ON THE VALIDITY OF THE JARQUE-BERA NORMALITY TEST IN CONDITIONALLY HETEROSKEDASTIC DYNAMIC REGRESSION MODELS

Gabriele Fiorentini, Enrique Sentana
and Giorgio Calzolari

ON THE VALIDITY OF THE JARQUE-BERA NORMALITY TEST IN CONDITIONALLY HETEROSKEDASTIC DYNAMIC REGRESSION MODELS

Abstract

We show that the Jarque-Bera test, originally devised for constant conditional variance models with no functional dependence between conditional mean and variance parameters, can be safely applied to a broad class of GARCH-M models, but not to all.

JEL Codes: C52, C15, C22
Keywords: Skewness, Kurtosis, ARCH, Moment Tests.

Gabriele Fiorentini
Università di Firenze
fiorentini@ds.unifi.it

Enrique Sentana
CEMFI
sentana@cemfi.es

Giorgio Calzolari
Università di Firenze
calzolar@ds.unifi.it

1 Introduction

In univariate, conditionally heteroskedastic, dynamic regression models, the dependent variable, y_t, is typically assumed to be generated by the following equations:

$$
\begin{aligned}
y_t &= \mu_t(\boldsymbol{\theta}_0) + \sigma_t(\boldsymbol{\theta}_0)\xi_t, \\
\mu_t(\boldsymbol{\theta}) &= \mu(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}), \\
\sigma_t(\boldsymbol{\theta}) &= \sigma(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}),
\end{aligned}
\tag{1}
$$

where $\mu()$ and $\sigma()$ are two functions known up to the $p \times 1$ vector of true parameter values $\boldsymbol{\theta}_0$, $\mathbf{z}_t$ are k contemporaneous conditioning variables, I_{t-1} denotes the information set available at $t-1$, which contains past values of y_t and $\mathbf{z}_t$, and ξ_t is a martingale difference sequence satisfying $E(\xi_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = 0$ and $V(\xi_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = 1$. As a consequence, $E(y_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mu_t(\boldsymbol{\theta}_0)$ and $V(y_t | \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \sigma_t^2(\boldsymbol{\theta}_0)$.

The most common method of estimation for these models is a Gaussian pseudo-maximum likelihood procedure, in which the estimator, $\tilde{\boldsymbol{\theta}}_T$ say, is obtained by maximising the criterion function $\sum_{t=1}^{T} l_t(\boldsymbol{\theta})$, where $l_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma_t^2(\boldsymbol{\theta}) - \frac{1}{2}\xi_t^2(\boldsymbol{\theta})$, $\xi_t(\boldsymbol{\theta}) = [y_t - \mu_t(\boldsymbol{\theta})]/\sigma(\boldsymbol{\theta})$. This objective function becomes a proper log-likelihood function if and only if the distribution of ξ_t given $\mathbf{z}_t$, I_{t-1} and $\boldsymbol{\theta}_0$ is $N(0,1)$. However, an important property of $\tilde{\boldsymbol{\theta}}_T$ is that it remains root-T consistent with a limiting Gaussian distribution when the conditional mean and variance functions are correctly specified, even though the assumption of conditional normality may be violated (see Bollerslev and Wooldridge (1992)). The proof is based on the fact that under correct specification of $\mu_t(\boldsymbol{\theta})$ and $\sigma_t(\boldsymbol{\theta})$, the pseudo log-likelihood score, $\mathbf{s}_t(\boldsymbol{\theta}) = \partial l_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, becomes a vector martingale difference sequence when evaluated at $\boldsymbol{\theta}_0$.

Despite this property, empirical researchers routinely apply the so-called Jarque-Bera (JB) normality test to (1) in order to assess if the conditional distribution of the observed series is indeed normal (see Jarque and Bera (1980) and Bera and Jarque (1981)). Their test was originally developed for the special case in which the conditional variance is constant ($= \omega$, say) and does not affect $\mu_t(\boldsymbol{\theta})$, and the conditional mean parameters, $\boldsymbol{\delta}$ say, and ω are variation free. It is based on the

following statistics:

$$JB_T^N = \left(JB_T^S\right)^2 + \left(JB_T^K\right)^2,$$

$$JB_T^S = \sqrt{\frac{T}{6}}\left[m_{3T}(\tilde{\boldsymbol{\theta}}_T) - 3m_{1T}(\tilde{\boldsymbol{\theta}}_T)\right],$$

$$JB_T^K = \sqrt{\frac{T}{24}}\left[m_{4T}(\tilde{\boldsymbol{\theta}}_T) - 3\right],$$

