1993) and Lundbergh and Teräsvirta (2002). This simulation exercise also provides us with the opportunity to investigate the effect of conditional mean misspecification on various parametric LM and CM tests of conditional variance in the regression context which has not been done in the literature before. Finally, we illustrate the testing procedures with the help of the S&P 500 data.

The remainder of this Chapter is organized as follows. In Section 4.2 we provide the null model and moment conditions considered in this Chapter. The PICM testing framework is discussed next which is followed by the Escanciano's IGS test and wild bootstrap scheme suggested by him. In Section 4.5 the limitations of Escanciano's FDWB procedure while applying for our GARCH regression model are pointed out and we put forward a modified bootstrap scheme. The parametric CM tests of the conditional variance specification are briefly introduced in Section 4.6. Finally, we present the simulation evidence in Section 4.7 and an empirical application in Section 4.8. All proofs are relegated to the appendix. In the remaining A^c is the complex conjugate of A, $\|A\|_M$ denotes the weighted norm $A'MA^c$ for a positive definite matrix M and a complex vector A.

4.2 The Null Model and Moment Conditions

Consider a $\{(y_t, X_t')'\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process on a probability space $(\Omega, \mathcal{F}, P)$ where y_t is the dependent variable and $Z_{t-1} = (y_{t-1}, X_t')' \in \Re^{1+m}, m \in \mathbb{N}$, is the explanatory random vector containing lagged values of y_t and possibly other variables. Suppose $\mathcal{I}_{t-1} = \left(Z_{t-1}', Z_{t-2}', \cdots\right)'$ is the information set at time $t-1$ and $\mathcal{F}_{t-1} = \sigma\left(Z_{t-1}', Z_{t-2}', \cdots\right)$ is the $\sigma-$field generated by the past information up to and including time $t-1$. Then define the conditional mean and variance $m(\mathcal{I}_{t-1}) = \mathrm{E}\left[y_t | \mathcal{I}_{t-1}\right]$, and $h(\mathcal{I}_{t-1}) = Var\left[y_t | \mathcal{I}_{t-1}\right]$, respectively, and standardized errors $\zeta_t = \frac{(y_t - m(\mathcal{I}_{t-1}))}{\sqrt{h(\mathcal{I}_{t-1})}}, t \in \mathbb{Z}$. We consider the following parametric model

$$\begin{aligned} y_t &= m\left(\mathcal{I}_{t-1,q}, \varphi_0\right) + \varepsilon_{0t}, \\ \varepsilon_{0t} &= \sqrt{h\left(\mathcal{I}_{t-1,q}, \eta_0\right)}\zeta_t, \end{aligned} \tag{4.3}$$

where ζ_t is i.i.d. $(0,1)$, $\mathcal{I}_{t-1,q} = \{Z_s\}_{s=1}^{t-q}$, $q < \infty$, $q \in \mathbb{N}$ and $\theta_0 = (\varphi_0', \eta_0')' \in \Theta \subset \Re^p$ where φ_0 and η_0 represent the true conditional mean and conditional variance parameters, respectively. This specification is quite general and includes linear ARMA-ARCH, ARMA-GARCH as well as nonlinear conditional mean (e.g., GARCH-in-Mean, bilinear) and non-linear and asymmetric variance models (e.g., GJR GARCH, EGARCH, STGARCH).

To focus our discussion we will consider the AR(1)-GARCH$(1,1)$ process, where

$$\begin{aligned} m\left(\mathcal{I}_{t-1,q},\varphi_0\right) &= W_t\varphi_0 = \varphi_{00}+\varphi_{01}y_{t-1}, \\ h\left(\mathcal{I}_{t-1,q},\eta_0\right) &= \eta_0' s_{0,t-1} = \alpha_{00}+\alpha_{01}\varepsilon_{0,t-1}^2+\beta_{01}h_{0,t-1}, \end{aligned} \tag{4.4}$$

with $W_t = \left(1, y_{t-1}\right), \ \varphi' = \left(\varphi_0, \varphi_1\right), \ s_{t-1} = \left(1, \varepsilon_{t-1}^2, h_{t-1}\right)', \ \eta = \left(\alpha_0, \alpha_1, \beta_1\right)'.$ For notational convenience we write $m_t \equiv m\left(\mathcal{I}_{t-1,q},\varphi\right)$ and $h_t \equiv h\left(\mathcal{I}_{t-1,q},\eta\right)$. Under correct model specification, $\{\varepsilon_{0t}\}$ is MDS wrt $\mathcal{F}_t$, with zero mean and conditional variance h_t. That is the correct joint specification is tantamount to saying

$$H_0 : \mathrm{E}\left[e_{0,1t}|\mathcal{I}_{t-1}\right] = 0 \text{ a.s. } \text{ and } \mathrm{E}\left[e_{0,2t}|\mathcal{I}_{t-1}\right] = 0 \text{ a.s. for some } \theta_0 \in \Theta, \tag{4.5}$$

where $e_{0,1t} \equiv \varepsilon_{0t} = Y_t - m_{0t}$ and $e_{0,2t} \equiv \varepsilon_{0t}^2 - h_{0t}$. Or, more compactly,

$$H_0 : \mathrm{E}\left[e_{0,t}|\mathcal{I}_{t-1}\right] = 0 \text{ a.s. for some } \theta_0 \in \Theta, \tag{4.6}$$

where $e_{0,t} \equiv e_t\left(\theta_0\right) = \left(e_{0,1t}, e_{0,2t}\right)'$. It is important to note that first conditional moment restriction (CMR) corresponds to adequacy of the conditional mean whereas both CMRs are necessary for correct specification of conditional variance.

We assume that our model satisfies the following regularity conditions:

Assumption 4.2.1 $\{y_t, X_t\}_{t\in\mathbb{Z}}$ *is a strictly stationary and ergodic process.*

Assumption 4.2.2 $\mathrm{E}\left[e_{0,1t}^2\right] = \mathrm{E}\left[\left(y_t - m_{0t}\right)^2\right] < \infty$, *and* $\mathrm{E}\left[e_{0,2t}^2\right] = \mathrm{E}\left[e_{0,1t}^2 - h_{0t}\right]^2 < \infty$.

Assumption 4.2.3 *Let* Θ_0 *be a small convex neighborhood of* θ_0. *The functions* $m\left(\mathcal{I}_{t-1,q},.\right)$ *and* $h\left(\mathcal{I}_{t-1,q},.\right)$ *are twice continuously differentiable (a.s.)*

wrt $\theta \in \Theta_0$. *Also,* $\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left\|g'_{jt}(\theta)\right\|\right] < \infty$, $j = 1,2$ *where* $g'_{1t}(\theta) = \frac{\partial e_{1t}}{\partial \theta'} = -\frac{\partial m_t}{\partial \theta'}$, $g'_{2t}(\theta) = \frac{\partial e_{2t}}{\partial \theta'} = 2e_{1t}(\theta)\, g'_{1t}(\theta) - \frac{\partial h_t}{\partial \theta'}$.

Assumption 4.2.4 *The parameter space* Θ *is compact in* $\Re^p$ *and* θ_0 *belongs to the interior of* Θ.

Assumption 4.2.5 *The observed information set at time* t, $\widehat{\mathcal{I}}_t$, *may contain some initial values and satisfies*

$$\left(\sum_{t=1}^{T}\left[\mathrm{E}\sup_{\theta\in\Theta_0}\left\|m\left(\widehat{\mathcal{I}}_{t-1},\theta\right) - m\left(\mathcal{I}_{t-1},\theta\right)\right\|^2\right]^{1/2}\right)^2 = o(T),$$

$$\left(\sum_{t=1}^{T}\left[\mathrm{E}\sup_{\theta\in\Theta_0}\left\|\left(Y_t - m\left(\widehat{\mathcal{I}}_{t-1},\theta\right)\right)^2 - \left(Y_t - m\left(\mathcal{I}_{t-1},\theta\right)\right)^2\right\|^2\right]^{1/2}\right)^2 = o(T),$$

$$\left(\sum_{t=1}^{T}\left[\mathrm{E}\sup_{\theta\in\Theta_0}\left\|h\left(\widehat{\mathcal{I}}_{t-1},\theta\right) - h\left(\mathcal{I}_{t-1},\theta\right)\right\|^2\right]^{1/2}\right)^2 = o(T).$$

These assumptions are fairly general and considerably weaker compared to the related conditions made in the literature. Assumptions 4.2.1 and 4.2.2 do not involve any mixing or asymptotic independence assumption as opposed to the mixing Assumption A.1 in Hong and Lee (2003), yet they allow for a long memory process. Assumption 4.2.3 and 4.2.4 are standard in the literature see, e.g., Escanciano (2006a) and satisfied in our model. Finally, Assumption 4.2.5 is a start-up value condition and similar in spirit to Assumption A4 in Hong and Lee (2003), which ensures that the impacts of initial values are asymptotically negligible. This condition holds for many time series models including ARMA-GARCH models; see Francq and Zakoian (2004).

4.3 The Projected ICM (PICM)Test

4.3.1 Test Statistics and Limit Distribution

The underlying idea of the ICM tests is to characterize the CMR under consideration by an infinite number of unconditional moment restrictions. More specifically, for a moment condition of the form $\mathrm{E}\left[e_t(\theta_0)|\, Z_{t-1}\right]$, we have (by the law of iterated expectation)

$$\mathrm{E}\left[e_t(\theta_0)\, w(Z_{t-1}, x)\right] = \int\limits_{(-\infty,x]} \mathrm{E}\left[e_t(\theta_0)|\, Z_{t-1} = x\right] \mathrm{d}\, P_z,\ x \in \Re^{1+m}, \qquad (4.7)$$

where P_z is a stationary probability measure of Z_{t-1} and $w(Z_{t-1}, x)$ is some weighting function such that the above equivalence holds (see Lemma 4.1 for the sufficient conditions for the weight function). Now from (4.7) and Billingsley (1995, Theorem 16.10iii), we have

$$\mathrm{E}\left[e_t(\theta_0)|\, Z_{t-1}\right] \equiv 0 \text{ a.s.} \Leftrightarrow \mathrm{E}\left[e_t(\theta_0)\, w(Z_{t-1}, x)\right] \equiv 0.$$

Within this framework, the null hypothesis can be written as:

$$H_0 : \mathrm{E}\left[e_t(\theta_0)\, w(Z_{t-1}, x)\right] = 0,\ \forall x \in \Re^{1+m} \text{ and some } \theta_0 \in \Theta,$$

against the class of nonparametric alternatives

$$H_A : P_z\left[\mathrm{E}\left[e_t(\theta_0)\, w(Z_{t-1}, x)\right] \neq 0\right] > 0,\ \forall\ \theta \in \Theta.$$

Lemma 4.1 *Let $\mathcal{C}_b(\Re^{1+m})$ be the space of all bounded, continuous, complex valued functions on $\Re^{1+m}$. Then, any of the following conditions are sufficient for the class of function $\mathcal{W} = \{w(Z, x) : x \in \Pi \subset [-\infty, \infty]^s\}$, Z is a random variable with the same dimension as Z_t, $t \in \mathbb{Z}$, Π is the nuisance parameter space with dimension s which depends on the particular family $\mathcal{W}$ used to satisfy the equivalence in (4.7):*

1. *$\mathcal{W} \subset \mathcal{C}_b(\Re^{1+m})$ is a vector lattice that contains the constant functions and separate points of $\mathbb{R}^{1+m}$.*

2. $\mathcal{W} \subset \mathcal{C}_b\left(\Re^{1+m}\right)$ *is a algebra that contains the constant functions and separate points of* $\Re^{1+m}$.

3. $\mathcal{W} = \{w\left(x'Z\right) : x \in \Pi\}$ *and is a non-polynomial analytical function.*

4. $\mathcal{W} = \{1\left(Z \in \mathcal{B}_x\right) : x \in \Pi\}$ *and* $\{\mathcal{B}_x\}_{x\in\Pi}$ *is a separating class of Borel sets of* $\Re^{1+m}$.

The most commonly used examples of w are the indicator weight function $w\left(Z_{t-1}, x\right) = 1\left(Z_{t-1} \leq x\right)$ with $x \in \Pi_{ind} = \left[-\infty, \infty\right]^{m+1}$ (Stute and Zhu, 2002) and the complex exponential function $w\left(Z_{t-1}, x\right) = \exp\left(ix'Z_{t-1}\right)$ with $i = \sqrt{-1}$ and $x \in \Pi_{\exp} = \Re^{m+1}$ (Bierens, 1982). Different families w have different power properties. "Optimal" choices for w depend on the true alternative at hand and the function used to measure orthogonality restrictions.

Now define the classical marked empirical process as

$$R_T\left(x, \theta\right) = T^{-1/2} \sum_{t=1}^{T} \left[e_t(\theta) w\left(Z_{t-1}, x\right)\right].$$

In the field of econometrics and statistics inference, there is a long tradition of using processes like $R_T\left(x, \theta\right)$, see, e.g., Bierens (1982), Stute (1997), Koul and Stute (1999), Escanciano (2007a), among others. Note that θ_0 is a nuisance parameters in the construction of the test and most existing tests do not acknowledge this fact. Since θ_0 is unknown, deviations in the direction of the score function cannot be differentiated from local deviation of θ_0 (i.e., deviations within the parametric model) which may result in tests with low power. To address this, for any $\sqrt{T}$-consistent estimator $\hat{\theta}$, Escanciano (2009) suggested a *projected* marked empirical process as follows:

$$R_T^1\left(x, \hat{\theta}\right) = T^{-1/2} \sum_{t=1}^{T} \left[w\left(Z_{t-1}, x\right) I_e - G'\left(x, \hat{\theta}\right) \Xi^{-1} g\left(Z_{t-1}, \hat{\theta}\right)\right] \hat{e}_t, \quad (4.8)$$

where $\hat{e}_t \equiv e_t(\hat{\theta})$, I_e is the identity matrix with same dimension of $\hat{e}_t$,

$$\begin{aligned} g\left(Z_{t-1},\theta\right) &= \frac{\partial e_t(\theta)}{\partial \theta}, \\ \Xi &= \mathrm{E}\left[g\left(Z_{t-1},\theta_0\right) g\left(Z_{t-1},\theta_0\right)'\right], \text{ and} \\ G\left(x,\theta\right) &= \mathrm{E}\left[g\left(Z_{t-1},\theta\right) w\left(Z_{t-1},x\right)\right]. \end{aligned}$$

Because of the amalgamation of the score information, tests constructed based on (4.8) do not waste power due to the local deviation of θ_0. Note that, under some regularity conditions,

$$\sup_{x\in\Re^{1+m}} \left\| R_T^1\left(x,\hat{\theta}\right) - R_T^1\left(x,\theta_0\right)\right\| = o_p\left(1\right). \tag{4.9}$$

Because of this key property, implementation of the test neither requires the asymptotic distribution of the estimator nor θ to be estimated in the each bootstrap world.[1]

Then the test statistic is a continuous functional of the feasible projected marked empirical process:

$$\hat{R}_T^1\left(x\right) = T^{-1/2}\sum_{t=1}^{T}\left[w\left(Z_{t-1},x\right)I_e - \hat{G}'\left(x,\hat{\theta}\right)\hat{\Xi}^{-1} g\left(Z_{t-1},\hat{\theta}\right)\right]\hat{e}_t, \tag{4.10}$$

with

$$\begin{aligned} \hat{G}\left(x,\hat{\theta}\right) &= T^{-1}\sum_{t=1}^{T}\left[g\left(Z_{t-1},\hat{\theta}\right) w\left(Z_{t-1},x\right)\right], \text{ and} \\ \hat{\Xi} &: = T^{-1}\sum_{t=1}^{T}\left[g\left(Z_{t-1},\hat{\theta}\right) g\left(Z_{t-1},\hat{\theta}\right)'\right]. \end{aligned}$$

Using the Cramer-von Mises (CvM) norm, the test statistic becomes:

$$CvM_T := \int_{\Re^{1+m}} \left\|\hat{R}_T^1\left(x\right)\right\|^2 dF_{T,Z}\left(x\right), \tag{4.11}$$

[1] As a matter of fact, any test statistics of the form $T^{-1/2}\sum_{t=1}^{T} a\left(Z_t\right)\hat{e}_t(\hat{\theta})$, with $\mathrm{E}\left[a\left(Z_t\right) g(Z_t,\theta_0)\right] = 0$, satisfies (4.9).

where $F_{T,Z}$ is the Empirical Distribution Function (EDF) of $\{Z_{t-1}\}_{t=1}^{T}$. For the limit distribution of $\hat{R}_T^1(x)$, in addition to the Assumptions 4.2.1 - 4.2.5, we require the following assumptions:

Assumption 4.3.1 *The derivatives* $g(Z_{t-1}, \theta)$ *satisfies*

$$\mathrm{E}\left[\sup_{\theta\in\Theta_0} \|g(Z_{t-1}, \theta)\|^2\right] < \infty, \quad \mathrm{E}\left[\|g(Z_{t-1}, \theta_0)\, e_t(\theta_0)\|^2\right] < \infty,$$

and $\mathrm{E}\left[g(Z_{t-1}, \theta)\, g'(Z_{t-1}, \theta)\right]$ *is positive definite in* Θ_0.

Assumption 4.3.2 $\left\|\widetilde{G}_t(x_1) - \widetilde{G}_t(x_2)\right\| \leq C_t \|x_1 - x_2\|^{s_1}$ *for each* $(x_1, x_2) \in \Re^{1+m} \times \Re^{1+m}$, *for some* $s_1 > 0$ *and a generic stationary sequence* C_t *with* $\mathrm{E}[C_t] < \infty$ *where* $\widetilde{G}_t(x_1) = \mathrm{E}\left[\mathrm{E}\left[\sup_{\theta\in\Theta_0} \|e_t^2(\theta_0)\,|Z_{t-1}\|\right] w(Z_{t-1}, x)\,|\mathcal{F}_{t-1}\right].$

Assumption 4.3.3 $\sqrt{T}\left(\hat{\theta} - \theta_0\right) = O_p(1)$.

Assumption 4.3.2 requires the existence of conditional Lipschitz moments while by Assumption 4.3.3 we only need any $\sqrt{T}$ consistent estimator of θ. Then following Theorem and Corollary provide the limit distribution of the process $\hat{R}_T^1(x)$ and CvM test statistics (for proof, see Escanciano, 2009):

Theorem 4.1 *Under the above assumptions,*

$$\sup_{x\in\Re^{1+m}} \left\|\hat{R}_T^1(x) - R_T^1(x, \theta_0)\right\| = o_p(1).$$

Corollary 4.1 *Under the above assumptions,*

$$\begin{aligned} \hat{R}_T^1(x) &\Rightarrow R_\infty^1, \\ CvM_T &\Rightarrow CvM_\infty := \int_{R^{1+m}} \left\|R_\infty^1(x)\right\|^2 dF_Z(x), \end{aligned}$$

where R_∞^1 *is a Gaussian process with zero mean and covariance function* $\mathrm{E}\left[K e_t(\theta_0)\, e_t(\theta_0)' K'\right]$ *with* $K = w(Z_{t-1}, x)\, I_e - G'(., \theta_0)\, \Gamma^{-1} g(Z_{t-1}, \theta_0)$.

4.3.2 The PICM Tests of the GARCH Regression Model

Since the asymptotic null distribution of $\hat{R}_T^1(x)$ depends on the DGP and null hypothesis in a complicated way, some approximation is required to obtain the critical values. More specifically, the unknown limiting null distribution of $CvM_T = \psi\left(\hat{R}_T^1(x)\right)$; i.e., the distribution of $\psi\left(R_\infty^1(x)\right)$, is approximated by the bootstrap distribution of $\psi\left(\hat{R}_T^{*1}(x)\right)$ where $\hat{R}_T^{*1}(x)$ is some bootstrap version of $\hat{R}_T^1(x)$. Although there are some suggestions available in the literature,[2] the most popular choice is the wild bootstrap technique which is used in a variety of problems; see, e.g., Stute et al. (1998), Whang (2000), and Escanciano (2007a), among others.

Definition 4.1 *A wild bootstrap involves adding together an estimated predicted part, which serves as a bootstrap world conditional mean, and a bootstrap error term which allow for heteroskedasticity of unknown form. Consider the regression model: $y_t = x_t'\beta + \varepsilon_t$. A typical observation for a wild bootstrap scheme for this regression model can be written as*

$$y_t^* = x_t'\hat{\beta} + \varepsilon_t^*,$$

where $\hat{\beta}$ is an estimator of β and $\varepsilon_t^* = f(\hat{\varepsilon}_t)U_t$ is the bootstrap error term in which $f(\hat{\varepsilon}_t)$ is some function of OLS residuals $\hat{\varepsilon}_t = y_t - x_t'\hat{\beta}$ ($\hat{\varepsilon}_t$ can possibly be obtained using a different estimator, say $\widetilde{\beta}$, other than $\hat{\beta}$) and U_t is a mutually independent drawing from a pick distribution which itself is completely independent of the data (y_t, x_t') with $\mathrm{E}(U_t) = 0$, $Var(U_t) = 1$ and has bounded support. The two most popular choices of U_t are Mammen's (1993) i.i.d. Bernoulli variates with

$$\Pr\left(U_t = \frac{1}{2}\left(1-\sqrt{5}\right)\right) = \frac{1+\sqrt{5}}{2\sqrt{5}},\ \Pr\left(U_t = \frac{1}{2}\left(1+\sqrt{5}\right)\right) = 1 - \frac{1+\sqrt{5}}{2\sqrt{5}}, \tag{4.12}$$

[2] For example, the Hansen's (1996) conditional p-value method, the Khmaladze's (1981) martingale transformation, upper bounds for the critical values (Bierens and Ploberger, 1997).

and Davidson and Flachaire's (2001,2008) Rademacher distribution with

$$\Pr(U_t = 1) = \Pr(U_t = -1) = 0.5. \tag{4.13}$$

Note that the wild bootstrap method requires a separable moment function ε_t so that the dependent variable can be recovered in an additive manner (e.g., in a regression model, $\varepsilon_t = y_t - x_t'\hat{\beta}$ so that $y_t = x_t'\hat{\beta} + \varepsilon_t$) and involves the estimation of θ_0 in each bootstrap replication.

Escanciano (2009) suggested a simple multiplier type approximation of $\hat{R}_T^1(x)$ as

$$\hat{R}_T^{*1}(x) = T^{-1/2} \sum_{t=1}^{T} \left[w(Z_{t-1}, x) I_e - \hat{G}'\left(x, \hat{\theta}\right) \hat{\Xi}^{-1} g\left(Z_{t-1}, \hat{\theta}\right) \right] U_t \hat{e}_t(\hat{\theta}),$$

where $\{U_t\}_{t=1}^T$ is a drawing from a pick distribution. This allows the execution of the tests without estimating the parameters in the bootstrap world (see Chapter 2.9 in van der Vaart and Wellner, 1996 for a discussion on multiplier central limit Theorem). This can be seen as an extension of the conventional wild bootstrap and can be applied to possibly non-separable CMR. For any continuous functional $\psi(.)$, Escanciano (2009) proved the consistency of this bootstrap procedure by showing that $\psi\left(\hat{R}_T^{*1}(x)\right) \xrightarrow{d} \psi(R_\infty^1)$.

Escanciano (2009) does mention about GARCH regression models in his paper (see Example 2 in Escanciano, 2009), however details are not provided there. For the AR-GARCH regression model defined in (4.4), to obtain the marginal tests corresponding to the conditional mean and the conditional variance, we construct separate projection for each element of $\hat{e}_t = (\hat{e}_{1t}, \hat{e}_{2t})'$ and then sum the resulting test statistics to obtain the joint specification tests. That is, for $\hat{e}_t = \{\hat{e}_{jt}\}$, $j = 1, 2$ we first construct

$$\hat{R}_{Tj}^1(x) = T^{-1/2} \sum_{t=1}^{T} \left[w(Z_{t-1}, x) I_e - \hat{G}_j'\left(x, \hat{\theta}\right) \hat{\Xi}_j^{-1} g_j\left(Z_{t-1}, \hat{\theta}\right) \right] \hat{e}_{jt},$$

with

$$
\begin{aligned}
g_j\left(Z_{t-1},\hat{\theta}\right) &= \left.\frac{\partial e_{jt}}{\partial \theta}\right|_{\theta=\hat{\theta}}, \\
\hat{G}_j\left(x,\hat{\theta}\right) &= T^{-1}\sum_{t=1}^{T}\left[g_j\left(Z_{t-1},\hat{\theta}\right)w\left(Z_{t-1},x\right)\right], \text{ and} \\
\hat{\Xi}_j &= T^{-1}\sum_{t=1}^{T}\left[g_j\left(Z_{t-1},\hat{\theta}\right)g_j\left(Z_{t-1},\hat{\theta}\right)'\right].
\end{aligned}
$$

The step by step PICM testing procedure is given below:

1. Given the information set $\mathcal{I}_{t-1}=\left(Z'_{t-1},Z'_{t-2},\cdots\right)'$ at time $t-1$, construct the $(T\times T)$ weight matrix $\mathcal{W}$. For example, the (r,s)-th element of indicator weight matrix is $\mathcal{W}_{r,s}=1\left(Z_{r-1}\leq Z_{s-1}\right),\ r,s=1,\cdots,T$.

2. Estimate $\hat{\theta}$ and $\hat{e}_t=\left(\hat{e}_{1t},\hat{e}_{2t}\right)'$ by the QMLE.

3. For $i=1,2$; construct the matrix of derivative $\widetilde{G}_i$ with rows $g_i\left(Z_t,\hat{\theta}\right)'$ where

$$
\begin{aligned}
g_1\left(Z_t,\theta\right) &= \frac{\partial e_{1t}(\theta)}{\partial\theta}=\left(\frac{\partial e_{1t}(\theta)}{\partial\varphi'},\frac{\partial e_{1t}(\theta)}{\partial\eta'}\right)' \\
&= \left(-W_t,\ 0\right)',
\end{aligned}
$$

and

$$
\begin{aligned}
g_2\left(Z_t,\theta\right) &= \frac{\partial e_{2t}(\theta)}{\partial\theta}=\left(\frac{\partial e_{2t}(\theta)}{\partial\varphi'},\frac{\partial e_{2t}(\theta)}{\partial\eta'}\right)' \\
&= \left(\left(-2W_te_{1t}-\frac{\partial h_t}{\partial\varphi'}\right),\ -\frac{\partial h_t}{\partial\eta'}\right)',
\end{aligned}
$$

Remark 4.1 *For the GARCH(1,1) case,* $\frac{\partial h_t}{\partial\varphi}=-2\alpha_1\varepsilon_{t-1}W_{t-1}+\beta_1\frac{\partial h_{t-1}}{\partial\varphi}$ *and* $\frac{\partial h_t}{\partial\eta}=s_{t-1}+\beta_1\frac{\partial h_{t-1}}{\partial\eta}$ *which can be obtained by recursions. For the GARCH (p,q) case this can be generalized easily.*

4. Regress $\hat{e}_i$ on $\widetilde{G}_i$, $i=1,2$ (deleting the columns containing only zeros);

and obtain the residuals as $\widetilde{\widetilde{e}}_i \equiv M_{\widetilde{G}_i}\hat{e}_i$, where $M_{\widetilde{G}_i} = I_T - \widetilde{G}_i \left(\widetilde{G}_i'\widetilde{G}_i\right)^{-1} \widetilde{G}_i'$ is the usual Projection matrix.

