Testing for Reduction to Random Walk in Autoregressive Conditional Heteroskedasticity Models

CLAUDIA KLÜPPELBERG[1], ROSS A. MALLER[2],
MARK VAN DE VYVER[3], DERICK WEE[4]

[1] *Center for Mathematical Sciences, Munich University of Technology, D-80290 München, Germany*
cklu@ma.tum.de, www.ma.tum.de/stat/
[2] *Department of Accounting and Finance, University of Western Australia, Nedlands Western Australia 6907*
rmaller@ecel.uwa.edu.au, www.ecel.uwa.edu.au/rmaller/Welcome.html/
[3] *Department of Accounting and Finance, University of Western Australia, Nedlands Western Australia 6907*
mvdv@bigfoot.com, www.bigfoot.com/ mvdv
[4] *Department of Accounting and Finance, University of Western Australia, Nedlands Western Australia 6907*
derickwee@hotmail.com

Summary: The AR–ARCH and AR–GARCH models, which allow for conditional heteroskedasticity and autoregression, reduce to random walk or white noise for some values of the parameters. We consider generalised versions of the AR-ARCH(1) and AR–GARCH(1,1) models, and, under mild assumptions, calculate the asymptotic distributions of pseudo-likelihood ratio statistics for testing hypotheses that reflect these reductions. These hypotheses are of two kinds: the conditional volatility parameters may take their boundary values of zero, or the autoregressive component may take the form of a unit root process or not in fact be present. The limiting distributions of the resulting test statistics can be expressed in terms of functionals of Brownian motion related to the Dickey-Fuller statistic, together with independent chi-square components. The finite sample performances of the test statistics are assessed by simulations, and percentiles are tabulated. The results have applications in the analysis of financial time series and random coefficient models.

Keywords: AR–ARCH and AR–GARCH models, conditional heteroskedasticity, autoregression, unit root, Dickey–Fuller test, pseudo-likelihood ratio test.

JEL Subject Classifications: C12, C13, C22

1. INTRODUCTION

This paper derives the large-sample distribution of pseudo-likelihood ratio tests for an autoregressive (AR) model which exhibits generalised autoregressive conditional heteroskedasticity (GARCH) type behaviour. The null hypotheses considered are:

(i) the AR structure has a unit root; or

(ii) the model has no autoregressive structure; and

(iii) there is no heteroskedasticity.

Simulation experiments show the asymptotic theory to be fairly accurate once the sample size exceeds 100 or so.

The processes we consider are of special interest for asset price modelling, or, more generally, for econometric data modelling. They take the form

$$X_i = \alpha X_{i-1} + \varepsilon_i \sigma_i, \ i = 2, 3, \ldots, n, \tag{1.1}$$

where the residuals ε_i are i.i.d. with expectation 0, variance 1 and a finite fourth moment (we assume only this throughout), and the variance function satisfies

$$\sigma_i^2 = \beta + \lambda g_{i-1} + \delta \sigma_{i-1}^2, \ i = 2, 3, \ldots, n \tag{1.2}$$

(with $\sigma_1 = \beta$). In (1.1) and (1.2), α, β, λ and δ are parameters with $\beta > 0$, $\lambda \geq 0$ and $\delta \geq 0$, and g_{i-1} is a non-negative, measurable function of $\{X_{i-1}, X_{i-2}, \ldots, X_1\}$ alone. We will analyse in detail the versions:

$$g_{i-1} = \varepsilon_{i-1}^2 \sigma_{i-1}^2 = (X_{i-1} - \alpha X_{i-2})^2 \tag{1.3}$$

and

$$g_{i-1} = X_{i-1}^2, \tag{1.4}$$

and a further version,

$$g_{i-1} = \varepsilon_{i-1}^2, \tag{1.5}$$

will be included in the simulations. (1.1) – (1.3) specify the well known AR–GARCH(1,1) model (if $\delta > 0$; otherwise an AR–ARCH(1) model). Thus the model can be thought of as a generalised autoregression, possibly with unit root, and with GARCH effect included. In the unit root case the specification (1.3) defines a GARCH(1,1) process in the differenced series $X_i - X_{i-1}$. On the other hand, (1.4) or some other formulation of the conditional variance may be indicated by the data. Nicholls and Pagan (1985, p. 444) suggest this model in connection with varying coefficient regressions. We give a unified approach to testing hypotheses concerning (1.1) and (1.2), confining ourselves mainly to (1.3) and (1.4) for illustration throughout.

The cases when $\alpha = 1$ or $\alpha = 0$ in (1.1) are of special interest, and we may wish to test these hypotheses as a potential simplification of the data. A further useful simplification occurs when $\lambda = \delta = 0$ in (1.2), because then (1.1) reduces to random walk for X_i in the case $\alpha = 1$, or to i.i.d. increments X_i (thus, to a random walk for $X_1 + \cdots + X_i$) in the case $\alpha = 0$. There is a strong motivation for such simplifications in the pricing of claims contingent on the value of the asset, in particular, because option pricing models

based on GARCH models are challanging to implement whereas for a Gaussian random walk the well understood Black-Scholes or binomial tree based pricing methodologies apply. Consequently it is worthwhile to consider the following program. Fit the model (1.1)–(1.2), with (1.3), (1.4), (1.5), or some other specification of g_{i-1}, with a view to testing if the data is unit root ($\alpha = 1$), or has no autoregressive component ($\alpha = 0$); or if there is in fact no conditional heteroskedasticity ($\lambda = \delta = 0$), so the data is sufficiently well modelled by random walk, by i.i.d. white noise, or as an AR(1) model.

In this paper we work out the asymptotic distributions, as sample size n tends to infinity, of pseudo-likelihood ratio statistics for testing the above hypotheses, after constructing estimating equations from a pseudo-likelihood under which the ε_i are assumed i.i.d. with a standard normal distribution. The large-sample distributions are given as combinations of integrals of a standard Brownian motion (SBM) (or, in one case, as a mixture of chi-square and truncated chi-square random variables), but are quite easy to use when the residuals have no skewness and kurtosis 3, for example.

In the case of ARCH, part of our program has been considered by Demos and Sentana (1998), who conjectured a limiting result when g_{i-1} satisfies (1.3) and $\alpha = 0$ (so the AR component is not present), and when the ε_i are standard normal. (They also considered higher order ARCH and GARCH models but we restrict ourselves to the first order cases since the main ideas are apparent for them.) They draw attention to the extra power available by taking into account the "boundary hypothesis" nature of the tests, as our analysis does.

A systematic investigation of a wide class of boundary hypothesis tests is in Andrews (1999, 2001), and various of his earlier papers, but his results do not include those given here, as we discuss in more detail in Section 3 below. Our models are more general in allowing the AR component, and also in allowing a general function g_{i-1} in (1.2). With emphasis on the AR rather than the ARCH/GARCH component, our results can be regarded as testing for ARCH/GARCH type effects in an autoregressive process, possibly with unit root. Seo (1999) derives some large-sample results for t–ratios from this point of view, but does not consider the boundary hypotheses for reduction to random walk. Davis and Dunsmuir (1996) give an asymptotic analysis of the MA(1) process when it is in or near the unit root case, using related ideas (see also Davis and Mikosch (1998)).

For general background on ARCH and GARCH we refer to Engle (1982, 1995), Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle and Nelson (1994) and Gouriéroux (1997); see also Shephard (1996) for a good discussion of statistical aspects of the models.

The remainder of the paper is laid out as follows. Section 2 sets out the assumptions made and outlines the methods we will use. Section 3 states the main results in the form of Theorem 3.1. To see how well the asymptotics work in finite samples, we report some simulations and tabulate percentiles of the test statistics, in Section 4. Some discussion of the results and their applications is in Section 5. Proofs are in the appendices. We restrict ourselves throughout to an AR–ARCH type of alternative ($\delta = 0$ in (1.2)), since, as for the ARCH/GARCH case, this will suffice in general, as we show in Appendix C.

2. MODEL ASSUMPTIONS AND METHODOLOGY

In (1.1), the ε_i depend on the values of the unknown parameters, though this is suppressed

in the notation. When the parameters take their "true" values, denoted by a subscript 0, the ARCH version of the model will be written as

$$X_i = \alpha_0 X_{i-1} + e_i \sqrt{\beta_0 + \lambda_0 g_{i-1}},\ i = 2, 3, \ldots, n, \tag{2.1}$$

with $X_1 = e_1$. Here e_i is the value of ε_i when the parameters take their true values. The e_i are i.i.d. random variables (rvs) with expectation 0, variance 1 and finite third and fourth moments μ_3 and μ_4. Collect the parameters into a parameter vector $\theta = (\alpha, \beta, \lambda)$, with "true" value $\theta_0 = (\alpha_0, \beta_0, \lambda_0)$. θ and θ_0 lie in the parameter space $\Theta = \mathbb{R} \times (0, \infty) \times [0, \infty)$. The first hypothesis we consider is

$$H_0^{(1)} : \theta_0 \in \Omega^{(1)} := \{1\} \times (0, \infty) \times \{0\},\ \text{i.e., } \alpha_0 = 1,\ \beta_0 > 0,\ \lambda_0 = 0. \tag{2.2}$$

Under $H_0^{(1)}$, the process X_i is a mean zero random walk, whatever the choice of g_i. As the "maintained" hypothesis, we take

$$H_1 : \theta_0 \in \tau := \mathbb{R} \times (0, \infty) \times [0, \infty),\ \text{i.e., } \alpha_0 \in \mathbb{R},\ \beta_0 > 0,\ \lambda_0 \geq 0. \tag{2.3}$$

(2.2) is a boundary hypothesis relative to (2.3), as discussed in Vu and Zhou (1997), Demos and Sentana (1998) and Andrews (1999, 2001). However, the results and methods in the latter papers do not apply directly to our problem, as we discuss in detail later. We proceed by modifying the methods of Vu and Zhou (1997) (hereafter, "VZ"), whose general formulation and method of approach lends itself easily to the further generalisation required here. We will show that, with probability approaching 1 as $n \to \infty$ (WPA1), there exist maximisers $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n$ in $\Omega^{(1)}$ and τ of the log-pseudo-likelihood function $\mathcal{L}_n(\theta)$ (taken conditional on X_1) which are consistent for θ_0, and we derive the asymptotic distribution (as $n \to \infty$) of the "deviance" (minus twice log-pseudo-likelihood ratio) statistic

$$d_n^{(1)} := -2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\hat{\theta}_n)) \tag{2.4}$$

for testing $H_0^{(1)}$ versus H_1. This deals with the unit root (random walk) aspect. For the white noise version, we consider instead the null hypothesis

$$H_0^{(2)} : \theta_0 \in \Omega^{(2)} := \{0\} \times (0, \infty) \times \{0\},\ \text{i.e., } \alpha_0 = 0,\ \beta_0 > 0,\ \lambda_0 = 0, \tag{2.5}$$

under which the X_i become the *increments* of a random walk (i.e., i.i.d. rvs), to be tested against H_1. A similar analysis can be made in this case, resulting in maximisers $\hat{\theta}_n^{(2)}$ and $\hat{\theta}_n$ of $\mathcal{L}_n(\theta)$ under $H_0^{(2)}$ and H_1. Again, under the assumptions made so far, we can derive the asymptotic distribution of

$$d_n^{(2)} = -2(\mathcal{L}_n(\hat{\theta}_n^{(2)}) - \mathcal{L}_n(\hat{\theta}_n)), \tag{2.6}$$

the deviance statistic for testing $H_0^{(2)}$ versus H_1. These results are stated next.

3. MAIN THEORETICAL RESULTS

Our results can be summarised as:

Theorem 3.1. *Suppose the X_i satisfy (2.1) for i.i.d. e_i with expectation 0, variance 1 and finite third and fourth moments, and that g_{i-1} satisfies (1.3).*

(i) Suppose further that $H_0^{(1)}$ as specified in (2.2) holds. Then $d_n^{(1)}$ as defined in (2.4) satisfies

$$d_n^{(1)} \xrightarrow{D} \frac{(W^2(1)-1)^2}{4\int_0^1 W^2(y)dy} + Z^2 I(Z \geq 0), \text{ as } n \to \infty, \tag{3.1}$$

where $W(\cdot)$ is an SBM on $[0,1]$, $Z = \sqrt{\mu_4 - 1}N/\sqrt{2}$, with N a standard normal random variable independent of $W(\cdot)$, and $I(Z \geq 0)$ is 1 if $Z \geq 0$ and 0 otherwise.

(ii) Suppose instead that $H_0^{(2)}$ as specified in (2.5) holds. Then the deviance statistic $d_n^{(2)}$ for testing $H_0^{(2)}$ versus H_1 satisfies

$$d_n^{(2)} \xrightarrow{D} N_1^2 + Z^2 I(Z \geq 0), \text{ as } n \to \infty, \tag{3.2}$$

where

$$Z = \frac{\mu_3^2}{\sqrt{2(\mu_4 - 1)}} N_1 + \sqrt{\frac{(\mu_4-1)^2 - \mu_3^4}{2(\mu_4 - 1)}} N_2,$$

and N_1 and N_2 are independent standard normal rvs.