where

$$m_{jT}(\tilde{\boldsymbol{\theta}}_T) = \frac{1}{T}\sum_{t=1}^{T}\xi_t^j(\tilde{\boldsymbol{\theta}}_T)$$

is the jth non-central empirical moment of the estimated standardised innovations $\xi_t^j(\tilde{\boldsymbol{\theta}}_T)$. As noted by these authors, $m_{1T}(\tilde{\boldsymbol{\theta}}_T) = 0 \ \forall T$ if the regression function $\mu_t(\boldsymbol{\theta})$ includes a constant term, in which case the expression for JB_T^S simplifies slightly to:

$$JB_T^{S'} = \sqrt{\frac{T}{6}}m_{3T}(\tilde{\boldsymbol{\theta}}_T), \tag{2}$$

which is the formula presented in many Econometrics textbooks for the special case in which $\mu_t(\boldsymbol{\theta})$ is constant.

The JB test was formally derived as a Lagrange Multiplier (LM) test of normality of the regression residuals versus the alternative that the (conditional) distribution of ξ_t belongs to the Pearson family. A closely related test was proposed by Kiefer and Salmon (1983) (KS), who developed an LM test for normality against a Hermite polynomial expansion of the (conditional) density of ξ_t. If one concentrates on the first four terms of such an expansion, this alternative test is based on the following statistics:

$$KS_T^N = \left(KS_T^S\right)^2 + \left(KS_T^K\right)^2,$$

$$KS_T^S = JB_T^S = JB_T^{S'} - \sqrt{\frac{3T}{2}}m_{1T}(\tilde{\boldsymbol{\theta}}_T)$$

$$KS_T^K = \sqrt{\frac{T}{24}}\left[m_{4T}(\tilde{\boldsymbol{\theta}}_T) - 6m_{2T}(\tilde{\boldsymbol{\theta}}_T) + 3\right] = JB_T^K - \sqrt{\frac{3T}{2}}\left[m_{2T}(\tilde{\boldsymbol{\theta}}_T) - 1\right]$$

If the conditional variance is constant, *and* there is functional independence between conditional mean and variance parameters, as assumed by all these authors, then we have that $m_{2T}(\tilde{\boldsymbol{\theta}}_T) = 1 \ \forall T$, so that both tests *numerically* coincide. Under

these special maintained assumptions, it can also be shown that the joint asymptotic distribution of KS_T^S and KS_T^K is spherical Gaussian when ξ_t given $\mathbf{z}_t, I_{t-1}$ and $\boldsymbol{\theta}_0$ is $N(0,1)$, which implies that both KS_T^N and JB_T^N have a null asymptotic chi-square distribution with 2 degrees of freedom.

2 The main results

As mentioned by Davidson and MacKinnon (1993) in section 16.7 of their textbook, while the asymptotic distribution of KS_T^S and KS_T^K remains valid in models in which $\sigma_t^2(\boldsymbol{\theta})$ is not constant, the same is not necessarily true of JB_T^K. Intuitively, the reason is that if we regard KS_T^K as a moment test based on the condition

$$E\left\{\left[\xi_t^4(\boldsymbol{\theta}) - 3\right] - 6\left[\xi_t^2(\boldsymbol{\theta}) - 1\right]\right\} = 0,$$

the inclusion of the term $-6\left[\xi_t^2(\boldsymbol{\theta}) - 1\right]$ is precisely what makes KS_T^K orthogonal to all the elements of the Gaussian pseudo-ML score $\mathbf{s}_t(\boldsymbol{\theta})$. The same point has recently been made rather forcefully by Bontemps and Meddahi (2002), who formally prove this result in the more general case in which there is no separation between the parameters affecting the mean and variance functions, and $\xi_t(\boldsymbol{\theta})$ is evaluated at some root-T consistent estimator of $\boldsymbol{\theta}$, possibly different from the pseudo-ML estimator $\tilde{\boldsymbol{\theta}}_T$.

