Models of Financial Return With Time-Varying Zero Probability

Sucarrat, Genaro and Grønneberg, Steffen

BI Norwegian Business School, BI Norwegian Business School

17 January 2016

Models of Financial Return With Time-Varying Zero Probability*

Genaro Sucarrat†and Steffen Grønneberg‡

17 January 2016

Abstract

The probability of an observed financial return being equal to zero is not necessarily zero. This can be due to price discreteness or rounding error, liquidity issues (*e.g.* low trading volume), market closures, data issues (*e.g.* data imputation due to missing values), characteristics specific to the market, and so on. Moreover, the zero probability may change and depend on market conditions. In standard models of return volatility, however, *e.g.* ARCH, SV and continuous time models, the zero probability is zero, constant or both. We propose a new class of models that allows for a time-varying zero probability, and which can be combined with standard models of return volatility: They are nested and obtained as special cases when the zero probability is constant and equal to zero. Another attraction is that the return properties of the new class (*e.g.* volatility, skewness, kurtosis, Value-at-Risk, Expected Shortfall) are obtained as functions of the underlying volatility model. The new class allows for autoregressive conditional dynamics in both the zero probability and volatility specifications, and for additional covariates. Simulations show parameter and risk estimates are biased if zeros are not appropriately handled, and an application illustrates that risk-estimates can be substantially biased in practice if the time-varying zero probability is not accommodated.

JEL Classification: C01, C22, C32, C51, C52, C58
Keywords: Financial return, volatility, zero-inflated return, GARCH, log-GARCH, ACL

*We are grateful to participants at the SNDE Annual Symposium 2015 (Oslo) and IAAE Conference (Thessaloniki) for useful comments, suggestions and questions.

†Corresponding author. Department of Economics, BI Norwegian Business School. Email: genaro.sucarrat@bi.no. Webpage: http://www.sucarrat.net/

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

Contents:

1 Introduction

It is well-known that the probability of an observed financial return being equal to zero is not necessarily zero. This can be due to price discreteness and/or rounding error, liquidity issues (*e.g.* low trading volume), market closures, data issues (*e.g.* data imputation due to missing values), characteristics specific to the market, and so on. Moreover, the zero probability may change and depend on market conditions. In standard models of financial return volatility, however, the probability of a zero return is either zero, or non-zero but constant. Examples include the Autoregressive Conditional Heteroscedasticity (ARCH) class of models of Engle (1982), the Stochastic Volatility (SV) class of models (see Shephard (2005)) and continuous time models (*e.g.* Brownian motion).[1] Hausman et al. (1992) relaxed the constancy assumption by allowing the zero probability to depend on other conditioning variables (*e.g.* volume, duration and past returns) in a probit framework. This was then extended in two different directions by Engle and Russell (1998), and Russell and Engle (2005), respectively. The latter in particular provides a comprehensive framework, since there price-changes are modelled by an Autoregressive Conditional Multinomial

[1] Bauwens et al. (2012) provides a recent survey of these models.

(ACM) model coupled with a continuous time model of the durations between trades. However, as pointed out by Liesenfeld et al. (2006), there are several limitations and drawbacks with this approach. Instead, they propose a dynamic integer count model, which is extended to the multivariate case in Bien et al. (2011). Finally, Rydberg and Shephard (2003) propose a framework in which the price increment is decomposed multiplicatively into three components: Activity, direction and integer magnitude.

Even though discrete models in many cases may provide a more accurate characterisation of observed returns, the most common models in empirical practice – *e.g.* ARCH, SV and continuous time models – are continuous. Arguably, the discreteness-point that causes the biggest problem for continuous models is located at zero. This is because zero often is the most frequently observed single value – particularly in intraday data, and because its probability is often time-varying and dependent on random or non-random events (*e.g.* periodicity), or both. A time-varying zero probability invalidates the parameter and risk estimates of continuous models, since the underlying estimation theory relies on the assumption that the conditional density is identical over time. We propose a new class of financial return models that allows for a time-varying conditional probability of a zero return. The new class decomposes return multiplicatively into a continuous part, which can be specified in terms of common volatility models, and a discrete part at zero. Standard volatility models (*e.g.* ARCH, SV and continuous time models) are therefore nested and obtained as special cases when the zero probability is constant and equal to zero. Hautsch et al. (2013) proposed a model for positively valued variables (*e.g.* volume) that uses a similar decomposition to ours. However, their dynamics is governed by a (restricted) log-GARCH specification, and by a specific conditional density. Our model is much more general. The volatility dynamics need not be specified as a log-GARCH, and the continuous density (in squared return) need not be a Generalised F. In fact, their model is nested and obtained as a very specific case in our model class, when the log-volatility dynamics is interpreted as a Multiplicative Error Model (MEM) (see Brownlees et al. (2012) for a recent survey of MEM models). Another attraction of our model class is that many return properties (*e.g.* conditional volatility, return skewness, Value-at-Risk and Expected Shortfall) are readily obtained as functions of the underlying volatility model. Moreover, our model allows for autoregressive conditional dynamics in both the zero probability and volatility specifications, and for a two-way feedback between the two. In the absence of a feedback effect from volatility to the zero probability specification, then estimation becomes particularly simple, since the model of zero probability and the model of volatility can then be estimated separately. The model is readily extended to include additional conditioning variables (*e.g.* leverage, volume, duration, spreads, volatility proxies, periodicity/seasonality terms, etc.) in the zero probability or volatility specifications, or in both, and by introducing new endogenous variables (*e.g.* volume, durations, spreads, volatility proxies, etc.) to form a complete dynamic system. Simulations show that common volatility models are inconsistently estimated by common methods if the zero probability is time-varying, and that estimates of risk (*i.e.* conditional volatility, Value-at-Risk and Expected Shortfall) are biased upwards. Finally, an empirical illustration shows that risk estimates can be substantially biased in practice if the time-varying zero probability is not accommodated appropriately.

