Finite sample and optimal adaptive inference in possibly nonstationary general volatility models with gaussian or heavy-tailed errors [1]

Jean-Marie Dufour [2]
Université de Montréal

Emma M. Iglesias [3]
Michigan State University

April 2005

[1]This project was supported financially by an Economic and Social Research Council (UK) grant (Award number: T026 27 1238) and by Generalitat Valenciana (Spanish government). The second author also thanks the hospitality of the Department of Economics at the University of Montreal during several research visits.

[2]Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal. C. P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour

[3]Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA. e-mail: iglesia5@msu.edu

ABSTRACT

In this paper, we develop a general framework for optimal testing in volatility models from a finite-sample perspective. We first describe the general structure of point-optimal tests in the context of a general class of volatility models, which includes ARCH, GARCH and stochastic volatility models. Then we propose an adaptive methodology [the split-sample Monte Carlo adaptive optimal (SSMCAO) technique], based on combining (maximized) Monte Carlo tests with split-sample methods, which allows one to search for the best point alternative. We get in this way approximate point-optimal tests under the true DGP under consideration (adaptive point-optimal tests), such the power of the test converges to the power envelope as the sample size increases. Further, the proposed procedure does not require the existence of moments and can be applied to any type of distribution for the errors, including heavy-tailed ones. Stationary, unit root as well as non-stationary volatility processes are allowed. Under parametric distributional assumptions, the level of the test is perfectly controlled.

Although the theory presented covers a wide spectrum of volatility models, inference on GARCH models is studied in greater detail. We consider the general problem of testing any possible set of coefficient values in GARCH models, with Gaussian and non-Gaussian errors. Both Engle-type, local best invariant and point-optimal tests are studied. Special problems considered include the hypothesis of no-GARCH effects and integrated models. We show that the method suggested provides provably valid tests in both finite and large samples, in cases where standard asymptotic and bootstrap methods fail in the presence of heavy-tailed errors [as shown by [Hall and Yao, 2003]]. When it is possible to find a consistent estimate, the procedure touches the power envelope asymptotically. The performance of the proposed tests, with both Gaussian and non-Gaussian errors, is analyzed in a simulation experiment. Our results show that the proposed procedures work well from the viewpoints of size and power. In many cases, spectacular power gains provided are achieved. The tests also exhibit good behaviour outside the stationarity region. Finally, the technique is illustrated by an application to a model of US inflation.

Key words: hypothesis testing; testability; GARCH; asymptotic theory; exact inference; fat tails; nonstationarity; conditional heteroskedasticity; unit root; optimal testing; Monte Carlo test; Maximised Monte Carlo test; Split Sample Monte Carlo Adaptive Optimal Tests.

Journal of Economic Literature classification: C12; C22; C15.

Contents

1 Introduction

In this paper we provide a framework for exact optimal testing in volatility models. The use of exact theory, allows for the possibility of having a very flexible setting in the models where it is applied. Our proposal has the advantage of not requiring the existence of moments, and that it works straightforwardly both in the stationary, unit root and nonstationary regions. Our main novel contribution to the literature, is to propose an adaptive procedure that allows to search for the point alternative, at the same time that size is fully controlled in the point optimal test. We start by characterising both local best invariant tests [extending the work of [King and Hillier, 1985] and [King, 1980]] and approximate point optimal tests in cases where the model can have a markovian structure or not. Because we deal with starting values as nuisance parameters [following [Mueller and Elliott, 2003]], we can get exact results in a straightforward way. Our approach proposes a combination of Monte Carlo tests, Maximised Monte Carlo tests with a splitting technique. In optimal testing, the most important issue is how to select the point alternative to use. We propose two possible mechanisms for that: either to use a consistent estimate, or, when that is not possible, to search for the point that maximises asymptotically the empirical frequency of rejection at the same time that size is controlled regardless of the sample size. We show the performance of our approach by specilising our results on developing finite sample inference procedures as well as optimal tests for GARCH models. Specially, we concentrate on testing any value of the conditional heteroscedastic coefficients. This also allows us to test for the presence of GARCH effects and the unit root case [with the IGARCH process], and we compare our results with other available procedures. There are already available several tests for accounting for ARCH and GARCH effects: see for example [Engle, 1982], [Lee, 1991] and [Lee and King, 1993]. More recently, [Dufour et al., 2004] have proposed to improve the properties of the inference by exploiting the use of Monte Carlo (MC) techniques [see [Dufour and Kiviet, 1996] and [Dufour and Kiviet, 1998] for more details]. In this paper we propose also the use of MC techniques, although we go further, and we present new procedures for testing any value of coefficients of much more general models in the mean and in the conditional variance equation. Among others, we propose a point optimal test. We also use the Maximised Monte Carlo (MMC) technique to deal with nuisance parameters [see [Dufour, 2004] for more details]. As the main innovation of this paper, we propose a combination of Monte Carlo tests with splitting the sample in two steps, and we show its optimality properties. If the sample is divided into two and we condition on the history of the first part of the sample [order matters in this case], we get a full control of the size after step 2 of the procedure.

In the case of testing the null of IGARCH(1,1), [Lumsdaine, 1995] gets results that Wald tests seem to have the best size, although the standard Lagrange multiplier statistic is badly oversized. At the same time, versions of the LM that are robust to possible nonnormality of the data perform only marginally better. In any case, [Lumsdaine, 1995] reports that in general, from her simulations, the Lagrange multiplier, likelihood ratio and Wald do not behave very well in small samples. Our framework also allows us to test for this hypothesis and with the MC and the MMC techniques we will be able to control for the size.

We also show through simulation that the tests developed in this paper can be applied both in the framework of gaussian and non-gaussian errors; including errors following a t-distribution with very low degrees of freedom [situations that are very important in practice, and where asymptotic theory and standard bootstrap methods break down]. [Hall and Yao, 2003] report problems with the conservatism of their subsampling technique when the tails of the errors are very light and problems with the anticonservatism for heavy-tails [in their study, for sample sizes of 1000 and for nominal levels of 0.10, they tend to get size distorsions of around 0.30]. Due to the fact that our procedure controls the size and the exactness of the test, this makes our proposal more attractive than the subsampling technique of [Hall and Yao, 2003]. Moreover, apart from controlling for the size, our proposal maximises the empirical frequency of rejection asymptotically through the structure of the approximate point optimal test, providing very good power in finite samples, as we will show through simulation. That implies that the procedures developed in this paper can be used by practicioners in any type of scenarios: including gaussian and non-gaussian errors without being worried about the existence of moments. Besides, [Hall and Yao, 2003] only show results for sample sizes 1000 and 500 and even already in those cases their procedure has size problems. We will show the good perfomance of our procedure even for sample sizes of 50.

[Jensen and Rahbek, 004a], [Jensen and Rahbek, 004b] have shown recently that the QMLE is always asymptotically normal provided that the fourth moment of the innovation process is finite in GARCH models. However, although this shows what happens asymptotically, it is still unknown in the literature which is the effect of being in a non-stationarity region in finite samples. In this paper we also consider cases where we are in a non-stationary region, and we will provide the behaviour of the tests in this framework. Again in this case, the existence of moments is crucial for the result of [Jensen and Rahbek, 004a],[Jensen and Rahbek, 004b] [their result does not allow to deal with non-stationary regions and fat tails at the same time] while our tests [through the MC and the MMC] can work both in the non-stationary region and/or in those cases where moments do not exist [with very fat tails].

Finally, making use of the MMC technique, and dealing with starting values we can afford the existence of a very general framework both in the mean and in the conditional variance. With asymptotic approximations, this would change the framework of the test, while the MMC technique takes that into account directly. We also make operational our procedures for testing the null of a sub-group of the GARCH coefficients equal to a value by making use of the Maximised Monte Carlo (MMC) technique.

The inmediate applications of the results in this paper are several. Among them, first, from our inference procedures we can construct confidence intervals for conditional volatility. Second, we can retrieve from there predictions of volatility through the confidence intervals and point predictions. Third, from our method we can get predictions of the underlying variables.

The plan of the paper is as follows. We first present our general setting and how we can get both exact and approximate point optimal tests. Section 3 develops the new methodology of exact splitting: the Split Sample Monte Carlo Adaptive Optimal (SSMCAO) technique and its minimised (MSSMCAO) version. Later, in Sections 4 and 5 we specialise our results in the case of GARCH and ARCH models. We also carry out a simulation study to find out about the size and power properties of the proposed test procedures. We also show the usefulness of our test in practice, when we re-visit the analysis of the US implicit price deflator for GNP. Finally, in Section 6 we conclude.

2 Optimal inference in volatility models

We consider both local best invariant tests and approximate point optimal tests in four scenarios: local best invariant tests [extending the work of [King and Hillier, 1985] and [King, 1980]] when the parameters in a linear mean equation are estimated or not, and approximate point optimal tests where the parameters in the mean equation are estimated and both when the mean equation is linear or not.

2.1 Exact optimal and locally best invariant tests with a known linear regression in the mean equation

We start by presenting an exact point optimal test that will be the local best invariant test in the sense of [King and Hillier, 1985] and [King, 1980] when the parameters in the mean equation are known.

Let's suppose the model:

$$y_t = x_t\acute{}\beta + \varepsilon_t, t = 1, ...T \text{ or } y = X\beta + \varepsilon \tag{1}$$

$$h_t = E\left(\varepsilon_t^2 / J_{t-1} : \Phi\right);\ \varepsilon_t = \sqrt{h_t} u_t \tag{2}$$

where u_t can follow any type of distribution that can fit our data, the information J_{t-1} contains at least the minimal natural filtration associated to the process ε_{t-1} : $I_{t-1} = \sigma(\varepsilon_\tau, \tau \leq t-1)$. y is $T \times 1$, X is $T \times k$, fixed, and of rank $k \leq T$. Many models in the econometrics literature deal with those characteristics including the general clase of square-root stochastic autoregressive models, stochastic volatility models or GARCH [see e.g. [Andersen, 1994] and [Meddahi and Renault, 2004] for more details]. This specification would of course include models that could specify asymmetries, leverage effects, or any other specification typically used nowadays in the conditional variance equation. We also assume first that β is known.

We characterise now the exact optimal test in the next theorem. We assume a t-distribution and conditional normality to derive the test in closed form, although any generalization to a different distributional assumption would be straightforward [we could allow for example, for the skewed t-quasi likelihood function (see [Fernandez and Steel, 1998] and [Hansen, 1994]), or for the $\alpha-$stable distribution (see eg. [Garcia and Veredas, 2004])]:

Theorem 2.1 *Exact point optimal tests. Suppose that $y_t = x_t'\beta + \varepsilon_t$, β is known and $h_t = E(\varepsilon_t^2/J_{t-1} : \Phi)$. Then an exact point optimal test at $\Phi = \Phi^1$ for the null of $H_0 : \Phi = \overline{\Phi}$, under any distribution function of u_t is given by:*

$$LR\left(\Phi^1, \overline{\Phi}\right) = -2\left[l_T\left(\varepsilon_t\left(\Phi^1\right), \Phi^1\right) - l_T\left(\varepsilon_t\left(\overline{\Phi}\right), \overline{\Phi}\right)\right]$$

where l_T will be the likelihood function of the distribution of u_t. In the special case of the t-distribution:

$$l_T(\Phi) = \ln\left[\Gamma\left(\frac{v+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{v}{2}\right) - 0.5\ln\left[\pi(v-2)\right]\right]$$

$$-0.5\sum_{t=1}^{T}\left[\ln\left(E\left(\varepsilon_t^2/J_{t-1} : \Phi\right)\right) + (1+v)\ln\left(1 + \frac{\varepsilon_t^2}{(v-2)\,E\left(\varepsilon_t^2/J_{t-1} : \Phi\right)}\right)\right]$$

under conditional normality is given by:

$$LR\left(\Phi^1, \overline{\Phi}\right) = \sum_{t=1}^{T}\ln\left(\frac{E\left(\varepsilon_t^2/J_{t-1} : \Phi^1\right)}{E\left(\varepsilon_t^2/J_{t-1} : \overline{\Phi}\right)}\right) + \sum_{t=1}^{T}\left(\frac{\varepsilon_t^2}{E\left(\varepsilon_t^2/J_{t-1} : \Phi^1\right)} - \frac{\varepsilon_t^2}{E\left(\varepsilon_t^2/J_{t-1} : \overline{\Phi}\right)}\right)$$

Proof. Given in Appendix 1. ■

2.2 Exact optimal and locally best invariant tests when a linear regression is estimated in the mean equation

To extend the previous setting to the case where β is estimated, we first get the local best invariant test.

Following [King and Hillier, 1985] and [King, 1980], the previous problem is invariant under transformations of the form:

$$y \rightarrow \gamma_0 y + X\widetilde{\gamma}$$

where γ_0 is a positive scalar and $\widetilde{\gamma}$ is a real $k \times 1$ vector.