Nevertheless, the exclusion from JB_T^K of the additional term

$$\sqrt{\frac{3T}{2}}\left[m_{2T}(\tilde{\boldsymbol{\theta}}_T) - 1\right] \tag{3}$$

does not necessarily lead to asymptotic size distortions when it is not identically zero. In particular, there will be no size distortions if (3) is $o_p(1)$. The following result establishes a necessary and sufficient condition for this to happen:

Proposition 1 *Under the null of conditional Gaussianity*

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left[\xi_t^2(\tilde{\boldsymbol{\theta}}_T) - 1\right] = o_p(1)$$

if and only if

$$\xi_t^2(\boldsymbol{\theta}_0) - 1 = w_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0)'\mathbf{s}_t(\boldsymbol{\theta}_0),$$

3

where

$$w_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \mathcal{I}^{-1}(\boldsymbol{\theta}_0) E \left\{ \left[\xi_t^2(\boldsymbol{\theta}_0) - 1 \right] \mathbf{s}_t(\boldsymbol{\theta}_0) \middle| \boldsymbol{\theta}_0 \right\},$$

and

$$\mathcal{I}(\boldsymbol{\theta}_0) = E \left[\mathbf{s}_t(\boldsymbol{\theta}_0) \mathbf{s}_t'(\boldsymbol{\theta}_0) \middle| \boldsymbol{\theta}_0 \right]$$

is the asymptotic information matrix.

Given that the above condition involves a rather complicated system of nonlinear differential equations, it is not possible to explicitly characterise which models for $\mu_t(\boldsymbol{\theta})$ and $\sigma_t^2(\boldsymbol{\theta})$ will satisfy it, so one has to proceed on a model by model basis. It turns out that such a condition is satisfied for the family of GARCH-M models analyzed by Hentschel (1995). More formally,

Proposition 2 *Consider the following stochastic process for y_t:*

$$y_t = \mu_t(\boldsymbol{\theta}_0) + \sigma_t(\boldsymbol{\theta}_0)\xi_t$$

$$\mu_t(\boldsymbol{\theta}) = \pi + \gamma \sigma_t^2(\boldsymbol{\theta})$$

$$\left. \begin{array}{l} \sigma_t^2(\boldsymbol{\theta}) = p_t^{2/\lambda}(\boldsymbol{\theta}) \\ p_t(\boldsymbol{\theta}) = (\lambda\omega + 1 - \beta) + p_{t-1} \left\{ \lambda\alpha f^v \left[\xi_{t-1}(\boldsymbol{\theta}) \right] + \beta \right\} \end{array} \right\} \ if \ \lambda \neq 0$$

$$\ln \sigma_t^2(\boldsymbol{\theta}) = 2\omega + 2\alpha f^v \left[\xi_{t-1}(\boldsymbol{\theta}) \right] + \beta \ln \sigma_{t-1}^2(\boldsymbol{\theta}) \ if \ \lambda = 0$$

$$f \left[\xi_{t-1}(\boldsymbol{\theta}) \right] = \sqrt{\left[\xi_{t-1}(\boldsymbol{\theta}) - b \right]^2 + \Delta} - c \left[\xi_{t-1}(\boldsymbol{\theta}) - b \right]$$

$$\xi_t | I_{t-1} \sim N(0, 1).$$

Then, the condition in Proposition 1 is satisfied in the limit as $\Delta \to 0$.

In this respect, note that Δ is simply a small positive number used to approximate the absolute value function by means of a rotated hyperbola, so that $\sigma_t^2(\boldsymbol{\theta})$ is everywhere differentiable, including at $\xi_{t-1}(\boldsymbol{\theta}) = b$.

Hentschel's (1995) family of models is remarkably rich, and nests many popular examples in the literature, including the standard GARCH-M ($\lambda = v = 2$, $b = c = 0$), the NAGARCH-M ($\lambda = v = 2$, $b \neq 0$, $c = 0$), the GJR GARCH-M ($\lambda = v = 2$, $c \neq 0$, $b = 0$), the APARCH-M ($\lambda = v \neq 0$, $b = 0$, $|c| \leq 1$), the absolute value GARCH-M ($\lambda = v = 1$, $b \neq 0$, $|c| \leq 1$) and the EGARCH-M model ($\lambda = 0, v = 1$)). In

contrast, the Quadratic GARCH-M model of Sentana (1995) cannot be nested in his framework. Nevertheless, an argument similar to the one used to prove Proposition 2 shows that (3) is also $o_p(1)$ in that case.

3   Conclusions

We have shown that the JB normality test, originally devised for constant conditional variance models with no functional dependence between conditional mean and variance parameters, can be safely applied to the broad class of GARCH-M models discussed by Hentschel (1995), as well as to the Quadratic GARCH-M model of Sentana (1995). Nevertheless, apart from the obvious situation in which $\Delta > 0$, it is possible to find examples of other ARCH models in which such a condition is not satisfied (for instance, the symmetric variant of the EGARCH model proposed in chapter 13 of Barndorf-Nielsen and Shephard (2001), in which $f\left[\xi_{t-1}(\boldsymbol{\theta})\right]$ effectively takes the form $\Phi^{-1}\left\{F_1\left[\xi_{t-1}^2(\boldsymbol{\theta})\right]\right\}$, where $F_1(.)$ is the cumulative distribution function (cdf) of a chi-square random variable with one degree of freedom, and $\Phi^{-1}(.)$ is the inverse cdf of a standard normal).