(iii) Alternatively, suppose g_{i-1} satisfies (1.4). Then, under $H_0^{(2)}$, (3.2) remains true as stated, while, under $H_0^{(1)}$,

$$d_n^{(1)} \xrightarrow{D} \frac{(W_1^2(1)-1)^2}{4\int_0^1 W_1^2(y)dy} + Z^2 I(Z \geq 0), \text{ as } n \to \infty, \tag{3.3}$$

where

$$Z = \frac{1}{\sqrt{2(J_4 - J_2^2)}} \left\{ -J_2 \left(\mu_3 W_1(1) + \sqrt{\mu_4 - 1 - \mu_3^2}\, W_2(1) \right) \right.$$
$$\left. + \mu_3 \int_0^1 W_1^2(y) dW_1(y) + \sqrt{\mu_4 - 1 - \mu_3^2} \int_0^1 W_1^2(y) dW_2(y) \right\}, \tag{3.4}$$

with $(W_1(\cdot), W_2(\cdot))$ an SBM in two dimensions, and $J_k := \int_0^1 W_1^k(y)dy$, $k \geq 1$.

When $\mu_3 = 0$ and $\mu_4 = 3$, as is the case for example when the e_i are $N(0,1)$, and g_{i-1} satisfies (1.3) or (1.4), Z in (3.1), (3.2) and (3.4) is standard normal. In this case

$$d_n^{(2)} \xrightarrow{D} N_1^2 + N_2^2 I(N_2 \geq 0), \text{ as } n \to \infty, \tag{3.5}$$

where N_1 and N_2 are independent standard normal rvs.

Remarks. (i) In (3.1) and elsewhere, "$\xrightarrow{D}$" means convergence in distribution.

(ii) We show in Appendix C below that the results of Theorem 3.1 remain true as stated for the GARCH(1,1) model.

(iii) Andrews (1999, 2001) gives a very extensive and intensive analysis of the boundary hypothesis testing problem for a wide class of models. While our approach is similar in some respects to his, there are some crucial differences in the models we treat, and our results are quite different. In both cases (his Assumption 3, our (B.5) below) an important calculation is to establish the joint convergence of the normed first and second derivative matrices of the log-pseudo-likelihood to a limiting random vector. One of our main contributions, and difficulties, is to find this limit for the cases we consider. (A somewhat similar approach is used by Davis and Dunsmuir (1996) in the MA(1) model.) Andrews (2001) considers a GARCH(1,1) model and gives a result comparable to (3.5) (but without the chi-square component corresponding to α, the AR parameter). Andrews, like us, forms a pseudo-likelihood from i.i.d. $N(0,1)$ rvs. He does not include the unit root case (3.1) or (3.3), and this extension to the non-stationary case requires significantly different calculations. One point of distinction is that the limit of the normed second derivative matrix for an ordinary GARCH model is non-random, which allows the use of Andrews' Assumption 3*. But this is not general enough for (3.1) and (3.3), and we need to bring in extra theory – we use results concerning convergence of stochastic integrals of Kurtz and Protter (1991) – to handle it. Andrews' calculations for the ordinary GARCH model rely heavily on the stationarity assumption.

To keep the paper mostly self-contained we give a separate development which, furthermore, for our specific setup, avoids some of the restrictions Andrews and others in this area have to apply in general. A common assumption, for example, which seems essential to approaches such as that of Andrews, is that the parameters be restricted to a compact parameter space. Our analysis avoids this requirement, which in effect introduces further, unnecessary bounding parameters which may be quite artificial in some situations. We allow the extra generality of the g_{i-1} in (1.2). On the other hand Andrews' model is more general in some respects; he includes a stationary sequence of regressor variables in the mean equation of the model, and the innovations are required only to be stationary martingale differences, whereas ours are assumed i.i.d. In this regard, our assumptions are kept simple so as to give a clear development but our methods can be generalised to these situations. Andrews also allows for an extra GARCH-AR parameter (i.e., in the variance, rather than the mean, equation), π, which disappears under the null and is handled by methods of Davies (1977, 1987). He considers test statistics other than log-pseudo-likelihood ratio statistics (Wald and score tests) and again obtains quite general results. No doubt our methods could be applied similarly for these cases too.

4. SIMULATIONS OF AR–ARCH AND AR–GARCH MODELS

To assess the asymptotic results, we evaluated the finite sample percentiles of the deviance statistic for various sample sizes, for various models, by conducting a Monte Carlo study. Simulations were done with the software package Mathematica (Wolfram Research (1999)), using all the standard packages, with the exception of the routine used to maximise the log-pseudo-likelihood. For this we used the package Global Optimization (Loehl (1999)). A detailed description of our experiment is in Van De Vyver (2001). Briefly, for each of

the two series satisfying (2.1) with $e_i \sim N(0,1)$, and for each of 3 choices of g_{i-1}, namely, (1.3), (1.4), and (1.5), we simulated 1000 series of length n, where $n = 100, 500, 1000$, according to the two null hypotheses:

$H_0^{(1)}$: Random Walk; i.e., a partial sum of i.i.d. $N(0,1)$ rvs;

$H_0^{(2)}$: Gaussian white noise, i.e., a sequence of i.i.d. $N(0,1)$ rvs.

Thus, for $H_0^{(1)}$ the true parameter values were $\alpha_0 = 1$ and $\beta_0 = 1$, and for $H_0^{(2)}$ they were $\alpha_0 = 0$ and $\beta_0 = 1$ (with $\lambda_0 = \delta_0 = 0$ in both cases). For each series the likelihood was maximised under the null and alternate models and the value of the deviance statistic calculated. Percentiles of the finite-sample distributions of the d_n were then found.

Our experience confirmed others' observations, that maximisation of the log-pseudo-likelihood for ARCH/GARCH type models is very sensitive to the initial parameter estimates used. To minimise the effect of this, for each replication the parameter vector was estimated five times, each time from a randomly generated set of starting parameter estimates. The null and alternative models were simultaneously estimated for each starting value.

The only constraints imposed on the optimization routine were the non-negativity of β, λ and δ (where applicable), and a conditional variance stationarity constraint ($\lambda + \delta \leq 1$) in the case of the GARCH model. Additional initial values of σ and ε are required in calculating the variance equation. Two candidates for the initial variance estimate are the sample variance and the theoretical expected value, $\beta/(1-\lambda-\delta)$. For series of length $n = 100$ and 1000 we found no difference in the deviance statistics or parameter estimates obtained using either starting method (to at least the tenth decimal place).

From the 1000 replications, we estimated the 90th, 95th and 99th percentiles of the empirical distribution of the deviances from the appropriate order statistics, and these are given in Table 4.1, along with the asymptotic values calculated from (3.1), (3.2) or (3.3) (in the "∞" row) for comparison. We report in Table 4.1 results also for the choice of g_i in (1.5); the theory for this can be developed by methods similar to those in Section 6.

Consistent with other reports in the literature, we found the "standard" Mathematica Time Series package to have some difficulty in finding the maximum of the likelihood function, for some samples. This is apparently intrinsic. For details related to specific software issues see Fiorentini, Calozari and Panattoni (1996), McCullough and Renfro (1998), Jerrell (2000) and Brooks, Burke and Persand (2001). Zumbach (2000) shows analytically, via a change of co-ordinates, that the difficulties evident in practical applications of (1.3) are due to a nearly degenerate property of the model. However, our procedure always resulted in reasonable estimates. For users of other programs, we mention here that some of our "difficult" samples were handled straightforwardly by E-Views, or SAS, or both, while we often found samples which our procedure handled but were not dealt with correctly by one or another of these programs. Some forms of (1.1) and (1.2) were more difficult to estimate than others. This suggests that the implications of Zumbach's (2000) analysis may depend on the form of the (G)ARCH model assumed. The optimization algorithm we employ seems well suited to these types of models and may usefully be employed in other, similar, situations.

INSERT TABLE 4.1

Simulation Results

The percentiles given in Table 4.1 suffice for our purposes, but are for illustrative rather than for testing purposes; while the individual deviances were determined to three-figure accuracy by the program, the percentiles are not estimated to that order of precision. They were estimated directly from the empirical distribution function, but since the models under consideration are heavy-tailed, the high quantiles will be strongly influenced by extreme peaks of the time series. An alternative estimation procedure using extreme value theory and taking the size of the peaks into account would be preferable for general use. For a presentation and discussion of this kind of methodology for stationary time series we refer to Borkovec (2000), Embrechts et al. (1997, Chapter 6 and Section 8.1), and references therein. Fuller simulations and estimations for our setup are reported in Van De Vyver (2001). (See also Beletski, May and Szimayer (2001) and Gleisberg (2001).) The standard errors of the estimated percentiles in Table 4.1, calculated using normal approximations, vary from about 10% of the tabulated values for the 90th and 95th to about 20% of the values for the 99th percentiles.

Perhaps the most striking conclusion to be drawn from the table is that when the innovations are normally distributed, the form of g_i makes little difference to the percentiles of the distribution of the deviance statistic either in finite series of length greater than about 100, or asymptotically, at least when the underlying distribution of the series is given by the null: random walk or white noise. (The form of g_i may well be of importance in the non-null case and/or when the residuals do not have $\mu_3 = 0$ and $\mu_4 = 3$.)

We note a somewhat closer correspondence between finite sample and asymptotic results, for the series lengths considered, for ARCH than for GARCH models, as might be expected. The length of the series, n, plays the expected role, with the results for $n = 1000$ somewhat closer to the asymptotic values than for $n = 100$, in most cases.

Calibrating Time Series

Typical applications to financial data sets of our models would take the X_i to be $\ln(S_i)$, or $\ln(S_i/S_{i-1})$, where S_i is an asset price or exchange rate series. (1.1) – (1.2) with $\alpha = 0$ and $g_i = X_i^2 = (\ln(S_i/S_{i-1}))^2$ represents the GARCH(1,1) model intensively studied in the econometrics and stochastics literature. The tail behaviour of the stationary distribution, the extremal behaviour and the limiting behaviour of the AR-ARCH(1) model have been investigated in Borkovec and Klüppelberg (2001), Borkovec (2000,2001) and Embrechts et al. (1997, Section 8.4). Bollerslev, Engle and Nelson (1994) provide an interesting survey of some time series analyses that use economic and financial data sets, illustrating the types of data properties that the various models are intended to capture. Pagan (1996, Section 5) provides an informative discussion of considerations needed in establishing that the alternate as well as the null model is economically sensible. Our methods are able to accomodate these kinds of models; for example, the models of Duan (1995) and Ritchken and Trevor (1999).

Stationarity

Stationarity of the processes we consider is not an issue in the context of this paper as our working makes clear. But in practice, if the null hypothesis is rejected and a conditional

heteroskedasticity model of the AR-ARCH or AR-GARCH type is thought appropriate for the data, then the precision of parameter estimation and questions of stationarity or otherwise become important. Parameter estimation in these models has been considered for example in Ling and Li (1998) and their references, and the stationarity of the models has been resolved for the AR(1) process with ARCH(1) errors in Borkovec and Klüppelberg (2001).

Dependence on Parameters

The distributions in (3.1), (3.2) and (3.3) in general depend on the parameters μ_3 and μ_4, which must be estimated from data if the result is to be used in this generality. The first summand on the righthand side of (3.1) is the square of the Dickey-Fuller "t"–statistic for testing the unit root (random walk) hypothesis in an autoregressive process. This of course is the contribution of the AR in the AR–ARCH type setup. When the distribution of the underlying residuals is normal, or, more generally, $\mu_3 = 0$ and $\mu_4 = 3$, we get for the righthand side of (3.2) a χ^2_1 rv plus an independent 50–50 mixture of χ^2_1 and a point mass at 0, consistent with the conjecture of Demos and Sentana (1998).

ACKNOWLEDGEMENTS

We are grateful to David Hendry, Adrian Pagan and other participants in the Queensland Finance Conference (Brisbane, October 1999) for encouragement and helpful comments. Thanks also to Alexander Szimayer, Anke Gleisberg, two referees and the Editor for an especially careful reading of the paper and constructive suggestions. The first author gratefully acknowledges the hospitality of the Department of Accounting and Finance at the University of Western Australia during a visit to that Department.

REFERENCES

Aitchison, J., and Silvey, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. Annals of Mathematical Statistics 29 813–828.

Andrews, D.W.K. (1999). Estimation when a parameter is on the boundary. Econometrica 67, 1341–1383.

Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica, 69, 683–784.

Beletski, T., May, A. and Szimayer, A. (2001) Testing for conditional heteroscedasticity in financial time series. Preprint. Research Center caesar, Bonn.

Bollerslev, T., Chou, R. Y., Kroner, K. F. (1992). ARCH modeling in finance, Journal of Econometrics, 59, 5–59.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). ARCH models. In: R. F. Engle and D. McFadden, Eds, The Handbook of Econometrics, Vol 4, pp. 2959–3038, North Holland, Amsterdam.

Borkovec, M. (2000). Extremal behavior of the autoregressive process with ARCH(1) errors, Stochastic Processes and their Applications, 85, 189–207.

Borkovec, M. (2001). Asymptotic behavior of the sample autocovariance and autocorrelation function of the AR(1) process with ARCH(1) errors. Bernoulli, 7, 847-872.

Borkovec, M., and Klüppelberg, C. (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors, Annals of Applied Probability, 11, 1220-1241.

Brooks, C., S. P. Burke, and G. Persand (2001). Benchmarks and the accuracy of GARCH Model Estimation. International Journal of Forecasting, 17, 45-56.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254.

Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Davis, R. A. and Dunsmuir, W.T.M. (1996). Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. Econometric Theory, 12, 1–29.

Davis, R. A. and Mikosch, T. (1998). Gaussian likelihood based inference for non-invertible MA(1) processes with α noise. Stochastic Processes and Applications, 77, 99– 122.

