Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown

Sucarrat, Genaro and Grønneberg, Steffen and Escribano, Alvaro

BI Norwegian Business School, BI Norwegian Business School, Universidad Carlos III de Madrid

11 August 2013

Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown*

Genaro Sucarrat[†], Steffen Grønneberg[‡] and Álvaro Escribano[§]

11 August 2013

Abstract

Exponential models of Autoregressive Conditional Heteroscedasticity (ARCH) enable richer dynamics (e.g. contrarian or cyclical), provide greater robustness to jumps and outliers, and guarantee the positivity of volatility. The latter is not guaranteed in ordinary ARCH models, in particular when additional exogenous or predetermined variables ("X") are included in the volatility specification. Here, we propose estimation and inference methods for univariate and multivariate Generalised log-ARCH-X (i.e. log-GARCH-X) models when the conditional density is not known via (V)ARMA-X representations. The multivariate specification allows for volatility feedback across equations, and time-varying correlations can be fitted in a subsequent step. Finally, our empirical applications on electricity prices show that the model-class is particularly useful when the X-vector is high-dimensional.

JEL Classification: C22, C32, C51, C52
Keywords: ARCH, exponential GARCH, log-GARCH, ARMA-X, Multivariate GARCH

*An earlier version of this paper was entitled "The Power Log-GARCH Model", see Sucarrat and Escribano (2010). We are grateful to Jonas Andersson, Luc Bauwens, Christian Francq, Andrew Harvey, Emma Iglesias, Sebastien Laurent, Enrique Sentana and seminar and conference participants at Université de Lille, BI Norwegian Business School, Universidad Carlos III de Madrid, BI Norwegian Business School (Oslo), IHS (Vienna), 2nd. Rimini Time Series Workshop (2013), 21st. SNDE Symposium (Milan, 2013), ESEM 2011 (Oslo), Interdisciplinary Workshop in Louvain-la-Neuve 2011, CFE conference 2010 (London), FIBE 2010 (Bergen), IWAP 2010 (Madrid) and Foro de Finanzas 2010 (Elche) for useful comments, suggestions and questions. Funding from the 6th. European Community Framework Programme, MICIN ECO2009-08308 and from The Bank of Spain Excellence Program is gratefully acknowledged.

[†]Corresponding author. Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Email genaro.sucarrat@bi.no, phone +47+46410779, fax +47+23264788. Webpage: http://www.sucarrat.net/

[‡]Department of Economics, BI Norwegian Business School. Email: steffeng@gmail.com.

[§]Department of Economics, Universidad Carlos III de Madrid (Spain)

1 Introduction

The Autoregressive Conditional Heteroscedasticity (ARCH) class of models due to Engle (1982) is useful in a wide range of empirical applications. In finance in particular, it has been extensively used to model the clustering of large (in absolute value) financial returns. Engle (1982) himself, however, originally motivated the class as useful in modelling the time-varying conditional uncertainty (i.e. conditional variance) of economic variables in general, and of UK inflation in particular. Other areas of application include, amongst other, temperature data (e.g. Franses et al. (2001)) and electricity prices (e.g. Koopman et al. (2007)). More generally, ARCH models can also be used to improve the estimation and inference in the conditional mean by means of Feasible Generalised Least Squares (FGLS) methods (Hamilton (2010)). For recent surveys of ARCH models, see Francq and Zakoïan (2010a) and Terasvirta (2009, 2012).

Within the ARCH class of models exponential versions are of special interest. This is because they enable richer autoregressive volatility dynamics (e.g. contrarian or cyclical) compared with non-exponential ARCH models, and because their fitted values of volatility are guaranteed to be positive. The latter is not necessarily the case for ordinary (i.e. non-exponential) ARCH models, in particular when additional exogenous or predetermined variables ("X") are included in the volatility equation. In fact, the greater the dimension of X, the more restrictions are needed in order to ensure positivity. Another desirable property, which is shared by the log-GARCH model and the Beta-t-EGARCH model of Harvey (2013) but not by Nelson's (1991) EGARCH (nor by ordinary GARCH models), is that volatility forecasts are more robust to jumps and outliers (this is illustrated in the first empirical application in section 4). Robustness is important in order to avoid volatility forecast failure subsequent to jumps and outliers. The log-GARCH class of models can be viewed as a dynamic version of Harvey's (1976) multiplicative heteroscedasticity model,[1] and was first proposed independently by Pantula (1986), Geweke (1986) and Milhøj (1987). Engle and Bollerslev (1986) argued against log-ARCH models because of the possibility of applying the log-operator (in the log-ARCH terms) on zero-values, which occurs whenever the error term in a regression equals zero. A solution to this problem, however, is provided in Sucarrat and Escribano (2013) for the case where the zero-probability is zero (e.g. because zeros are due to discreteness or missing values), and in Sucarrat (2013) for the case where the zero-probability is not zero and possibly time-varying. Another issue that has been cited in the literature (e.g. Teräsvirta (2009)), is that the first unconditional autocorrelations of the squared errors – a measure of volatility persistence – can be unreasonably high. But this only occurs in very specific cases. The log-GARCH class allows for a much larger range of autocorrelation patterns than ordinary GARCH models (see table 1), since the autocorrelation pattern depends on the shape of the conditional density (the more fat-tailed, the lower correlations) in addition to the persistence parameters.

[1] In some statistical softwares, e.g. JMP (2013), the multiplicative heteroscedasticity model is referred to as the log-variance model.

The assumption that the conditional density is unknown is particularly convenient from a practitioner's point of view, since the user then does not need to worry about changing the conditional density from application to application, or alternatively to work with a sufficiently general density that will often make estimation and inference numerically more challenging. This explains the attraction of Quasi Maximum Likelihood Estimators (QMLEs). In the univariate case consistency and asymptotic normality of QMLE for ARCH models and a GARCH(1,1) model, respectively, were first proved by Weiss (1986) and Lumsdaine (1996), while Berkes et al. (2003) and Francq and Zakoïan (2004) established more general results under milder conditions. To the best of our knowledge, there are no results on GARCH-X models. For exponential ARCH models, by contrast, most of the univariate results are either very limited or hold under unreasonable assumptions. In the multivariate case, asymptotic QML results have been established for the BEKK model of by Comte and Lieberman (2003), for an ARMA-GARCH with constant conditional correlations (CCCs) by Ling and McAleer (2003), for a VEC model by Hafner and Preminger (2009), and for a multivariate GARCH with CCCs by Francq and Zakoïan (2010b).[2] For exponential ARCH models, by contrast, there are no multivariate results when the conditional density is unknown, and the univariate results are either very limited or hold under unreasonable assumptions. Straumann and Mikosch (2006, p. 2452) prove consistency of the Gaussian QML estimator for Nelson's (1991) univariate EGARCH(1,1). But the result of Straumann and Mikosch is limited in that it does not apply to higher order EGARCH models, nor to multivariate versions. Also, their result is limited to estimator consistency; they do not prove asymptotic normality. However, in a recent (but unpublished) paper Wintenberger (2012) provides sufficient conditions for asymptotic normality under the condition of continuous invertibility, and under restrictive assumptions on the parameter space. Kristensen and Rahbek (2009) prove that QML is consistent for a class of univariate non-linear ARCH models that includes the log-ARCH(P) family. But their result does not apply to models that includes log-GARCH terms, nor to multivariate versions. Recently, Francq et al. (2012) prove consistency and asymptotic normality of an asymmetric version of the univariate log-GARCH(P, Q) model. But methods for multivariate models are not put forward, and zero errors cannot be handled satisfactorily since estimation is not via the ARMA representation, see Sucarrat and Escribano (2013). Finally, Kawakatsu (2006) has proposed a multivariate exponential ARCH model, the matrix exponential GARCH, which contains a multivariate version of Nelson's 1991 model. But estimation and inference results for the case where the conditional density is unknown have yet to be provided, and general conditions for the existence of its unconditional error moments are not available.

This paper makes at least five contributions. The first and most important is a theoretical result (Theorem 1), which we believe will have important consequences for empirical practice. It is well-known that all the coefficients apart from the

[2] Jeantheau (1998) established general conditions for strong consistency for QML estimation of multivariate GARCH models. However, as pointed out by Ling and McAleer (2003, p. 281), his results are based on the unrealistic assumption that the initial values are known.

log-volatility intercept in a univariate log-GARCH specification can be estimated consistently (under suitable assumptions) via an ARMA representation, see for example Psaradakis and Tzavalis (1999), and Francq and Zakoïan (2006). However, the estimate of the log-volatility intercept will be asymptotically biased, and the bias is made up of a log-moment expression that depends on the unknown density of the conditional error. We propose a simple estimator of the log-moment expression made up of the empirical residuals of the ARMA regression, and prove its consistency and asymptotic normality under mild assumptions. The proof relies on a result by Yu (2007), which means that the result holds for a range of ARMA estimators, including the Gaussian QMLE. Moreover, asymptotically our estimator is as efficient as if the true errors were used instead of the residuals. Consequentially, we prove the consistency and asymptotic normality of univariate log-GARCH(P, Q) models via the ARMA(P, Q) representation.

The second contribution of the paper is a consequence of the first. The addition of exogenous or predermined conditioning variables yields the log-GARCH-X model, which has a corresponding ARMA-X representation (see Subsection 2.4). The relation between the ARMA coefficients and the log-GARCH coefficients are not affected by the additional X-variables, so the ARMA error term retains the same structure as before. Consistent estimation of the ARMA-X representation will thus produce exactly the same bias as earlier. Accordingly, a reasonable conjecture is that the bias correction procedure described above will also be valid for ARMA-X models. This means a vast literature of already established time-series results and practices is likely to be available for the further development and study of log-GARCH-X models.

In a third contribution (Section 3) we propose a multivariate log-GARCH-X model that admits time-varying conditional correlations. The model is truly multivariate – and not only a collection of separate single equation specifications – in that it admits volatility feedback across the equations, via both lagged log-ARCH and lagged log-GARCH terms. Also, since the positivity of the volatilites is guaranteed due to the exponential specification, restrictive assumptions are not needed in order to ensure the positive definiteness of the (possibly) time-varying covariance matrix of the errors. The multivariate log-GARCH-X model has a VARMA-X representation with a vector of error-terms. The vector is either IID, which corresponds to the Constant Conditional Correlation (CCC) case, or independent but non-identical (ID), which corresponds to the time-varying correlations case. But even in the latter case each entry in the vector of errors is marginally IID. So the bias-correction from the univariate case can be used equation-by-equation subsequent to the estimation of the VARMA-X representation. Again, due to the structure of the problem, the bias-correction procedure is likely to hold under mild assumptions.

A fourth contribution (Propositions 1 and 2) consists of proving that the unconditional moments of the error exist for a much larger class of densities than the EGARCH of Nelson (1991). The existence of these moments are unlikely to have a bearing upon the estimation and inference methods that we propose. However, they are important for economic interpretation and analysis when the errors are interpreted as (de-meaned) returns, as is often the case in finance. In exponential ARCH models the existence of unconditional error moments (i.e. de-meaned returns) depend on the shape of the conditional density, and the error moments of Nelson's (1991) EGARCH do generally not exist for t-distributed errors, see Nelson (1991, p. 365).[3] In fact, in the presence of volatility asymmetry (i.e. "leverage") the problem is exacerbated. We also provide sufficient conditions for the existence of the unconditional error moments in the multivariate case.

The fifth contribution of this paper comprises three empirical applications to electricity prices. In the first we illustrate the robustness of the model-class to large values (i.e. "outliers" or "jumps"), a common feature of electricity price returns. In the next two we illustrate the versatility and flexibility of the model class to accommodate a large number of exogenous or predetermined variables in the log-variance, for both univariate and multivariate specifications.

The rest of the paper is organised as follows. The next section, section 2, presents the univariate log-GARCH model, contains our main theoretical result (Theorem 1) and reports the results from a set of Monte Carlo simulations. The relation between the univariate log-GARCH model and its ARMA representation is also set out in this section, and it is showed that the addition of exogenous and predetermined variables does not alter this relation. Section 3 show how the ideas extend to the multivariate case, whereas section 4 contains the empirical applications. Section 5 concludes, whereas the subsequent appendices contain various supporting information, including proofs. Tables and figures are located at the end.

2 Univariate log-GARCH

2.1 Notation and specification

The univariate log-GARCH(P, Q) model is given by

$$\begin{aligned} \epsilon_t &= \sigma_t z_t, \quad z_t \sim IID(0,1), \quad Prob(z_t = 0) = 0, \quad \sigma_t > 0, & (1) \\ \ln \sigma_t^2 &= \alpha_0 + \sum_{p=1}^{P} \alpha_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \ln \sigma_{t-q}^2, \quad t \in \mathbb{Z}, & (2) \end{aligned}$$

where P is the ARCH order and Q is the GARCH order. In finance, ϵ_t is often interpreted as return or de-meaned return, but more generally it is simply the error in a regression model. Denoting $P^* = \max\{P, Q\}$, if the roots of the lag polynomial $1 - (\alpha_1 + \beta_1)L - \cdots - (\alpha_{P^*} + \beta_{P^*})L^{P^*}$ are all greater than 1 in modulus and if $|\mathbb{E}(\ln z_t^2)| < \infty$, then $\{\ln \sigma_t^2\}$ is stable. For common densities like the Student's t with degrees of freedom greater than 2, and the Generalised Error Distribution (GED) with shape parameter greater than 1, then ϵ_t will generally be stable as well if $\ln \sigma_t^2$ is stable, see propositions 1 and 2. Practitioners are often interested in the

[3] The exponential versions of the models of Engle and Maricucci (2006, equation (16) on page 14), and of Hansen et al. (2012), are likely to be affected in a smiliar way, since they include ARCH terms similar to those of Nelson.

dynamics of other powers than the 2nd., e.g. the 1st. power (i.e. the conditional standard deviation). For that purpose it should be noted that the dth. power log-GARCH(P,Q) can be written as

$$\ln \sigma_t^d = \alpha_{0,d} + \sum_{p=1}^{P} \alpha_p \ln |\epsilon_{t-p}|^d + \sum_{q=1}^{Q} \beta_q \ln \sigma_{t-q}^d, \quad d > 0, \tag{3}$$

where $\alpha_{0,d} = \alpha_0 d/2$. This means that a complete analysis of the dth. power log-GARCH model can be undertaken almost entirely in terms of the $d = 2$ representation.

Formulas that facilitate the computations of the unconditional moments, autocovariances and autocrrelations of $\{\epsilon_t\}$ for the log-GARCH(1,1), are contained in Appendix B. Table 1 contains unconditional autocorrelations for a small selection of empirically relevant parameter values. As is clear from the table, the log-GARCH(1,1) accommodates a broader range of persistency structures than the ordinary GARCH(1,1). In particular, in contrast to the ordinary GARCH(1,1) model, the unconditional autocorrelations of log-GARCH(1,1) models depend on the distribution of z_t: The more fat-tailed, the weaker correlations. Also, the log-GARCH(1,1) is capable of generating both weaker and stronger autocorrelations than the GARCH(1,1), and autocorrelation functions that decline either more rapidly or more slowly.

2.2 The ARMA representation

If $|\mathbb{E}(\ln z_t^2)| < \infty$ and $\mathbb{E}[(\ln z_t^2)^2] < \infty$, then the log-GARCH$(P,Q)$ model (1)-(2) admits the ARMA(P,Q) representation

$$\ln \epsilon_t^2 = \phi_0 + \sum_{p=1}^{P} \phi_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \theta_q u_{t-q} + u_t, \quad u_t \sim IID(0, \sigma_u^2), \qquad t \in \mathbb{Z} \tag{4}$$

almost surely with $\sigma_u^2 < \infty$, where

$$\begin{aligned}
\phi_0 &= \alpha_0 + (1 - \sum_{q=1}^{Q} \beta_q) \cdot \mathbb{E}(\ln z_t^2) && (5)\\
\phi_p &= \alpha_p + \beta_p, \quad 1 \leq p \leq P, && (6)\\
\theta_1 &= -\beta_1, \quad 1 \leq q \leq Q, && (7)\\
u_t &= \ln z_t^2 - \mathbb{E}(\ln z_t^2). && (8)
\end{aligned}$$

Consistent and asymptotically normal estimates of all the ARMA parameters – and hence all the log-GARCH parameters except the log-volatility intercept α_0 – is thus readily obtained via usual ARMA estimation methods subject to appropriate assumptions, see e.g. Brockwell and Davis (2006). In order to obtain an estimate of α_0 the most common solutions have been to either impose restrictive assumptions regarding the distribution of z_t (say, normality, see e.g. Psaradakis and Tzavalis (1999)), or to use an *ex post* scale-adjustment (see e.g. Bauwens and Sucarrat (2010), and Sucarrat and Escribano (2012)). What Theorem 1 below states is that a slightly modified version of an *ex post* scale-adjustment provides a consistent and asymptotically normal estimate of $\mathbb{E}(\ln z_t^2)$ for a range of ARMA estimators.

