Unbiased Estimation of Log-GARCH Models in the Presence of Zero Returns

Sucarrat, Genaro and Escribano, Alvaro

BI Norwegian Business School, Universidad Carlos III de Madrid

9 September 2013

Unbiased Estimation of Log-GARCH Models in the Presence of Zero Returns [1]

Genaro Sucarrat[2] and Alvaro Escribano[3]

1 October 2014

Abstract

A critique that has been directed towards the log-GARCH model is that its log-volatility specification does not exist in the presence of zero returns. A common "remedy" is to replace the zeros with a small (in the absolute sense) non-zero value. However, this renders estimation asymptotically biased. Here, we propose a solution to the case where the true return is equal to zero with probability zero. In this case zero returns may be observed because of non-trading, measurement error (e.g. due to rounding), missing values and other data issues. The solution we propose treats the zeros as missing values and handles these by combining estimation via the ARMA representation with an Expectation-Maximisation (EM) type algorithm. An extensive number of simulations confirm the conjectured asymptotic properties of the bias-correcting algorithm, and several empirical applications illustrate that it can make a substantial difference in practice.

JEL Classification: C22, C58

Keywords: ARCH, exponential GARCH, log-GARCH, ARMA, Expectation- Maximisation (EM)

[1] We are grateful to participants at the ISNPS 2014 (Cadiz) and NSVCM 2014 (Paderborn) conferences for useful comments, suggestions and questions.

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

[3] Department of Economics, Universidad Carlos III de Madrid (Spain), email: alvaroe@eco.uc3m.es.

1 Introduction

Models in the Autoregressive Conditional Heteroscedasticity (ARCH) class due to Engle (1982) have been extensively used to model the time-varying volatility of financial return (see Francq and Zakoïan (2010) for a survey of ARCH models). In particular, the first-order Generalised ARCH model of Bollerslev (1986), *i.e.* the GARCH(1,1), has established itself as an almost unquestionable benchmark. Pantula (1986), Geweke (1986) and Milhøj (1987) independently proposed specifications within the log-ARCH class of models as an alternative to non-exponential ARCH models. Their main motivation was to ensure the positivity of fitted volatilities – this is not guaranteed in non-exponential ARCH models (in particular when additional exogenous or predetermined conditioning information is added), and to allow for richer dynamics (*e.g.* negative ARCH parameters for cyclical or contrarian dynamics). However, Pantula (1986, p. 73) also stressed that it enables tests for integrated log-variance via Dickey and Fuller (1979) tests for unit roots. Engle and Bollerslev (1986) argued against log-ARCH models because of the possibility of applying the log-operator on zero-values.[4] This occurs whenever the return or mean-corrected return equals zero. Subsequently Nelson (1991) proposed an alternative exponential ARCH specification, the EGARCH model, where the problem is sidestepped by replacing the problematic term with an expression that does not involve the log operator. This solution, however, comes at a considerable cost: Restrictive assumptions and complicated conditions are needed to ensure that the Quasi Maximum Likelihood Estimator (QMLE) provides Consistent and Asymptotically Normal (CAN) estimates (Wintenberger (2013)), and unconditional moments (*e.g.* the unconditional variance of returns) will generally not exist for t-distributed densities (see condition (A1.6) and the subsequent discussion in Nelson (1991, p. 365)).

Zero returns occur in two different types of situations. In the first the zero-probability of actual return is zero, but zeros are nevertheless observed due to, say, non-trading, discreteness approximation error (*e.g.* rounding error), missing values and other data issues. For example, missing quotes or transaction prices are typically replaced by the previous observation, which in many cases results in an observed zero return even though the actual one is non-zero. Similarly, financial prices are usually quoted with a few digits only (typically two), so financial returns are thus often measured as zero even though the true return is non-zero. This leads to the observation that an asset often exhibit more zeros when low-priced (in nominal terms), since a tick then corresponds to a higher return than when highly-priced. Accordingly, one may argue that zeros should be treated as missing values instead of zeros. Finally, impulse dummies are sometimes used to mean-correct returns in the conditional mean. This leads to mean-corrected returns equal to

[4] Another critique that has been directed towards the log-GARCH (*e.g.* Teräsvirta (2009)) is that the first unconditional autocorrelations of the squared returns, 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, 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.

zero. When the impulse dummies are intended to neutralise the effect of large outliers or "jumps" – this is often the motivation in macroeconomics and finance, then one may argue that the zeros should be treated as missing observations of actual (mean-corrected) returns. The second type of situation in which zero returns occur is when the zero-probability of actual return is truly non-zero. In the conclusions we outline an extension in which the results of this paper can be used to develop a framework for this situation. For now, however, our focus is exclusively on the first type of situation in which the zero probability is zero.

Estimators that do not rely on a specific distribution (*e.g.* QMLEs) are greatly appreciated by practitioners, since then one needs not change the conditional density from application to application, or alternatively to use a sufficiently general and extra-parametrised density that makes estimation and inference more challenging. Two types of such estimators have been proposed for the log-GARCH model, one "Standard" and one based on the ARMA representation. Francq et al. (2013) prove CAN under mild assumptions for a QMLE of the first type. In their estimator the density of the conditional (mean-corrected) return is used for estimation, hence their estimator being Standard. For the second type, Sucarrat et al. (2014) show that CAN estimation of the ARMA representation of the log-GARCH model provides CAN estimates of the log-GARCH parameters, as long as the intercept bias in the log-volatility specification induced by the ARMA representation is appropriately adjusted for. In particular, both the Gaussian QMLE and the Least Squares Estimator (LSE), the two most common ARMA estimators, are applicable. Subsequently, another ARMA-based estimator was proposed by Francq and Sucarrat (2013). Their QMLE uses the centred exponential chi-squared distribution as instrumental density, which is more efficient when the conditional error is normal or close to normal. Both the Standard and the ARMA-based estimators are valid under mild assumptions, both types rely on the assumption that the probability of a zero (mean-corrected) return is zero and both types produce asymptotically biased estimates in the presence of zeros.

This paper makes three contributions. First, in Sections 2.2 and 2.3, we develop a framework for observed zero returns and propose an algorithm for unbiased estimation in their presence. The algorithm treats zeros as missing observations and replaces them with estimates of their conditional expectation. The algorithm is computationally simple and straightforwardly implemented with ARMA-based estimators, but it cannot be combined with the Standard QMLE. Second, we undertake an extensive set of Monte Carlo simulations (Section 3) to shed light on the effect of zeros, and to study the properties of our algorithm. If the algorithm we propose is not used when zeros are observed, then the simulations show that the downwards bias of volatility increases with the number of zeros, that the reaction to shocks is underestimated and that the empirical standard errors are larger. The extent of these features depend on the parameter values, on the exact value used to replace zeros, on whether the conditional density is fat-tailed or not and on the type of estimator. By contrast, if the algorithm we propose is used, then the simulations show that the estimates are unbiased, and that the sample-size adjusted empirical standard error correspond well to their asymptotic counterparts. An additional study suggests that we can usefully include estimated zero-probabilities as conditioning variables in a log-GARCH-X model. Third, several empirical applications (Section 4) illustrate that the parameter estimates and the fitted conditional standard deviations can differ substantially in practice if zeros are not appropriately handled, and a case study of the Apple stock sheds further light on the relationship between return volatility, zeros and volume.

The rest of the paper is organised as follows. The next section, Section 2, provides an overview of the log-GARCH model and how estimation via the ARMA-representation is implemented. The section also contains the underlying framework that we rely upon and the details of our algorithm. Section 3 contains the Monte Carlo studies. Section 4 contains the empirical section, whereas Section 5 concludes. Tables and Figures are located at the end.

2 Model, framework and algorithm

2.1 The log-GARCH model

If ϵ_t denotes financial return (possibly mean-corrected), then the log-GARCH(p, q) model is given by

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0,1), \quad Prob(z_t = 0) = 0, \quad \sigma_t > 0, \tag{1}$$

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \tag{2}$$

where p is the ARCH order and q is the GARCH order. In the context of log-GARCH models, the socalled inlier issue (see Breidt and Carriquiry (1996) for a discussion in a Stochastic Volatility (SV) context) amounts to whether $E(\ln z_t^2)$ exists. For the Student's t density and for the Generalised Error Distribution (GED), the two most common distributions in finance, $E(\ln z_t^2)$ generally exists. Francq et al. (2013) provide general conditions for the existence of log-moments.

It is well-known that (2) admits an ARMA representation, see *e.g.* Pantula (1986), Psaradakis and Tzavalis (1999) and Francq and Zakoïan (2006). Specifically, if $|E(\ln z_t^2)| < \infty$, then adding $\ln z_t^2$ to each side of (2), and then adding and subtracting $E(\ln z_t^2) \cdot (1 - \sum_{j=1}^{q} \beta_j)$ to the right-hand side, yields (by re-arranging the terms)

$$\ln \epsilon_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i \ln \epsilon_{t-i}^2 + \sum_{j=1}^{q} \theta_j u_{t-j} + u_t, \tag{3}$$

where

$$\begin{aligned}\phi_0 &= \alpha_0 + (1 - \sum_{j=1}^{q} \beta_j) \cdot E(\ln z_t^2), &(4)\\ \phi_p &= \alpha_i + \beta_i, \qquad 1 \leq i \leq p, &(5)\\ \theta_j &= -\beta_j, \qquad 0 \leq j \leq q, &(6)\\ u_t &= \ln z_t^2 - E(\ln z_t^2), \quad u_t \sim IID(0, \sigma_u^2). &(7)\end{aligned}$$

Moreover, if $E[(\ln z_t^2)^2] < \infty$, then $\sigma_u^2 < \infty$. Denoting $p^* = \max\{p, q\}$, then both (2) and (3) admit a strictly stationary solution if the AR polynomial $\mathcal{A}(L) = 1 - \phi_1 L - \cdots - \phi_{p^*} L^{p^*}$ satisfies $\mathcal{A}(z) \neq 0$ when $|z| \leq 1$. In other words, consistent and asymptotically normal estimates of all the ARMA parameters – and hence all the log-GARCH parameters except the log-volatility intercept α_0 – are thus readily obtained via usual ARMA estimation methods (*e.g.* the Gaussian QMLE or the LSE) subject to the usual ARMA assumptions (*i.e.* stationarity, invertibility, etc.), see *e.g.* Brockwell and Davis (2006 [1991]). For a consistent estimate of α_0, however, a consistent estimate of $E(\ln z_t^2)$ is needed. Sucarrat et al. (2014) show that

$$-\ln[T^{-1} \sum_{t=1}^{T} \exp(\hat{u}_t - \overline{\hat{u}}_T)] \tag{8}$$

provides a consistent and asymptotically normal estimate of $E(\ln z_t^2)$,[5] where $\hat{u}_t$ is the ARMA residuals and $\overline{\hat{u}}_T = T^{-1} \sum_{t=1}^{T} \hat{u}_t$. As a consequence, all the log-GARCH(p, q) parameters can be estimated consistently via the ARMA representation for a range of ARMA estimators, including the Gaussian QMLE and the LSE. Additional terms, *e.g.* asymmetry/leverage terms, or exogenous or predetermined conditioning information (*i.e.* "X"), can also be added without affecting the relationship between the log-GARCH and ARMA parameters, nor the structure of the bias-correction procedure. Francq and Sucarrat (2013) propose another ARMA-based QMLE that uses the centred exponential chi-squared distribution instead of the Gaussian as instrumental density. The motivation for this estimator is that it is asymptotically more efficient than the Gaussian ARMA-QMLE (and the LSE) when the conditional error z_t is normal or close to normal. In the (empirical) presence of zeros, however, both are biased if zeros are replaced with non-zero values.

The Standard QMLE of Francq et al. (2013) undertakes estimation in terms of the conditional density of ϵ_t. Also, they use a slightly different version of the (symmetric)

[5] For consistency, the key assumption of the estimator is that $T^{-1} \sum_{t=1}^{T} \exp(\hat{u}_t - \overline{\hat{u}}_T) - T^{-1} \sum_{t=1}^{T} \exp(u_t - \overline{u}_T) = o_P(1)$. For asymptotic normality, the estimator also relies on the stronger condition $\sqrt{T} \sum_{t=1}^{T} \exp(\hat{u}_t - \overline{\hat{u}}_T) - \sqrt{T} \sum_{t=1}^{T} \exp(u_t - \overline{u}_T) = o_P(1)$, together with the additional moment assumptions $E(z_t^4) < \infty$ and $E[(\ln z_t^2)^2] < \infty$. If this holds, then the asymptotic variance of the estimator is $Var(z_t^2 - \ln z_t^2)$, see Sucarrat et al. (2014).

log-volatility specification. In their setup (2) is replaced by

$$\ln \sigma_t^2 \quad = \quad \alpha_0 + \sum_{i=1}^{p} \alpha_i I_{\{z_{t-i} \neq 0\}} \ln \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2, \tag{9}$$

where $I_{\{z_{t-i} \neq 0\}}$ is an indicator function equal to 0 if $z_{t-i} = 0$, and 1 otherwise. Of course, theoretically (2) and (9) are equal at t with probability 1. In empirical practice, however, (9) avoids the problem of possibly applying the natural logarithm operator on zero values. Nevertheless, since the Standard QMLE also relies on the assumption $Prob(z_t = 0) = 0$, the empirical presence of zeros also leads to biased estimates.