The rest of the paper is organised as follows. Section 2 presents the new class and derives some general properties. Section 3 proposes specific models of the zero probability and of the volatility, and a joint model that allows for a two-way feedback. Section 4 contains a Monte Carlo study of the parameter and risk estimation bias induced in some common models of volatility, when the time-varying zero probability is not appropriately accommodated. Section 5 contains our empirical application, whereas Section 6 concludes. The Appendix contains auxiliary derivations, and additional material and simulations. Tables and figures are located at the end.

2 Financial returns with time-varying zero probability

2.1 The ordinary model of return

The ordinary model of a financial return r_t (possibly mean-corrected) is given by

$$r_t = \sigma_t w_t, \quad w_t \sim IID(0, \sigma_w^2), \quad P_{t-1}(w_t = 0) = 0, \quad t \in \mathbb{Z}, \tag{1}$$

where $\sigma_t > 0$ is a time-varying scale or volatility (that needs not equal the conditional standard deviation), $w_t \in \mathbb{R}$ is an Independentically and Identically Distributed (IID) innovation conditional on the past $\mathcal{I}_{t-1}$ and $P_{t-1}(w_t = 0)$ is the zero probability of w_t conditional on the past. The subscript $t-1$ is thus notational shorthand for conditioning on past information $\mathcal{I}_{t-1}$. We refer to (1) as an "ordinary" model of return, since the zero probability of return r_t is 0 and constant. An example of an ordinary model is the GARCH(1,1) of Bollerslev (1986), where

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad w_t \sim IID(0, \sigma_w^2 = 1). \tag{2}$$

Another example is the Stochastic Volatility (SV) model, where

$$\ln \sigma_t^2 = \alpha_0 + \beta_1 \ln \sigma_{t-1}^2 + \eta v_{t-1}, \quad v_t \sim IID(0, \sigma_v^2), \tag{3}$$

with v_i being independent of w_j for all pairs i, j. Other examples include quadratic variation (*e.g.* Brownian motion) and other continuous time notions of volatility, the log-GARCH class proposed independently by Geweke (1986), Pantula (1986) and Milhøj (1987), the EGARCH model of Nelson (1991), the mixed data sampling (MIDAS) regression of Ghysels et al. (2006), and the Dynamic Conditional Score (DCS)/Generalised Autoregressive Conditional Score (GAS) models of Harvey (2013) and Creal et al. (2013).

2.2 The model of return with time-varying zero probability

The model of return with time-varying conditional zero probability is given by

$$\begin{aligned} r_t &= \sigma_t z_t, \quad z_t = w_t I_t \pi_{1t}^{-1/2}, \quad w_t \sim IID(0, \sigma_w^2), \quad P_{t-1}(w_t = 0) = 0, && (4)\\ I_t &\in \{0, 1\}, \quad \pi_{1t} = P_{t-1}(I_t = 1), \quad 0 < \pi_{1t} \leq 1, \quad I_t \perp w_t. && (5) \end{aligned}$$

The variable I_t determines whether return r_t is zero or not. If $I_t = 1$, then $r_t \neq 0$ with probability 1, and if $I_t = 0$, then $r_t = 0$. The probability of a zero return conditional on the past is thus $\pi_{0t} = 1 - \pi_{1t}$. 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.* $\pi_{0t} = 1 - \pi_{1t}$). The symbolism $I_t \perp w_t$ means I_t and w_t are independent at t conditional on the past. The motivation for letting π_{1t} enter the way it does in z_t, is that this ensures that $Var_{t-1}(z) = \sigma_w^2$ (see Property 2 in Section 2.3).

In (4), σ_t can be specified as a wide range of volatility models in terms of the zero-adjusted return

$$\widetilde{r}_t = \sigma_t w_t. \tag{6}$$

We refer to this quantity as "zero-adjusted" return, since $\widetilde{r}_t = r_t \pi_{1t}^{1/2}$ whenever $I_t = 1$. For example, the GARCH(1,1) model in terms of zero-adjusted return is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{7}$$

whereas the zero-adjusted log-GARCH(1,1) model is given by

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \widetilde{r}_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2. \tag{8}$$

In both cases their ordinary counterparts are obtained as special cases when π_{1t} is constant and equal to 1. In empirical practice we observe r_t rather than the zero-adjusted return $\widetilde{r}_t$. But for a given set of values of π_{1t} (or estimates of π_{1t}, rather), we can obtain $\widetilde{r}_t$ (or an estimate of $\widetilde{r}_t$, rather) whenever $I_t = 1$, since then we have $\widetilde{r}_t = r_t \pi_{1t}^{1/2}$. Whenever $I_t = 0$, the zero-adjusted return $\widetilde{r}_t$ will be unobserved or "missing". Nevertheless, algorithms that handles missing values can be used for estimation and inference. Details of how we implement estimation with missing values is given in Section 3.2 and in the Appendix . An alternative way of specifying σ_t, which avoids the need for an algorithm that handles missing values, is to let the zero-adjusted return enter the volatility equation only at non-zero locations. This is for example the strategy employed by Hautsch et al. (2013) in their model of positively-valued variables (*e.g.* volume). For example, a simplified version of their log-GARCH(1,1) specification is $\ln \sigma_t^2 = \alpha_0 + \alpha_1 (\ln \widetilde{r}_{t-1}^2) I_{t-1} + \ln \sigma_{t-1}^2$.