To obtain an understanding of the motivation behind the scale-adjustment, consider writing (1) as

$$\epsilon_t = \sigma_t^* z_t^*, \quad z_t^* \sim IID(0, \sigma_{z^*}^2),$$

where σ_t^* is a time-varying scale not necessarily equal to the standard deviation, and where z_t^* does not necessarily have unit variance. Of course, by construction we have $\sigma_t = \sigma_t^* \sigma_{z^*}$ and $z_t = z_t^*/\sigma_{z^*}$. Next, suppose a log-scale specification (e.g. an ARMA specification contained in (4)) is fitted to $\ln \epsilon_t^2$, with $\ln \hat{\sigma}_t^{*2}$ denoting the fitted value of the ARMA specification such that $\hat{\sigma}_t^* = \exp(\ln \hat{\sigma}_t^*)$, and with the ARMA residual defined as $\hat{u}_t = \ln \epsilon_t^2 - \ln \hat{\sigma}_t^{*2}$. In order to obtain an estimate of the time-varying conditional standard deviation, which is needed for comparison with most other volatility models, then it is natural to consider adjusting $\hat{\sigma}_t^*$ by multiplying it with an estimate of σ_{z^*}, say, the sample standard deviation of the standardised residuals $\{\hat{z}_t^*\}$. Although this argument is fine heuristically, it may not be apparent what underlying magnitude the adjustment in fact estimates, nor may it be straightforward to obtain the limiting properties of the adjustment under suitable conditions. In the log-GARCH model, however, the log of the scale-adjustment provides an estimate of $-\mathbb{E}\ln(z_t^2)$. To see this consider the scale adjustment and its approximation:

$$\hat{\sigma}_{z^*}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{z}_t^* - \bar{\hat{z}}_t^*)^2 \quad \approx \quad \frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t^*)^2 = \frac{1}{T}\sum_{t=1}^{T}\exp(\hat{u}_t). \tag{9}$$

The population analogue of the final expression on the right is $\mathbb{E}[\exp(u_t)]$. Taking the log of $\mathbb{E}[\exp(u_t)]$ gives

$$\begin{aligned}
\ln \mathbb{E}[\exp(u_t)] &= \ln \mathbb{E}[\exp(\ln z_t^2) \cdot \exp(-\mathbb{E}(\ln z_t^2))] \\
&= \ln \left[\mathbb{E}(z_t^2) \cdot \exp(-\mathbb{E}(\ln z_t^2))\right] \\
&= \ln \left[\exp(-\mathbb{E}(\ln z_t^2))\right] \\
&= -\mathbb{E}(\ln z_t^2),
\end{aligned}$$

under the assumption that $\mathbb{E}(z_t^2) = 1$. This suggests that

$$-\ln \left[\frac{1}{T}\sum_{t=1}^{T}\exp(\hat{u}_t)\right] \longrightarrow \mathbb{E}(\ln z_t^2) \tag{10}$$

due to the continuity of the natural log operator. However, the expression $(1/T)\sum_{t=1}^{T}\exp(\hat{u}_t)$ involves the ARMA residuals $\{\hat{u}\}$, which means that the standard law of large numbers cannot be applied. But a result by Yu (2007) enables us to show that a slightly modified scale adjustment provides a consistent, asymptotically normal and efficient estimate of $\mathbb{E}(\ln z_t^2)$ for a range of ARMA estimators.

Our result relies on the following assumptions:

A1: $|\mathbb{E}(\ln z_t^2)| < \infty$.

A2: The ARMA representation (4) has no common roots, and all the roots are outside the unit circle of the complex plane.

A3: Let $\hat{\phi}_p$, $\hat{\theta}_q$ be some given estimators for ϕ_p, θ_q, where $0 \leq p \leq P$ and $1 \leq q \leq Q$, such that:

- *a*) The assumptions underlying the estimators are fulfilled and compatible with **A1** and **A2**
- *b*) $\sqrt{T}(\hat{\phi}_p - \phi_p)$ and $\sqrt{T}(\hat{\theta}_q - \theta_q)$ are all $O_P(1)$

A4: Let $\delta_M = \sup\{\delta : \mathbb{E}(z_t^{2\delta}) < \infty\}$:

- *a*) $\delta_M > 1$
- *b*) $\delta_M > 2$ and $|\mathbb{E}[(\ln z_t^2)^2]| < \infty$.

Assumption **A1** is necessary for the ARMA representation (4) to exist. Because $\ln x \leq x - 1$, we have $\mathbb{E} \ln |z_t| < \mathbb{E}|z_t| - 1$. Since we assume z_t to be standardized, z_t has at least two moments. However, a lower bound for $\mathbb{E} \ln |z_t|$ depends on the distribution of z_t near zero. If the density of z_t given by f_{z_t} is bounded near zero, this expectation always exists, as $|\int_{-l}^{l} (\log |x|) f_{z_t}(x)\, \mathrm{d}x| \leq \sup_{-l \leq x \leq l} f_{z_t}(x) | \int_{-l}^{l} \log |x|\, \mathrm{d}x| < \infty$ for any $0 \leq l < \infty$. However, if the density is unbounded near zero, then more care is needed, and $|\mathbb{E}(\ln z_t^2)|$ may even be infinite if there is sufficient probability mass close to zero. For the two most commonly used densities of z_t in finance, i.e. $N(0,1)$ and the t, $\mathbb{E}(\ln z_t^2)$ is finite. In **A2** the no common roots assumption implies in this context that $P \geq Q$. For example, in the ARMA representation of a log-GARCH(1,2) we get that $\phi_2 = -\theta_2$, i.e. a common root. The additional assumption that all the roots are outside the unit circle implies that all the roots of the lag polynomial $1 - (\alpha_1 + \beta_1)L - \cdots - (\alpha_{P^*} + \beta_{P^*})L^{P^*}$ are all greater than 1 in modulus, that is, that $\ln \sigma_t^2$ in (2) is stable. The consistency and asymptotic normality of our estimator of $\mathbb{E}(\ln z_t^2)$ will hold for several ARMA estimators. Assumption **A3** simply states that the assumptions – whatever they are – that underly these ARMA estimators hold, that they are compatible with **A1** and **A2**, and that they satisfy **A3** b). In that regard, it should be noted that **A3**b) is implied by asymptotic normality, but the condtition does not require it. In **A4** condition a) will be required for our consistency result, whereas b) will also be required for the asymptotic normality result. Of course, *b*) implies *a*). Whether $|\mathbb{E}[(\ln z_t^2)^2]| < \infty$ holds depends on the density of z_t near zero, see the discussion in **A1**. Of course, if $|\mathbb{E}[(\ln z_t^2)^2]| < \infty$, then **A1** also holds. As will become apparent by our later arguments (see the proof of Lemma 1 for details), if u_t were observable, then a necessary and sufficient condition for consistency and asymptotic normality, respectively, would be that $\delta_M \geq 1$ and $\delta_M \geq 2$, respectively. In other words, the conditions $\delta_M > 1$ and $\delta_M > 2$ are very close to the weakest possible conditions.

The following lemma is an important step towards our main result.

Lemma 1. Suppose (1)-(2) and assumptions **A1** – **A3** hold, and let $\hat{u}_t$ for $t = 1, 2, \ldots, T$ denote the ARMA-residual when using some estimators $\hat{\phi}_p$, $\hat{\theta}_q$ for ϕ_p, θ_q, where $0 \leq p \leq P$ and $1 \leq q \leq Q$, in estimating the ARMA representation (4). Denoting $\overline{\hat{u}}_T$ and $\overline{u}_T$ as the empirical averages of $\hat{u}_t$ and u_t, respectively:

a) If **A4***a*) holds, then $\displaystyle \frac{1}{T}\sum_{t=1}^{T}\exp(\hat{u}_t - \overline{\hat{u}}_T) - \frac{1}{T}\sum_{t=1}^{T}\exp(u_t - \overline{u}_T) = o_P(1)$

b) If **A4***b*) holds, then $\displaystyle \sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T}\exp(\hat{u}_t - \overline{\hat{u}}_T) - \frac{1}{T}\sum_{t=1}^{T}\exp(u_t - \overline{u}_T)\right] = o_P(1)$

Proof. See Appendix C.

The intuition of the Lemma is that it provides sufficient conditions under which sums of exponentials like $(1/T)\sum_t \exp(\hat{u}_t - \overline{\hat{u}}_T)$ can be treated *as if* we observe the actual errors $\{u_t\}$. It should be noted that estimation of this term implies non-trivial changes in the behaviour of the residuals, and so Lemma 1 will not be valid if the residuals are not mean corrected with $\overline{\hat{u}}_T$. Nevertheless, in some cases, e.g. when OLS is used to estimate the AR(P) representation of a log-ARCH(P) specification, then $\overline{\hat{u}}_T$ will by construction be zero. Our main theoretical result now follows.

Theorem 1. Suppose (1)-(2) and assumptions **A1** – **A3** hold:

a) If assumption **A4***a*) also holds, then

$$-\ln\left[\frac{1}{T}\sum_{t=1}^{T}\exp(\widehat{u}_t - \overline{\widehat{u}}_T)\right] \xrightarrow{P} \mathbb{E}(\ln z_t^2). \tag{11}$$

b) If assumption **A4***b*) also holds, then

$$\sqrt{T}\left[-\ln\left[\frac{1}{T}\sum_{t=1}^{T}\exp(\widehat{u}_t - \overline{\widehat{u}}_T)\right] - \mathbb{E}(\ln z_t^2)\right] \xrightarrow{d} N(0, \zeta^2), \tag{12}$$

where

$$\zeta^2 = \mathbb{E}[(\ln z_t^2)^2] - [\mathbb{E}(\ln z_t^2)]^2 + (\mathbb{E}(z_t^4) - 1) - 2\mathbb{E}[(\ln z_t^2)z_t^2] + 2\mathbb{E}(\ln z_t^2). \tag{13}$$

Proof. The consistency result in a) follows straightforwardly from a) in Lemma 1 due to the continuity of the log-transformation. The proof of the asymptotic normality result b) is given in Appendix C. □

The practical implication of Theorem 1 is that the residuals from a range of estimators of the ARMA representation (4) – including Gaussian QML – can be plugged into the formula in order to obtain a consistent estimate of $\mathbb{E}(\ln z_t^2)$. Moreover, the asymptotic variance ζ^2 is in fact equal to the situation where we have access to the actual errors $\{u_t\}$.

An extensive set of Monte Carlo simulations were performed, of which Tables 2 and 3 only report a small subset. They succinctly summarise the main insights from the simulations. Both tables report the results of simulations when Gaussian QML is used to estimate the ARMA(1,1) representation of a log-GARCH(1,1). In the first table z_t is $N(0,1)$ and in the second z_t is standardised $t(5)$. The simulations suggest the empirical standard errors and correlations coincide with their asymptotic counterparts, although – as expected – a larger number of observations is needed as the persistence parameter $\phi_1 = \alpha_1 + \beta_1$ approaches 1, and when α_1 goes towards zero (i.e. the common root situation). Also, when ϕ_1 approaches 1, the finite sample bias of the estimator of $\mathbb{E}(\ln z_t^2)$ increases and the estimate becomes more imprecise.

2.3 Existence of unconditional moments of ϵ_t

The existence of the unconditional moments of ϵ_t do not have have any bearing upon the estimation and inference methods that we propose. However, they are important for the economic interpretation and analysis of the unconditional errors when they are interpreted as (de-meaned) returns, as is often the case in finance. In particular, their existence are needed for the computation of the autocorrelation function of squared (de-meaned) returns ϵ_t^2. A shortcoming in Nelson's (1991) EGARCH model is that its unconditional variance (and other, higher order unconditional moments) of ϵ_t may not exist for common distributions of the standardised innovations z_t. For example, if $z_t \overset{IID}{\sim} t(\nu), \nu > 2$ in Nelson's EGARCH(1,1) with log-volatility specification

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \left(|z_{t-1}| - E|z_{t-1}|\right) + \lambda z_{t-1} + \beta_1 \ln \sigma_{t-1}^2,$$

then the empirically unreasonable assumption $\alpha_1 < 0$ (i.e. a negative impact of ARCH) is a necessary condition for the existence of the unconditional variance of ϵ_t, see condition (A1.6) and the subsequent discussion in Nelson (1991, p. 365). Moreover, if $\lambda \neq 0$ (i.e. there is a leverage effect), then α_1 has to be even more negative for the unconditional variance of ϵ_t to exist. These are the shortcomings that prompted the work by Harvey (2013) on the Beta-t-EGARCH model.

In the log-GARCH(1,1) with $z_t \sim t(\nu), \nu > 2$, by contrast, the unconditional variance of ϵ_t will generally exist, regardless of the signs of the ARCH and GARCH coefficients α_1 and β_1. The following proposition provides a set of sufficient conditions for the existence of the unconditional moments of ϵ_t.

Proposition 1. Consider the log-GARCH(1,1) specification

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2$$

with $|\alpha_1 + \beta_1| < 1$:

a) If $z_t \overset{IID}{\sim} GED(\tau), \tau > 1$ and $2\alpha_1(\alpha_1+\beta_1)^{i-1} > -1$ for each $i = 1, 2, \ldots$, then, for $s > -1$, $\mathbb{E}(\epsilon_t^s) < \infty$ and is given by equation (35) in Appendix B

b) If $z_t \overset{IID}{\sim} t(\nu), \nu \neq 2$, $s < \nu$ and if $2\alpha_1(\alpha_1+\beta_1)^{i-1} \in (-1, \nu)$ for each $i = 1, 2, \ldots, t$, then, for $s > -1$, $\mathbb{E}(\epsilon_t^s) < \infty$ and is given by equation (35) in Appendix B

Proof. See Appendix D.

In practice, the conditions in Proposition 1 are very weak and will generally be satisfied, since the typical estimates of α_1 and β_1 are about 0.05 and 0.90, respectively. In particular, suppose $\tau > 1$ if z_t is $GED(\tau)$, and that $\nu > 2$ if z_t is $t(\nu)$. If $|\alpha_1 + \beta_1| < 1$ (which is the case in the majority of situations) and if $\alpha_1 \in (-0.5, 2)$ (which is almost always the case), then $2\alpha_1(\alpha_1+\beta_1)^{i-1}$ takes values in $(-1, 2)$ for all $i = 1, 2, \ldots$ Hence, $\mathbb{E}(\epsilon_t^s) < \infty$ will generally hold. A set of sufficient conditions for more general univariate log-GARCH specifications is provided by setting $M = 1$ in Proposition 2 in section 3.

2.4 Log-GARCH-X

Additional exogenous or predetermined variables ("X") can be added linearly or nonlinearly to the log-volatility specification $\ln \sigma_t^2$ without affecting the relationship between the log-GARCH coefficients and the ARMA coefficients. Specifically, let the log-GARCH-X model be given by

$$\ln \sigma_t^2 = \alpha_0 + \sum_{p=1}^{P} \alpha_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \ln \sigma_{t-q}^2 + g(\lambda, x_t), \tag{14}$$

where g is a linear or nonlinear function of the exogenous or predetermined variables x_t and a parameter vector λ. The index t in x_t does not necessarily mean that all (or any) of its elements are contemporaneous. If $|\mathbb{E}(\ln z_t^2)| < \infty$ and if $\mathbb{E}[(\ln z_t^2)^2] < \infty$, then the ARMA-X representation of (14) exists almost surely and is given by

$$\ln \epsilon_t^2 = \phi_0 + \sum_{p=1}^{P} \phi_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \theta_q u_{t-q} + g(\lambda, x_t) + u_t, \quad u_t \sim IID(0, \sigma_u^2), \tag{15}$$

where the ARMA coefficients are defined as before, i.e. by (5)-(7), and where u_t is the same as earlier, i.e. $u_t = \ln z_t^2 - \mathbb{E}(\ln z_t^2)$. Rigorously derived estimation and inference results, which we do not provide here, would require precise assumptions on the behaviour of x_t, see for example Hannan and Deistler (2012, chapter 4). However, if all the ARMA-X parameters are estimated consistently, then a reasonable conjecture is that (11) provides a consistent estimate of $\mathbb{E}(\ln z_t^2)$, and hence that all the log-GARCH parameters can be estimated consistently.

3 Multivariate log-GARCH

3.1 Notation and specification

The M-dimensional log-GARCH model is given by

$$\epsilon_t \sim ID(0, H_t), \quad t \in \mathbb{Z}, \tag{16}$$
$$D_t^2 = \operatorname{diag}\left\{\sigma_{m,t}^2\right\}, \quad m = 1, \ldots, M, \tag{17}$$
$$z_t = D_t^{-1}\epsilon_t, \quad z_{m,t} \sim IID(0,1), \quad Prob(z_t = 0) = 0, \tag{18}$$

where ϵ_t, σ_t^2 and z_t are $M \times 1$ vectors, and where H_t and D_t are $M \times M$ matrices. Equation (16) means ϵ_t is independent with mean zero and a time-varying conditional covariance matrix H_t. The IID assumption in equation (18) states that each marginal series $\{z_{m,t}\}$ is $IID(0,1)$. Marginal identicality is a key characteristic of the ARCH class of models, and is needed for the formula in Theorem 1 to be applicable after estimation via the VARMA representation. An implication of (18) is that $z_t \sim ID(0, R_t)$, where R_t is both the conditional covariance and correlation matrix – possibly time-varying – of z_t. In other words, the vector z_t is ID but not necessarily IID, even though each marginal series $\{z_{mt}\}$ is IID. Estimation of the volatilities D_t^2 does not require that the off-diagonals of H_t (i.e. the covariances) are specified explicitly. Nor do we need to assume that ϵ_t is distributed according to a certain density, say, the normal.