2.2 Observed zeros – a framework

If ϵ_t denotes the actual or true return, then the observed return $\tilde{\epsilon}_t$ is given by

$$\tilde{\epsilon}_t = \epsilon_t I_t, \qquad I_t \in \{0, 1\}, \qquad \pi_{0t} = Prob_{t-1}(I_t = 0) \geq 0, \tag{10}$$

where π_{0t} denotes the (possibly) time-varying zero-probability conditional on the past.[6] Of course, this means $Prob_{t-1}(I_t = 1) = \pi_{1t} = 1 - \pi_{0t}$. To fix ideas we will specify the zero probability as a dynamic logit model. However, our framework is by no means restricted to this class of models. In the simplest case, therefore, when I_t is IID, we have that $h_t = \rho_0$, where $h_t = \ln(\pi_{1t}/\pi_{0t})$. For convenience we will sometimes refer to π_{1t} (and transformations thereof, *e.g.* $h_t = \ln(\pi_{1t}/\pi_{0t})$) as the zero-probability, since π_{0t} can straightforwardly be obtained via π_{1t} (and transformations thereof, *e.g.* $\pi_{0t} = 1 - \pi_{1t}$ where $\pi_{1t} = 1/(1 + \exp(-h_t))$.

The process that determines true return ϵ_t we refer to as the Data Generating Process (DGP), whereas the process that determines I_t we refer to as the Zero Generating Process (ZGP). The algorithm we propose requires that these two processes are sufficiently unrelated. Let $\mathcal{I}_{t-1}$ denote the set of past information, and let $\mathcal{I}_{t-1}^z = \{z_{t-1}, z_{t-2}, \ldots\}$ and $\mathcal{I}_{t-1}^I = \{I_{t-1}, I_{t-2}, \ldots\}$ with both $\mathcal{I}_{t-1}^z \subset \mathcal{I}_{t-1}$ and $\mathcal{I}_{t-1}^I \subset \mathcal{I}_{t-1}$. Since the algorithm we propose replaces the missing values with the conditional expectations of the ARMA representation whenever $I_t = 0$, we need that

$$E_{t-1}(\ln \epsilon_t^2) = E(\ln \epsilon_t^2 | I_t, \mathcal{I}_{t-1}), \tag{11}$$

where $E_{t-1}(\ln \epsilon_t^2)$ is the expectation of $\ln \epsilon_t^2$ conditional on the past associated with the ARMA representation. Schafer and Graham (2002) (see also Rubin (1976)) distinguish between three cases, and we may provide a similar distinction adapted to the current setting:

[6] A straightforward extension is to distinguish between the reason for the zeros. For example, observed return could be specified as, say, $\tilde{\epsilon}_t = \epsilon_t I_t^0 I_t^m$, where I_t^m is and indicator equal to zero when zero is truly due to missing observations and one otherwise, and where I_t^0 is equal to zero when return is zero but nonmissing and one otherwise.

1. Missing Completely at Random (MCAR): z_i and I_j are independent for all pairs i, j
2. Missing at Random (MAR): z_t and I_t are contemporaneously independent conditional on $\mathcal{I}_{t-1}$
3. Missing Not at Random (MNAR): z_t and I_t are contemporaneously dependent

The algorithm we propose below will be valid for the MCAR and – by a straightforward extension – MAR cases, but not necessarily for the MNAR case.

In the MCAR case the ZGP is entirely independent of the DGP. Accordingly, condition (11) will hold if $\mathcal{I}_{t-1} = \mathcal{I}^z_{t-1} \cup \mathcal{I}^I_{t-1}$, and if the ARMA representation (3) is stationary and invertible, since then $E_{t-1}(\ln \epsilon_t^2) = E(\ln \epsilon_t^2 | \mathcal{I}^z_{t-1}) = \psi_0 + \sum_{i=1}^{\infty} \psi_i \ln z_{t-i}^2 + E(\ln z_t^2)$. A straightforward example of MCAR is when ϵ_t is governed by (1)-(2) and when I_t is IID. Another example is when I_t is independent over time, but with a trend in the zero probability, *e.g.* $h_t = \rho_0 + \rho_1 T$. This model is of special interest, since it provides a simple description of a steady decrease (or increase) in the zero probability over time without an effect on return variability. We will return to variations of this model in Sections 3.3 and 4.2.

In the MAR case the log-volatility $\ln \sigma_t^2$ can depend on past values of I_t. If this is the case, then condition (11) will not hold for the ARMA representation (3). However, the log-GARCH model – and hence the ARMA representation – can readily be extended to allow past values of I_t to have an effect on $\ln \sigma_t^2$ by means of a log-GARCH-X specification. The log-GARCH-X model is given by

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 + g(\lambda, x_{t-1}), \tag{12}$$

where g is a linear or nonlinear function of the exogenous or predetermined variables x_{t-1}, and a parameter vector λ. If $Prob(z_t = 0) = 0$ and $|E(\ln z_t^2)| < \infty$, then (12) admits the ARMA-X representation

$$\ln \epsilon_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i \ln \epsilon_{t-i}^2 + \sum_{j=1}^{q} \theta_j u_{t-j} + g(\lambda, x_{t-1}) + u_t, \tag{13}$$

where the ARMA coefficients are related to the log-GARCH coefficients in the same ways as before, *i.e.* by (4)-(6), and where u_t is the same as earlier, *i.e.* $u_t = \ln z_t^2 - E(\ln z_t^2)$. Accordingly, with suitable assumptions on the x_t, the log-GARCH-X model can be estimated via the ARMA-X representation if the assumption in footnote 5 holds, see Sucarrat et al. (2014). Examples of X-variables of interest include leverage, volatility proxies, volume and past values of I_t (or transformations thereof, *e.g.* past values of h_t). The log-GARCH-X specification is therefore particularly interesting in the current context, since it provides a framework in which the effects of past observed zeros, volume, etc. on volatility can be studied jointly. Let the set of past information now be given by $\mathcal{I}_{t-1} = \mathcal{I}^z_{t-1} \cup \mathcal{I}^X_{t-1}$, where $\mathcal{I}^z_{t-1} = \{z_{t-1}, z_{t-2}, \dots\}$, $\mathcal{I}^X_{t-1} = \{x_{t-1}, x_{t-2}, \dots\}$ and $\mathcal{I}^I_{t-1} \subset \mathcal{I}^X_{t-1}$.

In other words, the conditioning X-vector contains the past information associated with the zero-process. If $E_{t-1}(\ln \epsilon_t^2) = E(\ln \epsilon_t^2|\mathcal{I}_{t-1}) = \phi_0 + \sum_{i=1}^{p} \phi_i \ln \epsilon_{t-i}^2 + \sum_{j=1}^{q} \theta_j u_{t-j} + g(\lambda, x_{t-1})$, then condition (11) holds.

In the MNAR case condition (11) may not hold, since z_t and I_t are contemporaneously dependent. This implies that also ϵ_t and I_t are contemporaneously dependent. If observed zeros are due to discrete prices or rounding errors, then this seems to suggest that missingness depends on the true return being close to zero, *i.e.* that ϵ_t and I_t are contemporaneously dependent. However, this is not clear, see Campbell et al. (1997, Section 3.3) for a discussion of discrete pricing. Define the true log-return in percent as $\epsilon_t = (\ln P_t - \ln P_{t-1}) \cdot 100$, so that the true price at t is given by $P_t = P_{t-1} \cdot e^{\epsilon_t/100}$. A zero return is thus observed whenever the rounded price $\widetilde{P}_t$ is equal to the rounded price $\widetilde{P}_{t-1}$. Clearly the occurrence of zeros depends on a range of factors, including the degree of discreteness (*e.g.* the tick-size), the nominal level of the price (*i.e.* the lower price, the more likely a zero will be observed),[7] the level of conditional volatility (*i.e.* the higher, the less likely a zero will be observed), the dynamics of volatility (*e.g.* the sensitivity to shocks) and the value of $|z_t|$. Accordingly, zeros due to rounding are not necessarily due to small values of $|z_t|$. However, if they are, then condition (11) may not hold, since this then implies a contemporaneous dependence between z_t and I_t.

2.3 Algorithm

The actual return ϵ_t is correctly observed whenever $I_t = 1$ in (10). Whenever $I_t = 0$, then the actual return ϵ_t is incorrectly observed or "missing". A common approach to missing observations in an ARMA context are state-space models, see *e.g.* Jones (1980), Shumway and Stoffer (1982), Kohn and Ansley (1986), and Gomez and Maravall (1994). In each of these, however, the density of the missing values – or the densities in question – is assumed known. In other words, they are not amenable to a "QML" type interpretation. Another common approach to missing observations is the Expectation-Maximisation (EM) algorithm popularised by Dempster et al. (1977). There, missing values are handled in what they characterised as two separate steps: The Expectation or E-step and the Maximisation or M-step. (Shumway and Stoffer (1982) combines the state-space approach with the EM-algorithm.) The algorithm we propose is in the spirit of the EM-algorithm, since it replaces the missing values with the conditional expectation. However, the separation between the E and M steps is not as sharp as usual. The EM-like algorithm that we propose holds several advantages over state-space approaches. In particular, we do not need that z_t (nor $\ln z_t^2$) is distributed according to a certain density, and our algorithm is simpler both conceptually and computationally. To the best of our knowledge, however, there is no complete and rigorous proof under mild assumptions that ensures that our algorithm in combination with, say, the Gaussian QMLE (or the LSE) provides consistent estimates of an ARMA model (although Brockwell and Davis (2002,

[7] For example, suppose the true return at t is +0.25%. If prices are rounded to two decimals, then $P_{t-1} = 1$ will result in an observed zero return at t, whereas for $P_{t-1} = 10$ it will not.

pp. 289-290) suggest how such a proof could proceed in terms of the exact likelihood). Hence, we outline our algorithm in some detail before we discuss the key assumption that it relies on. For concreteness we outline the algorithm for the Gaussian QMLE, but the algorithm can straightforwardly be adapted for use with the LSE as well. The next section, Section 3, contains simulations that support our conjecture that the asymptotic properties of the Gaussian QMLE are retained in the presence of zeros.

Let $\widehat{\phi}_0^{(k)}, \widehat{\phi}_1^{(k)}, \ldots, \widehat{\phi}_p^{(k)}$ and $\widehat{\theta}_1^{(k)}, \ldots, \widehat{\theta}_q^{(k)}$ denote the parameter estimates of the ARMA representation (3) after k iterations with some numerical method (*e.g.* Newton-Raphson). The initial values are given at $k = 0$. If there are no observed zeros, then at the kth. iteration the numerical method thus proceeds in the usual way:

1. Compute, recursively, for $t = 1, \ldots, T$:

$$\widehat{u}_t^{(k-1)} = \ln \epsilon_t^2 - \widehat{\phi}_0^{(k-1)} - \sum_{i=1}^{p} \widehat{\phi}_i^{(k-1)} \ln \epsilon_{t-i}^2 - \sum_{j=1}^{q} \widehat{\theta}_j^{(k-1)} \widehat{u}_{t-j}^{(k-1)}, \tag{14}$$

   where $\widehat{u}_t^{(k-1)}$ is an estimate of u_t.

2. Compute the log-likelihood $\sum_{t=1}^{T} \ln f_u(\widehat{u}_t^{(k-1)})$, where f_u is the Gaussian density, and other quantities (*e.g.* the gradient and/or Hessian) needed by the numerical method to generate $\widehat{\phi}_0^{(k)}, \widehat{\phi}_1^{(k)}, \ldots, \widehat{\phi}_p^{(k)}$ and $\widehat{\theta}_1^{(k)}, \ldots, \widehat{\theta}_q^{(k)}$.

The algorithm we propose modifies this procedure in several ways. Let τ denote the locations of the non-zero values of ϵ_t, and let T^* denote the number of non-zero values. The kth. iteration is now modified as follows:

1. Compute, recursively, for $t = 1, \ldots, T$:

   a) $$\overline{\ln \epsilon_t^2}^{(k-1)} = \begin{cases} \ln \epsilon_t^2 & \text{if } t \in \tau \\ \widehat{E}_{t-1}(\ln \epsilon_t^2)^{(k-1)} & \text{if } t \notin \tau \end{cases} \tag{15}$$

   where

$$\widehat{E}_{t-1}(\ln \epsilon_t^2)^{(k-1)} = \widehat{\phi}_0^{(k-1)} + \sum_{i=1}^{p} \widehat{\phi}_i^{(k-1)} \overline{\ln \epsilon^2}_{t-i}^{(k-1)} + \sum_{j=1}^{q} \widehat{\theta}_j^{(k-1)} \widehat{u}_{t-j}^{(k-1)} \tag{16}$$

   b) $$\widehat{u}_t^{(k-1)} = \overline{\ln \epsilon_t^2}^{(k-1)} - \widehat{\phi}_0^{(k-1)} - \sum_{i=1}^{p} \widehat{\phi}_i^{(k-1)} \overline{\ln \epsilon^2}_{t-i}^{(k-1)} - \sum_{j=1}^{q} \widehat{\theta}_j^{(k-1)} \widehat{u}_{t-j}^{(k-1)}. \tag{17}$$

2. Compute the log-likelihood $\sum_{t\in\tau} \ln f_u(\widehat{u}_t^{(k-1)})$, where f_u is the Gaussian density, and other quantities (*e.g.* the gradient and/or Hessian) needed by the numerical method to generate $\widehat{\phi}_0^{(k)}, \widehat{\phi}_1^{(k)}, \ldots, \widehat{\phi}_p^{(k)}$ and $\widehat{\theta}_1^{(k)}, \ldots, \widehat{\theta}_q^{(k)}$.