5. Calculate $CvM_{T,i} = T^{-2}\widetilde{\widetilde{e}}_i'\mathcal{W}\mathcal{W}'\widetilde{\widetilde{e}}_i = T^{-2}\hat{e}_i'PP'\hat{e}_i$, where $P \equiv M_{\widetilde{G}_i}\mathcal{W}$.

6. In the bootstrap world, generate $\hat{e}_i^* = \{U_t\hat{e}_{it}\}_{t=1}^T$, where $\{U_t\}_{t=1}^T$ is a sequence of i.i.d. draws from a pick distribution, and $CvM_{Ti}^* = T^{-2}\hat{e}_i'^*PP'\hat{e}_i^*$.

7. For joint test, obtain $CvM_{T,J} = \sum_{i=1}^2 CvM_{T,i}$ and $CvM_{T,J}^* = \sum_{i=1}^2 CvM_{T,i}^*$.

8. Reject H_0 at $100\alpha\%$ when bootstrap p-value $p_T^* < \alpha$, where $p_T^* = P\left(CvM_T^* \geq CvM_T \middle| \{y_t, \mathcal{I}_{t-1}\}_{t=1}^T\right)$. In practice, the bootstrap p-value is computed as

$$p_T^* = \frac{\#\left\{CvM_{T,b}^* \geq CvM_T\right\}}{B},$$

where $\#\{A\}$ denotes the number of times that event A occurs, $CvM_{T,b}^*$, $b = 1, \cdots, B$ are the bootstrap realizations of the test statistics and B is the number of bootstrap replications.

4.4 The Integrated Generalized Spectral (IGS) Tests

Most financial data shows highly persistent volatility suggesting that the conditioning set for variance specifications should contain long lags. Again for large lags d, classical consistent tests are affected by "curse of dimensionality". For example, de Jong's (1996) generalization of Bierens (1982) test for $d \to \infty$ as $T \to \infty$ suffers from two drawbacks: it requires numerical integration with dimension T and loss of degrees of freedom due to introducing many lags. In addition popular conditional variance models (such as GARCH, ARCH(∞)) are non-markovian. The IGS testing approach, introduced by Hong (1999) in a non-linear time series framework, is particularly useful when dealing with infinite-dimensional conditioning sets and non-Markovian processes. Hong and Lee (2003, 2005) extended this idea to test the null for processes having conditional dependence at second and higher conditional moments. Escanciano and Velasco (2006) proposed generalized spectral tests for Martingale Difference Hypothesis (MDH) which, unlike Hong and Lee tests, do not depend on kernel and bandwidth parameter and do not require the existence of fourth moment. We introduce the idea of generalized spectral density below.

Definition 4.2 *Generalized Spectral Density (Hong, 1999): For a strictly stationary time series* $\{e_t\}$ *consider the spectrum of the transformed series* $\{e^{iue_t}\}$, *where* $i = \sqrt{-1}$, $u \in (-\infty, \infty)$. *The covariance between* e^{iue_t} *and* $e^{ive_{t-j}}$ *is given by*

$$\begin{aligned} \sigma_j(u,v) &= cov\left(e^{iue_t}, e^{ive_{t-j}}\right) \quad j = 0, \pm 1, \cdots, \\ &= \mathrm{E}\left[e^{i(ue_t + ve_{t-j})}\right] - \mathrm{E}\left[e^{iue_t}\right]\mathrm{E}\left[e^{ive_{t-j}}\right], \end{aligned}$$

where the first component is the joint and second is the product of marginal characteristic functions of (e_t, e_{t-j}). *Thus* $\sigma_j(u,v) = 0$ *for all* $(u,v) \in \Re^2$ *iff* e_t *and* e_j *are independent. Assuming* $\sup_{(u,v)\in\Re^2} \sum_{j=-\infty}^{\infty} \|\sigma_j(u,v)\| < \infty$, *the Fourier transform of* $\sigma_j(u,v)$ *exists:*

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v)\, e^{-ij\omega}, \quad \omega = [-\pi, \pi], \tag{4.14}$$

which contains the same information as contained in $\sigma_j(u,v)$ *and important to note that no moment condition on* e_t *is required. However, when* $var(e_t)$ *exists, the conventional spectral density is obtained from (4.14) as:*

$$-\frac{\partial^2 f(\omega,u,v)}{\partial u \partial v} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} cov(e_t, e_{t-j})\, e^{-ij\omega}.$$

For this reason Hong (1999) termed (4.14) as generalized spectral density. Note that when $\{e_t\}$ *is i.i.d.,* $f(\omega,u,v)$ *becomes a flat generalized spectrum:* $f_0(\omega,u,v) = \frac{1}{2\pi}\sigma_0(u,v)$, $\omega = [-\pi,\pi]$. *Therefore any deviation of* $f(\omega,u,v)$ *from* $f_0(\omega,u,v)$ *provides evidence of serial dependence of* $\{e_t\}$.

Now we briefly outline Escanciano's (2008) joint and marginal IGS tests which apply a pair-wise generalized spectrum approach. Under H_0, we write the joint CMR as:

$$\gamma_j(\theta_0) = \mathrm{E}\left[e_t(\theta_0)\,|Z_{t-j}\right] = 0 \text{ a.s. } \forall j \geq 1 \text{ for some } \theta_0 \in \Theta \subset \Re^p. \tag{4.15}$$

By appropriately choosing a weight function $w(Z_{t-j}, x)$, (4.15) can be written as:

$$\gamma_{j,w}(x, \theta_0) = \mathrm{E}\left[e_t(\theta_0)\, w(Z_{t-j}, x)\right] = 0 \text{ a.e. in } \Pi \subset [-\infty, \infty]^s, \forall j \geq 1, \tag{4.16}$$

where Π is the nuisance parameter space with dimension s which depends on the particular family $\mathcal{W}$ used. To consider simultaneously all dependence measures, define $\gamma_{-j,w}(., \theta_0) = \gamma_{j,w}(., \theta_0)$ for $j \geq 1$ and $\gamma_{0,w}(., \theta_0) = \mathrm{E}\left[e_t(\theta_0)\, w(Z_t, x)\right]$. Then the Fourier transform of the functions $\left\{\gamma_{j,w}(., \theta_0)\right\}_{j=-\infty}^{\infty}$ is

$$f_w(u, x, \theta_0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_{j,w}(x, \theta_0)\, e^{-iju}, \ \forall u \in [-\pi, \pi], \ x \in \Pi. \tag{4.17}$$

Under H_0, it becomes $f_w(u, x, \theta_0) \equiv f_{0,w}(x, \theta_0) = (2\pi)^{-1} \zeta_{0,w}(x, \theta_0)$ and serves as the basis to test the hypothesis (4.5). To avoid the nonparametric smoothed estimation of (4.17) (as proposed by Hong and Lee, 2003 based on Parzen's (1957) smoothed kernel estimators), a generalized spectral distribution function is used based on the dependence measure $\left\{\zeta_{j,w}(., \theta_0)\right\}_{j=-\infty}^{\infty}$ and the test is based on the integral of $f_w(u, x, \theta_0)$:

$$\begin{aligned} Q_w(\tau, x, \theta_0) &= 2\int_0^{\tau\pi} f_w(u, x, \theta_0)\, du, \quad \forall \tau \in [0, 1], \ x \in \Pi \\ &= \gamma_{0,w}(x, \theta_0)\, \tau + 2\sum_{j=1}^{\infty} \gamma_{j,w}(x, \theta_0) \frac{\sin j\pi\tau}{j\pi}. \end{aligned} \tag{4.18}$$

For a sample $\{Y_t, \mathcal{I}_{t-1,q}\}_{t=1}^{T}$, let $\hat{\theta}$ be a $\sqrt{T}$-consistent estimator for θ_0 (e.g., the QMLE) and let

$$\hat{e}_{1t} \equiv e_{1t}\left(\hat{\theta}\right),\ \hat{e}_{2t} \equiv e_{2t}\left(\hat{\theta}\right) \text{ and } \hat{e}_t = \left(\hat{e}_{1t}, \hat{e}_{2t}\right)'. \tag{4.19}$$

Then, sample analogues of (4.16) and (4.18) are, respectively,

$$\hat{\gamma}_{j,w}\left(x, \hat{\theta}\right) = \frac{1}{T_j}\sum_{t=j}^{T} \hat{e}_t w\left(Z_{t-j}, x\right),\ T_j = T - j + 1, \tag{4.20}$$

$$\hat{Q}_w\left(\tau, x, \hat{\theta}\right) = \hat{\gamma}_{0,w}\left(x, \hat{\theta}\right)\tau + 2\sum_{j=1}^{T} \hat{\gamma}_{j,w}\left(x, \hat{\theta}\right)\left(\frac{T_j}{T}\right)^{1/2} \frac{\sin j\pi\tau}{j\pi}, \tag{4.21}$$

where $\left(\frac{T_j}{T}\right)^{1/2}$ is a finite sample correction factor to put lesser weight for larger lag, which has no effect on asymptotic theory but provides better finite sample performance. Under H_0, $Q_w\left(\tau, x, \theta_0\right) = \gamma_{0,w}\left(x, \theta_0\right)\tau$ suggesting a test based on the distance between $\hat{Q}_w\left(\tau, x, \hat{\theta}\right)$ and $\hat{Q}_{0,w}\left(\tau, x, \hat{\theta}\right) = \hat{\gamma}_{0,w}\left(x, \hat{\theta}\right)\tau$. We define the marked empirical process $R_{T,w}\left(\tau, x, \hat{\theta}\right)$ as

$$\begin{aligned} R_{T,w}\left(\tau, x, \hat{\theta}\right) &= \left(\frac{T}{2}\right)^{1/2}\left[\hat{Q}_w\left(\tau, x, \hat{\theta}\right) - \hat{Q}_{0,w}\left(\tau, x, \hat{\theta}\right)\right] \\ &= \sum_{j=1}^{T} T_j^{1/2}\hat{\gamma}_{j,w}\left(x, \hat{\theta}\right)\frac{\sqrt{2}\sin j\pi\tau}{j\pi}. \end{aligned} \tag{4.22}$$

To evaluate the distance from $R_{T,w}$ to zero we need to consider some norms; e.g., Cramer-von Mises (CvM) and Kolmogorov-Smirnov (KS) functionals. Using the CvM norm, the joint specification test statistics are given by

$$\begin{aligned} J_{T,w}^2 &\equiv J_{T,w}^2\left(\hat{\theta}_T\right) = \int_{\Pi'} \left\|R_{T,w}\left(\tau, x, \hat{\theta}\right)\right\|_M^2 W(\mathrm{d}\,x)\,\mathrm{d}\,\tau \\ &= \sum_{j=1}^{T} \frac{T_j}{(j\pi)^2}\int_{\Pi} \left\|\hat{\gamma}_{j,w}\left(x, \hat{\theta}\right)\right\|_M^2 W(\mathrm{d}\,x), \end{aligned}$$

where $\Pi' = [0,1] \times \Pi$, $W(.)$ is an integrating function depending on the weight family $\mathcal{W}$ and M is a 2×2 psd matrix with rows $(m_1, 0)$ and $(0, m_2)$ to obtain marginal components from the joint test (see Assumption 4.4.2). For example, $m_1 = 1$, and $m_2 = 0$, leads to marginal test for mean specification. For any $\sqrt{T}$-consistent estimators $\hat{\theta}$, with indicator weight function $w(Z_{t-j}, x) = 1(Z_{t-j} \leq x)$ and $W(.) = F_T(.)$ where $F_T(.)$ is the empirical distribution function of $\{Z_{t-1}\}_{t=1}^{T}$, the test statistic has the following simple form:

$$J_{T,I}^2 = \sum_{j=1}^{T} \frac{T_j}{T(j\pi)^2} \sum_{t=1}^{T} \left\{ m_1 \hat{\sigma}_{1e}^{-2} \hat{\gamma}_{I,j,m}^2 \left(Z_{t-1}, \hat{\theta}\right) + m_2 \hat{\sigma}_{2e}^{-2} \hat{\gamma}_{I,j,v}^2 \left(Z_{t-1}, \hat{\theta}\right) \right\}, \tag{4.23}$$

where $T_j = T - j + 1$, $\hat{\sigma}_{ie}^2 = T^{-1} \sum_{t=1}^{T} \hat{e}_{it}^2$, $i = 1, 2$ and $\hat{\gamma}_{I,j} = \left(\hat{\gamma}_{I,j,m}, \hat{\gamma}_{I,j,v}\right)'$, with

$$\begin{aligned}
\hat{\gamma}_{I,j,m} \left(Z_{t-1}, \hat{\theta}\right) &= \frac{1}{T_j} \sum_{t=j}^{T} \hat{e}_{1t} w(Z_{t-j}, Z_{t-1}), \\
\hat{\gamma}_{I,j,v} \left(Z_{t-1}, \hat{\theta}\right) &= \frac{1}{T_j} \sum_{t=j}^{T} \hat{e}_{2t} w(Z_{t-j}, Z_{t-1}).
\end{aligned}$$

Note that $(m_1, m_2) = (1, 1)$ gives the joint test whilst setting $(m_1, m_2) = (1, 0)$ and $(m_1, m_2) = (0, 1)$ give the marginal tests for conditional mean $\left(D_{T,I,m}^2\right)$ and conditional variance $\left(D_{T,I,v}^2\right)$, respectively; i.e.,

$$D_{T,I,m}^2 = \sum_{j=1}^{T} \frac{T_j}{T(j\pi)^2} \sum_{t=1}^{T} m_1 \hat{\sigma}_{1e}^{-2} \hat{\gamma}_{I,j,m}^2 \left(Z_{t-1}, \hat{\theta}\right), \tag{4.24}$$

$$D_{T,I,v}^2 = \sum_{j=1}^{T} \frac{T_j}{T(j\pi)^2} \sum_{t=1}^{T} m_2 \hat{\sigma}_{2e}^{-2} \hat{\gamma}_{I,j,v}^2 \left(Z_{t-1}, \hat{\theta}\right). \tag{4.25}$$

Similarly with the complex exponential weight function $w(Z_{t-j}, x) = \exp(ix'Z_{t-j})$ and $W(dx) = \varphi(x)\,dx$ where $\varphi(x)$ is the standard normal density, the test statistic can be expressed as

$$J_{T,C}^2 = \sum_{j=1}^{T} \frac{T_j^{-1}}{(j\pi)^2} \sum_{t=1}^{T} \sum_{s=j}^{T} \left\{ \frac{m_1}{\hat{\sigma}_{1e}^2} \hat{e}_{1t}\hat{e}_{1s} + \frac{m_2}{\hat{\sigma}_{2e}^2} \hat{e}_{2t}\hat{e}_{2s} \right\} \exp\left(-0.5\left(Z_{t-j} - Z_{s-j}\right)^2\right), \tag{4.26}$$

and analogously $D_{T,C,m}^2$ and $D_{T,C,v}^2$ are defined as

$$D_{T,C,m}^2 = \sum_{j=1}^{T} \frac{T_j^{-1}}{(j\pi)^2} \sum_{t=1}^{T} \sum_{s=j}^{T} \frac{m_1}{\hat{\sigma}_{1e}^2} \hat{e}_{1t}\hat{e}_{1s} \exp\left(-0.5\left(Z_{t-j} - Z_{s-j}\right)^2\right), \tag{4.27}$$

$$D_{T,C,v}^2 = \sum_{j=1}^{T} \frac{T_j^{-1}}{(j\pi)^2} \sum_{t=1}^{T} \sum_{s=j}^{T} \frac{m_2}{\hat{\sigma}_{2e}^2} \hat{e}_{2t}\hat{e}_{2s} \exp\left(-0.5\left(Z_{t-j} - Z_{s-j}\right)^2\right). \tag{4.28}$$

4.4.1 Asymptotic Null Distribution and Bootstrap Approximation

To establish the asymptotic theory, in addition to the Assumptions 4.2.1 - 4.2.5, Escanciano (2007b, 2008) made the following assumptions:

Assumption 4.4.1 *Under H_0, $\hat{\theta}$ satisfies the asymptotic Bahadur expansion*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) = T^{-1/2} \sum_{t=1}^{T} \varrho\left(\mathcal{I}_{t-1,q}, \theta_0\right) e_t\left(\theta_0\right) + o_p(1),$$

where $\varrho(.)$ is such that $\mathrm{E}\left[\varrho\left(\mathcal{I}_{t-1,q}, \theta_0\right) e_t\left(\theta_0\right) e_t'\left(\theta_0\right) \varrho'\left(\mathcal{I}_{t-1,q}, \theta_0\right)\right]$ exists and positive definite.

Assumption 4.4.2 *The integrating function $W(.)$ is a probability density function absolutely continuous wrt Lebesgue measure. M is 2×2 psd matrix. The weight function $w(.)$ is such that the equivalence between (4.15) and (4.16) holds and it is uniformly bounded on compacta. Also, $w(.)$ satisfies the following Uniform Law of Large Number (ULLN)*

$$\sup_{x \in \Pi_c} T^{-1} \left\| \sum_{t=1}^{n} \widetilde{\upsilon}_t w\left(\upsilon_t, x\right) - \mathrm{E}\left[\widetilde{\upsilon}_t w\left(\upsilon_t, x\right)\right] \right\| \to 0, \text{ as,}$$

whenever $\{(\widetilde{\upsilon}_t, \upsilon_t), \ t = 0, \pm 1, \cdots\}$ *is strictly stationary and ergodic process with* $\widetilde{\upsilon}_t \in \Re$, $\upsilon_t \in \Re^{1+m}$, $\mathrm{E}[\gamma_1] < \infty$, *and* Π_c *is any compact subset of* $\Pi \subset [-\infty, \infty]^s$.

Assumption 4.4.1 is satisfied under mild conditions for most estimators. Conditions for the local QMLE under martingale conditions have been established in Lee and Hansen (1994). The following Lemma shows that the QMLE $\hat{\theta}$ indeed satisfies Assumption 4.4.1.

Lemma 4.2 *The QMLE* $\hat{\theta} = (\hat{\varphi}', \hat{\eta}')'$ *of (4.3) satisfies*

$$\sqrt{T}(\hat{\theta} - \theta_0) = T^{-1/2} \sum_{t=1}^{T} \varrho_t(\theta_0) e_t(\theta_0) + o_p(1),$$

and $E[\varrho_t(\theta_0) e_t(\theta_0) e_t'(\theta_0) \varrho_t'(\theta_0)]$ *is finite and positive definite, where* $J_{\theta\theta}$ *is the negative of the expected Hessian and*

$$\varrho_t(\theta_0) = J_{\theta\theta}^{-1} h_{0t}^{-1} \begin{bmatrix} f_{0t} & \frac{1}{2} c_{0t} \\ 0 & \frac{1}{2} x_{0t} \end{bmatrix}.$$

with $f_t = \dfrac{\partial m_t}{\partial \varphi}$, $x_t = \dfrac{1}{h_t} \dfrac{\partial h_t}{\partial \eta}$, $c_t = \dfrac{1}{h_t} \dfrac{\partial h_t}{\partial \varphi}$ *and* $f_{0t} = f_t(\varphi_0)$, $c_{0t} = c_t(\theta_0)$, *and* $x_{0t} = x_t(\theta_0)$.

Given these assumptions the limit distribution of $J_{T,w}^2$ under H_0 can be given as $J_{T,w}^2 \xrightarrow{d} J_{\infty,w}^2 = \int |R_w(\tau, x, \theta_0)|_M^2 \, W(\mathrm{d}\,x)\, \mathrm{d}\,\tau$ (for details and proof, see Escanciano 2007b, 2008). To perform the IGS tests, Escanciano (2008) suggested the following FDWB procedure to approximate $R_{T,w}^*\left(\tau, x, \hat{\theta}^*\right) = \sum_{j=1}^{T} T_j^{1/2} \hat{\gamma}_j^*(x) \frac{\sqrt{2}\sin j\pi\tau}{j\pi}$ with

$$\begin{aligned} \hat{\gamma}_j^*(x) &= \left(\hat{\gamma}_{j,m}^*(x), \hat{\gamma}_{j,v}^*(x)\right)' \\ &= \left(\frac{1}{T_j} \sum_{t=j}^{T} \hat{e}_{1t}^* w(Z_{t-j}, x), \frac{1}{T_j} \sum_{t=j}^{T} \hat{e}_{2t}^* w(Z_{t-j}, x)\right)', \end{aligned}$$

where $\hat{e}_t^* = (\hat{e}_{1t}^*, \hat{e}_{2t}^*)'$ are obtained from the following algorithm:

A. Estimate the original model (here by the QMLE) and obtain $\hat{\theta}$, $\hat{m}_t$, $\hat{h}_t$ and $\hat{e}_t$.

B. Generate wild bootstrap residuals as $\hat{\varepsilon}^*_{1t} = \hat{e}_{1t}U_t$, and $\hat{\varepsilon}^*_{2t} = \hat{e}_{2t}U_t$ for $1 \leq t \leq T$ where $\{U_t\}_{t=1}^T$ is a sequence of i.i.d. draws from a pick distribution.

C. Given $\hat{\theta}$, $\hat{\varepsilon}^*_{1t}$ and $\hat{\varepsilon}^*_{2t}$, generate fixed design bootstrap data according to

$$Y^*_{1t} = \hat{m}_t + \hat{\varepsilon}^*_{1t}, \quad Y^*_{2t} = \hat{h}_t + \hat{\varepsilon}^*_{2t} \text{ for } 1 \leq t \leq T.$$

D. Compute $\hat{\theta}^*$ from the bootstrap data $\left\{Y^*_{1t}, Y^*_{2t}, \mathcal{I}'_{t-1,q}\right\}_{t=1}^T$ to construct $\hat{e}^*_{1t} = Y^*_{1t} - \hat{m}^*_t$, $\hat{e}^*_{2t} = Y^*_{2t} - \hat{h}^*_t$ for $1 \leq t \leq T$ where $\hat{m}^*_t \equiv m\left(\mathcal{I}_{t-1,q}, \hat{\theta}^*\right)$ and $\hat{h}^*_t \equiv \hat{h}^*_t\left(\mathcal{I}_{t-1,q}, \hat{\theta}^*\right)$.

The consistency of this FDWB procedure is proved in Escanciano (2007b) under previously stated assumptions A1-A7 and the following conditions on $\hat{\theta}^*$:

Assumption 4.4.3 *The estimator $\hat{\theta}^*$ satisfies the asymptotic expansion*

$$\sqrt{T}\left(\hat{\theta}^* - \hat{\theta}\right) = T^{-1/2}\sum_{t=1}^{T} U_t \varrho\left(\mathcal{I}_{t-1,q}, \hat{\theta}\right)\hat{e}_t + o_p(1).$$

4.5 Problems with Escanciano's FDWB and A Modified Testing Procedure

Since we obtain $\hat{\theta}$ and $\hat{e}_t$ by the QMLE in the real world, ideally we would like to mimic the same estimation procedure in the bootstrap world. We can apply the QMLE in the bootstrap world for the GARCH models by adapting the model based bootstrap used in Pascual et al. (2006) and Christoffersen and Gonclaves (2005). However, this will be computationally costly and cannot be performed with the standard software which may discourage the applied researcher to use these tests. Unfortunately with Escanciano's

FDWB scheme, which is an easier alternative to model based bootstrap, it is problematic to employ the QMLE in the bootstrap world.

This FDWB method, as we will see later, involves generating two separate bootstrap data samples: one provides only the conditional mean structure (and does not include any information about conditional variance structure) while the other provides conditional variance structure. Therefore, for GARCH regression models the QML estimation is not possible in the bootstrap world (since we need a single dependent variable containing both mean and variance structure) whereas we still require the QMLE in the real world to estimate conditional variance parameters. To appreciate the problem associated with this procedure, note that in step C, Y_{1t}^* and Y_{2t}^* provide the conditional mean and variance structure separately in the bootstrap world; as opposed to the single variable y_t in the real world which contain both mean and variance information. The presence of two dependent variables thus restricts the use of the QMLE to compute $\hat{\theta}^*$ in step D.

Remark 4.2 *The nature of the null DGPs considered in Escanciano's (2008) study is the reason for the FDWB working in his simulation study. The null DGPs are:*

$$\begin{aligned} &\text{DGP1:} \qquad y_t = \sqrt{h_t}u_t;\ h_t = a + by_{t-1}^2. \\ &\text{DGP2:} \qquad y_t = ay_{t-1} + \sqrt{h_t}u_t;\ h_t = b + cy_{t-1}^2. \end{aligned}$$

The first one is an ARCH process with no conditional mean and the second one is an AR(1) regression model with conditional heteroskedastic (1) error (in short AR(1)-CH(1) model). Note that unlike our AR-GARCH null model, neither of these DGPs involves lagged unobserved variables, such as ε_{t-1} *or* h_{t-1}*, making it possible to estimate the conditional variance parameters by the OLS both in real and bootstrap world. The author also indicates the application of least squares estimator in his simulation study (Escanciano, 2008, p.82). In addition, we remark that it is not verified whether the OLS estimators satisfy Assumption 4.4.3 in his study.*

Remark 4.3 *Escanciano (2008) further illustrates the IGS test with an em-*

pirical application to the S&P 500 data. In particular, he fits the AR (1)-GARCH(1,1) to the data and finds the evidence that the conditional mean is well specified whereas the conditional variance is misspecified. Though the author does mention that in the real world the QMLE is used to obtain parameter estimates, it is, however, not clear how the parameter estimates in the bootstrap world are obtained. Given his FDWB algorithm, we assume that the OLS is used to obtain $\hat{\theta}^*$.

We have examined the consequences of ignoring this problem of obtaining $\hat{\theta}^*$ by the QMLE and employ Escanciano's FDWB scheme. In particular, in the real world $\hat{\theta}$ is obtained by the QMLE and sample moment conditions are obtained through these estimates. Then in bootstrap world, in step C, the OLS is applied to Y_{1t}^* on $(1,\ \hat{m}_t)$ and Y_{2t}^* on $\left(1,\ \hat{h}_t\right)$ to get bootstrapped moment conditions. Table 4.1 reports the size of the IGS tests with a AR(1)-GARCH(1,1) null model (for details of the DGP, see Section 4.7). Since we are performing two different estimation techniques in the real and bootstrap world, the poor size performance of the IGS tests with this procedure is demonstrated in Monte Carlo experiments as expected which suggests that we need some modifications in the testing procedure.

Table 4.1: Empirical size at 5 per cent nominal level using Escanciano's FDWB: Normal errors

	E2		
T	100	200	300
$D^2_{I,m}$	14.80	21.50	26.20
$D^2_{I,v}$	11.20	8.30	8.70
J^2_I	14.90	13.90	18.10
$D^2_{C,m}$	8.90	10.40	12.00
$D^2_{C,v}$	10.20	7.10	5.70
J^2_C	11.90	9.30	9.50

Notes: 1. Data is generated by the DGP E2 (AR(1)-GARCH(1,1)) as detailed in (4.29).

2.Values in the Table represent the empirical rejection frequencies, against 5% nominal level, of the null (4.5).