In addition, we can combine the expression for $\mathbf{s}_t(\boldsymbol{\theta})$ in the Appendix with the analogue of Proposition 1 for $\xi_t(\boldsymbol{\theta})$ to prove that the asymptotic distribution of $\sqrt{T}m_{1T}(\tilde{\boldsymbol{\theta}}_T)$ will not be $o_p(1)$ when $\sigma_t^2(\boldsymbol{\theta})$ is time-varying. As a result, the test statistic $JB_T^{S'}$ in (2) will be incorrectly sized in Hentschel's model despite the fact that $\mu_t(\boldsymbol{\theta})$ includes a constant term.

Therefore, our recommendation would be to use the version proposed by KS despite the fact that the asymptotic size of the JB normality test for regression residuals commonly employed by practitioners is often correct, because the limiting null distribution of KS_T^N never depends on the particular parametrisation used, and the additional computational cost is negligible.

5

References

Arellano, M. (1991): "Moment testing with non-ML estimators", mimeo, CEMFI.

Barndorf-Nielsen, O. E. and Shephard, N. (2001): *Lévy-based dynamic models for financial economics*, in preparation, Nuffield College, Oxford.

Bera, A.K. and Jarque, C.M. (1981): "An efficient large-sample test for normality of observations and regression residuals", Australian National University Working Paper in Economics and Econometrics 40.

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.

Bontempts, C. and Meddahi, N. (2002): "Testing normality: a GMM approach", CIRANO Working Paper 2002s-63.

Davidson R. and MacKinnon, J.G. (1993): *Estimation and inference in econometrics*, Oxford University Press, Oxford.

Hentschel, L. (1995): "All in the family: nesting symmetric and asymmetric GARCH models", *Journal of Financial Economics* 39, 71-104.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economic Letters*, 6, 255-259.

Kiefer, N.M. and Salmon, M. (1983): "Testing normality in econometric models", *Economic Letters* 11, 123-127.

Sentana, E. (1995): "Quadratic ARCH models", *Review of Economic Studies* 62, 639-661.

Appendix

Proofs

Proposition 1

A straightforward application of the results in Arellano (1991) implies that

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left[\xi_t^2(\tilde{\boldsymbol{\theta}}_T)-1\right] = \frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left[\xi_t^2(\boldsymbol{\theta}_0)-1+\Phi'(\boldsymbol{\theta}_0)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{s}_t(\boldsymbol{\theta}_0)\right]+o_p(1),$$

where

$$\Phi(\boldsymbol{\theta}_0) = E\left[\left.\frac{\partial \xi_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right|\boldsymbol{\theta}_0\right] = -E\left\{\left[\xi_t^2(\boldsymbol{\theta}_0)-1\right]\mathbf{s}_t(\boldsymbol{\theta}_0)\middle|\boldsymbol{\theta}_0\right\}.$$

But since

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\xi_t^2(\boldsymbol{\theta}_0)-1\\\mathbf{s}_t(\boldsymbol{\theta}_0)\end{array}\right] \xrightarrow{d} N\left\{\left(\begin{array}{c}0\\\mathbf{0}\end{array}\right),\left[\begin{array}{cc}2 & -\Phi'(\boldsymbol{\theta}_0)\\-\Phi(\boldsymbol{\theta}_0) & \mathcal{I}(\boldsymbol{\theta}_0)\end{array}\right]\right\}$$

under the null of conditional normality, then (3) will be $o_p(1)$ if and only if $\xi_t^2(\boldsymbol{\theta}_0)-1$ can be written as an exact linear combination of $\mathbf{s}_t(\boldsymbol{\theta}_0)$.

Proposition 2

We prove first the general case in which $\lambda \neq 0$ and $\lambda \neq \upsilon$. Later on, we prove those special cases. As a by-product, we also provide analytical expressions for the derivatives of the conditional mean and variance functions in Hentschel's model with respect to the different parameters, which can be used to speed up the computations and provide more reliable standard errors and test statistics.