Step 1.a) means the value of $\ln \epsilon_t^2$ is replaced by an estimate of the conditional expectation $E_{t-1}(\ln \epsilon_t^2)$ at zero locations. This estimate does not rely on any specific assumption on the density of $\ln z_t^2$ (say, Gaussianity), nor on the density of z_t, apart from $Prob(z_t = 0) = 0$ and the existence of $E(\ln z_t^2)$. However, it is worth noting that, at the population level, the value $E_{t-1}(\ln \epsilon_t^2)$ is not only the value that minimises the forecast error variance, it is also the value that maximises the conditional expected log-likelihood $E_{t-1}(\ln f_u(u_t))$ at t if f_u is the Gaussian density. In other words, if the Gaussian distribution is used as the instrumental density in a QMLE, then there is a clear link to the EM-algorithm, where missing values are estimated by maximising the expected log-likelihood conditional on the observed data, see Dempster et al. (1977, page 6). In Step 1.b) the recursion value $\widehat{u}_t^{(k-1)}$ is, by construction, equal to 0 at the zero-locations. This has implications for Step 2, where the symbolism $t \in \tau$ means the log-likelihood only includes contributions from non-zero locations. An important practical implication of this is that likelihood comparisons with competing models should be in terms of the average log-likelihood with division by T^* rather than T. After estimation of the ARMA representation the ARMA residuals $\widehat{u}_t$ at non-zero locations are used to estimate $E(\ln z_t^2)$ with (8). Next, estimates of the log-GARCH parameters are obtained via the formulas in (4)-(6). The algorithm we have outlined is valid for the MCAR case. However, by straightforward modifications, *i.e.* replacing the log-GARCH and ARMA expressions with log-GARCH-X and ARMA-X expressions, respectively, the algorithm can also be applied to log-GARCH-X models when zeros occur according to the MAR scheme. In the Standard QMLE the algorithm is not applicable. The reason for this is that an estimate of $\ln \epsilon_t^2$ is needed as a replacement for the missing observations in the recursion of the log-volatility specification (9), and this is not provided by the estimator when it is interpreted as a QMLE. If the Standard QMLE is interpreted as an *exact* MLE, however, then a similar algorithm to the one above can be used. In that case z_t is standard normal and $E(\ln z_t^2) = -1.27$.

It is well known that both the Gaussian QMLE and the LSE produce consistent and asymptotically normal estimates of the ARMA parameters under mild assumptions when there are no missing values, see *e.g.* Hannan (1973), and Brockwell and Davis (2006 [1991]) (Francq et al. (2011), and the references therein, contain more recent and general results). In particular, it is not required that the initial values $\widehat{u}_0, \widehat{u}_{-1}, \ldots$ in the recursion (14) are equal to their true values $u_0, u_{-1}, \ldots$. This can be referred to as an "irrelevance of initial values" condition. If $\widetilde{u}_t$ denotes the ARMA error at the true parameter values but having started at $t = 0$ (*i.e.* not in the infinite past) with some arbitrary initial values, then the irrelevance of initial values condition means the difference between u_t and $\widetilde{u}_t$ becomes sufficiently small in some appropriate sense as $t \to \infty$. Heuristically, our algorithm can be viewed as repeatedly (instead of only once) creating an initial value issue, since the true value of $\ln \epsilon_t^2$ is replaced by its conditional expectation whenever $I_t = 0$. In other words, whenever $I_t = 0$, then the $\widetilde{u}_t$ is perturbed away from u_t. This suggests zeros cannot occur too often, since – heuristically – $\widetilde{u}_t$ may need sufficient time to converge back towards u_t before it is perturbed away again. Otherwise the cumulated difference may not be asymptotically irrelevant. How large the zero-probability can be before the cumulated discrepancy becomes relevant (in some appropriate sense), however, is not clear. The perturbation is minimal in the conditional variance sense, since the missing values are replaced by their conditional expectations. So a reasonable conjecture is that the probability can be sufficiently large to be of practical interest (in daily financial data the zero-proportion is usually between 0 and 0.05). This is certainly supported by the simulations in Section 3, and – in fact – cursory simulations (not reported but available on request) suggest the zero-probability can be at least as large as 0.5 for large datasets.

3 A Monte Carlo study

3.1 Effect of zeros

To shed light on the effect of observed zeros on parameter estimates we compare the Standard QMLE and the Gaussian ARMA-QMLE in a simulation experiment. In the experiment the DGP of return ϵ_t is given by the log-GARCH(1,1) specification

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0,1), \quad Prob(z_t = 0) = 0, \tag{18}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2, \tag{19}$$

for empirically relevant combinations of the parameters α_0, α_1 and β_1. These combinations are referred to as A, B and C. The zero probability is constant over time and equal to either 0, 0.05, 0.10 or 0.20. In other words, the DGP and the ZGP are entirely independent, *i.e.* the MCAR case (we relax this assumption in Section 3.3).

For the Gaussian ARMA-QMLE estimation is undertaken with the adjusted return

$$\tilde{\tilde{\epsilon}}_t = \begin{cases} \epsilon_t \text{ if } I_t = 1, \\ c \text{ if } I_t = 0, \quad c > 0. \end{cases} \tag{20}$$

The log-volatility specification is thus $\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \tilde{\tilde{\epsilon}}_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2$. Clearly the choice of c will influence the results. In particular, among the natural choices of c, *i.e.* values between the numerical minimum of the statistical software in question and, say, 0.1, the closer to zero, the larger the bias is likely to be in our experience. Moreover, the smaller the value, the more often numerical issues (*e.g.* non-convergence) are encountered. As an intermediate choice we therefore choose $c = 0.01$ for the simulations. For the Standard QMLE the zeros of observed return $\tilde{\epsilon}_t$ are simply not included in the recursion because of the indicator function in the log-volatility specification (9). It is worth noting, however, that this is equivalent of setting $c = 1$ in (20). In other words, any difference in simulation result is not only due to the estimator, but also due to the different value of c.

Table 1 contains the results, and Figures 1-3 provides a comparison with the algorithm (see Section 3.2). For the Standard QMLE the effect of zeros is straightforward: The higher the zero probability, the greater the bias, and the bias is almost invariably equal or higher when the conditional density is fat-tailed (*i.e.* $t(5)$). The log-volatility intercept α_0 is biased downwards, which means volatility will generally be biased downwards in the presence of zeros. The ARCH parameter α_1, which controls the impact of shocks on volatility, is also biased downwards. The presence of zeros thus means volatility will be under-responsive to shocks. This effect is exacerbated by the upward bias of β_1, since this parameter controls the effect of the long-term component of volatility. Finally, increasing the zero probability increases the standard errors.

For the Gaussian ARMA-QMLE the biases are generally bigger compared with the Standard QMLE, but not always as straightforward. This is most readily seen in the Figures. Higher zero probability means larger negative bias for both α_0 and α_1, although the bias is not always larger for α_0 when compared with those of the Standard QMLE. For β_1 the effect of zeros is more complex since the bias can change sign. Finally, also for the Gaussian ARMA-QMLE do the empirical standard increase when the zero probability increases, but the increase can be much bigger.

3.2 Properties of algorithm

To study the properties of the Gaussian ARMA-QMLE in combination with our algorithm we conduct two experiments. The first is similar to the one in the previous subsection in that the ZGP is IID, whereas in the second the zero probability decreases over time. Both experiments corresponds to the MCAR case, where the ZGP is entirely independent of the DGP.

The results of the first experiment are contained in Panel 1 of Table 2, and Figures 1-3 compare the finite sample bias with those of Section 3.1. It is clear that the algorithm corrects the bias for all three parameters. The finite sample biases increase slightly (and more so for $z_t \sim t(5)$) as the zero probability increases, but this is not surprising since more observations are lost by treating zeros as missing values when the zero probability increases. The empirical standard errors are virtually unaffected as the the zero probability increases, which is in stark contrast to the QMLEs without the algorithm. Moreover, the empirical standard error correspond well to their (adjusted) asymptotic counterparts. Finally, compared with the Standard QMLE the finite sample bias is substantially smaller for the location-parameter α_0, *i.e.* the most important parameter in determining the level of volatility.

In the second experiment we study the properties of our algorithm in combination with the Gaussian QMLE when the zero probability is steadily decreasing. This is in line with the empirical observation that the zero probability falls over time, *e.g.* due to increased liquidity and/or volume, reduced discreteness/smaller ticks, higher nominal prices and other changes in how markets operate. As earlier we specify the DGP as a log-GARCH(1,1) with the same parameter values as earlier, *i.e.* A, B and C. The ZGP is entirely independent of the DGP, with the zero probability steadily falling according to a dynamic logit model. Let "relative" time be given by t/T for $t = 0, 1, 2, \ldots, T$ such that $t/T \in [0, 1]$. The ZDP is then given by the dynamic logit-model

$$h_t = \rho_0 + \rho_1 \cdot (t/T), \tag{21}$$

where $h_t = \ln(\pi_{1t}/\pi_{0t})$, $\pi_{1t} = 1/(1 + \exp(-h_t))$ and $\rho_0 = 1.9$, $\rho_1 = 3.4$. The values of ρ_0 and ρ_1 are chosen on the basis of the empirical estimates in Section 4.2. This means the zero-probability is given by $\pi_{0t} = 1 - \pi_{1t} = 0.130$ at the beginning of the sample (*i.e.* $t = 0$), and by $\pi_{0t} = 0.005$ at the end (*i.e.* $t = T$).

The results are contained in Panel 2 of Table 2. On average, the zero proportion $\hat{\pi}_0$ produced by the ZGP given by (21) is about 0.04, and the properties of the Gaussian ARMA-QMLE w/algorithm is therefore virtually identical to the results in Panel 1 when $\pi_0 = 0.05$. The biases are corrected, and the empirical standard errors correspond well to their (adjusted) asymptotic counterparts.

3.3 Can we condition on estimated zero-probabilities?

An X-variable of special interest is the zero-probability itself (or transformations thereof). Of course, in empirical practice one is unlikely to have access to the true zero-probabilities. To shed light on whether one in practice may (usefully) try to model volatility in terms of past zero-probabilities, we device an experiment where the true return depends on (transformations of) past zero probabilities (*i.e.* we are in the MAR case). Specifically, the ZDP is given by (21), but the DGP of true return ϵ_t is given by the log-GARCH(1,1)-X specification

$$\ln \sigma_t^2 \quad = \quad \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2 + \lambda h_{t-1}, \tag{22}$$

where $\lambda = -0.1$, and where the combinations of α_0, α_1 and β_1 are the same as in the previous simulations (*i.e.* A, B and C). Table 3 contains the simulation results for known values of h_t (Panel 1) and for estimated values (Panel 2). The results are very similar, even for small (in financial contexts) samples, which suggests estimates of h_t as conditioning variables are capable of proxying their true values reasonably well.

4 Empirical applications

This section contains two empirical applications. In the first our objective is simply to illustrate how much volatility estimates can differ if zero-returns are not properly handled. In the second we illustrate how a log-GARCH-X model can be exploited to shed further light on the relationship between return volatility, observed zeros and volume.

4.1 The effect of zero-returns on volatility

We compare the difference in parameter estimates and fitted conditional standard deviations for six daily financial returns: The Apple and Ekornes stocks (more information on Ekornes shortly), the Standard and Poor's 500 stock market index (SP500), the EUR/USD exchange rate, the WTI oil price and the London gold price. This small selection of returns accounts for a variety of market characteristics. For example, whereas the EUR/USD is traded in a global market almost continuously 24-hours a day and seven days a week – possibly with thousands of trades per second, the London Gold price is only fixed twice a day, and presumably not on Bank holidays and in weekends. Our interest in the Ekornes stock price return is due to its relatively large proportion of zeros (about 19%), and the main reason for the zeros is non-trading (*i.e.* a zero volume). Ekornes is a leading Nordic furniture manucturer listed on the Oslo Stock Exchange. It can be described as a medium-sized company in international terms, since its current market value is approximately 300 million euros. The sources of the data are Yahoo Finance (http://finance.yahoo.com) for the Apple, Ekornes and SP500 series, the European Central Bank (http://www.ecb.int/) for the EUR/USD series, the US Energy Information Agency (http://www.eia.gov/) for the WTI crude oil price (in USD) per barrel series and Kitco (http://www.kitco.com/) for the London afternoon (*i.e.* PM) gold price series.

The sample dates and the descriptive statistics of the returns are contained in Table 4, whereas Figure 4 contains graphs of the returns. They confirm that the returns exhibit the usual properties of excess kurtosis compared with the normal, and ARCH as measured by the first and second order serial correlations in the squared return. The number of zeros varies from only 2 observations (about 0.1% of the sample) for SP500 to 667 observations (about 19% of the sample) for Ekornes. The reasons for each zero are likely to differ substantially both within and across markets. We do not try to identify these reasons, since our main objective is to illustrate how the estimates and fitted conditional standard deviations differ according to estimation method.

Table 5 contains the estimates of the log-GARCH(1,1) specification

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0,1), \quad Pr(z_t = 0) = 0, \tag{23}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2, \tag{24}$$

where ϵ_t is the log-return in percent (*i.e.*, the log-difference of the financial price multiplied by 100). Unsurprisingly the biggest numerical differences in the parameter estimates are produced by Ekornes (highest number of zeros, 19% of the sample), and the smallest are produced by SP500 (only two zeros, 0.1% of the sample). The estimates of the intercept α_0, which controls the unconditional variance, are always higher for the algorithm, apart from Ekornes. This is somewhat surprising, due to the high number of zeros for Ekornes. There clearly seems to be some interaction with the persistence parameter β_1, since it is unusually low, 0.784, when zeros are replaced with the minimum of non-zero returns. Similarly, the algorithm estimate of β_1 for Ekornes is the only one that is higher (and substantially so since it is 0.943). In the other five cases the 0-adj estimates are slightly higher. With respect to the estimate of the ARCH parameter α_1, which controls the short-term impact of shocks or large (in absolute value) returns at $t-1$, the estimates of the algorithm are substantially higher except for oil where they are only slightly higher.