3. T is the sample size and results are based on 1,000 Monte Carlo simulations and 300 Bootstrap replications.

In the next subsection we propose a simple bootstrap procedure which provide a solution to this problem.

4.5.1 Modified Test Procedure

To implement the IGS tests for AR-GARCH regression model as defined in (4.4), the idea put forward in this Chapter is a simple one: after the QMLE estimation to obtain $\hat{\theta}$ and $\hat{h}_t$, we consider $\hat{h}_t$ as observed and introduce a set of auxiliary OLS regressions (possibly nonlinear in case of nonlinear specification for mean function) to obtain the moment conditions in the real world; and finally mimicing the same OLS regressions in the bootstrap world. In what follows "hats ($\hat{}$)" denotes the QMLE while "tilda ($\tilde{}$)" denotes the OLS estimation. A step-by-step discussion of the proposed testing procedure for AR(1)-GARCH(1,1) is given below, which can be generalized for nonlinear mean function and higher order/extension of GARCH models in an obvious way.

Real World Estimation

1. Estimate the original model by the QMLE and obtain $\hat{\theta} = (\hat{\varphi}', \hat{\eta}')'$, $\hat{m}_t$, $\hat{h}_t$.

2. Since we have linear conditional mean specification (e.g., AR(1)), regress y_t on a constant and y_{t-1} to obtain

$$\begin{aligned} \tilde{v}_{1t} &= y_t - \tilde{\varphi}_0 - \tilde{\varphi}_1 y_{t-1} = y_t - W_t' \tilde{\varphi} \\ &\equiv y_t - \tilde{m}_t, \end{aligned}$$

   where $\tilde{\varphi} = (\tilde{\varphi}_0, \tilde{\varphi}_1)'$ are OLS estimators from the regression y_t on $W_t = (1, y_{t-1})'$ and $\tilde{m}_t = W_t' \tilde{\varphi}$. See Remark 4.4 for nonlinear mean function.

3. Define $\hat{z}_t = \left(1, \hat{h}_t\right)'$. Then obtain the second sample moment condition as the residual from an OLS regression of $\tilde{v}_{1t}^2$ on $\hat{z}_t$; i.e.,

$$\begin{aligned} \tilde{v}_{2t} &= \tilde{v}_{1t}^2 - \tilde{\beta}_0 - \tilde{\beta}_1 \hat{h}_t = \tilde{v}_{1t}^2 - \hat{z}_t' \tilde{\beta} \\ &\equiv \tilde{v}_{1t}^2 - \tilde{h}_t, \end{aligned}$$

   where $\tilde{\beta} = \left(\tilde{\beta}_0, \tilde{\beta}_1\right)'$ are OLS estimators and $\tilde{h}_t = \hat{z}_t' \tilde{\beta}$.

Bootstrap World

1. Generate WB residuals :

$$\hat{\varepsilon}^*_{1t} = \tilde{v}_{1t} U_t,$$
$$\hat{\varepsilon}^*_{2t} = \tilde{v}_{2t} U_t,$$

   where $\{U_t\}$ a sequence of i.i.d. r.v.s with zero mean and unit variance, bounded support and independent of the sequence $\{y_t, \mathcal{I}_{t-1,q}\}_{t=1}^T$.

2. Generate bootstrap data:

$$Y^*_{1t} = \tilde{m}_t + \hat{\varepsilon}^*_{1t},$$
$$Y^*_{2t} = \tilde{h}_t + \hat{\varepsilon}^*_{2t}.$$

3. Compute $\tilde{\varphi}^* = (\tilde{\varphi}^*_0, \tilde{\varphi}^*_1)'$ from $\{Y^*_{1t}, \mathcal{I}_{t-1,q}\}$ by an OLS regression of Y^*_{1t} on W_t and subsequently first moment condition in the bootstrap world $\tilde{v}^*_{1t} = Y^*_{1t} - \tilde{\varphi}^*_0 - \tilde{\varphi}^*_1 y_{t-1} = Y^*_{1t} - W_t' \tilde{\varphi}^*$.

4. Compute $\tilde{\beta}^* = \left(\tilde{\beta}^*_0, \tilde{\beta}^*_1\right)'$ from $\{Y^*_{2t}, \mathcal{I}_{t-1,q}\}$ by an OLS regression of Y^*_{2t} on $\hat{z}_t$ and obtain $\tilde{v}^*_{2t} = Y^*_{2t} - \tilde{\beta}^*_0 - \tilde{\beta}^*_1 \hat{h}_t = Y^*_{2t} - \hat{z}_t' \tilde{\beta}^*$.

Remark 4.4 *For non-linear conditional mean function, one can employ a non-linear least squares (NLS) method in Step 2 to estimate $\tilde{m}_t$ and $\tilde{v}_{1t}$. Note that we need to perform the same NLS estimation in the bootstrap world for $\tilde{v}^*_{1t}$. Alternatively we can avoid the NLSE by using the $\hat{m}_t$ (obtained from the QMLE estimation at Step 1) to estimate $\tilde{v}_{1t}$ as the residual from a OLS regression of y_t on $(1, \hat{m}_t)$. And then follow the above algorithm in bootstrap world.*

Remark 4.5 *Imitating the GARCH process, we can adopt a slightly different specification for the auxiliary regression in step 3 (real world) to obtain $\tilde{v}_{2t}$ and subsequently in step 4 (bootstrap world) for $\tilde{v}^*_{2t}$ (all other steps remain same). In particular, in the Step 3 (real world) compute $\tilde{\beta} = \left(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2\right)'$ by an OLS regression of $\tilde{v}^2_{1t}$ on $\hat{z}_t = \left(1, \tilde{v}^2_{1,t-1}, \hat{h}_{t-1}\right)'$ and subsequently obtain* *the second moment condition as* $\tilde{v}_{2t} = \tilde{v}_{1t}^2 - \tilde{\beta}_0 - \tilde{\beta}_1 \tilde{v}_{1,t-1}^2 - \tilde{\beta}_2 \hat{h}_{t-1}$*. Note that in this case, with this new definition of* $\tilde{\beta}$ *and* $\hat{z}_t'$*, we have* $\tilde{h}_t = \hat{z}_t'\tilde{\beta}$ *and this will be used in Step 2 (bootstrap world) to generate* Y_{2t}^**. Similarly in Step 4 (bootstrap world), compute* $\tilde{\beta}^* = \left(\tilde{\beta}_0^*, \tilde{\beta}_1^*, \tilde{\beta}_2^*\right)'$ *by an OLS regression of* Y_{2t}^* *on* $\hat{z}_t = \left(1, \tilde{v}_{1,t-1}^2, \hat{h}_{t-1}\right)'$ *and subsequently obtain the second moment condition as* $\tilde{v}_{2t}^* = Y_{2t}^* - \hat{z}_t'\tilde{\beta}^*$*. In an analogous way to Theorem 4.2 (below), it can be shown that in this case also* $\tilde{\beta}$ *satisfies the asymptotic Bahadur expansion.*

The next theorem shows that for the above testing procedure $\tilde{\varphi}$ and $\tilde{\beta}$ in the real world satisfy the asymptotic Bahadur expansion as stated in Assumption 4.4.1.

Theorem 4.2 *Under the stated regularity conditions (Assumptions 4.2.1-4.2.5),*

$$\sqrt{T}\begin{pmatrix} \tilde{\varphi} - \varphi_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} = T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix} q_t(\theta_0) \\ p_t(\theta_0) \end{bmatrix} e_t(\theta_0) + o_p(1),$$

where $\tilde{\varphi} = (\tilde{\varphi}_0, \tilde{\varphi}_1)'$, $\tilde{\beta} = \left(\tilde{\beta}_0, \tilde{\beta}_1\right)'$, $\theta_0' = (\varphi_0', \eta_0')$, *and the expressions for* $q_t(\theta_0)$ *and* $p_t(\theta_0)$ *are provided in the proof.*

The asymptotic analysis shows that, although in the bootstrap world $\tilde{\varphi}^*$ satisfies the sufficient asymptotic expansion, unfortunately $\tilde{\beta}^*$ does not meet the sufficient conditions; i.e., Assumption 4.4.3 (see Appendix for the details). It is found that $\tilde{\beta}^*$ would have satisfied the Bahadur expansion if we could use the true $z_{0t} = (1, h_{0t})'$ instead of $\hat{z}_t = \left(1, \hat{h}_t\right)'$. Since h_{0t} is not observable, we are forced to use $\hat{h}_t$. Therefore, strictly speaking the proposed bootstrap procedure does not satisfy the sufficient conditions. However, this does not mean the procedure is necessarily invalid. We assess the potential validity of our procedure via a simulation study. On the other hand it is an easily implementable solution as opposed to the full parametric model based bootstrap. It is, therefore, worthwhile to investigate the finite sample performance of the IGS tests with our proposed bootstrap scheme.

4.6 CM Tests of the GARCH Specification

In this section we will briefly discuss the parametric CM tests considered in Halunga and Orme (2009) which will be used in our simulation. Assuming the correct specification for the conditional mean, the general CM testing framework of Halunga and Orme (2009) based on the idea that under a correct GARCH specification the squared standardized residuals ζ_t^2 should be serially uncorrelated with any function of the past information:

$$H_0 : \mathrm{E}\left[\left(\zeta_t^2 - 1\right) r_t\left(\theta_0\right)\right] = 0,$$

where is $r_t\left(\theta_0\right)$ is $\mathcal{F}_{t-1}$ measurable. Then the generic CM test indicator is

$$\delta\left(\hat{\theta}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\hat{\zeta}_t^2 - 1\right)\hat{r}_t\right],$$

and the generic form of the test statistic is

$$T = n\delta\left(\hat{\theta}\right)' \hat{\Sigma}_T^{-1}\delta\left(\hat{\theta}\right),$$

where $\hat{\Sigma}_T = \Sigma + o_p\left(1\right)$, Σ is the asymptotic variance-covariance matrix of $\delta\left(\hat{\theta}\right)$ which has a χ_m^2 limiting distribution under the null (see Halunga and Orme, 2009).

Halunga and Orme (2009) also analyzed the Engle and Ng (1993) asymmetry and Lundbergh and Teräsvirta (2002) non-linearity tests and showed that these tests are asymptotically invalid as these do not take account of the asymptotically non-negligible estimation effects from the correct specification of the conditional mean function. Halunga and Orme (2009) also suggested two alternative asymptotically valid tests of asymmetry and non-linearity.

For a AR(1)-GARCH (1,1) regression process; i.e., $y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t,\ h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, in our study we consider the following four parametric CM tests all of which has a χ^2 limit distribution:

1. The Engle and Ng asymmetry (negative size bias) test (T_{EN}) with $\hat{r}_{1t} = \left[1\left(\hat{\varepsilon}_{t-1} \leq 0\right)\right]\hat{\varepsilon}_{t-1}$,

2. The Lundbergh and Teräsvirta non-linearity test (T_{LT}) with $\hat{r}_{2t} = \hat{\varepsilon}^3_{t-1}$,

3. The Halunga and Orme asymmetry (T_A) test with

$$\hat{r}_{3t} = \frac{1}{\hat{h}_t} \sum_{i=0}^{t-1} \hat{\beta}_1^i \left[1\left(\hat{e}_{2,t-1-i} \leq 0\right)\right] \hat{\varepsilon}_{t-1-i},$$

4. The Halunga and Orme nonlinearity (T_N) test with $\hat{r}_{4t} = \frac{1}{\hat{h}_t} \sum_{i=0}^{t-1} \hat{\beta}_1^i \hat{\varepsilon}^3_{t-1-i}$.

The corresponding expressions of $\hat{\Sigma}_T$ for these tests and other details are presented in Halunga and Orme (2009). To be specific T_{EN} and T_{LT} employ expression given by equation (14) in their paper (Halunga and Orme, 2009, p. 375), whereas T_A and T_N employ expressions given by equation (13) and (15), respectively (Halunga and Orme, 2009, p. 375).

4.7 Monte Carlo Experiments

In this section the finite sample performance of previously discussed two nonparametric testing procedures (the PICM and IGS) and four parametric CM tests (T_{LT}, T_{EN}, T_A and T_N) are compared. For both nonparametric testing procedures, we consider two family of weight functions, namely, the indicator and exponential weight functions and we set $Z_{t-1} = y_{t-1}$. The joint and marginal mean and variance IGS tests based on indicator weight function J_I^2, $D_{I,m}^2$, $D_{I,v}^2$, and complex exponential weight functions J_C^2, $D_{C,m}^2$ and $D_{C,v}^2$ are given in (4.23) -(4.28). These IGS tests are constructed employing the proposed modified bootstrap scheme. The alternative specification of the conditional variance auxiliary regression (as mentioned in Remark 4.5) is also considered in the simulation, however the results are qualitatively similar to the former one and to save space we do not report them here.[3] The PICM joint and marginal mean and variance tests with indicator weight are denoted by $C_{I,J}^2$, $C_{I,m}^2$ and $C_{I,v}^2$, respectively, while the corresponding tests with exponential weight are denoted by $C_{C,J}^2$, $C_{C,m}^2$ and $C_{C,v}^2$, respectively. All

[3]These can be obtained from the author upon request.

experiments are done with 1000 Monte Carlo replications and for nonparametric tests 300 bootstrap samples are generated. We consider the sample size T = 100, 200, 300 and 500, after discarding the first 200 observations from the sample to offset any initial value effect. For generating the bootstrap data, we consider the Rademacher distribution given in (4.13).[4] All simulations are programmed in GAUSS.

4.7.1 Size and Robustness to Non-normality

For size experiments we consider the following AR(1)-GARCH(1,1) null models:

$$\begin{aligned} \text{E1} &: Y_t = 1 + 0.1Y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\ h_t &= 0.20 + 0.05\varepsilon_{t-1}^2 + 0.75h_{t-1},\ \zeta_t \sim N(0,1). \\ \text{E2} &: Y_t = 1 + 0.1Y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\ h_t &= 0.01 + 0.09\varepsilon_{t-1}^2 + 0.90h_{t-1},\ \ \zeta_t \sim N(0,1). \end{aligned} \tag{4.29}$$

E1 and E2 correspond to low persistent and high persistent volatility process with $\alpha + \beta = 0.80$ and $\alpha + \beta = 0.99$, respectively, where $\zeta_t \sim N(0,1)$. The parameter values are standard in the literature and are used by Halunga and Orme (2009) and Engle and Ng (1993) among others.

Table 4.2 displays the size of the various tests for a nominal size of 5% and for $T = 100, 200, 300$ and 500, where the null DGPs are E1 and E2 with $\zeta_t \sim N(0,1)$. The parametric CM tests perform poorly for small sample size but size distortions decrease as T increases except T_{LT}, which is the worst performer in terms of size. This finding is similar to Halunga and Orme study where even for $T = 1000$, T_{LT} is significantly undersized. On the other hand the empirical sizes of the IGS tests, for both high and low persistent GARCH process and with both weight functions, are close to the nominal level. Even for very small size, e.g., $T = 100$ and $T = 200$, these tests, in

[4] Similar conclusions are obtained by using the Mammen's distribution as given in (4.12), hence we do not report them.

Table 4.2: Empirical size at 5 per cent nominal level with the Normal errors

	E1				E2			
T	100	200	300	500	100	200	300	500
T_{LT}	0.70	2.60	3.50	2.70	1.70	2.60	2.10	1.80
T_N	1.20	3.60	4.10	4.80	3.80	4.90	3.20	5.10
T_{EN}	3.40	3.70	5.90	4.90	4.50	5.80	4.70	3.70
T_A	12.50	7.70	6.30	6.80	13.00	8.20	8.70	6.60
IGS tests								
$D^2_{I,m}$	6.80	6.30	4.90	5.30	5.40	4.50	5.60	7.00
$D^2_{I,v}$	4.80	3.80	5.90	5.70	5.60	5.60	5.50	4.40
J^2_I	4.60	4.80	5.50	5.90	5.50	6.20	5.70	4.90
$D^2_{C,m}$	7.80	6.00	4.60	5.10	5.50	5.20	6.30	7.20
$D^2_{C,v}$	4.60	3.80	4.80	5.00	5.50	5.10	4.60	3.90
J^2_C	5.00	4.00	5.60	5.30	6.20	5.30	4.30	4.90
PICM tests								
$C^2_{I,m}$	6.30	6.30	5.50	5.20	5.60	4.10	5.20	4.40
$C^2_{I,v}$	7.50	6.70	5.30	4.50	8.50	6.80	6.60	5.80
$C^2_{I,J}$	8.00	6.80	5.30	4.90	8.40	7.10	6.80	5.90
$C^2_{C,m}$	4.80	5.10	5.40	5.60	6.00	4.90	5.10	3.40
$C^2_{C,v}$	6.80	6.60	5.40	4.60	7.50	6.90	6.50	5.30
$C^2_{C,J}$	7.30	6.70	5.60	5.10	7.60	6.60	6.80	4.70

Notes: 1. Data is generated by AR(1)-GARCH(1,1) DGP as detailed in (4.29).

2.Values in the Table represent the empirical rejection frequencies, against 5% nominal level, of the null (4.5).

3. The first, second and third block correspond to the parametric CM, IGS and PICM tests, respectively.

4. T is the sample size and results are based on 1,000 Monte Carlo simulations.

The IGS and PICM tests are based on 300 Bootstrap replications.

general, demonstrate excellent size property. The PICM tests are slightly oversized for smaller sample size, but they perform much better than the parametric tests.

To investigate the robustness of these tests to non-normality, Table 4.3 reports the size, again against 5% nominal level, for E1 and E2 where $\zeta_t \sim t(5)$ and $\zeta_t \sim t(3)$. The performances of the parametric CM tests are worse in this case compared to Gaussian error. T_{LT}, T_N and T_{EN} are undersized while T_A is significantly oversized for both DGPs, although for $T = 500$, T_A and T_N size distortions decrease, as expected. It should be noted that Halunga and Orme (2009) reports that for $T = 1000$, T_A and T_N show satisfactory size under non-normality, however our findings reveal that for smaller sample size these are less robust to non-normality. On the other hand both nonparametric tests display robust size property under non-normality, even for very small T.

4.7.2 Power

For the power experiments we consider 3 types of misspecified models, namely, correct specification for mean but misspecified variance (P1, P2 and P3), misspecified mean and correct specification for variance (P4 and P5) and both

Table 4.3: Empirical size at 5 per cent level with the t(5) and t(3) standardized errors

	E1				E2			
T	100	200	300	500	100	200	300	500
				$\zeta_t \sim t(5)$				
T_{LT}	0.90	1.00	0.90	1.30	1.60	0.50	1.00	0.80
T_N	2.60	2.40	2.10	4.10	1.90	2.00	2.60	4.20
T_{EN}	2.80	2.10	3.20	3.80	2.30	2.80	2.70	2.10
T_A	14.00	10.30	8.00	6.70	16.10	10.30	8.50	9.20
				IGS tests				
$D^2_{I,m}$	5.50	5.70	5.30	4.10	5.60	6.10	5.80	6.80
$D^2_{I,v}$	5.80	5.00	5.00	6.70	4.00	5.30	5.90	4.90
J^2_I	5.80	4.90	5.30	6.20	4.30	5.60	6.00	5.20
$D^2_{C,m}$	5.00	5.40	6.10	6.10	6.50	7.00	6.00	5.20
$D^2_{C,v}$	5.90	5.30	4.90	6.30	4.70	6.30	5.30	5.80
J^2_C	6.40	5.20	5.40	5.40	4.60	6.60	5.70	5.60
				PICM tests				
$C^2_{I,m}$	7.10	5.20	6.30	4.20	4.40	5.70	5.80	5.50
$C^2_{I,v}$	6.30	6.60	7.00	5.40	8.60	6.70	6.80	6.40
$C^2_{I,J}$	6.90	6.60	6.70	5.50	8.50	6.90	7.10	6.30
$C^2_{C,m}$	5.80	5.40	5.20	4.30	4.80	6.60	5.30	5.50
$C^2_{C,v}$	6.10	6.70	6.90	4.90	9.10	6.00	5.30	5.40
$C^2_{C,J}$	7.00	6.40	7.20	4.50	8.50	6.40	6.30	5.40
				$\zeta_t \sim t(3)$				
T_{LT}	0.50	0.60	0.90	0.70	0.50	0.90	0.30	0.70
T_N	0.60	2.10	1.50	1.60	1.00	2.00	1.20	1.80
T_{EN}	1.00	2.40	1.30	1.70	1.70	1.90	1.50	2.10
T_A	15.90	14.40	10.80	4.80	16.40	11.10	9.70	9.60
				IGS tests				
$D^2_{I,m}$	5.60	5.20	5.80	5.60	5.50	6.80	6.20	6.10
$D^2_{I,v}$	4.10	6.40	4.90	5.00	5.30	5.60	4.40	5.20
J^2_I	5.10	5.90	4.00	5.90	5.40	7.70	5.40	5.30
$D^2_{C,m}$	6.00	5.80	5.10	6.60	6.00	6.30	6.40	6.20
$D^2_{C,v}$	5.00	5.40	4.30	4.60	6.00	5.80	5.10	5.00
J^2_C	5.70	6.60	5.00	5.30	6.70	7.00	6.40	5.80
				PICM tests				
$C^2_{I,m}$	5.40	5.10	5.30	4.80	6.10	5.70	5.40	5.40
$C^2_{I,v}$	4.90	5.40	5.90	5.50	5.70	7.50	5.40	6.50
$C^2_{I,J}$	5.40	5.80	6.00	5.60	6.00	7.40	6.20	7.00
$C^2_{C,m}$	5.80	4.80	5.40	5.10	5.40	4.80	6.20	5.00
$C^2_{C,v}$	5.60	4.60	6.20	5.30	7.80	7.30	5.70	4.50
$C^2_{C,J}$	6.80	4.50	5.30	5.20	8.50	7.00	6.10	4.80

Notes: See notes 1-4 of Table 4.2.

conditional mean and variance are misspecified (P6):

$$\begin{aligned}
\text{P1} &: y_t = 1 + 0.1y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\
h_t &= 0.005 + 0.28[|\varepsilon_{t-1}| - 0.23\varepsilon_{t-1}]^2 + 0.7h_{t-1},\ \zeta_t \sim N(0,1). \\
\text{P2} &: y_t = 1 + 0.1y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\
h_t &= -0.23 + 0.9\log(h_{t-1}) + 0.25\left[|\zeta_{t-1}| - 0.3\zeta_{t-1}\right],\ \zeta_t \sim N(0,1). \\
\text{P3} &: y_t = 1 + 0.1y_{t-1} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\
h_t &= 0.9 + 0.1y_{t-1}^2,\ \ \zeta_t \sim N(0,1). \\
\text{P4} &: y_t = 1 + 0.1y_{t-1} + 1.5\sqrt{h_t} + \varepsilon_t, \varepsilon_t = \sqrt{h_t}\zeta_t, \\
h_t &= 0.2 + 0.05\varepsilon_{t-1}^2 + 0.75h_{t-1},\ \zeta_t \sim N(0,1). \\
\text{P5} &: y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\varepsilon_{t-1} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\
h_t &= 0.2 + 0.05\varepsilon_{t-1}^2 + 0.75h_{t-1},\ \zeta_t \sim N(0,1). \\
\text{P6} &: y_t = 1 + 0.1y_{t-1} + 1.5\sqrt{h_t} + \varepsilon_t,\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\
h_t &= 0.005 + 0.28[|\varepsilon_{t-1}| - 0.23\varepsilon_{t-1}]^2 + 0.7h_{t-1},\ \zeta_t \sim N(0,1). \qquad (4.30)
\end{aligned}$$

Among the alternative models corresponding to misspecified variance but correct mean functions, P1 is the AR(1) - GJR(1,1), P2 is the AR(1) - EGARCH(1,1), and P3 is the AR(1) - Conditional heteroskedasticity (CH(1)) model. While P4 (GARCH-in-mean - Null GARCH(1,1)) and P5 (bilinear AR(2) -Null GARCH(1,1)) are the two misspecified conditional mean with correct variance function DGPs. Finally, for P6 (GARCH-in-mean - GJR(1,1) GARCH) both functions are misspecified. The parameter values are again common in the literature; e.g., Halunga and Orme (2009), Lundbergh and Teräsvirta (2002), Engle and Ng (1993), Escanciano (2009), Becker and Hurn (2009) used these alternative models and parameters in their simulation experiments.

Table 4.4 reports the empirical power of the tests for misspecified variance with correct mean models where the nominal size is again 5%. A number of interesting issues to be noticed here. Firstly, in this case it is expected that for the nonparametric tests the joint and marginal variance component (i.e., J^2 and D_v^2 in case of the IGS and C_J^2 and C_v^2 in case of the PICM) would pick up the misspecification in conditional variance while this would have no impact on D_m^2 and C_m^2. The power property indeed reflects this fact as the empirical rejection frequencies of D_m^2 and C_m^2 are close to nominal level of 5%, while the power of D_v^2 and C_v^2 (J^2 and C_J^2) indicate that they pick up the misspecification increasingly as sample size grow. Secondly, the parametric CM tests, particularly T_N and T_A, show very good power properties even with a moderate $T = 500$ and T_{LT} lacks in power in all three cases. Once again, these are supported by the results of Halunga and Orme. Thirdly, for EGARCH(1,1) and CH(1) alternative models (i.e., P2 and P3) the IGS tests demonstrate equally impressive (even better in case of P3) power compared to parametric ones. The PICM tests display slightly lower power than the IGS tests for P1 and P2. However, note that even with $T = 500$ the power of both nonparametric tests is relatively weak in case of P1 (GJR alternative); e.g., below 30%. It is worthwhile to note that the parametric CM tests also perform relatively poorly for GJR alternative. With our small to moderate sample size this is not unusual though, as Engle and Ng (1993, p. 1762) also observe weak power for small sample size and concluded "This weakness is expected as both asymmetric effect and time-varying variance are hard to detect in small samples". Important to note that the power is increasing with T and for large sample sizes (which is the case in most real life situation) we could expect that the power would increase substantially for this type of alternative. Finally, the tests based on indicator weight function generally perform slightly better compared to complex exponential weight function.

Next, Table 4.5 displays the simulated power, against 5% nominal level, when the data is generated by models P4 and P5 (misspecified mean but correct variance) and P6 (both mean and variance are misspecified). For P4 and P5, one would expect that D_m^2 and C_m^2 (J^2 and C_J^2) would pick up the misspecification in conditional mean; and ideally want that D_v^2 and C_v^2 to be robust to this type of misspecification. We cannot, however, be sure about the rejection frequencies of marginal variance tests (i.e., D_v^2 and C_v^2) as the conditional variance specification depends on conditional mean and thereby they may pick the misspecification in mean despite the correct specification of variance. From our Monte Carlo experiments, we can see that J^2 and

D_m^2 (in case of the IGS tests) and C_J^2 and C_m^2 (in case of the PICM tests) demonstrate excellent power with both weight functions. On the other hand the rejection frequencies for D_v^2 and C_v^2 increase with sample size. However, the rejection frequencies for D_v^2 and C_v^2 are well below compared to $D_m^2$and C_m^2, respectively.