The general case

In order to obtain the Gaussian pseudo log-likelihood score, we need the derivatives of the conditional mean and variance functions with respect to the nine-dimensional parameter vector $\boldsymbol{\theta} = (\pi, \gamma, \omega, \alpha, b, c, \beta, \upsilon, \lambda)'$. In this respect, it is important to note that the partial derivatives of $\mu_t(\boldsymbol{\theta})$ with respect to all the parameters except π and γ will be given by the following expression:

$$\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \theta_j} = \gamma \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_j} = \gamma \sigma_t^2 \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_j} \quad j = 3, \dots 9,$$

while the partial derivatives with respect to π and γ will be

$$
\begin{aligned}
\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \pi} &= 1 + \gamma \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \pi} = 1 + \gamma \sigma_t^2(\boldsymbol{\theta}) \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \pi}, \\
\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \gamma} &= \sigma_t^2(\boldsymbol{\theta}) + \gamma \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \gamma} = \sigma_t^2(\boldsymbol{\theta}) + \gamma \sigma_t^2(\boldsymbol{\theta}) \frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \gamma}.
\end{aligned}
$$

Similarly, the partial derivatives of $\sigma_t^2(\boldsymbol{\theta})$ with respect to $\pi, \gamma, \omega, \alpha, b, c, \beta$ and υ will be given by the expression

$$
\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_j} = \frac{2}{\lambda} p_t^{\frac{2}{\lambda}-1}(\boldsymbol{\theta}) \frac{\partial p_t(\boldsymbol{\theta})}{\partial \theta_j} = \frac{2}{\lambda} \sigma_t^2(\boldsymbol{\theta}) \frac{\partial p_t(\boldsymbol{\theta})}{\partial \theta_j} \frac{1}{p_t(\boldsymbol{\theta})} \quad j = 1, \ldots 8,
$$

so that

$$
\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_j} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} = \frac{2}{\lambda} \frac{\partial \ln p_t(\boldsymbol{\theta})}{\partial \theta_j} \quad j = 1, \ldots 8.
$$

However, the partial derivative with respect to λ will be given by

$$
\begin{aligned}
\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \lambda} &= \frac{2}{\lambda} p_t^{2/\lambda}(\boldsymbol{\theta}) \left(-\frac{1}{\lambda} \ln \left[p_t(\boldsymbol{\theta}) \right] + \frac{\partial p_t(\boldsymbol{\theta})}{\partial \lambda} \frac{1}{p_t(\boldsymbol{\theta})} \right) \\
&= \frac{2}{\lambda} \sigma_t^2(\boldsymbol{\theta}) \left(-\frac{1}{\lambda} \ln \left[p_t(\boldsymbol{\theta}) \right] + \frac{\partial p_t(\boldsymbol{\theta})}{\partial \lambda} \frac{1}{p_t(\boldsymbol{\theta})} \right),
\end{aligned}
$$

from where

$$
\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \lambda} = \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \lambda} \frac{1}{\sigma_t^2(\boldsymbol{\theta})} = \frac{2}{\lambda} \left(-\frac{1}{\lambda} \ln \left[p_t(\boldsymbol{\theta}) \right] + \frac{\partial \ln p_t(\boldsymbol{\theta})}{\partial \lambda} \right).
$$

Now, if we exploit the fact that $\sigma_t^\lambda(\boldsymbol{\theta}) = p_t(\boldsymbol{\theta})$, and

$$
\frac{\partial \ln \sigma_{t-1}(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial \sigma_{t-1}(\boldsymbol{\theta})}{\partial \theta_j} \frac{1}{\sigma_{t-1}(\boldsymbol{\theta})} = \frac{1}{2} \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \theta_j} \quad \forall j,
$$

so that

$$
\lambda \sigma_{t-1}^\lambda(\boldsymbol{\theta}) \frac{\partial \sigma_{t-1}(\boldsymbol{\theta})}{\partial \theta_j} \frac{1}{\sigma_{t-1}(\boldsymbol{\theta})} = \frac{1}{2} \lambda p_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \theta_j} \quad \forall j,
$$

we will have that

$$
\frac{\partial p_t(\boldsymbol{\theta})}{\partial \pi} = -\alpha \lambda \upsilon p_{t-1}(\boldsymbol{\theta}) f^{\upsilon-1} \left[\xi_{t-1}(\boldsymbol{\theta}) \right] f' \left[\xi_{t-1}(\boldsymbol{\theta}) \right] \frac{1}{\sigma_{t-1}(\boldsymbol{\theta})} + g_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \pi},
$$

where

$$
g_{t-1}(\boldsymbol{\theta}) = \frac{1}{2} \lambda p_{t-1}(\boldsymbol{\theta}) \left[\begin{array}{c} \alpha \lambda f^\upsilon \left[\xi_{t-1}(\boldsymbol{\theta}) \right] + \beta \\ -\alpha \upsilon \left[\xi_{t-1}(\boldsymbol{\theta}) + 2\gamma \sigma_{t-1}(\boldsymbol{\theta}) \right] f^{\upsilon-1} \left[\xi_{t-1}(\boldsymbol{\theta}) \right] f' \left[\xi_{t-1}(\boldsymbol{\theta}) \right] \end{array} \right],
$$