Descriptive statistics and graphs of the fitted conditional standard deviations, their differences and their ratios, are contained in Table 6 and Figures 5-7. They clearly suggest that estimation method can matter a lot, both nominally and in relative terms. For example, for Apple the algorithm yields fitted conditional standard deviations that are at most 2.15 times higher, and the maximum nominal difference is 1.9. Such numerical differences can make a huge difference in risk analysis and asset pricing. The Apple graphs also reveals what seems to be an inverse tendency. In the beginning of the sample the algorithm produces higher fitted conditional standard deviations. However, this is reversed in the second part of the sample. A possible reason is that there are fewer zeros in the second part of the sample (see the graph of I_t in Figure 8). For most returns the average fitted conditional standard deviation is higher for the algorithm. This is most clearly seen in the graph of EUR/USD, where the fitted conditional standard deviations produced by the algorithm are clearly above almost everywhere. The only case where the average difference is not positive is oil. There, the average is approximately equal to zero. But the ratio graph clearly shows that, in relative terms, the algorithm occasionally produce values that are up to 66% higher. So all in all the comparison of fitted conditional standard deviations show that the algorithm generally produces higher values, and sometimes much higher.

4.2 Apple: Return volatility, zeros and volume

The proportion of observed zeros can change over time. Including past zero probabilities (and/or transformations thereof) and other variables to the X-vector in a log-GARCH-X model enables us therefore to study the impact on volatility in more detail. Here, we illustrate this by a simple case study of the Apple stock return.

Figure 8 contains graphs of the Apple stock price, return, I_t (*i.e.* $I_t = 0$ means return is zero on day t), volume (in USD) and log-volume. The graph of I_t clearly shows that the occurrence of zeros is less likely towards the end of the sample. Similarly, the volume graph reveals that volume is higher in the second half of the sample, and the price graph reveals that the nominal price is increasing over the sample. To shed light on whether the increase in volume or nominal price is indeed one of the reasons for the fall in zero-probability, we estimate the four dynamic logit models

$$h_t = \rho_0, \quad (25)$$
$$h_t = \rho_0 + \rho_1 \cdot (t/T), \quad (26)$$
$$h_t = \rho_0 + \rho_1 \ln V_t, \quad (27)$$
$$h_t = \rho_0 + \rho_1 \ln P_t, \quad (28)$$

where $h_t = \ln(\pi_{1t}/\pi_{0t})$ and $\pi_{1t} = 1/(1 + \exp(-h_t))$. In the first the zero-probability is constant, in the second the zero-probability is determined by a time-trend, in the third volume determines the zero-probability, whereas in the fourth the nominal price P_t determines the zero probability. To recall, the last model is motivated by the fact that higher (nominal) prices often results in fewer zeros than when an asset is low-priced for the same tick-size. Table 7 contains the estimation results. Unsurprisingly the latter three models produce a substantially higher log-likelihood than the first, and among the latter three the log-likelihood of the volume-model is slightly higher. This can be interpreted to suggest that increased volume rather than discreteness (*i.e.* fewer zero due to higher nominal prices) is the main reason for the downwards trend in the zero probability.

To shed light on whether zeros and volume have an effect on volatility, we estimate the four volatility models

$$\ln \sigma_t^2 = \alpha_0, \tag{29}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2, \tag{30}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2 + \lambda_1 \ln V_{t-1}, \tag{31}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2 + \lambda_1 \ln V_{t-1} + \lambda_2 I_{t-1} + \lambda_3 \widehat{h}_{t-1}, \tag{32}$$

where $\widehat{h}_t$ are the fitted values of (27). Table 8 contains the estimation results. Unsurprisingly the constant volatility model has a much smaller log-likelihood than the three other models, and fares worse according to the Schwarz (1978) Information Criterion (BIC). Maybe somewhat surprisingly, however, is that the lags of volume, I_t and $\widehat{h}_t$ do not improve the fit. (The reason the log-likelihoods are lower for the last two specifications even though they contain more terms, is that the comparison is in terms of $\sum_{t\in\tau} \ln f(\epsilon_t; \hat{\sigma}_t)$ rather than the log-likelihood of the ARMA-X representation, *i.e.* $\sum_{t\in\tau} \ln f(\hat{u}_t; \hat{\sigma}_u)$, where f is the normal density.) Also, the standard errors are high relative to the parameter estimates, so t-tests with nulls $\lambda_1 = 0$, $\lambda_2 = 0$ and $\lambda_3 = 0$ do not reject at usual significance levels. All in all, then, this simple analysis does not suggest past volume nor past zeros or zero probabilities have an effect on volatility.

5 Conclusions

We propose an estimation procedure for log-GARCH models that is unbiased in the presence of zero returns. The procedure combines estimation via the ARMA representation with an Expectation-Maximisation (EM) type algorithm, a procedure that is not straightforwardly implemented with the Standard QMLE of Francq et al. (2013). The estimation procedure we propose distinguishes between true and observed return, and relies on the assumption that the true return is equal to zero with zero probability. This is compatible with observed return being zero due to missing values, non-trading, certain types of rounding or discreteness approximation error, impulse dummies in the mean specification to neutralise jumps or "outliers" and other data issues. In our framework the zero probability can be time-varying and conditionally dependent on the past, and volatility can depend on past zeros and zero probabilities. However, our framework may not be valid if the occurrence of a zero depends contemporaneously on the value of the de-volatilised return. Our Monte Carlo simulations show that volatility is generally underestimated when zeros are present if our proposed estimation procedure is not used, and that our estimation procedure corrects the bias with the empirical standard errors corresponding well to their asymptotic counterparts. The empirical illustrations confirm that volatility is generally underestimated when zeros are present, and that the impact of shocks on volatility is underestimated in the presence of zeros. In practice this means that the fitted conditional standard deviations are generally underestimated – sometimes substantially.

Finally, a case study of Apple return, whose zero-probability has been steadily declining, suggest that volatility does not depend on past zeros, nor on past zero-probabilities or the level of volume.

The results in this paper may be extended in at least four ways. First, if zeros are the result of measurement error of a true return whose zero probability is zero, then this also leads to biased estimates for other ARCH models, *e.g.* the GARCH of Bollerslev (1986), the EGARCH of Nelson (1991) and the Beta-t-EGARCH model of Harvey (2013). Estimation procedures in combination with an algorithm similar to the one we have proposed here can be used in all three classes. Second, unbiased QMLEs for univariate and multivariate log-MEM-X models can be devised, where MEM is short for Multiplicative Error Models, see Brownlees et al. (2012). MEM-models are particularly suited for non-negative financial data like volume, durations and trades, and because of its structure a QMLE for log-GARCH-X models is also a QMLE for log-MEM-X models. There are often zeros and/or missing values in volume, duration and trade data, see *e.g.* Hautsch et al. (2013). When these zeros can be viewed as a result of missing values, or if the distinction between a true and an observed value is appropriate, then the methods in this paper can be used to adjust for the bias induced by the zeros. Third, improved estimation efficiency may be achieved by extending the log-GARCH-X model into a joint bivariate system with $\ln \sigma_t^2$ and h_t as left-hand side variables, and/or by devising a (non-QML) density of the occurrence of zeros along the lines of Hautsch et al. (2013). Finally, the ideas in this paper can be used to develop an estimator in which the zero probability is truly non-zero, by appropriately scaling the return.

References

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

Breidt, F. J. and A. L. Carriquiry (1996). Improved Quasi-Maximum Likelihood Estimation for Stochastic Volatility Models. In J. Lee, W. Johnson, and A. Zellner (Eds.), *Modeling and Prediction: Honoring Seymour Geisser*. Berlin: Springer.

Brockwell, P. J. and R. A. Davis (2002). *Introduction to Time Series and Forecasting*. Springer. 2nd. Edition.

Brockwell, P. J. and R. A. Davis (2006). *Time Series: Theory and Methods*. New York: Springer. 2nd. Edition, first published in 1991.

Brownlees, C., F. Cipollini, and G. Gallo (2012). Multiplicative Error Models. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*, pp. 223–247. New Jersey: Wiley.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press: Princeton.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B 39*, 1–38.

Dickey, D. and W. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 426–431.

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

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

Francq, C., R. Roy, and A. Saidi (2011). Aasymptotic Properties of Weighted Least Squares Estimation in Weak PARMA Models. *Journal of Time Series Analysis 32*, 699–723.

Francq, C. and G. Sucarrat (2013). An Exponential Chi-Squared QMLE for Log-GARCH Models Via the ARMA Representation. http://mpra.ub.uni-muenchen.de/51783/.

Francq, C., O. Wintenberger, and J.-M. Zakoïan (2013). GARCH Models Without Positivity Constraints: Exponential or Log-GARCH? Forthcoming in *Journal of Econometrics*, http//dx.doi.org/10.1016/j.jeconom.2013.05.004.

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 (2010). *GARCH Models*. New York: Marcel Dekker.

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

Gomez, V. and A. Maravall (1994). Estimation, Prediction, and Interpolation for Nonstationary Series with the Kalman Filter. *Journal of the American Statistical Association 89*, 611–624.

Hannan, E. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability 10*, 130–145.

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

Hautsch, N., P. Malec, and M. Schienle (2013). Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics*. Forthcoming. http://jfec.oxfordjournals.org/content/early/2013/03/29/jjfinec.nbt002.

Jones, R. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics 22*, 389–395.

Kohn, R. and C. Ansley (1986). Estimation, Prediction, and Interpolation for ARIMA Models With Missing Data. *Journal of the American Statistical Association 81*, 751–761.

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

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.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubin, D. (1976). Inference and Missing Data. *Biometrica 63*, 581–592.

Schafer, J. and J. Graham (2002). Missing Data: Our View of the State of the Art. *Psychological Method 7*, 147–177.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics 6*, 461–464.

Shumway, R. and D. Stoffer (1982). An approach to time series smoothing and forecasting using the EM-algorithm. *Journal of Time Series Analysis 3*, 254–264.

Sucarrat, G. (2014). *lgarch: Simulation and estimation of log-GARCH models*. http://cran.r-project.org/web/packages/lgarch/.

Sucarrat, G., S. Grønneberg, and Á. Escribano (2014). Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown. MPRA Paper No. 57238. Online at http://mpra.ub.uni-muenchen.de/57238/.

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.

Wintenberger, O. (2013). Continuous Invertibility and Stable QML Estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics 40*, 846–867.

Table 1: The Standard QMLE and the Gaussian ARMA-QMLE when zero returns are replaced by $c > 0$ (see Section 3.1)

	DGP (ID: α_0,α_1,β_1)	π_0	Standard QMLE: $m(\hat{\alpha}_0)$	$se(\hat{\alpha}_0)$	$m(\hat{\alpha}_1)$	$se(\hat{\alpha}_1)$	$m(\hat{\beta}_1)$	$se(\hat{\beta}_1)$	Gaussian ARMA-QMLE: $m(\hat{\alpha}_0)$	$se(\hat{\alpha}_0)$	$m(\hat{\alpha}_1)$	$se(\hat{\alpha}_1)$	$m(\hat{\beta}_1)$	$se(\hat{\beta}_1)$
$N(0,1)$:	A: 0, 0.10, 0.80	0.00	-0.001	0.009	0.100	0.004	0.799	0.011	-0.003	0.015	0.100	0.007	0.798	0.017
		0.05	-0.020	0.010	0.097	0.005	0.807	0.012	-0.019	0.018	0.071	0.007	0.826	0.021
		0.10	-0.040	0.012	0.094	0.005	0.813	0.013	-0.033	0.023	0.055	0.007	0.843	0.025
		0.20	-0.076	0.015	0.089	0.005	0.826	0.013	-0.195	0.579	0.034	0.010	0.767	0.403
	B: 0, 0.05, 0.90	0.00	-0.001	0.006	0.050	0.003	0.899	0.008	-0.001	0.010	0.050	0.005	0.899	0.012
		0.05	-0.011	0.007	0.048	0.004	0.903	0.009	-0.012	0.014	0.036	0.005	0.912	0.016
		0.10	-0.020	0.008	0.047	0.004	0.907	0.009	-0.035	0.182	0.028	0.005	0.908	0.140
		0.20	-0.039	0.011	0.044	0.004	0.913	0.010	-0.225	0.662	0.018	0.007	0.797	0.450
	C: 0, 0.03, 0.95	0.00	-0.001	0.005	0.030	0.002	0.949	0.005	-0.002	0.008	0.030	0.004	0.949	0.007
		0.05	-0.007	0.006	0.028	0.002	0.952	0.005	-0.007	0.010	0.023	0.004	0.955	0.009
		0.10	-0.013	0.007	0.027	0.003	0.955	0.005	-0.014	0.117	0.019	0.004	0.958	0.061
		0.20	-0.024	0.009	0.025	0.003	0.959	0.006	-0.138	0.600	0.013	0.005	0.906	0.286
$t(5)$:	A: 0, 0.10, 0.80	0.00	-0.002	0.019	0.100	0.008	0.798	0.019	-0.002	0.018	0.100	0.007	0.798	0.017
		0.05	-0.026	0.021	0.095	0.007	0.807	0.018	-0.021	0.022	0.076	0.007	0.822	0.020
		0.10	-0.048	0.024	0.091	0.008	0.815	0.021	-0.038	0.026	0.061	0.007	0.836	0.023
		0.20	-0.091	0.029	0.084	0.008	0.828	0.023	-0.080	0.156	0.042	0.007	0.847	0.093
	B: 0, 0.05, 0.90	0.00	-0.001	0.015	0.050	0.006	0.899	0.015	-0.002	0.013	0.050	0.005	0.898	0.013
		0.05	-0.015	0.017	0.048	0.006	0.902	0.017	-0.013	0.016	0.038	0.005	0.910	0.014
		0.10	-0.026	0.019	0.046	0.006	0.906	0.017	-0.024	0.020	0.031	0.005	0.916	0.018
		0.20	-0.048	0.023	0.042	0.007	0.913	0.019	-0.083	0.317	0.021	0.006	0.900	0.182
	C: 0, 0.03, 0.95	0.00	-0.001	0.012	0.030	0.004	0.949	0.009	-0.002	0.010	0.030	0.004	0.949	0.007
		0.05	-0.010	0.012	0.028	0.004	0.953	0.009	-0.007	0.012	0.024	0.004	0.954	0.008
		0.10	-0.018	0.015	0.026	0.004	0.955	0.009	-0.016	0.140	0.021	0.004	0.956	0.059
		0.20	-0.031	0.020	0.023	0.004	0.959	0.011	-0.175	0.795	0.015	0.005	0.901	0.314