If $\ln \sigma_t^2$ denotes the $M \times 1$ vector resulting from applying the log on σ_t^2, then the M-dimensional log-volatility specification is given by

$$\ln \sigma_t^2 = \alpha_0 + \sum_{p=1}^{P} \alpha_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \ln \sigma_{t-q}^2, \qquad P \geq Q, \tag{19}$$

where

$$\ln \sigma_t^2 = \begin{pmatrix} \ln \sigma_{1,t}^2 \\ \vdots \\ \ln \sigma_{m,t}^2 \\ \vdots \\ \ln \sigma_{M,t}^2 \end{pmatrix}, \quad \alpha_0 = \begin{pmatrix} \alpha_{1.0} \\ \vdots \\ \alpha_{m.0} \\ \vdots \\ \alpha_{M.0} \end{pmatrix}, \quad \alpha_p = \begin{pmatrix} \alpha_{11.p} & \cdots & \alpha_{1m.p} & \cdots & \alpha_{1M.p} \\ \vdots & \ddots & \vdots & & \vdots \\ \alpha_{m1.p} & \cdots & \alpha_{mm.p} & \cdots & \alpha_{mM.p} \\ \vdots & & \vdots & \ddots & \vdots \\ \alpha_{M1.p} & \cdots & \alpha_{Mm.p} & \cdots & \alpha_{MM.p} \end{pmatrix},$$

$$\ln \epsilon_{t-p}^2 = \begin{pmatrix} \ln \epsilon_{1,t-p}^2 \\ \vdots \\ \ln \epsilon_{m,t-p}^2 \\ \vdots \\ \ln \epsilon_{M,t-p}^2 \end{pmatrix}, \quad \beta_q = \begin{pmatrix} \beta_{11.q} & \cdots & \beta_{1m.q} & \cdots & \beta_{1M.q} \\ \vdots & \ddots & \vdots & & \vdots \\ \beta_{m1.q} & \cdots & \beta_{mm.q} & \cdots & \beta_{mM.q} \\ \vdots & & \vdots & \ddots & \vdots \\ \beta_{M1.q} & \cdots & \beta_{Mm.q} & \cdots & \beta_{MM.q} \end{pmatrix}.$$

The multivariate log-GARCH model is thus *not* simply a collection of univariate log-GARCH models, since any equation in the the multivariate log-GARCH(P, Q)

admits feedback from any of the other $M-1$ equations via the lagged log-ARCH and log-GARCH terms. For example, a two-dimensional log-ARCH(1) specification is given by

$$\begin{aligned}\ln\sigma_{1,t}^2 &= \alpha_{1.0}+\alpha_{11.1}\ln\epsilon_{1,t-1}^2+\alpha_{12.1}\ln\epsilon_{2,t-1}^2\\ \ln\sigma_{2,t}^2 &= \alpha_{2.0}+\alpha_{21.1}\ln\epsilon_{1,t-1}^2+\alpha_{22.1}\ln\epsilon_{2,t-1}^2,\end{aligned}$$

whereas the specification of a two-dimensional log-GARCH(2,1) is given by

$$\begin{aligned}\ln\sigma_{1,t}^2 &= \alpha_{1.0}+\alpha_{11.1}\ln\epsilon_{1,t-1}^2+\alpha_{12.1}\ln\epsilon_{2,t-1}^2+\alpha_{11.2}\ln\epsilon_{2,t-2}^2\\ &\quad+\alpha_{12.2}\ln\epsilon_{2,t-2}^2+\beta_{11,1}\ln\sigma_{1,t-1}^2+\beta_{12,1}\ln\sigma_{2,t-1}^2\\ \ln\sigma_{2,t}^2 &= \alpha_{2.0}+\alpha_{21.1}\ln\epsilon_{1,t-1}^2+\alpha_{22.1}\ln\epsilon_{2,t-1}^2+\alpha_{21.2}\ln\epsilon_{2,t-2}^2\\ &\quad+\alpha_{22.2}\ln\epsilon_{2,t-1}^2+\beta_{21,1}\ln\sigma_{1,t-1}^2+\beta_{22,1}\ln\sigma_{2,t-1}^2.\end{aligned}$$

And so on.

A drawback with Nelson's (1991) EGARCH model is that the unconditional moments of ϵ_t generally do not exist for certain conditional densities, e.g. the standardised t. By contrast, the proposition below provides a general set of non-restrictive sufficient conditions for the existence of the unconditional moments of ϵ_t. Again, it should be noted that the existence or non-existence of the unconditional moments of ϵ_t will usually have no bearing upon the estimation and inference methods that we propose.

Proposition 2. Consider an M-dimensional log-GARCH(P,Q) model with $P\geq Q$ that admits the representation $\ln\sigma_t^2=\Psi_0+\sum_{i=1}^{\infty}\Psi_i\ln z_{t-i}^2$ with $\{\Psi_i\}$ being an absolutely summable sequence of $(M\times M)$ matrices. Then the sth. unconditional moment $E(\epsilon_{m,t}^s)=\exp(s2^{-1}\psi_{m,0})\cdot\prod_{i=1}^{\infty}E\left[|z_{1,t-i}|^{s\psi_{i,m1}}|z_{2,t-i}|^{s\psi_{i,m2}}\cdots\ |z_{M,t-i}|^{s\psi_{i,mM}}\right]$, $s\in\{1,2,\ldots\}$, of variable $m\in\{1,\ldots,M\}$ exists if $|E(z_{m,t}^s)|<\infty$ and if $E\left[|z_{1,t-i}|^{s\psi_{i,m1}}|z_{2,t-i}|^{s\psi_{i,m2}}\cdots|z_{M,t-i}|^{s\psi_{i,mM}}\right]<\infty$ for each i

Proof. See Appendix D.

In practice, the natural condition to check is whether all the eigenvalues of the $(M\times M)$ matrix $\sum_{p=1}^{P^*}(\alpha_p+\beta_p)$, where $P^*=\max\{P,Q\}$, are smaller than 1 in modulus. If this is the case, then $\{\Psi_i\}$ is absolutely summable. Whether the second set of conditions is satisfied or not, that is, $|E(z_{m,t}^s)|<\infty$ for each m and $E\left[|z_{1,t-i}|^{s\psi_{i,m1}}|z_{2,t-i}|^{s\psi_{i,m2}}\cdots|z_{M,t-i}|^{s\psi_{i,mM}}\right]<\infty$ for each i, will depend on the distribution of z_t.

3.2 Estimation and inference via the VARMA representation

If $|\mathbb{E}(\ln z_t^2)| < \infty$, then the VARMA$(P, Q)$ representation of the M-dimensional log-GARCH(P, Q) model (19) exists almost surely and is given by

$$\ln \epsilon_t^2 \;=\; \phi_0 + \sum_{p=1}^{P} \phi_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \theta_q u_{t-q} + u_t, \quad u_t \sim ID(0, \Sigma_t), \tag{20}$$

where

$$\phi_0 = \alpha_0 + (I_M - \sum_{q=1}^{Q} \beta_q \cdot \mathbb{E}(\ln z_t^2)), \quad \phi_p = \alpha_p + \beta_p, \quad \theta_q = -\beta_q, \tag{21}$$

$$u_t = \ln z_t - \mathbb{E}(\ln z_t^2), \quad \ln z_t^2 = \begin{pmatrix} \ln z_{1,t}^2 \\ \vdots \\ \ln z_{m,t}^2 \\ \vdots \\ \ln z_{M,t}^2 \end{pmatrix}, \quad \mathbb{E}(\ln z_t^2) = \begin{pmatrix} \mathbb{E}(\ln z_{1,t}^2) \\ \vdots \\ \mathbb{E}(\ln z_{m,t}^2) \\ \vdots \\ \mathbb{E}(\ln z_{M,t}^2) \end{pmatrix}. \tag{22}$$

Additional assumptions are needed for Σ_t to be finite, but not for the VARMA representation to exist almost surely. Although $\{u_t\}$ is only an independent and not necessarily an identical zero-mean vector process, each marginal process $\{u_{mt}\}$, $m = 1, \ldots, M$, is IID due to the assumption in (18). In the special case where the vector z_t is IID, which implies that the conditional correlations are constant, then the vector u_t is IID as well. In this case it is well known that multivariate Gaussian QML provides consistent and asymptotically normal estimates of the VARMA coefficients under suitable assumptions, see e.g. Lütkepohl (2005). Accordingly, consistent estimation and asymptotically normal inference regarding all the log-GARCH coefficients (apart from the log-volatility intercept α_0) is available as well. But in order to obtain a consistent estimate of α_0, an estimate of the vector $\mathbb{E}(\ln z_t^2)$ is needed. Since the process $\{u_{m,t}\}$ is marginally IID for $m = 1, \ldots, M$, a reasonable conjecture is that equation-by-equation application of the formula in Theorem 1 after estimation of the VARMA representation will provide consistent estimates of each element in $\mathbb{E}(\ln z_t^2)$.

In the more general case where the vector z_t is only ID, which is implied by time-varying correlations, then the vector u_t is only ID as well. This corresponds to a VARMA model with a heteroscedastic error u_t. Fewer QML results are available in this case, e.g. Bardet and Wintenberger (2009). But a natural conjecture – on which widespread statistical practice is based – is that multivariate Gaussian QML estimation will provide consistent estimates of the VARMA coefficients (under suitable conditions). If this is indeed the case, another natural conjecture is that equation-by-equation application of the formula in Theorem 1 will provide consistent estimates of each element in $\mathbb{E}(\ln z_t^2)$ also here.

3.3 Multivariate log-GARCH-X

Just as in the univariate case, the multivariate log-GARCH model permits exogenous or predetermined conditioning variables in each of the M equations. Specifically, write the multivariate log-GARCH-X specification as

$$\ln \sigma_t^2 = \alpha_0 + \sum_{p=1}^{P} \alpha_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \ln \sigma_{t-q}^2 + \sum_{k=0}^{K} \lambda_k x_{t-k}, \tag{23}$$

where x_t is an $L \times 1$ vector of predetermined or exogenous variables, and where $\lambda_0, \lambda_1, \ldots, \lambda_K$ are $M \times L$ matrices. Here, for notational simplicity, we let the predetermined or exogenous variables enter linearly. However, in principle they could instead enter non-linearly, just as in the univariate case. The VARMA-X representation is given by

$$\ln \epsilon_t^2 \;\; = \;\; \phi_0 + \sum_{p=1}^{P} \phi_p \ln \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \theta_q u_{t-q} + \sum_{k=0}^{K} \lambda_k x_{t-k} + u_t, \quad u_t \sim ID(0, \Sigma_t),$$

with the VARMA coefficients and u_t defined as before (i.e. by (21) and (22)). In other words, the relation between the VARMA coefficients and the log-GARCH coefficients are not affected when $\sum_{k=0}^{K} \lambda_k x_{t-k}$ is added to (23). Again, a natural conjecture to make is that multivariate Gaussian QML via the VARMA-X representation will provide consistent estimates and asymptotically normal inference of the VARMA-X coefficients (and therefore of the log-GARCH-X coefficients apart from the log-volatility intercept), and that the formula from Theorem 1 can be used equation-by-equation to estimate each element in $\mathbb{E}(\ln z_t^2)$ in order to obtain an estimate of α_0.

3.4 Time-varying correlations

Estimation of the volatilities D_t^2 does not require that the off-diagonals of H_t are specified explicitly. Accordingly, estimation of the (possibly) time-varying covariances for some application would have to be undertaken subsequently. Arguably the most common approach of this type is Engle's (2002) Dynamic Conditional Correlations (DCC) model. In the model (16)-(18) this amounts to specifying H_t as

$$H_t \;\; = \;\; D_t R_t D_t, \tag{24}$$

$$R_t \;\; = \;\; \text{diag}\left\{Q_t\right\}^{-1} Q_t \text{diag}\left\{Q_t\right\}^{-1}, \tag{25}$$

where R_t and Q_t are $M \times M$ matrices. There is a range of possibilities for the specification of Q_t in equation (24), see e.g. Bauwens et al. (2006, pp. 89-91) and the discussion in Engle (2002). The most commonly used specification, however, is the one that specifies the entries of Q_t in a GARCH(1,1) like manner,

$$q_{ij,t} = \bar{\rho}_{ij} + a(z_{i,t-1} z_{j,t-1} - \bar{\rho}_{ij}) + b(q_{ij,t-1} - \bar{\rho}_{ij}), \tag{26}$$

where $\bar{\rho}_{ij}$ is the unconditional correlation between $z_{i,t}$ and $z_{j,t}$. This specification is commonly referred to as the DCC of Engle (2002), and this is the specification used in the empirical section for the multivariate application.

4 Modelling the uncertainty of short-term electricity prices

Sucarrat and Escribano (2012) explicitly relies on the methods of this paper. There, in the empirical aplication, a log-ARCH-X model is used to model exchange rate volatility and stock price volatility, respectively. Here, our focus is on short-term electricity prices.

Short-term electricity price modelling and forecasting is of great importance for energy market participants on both the supply and on the demand side. On the supply side producers need forecasts of prices and the time-varying uncertainty associated with those forecasts, in order to appropriately determine the price and production levels. On the demand side consumers and speculators need the same type of information in order to decide when and where to produce, in order to speculate and/or hedge against adverse price changes, and for risk management purposes.

Daily electricity prices are characterised by strong autoregressive persistence, day-of-the week effects, seasonal effects, large spikes or jumps, fast mean-reversal and ARCH. Koopman et al. (2007), and Escribano et al. (2011) have proposed univariate models that contain some or all of these features. However, in neither of these models is the volatility specification – a GARCH(1,1) – robust to the large spikes that is a common characteristic of electricity prices. Nor are they flexible enough to accommodate a complex and rich heteroscedasticity dynamics similar to that of the mean specification without imposing very strong parameter restrictions. Nor are multivariate versions available, which are needed when, say, a factory considers shifting its production away from peak hours to off-peak hours. Also, automated model selection is infeasible in practice due to computational complexity and positivity constraints, and this problem is compounded in the multivariate case. The log-GARCH-X class of models, by contrast, remedies these deficiencies and permits a robust, flexible and rich characterisation of the volatility dynamics. We illustrate this in three applications. In the first the robustness to spikes is illustrated by comparing a log-GARCH(1,1) model with a GARCH(1,1) and with an EGARCH(1,1) specification. In the second application the flexibility and computational simplicity of the log-ARCH-X model is illustrated by undertaking automated model selection of a starting model with 33 deletable regressors. In the third application a similar exercise is performed for a bivariate model of the relationship between peak and off-peak spot daily electricity prices.

4.1 Robustness to price spikes: The log-GARCH(1,1)

In order to illustrate the robustness to spikes of the log-GARCH(1,1) model we revisit the Spanish electricity price data in Escribano et al. (2011). The part of the data that we revisit spans the period 1 January 1998 to 31 December 2003 ($T = 2191$ observations), see the upper two graphs of Figure 1. If $r_t = \Delta \ln S_t$ denotes the log-return of the daily Spanish spot electricity price S_t, then our empirical model of the conditional mean is given by (t-ratios in parentheses and p-values in square brackets):[4]

$$
\begin{aligned}
\hat{r}_t \;=\; & \underset{(13.5)}{0.088} - \underset{(-21.0)}{0.050} \cdot (8r_{t-1} + 6r_{t-2} + 4r_{t-3} + 2r_{t-4} + 2r_{t-5} + 2r_{t-6} \\
& + r_{t-9} + r_{t-10} + r_{t-11} + r_{t-12} + r_{t-13} + r_{t-14}) \\
& + \underset{(10.4)}{0.063} \cdot (r_{t-7} + r_{t-14} + r_{t-21} + r_{t-28} + r_{t-35}) \\
& + \underset{(4.03)}{0.142} I_{\{r_{t-1}<-0.5\}} + \underset{(2.29)}{0.077} I_{\{r_{t-2}<-0.5\}} \\
& - \underset{(-15.2)}{0.051} (Tue_t + Wed_t + Thu_t + 2Fri_t + 3Sat_t + 4Sun_t) \\
& - \underset{(-2.75)}{0.028} Dec_t
\end{aligned}
\tag{27}
$$

AR_1	AR_2	AR_7	AR_{14}	$ARCH_1$	$ARCH_2$	$ARCH_7$	$ARCH_{14}$	R^2
−0.03	0.01	−0.00	−0.01	0.23	0.05	0.12	0.05	0.56
[0.22]	[0.39]	[0.43]	[0.25]	[0.00]	[0.00]	[0.00]	[0.00]	

The threshold variables $I_{\{r_{t-1}<-0.5\}}$ and $I_{\{r_{t-2}<-0.5\}}$ seek to capture the strong and often almost immediate reversal effect when r_t drops by more than 0.5 points. The variables Tue_t to Sun_t are day-of-the-week dummies, whereas Dec_t is a month-of-the-year dummy. The lag structure suggests a negative but declining effect of previous intra-week days, whereas the effect of lag-multiples of 7—a periodicity effect—is positive. The negative effects of the day-of-the week dummies suggests that electricity prices tend to be lower throughout the week in comparison with Mondays, the more so on Saturdays and Sundays, and similarly for the month of December. AR and ARCH are the autocorrelations of the residuals and squared residuals, respectively, with the p-value of a Ljung and Box (1979) test in square brackets. The model is well-specified in the sense that the AR tests show no sign of autocorrelation in the residuals. However, the ARCH tests show clear signs of

[4]The model was obtained by means of General-to-Specific (GETS) model selection with the R package AutoSEARCH, see Sucarrat (2012), using the White (1980) coefficient covariance matrix for the regressor and parsimonious encompassing tests. The starting model contained an intercept (which was excluded from deletion), lags 1 to 14, 21, 28 and 35, six day-of-the-week dummies, eleven month-of-the-year dummies and the four threshold variables $I_{\{r_{t-1}<-0.5\}}, I_{\{r_{t-2}<-0.5\}}, I_{\{r_{t-1}>0.5\}}, I_{\{r_{t-2}>0.5\}}$ intended to capture the reversal effect of prior spikes, negative and positive. After simplification economically meaningful restrictions on the parameters were imposed in order to make the model more interpretable. Formally, our starting model had the form $r_t = \gamma_0 + \sum_{m \in M} \gamma_m r_{t-m} + \sum_{n=1}^{21} \eta_n x_{nt} + \epsilon_t$ with $M = \{1, \ldots, 14, 21, 28, 35\}$ and $\epsilon_t = \sigma_t z_t, z_t \sim IID(0,1)$, and estimation was undertaken with OLS.