2.3 Some general properties

An attractive feature of the model (4)–(5) is that many of its properties follow straightforwardly (assuming the quantities in question exist) as a function of the underlying models of volatility and zero probability. These properties are:

1. Ordinary models of return are obtained when the zero probability is constant and equal to zero, *i.e.* when $\pi_{1t} = 1$, for all t.

2. Although z_t is not IID conditional on the past when π_{1t} is non-constant, z_t does have a constant conditional mean $E_{t-1}(z_t)$ equal to zero, and a constant conditional second moment $Var_{t-1}(z_t)$ equal to σ_w^2. As a consequence, the return series $\{r_t\}$ remains a martingale difference sequence even though π_{1t} is non-constant.

3. Both the conditional and unconditional second-moment properties of the underlying volatility model are retained:

$$Var_{t-1}(r_t) = \sigma_t^2 \sigma_w^2, \quad \text{and} \quad Var(r_t) = E(\sigma_t^2)\sigma_w^2. \tag{9}$$

   In a zero-adjusted stationary GARCH(1,1), for example, where the identifiability condition is $\sigma_w^2 = 1$, we have that $Var_{t-1}(r_t) = \sigma_t^2$ and $Var(r_t) = \alpha_0/(1 - \alpha_1 - \beta_1)$. This holds regardless of whether π_{1t} is constant or time-varying.

4. The conditional second-moment assuming return is non-zero is given by

$$Var(r_t | I_t = 1, \mathcal{I}_{t-1}) = \sigma_t^2 \pi_{1t}^{-1} \sigma_w^2. \tag{10}$$

   In other words, conditional volatility is scaled upwards under the assumption that return is non-zero, and the more so the higher the conditional zero probability π_{0t}.

5. The sth. conditional moment of return is given by

$$E_{t-1}(r_t^s) = \sigma_t^s \pi_{1t}^{(2-s)/2} E(w_t^s). \tag{11}$$

   Higher order (*i.e.* $s > 2$) conditional moments (in absolute value) are thus scaled upwards by positive conditional zero probabilities, whereas the opposite is the case for lower order (*i.e.* $s < 2$) conditional moments (in absolute value). In particular, conditional skewness ($s = 3$) and conditional kurtosis ($s = 4$) become more pronounced when $\pi_{0t} > 0$, whereas $E_{t-1}(r_t)$ in absolute value is usually scaled downwards since $E|w_t| < 1$ for most densities of empirical relevance.

6. If there is no feedback between π_{1t} and σ_t, and if their unconditional moments exist, then the sth. unconditional moment is

$$E(r_t^s) = E(\sigma_t^s) E(\pi_{1t}^{(2-s)/2}) E(w_t^s). \tag{12}$$

   The effect is thus similar to that of conditional moments: Higher order (*i.e.* $s > 2$) unconditional moments (in absolute value) are scaled upwards by positive zero probabilities, whereas the opposite is the case for lower order (*i.e.* $s < 2$) unconditional moments (in absolute value)

7. The probability density function (pdf) and cumulative distribution function (cdf) of z_t conditional on the past are, respectively, given by

$$f_z(z_t) = \begin{cases} \pi_{1t}^{3/2} f_w(z_t \pi_{1t}^{1/2}) \text{ if } z_t \neq 0, \\ (1 - \pi_{1t}) \text{ if } z_t = 0, \end{cases} \tag{13}$$

$$F_z(z_t) = \pi_{1t} F_w(z_t \pi_{1t}^{1/2}) + 1_{\{z_t \geq 0\}}(1 - \pi_{1t}), \tag{14}$$

   where f_w is the pdf of w_t conditional on the past, F_w is the cdf of w_t conditional on the past and $1_{\{z_t \geq 0\}}$ is an indicator function equal to 1 if $z_t \geq 0$ and 0 otherwise.

8. The pdf and cdf of r_t conditional on the past are

$$f_r(r_t) = \begin{cases} \pi_{1t}^{3/2} f_{\widetilde{r}}(r_t \pi_{1t}^{1/2}) \text{ if } r_t \neq 0, \\ (1-\pi_{1t}) \text{ if } r_t = 0, \end{cases} \tag{15}$$

$$F_r(r_t) = \pi_{1t} F_{\widetilde{r}}(r_t \pi_{1t}^{1/2}) + 1_{\{r_t \geq 0\}}(1-\pi_{1t}), \tag{16}$$

where $f_{\widetilde{r}}(\widetilde{r}_t) = \sigma_t^{-1} f_w(\widetilde{r}_t/\sigma_t)$ is the pdf of $\widetilde{r}_t$ conditional on the past, $F_{\widetilde{r}}(\widetilde{r}_t) = F_w(\widetilde{r}_t/\sigma_t)$ is the cdf of $\widetilde{r}_t$ conditional on the past and $1_{\{r_t \geq 0\}}$ is an indicator function equal to 1 if $r_t \geq 0$ and 0 otherwise.