8

and

$$f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] = \frac{\xi_{t-1}(\boldsymbol{\theta}) - b}{\sqrt{\left[\xi_{t-1}(\boldsymbol{\theta}) - b\right]^2 + \Delta}} - c.$$

In this respect, note that the non-differentiability of the absolute value function is reflected in the fact that

$$\lim_{\Delta \to 0} f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] = \begin{cases} 1 - c & \text{if} \quad \xi_{t-1}(\boldsymbol{\theta}) - b > 0 \\ -c & \text{if} \quad \xi_{t-1}(\boldsymbol{\theta}) - b = 0 \\ -1 - c & \text{if} \quad \xi_{t-1}(\boldsymbol{\theta}) - b < 0 \end{cases}.$$

Similarly

$$\frac{\partial p_t(\boldsymbol{\theta})}{\partial \gamma} = -\alpha \lambda \upsilon p_{t-1}(\boldsymbol{\theta}) f^{\upsilon-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] \sigma_{t-1}(\boldsymbol{\theta}) + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \gamma},$$

$$\begin{aligned}
\frac{\partial p_t(\boldsymbol{\theta})}{\partial \omega} &= \lambda + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \omega}, \\
\frac{\partial p_t(\boldsymbol{\theta})}{\partial \alpha} &= \lambda p_{t-1}(\boldsymbol{\theta}) f^{\upsilon}\left[\xi_{t-1}(\boldsymbol{\theta})\right] + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \alpha}, \\
\frac{\partial p_t(\boldsymbol{\theta})}{\partial b} &= -\alpha \upsilon \lambda p_{t-1}(\boldsymbol{\theta}) f^{\upsilon-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial b}, \\
\frac{\partial p_t(\boldsymbol{\theta})}{\partial c} &= -\alpha \upsilon \lambda \left[\xi_{t-1}(\boldsymbol{\theta}) - b\right] p_{t-1}(\boldsymbol{\theta}) f^{\upsilon-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial c}, \\
\frac{\partial p_t(\boldsymbol{\theta})}{\partial \beta} &= p_{t-1}(\boldsymbol{\theta}) - 1 + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \beta}.
\end{aligned}$$

The derivatives with respect to υ and λ are slightly trickier

$$\begin{aligned}
\frac{\partial p_t(\boldsymbol{\theta})}{\partial \upsilon} &= p_{t-1}(\boldsymbol{\theta}) \alpha \lambda f^{\upsilon}\left[\xi_{t-1}(\boldsymbol{\theta})\right] \ln f\left[\xi_{t-1}(\boldsymbol{\theta})\right] + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \upsilon}, \\
\frac{\partial p_t(\boldsymbol{\theta})}{\partial \lambda} &= \omega + p_{t-1}(\boldsymbol{\theta})\left[\ln\left[p_{t-1}(\boldsymbol{\theta})\right] \cdot \left\{\alpha \lambda f^{\upsilon}\left[\xi_{t-1}(\boldsymbol{\theta})\right] + \beta\right\}/\lambda + \alpha f^{\upsilon}\left[\xi_{t-1}(\boldsymbol{\theta})\right]\right] \\
&\quad + g_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \lambda}.
\end{aligned}$$

According to Proposition 1, we need to find a time-invariant linear combination of the score that equals $\xi_t^2(\boldsymbol{\theta}) - 1$ for all t in order for the kurtosis component of the usual Jarque-Bera test to show no size distortions. But since

$$\mathbf{s}_t(\boldsymbol{\theta}) = \frac{1}{\sigma_t(\boldsymbol{\theta})}\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\xi_t(\boldsymbol{\theta}) + \frac{1}{2}\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left[\xi_t^2(\boldsymbol{\theta}) - 1\right],$$

and both $\mu_t(\boldsymbol{\theta})$ and $\sigma_t^2(\boldsymbol{\theta})$ depend exclusively on information known at time $t-1$, it must be the case that such a linear combination, provided that it exists, will satisfy the following system of simultaneous partial differential equations:

$$\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0,$$
$$\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 1,$$