Demos, A. and Sentana, E. (1998). Testing for GARCH effects: a one-sided approach. Journal of Econometrics 86, 97–127.

Duan, J. C. (1995). The GARCH option pricing model, Mathematical Finance, 5, 13-32.

Durrett, R. (1991). *Probability: Theory and Examples* Wadsworth, California.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R. F. (Ed.) (1995). *ARCH: Selected Readings.* Oxford Univ. Press.

Fiorentini, G., G. Calzolari, and L. Panattoni, (1996). Analytic Derivatives and the Computation of Garch Estimates, Journal of Applied Econometrics, 11, 399-417.

Gleisberg, A. (2001). ARCH und GARCH Modelle im Test gegen das Black-Scholes Modell. Diplomarbeit, Department of Mathematics, Technical University Darmstadt, Darmstadt, Germany.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer Series in Statistics. Springer-Verlag, New York.

Jerrell, M., 2000, Applications of public domain global optimization software to difficult econometric functions, Arizona: Northern Arizona University, 1-22.

Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. Annals of Probability 19, 1035–1070.

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

Ling, S. and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscdastic errors. Annals of Statistics, 26, 84–125.

Loehl Enterprises (1998) Global Optimization: Global Nonlinear Optimization using Mathematica, v2.4, Naperville, IL.

McCullough, B. D., and C. G. Renfro, (1998). Benchmarks and software standards: A case study of GARCH procedures. Journal of Economic and Social Measurement, 25, 59-71.

Nicholls, D. F. and Pagan, A. R. (1985). Varying coefficient regression. E. J. Hannan, P. R. Krishnaiah and M. M. Rao (Eds), Handbook of Statistics, 5, pp. 413–449. North Holland, Amsterdam.

Pagan, A. (1996). The econometrics of financial markets. Journal of Empirical Finance, 3, 15-102.

Ritchken, P. and Trevor, R., (1999). Pricing options under generalised GARCH and stochastic volatility processes. Journal of Finance, 54, 377–402.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. Journal of Econometrics 91, 113–144.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: O. E. Barndorff-Nielsen, D. R. Cox and D. V. Hinkley (Eds), *Time Series Models in Econometrics, Finance and other Fields*, pp. 1-67. Chapman Hall.

Silvey, S. D. (1959). The Lagrangian multiplier test. Annals of Mathematical Statistics 30 389–407.

Van De Vyver, M. (2001). Estimating Latent Variable Models of Returns to Equity Holders: Theory and Evidence. PhD thesis, Department of Accounting and Finance, University of Western Australia.

Vu, H. T. V. and Zhou, X. (1997). Generalisation of likelihood ratio tests under nonstandard conditions. Annals of Statistics 25, 897–916.

Wolfram Research Inc., (1999). Mathematica, Version 4.0.0.1, Champaign, IL.

Zumbach, G. O. (2000). The pitfalls in fitting GARCH(1,1) processes, in: C. L. Dunis (Ed.), Studies in Computational Finance, Kluwer, 179-200.

APPENDICES: PROOFS

As outlined in Section 2, we follow the VZ approach to prove Theorem 3.1. Under certain conditions on the model and a "maximising function" (which in our case will be the log-pseudo–likelihood function $\mathcal{L}_n(\theta)$), VZ show that, WPA1 as $n \to \infty$, there exist unique maximisers of $\mathcal{L}_n(\theta)$ in the null and alternate spaces which are consistent for θ_0, and derive the asymptotic distribution of the deviance statistic. The VZ "A" conditions, specifically: (A1), existence and continuity of first and second derivatives of $\mathcal{L}_n(\theta)$, taken to be one-sided at boundary points, and (A2), there is a closed cone $C_{\Omega^{(1)}}$ with vertex at θ_0 such that $C_{\Omega^{(1)}} \cap \mathcal{N} = \Omega^{(1)} \cap \mathcal{N}$ for a closed neighbourhood $\mathcal{N}$ in $\mathbb{R}^3$ of θ_0, and similarly for $\Omega^{(2)}$ and τ, are satisfied for our setup. However, some of the "B" conditions are not satisfied, so we must modify their methods. The approach given here can be used in other, similar, situations as well, so we keep the exposition general for as long as possible, only resorting to the specifics of the models when necessary.

Proofs are collected in the following four appendices. Appendix A lists the derivatives of the log-pseudo-likelihood, and some related quantities, Appendix B contains basic lemmas, Appendix C has the main parts of the proof of Theorem 3.1 and Appendix D verifies condition (B.6) of Lemma B.1 for $H_0^{(1)}$ and g_{i-1} satisfying (1.3).

APPENDIX A: DERIVATIVES.

The log-pseudo-likelihood of the observables $X_n, X_{n-1}, \ldots, X_2$, conditional on X_1, is

$$\mathcal{L}_n(\theta) = -\frac{1}{2}\sum_{i=2}^{n}\ln(\sigma_i^2) - \frac{1}{2}\sum_{i=2}^{n}\varepsilon_i^2 - \frac{1}{2}(n-1)\ln(2\pi) \qquad (A.1)$$

where

$$\sigma_i^2 = \beta + \lambda g_{i-1},\ i = 2, 3, \ldots, \qquad (A.2)$$

and, throughout, we use the relation $\varepsilon_i = (X_i - \alpha X_{i-1})/\sigma_i$. Straightforward differentiation, allowing for g_{i-1} to depend on α, β and/or λ in general, gives

$$\frac{\partial\varepsilon_i^2}{\partial\alpha} = -\frac{2X_{i-1}\varepsilon_i}{\sigma_i} - \frac{\varepsilon_i^2}{\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\alpha}, \quad \frac{\partial\varepsilon_i^2}{\partial\beta} = -\frac{\varepsilon_i^2}{\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\beta}, \quad \text{and} \quad \frac{\partial\varepsilon_i^2}{\partial\lambda} = -\frac{\varepsilon_i^2}{\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\lambda}. \qquad (A.3)$$

Let

$$S_n(\theta) = \frac{\partial\mathcal{L}_n(\theta)}{\partial\theta} \quad \text{and} \quad \mathcal{F}_n(\theta) = -\frac{\partial^2\mathcal{L}_n(\theta)}{\partial\theta^2}$$

be the first derivative vector and negative second derivative matrix of $\mathcal{L}_n(\theta)$. From (A.1) – (A.3) we get for the elements of $S_n(\theta)$:

$$\frac{\partial\mathcal{L}_n}{\partial\alpha} = \frac{1}{2}\sum_{i=2}^{n}\left(\frac{2X_{i-1}\varepsilon_i}{\sigma_i} + \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\alpha}\right), \quad \frac{\partial\mathcal{L}_n}{\partial\beta} = \frac{1}{2}\sum_{i=2}^{n}\frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\beta} \qquad (A.4a)$$

and

$$\frac{\partial\mathcal{L}_n}{\partial\lambda} = \frac{1}{2}\sum_{i=2}^{n}\frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial\sigma_i^2}{\partial\lambda}. \qquad (A.4b)$$

Noting that

$$\frac{\partial}{\partial\alpha}\left(\frac{\varepsilon_i^2-1}{\sigma_i^2}\right) = -\frac{2X_{i-1}\varepsilon_i}{\sigma_i^3} - \frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial\sigma_i^2}{\partial\alpha}, \qquad (A.5a)$$

$$\frac{\partial}{\partial\beta}\left(\frac{\varepsilon_i^2-1}{\sigma_i^2}\right) = -\frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial\sigma_i^2}{\partial\beta} \quad \text{and} \quad \frac{\partial}{\partial\lambda}\left(\frac{\varepsilon_i^2-1}{\sigma_i^2}\right) = -\frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial\sigma_i^2}{\partial\lambda}, \qquad (A.5b)$$

and with $\mathcal{F}_n(\theta) = [\mathcal{F}_n^{rs}(\theta)]$, $r, s, = 1, 2, 3$, we then get from (A.4)

$$\mathcal{F}_n^{11}(\theta) = -\frac{\partial^2\mathcal{L}_n}{\partial\alpha^2} = \frac{1}{2}\sum_{i=2}^{n}\left(\frac{2X_{i-1}^2}{\sigma_i^2} + \left(\frac{4X_{i-1}\varepsilon_i}{\sigma_i^3} + \frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial\sigma_i^2}{\partial\alpha}\right)\frac{\partial\sigma_i^2}{\partial\alpha} - \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial^2\sigma_i^2}{\partial\alpha^2}\right) \qquad (A.6a)$$

$$\mathcal{F}_n^{12}(\theta) = -\frac{\partial^2\mathcal{L}_n}{\partial\alpha\partial\beta} = \frac{1}{2}\sum_{i=2}^{n}\left(\left(\frac{2\varepsilon_i X_{i-1}}{\sigma_i^3} + \frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial\sigma_i^2}{\partial\alpha}\right)\frac{\partial\sigma_i^2}{\partial\beta} - \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial^2\sigma_i^2}{\partial\alpha\partial\beta}\right), \qquad (A.6b)$$

$$\mathcal{F}_n^{13}(\theta) = -\frac{\partial^2\mathcal{L}_n}{\partial\alpha\partial\lambda} = \frac{1}{2}\sum_{i=2}^{n}\left(\left(\frac{2\varepsilon_i X_{i-1}}{\sigma_i^3} + \frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial\sigma_i^2}{\partial\alpha}\right)\frac{\partial\sigma_i^2}{\partial\lambda} - \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial^2\sigma_i^2}{\partial\alpha\partial\lambda}\right), \qquad (A.6c)$$

$$\mathcal{F}_n^{22}(\theta) = -\frac{\partial^2 \mathcal{L}_n}{\partial \beta^2} = \frac{1}{2}\sum_{i=2}^{n}\left\{\frac{(2\varepsilon_i^2-1)}{\sigma_i^4}(\frac{\partial \sigma_i^2}{\partial \beta})^2 - \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial^2\sigma_i^2}{\partial\beta^2}\right\}, \qquad (A.6d)$$

$$\mathcal{F}_n^{23}(\theta) = -\frac{\partial^2 \mathcal{L}_n}{\partial \beta \partial \lambda} = \frac{1}{2}\sum_{i=2}^{n}\left\{\frac{(2\varepsilon_i^2-1)}{\sigma_i^4}\frac{\partial \sigma_i^2}{\partial \beta}\frac{\partial \sigma_i^2}{\partial \lambda} - \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial^2\sigma_i^2}{\partial\beta\partial\lambda}\right\}, \qquad (A.6e)$$

and

$$\mathcal{F}_n^{33}(\theta) = -\frac{\partial^2 \mathcal{L}_n}{\partial \lambda^2} = \frac{1}{2}\sum_{i=2}^{n}\left\{\frac{(2\varepsilon_i^2-1)}{\sigma_i^4}(\frac{\partial \sigma_i^2}{\partial \lambda})^2 - \frac{(\varepsilon_i^2-1)}{\sigma_i^2}\frac{\partial^2\sigma_i^2}{\partial\lambda^2}\right\}. \qquad (A.6f)$$

Also, from (A.2)

$$\frac{\partial \sigma_i^2}{\partial \alpha}\Big|_{\theta_0} = \lambda\frac{\partial g_{i-1}}{\partial \alpha}\Big|_{\theta_0} = 0 = \frac{\partial^2 \sigma_i^2}{\partial \alpha^2}\Big|_{\theta_0} = \frac{\partial^2 \sigma_i^2}{\partial \alpha \partial \beta}\Big|_{\theta_0}, \quad \frac{\partial^2 \sigma_i^2}{\partial \alpha \partial \lambda}\Big|_{\theta_0} = \frac{\partial g_{i-1,0}}{\partial \alpha}, \qquad (A.7a)$$

$$\frac{\partial \sigma_i^2}{\partial \beta}\Big|_{\theta_0} = (1+\lambda\frac{\partial g_{i-1}}{\partial \beta})\Big|_{\theta_0} = 1, \quad \frac{\partial \sigma_i^2}{\partial \lambda}\Big|_{\theta_0} = (g_{i-1}+\lambda\frac{\partial g_{i-1}}{\partial \lambda})\Big|_{\theta_0} = g_{i-1,0}, \qquad (A.7b)$$

$$\frac{\partial^2 \sigma_i^2}{\partial \beta^2}\Big|_{\theta_0} = 0, \quad \frac{\partial^2 \sigma_i^2}{\partial \beta \partial \lambda}\Big|_{\theta_0} = \frac{\partial g_{i-1,0}}{\partial \beta}, \quad \text{and} \quad \frac{\partial^2 \sigma_i^2}{\partial \lambda^2}\Big|_{\theta_0} = 2\frac{\partial g_{i-1,0}}{\partial \lambda}. \qquad (A.7c)$$

APPENDIX B: BASIC LEMMAS.

The next lemma provides the basis for Theorem 3.1. In what follows, $O_P(1)$ denotes a random variable bounded in probability (relatively compact) as $n \to \infty$, and $o_P(1)$ denotes a random variable which tends to 0 in probability as $n \to \infty$. $\lambda_{\min}(M)$ ($\lambda_{\max}(M)$) denotes the minimum (maximum) eigenvalue, and $M^{1/2}$ denotes the left Cholesky square root, of a positive definite matrix M. "T" denotes a transpose. For $n \geq 1$ and $A > 0$, $N_n(A)$ denotes the neighbourhood

$$N_n(A) = \left\{\theta : (\theta-\theta_0)^T G_n(\theta-\theta_0) \leq A^2\right\}. \qquad (B.1)$$

In Lemma B.1, we write Ω for $\Omega^{(1)}$ or $\Omega^{(2)}$ and d_n for $d_n^{(1)}$ or $d_n^{(2)}$.