Let $M = I_T - X\left(X'X\right)^{-1} X'$, $z = My$, and P_1 be an $(T-k) \times T$ matrix such that $P_1 P_1' = I_{T-k}$ and $M = P_1' P_1$. We also define P_2 as an $k \times T$ matrix where $P = (P_1, P_2)'$, $PP' = I_T$, $P_2 P_2' = I_k$, $P_1 P_2' = 0$. Then:

Theorem 2.2 *Locally best invariant tests.Suppose that $y_t = x_t'\beta + \varepsilon_t, \beta$ is estimated and $h_t = E\left(\varepsilon_t^2 / J_{t-1} : \Phi\right)$. Then a local best invariant test for $H_0 : \Phi = 0$ against $H_1 : \Phi > 0$ is that with critical regions of the form:*

$$\frac{\partial \ln f(v;\Phi)}{\partial \Phi}|_{\Phi=0} > c_2$$

where c_2 is a suitable constant and $v = \frac{P_1 z}{(z'P_1'P_1 z)^{1/2}}$ is the maximal invariant under the above group of transformations. Besides:

$$f(v;\Phi) = \int f(\varepsilon, s; \Phi)\, ds$$

where $s = \frac{P_2 z}{(z'P_2'P_2 z)^{1/2}}$ and:

$$f(\varepsilon;\Phi) = f(\varepsilon_T / I_{T-1};\Phi)\, f(\varepsilon_{T-1}/I_{T-2};\Phi)\, f(\varepsilon_{T-2}/I_{T-3};\Phi) \ldots . f(\varepsilon_1;\Phi)$$

where $f(\varepsilon_1)$ is treated as a nuisance parameter to estimate.

To obtain the local best uniform invariant test yields critical regions w_0 of the form:

$$\frac{\partial^2 f(v;\Phi)}{\partial \Phi \partial \Phi'}|_{\Phi=0} > k_1 f(v;\Phi)|_{\Phi=0} + k_2 \frac{\partial \ln f(v;\Phi)}{\partial \Phi}|_{\Phi=0}$$

where k_1 and k_2 are constants so that w_0 satisfies the size condition:

$$\Pr\{v \in w_0 | \Phi = 0\} = \alpha$$

and the local unbiasedness condition:

$$\frac{\partial \Pr\{v \in w_0\}}{\partial \Phi}|_{\Phi=0} = 0$$

Proof. Based on [King and Hillier, 1985]. ■

Theorem 2.2 makes our proposal difficult to be operative from the applied point of view, unless an extensive simulation exercise is carried out to evaluate the integral. In relation to the exact optimal test, the expression is exactly the same as the one given in Theorem 2.1, due to the fact that we are using residual-based tests [see e.g. [Dufour et al., 2004] for more details]. Then we can justify that the previous test is invariant to the choice of the intercept in the conditional variance and in the mean equation, and to the parameters of any number of exogenous variables that are included in the mean equation. We justify this in the following proposition and corollary:

Proposition 2.3 *(Pivotality of a statistic). Under (1) and (2), let* $S(y,X) = (S_1(y,X), S_2(y,X), ..., S_m(y,X))'$ *be any vector of real-valued statistics such that*

$$S(cy + Xd, X) = S(y,X), \forall c, d \in R^k.$$

Then, for any positive constant $\sqrt{h_0} > 0$, *we can write*

$$S(y,X) = S\left(\varepsilon/\sqrt{h_0}, X\right)$$

and the conditional distribution of $S(y,X)$*, given* X*, is completely determined by the matrix* X *and the conditional distribution of* $\varepsilon/\sqrt{h_0} = \Delta\eta/\sqrt{h_0}$ *given* X*, where* $\Delta = diag\,(\sqrt{h_t}, t = 1, ..., T)$*. In particular, under* $H_0 : \Theta = \overline{\Theta}, \Phi = \overline{\Phi}$*, we have:*

$$S(y,X) = S(\eta, X)$$

where $\eta = \varepsilon/\sqrt{h_t}$*, and the conditional distribution of* $S(y,X)$*, given X, is completely determined by the matrix X and the conditional distribution of* η *given X.*

Proof. Taking $c = 1/\sqrt{h_0}$ and $d = -\beta/\sqrt{h_0}$, then:

$$cy + Xd = (X\beta + \varepsilon)/\sqrt{h_0} - X\beta/\sqrt{h_0} = \varepsilon/\sqrt{h_0}.$$

From (2) then, and under $H_0 : \Theta = \overline{\Theta}, \Phi = \overline{\Phi}$, we have that $\varepsilon = \Delta\eta = \sqrt{h_t}\eta$, and then, taking $\sqrt{h_0} = \sqrt{h_t}$, we get finally $\varepsilon/\sqrt{h_0} = \eta$ and $S(y,X) = S(\eta,X)$. ■

It is also necessary to prove that the pivotality characteristic also holds because in our case our statistics are scale-invariant functions of OLS residuals. We prove this in the next Corollary:

Corollary 2.4 *(Pivotal property of residual-based statistics). Under (1) and (2), let* $S(y,X) = (S_1(y,X), S_2(y,X), ..., S_m(y,X))'$ *be any vector of real-valued statistics such that*

$$S(y,X) = \overline{S}(A(X)y, X)$$

where $A(X)$ is any $n \times k$ matrix $(n \geq 1)$ such that

$$A(X)X = 0$$

and $\overline{S}(A(X)y, X)$ satisfies the scale invariance condition

$$\overline{S}(cA(X)y, X) = \overline{S}(A(X)y, X), forallc > 0.$$

Then for any positive constant $\sqrt{h_0} > 0$, we can write

$$S(y,X) = \overline{S}\left(A(X)\varepsilon/\sqrt{h_0}, X\right)$$

and the conditional distribution of $S(y,X)$, given X, is completely determined by the matrix X jointly with the conditional distribution of $A(X)\varepsilon/\sqrt{h_0}$ given X.

Because the previous expressions do not have a straightforward representation to apply in practice, we proceed in the next section to apply an approximate point optimal test.

2.3 Approximate optimal tests when the mean equation has a linear regression

Let's suppose that our setting is again the one given in the previous section where β is estimated. The approximate point optimal test will be given by the same expression as in Theorem 2.2, but where ε is replaced by $\widehat{\varepsilon}$ [the residuals], and again, we can use the invariance property given in Proposition 2.3 and Corollary 2.4.

2.4 Approximate optimal tests when the mean equation is nonlinear

This section contains the more general case that our theory covers. We start now by characterising the general framework of approximate point optimal tests. We allow for the existence of any process of the form:

$$y_t = f(\mu_t : \Theta) + \varepsilon_t$$

$$h_t = E\left(\varepsilon_t^2 / J_{t-1} : \Phi\right); \varepsilon_t = \sqrt{h_t} u_t$$

where u_t can follow any type of distributional function, the information J_{t-1} contains at least the minimal natural filtration associated to the process ε_{t-1} :

$I_{t-1} = \sigma(\varepsilon_\tau, \tau \leq t-1)$. $f(\mu_t : \Theta)$ may contain exogenous variables, ARMA or another [possibly nonlinear] processes, and Θ and Φ are the parameter vectors of interest. The main theorem states a general framework of approximate point optimal tests for any type of possibly non-linear process that would fit in the previous specification. Again, this specification would of course include models that would specify asymmetries, leverage effects, or any other specification typically used nowadays. Many models in the econometrics literature deal with those characteristics including the general class of square-root stochastic autoregressive models, stochastic volatility models or GARCH [see e.g. [Andersen, 1994] and [Meddahi and Renault, 2004] for more details]. We state the theorem in the context of the conditional normal distribution for $\frac{\varepsilon_t}{\sqrt{h_t}}$ for sake of simplicity, although any generalization to a different distributional assumption would be straightforward [we could allow for example, for the skewed t-quasi likelihood function (see [Fernandez and Steel, 1998] and [Hansen, 1994]), or for the $\alpha-$stable distribution (see eg. [Garcia and Veredas, 2004])], where using a pretest-estimator, first we would find the best distribution that conditionally would accommodate our data, and later we would apply in a second stage our approximate point optimal test where we would deal with the nuisance parameters with the MMC technique.

Theorem 2.5 *Approximate point optimal tests. Suppose that* $y_t = f(\mu_t : \Theta) + \varepsilon_t$ *and* $h_t = E\left(\varepsilon_t^2 / J_{t-1} : \Phi\right); \varepsilon_t = \sqrt{h_t} u_t$. *Then an approximate point optimal test at* $\Theta = \Theta^1, \Phi = \Phi^1$ *for the null of* $H_0 : \Theta = \overline{\Theta}, \Phi = \overline{\Phi}$*, is given by:*

$$LR\left(\overline{\Theta}, \overline{\Phi}, \Theta^1, \Phi^1\right) = -2\left[l_T\left(\varepsilon_t\left(\Theta^1, \Phi^1\right), \Theta^1, \Phi^1\right) - l_T\left(\varepsilon_t\left(\overline{\Theta}, \overline{\Phi}\right), \overline{\Theta}, \overline{\Phi}\right)\right]$$

under conditional normality is given by:

$$LR\left(\overline{\Theta}, \overline{\Phi}, \Theta^1, \Phi^1\right) = \sum_{t=1}^{T} \ln\left(\frac{E\left(\varepsilon_t^2 / J_{t-1} : \Phi^1\right)}{E\left(\varepsilon_t^2 / J_{t-1} : \overline{\Phi}\right)}\right) + \sum_{t=1}^{T}\left(\frac{\varepsilon_t^2\left(\Theta^1\right)}{E\left(\varepsilon_t^2 / J_{t-1} : \Phi^1\right)} - \frac{\varepsilon_t^2\left(\overline{\Theta}\right)}{E\left(\varepsilon_t^2 / J_{t-1} : \overline{\Phi}\right)}\right)$$

Proof. Given in Appendix 1. ■

To understand better Theorem 2.5, let's start by clasifying the type of models that have the property of autoregression of the variance in two cases:

1. Those models that can be expressed as markovian processes. An example of this case would be the AR(p) or ARCH(p) models. In this case, we do not need to deal with nuisance parameters to test for the whole parameter vector, so we can simply use the Monte Carlo technique [see [Dufour and Kiviet, 1996], [Dufour and Kiviet, 1998], [Dufour et al., 2004] for more details]. If the practicioner would be interested to test null hyphotesis different from the whole parameter vector, it would be possible again thanks to the MMC technique. We see later more in detail the case of the ARCH(p) model.

2. Those models that cannot be expressed as markovian processes. A typical example of this case is the ARMA(p,q) or the GARCH(p,q) model or combinations of both. We can express these models as a function of parameters and starting values. This is one of the novel approaches we propose in this paper: in order to provide provably valid inference of the previous point optimal test, we propose to deal with the initial values as nuisance parameters. Due to the non-standard distribution of the statistic, we propose to retrieve the critical values of the point optimal test through the MMC technique [see [Dufour, 2004] for more details]. We will see more in detail in the next section how we can deal with the GARCH(p,q) model, and how a much simpler expression of the point optimal test can be found. This approach allows for enough flexibility to consider non-stationary scenarios as well as regions of the parameter space where the moments do not exist [a crucial issue in volatility models].

3 Split-sample monte carlo adaptive optimal (SSMCAO) tests

Because our proposal includes the use of exact approximate point optimal tests, a primary concern is to find a methodology to find the point to optimise the test. There are already several alternatives available in the literature: one could look for the middle of the parameter space [as in [Elliott et al., 1996]] or to maximise power against a weighted average of alternatives [[Andrews and Ploberger, 1994]]. Our proposal includes the novelty of selecting the point to optimise the test from the own data directly: the Split-Sample Monte Carlo Adaptive Optimal (SSMCAO) tests. The two stages procedure is described below:

Let's suppose that $\theta^1 = \left(\theta_1^1, \theta_2^1, ..., \theta_s^1\right)'$ are the possible points to optimise the point optimal test, and let's suppose that we do not have nuisance parameters [otherwise instead of applying the Monte Carlo technique, we would use the Maximised Monte Carlo technique]. Let's suppose we have an original sample size "T". Let's divide T in two subsamples: "q" the first and "p" the second. We will use q for the first step and p for the second step [the order is crucial when dealing with dynamic models].

STEP 1:

1. In case asymptotic theory would work, one possibility is to find a consistent estimator for the point to optimise the test. We would estimate the model to

find a consistent estimate $\left(\theta^{1*}\right)$ with the first part of the sample q, and later we move to step 2.