ARCH. Graphs of the residuals and absolute residuals, respectively, are contained at the bottom of Figure 1.

The most popular model of ARCH in the residuals is the GARCH(1,1) of Bollerslev (1986), but it may not be the most appropriate model when modelling the uncertainty of electricity prices. Electricity prices occasionally exhibit large spikes or jumps with almost immediate and subsequent reversal to its normal level, and it is well known that the GARCH(1,1) is not robust to such spikes, see for example Carnero et al. (2007). In order to compare the conditional standard deviations, a GARCH(1,1), a log-GARCH(1,1) and an EGARCH(1,1) without leverage were fitted to the residuals of (27). The estimates of the models are:[5]

$$\begin{aligned}
\text{GARCH(1,1):} \quad & \hat{\sigma}_t^2 = 0.002 + 0.144\hat{\epsilon}_{t-1}^2 + 0.758\hat{\sigma}_{t-1}^2 \\
\text{EGARCH(1,1):} \quad & \ln \hat{\sigma}_t^2 = -0.710 + 0.289\left|\frac{\hat{\epsilon}_{t-1}}{\hat{\sigma}_{t-1}}\right| + 0.879 \ln \hat{\sigma}_{t-1}^2 \\
\text{log-GARCH(1,1):} \quad & \ln \hat{\sigma}_t^2 = -0.067 + 0.063 \ln \hat{\epsilon}_{t-1}^2 + 0.894 \ln \hat{\sigma}_{t-1}^2
\end{aligned}$$

Figure 2 compares the fitted conditional standard deviations. The upper graph in the figure may be interpreted to suggest that the conditional standard deviations of the three models are, on the whole, very similar. However, a closer look shows that they are occasionally very different. For example, in the bottom left graph we see that the three conditional SDs are very similar until 2 March 2002 when the residual experiences a 0.89 point drop. In financial terms this is a huge number, since it would correpsond to a 89% drop if log-differences were an accurate approximation to relative changes. On the subsequent day a new large residual of 0.52 points occurs, but thereafter the residuals remain relatively low (in absolute value). The log-GARCH SDs are not much affected by the large (in absolute value) residual, since the effect of a shock works through the logarithm operator. The GARCH, by contrast, produces SD forecast that are twice as high for the subsequent days compared with the log-GARCH. Moreover, the GARCH needs almost two weeks in order to return to the level of the log-GARCH. The EGARCH is an intermediate case, since it reacts almost as abruptly as the GARCH, but it needs substantially less time for the reversal. Similarly, on 6 November 2002 another large spike occurs. The GARCH SD more than doubles from about 17 to about 42 points, and rises further one more day before it needs almost two weeks to return back to the level of the two other models. By contrast, for the log-GARCH the effect of the shock is dampened, so the SDs remain much closer to the values of the absolute residuals. Again, the EGARCH is an intermediate case in that it does not react as abruptly as the GARCH, and in that it does not need as many days for the reversal.

[5] The GARCH and EGARCH models were estimated by means of Gaussian QML in the standardised error z_t, whereas the log-GARCH model was estimated by means of Gaussian QML in u_t via the ARMA representation.

4.2 The log-ARCH-X: Flexible modelling without computational difficulties

Electricity prices are strongly characterised by periodicity effects, so it seems reasonable that the same may hold for the price uncertainty (i.e. volatility). Adding many exogenous or predetermined variables to the GARCH, EGARCH and log-GARCH models, however, presents numerical challenges due to the complexity of the estimation problem. One solution, which is pursued by Koopman et al. (2007, p. 20), is to decompose the periodicity and ARCH multiplicatively, and then estimate the two effects separately. However, this assumes the two effects have no bearing upon each other. By contrast, periodicity, ARCH and other effects can be modelled jointly within the log-ARCH-X model. In fact, the model of Koopman et al. (2007) is nested in the log-ARCH-X if one of the conditioning variables is defined as the log of the GARCH volatility prediction.

Our starting point is a general volatility model that is fitted to the same residuals as earlier (i.e. the residuals of (27)):

$$\begin{aligned}\epsilon_t &= \sigma_t z_t, \quad z_t \sim IID(0,1), \quad \sigma_t > 0, \\ \ln \sigma_t^2 &= \alpha_0 + \sum_{p \in P} \alpha_p \ln \epsilon_{t-p}^2 + \lambda_1 \ln EqWMA(7)_{t-1} + \sum_{l=1}^{21} \lambda_l x_{lt}.\end{aligned}$$

The log-ARCH lags are $P = \{1, \ldots, 7, 14, 21, 28, 35\}$, $EqWMA(7)_{t-1}$ is a volatility proxy[6] made up of a rolling average of $\epsilon_{t-1}^2, \ldots, \epsilon_{t-7}^2$ and the 21 x_{nt} variables are the same as in the mean specification (see footnote 4). This means the general unrestricted log-volatility specification contains a total of 33 deletable regressors, and one regressor (the intercept) which is restricted from deletion in the specification search. Automated GETS multi-path model selection with AutoSEARCH (Sucarrat 2012) yields a parsimonious model, which we further simplify by imposing economically meaningful parameter restrictions. The end result is (t-ratios in parentheses and p-values in square brackets):

$$\begin{aligned}\ln \hat{\sigma}_t^2 &= -1.502 - \underset{(-1.62)}{0.037} \ln \hat{\epsilon}_{t-3}^2 - \underset{(2.75)}{0.063} \ln \hat{\epsilon}_{t-4}^2 - \underset{(-1.13)}{0.026} \ln \hat{\epsilon}_{t-6}^2 \\ &\quad + \underset{(11.2)}{0.677} \ln EqWMA(7)_{t-1} - \underset{(-5.49)}{0.086} (7Tue_t + 5Thu_t + 7Fri_t) \\ &\quad - \underset{(-1.72)}{0.064} (4Mar_t + 3Nov_t) \end{aligned} \tag{28}$$

AR_1	AR_2	AR_7	AR_{14}	$ARCH_1$	$ARCH_2$	$ARCH_7$	$ARCH_{14}$
−0.01 [0.63]	0.01 [0.75]	0.00 [0.80]	−0.00 [0.89]	0.04 [0.10]	−0.02 [0.17]	0.01 [0.62]	0.02 [0.87]

The model is well-specified in the sense that the AR and ARCH tests exhibit little or no signs of autocorrelation in the standardised residuals, and in the squared standardised residuals (see also the bottom graph of Figure 1). It is worth underlining

[6]The fitted volatilities of the GARCH, EGARCH and log-GARCH models were also considered as volatility proxies. However, they did not lead to final models free from ARCH.

that since the R^2 of the conditional mean specification (27) is substantially different from zero, the log-volatility specification should not be interpreted as a measure of price variability as is common in financial econometrics. Rather, the log-volatility specification acquires the more traditional interpretation of a measure of the time-varying inaccuracy of the mean specification: The lower $\ln \sigma_t^2$ is, the more accurate is the mean specification. The negative effect of the log-ARCH terms suggests that there is some sort of cyclical behaviour in the uncertainty. Two of the log-ARCH terms are not significant at 5%, and this means they have been retained because their deletion would induce serial correlation in either the standardised residuals or in the squared standardised residuals or in both (see Sucarrat and Escribano (2012) for the details of GETS model selection). The lagged impact of the log of the volatility proxy $\text{EqWMA}(7)_t$ is positive and about 0.68, which suggests a notable degree of persistence. The negative effects of the day-of-the-week dummies and of the two month-of-the-year dummies, means that the mean specification tend to be more precise on Tuesdays, Thursdays and Fridays, and in March and November.

4.3 Multivariate electricity price volatility modelling

Multivariate ARCH models are plagued by the curse of dimensionality due to the number of parameters that has to be estimated. Here, we show that a rich, multivariate model contained in the log-GARCH-X class can be straightforwardly estimated equation-by-equation with OLS, thus avoiding numerical estimation issues, even when the number of observations is relatively small. Next, the same automated GETS model selection methods that were used in the previous subsection are applied. The end result is a parsimonious, multivariate model and, in the process, a test for Granger-causality in the volatilities. Finally, we fit the DCC model of Engle (2002) to the time-varying correlations.

The data of our multivariate analysis consist of the daily peak and off-peak spot electricity price from 1 January 2010 to 19 September 2012 ($T = 993$ observations) for the Oslo area in Norway.[7] Factories, companies and other institutions with substantial electricity consumption may want to shift part of their activity to and from peak hours for efficient cost management, since the difference between peak and off-peak prices can be very large at times (see the bottom graph of Figure 3). As an aid in the decision-making process, forecasts of future prices and price uncertainty (volatility) can therefore be of great usefulness. The daily peak spot price $S_{1,t}$ is computed as the average of the spot prices during peak hours, that is, $S_{1,t} = (S_{t(8am)} + \cdots + S_{t(9pm)})/14$, whereas the daily off-peak spot price $S_{2,t}$ is computed as the average of the spot prices during off-peak hours, that is, $S_{2,t} = (S_{t(0am)} + \cdots + S_{t(7am)} + S_{t(10pm)} + S_{t(11pm)})/10$. Note that $S_{t(8am)}$ should be interpreted as the electricity price from 8am to 9am, $S_{t(9am)}$ should be interpreted as the electricity price from 9am to 10am, and so on. Graphs of $S_{1,t}, S_{2,t}$ and their

[7] The source of the data is http://www.nordpoolspot.com/, and the sample was determined by availability: Observations prior to the sample period are not available, and the data were downloaded 20 September 2012.

log-returns are contained in Figures 3 and 4. The price and returns figures exhibit the usual characteristics of electricity prices, namely that the price variability is substantially larger than those of financial prices (say, stocks, stock indices and exchange rates), and that big jumps with fast reversion occurs relatively frequently.

Conditional mean specifications are obtained in a similar way to above.[8] In terms of R^2, the simplified conditional mean specifications explain about 48% and 47% of the variation in the peak and off-peak log-returns, respectively. Most of this explanatory and predictive power comes from own AR-dynamics, asymmetry terms and the lagged error-correction term.

The resulting set of residuals of the simplified VECM models are used for the modelling of the log-volatility specifications. First, the following two-dimensional general model is formulated:

$$\begin{aligned}
\ln \sigma_{1,t}^2 &= \alpha_{1,0} + \sum_{p\in P} \alpha_{11,p} \ln \epsilon_{1,t-p}^2 + \sum_{p\in P} \alpha_{12,p} \ln \epsilon_{2,t-p}^2 + \sum_{k\in K} \lambda_{11,k} \ln EqWMA(k)_{1,t-1} \\
&\quad + \sum_{k\in K} \lambda_{12,k} \ln EqWMA(k)_{2,t-1} + \sum_{p\in P} +(\ln \lambda_{13,p}^2) I_{\{\epsilon_{1,t-p}<0\}} \ln \epsilon_{1,t-p}^2 \\
&\quad + \sum_{l\in L} \omega_{1,l} x_{1l,t} \\
\ln \sigma_{2,t}^2 &= \alpha_{2,0} + \sum_{p\in P} \alpha_{21,p} \ln \epsilon_{1,t-p}^2 + \sum_{p\in P} \alpha_{22,p} \ln \epsilon_{2,t-p}^2 + \sum_{k\in K} \lambda_{21,k} \ln EqWMA(k)_{1,t-1} \\
&\quad + \sum_{k\in K} \lambda_{22,k} \ln EqWMA(k)_{2,t-1} + \sum_{p\in P} +(\ln \lambda_{23,p}^2) I_{\{\epsilon_{2,t-p}<0\}} \ln \epsilon_{2,t-p}^2 \\
&\quad + \sum_{l\in L} \omega_{2,l} x_{2l,t},
\end{aligned}$$

where $P = \{1, \ldots, 14, 21, 28, 35\}$, $K = \{3, 7, 14, 21, 28, 35\}$ and the $\{x_{(\cdot)1,t}, \ldots, x_{(\cdot)17,t}\}$ are 11 month-of-the-year dummies and 6 day-of-the-week dummies. This amounts to a total of 62 regressors (intercept included) in each log-volatility specification. Next, automated GETS model selection is undertaken separately for each equation,

[8] First, a general unrestricted Vector Error Correction Model (VECM) is formulated and estimated. Each of the two VECMs contain AR-lags (1 to 14, 21, 28 and 35) of each return, daily and monthly impulse dummies, a lagged error-correction term, two "GARCH-in-mean" proxies and several terms to capture various types of asymmetry. This amounts to a total of 59 regressors (intercept included) in each of the two specifications. Next, simplification of each conditional mean specification is undertaken by means of GETS model selection using a significance level of 5% (regressors, parsimonious encompassing tests and overall diagnostics) and White (1980) coefficient covariance matrix.

which ultimately yields (t-ratios in parentheses and p-values in square brackets)

$$\begin{aligned}
\ln \hat{\sigma}^2_{1,t} &= -1.573 + \underset{(2.02)}{0.071} \ln \hat{\epsilon}^2_{1,t-2} + \underset{(2.06)}{0.073} \ln \hat{\epsilon}^2_{1,t-3} - \underset{(-2.64)}{0.093} \ln \hat{\epsilon}^2_{1,t-5} - \underset{(-2.38)}{0.082} \ln \hat{\epsilon}^2_{1,t-6} \\
&\quad + \underset{(3.23)}{0.101} \ln \hat{\epsilon}^2_{1,t-21} - \underset{(-2.34)}{0.092} \ln \hat{\epsilon}^2_{2,t-3} + \underset{(4.89)}{0.450} \ln EqWMA(7)_{1,t-1} \\
&\quad + \underset{(2.71)}{0.184} \ln EqWMA(3)_{2,t-1} - \underset{(-2.45)}{0.447} Sat_t \qquad (29) \\
\ln \hat{\sigma}^2_{2,t} &= -1.947 + \underset{(8.77)}{0.429} \ln EqWMA(3)_{2,t-1} + \underset{(2.89)}{0.087} \ln \hat{\epsilon}^2_{1,t-4} + \underset{(4.12)}{0.756} Sun_t, \qquad (30)
\end{aligned}$$

	AR_1	AR_2	AR_7	AR_{14}	$ARCH_1$	$ARCH_2$	$ARCH_7$	$ARCH_{14}$
1:	0.03 [0.42]	−0.01 [0.66]	0.07 [0.25]	−0.01 [0.53]	0.04 [0.18]	0.02 [0.33]	0.01 [0.67]	0.07 [0.70]
2:	0.02 [0.47]	−0.02 [0.63]	0.01 [0.96]	−0.04 [0.92]	0.05 [0.15]	0.04 [0.18]	0.00 [0.66]	0.10 [0.38]

The AR and ARCH tests of the standardised residuals suggest that the log-volatility specifications are well-specified. The simplified log-volatility specifications suggest there is Granger-causality in the log-volatilities, since there are feedback effects in both equations: The volatility of peak price returns depends on the 3rd. order log-ARCH of off-peak prices, and on the lagged 3-period moving average of past squared residuals, $EqWMA(3)_{t-1}$ (i.e. a volatility proxy), whereas the volatility of off-peak price returns depend on the 4th. order log-ARCH term of peak returns. In terms of the estimated coefficient sizes, the effect is greater from off-peak volatility to peak volatility than opposite. None of volatility asymmetry effects are retained, which may not be a surprise, since such effects are particularly associated with the stock market. Finally, the retention of one day-of-the-week dummy in each of the log-volatility specification (Saturday for peak and Sunday for off-peak), suggests that the precision of the return forecasts varies across the week.

In the next step the standardised residuals $\hat{z}_{1,t}$ and $\hat{z}_{2,t}$ are used to fit a DCC model of the Engle (2002) type.[9] Estimation is undertaken by ML using a bivariate Gaussian density, and this gives

$$\hat{q}_{ij,t} = \bar{\rho} + \underset{(0.017)}{0.036}(\hat{z}_{i,t-1}\hat{z}_{j,t-1} - \bar{\rho}) + \underset{(0.226)}{0.551}(\hat{q}_{ij,t-1} - \bar{q}_{ij,2}),$$

where $\bar{\rho} = 0.4021$, and where the fitted time-varying conditional correlations $\hat{\rho}_t$ are computed as $\hat{q}_{12,t}/\sqrt{\hat{q}_{11,t}\hat{q}_{22,t}}$. Figure 7 contains a graph of the fitted conditional correlations. Most of the correlations are close to the unconditional estimate of about 0.4. However, occasionally, in relation with large electricity price jumps, a similar abrupt and large movement occurs in the conditional correlations. Partly as a consequence of this, the estimate of persistence, $0.036 + 0.551 = 0.587$, is relatively low in comparison with financial price returns (say, stocks, stock indices and exchange rates).