where $\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is a vector of time-invariant weights, which can, nevertheless, be arbitrary functions of $\boldsymbol{\theta}$. But since

$$\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = w_\pi(\boldsymbol{\theta}) + w_\gamma(\boldsymbol{\theta})\sigma_t^2(\boldsymbol{\theta}) + \gamma \sigma_t^2(\boldsymbol{\theta})\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

it is easy to verify that if we choose $w_\pi(\boldsymbol{\theta}) = 0$ and $w_\gamma(\boldsymbol{\theta}) = -\gamma$, the first condition will be satisfied provided the second one is. In this respect, we shall next prove that the linear combination:

$$\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = (0, -\gamma, \frac{\lambda\omega + 1 - \beta}{2}, \frac{\alpha v}{2}, -\frac{b}{2}, 0, 0, 0, 0)$$

will become arbitrarily close to satisfying such a requirement as $\Delta \to 0$.

We will do so recursively, by proving that $\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\partial \ln \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 1$ if the same linear combination was 1 in $t-1$, under the implicit assumption that $\ln \sigma_0^2(\boldsymbol{\theta})$ has been parametrised in such a way that $\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\partial \ln \sigma_0^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 1$. But first, let us look at the linear combination $\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\partial p_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$, which will be given by

$$\gamma\alpha\lambda v p_{t-1}(\boldsymbol{\theta})f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right]f'\left[\xi_{t-1}(\boldsymbol{\theta})\right]\sigma_{t-1}(\boldsymbol{\theta})$$
$$+\frac{\lambda}{2}\left\{\begin{array}{c}\lambda\omega + 1 - \beta + \alpha v p_{t-1}(\boldsymbol{\theta})f^v\left[\xi_{t-1}(\boldsymbol{\theta})\right] \\ +b\alpha v p_{t-1}(\boldsymbol{\theta})f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right]f'\left[\xi_{t-1}(\boldsymbol{\theta})\right]\end{array}\right\} + g_{t-1}(\boldsymbol{\theta})$$
$$= \frac{\lambda}{2}\left\{\lambda\omega + 1 - \beta + p_{t-1}(\boldsymbol{\theta})\left[\alpha\lambda f^v\left[\xi_{t-1}(\boldsymbol{\theta})\right] + \beta\right]\right\}$$
$$+\frac{\lambda}{2}p_{t-1}(\boldsymbol{\theta})\alpha v f^{v-1}\left(\xi_{t-1}\right)\left\{f^v\left[\xi_{t-1}(\boldsymbol{\theta})\right] - \left(\left[\xi_{t-1}(\boldsymbol{\theta})\right] - b\right)f'\left[\xi_{t-1}(\boldsymbol{\theta})\right]\right\}.$$

Given that
$$f\left[\xi_{t-1}(\boldsymbol{\theta})\right] - \left[\xi_{t-1}(\boldsymbol{\theta}) - b\right]f'\left[\xi_{t-1}(\boldsymbol{\theta})\right]$$
converges to 0 as $\Delta \to 0$ regardless of the sign of $\xi_{t-1}(\boldsymbol{\theta}) - b$, then the above expression can be re-written in the limit as:

$$\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial p_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\lambda}{2}\left\{\lambda\omega + 1 - \beta + p_{t-1}(\boldsymbol{\theta})\left[\alpha\lambda f^v\left[\xi_{t-1}(\boldsymbol{\theta})\right] + \beta\right]\right\} = \frac{\lambda}{2}p_t(\boldsymbol{\theta}),$$

which means that $\mathbf{w}'_{\boldsymbol{\theta}}\partial \ln p_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = \lambda/2$. But since $w_\lambda(\boldsymbol{\theta}) = 0$ in our case, then

$$\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{\lambda}\mathbf{w}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\frac{\partial \ln p_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 1$$

as required. $\qquad\square$

The case $v = \lambda$

In this context, a straightforward application of the chain rule yields

$$\frac{\partial p_t(\boldsymbol{\theta})}{\partial \lambda} = \omega + p_{t-1}(\boldsymbol{\theta}) \left\{ \begin{array}{c} \ln\left[p_{t-1}(\boldsymbol{\theta})\right] \cdot \left\{ \alpha \lambda f^\lambda \left[\xi_{t-1}(\boldsymbol{\theta})\right] + \beta \right\} / \lambda \\ + \alpha f^\lambda \left[\xi_{t-1}(\boldsymbol{\theta})\right] + \alpha \lambda f^v \left[\xi_{t-1}(\boldsymbol{\theta})\right] \ln f \left[\xi_{t-1}(\boldsymbol{\theta})\right] \end{array} \right\}$$
$$+ g_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \lambda},$$

where $g_{t-1}(\boldsymbol{\theta})$ must be evaluated at $v = \lambda$.