2. However, we want to generalise this result and to make it general enough to work as well when asymptotics breaks down and it is not possible or extremely difficult to get a consistent estimate [and to be able to deal with situations like in [Hall and Yao, 2003]]. In this case, we propose to minimise the empirical p-value from the first subsample in order to find the chosen θ^{1*} [the point alternative] that it will be used in step 2 with the second subsample. This originates a Minimised Split-Sample Monte Carlo Adaptive Optimal (MSSMCAO) test. The idea is based on chosing the θ^{1*} that produces the higher frequency of rejection from our sample. In more detail:

Select a first point to optimise the test θ_1^1 and compute the corresponding statistic $S_0\left(y_t, \theta_1^1, \theta_0\right)$ for the first subsample of size p. Then retrieve a simulated p-value $\widehat{p}_{N1}$ for θ_1^1 through the Monte Carlo technique [in case of the presence of nuisance parameters, we would use the MMC technique]:

$$\widehat{G}_{N1}\left(S_0\left(y_t, \theta_1^1, \theta_0\right), \theta_0, \theta_1^1\right) = \frac{\sum_{i=1}^{N} 1\left(S_i\left(\theta_0, \theta_1^1, \theta_0\right) - S_0\left(y_t, \theta_1^1, \theta_0\right)\right)}{N}$$

where $S_1, ..., S_N$ are simulated from the null hypothesis. y_t is our available data for each subsample. And compute:

$$\widehat{p}_{N1}\left(S_0\left(y_t, \theta_1^1, \theta_0\right), \theta_0, \theta_1^1\right) = \frac{N\widehat{G}_{N1}\left(S_0\left(y_t, \theta_1^1, \theta_0\right), \theta_0, \theta_1^1\right) + 1}{N+1}$$

Repeat the procedure with θ_2^1 :

$$\widehat{G}_{N2}\left(S_0\left(y_t, \theta_2^1, \theta_0\right), \theta_0, \theta_2^1\right) = \frac{\sum_{i=1}^{N} 1\left(S_i\left(\theta_0, \theta_2^1, \theta_0\right) - S_{0j}\left(y_t, \theta_2^1, \theta_0\right)\right)}{N}$$

where $S_1, ..., S_N$ are simulated from the null hypothesis. And compute:

$$\widehat{p}_{N2}\left(S_0\left(y_t, \theta_2^1, \theta_0\right), \theta_0, \theta_2^1\right) = \frac{N\widehat{G}_{N2}\left(S_{0j}\left(y_t, \theta_2^1, \theta_0\right), \theta_0, \theta_2^1\right) + 1}{N+1}$$

Finally, with the θ_i^1, $i = 1, ..., s$ different possible points to optimise the test, compute:

$$\widehat{p}_{Ni}\left(S_0\left(y_t,\theta_i^1,\theta_0\right),\theta_0,\theta_2^1\right)=\frac{N\widehat{G}_{Ni}\left(S_0\left(y_t,\theta_i^1,\theta_0\right),\theta_0,\theta_i^1\right)+1}{N+1}$$

Finally the criterium is to choose:

$$\theta^{1*}=\arg\min_{\theta^1}\left\{\widehat{p}_N\left(S_0\left(y_t,\theta^1,\theta_0\right),\theta_0,\theta^1\right)\right\},\quad \theta^1=\left(\theta_1^1,\theta_2^1,...,\theta_s^1\right)'$$

$\widehat{p}_N\left(S_0\left(y_t,\theta^1,\theta_0\right),\theta_0,\theta^1\right)$ is the empirical p-value. The control for the size is preserved through the sample splitting. And when the data comes from the alternative hypothesis, we want to choose θ^{1*} in such a way that our sample rejects the null hypothesis as much as we can [the smallest empirical p-value].

In practice, we propose to apply the simulated annealing algorithm [see eg. [Goffe and Rogers, 1994] for more details] to search on the parameter space of θ_i^1 till we reach the minimum of the empirical p-value in the first step.

STEP 2. With θ^{1*} from STEP 1, finally we apply to the remaining sample "p" the Monte Carlo technique. This second step allows to control size both in large and finite samples because we condition to the history of the data in the first part of the sample.

$$\widehat{G}_N\left(S_0\left(y_t,\theta^{1*},\theta_0\right),\theta_0,\theta^{1*}\right)=\frac{\sum_{i=1}^{N}1\left(S_i\left(\theta_0,\theta^{1*},\theta_0\right)-S_0\left(y_t,\theta^{1*},\theta_0\right)\right)}{N}$$

where $S_1,...,S_N$ are simulated from the null hypothesis. And compute:

$$\widehat{p}_N\left(S_0\left(y_t,\theta^{1*},\theta_0\right),\theta_0,\theta^{1*}\right)=\frac{N\widehat{G}_N\left(S_0\left(y_t,\theta^{1*},\theta_0\right),\theta_0,\theta^{1*}\right)+1}{N+1}$$

The proposed procedure holds the next properties:

Theorem 3.1 *Split maximised technique. (a) The procedure defined in STEP 1 and STEP 2, is provably valid both in finite and large samples [both when finding a consistent estimate of θ^1 and when minimising the empirical p-value].*

(b) The procedure defined in STEP 1 and STEP 2, when θ^1 is chosen as a consistent estimate of the θ from where the sample comes, it is the most powerful test and it touches the power envelope asymptotically.

(c) The procedure defined in STEP 1 and STEP 2, when θ^1 is chosen as the minimum empirical p-value from the sample, asympotically it maximises the probability of rejection from the sample.

Proof. (a): by assuming that in step 2 we condition on the history of the first sample of the data [from 1 to q], the procedure controls size in finite samples and asymptotically. Here the order matters, and, because of the possible recursive nature of the data, it is necessary that we split the sample into two, and we use the first sample for step 1 and the second for step 2.

(b): Because of the consistency of the estimate, asymptotically and based on a mean-value expansion, the procedure touches the power envelope.

(c): by assuming that $q/T \to 0$, along with $q \to \infty$, $T \to \infty$, [the same as the requirements in [Hall and Yao, 2003]], by construction θ^{1*} is chosen such that it maximises the probability of rejection from the sample. ■

4 The GARCH(p,q) case

4.1 The setting and the tests

We apply now our SSMCAO test to GARCH models. We consider the case of univariate GARCH models, although our methodology may be applied to multivariate settings (see eg. [Cajigas and Urga, 2004]) and other type of models. The GARCH(p,q) model will be an important example of the flexibility that our procedure offers and the important applicability in practice. Our procedure allows to deal many problems that cannot be solved nowadays by asymptotic theory or standard bootstrap procedures. We consider the model:

$$y_t = x_t' \beta + \varepsilon_t \tag{3}$$

$$\varepsilon_t = \sqrt{h_t} u_t, t = 1, ...T, \; h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \sum_{i=1}^{p} \theta_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-j} \tag{4}$$

To have a better understanding, let's consider first the case where we only deal with exogenous in the mean equation and we are interested in the problem of testing any possible set of values for the GARCH coefficients:

$$H_0 : \theta_i = \overline{\theta}_i, \gamma_j = \overline{\gamma}_j \tag{5}$$

Again, the case of a conditional gaussian distribution is a special case, although we can allow for possibly any other distribution for u_t. Although the example that we follow in this section implies that we are not dealing with nuisance parameters in the mean equation, our framework allows for that by using the MMC technique.

As we already said in the previous section, due to the fact that we are using residual-based tests [see e.g. [Dufour et al., 2004] for more details], we can justify that the tests are invariant to the choice of the intercept in the conditional variance and in the mean equation, and to the parameters of any number of exogenous variables that are included in the mean equation.

Although we do not have nuisance parameters in the mean equation, we have still the initial values in the conditional variance equation. In this paper we deal with them as nuisance parameters in order to get provably valid inference [following [**?**]]. In this case, the point optimal test would take a much simpler form than the one in Theorem 2.5:

Theorem 4.1 *Under conditional normality. Suppose that* $y_t = x_t\acute{}\beta + \varepsilon_t$ *and* $h_t = E\left(\varepsilon_t^2/I_{t-1}\right) = \theta_0 + \sum_{i=1}^{p}\theta_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\gamma_j h_{t-j}$*. Then a point optimal test at* $\theta_i = \theta_i^1, \gamma_j = \gamma_j^1$ *for the null of* $H_0 : \theta_i = \overline{\theta}_i, \gamma_j = \overline{\gamma}_j$*, under conditional normality is given by:*

$$LR\left(\overline{\theta}_i, \overline{\gamma}_j, \theta_i^1, \gamma_j^1, h_0, \varepsilon_0\right) = \sum_{t=1}^{T}\ln\left(\overline{h}_t\right) + \sum_{t=1}^{T}\left(\frac{1}{\overline{h}_t} - 1\right)\frac{\varepsilon_t^2}{\theta_0 + \sum_{i=1}^{p}\overline{\theta}_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\overline{\gamma}_j h_{t-j}}$$

where $\overline{h}_t = \left(\frac{\theta_0 + \sum_{i=1}^{p}\theta_i^1\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\gamma_j^1 h_{t-j}}{\theta_0 + \sum_{i=1}^{p}\overline{\theta}_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\overline{\gamma}_j h_{t-j}}\right)$.

Proof. Given in Appendix 2. ■

h_0 and ε_0 are treated as nuisance parameters through the MMC technique. We can include more complicated structures in the mean and the conditional variance equation, and using the MMC technique. [Hall and Yao, 2003], [Jensen and Rahbek, 004a] and [Jensen and Rahbek, 004b] have brought into consideration an important issue in practice and it is the behaviour of these models in the non-stationary region and in the case of fat tails. Our framework allows straightforwardly to consider their framework, and in a much more general setting, because we allow for a full iteraction between mean and conditional variance equation, while their framework is much more restrictive. The procedure in [Hall and Yao, 2003] only shows the good asymptotic properties, although later in finite samples, there are strong size distorsions and dependence of the choice of the subsample size. We proceed now to provide some simulation results of our proposal.

4.2 Simulation results

We consider the GARCH(1,1) and we analyse the settting [to follow the simulation results given in [Hall and Yao, 2003]]:

$y_t = \varepsilon_t$, where $h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \gamma_1 h_{t-1}$

The simulation results are given in Appendix 3. We have selected two null hypothesis: $\theta_1 = 0.5$ and $\gamma_1 = 0.4$ [the same null analysed in [Hall and Yao, 2003]] and a non-stationary region null hypothesis to see the consequences of that [to explore the setting analysed in [Jensen and Rahbek, 004a], [Jensen and Rahbek, 004b]]. In order to check the robustness of our result to the distributional error, we show the results both when the normal distribution is the correct one, and when the errors are finally distributed with a t-distribution with 3 degrees of freedom. We want to analyse the case where the true errors come from a t-distribution with 3 degrees of freedom to compare our results with those of [Hall and Yao, 2003], and because in this case to find a consistent estimate for the GARCH parameters it is a difficult task, we use our SSMCAO procedure by finding the point alternative in relation to the minimum p-value in the first stage, as well as the result we would get through QML. SSMCAO when we use the minimum p-value from the first part of the sample, produces the Minimised SSMCAO (MSSMCAO) (in Appendix 3, this is denoted SSMMCAO m-p). Because we treat the initial values as nuisance parameters, we have to use instead of the MC technique, the MMC technique. And when this is combined with our split sample technique and the minimum p-value, this produces the Minimised Split-Sample Maximised Monte Carlo Adaptive Optimal (MSSMMCAO) test . We also show the results of the SSMMCAO when we use the QML estimator to optimise the test in the second stage (even although for t(3) errors, the estimate is not consistent). Besides, we also show the result of the point optimal test when the full sample size is used both for the first and the second stage (PO(full sample)).

[Hall and Yao, 2003] only offered results about the size of their test. We first see how our test procedure fully controls the size [in all the cases where we split the sample, except when the full sample is used for both the first and second stage]. We also show the performance of our test in relation to power. We see how indeed the power of the point optimal test is very sensitive against the alternative to which it is maximised (in some cases, where the wrong alternative in the point optimal test is used, the power can be very low, and even as well when the point is very close to the true one), and how our proposal of splitting the sample offers very good finite sample results when the researcher does not a any a priori information about where to optimise the approximate point optimal test. Our proposal has very good power both in the cases of the stationary and the non-stationary region and it is quite close to the power envelope. For our split sample technique we have used 1000 replications, the MMC technique in STEP 1 to deal with the starting values as nuisance parameters, and when the sample size is 50, we have used 25 for the first step and 25 for the second step. In case the total size was 200, we have used 50 for the first step and 150 for the second step. A general rule that is supported through simulations, it is that it is advisable to use around the 20% of the sample for step 1 and 80% for step 2. We see how the procedure of minimising the p-value always have very good results. The procedure of using the consistent estimate has very good results under normality (when the estimator is consistent), although the MSSMMCAO technique is always better when the QML does not offer a consistent estimator. In any case, the SSMMCAO technique clearly outperforms the subsampling technique of Hall and Yao (2003) (controls for size and it is very close to the power envelope). The procedure of using the full sample for both stages offers large size distorsions.

We also show the results of using retrieving the QMLE from the whole sample size and to use it to optimise a point optimal test in the whole sample size again. It is shown that the test is not exact (there are size distorsions) and the power increments do not overcome the size distorsions (what supports the use of our split sample technique).

5 The ARCH(p) case

5.1 Alternative test statistics for any value of the ARCH coefficients

We proceed now to propose alternative tests for any value of the AR(q)-ARCH(p) coefficients. We could also allow for more complicated ARMA-GARCH processes, but in order to keep the notation simple, we specialise this section in the AR(q)-ARCH(p) case. We consider the model:

$$y_t = x_t\acute{}\beta + \sum_{i=1}^{q} \phi_i y_{t-i} + \varepsilon_t \tag{6}$$

$$\varepsilon_t = \sqrt{h_t} u_t, t = 1, ...T,\ h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + ... + \theta_p \varepsilon_{t-p}^2 \tag{7}$$

where $x_t = (x_{t1}, x_{t2}, ..., x_{tk})\acute{}$, $X \equiv [x_1, ..., x_T]\acute{}$ is a full-column rank $T \times k$ matrix, $\beta = (\beta_1, ..., \beta_k)\acute{}$ is a $k \times 1$ vector of unknown coefficients, $\sqrt{h_1}, ..., \sqrt{h_T}$ are [possibly random] scale parameters, and $u_t = (u_1, ..., u_T)\acute{}$ is a random vector. The case of a conditional gaussian distribution is a special case, although we can allow for possibly any other distribution.