Lemma B.1. *(i) Suppose there is a deterministic diagonal nonsingular matrix G_n with $\lambda_{\min}(G_n) \to \infty$ ($n \to \infty$) such that*

$$G_n^{-1/2} S_n(\theta_0) = O_P(1) \qquad (B.2)$$

and

$$\lim_{c\to 0+}\limsup_{A\to\infty}\limsup_{n\to\infty} P\{\inf_{\theta\in N_n(A)} \lambda_{\min}(G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2}) \leq c\} = 0. \qquad (B.3)$$

Then there are pseudo-maximum likelihood estimators (pseudo-MLEs) $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n$ which for each $A > 0$ uniquely maximise $\mathcal{L}_n(\theta)$ on $N_n(A) \cap \Omega$ and $N_n(A) \cap \tau$, respectively, on an event which has probability approaching 1 as $n \to \infty$ and $A \to \infty$. These estimators are consistent for θ_0.

(ii) Suppose in addition that

$$G_n^{-1/2}\mathcal{F}_n^{1/2}(\theta_0) = Y_n + o_P(1), \tag{B.4}$$

where Y_n is a lower triangular matrix with positive diagonal elements, having, apart from the diagonal element, zeroes in the first column; that

$$\left(G_n^{-1/2}S_n(\theta_0), Y_n\right) \xrightarrow{D} (S,\ Y), \tag{B.5}$$

as $n \to \infty$, for some a.s. finite random vector S and a.s. finite, nonsingular, matrix Y (in (B.5), joint convergence of all elements is required); and that, as $n \to \infty$, for each $A > 0$,

$$\sup_{\theta \in N_n(A)} \left\| G_n^{-1/2}\left(\mathcal{F}_n(\theta) - \mathcal{F}_n(\theta_0)\right) G_n^{-1/2} \right\| \xrightarrow{P} 0. \tag{B.6}$$

Then, as $n \to \infty$,

$$\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0) \xrightarrow{D} Y^{-1}S = Z, \tag{B.7}$$

for a finite random vector $Z = (Z_1, Z_2, Z_3)$, and

$$d_n = 2(\mathcal{L}_n(\hat{\theta}_n) - \mathcal{L}_n(\hat{\theta}_n^{(1)})) \xrightarrow{D} Z_1^2 + Z_3^2 I(Z_3 \geq 0). \tag{B.8}$$

Proof of Lemma B.1: As mentioned, we follow the proof of Theorem 2.1 of VZ, replacing (B3') of VZ (note that VZ do not use their (B1) in the proof of their Theorem 2.1), which does not hold here, with the weaker condition (B.6). The proof of Theorem 2.1 of VZ can then be carried through with minor modifications to find $\hat{\theta}_n^{(1)}$. Potentially, $\hat{\theta}_n^{(1)}$ depends on A but a standard argument in analysis allows us to choose $\hat{\theta}_n^{(1)}$ not depending on A, still with $\hat{\theta}_n^{(1)} \in N_n(A)$, on an event which has probability approaching 1 as $n \to \infty$ and $A \to \infty$. This is our pseudo-MLE of θ_0 in Ω. Similarly, we can get a pseudo-MLE $\hat{\theta}_n$ in τ, and these are clearly consistent for θ_0. This establishes Part (i) of Lemma B.1.

For Part (ii), (B.7) follows immediately from (B.4) and (B.5). To prove (B.8), write

$$\begin{aligned} 2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\theta_0)) &= 2(\hat{\theta}_n^{(1)} - \theta_0)^T S_n(\theta_0) - (\hat{\theta}_n^{(1)} - \theta_0)^T \mathcal{F}_n(\widetilde{\theta}_n)(\hat{\theta}_n^{(1)} - \theta_0) \\ &= h_n(\hat{\theta}_n^{(1)}) + r_n(\hat{\theta}_n^{(1)}), \end{aligned} \tag{B.9}$$

where $\widetilde{\theta}_n = a\theta_0 + (1-a)\hat{\theta}_n^{(1)}$ for some $0 \leq a \leq 1$,

$$h_n(\theta) = -|\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0) - \mathcal{F}_n^{T/2}(\theta_0)(\theta - \theta_0)|^2 + S_n^T(\theta_0)\mathcal{F}_n^{-1}(\theta_0)S_n(\theta_0) \tag{B.10}$$

and

$$r_n(\theta) = (\theta - \theta_0)^T \left(\mathcal{F}_n(\theta_0) - \mathcal{F}_n(\widetilde{\theta}_n)\right)(\theta - \theta_0).$$

In view of the preceding, given $\varepsilon > 0$, we can take $A \geq A_0(\varepsilon)$ and $n \geq n_0(\varepsilon, A)$ such that the event

$$E_n := \{\hat{\theta}_n^{(1)} \in N_n(A),\ \hat{\theta}_n \in N_n(A),\ |\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0)|^2 \leq A^2\lambda_{\min}(Y_n^T Y_n)/4\}$$

has probability at least $1 - \varepsilon$. Suppose E_n occurs. In (B.9) we then have $\widetilde{\theta}_n \in N_n(A)$ WPA1, and consequently from (B.1) and (B.6) we get $r_n(\hat{\theta}_n^{(1)}) \xrightarrow{P} 0$. (B.9) then gives

$$2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\theta_0)) = h_n(\hat{\theta}_n^{(1)}) + o_P(1). \tag{B.11}$$

If $\theta \in N_n(A)$, so that $|G_n^{1/2}(\theta - \theta_0)| \leq A$, we have from (B.4) that

$$\mathcal{F}_n^{T/2}(\theta_0)(\theta - \theta_0) = Y_n^T G_n^{1/2}(\theta - \theta_0) + o_p(1).$$

As in VZ we then get from (B.11)

$$\begin{aligned}
&2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\theta_0)) \\
&= -\inf_{\theta \in N_n(A) \cap \Omega} |\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0) - \mathcal{F}_n^{T/2}(\theta_0)(\theta - \theta_0)| + S_n^T(\theta_0)\mathcal{F}_n^{-1}(\theta_0)S_n(\theta_0) + o_P(1) \\
&= -\inf_{\theta \in N_n(A) \cap \Omega} |\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0) - Y_n^T G_n^{1/2}(\theta - \theta_0)| + S_n^T(\theta_0)\mathcal{F}_n^{-1}(\theta_0)S_n(\theta_0) + o_P(1).
\end{aligned} \tag{B.12}$$

(B.12) differs from VZ in that the norming is by random matrices $\mathcal{F}_n(\theta_0)$ and Y_n rather than by G_n. Nevertheless, proceeding, we transform from θ to $\theta' = Y_n^T G_n^{1/2}(\theta - \theta_0)$, so that $N_n(A)$ transforms to $N_n'(A)$, say. Now (A2) of VZ requires a cone C_Ω with vertex at $\theta_0 \in \Omega$ which coincides with Ω on a neighbourhood of θ_0, i.e, $C_\Omega \cap N_n(A) = \Omega \cap N_n(A)$. We can take

$$C_\Omega = \theta_0 + \{0\} \times \mathbb{R} \times \{0\},$$

which transforms via θ' to

$$\{Y_n^T G_n^{1/2}(\theta - \theta_0) : \theta \in C_\Omega\} = \{0\} \times \mathbb{R} \times \{0\} = \widetilde{C}_\Omega,\ \text{say}.$$

This holds because of the conformation of Y_n; apart from the positive diagonal elements, only the $(3, 2)$ element of Y_n may be nonzero; and G_n is diagonal with positive elements. Note further that if

$$|\theta'|^2 \leq A^2\lambda_{\min}(Y_n^T Y_n),$$

then θ such that $\theta' = Y_n^T G_n^{1/2}(\theta - \theta_0)$ satisfies

$$\begin{aligned}
&(\theta - \theta_0)^T G_n(\theta - \theta_0) = (\theta - \theta_0)^T G_n^{1/2} Y_n (Y_n^T Y_n)^{-1} Y_n^T G_n^{1/2}(\theta - \theta_0) \\
&= (\theta')^T (Y_n^T Y_n)^{-1}\theta' \leq |\theta'|^2 \lambda_{\max}(Y_n^T Y_n)^{-1} \leq A^2,
\end{aligned}$$

so $\theta \in N_n(A)$. Thus $\theta' \in N_n'(A)$, so

$$N_n'(A) \supseteq \{\theta' : |\theta'|^2 \leq A^2\lambda_{\min}(Y_n^T Y_n)\}. \tag{B.13}$$

From (B.12)

$$2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\theta_0))$$
$$= - \inf_{\theta' \in N_n'(A) \cap \widetilde{C}_\Omega} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta'| + S_n^T(\theta_0) \mathcal{F}_n^{-1}(\theta_0) S_n(\theta_0) + o_P(1). \quad (B.14)$$

Let $\dot{\theta}_n \in \widetilde{C}_\Omega$ be such that

$$|\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \dot{\theta}_n| = \inf_{\theta' \in \widetilde{C}_\Omega} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta'|.$$

Since $\widetilde{C}_\Omega$ contains 0,

$$\inf_{\theta' \in \widetilde{C}_\Omega} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta'|^2 \le |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0)|^2 \le A^2 \lambda_{\min}(Y_n^T Y_n)/4,$$

on E_n. It follows that $|\dot{\theta}_n|^2 \le A^2 \lambda_{\min}(Y_n^T Y_n)$, so $\dot{\theta}_n \in N_n'(A)$. Thus from (B.14)

$$2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\theta_0))$$
$$= - \inf_{\theta' \in \widetilde{C}_\Omega} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta'| + S_n^T(\theta_0) \mathcal{F}_n^{-1}(\theta_0) S_n(\theta_0) + o_P(1). \quad (B.15)$$

A similar analysis for $\hat{\theta}_n$ gives

$$2(\mathcal{L}_n(\hat{\theta}_n) - \mathcal{L}_n(\theta_0))$$
$$= - \inf_{\theta' \in \widetilde{C}_\tau} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta'| + S_n^T(\theta_0) \mathcal{F}_n^{-1}(\theta_0) S_n(\theta_0) + o_P(1), \quad (B.16)$$

where $\widetilde{C}_\tau = \mathbb{R}^2 \times [0, \infty)$; notice that $C_\tau = \theta_0 + \mathbb{R}^2 \times [0, \infty)$ transforms to $\widetilde{C}_\tau$ via $\theta' = Y_n^T G_n^{1/2} (\theta - \theta_0)$, again because of the conformation of Y_n and G_n. Subtracting (B.15) from (B.16) we get

$$d_n = \inf_{\theta \in \widetilde{C}_\Omega} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta|^2 - \inf_{\theta \in \widetilde{C}_\tau} |\mathcal{F}_n^{-1/2}(\theta_0) S_n(\theta_0) - \theta|^2 + o_P(1)$$
$$\stackrel{D}{\to} \inf_{\theta \in \widetilde{C}_\Omega} |Z - \theta|^2 - \inf_{\theta \in \widetilde{C}_\tau} |Z - \theta|^2 \quad (B.17)$$

with $Z = (Z_1, Z_2, Z_3)$ as in (B.7). Hence

$$d_n \stackrel{D}{\to} \inf_{\theta \in \{0\} \times \mathbf{R} \times \{0\}} |Z - \theta|^2 - \inf_{\theta \in \mathbf{R}^2 \times [0, \infty)} |Z - \theta|^2$$
$$= Z_1^2 + Z_3^2 - Z_3^2 I(Z_3 < 0) = Z_1^2 + Z_3^2 I(Z_3 \ge 0). \quad (B.18)$$

This is (B.8), and completes the proof of Lemma B.1.

The method of proof of Theorem 3.1 will be to verify the conditions of Lemma B.1 for the model version and null hypothesis currently assumed. Now (B.2) follows from (B.5),

and (B.3) follows from (B.4), (B.5) and (B.6) because Y, as a positive definite matrix, has no mass at 0. The verification of (B.6) under the assumptions of Theorem 3.1 consists of routine but tedious calculations and is relegated to Appendix D. It then only remains to demonstrate (B.4) and (B.5).