[9] For computational convenience we do not use the bias-corrected estimator of Aielli (2009), since it is likely to produce almost identical results.

5 Conclusions

We have proposed estimation and inference methods for univariate and multivariate log-GARCH-X models via (V)ARMA-X representations. Estimation of log-GARCH-X models via the (V)ARMA-X representation induces a bias in the log-volatility intercept made up of a log-moment expression that depends on the conditional density. We proposed an estimator of the log-moment expression and proved its consistency, asymptotic normality and asymptotic efficiency for a range of ARMA estimators in the univariate log-GARCH case. Due to the structure of the problem the bias-correction procedure is likely to also hold for univariate log-GARCH-X models and – equation-by-equation – for multivariate log-GARCH-X models. We have also showed that the unconditional moments of the log-GARCH model exist under much weaker assumptions than for Nelson's (1991) EGARCH. Finally, our empirical applications on electricity prices show that the methods are particularly useful when the volatility dynamics are complex and affected by a number of factors, and when the conditional density is fat-tailed and/or contains jumps or outliers.

This paper is part of a larger research agenda. Sucarrat and Escribano (2012) rely explicitly on Theorem 1, whereas Bauwens and Sucarrat (2010) is a precursor to that paper. These two papers led to the development of `AutoSEARCH`, an *R* package for automated General-to-Specific (Gets) model selection of log-ARCH-X models (see Sucarrat (2012)). An early critique of the log-ARCH class of models was that the log-ARCH terms in the log-volatility specification may not exist, since the errors of a regression in empirical practice can be zero. This problem, however, is solved in Sucarrat and Escribano (2013), and in Sucarrat (2013).

References

Aielli, G. P. (2009). Dynamic Conditional Correlations: On Properties and Eestimation. http://ssrn.com/abstract=1507743.

Bai, J. (1994). Weak convergence of the sequential empirical process of residuals in arma models. *The Annals of Statistics 22*, 2051–2061.

Bardet, J.-M. and O. Wintenberger (2009). Asymptotic normality of the quasi maximum likelihood estimator for multidimensional causal processes. Unpublished working paper.

Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics 21*, 79–109.

Bauwens, L. and G. Sucarrat (2010). General to Specific Modelling of Exchange Rate Volatility: A Forecast Evaluation. *International Journal of Forecasting 26*, 885–907.

Berkes, I., L. Horvath, and P. Kokoszka (2003). GARCH processes: structure and estimation. *Bernoulli 9*, 201–227.

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

Brockwell, P. J. and R. A. Davis (2006). *Time Series: Theory and Methods*. New York: Springer. 2nd. Edition.

Carnero, M. A., D. Pena, and E. Ruiz (2007). Effects of outliers on the identification and estimation of GARCH models. *Journal of Time Series Analysis 28*, pp. 471–497.

Comte, F. and O. Lieberman (2003). Asymptotic Theory for Multivariate GARCH Processes. *Journal of Multivariate Analysis 84*, 61–84.

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations. *Econometrica 50*, 987–1008.

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics 20*, 339–350.

Engle, R. and J. Maricucci (2006). A long-run Pure Variance Common Features model for the common volatilities of the Dow Jones. *Journal of Econometrics 132*, 7–42.

Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews 5*, 1–50.

Escribano, A., , J. I. Peña, and P. Villaplana (2011). Modelling Electricity Prices: International Evidence. *Oxford Bulletin of Economics and Statistics 73*, 622–650.

Francq, C., O. Wintenberger, and J.-M. Zakoïan (2012). GARCH Models Without Positivity Constraints: Exponential or Log-GARCH? To appear in *Journal of Econometrics*, http//dx.doi.org/10.1016/j.jeconom.2013.05.004.

Francq, C. and J.-M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10*, 605–637.

Francq, C. and J.-M. Zakoïan (2006). Linear-representation Based Estimation of Stochastic Volatility Models. *Scandinavian Journal of Statistics 33*, 785–806.

Francq, C. and J.-M. Zakoïan (2010a). *GARCH Models*. New York: Marcel Dekker.

Francq, C. and J.-M. Zakoïan (2010b). QML estimation of a class of multivariate GARCH models without moment conditions on the observed process. Unpublished working paper.

Franses, P. H., J. Neele, and D. Van Dijk (2001). Modelling asymmetric volatility in weekly Dutch temperature data. *Environmental Modeling and Software 16*, 131–137.

Geweke, J. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 57–61.

Gradshteyn, I. S. and I. M. Ryzhik (2007). *Table of Integrals, Series and Products*. New York: Academic Press. Seventh edition. Available via http://books.google.com/.

Grønneberg, S. and G. Sucarrat (2013). Asymptotic Theory for Univariate and Multivariate Log-GARCH-X Models. Work in progress.

Hafner, C. and A. Preminger (2009). Asymptotic theory for a factor GARCH model. *Econometric Theory 25*, 336–363.

Hamilton, J. D. (2010). Macroeconomics and ARCH. In T. Bollerslev, J. R. Russell, and M. Watson (Eds.), *Festschrift in Honor of Robert F. Engle*. Oxford: Oxford University Press.

Hannan, E. and M. Deistler (2012). *The statistical theory of linear systems*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). Originally published in 1988 by Wiley, New York.

Hansen, P. R., Z. Huan, and H. H. Shek (2012). Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics 27*, 877–906.

Harvey, A. C. (1976). Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica 44*, 461–465.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. New York: Cambridge University Press.

Harvey, A. C. and T. Chakravarty (2008). Beta-t-(E)GARCH. Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge.

Harvey, A. C. and N. Shephard (1996). Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business and Economic Statistics 14*, 429–434.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*, pp. 70–86.

Kawakatsu, H. (2006). Matrix exponential GARCH. *Journal of Econometrics 134*, 95–128.

Koopman, S. J., M. Ooms, and M. A. Carnero (2007). Periodic Seasonal REG-ARFIMA-GARCH Models for Daily Electricity Spot Prices. *Journal of the American Statistical Association 102*, 16–27.

Kristensen, D. and A. Rahbek (2009). Asymptotics of the QMLE for Non-Linear ARCH Models. *Journal of Time Series Econometrics 1*, Issue 1, Article 2. Available at: http://www.bepress.com/jtse/vol1/iss1/art2.

Ling, S. and M. McAleer (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory 19*, 280–310.

Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika 66*, 265–270.

Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica 64*, 575–596.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Milhøj, A. (1987). A Multiplicative Parametrization of ARCH Models. Research Report 101, University of Copenhagen: Institute of Statistics.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica 59*, 347–370.

Pantula, S. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 71–73.

Psaradakis, Z. and E. Tzavalis (1999). On regression-based tests for persistence in logarithmic volatility models. *Econometric Reviews 18*, 441–448.

SAS Institute Inc. (2013). *JMP Version 10*. Cary, NC: SAS Institute Inc.

Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach. *The Annals of Statistics 34*, 2449–2495.

Sucarrat, G. (2012). *AutoSEARCH: General-to-Specific (GETS) Model Selection*. R package version 1.2.

Sucarrat, G. (2013). Models of Zero-Augmented Financial Return with Time-Varying Zero-Probabilities. Work in progress.

Sucarrat, G. and Á. Escribano (2010). The Power Log-GARCH Model. Universidad Carlos III de Madrid Working Paper 10-13 in the Economic Series, June 2010. http://e-archivo.uc3m.es/bitstream/10016/8793/1/we1013.pdf.

Sucarrat, G. and Á. Escribano (2012). Automated Model Selection in Finance: General-to-Specific Modelling of the Mean and Volatility Specifications. *Oxford Bulletin of Economics and Statistics 74*, 716–735.

Sucarrat, G. and Á. Escribano (2013). Unbiased QML Estimation of Log-GARCH Models in the Presence of Zero Errors. Work in progress.

Teräsvirta, T. (2009). An introduction to univariate GARCH models. In T. Mikosch, T. Kreiss, J.-P. Davis, R. Andersen, and T. Gustav (Eds.), *Handbook of Financial Time Series*. Berlin: Springer.

Teräsvirta, T. (2012). Nonlinear Models for Autoregressive Conditional Heteroskedasticity. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*. New Jersey: Wiley.

Weiss, A. (1986). Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory 2*, 107–131.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. *Econometrica 48*, 817–838.

Wintenberger, O. (2012). QML estimation of the continuously invertible EGARCH(1,1) model. Unpublished working paper.

Yu, H. (2007). High moment partial sum processes of residuals in ARMA models and their applications. *Journal of Time Series Analysis 28*, 72–91.

A Closed form expressions of $\mathbb{E}|z_t|^c$ when $z_t \sim GED(\tau)$ and when $z_t \sim t(\nu)$

The expectation of the absolute value of a GED variate ε raised to the power c is readily available, since it can be showed that $|\varepsilon|^c$ is $Gamma(1/2, \tau)$ distributed where τ is the GED shape parameter ($\tau = 2$ yields the standard normal), see Harvey and Chakravarty (2008). Accordingly:

$$\mathbb{E}|\varepsilon|^c = \frac{2^{c/\tau}\Gamma[(c+1)/\tau]}{\Gamma(1/\tau)}, \quad c > -1, \quad \tau > 0. \tag{31}$$

In particular, $Var(\varepsilon) = \mathbb{E}|\varepsilon|^2 = 2^{2/\tau}\Gamma(3/\tau)/\Gamma(1/\tau)$, and so for the standardised (zero-mean, unit variance) GED variate $z = \varepsilon/\sqrt{Var(\varepsilon)}$ we obtain:

$$\mathbb{E}|z|^c = \frac{\Gamma(1/\tau)^{c/2}\Gamma[(c+1)/\tau]}{\Gamma(3/\tau)^{c/2}\Gamma(1/\tau)}, \quad c > -1, \quad \tau > 0. \tag{32}$$

Using the property that a t-variate with $\nu > -1$ degrees of freedom can be written as $X\nu^{1/2}/Y_\nu^{1/2}$ where X is a standard normal and Y_ν is a Chi-squared with ν degrees of freedom, and where X and Y are independent, then the expectation of the absolute value of a t variate ε is:

$$\mathbb{E}|\varepsilon|^c = \frac{\nu^{c/2}\Gamma(c/2+1/2)\Gamma(-c/2+\nu/2)}{\Gamma(1/2)\Gamma(\nu/2)}, \quad -1 < c < \nu, \tag{33}$$

see Harvey and Shephard (1996, p. 434). Next, since $Var(\varepsilon) = \nu/(\nu-2)$ we obtain (by setting $z = \varepsilon/\sqrt{Var(\varepsilon)}$):

$$\mathbb{E}|z|^c = \frac{(\nu-2)^{c/2}\Gamma(c/2+1/2)\Gamma(-c/2+\nu/2)}{\Gamma(1/2)\Gamma(\nu/2)}, \quad -1 < c < \nu, \quad \nu \neq 2. \tag{34}$$

B $\mathbb{E}(\epsilon_t^s)$ and $\mathbb{E}(\epsilon_t^2\epsilon_{t-j}^2)$ for the univariate log-GARCH(1,1) model

For the univariate log-GARCH(1,1) model the unconditional variance of $\{\epsilon_t\}$, and the autocovariances and autocorrelations of $\{\epsilon_t^2\}$, are all made up of $\mathbb{E}(\epsilon_t^2)$, $\mathbb{E}(\epsilon_t^2\epsilon_{t-j}^2)$ and $\mathbb{E}(\epsilon_t^4)$. If $|\alpha_1+\beta_1| < 1$, then the sth. unconditional moment $\mathbb{E}(\epsilon_t^s)$ (assuming it exists) is given by

$$\begin{aligned}\mathbb{E}(\epsilon_t^s) &= \mathbb{E}(z_t^s)\cdot\exp\left(\frac{s}{2}\cdot\frac{\alpha_0}{1-\alpha_1-\beta_1}\right)\\ &\quad\cdot\prod_{i=1}^{\infty}\mathbb{E}\left(|z_{t-i}|^{s\alpha_1(\alpha_1+\beta_1)^{i-1}}\right),\end{aligned} \tag{35}$$

whereas for $j = 1, 2, \ldots$ the formula for $\mathbb{E}(\epsilon_t^2\epsilon_{t-j}^2)$ is

$$\begin{aligned}\mathbb{E}(\epsilon_t^2\epsilon_{t-j}^2) &= \exp\left[\frac{\alpha_0(1+(\alpha_1+\beta_1)^j)}{(1-\alpha_1-\beta_1)} + \sum_{i=1}^{j}\alpha_0(\alpha_1+\beta_1)^{i-1}\right]\\ &\quad\cdot\prod_{i=1}^{j}\mathbb{E}\left((z_{t-i}^2)^{\alpha_1(\alpha_1+\beta_1)^{i-1}}\right)\\ &\quad\cdot\prod_{i=1}^{\infty}\mathbb{E}\left((z_{t-j-i}^2)^{\alpha_1(\alpha_1+\beta_1)^{i-1}\cdot[1+(\alpha_1+\beta_1)^j]}\right).\end{aligned} \tag{36}$$

C Proofs of Lemma 1 and Theorem 1

C.1 Proof of Lemma 1

Proof of Lemma 1. We will only prove the second, more fine tuned statement, since the proof of the first statement follows along the same lines. Let

$$d_t = d_{t,T} := (\hat{u}_t - \bar{\hat{u}}_T) - (u_t - \bar{u}_T).$$

We have that

$$\sqrt{T}\Delta_T := \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[e^{\hat{u}_t-\bar{\hat{u}}_t} - e^{u_t-\bar{u}_t}\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}e^{u_t-\bar{u}_t}(e^{d_t}-1).$$

Choose a δ with $1 < \delta \leq \delta_M$ so that the Hölder conjugate of δ given by $\gamma = \delta_M/(\delta_M - 1)$ is an integer divisible by two. This can always be arranged, since $x/(x-1)$ goes to infinity when $x \to 1^+$. By the Hölder inequality, we have the basic bound

$$|\sqrt{T}\Delta_T| = \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e^{u_t - \bar{u}_t}(e^{d_t} - 1) \right| \leq \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} |e^{\delta u_t - \delta \bar{u}_t}| \right)^{1/\delta} Q_T^{1/\gamma}.$$

where $Q_T = T^{-1/2} \sum_{t=1}^{T} |e^{d_t} - 1|^\gamma$. We hence have

$$\begin{aligned}
(\sqrt{T}\Delta_T)^\delta &\leq \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} e^{\delta u_t - \delta \bar{u}_t} - \mathbb{E}e^{\delta u_1} + \mathbb{E}e^{\delta u_1} \right) Q_T^{\delta/\gamma} \\
&= \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} e^{\delta u_t - \delta \bar{u}_t} - \mathbb{E}e^{\delta u_1} \right) Q_T^{\delta/\gamma} + Q_T^{\delta/\gamma} \mathbb{E}e^{\delta u_1}.
\end{aligned}$$

The assumed existence of δ_M enables us to use the Central Limit Theorem and Slutsky's Theorem to conclude that because $\delta_M \geq \delta$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(e^{\delta u_t - \delta \bar{u}_t} - \mathbb{E}e^{\delta u_1} \right)$$

is asymptotically normal, and hence $O_P(1)$. We can therefore conclude that $\sqrt{T}\Delta_T = o_P(1)$ if we show that $Q_T = o_P(1)$.

Because Yu (2007) proves that $T^{-1/2} \sum_{t=1}^{T} d_t^\gamma = o_P(1)$, we wish to connect $\exp(d_t) - 1$ to $|d_t|$ through the inequality

$$e^x - 1 \leq \frac{7}{4}|x|, \tag{37}$$

which is valid when $|x| < 1$. By the structure of the ARMA-residuals revealed in Yu (2007), this is clearly justified as $T \to \infty$. However, the exact result that we need is not contained in Yu (2007), but will be given by Lemma 2 which is stated below. For T' as in the Lemma and for $Q \in \mathbb{N}$ with $T' + Q < T$ we have that

$$Q_T = T^{-1/2} \sum_{t=1}^{T} |e^{d_t} - 1|^\gamma = T^{-1/2} \sum_{t=1}^{T'+Q-1} |e^{d_t} - 1|^\gamma + T^{-1/2} \sum_{t=T'+Q}^{T} |e^{d_t} - 1|^\gamma.$$

Clearly, $Q_{T'+Q,T} := T^{-1/2} \sum_{t=1}^{T'+Q-1} |e^{d_t} - 1|^\gamma$ is non-negative and $o_P(1)$. For an $\varepsilon > 0$, we have that

$$\begin{aligned}
P(Q_T > \varepsilon) \leq P(Q_T > \varepsilon, |M_1| < T' + Q, |M_2| < \sqrt{T}) \\
+ P(|M_1| \geq T' + Q) + P(|M_2| \geq \sqrt{T}).
\end{aligned}$$

Because M_1 and M_2 are random variables, and hence finite, we can make $P(|M_1| \geq T' + Q) + P(|M_2| \geq \sqrt{T})$ arbitrarily small by increasing T and Q. By the non-negativity of $Q_{T'+Q,T}$, we have that

$$
\begin{aligned}
P&\left(Q_T > \varepsilon, |M_1| < T' + Q, |M_2| < \sqrt{T}\right) \\
&\leq P\left(Q_{T'+Q,T} > \varepsilon/2, |M_1| < T' + Q, |M_2| < \sqrt{T}\right) \\
&+ P\left(T^{-1/2} \sum_{t=T'+Q}^{T} |e^{d_t} - 1|^\gamma > \varepsilon/2, |M_1| < T' + Q, |M_2| < \sqrt{T}\right).
\end{aligned}
$$

The first term on the right hand side of the above inequality converges to zero as $T \to \infty$, so we only need to deal with the second term. Here eq. (37) is valid, and so

$$
\begin{aligned}
P&\left(T^{-1/2} \sum_{t=T'+Q}^{T} |e^{d_t} - 1|^\gamma > \varepsilon/2, |M_1| < T' + Q, |M_2| < \sqrt{T}\right) \\
&\leq P\left((7/8)^\gamma T^{-1/2} \sum_{t=T'+Q}^{T} |d_t|^\gamma > \varepsilon/2, |M_1| < T' + Q, |M_2| < \sqrt{T}\right) \\
&\leq P\left(T^{-1/2} \sum_{t=1}^{T} |d_t|^\gamma > (8/7)^\gamma \varepsilon/2\right),
\end{aligned}
$$

which converges to zero by Theorem 3 in Yu (2007). □

Lemma 2. Under Assumptions **A1** – **A3**, there exists a $T' \in \mathbb{N}$ and random variables M_1, M_2 with $|d_{t,T}| < M_1/t + M_2/\sqrt{T}$ for all $T \geq T'$.