9. If F_w is strictly increasing and $c \in (0,1)$, then the cth. quantile of z_t and r_t conditional on the past are

$$z_{c,t} = F_z^{-1}(c) = \begin{cases} \pi_{1t}^{-1/2} F_w^{-1}(c/\pi_{1t}) & \text{if } c < F_w(0)\pi_{1t} \\ 0 & \text{if } F_w(0)\pi_{1t} \leq c < F_w(0)\pi_{1t} + \pi_{0t} \\ \pi_{1t}^{-1/2} F_w^{-1}\left[\frac{(c-\pi_{0t})}{\pi_{1t}}\right] & \text{if } c \geq F_w(0)\pi_{1t} + \pi_{0t}, \end{cases} \tag{17}$$

$$r_{c,t} = F_r^{-1}(c) = \begin{cases} \pi_{1t}^{-1/2} F_{\widetilde{r}}^{-1}(c/\pi_{1t}) & \text{if } c < F_{\widetilde{r}}(0)\pi_{1t} \\ 0 & \text{if } F_{\widetilde{r}}(0)\pi_{1t} \leq c < F_{\widetilde{r}}(0)\pi_{1t} + \pi_{0t} \\ \pi_{1t}^{-1/2} F_{\widetilde{r}}^{-1}\left[\frac{(c-\pi_{0t})}{\pi_{1t}}\right] & \text{if } c \geq F_{\widetilde{r}}(0)\pi_{1t} + \pi_{0t}. \end{cases} \tag{18}$$

The *conditional* $(100 \cdot c)\%$ Value-at-Risk (VaR$_c$) of z_t and r_t, respectively, are therefore defined as $-z_{c,t}$ and $-r_{c,t}$. Note that, since $F_{\widetilde{r}}^{-1}(x) = \sigma_t F_w^{-1}(x)$, equation (18) can be written as

$$r_{c,t} = \sigma_t F_z^{-1}(c). \tag{19}$$

This is particularly convenient when the density of $\widetilde{r}_t$ is unknown (*e.g.* when estimation is by QML).

10. If F_w is strictly increasing and $c > 0$, then the $(100 \cdot c)\%$ Expected Shortfall (ES$_c$) of z_t and r_t *conditional* on the past are given by $-E_{t-1}(z_t|z_t \leq z_{c,t})$ and $-E_{t-1}(r_t|r_t \leq r_{c,t})$, respectively, where

$$E_{t-1}(z_t|z_t \leq z_{c,t}) = \frac{\pi_{1t}}{c} E_{t-1}\left(w_t 1_{\{w_t \leq F_w^{-1}(c/\pi_{1t})\}}\right) \quad \text{if } c < F_w(0)\pi_{1t}, \tag{20}$$

$$E_{t-1}(r_t|r_t \leq r_{c,t}) = \frac{\pi_{1t}}{c} E_{t-1}\left(\widetilde{r}_t 1_{\{\widetilde{r}_t \leq F_{\widetilde{r}}^{-1}(c/\pi_{1t})\}}\right) \quad \text{if } c < F_{\widetilde{r}}(0)\pi_{1t}. \tag{21}$$

Note that, since $F_{\widetilde{r}}^{-1}(x) = \sigma_t F_w^{-1}(x)$, equation (21) can also be written as

$$E_{t-1}(r_t|r_t \leq r_{c,t}) = \sigma_t E_{t-1}(z_t|z_t \leq z_{c,t}) \quad \text{if} \quad c < F_{\widetilde{r}}(0)\pi_{1t}. \tag{22}$$

This formulation is particularly convenient when the density of $\widetilde{r}_t$ is unknown (*e.g.* when estimation is by QML).

3 Models of π_{1t} and σ_t

If $0 < \pi_{1t} < 1$, then it follows from (15) that the log-likelihood at t conditional on the past can be written as

$$\begin{aligned} \ln f_r(r_t) &= I_t \ln f_{\tilde{r}}(r_t \pi_{1t}^{1/2}) + I_t \ln \pi_{1t}^{3/2} + (1 - I_t) \ln(1 - \pi_{1t}), && (23) \\ &= I_t \ln f_{\tilde{r}}(\widetilde{r}_t) + I_t \ln \pi_{1t} + (1 - I_t) \ln(1 - \pi_{1t}), && (24) \end{aligned}$$

since $f_{\tilde{r}}(\widetilde{r}_t) = \pi_{1t}^{-1/2} f_{\tilde{r}}(r_t \pi_{1t}^{1/2})$ when $I_t = 1$. The total log-likelihood is therefore given by $\sum_{t=1}^{n} \ln f_r(r_t) = LogL_\sigma + LogL_\pi$, where

$$LogL_\sigma = \sum_{t=1}^{n} I_t \ln f_{\tilde{r}}(\widetilde{r}_t) \quad \text{and} \quad LogL_\pi = \sum_{t=1}^{n} I_t \ln \pi_{1t} + (1 - I_t) \ln(1 - \pi_{1t}). \tag{25}$$

When there is no feedback from σ_t to π_{1t}, then the models of σ_t and π_{1t}, respectively, can be estimated in two separate steps. First, π_{1t} can be estimated by maximising $LogL_\pi$. Second, the fitted values of π_{1t} can be used to generate estimates of $\widetilde{r}_t$, which can subsequently be used to estimate σ_t by maximising $LogL_\sigma$. In the next two subsections we consider such models that can be estimated in two separate steps. Then, in subsection 3.3, we propose a joint model of π_{1t} and σ_t with feedback effects.

3.1 Models of π_{1t}

The model (4)-(5) admits a wide range of specifications of π_{1t}. In the simplest, π_{1t} is constant and can be estimated by computing the fraction of non-zero returns. Another simple specification, which may be particularly useful in the presence of periodic, say, intraday zero probabilities, is $\pi_{1t} = \sum_{i=1}^{24} \phi_i d_{it}$, where the d_{it}'s are dummy variables associated with the hours of the day. Then each ϕ_i is readily estimated by computing the fraction of non-zero returns in each hour of the day. The logistic representation of this model is $h_t = \rho_0 + \sum_{i=2}^{24} \lambda_i d_{it}$, where $\pi_{1t} = 1/(1 + \exp(-h_t))$. A third class of models that is straightforwardly estimated is one in which π_{1t} is a function of a deterministic time-trend, say, $h_t = \rho_0 + \lambda t$. The motivation for this specification is that market developments (*e.g.* the influx of high-frequency algorithmic trading, increased trading volume, increased quoting frequency, lower tick-size, etc.) may have reduced the probability of zeros in a gradual and monotonous way.