Before embarking on these calculations, we find a simplified asymptotic representation for $G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2}$ which holds for all the versions we consider. Inspection of the derivatives listed in Appendix A suggests taking for G_n a diagonal matrix of the form

$$G_n = \mathrm{diag}(a_n,\ n,\ b_n), \tag{B.19}$$

where $a_n \to \infty$ and $b_n \to \infty$ are deterministic sequences whose definitions will depend on the choice of H_0, and on the choice of the sequence g_{i-1} in (1.2). For $H_0^{(1)}$ we take $a_n = n^2$, and for $H_0^{(2)}$, $a_n = n$. Under (1.3), take $b_n = n$, while $b_n = n^3$ is appropriate under (1.4), as we will see. With these choices, the following will be shown to hold:

Lemma B.2. *Assume $H_0^{(1)}$ or $H_0^{(2)}$ for either model specification (1.3) or (1.4), and suppose that, as $n \to \infty$, each of*

$$\frac{1}{\sqrt{nb_n}}\sum_{i=2}^{n}(e_i^2-1)g_{i-1,0}, \quad \frac{1}{b_n}\sum_{i=2}^{n}(e_i^2-1)g_{i-1,0}^2, \tag{B.20a}$$

$$\frac{1}{\sqrt{a_nb_n}}\sum_{i=2}^{n}(e_i^2-1)\frac{\partial g_{i-1,0}}{\partial\alpha}, \quad \frac{1}{\sqrt{nb_n}}\sum_{i=2}^{n}(e_i^2-1)\frac{\partial g_{i-1,0}}{\partial\beta}, \quad \frac{1}{b_n}\sum_{i=2}^{n}(e_i^2-1)\frac{\partial g_{i-1,0}}{\partial\lambda} \tag{B.20b}$$

and

$$\frac{1}{\sqrt{a_nb_n}}\sum_{i=2}^{n}e_iX_{i-1}g_{i-1,0} \tag{B.20c}$$

is $o_P(1)$, where $g_{i-1,0}$ is the value of g_{i-1} under the current null hypothesis. Then

$$G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2} = \begin{bmatrix} \frac{1}{\beta_0 a_n}\sum_{i=2}^{n}X_{i-1}^2 & 0 & 0 \\ & \frac{1}{2\beta_0^2} & \frac{1}{2\beta_0^2\sqrt{nb_n}}\sum_{i=2}^{n}g_{i-1,0} \\ & & \frac{1}{2\beta_0^2 b_n}\sum_{i=2}^{n}g_{i-1,0}^2 \end{bmatrix} + o_P(1) \tag{B.21}$$

(with the remaining elements filled in by symmetry).

Proof of Lemma B.2: Pre– and post– multiplying $\mathcal{F}_n(\theta)$ by $G_n^{-1/2}$, where G_n is given

in (B.19), gives

$$G_n^{-1/2}\mathcal{F}_n(\theta)G_n^{-1/2} = \begin{bmatrix} \dfrac{\mathcal{F}_n^{11}(\theta)}{a_n} & \dfrac{\mathcal{F}_n^{12}(\theta)}{\sqrt{na_n}} & \dfrac{\mathcal{F}_n^{13}(\theta)}{\sqrt{a_nb_n}} \\ & \dfrac{\mathcal{F}_n^{22}(\theta)}{n} & \dfrac{\mathcal{F}_n^{23}(\theta)}{\sqrt{nb_n}} \\ & & \dfrac{\mathcal{F}_n^{33}(\theta)}{b_n} \end{bmatrix} \qquad (B.22)$$

(with the remaining elements filled in by symmetry).

Now, (A.6a) and (A.7a) mean that, in all cases,

$$\frac{\mathcal{F}_n^{11}(\theta_0)}{a_n} = \frac{1}{\beta_0 a_n}\sum_{i=2}^{n} X_{i-1}^2,$$

as required for (B.21).

Under $H_0^{(1)}$, X_{i-1} is a mean 0 finite variance random walk and we take $a_n = n^2$, so by (A.6b), (A.7a) and (A.7b)

$$\frac{\mathcal{F}_n^{12}(\theta_0)}{\sqrt{na_n}} = \frac{1}{\beta_0^{3/2}n^{3/2}}\sum_{i=2}^{n} e_i X_{i-1} = O_P(1/\sqrt{n}) = o_P(1). \qquad (B.23a)$$

This follows by noting that e_iX_{i-1} is a martingale difference sequence with respect to $\mathcal{H}_i$, the σ–field generated by the sequence $\{e_i, e_{i-1}, \ldots, e_1\}$, having variance $\mathrm{Var}(e_iX_{i-1}) = EX_{i-1}^2 = (i-1)\beta_0$. Alternatively, under $H_0^{(2)}$, $X_{i-1} = \sqrt{\beta_0}e_{i-1}$, e_ie_{i-1} is a mean 0 stationary sequence, and we take $a_n = n$, so by the ergodic theorem,

$$\frac{\mathcal{F}_n^{12}(\theta_0)}{\sqrt{na_n}} = \frac{1}{\beta_0 n}\sum_{i=2}^{n} e_i e_{i-1} = o_P(1), \qquad (B.23b)$$

again. This deals with the $(1,2)$ element in (B.21).

Now assume (B.20). From (A.6c), (A.7), (B.20a) and (B.20c) we have in all cases

$$\frac{\mathcal{F}_n^{13}(\theta_0)}{\sqrt{a_nb_n}} = \frac{1}{\beta_0^{3/2}\sqrt{a_nb_n}}\sum_{i=2}^{n} e_i X_{i-1} g_{i-1,0} - \frac{1}{2\beta_0\sqrt{a_nb_n}}\sum_{i=2}^{n}(e_i^2-1)\frac{\partial g_{i-1,0}}{\partial\alpha} = o_P(1). \quad (B.24)$$

Next use $2e_i^2 - 1 = 1 + (2e_i^2 - 2)$ and the ergodic theorem in (A.6d) to get

$$\frac{\mathcal{F}_n^{22}(\theta_0)}{n} = \frac{1}{2\beta_0^2 n}\sum_{i=2}^{n}(2e_i^2-1) = \frac{n-1}{2\beta_0^2 n} + \frac{1}{\beta_0^2 n}\sum_{i=2}^{n}(e_i^2-1) = \frac{1}{2\beta_0^2} + o_P(1), \qquad (B.25)$$

and similarly, using (A.6e) and (A.7),

$$\frac{\mathcal{F}_n^{23}(\theta_0)}{\sqrt{nb_n}} = \frac{1}{2\beta_0^2\sqrt{nb_n}}\sum_{i=2}^n g_{i-1,0} + \frac{1}{\beta_0^2\sqrt{nb_n}}\sum_{i=2}^n (e_i^2-1)g_{i-1,0} - \frac{1}{2\beta_0\sqrt{nb_n}}\sum_{i=2}^n (e_i^2-1)\frac{\partial g_{i-1,0}}{\partial\beta}$$
$$= \frac{1}{2\beta_0^2\sqrt{nb_n}}\sum_{i=2}^n g_{i-1,0} + o_P(1) \quad \text{(by (B.20a) and (B.20b))}. \tag{B.26}$$

Using (A.6f),

$$\frac{\mathcal{F}_n^{33}(\theta_0)}{b_n} = \frac{1}{2\beta_0^2 b_n}\sum_{i=2}^n g_{i-1,0}^2 + \frac{1}{\beta_0^2 b_n}\sum_{i=2}^n (e_i^2-1)g_{i-1,0}^2 - \frac{1}{2\beta_0 b_n}\sum_{i=2}^n (e_i^2-1)\frac{\partial g_{i-1,0}}{\partial\lambda}$$
$$= \frac{1}{2\beta_0^2 b_n}\sum_{i=2}^n g_{i-1}^2 + o_P(1) \quad \text{(by (B.20a) and (B.20b))}. \tag{B.27}$$

Substituting (B.23) – (B.27) in (B.22) gives (B.21).

APPENDIX C: PROOF OF THEOREM 3.1.

Case 1: Suppose (1.3) holds, so $g_{i-1,0} = e_{i-1}^2\sigma_{i-1}^2 = (X_{i-1}-\alpha X_{i-2})^2$, and suppose Case 1(a): $H_0^{(1)} : \alpha_0 = 1, \lambda_0 = 0$.

Under $H_0^{(1)}$ we have $\sigma_i^2 = \sigma_{i0}^2 = \beta_0$ and X_i is a random walk whose increments have expectation 0, variance β_0 and finite third and fourth moments μ_3 and μ_4. Take $a_n = n^2$ and $b_n = n$. First we check (B.20). For (B.20a),

$$\frac{1}{n}\sum_{i=2}^n (e_i^2-1)e_{i-1}^2 \xrightarrow{P} 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=2}^n (e_i^2-1)e_{i-1}^4 \xrightarrow{P} 0 \tag{C.1}$$

hold by the ergodic theorem, as $\{(e_i^2-1)e_{i-1}^2\}$ and $\{(e_i^2-1)e_{i-1}^4\}$ are mean 0 stationary sequences. Next, we have

$$\frac{\partial g_{i-1}}{\partial\alpha}\Big|_{\theta_0} = -2X_{i-2}(X_{i-1}-\alpha X_{i-2})\Big|_{\theta_0} = -2X_{i-2}e_{i-1}\sqrt{\beta_0} \text{ and } \frac{\partial g_{i-1}}{\partial\beta}\Big|_{\theta_0} = 0 = \frac{\partial g_{i-1}}{\partial\lambda}\Big|_{\theta_0},$$

so for (B.20b) we need

$$\frac{1}{n^{3/2}}\sum_{i=2}^n (e_i^2-1)X_{i-2}e_{i-1} = o_P(1). \tag{C.2}$$

The sum in (C.2) is a martingale with respect to $\mathcal{H}_i$, and an easy calculation of its variance establishes the required convergence via Chebychev's inequality. For (B.20c), we use:

Lemma C.1. *Assume e_i are i.i.d. rvs with expectation 0, variance 1 and finite fourth moments, and $X_i = \sum_1^i e_j$. Then*

$$\frac{1}{n^{3/2}}\sum_{i=2}^n e_i X_{i-1}e_{i-1}^2 \xrightarrow{P} 0. \tag{C.3}$$

Proof of Lemma C.1: Using Chebychev's inequality, we see that the normed martingale $\sum_2^n e_i X_{i-1}/n$ is $O_P(1)$ (i.e., stochastically bounded), hence

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i X_{i-1} e_{i-1}^2 = \frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i X_{i-1}(e_{i-1}^2-1) + \frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i X_{i-1} \\
&= \frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i (X_{i-2}+e_{i-1})(e_{i-1}^2-1) + O_P(1/\sqrt{n}) \\
&= \frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i X_{i-2}(e_{i-1}^2-1) + \frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i e_{i-1}(e_{i-1}^2-1) + o_P(1) \\
&= \frac{1}{n^{3/2}}\sum_{i=2}^{n} e_i X_{i-2}(e_{i-1}^2-1) + o_P(1).
\end{aligned}$$

The last equality follows from the ergodic theorem, since the sequence $\{e_i e_{i-1}(e_{i-1}^2-1)\}$ is stationary with mean 0. Now the sequence $\{e_i X_{i-2}(e_{i-1}^2-1), \mathcal{H}_{i-1}\}$ is a martingale difference sequence with variance proportional to $i-2$, so the last term is $o_P(1)$, and thus (C.3) holds.

From (C.1) – (C.3), (B.20) and thus (B.21) follows by Lemma B.2.

Now we demonstrate (B.4). With an application of the weak law of large numbers, (B.21) leads in the present case to

$$G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2} = \begin{bmatrix} \frac{1}{\beta_0 n^2}\sum_{i=2}^{n} X_{i-1}^2 & 0 & 0 \\ 0 & \frac{1}{2\beta_0^2} & \frac{1}{2\beta_0} \\ 0 & \frac{1}{2\beta_0} & \frac{\mu_4}{2} \end{bmatrix} + o_P(1). \qquad (C.4)$$

Taking the Cholesky square root we get

$$G_n^{-1/2}\mathcal{F}_n^{1/2}(\theta_0) = \begin{bmatrix} \sqrt{\frac{1}{\beta_0 n^2}\sum_{i=2}^{n} X_{i-1}^2} & 0 & 0 \\ 0 & \frac{1}{\beta_0\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{\frac{\mu_4-1}{2}} \end{bmatrix} + o_P(1), \qquad (C.5)$$

which is of the form required in (B.4).

Finally we check (B.5). From (A.4) and (A.7) of Appendix A we obtain

$$G_n^{-1/2} S_n(\theta_0) = \sum_{i=2}^{n} \begin{bmatrix} \dfrac{e_i X_{i-1}}{n\sqrt{\beta_0}} \\ \dfrac{e_i^2 - 1}{2\beta_0\sqrt{n}} \\ \dfrac{(e_i^2-1)e_{i-1}^2}{2\sqrt{n}} \end{bmatrix}. \qquad (C.6)$$

From this and (C.5) we see that it will suffice to find the joint limiting distribution of

$$\left(\frac{1}{n\sqrt{\beta_0}} \sum_{i=2}^{n} e_i X_{i-1},\ \frac{1}{2\beta_0\sqrt{n}} \sum_{i=2}^{n} (e_i^2 - 1),\ \frac{1}{2\sqrt{n}} \sum_{i=2}^{n} (e_i^2 - 1) e_{i-1}^2,\ \frac{1}{\beta_0 n^2} \sum_{i=2}^{n} X_{i-1}^2 \right). \quad (C.7)$$

To deal with (C.7), consider the matrix product

$$\begin{bmatrix} \dfrac{X_{i-1}}{\sqrt{n\beta_0}} & 0 & 0 \\ 0 & \dfrac{1}{2\beta_0} & 0 \\ 0 & 0 & \dfrac{1}{2} \end{bmatrix} \begin{bmatrix} e_i/\sqrt{n} \\ (e_i^2 - 1)/\sqrt{n} \\ (e_i^2 - 1)e_{i-1}^2/\sqrt{n} \end{bmatrix} = A_{in} B_{in}, \text{ say}, \qquad (C.8)$$

which forms the summand in (C.6). For $0 \le t \le 1$ let

$$A_n(t) = A_{[nt],n} \quad \text{and} \quad B_n(t) = \sum_{i=1}^{[nt]} B_{in}$$

($[\cdot]$ denotes the integer part), then, with $A_n^{11}(t)$ as the $(1,1)$ element of $A_n(t)$,

$$G_n^{-1/2} S_n(\theta_0) = \int_{2/n}^{1} A_n(t) dB_n(t) \quad \text{and} \quad \frac{1}{\beta_0 n^2} \sum_{i=2}^{n} X_{i-1}^2 = \int_{1/n}^{1} (A_n^{11}(t))^2 dt. \qquad (C.9)$$