Let's consider first the case where we only deal with exogenous in the mean equation. We are interested in the problem of testing any possible set of values for the ARCH(p) coefficients:

$$H_0 : h_t = \overline{h}_t = \overline{\theta}_0 + \overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2 \tag{8}$$

We stress the fact that our procedures allow as well to test the null hypothesis of no-ARCH and the integrated ARCH cases. Our scenario using the MC technique allows us as well the introduction of exogenous variables in the mean equation, and we also show later in the simulation study that we can allow for the presence of normal or non normal errors (including those cases where asymptotic theory even breaks down [see [Hall and Yao, 2003]].

In the case we wanted to test sub-vector coefficients of the ARCH process, we could still preserve the exactness of the test by dealing with the nuisance parameters through the MMC technique. The same type of methodology would be used in case we would need to incorporate dynamics in the mean equation.

5.2 Extension of the Engle test under a null hypothesis different from the one of no-ARCH effects

A first alternative we consider is what we name to be an extension of "an Engle-type test". The original [Engle, 1982] test was designed as an LM test:

$$h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + ... + \theta_p \varepsilon_{t-p}^2$$

where the null hypothesis is $H_0 : \theta_1 = \theta_2 = ... = \theta_p = 0$, h_t is the conditional variance, and the test is formulated by TR^2 where T is the sample size and R^2 is the determination coefficient of a regression of OLS residuals $\widehat{\varepsilon}_t^2$ on a constant and $\widehat{\varepsilon}_{t-i}^2$ for $i = 1, ..., p$. The test is distributed as a χ^2 with p degrees of freedom.

In this paper we propose the next extension:

Theorem 5.1 *Suppose that* $y_t = x_t'\beta + \varepsilon_t$ *and* $h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + ... + \theta_p \varepsilon_{t-p}^2$, *then, an extension of the [Engle, 1982] test to any possible set of values of the ARCH coefficients* $H_0 : \theta_1 = \overline{\theta}_1, \theta_2 = \overline{\theta}_2, ..., \theta_p = \overline{\theta}_p$, *is given by* $TR^2 \sim \chi_p^2$, *where* $\overline{h}_t = \overline{\theta}_0 + \overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2$ *and* R^2 *is the determination coefficient coming from the regression:*

$$\varepsilon_t^2 = 2\overline{h}_t^2 \left[\gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 + ... + \gamma_p \varepsilon_{t-p}^2\right] + \overline{h}_t + v_t.$$

and it is asymptotically normally distributed as a χ_p^2. *Besides, the statistic is pivotal in finite samples.*

Proof. Given in Appendix 4, and a direct consequence of Proposition 2.3 and Corollary 2.4. ■

In practice, the test would imply to take the residuals $\widehat{\varepsilon}_t^2$, to compute the dependent variable $\left(\frac{\widehat{\varepsilon}_t^2}{2\left(\overline{\theta}_0 + \overline{\theta}_1 \widehat{\varepsilon}_{t-1}^2 + ... + \overline{\theta}_p \widehat{\varepsilon}_{t-p}^2\right)^2} - \frac{1}{2\left(\overline{\theta}_0 + \overline{\theta}_1 \widehat{\varepsilon}_{t-1}^2 + ... + \overline{\theta}_p \widehat{\varepsilon}_{t-p}^2\right)}\right)$, and to regress this depend variable on a constant and $\widehat{\varepsilon}_{t-i}^2$ for $i = 1, ..., p$. The test is distributed as a χ_p^2.

5.3 A point optimal test

To find out the most powerful test for ARCH processes, we develop now a point optimal test.

Theorem 5.2 *Suppose that* $y_t = x_t \acute{} \beta + \sum_{i=1}^{q} \phi_i y_{t-i} + \varepsilon_t$ *and* $h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + ... + \theta_p \varepsilon_{t-p}^2$. *Then a point optimal test at* $\theta_1 = \theta_1^1, ..., \theta_p = \theta_p^1$ *for the null of* $H_0 : \theta_1 = \overline{\theta}_1, \theta_2 = \overline{\theta}_2, ..., \theta_p = \overline{\theta}_p$, *under conditional normality is given by:*

$$LR\left(\overline{\theta}_1, ..., \overline{\theta}_p, \theta_1^1, ..., \theta_p^1\right) = \sum_{t=1}^{T} \ln\left(\overline{h}_t\right) + \sum_{t=1}^{T} \left(\frac{1}{\overline{h}_t} - 1\right) \frac{\varepsilon_t^2}{\left(1 + \overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2\right)}$$

where $\overline{h}_t = \left(\frac{\left(1+\theta_1^1 y_{t-1}^2 + ... + \theta_p^1 y_{t-p}^2\right)}{\left(1+\overline{\theta}_1 y_{t-1}^2 + ... + \overline{\theta}_p y_{t-p}^2\right)}\right)$.

Proof. It comes as a special case from the GARCH(p,q) model. ■

In practice, the test implies the following. Take the residuals $\widehat{\varepsilon}_t^2$, and compute the test statistic:

$$\sum_{t=1}^{T} \ln\left(\frac{\left(1+\theta_1^1 \widehat{\varepsilon}_{t-1}^2 + ... + \theta_p^1 \widehat{\varepsilon}_{t-p}^2\right)}{\left(1+\overline{\theta}_1 \widehat{\varepsilon}_{t-1}^2 + ... + \overline{\theta}_p \widehat{\varepsilon}_{t-p}^2\right)}\right) + \sum_{t=1}^{T} \left(\frac{\left(1+\overline{\theta}_1 \widehat{\varepsilon}_{t-1}^2 + ... + \overline{\theta}_p \widehat{\varepsilon}_{t-p}^2\right)}{\left(1+\theta_1^1 \widehat{\varepsilon}_{t-1}^2 + ... + \theta_p^1 \widehat{\varepsilon}_{t-p}^2\right)} - 1\right) \frac{\widehat{\varepsilon}_t^2}{\left(1+\overline{\theta}_1 \widehat{\varepsilon}_{t-1}^2 + ... + \overline{\theta}_p \widehat{\varepsilon}_{t-p}^2\right)}$$

Due to the non-standard distribution of the statistic, in this paper we propose to retrieve the critical values through the MC technique [see [Dufour and Kiviet, 1996], [Dufour and Kiviet, 1998] and [Dufour et al., 2004] for more details]. In case we would have an AR(p) process in the mean equation, we could use the MMC technique to treat those coefficients as nuisance parameters.

5.4 An alternative test

Another possible test is constructed by exploiting the use of pivotal properties of ARCH processes:

Theorem 5.3 *Suppose that* $y_t = x_t \acute{} \beta + \varepsilon_t$ *and* $h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + ... + \theta_p \varepsilon_{t-p}^2$, *then, a test for the null of* $H_0 : \theta_1 = \overline{\theta}_1, \theta_2 = \overline{\theta}_2, ..., \theta_p = \overline{\theta}_p$, *is given as:*

$$\frac{\varepsilon_t^2}{\left(1+\overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2\right)} - 1 = \theta_1^* \frac{\varepsilon_{t-1}^2}{\left(1+\overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2\right)} + ... + \theta_p^* \frac{\varepsilon_{t-p}^2}{\left(1+\overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2\right)} + w_t$$

This implies:

$$\varepsilon_t^2 - \left(1 + \overline{\theta}_1 \varepsilon_{t-1}^2 + ... + \overline{\theta}_p \varepsilon_{t-p}^2\right) = \theta_1^* \varepsilon_{t-1}^2 + ... + \theta_p^* \varepsilon_{t-p}^2 + w_t$$

where $\theta_i^* = \left(\theta_i - \overline{\theta}_i\right), \forall i = 1, .., p$. *The test can be re-written as an F-type test for the null of* $H_0 : \theta_1^* = 0, \theta_2^* = 0, ..., \theta_p^* = 0$. *The statistic is pivotal in finite samples.*

Proof. Given in Appendix 5, and a direct consequence of Proposition 2.3 and Corollary 2.4. ■

In practice, the test implies the following. Take the residuals $\widehat{\varepsilon}_t^2$, and regress $\frac{\widehat{\varepsilon}_t^2}{\left(1+\overline{\theta}_1\widehat{\varepsilon}_{t-1}^2+...+\overline{\theta}_p\widehat{\varepsilon}_{t-p}^2\right)}$ on $\frac{\widehat{\varepsilon}_{t-i}^2}{\left(1+\overline{\theta}_1\widehat{\varepsilon}_{t-1}^2+...+\overline{\theta}_p\widehat{\varepsilon}_{t-p}^2\right)}$ for $i = 1, ..., p$. The critical values can be obtain from the asymptotic theory from an F-distribution or through the MC technique.

Finally, it is important to stress that we can apply these tests not only in the presence of pivotality when we test for the null of the full ARCH coefficients vector [using the MC technique] equal to a value; but also, using the Maximised Monte Carlo (MMC) technique, we can handle those cases where we loose the pivotal property when we test only for a sub-vector of those ARCH coefficients equal to a value [see [Dufour, 2004] for more details)]. The MMC also allows for the presence of processes in the mean equation.

5.5 Simulation results

The main objective of this section is to show the poor finite sample properties that we can get when we use asymptotic theory for testing procedures regardless if we have fat tails or not and if we are in a nonstationary region or not. We proceed now to compare our different test procedures in the context of an ARCH(2) process under normal, t(5) and t(3) errors. This model was already analysed in [Hall and Yao, 2003] and it was chosen for comparative purposes. The model is then given by:

$y_t = \varepsilon_t$, where $E\left(\varepsilon_t^2/I_{t-1}\right) = \theta_0 + \theta_1\varepsilon_{t-1}^2 + \theta_2\varepsilon_{t-2}^2$. We take $\theta_0 = 0.81$.

We consider seven different types of null hypothesis. The results are given in Appendix 6. We will present the results both using asymptotic theory [when it is available] and the MC and MMC techniques [see [Dufour and Kiviet, 1996], [Dufour and Kiviet, 1998] and [Dufour et al., 2004] for more details]. While in [Hall and Yao, 2003] they only reported results for sample sizes 500 and 1000 and for size, we will report results for sample sizes 500, 200 and 50, to show the good finite sample properties of our procedures. We also consider a much more variety of null and alternative hypothesis than those given in [Hall and Yao, 2003]. In this last paper, they report that with their proposed subsampling technique, light tails in the distribution of the errors tend to produce relatively conservative confidence intervals. On the other hand, for extreme heavy-tailed-errors (t-3) the anticonservatism becames a problem. Our procedure allows to control for the size, and besides, we will provide clear results about the power properties. The objective of this simulation study is again to show the robustness of our procedure to the distributional assumptions of the errors and having fat tails. We also compare our three test proposals, and which are the consequences of relying on the asymptotic approximations or using the exact distribution. Also, [Hall and Yao, 2003] only showed the results for size, while we will report results for size and power.

The results for the Monte Carlo technique are based on 40000 replications, and N=99. The results for asymptotics where carried out with 40000 replications. The results for the simulated annealing [for the Maximised Monte Carlo technique] have been based on 1000 replications and N=99.

We cover 7 different types of nulls in the simulation results. All tests are very conservative in relation to the size distorsion.We proceed now to comment in detail results in each of the null hypothesis:

1) the first one, is the case of the IARCH(2) process. Here specially, the advantages of the point optimal test are huge. The point optimal test evaluated in the middle of the parameter space seems a very good alternative [following [Elliott et al., 1996]]. We also check that the procedure does not loose important power when the residuals are not gaussian. In this case, for alternatives with low values of the coefficients, to set the point optimal test to the middle value seems not to have good power properties. In this case, it would more advisable to use a small fraction of the sample [at the beginning] to estimate a plausible value for the alternative, and the rest to perform the test by using our splitting sample technique. Test 2 also has a good performance.

2) the second one is where θ_1=0.98 and θ_2=0.01; namely, it is similar to an ARCH(1) process [because θ_2 is very small], and there, any point optimal test seems to be doing it ok, what means that to set the point to optimize the test equal to (0.49, 0.49) seems to be a good suggestion [in an ARCH(1), then for nulls of the type near the unit root it would be a good idea to set the point optimal test to maximize in 0.49]. Test 2 also provides a good performance.

3) The third case is when the null is quite close to that recommendation of setting the point to the middle of the parameter space. Here the recommendation to the researcher is to follow again the route of maximing power through our split sample technique [as the proposal in this paper], since for this type of nulls, the point optimal test has very different powers depending on when it is the true alternative. Test 2 has a bad performance power in this case.

4) The fourth null covers the case of very low values of θ_1 and θ_2, and there a good point to optimize the test is again the recommendation of [Elliott et al., 1996]: (0.49, 0.49) for the ARCH(2) [and for the ARCH(1), it would be in 0.49].

5) The fifth case covers the case of testing for ARCH effects. Here, the Engle test has a much better performance than for the null of any other value. Even so, our point optimal test has superior properties to the [Engle, 1982] test, specially for low value alternatives and low degrees of freedom as well. Our point optimal test also has very good power when the distribution of the innovation process has

very fat tails.

6) The sixth null relates to the case where we would be interested in testing for subsets of the whole set of ARCH coefficients. We see how the MMC technique also makes our tests operational in case we want to test for a null hypothesis different from the one of the whole coefficient ARCH vector equal to a value. The MMC technique was carried out using the simulated annealing algorithm. The results show that even for sample sizes of 50, the point optimal test although it decreases the power for alternative hypothesis quite close to the null, still has good power, after optimising the p-value out of all possible values for the nuisance parameters.