The parametric CM tests inherently assume a correct specification of mean function and it is observed that all these tests (except T_{LT}) pick up the misspecification in mean and their rejection frequencies are much higher than D_v^2 and C_v^2. T_{EN} is mostly affected by the misspecification in mean followed by T_N and T_A, whereas T_{LT} surprisingly seems to be insensitive to the conditional mean misspecification. This finding confirms that the parametric CM tests are not robust to the misspecification in mean function and in the presence of mean misspecification these tests erroneously over-reject the null of correct variance specification.

Finally when both mean and variance are misspecified (P6), evidence shows that except T_{LT} all tests, parametric and nonparametric, pick up the misspecification. Since we are using a GJR alternative model for conditional variance, marginal variance tests show relatively low power compared to joint and marginal mean tests.

Table 4.4: Empirical power against 5 per cent nominal level for the DGPs P1, P2 and P3

	P1				P2				P3			
T	100	200	300	500	100	200	300	500	100	200	300	500
T_{LT}	1.80	2.40	1.80	3.80	2.50	5.10	6.90	11.80	5.80	12.60	20.60	39.40
T_N	8.60	21.70	31.60	53.10	15.10	47.50	72.00	92.70	8.30	16.60	27.40	47.30
T_{EN}	4.80	9.10	11.80	18.30	17.40	42.20	62.50	87.40	18.80	37.90	54.00	78.70
T_A	16.60	24.40	34.70	52.10	15.90	26.30	36.60	53.90	23.40	34.60	48.80	70.80
					IGS tests							
$D^2_{I,m}$	5.70	7.50	5.80	7.30	5.70	6.70	6.00	7.80	5.30	5.90	5.20	6.20
$D^2_{I,v}$	8.00	12.10	16.70	25.70	24.00	52.30	70.30	92.50	22.50	41.60	58.50	80.60
J^2_I	7.30	13.20	17.00	24.10	22.60	49.40	65.50	87.30	21.80	40.30	58.60	80.50
$D^2_{C,m}$	5.70	5.70	4.30	6.70	7.50	7.20	5.60	7.90	6.70	6.20	5.30	7.00
$D^2_{C,v}$	9.20	12.50	17.30	24.00	21.40	49.20	67.90	86.50	19.80	36.60	53.70	77.40
J^2_C	8.10	12.90	16.40	23.80	17.60	37.70	49.70	71.90	16.10	32.50	47.40	72.40
					PICM tests							
$C^2_{I,m}$	6.00	5.80	5.30	4.50	5.70	4.30	5.60	5.20	5.90	4.90	5.60	5.60
$C^2_{I,v}$	10.10	13.60	15.70	20.10	19.10	32.50	45.50	61.90	24.90	42.40	58.00	81.10
$C^2_{I,J}$	6.40	8.90	10.20	13.20	18.80	31.70	45.30	62.20	24.60	42.20	57.30	81.20
$C^2_{C,m}$	5.20	5.20	4.80	4.30	5.90	5.00	5.60	6.00	7.10	5.60	4.70	7.00
$C^2_{C,v}$	10.10	12.40	14.30	18.50	14.00	24.30	31.30	40.80	21.30	41.70	54.80	78.20
$C^2_{C,J}$	9.50	11.80	13.10	17.10	13.30	22.80	30.80	39.90	21.50	41.50	54.20	77.60

Notes: 1. The DGPs P1, P2 and P3 correspond to the AR(1)-GJR(1,1), AR(1) - EGARCH (1,1), and AR(1)- CH(1) model, respectively with the parameter values detailed in (4.30).

2. Values in the Table represent the empirical rejection frequencies, against 5% nominal level, of the null (4.5).

3.The first, second and third block correspond to the parametric CM, IGS and PICM tests, respectively.

4. T is the sample size and results are based on 1,000 Monte Carlo simulations and 300 Bootstrap replications (for IGS and PICM tests).

Table 4.5: Empirical power against 5 per cent nominal level for the DGPs P4, P5 and P6

	P4				P5				P6			
T	100	200	300	500	100	200	300	500	100	200	300	500
T_{LT}	1.50	1.90	3.00	3.80	0.80	1.10	1.10	0.90	2.10	4.00	8.40	11.00
T_N	17.70	34.10	48.40	65.20	11.30	20.20	30.30	49.90	7.30	17.90	29.70	51.00
T_{EN}	40.00	60.70	74.40	88.20	6.80	13.00	19.40	34.10	18.20	38.90	56.10	77.10
T_A	14.80	20.90	29.80	39.80	13.00	21.90	33.90	50.50	18.30	31.00	45.70	65.00
IGS tests												
$D^2_{I,m}$	68.30	93.00	97.40	98.90	84.30	93.10	93.80	96.40	43.00	83.50	96.40	99.90
$D^2_{I,v}$	7.00	17.80	28.20	49.50	18.70	25.10	31.50	42.40	6.60	16.60	28.90	52.60
J^2_I	39.80	79.20	94.80	98.50	55.70	88.40	93.40	97.40	20.10	57.10	84.30	99.00
$D^2_{C,m}$	74.10	94.30	98.50	99.50	65.50	89.40	94.50	97.90	32.80	69.30	90.20	99.60
$D^2_{C,v}$	16.70	36.60	57.80	76.20	18.20	24.90	28.90	39.50	8.80	18.70	32.10	51.40
J^2_C	61.90	92.70	98.10	99.60	67.10	90.60	93.90	97.50	61.90	48.80	74.80	95.40
PICM tests												
$C^2_{I,m}$	68.80	91.50	95.40	98.30	46.50	69.50	80.60	89.30	37.80	73.80	90.40	98.40
$C^2_{I,v}$	12.30	30.60	47.00	69.20	31.70	52.10	66.90	81.10	5.90	8.30	9.40	15.50
$C^2_{I,J}$	28.30	48.70	60.60	76.00	35.80	58.60	72.80	86.20	36.20	69.90	85.00	94.70
$C^2_{C,m}$	67.20	91.70	96.90	98.90	48.80	78.30	91.10	97.10	39.20	72.90	86.10	95.90
$C^2_{C,v}$	3.40	7.50	16.10	36.30	21.50	34.60	43.10	56.40	3.40	3.40	4.30	4.50
$C^2_{C,J}$	18.40	28.10	37.40	55.80	33.80	52.90	62.00	70.50	33.80	62.20	76.00	84.80

Notes: 1. The DGPs P4, P5 and P6 correspond to the GARCH-in-mean-GARCH(1,1), bilinear AR(2)-GARCH(1,1), and GARCH-in-mean-GJR(1,1) GARCH model, respectively, with the parameter values detailed in (4.30).

2. Also see notes 2-4 of Table 4.4.

4.8 Empirical Illustration

In this Section, we illustrate the nonparametric and parametric testing methodology to the famous and extensively used S&P 500 daily stock index, which is also a representative set of data for which substantial number of studies used GARCH regression models (see Bollerslev, 1992 and references therein). We consider the daily index data covering the time period July 1, 2004 to July 30, 2010. Therefore we have a sample of 1531 observations. We further subdivide the whole sample period into three two-years sub-period: July1, 2004 to June 30, 2006 (505 observations), July 3, 2006 to June 30, 2008 (501 observations) and July 1, 2008 to July 30, 2010 (525 observations). The choice of these sample periods is motivated from the financial crisis of 2007.

We want to examine the dynamics of the S&P 500 by fitting an AR(1)-GARCH(1,1) model to the data (log returns) and applying our tests to make inference of the null specification. The QML estimates of the parameters along with their standard errors are presented in Table 4.6. Next we apply the four parametric CM tests and two nonparametric tests (i.e., the IGS and PICM tests) and the results are given in Table 4.7. The parametric CM tests do not test the mean specification and we can see a clear disagreement between the nonlinearity tests of Lundbergh and Teräsvirta (2002) and Halunga and Orme (2009) and between asymmetry tests of Engle and Ng (1993) and Halunga and Orme (2009). T_{LT} and T_{EN} do not reject the null of correct conditional variance specification in any of three sub-periods and full samples whereas T_A and T_N reject the null in all periods under consideration; except T_N for 2006-08 in which case we cannot reject the null. The nonparametric tests, on the other hand, give us the scope to test the mean and variance specification simultaneously. It can be seen that for all periods under consideration, the AR(1) specification cannot be rejected as revealed by very large p-values of D_m^2 and C_m^2. This implies that the parametric CM tests are not adversely affected from mean misspecification in this case.

In terms of marginal variance specification tests, both the IGS and PICM tests are mostly in agreement except 2006-08 period. For the overall period 2004-2010, D_v^2 and C_v^2 strongly reject the GARCH(1,1) specification with p-

Table 4.6: The QML parameter estimates of the AR(1)-GARCH(1,1) model for different time periods

	2004-06	2006-08	2008-10	2004-10
$\hat{\varphi}_0$	0.027	0.049	0.085	0.046*
	(0.029)	(0.037)	(0.056)	(0.021)
$\hat{\varphi}_1$	-0.044	-0.105*	-0.089	-0.081*
	(0.046)	(0.048)	(0.047)	(0.027)
$\hat{\alpha}_0$	0.030	0.012*	0.026	0.013*
	(0.020)	(0.006)	(0.015)	(0.004)
$\hat{\alpha}_1$	0.039	0.060*	0.120*	0.084*
	(0.021)	(0.016)	(0.014)	(0.012)
$\hat{\beta}_1$	0.895*	0.930*	0.878*	0.907*
	(0.054)	(0.019)	(0.02)	(0.012)
T	505	501	525	1531

Notes: 1. Figures in the parenthesis are the standard errors

2.* denotes significant at 5% level

3. T is the sample size

values 0.01 and 0.00, respectively. The joint tests also reject the correct joint specification for this period. The results indicate that a AR(1)-GARCH(1,1) model is adequate representation for the first sub-period 2004-06 (i.e., before the financial crisis of 2007) with p-values 0.11, 0.12, 0.29 and 0.21 for D_v^2, J^2, C_v^2 and C_J^2, respectively. Similarly, both the IGS and PICM tests strongly reject the null GARCH(1,1) model for 2008-10 when the financial crisis is in place. However, for 2006-08 the PICM tests reject the null of correct variance and joint specification quite strongly with p-values 0.62 and 0.68, respectively, while the corresponding p-values for the IGS tests are 0.047 (D_v^2) and 0.077 (J^2) which are in borderline of acceptance and rejection region.

In summary, we can see that the AR(1) is an adequate representation of the conditional mean specification for all time periods considered here, though the GARCH (1,1) fits well only in 2004-06. Note that, the nonparametric tests tell us that there is something wrong in the specification (i.e., they are omnibus tests) but do not direct us to the correct model specification. Here comes the importance of parametric test which assume a specific parametric alternative in their construction. In this sense we consider these

Table 4.7: The p-values of various tests for various time periods

	2004-06	2006-08	2008-10	2004-10
T_{LT}	0.1486	0.7700	0.5981	0.5340
T_N	0.0002	0.1276	0.0205	0.0002
T_{EN}	0.5536	0.7025	0.7549	0.8940
T_A	0.0002	0.0000	0.0009	0.0008
		IGS Tests		
D_m^2	0.2600	0.8200	0.4667	0.4500
D_v^2	0.1100	0.0467	0.0033	0.0100
J^2	0.1200	0.0767	0.0100	0.0167
		PICM tests		
C_m	0.1367	0.6733	0.8167	0.7233
C_v	0.2933	0.6233	0.0033	0.0000
C_J	0.2133	0.6833	0.0033	0.0000

nonparametric and parametric testing procedures as complimentary not competing. However, the contradiction among the parametric CM tests about the variance specification is indeed a confusing issue. For example, in our case the asymmetry test of Engle and Ng (1993) never reject the null with very high p-values whereas the Halunga and Orme (2009) T_A strongly rejects GARCH (1,1) in all periods. As noted by Halunga and Orme (2009), the Engle and Ng (1993) and Lundberg and Teräsvirta (2002) tests neglect the recursive behavior of the alternative under consideration and therefore they may lack power. In this particular case, since the nonparametric tests also suggest a misspecification, we need to modify the conditional variance specification.

4.9 Concluding Remarks

In this Chapter, we investigate the nonparametric simultaneous joint and marginal conditional mean and conditional variance specification testing of the null AR-GARCH model. We explicitly demonstrate how to perform the IGS and PICM tests in the model. In particular, we propose a modified wild bootstrap procedure for the IGS tests which performs well in our Monte Carlo study. A number of parametric CM tests for conditional variance, which implicitly assumes a correct conditional mean specification, are also considered and our Monte Carlo simulation confirms that these tests are indeed sensitive to misspecified mean function. The empirical application with the S&P 500 data also highlights the usefulness of the marginal and joint testing within the nonparametric framework.

Our simulation experiments reveal that both the IGS and PICM tests have satisfactory size and impressive robustness to non-normality. The parametric CM tests suffer from size distortion for small T and the distortion is greater under non-normality. Except T_{LT}, size property of other three parametric tests improve as T increase. Further research could focus on using the proposed bootstrap scheme to improve the size properties of the parametric tests.

We want to stress that the inability of our proposed bootstrap procedure to strictly satisfy the identified sufficient conditions does not necessarily mean our test is asymptotically invalid. Our Monte Carlo evidence shows that the size performance of the IGS tests outperforms the parametric tests, particularly T_{LT}, in small samples and, in absolute terms, has satisfactory significance level. The excellent size property under variety of distributional assumptions may indicate that the identified sufficient conditions are too stringent/restrictive and there exists a weaker set of conditions under which the tests are implementable. Searching such less restrictive necessary and sufficient conditions for this test is left for future research.

The power analysis indicates that for correct mean but misspecified variance models, the marginal nonparametric tests demonstrate the ability to identify the source of misspecification with the rejection frequencies of marginal mean tests (D_m^2 and C_m^2) close to the nominal significance level whereas the joint and marginal variance tests pick up the misspecification. The parametric CM tests (barring T_{LT}) also show excellent power in this case. We also note that for GJR alternative (P1) tests have relatively low power, the IGS tests in general demonstrate better power compared to the PICM tests and the indicator weight function performs slightly better than exponential weight function.

For misspecified mean and correct variance, the parametric CM tests incorrectly but unsurprisingly over-reject the null of correct variance specification. In case of the nonparametric tests, D_m^2 and C_m^2 (and their corresponding joint tests) display excellent empirical power. As expected the marginal variance tests pick some of the misspecification through the channel of conditional mean and the rejection frequencies increase as T increase. The PICM tests, in general, suffer more with relatively higher rejection rate. However, in this case D_m^2 and C_m^2 reject significantly more often than D_v^2 and C_v^2. Our suggestion is that whenever D_m^2 or C_m^2 rejects the null one has to revise the mean specification first until the tests provide evidence against mean misspecification and then examine the variance specification.

Finally, the IGS tests can easily be applied to check the adequacy of other extensions of GARCH models without any further modifications. The PICM tests are much quicker and easy to implement for our AR-GARCH model, however for extensions of GARCH models one needs to find the first partial derivatives of the moment conditions under consideration.

Appendix

4.A Appendix A: Proofs

4.A.1 Proof of Lemma 4.2

Proof. Write $\theta' = (\varphi', \eta')$, and let $\hat{\theta}$ be the QMLE for θ_0. Define $f_t = \dfrac{\partial m_t}{\partial \varphi}$, $x_t = \dfrac{1}{h_t}\dfrac{\partial h_t}{\partial \eta}$ and $c_t = \dfrac{1}{h_t}\dfrac{\partial h_t}{\partial \varphi}$. Halunga and Orme (2009) showed that

$$\sqrt{T}(\hat{\theta} - \theta_0) = J_{\theta\theta}^{-1}\sqrt{T} D_{\theta T}(\theta_0) + o_p(1), \tag{4.31}$$

where $D_{\theta T}(\theta_0) = (D_{\varphi T}(\theta), D_{\eta T}(\theta))'$, $D_{\varphi T}(\theta) = T^{-1}\sum_{t=1}^{T}\left\{\dfrac{\varepsilon_t f_t}{h_t} + \dfrac{1}{2}\left(\dfrac{\varepsilon_t^2}{h_t} - 1\right)c_t\right\}$ and $D_{\eta T}(\theta) = T^{-1}\dfrac{1}{2}\sum_{t=1}^{T}\left(\dfrac{\varepsilon_t^2}{h_t} - 1\right)x_t$.

Given regularity conditions, $\hat{\theta} \xrightarrow{p} \theta_0$ and

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, J_{\theta\theta}^{-1}\Omega_{\theta\theta}J_{\theta\theta}^{-1}\right),$$

where $J_{\theta\theta}$ and $\Omega_{\theta\theta}$ are both finite and positive definite, and are defined in Halunga and Orme (2009, Theorem 1).

Define $e_{1t}(\varphi) \equiv \varepsilon_t = y_t - m_t$, $e_{2t} \equiv \varepsilon_t^2 - h_t = e_{1t}^2(\varphi) - \eta' s_{t-1}(\theta)$, and $e_t(\theta) = (e_{1t}(\theta), e_{2t}(\theta))'$, so that

$$\begin{aligned} D_{\varphi T}(\theta) &= T^{-1}\sum_{t=1}^{T} h_t^{-1}\left(f_t, \frac{1}{2}c_t\right)e_t(\theta), \text{ and} \\ D_{\eta T}(\theta) &= T^{-1}\sum_{t=1}^{T} h_t^{-1}\left(0, \frac{1}{2}x_t\right)e_t(\theta). \end{aligned}$$

Now, it is straightforward to show that

$$\sqrt{T}(\hat{\theta} - \theta_0) = T^{-1/2}\sum_{t=1}^{T}\varrho_t(\theta_0)e_{0t} + o_p(1), \tag{4.32}$$

and $E\left[\varrho_{0t}e_{0t}e_{0t}'\varrho_{0t}'\right] = J_{\theta\theta}^{-1}\Omega_{\theta\theta}J_{\theta\theta}^{-1}$ is finite and positive definite, where $e_{0t} = e_t(\theta_0)$ and $\varrho_{0t} \equiv \varrho_t(\theta_0) = J_{\theta\theta}^{-1}h_{0t}^{-1}\begin{bmatrix} f_{0t} & \frac{1}{2}c_{0t} \\ 0 & \frac{1}{2}x_{0t} \end{bmatrix}$, $f_{0t} = f_t(\varphi_0)$, $c_{0t} = c_t(\theta_0)$, and $x_{0t} = x_t(\theta_0)$. ■

4.A.2 Proof of Theorem 4.2

Proof. For simplicity of exposition, we will assume a linear mean specification so that $m(w_t;\varphi) = w_t'\varphi$. In the test procedures, the estimators $\tilde{\varphi}$ and $\tilde{\beta}$ are used as follows:

(a) $\tilde{\varphi}$ is obtained from a (possibly non-linear) least squares regression of y_t on w_t' and residuals $\tilde{v}_{1t} \equiv v_{1t}(\tilde{\varphi}) = y_t - w_t'\tilde{\varphi}$ are obtained and used in the construction of the various test statistics (rather than $e_{1t}(\hat{\varphi})$), for the construction of $\tilde{\beta}$, next, and when implementing a wild bootstrap scheme (see Section 4.A.3, below).

(b) Let $\hat{h}_t = h_t(\hat{\theta})$ be constructed using the QMLE, $\hat{\theta}$, and define $\hat{z}_t' = \left(1, \hat{h}_t\right)$. Then $\tilde{\beta}$ is obtained as the (2×1) least squares parameter estimator following a regression of $\tilde{v}_{1t}^2$ on $\hat{z}_t'$. Following this regression the residuals $\tilde{v}_{1t}^2 - \hat{z}_t'\tilde{\beta}$ are obtained and used in the construction of the various test statistics (rather than $e_{2t}(\hat{\theta})$), and when implementing a wild bootstrap scheme (see Section 4.A.3, below).

For the above estimators, the corresponding "true" parameter values are φ_0 and $\beta_0 \equiv (0,1)'$, respectively. Ideally, to obtain $\tilde{\beta}$, we would like to regress $\tilde{v}_{1t}^2$ on $z_{0t}' = (1, h_{0t})$, but h_{0t} is unobservable so we use $\hat{h}_t$ instead. Because of this, the residual associated with the estimation of $\tilde{\beta}$ depends upon $\hat{\theta}$, through $\hat{h}_t$, so this must be taken into account. In addition, the moment errors must be defined in terms of the parameters being estimated in the least squares procedures. Accordingly, let $\lambda = (\varphi', \beta')'$ and define the following "second moment" error

$$v_{2t}(\lambda) = v_{1t}^2(\varphi) - \hat{z}_t'\beta = v_{1t}^2(\varphi) - z_{0t}'\beta - (\hat{z}_t - z_{0t})'\beta.$$

The corresponding residual would then be

$$\upsilon_{2t}(\tilde{\lambda}) = \upsilon_{1t}^2(\tilde{\varphi}) - \hat{z}_t'\tilde{\beta} = \upsilon_{1t}^2(\tilde{\varphi}) - z_{0t}'\tilde{\beta} - (\hat{z}_t - z_{0t})'\,\tilde{\beta}.$$

It is important to make the distinction between the υ_{jt} used here and the e_{jt} defined previously, $j = 1, 2$, because although it is true that $\upsilon_{1t}(\varphi) \equiv e_{1t}(\varphi)$, for all φ, it is not true that $\upsilon_{2t}(\lambda) = \upsilon_{1t}^2(\varphi) - \hat{z}_t'\beta$ is the same as $e_{2t}(\theta)$. In the ensuing analysis it will be useful to define $\upsilon_t(\lambda) = (\upsilon_{1t}(\varphi), \upsilon_{2t}(\lambda))'$.

First consider $\tilde{\varphi}$. The least squares regression of y_t on w_t' yields $\tilde{\varphi}$, which satisfies

$$\begin{aligned}\sqrt{T}\,(\tilde{\varphi} - \varphi_0) &= \hat{Q}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t\upsilon_{1t}(\varphi_0) + o_p(1)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} q_t\upsilon_t(\lambda_0) + o_p(1),\end{aligned} \tag{4.33}$$

where $\hat{Q}_T = T^{-1}\sum_{t=1}^{T} w_t w_t'$, $q_t = Q^{-1}[w_t,\ 0]$, $Q = \mathrm{E}[w_t w_t']$, and $\mathrm{E}[q_t\upsilon_t(\lambda_0)\upsilon_t'(\lambda_0)q_t'] = Q^{-1}\,\mathrm{E}[\varepsilon_{0t}^2 w_t w_t']\,Q^{-1}$ is finite and positive definite. From this, residuals $\tilde{\upsilon}_{1t} \equiv \upsilon_{1t}(\tilde{\varphi}) = y_t - w_t'\tilde{\varphi}$ are obtained and used in the construction of the various test statistics, for the construction of $\tilde{\beta}$.

Next consider $\tilde{\beta}$. Now regress $\tilde{\upsilon}_{1t}^2 = (y_t - w_t'\tilde{\varphi})^2$ on $\hat{z}_t'$ to give

$$\tilde{\beta} = \hat{V}_T^{-1}T^{-1}\sum_{t=1}^{T}\hat{z}_t\tilde{\upsilon}_{1t}^2, \tag{4.34}$$

where $\hat{V}_T = T^{-1}\sum_{t=1}^{T}\hat{z}_t\hat{z}_t'$.

Note that:

$$\tilde{\upsilon}_{1t}^2 = \upsilon_{1t}^2(\varphi_0) + \tilde{\xi}_t, \text{ where } \tilde{\xi}_t = 2w_t'\upsilon_{1t}(\varphi_0)\,(\tilde{\varphi} - \varphi_0) + (\tilde{\varphi} - \varphi_0)'\,w_t w_t'\,(\tilde{\varphi} - \varphi_0)\,,$$

and

$$\upsilon_{1t}^2(\varphi_0) = \hat{z}_t'\beta_0 + \upsilon_{2t}(\lambda_0) = \hat{z}_t'\beta_0 + \left(\upsilon_{1t}^2(\varphi_0) - h_{0t}\right) - (\hat{z}_t - z_{0t})'\,\beta_0,$$

where, recall, $\beta_0 \equiv (0,1)'$, and $h_{0t} = z'_{0t}\beta_0$. Using these in (4.34), we obtain

$$\begin{aligned}\tilde{\beta} &= \hat{V}_T^{-1}T^{-1}\sum_{t=1}^{T}\hat{z}_t\left(\upsilon_{1t}^2(\varphi_0)+\tilde{\xi}_t\right)\\ &= \beta_0+\hat{V}_T^{-1}T^{-1}\sum_{t=1}^{T}\hat{z}_t\left(\upsilon_{1t}^2(\varphi_0)-z'_{0t}\beta_0-(\hat{z}_t-z_t)'\beta_0+\tilde{\xi}_t\right),\end{aligned}$$

so that

$$\begin{aligned}\sqrt{T}(\tilde{\beta}-\beta_0) &= \hat{V}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{z}_t\left(\upsilon_{1t}^2(\varphi_0)-z'_{0t}\beta_0-(\hat{z}_t-z_{0t})'\beta_0\right)+o_p(1)\\ &= \hat{V}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{z}_t\upsilon_{2t}(\lambda_0)+o_p(1), \qquad (4.35)\end{aligned}$$

which exploits $\sqrt{T}$ consistency of $\hat{\varphi}$ and $\tilde{\varphi}$ and a ULLN which ensures that $\hat{V}_T = O_p(1)$ and $T^{-1}\sum_{t=1}^{T}\hat{z}_t\tilde{\xi}_t = o_p(1)$.

However, $\upsilon_{2t}(\lambda_0) = \upsilon_{1t}^2(\varphi_0) - \hat{z}'_t\beta_0$ depends on $\hat{\theta}$, through $\hat{z}_t$, and this must be taken into account when analyzing the asymptotic sampling distribution of $\sqrt{T}(\tilde{\beta}-\beta_0)$.