Now $\{B_{in}\}_{1 \le i \le n}$ are the terms of a martingale triangular array with respect to $\{\mathcal{H}_i\}$, so by Theorem 2.2 of Kurtz and Protter (1991), from the joint convergence of $(A_n(t), B_n(t))$ to a limit process, $(A(t), B(t))$, say, we can deduce the joint convergence in the Skorohod topology on $[0,1]$ of

$$\left(A_n(t), B_n(t), \int_0^t A_n(s) dB_n(s) \right) \quad \text{to} \quad \left(A(t), B(t), \int_0^t A(s) dB(s) \right). \qquad (C.10)$$

This will give the required joint convergence in (C.7). Since $X_i = \sum_1^i e_j$, the joint convergence of $A_n(\cdot)$ and $B_n(\cdot)$ will follow from that of

$$\left(\frac{1}{\sqrt{n}} \sum_{i=2}^{[nt]} e_i, \ \frac{1}{\sqrt{n}} \sum_{i=2}^{[nt]} (e_i^2 - 1), \ \frac{1}{\sqrt{n}} \sum_{i=2}^{[nt]} (e_i^2 - 1) e_{i-1}^2 \right), \tag{C.11}$$

in $D^3[0,1]$, and we deduce this from a martingale invariance principle (e.g., Durrett (1991, p. 374)) as follows. Take a linear combination

$$Y_i = u_1 e_i + u_2 (e_i^2 - 1) + u_3 (e_i^2 - 1) e_{i-1}^2, \tag{C.12}$$

where $u = (u_1, u_2, u_3)$ are constants with $u_1^2 + u_2^2 + u_3^2 = 1$. Y_i is a martingale difference sequence with respect to $\mathcal{H}_i$, and by the weak law of large numbers

$$\begin{aligned}
\frac{1}{n} \sum_{i=2}^{n} E(Y_i^2 | \mathcal{H}_{i-1}) &= \frac{1}{n} \sum_{i=2}^{n} \left\{ u_1^2 E e_1^2 + u_2^2 E(e_1^2 - 1)^2 + 2u_1 u_2 E(e_1 (e_1^2 - 1)) \right. \\
&\quad \left. + u_3^2 E(e_1^2 - 1)^2 e_{i-1}^4 + 2u_1 u_3 E(e_1 (e_1^2 - 1)) e_{i-1}^2 + 2u_2 u_3 E(e_1^2 - 1)^2 e_{i-1}^2 \right\} \\
&= \frac{1}{n} \sum_{i=2}^{n} \left\{ u_1^2 + u_2^2 (\mu_4 - 1) + 2u_1 u_2 \mu_3 + u_3^2 (\mu_4 - 1) e_{i-1}^4 + 2u_1 u_3 \mu_3 e_{i-1}^2 + 2u_2 u_3 (\mu_4 - 1) e_{i-1}^2 \right\} \\
&\xrightarrow{P} u_1^2 + u_2^2 (\mu_4 - 1) + 2u_1 u_2 \mu_3 + u_3^2 (\mu_4 - 1) \mu_4 + 2u_1 u_3 \mu_3 + 2u_2 u_3 (\mu_4 - 1).
\end{aligned} \tag{C.13}$$

The last expression is $u^T M_3 u$, where

$$M_3 = \begin{bmatrix} 1 & \mu_3 & \mu_3 \\ \mu_3 & \mu_4 - 1 & \mu_4 - 1 \\ \mu_3 & \mu_4 - 1 & \mu_4 (\mu_4 - 1) \end{bmatrix}. \tag{C.14}$$

We need also to check the Lindeberg condition in the form

$$\frac{1}{n} \sum_{i=2}^{n} E(Y_i^2 I(|Y_i| > \varepsilon \sqrt{n}) | \mathcal{H}_{i-1}) \xrightarrow{P} 0, \text{ for each } \varepsilon > 0,$$

and this follows easily by similar calculation as in (C.13). We conclude that the vector in (C.11) has asymptotic distribution that of $M_3^{1/2} \widetilde{W}_3(t)$, where $\widetilde{W}_3(t)$ is SBM in 3 dimensions, $\widetilde{W}_3(t) = (W_1(t), W_2(t), W_3(t))$, say. Consequently, calculating the Cholesky square root of M_3, we see that $G_n^{-1/2} S_n(\theta_0)$ converges in distribution to

$$\int_0^1 A(t) M_3^{1/2} d\widetilde{W}_3(t) = \begin{bmatrix} \displaystyle\int_0^1 W_1(y) dW_1(y) \\ \displaystyle\frac{\mu_3}{2\beta_0} W_1(1) + \frac{\sqrt{\mu_4 - 1 - \mu_3^2}}{2\beta_0} \ W_2(1) \\ \displaystyle\frac{\mu_3}{2} \ W_2(1) + \frac{\sqrt{\mu_4 - 1 - \mu_3^2}}{2} \ W_2(1) + \left(\frac{\mu_4 - 1}{2} \right) W_3(1) \end{bmatrix}. \tag{C.15}$$

Since $A_n^{11}(t)$ converges weakly to $W_1(t)$ (see (C.9)), the limit in distribution of (C.5) is given by replacing the $(1,1)$ element of the matrix with $\sqrt{J_2}$, where

$$J_k = \int_0^1 W_1^k(y)dy \tag{C.16}$$

(as in (3.4)). Taking the inverse of the limit of the matrix in (C.5) gives

$$\left(G_n^{-1/2}\mathcal{F}_n^{1/2}(\theta_0)\right)^{-1} \xrightarrow{D} \begin{bmatrix} \frac{1}{\sqrt{J_2}} & 0 & 0 \\ 0 & \sqrt{2}\beta_0 & 0 \\ 0 & \frac{-\beta_0\sqrt{2}}{\sqrt{\mu_4-1}} & \sqrt{\frac{2}{\mu_4-1}} \end{bmatrix}. \tag{C.17}$$

Pre-multiplying the matrix in (C.15) by the one in (C.17) gives (since we have joint convergence of all elements)

$$\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0) \xrightarrow{D} \begin{bmatrix} \int_0^1 W_1(y)dW_1(y)/\sqrt{J_2} \\ \frac{\mu_3}{\sqrt{2}}W_1(1) + \sqrt{\frac{\mu_4-1-\mu_3^2}{2}}\,W_2(1) \\ \sqrt{\frac{\mu_4-1}{2}}\,W_3(1) \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}, \tag{C.18}$$

from which we identify the Z_3 needed in (B.8) as

$$Z_3 = \sqrt{\mu_4-1}\,W_3(1)/\sqrt{2}. \tag{C.19}$$

This is distributed independently of Z_1, the first element in (C.18). Thus the limiting distribution of $d_n^{(1)}$ can be found from (B.8) as

$$\frac{\left(\int_0^1 W_1(y)dW_1(y)\right)^2}{\int_0^1 W_1^2(y)dy} + \frac{1}{2}(\mu_4-1)W_3^2(1)I(W_3(1)\le 0), \tag{C.20}$$

which is of the form required in (3.1). This proves (3.1), subject to verifying (B.6), which is done in Appendix D.

Case 1: Suppose (1.3) holds, so $g_{i-1,0} = e_{i-1}^2\sigma_{i-1}^2 = (X_{i-1}-\alpha X_{i-2})^2$, and suppose Case 1(b): $H_0^{(2)} : \alpha_0 = 0 = \lambda_0$.

Next assume (1.3) and (2.5), so that $X_i = e_i\sigma_{i0} = e_i\sqrt{\beta_0}$ is now an i.i.d. sequence. Only some minor modifications to the previous analysis are needed. Again take $b_n = n$,

but this time let $a_n = n$. (B.20a) holds as before. (B.20b) requires, instead of (C.3), $\sum_2^n (e_i^2 - 1)e_{i-2}e_{i-1}^3/n \xrightarrow{P} 0$, and (B.20c) requires $\sum_2^n e_i e_{i-1}^3/n \xrightarrow{P} 0$. These hold by the ergodic theorem. Instead of (C.6), use

$$G_n^{-1/2} S_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=2}^{n} \begin{bmatrix} e_i e_{i-1} \\ (e_i^2 - 1)/2\beta_0 \\ (e_i^2 - 1)e_{i-1}^2/2 \end{bmatrix}, \tag{C.21}$$

and, from (B.21), in place of (C.4) write, by virtue of the weak law of large numbers,

$$G_n^{-1/2} \mathcal{F}_n(\theta_0) G_n^{-1/2} = \begin{bmatrix} 1 & 0 \\ 0 & F_2 \end{bmatrix} + o_P(1) \tag{C.22}$$

where F_2 is the lower right 2×2 block in (C.4). Again, the Cholesky square root of this is of the form required in (B.4). In place of (C.12) use the martingale difference sequence

$$Y_i = u_1 e_i e_{i-1} + \frac{u_2}{2\beta_0}(e_i^2 - 1) + \frac{1}{2} u_3 (e_i^2 - 1) e_{i-1}^2,$$

and with an analysis similar to (C.13) we obtain

$$G_n^{-1/2} S_n(\theta_0) \xrightarrow{D} \widetilde{M}_3^{1/2} \widetilde{W}_3(1),$$

where

$$\widetilde{M}_3 = \begin{bmatrix} 1 & 0 & \dfrac{\mu_3^2}{2} \\ 0 & \dfrac{\mu_4 - 1}{4\beta_0^2} & \dfrac{\mu_4 - 1}{4\beta_0} \\ \dfrac{\mu_3^2}{2} & \dfrac{\mu_4 - 1}{4\beta_0} & \dfrac{\mu_4(\mu_4 - 1)}{4} \end{bmatrix}, \tag{C.23}$$

and $\widetilde{W}_3 = (W_1(t), W_2(t), W_3(t))$ is SBM in 3 dimensions. We find that Z satisfies

$$Z_1 = W_1(1), \quad Z_2 = \sqrt{(\mu_4 - 1)/2}\; W_2(1), \tag{C.24a}$$

$$Z_3 = \frac{\mu_3^2}{\sqrt{2(\mu_4 - 1)}}\; W_1(1) + \sqrt{\frac{(\mu_4 - 1)^2 - \mu_3^4}{2(\mu_4 - 1)}}\; W_3(1). \tag{C.24b}$$

Hence (3.2) of Theorem 3.1 follows if we verify (B.6) for the present setup. We omit the details of this.

Case 2: Suppose (1.4) holds, so $g_{i-1} = X_{i-1}^2$, and suppose

Case 2(a): $H_0^{(1)} : \alpha_0 = 1, \lambda_0 = 0$**.**

The appropriate b_n is $b_n = n^3$ and for $H_0^{(1)}$ we take $a_n = n^2$, so

$$G_n = \mathrm{diag}(n^2,\ n,\ n^3). \tag{C.25}$$

(B.20) holds here immediately as an application of the functional central limit theorem (note that the partial derivatives of g_{i-1} are 0 in this case), so (B.21) holds and we have

$$G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2} = \begin{bmatrix} \frac{1}{\beta_0 n^2}\sum_{i=2}^{n} X_{i-1}^2 & 0 & 0 \\ & \frac{1}{2\beta_0^2} & \frac{1}{2\beta_0^2 n^2}\sum_{i=2}^{n} X_{i-1}^2 \\ & & \frac{1}{2\beta_0^2 n^3}\sum_{i=2}^{n} X_{i-1}^4 \end{bmatrix} + o_P(1) \tag{C.26}$$

(with the remaining elements filled in by symmetry). Under $H_0^{(1)}$ the first derivatives satisfy

$$G_n^{-1/2}S_n(\theta_0) = \sum_{i=2}^{n} \begin{bmatrix} \frac{e_i X_{i-1}}{n\sqrt{\beta_0}} \\ \frac{e_i^2 - 1}{2\beta_0\sqrt{n}} \\ \frac{(e_i^2-1)X_{i-1}^2}{2\beta_0 n^{3/2}} \end{bmatrix}, \tag{C.27}$$

and differ from (C.6) only in the third element. From (C.26) and (C.27) we see that we need the joint asymptotic distribution of

$$\left(\frac{1}{n}\sum_{i=2}^{n} e_i X_{i-1},\ \frac{1}{\sqrt{n}}\sum_{i=2}^{n}(e_i^2-1),\ \frac{1}{n^{3/2}}\sum_{i=2}^{n}(e_i^2-1)X_{i-1}^2,\ \frac{1}{n^2}\sum_{i=2}^{n} X_{i-1}^2,\ \frac{1}{n^3}\sum_{i=2}^{n} X_{i-1}^4\right). \tag{C.28}$$

Instead of (C.8), this time consider

$$\begin{bmatrix} \frac{X_{i-1}}{\sqrt{\beta_0 n}} & 0 \\ 0 & \frac{1}{2\beta_0} \\ 0 & \frac{X_{i-1}^2}{2\beta_0 n} \end{bmatrix} \begin{bmatrix} e_i/\sqrt{n} \\ (e_i^2-1)/\sqrt{n} \end{bmatrix} = \widetilde{A}_{in}\widetilde{B}_{in},\ \text{say.} \tag{C.29}$$