7) In the seventh case, we consider a non-stationary case where we show the behaviour of our point optimal case against the asymptotic alternatives [in the context of [Jensen and Rahbek, 004a], [Jensen and Rahbek, 004b]]. They have proved that in the setting of a non-stationary region, the QMLE is still asymptotically normal. However, there are still no results in the literature that show what happens in small samples. We provide results in this context both when the fourth moment of the innovation process exists and when it does not exist. We show that the behaviour of our point optimal test in this setting has very good power regardless if the errors have very fat tails or not [at the same time of controlling for the size]. The extension of the Engle test and the other test we propose have very low power in some of the alternatives. This possibility of very good behaviour of point optimal tests outside the stationarity region was already shown in [Dufour and King, 1991].

So, from the simulation results, we advice the practitioners that it is possible to use as a rule of thumb in some cases the value of the point alternative equal to the middle of the parameter space; although when it is possible, it is always better to find a consistent estimate [when possible] or to split the sample size [as we did in the previous section], and to use a small fraction of the sample (at the beginning) to estimate a plausible value for the alternative [as it is suggested in this paper]. We also show the behaviour of the point optimal test under non-normal-t(3) errors, and the findings indicate good power in these cases [allowing also for control of the size and then, improving on [Hall and Yao, 2003]].

5.6 Application of an ARCH model for US inflation

In this section we re-visit the analysis of the Implicit price deflator for GNP done by [Engle and Kraft,] and also reported by [Bollerslev, 1986]. The series is also analysed in [Greene, 2000] [page 809]. The data corresponds to quaterly observations on the implicit price deflator for GNP from 1948.II to 1983.IV. We have obtained the data from the U.S. Department of Commerce: Bureau of Economic Analysis.

For this data, [Engle and Kraft,] selected an AR(4)-ARCH(8) model such as:

$$\widehat{\pi}_t = 0.138 + 0.423\pi_{t-1} + 0.222\pi_{t-2} + 0.377\pi_{t-3} - 0.175\pi_{t-4}$$

$$(0.059)\ (0.081) \qquad (0.108) \qquad (0.078) \qquad (0.104)$$

$$\widehat{h}_t = 0.058 + 0.808 \sum_{i=1}^{8} \left(\frac{9-i}{36} \right) \varepsilon_{t-i}^2 \tag{9}$$

$$(0.033) \qquad\qquad (0.265)$$

where:

$$\pi_t = 100 \ln \frac{P_t}{P_{t-1}}$$

Asymptotic standard errors are given in brackets. Acording to the previous study, the values in that model decline linearly from 0.179 to 0.022. The normalised residuals from this model show no evidence of autocorrelation, nor do their squares.

We first carried out the analysis of the series for the same time period 1948.II to 1983.IV using an AR(4) model in the mean equation and using asymptotic [Newey and West, 1987] HAC standard errors. The results were:

$$\widehat{\pi}_t = 0.182 + 0.595\pi_{t-1} + 0.147\pi_{t-2} + 0.144\pi_{t-3} - 0.075\pi_{t-4}$$

$$(0.090)\ (0.106) \qquad (0.110) \qquad (0.131) \qquad (0.110)$$

The same that happens in the study of [Engle and Kraft,], when we test for ARCH effects at lags 1, 4 and 8, we reject the null hypothesis both using the asymptotic LM test of [Engle, 1982] and when we apply our point optimal test using the MSSMMCAO test [in all the cases giving very small p-values of 0.000]. To use our point optimal test in this framework, we have applied the MMC technique where we have kept the AR coefficients in the mean equation as nuisance parameters, in order to obtain provably valid results. We have used the point to optimise looking for it as in our split sample technique. Then we proceed as the previous study to fit an unrestricted ARCH(8) model where we have used asymptotic [Bollerslev and Wooldridge, 1992] robust standard errors, and we obtain:

$$\widehat{\pi}_t = 0.136 + 0.544\pi_{t-1} + 0.171\pi_{t-2} + 0.158\pi_{t-3} - 0.011\pi_{t-4}$$

$$(0.040)\ (0.084)\quad (0.094)\quad (0.076)\quad (0.081)$$

$$\widehat{h}_t = 0.064 + 0.166\varepsilon_{t-1}^2 + 0.109\varepsilon_{t-2}^2 - 0.021\varepsilon_{t-3}^2 + 0.120\varepsilon_{t-4}^2 + 0.033\varepsilon_{t-5}^2$$

$$(0.015)\quad (0.206)\quad (0.109)\quad (0.095)\quad (0.111)\quad (0.104)$$

$$+0.189\varepsilon_{t-6}^2 + 0.051\varepsilon_{t-7}^2 - 0.050\varepsilon_{t-8}^2$$

$$(0.114)\quad (0.123)\quad (0.048)$$

As [Greene, 2000] says, the linear restriction of the linear lag model given in (9) on the unrestricted ARCH(8) model appears not to be statistically significant. We indeed tested for the existence of the linear restriction in (9) to see if it holds in our model, and both using an asymptotic test and our point optimal test we reject the linear restriction with a p-value of 0.000 in both cases. However, one interesting puzzle is why although with the Engle-test and with our point optimal test we reject the null of no-ARCH effects of order 8, when we fit the previous model we obtain that all the coefficients in the ARCH(8) are not individually statistically significant according to the asymptotic results. We proceed then to apply our point optimal test using the MSSMMCAO technique to test if the ARCH coefficients in the previous model where individually statistically significant, and we found in the 8 cases a p-value around 0.000 rejecting the null hypothesis that each of the coefficients are individually equal to zero. In this case, it is proved that asymptotic results give pretty bad inference for the individual statistically significance for this example.

6 Conclusion

In this paper we have provided a general framework for exact optimal testing. We propose a new adaptive methodology that allows to search the point alternative by combining [Maximised] Monte Carlo tests and splitting the sample: the SSMCAO technique and its minimized version (MSSMCAO) when we do not have a consistent estimate available. We have specialised our results to find the best procedures for practitioners to test for any value of GARCH coefficients. We have developed tests, including point optimal tests and another tests where we provide the asymptotic distribution, and we have provided evidence of their performance both in the case of normal errors, very fat tails [to compare with [Hall and Yao, 2003]], and/or in a non-stationary region [to cover as well the framework of [Jensen and Rahbek, 004a], [Jensen and Rahbek, 004b]]. We have shown mainly that, while the SSMCAO tests can have a very good performance in all these cases in terms of power and size, asymptotic theory can provide very poor results; both regardless if we are in the situation of heavy tails or not, and in the stationary or in the nonstationary case. We have also shown that our tests can be made operational not only in the case where pivotality is guaranteed, but also when we test for the null of a sub-vector of the whole coefficient vector in the GARCH structure or when an ARMA processes are introduced in the mean equation without loosing the exactness. We have also provided evidence of how our split sample proposal is very useful when the researcher does not have any a priori knowledge of where the point alternative is. Finally, when we cannot find a consistent estimate, we extend our SSMCAO test to the minimised version (MSSMCAO) where the empirical p-value is minimised in the first stage. We show that under the case of moments non-existing [such as in [Hall and Yao, 2003]] this is the best alternative (allowing for a full control of the size)]. Finally, we have also reported the usefulness of our methodology by applying our procedure to the US inflation.

In summary, a general rule to advice to practitioners is that in some circumnstances and for some null hypothesis is good enough to use our point optimal test setting the point to optimise equal to the middle value of the parameter space [following the recommendation of [Elliott et al., 1996]]. However, sometimes, as we have shown in our simulations, this is not a good strategy, and the best recommendation, is to use our SSMCAO to allow the sample to select the point (or MSSMCAO if moments do not exist). Our simulations show that our (M)SSMCAO test has good properties in finite samples regardless if the errors are gaussian or not [including if the errors have very fat-tails] and/or if the process is in the stationarity region or not.

Appendix

A Appendix 1

Here we provide a proof for Theorems 2.2 and 2.5. Theorem 2.5 is a generalisation of Theorem 2.2, so our proof will be referred explicitly to the general case.

Let's suppose a general process:

$$\frac{\varepsilon_t\left(\Theta\right)}{E\left(\varepsilon_t^2/J_{t-1}:\Phi\right)^{1/2}}$$

Under H_0: $\Theta=\overline{\Theta}, \Phi=\overline{\Phi}$,

$$\frac{\varepsilon_t\left(\overline{\Theta}\right)}{E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)^{1/2}}\equiv\varepsilon_t\left(\overline{\Theta},\overline{\Phi}\right)$$

Under H_1: $\Theta=\Theta^1, \Phi=\Phi^1$

$$\frac{\varepsilon_t\left(\Theta^1\right)}{E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)^{1/2}}=\varepsilon_t\left(\Theta^1,\Phi^1\right)$$

Let

$$\varepsilon_t\left(\overline{\Theta},\overline{\Phi}\right)=\left(\varepsilon_1\left(\overline{\Theta},\overline{\Phi}\right),...,\varepsilon_T\left(\overline{\Theta},\overline{\Phi}\right)\right)$$

Then

$$l_T\left(\varepsilon_t\left(\overline{\Theta},\overline{\Phi}\right),\overline{\Theta},\overline{\Phi}\right)=-\frac{T}{2}\ln 2\pi-\frac{1}{2}\sum_{t=1}^{T}\ln\left(E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)\right)-\frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2\left(\overline{\Theta}\right)}{E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)}$$

$$l_T\left(\varepsilon_t\left(\Theta^1,\Phi^1\right),\Theta^1,\Phi^1\right)=-\frac{T}{2}\ln 2\pi-\frac{1}{2}\sum_{t=1}^{T}\ln\left(E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)\right)-\frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2\left(\Theta^1\right)}{E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)}$$

So:

$$\begin{aligned}&LR\left(\overline{\Theta},\overline{\Phi},\Theta^1,\Phi^1\right)=-2\left[l_T\left(\varepsilon_t\left(\Theta^1,\Phi^1\right),\Theta^1,\Phi^1\right)-l_T\left(\varepsilon_t\left(\overline{\Theta},\overline{\Phi}\right),\overline{\Theta},\overline{\Phi}\right)\right]=\\&-2\left[-\tfrac{1}{2}\textstyle\sum_{t=1}^{T}\ln\left(\frac{E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)}{E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)}\right)-\tfrac{1}{2}\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2\left(\Theta^1\right)}{E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)}-\frac{\varepsilon_t^2\left(\overline{\Theta}\right)}{E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)}\right)\right]=\\&\textstyle\sum_{t=1}^{T}\ln\left(\frac{E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)}{E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)}\right)+\sum_{t=1}^{T}\left(\frac{\varepsilon_t^2\left(\Theta^1\right)}{E\left(\varepsilon_t^2/J_{t-1}:\Phi^1\right)}-\frac{\varepsilon_t^2\left(\overline{\Theta}\right)}{E\left(\varepsilon_t^2/J_{t-1}:\overline{\Phi}\right)}\right).\end{aligned}$$

B Appendix 2

Let's suppose a simple GARCH(p,q) process:

$$\varepsilon_t = \left(\theta_0 + \sum_{i=1}^{p} \theta_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-j}\right)^{1/2} v_t$$

Under H_0: $\theta_i = \overline{\theta}_i, \gamma_j = \overline{\gamma}_j$

$$v_t = \frac{\varepsilon_t}{\left(\theta_0 + \sum_{i=1}^{p} \overline{\theta}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \overline{\gamma}_j h_{t-j}\right)^{1/2}} \equiv \varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right)$$

Under H_1: $\theta_i = \theta_i^1, \gamma_j = \gamma_j^1$

$$v_t = \frac{\varepsilon_t}{\left(\theta_0 + \sum_{i=1}^{p} \theta_i^1 \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j^1 h_{t-j}\right)^{1/2}}$$

and

$$\varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right) = \frac{\left(\theta_0 + \sum_{i=1}^{p} \theta_i^1 \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j^1 h_{t-j}\right)^{1/2}}{\left(\theta_0 + \sum_{i=1}^{p} \overline{\theta}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \overline{\gamma}_j h_{t-j}\right)^{1/2}} v_t = \overline{h}_t\left(\overline{\theta}_i, \overline{\gamma}_j, \theta_i^1, \gamma_j^1, h_0, \varepsilon_0\right)^{1/2} v_t$$

where:

$$\overline{h}_t\left(\overline{\theta}_i, \overline{\gamma}_j, \theta_i^1, \gamma_j^1, h_0, \varepsilon_0\right) = \frac{\left(\theta_0 + \sum_{i=1}^{p} \theta_i^1 \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j^1 h_{t-j}\right)}{\left(\theta_0 + \sum_{i=1}^{p} \overline{\theta}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \overline{\gamma}_j h_{t-j}\right)} \equiv \overline{h}_t$$

Let

$$\varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right) = \left(\varepsilon_1\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right), ..., \varepsilon_T\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right)\right)$$

Then

$$l_T\left(\varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right), \overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right)^2$$

$$l_T\left(\varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right), \theta_i^1, \gamma_j^1, h_0, \varepsilon_0\right) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T} \ln\left(\overline{h}_t\right) - \frac{1}{2}\sum_{t=1}^{T} \frac{\varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right)^2}{\overline{h}_t}$$