Firstly, then, since $\hat{h}_t - h_{0t} = \bar{h}_t\,(\bar{c}'_t,\ \bar{x}'_t)\left(\hat{\theta}-\theta_0\right)$, where here a "bar" indicates evaluation at the mean value $\bar{\theta}$, we have

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{h}_t\upsilon_{2t}(\lambda_0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_{0t}\upsilon_{2t}(\lambda_0)+T^{-1}\sum_{t=1}^{T}\upsilon_{2t}(\lambda_0)\bar{h}_t\,(\bar{c}'_t,\ \bar{x}'_t)\sqrt{T}\left(\hat{\theta}-\theta_0\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}h_{0t}\upsilon_{2t}(\lambda_0)+o_p(T^{-1/2}),\end{aligned}$$

since by consistency $\hat{\theta}$ and a ULLN, $T^{-1}\sum_{t=1}^{T}\upsilon_{2t}(\lambda_0)\bar{h}_t\,(\bar{c}'_t,\ \bar{x}'_t) = o_p(1)$. Thus

$$T^{-1/2}\sum_{t=1}^{T}\hat{z}_t\upsilon_{2t}(\lambda_0) = T^{-1/2}\sum_{t=1}^{T}z_{0t}\upsilon_{2t}(\lambda_0)+o_p(1),$$

and substituting this into (4.35) yields$\sqrt{T}(\tilde{\beta}-\beta_0) = \hat{V}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_{0t}\upsilon_{2t}(\lambda_0)$, so that

$$\begin{aligned}\sqrt{T}(\tilde{\beta}-\beta_0) &= \hat{V}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{0t}\left(\upsilon_{1t}^2(\varphi_0)-h_{0t}\right) \\ &\quad -\hat{V}_T^{-1}\sqrt{T}\left(\hat{V}_T - T^{-1}\sum_{t=1}^{T}\hat{z}_t z_{0t}'\right)\beta_0 + o_p(1). \qquad (4.36)\end{aligned}$$

Now, consider how we might express the second term. We have

$$\begin{aligned}\sqrt{T}\left(\hat{V}_T - T^{-1}\sum_{t=1}^{T}\hat{z}_t z_{0t}'\right)\boldsymbol{\beta}_0 &= \left(T^{-1}\sum_{t=1}^{T}\hat{z}_t\left(\hat{z}_t - z_{0t}\right)'\right)\begin{pmatrix}0\\1\end{pmatrix} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{z}_t\left(\hat{h}_t - h_{0t}\right) \\ &= \left(T^{-1}\sum_{t=1}^{T}\hat{z}_t\bar{h}_t\left(\bar{c}_t',\ \bar{x}_t'\right)\right)\sqrt{T}\left(\hat{\theta}-\theta_0\right) \\ &= A\sqrt{T}\left(\hat{\theta}-\theta_0\right) + o_p(1),\end{aligned}$$

where consistency of $\hat{\theta}$ and a ULLN will ensure that $T^{-1}\sum_{t=1}^{T}\hat{z}_t\bar{h}_t\left(\bar{c}_t',\ \bar{x}_t'\right) = A + o_p(1)$. In addition, given sufficient regularity $\hat{V}_T = V + o_p(1)$, where $V = \mathrm{E}\left[z_{0t}z_{0t}'\right]$.

Substituting these results and (4.32) into (4.36) we obtain the following expression

$$\begin{aligned}\sqrt{T}(\tilde{\beta}-\beta_0) &= V^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{0t}\left(\upsilon_{1t}^2(\varphi_0)-h_{0t}\right) - V^{-1}A\sqrt{T}\left(\hat{\theta}-\theta_0\right) + o_p(1) \\ &= V^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{0t}\left(\upsilon_{1t}^2(\varphi_0)-h_{0t}\right) - V^{-1}A\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varrho_{0t}e_{0t} + o_p(1).\end{aligned}$$

Then, we have the following asymptotic expansion

$$\sqrt{T}\begin{pmatrix}\tilde{\varphi}-\varphi_0\\ \tilde{\beta}-\beta_0\end{pmatrix} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\begin{bmatrix}q_t\\ \overline{p}_t\end{bmatrix}\begin{bmatrix}\upsilon_{1t}(\varphi_0)\\ \upsilon_{1t}^2(\varphi_0)-z_{0t}'\beta_0\end{bmatrix}\right\} - \nu_T(\theta_0) + o_p(1), \qquad (4.37)$$

where $\overline{p}_t = V^{-1}(0, z_{0t})$ and $\nu_T(\theta_0) = \begin{bmatrix} 0 \\ V^{-1}A\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varrho_{0t}e_{0t} \end{bmatrix}$. This expansion is crucial when comparison is made to the corresponding bootstrap expansion.

Finally, however, since $\upsilon_{1t}(\varphi_0) \equiv e_{1t}(\varphi_0)$ and $\upsilon_{1t}^2(\varphi_0) - z_{0t}'\beta_0 \equiv e_{2t}(\theta_0)$, we can write

$$\sqrt{T}\begin{pmatrix} \tilde{\varphi} - \varphi_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\begin{bmatrix} q_t \\ p_t \end{bmatrix}\begin{bmatrix} e_{1t}(\varphi_0) \\ e_{2t}(\theta_0) \end{bmatrix}\right\} + o_p(1)$$

where $p_t = V^{-1}([0, z_{0t}] - A\varrho_{0t})$. ■

4.A.3 Analysis of the Wild Bootstrap Scheme

The bootstrap estimator, denoted $\tilde{\lambda}^*$, is obtained from bootstrap data generated as follows.

Following the least squares procedures, defined above, which yield $\tilde{\varphi}$ and $\tilde{\beta}$, respectively, generate the following bootstrap data, where U_t are i.i.d. random variables with zero mean and unit variance:

1. $Y_{1t}^* = w_t'\tilde{\varphi} + \varepsilon_{1t}^*$, where $\varepsilon_{1t}^* = U_t\tilde{\upsilon}_{1t}$ and $\tilde{\upsilon}_{1t} = \upsilon_{1t}(\tilde{\varphi})$.
2. $Y_{2t}^* = \hat{z}_t'\tilde{\beta} + \varepsilon_{2t}^*$, where $\varepsilon_{1t}^* = U_t\tilde{\upsilon}_{2t}$ and $\tilde{\upsilon}_{2t} = \upsilon_{2t}(\tilde{\lambda}) = \upsilon_{1t}^2(\tilde{\varphi}) - \hat{z}_t'\tilde{\beta}$.

To obtain $\tilde{\varphi}^*$ regress Y_{1t}^* on w_t to obtain

$$\begin{aligned} \sqrt{T}(\tilde{\varphi}^* - \tilde{\varphi}) &= \left(T^{-1}\sum_{t=1}^{T}w_tw_t'\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_t\varepsilon_{1t}^* \\ &= \hat{Q}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_tU_t\upsilon_{1t}(\tilde{\varphi}), \end{aligned} \tag{4.38}$$

where we have used the fact that conditionally on the sample data, in the bootstrap world $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_t\varepsilon_{1t}^*$ is bounded in probability. In particular if $E^*(.)$ and $var^*(.)$ denote expectation and variance in the bootstrap world, respectively, conditional on the sample data, then $E^*(w_t\varepsilon_{1t}^*) = 0$, because

$\mathrm{E}^*(U_t) = 0$, and

$$var^* \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \varepsilon_{1t}^* \right) = \frac{1}{T} \sum_{t=1}^{T} \upsilon_{1t}^2(\tilde{\varphi}) w_t w_t'.$$

Note that the expansion (4.38) agrees with the expansion for $\sqrt{T}(\tilde{\varphi} - \varphi_0)$ but with $U_t \upsilon_{1t}(\tilde{\varphi})$ replacing $\upsilon_{1t}(\varphi_0)$ in the right hand side.

To obtain $\tilde{\beta}^*$, regress Y_{2t}^* on $\hat{z}_t$ to obtain

$$\begin{aligned} \sqrt{T}\left(\tilde{\beta}^* - \tilde{\beta}\right) &= \hat{V}_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{z}_t \varepsilon_{2t}^* \\ &= \hat{V}_T^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{z}_t U_t \upsilon_{2t}(\tilde{\lambda}), \end{aligned} \tag{4.39}$$

where, again, conditionally on the sample data, in the bootstrap world $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \varepsilon_{2t}^*$ is bounded in probability, with

$$var^* \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{z}_t \varepsilon_{2t}^* \right) = \frac{1}{T} \sum_{t=1}^{T} \upsilon_{2t}^2(\tilde{\lambda}) \hat{z}_t \hat{z}_t'.$$

Therefore, although (4.38) is of the same form as (4.33) but with $\tilde{\lambda}$ replacing λ_0, in the bootstrap world $\hat{z}_t$ will be regarded as fixed, in the definition of $\upsilon_{2t}(\tilde{\lambda})$ for each simulation of U_t, rather than varying as it is in the definition of $\upsilon_{2t}(\lambda_0) = \upsilon_{1t}^2(\varphi_0) - \hat{z}_t' \beta_0$ employed in (4.35).

If we were able to observe h_{0t} and so could employ z_{0t} rather than $\hat{z}_t$ there would be no such discrepancy. However, as we do use $\hat{z}_t$ strictly speaking the bootstrap procedure does not satisfy the sufficient conditions.

Chapter 5

Simultaneous Nonparametric Moment Tests of Multivariate GARCH Models

5.1 Introduction

Multivariate volatility modelling and forecasting have played an important role in economics and finance including risk management, asset pricing theory, portfolio selection and deciding monetary policy. Investigating the adequacy of the multivariate GARCH, or MGARCH, specification and particularly the interaction of economic and financial variables (e.g., among financial markets or assets, between GDP growth and asset prices) by examining the conditional covariance structure is of paramount importance. Chapter 2 reviewed some of the most popular MGARCH models including VEC, BEKK, CCC and DCC models. As emphasized by Kroner and Ng (1998), the literature on MGARCH models reveals that despite the escalating attention in designing various MGARCH models, relatively little interest has been paid to construct tests for the adequacy of these models.

The available parametric diagnostic tests of MGARCH models can be broadly distinguished into two groups, namely univariate tests applied independently to each variable and multivariate tests applied to the whole vector data series. Although the univariate tests shed some light, due to the dependence among the variables, these univariate test statistics are not independent which in turn raises the issue of size control while combining test decisions over all equations. Hence it is desirable to test all the equations jointly (Dufour et al., 2003). As noted by Tse (2002), the multivariate diagnostic tests proposed in the literature for MGARCH models can be divided into three categories: Box-Pierce-Ljung type portmanteau tests (e.g., Ling and Li, 1997), residual-based diagnostics (e.g., Tse, 2002) and LM type tests (e.g., Bollerslev et al., 1988; Engle and Kroner, 1995, Sentana and Fiorentini, 2001). However, these tests can be seen as *general* misspecification tests in the sense that they are useful to check the adequacy of an estimated model and are not designed to test the model against a particular parametric alternative. Thus if the null of the correct specification is rejected one cannot detect the source of misspecification; i.e., whether the misspecification is due to conditional mean, conditional variance, covariance or any combination of them without employing other testing device. Another group of diagnostic tests consider the adequacy of the correlation assumption (e.g., CCC and ECCC) instead of directly checking the covariance structure of the MGARCH model (see Chapter 3 for a discussion on these tests). However, to the best of our knowledge there is no test available in the literature which consider the conditional covariance between variables directly.

On the other hand only a few studies have investigated the issue of testing MGARCH models nonparametrically. For example, to detect multivariate ARCH effects, Duchesne and Lalancette (2003) proposed a test using a spectral approach based on a comparison, via a quadratic norm, between the uniform density and a kernel-based spectral density estimator of the squared and cross products of residuals. By extending Hong's (1999) univariate generalized spectral analysis, Hong and Lee (2005) proposed a class of omnibus testing procedures for univariate and multivariate volatility models. McCloud and Hong (2008) suggested a class of generally applicable specification tests for constant and dynamic structures of conditional correlations in MGARCH models, which again can be seen as an extension of Hong's (1999) univariate generalized spectral tests. These tests are consistent with standard asymptotic null distributions and importantly all of them are based on a *local* approach; i.e., they require a nonparametric smoothing of the data. We have not found any study which deals the integrated approach of consistent testing of MGARCH models. In addition, simultaneous nonparametric testing of conditional mean, variance and covariance has yet not been investigated in the literature.

5.1.1 Contributions and Plan of the Chapter

In this chapter we extend the nonparametric testing procedures of the univariate GARCH models discussed in Chapter 4, namely the Projected Integrated Conditional Moment (PICM) tests and Integrated Generalized Spectral (IGS) tests, to the multivariate context. Specifically this Chapter will investigate whether such extensions are useful for assessing the conditional mean, variance and covariance specification simultaneously. Unlike the parametric tests of the correlation assumption which consider a particular null correlation model in advance (e.g., the CM tests considered in Chapter 3, the Tse's (2000) LM test, the Nakatini and Teräsvirta (2009) LM test, the Bera and Kim (2002) IM test), these nonparametric tests are applicable to any MGARCH model. In this study, to focus the discussion, we illustrate the procedure with the CCC-GARCH model as one of the leading cases of interest. This can be generalized to other MGARCH specifications. To implement the IGS tests we extend the wild bootstrap technique proposed in the previous Chapter to accommodate the conditional covariance structure. Via a Monte Carlo study, we assess the ability of this bootstrap procedure to control size and report on its power properties.

The remaining of the Chapter is organized as follows: the null model and moment conditions are presented in next section. The multivariate extensions of the PICM and IGS tests are discussed next. Finally we investigate the finite sample properties of the nonparametric tests via Monte Carlo experiments.

5.2 The Null Model and Moment Conditions

Suppose $\{(y_t, X_t')'\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process on a probability space $(\Omega, \mathcal{F}, P)$ with the $(N\times 1)$ vector of dependent variables $y_t = (y_{1t}, \cdots, y_{Nt})'$ and $Z_{t-1} = (y_{t-1}, X_t')' \in \mathbb{R}^m, m \in \mathbb{N}$, is the explanatory random vectors containing lagged values of y_t and possibly other variables. The conditioning set at time $t-1$ is denoted by $\mathcal{I}_{t-1} = \left(Z_{t-1}', Z_{t-2}', \cdots\right)'$ and $\mathcal{F}_{t-1} = \sigma\left(Z_{t-1}', Z_{t-2}', \cdots\right)$ is the $\sigma-$field generated by the past information up to and including time $t-1$. Note that $Z_{t-1} = (Z_{1,t-1}, \cdots, Z_{N,t-1})$ where $Z_{i,t-1} = (y_{i,t-1}, X_{it}')'$.

Define the conditional mean and conditional variance-covariance $m\left(\mathcal{I}_{t-1}\right) \equiv m_t = \mathrm{E}\left[y_t | \mathcal{I}_{t-1}\right]$, $H\left(\mathcal{I}_{t-1}\right) \equiv H_t = cov\left[y_t y_t' | \mathcal{I}_{t-1}\right]$. Then for the i-th dependent variable y_{it}, $i = 1, \cdots N$, $t \in \mathbb{Z}$, we have the conditional mean, variance and standardized error $m_i\left(\mathcal{I}_{t-1}\right) = \mathrm{E}\left[y_{it} | \mathcal{I}_{t-1}\right]$, $h_i\left(\mathcal{I}_{t-1}\right) \equiv h_{it} = Var\left[y_{it} | \mathcal{I}_{t-1}\right]$ and $\zeta_{it} = \frac{(y_{it} - m_i(\mathcal{I}_{t-1}))}{\sqrt{h_i(\mathcal{I}_{t-1})}}$, respectively. The conditional covariance between y_{it} and y_{jt} is defined as

$$h_{ij}\left(\mathcal{I}_{t-1}\right) \equiv h_{ijt} = cov\left[y_{it} y_{jt} | \mathcal{I}_{t-1}\right],\ i \neq j.$$

Hence the conditional variance-covariance matrix is given by

$$\begin{aligned} H_t\left(\mathcal{I}_{t-1}\right) &= cov\left[y_t y_t' | \mathcal{I}_{t-1}\right] \\ &= \begin{bmatrix} h_1\left(\mathcal{I}_{t-1}\right) & \cdots & h_{1i}\left(\mathcal{I}_{t-1}\right) & \cdots & h_{1N}\left(\mathcal{I}_{t-1}\right) \\ \vdots & & & & \vdots \\ h_{i1}\left(\mathcal{I}_{t-1}\right) & \cdots & h_i\left(\mathcal{I}_{t-1}\right) & \cdots & h_{iN}\left(\mathcal{I}_{t-1}\right) \\ \vdots & & & & \vdots \\ h_{N1}\left(\mathcal{I}_{t-1}\right) & \cdots & h_{Ni}\left(\mathcal{I}_{t-1}\right) & \cdots & h_N\left(\mathcal{I}_{t-1}\right) \end{bmatrix}. \end{aligned}$$

In this Chapter we consider the following parametric MGARCH specification:

$$\begin{aligned} y_t &= m\left(\mathcal{I}_{t-1,q}; \varphi_0\right) + \varepsilon_{0t}, \quad t = 1, ..., T, \\ \varepsilon_{0t} &= H_t^{1/2}(\varpi_0)\xi_t, \end{aligned} \tag{5.1}$$

where $\mathcal{I}_{t-1,q} = \{Z_{s-1}\}_{s=1}^{t-q}$, $q < \infty$ and $q \in \mathbb{N}$, $\varphi \subset \Re^{NK}$ is a $(NK \times 1)$ vector of conditional mean parameters, $H_t^{1/2}$is a $N \times N$ positive definite matrix such that $H_t(\varpi_0) = Var(\varepsilon_{0t}|\mathcal{I}_{t-1})$ and $H_t^{1/2}$ may be obtained from Cholesky factorization, $\varpi \in \Theta \subset \Re^p$ is the vector of unknown parameters which includes φ, and ξ_t is an i.i.d. process with $\mathrm{E}(\xi_t) = 0$ and $\mathrm{E}(\xi_t\xi_t') = I_N$. This specification is quite general and covers various well-known MGARCH models such as BEKK, Constant Conditional Correlation (CCC), Dynamic Conditional Correlation (DCC) models and many others.

We aim to test simultaneously the conditional mean, variance and covariance specification of the model (5.1). To focus the discussion, we consider a bivariate AR(1)-CCC GARCH model as a special case of model (5.1); i.e., for $i = 1, 2$,

$$\begin{aligned} y_{it} &= m_{it}+\varepsilon_{it} = W_{it}\varphi_i + \varepsilon_{it} = \varphi_{i0} + \varphi_{i1}y_{i,t-1}+\varepsilon_{it}, \\ \varepsilon_t &= H_t^{1/2}\xi_t,\ \varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})', \\ Var(\varepsilon_t|\mathcal{I}_{t-1}) &= H_t = D_t\Gamma D_t,\ D_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 \\ 0 & \sqrt{h_{2t}} \end{bmatrix}, \Gamma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \\ \mathrm{E}\left[\varepsilon_{it}^2|\mathcal{I}_{t-1}\right] &= h_{it} = \eta_i's_{i,t-1} = \alpha_{i0} + \alpha_{i1}\varepsilon_{i,t-1}^2 + \beta_{i1}h_{i,t-1}, \\ \mathrm{E}\left[\varepsilon_{it}\varepsilon_{jt}|\mathcal{I}_{t-1}\right] &= h_{ij,t} = \rho\sqrt{h_{1t}h_{2t}}. \end{aligned} \tag{5.2}$$

Here $\varpi = (\varphi', \eta', \rho)' \in \Theta$. For notational convenience we drop the 0 subscript to denote the true parameters and write $H_t(\varpi) = H_t$, $m_{it} \equiv m_i(\mathcal{I}_{t-1,q}, \varphi_i)$ and $h_{it} \equiv h_i(\mathcal{I}_{t-1,q}, \eta_i)$ where $\eta_i' = (\alpha_{i0}, \alpha_{i1}, \beta_{i1})$ and $s_{i,t-1} = \left(1, \varepsilon_{i,t-1}^2, h_{i,t-1}\right)'$. In this bivariate case we have, therefore, a set of five moment conditions and the correct joint specification is tantamount to

$$H_0 : \begin{cases} \mathrm{E}\left[e_{0,i1t}|\mathcal{I}_{t-1}\right] = 0 \text{ a.s.}, i = 1, 2 \\ \mathrm{E}\left[e_{0,i2t}|\mathcal{I}_{t-1}\right] = 0 \text{ a.s.}, i = 1, 2 \\ \mathrm{E}\left[e_{0,3t}|\mathcal{I}_{t-1}\right] = 0 \text{ a.s.}, \end{cases} \text{, for some } \varpi_0 \in \Theta, \tag{5.3}$$

where $e_{0,i1t} \equiv \varepsilon_{0,it} = y_{it} - m_{0,it}$, $e_{0,i2t} = \varepsilon_{0,it}^2 - h_{0,it}$ for $i = 1, 2$ and $e_{0,3t} = \varepsilon_{0,1t}\varepsilon_{0,2t} - h_{0,12t} = \varepsilon_{0,1t}\varepsilon_{0,2t} - \rho_0\sqrt{h_{01t}h_{02t}}$ correspond to the conditional mean, conditional variance and conditional covariance specification, respectively. In a condensed form we can write:

$$H_0 : \mathrm{E}\left[e_{0t}|\mathcal{I}_{t-1}\right] = 0, \tag{5.4}$$

where $e_{0t} \equiv e_t(\varpi_0) = (e_{0,i1t}, e_{0,i2t}, e_{0,3t})'$, $i = 1,2$ is a (5×1) vector.

Remark 5.1 *Note that we are not testing the correlation structure, instead we are testing the covariance structure of the model. Testing the correlation structure is indeed possible within this framework; e.g., for a time varying correlation model* $e_{0,3t}$ *can be written as:*

$$e_{0,3t} = \frac{\varepsilon_{0,1t}\varepsilon_{0,2t}}{\sqrt{h_{01t}h_{02t}}} - \rho_{0,t}.$$

If one is interested in testing the adequacy of a proposed dynamic correlation structure, the PICM and IGS tests, extending the FDWB proposed in Chapter 4, can be implemented. However, we cannot implement the tests in the case of the constant conditional correlation model as there is no variation in ρ *over time. To be precise, in this case* $\frac{\partial e_{3t}(\varpi)}{\partial \rho} = -1$ *implying that the PICM test cannot be performed (see step 3 and 4; Section 5.3). Similarly, for the IGS tests we cannot obtain the correlation sample moment condition from the residuals of the OLS regression in see step 3 for the real world estimation (Section 5.4.1).*

We assume that our model (5.1) satisfies the regularity conditions for the consistency and asymptotic normality given in Chapter 3 (see Assumptions 3.2.1-3.2.6). In addition, we require the existence of the following moments, $\mathrm{E}\left[e_{0,i1t}^2\right] = \mathrm{E}\left[(y_{it} - m_{0,it})^2\right] < \infty$, $\mathrm{E}\left[e_{0,i2t}^2\right] = \mathrm{E}\left[e_{0,i1t}^2 - h_{0,it}\right]^2 < \infty$, $i = 1,2$, and $\mathrm{E}\left[e_{0,3t}^2\right] = \mathrm{E}\left[\varepsilon_{0,1t}\varepsilon_{0,2t} - h_{0,12t}\right]^2 < \infty$ and we assume that our model satisfies a start-up value condition which ensures that the impacts of initial values are asymptotically negligible.

5.3 The PICM Tests for MGARCH Models

In this section we extend the univariate projected ICM testing framework discussed in Chapter 4 to the MGARCH model with reference to (5.2). For the PICM test, in addition to Assumptions 3.2.1-3.2.6, we assume that our model satisfies Assumptions 4.3.1-4.3.3. Recall that the feasible projected marked empirical process and test statistic based on the CvM norm is given in (4.10) and (4.11), respectively. To implement this test we need to approximate the asymptotic null distribution of the CVM test statistic by some bootstrap procedure (e.g., wild bootstrap), because of (4.9) we do not require to estimate the model parameters in each bootstrap.

To be specific, for the bivariate AR-CCC GARCH regression model defined in (5.2), we require five PICM test statistics corresponding to five null moment conditions defined in (5.3). Unlike the univariate case, note that we are required to construct three weight matrices: one for the mean and variance moment conditions corresponding to the first variable (y_{1t}), one for the mean and variance moment conditions corresponding to second variable (y_{2t}), and the remaining one is required for testing the covariance moment condition. The step by step PICM testing procedure is given below:

1. Given the information set $\mathcal{I}_{t-1} = \left(Z'_{t-1}, Z'_{t-2}, \cdots\right)'$ at time $t-1$ where $Z_{t-1} = (Z_{1,t-1}, Z_{2,t-1})$ and $Z_{i,t-1} = (y_{i,t-1}, X'_{it})'$, $i = 1, 2$, construct three $(T \times T)$ indicator function weight matrices $\mathcal{W}_1$, $\mathcal{W}_2$ and $\mathcal{W}_3$ with the corresponding (r, s)-th element $1\left(Z_{1,r-1} \leq Z_{1,s-1}\right)$, $1\left(Z_{2,r-1} \leq Z_{2,s-1}\right)$, and $1\left(Z_{r-1} \leq Z_{s-1}\right)$, $r, s = 1, \cdots, T$, respectively.

2. Estimate $\hat{\varpi}$ and $\hat{e}_t \equiv e_t(\hat{\varpi}) = (\hat{e}_{11t}, \hat{e}_{21t}, \hat{e}_{12t}, \hat{e}_{22t}, \hat{e}_{3t})'$ by the QMLE. It is to noted here that we can employ both the FQMLE and PQMLE to estimate the model which will not have any effect on the testing procedure; we only require the $\sqrt{T}$-consistency of $\hat{\varpi}$.

3. For $i, k = 1, 2$, $k \neq i$ and $j = 1, 2$, construct the matrix of derivatives

$\widetilde{G}_{ij}$ with rows $g_{ij}\left(Z_{it},\hat{\varpi}\right)'$ where

$$\begin{aligned} g_{i1}\left(Z_{it},\varpi\right) &= \frac{\partial e_{i1t}\left(\varpi\right)}{\partial\varpi} \\ &= \left(\frac{\partial e_{i1t}\left(\varpi\right)}{\partial\varphi_i'},\frac{\partial e_{i1t}\left(\varpi\right)}{\partial\varphi_k'},\frac{\partial e_{i1t}\left(\varpi\right)}{\partial\eta_i'},\frac{\partial e_{i1t}\left(\varpi\right)}{\partial\eta_k'},\frac{\partial e_{i1t}\left(\varpi\right)}{\partial\rho}\right)' \\ &= -\left(W_{it},\ 0,\ 0,\ 0,\ 0\right)', \end{aligned}$$

and

$$\begin{aligned} g_{i2}\left(Z_{it},\varpi\right) &= \frac{\partial e_{i2t}\left(\varpi\right)}{\partial\varpi} \\ &= \left(\frac{\partial e_{i2t}\left(\varpi\right)}{\partial\varphi_i'},\frac{\partial e_{i2t}\left(\varpi\right)}{\partial\varphi_k'},\frac{\partial e_{i2t}\left(\varpi\right)}{\partial\eta_i'},\frac{\partial e_{i2t}\left(\varpi\right)}{\partial\eta_k'},\frac{\partial e_{i2t}\left(\varpi\right)}{\partial\rho}\right)' \\ &= -\left(\left(2W_{it}e_{i1t}+\frac{\partial h_{it}}{\partial\varphi_i'}\right),\ 0,\ \frac{\partial h_{it}}{\partial\eta_i'},\ 0,\ 0\right)', \end{aligned}$$

Remark 5.2 *For the GARCH(1,1) case,* $\frac{\partial h_{it}}{\partial\varphi_i}=-2\alpha_{i1}\varepsilon_{i,t-1}W_{i,t-1}+\beta_{i1}\frac{\partial h_{i,t-1}}{\partial\varphi_i}$ *and* $\frac{\partial h_{it}}{\partial\eta_i}=s_{i,t-1}+\beta_{i1}\frac{\partial h_{i,t-1}}{\partial\eta_i}$ *are obtained by recursions. It is easy to generalize for higher dimensional GARCH models..*

Define $c_{it}=\frac{1}{h_{it}}\frac{\partial h_t}{\partial\varphi_i}$ and $x_{it}=\frac{1}{h_{it}}\frac{\partial h_t}{\partial\eta_i}$. For $\hat{e}_{3t}$ construct the matrix of derivative $\widetilde{G}_3$ with rows $g_3\left(Z_t,\hat{\varpi}\right)'$ where

$$\begin{aligned} g_3\left(Z_t,\varpi\right) &= \frac{\partial e_{3t}\left(\varpi\right)}{\partial\varpi} \\ &= \left(\frac{\partial e_{3t}\left(\varpi\right)}{\partial\varphi_i'},\frac{\partial e_{3t}\left(\varpi\right)}{\partial\varphi_k'},\frac{\partial e_{3t}\left(\varpi\right)}{\partial\eta_i'},\frac{\partial e_{3t}\left(\varpi\right)}{\partial\eta_k'},\frac{\partial e_{3t}\left(\varpi\right)}{\partial\rho}\right)', \end{aligned}$$

with

$$\begin{aligned} \frac{\partial e_{3t}\left(\varpi\right)}{\partial\varphi_i'} &= -\left(\left(w_{it}\varepsilon_{kt}\right)+\frac{1}{2}\sqrt{h_{it}h_{kt}}c_{it}\right),\ i,k=1,2 \text{ and } k\neq i, \\ \frac{\partial e_{3t}\left(\varpi\right)}{\partial\eta_i'} &= -\left(\frac{1}{2}\sqrt{h_{it}h_{kt}}x_{it}\right),\ i,k=1,2 \text{ and } k\neq i, \\ \frac{\partial e_{3t}\left(\varpi\right)}{\partial\rho} &= -\sqrt{h_{it}h_{kt}},\ i,k=1,2 \text{ and } k\neq i. \end{aligned}$$

4. For $i, j = 1, 2$, regress $\hat{e}_{ij}$ on $\widetilde{G}_{ij}$ (after deleting the columns containing zeros only in $\widetilde{G}_{ij}$) and $\hat{e}_3$ on $\widetilde{G}_3$ to obtain five series of residuals, respectively, as $\widetilde{\widetilde{e}}_{ij} \equiv M_{\widetilde{G}_{ij}} \hat{e}_{ij}(\hat{\varpi})$ and $\widetilde{\widetilde{e}}_3 \equiv M_{\widetilde{G}_3} \hat{e}_3(\hat{\varpi})$ where $M_{\widetilde{G}} = I_T - \widetilde{G}\left(\widetilde{G}'\widetilde{G}\right)^{-1}\widetilde{G}'$ is the usual projection matrix.