DGP, $\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2$ with $z_t \sim IID(0,1)$ and $T = 10000$. $N(0,1)$, z_t is standard normal. $t(5)$, z_t is standardised t with 5 degrees of freedom. ID, parameter identifier (*i.e.* A, B or C). π_0, the zero probability associated with $I_t \sim IID$, $I_t \in \{0,1\}$. Standard QMLE, zeros replaced with $c = 1$. Gaussian ARMA-QMLE, zeros replaced with $c = 0.01$. $m(\cdot)$, sample average of estimates. $se(\cdot)$, sample standard deviation of estimates (division by S, not by $S-1$, where $S = 1000$ is the number of Monte Carlo simulations). All computations in R version 3.1.0, see R Core Team (2014). Simulation and part of the estimation (the Gaussian ARMA-QMLE) with `lgarch` package version 0.5 (Sucarrat (2014)).

Table 2: The Gaussian ARMA-QMLE w/algorithm in the presence of zeros (see Section 3.2)

Panel 1 ($I_t \sim IID$, $\pi_0 = Prob(I_t = 0)$):

	DGP (ID: α_0,α_1,β_1)	π_0	$m(\hat{\alpha}_0)$	$se(\hat{\alpha}_0)$	$m(\hat{\alpha}_1)$	$se(\hat{\alpha}_1)$	$ase(\hat{\alpha}_1)$	$m(\hat{\beta}_1)$	$se(\hat{\beta}_1)$	$ase(\hat{\beta}_1)$	$m(\hat{E}(\ln z_t^2))$	$se(\hat{E}(\ln z_t^2))$	$ase(\hat{E}(\ln z_t^2))$
$N(0,1)$:	A: 0, 0.10, 0.80	0.00	-0.003	0.015	0.100	0.007	0.007	0.798	0.017	0.017	-1.270	0.018	0.017
		0.05	0.000	0.014	0.101	0.007	0.007	0.797	0.017	0.017	-1.274	0.017	0.018
		0.10	0.003	0.015	0.102	0.008	0.007	0.796	0.018	0.018	-1.277	0.018	0.018
		0.20	0.007	0.015	0.105	0.008	0.008	0.794	0.020	0.019	-1.286	0.019	0.019
	B: 0, 0.05, 0.90	0.00	-0.001	0.010	0.050	0.005	0.005	0.899	0.012	0.013	-1.270	0.017	0.017
		0.05	-0.001	0.011	0.050	0.006	0.005	0.898	0.014	0.013	-1.272	0.017	0.018
		0.10	0.000	0.011	0.051	0.006	0.005	0.898	0.015	0.013	-1.272	0.019	0.018
		0.20	0.003	0.011	0.052	0.006	0.006	0.897	0.015	0.014	-1.277	0.020	0.019
	C: 0, 0.03, 0.95	0.00	-0.002	0.008	0.030	0.004	0.004	0.949	0.007	0.007	-1.270	0.017	0.017
		0.05	-0.002	0.009	0.031	0.004	0.004	0.948	0.008	0.007	-1.272	0.018	0.018
		0.10	-0.001	0.009	0.031	0.004	0.004	0.948	0.008	0.008	-1.273	0.018	0.018
		0.20	0.000	0.009	0.032	0.005	0.004	0.947	0.009	0.008	-1.275	0.019	0.019
$t(5)$:	A: 0, 0.10, 0.80	0.00	-0.002	0.018	0.100	0.007	0.007	0.798	0.017	0.017	-1.568	0.027	0.028
		0.05	0.000	0.019	0.101	0.007	0.007	0.797	0.018	0.017	-1.572	0.030	0.028
		0.10	0.002	0.019	0.103	0.008	0.007	0.795	0.018	0.018	-1.575	0.028	0.029
		0.20	0.009	0.019	0.105	0.009	0.008	0.795	0.020	0.019	-1.582	0.031	0.031
	B: 0, 0.05, 0.90	0.00	-0.002	0.013	0.050	0.005	0.005	0.898	0.013	0.013	-1.568	0.027	0.028
		0.05	-0.001	0.013	0.051	0.005	0.005	0.897	0.014	0.013	-1.568	0.033	0.028
		0.10	0.000	0.013	0.051	0.006	0.005	0.897	0.014	0.013	-1.571	0.029	0.029
		0.20	0.001	0.014	0.052	0.006	0.006	0.896	0.015	0.014	-1.575	0.029	0.031
	C: 0, 0.03, 0.95	0.00	-0.002	0.010	0.030	0.004	0.004	0.949	0.007	0.007	-1.567	0.027	0.028
		0.05	-0.002	0.010	0.030	0.004	0.004	0.949	0.008	0.007	-1.569	0.029	0.028
		0.10	-0.002	0.010	0.031	0.004	0.004	0.948	0.008	0.008	-1.569	0.029	0.029
		0.20	0.000	0.011	0.032	0.004	0.004	0.947	0.008	0.008	-1.574	0.031	0.031

Panel 2 ($h_t = 1.9 + 3.4 \cdot (t/T)$):

	DGP (ID: α_0,α_1,β_1)	$\hat{\pi}_0$	$m(\hat{\alpha}_0)$	$se(\hat{\alpha}_0)$	$m(\hat{\alpha}_1)$	$se(\hat{\alpha}_1)$	$ase(\hat{\alpha}_1)$	$m(\hat{\beta}_1)$	$se(\hat{\beta}_1)$	$ase(\hat{\beta}_1)$	$m(\hat{E}(\ln z_t^2))$	$se(\hat{E}(\ln z_t^2))$	$ase(\hat{E}(\ln z_t^2))$
$N(0,1)$:	A: 0, 0.10, 0.80	0.040	0.000	0.014	0.101	0.007	0.007	0.797	0.017	0.017	-1.274	0.018	0.017
	B: 0, 0.05, 0.90	0.039	-0.001	0.010	0.050	0.005	0.005	0.898	0.013	0.013	-1.272	0.017	0.017
	C: 0, 0.03, 0.95	0.039	-0.002	0.009	0.030	0.004	0.004	0.949	0.008	0.007	-1.270	0.017	0.017
$t(5)$:	A: 0, 0.10, 0.80	0.040	0.000	0.018	0.101	0.007	0.007	0.798	0.017	0.017	-1.569	0.031	0.028
	B: 0, 0.05, 0.90	0.040	-0.002	0.013	0.051	0.005	0.005	0.898	0.013	0.013	-1.569	0.029	0.028
	C: 0, 0.03, 0.95	0.040	-0.002	0.011	0.031	0.004	0.004	0.948	0.008	0.007	-1.567	0.028	0.028

DGP, $\ln\sigma_t^2 = \alpha_0 + \alpha_1 \ln\epsilon_{t-1}^2 + \beta_1 \ln\sigma_{t-1}^2$ with $z_t \sim IID(0,1)$ and $T = 10000$. $N(0,1)$, z_t is standard normal with $E(\ln z_t^2) \approx -1.27$ and $Var(z_t^2 - \ln z_t^2) \approx 2.93$. $t(5)$, z_t is standardised t with 5 degrees of freedom, $E(\ln z_t^2) \approx -1.57$ and $Var(z_t^2 - \ln z_t^2) \approx 7.63$. ID, parameter identifier (*i.e.* A, B or C). In Panel 2, $\hat{\pi}_0$ is the average proportion of zeros. $m(\cdot)$, sample average of the Monte Carlo estimates. $se(\cdot)$, sample standard deviation of the Monte Carlo estimates (division by S, not by $S-1$, where $S = 1000$ is the number of Monte Carlo simulations). $ase(x)$, asymptotic standard error of x, computed as $\sqrt{av(x)}/\sqrt{T^*}$, where $av(x)$ is the asymptotic variance of x and $T^* = (1-\pi_0)T$ in Panel 1 ($T^* = (1-\hat{\pi}_0)T$ in Panel 2). The expressions of $av(\hat{\alpha}_1)$ and $av(\hat{\beta}_1)$ are based on the ARMA(1,1) formulas in Brockwell and Davis (2006, pp. 259-260), whereas $av(\hat{E}(\ln z_t^2)) = Var(z_t^2 - \ln z_t^2)$ is derived in Sucarrat et al. (2014). Simulation and estimation with the `lgarch` package version 0.5 (Sucarrat (2014)) under R version 3.1.0, see R Core Team (2014).

Table 3: The Gaussian ARMA-QMLE w/algorithm when h_{t-1} is a conditioning variable (see Section 3.3)

Panel 1 (h_t known):

DGP (ID: α_0,α_1,β_1)	T	$m(\hat{\alpha}_0)$	$se(\hat{\alpha}_0)$	$m(\hat{\alpha}_1)$	$se(\hat{\alpha}_1)$	$m(\hat{\beta}_1)$	$se(\hat{\beta}_1)$	$m(\hat{\lambda})$	$se(\hat{\lambda})$
A: 0, 0.10, 0.80	1000	-0.025	0.096	0.101	0.023	0.780	0.060	-0.119	0.049
	2000	-0.011	0.057	0.100	0.016	0.789	0.043	-0.111	0.035
	5000	-0.004	0.032	0.100	0.011	0.795	0.026	-0.105	0.020
	10000	-0.001	0.022	0.101	0.007	0.798	0.018	-0.102	0.014
B: 0, 0.05, 0.90	1000	-0.021	0.055	0.046	0.018	0.892	0.041	-0.122	0.057
	2000	-0.010	0.039	0.050	0.011	0.893	0.028	-0.115	0.041
	5000	-0.003	0.020	0.050	0.008	0.898	0.018	-0.105	0.025
	10000	-0.002	0.014	0.050	0.005	0.898	0.012	-0.103	0.018
C: 0, 0.03, 0.95	1000	-0.013	0.025	0.022	0.014	0.957	0.015	-0.103	0.016
	2000	-0.006	0.017	0.027	0.008	0.952	0.009	-0.104	0.015
	5000	-0.002	0.010	0.029	0.005	0.950	0.006	-0.104	0.014
	10000	-0.001	0.008	0.030	0.003	0.950	0.005	-0.103	0.013

Panel 2 (h_t unknown, but estimated in a prior step):

DGP (ID: α_0,α_1,β_1)	T	$m(\hat{\alpha}_0)$	$se(\hat{\alpha}_0)$	$m(\hat{\alpha}_1)$	$se(\hat{\alpha}_1)$	$m(\hat{\beta}_1)$	$se(\hat{\beta}_1)$	$m(\hat{\lambda})$	$se(\hat{\lambda})$
A: 0, 0.10, 0.80	1000	-0.007	0.139	0.101	0.023	0.779	0.062	-0.124	0.066
	2000	-0.004	0.078	0.101	0.016	0.790	0.041	-0.111	0.039
	5000	-0.002	0.045	0.100	0.010	0.796	0.025	-0.104	0.022
	10000	-0.001	0.031	0.101	0.007	0.797	0.017	-0.102	0.015
B: 0, 0.05, 0.90	1000	-0.020	0.118	0.047	0.018	0.889	0.045	-0.127	0.075
	2000	-0.003	0.077	0.049	0.012	0.893	0.033	-0.117	0.052
	5000	-0.002	0.038	0.050	0.008	0.896	0.018	-0.108	0.028
	10000	0.000	0.024	0.051	0.005	0.898	0.013	-0.104	0.020
C: 0, 0.03, 0.95	1000	-0.003	0.084	0.021	0.014	0.958	0.015	-0.106	0.030
	2000	-0.004	0.050	0.027	0.008	0.952	0.009	-0.104	0.022
	5000	0.000	0.031	0.029	0.005	0.951	0.006	-0.103	0.017
	10000	0.000	0.021	0.030	0.004	0.950	0.005	-0.103	0.015

DGP, $\ln\sigma_t^2 = \alpha_0 + \alpha_1 \ln\epsilon_{t-1}^2 + \beta_1 \ln\sigma_{t-1}^2 + \lambda h_{t-1}$ with $z_t \overset{IID}{\sim} N(0,1)$, $\lambda = -0.10$, $h_t = 1.9 + 3.4 \cdot (t/T)$ and $\pi_{1t} = 1/(1+\exp(-h_t))$. ID, experiment identifier (*i.e.* A, B or C). In Panel 1, the parameters are estimated under the assumption that h_t is known. In Panel 2, h_t is first estimated, then the estimate $\hat{h}_t$ is used as a proxy for h_t. $m(\cdot)$, sample average of the Monte Carlo estimates. $se(\cdot)$, sample standard deviation of the Monte Carlo estimates (division by S, not by $S-1$, where $S = 1000$ is the number of Monte Carlo simulations). Simulation and estimation with the `lgarch` package version 0.4 (Sucarrat (2014)) under *R* version 3.1.0, see R Core Team (2014).