So:

$$\begin{aligned} &LR\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0, \theta_i^1, \gamma_j^1\right) \\ &= -2\left[l_t\left(v\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right), \theta_i^1, \gamma_j^1, h_0, \varepsilon_0\right) - l_t\left(v\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right), \overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right)\right] \\ &= \sum_{t=1}^{T}\left[\ln\left(\overline{h}_t\right) + \left(\frac{1}{\overline{h}_t} - 1\right)\varepsilon_t\left(\overline{\theta}_i, \overline{\gamma}_j, h_0, \varepsilon_0\right)^2\right] \end{aligned}$$

Appendix 3

Null 1: $H_0 : \theta_1 = 0.5; \gamma_1 = 0.4$

		PO (0.6,0.7)		PO (0,0)		PO (0.3,0.2)		PO (0.2,0.3)		PO (0.05,0.9)		PO (0.1,0.9)		PO (1,0)		PO (SSMMCAO technique)				PO(full sample)	
		MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	m–p-t3	m–p-n	QML-t3	QML-n	QML-t3	QML-n
Size	T=200	0.045	0.047	0.053	0.054	0.043	0.056	0.051	0.059	0.049	0.052	0.045	0.047	0.059	0.061	0.052	0.051	0.053	0.057	0.078	0.077
	T=50	0.042	0.053	0.052	0.048	0.045	0.057	0.049	0.056	0.047	0.057	0.051	0.058	0.058	0.049	0.057	0.043	0.061	0.052	0.089	0.084
H(0,0)	T=200	0.004	0.005	1.000	1.000	0.403	0.605	0.398	0.609	0.403	0.506	0.421	0.403	0.702	0.813	0.995	1.000	0.996	1.000	0.999	1.000
	T=50	0.010	0.001	0.960	1.000	0.201	0.430	0.204	0.426	0.201	0.322	0.213	0.312	0.639	0.705	0.934	0.998	0.920	1.000	0.945	1.000
H(0.6,0.7)	T=200	0.740	1.000	0.002	0.005	0.251	0.300	0.301	0.351	0.311	0.312	0.413	0.515	0.502	0.701	0.645	0.988	0.500	1.000	0.603	1.000
	T=50	0.470	0.999	0.003	0.003	0.190	0.251	0.202	0.253	0.191	0.225	0.222	0.321	0.251	0.434	0.440	0.770	0.401	0.998	0.438	1.000
H(0.3,0.2)	T=200	0.261	0.351	0.196	0.250	0.842	1.000	0.791	1.000	0.613	0.903	0.504	0.809	0.401	0.512	0.815	0.913	0.813	1.000	0.822	1.000
	T=50	0.191	0.256	0.096	0.110	0.713	0.990	0.632	0.704	0.504	0.605	0.473	0.525	0.215	0.322	0.691	0.807	0.688	0.972	0.703	0.981
H(0.2,0.3)	T=200	0.312	0.512	0.502	0.601	0.521	0.902	0.871	1.000	0.571	0.913	0.501	0.908	0.413	0.891	0.851	1.000	0.832	1.000	0.853	1.000
	T=50	0.095	0.321	0.312	0.413	0.413	0.513	0.639	0.813	0.312	0.612	0.413	0.504	0.315	0.413	0.612	0.791	0.503	0.801	0.615	0.811
H(0.05,0.9)	T=200	0.231	0.421	0.109	0.195	0.301	0.504	0.304	0.507	0.649	1.000	0.491	0.998	0.195	0.271	0.631	1.000	0.591	1.000	0.629	1.000
	T=50	0.203	0.312	0.091	0.113	0.271	0.371	0.251	0.431	0.512	0.721	0.471	0.691	0.154	0.197	0.479	0.669	0.391	0.691	0.451	0.722
H(0.1,0.9)	T=200	0.371	0.541	0.151	0.242	0.451	0.491	0.401	0.522	0.613	0.898	0.750	1.000	0.512	0.613	0.729	1.000	0.691	1.000	0.731	1.000
	T=50	0.213	0.322	0.091	0.129	0.371	0.403	0.293	0.381	0.421	0.713	0.621	0.813	0.401	0.524	0.599	0.776	0.502	0.803	0.551	0.809
H(1,0)	T=200	0.291	0.691	0.191	0.512	0.613	0.771	0.412	0.721	0.211	0.321	0.391	0.401	0.871	1.000	0.853	1.000	0.813	1.000	0.849	1.000
	T=50	0.197	0.473	0.163	0.321	0.421	0.612	0.321	0.428	0.191	0.199	0.311	0.391	0.792	0.893	0.769	0.861	0.712	0.879	0.771	0.894

H(a, b) means the values $H_A : \theta_1 = a; \gamma_1 = b$

.

Null 2: $H_0 : \theta_1 = 0.5; \gamma_1 = 0.6$

		PO (0.5,0.4)		PO (0,0)		PO (0.3,0.2)		PO (0.2,0.3)		PO (0.05,0.9)		PO (0.1,0.9)		PO (1,0)		PO (SSMMCAO technique)				PO(full sample)	
		MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	MMC-t3	MMC-n	m–p-t3	m–p-n	QML-t3	QML-n	QML-t3	QML-n
Size	T=200	0.054	0.057	0.054	0.052	0.045	0.049	0.042	0.054	0.046	0.055	0.054	0.047	0.052	0.054	0.050	0.051	0.051	0.056	0.071	0.079
	T=50	0.052	0.056	0.053	0.055	0.048	0.056	0.053	0.049	0.045	0.047	0.055	0.051	0.053	0.055	0.049	0.053	0.056	0.052	0.081	0.084
H(0,0)	T=200	0.970	1.000	1.000	1.000	0.902	1.000	0.892	1.000	0.704	1.000	0.850	1.000	0.860	0.998	0.997	1.000	0.890	1.000	0.991	1.000
	T=50	0.870	0.800	0.988	1.000	0.852	0.709	0.853	0.702	0.691	0.987	0.801	0.911	0.871	0.913	0.981	1.000	0.882	1.000	0.901	1.000
H(0.5,0.4)	T=200	0.659	1.000	0.502	0.900	0.501	0.893	0.498	0.754	0.503	0.622	0.612	0.703	0.514	0.901	0.640	0.979	0.520	1.000	0.620	1.000
	T=50	0.400	0.890	0.349	0.615	0.356	0.412	0.341	0.391	0.298	0.430	0.301	0.451	0.352	0.546	0.390	0.772	0.350	0.870	0.360	0.889
H(0.3,0.2)	T=200	0.430	0.711	0.530	0.632	0.903	1.000	0.851	1.000	0.841	1.000	0.849	1.000	0.605	0.809	0.882	1.000	0.870	1.000	0.880	1.000
	T=50	0.212	0.504	0.323	0.405	0.702	0.903	0.582	0.792	0.571	0.703	0.582	0.699	0.359	0.603	0.653	0.805	0.590	0.867	0.620	0.873
H(0.2,0.3)	T=200	0.429	0.693	0.603	0.704	0.691	1.000	0.907	1.000	0.676	1.000	0.695	1.000	0.573	0.971	0.891	1.000	0.704	1.000	0.871	1.000
	T=50	0.193	0.579	0.524	0.609	0.591	0.803	0.751	0.912	0.599	0.792	0.603	0.818	0.403	0.605	0.704	0.821	0.509	0.870	0.659	0.892
H(0.05,0.9)	T=200	0.401	0.543	0.106	0.205	0.431	0.641	0.454	0.649	0.741	1.000	0.698	1.000	0.209	0.306	0.713	1.000	0.693	1.000	0.701	1.000
	T=50	0.329	0.402	0.095	0.159	0.322	0.491	0.325	0.502	0.609	0.805	0.572	0.751	0.163	0.205	0.591	0.759	0.582	0.791	0.592	0.804
H(0.1,0.9)	T=200	0.403	0.602	0.192	0.251	0.502	0.651	0.506	0.651	0.742	1.000	0.803	1.000	0.611	0.831	0.761	1.000	0.702	1.000	0.759	1.000
	T=50	0.319	0.451	0.103	0.163	0.403	0.502	0.393	0.592	0.681	0.859	0.704	0.902	0.503	0.630	0.692	0.849	0.672	0.861	0.697	0.892
H(1,0)	T=200	0.531	0.703	0.264	0.612	0.705	0.884	0.653	0.873	0.391	0.403	0.403	0.474	0.951	1.000	0.943	1.000	0.904	1.000	0.932	1.000
	T=50	0.252	0.509	0.231	0.503	0.621	0.703	0.594	0.692	0.352	0.391	0.391	0.401	0.870	0.973	0.851	0.943	0.803	0.961	0.824	0.968

n: denotes normal distribution

t3: denotes t-distribution with 3 degress of freedom

D Appendix 4

The construction of (what we denote) an Engle test for the null of any possible set of values for the ARCH coefficients, implies the following. Let's suppose an ARCH(p) model:

$$y_t = x_t{'}\beta + \sum_{i=1}^{q} \phi_i y_{t-i} + \varepsilon_t$$

$$h_t = E\left(\varepsilon_t^2 / I_{t-1}\right) = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + ... + \theta_p \varepsilon_{t-p}^2$$

The log-likelihood:

$$L \propto -\frac{1}{2} \sum_{t=1}^{T} \log h_t - \frac{1}{2} \sum_{t=1}^{T} \frac{\varepsilon_t^2}{h_t}$$

So the gradient (grad) and the hessian (hes) are:

$$grad = \left[\frac{\varepsilon_t^2}{2h_t^2} - \frac{1}{2h_t}\right] Z$$

$$hes = \left[\frac{1}{2h_t^2} - \frac{\varepsilon_t^2}{h_t^3}\right] Z Z'$$

where Z is the vector $Z = \left(1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ..., \varepsilon_{t-p}^2\right)'$.

Under any null hypothesis, we have to get the R^2 coming from regressing a column of ones on the derivatives of the log likelihood function computed at the restricted estimator.

Following Engle (1982), we can denote f^0 to be the first part of the gradient evaluated at the restricted estimator under the null:

$$f^0 = \left[\frac{\varepsilon_t^2}{2h_t^2} - \frac{1}{2h_t}\right]_0$$

So, the Engle test comes from TR^2, where R^2 comes from regressing f^0 on Z.

In the special case where the null is no-ARCH effects, since (following Engle (1982)) adding a constant and multiplying by a scalar won't change the R^2 of a regression, this will be the R^2 coming from regressing ε_t^2 on an intercept and p lagged values of ε_t^2.

In the general case of any other null except the no-ARCH effects one, h_t evaluated at a restricted estimator is going to contain lagged ε_t^2 and the values we are testing. The test in this case can be interpreted as regressing under the null:

$$\left[\frac{\varepsilon_t^2}{2h_t^2} - \frac{1}{2h_t}\right]_0 = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 + ... + \gamma_p \varepsilon_{t-p}^2 + v_t \qquad \text{(A.3.1)}$$

Or:

$$\varepsilon_t^2 = 2h_t^2 \left[\gamma_0 + \gamma_1\varepsilon_{t-1}^2 + \gamma_2\varepsilon_{t-2}^2 + ... + \gamma_p\varepsilon_{t-p}^2\right] + h_t + v_t$$

So running (A.3.1), can be equivalent to run as well a regression of ε_t^2 under the null, on $2h_t^2$ times the usual regression terms in the Engle test plus the h_t (everything evaluated under the null). Implicitly, because h_t has ε_{t-i}^2 inside, in order to construct the confidence sets, this may have some equivalent representation to running ε_t^2 under the null on a constant, as many lags of ε_t as the order of the ARCH(p) we are testing, raised to six, to four and to two (so, terms of the form $\varepsilon_{t-i}^6, \varepsilon_{t-i}^4, \varepsilon_{t-i}^2, \forall i = 1, ..., p)$, and cross products of as many lags of ε_t as the order of the ARCH(p) we are testing raised to four and two $($so terms of the form $\varepsilon_{t-i}^4\varepsilon_{t-j}^2, \forall i \neq j)$, plus cross products of as many lags of ε_t as order of the ARCH(p) we are testing raised to two (so terms of the form $\varepsilon_{t-i}^2\varepsilon_{t-j}^2, \forall i \neq j)$. But with the "problem" of multiplying in each case the previous terms by the values we are testing under the null.