This can be written in the form $\int \widetilde{A}_n(t)d\widetilde{B}_n(t)$ in a similar way as before. Once again the required joint convergence will be obtained if $(\widetilde{A}_n, \widetilde{B}_n) \xrightarrow{D} (\widetilde{A}(t), \widetilde{B}(t))$, say. But this is

immediate from the convergence of the normed martingale

$$\frac{1}{\sqrt{n}}\sum_{i=2}^{[nt]}(e_i,\ e_i^2-1) \tag{C.30}$$

to $\widetilde{M}_2^{1/2}\widetilde{W}_2(t)$, where $\widetilde{W}_2(t)=(W_1(t),W_2(t))$ is an SBM in 2 dimensions and $\widetilde{M}_2=\mathrm{Var}(e_1,e_1^2-1)$ is the upper left 2×2 block in (C.14). Calculations similar to (C.15) give

$$G_n^{-1/2}S_n(\theta_0)\xrightarrow{D}\begin{bmatrix}\displaystyle\int_0^1 W_1(y)dW_1(y)\\ \displaystyle\frac{\mu_3}{2\beta_0}W_1(1)+\frac{\sqrt{\mu_4-1-\mu_3^2}}{2\beta_0}\,W_2(1)\\ \displaystyle\frac{\mu_3}{2}\int_0^1 W_1^2(y)dW_1(y)+\frac{\sqrt{\mu_4-1-\mu_3^2}}{2}\int_0^1 W_1^2(y)dW_2(y)\end{bmatrix}. \tag{C.31}$$

Once again we have the convergence of the first component of $\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0)$ to Z_1, as in (C.18). For the (2,3) components, note that the lower right 2×2 block of $G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2}$ in (C.26) converges (jointly with the other quantities) to

$$\frac{1}{2}\begin{bmatrix}\dfrac{1}{\beta_0^2} & \dfrac{1}{\beta_0}\displaystyle\int_0^1 W_1^2(y)dy\\ \dfrac{1}{\beta_0}\displaystyle\int_0^1 W_1^2(y)dy & \displaystyle\int_0^1 W_1^4(y)dy\end{bmatrix}=\frac{1}{2}\begin{bmatrix}\dfrac{1}{\beta_0^2} & \dfrac{J_2}{\beta_0}\\ \dfrac{J_2}{\beta_0} & J_4\end{bmatrix}, \tag{C.32}$$

where J_k is given by (C.16). The matrix in (C.32) has inverse Cholesky square root

$$\begin{bmatrix}\beta_0\sqrt{2} & 0\\ \dfrac{-\sqrt{2}\beta_0 J_2}{\sqrt{J_4-J_2^2}} & \sqrt{\dfrac{2}{J_4-J_2^2}}\end{bmatrix}, \tag{C.33}$$

and consequently

$$\left(G_n^{-1/2}\mathcal{F}_n^{1/2}(\theta_0)\right)^{-1}\xrightarrow{D}\begin{bmatrix}\dfrac{1}{\sqrt{J_2}} & 0 & 0\\ 0 & \sqrt{2}\beta_0 & 0\\ 0 & \dfrac{-\beta_0\sqrt{2}J_2}{\sqrt{J_4-J_2^2}} & \sqrt{\dfrac{2}{J_4-J_2^2}}\end{bmatrix}. \tag{C.34}$$

Thus $\mathcal{F}_n^{-1/2}(\theta_0)S_n(\theta_0)$ converges in distribution to (Z_1,Z_2,Z_3), where Z_1 is as in (C.18),

$$Z_2=\frac{\mu_3}{\sqrt{2}}W_1(1)+\sqrt{\frac{\mu_4-1-\mu_3^2}{2}}\,W_2(1) \tag{C.35}$$

and (with J_k as in (C.16))

$$Z_3 = \frac{1}{\sqrt{2(J_4 - J_2^2)}} \left\{ -J_2 \left(\mu_3 W_1(1) + \sqrt{\mu_4 - 1 - \mu_3^2}\, W_2(1) \right) \right.$$

$$\left. + \mu_3 \int_0^1 W_1^2(y) dW_1(y) + \sqrt{\mu_4 - 1 - \mu_3^2} \int_0^1 W_1^2(y) dW_2(y) \right\}. \qquad (C.36)$$

From this we obtain Z_1 and Z_3 as required for (B.8) and then (3.3) follows.

When $\mu_3 = 0$ and $\mu_4 = 3$, as they are in the simulations reported in Section 5, Z_3 is standard normal. We leave this for the reader to check.

Case 2: Suppose (1.4) holds, so $g_{i-1} = X_{i-1}^2$, and suppose
Case 2(b): $H_0^{(2)} : \alpha_0 = 0 = \lambda_0$.

Under $H_0^{(2)}$, $X_i = e_i\sqrt{\beta_0}$, so $g_{i-1} = X_{i-1}^2 = e_{i-1}^2\beta_0$ and this model is exactly the same as for Case 1: $g_{i-1} = e_{i-1}^2\beta_0$. So we again arrive at (3.2) for the limiting distribution.

Remarks: The AR–GARCH Case. Lee (1991) observed that the Lagrange multiplier test, based on the log-pseudo-likelihood, for the hypothesis $\lambda_0 = 0 = \delta_0$ in GARCH, is the same as for the hypothesis $\lambda_0 = 0$ in ARCH. This is because when the true model is random walk, a GARCH model is over-parameterised in that the β and δ parameters cannot be separately identified. The same is true for the general conditional heteroskedasticity model in (1.2) (including the AR–GARCH model) under the mild conditions we have imposed. For the model specified by (1.1)–(1.3) (now with δ included), and assuming (B.20), we have instead of (B.21), for $G_n^{-1/2}\mathcal{F}_n(\theta_0)G_n^{-1/2}$ the 4×4 matrix

$$2 \begin{bmatrix} \frac{1}{\beta_0 a_n}\sum_{i=2}^n X_{i-1}^2 & 0 & 0 & 0 \\ & \frac{1}{\beta_0^2} & \frac{1}{\beta_0^2\sqrt{nb_n}}\sum_{i=2}^n g_{i-1,0} & \frac{1}{\beta_0} \\ & & \frac{1}{\beta_0^2 b_n}\sum_{i=2}^n g_{i-1,0}^2 & \frac{1}{\beta_0^2\sqrt{nb_n}}\sum_{i=2}^n g_{i-1,0} \\ & & & 1 \end{bmatrix} + o_P(1)$$

(with the remaining elements filled in by symmetry). In the limit as $n \to \infty$ this is singular, WPA1, due to the asymptotic linear dependence between the β and δ parameters. Just as in Lee (1991) we can now use the theory of Aitchison and Silvey (1958) and Silvey (1959) to show that the limiting distributions of the d_n statistics for testing $H_0^{(1)} : \alpha_0 = 1, \lambda_0 = 0 = \delta_0$ or $H_0^{(2)} : \alpha_0 = 0 = \lambda_0 = \delta_0$ are still given by (3.1), (3.2), or (3.3), in the respective cases.

APPENDIX D: VERIFYING (B.6) WHEN $H_0^{(1)}$ HOLDS AND $g_{i-1} = \varepsilon_{i-1}^2\sigma_{i-1}^2$.

Let $\theta \in N_n(A) \cap \tau$ (see (2.3) and (B.1)) with $A > 0$ and take

$$G_n = \mathrm{diag}(a_n, n, b_n) = \mathrm{diag}(n^2, n, n). \tag{D.1}$$

Thus, using (B.1) and with $\theta_0 = (1, \beta_0, 0)$,

$$n^2(\alpha - 1)^2 + n(\beta - \beta_0)^2 + n\lambda^2 \le A^2. \tag{D.2}$$

Now

$$\sigma_i^2 = \beta + \lambda g_{i-1} = \beta + \lambda\varepsilon_{i-1}^2\sigma_{i-1}^2 = \beta + \lambda(X_{i-1} - \alpha X_{i-2})^2, \ i = 3, 4, \ldots \tag{D.3}$$

has

$$\frac{\partial\sigma_i^2}{\partial\alpha} = -2\lambda X_{i-2}(X_{i-1} - \alpha X_{i-2}) = -2\lambda X_{i-2}\varepsilon_{i-1}\sigma_{i-1} \quad \text{and} \quad \frac{\partial^2\sigma_i^2}{\partial\alpha^2} = 2\lambda X_{i-2}^2.$$

So we have from (A.6a)

$$\frac{1}{n^2}\Big|\mathcal{F}_n^{11}(\theta) - \mathcal{F}_n^{11}(\theta_0)\Big| \le \frac{1}{n^2}\sum_{i=2}^n \Big|\frac{X_{i-1}^2}{\sigma_i^2} - \frac{X_{i-1}^2}{\sigma_{i0}^2}\Big|$$
$$+ \frac{\lambda}{n^2}\Big|\sum_{i=2}^n \left(\frac{4X_{i-1}\varepsilon_i}{\sigma_i^3} - \frac{2\lambda(2\varepsilon_i^2 - 1)X_{i-2}\varepsilon_{i-1}\sigma_{i-1}}{\sigma_i^4}\right) X_{i-2}\varepsilon_{i-1}\sigma_{i-1}\Big| + \frac{\lambda}{n^2}\Big|\sum_{i=2}^n \frac{(\varepsilon_i^2 - 1)X_{i-2}^2}{\sigma_i^2}\Big|. \tag{D.4}$$

By (D.2), $|\alpha - 1| \le A/n$, $|\beta - \beta_0| \le A/\sqrt{n}$, and $\lambda \le A/\sqrt{n}$. Assume n is so large that $A/\sqrt{n} \le \beta_0/2$, then $\beta \ge \beta_0 - A/\sqrt{n} \ge \beta_0/2$, and so $\sigma_i^2 \ge \beta_0/2$ (almost surely – we will omit this qualifier from now on). This means that any σ_i in a denominator is bounded away from 0. As a mean 0 finite variance random walk, the X_i satisfy $\max_{1\le i\le n}|X_i| = O_P(\sqrt{n})$ as $n \to \infty$, and as an i.i.d. sequence with finite fourth moment, the e_i satisfy $\max_{1\le i\le n} e_i^4/n \xrightarrow{P} 0$, thus $\max_{1\le i\le n}|e_i| = o_P(n^{1/4})$. Consequently

$$\varepsilon_{i-1}\sigma_{i-1} = X_{i-1} - \alpha X_{i-2} = \sqrt{\beta_0}e_{i-1} + (1-\alpha)X_{i-2} = \sqrt{\beta_0}e_{i-1} + O_P(1/\sqrt{n}) = o_P(n^{1/4}), \tag{D.5}$$

uniformly in $1 \le i \le n$. This also shows that $\max_{1\le i\le n}\varepsilon_i^2 = o_P(\sqrt{n})$. It follows from (D.5) that

$$\sigma_i^2 - \sigma_{i0}^2 = \beta - \beta_0 + \lambda\varepsilon_{i-1}^2\sigma_{i-1}^2 = O_P(1/\sqrt{n}) + o_P(1) = o_P(1), \tag{D.6}$$

thus, also,

$$\Big|\frac{1}{\sigma_i^2} - \frac{1}{\sigma_{i0}^2}\Big| = \frac{|\sigma_i^2 - \sigma_{i0}^2|}{\sigma_i^2\sigma_{i0}^2} \le \frac{2|\sigma_i^2 - \sigma_{i0}^2|}{\beta_0^2} = o_P(1), \tag{D.7}$$

uniformly in $1 \le i \le n$. These show that the first sum on the righthand side of (D.4) is $o_P(1)$. The second and third sums on the righthand side of (D.4) are also $o_P(1)$, using the estimates obtained above. This deals with the $(1,1)$ term.

(D.3) shows that $\partial\sigma_i^2/\partial\beta = 1$, so (A.6b) together with $\max_{1\le i\le n}\varepsilon_i^2 = o_P(\sqrt{n})$ gives

$$\begin{aligned}\frac{1}{n^{3/2}}\Big|\mathcal{F}_n^{12}(\theta)-\mathcal{F}_n^{12}(\theta_0)\Big| &\le \frac{1}{n^{3/2}}\Big|\sum_{i=2}^n\left(\frac{X_i-\alpha X_{i-1}}{\sigma_i^4}-\frac{\sqrt{\beta_0}e_i}{\sigma_{i0}^4}\right)X_{i-1}\Big| \\ &\quad+\frac{\lambda}{n^{3/2}}\Big|\sum_{i=2}^n\frac{(2\varepsilon_i^2-1)X_{i-2}\varepsilon_{i-1}\sigma_{i-1}}{\sigma_i^4}\Big| \\ &\le \frac{1}{n^{3/2}}\Big|\sum_{i=2}^n X_{i-1}\frac{(X_i-\alpha X_{i-1}-\sqrt{\beta_0}e_i)}{\sigma_i^4}\Big|+\frac{\sqrt{\beta_0}}{n^{3/2}}\Big|\sum_{i=2}^n X_{i-1}e_i\left(\frac{1}{\sigma_i^4}-\frac{1}{\sigma_{i0}^4}\right)\Big| \\ &\quad+\frac{A}{n^2}\Big|\sum_{i=2}^n\frac{(2\varepsilon_i^2-1)X_{i-2}\varepsilon_{i-1}\sigma_{i-1}}{\sigma_i^4}\Big|. \qquad (D.8)\end{aligned}$$

From (D.7) we get

$$\Big|\frac{1}{\sigma_i^4}-\frac{1}{\sigma_{i0}^4}\Big| = o_P(1), \qquad (D.9)$$

uniformly in $1\le i\le n$, so the second term on the righthand side of (D.8) is

$$o_P\left(\sum_{i=2}^n|e_i|/n\right) = o_P(1), \qquad (D.10)$$

by the weak law of large numbers. For the first and third terms we need

$$\varepsilon_i\sigma_i-\sqrt{\beta_0}e_i = X_i-\alpha X_{i-1}-\sqrt{\beta_0}e_i = (1-\alpha)X_{i-1} = O_P(1/\sqrt{n}), \qquad (D.11)$$

which gives $\varepsilon_i = e_i\sqrt{\beta_0}/\sigma_i + o_P(1)$. Thus the first term on the righthand side of (D.8) is $O_P(1/\sqrt{n}) = o_P(1)$. The third term is $O_P(n^{-3/2}\sum_{i=2}^n e_i^2|e_{i-1}|)$, which is $O_P(n^{-1/2})$ by the ergodic theorem. This deals with the $(1,2)$ term.