5. Using the residuals from step 4 and corresponding weight matrix from step 1, calculate five CVM test statistics which have a general form $CvM_T = T^{-2}\widetilde{\widetilde{e}}'\mathcal{W}\mathcal{W}'\widetilde{\widetilde{e}} = T^{-2}\hat{e}'PP'\hat{e}$, where $P \equiv M_{\widetilde{G}}\mathcal{W}$.

6. In the bootstrap world, generate five series of bootstrapped residuals $\hat{e}^*_{ijt} = \{U_t \hat{e}_{ijt}\}_{t=1}^T$, $i, j = 1, 2$, and $\hat{e}^*_{3t} = \{U_t \hat{e}_{3t}\}_{t=1}^T$, where $\{U_t\}_{t=1}^T$ is a sequence of i.i.d. draws from a pick distribution.[1]

7. Calculate five bootstrapped CVM test statistics with the general form $CvM^*_T = T^{-2}\hat{e}'^*PP'\hat{e}^*$.

8. For joint test, obtain $CvM_{T,J} = \sum_{i=1}^5 CvM_{T,i}$ and $CvM^*_{T,J} = \sum_{i=1}^5 CvM^*_{T,i}$.

9. Reject H_0 at $100\alpha\%$ when bootstrap p-value $p^*_T < \alpha$ where p^*_T is calculated in a similar way as given in step 8 in Section 4.3.2 (p. 156-157).

5.4 The Integrated Generalized Spectral (IGS) Test

Now we investigate the Escanciano's (2008) IGS tests in the present multivariate context. With some appropriate parametric family of weight functions $\mathcal{W} = \{w(Z_{t-l}, x) : x \in \Pi \subset [-\infty, \infty]^s\}$ where Π is the nuisance parameter space with dimension s which depends on the particular family $\mathcal{W}$ used, the joint null (5.2) can be written as:

$$\gamma_{l,w}(x, \varpi_0) = \mathrm{E}\left[e_t(\varpi_0) w(Z_{t-l}, x)\right] = 0 \text{ a.e. in } \Pi \subset [-\infty, \infty]^s, \forall l \geq 1,$$

[1] To generate bootstrapped residuals, in Monte Carlo simulation we have also tried with draws from three separate pick distributions $\{U_{1t}\}_{t=1}^T$ (for the mean and variance moment conditions corresponding to y_{1t}), $\{U_{2t}\}_{t=1}^T$ (for the mean and variance moment conditions corresponding to y_{2t}) and $\{U_{3t}\}_{t=1}^T$ (for the covariance moment condition), and no significant difference in results was observed.

where $e_t(\varpi_0) = (e_{11t}(\varpi_0), e_{21t}(\varpi_0), e_{21t}(\varpi_0), e_{22t}(\varpi_0), e_{3t}(\varpi_0))'$. For a sample $\{y_t, \mathcal{I}_{t-1,q}\}_{t=1}^{T}$, the QMLE $\hat{\varpi}$ is $\sqrt{T}$-consistent estimator for ϖ_0 and $\hat{e}_{ijt} \equiv \hat{e}_{ijt}(\hat{\varpi})$ for $i, j = 1, 2$ and $\hat{e}_{3t} \equiv \hat{e}_{3t}(\hat{\varpi})$. We assume that $\hat{\varpi}$ satisfies the following asymptotic expansion:

Assumption 5.4.1 *Under H_0, $\hat{\varpi}$ satisfies the asymptotic Bahadur expansion*

$$\sqrt{T}(\hat{\varpi} - \varpi_0) = T^{-1/2} \sum_{t=1}^{T} \varrho(\mathcal{I}_{t-1,q}, \varpi_0) e_t(\varpi_0) + o_p(1), \tag{5.5}$$

where $\varrho(.)$ is such that $\mathrm{E}[\varrho(\mathcal{I}_{t-1,q}, \varpi_0) e_t(\varpi_0) e_t'(\theta_0) \varrho'(\mathcal{I}_{t-1,q}, \varpi_0)]$ exists and positive definite.

Considering the CvM norm, the Escanciano's (2008) joint IGS specification test statistics are given by (see Chapter 4 for details)

$$\begin{aligned} J_{T,w}^2 &\equiv J_{T,w}^2(\hat{\varpi}) = \int_{\Pi'} |R_{T,w}(\tau, x, \hat{\varpi})|_M^2 W(\mathrm{d}x)\,\mathrm{d}\tau \\ &= \sum_{l=1}^{n} \frac{T_l}{(l\pi)^2} \int_{\Pi} \left|\hat{\gamma}_{l,w}(x, \hat{\varpi})\right|_M^2 W(\mathrm{d}x), \end{aligned}$$

where $T_l = T - l + 1$, $R_{T,w}(\tau, x, \hat{\varpi})$ is defined in (4.22), $\Pi' = [0,1] \times \Pi$, $W(.)$ is an integrating function depending on the weight family $\mathcal{W}$ and here for our bivariate model, M is a 5×5 (in general a square matrix of order $2N + \frac{N(N-1)}{2}$) positive semidefinite matrix with

$$M = \begin{bmatrix} m_{11} & 0 & 0 & 0 & 0 \\ 0 & m_{12} & 0 & 0 & 0 \\ 0 & 0 & m_{21} & 0 & 0 \\ 0 & 0 & 0 & m_{22} & 0 \\ 0 & 0 & 0 & 0 & m_3 \end{bmatrix},$$

where m_{ij}, $i, j = 1, 2$ corresponds to j-th moment restriction for the i-th variable and m_3 corresponds to the covariance moment condition. Hence, we can obtain the marginal components by appropriately choosing the values of m's. For example, If M is an identity matrix of order 5, then the joint test is obtained; if $m_{11} = 1$, and setting other m's equal to 0 leads to the marginal test for mean specification for the first variable and so on.

Similar to the PICM tests, define three indicator weight functions corresponding to three conditioning sets namely, $w(Z_{i,t-l}, x) = 1(Z_{i,t-l} \le x)$ for $i = 1, 2$ and $w(Z_{t-l}, x) = 1(Z_{t-l} \le x)$. Now, for any $\sqrt{T}$-consistent estimators $\hat{\varpi}$, with these indicator weight functions and $W(.) = F_T(.)$ where $F_T(.)$ is the empirical distribution function, the test statistic has the following form:

$$J_{T,I}^2 = \sum_{l=1}^{T} \frac{T_l}{T(l\pi)^2} \sum_{t=1}^{T} \left[\begin{array}{c} \sum_{i=1}^2 \left\{ m_{i1}\hat{\sigma}_{i1}^{-2}\hat{\gamma}_{i,l,m}^2 (Z_{i,t-1}, \hat{\varpi}) + m_{i2}\hat{\sigma}_{i2}^{-2}\hat{\gamma}_{i,l,v}^2 (Z_{i,t-1}, \hat{\varpi}) \right\} \\ + m_3 \hat{\sigma}_3^{-2} \hat{\gamma}_{i,l,c}^2 (Z_{t-1}, \hat{\varpi}) \end{array} \right], \tag{5.6}$$

where

$$\begin{aligned} \hat{\gamma}_{i,l,m}(Z_{i,t-1}, \hat{\varpi}) &= \frac{1}{T_l}\sum_{t=l}^{T} \hat{e}_{i1t} w(Z_{i,t-l}, Z_{i,t-1}), \\ \hat{\gamma}_{i,l,v}(Z_{i,t-1}, \hat{\varpi}) &= \frac{1}{T_l}\sum_{t=l}^{T} \hat{e}_{i2t} w(Z_{i,t-l}, Z_{i,t-1}), \\ \hat{\gamma}_{i,l,c}(Z_{t-1}, \hat{\varpi}) &= \frac{1}{T_l}\sum_{t=l}^{T} \hat{e}_{3t} w(Z_{t-l}, Z_{t-1}), \\ \hat{\sigma}_{ij}^2 &= T^{-1}\sum_{t=1}^{T} \hat{e}_{ijt}^2, \text{ for } i,j = 1,2 \text{ and} \\ \hat{\sigma}_3^2 &= T^{-1}\sum_{t=1}^{T} \hat{e}_{3t}^2. \end{aligned}$$

The five marginal components are obtained as follows:

1. Setting $m_{i1} = 1$ and other m's equal to zero gives marginal mean component for the i-th variable, $D_{i,T,m}^2$, $i = 1, 2$.

2. Setting $m_{i2} = 1$ and other m's equal to zero gives marginal variance component for the i-th variable, $D_{i,T,v}^2$, $i = 1, 2$.

3. Setting $m_3 = 1$ and other m's equal to zero gives marginal covariance component for the first and second variable, $D_{3,T,c}^2$.

Following Escanciano (2008), we can also obtain the test statistics with complex exponential weight function $w(Z_{t-j}, x) = \exp(ix'Z_{t-j})$ and $W(dx) = \varphi(x)\, dx$ where $\varphi(x)$ is the standard normal density. However, in this Chapter we will not pursue this further. Recall that the simulation exercise in Chapter 4 (and also in other studies; e.g., Escanciano, 2008) shows no significant difference in results between test statistics based on indicator and exponential weight functions.

5.4.1 The FDWB Procedure

In the previous Chapter we have seen that the limiting null distribution needs to be approximated and bootstrap procedure, particularly the wild bootstrap procedure, is by far the most popular choice. To the best of our knowledge, the wild bootstrapping for multivariate GARCH models has not been considered in literature. To implement the IGS tests for the MGARCH regression model as defined in (5.1), we extend the proposed testing procedure suggested in Chapter 4. In particular, after the QMLE estimation of $\hat{\varpi}$ and $\hat{H}_t$, we consider $\hat{H}_t$ as observed. Then with the help of a set of auxiliary OLS regressions we obtain the estimates of moment conditions in the real world and given the real world information the same OLS regressions are run in the bootstrap world. In what follows "hats ($\hat{}$)" denotes QMLE while "tilda ($\tilde{}$)" denotes OLS estimator. The proposed bootstrap algorithm for AR(1)-CCC GARCH(1,1) model is given below:

Real World Estimation

1. Estimate the original model by the QMLE and obtain $\hat{\varpi} = (\hat{\varphi}', \hat{\eta}', \hat{\rho})'$, $\hat{m}_{it}$, $\hat{H}_t \Rightarrow \hat{h}_{it}$, $i = 1, 2$ and $\hat{h}_{12t} = \hat{\rho}\sqrt{h_{1t}h_{2t}}$.

2. For $i = 1, 2$ estimate the mean and variance sample moment condition as

$$\begin{aligned} \tilde{v}_{i1t} &= y_{it} - \tilde{\varphi}_{i0} - \tilde{\varphi}_{i1} y_{i,t-1} = y_{it} - W_{it}'\tilde{\varphi}_i \\ &\equiv y_{it} - \tilde{m}_{it}, \end{aligned}$$

and

$$\begin{aligned}\tilde{v}_{i2t} &= \tilde{v}_{i1t}^2 - \tilde{\beta}_{i0} - \tilde{\beta}_{i1}\hat{h}_{it} = \tilde{v}_{i1t}^2 - \hat{z}_{it}'\tilde{\beta}_i \\ &\equiv \tilde{v}_{i1t}^2 - \tilde{h}_{it},\end{aligned}$$

respectively, where $\tilde{\varphi}_i = (\tilde{\varphi}_{i0}, \tilde{\varphi}_{i1})'$ and $\tilde{\beta}_i = \left(\tilde{\beta}_{i0}, \tilde{\beta}_{i1}\right)'$ are OLS estimators from the regression y_{it} on $W_{it} = (1, y_{i,t-1})'$ and $\tilde{v}_{i1t}^2$ on $\hat{z}_{it} = \left(1, \hat{h}_{it}\right)'$, respectively, $\tilde{m}_{it} = W_{it}'\tilde{\varphi}_i$ and $\tilde{h}_{it} = \hat{z}_{it}'\tilde{\beta}_i$.

3. Obtain the covariance sample moment condition as the residual from the OLS regression of $(\tilde{v}_{i1t}\tilde{v}_{i2t})$ on $\hat{z}_{3t} = \left(1, \hat{h}_{12t}\right)'$; i.e.,

$$\begin{aligned}\tilde{v}_{3t} &= \tilde{v}_{i1t}\tilde{v}_{i2t} - \tilde{\gamma}_0 - \tilde{\gamma}_1\hat{h}_{12t} = \tilde{v}_{i1t}\tilde{v}_{i2t} - \hat{z}_{3t}'\tilde{\gamma} \\ &\equiv \tilde{v}_{i1t}\tilde{v}_{i2t} - \tilde{h}_{12t},\end{aligned}$$

where $\tilde{\gamma} = \left(\tilde{\gamma}_{0,}\tilde{\gamma}_1\right)'$ and $\tilde{h}_{12t} = \hat{z}_{3t}'\tilde{\gamma}$.

Bootstrap World

1. Generate WB residuals :

$$\begin{aligned}\hat{\varepsilon}_{i1t}^* &= \tilde{v}_{i1t}U_t,\ i = 1, 2, \\ \hat{\varepsilon}_{i2t}^* &= \tilde{v}_{i2t}U_t,\ i = 1, 2, \\ \hat{\varepsilon}_{3t}^* &= \tilde{v}_{3t}U_t,\end{aligned}$$

where $\{U_t\}$ a sequence of i.i.d. r.v.s with zero mean and unit variance, bounded support and independent of the sequence $\{y_t, \mathcal{I}_{t-1,q}\}_{t=1}^T$.

2. Generate bootstrap data:

$$\begin{aligned}Y_{i1t}^* &= \tilde{m}_{it} + \hat{\varepsilon}_{i1t}^*,\ i = 1, 2 \\ Y_{i2t}^* &= \tilde{h}_{it} + \hat{\varepsilon}_{i2t}^*,\ i = 1, 2 \\ Y_{3t}^* &= \tilde{h}_{12t} + \hat{\varepsilon}_{3t}^*.\end{aligned}$$

3. For $i = 1, 2$, Compute $\tilde{\varphi}_i^* = (\tilde{\varphi}_{i0}^*, \tilde{\varphi}_{i1}^*)'$ from the OLS regression of Y_{i1t}^* on W_{it} and subsequently first moment condition in bootstrap world $\tilde{v}_{i1t}^* = Y_{i1t}^* - \tilde{\varphi}_{i0}^* - \tilde{\varphi}_{i1}^* y_{it-1} = Y_{i1t}^* - W_{it}'\tilde{\varphi}_i^*$.

4. Similarly, for $i = 1, 2$, Compute $\tilde{\beta}_i^* = \left(\tilde{\beta}_{i0}^*, \tilde{\beta}_{i1}^*\right)'$ from the OLS regression of Y_{i2t}^* on $\hat{z}_{it}$ and obtain $\tilde{v}_{i2t}^* = Y_{i2t}^* - \tilde{\beta}_{i0}^* - \tilde{\beta}_{i1}^* \hat{h}_{it} = Y_{i2t}^* - \hat{z}_{it}\tilde{\beta}_i^*$.

5. Compute $\tilde{\gamma}^* = (\tilde{\gamma}_0^*, \tilde{\gamma}_1^*)'$ from the OLS regression of Y_{3t}^* on $\hat{z}_{3t}$ and obtain $\tilde{v}_{3t}^* = Y_{3t}^* - \tilde{\gamma}_0^* - \tilde{\gamma}_1^* \hat{h}_{12t} = Y_{3t}^* - \hat{z}_{3t}'\tilde{\gamma}^*$.

Similar to PICM tests, Monte Carlo experiments were conducted with three separate pick distributions and again no significant difference was found.

Suppose $\theta = \left(\varphi', \beta'_, \gamma'\right)'$ be the collection of all parameters from the above auxiliary regressions where $\varphi = \left(\varphi_{10}, \varphi_{11,}\varphi_{20}, \varphi_{21}\right)'$, $\beta = \left(\beta_{10}, \beta_{11,}\beta_{20}, \beta_{21}\right)'$ and $\gamma = (\gamma_0, \gamma_1)'$. Also denote the vector of true parameters as $\theta_0 = \left(\varphi_0', \beta_{0,}'\gamma_0'\right)'$. Similarly define the OLS estimators in the real world and bootstrap world as $\tilde{\theta} = \left(\tilde{\varphi}', \tilde{\beta}'_, \tilde{\gamma}'\right)'$ and $\tilde{\theta}^* = \left(\tilde{\varphi}^{*\prime}, \tilde{\beta}^{*\prime}_, \tilde{\gamma}^{*\prime}\right)'$, respectively. The following theorem shows that with the above testing procedure $\tilde{\theta}$ satisfies the asymptotic Bahadur expansion as stated in Assumption 5.5.

Theorem 5.1 *Under the Assumptions mentioned in 5.2 and 5.4.1, it can be shown that*

$$\sqrt{T}\left(\tilde{\theta} - \theta_0\right) = \sqrt{T}\begin{pmatrix} \tilde{\varphi} - \varphi_0 \\ \tilde{\beta} - \beta_0 \\ \tilde{\gamma} - \gamma_0 \end{pmatrix} = T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix} q_t(\varpi_0) \\ p_t(\varpi_0) \\ r_t(\varpi_0) \end{bmatrix} e_t(\varpi_0) + o_p(1),$$

where $\varpi_0' = (\varphi_0', \eta_0', \rho)$ *and the expressions for* $q_t(\varpi_0)$, $p_t(\varpi_0)$ *and* $r_t(\varpi_0)$ *are provided in the proof.*

However, since the true variance-covariance matrix $H_{0,t}$ is unobserved we cannot use $h_{0,it}$, $i = 1, 2$ and $h_{0,12t}$ to generate the bootstrap data, instead we have to use $\hat{h}_{it}$, $i = 1, 2$ and $\hat{h}_{12t}$. Therefore, strictly speaking, the asymptotic requirements are not fulfilled in the bootstrap world. But the simplicity of the procedure and the encouraging size performance of the similar procedure in the case of univariate GARCH model in previous Chapter, motivate us to examine the performance of the IGS tests using this bootstrap scheme to the multivariate context.

5.5 Monte Carlo Evidence

We investigate the finite sample performance of the above discussed two non-parametric tests namely the PICM and IGS tests in a bivariate context. We set $Z_{i,t-1} = y_{i,t-1}$; $i = 1, 2$ and $Z_{t-1} = y_{1,t-1}y_{2,t-1}$. Both tests are constructed based on the indicator family of weight function; i.e., the three weight functions constructed as $w_i(y_{i,t-1}, x) = 1(y_{i,t-1} \leq x)$ (for mean and variance of the i-th variable, $i = 1, 2$) whilst $w_3(y_{1,t-1}y_{2,t-1}, x) = 1(y_{1,t-1}y_{2,t-1} \leq x)$ (for covariance). A total of 12 test statistics considered in this experiment which are summarized in the Table 5.1. All experiments are done with 1000 Monte Carlo replications and programmed in GAUSS. For each replication 300 bootstrap samples are generated by using the Rademacher distribution given in (4.13). The sample size is considered $T = 100$, 200, 300 and 500, after discarding the first 200 observations from the sample to offset any initial value effect.

Table 5.1: The marginal and joint PICM and IGS tests under consideration

Tests	IGS tests	PICM tests
Marginal mean	$D^2_{i,m}$ $(i = 1, 2)$	$C^2_{i,m}$ $(i = 1, 2)$
Marginal variance	$D^2_{i,v}$ $(i = 1, 2)$	$C^2_{i,v}$ $(i = 1, 2)$
Marginal covariance	$D^2_{12,c}$	$C^2_{12,c}$
Joint	J^2	C^2

We consider the following bivariate AR(1)-CCC GARCH null model in our simulation exercise:

$$\begin{aligned}
y_{it} &= \varphi_{i0} + \varphi_{i1} y_{i,t-1} + \varepsilon_{it}, \\
\varepsilon_t &= H_t^{1/2} \xi_t, \ \varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})', \\
Var(\varepsilon_t | \mathcal{I}_{t-1}) &= H_t = D_t \Gamma D_t, \ D_t = \begin{bmatrix} \sqrt{h_{1t}} & 0 \\ 0 & \sqrt{h_{2t}} \end{bmatrix}, \Gamma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \\
\mathrm{E}\left[\varepsilon_{it}^2 | \mathcal{I}_{t-1}\right] &= h_{it} = \eta_i' s_{i,t-1} = \alpha_{i0} + \alpha_{i1} \varepsilon_{i,t-1}^2 + \beta_{i1} h_{i,t-1}, \\
\mathrm{E}\left[\varepsilon_{it} \varepsilon_{jt} | \mathcal{I}_{t-1}\right] &= h_{ij,t} = \rho \sqrt{h_{1t} h_{2t}}. \qquad (5.7)
\end{aligned}$$

The model can be estimated either using the PQMLE or FQMLE method. We have conducted the experiments with both methods and the conclusions are qualitatively similar. To save space here we only report the results for PQMLE.

5.5.1 Finite Sample Property: Size

In Chapter 3 we have seen that the univariate volatility does not have any significant impact on the finite sample performance whereas the size property is sensitive to the correlation parameter. Therefore here we contemplate only the high persistence GARCH(1,1) for univariate volatility whereas to assess the robustness of these tests to the correlation parameters both low and high correlation parameters are considered. In addition, to investigate the robustness to the departure from non-normality, data are generated both under the Normal distribution and standardized Student's t-distribution. The details of the DGPs are presented in Table 5.2.

Table 5.3 reports the actual rejection frequencies, when the null of AR-CCC is true, for a nominal size of 5%. It can be seen that both the PICM and IGS (marginal and joint) tests display satisfactory size property. Moreover, unlike the parametric CM tests of the CCC assumption (cf. Chapter 3), in this case the correlation parameter does not play any significant role. The IGS tests perform relatively better in terms of size. Similar to the GARCH model (cf. Chapter 4), these tests demonstrate impressive robustness under non-normality (see Table 5.4).

Table 5.2: True parameter values for size simulation

	E1	E2	E3	E4
$\varphi_1' = (\varphi_{10}, \varphi_{11})$		$(1.00, 0.10)$		
$\eta_1' = (\alpha_{10}, \alpha_{11}, \beta_{11})$		$(0.01, 0.15, 0.80)$		
$\varphi_2' = (\varphi_{20}, \varphi_{21})$		$(1.00, 0.50)$		
$\eta_2' = (\alpha_{20}, \alpha_{21}, \beta_{21})$		$(0.05, 0.20, 0.70)$		
ρ	0.20	0.80	0.20	0.80
Error distribution	Normal	Normal	$t(8)$	$t(8)$

Note: The parameter vectors $\varphi_1, \eta_1, \varphi_2, \eta_2$ and ρ refer to the AR(1)-CCC-GARCH(1,1) model in (5.7).

Table 5.3: Finite sample size at 5 per cent level: Normal errors

	E1				E2			
T	100	200	300	500	100	200	300	500
				PICM Tests				
$C^2_{1,m}$	5.20	5.50	6.90	5.90	6.00	4.70	5.00	5.90
$C^2_{1,v}$	7.50	7.30	5.10	5.40	7.60	5.90	6.10	5.40
$C^2_{2,m}$	5.60	5.30	4.40	5.60	5.70	6.00	3.00	5.10
$C^2_{2,v}$	7.20	6.60	6.80	5.90	6.80	8.00	6.60	6.10
$C^2_{12,c}$	7.10	7.10	4.60	5.40	6.20	5.50	5.60	4.90
C^2	7.40	7.50	5.60	5.90	7.80	6.10	5.80	5.70
				IGS Tests				
$D^2_{1,m}$	5.30	5.70	5.80	4.00	5.90	4.50	5.60	4.00
$D^2_{1,v}$	4.80	5.40	4.00	4.40	5.00	4.60	5.10	4.40
$D^2_{2,m}$	5.50	5.70	5.70	4.20	5.00	7.10	3.80	4.80
$D^2_{2,v}$	5.20	5.00	5.60	6.10	4.50	4.80	4.80	5.50
$D^2_{12,c}$	6.20	5.30	4.70	4.40	5.00	4.60	4.70	4.50
J^2	5.50	6.00	4.60	4.70	4.30	5.60	5.60	4.70

Notes: 1. Data generated according to the AR(1)-CCC-GARCH(1,1) DGP in (5.7) with parameter values as detailed in Table 5.2. Various tests are defined in Table 5.1.

2. Values in the Table are the empirical rejection frequencies of the null (5.3)

3. T is the sample size and results are based on 1000 Monte Carlo simulations and 300 bootstrap replications.

Table 5.4: Finite sample size at 5 per cent level: t(8) errors

T	E3 100	E3 200	E3 300	E3 500	E4 100	E4 200	E4 300	E4 500
PICM Tests								
$C^2_{1,m}$	4.80	4.70	6.20	5.70	4.60	5.30	5.10	4.40
$C^2_{1,v}$	7.50	7.50	6.00	5.60	6.90	6.70	6.00	5.80
$C^2_{2,m}$	4.70	6.40	4.60	4.90	3.80	4.10	4.50	4.50
$C^2_{2,v}$	7.80	7.60	6.10	5.90	8.60	7.40	6.90	6.50
$C^2_{12,c}$	7.70	6.20	6.40	5.90	6.90	5.80	4.70	5.60
C^2	7.50	7.60	6.10	5.30	6.80	6.60	6.30	5.80
IGS Tests								
$D^2_{1,m}$	5.80	5.80	5.70	5.50	4.70	6.80	5.60	5.10
$D^2_{1,v}$	5.00	5.50	5.20	5.10	5.80	5.80	4.70	6.20
$D^2_{2,m}$	5.40	6.30	5.60	5.70	5.00	4.90	5.70	5.70
$D^2_{2,v}$	5.00	5.90	5.10	5.20	5.90	5.00	5.00	5.80
$D^2_{12,c}$	6.30	5.50	4.70	5.10	5.50	5.70	4.50	5.20
J^2	5.00	6.10	5.30	5.20	6.40	6.70	5.70	6.30

Notes: see notes 1-3 of Table 5.3.