E Appendix 5

Let us consider the expression:

$$v_t\left(\overline{\theta}_1\right) = \frac{\varepsilon_t}{\left(\theta_0 + \overline{\theta}_1\varepsilon_{t-1}^2 + ... + \overline{\theta}_p\varepsilon_{t-p}^2\right)^{1/2}}$$

where, for the case of the ARCH(1):

$$y_t = \varepsilon_t$$

$$h_t = E\left(\varepsilon_t^2/I_{t-1}\right) = \theta_0 + \theta_1\varepsilon_{t-1}^2$$

$$E\left(\ v_t^2\left(\overline{\theta}_1\right)\right) = \frac{E\left(y_t/I_{t-1}\right)}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)} = \frac{\left(\theta_0 + \theta_1 y_{t-1}^2\right)}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)}$$

So:

$$E\left(\ v_t^2\left(\overline{\theta}_1\right) - \frac{1}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)}/I_{t-1}\right) = \frac{\theta_1 y_{t-1}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)}$$

$$E\left(\ v_t^2\left(\overline{\theta}_1\right) - 1\right) = \left(\theta_1 - \overline{\theta}_1\right)\frac{y_{t-1}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)}$$

So finally in this case we will regress:

$$\frac{y_t^2}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)} - 1 = \theta_1^* \frac{y_{t-1}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2\right)} + w_t$$

In the general case of an ARCH(p) model:

$$E\left(\ v_t^2\left(\overline{\theta}_1, ..., \overline{\theta}_p\right) - 1\right) = \left(\theta_1 - \overline{\theta}_1\right)\frac{y_{t-1}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2 + ... + \overline{\theta}_p y_{t-p}^2\right)} + ...$$

$$+ \left(\theta_p - \overline{\theta}_p\right)\frac{y_{t-p}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2 + ... + \overline{\theta}_p y_{t-p}^2\right)}$$

So the testing would imply to regress:

$$\frac{y_t^2}{\left(1+\overline{\theta}_1 y_{t-1}^2 + ... + \overline{\theta}_p y_{t-p}^2\right)} - 1 =$$

$$\theta_1^* \frac{y_{t-1}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2 + ... + \overline{\theta}_p y_{t-p}^2\right)} + ... + \theta_p^* \frac{y_{t-p}^2}{\left(1+\overline{\theta}_1 y_{t-1}^2 + ... + \overline{\theta}_p y_{t-p}^2\right)} + w_t$$

F Appendix 6

Null 1: $H_0 : \theta_1 = 0.81; \theta_2 = 0.19$

	Engle-normal		Engle-t(5)		Engle-t(3)		T2-normal		T2-t(5)	
	AS	MC	AS	MC	AS	MC	AS	MC	AS	MC
Size: simulating under the null										
T=500	0.002	0.050	0.001	0.048		0.051	0.018	0.047	0.020	0.050
T=200	0.002	0.050	0.001	0.047		0.047	0.017	0.050	0.021	0.049
T=50	0.002	0.047	0.001	0.050		0.048	0.014	0.050	0.018	0.051
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$										
T=500	0.004	0.780	0.001	0.565		0.700	0.963	1.000	0.438	0.845
T=200	0.003	0.560	0.001	0.420		0.555	0.690	1.000	0.192	0.460
T=50	0.006	0.145	0.001	0.130		0.095	0.101	0.315	0.050	0.145
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$										
T=500	0.001	0.395	0.001	0.145		0.115	0.824	1.000	0.554	1.000
T=200	0.003	0.280	0.001	0.140		0.075	0.410	1.000	0.298	0.955
T=50	0.003	0.095	0.002	0.065		0.055	0.095	0.245	0.093	0.255
Power: simulating under $H_A : \theta_1 = 0.01; \theta_2 = 0.98$										
T=500	0.011	1.000	0.001	0.370		0.220	1.000	1.000	0.997	1.000
T=200	0.003	1.000	0.001	0.095		0.210	0.999	1.000	0.967	1.000
T=50	0.009	0.170	0.001	0.085		0.095	0.582	1.000	0.485	1.000

	T2-t(3)		PO (0.01,0.98)		PO (0.49,0.49)		PO (0.16,0.16)	
	AS	MC	MC-t(3)	MC-normal	MC-t(3)	MC-normal	MC-t(3)	MC-normal
Size: simulating under the null								
T=500		0.050	0.048	0.051	0.050	0.051	0.050	0.051
T=200		0.049	0.050	0.049	0.049	0.049	0.054	0.049
T=50		0.050	0.047	0.050	0.050	0.049	0.050	0.048
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$								
T=500		0.330	1.000	1.000	0.060	1.000	1.000	1.000
T=200		0.230	1.000	1.000	0.050	0.755	1.000	1.000
T=50		0.095	0.500	0.580	0.040	0.090	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$								
T=500		0.795	0.535	1.000	0.325	1.000	0.600	1.000
T=200		0.625	0.335	1.000	0.210	1.000	0.270	0.495
T=50		0.175	0.190	0.320	0.140	0.485	0.110	0.150
Power: simulating under $H_A : \theta_1 = 0.01; \theta_2 = 0.98$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		1.000	1.000	1.000	1.000	1.000	0.485	0.705

AS: Asymptotic. MC: Monte Carlo. PO(θ_1^1,θ_2^1): Point optimal. T2: Test in Theorem 9.

Null 2: $H_0 : \theta_1 = 0.98; \theta_2 = 0.01$

	Engle-normal		Engle-t(5)		Engle-t(3)		T2-normal		T2-t(5)	
	AS	MC	AS	MC	AS	MC	AS	MC	AS	MC
Size: simulating under the null										
T=500	0.003	0.054	0.001	0.047		0.049	0.023	0.049	0.022	0.049
T=200	0.001	0.054	0.001	0.049		0.055	0.026	0.052	0.024	0.053
T=50	0.001	0.053	0.001	0.054		0.055	0.025	0.055	0.027	0.050
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$										
T=500	0.456	1.000	0.089	1.000		1.000	0.999	1.000	0.885	1.000
T=200	0.171	1.000	0.021	1.000		1.000	0.909	1.000	0.625	1.000
T=50	0.012	0.395	0.003	0.465		0.410	0.261	0.560	0.232	0.540
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$										
T=500	0.029	0.255	0.001	0.110		0.085	0.999	1.000	0.988	1.000
T=200	0.008	0.195	0.001	0.145		0.050	0.905	1.000	0.861	1.000
T=50	0.002	0.110	0.001	0.100		0.070	0.395	1.000	0.431	1.000
Power: simulating under $H_A : \theta_1 = 0.01; \theta_2 = 0.98$										
T=500	0.003	0.135	0.001	0.135		0.150	1.000	1.000	0.999	1.000
T=200	0.005	0.125	0.001	0.085		0.095	0.995	1.000	0.977	1.000
T=50	0.002	0.130	0.001	0.110		0.145	0.662	1.000	0.613	1.000

	T2-t(3)		PO (0.01,0.98)		PO (0.49,0.49)		PO (0.16,0.16)	
	AS	MC	MC-t(3)	MC-normal	MC-t(3)	MC-normal	MC-t(3)	MC-normal
Size: simulating under the null								
T=500		0.051	0.055	0.049	0.051	0.050	0.051	0.051
T=200		0.051	0.047	0.050	0.049	0.049	0.048	0.049
T=50		0.049	0.051	0.050	0.051	0.048	0.051	0.052
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$								
T=500		1.000	1.000	1.000	0.860	1.000	1.000	1.000
T=200		1.000	1.000	1.000	0.440	1.000	1.000	1.000
T=50		0.545	0.765	1.000	0.250	0.785	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		0.980	0.930	1.000	1.000	1.000	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.01; \theta_2 = 0.98$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		1.000	1.000	1.000	1.000	1.000	1.000	1.000

Null 3: $H_0 : \theta_1 = 0.5; \theta_2 = 0.4$

	Engle-normal		Engle-t(5)		Engle-t(3)		T2-normal		T2-t(5)	
	AS	MC	AS	MC	AS	MC	AS	MC	AS	MC
Size: simulating under the null										
T=500	0.078	0.050	0.001	0.051		0.052	0.016	0.044	0.019	0.047
T=200	0.035	0.047	0.001	0.055		0.050	0.016	0.051	0.017	0.053
T=50	0.013	0.048	0.002	0.052		0.051	0.012	0.049	0.012	0.047
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$										
T=500	0.607	1.000	0.137	1.000		1.000	0.784	1.000	0.015	0.045
T=200	0.365	0.895	0.049	1.000		1.000	0.351	1.000	0.010	0.035
T=50	0.070	0.310	0.001	0.780		0.305	0.032	0.135	0.008	0.040
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.4$										
T=500	0.393	0.480	0.023	0.600		0.570	0.762	1.000	0.214	0.400
T=200	0.177	0.410	0.012	0.545		0.450	0.293	1.000	0.101	0.320
T=50	0.041	0.165	0.007	0.145		0.175	0.040	0.135	0.028	0.105
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.64$										
T=500	0.094	0.040	0.001	0.110		0.085	0.981	1.000	0.804	1.000
T=200	0.041	0.085	0.001	0.105		0.115	0.656	1.000	0.654	1.000
T=50	0.019	0.045	0.001	0.080		0.095	0.118	0.400	0.093	0.435
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.16$										
T=500	0.237	0.215	0.004	0.240		0.210	0.364	0.515	0.071	0.255
T=200	0.037	0.135	0.003	0.225		0.185	0.094	0.305	0.040	0.150
T=50	0.024	0.209	0.004	0.095		0.100	0.015	0.060	0.011	0.040
Power: simulating under $H_A : \theta_1 = 0.9; \theta_2 = 0.09$										
T=500	0.005	0.015	0.001	0.015		0.025	0.995	1.000	0.824	1.000
T=200	0.004	0.010	0.001	0.005		0.035	0.711	1.000	0.449	1.000
T=50	0.005	0.020	0.001	0.025		0.060	0.092	0.315	0.061	0.220
Power: simulating under $H_A : \theta_1 = 0.09; \theta_2 = 0.9$										
T=500	0.007	0.020	0.001	0.010		0.015	1.000	1.000	0.984	1.000
T=200	0.007	0.005	0.001	0.010		0.030	0.964	1.000	0.843	1.000
T=50	0.008	0.040	0.002	0.030		0.060	0.324	1.000	0.241	0.860

	T2-t(3)		PO (0.01,0.98)		PO (0.49,0.49)		PO (0.16,0.16)	
Size: simulating under the null								
	AS	MC	MC-t(3)	MC-normal	MC-t(3)	MC-normal	MC-t(3)	MC-normal
T=500		0.048	0.051	0.050	0.051	0.051	0.050	0.049
T=200		0.048	0.051	0.048	0.050	0.051	0.051	0.048
T=50		0.053	0.046	0.051	0.047	0.051	0.048	0.052
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$								
T=500		0.015	1.000	1.000	0.015	0.010	1.000	1.000
T=200		0.010	1.000	1.000	0.010	0.006	1.000	1.000
T=50		0.020	0.165	0.070	0.010	0.005	1.000	0.770
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.4$								
T=500		0.255	1.000	1.000	0.030	0.025	1.000	1.000
T=200		0.210	1.000	1.000	0.035	0.015	0.710	1.000
T=50		0.095	0.250	0.275	0.030	0.015	0.375	0.310
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.64$								
T=500		1.000	1.000	1.000	0.175	1.000	0.250	0.315
T=200		1.000	1.000	1.000	0.150	1.000	0.160	0.175
T=50		0.260	0.370	0.815	0.120	0.380	0.100	0.085
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.16$								
T=500		0.160	0.380	0.070	0.015	0.010	1.000	1.000
T=200		0.095	0.145	0.030	0.015	0.005	0.690	0.555
T=50		0.040	0.035	0.015	0.020	0.001	0.185	0.130
Power: simulating under $H_A : \theta_1 = 0.9; \theta_2 = 0.09$								
T=500		1.000	0.010	0.005	0.025	0.015	0.010	0.025
T=200		0.970	0.010	0.002	0.025	0.011	0.015	0.022
T=50		0.150	0.010	0.001	0.025	0.010	0.030	0.020
Power: simulating under $H_A : \theta_1 = 0.09; \theta_2 = 0.9$								
T=500		1.000	1.000	1.000	0.960	1.000	0.010	0.040
T=200		1.000	1.000	1.000	0.585	1.000	0.015	0.036
T=50		0.790	0.860	1.000	0.290	1.000	0.040	0.035

Null 4: $H_0 : \theta_1 = 0.16; \theta_2 = 0.25$

	Engle-normal		Engle-t(5)		Engle-t(3)		T2-normal		T2-t(5)	
	AS	MC	AS	MC	AS	MC	AS	MC	AS	MC
Size: simulating under the null										
T=500	0.077	0.049	0.016	0.051		0.045	0.017	0.049	0.022	0.047
T=200	0.002	0.048	0.001	0.049		0.056	0.018	0.047	0.028	0.051
T=50	0.003	0.046	0.001	0.053		0.045	0.014	0.041	0.019	0.052
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.16$										
T=500	0.019	0.020	0.001	0.005		0.005	0.919	1.000	0.685	1.000
T=200	0.001	0.005	0.001	0.005		0.020	0.526	1.000	0.399	1.000
T=50	0.001	0.020	0.001	0.025		0.015	0.109	0.380	0.102	0.355
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$										
T=500	0.003	0.020	0.001	0.001		0.001	0.863	1.000	0.602	1.000
T=200	0.001	0.015	0.001	0.001		0.001	0.466	1.000	0.313	0.910
T=50	0.001	0.030	0.001	0.001		0.001	0.090	0.220	0.076	0.195
Power: simulating under $H_A : \theta_1 = 0.9; \theta_2 = 0.09$										
T=500	0.001	0.001	0.001	0.001		0.001	1.000	1.000	0.985	1.000
T=200	0.001	0.001	0.001	0.001		0.001	0.970	1.000	0.846	1.000
T=50	0.001	0.005	0.001	0.020		0.010	0.361	1.000	0.261	0.690
Power: simulating under $H_A : \theta_1 = 0.09; \theta_2 = 0.9$										
T=500	0.005	0.015	0.001	0.001		0.001	0.999	1.000	0.965	1.000
T=200	0.001	0.001	0.001	0.005		0.001	0.912	1.000	0.729	1.000
T=50	0.001	0.005	0.001	0.020		0.010	0.254	0.995	0.172	0.400