In many situations the models just described are not sufficient to adequately capture the zero probability dynamics. The π_{1t} may be autoregressively dependent and/or determined by other variables (*e.g.* volume, news, etc.), either instead of or in addition to periodicity effects and trends. For computational simplicity one could consider modelling all this in a linear probability model with I_t as the left-hand side variable, with the usual problems that this entails (*e.g.* fitted probabilities outside the unit interval). Another option, which avoids the drawbacks of the linear probability model, is the dynamic logit model proposed by Hautsch et al. (2013) for trading volume. This model is a special case of the Autoregressive Conditional Multinomial

(ACM) model by Russell and Engle (2005), and its specification is

$$h_t = \rho_0 + \sum_{k=1}^{K} \rho_k s_{t-k} + \sum_{l=1}^{L} \zeta_l h_{t-l}, \quad s_t = \frac{I_t - \pi_{1t}}{\sqrt{\pi_{1t}(1-\pi_{1t})}}, \quad s_t \in \mathbb{R}, \tag{26}$$

where s_t conditional on the past is $IID(0,1)$. We will henceforth refer to (26) as an Autoregressive Conditional Logit (ACL) model. The $\{h_t\}$ is an ARMA process, so in the first order case, *i.e.* $h_t = \rho_0 + \rho_1 s_{t-1} + \zeta_1 h_{t-1}$, each parameter has the usual ARMA-like interpretation. The ρ_0 controls the level of the unconditional probability $E(\pi_{1t})$, whereas the ρ_1 controls the impact of a shock s_t: The larger in absolute value, the greater the departure from recent values of π_{1t}. The ζ_1 is a persistence parameter ($|\zeta_1| < 1$ entails stability): The closer to 1, the higher persistence of π_{1t}. Finally, the ACL can be augmented with additional conditioning information or covariates by including them in the h_t specifiction, yielding an ACL-X model of π_{1t}.

3.2 Models of σ_t

When there is no feedback from σ_t to π_{1t}, then the former can be estimated in a second step conditional on the estimates of π_{1t}. One strategy in formulating a specification for σ_t, or $\ln \sigma_t$, is to simply skip the zeros. An example of this is Francq et al. (2013), where the log-GARCH(1,1) specification (without asymmetry) is $\ln \sigma_t^2 = \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 I_{t-1} + \beta_1 \ln \sigma_{t-1}^2$. This, in fact, this is equivalent to replacing a zero on $\widetilde{r}_t$ with the value 1 (rather than a small non-negative value, which is the more common approach), and may not the best solution empirically since it creates a "jumpy" or erratic contribution from the log-ARCH term. A similar effect is induced in the GARCH model if the ARCH-term $\widetilde{r}_{t-1}^2$ is replaced by $\widetilde{r}_{t-1}^2 I_{t-1}$. Moreover, if the model relies on the assumption that $P_{t-1}(w_t = 0) = 0$, then parameter estimates will in fact be biased, see Sucarrat and Escribano (2013). An alternative strategy is to treat zeros as "missing values". The main advantage with this approach is that the contribution of the ARCH-term is less erratic, and that the properties of the volatility model in question carry over more straighforwardly. In particular, the properties derived in Section 2.3 are easier to exploit.

We propose the following general procedure to estimate models of σ_t while treating $\widetilde{r}_t$ as missing at zero locations:

1. Record the locations at which the observed return r_t is zero and non-zero, respectively. Use these locations to estimate π_{1t} by maximising $LogL_\pi$.

2. Obtain an estimate of $\widetilde{r}_t$ by multiplying r_t with $\widehat{\pi}_{1t}^{1/2}$, where $\widehat{\pi}_{1t}$ is the fitted value of π_{1t} from Step 1. At zero locations the zero-adjusted return $\widetilde{r}_t$ is unobserved or "missing".

3. Use an estimation procedure that handles missing values to estimate the volatility model σ_t by maximising $LogL_\sigma$.

Sucarrat and Escribano (2013) propose an algorithm for the log-GARCH model where missing values are replaced by estimates of the conditional expectation. If Gaussian (Q)ML is used for estimation, then this can be viewed as a variant of the Expectation

Maximisation (EM) algorithm. A similar algorithm can be devised for many additional volatility models, including the GARCH model. The Appendix contains the details of this algorithm, whereas Section 4.2 contains a small simulation study that compares its accuracy with ordinary methods.

Additional conditioning variables or covariates ("X"), *e.g.* past values of leverage, volume, duration and periodicity/seasonality terms, can be added to the GARCH and log-GARCH specifications, respectively, to form GARCH-X and log-GARCH-X models, see Han and Kristensen (2014), Francq and Thieu (2015), Sucarrat et al. (2015), and Francq and Sucarrat (2015). A zero-adjusted version of the log-GARCH-X model constitutes a particularly attractive alternative, since it does not impose any negativity restrictions on the parameters, and since estimation, inference and missing values can be handled via its ARMA-X representation.