But given that in the previous section we have proved that in general $w_v = w_\lambda = 0$, then it is clear that in this case we can achieve the same outcome with $w_\lambda = 0$ alone. $\square$

The case $\lambda = 0$

The only remaining example to analyze is the limiting case of $\lambda \to 0$, which converges to an EGARCH-M model for $v = 1$. More specifically, the log of the conditional variance function will be given by

$$\ln \sigma_t^2(\boldsymbol{\theta}) = 2\omega + 2\alpha f^v \left[\xi_{t-1}(\boldsymbol{\theta})\right] + \beta \ln \sigma_{t-1}^2(\boldsymbol{\theta}).$$

Tedious but otherwise straightforward algebra shows that in this case, the appropriate linear combination will be given by the vector

$$\mathbf{w}_\theta' = \left(0, -\gamma, \frac{1-\beta}{2}, \frac{\alpha v}{2}, -\frac{b}{2}, 0, 0, 0 \right),$$

which is precisely the limit as $\lambda \to 0$ of the vector obtained for the general case.

To prove it formally, let us obtain all the required derivatives. In particular, if we call

$$h_{t-1}(\boldsymbol{\theta}) = \beta - \alpha v \left[\xi_{t-1}(\boldsymbol{\theta}) + 2\gamma \sigma_{t-1}(\boldsymbol{\theta})\right] f^{v-1} \left[\xi_{t-1}(\boldsymbol{\theta})\right] f' \left[\xi_{t-1}(\boldsymbol{\theta})\right],$$

then

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \pi} = -2\alpha v f^{v-1} \left[\xi_{t-1}(\boldsymbol{\theta})\right] f' \left[\xi_{t-1}(\boldsymbol{\theta})\right] \frac{1}{\sigma_{t-1}(\boldsymbol{\theta})} + h_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \pi},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \gamma} = -2\alpha v f^{v-1} \left[\xi_{t-1}(\boldsymbol{\theta})\right] f' \left[\xi_{t-1}(\boldsymbol{\theta})\right] \sigma_{t-1}(\boldsymbol{\theta}) + h_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \gamma},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \omega} = 2 + h_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \omega},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \alpha} = 2 f^v \left[\xi_{t-1}(\boldsymbol{\theta})\right] + h_{t-1}(\boldsymbol{\theta}) \frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \alpha},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial b} = -2\alpha v f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] + h_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial b},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial c} = -2\alpha v \left[\xi_{t-1}(\boldsymbol{\theta}) - b\right] f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] + h_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial c},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \beta} = \ln \sigma_{t-1}^2(\boldsymbol{\theta}) + h_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \beta},$$

$$\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial v} = 2\alpha f^v \left[\xi_{t-1}(\boldsymbol{\theta})\right] \ln f \left[\xi_{t-1}(\boldsymbol{\theta})\right] + h_{t-1}(\boldsymbol{\theta})\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial v}.$$

Hence,

$$\mathbf{w}_{\boldsymbol{\theta}}'\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\gamma\alpha v f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] \sigma_{t-1}(\boldsymbol{\theta}) + 1 - \beta + \alpha v f^v \left[\xi_{t-1}(\boldsymbol{\theta})\right]$$

$$+ \alpha v b f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right] f'\left[\xi_{t-1}(\boldsymbol{\theta})\right] + h_{t-1}(\boldsymbol{\theta})\mathbf{w}_{\boldsymbol{\theta}}'\frac{\partial \ln \sigma_{t-1}^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

But if we assume once more that $\ln \sigma_0^2(\boldsymbol{\theta})$ has been parametrised in such a way that $\mathbf{w}_{\boldsymbol{\theta}}'(\boldsymbol{\theta})\partial \ln \sigma_0^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 1$, then we are left with:

$$\mathbf{w}_{\boldsymbol{\theta}}'\frac{\partial \ln \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 1 + \alpha v f^{v-1}\left[\xi_{t-1}(\boldsymbol{\theta})\right]\left\{f\left[\xi_{t-1}(\boldsymbol{\theta})\right] - \left[\xi_{t-1}(\boldsymbol{\theta}) - b\right] f'\left[\xi_{t-1}(\boldsymbol{\theta})\right]\right\},$$

which again goes to 1 as $\Delta \to 0$ regardless of the sign of $\xi_{t-1}(\boldsymbol{\theta}) - b$. $\qquad\square$