	T2-t(3)		PO (0.01,0.98)		PO (0.49,0.49)		PO (0.16,0.16)	
	AS	MC	MC-t(3)	MC-normal	MC-t(3)	MC-normal	MC-t(3)	MC-normal
Size: simulating under the null								
T=500		0.045	0.051	0.051	0.051	0.051	0.051	0.049
T=200		0.051	0.046	0.050	0.051	0.050	0.048	0.049
T=50		0.045	0.051	0.050	0.048	0.049	0.053	0.051
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.16$								
T=500		0.975	0.035	0.030	0.375	1.000	0.010	0.060
T=200		0.630	0.035	0.025	0.320	1.000	0.015	0.055
T=50		0.180	0.045	0.005	0.250	0.950	0.025	0.050
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$								
T=500		0.495	0.045	0.525	1.000	1.000	0.003	0.015
T=200		0.345	0.065	0.480	0.915	1.000	0.001	0.011
T=50		0.185	0.080	0.335	0.425	1.000	0.004	0.010
Power: simulating under $H_A : \theta_1 = 0.9; \theta_2 = 0.09$								
T=500		1.000	0.030	0.350	1.000	1.000	0.010	0.035
T=200		1.000	0.045	0.325	1.000	1.000	0.005	0.030
T=50		0.465	0.070	0.320	0.870	1.000	0.005	0.025
Power: simulating under $H_A : \theta_1 = 0.09; \theta_2 = 0.9$								
T=500		1.000	1.000	1.000	1.000	1.000	0.002	0.010
T=200		1.000	1.000	1.000	1.000	1.000	0.001	0.005
T=50		0.370	0.645	1.000	0.460	1.000	0.001	0.002

Null 5: $H_0 : \theta_1 = 0; \theta_2 = 0$

	Engle-normal		Engle-t(5)		Engle-t(3)		T2-normal		T2-t(5)	
	AS	MC	AS	MC	AS	MC	AS	MC	AS	MC
Size: simulating under the null										
T=500	0.045	0.052	0.045	0.048		0.047	0.023	0.049	0.016	0.046
T=200	0.043	0.051	0.041	0.053		0.050	0.023	0.048	0.028	0.048
T=50	0.039	0.050	0.033	0.048		0.050	0.019	0.047	0.025	0.050
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$										
T=500	0.957	1.000	0.984	1.000		1.000	0.923	1.000	0.873	1.000
T=200	0.654	1.000	0.815	1.000		1.000	0.620	1.000	0.650	1.000
T=50	0.191	0.375	0.288	0.665		0.960	0.170	0.575	0.241	0.700
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$										
T=500	1.000	1.000	0.995	1.000		1.000	0.999	1.000	0.900	1.000
T=200	0.991	1.000	0.976	1.000		1.000	0.961	1.000	0.834	1.000
T=50	0.528	1.000	0.594	1.000		1.000	0.541	1.000	0.488	1.000
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.16$										
T=500	1.000	1.000	0.995	1.000		1.000	0.998	1.000	0.980	1.000
T=200	0.698	1.000	0.962	1.000		1.000	0.926	1.000	0.886	1.000
T=50	0.433	1.000	0.512	1.000		1.000	0.387	1.000	0.420	1.000
Power: simulating under $H_A : \theta_1 = 0.09; \theta_2 = 0.9$										
T=500	0.999	1.000	0.976	1.000		1.000	0.999	1.000	0.976	1.000
T=200	0.984	1.000	0.949	1.000		1.000	0.509	1.000	0.888	1.000
T=50	0.568	1.000	0.543	1.000		1.000	0.468	1.000	0.443	1.000
Power: simulating under $H_A : \theta_1 = 0.9; \theta_2 = 0.09$										
T=500	1.000	1.000	0.990	1.000		1.000	0.998	1.000	0.981	1.000
T=200	0.990	1.000	0.961	1.000		1.000	0.957	1.000	0.906	1.000
T=50	0.558	1.000	0.580	1.000		1.000	0.490	1.000	0.481	1.000

	T2-t(3)		PO (0.01,0.98)		PO (0.49,0.49)		PO (0.16,0.16)	
	AS	MC	MC-t(3)	MC-normal	MC-t(3)	MC-normal	MC-t(3)	MC-normal
Size: simulating under the null								
T=500		0.051	0.051	0.050	0.051	0.051	0.052	0.049
T=200		0.046	0.051	0.053	0.053	0.050	0.050	0.048
T=50		0.055	0.046	0.048	0.051	0.049	0.051	0.051
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		0.790	1.000	1.000	1.000	1.000	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		1.000	1.000	1.000	1.000	1.000	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.16$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		1.000	1.000	1.000	1.000	1.000	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.09; \theta_2 = 0.9$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		1.000	1.000	1.000	1.000	1.000	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.9; \theta_2 = 0.09$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		1.000	1.000	1.000	1.000	1.000	1.000	1.000

Null 6: $H_0 : \theta_1 = 0$
Point optimal test set to (0.49,0.49) using MMC with t(3) and normal.

	Size		Power $\theta_1 = 0.9$		Power $\theta_1 = 0.5$		Power $\theta_1 = 0.16$	
	MMC-t(3)	MMC-nor	MMC-t(3)	MMC-nor	MMC-t(3)	MMC-nor	MMC-t(3)	MMC-nor
T=50	0.059	0.060	0.910	0.940	0.820	0.840	0.340	0.400

Null 7: $H_0 : \theta_1 = 0.81; \theta_2 = 0.49$

	Engle-normal		Engle-t(5)		Engle-t(3)		T2-normal		T2-t(5)	
	AS	MC	AS	MC	AS	MC	AS	MC	AS	MC
Size: simulating under the null										
T=500	0.003	0.049	0.002	0.051		0.051	0.020	0.050	0.017	0.052
T=200	0.002	0.049	0.001	0.052		0.052	0.016	0.049	0.015	0.053
T=50	0.002	0.051	0.001	0.051		0.048	0.014	0.052	0.013	0.049
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$										
T=500	0.007	1.000	0.002	1.000		1.000	0.956	1.000	0.505	0.950
T=200	0.007	0.720	0.002	0.695		1.000	0.723	1.000	0.137	0.535
T=50	0.009	0.180	0.002	0.185		0.225	0.111	0.390	0.018	0.065
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$										
T=500	0.003	0.505	0.001	0.180		0.655	0.363	0.565	0.018	0.070
T=200	0.001	0.385	0.001	0.150		0.400	0.108	0.355	0.016	0.050
T=50	0.005	0.120	0.001	0.080		0.115	0.022	0.110	0.012	0.040
Power: simulating under $H_A : \theta_1 = 0.01; \theta_2 = 0.98$										
T=500	0.002	0.600	0.001	0.430		0.975	1.000	1.000	0.989	1.000
T=200	0.001	0.360	0.001	0.360		0.695	0.992	1.000	0.910	1.000
T=50	0.004	0.090	0.001	0.100		0.125	0.417	1.000	0.311	1.000
Power: simulating under $H_A : \theta_1 = 0.49; \theta_2 = 0.81$										
T=500	0.001	0.080	0.002	0.045		0.040	0.927	1.000	0.724	1.000
T=200	0.001	0.090	0.001	0.075		0.050	0.513	1.000	0.348	1.000
T=50	0.001	0.060	0.001	0.040		0.085	0.099	0.455	0.077	0.265
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 1$										
T=500	0.003	0.160	0.002	0.175		0.275	1.000	1.000	0.984	1.000
T=200	0.003	0.125	0.001	0.185		0.240	0.970	1.000	0.832	1.000
T=50	0.003	0.085	0.001	0.095		0.100	0.323	1.000	0.231	0.840

	T2-t(3)		PO (0.01,0.98)		PO (0.49,0.49)		PO (0.16,0.16)	
	AS	MC	MC-t(3)	MC-normal	MC-t(3)	MC-normal	MC-t(3)	MC-normal
Size: simulating under the null								
T=500		0.050	0.052	0.049	0.051	0.053	0.049	0.047
T=200		0.051	0.053	0.052	0.049	0.052	0.049	0.052
T=50		0.049	0.048	0.051	0.048	0.050	0.052	0.049
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 0.16$								
T=500		0.050	1.000	1.000	1.000	1.000	1.000	1.000
T=200		0.075	1.000	1.000	1.000	1.000	1.000	1.000
T=50		0.035	1.000	1.000	1.000	1.000	1.000	1.000
Power: simulating under $H_A : \theta_1 = 0.5; \theta_2 = 0.4$								
T=500		0.055	1.000	1.000	1.000	1.000	1.000	1.000
T=200		0.075	1.000	1.000	1.000	1.000	1.000	1.000
T=50		0.040	0.280	0.220	0.370	0.285	0.390	0.355
Power: simulating under $H_A : \theta_1 = 0.01; \theta_2 = 0.98$								
T=500		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=200		1.000	1.000	1.000	1.000	1.000	1.000	1.000
T=50		0.920	1.000	1.000	1.000	1.000	0.556	0.390
Power: simulating under $H_A : \theta_1 = 0.49; \theta_2 = 0.81$								
T=500		1.000	0.290	0.860	0.540	0.500	0.055	0.050
T=200		0.725	0.165	0.315	0.255	0.515	0.055	0.065
T=50		0.300	0.135	0.220	0.115	0.190	0.055	0.060
Power: simulating under $H_A : \theta_1 = 0.16; \theta_2 = 1$								
T=500		1.000	1.000	1.000	1.000	1.000	0.825	0.800
T=200		1.000	1.000	1.000	1.000	1.000	0.425	0.415
T=50		0.625	0.905	1.000	0.705	1.000	0.175	0.155

References

[Andersen, 1994] Andersen, T. G. (1994). Stochastic autoregressive volatility: A framework for volatility modeling. *Mathematical Finance*, 4:75–102.

[Andrews and Ploberger, 1994] Andrews, D. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327.

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

[Cajigas and Urga, 2004] Cajigas, J. P. and Urga, G. (2004). Non-normal elliptical distributions and dynamic conditional correlation models. *Working paper*.

[Dufour and King, 1991] Dufour, J. and King, M. (1991). Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or non-stationary AR(1) errors. *Journal of Econometrics*, 47:115–143.

[Dufour, 2004] Dufour, J. M. (2004). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. *Journal of Econometrics, forthcoming*.

[Dufour et al., 2004] Dufour, J.-M., Khalaf, L., Bernard, J.-T., and Genest, I. (2004). Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. *Journal of Econometrics*, 122(2):317–347.

[Dufour and Kiviet, 1996] Dufour, J.-M. and Kiviet, J. F. (1996). Exact tests for structural change in first-order dynamic models. *Journal of Econometrics*, 70:39–68.

[Dufour and Kiviet, 1998] Dufour, J.-M. and Kiviet, J. F. (1998). Exact inference methods for first-order autoregressive distributed lag models. *Econometrica*, 66:79–104.

[Elliott et al., 1996] Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836.

[Engle and Kraft,] Engle, R. and Kraft, D. Multiperiod forecast error variances of inflation estimated from ARCH models. *In A. Zellner, ed., Applied Time Series Analysis of Economic Data. Washington D.C.: Bureau of the Census*.

[Engle, 1982] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1008.

[Fernandez and Steel, 1998] Fernandez, C. and Steel, M. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93:359–371.

[Garcia and Veredas, 2004] Garcia, R., R. E. and Veredas, D. (2004). Estimation of stable distributions by indirect inference. *Mimeo*, pages –.

[Goffe and Rogers, 1994] Goffe, W.L., F. G. and Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60:65–99.

[Greene, 2000] Greene, W. H. (2000). *Econometric Analysis*. Prentice-Hall, fourth edition.

[Hall and Yao, 2003] Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, 71:285–317.

[Hansen, 1994] Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35:705–730.

[Jensen and Rahbek, 004a] Jensen, S. T. and Rahbek, A. (2004a). Asymptotic normality of the QMLE of ARCH in the nonstationary case. *Econometrica*, 72(2):641–646.

[Jensen and Rahbek, 004b] Jensen, S. T. and Rahbek, A. (2004b). Asymptotic inference for nonstationary GARCH. *Econometric Theory*, 20(6):1203–1226.

[King, 1980] King, M. L. (1980). Robust tests for spherical symmetry and their application to least squares regression. *The Annals of Statistics*, 8(6):1265–1271.

[King and Hillier, 1985] King, M. L. and Hillier, G. H. (1985). Locally best invariant tests of the error covariance matrix of the linear regression model. *Journal of the Royal Statistical Society, Series B*, 47(1):98–102.

[Lee and King, 1993] Lee, J. H. and King, M. L. (1993). A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics*, 11:17–27. Correction 12 (1994), 139.

[Lee, 1991] Lee, J. H. H. (1991). A Lagrange multiplier test for GARCH models. *Economics Letters*, 37:265–271.

[Lumsdaine, 1995] Lumsdaine, R. (1995). Finite-sample properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, 13:1–10.

[Meddahi and Renault, 2004] Meddahi, N. and Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics*, 119:355–379.

[Mueller and Elliott, 2003] Mueller, U. and Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica*, 71:1269–1286.

[Newey and West, 1987] Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.