3.3 A joint model of π_{1t} and σ_t with two-way feedback

If there is feedback from σ_t to π_{1t}, then the models of π_{1t} and σ_t cannot be estimated separately. Here, we propose a joint model with two-way feedback together with a QML estimation procedure. The model is a combination of the log-GARCH model and the ACL model, so we refer to it as a log-GARCH-ACL model. In general form the model is

$$\begin{aligned}
\ln \sigma_t^2 &= \alpha_0 + \sum_{p=1}^{P} \alpha_{1p} \ln \widetilde{r}_{t-p}^2 + \sum_{k=1}^{K} \rho_{1k} s_{t-k} + \sum_{q=1}^{Q} \beta_{1q} \ln \sigma_{t-q}^2 + \sum_{l=1}^{L} \zeta_{1l} h_{t-l} && (27)\\
h_t &= \rho_0 + \sum_{p=1}^{P} \alpha_{2p} \ln \widetilde{r}_{t-p}^2 + \sum_{k=1}^{K} \rho_{2k} s_{t-k} + \sum_{q=1}^{Q} \beta_{2q} \ln \sigma_{t-q}^2 + \sum_{l=1}^{L} \zeta_{2l} h_{t-l}. && (28)
\end{aligned}$$

The exponential specifications ensure the positivity of σ_t and π_{1t}, and enable more flexible dynamics (*e.g.* parameters can be negative). Also, the logit-specification ensures that $\pi_{1t} \in (0, 1)$. The motivation for the log-GARCH specification for $\ln \sigma_t^2$ instead of, say, the EGARCH of Nelson (1991), is that the latter is not amenable to QML estimation and inference in the presence of missing values of $\widetilde{r}_t$.[2] Also, it is not clear that an ML-based procedure for the EGARCH would yield consistent estimates due to invertibility issues, see Wintenberger (2013), since the invertibility-problem is in fact compounded in the presence of missing values.

For notational economy we will hereafter work with the first order specification only. The first order version of the model can be written as

$$y_t = \omega + A_1 v_{t-1} + B_1 y_{t-1}, \tag{29}$$

where $y_t = (\ln \sigma_t^2, h_t)'$, $\omega = (\alpha_{10}, \rho_{10})'$, $v_t = (\ln \widetilde{r}_t^2, s_t)'$,

$$A_1 = \begin{pmatrix} \alpha_{11} & \rho_{11} \\ \alpha_{21} & \rho_{21} \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} \beta_{11} & \zeta_{11} \\ \beta_{21} & \zeta_{21} \end{pmatrix}. \tag{30}$$

If $|E(\ln w_t^2)| < \infty$, which is usually the case for the most commonly used densities in

[2]The EGARCH of Nelson (1991) requires exact ML methods for the algorithm to be applicable.

finance (*e.g.* the Student's t and the GED), then a stability condition of the system is that all the solutions of $|1_2 - (A_1 \circ D + B_1)c| = 0$ are greater than one in modulus, where

$$D = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \tag{31}$$

and $A_1 \circ D$ is the Hadamard product of A_1 and D. The estimation procedure that we propose is based on QML estimation of the VARMA representation. The motivation for this is to get the system into a form such that missing values can be handled with the algorithm outlined in Section 3.1. The details of the estimation procedure, together with simulations, are contained in the Appendix .

Conditioning variables or covariates ("X") can be added to either (27) or (28), or both, since the transformation to the VARMA representation needed for estimation is not affected by the additional variables. The model can also be extended with additional endogenous variables like (say) volume, volatility proxies, volatilities from other return series, and so on. Specifically, let $y_t = (\ln \sigma_t^2, h_t, \ln \sigma_{3t}^2, \ldots, \ln \sigma_{Mt}^2)'$ and x_t be the vectors of endogenous and exogenous variables, respectively, where the time index in x_t does not necessarily mean that all (or any) of the exogenous terms are contemporaneous. The $\sigma_{3t}^2, \ldots, \sigma_{Mt}^2$ are either return-volatilities or the conditional expectations of the square of a positively valued variable, *e.g.* volume. The first order version of the system (generalisation to higher orders is straightforward) can then be written as

$$y_t = \omega + A_1 v_{t-1} + B_1 y_{t-1} + C_1 x_t, \tag{32}$$

where $\omega = (\alpha_{10}, \rho_{10}, \alpha_{30}, \ldots, \alpha_{M0})'$, $v_t = (\ln \widetilde{r}_{1t}^2, s_t, \ln \widetilde{r}_{3t}^2, \ldots, \ln \widetilde{r}_{Mt}^2,)'$, and where A_1, B_1 and C are appropriately sized coefficient matrices. The additional variables $\widetilde{r}_{3t}, \ldots, \widetilde{r}_{Mt}$ either take on values in the whole real space (*e.g.* return) or are positively-valued (*e.g.* volume or duration). In the latter case the $\ln \widetilde{r}_t^2$ specification is a logarithmic Multiplicative Error Model (MEM), see Brownlees et al. (2012) for a survey of MEMs. The stability properties of the system can be investigated in the same way as earlier, and QML estimation and inference procedures are available via the VARMA-X representation.

4 Simulations

4.1 The effect on parameter and risk estimates

If the zero probability is time-varying, then the estimation theory of common volatility models is not valid. Here, we study how this affects parameter and risk estimates. In the simulations the Data Generating Process (DGP) of return is given by

$$r_t = \sigma_t I_t w_t \pi_{1t}^{-1/2}, \quad w_t \sim N(0,1), \quad t = 1, \ldots, n = 10000, \tag{33}$$

where the 0-DGP is governed by a deterministic trend equal to

$$\pi_{1t} = 1/(1 + \exp(-h_t)), \quad h_t = \rho_0 + \lambda t^*, \quad t^* = t/n. \tag{34}$$

The term $t^* = t/n$ is thus "relative" time with $t^* \in (0, 1]$. We use three parameter configurations for the 0-DGP: $(\rho_0, \lambda) = (\infty, 0)$, $(\rho_0, \lambda) = (0.1, 3)$ and $(\rho_0, \lambda) = (0.2, 3)$. These yield fractions of zeros over the sample equal to 0, 0.1 and 0.2, respectively. The DGPs of the GARCH and log-GARCH models, respectively, are given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \sigma_{t-1}^2, \tag{35}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \widetilde{r}_{t-1}^2 + \ln \sigma_{t-1}^2, \tag{36}$$

with $(\alpha_0, \alpha_1, \beta_1) = (0.02, 0.1, 0.8)$ in each. In both cases estimation proceeds by replacing $\widetilde{r}_t^2$ with r_t^2 in the recursions. For the log-GARCH, whenever $r_t^2 = 0$, its value is set to 1 (*i.e.* the specification of Francq et al. (2013), but without asymmetry). Estimation of the GARCH model is by Gaussian QML, whereas estimation of the log-GARCH is by Gaussian QML via the ARMA-representation, see Sucarrat et al. (2015).