5.5.2 Finite Sample Property: Power

For the power experiments we consider 3 types of misspecified models, namely, only misspecified mean (DGP P1a, P1b, P1c and P1d), only misspecified variance (DGP P2a, P2b, P2c and P2d), and only misspecified covariance (correlation) models (DGP P3a, P3b, P4a and P4b). To investigate the effect of the correlation parameters on the tests, when the correlation structure is correctly specified (i.e., constant) but either the mean or variance is misspecified, both high correlation ($\rho = 0.80$) and low correlation ($\rho = 0.20$) are considered. The details of the DGPs and the parameter values are given in Table 5.5 and 5.6, respectively. The true parameter values are taken from the existing literature.

Table 5.7 displays the actual rejection frequencies against 5% nominal level, when the mean function is misspecified. The results are very encouraging: when only m_{2t} is misspecified (i.e., P1a and P1C) the corresponding marginal mean components of the PICM and IGS tests (i.e., $C^2_{2,m}$ and $D^2_{2,m}$ respectively) and the joint tests pick the misspecification. As we have seen in Chapter 4 even with the correct variance specification but with the misspecified mean function, the marginal variance tests over-rejects the null of correct specification of variance due to the misspecified mean function, in this case $C^2_{2,v}$ and $D^2_{2,v}$ also pick this misspecification. However, it is important to note that in this case the marginal components corresponding to the first variable (i.e. $C^2_{1,m}$, $C^2_{1,v}$, $D^2_{1,m}$,and $D^2_{1,v}$) and covariance structure ($C^2_{12,c}$ and $D^2_{12,c}$) are not affected. When both the mean functions m_{1t} and m_{2t} are wrongly specified, the marginal mean and joint tests demonstrate the power by picking up the misspecification. Finally, the power of these tests are not sensitive to the correlation parameters.

Similar conclusions are arrived at when the variance functions are misspecified (see Table 5.8). The marginal variance tests ($C^2_{2,v}$ and $D^2_{2,v}$ for P2a and P2c whilst $C^2_{1,v}$, $C^2_{2,v}$, $D^2_{1,v}$,and $D^2_{2,v}$ for P2b and P2d) pick the misspecified variance equation(s) leaving other marginal components unaffected. Again the tests are robust to the variation in correlation parameter.

Finally, Table 5.9 presents the actual rejection frequencies (against 5% nominal level) when the covariance structure is misspecified. The tests demonstrate a mixed performance here. First note that in this case, as expected, the mean and variance marginal tests are not affected by covariance misspecification. Secondly, for the DCC alternative models (i.e., P4a and P4b) both $C^2_{12,c}$ and $D^2_{12,c}$ show reasonable power. However, for the BEKK DGPs, particularly for P3a with lesser variability in the true correlation parameter, both the PICM and IGS tests lack power. The IGS tests demonstrate relatively better power in the case of DCC alternatives compared to the PICM test; on the other hand for P3b (BEKK DGP with relatively higher variation in correlation) the $C^2_{12,c}$ outperforms the $D^2_{12,c}$. Both tests display better power with the DGPs with higher variability in ρ.

Table 5.5: The DGPs for power experiments

	First variable (y_{1t})		Second variable (y_{2t})		Correlation
	m_{1t}	h_{1t}	m_{2t}	h_{2t}	ρ
P1a	AR	GARCH	GARCH in mean	GARCH	0.20
P1b	Bilinear	GARCH	GARCH in mean	GARCH	0.20
P1c	AR	GARCH	GARCH in mean	GARCH	0.80
P1d	Bilinear	GARCH	GARCH in mean	GARCH	0.80
P2a	AR	GARCH	AR	EGARCH	0.20
P2b	AR	EGARCH	AR	TGARCH	0.20
P2c	AR	GARCH	AR	EGARCH	0.80
P2d	AR	EGARCH	AR	TGARCH	0.80
P3a	AR	GARCH	AR	GARCH	BEKK
P3b	AR	GARCH	AR	GARCH	BEKK
P4a	AR	GARCH	AR	GARCH	DCC
P4b	AR	GARCH	AR	GARCH	DCC
			$\xi_t \sim N(0,1)$		

Table 5.6: True parameter values for power simulation

	True parameter values
Mean specification	
AR	$\varphi_1' = (1.00, 0.10)\,;\ \varphi_2' = (1.00, 0.50)\,;$
GARCH in mean	$y_{it} = 1 + 0.1 y_{i,t-1} + 1.5\sqrt{h_{it}} + \varepsilon_{it};$
Bilinear	$y_{it} = 0.4 y_{i,t-1} - 0.3 y_{i,t-2} + 0.5 y_{i,t-1}\varepsilon_{i,t-1} + \varepsilon_{it};$
Variance specification	
GARCH	$\eta_1' = (0.01, 0.15, 0.80)\,;\ \eta_2' = (0.05, 0.20, 0.70)\,;$
EGARCH	$h_{it} = -0.23 + 0.9\log(h_{i,t-1}) + 0.25\left[\lvert\zeta_{i,t-1}\rvert - 0.3\zeta_{i,t-1}\right];$
TGARCH	$\sqrt{h_{it}} = 0.07 + 0.0811 - I_{t-1})\lvert\varepsilon_{i,t-1}\rvert + 0.193 I_{t-1}\lvert\varepsilon_{i,t-1}\rvert + 0.831\sqrt{h_{i,t-1}};$
Covariance/Correlation specification	
BEKK	$H_t = C_B + A_B'\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right)A_B + B_B' H_{t-1} B_B;$
BEKK (P3a)	$A_B = \begin{bmatrix} 0.30 & 0.10 \\ 0.10 & 0.30 \end{bmatrix};\ B_B = \begin{bmatrix} 0.60 & 0.20 \\ 0.20 & 0.60 \end{bmatrix};\ C_B = \begin{bmatrix} 0.20 & 0.10 \\ 0.10 & 0.20 \end{bmatrix};$
BEKK (P3b)	$A_B = \begin{bmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{bmatrix};\ B_B = \begin{bmatrix} 0.40 & 0.20 \\ 0.20 & 0.40 \end{bmatrix};\ C_B = \begin{bmatrix} 0.20 & 0.04 \\ 0.04 & 0.20 \end{bmatrix};$
DCC	$\Gamma_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2};\ Q_t = (1 - \widetilde{\alpha} - \widetilde{\beta})\overline{Q} + \widetilde{\alpha}\zeta_{t-1}\zeta_{t-1}' + \widetilde{\beta} Q_{t-1};$
DCC (P4a)	$\widetilde{\alpha} = 0.05;\ \widetilde{\beta} = 0.90;\ \overline{Q} = \begin{bmatrix} 1.00 & 0.60 \\ 0.60 & 1.00 \end{bmatrix};$
DCC (P4b)	$\widetilde{\alpha} = 0.10;\ \widetilde{\beta} = 0.85;\ \overline{Q} = \begin{bmatrix} 1.00 & 0.60 \\ 0.60 & 1.00 \end{bmatrix}.$

Table 5.7: Empirical power against 5 per cent level: Misspecified mean equation(s)

T	p1a				p1b				p1c				p1d			
	100	200	300	500	100	200	300	500	100	200	300	500	100	200	300	500
PICM Tests																
$C^2_{1,m}$	5.20	5.40	5.80	5.90	43.40	69.50	82.60	88.40	6.00	6.50	6.70	5.30	44.70	70.10	82.30	87.10
$C^2_{1,v}$	8.80	5.90	5.30	5.40	31.50	48.50	64.50	79.50	7.30	6.10	6.10	7.00	29.10	52.80	66.10	79.10
$C^2_{2,m}$	19.10	36.30	54.20	78.20	18.30	39.60	54.20	76.50	19.20	34.20	52.40	79.00	16.10	35.60	51.90	77.40
$C^2_{2,v}$	9.10	10.40	12.80	13.60	9.30	10.60	13.30	14.30	8.10	9.00	10.50	15.60	10.30	10.70	12.50	16.30
$C^2_{12,c}$	6.90	7.20	5.20	5.50	7.40	6.90	6.30	5.70	6.50	5.00	4.50	3.40	5.70	4.10	5.20	4.30
C^2	10.20	7.70	5.90	7.50	36.60	58.50	73.00	85.20	8.90	6.80	7.10	8.50	34.50	61.80	73.70	84.40
IGS Tests																
$D^2_{1,m}$	5.50	5.20	4.80	4.00	80.10	91.50	95.30	95.90	5.30	6.60	5.40	5.60	82.30	91.20	95.10	95.70
$D^2_{1,v}$	5.90	4.60	4.50	4.40	16.30	23.90	31.30	42.70	5.20	5.00	4.90	6.40	16.90	23.20	31.60	42.70
$D^2_{2,m}$	18.30	38.60	61.20	87.00	18.30	40.60	59.70	85.00	17.20	36.10	57.10	86.00	17.30	39.10	56.10	86.00
$D^2_{2,v}$	4.80	4.90	6.20	8.00	4.00	3.40	4.70	8.70	3.30	4.60	4.10	9.50	4.40	3.70	4.80	9.60
$D^2_{12,c}$	4.70	6.80	5.60	3.50	6.60	5.40	4.70	5.50	5.30	4.50	5.80	6.70	5.30	5.80	7.00	7.70
J^2	7.60	10.20	13.60	27.60	38.50	80.10	94.70	97.80	6.60	9.30	13.70	31.40	39.20	79.30	96.20	97.80

Notes: For the definition of the DGPs and associated parameter values see Table 5.5 and 5.6. Various tests are defined in Table 5.1. Also see notes 2 and 3 of Table 5.3.

Table 5.8: Empirical power against 5 per cent level: Misspecified univariate volatility equation(s)

	P2a				P2b				P2c				P2d			
T	100	200	300	500	100	200	300	500	100	200	300	500	100	200	300	500
PICM Tests																
$C^2_{1,m}$	4.40	4.90	7.20	5.60	7.10	5.90	5.10	5.20	6.80	5.10	5.20	6.00	4.80	4.90	5.60	5.20
$C^2_{1,v}$	7.40	4.90	5.20	5.10	17.10	32.70	45.60	59.40	8.00	6.60	5.70	6.90	16.40	32.40	45.60	59.40
$C^2_{2,m}$	4.80	4.70	6.30	5.20	5.20	4.50	5.00	5.20	4.60	5.00	4.20	4.70	5.00	5.00	4.60	4.60
$C^2_{2,v}$	16.00	32.80	47.40	68.50	8.50	12.70	18.30	25.00	13.10	22.90	31.50	47.60	10.00	12.00	18.20	27.00
$C^2_{12,c}$	7.60	6.50	6.50	5.70	5.80	6.30	5.00	5.40	8.90	8.80	10.70	13.80	4.30	5.40	4.50	5.00
C^2	14.90	27.30	42.50	57.50	15.80	28.70	43.20	58.30	14.30	21.50	30.30	44.90	15.10	27.10	39.70	56.80
IGS Tests																
$D^2_{1,m}$	4.70	6.30	6.10	5.30	6.00	6.80	8.30	6.90	5.80	5.60	5.90	6.10	5.90	6.40	8.60	6.90
$D^2_{1,v}$	5.10	4.80	4.10	4.90	20.50	48.80	71.50	88.70	5.40	4.80	5.00	5.70	21.80	49.00	70.50	88.70
$D^2_{2,m}$	5.80	6.60	7.70	5.50	4.60	5.30	6.10	6.20	5.90	5.50	5.30	6.80	5.30	4.80	5.80	5.10
$D^2_{2,v}$	20.10	51.20	68.60	75.80	9.10	12.60	21.60	28.20	14.20	29.90	41.80	67.70	9.50	13.80	20.60	30.70
$D^2_{12,c}$	5.60	4.40	5.80	5.30	4.70	4.50	4.50	6.70	4.40	5.70	4.00	4.10	4.20	4.60	4.90	4.20
J^2	14.00	31.70	48.30	58.50	17.40	37.20	60.10	80.80	9.60	15.60	21.90	39.10	17.10	38.20	56.60	78.10

Notes: For the definition of the DGPs and associated parameter values see Table 5.5 and 5.6. Various tests are defined in Table 5.1. Also see notes 2 and 3 of Table 5.3.

Table 5.9: Empirical power against 5 per cent level: Misspecified covariance structure

	P3a				P3b				P4a				P4b			
T	100	200	300	500	100	200	300	500	100	200	300	500	100	200	300	500
PICM Tests																
$C^2_{1,m}$	6.00	5.20	6.50	5.60	5.50	4.90	4.70	5.50	6.70	5.50	5.80	5.80	4.20	5.80	5.70	5.80
$C^2_{1,v}$	7.10	6.30	5.50	5.60	6.50	6.30	8.30	6.10	6.60	8.10	4.90	5.40	7.50	6.00	6.50	5.40
$C^2_{2,m}$	5.00	5.90	5.70	4.80	6.00	5.70	4.40	5.70	4.60	5.90	4.70	6.00	4.90	5.10	3.90	5.60
$C^2_{2,v}$	7.90	7.00	7.60	7.70	6.80	7.60	7.10	7.90	9.60	7.60	7.20	6.50	9.20	7.60	7.00	8.00
$C^2_{12,c}$	5.70	4.70	4.90	6.30	8.00	18.30	19.50	29.10	5.50	7.10	12.00	15.90	8.00	13.00	16.50	23.00
C^2	10.60	9.50	7.90	7.90	10.60	8.90	11.30	11.20	7.00	8.50	5.00	6.10	7.80	6.40	6.60	6.10
IGS Tests																
$D^2_{1,m}$	6.20	6.00	5.30	6.90	5.70	6.40	6.10	6.00	6.30	6.00	5.70	4.10	4.30	5.10	5.70	4.10
$D^2_{1,v}$	5.90	4.80	4.80	4.00	4.70	4.60	5.10	4.00	5.20	5.90	4.20	4.40	5.00	5.40	5.40	4.40
$D^2_{2,m}$	5.30	5.80	5.90	4.80	6.30	5.50	4.10	5.80	4.80	5.50	4.40	4.90	4.40	5.70	4.10	4.50
$D^2_{2,v}$	5.90	5.40	4.10	5.10	5.30	5.00	4.40	5.20	6.20	5.40	5.60	4.10	5.90	5.30	6.10	5.00
$D^2_{12,c}$	4.50	4.90	6.10	4.70	6.10	6.60	7.90	9.30	8.10	11.70	14.20	24.00	11.30	17.00	23.50	32.70
J^2	7.60	6.80	7.00	6.10	6.10	5.70	6.70	7.30	6.40	9.30	11.60	14.50	9.80	12.10	14.00	18.30
Range of true correlation (ρ) in simulated sample*																
max	0.325	0.344	0.332	0.323	0.759	0.795	0.788	0.816	1.323	1.424	1.525	1.473	1.380	1.472	1.559	1.563
min	0.119	0.140	0.160	0.167	0.391	0.437	0.478	0.535	0.190	0.297	0.414	0.530	0.238	0.408	0.578	0.707
average	0.184	0.206	0.217	0.231	0.548	0.591	0.613	0.638	0.698	0.776	0.853	0.951	0.858	0.928	0.998	1.082

Notes: For the definition of the DGPs and associated parameter values see Table 5.5 and 5.6. Various tests are defined in Table 5.1. Also see notes 2 and 3 of Table 5.3.

*As a measure of the variability of the true ρ, we calculate the range (i.e., maximum-minimum) of ρ in each simulated sample of T observations. These figures summarize the maximum, minimum and average ranges of ρ in the Monte Carlo samples of 1000.

5.6 Concluding Remarks

In this Chapter we extend the univariate PICM and IGS testing procedures in the multivariate GARCH context with the aim to simultaneously test the first and second moments of the model; i.e., the individual mean, the individual variance and the covariance structure among the variables. The FDWB procedure for the IGS tests, proposed in Chapter 4, is extended to accommodate the MGARCH model. We illustrate the testing methodology with the CCC-GARCH model for simplicity, but can be applied to other MGARCH models.

The Monte Carlo simulations show that the both the IGS and PICM marginal and joint tests have excellent size property irrespective of the correlation parameter. They also show impressive robustness to non-normality. Hence, in spite of the failure of the bootstrap procedure to satisfy the sufficient Bahadur asymptotic expansion, its ability to control the size well leads to a meaningful power comparison. Again, finding the weaker set of conditions under which this test work is a future research agenda.

In the case of individual conditional mean or variance equation(s) misspecification, both nonparametric tests can successfully identify the source of misspecification with high power. However, the power of the covariance marginal tests seems to depend on the true alternative models and the variability of the true correlation parameter. For the DCC alternatives both tests (particularly the IGS tests) show reasonable power even with our small to moderate sample size. However, for BEKK alternative the tests lack in power. The nonparametric testing of the MGARCH models is still in its early days and this study clearly shows the merit of the testing procedures in separating the sources of misspecification and the necessity of further research on devising more powerful nonparametric tests, particularly for covariance structure.

Appendix

5.A Appendix A: Proofs

5.A.1 Proof of Theorem 5.1

Proof. First note that the proof for individual conditional mean and conditional variance is provided in Chapter 4 (see Theorem 4.2). Using those results for the i-th variable, we have

$$\begin{aligned}\sqrt{T}\left(\tilde{\varphi}_i - \varphi_{i0}\right) &= T^{-1/2}\sum_{t=1}^{T} q_{it}(\varpi_0)e_{it}(\varpi_0) + o_p(1), \\ \sqrt{T}\left(\tilde{\beta}_i - \beta_{i0}\right) &= T^{-1/2}\sum_{t=1}^{T} p_{it}\left(\varpi_0\right)e_{it}(\varpi_0) + o_p(1),\end{aligned}$$

where $e_{it}(\varpi_0) = (e_{i1t}(\varpi_0), e_{i2t}(\varpi_0))\,,\ i = 1,2.$

Therefore we only require to show that $\tilde{\gamma}$ satisfies the asymptotic Bahadur expansion. Let $\hat{h}_{12t} = \hat{\rho}\sqrt{\hat{h}_{1t}\hat{h}_{2t}}$ be constructed using the QMLE, $\hat{\varpi}$, and define $\hat{z}'_{3t} = \left(1, \hat{h}_{12t}\right)$ with $z'_{3t} = (1, h_{12t})\,.$ Now regress $\tilde{v}_{11t}\tilde{v}_{21t} \equiv \tilde{\varepsilon}_{1t}\tilde{\varepsilon}_{2t} = (y_{1t} - \tilde{m}_{1t})(y_{2t} - \tilde{m}_{2t})$ on $\hat{z}'_{3t}$ to give

$$\tilde{\gamma} = \left(T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\hat{z}'_{3t}\right)^{-1} T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\tilde{v}_{11t}\tilde{v}_{21t}. \tag{5.8}$$

Note that: $\tilde{v}_{11t}\tilde{v}_{21t} = e_{11t}e_{21t} - \left(f'_{1t}\bar{e}_{21t},\ f'_{2t}\bar{e}_{11t}\right)(\tilde{\varphi}-\varphi_0) = e_{11t}e_{21t} + \bar{w}_t$ (say), where $f_{it} = \frac{\partial m_{it}}{\partial \varphi_i}$ and $\bar{e}_{i1t} \equiv \bar{\varepsilon}_{it} = \left(y_{it} - m\left(\mathcal{I}_{t-1,q}, \bar{\varphi}_i\right)\right)$ and $\bar{\varphi}_i$ is a mean value. Secondly, $e_{11t}e_{21t} = (e_{11t}e_{21t} - h_{12t}) + h_{12t} = e_{3t} + h_{12t} = e_{3t} + z'_{3t}\gamma_0 = e_{3t} + \hat{z}'_{3t}\gamma_0 - (\hat{z}_{3t} - z_{3t})'\,\gamma_0$, where $\gamma_0 = (0,1)'\,.$ Therefore, (5.8) can be written as:

$$\begin{aligned}
\tilde{\gamma} &= \left(T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\hat{z}_{3t}'\right)^{-1} T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\left(e_{11t}e_{21t}+\bar{w}_t\right) \\
&= \gamma_0 + \left(T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\hat{z}_{3t}'\right)^{-1} T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\left\{e_{3t}-\left(\hat{z}_{3t}-z_{3t}\right)'\gamma_0+\bar{w}_t\right\} \\
&= \gamma_0 + \left(T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\hat{z}_{3t}'\right)^{-1} T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\left\{e_{3t}-\left(\hat{z}_{3t}-z_{3t}\right)'\gamma_0\right\} \\
&\quad + o_p(T^{-1/2}), \qquad (5.9)
\end{aligned}$$

where the last line exploits $\sqrt{T}$ consistency of $\hat{\varpi}$ and $\tilde{\varphi}$ and a ULLN which ensures that $T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\hat{z}_{3t}' = O_p(1)$ and $T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\overline{f}_{it}'\bar{e}_{i1t} = o_p(1)$.

In what follows, write $\hat{V}_T = T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\hat{z}_{3t}'$. Given sufficient regularity $\hat{V}_T = V + o_p(1)$, where $V = \mathrm{E}\left[z_{3t}z_{3t}'\right]$. Since $\hat{h}_{12t} - h_{12t} = \bar{s}_t\left(\hat{\varpi}-\varpi_0\right)$, where $\bar{s}_t = \left(\frac{\partial h_{12t}}{\partial\varphi'}, \frac{\partial h_{12t}}{\partial\eta'}, \frac{\partial h_{12t}}{\partial\rho}\right)'\Big|_{\varpi=\bar{\varpi}}$, we have

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\hat{h}_{12t}e_{3t} &= T^{-1/2}\sum_{t=1}^{T}h_{12t}e_{3t} + T^{-1}\sum_{t=1}^{T}\bar{e}_{3t}\bar{s}_t\sqrt{T}\left(\hat{\varpi}-\varpi_0\right) \\
&= T^{-1/2}\sum_{t=1}^{T}h_{12t}e_{3t} + o_p(1),
\end{aligned}$$

since by consistency $\hat{\varpi}$ and a ULLN, $T^{-1}\sum_{t=1}^{T}\bar{e}_{3t}\bar{s}_t = o_p(1)$. Thus

$$T^{-1/2}\sum_{t=1}^{T}\hat{z}_{3t}e_{3t} = T^{-1/2}\sum_{t=1}^{T}z_{3t}e_{3t} + o_p(1),$$

substituting this into (5.9) yields

$$\sqrt{T}(\tilde{\gamma}-\gamma_0) = \hat{V}_T^{-1}T^{-1/2}\sum_{t=1}^{T}z_{3t}e_{3t} - \hat{V}_T^{-1}\sqrt{T}\left(\hat{V}_T - T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}z_{3t}'\right)\gamma_0 + o_p(1). \qquad (5.10)$$

Consider

$$
\begin{aligned}
\sqrt{T}\left(\hat{V}_T - T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}z_{3t}'\right)\gamma_0 &= \left(T^{-1/2}\sum_{t=1}^{T}\hat{z}_{3t}\left(\hat{z}_{3t}-z_{3t}\right)'\right)\begin{pmatrix}0\\1\end{pmatrix} \\
&= T^{-1/2}\sum_{t=1}^{T}\hat{z}_{3t}\left(\hat{h}_{12tt}-h_{12tt}\right) \\
&= \left(T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\bar{s}_t\right)\sqrt{T}\left(\hat{\varpi}-\varpi_0\right) \\
&= A\sqrt{T}\left(\hat{\varpi}-\varpi_0\right)+o_p(1),
\end{aligned}
$$

where consistency of $\hat{\varpi}$ and a ULLN will ensure that $T^{-1}\sum_{t=1}^{T}\hat{z}_{3t}\bar{s}_t = A + o_p(1)$.

Finally, substituting this result and (5.5) into (5.10) we obtain

$$
\begin{aligned}
\sqrt{T}(\tilde{\gamma}-\gamma_0) &= V^{-1}T^{-1/2}\sum_{t=1}^{T}z_{3t}e_{3t} - V^{-1}A\sqrt{T}\left(\hat{\varpi}-\varpi_0\right)+o_p(1) \\
&= V^{-1}T^{-1/2}\sum_{t=1}^{T}z_{3t}e_{3t} - V^{-1}AT^{-1/2}\sum_{t=1}^{T}\varrho_t(\varpi_0)e_t(\varpi_0)+o_p(1) \\
&= T^{-1/2}\sum_{t=1}^{T}r_t(\varpi_0)e_t(\varpi_0)+o_p(1),
\end{aligned}
$$

where $r_t\left(\varpi_0\right) = V^{-1}\left(\left[0,0,0,0,\ z_{3t}\right] - A\varrho_t(\varpi_0)\right)$ and $\mathrm{E}\left[r_t(\varpi_0)e_t(\varpi_0)e_t'(\varpi_0)r_t'(\varpi_0)\right]$ is finite and positive definite.

Putting all these together we have the following asymptotic expansion

$$
\sqrt{T}\begin{pmatrix}\tilde{\varphi}-\varphi_0\\ \tilde{\beta}-\beta_0\\ \tilde{\gamma}-\gamma_0\end{pmatrix} = T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix}q_t(\varpi_0)\\ p_t\left(\varpi_0\right)\\ r_t(\varpi_0)\end{bmatrix}e_t(\varpi_0)+o_p(1),
$$

which is of the required form. ■

Chapter 6

Conclusions

In this thesis we focus on the specification testing problems of GARCH regression models, which is a relatively under-researched topic compared to the voluminous literature in this field. This research, apart from a literature review of multivariate GARCH (MGARCH) models in Chapter 2, consists of three substantive contributions presented in Chapters 3, 4 and 5.

In Chapter 3 we have proposed a set of asymptotically valid parametric CM tests to test the *Constant Conditional Correlation (CCC)* assumption. These tests are easy to construct and can be implemented under both the Full QMLE (FQMLE) and Partial QMLE (PQMLE) frameworks. We want to emphasize that, to the best of our knowledge, tests based on the PQMLE are not available in the literature. Moreover, the robust (to non-normality) and OPG versions of these tests have been developed. The Tse's (2000) LM test, which is a FQMLE-OPG type test, has been analyzed and modified so that it can be accommodated within our CM testing framework. In addition a robust (to non-normality) version of the Tse's test has been proposed. We also provide a new and easily programmable expressions for the expected Hessian for the FQMLE. The major findings from the Monte Carlo experiments are:

- In general all tests have desirable size property and the robust versions manifest better size under non-normality.
- The true correlation parameter has a significant impact on empirical size of these tests (low correlation is associated with better size prop-

erty). The size is, however, not affected by the degree of univariate volatility persistence.

- The PQMLE based tests exhibit more robustness to non-normality compared to the FQMLE tests.