Table 4: Descriptive statistics of financial returns

	s^2	s^4	$ARCH_1$ $[p{-}val]$	$ARCH_2$ $[p{-}val]$	T	0s	$\hat{\pi}_0$
Apple (10 Sep. 1984 – 23 Aug. 2013):	9.25	55.03	7.12 [0.01]	14.77 [0.00]	7303	294	0.040
Ekornes (4 Jan. 2000 – 26 Aug. 2013):	5.70	10.32	54.01 [0.00]	228.20 [0.00]	3546	667	0.188
EUR/USD (5 Jan. 1999 – 23 Aug. 2013):	0.43	5.44	150.63 [0.00]	183.44 [0.00]	3751	32	0.009
SP500 (5 Jan. 1999 – 23 Aug. 2013):	1.73	10.30	143.10 [0.00]	669.52 [0.00]	3684	2	0.001
Oil (5 Apr. 1983 – 19 Aug. 2013):	5.72	18.80	160.60 [0.00]	252.96 [0.00]	7621	73	0.010
Gold (4 Jan. 2006 – 23 Aug. 2013):	1.85	7.29	10.94 [0.00]	20.35 [0.00]	1929	20	0.010

s^2, sample variance. s^4, sample kurtosis. $ARCH_1$ and $ARCH_2$, Ljung and Box (1979) test statistics of first-order and second-order serial correlation in the squared return. T, number of returns. 0s, number of zero returns in the sample. $\hat{\pi}_0$, proportion of zero returns in the sample. All computations in R version 3.1.0, see R Core Team (2014).

Table 5: Log-GARCH(1,1) estimates for six daily financial returns

	Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$se(\hat{\alpha}_1)$	$\hat{\beta}_1$	$se(\hat{\beta}_1)$
Apple:	0-adj	0.034	0.014	0.003	0.983	0.005
	Algo	0.048	0.029	0.004	0.967	0.005
Ekornes:	0-adj	0.374	0.087	0.016	0.784	0.059
	Algo	0.074	0.047	0.011	0.943	0.016
EUR/USD:	0-adj	0.022	0.019	0.004	0.976	0.005
	Algo	0.023	0.020	0.004	0.974	0.006
SP500:	0-adj	0.070	0.045	0.006	0.946	0.008
	Algo	0.071	0.046	0.006	0.946	0.008
Oil:	0-adj	0.074	0.043	0.004	0.951	0.005
	Algo	0.075	0.046	0.004	0.948	0.005
Gold:	0-adj	0.055	0.029	0.006	0.959	0.009
	Algo	0.058	0.033	0.006	0.956	0.009

Gaussian ARMA-QML estimation of the log-GARCH(1,1) specification $\ln\sigma_t^2 = \alpha_0 + \alpha_1 \ln\epsilon_{t-1}^2 + \beta\ln\sigma_{t-1}^2$. 0-adj, zero returns replaced by the minimum of the absolute non-zero value before estimation. Algo, estimation with algorithm (*i.e.* zeros not replaced, but treated as missing values). $se(\cdot)$, standard error of estimate (based on the numerically estimated Hessian of the ARMA representation). Estimation with the `lgarch` package version 0.5 (Sucarrat (2014)) under R version 3.1.0, see R Core Team (2014).

Table 6: Descriptive statistics of fitted conditional standard deviations

		Mean	s^2	Max	Min
Apple:	0-adj	2.971	0.753	6.348	1.510
	Algo	2.985	0.796	6.175	1.097
	Diff	0.015	0.356	1.902	-1.095
	Ratio	1.023	0.056	2.145	0.625
Ekornes:	0-adj	2.360	0.220	5.275	1.297
	Algo	2.572	0.390	5.465	1.349
	Diff	0.212	0.268	3.235	-0.933
	Ratio	1.104	0.057	2.608	0.656
EUR/USD:	0-adj	0.643	0.018	1.171	0.360
	Algo	0.645	0.017	1.164	0.365
	Diff	0.002	0.000	0.078	-0.028
	Ratio	1.005	0.001	1.167	0.954
SP500:	0-adj	1.191	0.327	4.730	0.437
	Algo	1.193	0.334	4.766	0.432
	Diff	0.003	0.001	0.392	-0.009
	Ratio	1.001	0.001	1.364	0.987
Oil:	0-adj	2.197	0.968	7.532	0.410
	Algo	2.189	0.903	7.422	0.431
	Diff	-0.008	0.018	1.042	-0.217
	Ratio	1.006	0.006	1.660	0.937
Gold:	0-adj	1.317	0.096	2.579	0.723
	Algo	1.326	0.120	2.878	0.674
	Diff	0.009	0.006	0.482	-0.069
	Ratio	1.002	0.003	1.255	0.919

Mean, sample average. s^2, sample variance. Max, maximum value. Min, minimum value. Diff, the difference between fitted conditional standard deviations: $\hat{\sigma}_{t,\text{Algo}} - \hat{\sigma}_{t,\text{0-adj}}$. Ratio, the ratio between fitted conditional standard deviations: $\hat{\sigma}_{t,\text{Algo}}/\hat{\sigma}_{t,\text{0-adj}}$. All computations in *R* version 3.1.0, see R Core Team (2014).

Table 7: Models of the zero probability of daily Apple returns

Model	$\hat{\rho}_0$	$se(\hat{\rho}_0)$	$\hat{\rho}_1$	$se(\hat{\rho}_1)$	k	$LogL$	BIC
$h_t = \rho_0$	3.171	0.060	–	–	1	-1232.465	0.339
$h_t = \rho_0 + \rho_1 \bar{t}$	1.870	0.094	3.437	0.263	2	-1123.887	0.310
$h_t = \rho_0 + \rho_1 \ln V_t$	-16.081	1.374	1.212	0.088	2	-1122.887	0.310
$h_t = \rho_0 + \rho_1 \ln P_t$	1.230	0.156	0.808	0.072	2	-1134.604	0.313

Dynamic logit models where $\pi_{1t} = 1/(1 + \exp(-h_t))$, $h_t = \ln(\pi_{1t}/\pi_{0t})$, V_t is the traded volume in USD and P_t is the stock price. $se(\cdot)$, standard error of estimate (computed as the square root of the diagonal of $-\hat{H}^{-1}$, where $\hat{H}$ is the numerically estimated Hessian). Estimation by maximum likelihood. k, number of parameters. $LogL$, the attained log-likelihood. BIC, the Schwarz (1978) information criterion computed in terms of the average log-likelihood $LogL/T$. All computations in R version 3.1.0, see R Core Team (2014).

Table 8: Models of Apple stock return volatility

Model	$\hat{\alpha}_0$	$\hat{\alpha}_1$ (s.e.)	$\hat{\beta}_1$ (s.e.)	$\hat{\lambda}_1$ (s.e.)	$\hat{\lambda}_2$ (s.e.)	$\hat{\lambda}_3$ (s.e.)	$LogL^*$	BIC
log-GARCH(0,0)	2.266	–	–	–	–	–	-17885.217	5.105
log-GARCH(1,1)	0.048	0.029 (0.004)	0.967 (0.005)	–	–	–	-17263.842	4.930
log-GARCH(1,1)-V	0.082	0.030 (0.004)	0.966 (0.005)	-0.002 (0.001)	–	–	-17270.645	4.933
log-GARCH(1,1)-V,I,h	5.700	0.030 (0.004)	0.966 (0.006)	-0.427 (0.764)	0.023 (0.029)	0.350 (0.631)	-17266.914	4.935

Gaussian ARMA-QML estimates w/algorithm of specifications contained in $\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \epsilon_{t-1}^2 + \beta \ln \sigma_{t-1}^2 + \lambda_1 \ln V_{t-1} + \lambda_2 I_{t-1} + \lambda_3 \widehat{h}_{t-1}$. $se(\cdot)$, standard error of estimate (based on the numerically estimated Hessian of the ARMA representation). $LogL^*$, log-likelihood computed as $\sum_{t \in \tau} \ln f_\epsilon(\epsilon_t; \hat{\sigma}_t)$, where f_ϵ is the normal density, $\hat{\sigma}_t$ is the fitted conditional standard deviation and τ is the set of non-zero locations. BIC, Schwarz (1978) information criterion computed in terms of the average log-likelihood $LogL^*/T^*$, where T^* is the number of non-zero returns. Estimation with the `lgarch` package version 0.4 (Sucarrat (2014)) under R version 3.1.0, see R Core Team (2014).

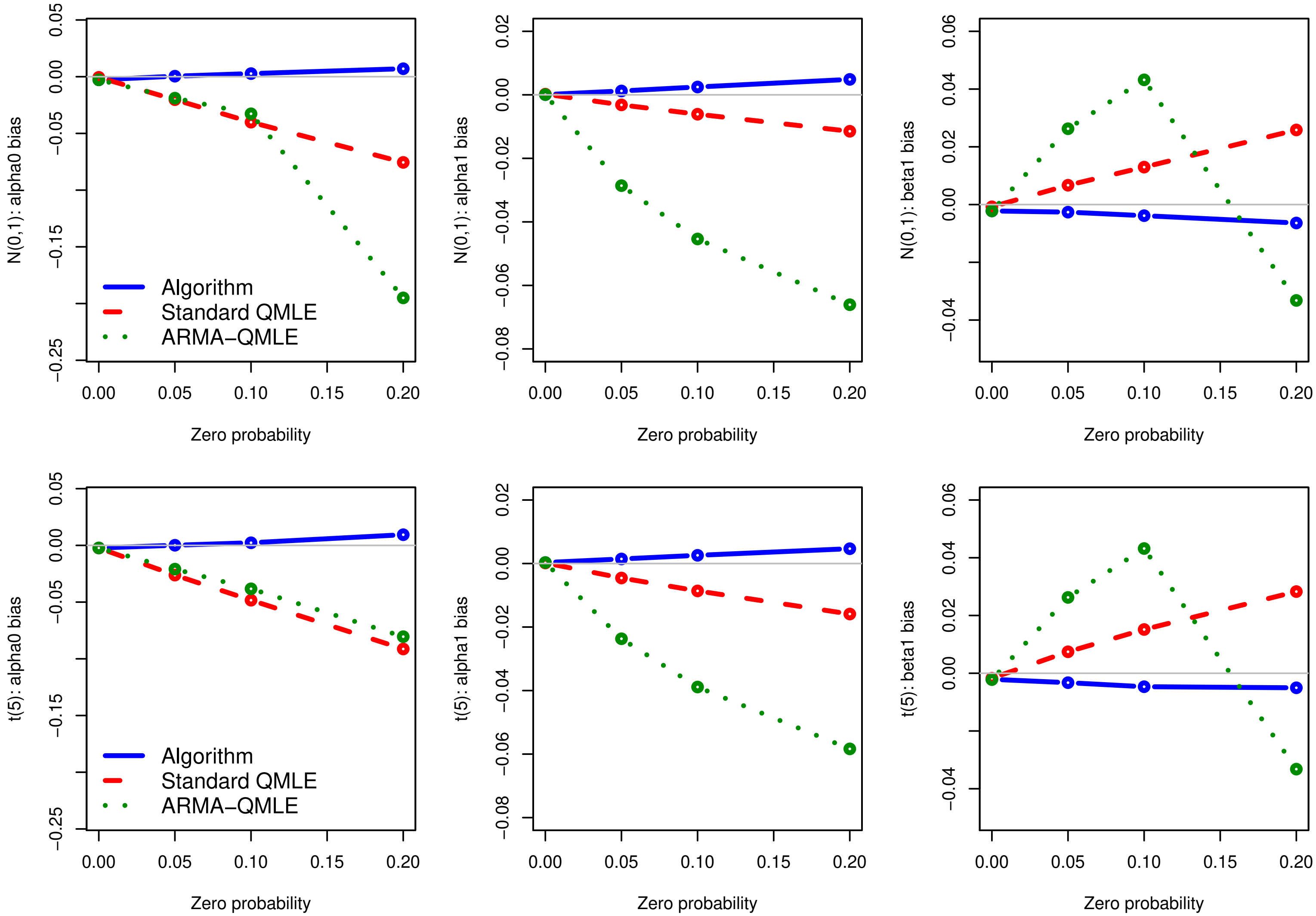

Figure 1: Bias (*estimate* − *true value*) of three estimators for parameter combination A. Solid blue line (Algorithm): Gaussian ARMA-QMLE w/algorithm, see Section 3.2. Dashed red line: Standard QMLE, see Section 3.1. Dotted green line: Gaussian ARMA-QMLE with zeros replaced by a small value, see Section 3.1