The upper row of graphs in Figure 1 contains the average parameter biases, where the bias in replication i is computed as $\textit{average estimate}_i - \textit{true}_i$ (the no. of replications is 1000). The general tendency is clear: The higher the proportion of zeros, the greater the bias. Seemingly, this is not the case for α_0 in the GARCH model. However, this is simply due to the y-scale of the graph, since closer inspection reveals that there is indeed a (small) upwards bias. Generally, the magnitude of the bias is smaller for the GARCH. The exact bias in the log-GARCH, however, depends on the value used to replace zeros. So a different replacement-value for zeros may in fact result in a smaller bias than for the GARCH. The middle and lower rows of graphs in Figure 1 contain the Mean Percentage Errors (MPEs) of the risk estimates, computed as $100 \cdot \sum_{t=1}^{n}(x_t - 1)$ in replication i, where $x_t = \textit{estimate}_t/\textit{true}_t$. For example, the volatility percentage error at t is $100 \cdot (x_t - 1)$ with $x_t = \widehat{\sigma}_t/\sigma_t$. The VaR and ES graphs are for a risk level equal to 1%. Again, the graphs show a clear tendency: The higher the proportion of zeros, the greater the bias. Moreover, the bias is always positive.[3]

Additional simulations with fat-tailed w_t, other values on $\alpha_0, \alpha_1, \beta_1$ and risk level (for VaR and ES), and with a different 0-DGP, produce similar results. These simulations are not reported, but are available on request.

4.2 A missing values algorithm

I order to study the finite sample bias of the algorithm outlined in Section 3.2, we undertake a simulation study similar to that above. The DGP is exactly the same, but estimation proceeds differently. In the GARCH, whenever $\widetilde{r}_t^2$ is zero, then it is replaced by an estimate of its conditional expectation $E_{t-1}(\widetilde{r}_t^2)$ at that point (see the Appendix). Estimation of the log-GARCH proceeds similarly, except that now it is

[3] This is in line with previous studies on the effect of discreteness. Gottlieb and Kalay (1985) found that variance estimates of daily stock returns were biased upwards due to discreteness, but their analysis assumed the nature of the discreteness was constant, and that the conditional density of returns was normal. Cho and Frees (1988) also found that volatility was overestimated in most cases, using a measure of how quickly prices change. Li and Mykland (2014) find that Realised Volatility (RV) is biased and overestimates volatility in the presence of discreteness, and that the bias is increasing in the sampling frequency.

$\ln \widetilde{r}_t^2$ that is replaced by an estimate of $E(\ln \widetilde{r}_t^2)$ whenever r_t is zero.

Figure 2 contains the parameter biases for the GARCH(1,1) and log-GARCH(1,1) models, respectively. A solid blue line stands for the bias produced by the algorithm, whereas a dotted red line stands for the bias of ordinary Gaussian QML estimation *without* zero-adjustment. The Figure confirms that the algorithm provides approximately unbiased estimates in finite samples in the presence of missing values. Nominally, the biases produced by the ordinary method may appear small. However, as we will see in the empirical applications, such small nominal differences in the parameters can produces large differences in the dynamics.

Additional simulations with fat-tailed w_t, other values on α_0, α_1 and β_1, and with a different 0-DGP, produce similar results. These simulations are not reported, but are available on request.

5 Empirical application

In order to shed light on how returns with time-varying zero probabilities affect volatility dynamics, Value-at-Risk (VaR) and Expected Shortfall (ES) in practice, we revisit three of the return series in Sucarrat and Escribano (2013). These series are of interest, since they exhibit a variety of zero probability characteristics. The three series are the daily Standard and Poor's 500 stock market index (SP500) return, the daily Apple stock price return and the daily Ekornes stock price return. The first two return series are well-known, whereas the third is a leading Nordic furniture manufacturer listed on the Oslo Stock Exchange. Ekornes is a medium-sized company in international terms, since its market value is approximately 300 million euros (at the end of the series). Our interest in Ekornes is due to its relatively large – for daily returns – proportion of zeros over the sample (about 19%). The source of the data is Yahoo Finance (`http://finance.yahoo.com`). All three returns are computed as $(\ln S_t - \ln S_{t-1}) \cdot 100$, where S_t is the index level or stock price at day t. Saturdays and Sundays, where returns are usually 0, are not included in our sample. Descriptive statistics are contained in the upper part of Table 2. The statistics confirm that the returns exhibit the usual properties of excess kurtosis compared with the normal, and ARCH as measured by first order serial correlation 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 middle part of Table 2 contains estimates of three dynamic logit models for each return. The three models are:

$$\begin{aligned}
\text{Constant:} \quad h_t &= \rho_0, \\
\text{Trend:} \quad h_t &= \rho_0 + \lambda t^*, \qquad t^* = t/n, \qquad t^* \in (0,1], \\
\text{ACL(1,1):} \quad h_t &= \rho_0 + \rho_1 s_{t-1} + \zeta_1 h_{t-1}.
\end{aligned}$$

In the first model the zero probability is constant, in the second it is governed by a deterministic trend (t^* is "relative time") and in the third it is an ACL(1,1). For SP500 returns, it is the first logit specification that fits the data best according to the Schwarz (1978) information criterion (SIC). Accordingly, we use its fitted values of π_{1t} to compute the zero-adjusted returns $\widetilde{r}_t$. For Apple and Ekornes returns the best model according to SIC is the ACL(1,1).