(D.3) shows that $\partial\sigma_i^2/\partial\lambda = \varepsilon_{i-1}^2\sigma_{i-1}^2$, so (A.6c) gives

$$\begin{aligned}&\frac{1}{n^{3/2}}\Big|\mathcal{F}_n^{13}(\theta)-\mathcal{F}_n^{13}(\theta_0)\Big| \\ &\le \frac{1}{n^{3/2}}\Big|\sum_{i=2}^n\left(\frac{\varepsilon_i\varepsilon_{i-1}^2\sigma_{i-1}^2}{\sigma_i^3}-\frac{e_ie_{i-1}^2\beta_0}{\sigma_{i0}^3}\right)X_{i-1}\Big|+\frac{\lambda}{n^{3/2}}\Big|\sum_{i=2}^n\frac{(2\varepsilon_i^2-1)X_{i-2}\varepsilon_{i-1}^3\sigma_{i-1}^3}{\sigma_i^4}\Big| \\ &+\frac{1}{n^{3/2}}\Big|\sum_{i=2}^n\left(\frac{(\varepsilon_i^2-1)\varepsilon_{i-1}\sigma_{i-1}}{\sigma_i^2}-\frac{(e_i^2-1)e_{i-1}\sqrt{\beta_0}}{\sigma_{i0}^2}\right)X_{i-2}\Big| \\ &\le \frac{1}{n^{3/2}}\Big|\sum_{i=2}^n\frac{(\varepsilon_i\varepsilon_{i-1}^2\sigma_{i-1}^2-e_ie_{i-1}^2\beta_0)X_{i-1}}{\sigma_i^3}\Big|+\frac{1}{n^{3/2}}\Big|\sum_{i=2}^n X_{i-1}e_ie_{i-1}^2\left(\frac{1}{\sigma_i^3}-\frac{1}{\sigma_{i0}^3}\right)\Big| \\ &+\frac{A}{n^2}\Big|\sum_{i=2}^n\frac{(2\varepsilon_i^2-1)X_{i-2}\varepsilon_{i-1}^3\sigma_{i-1}^3}{\sigma_i^4}\Big| \\ &\qquad+\frac{1}{n^{3/2}}\Big|\sum_{i=2}^n\left(\frac{(\varepsilon_i^2-1)\varepsilon_{i-1}\sigma_{i-1}}{\sigma_i^2}-\frac{(e_i^2-1)e_{i-1}\sqrt{\beta_0}}{\sigma_{i0}^2}\right)X_{i-2}\Big|. \qquad (D.12)\end{aligned}$$

Note that

$$\begin{aligned}\varepsilon_i^2 - e_i^2 &= ((X_i - \alpha X_{i-1})^2 - \sigma_i^2 e_i^2)/\sigma_i^2 \\ &= (1-\alpha)X_{i-1}(X_i - \alpha X_{i-1} + \sqrt{\beta_0}e_i)/\sigma_i^2 + (\sigma_{i0}^2 - \sigma_i^2)e_i^2/\sigma_i^2 \\ &= O_P(1/\sqrt{n})(\varepsilon_i\sigma_i + \sqrt{\beta_0}e_i) + e_i^2 o_P(1) = o_P(1)(1+e_i^2), \qquad (D.13)\end{aligned}$$

by (D.5). So $\varepsilon_i^2 = (1+o_P(1))e_i^2 + o_P(1)$. In the first term on the righthand side of (D.12), use

$$\varepsilon_i\varepsilon_{i-1}^2\sigma_{i-1}^2 - e_i e_{i-1}^2\beta_0 = (\varepsilon_i\varepsilon_{i-1}^2 - e_i e_{i-1}^2)\sigma_{i-1}^2 + e_i e_{i-1}^2(\sigma_{i-1}^2 - \beta_0),$$

(D.6), and, from (D.13),

$$\begin{aligned}&\varepsilon_i\varepsilon_{i-1}^2 - e_i e_{i-1}^2 = (\varepsilon_i - e_i)\varepsilon_{i-1}^2 + (\varepsilon_{i-1}^2 - e_{i-1}^2)e_i \\ &= (X_i - \alpha X_{i-1} - \sqrt{\beta_0}e_i)\varepsilon_{i-1}^2/\sigma_i + (\sigma_{i0} - \sigma_i)e_i\varepsilon_{i-1}^2/\sigma_i + o_P(1)(1+e_{i-1}^2)|e_i| \\ &= O_P(1/\sqrt{n})o_P(\sqrt{n}) + o_P(1)(1+o_P(1))|e_i|e_i^2 + o_P(1)(1+e_{i-1}^2)|e_i|.\end{aligned}$$

Since $X_{i-1} = O_P(\sqrt{n})$, uniformly in $1 \le i \le n$, we can sum over $1 \le i \le n$ and divide by $n^{3/2}$ to see that the first term on the righthand side of (D.12) is $o_P(1)$. The second term on the righthand side of (D.12) is $o_P(1)$, by a similar analysis as in (D.9) and (D.10). The third term on the righthand side of (D.12) is $o_P(1)$ by (D.13). The fourth term on the righthand side of (D.12) can be handled similarly to the first and second terms; note that

$$(\varepsilon_i^2 - 1)\varepsilon_{i-1} - (e_i^2 - 1)e_{i-1} = (\varepsilon_i^2 - e_i^2)\varepsilon_{i-1} + (e_i^2 - 1)(\varepsilon_{i-1} - e_{i-1}),$$

and use (D.13). Next, (D.3) shows that $\partial\sigma_i^2/\partial\beta = 1$, so $\partial^2\sigma_i^2/\partial\beta^2 = 0$ and (A.6d) gives

$$\begin{aligned}&\frac{1}{n}\Big|\mathcal{F}_n^{22}(\theta) - \mathcal{F}_n^{22}(\theta_0)\Big| = \frac{1}{2n}\Big|\sum_{i=2}^n\left(\frac{(2\varepsilon_i^2-1)}{\sigma_i^4} - \frac{(2e_i^2-1)}{\sigma_{i0}^4}\right)\Big| \\ &\le \frac{1}{n}\Big|\sum_{i=2}^n\frac{(\varepsilon_i^2 - e_i^2)}{\sigma_i^4}\Big| + \frac{1}{n}\Big|\sum_{i=2}^n(2e_i^2-1)\left(\frac{1}{\sigma_i^4} - \frac{1}{\sigma_{i0}^4}\right)\Big|. \qquad (D.14)\end{aligned}$$

The first term on the righthand side of (D.14) is $o_P(1)$, using (D.13). The second term on the righthand side of (D.14) is also $o_P(1)$, using (D.9).

For the $(2,3)$ component use (A.6e) to get

$$\begin{aligned}&\frac{1}{n}\Big|\mathcal{F}_n^{23}(\theta) - \mathcal{F}_n^{23}(\theta_0)\Big| = \frac{1}{2n}\Big|\sum_{i=2}^n\left(\frac{(2\varepsilon_i^2-1)\varepsilon_{i-1}^2\sigma_{i-1}^2}{\sigma_i^4} - \frac{(2e_i^2-1)e_i^2\beta_0}{\sigma_{i0}^4}\right)\Big| \\ &\le \frac{1}{n}\Big|\sum_{i=2}^n\frac{(2\varepsilon_i^2-1)\varepsilon_{i-1}^2\sigma_{i-1}^2 - (2e_i^2-1)e_{i-1}^2\beta_0}{\sigma_i^4} + \frac{\beta_0}{n}\Big|\sum_{i=2}^n(2e_i^2-1)e_{i-1}^2\left(\frac{1}{\sigma_i^4} - \frac{1}{\sigma_{i0}^4}\right)\Big|. \qquad (D.15)\end{aligned}$$

The second term on the righthand side of (D.15) is $o_P(1)$, by (D.9). For the first,

$$(2\varepsilon_i^2-1)\varepsilon_{i-1}^2\sigma_{i-1}^2-(2e_i^2-1)e_{i-1}^2\beta_0=2(\varepsilon_i^2-e_i^2)\varepsilon_{i-1}^2\sigma_{i-1}^2+(2e_i^2-1)(\varepsilon_{i-1}^2\sigma_{i-1}^2-e_{i-1}^2\beta_0)$$
$$=\big(2(\varepsilon_i^2-e_i^2)+(2e_i^2-1)\big)\,(\varepsilon_{i-1}^2\sigma_{i-1}^2-e_{i-1}^2\beta_0)+2(\varepsilon_i^2-e_i^2)e_{i-1}^2\beta_0. \qquad (D.16)$$

The last term is $o_P(1+e_{i-1}^2)$ by (D.13), and this will give a term of

$$o_P\left(1+\sum_{i=2}^n e_{i-1}^2/n\right)=o_P(1), \qquad (D.17)$$

in (D.15). Also, using (D.5) and (D.11),

$$\varepsilon_{i-1}^2\sigma_{i-1}^2-e_{i-1}^2\beta_0=(\varepsilon_{i-1}\sigma_{i-1}-\sqrt{\beta_0}e_i)(\varepsilon_{i-1}\sigma_{i-1}+\sqrt{\beta_0}e_i)=O_P(1/\sqrt{n})o_P(n^{1/4})=o_P(1). \qquad (D.18)$$

Thus the first term in (D.16) is of order $o_P(1)(1+e_{i-1}^2)$ and so gives rise to a $o_P(1)$ term as in (D.17).

Finally for the $(3,3)$ component, (A.6f) gives

$$\frac{1}{n}\Big|\mathcal{F}_n^{33}(\theta)-\mathcal{F}_n^{33}(\theta_0)\Big|=\frac{1}{2n}\Big|\sum_{i=2}^n\left(\frac{(2\varepsilon_i^2-1)\varepsilon_{i-1}^4\sigma_{i-1}^4}{\sigma_i^4}-\frac{(2e_i^2-1)e_{i-1}^4\beta_0^2}{\sigma_{i0}^4}\right)\Big|$$
$$\le\frac{1}{2n}\Big|\sum_{i=2}^n\frac{(2\varepsilon_i^2-1)\varepsilon_{i-1}^4\sigma_{i-1}^4-(2e_i^2-1)e_{i-1}^4\beta_0^2}{\sigma_i^4}+\frac{\beta_0^2}{2n}\Big|\sum_{i=2}^n(2e_i^2-1)e_{i-1}^4\left(\frac{1}{\sigma_i^4}-\frac{1}{\sigma_{i0}^4}\right)\Big|. \qquad (D.19)$$

The second term on the righthand side of (D.19) is treated as in (D.15), noting that

$$\frac{1}{n}\sum_{i=2}^n|2e_i^2-1|e_{i-1}^4\xrightarrow{P}E(|2e_2^2-1|e_1^4)$$

by the ergodic theorem. For the first,

$$(2\varepsilon_i^2-1)\varepsilon_{i-1}^4\sigma_{i-1}^4-(2e_i^2-1)e_{i-1}^4\beta_0^2=2(\varepsilon_i^2-e_i^2)\varepsilon_{i-1}^4\sigma_{i-1}^4+(2e_i^2-1)(\varepsilon_{i-1}^4\sigma_{i-1}^4-e_{i-1}^4\beta_0^2)$$
$$=\big(2(\varepsilon_i^2-e_i^2)+(2e_i^2-1)\big)\,(\varepsilon_{i-1}^4\sigma_{i-1}^4-e_{i-1}^4\beta_0^2)+(\varepsilon_i^2-e_i^2)e_{i-1}^4\beta_0^2. \qquad (D.20)$$

The last term is $o_P(e_{i-1}^4+e_i^2e_{i-1}^4)$ by (D.14) which will give an $o_P(1)$ contribution to (D.19) when averaged over $1\le i\le n$. Next, $\varepsilon_{i-1}^2\sigma_{i-1}^2=e_{i-1}^2\beta_0+o_P(1)$ by (D.5), so

$$\begin{aligned}\varepsilon_{i-1}^4\sigma_{i-1}^4-e_{i-1}^4\beta_0^2&=(\varepsilon_{i-1}^2\sigma_{i-1}^2-e_{i-1}^2\beta_0)(o_P(1)+e_{i-1}^2\beta_0)\\&=o_P(1)(o_P(1)+e_{i-1}^2\beta_0),\end{aligned}$$

uniformly in $1\le i\le n$, using (D.18) again. This will also give an $o_P(1)$ contribution to (D.19) when averaged over $1\le i\le n$. This completes the proof of (B.6) for this model.