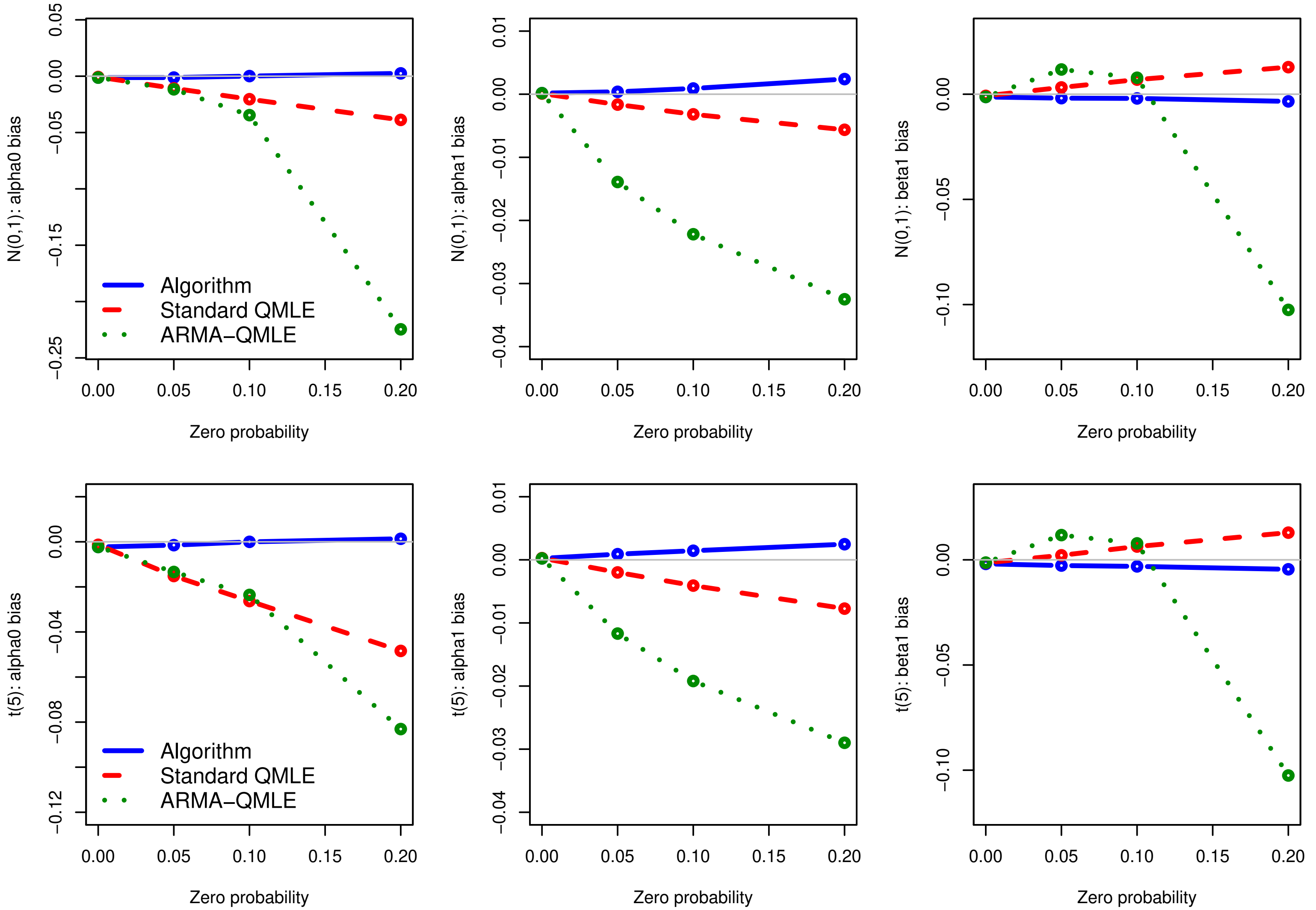

Figure 2: Bias (*estimate* − *true value*) of three estimators for parameter combination B. Solid blue line (Algorithm): Gaussian ARMA-QMLE w/algorithm, see Section 3.2. Dashed red line: Standard QMLE, see Section 3.1. Dotted green line: Gaussian ARMA-QMLE with zeros replaced by a small value, see Section 3.1

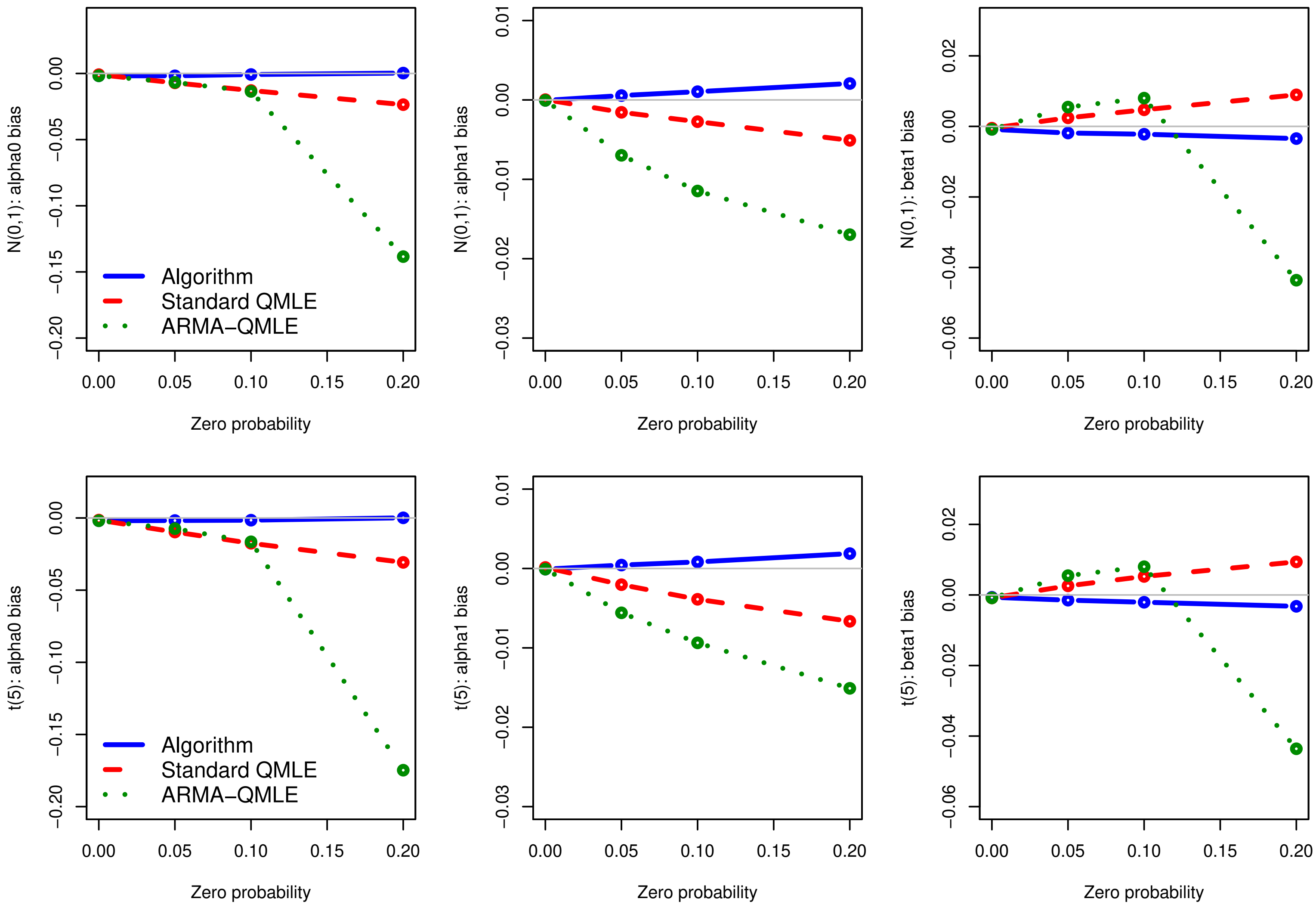

Figure 3: Bias (*estimate* − *true value*) of three estimators for parameter combination C. Solid blue line (Algorithm): Gaussian ARMA-QMLE w/algorithm, see Section 3.2. Dashed red line: Standard QMLE, see Section 3.1. Dotted green line: Gaussian ARMA-QMLE with zeros replaced by a small value, see Section 3.1

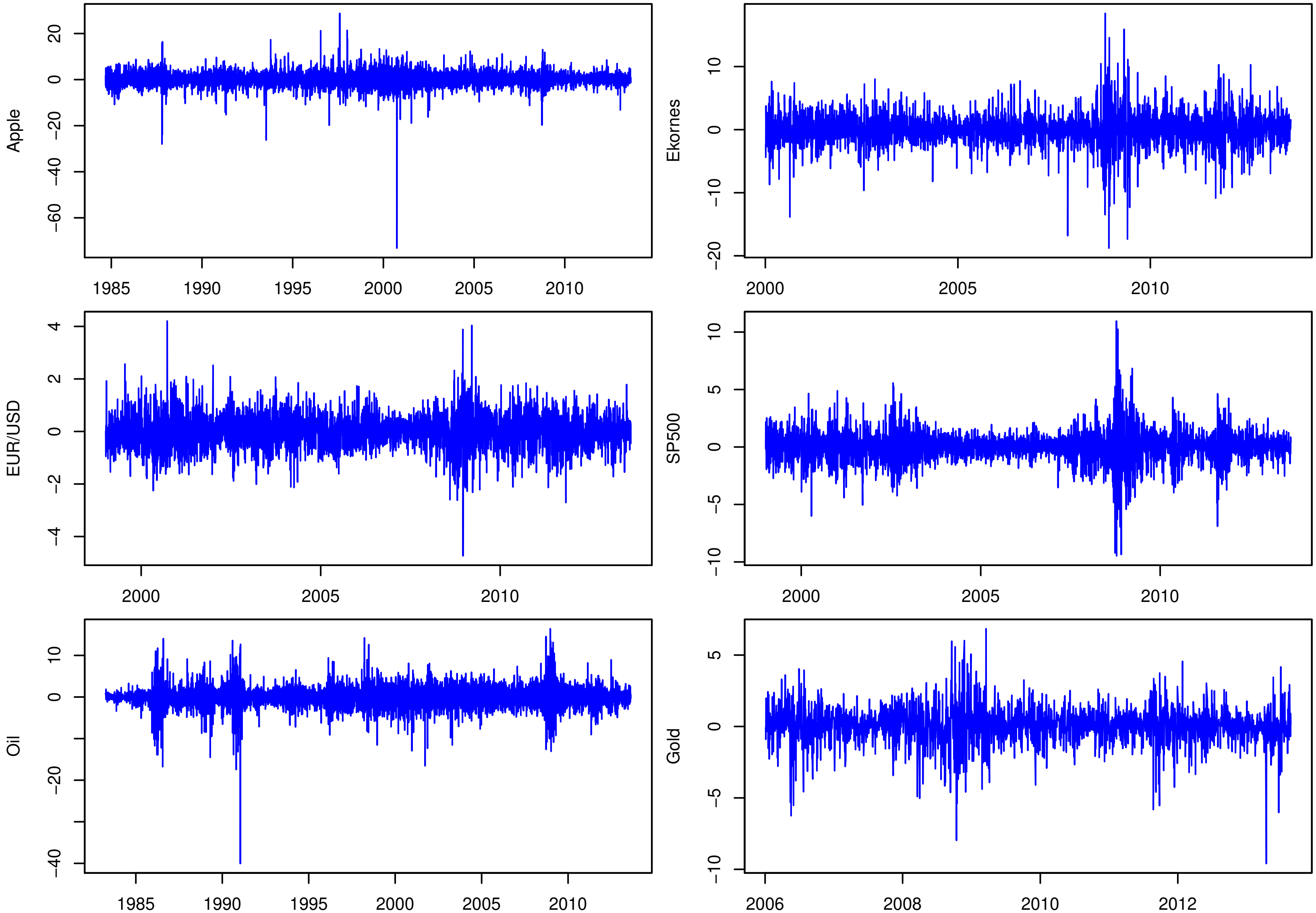

Figure 4: Daily financial log-returns in percent

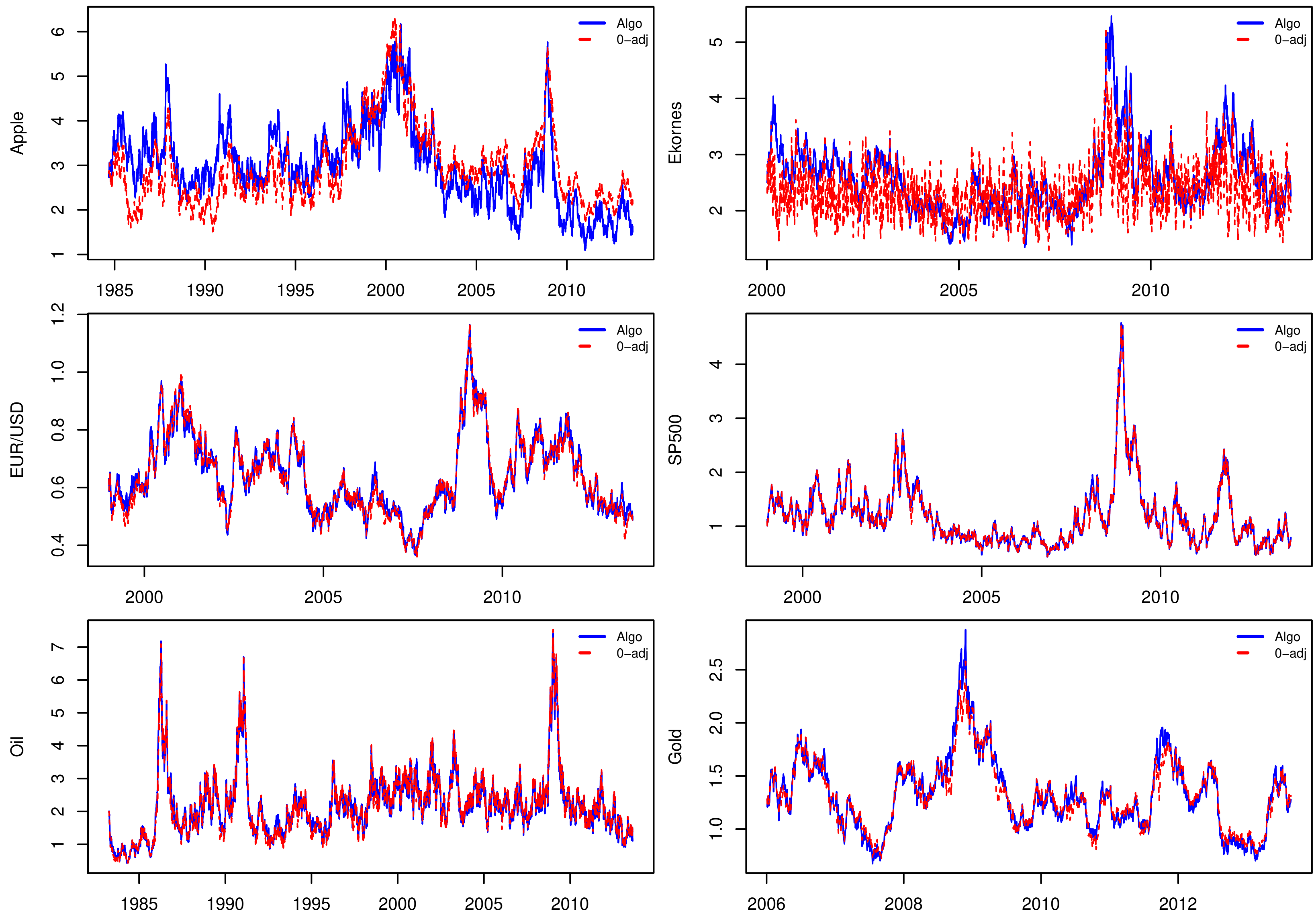

Figure 5: Fitted conditional standard deviations ($\hat{\sigma}_{t,Algo}$ solid blue line, $\hat{\sigma}_{t,\text{0-adj}}$ dashed red line)