Proof. Using the notation and basic results of Section 3 in Yu (2007), we have that

$$
d_t = d_{t,T} = (\hat{u}_t - \bar{\hat{u}}_T) - (u_t - \bar{u}_T) = \Lambda_t(\hat{a}, \hat{b}, \hat{c}) + \bar{\Lambda}_T(\hat{a}, \hat{b}, \hat{c})
$$

where

$$
(\hat{a}, \hat{b}, \hat{c}) = (\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{\phi} - \phi), \sqrt{n}(\hat{\mu} - \mu)).
$$

We have that $\bar{\Lambda}_T(x) = T^{-1} \sum_{t=1}^{T} \Lambda_t(x)$ is defined by a function $\Lambda : \mathbb{R}^{P+Q+1} \mapsto \mathbb{R}$. Let $x = (a, b, c)$ in which $a \in \mathbb{R}^Q$, $b \in \mathbb{R}^P$ and $c \in \mathbb{R}$. We have that

$$
\Lambda_t(x) = Y_t\left(\theta + \frac{1}{\sqrt{T}} a\right) - \frac{1}{\sqrt{T}} \xi_t(a, b) + \frac{1}{\sqrt{T}} Z_t(a, b, c)
$$

in which $\theta = (\theta_1, \theta_2, \dots, \theta_Q)$ and

$$
\begin{aligned}
Y_t(a) &= -\psi_t(a)u_0 - \{\psi_{t+1}(a) + \psi_t(a)a_1\}\, u_{-1} \\
&\quad - \cdots - \{\psi_{t+1-1}(a) + \psi_{t+q-2}(a)u_1 + \cdots + \psi_t(a)a_{q-1}\}\, u_{-q+1}, \\
\xi_t(a,b) &= \sum_{i=1}^{p} b_i \sum_{j=0}^{t-1} \psi_j \left(\theta + \frac{1}{\sqrt{T}}a\right) X_{t-i-j} + \sum_{i=1}^{q} a_i \sum_{j=0}^{t-1} \psi_j \left(\theta + \frac{1}{\sqrt{T}}a\right) u_{t-i-j}, \\
Z_t(a,b,c) &= -\left(1 - \sum_{i=1}^{p} \phi_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{p} b_i\right) c \sum_{j=0}^{t-1} \psi_j \left(\theta + \frac{1}{\sqrt{T}}a\right)
\end{aligned}
$$

where ψ_i are the coefficients of the power series expansion $1/\Phi(z) = \sum_{i=0}^{\infty} \psi_i(\theta) z^i$.

The basic logic of Bai (1994) and Yu (2007) is to use the assumed consistency of an estimator of the ARMA parameters to justify the assumption that for sufficiently large T, we have that for any $r > 0$, we eventually have $(\hat{a}, \hat{b}, \hat{c}) \in B_r$ where B_r is a ball of radius r in $\mathbb{R}^{P+Q+1}$. We therefore reach our conclusion by bounding the supremum of $|\Lambda_t(x) + \bar{\Lambda}_T(x)|$ over $x \in B_r$.

We will first deal with the terms in Y_t and ξ_t. The bounding of the corresponding terms in $\bar{\Lambda}_T(x)$ will then follow as an immediate consequence. The proof of Lemmas 3 and 4 in Yu (2007) shows that for sufficiently small $r > 0$ there exists some constants $M, \varepsilon > 0$, $0 < \beta < 1$ so that

$$
A_t := \sup_{|a| \le r} \left| Y_t \left(\theta + \frac{1}{\sqrt{T}}a\right) \right| \le \frac{M \max(|\theta| + \varepsilon, 1)}{1 - \beta} \beta^t \sum_{-q+1 \le i \le 0} |u_i|
$$

and

$$
\bar{A}_T := T^{-1} \sum_{t=1}^{T} A_t \le T^{-1} \frac{M \max(|\theta| + \varepsilon, 1)}{1 - \beta} \frac{1 - \beta^{T+1}}{1 - \beta} \sum_{-q+1 \le i \le 0} |u_i|,
$$

both of which are of the desired form. We also have that

$$
\begin{aligned}
B_t := \frac{1}{\sqrt{T}} \sup_{|u| \le r, |v| \le r} |\xi_t(a,b)| &\le \frac{bM}{\sqrt{T}} \sum_{j=0}^{t-1} \left\{ \sum_{i=1}^{p} \beta^j |X_{t-i-j}| + \sum_{i=1}^{q} \beta^j |u_{t-i-j}| \right\} \\
&\le \frac{bM}{\sqrt{T}} \sum_{j=0}^{\infty} \beta^j \left\{ \sum_{i=1}^{p \vee q} |X_{j-i+1}| + \sum_{i=1}^{p \vee q} |u_{j-i+1}| \right\}.
\end{aligned}
$$

Because X_t and u_t have finite first order moments, this provides the desired bound for both B_t and its average.

We now work with

$$\begin{aligned} C_t(x) &= Z_t(a,b,c) - \frac{1}{T}\sum_{s=1}^{T} Z_s(a,b,c) \\ &= D_T(x)\left[\sum_{j=0}^{t-1}\psi_j\left(\theta + \frac{1}{\sqrt{T}}a\right) - \frac{1}{T}\sum_{s=1}^{T}\sum_{k=0}^{s-1}\psi_k\left(\theta + \frac{1}{\sqrt{T}}a\right)\right], \end{aligned}$$

where

$$D_T(x) := \left(1 - \sum_{i=1}^{p}\phi_i - \frac{1}{\sqrt{T}}\sum_{i=1}^{p} b_i\right) c.$$

Notice that for any numbers $\{x_j : j \in \mathbb{Z}\}$, we have

$$\begin{aligned} \sum_{j=0}^{t-1} x_j - \frac{1}{T}\sum_{s=1}^{T}\sum_{k=0}^{s-1} x_k &= \sum_{j=0}^{t-1} x_j - \frac{1}{T}\sum_{\substack{1\le s\le T\\ 0\le k\le s-1}} x_k = \sum_{j=0}^{t-1} x_j - \frac{1}{T}\sum_{k=0}^{T-1}\sum_{s=k+1}^{T} x_k \\ &= \sum_{j=0}^{t-1} x_j - \frac{1}{T}\sum_{j=0}^{T}(T-j-1)x_j = -\sum_{j=t}^{T} x_j + \frac{1}{T}\sum_{j=0}^{T}(j+1)x_j \end{aligned}$$

so that Lemma 1 (i) of Yu (2007) implies that

$$\sup_{x\in B_r}|C_t(x)| \le \sup_{x\in B_r}|D_T(x)| M\left(\sum_{j=t}^{T}\beta^j + \frac{1}{T}\sum_{j=0}^{T}(j+1)\beta^j\right).$$

As $0 < \beta < 1$ we have

$$\begin{aligned} \sum_{j=t}^{T}\beta^j = \beta^t\sum_{j=0}^{T-t}\beta^j \le \beta^t\sum_{j=0}^{\infty}\beta_j = \frac{\beta^t}{1-\beta}, \\ \sum_{j=0}^{T}(j+1)\beta^j \le \sum_{j=0}^{\infty}(j+1)\beta^j = \frac{1}{(1-\beta)^2}, \end{aligned}$$

and can conclude that

$$\sup_{x\in B_r}|C_t(x)| \le M\left(\frac{\beta^t}{1-\beta} + \frac{1}{T}\frac{1}{(1-\beta)^2}\right)\sup_{x\in B_r}|D_T(x)|$$

which is also of the desired form as $\sup_{x\in B_r}|D_T(x)|$ is non-stochastic and bounded.
□

C.2 Proof of b) in Theorem 1

Proof of b) in Theorem 1. Lemma 1 and the smoothness of the logarithm function imply that

$$\hat{\tau}_T = -\ln\left[\frac{1}{T}\sum_{t=1}^{T}\exp(\widehat{u}_t - \overline{\widehat{u}}_T)\right] \quad \text{and} \quad \tilde{\tau}_T = -\ln\left[\frac{1}{T}\sum_{t=1}^{T}\exp(u_t - \overline{u}_T)\right]$$

have the same behaviour up to either $o_P(1)$ or $o_P(T^{-1/2})$. Let $\tau = \mathbb{E}\ln(z_1^2) = -\ln\mathbb{E}e^{u_t}$. Since b) in Lemma 1 holds, we have that

$$\sqrt{T}(\hat{\tau}_T - \tau) = \sqrt{T}(\tilde{\tau}_T - \tau) + o_P(1).$$

Slutsky's Theorem hence implies that we only need show that $\tilde{\Delta}_T := \sqrt{T}(\tilde{\tau}_T - \tau)$ is asymptotically normal. We have that

$$\tilde{\tau}_T = -\log\frac{1}{T}\sum_{t=1}^{T}e^{u_t - \bar{u}_T} = \bar{u}_T - \log\frac{1}{T}\sum_{t=1}^{T}e^{u_t}.$$

Hence,

$$\tilde{\Delta}_T = \sqrt{T}\bar{u}_T + \sqrt{T}\left[f\left(\frac{1}{T}\sum_{t=1}^{T}e^{u_t}\right) - f(\mathbb{E}e^{u_1})\right]$$

where $f(x) = -\ln x$, with $f'(x) = -1/|x|$. By the smoothness of f, the delta method implies that

$$\begin{aligned}
\tilde{\Delta}_T &= \sqrt{T}\bar{u}_T + f'(\mathbb{E}e^{u_1})\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T}e^{u_t} - \mathbb{E}e^{u_1}\right] + o_P(1) \\
&= (f'(\mathbb{E}e^{u_1}), 1)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} e^{u_t} - \mathbb{E}e^{u_1} \\ u_t \end{pmatrix} + o_P(1).
\end{aligned}$$

By the Multivariate Central Limit Theorem, we have that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} e^{u_t} - \mathbb{E}e^{u_1} \\ u_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \operatorname{Var} e^{u_1} & \mathbb{E}u_1 e^{u_1} \\ \mathbb{E}u_1 e^{u_1} & \operatorname{Var} u_1 \end{pmatrix}\right)$$

where we used that $\mathbb{E}u_1 = 0$ so that $\operatorname{Cov}(u_1, e^{u_1}) = \mathbb{E}u_1 e^{u_1}$. Hence,

$$\tilde{\Delta}_T \xrightarrow{d} f'(\mathbb{E}e^{u_1})X + Y.$$

which is mean zero normal with variance equal to

$$\begin{aligned}\zeta^2 &= \left(f'(\mathbb{E}e^{u_1})\right)^2 \operatorname{Var} X + \operatorname{Var} Y + 2f'(\mathbb{E}e^{u_1})\operatorname{Cov}(X,Y)\\ &= \frac{\operatorname{Var}[\exp(u_1)]}{[\mathbb{E}\exp(u_1)]^2} + \operatorname{Var}(u_1) - 2\frac{\mathbb{E}[u_1\exp(u_1)]}{\mathbb{E}\exp(u_1)}.\end{aligned}$$

Using the equalities

$$\begin{aligned}\operatorname{Var}(u_1) &= \mathbb{E}[(\ln z^2)^2] - [\mathbb{E}\ln(z^2)]^2\\ \operatorname{Var}[\exp(u_1)] &= \frac{1}{\{\exp[\mathbb{E}\ln(z^2)]\}^2}\cdot(\mathbb{E}z^4 - 1)\\ \mathbb{E}\exp(u_1) &= \frac{1}{\exp[\mathbb{E}\ln(z^2)]}\\ \mathbb{E}[u_1\exp(u_1)] &= \frac{1}{\exp[\mathbb{E}\ln(z^2)]}\cdot\left\{\mathbb{E}[(\ln z^2)z^2] - \mathbb{E}\ln(z^2)\right\}\end{aligned}$$

we see that

$$\zeta^2 = \mathbb{E}[(\ln z_t^2)^2] - [\mathbb{E}(\ln z_t^2)]^2 + (\mathbb{E}(z_t^4) - 1) - 2\mathbb{E}[(\ln z_t^2)z_t^2] + 2\mathbb{E}(\ln z_t^2).$$

The assumption **A4**b) implies that $|\mathbb{E}(z_t^4)| < \infty$, and the Cauchy-Schwarz inequality implies that $|\mathbb{E}[(\ln z^2)z^2]|^2 \leq (\mathbb{E}[(\ln z^2)^2])(\mathbb{E}z^4)$, whose right-hand side terms are assumed to be finite. Hence, ζ^2 is finite. □

D Proofs of propositions

The following result from Gradshteyn and Ryzhik (2007, section 0.25) will be used in the proofs of the remaining propositions.

Lemma 3. Suppose $\{a_i\}$ is a sequence of finite, positive and non-zero real numbers. A sufficient condition for the product $\prod_{i=1}^{\infty} a_i$ to converge to a finite, non-zero number is that the series $\sum_{i=1}^{\infty} |a_i - 1|$ converges.

Proof of Proposition 1. From equation (35) in Appendix B it follows that $\mathbb{E}(z_t^s)$ must be finite for $\mathbb{E}(\epsilon_t^s)$ to exist. This is the case for both $z_t \sim GED(\tau)$ and $z_t \sim t(\nu)$ under the conditions (i.e. $\tau > 1, \nu \neq 2$ and $s < \nu$) stated in the proposition. Next, $\mathbb{E}\left(|z_{t-i}|^{2\alpha_1(\alpha_1+\beta_1)^{i-1}}\right)$ must also be finite for each $i = 1, 2, \ldots$. For $z_t \sim GED(\tau), \tau > 1$, then $\mathbb{E}(|z_t|^c) < \infty$ for $c > -1$, see Appendix A. For $z_t \sim t(\nu), \nu > 2$, then $\mathbb{E}(|z_t|^c) < \infty$ for $-1 < c < \nu, c \neq 2$, see Appendix A. So if $|\alpha_1 + \beta_1| < 1$ and $2\alpha_1(\alpha_1+\beta_1)^{i-1} \in (-1, 2)$ for each i, then $\mathbb{E}\left(|z_{t-i}|^{2\alpha_1(\alpha_1+\beta_1)^{i-1}}\right) < \infty$ for each $i = 1, 2, \ldots$ Finally, due to Lemma 3, the infinite product converges and so $\mathbb{E}(\epsilon_t^s) < \infty$. □

Proof of Proposition 2. By definition, absolute summability of the matrix sequence $\{\Psi_i\}$ means $\sum_{i=1}^{\infty} |\psi_{i,mn}| < \infty$ for each $m, n \in \{1, 2, \ldots, M\}$. Next, a sufficient condition for an infinite product $\prod_{i=1}^{\infty} a_i$ to converge to a finite, nonzero number is that the series $\sum_{i=1}^{\infty} |a_i - 1|$ converges (Gradshteyn and Ryzhik (2007, section 0.25)). Since $\mathbb{E}\left[|z_{1,t-i}|^{s\psi_{i,m1}}|z_{2,t-i}|^{s\psi_{i,m2}} \cdots |z_{M,t-i}|^{s\psi_{i,mM}}\right] \to 1$ as $i \to \infty$ due to absolute summability, it follows that $|a_i - 1| \to 0$ as $i \to \infty$. Accordingly, if $a_i = \mathbb{E}\left[|z_{1,t-i}|^{s\psi_{i,m1}}|z_{2,t-i}|^{s\psi_{i,m2}} \cdots |z_{M,t-i}|^{s\psi_{i,mM}}\right] < \infty$ for each i, and if $|\mathbb{E}(z_{m,t}^s)| < \infty$, it follows that $\mathbb{E}(\epsilon_{mt}^s)$ exists. $\square$

Table 1: Unconditional autocorrelations of $\{\epsilon_t^2\}$ for the log-GARCH(1,1) model and for the GARCH(1,1) model, with $\alpha_0 = 0.005$, $\alpha_1 = 0.05$ and $\beta_1 \in \{0.9, 0.94\}$

	Log-GARCH(1,1), $\beta_1 = 0.9$			Log-GARCH(1,1), $\beta_1 = 0.94$			GARCH(1,1)	
Lag	N	$GED(1.1)$	$t(5)$	N	$GED(1.1)$	$t(5)$	$\beta_1 = 0.9$	$\beta_1 = 0.94$
1	0.096	0.062	0.031	0.215	0.135	0.070	0.073	0.155
2	0.091	0.058	0.029	0.212	0.133	0.069	0.069	0.153
3	0.086	0.055	0.028	0.209	0.132	0.068	0.065	0.152
4	0.081	0.052	0.026	0.207	0.130	0.067	0.062	0.150
5	0.077	0.049	0.025	0.204	0.128	0.066	0.059	0.149
6	0.073	0.047	0.023	0.201	0.126	0.065	0.056	0.147
7	0.069	0.044	0.022	0.199	0.124	0.065	0.053	0.146
8	0.066	0.042	0.021	0.196	0.123	0.064	0.051	0.144
9	0.062	0.040	0.020	0.194	0.121	0.063	0.048	0.143
10	0.059	0.038	0.019	0.191	0.119	0.062	0.046	0.142
11	0.056	0.036	0.018	0.189	0.118	0.061	0.043	0.140
12	0.053	0.034	0.017	0.187	0.116	0.060	0.041	0.139
13	0.050	0.032	0.016	0.184	0.114	0.060	0.039	0.137
14	0.048	0.030	0.015	0.182	0.113	0.059	0.037	0.136
15	0.045	0.029	0.014	0.180	0.111	0.058	0.035	0.135
16	0.043	0.027	0.014	0.177	0.110	0.057	0.034	0.133
17	0.041	0.026	0.013	0.175	0.108	0.057	0.032	0.132
18	0.039	0.024	0.012	0.173	0.107	0.056	0.030	0.131
19	0.037	0.023	0.012	0.171	0.105	0.055	0.029	0.129
20	0.035	0.022	0.011	0.169	0.104	0.054	0.027	0.128
21	0.033	0.021	0.010	0.167	0.103	0.054	0.026	0.127
22	0.031	0.020	0.010	0.165	0.101	0.053	0.025	0.125
23	0.030	0.019	0.009	0.163	0.100	0.052	0.023	0.124
24	0.028	0.018	0.009	0.161	0.099	0.052	0.022	0.123
25	0.027	0.017	0.008	0.159	0.097	0.051	0.021	0.122

The label N means $z_t \sim N(0,1)$, $GED(1,1)$ means $z_t \sim GED$ with shape parameter $\tau = 1.1$ ($\tau = 2$ gives N, $\tau < 2$ gives densities that are more fat-tailed than the normal), whereas $t(5)$ means z_t is standardised t with 5 degrees of freedom.