- When the assumption of the null model is violated by assuming misspecified univariate volatility structure (but maintaining the CCC assumption):

  - the size of these tests is not affected by volatility spillover effect,
  - however when one equation is misspecified and true correlation is high, all tests over-reject the null of the CCC assumption. The rejection rate is higher in the case of FCM tests as expected, as by construction these tests consider the individual volatility moment conditions as well.

- The power of these tests depends on the variability of the true correlation parameter. Tests based on the Tse's modified indicator and the FCM tests show excellent power, even in models with less dispersed correlations. The CCM tests, in general, show lower relatively power and particularly in models with less dispersed correlations they have limited power. In terms of power there is very little to choose between the OPG and robust; and between the FQMLE and PQMLE.

Earlier researches (see, e.g., Kyrtsou ,2008; Blake and Kapetanios, 2007; Lumsdaine and Ng, 1999) have provided ample evidence that the neglected misspecification of the mean may result in incorrect inferences from variance specification tests, which implicitly assume a correct conditional mean specification. In Chapter 4 we, therefore, turned our focus back to the univariate GARCH regression model and investigate the simultaneous (nonparametric) testing problem of conditional mean and variance. Two nonparametric testing procedures, namely the Projected Integrated Conditional Moment, or PICM, (Escanciano, 2009) and Integrated Generalized Spectral, or IGS, (Escanciano, 2008), are investigated in this study. Both of these are based on an integrated approach to consistent testing; i.e., they avoid the smoothing of the data in the one hand but require bootstrap approximation to obtain the critical values on the other. Bootstrapping of the GARCH model is indeed a computationally demanding and operationally problematic issue which may restrict the application of these tests in practice. The PICM tests avoids the estimation of the parameters in each bootstrap replication, which makes the implementation of the test much simpler. To implement the IGS tests, Escanciano (2008) suggested a FDWB procedure for conditional heteroskedastic models. We, however, showed that this procedure is not directly applicable in the case of null AR-GARCH model. We have illustrated in detail how to perform the IGS and PICM tests for our model. In particular, we have proposed a simple and easily implementable wild bootstrap scheme for the IGS tests of GARCH regression models. Unfortunately the bootstrap procedure does not strictly satisfy the sufficient conditions, but the IGS tests based on our scheme control the size very well which justify the power comparison. This also suggests that the results may be obtained under a weaker set of conditions. Our Monte Carlo simulation confirms that a number of parametric CM tests for the conditional variance, which inherently assume a correct conditional mean specification, are indeed sensitive to misspecified mean function. The empirical application with the S&P 500 data also highlights the usefulness of the marginal and joint testing within the nonparametric framework. Some major findings of the simulation exercise are:

- Both the IGS and PICM tests show excellent size and impressive robust performance under non-normality. The parametric CM tests suffer from size distortion for small T and the distortion is greater under non-normality.

- The power analysis indicates that for the correct mean but misspecified variance alternatives, the marginal nonparametric tests demonstrate the ability to identify the source of misspecification. The parametric CM tests (barring T_{LT}) also show excellent power in this case.

- For misspecified mean and correct variance alternatives, the parametric CM tests wrongly and unsurprisingly over-reject the null of correct variance specification. Both nonparametric marginal mean components show excellent empirical power. Although the marginal variance tests pick some of the misspecification through the channel of conditional mean and the rejection frequencies increase as T increase, this is much lower compared to mean component.

Finally, in Chapter 5 we have extended the nonparametric tests of Chapter 4 in a MGARCH context with the aim to test the mean and variance-covariance structure simultaneously. The FDWB procedure proposed for the GARCH models in Chapter 4 is also extended to the MGARCH case. The procedure is illustrated with the help of null AR - CCC GARCH model for simplicity. The Monte Carlo simulations show that:

- Both the IGS and PICM (marginal and joint) tests have excellent size property irrespective of the correlation parameter and they demonstrate impressive robustness under non-normality.
- In the case of individual conditional mean or variance equation(s) misspecification, both nonparametric tests can successfully identify the source of misspecification with high power.
- The power of the covariance marginal tests seems to depend on the true alternative models and the variability of the true correlation parameter.

A future research agenda includes 1) devising parametric CM tests for dynamic conditional correlation models which allow the PQMLE, 2) applying the proposed wild bootstrap procedure developed in Chapter 3 to the parametric CM tests (Halunga and Orme, 2009) of univariate GARCH models to improve their size property and compare the performance with a projected version of the tests, 3) finding the set of less restrictive necessary and sufficient conditions for the IGS tests of univariate GARCH models, and 4) developing powerful nonparametric tests for multivariate GARCH models.

Bibliography

[1] Alexander, C.O. and Chibumba, A.M., 1997. Multivariate orthogonal factor GARCH. Mimeo, University of Sussex.

[2] Andreou, E. and Ghysels, E., 2003. Tests for breaks in the conditional co-movements of asset returns. Statistica Sinica, 13(4), 1045–1074.

[3] Ang, A. and Chen, J., 2002. Asymmetric correlations of equity portfolios. Journal of Financial Economics, 63(3), 443-494.

[4] Aslanidis, N., Osborn, D. and Sensier, M., 2008. Co-movements between US and UK stock prices: the roles of macroeconomic information and time-varying conditional correlations. Manchester Economics Discussion Paper Series 0805.

[5] Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74(1), 3-30.

[6] Becker, R. and Hurn, S., 2009. Testing for nonlinearity in mean in the Presence of heteroskedasticity. Economic Analysis and Policy, 39(2), 311-326.

[7] Bera, A.K. and Higgins, M.L., 1993. ARCH models: properties, estimation and testing. Journal of Economic Surveys, 7(4), 305-366.

[8] Bera, A.K. and Higgins, M.L., 1997. ARCH and bilinearity as competing models for nonlinear dependence. Journal of Business & Economic Statistics, 15(1), 43–50.

[9] Bera, A.K. and Kim, S., 2002. Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns. Journal of Empirical Finance, 9(2), 171–195.

[10] Bera, A.K. and McKenzie, C.R., 1986. Alternative forms and properties of the score test. Journal of Applied Statististics, 13(1), 13–25.

[11] Berben, R.-P., and Jansen, W. J., 2005. Comovement in international equity markets: A sectoral view Journal of International Money and Finance, 24(5), 832-857.

[12] Bierens, H. J., 1982. Consistent model specification tests. Journal of Econometrics, 20(1), 105-134.

[13] Bierens, H.J., 1984. Model specification testing of time series regressions. Journal of Econometrics, 26(3), 323-353.

[14] Bierens, H.J. and Ploberger, W., 1997. Asymptotic theory of integrated conditional moment test. Econometrica, 65(5), 1129–1151.

[15] Billingsley, P., 1995. Probability and Measure (3rd ed.). Wiley New York.

[16] Billio, M. and Caporin, M., 2009. A generalized dynamic conditional correlation model for portfolio risk evaluation. Mathematics and Computers in Simulation, 79(8), 2566-2578

[17] Billo, M., Caporin, M. and Gobbo, M., 2006. Flexible dynamic conditional correlation multivariate GARCH models for asset allocation. Applied Financial Economics Letters, 2(2), 123–130.

[18] Blake, A.P. and Kapetanios, G. , 2007. Testing for ARCH in the presence of nonlinearity of unknown form in the conditional mean. Journal of Econometrics 137(2), 472–488.

[19] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

[20] Bollerslev, T., 1990. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Review of Economics and Statistics, 72(3), 498–505.

[21] Bollerslev, T., Chou, R.Y. and Kroner K.F., 1992. ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics, 52(1-2), 5-59.

[22] Bollerslev, T., Engle, R.F. and Nelson, D.B., 1994. ARCH models. In: Engle, R. and McFadden, D., eds. Handbook of Econometrics, North Holland Press: Amsterdam.

[23] Bollerslev, T, Engle, R.F. and Wooldridge, J. M., 1988. A capital asset pricing model with time varying covariances. Journal of Political Economy, 96(1), 116-131.

[24] Bollerslev, T. and Wooldridge, J.M., 1992. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances. Econometric Reviews, 11(2), 143–172.

[25] Butler, K.C. and Joaquin, D.C., 2002. Are the gains from international portfolio diversification exaggerated? The influence of downside risk in bear markets. Journal of International Money and Finance 21(7), 981-1011.

[26] Cappiello, L., Engle, R. F. and Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics, 4(4), 537-572.

[27] Chan, F. and McAleer, M., 2002. Maximum likelihood estimation of STAR and STAR-GARCH models: Theory and Monte Carlo evidence. Journal of Applied Econometrics, 17(5), 509–534.

[28] Chen, X. and Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. Journal of Econometrics, 91(2), 373–401.

[29] Chesher, A., 1984. Testing for neglected heterogeneity. Econometrica, 52(4), 865–872.

[30] Chesher, A. and Spady, R., 1991. Asymptotic expansions of the information matrix test statistics. Econometrica, 59(3), 787–816.

[31] Christodoulakis, G.A. and Satchell, S.E., 2002. Correlated ARCH: modelling the time-varying correlation between financial asset returns. European Journal of Operations Research, 139(2), 351-370.

[32] Christoffersen, P. and Goncalves, S., 2005. Estimation Risk in Financial Risk Management. Journal of Risk, 7(3), 1-28.

[33] Comte, F. and Lieberman, O., 2003. Asymptotic Theory for Multivariate GARCH Processes. Journal of Multivariate Analysis, 84(1), 61–84.

[34] Cox, D.R., 1983. Some remarks on overdispersion. Biometrika, 70(1), 269–274.

[35] Dagenais, M.G. and Dufour, J.M., 1991. Invariance, nonlinear models, and asymptotic tests. Econometrica, 59(6), 1601-1615.

[36] Davidson, R. and Flachaire, E., 2001. The wild bootstrap, tamed at last. Working Papers 1000, Queen's University, Department of Economics.

[37] Davidson, R. and Flachaire, E., 2008. The wild bootstrap, tamed at last. Journal of Econometrics, 146(1), 162-169.

[38] Davidson, R. and MacKinnon, J.G., 1983. Small sample properties of alternative forms of the Lagrange multiplier test. Economics Letters, 12(3-4), 269–275.

[39] de Jong, R., 1996. The Bierens test under data dependence. Journal of Econometrics, 72(1), 1-32.

[40] Delgado, M., Dominguez, M. and Lavergne, P., 2006. Consistent tests of conditional moment restrictions. Annales d'Economie et de Statistique, 81(1), 33-67.

[41] Dempster, A. P., Laird, N. M. and Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39(1), 1-38.

[42] Ding, Z. and Engle, R.F., 2001. Large scale conditional covariance matrix modeling estimation and testing. Academia Economic Papers 29(2), 157-184.

[43] Ding, Z., Granger, C.W.J. and Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance, 1(1), 83-106.

[44] Dufour J.M., Khalaf, L. and Beaulieu, M.C., 2003. Finite-sample diagnostics for multivariate regression with applications to linear asset pricing models. Working Paper No. 06-2003, CIREQ.

[45] Duchesne, P. and Lalancette, S., 2003. On testing for multivariate ARCH effects in vector time series models. Canadian Journal of Statistics, 31(3), 275-292.

[46] El Babsiri, M. and Zakoian, J-M., 2001. Contemporaneous asymmetry in GARCH processes. Journal of Econometrics, 101(2), 257-294.

[47] Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4), 987-1007.

[48] Engle, R.F., 2002. Dynamic conditional correlation: a simple class of multivariate GARCH models. Journal of Business and Economic Statistics, 20(3), 339–350.

[49] Engle, R. F., 2009a. High dimensional dynamic correlations. In: Castle, J. L. and Shephard, N., Eds. The methodology and practice of econometrics: papers in honour of David F Hendry. Oxford University Press. Chapter 5.

[50] Engle, R. F., 2009b. Anticipating Correlations: A New Paradigm for Risk Management. Princeton and Oxford: Princeton University Press.

[51] Engle, R. F. and Kelly, B., 2008. Dynamic equicorrelation. NYU Working Paper No. FIN-08-038.

[52] Engle, R.F. and Kroner, F.K., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory, 11(1), pp.122–150.

[53] Engle, R.F. and Mezrich, J., 1996. GARCH for groups. Risk 9(8), 36-40.

[54] Engle, R.F and Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance, 48(5), 1749–1778.

[55] Engle, R.F., Ng, V.K. and Rothschild, M., 1990. Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. Journal of Econometrics 45(1-2), 213-237.

[56] Engle, R.F. and Sheppard, K. 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. NBER Working Paper No. 8554.

[57] Engle, R.F and Sheppard, K., 2008. Evaluating the specification of covariance models for large portfolios. Working Paper, Department of Economics, University of Oxford.

[58] Escanciano, J.C., 2006a. Goodness-of-fit tests for linear and non-linear time series models, Journal of the American Statistical Association, 101(474), 531–541.

[59] Escanciano, J.C., 2006b.Consistent diagnostic test for regression models using projections. Econometric Theory, 22(6), 1030–1051.

[60] Escanciano, J.C., 2007a. Model checks using residual marked empirical processes, Statistica Sinica, 17(1), 115–138.

[61] Escanciano, J.C., 2007b. Joint and marginal diagnostic tests for conditional mean and variance specifications. CAEPR Working Paper No. 2007-009.

[62] Escanciano, J.C., 2008. Joint and marginal specification tests for conditional mean and variance models. Journal of Econometrics, 143(1), 74-87.

[63] Escanciano, J.C., 2009. Simple bootstrap tests for conditional moment restrictions. Mimeo, available at http://www.ecares.org/ecaresdocuments/seminars0809/escanciano.pdf.

[64] Escanciano, J.C. and Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. Journal of Econometrics, 134(1), 151–185.

[65] Fan, J. and Gijbels, I., 1996. Local polynomial modelling and its applications: Monographs on Statistics and Applied Probability 66 (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). Chapman & Hall, London, New York.

[66] Fan, Y. and Li, Q., 2000. Consistent model specification tests: Kernel based tests versus Bierens' ICM tests. Econometric Theory, 16(6), 1016-1041.

[67] Francq, C. and Zakoian, J.M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH. Bernoulli, 10(4), 605–637.

[68] Glosten, L.R., Jagannathan, R. and Runkle, D.E., 1993. On the relation between expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48(5), 1779-1801.

[69] Gourieroux, C., 1997. ARCH models and financial applications. Springer-Verlag, New York.

[70] Hafner, C. M., and Rombouts, J. V. K., 2007. Semiparametric multivariate volatility models. Econometric Theory, 23(2), 251–280.

[71] Hafner, C.M., van Dijk, D. and Franses, P.H., 2005. Semi-parametric modelling of correlation dynamics. In: Fomby, T., Hill, C. and Terrell, D. eds. Advances in Econometrics, vol. 20/A: 59–103. Amsterdam: Elsevier Sciences.

[72] Hafner, C.M. and Franses, P.H., 2003. A generalized dynamic conditional correlation model for many asset returns. Econometric Institute Report EI 2003-18, Erasmus University Rotterdam, Econometric Institute.

[73] Hafner, C.M. and Herwartz, H., 2008. Analytical quasi maximum likelihood inference in multivariate volatility models. Metrika, 67(2), 219-239.

[74] Halunga. A. and Orme, C.D., 2009. First-order asymptotic theory for misspecification tests of GARCH models. Econometric Theory, 25(2), 364-410.

[75] Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica, 64(2), 413–430

[76] He, C., and Teräsvirta, T., 2004. An extended constant conditional correlation GARCH model and its fourth-moment structure. Econometric Theory, 20(5), 904-926.

[77] Holly, A., 1982. A remark on Hausman's specification test. Econometrica, 50(3), 749-759.

[78] Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function, Journal of the American Statistical Association, 94 (448), 1201–1220.

[79] Hong, Y., and Lee, T.H., 2003. Diagnostic checking for adequacy of nonlinear time series models. Econometric Theory, 19(6), 1065–1121.

[80] Hong, Y., and Lee, Y. J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. Review of Economic Studies, 72(2), 499-541.

[81] Jeantheau, T., 1998. Strong consistency of estimators for multivariate ARCH models. Econometric Theory, 14(1), 70–86.

[82] Jondeau E. and Rockinger, M., 2006. The copula-GARCH model of conditional dependencies: an international stock-market application. Journal of International Money and Finance, 25(5), 827-853.

[83] Kariya, T., 1988. MTV model and its application to the prediction of stock prices. In: Pullila, T. and Puntanen, S., eds. Proceedings of the second International Tampere Conference in Statistics, University of Tampere, Finland.

[84] Kawakatsu, H., 2003. Cholesky factor GARCH. Mimeo, Quantitative Micro Software, Irvine, CA.

[85] Kawakatsu, H., 2006. Matrix exponential GARCH. Journal of Econometrics, 134(1), 95-128.

[86] Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory of Probability and its Applications, 26 (2), 240–257.

[87] Koul, H.L. and Stute, W., 1999. Nonparametric model checks for time series. Annals of Statistics, 27 (1), 204–236.

[88] Kroner, K.F. and Claessens, S., 1991. Optimal dynamic hedging portfolios and the currency composition of external debt. Journal of International Money and Finance, 10(1), 131-148.

[89] Kroner, F.K. and. Ng, V.K., 1998. Modelling asymmetric comovements of asset returns. The Review of Financial Studies, 11(4), 817-844.

[90] Kroner, K.F., and Sultan, J., 1991. Exchange rate volatility and time varying hedge ratios. In: Rhee, S.G. and Chang, R.P. eds, Pacific-Basin Capital Market Research, 2, Amsterdam: North-Holland, 397-412.

[91] Kroner, K.F., and Sultan, J., 1993. Time varying distribution and dynamic hedging with foreign currency futures. Journal of Financial and Quantitative Analysis, 28(4), 535-551.

[92] Kyrtsou, C., 2008. Re-examining the sources of heteroskedasticity: The paradigm of noisy chaotic models. Physica A, 387(27), 6785-6789.

[93] Laurent, S., Bauwens, L. and Rombouts, J. V. K., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics, 21(1), 79-109.

[94] Lee, J., 1991. A Lagrange multiplier test for GARCH models. Economics Letter, 37(3), 265-271.

[95] Lee, J., 2006. The comovement between output and prices: Evidence from a dynamic conditional correlation GARCH model. Economics Letters, 91(1), 110-116.

[96] Lee, S. and Hansen, B. E., 1994. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory, 10(1), 29-52.

[97] Lee, T., White, H. and Granger, C.W.J., 1993. Testing for neglected nonlinearity in time series models. Journal of Econometrics, 56(3), 269-290.

[98] Li, W.K. and Mak, T.K., 1994. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. Journal of Time Series Analysis, 15(6), 627–636.

[99] Lien, D. and Tse, Y.K., 1998. Hedging time-varying downside risk. Journal of Futures Markets, 18(2), 705 - 722.

[100] Lien, D., Tse, Y.K. and Tsui, A. K. C., 2002. Evaluating the hedging performance of the constant-correlation GARCH model. Applied Financial Economics, 12(11), 791-798.

[101] Ling, S., and Li, W. K., 1997. Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. Journal of Time Series Analysis, 18(5), 447–464

[102] Ling, S. and McAleer, M., 2003. Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory, 19(2), 280–310.

[103] Long, X. and Ullah, A., 2005. Nonparametric and semiparametric multivariate GARCH model. University of California, Riverside.

[104] Longin, F. and Solnik, B., 1995. Is the correlation in international equity returns constant: 1960–1990? Journal of International Money and Finance, 14(1), 3–26.

[105] Longin, F. and Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance, 56(2), 651-678.

[106] Lumsdaine, R.L. and Ng, S., 1999. Testing for ARCH in the presence of a possibly misspecified mean. Journal of Econometrics, 93(2), 257–279.

[107] Lundbergh, S. and Teräsvirta, T., 2002. Evaluating GARCH models. Journal of Econometrics, 110(2), 417–435.

[108] Magnus, J. R. and Neudecker, H., 1999. Matrix Differential Calculus. Wiley Series in Probability and Statistics (revised edition). Wiley.

[109] Mammen, E., 1993. Bootstrap and wild bootstrap for high-dimensional linear models. Annals of Statistics, 21(1), 255-285.

[110] McAleer, M., 2005. Automated inference and learning in modelling financial volatility. Econometric Theory, 21(1), 232-261.

[111] McAleer, M., Chan, F., Hoti, S. and Lieberman, O., 2008. Generalized autoregressive conditional correlation. Econometric Theory, 24(6), 1554-1583.

[112] McAleer, M., Hoti, S. and Chan, F., 2009. Structure and asymptotic theory for multivariate asymmetric conditional volatility. Econometric Reviews, 28(5), 422-440.

[113] McCloud, N. and Hong, Y., 2008. Testing the structure of conditional correlations in multivariate GARCH models: A generalized cross-spectrum approach. Mimeo. (available at http://www.wise.xmu.edu.cn/Activity/XMU-HUB2008/paper/McCLOUD.pdf)

[114] McLeod, A. I. and Li, W. K., 1983. Diagnostic checking ARMA time series models using squared residual autocorrelations. Journal of Time Series Analysis, 4(4), 269–273.

[115] Nakatani, T. and Teräsvirta, T., 2008. Appendix to Testing for volatility interactions in the Constant Conditional Correlation GARCH model. Department of Economic Statistics, Stockholm School of Economics, available at http://swopec.hhs.se/hastef/abs/hastef0649.htm.

[116] Nakatani, T. and Teräsvirta, T., 2009. Testing for volatility interactions in the Constant Conditional Correlation GARCH model. Econometrics Journal, 12(1), 147-163.

[117] Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 59(2), 349–370.

[118] Nelson, D.B. and Cao, C.Q., 1992. Inequality constraints in the univariate GARCH model. Journal of Business and Economic Statistics, 10(2), 229-235.

[119] Nelson, D. B. and Foster, D. P., 1994. Asymptotic filtering theory for univariate ARCH models. Econometrica, 62(1), 1-41.

[120] Newey, W.K., 1985. Maximum likelihood specification testing and conditional moment tests. Econometrica, 53(5), 1047–1070.

[121] Newey, W.K. and McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R. and McFadden, D., eds. The Handbook of Econometrics, vol. 4. North-Holland, Amsterdam.

[122] Ngatchou-Wandji, J., 2005. Checking nonlinear heteroscedastic time series models. Journal of Statistical Planning and Inference, 133(1), 33–68.

[123] Noureldin, D., Shephard, N. and Sheppard, K., 2011. Multivariate High-Frequency-Based Volatility (HEAVY) models. Department of Economics Discussion Paper Series, University of Oxford.

[124] Pagan, A., 1996. The econometrics of financial markets. Journal of Empirical Finance, 3(1), 15-102.

[125] Pagan, A. and Ullah, A., 1999. Nonparametric econometrics. Cambridge University Press, Cambridge, New York .

[126] Palm, F.C., 1996. GARCH models of volatility. In: Maddala, G.S. and Rao, C.R., eds. Handbook of Statistics. Elsevier Science:Amsterdam.

[127] Park, T. and Switzer, L., 1995. Time-varying distribution and the optimal hedge ratios for stock index futures. Applied Financial Economics, 5(3), 131–137.

[128] Parzen, E., 1957. On consistent estimates of the spectrum of a stationary time series. The Annals of Mathematical Statistics, 28(2), 329–348.

[129] Pascual, L., Romo, J. and Ruiz, E., 2006. Bootstrap prediction for returns and volatilities in GARCH models, Computational Statistics & Data Analysis, 50 (9), 2293–2312.

[130] Patton, A., 2006. Modelling asymmetric exchange rate dependence. International Economic Review, 47(2), 527-556.

[131] Pelletier, D., 2006. Regime switching for dynamic correlations. Journal of Econometrics, 131(1-2), 445-473

[132] Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), 341-360.

[133] Sentana, E., 1995. Quadratic ARCH models. Review of Economic Studies, 62(4), 639–661.

[134] Sentana, E. and Fiorentini, G., 2001. Identification, estimation and testing of conditionally heteroskedastic factor models. Journal of Econometrics, 102(2), 143-164.

[135] Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Hinkley, D.V., Cox, D.R. and Barndorff-Nielsen, O.E., eds.

Time Series Models in Econometrics, Finance and Other Fields. Chapman & Hall: London.

[136] Silvennoinen, A. and Teräsvirta, T., 2005. Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations. SSE/EFI Working Paper Series in Economics and Finance No. 577.

[137] Silvennoinen, A. and Teräsvirta, T., 2009a. Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. Journal of Financial Econometrics, 7 (4): 373-411.

[138] Silvennoinen, A. and Teräsvirta, T., 2009b. Multivariate GARCH Models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. eds, Handbook of Financial Time Series, part 1, Springer-Verlag, Berlin, 201-229.

[139] Stelzer, R., 2008. On the relation between the VEC and BEKK multivariate GARCH model. Econometric Theory, 24(4), 1131-1136.

[140] Stinchcombe, M. and White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory, 14(3), 295–325.

[141] Stute, W., 1997. Nonparametric model checks for regression. Annals of Statistics, 25(2), 613–641.

[142] Stute, W., Gonzalez-Manteiga, W. and Presedo-Quindimil, M., 1998. Bootstrap approximations in model checks for regression. Journal of the American Statistical Association, 93(441), 141–149.

[143] Stute, W. and Zhu, L.X., 2002. Model checks for generalized linear models. Scandinavian Journal of Statistics, 29(3), 535-545.

[144] Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models. Journal of Econometrics, 30(1-2), 415-443.

[145] Tse, Y.K., 2000. A test for constant correlations in a multivariate GARCH model. Journal of Econometrics, 98 (1), 107–127.

[146] Tse, Y.K., 2002. Residual-based diagnostics for conditional heteroscedasticity models. Econometrics Journal, 5(2), 358-374.

[147] Tse, Y.K and Tsui, A.K.C., 1999. A note on diagnosing multivariate conditional heteroscedasticity models. Journal of Time Series Analysis, 20(6), 679-691.

[148] Tse, Y.K and Tsui, A.K.C., 2002. A multivariate GARCH model with time-varying correlations. Journal of Business and Economic Statistics, 20(3), 351-362.

[149] Tsui, A. K. and Yu, Q., 1999. Constant conditional correlation in a bivariate garch model: Evidence from the stock market in china, Mathematics and Computers in Simulation, 48(4-6), 503-509.

[150] van der Vaart, A. and Wellner, J., 1996. Weak convergence and empirical processes (with applications to statistics). Springer-Verlag, New York.

[151] van der Weide, R., 2002. GO-GARCH: a multivariate generalized orthogonal GARCH model. Journal of Applied Econometrics, 17(5), 49-564.

[152] Vrontos, I.D., Dellaportas, P. and Politis, D.N., 2003. A full-factor multivariate GARCH model. Econometrics Journal, 6(2), 311-333.

[153] Weiss, A. A., 1986. Asymptotic theory for ARCH models: Estimation and Testing. Econometric Theory, 2(1), 107-131.

[154] Whang, Y., 2000. Consistent bootstrap tests of parametric regression functions. Journal of Econometrics, 98(1), 27-46.

[155] White, H., 1982. Maximum Likelihood Estimation of Misspecified Models. Econometrica, 50(1), 1-25.

[156] Wooldridge, J., 1990. A unified approach to robust, regression-based specification tests. Econometric Theory, 6(1), 17-43.

[157] Zakoian, J-M., 1994. Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 18(5), 931-955.