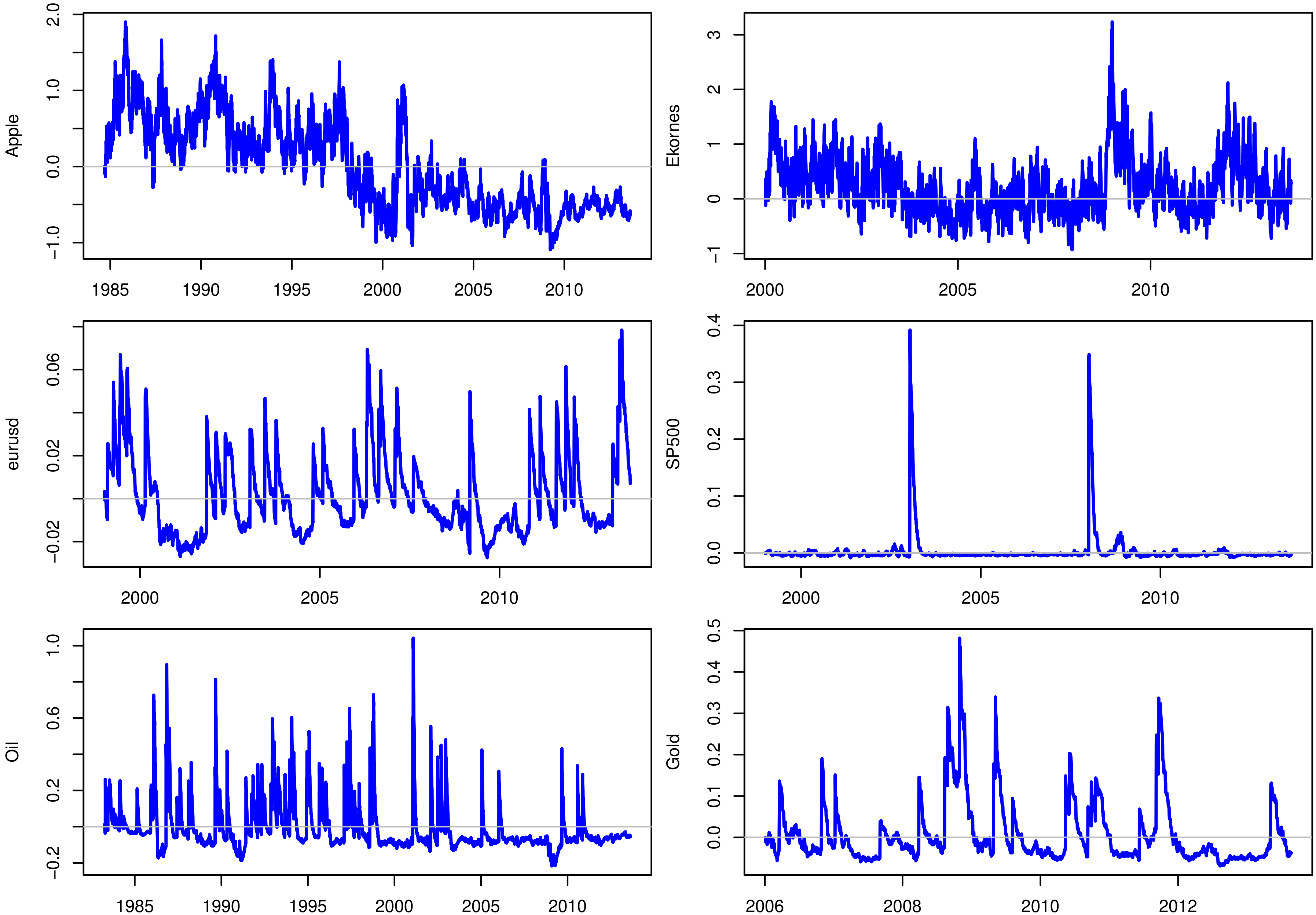

Figure 6: Difference of fitted conditional standard deviations (*i.e.* $\hat{\sigma}_{t,\text{Algo}} - \hat{\sigma}_{t,\text{0-adj}}$)

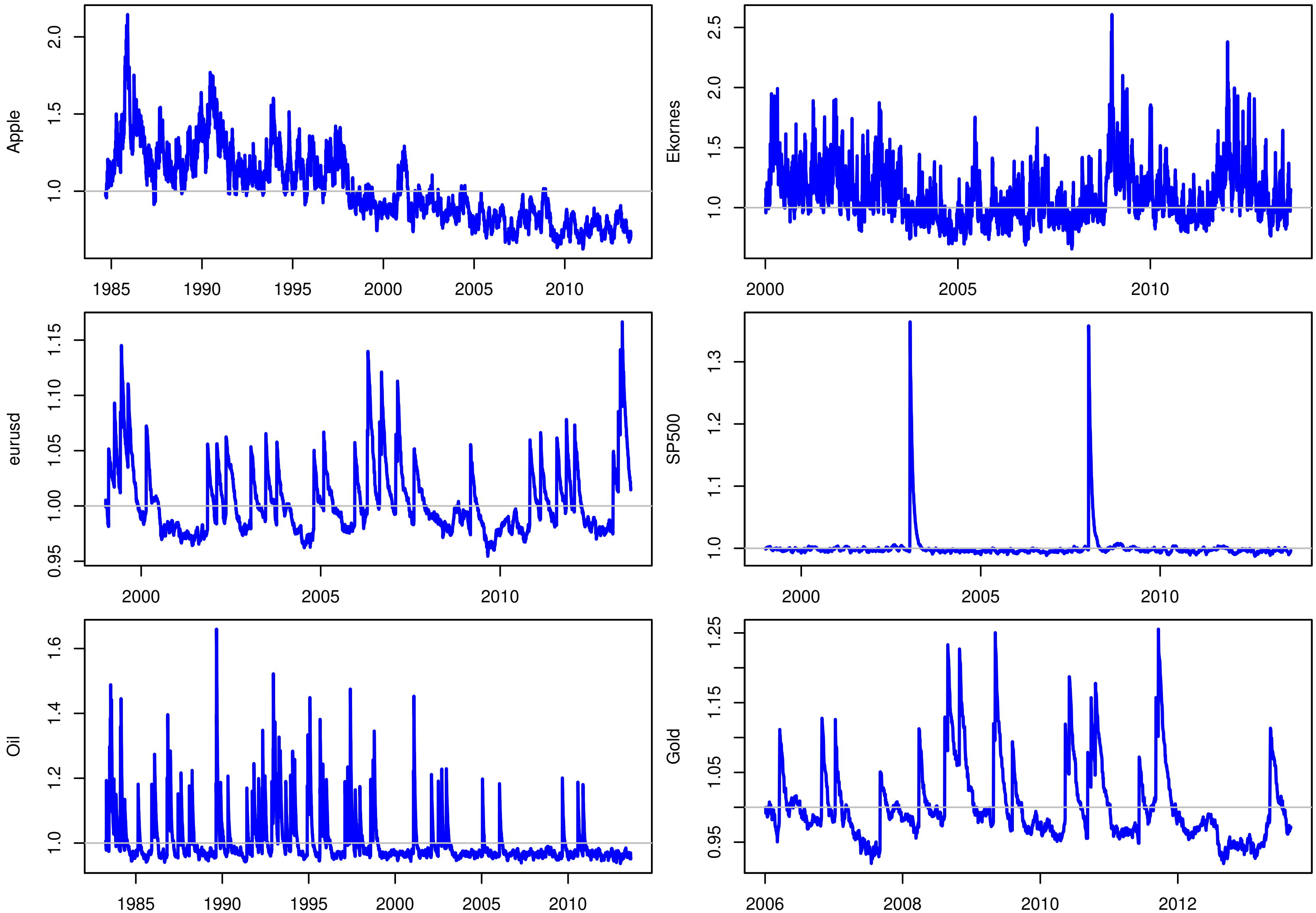

Figure 7: Ratios of fitted conditional standard deviations (*i.e.* $\hat{\sigma}_{t,\text{Algo}}/\hat{\sigma}_{t,\text{0-adj}}$)

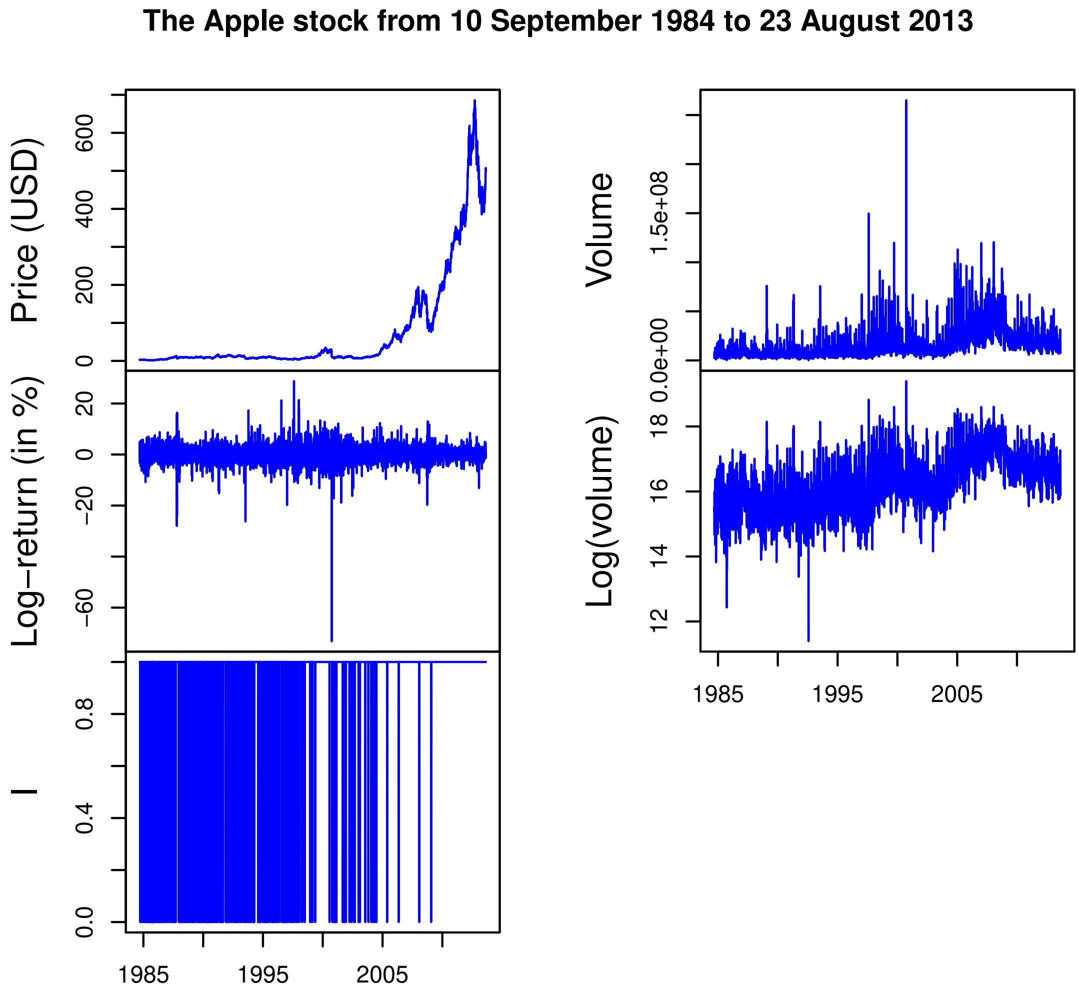

Figure 8: The Apple stock price (in USD), observed return, I_t, volume and log-volume from 10 September 1984 to 23 August 2013 ($T = 7303$ observations)