Table 2: Finite sample properties of Gaussian QML via the ARMA(1,1) representation using a $N(0, \sigma_u^2)$ density in the ARMA error u_t. DGP: Log-GARCH(1,1) with $z_t \sim N(0,1), \mathbb{E}(\ln z_t^2) = -1.27$

Sample size T:	1000	10000	100000	1000	10000	100000	1000	10000	100000
DGP $(\alpha_0, \alpha_1, \beta_1)$:	$0, 0.3, 0.1$	$0, 0.3, 0.1$	$0, 0.3, 0.1$	$0, 0.1, 0.8$	$0, 0.1, 0.8$	$0, 0.1, 0.8$	$0, 0.05, 0.94$	$0, 0.05, 0.94$	$0, 0.05, 0.94$
$m(\hat{\alpha}_0)$	-0.010	-0.001	0.000	-0.036	-0.001	0.000	-0.134	-0.004	0.000
$se(\hat{\alpha}_0)$	0.083	0.026	0.008	0.097	0.023	0.005	0.552	0.018	0.004
$m(\hat{\alpha}_1)$	0.299	0.300	0.300	0.104	0.101	0.100	0.055	0.051	0.050
$se(\hat{\alpha}_1)$	0.031	0.010	0.003	0.023	0.007	0.002	0.015	0.004	0.001
$ase(\hat{\alpha}_1)$	0.032	0.010	0.003	0.022	0.007	0.002	0.012	0.004	0.001
$m(\hat{\beta}_1)$	0.092	0.098	0.100	0.763	0.797	0.800	0.912	0.938	0.940
$se(\hat{\beta}_1)$	0.103	0.033	0.010	0.090	0.018	0.005	0.089	0.006	0.002
$ase(\hat{\beta}_1)$	0.101	0.032	0.010	0.053	0.017	0.005	0.015	0.005	0.001
$c(\hat{\alpha}_1, \hat{\beta}_1)$	-0.391	-0.375	-0.347	-0.506	-0.792	-0.774	-0.183	-0.934	-0.925
$ac(\hat{\alpha}_1, \hat{\beta}_1)$	-0.398	-0.398	-0.398	-0.778	-0.778	-0.778	-0.923	-0.923	-0.923
$m(\hat{\mathbb{E}}(\ln z_t^2))$	-1.267	-1.270	-1.270	-1.276	-1.275	-1.271	-1.292	-1.285	-1.273
$se(\hat{\mathbb{E}}(\ln z_t^2))$	0.054	0.017	0.006	0.060	0.079	0.007	0.101	0.159	0.038
$ase(\hat{\mathbb{E}}(\ln z_t^2))$	0.054	0.017	0.005	0.054	0.017	0.005	0.054	0.017	0.005

$m(x)$, sample mean of the estimates of estimator x. $se(x)$, sample standard deviation (division by R instead of $R-1$, where $R = 1000$ is the number of replications). $ase(x)$, asymptotic standard error of x (computed as $\sqrt{av(x)}/\sqrt{n}$, where $av(x)$ is the asymptotic variance of x). $c(x, y)$, sample correlation between estimators x and y. $ac(x, y)$, asymptotic correlation between x and y (computed as $acov(x, y)/\sqrt{av(x) \cdot av(y)}$, where $acov(x, y)$ is the asymptotic covariance between x and y). The asymptotic standard errors and correlations are those implied by Brockwell and Davis (2006, pp. 259-260).

Table 3: Finite sample properties of Gaussian QML via the ARMA(1,1) representation using a $N(0, \sigma_u^2)$ density in the ARMA error u_t. DGP: Log-GARCH(1,1) with $\eta_t \sim t(5), \mathbb{E}(\ln z_t^2) = -1.56$

Sample size T:	1000	10000	100000	1000	10000	100000	1000	10000	100000
DGP $(\alpha_0, \alpha_1, \beta_1)$:	$0, 0.3, 0.1$	$0, 0.3, 0.1$	$0, 0.3, 0.1$	$0, 0.1, 0.8$	$0, 0.1, 0.8$	$0, 0.1, 0.8$	$0, 0.05, 0.94$	$0, 0.05, 0.94$	$0, 0.05, 0.94$
$m(\hat{\alpha})$	-0.010	-0.004	-0.001	-0.039	-0.003	-0.001	-0.163	-0.006	0.000
$se(\hat{\alpha})$	0.112	0.034	0.012	0.113	0.019	0.006	0.354	0.035	0.009
$m(\hat{\alpha}_1)$	0.297	0.299	0.300	0.104	0.100	0.100	0.056	0.051	0.050
$se(\hat{\alpha}_1)$	0.032	0.010	0.003	0.024	0.007	0.002	0.015	0.004	0.001
$ase(\hat{\alpha}_1)$	0.032	0.010	0.003	0.022	0.007	0.002	0.012	0.004	0.001
$m(\hat{\beta}_1)$	0.096	0.098	0.100	0.768	0.797	0.800	0.911	0.938	0.940
$se(\hat{\beta}_1)$	0.099	0.031	0.010	0.087	0.017	0.005	0.055	0.007	0.002
$ase(\hat{\beta}_1)$	0.101	0.032	0.010	0.053	0.017	0.005	0.015	0.005	0.001
$c(\hat{\alpha}_1, \hat{\beta}_1)$	-0.396	-0.374	-0.400	-0.567	-0.773	-0.802	-0.631	-0.921	-0.936
$ac(\hat{\alpha}_1, \hat{\beta}_1)$	-0.398	-0.398	-0.398	-0.778	-0.778	-0.778	-0.923	-0.923	-0.923
$m(\hat{\mathbb{E}}(lnz_t^2))$	-1.568	-1.567	-1.568	-1.563	-1.570	-1.568	-1.596	-1.593	-1.577
$se(\hat{\mathbb{E}}(lnz_t^2))$	0.087	0.026	0.009	0.095	0.040	0.010	0.140	0.296	0.134
$ase(\hat{\mathbb{E}}(lnz_t^2))$	0.085	0.027	0.009	0.085	0.027	0.009	0.085	0.027	0.009

$m(x)$, sample mean of the estimates of estimator x. $se(x)$, sample standard deviation (division by R instead of $R-1$, where $R = 1000$ is the number of replications). $ase(x)$, asymptotic standard error of x (computed as $\sqrt{av(x)}/\sqrt{n}$, where $av(x)$ is the asymptotic variance of x). $c(x, y)$, sample correlation between estimators x and y. $ac(x, y)$, asymptotic correlation between x and y (computed as $acov(x, y)/\sqrt{av(x) \cdot av(y)}$, where $acov(x, y)$ is the asymptotic covariance between x and y). The asymptotic standard errors and correlations are those implied by Brockwell and Davis (2006, pp. 259-260).

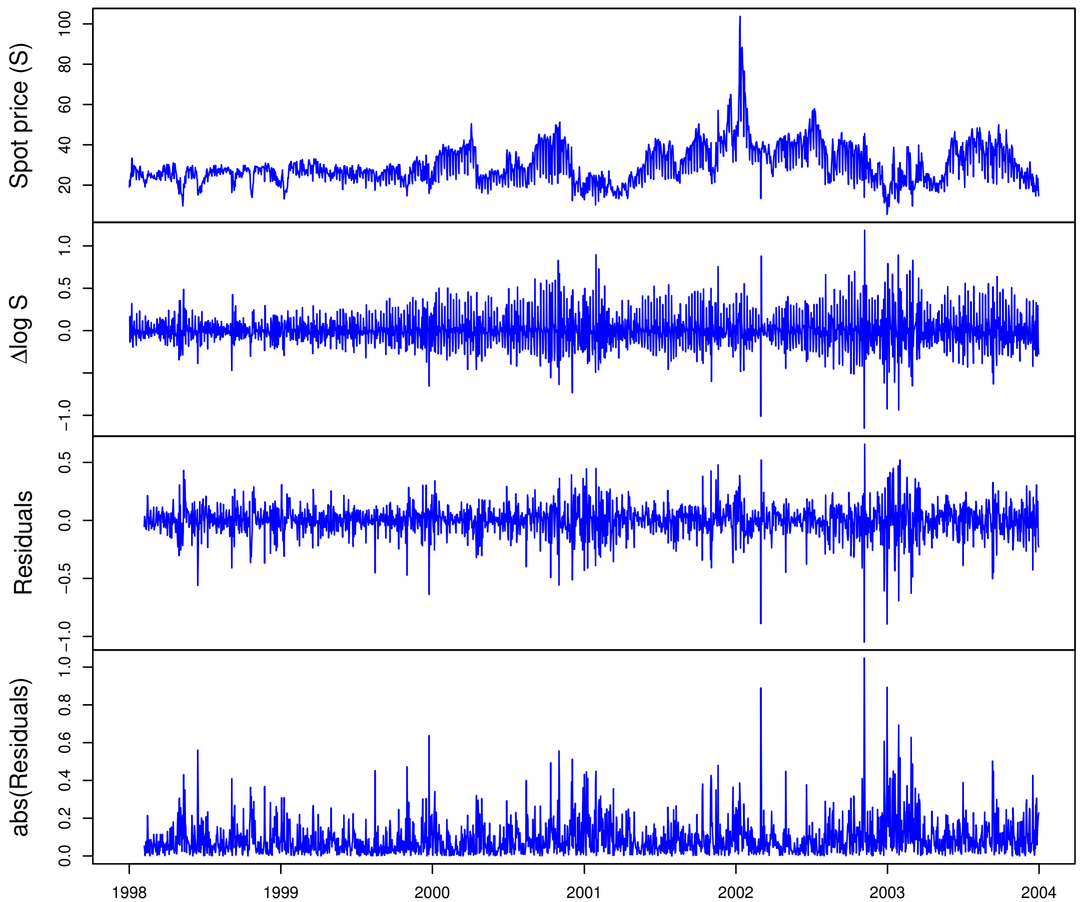

Figure 1: Daily electricity price (upper graph), log-returns (second graph), and the residuals (third graph) and absolute residuals (bottom graph) of the conditional mean model (27) for Spain (1 January 1998 - 31 December 2003).

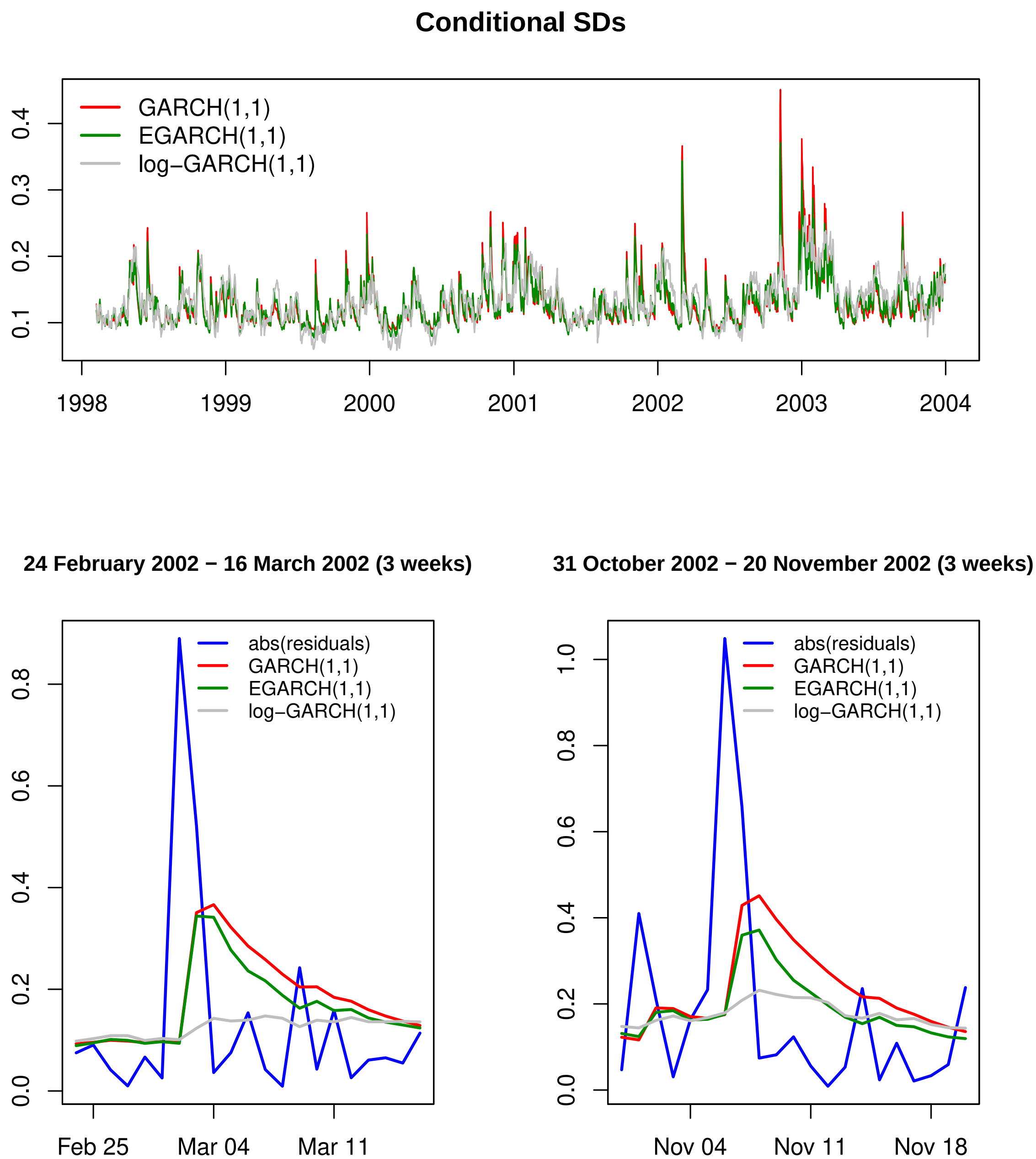

Figure 2: Conditional standard deviations (upper graph) of a GARCH(1,1), an EGARCH(1,1) and a log-GARCH(1,1) fitted to the residuals of the mean specification (27), and two snapshots (lower graphs) of the same conditional standard deviations against the absolute residuals

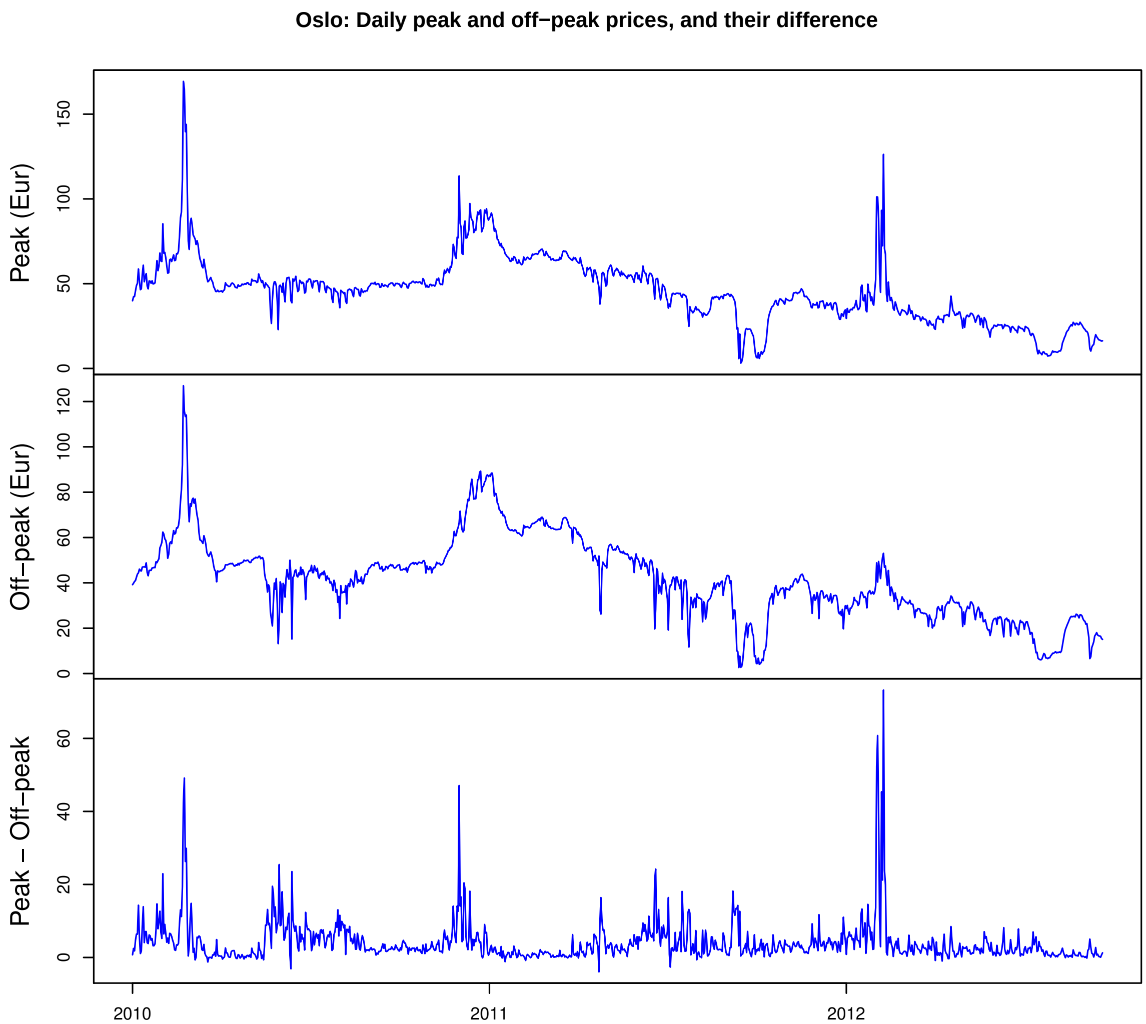

Figure 3: Daily peak and off-peak (and their difference) spot electricity prices (in Euros) for the Oslo area in Norway, 1 January 2010 - 19 September 2012 ($T = 993$ observations)

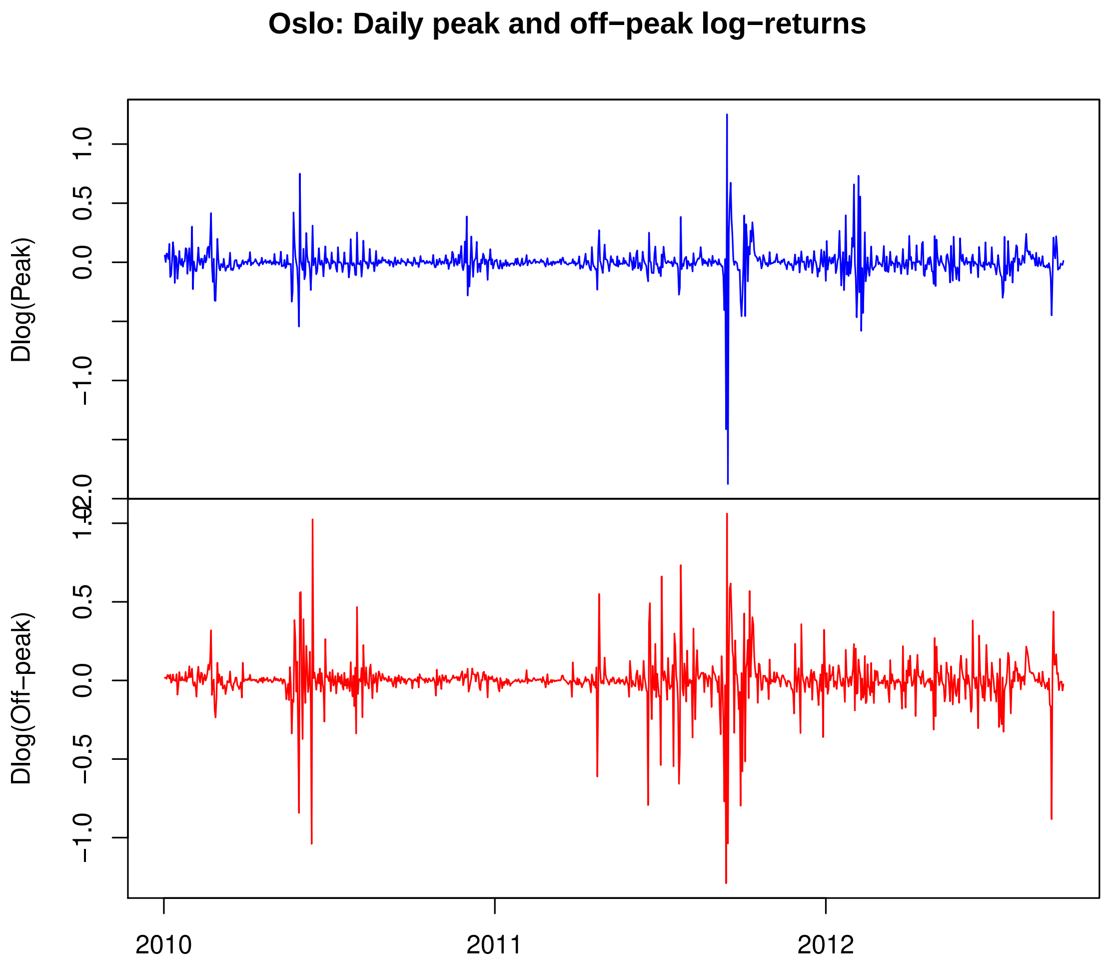

Figure 4: Daily peak and off-peak spot electricity price log-returns (in Euros) for the Oslo area in Norway, 1 January 2010 - 19 September 2012

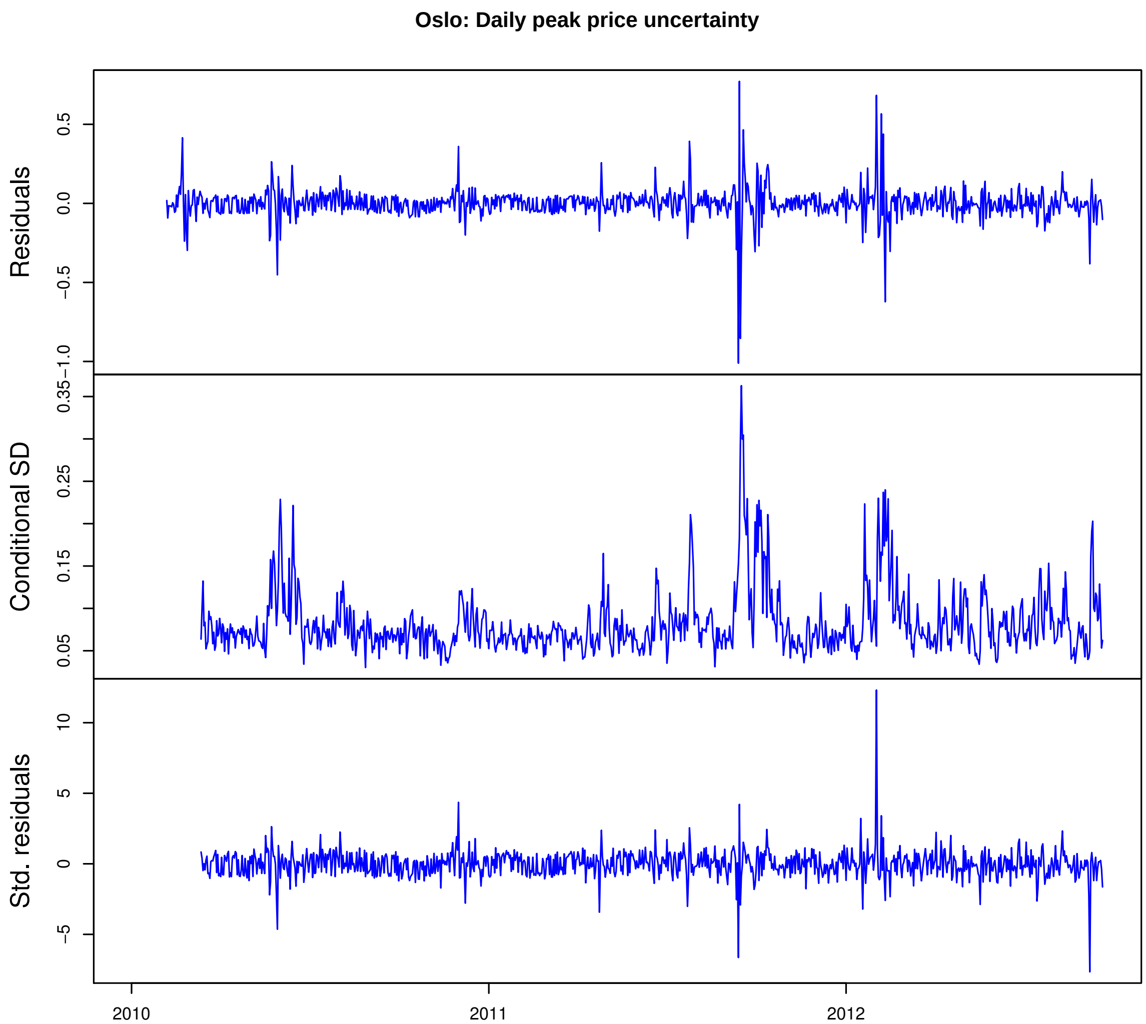

Figure 5: The residuals of the simplified conditional mean model of peak log-returns, and the fitted conditional standard deviations and standardised residuals of the simplified log-ARCH-X model (29)

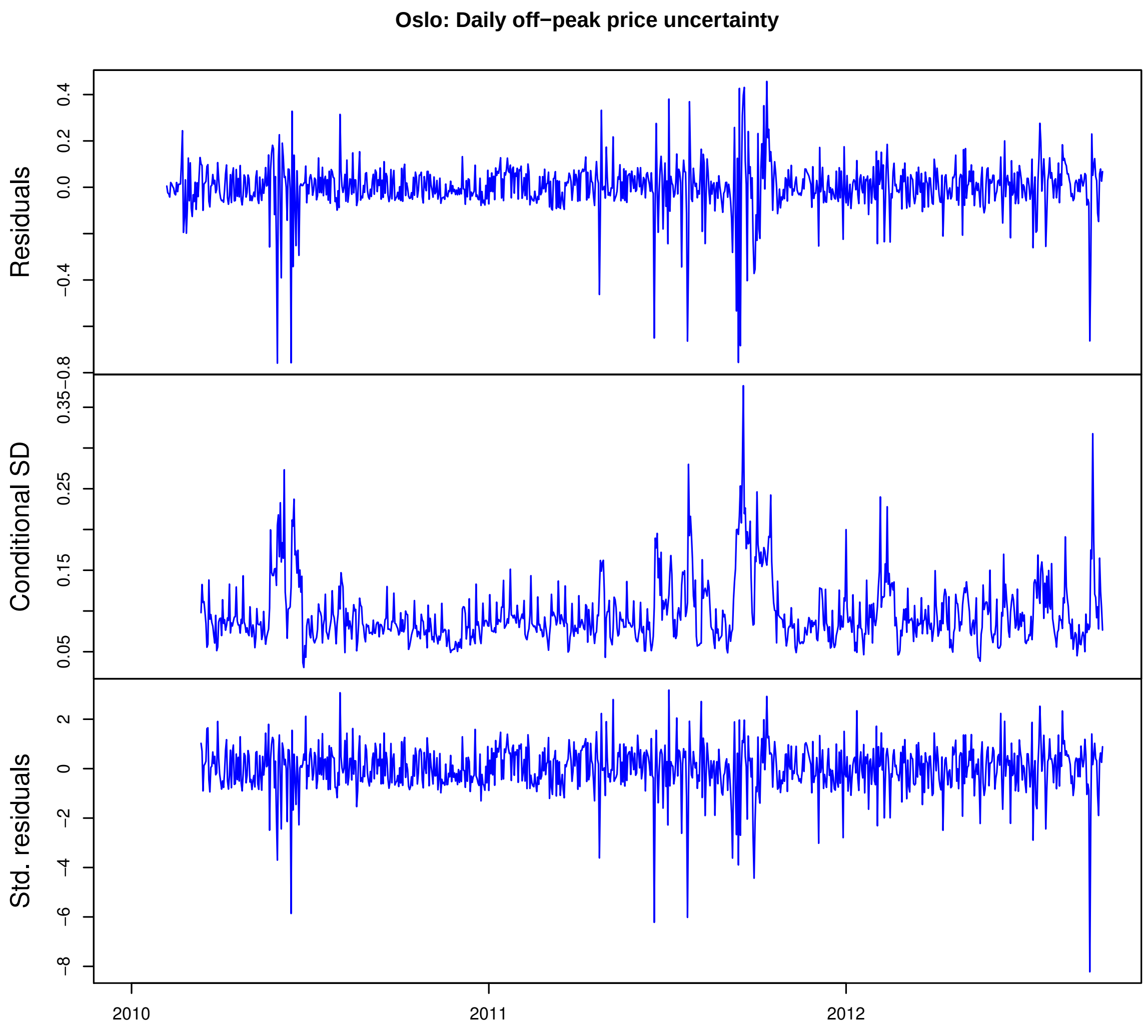

Figure 6: The residuals of the simplified conditional mean model of off-peak log-returns, and the fitted conditional standard deviations and standardised residuals of the simplified log-ARCH-X model (30)

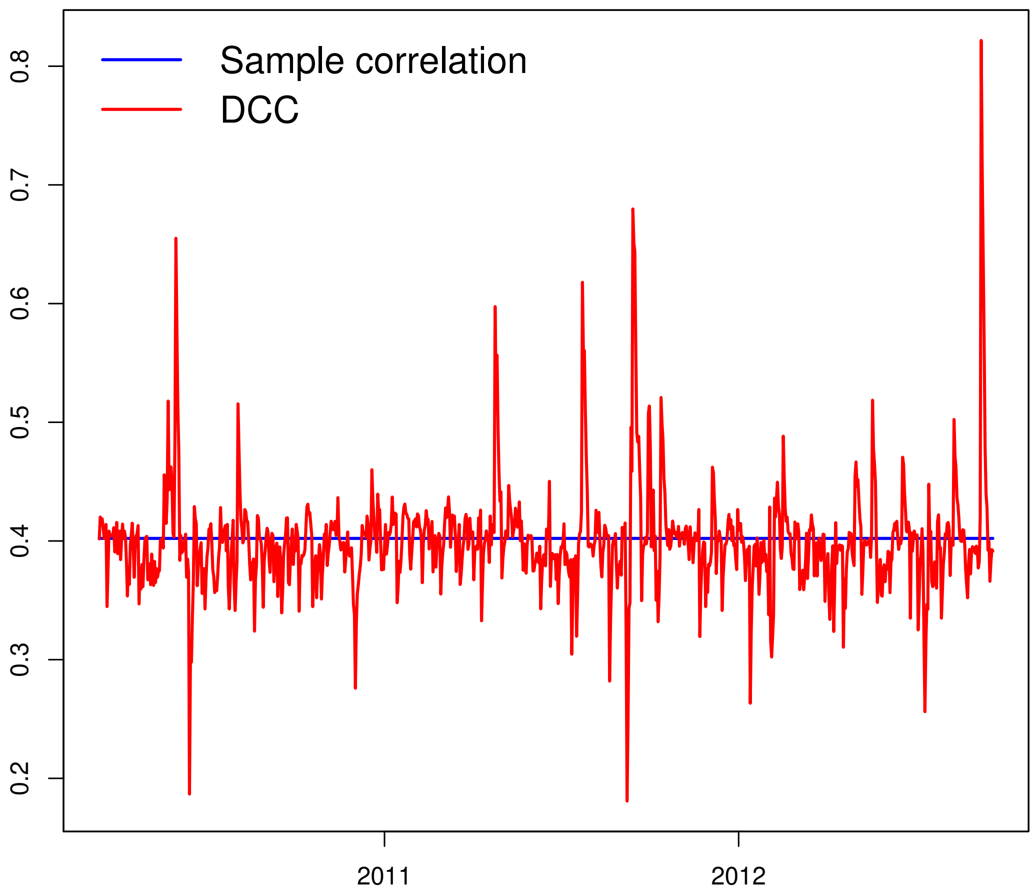

Figure 7: Constant (i.e. the sample correlation) and time-varying DCC correlations (Engle (2002)) of the standardised residuals $\hat{z}_{1,t}$ and $\hat{z}_{2,t}$ of (29)-(30)