The bottom part of Table 2 contains estimates of two GARCH(1,1) specifications for each return. These are

$$\begin{aligned} \text{Ordinary:} \qquad \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\ \text{0-adj:} \qquad \sigma_t^2 &= \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

Estimation of the Ordinary specification proceeds by Gaussian QML without adjustment of the observed returns r_t. Estimation of the second specification is also by Gaussian QML, but here the observed returns are replaced by estimates of the zero-adjusted ones, $\widetilde{r}_t$, treating zeros as missing values. For SP500, in which there are only 2 zeros, the two sets of estimates are virtually identical. For the two other returns, by contrast, the nominal differences vary between 0.003 and 0.007. These may appear small. However, as we will see shortly, these small nominal parameter differences – together with the different treatment of zeros – can lead to substantially different risk measure dynamics.

Figure 3 contains graphs of the fitted conditional zero probabilities $\widehat{\pi}_{0t}$ (upper row of graphs), and ratios of the conditional risk measures from the two estimation methods. The ratio of the fitted conditional standard deviations (second row of graphs) is computed as $\widehat{\sigma}_{t,\text{Ordinary}}/\widehat{\sigma}_{t,\text{0-adj}}$, *i.e.* the values from the Ordinary specification over those from the 0-adjusted specification. The VaR and ES ratios (third and fourth rows of graphs) are computed similarly. The first column of graphs are those of SP500. Unsurprisingly, since the SP500 return series only contain 2 zeros, and since the estimated parameters are virtually identical, the ratios are essentially equal to 1 throughout the sample. This is reflected in the Mean Percentage Errors (MPEs), computed as $n^{-1}\sum_{t=1}^{n}(x_t - 1)\cdot 100$, where x_t is the ratio in question. The second column of graphs are those of Apple. The fitted zero probability declines over the sample in a non-stationary fashion, and in the latter part it is essentially zero. This is reflected in the ratio between the conditional standard deviations. In the first part of the sample, until 1998 approximately, the conditional standard deviations of the Ordinary specification are about 3-5% higher, whereas in the latter part they are only about 1% higher (the MPE over the whole sample is 2.36%). Interestingly, there is not much difference between the specifications for the 1% VaR, and this holds throughout the sample. For the 1% ES, however, the values of the Ordinary specification start at about 8-10% higher. Then they fall steadily before stabilising towards the end at 3-4% lower than the values of the 0-adjusted specification. Finally, the third column of graphs are those of Ekornes. The evolution of the zero probabilities are very different from the others' in two ways. First, they are substantially higher throughout the sample. Second, in contrast to those of Apple, the dynamics appears stationary. This is reflected in the evolution of the ratios. The conditional standard deviations of the Ordinary method, for example, are – on average – 8.38% higher than those of the 0-adjustment method, and stably so throughout. Again, the 1% VaRs are, on average, very similar throughout the sample, and again the 1% ESs are different. However, this time the ESs are not only different, but substantially different, since the Ordinary specification yields a 1% ES that is on average 11.51% higher than that of the 0-adjusted specification – and this difference appears stationary throughout the sample.

All in all, the empirical comparison reveals that the biased parameter estimates of the Ordinary method can lead to substantially different values on two common risk measures, namely volatility and ES. Moreover, the extent and sign (i.e. whether it is higher or lower) of the magnitude depend on how big the zero probability is, and on the exact nature of the zero probability dynamics (*e.g.* whether it is stationary or not).

6 Conclusions

We propose a new class of financial return models that allows for a time-varying zero probability. A key feature of the new class is that standard volatility models (*e.g.* ARCH, SV and continuous time models) are nested and obtained as special cases when the zero probability is constant and equal to zero. Another attraction is that the properties of the new class (*e.g.* conditional volatility, skewness, kurtosis, Value-at-Risk, Expected Shortfall, etc.) are obtained as functions of the underlying volatility model. The new class allows for autoregressive conditional dynamics in both the zero probability and volatility specifications, and for a two-way feedback between the two. In the absence of a feedback effect from volatility to the zero probability specification, then estimation becomes particularly simple, since the model of zero probability and the model of volatility can then be estimated separately. Our empirical illustration shows that risk estimates can be substantially biased if the time-varying zero probability is not accommodated appropriately.

References

Ball, C. A. (1988). Estimation Bias Induced by Discrete Security Prices. *Journal of Finance 43*, 841–865.

Bauwens, L., C. Hafner, and S. Laurent (2012). *Handbook of Volatility Models and Their Applications*. New Jersey: Wiley.

Bien, K., I. Nolte, and W. Pohlmeier (2011). An inflated multivariate integer count hurdle model: an application to bid and ask quote dynamics. *Journal of Applied Econometrics 26*, 669–707.

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

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.

Cho, D. C. and E. W. Frees (1988). Estimating the Volatility of Discrete Stock Prices. *Journal of Finance 43*, 451–466.

Creal, D., S. J. Koopmans, and A. Lucas (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics 28*, 777–795.

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

Engle, R. F. and J. R. Russell (1998). Autoregressive Conditional Duration: A New Model of Irregularly Spaced Transaction Data. *Econometrica 66*, 1127–1162.

Francq, C. and G. Sucarrat (2015). An Equation-by-Equation Estimator of a Multivariate Log-GARCH-X Model of Financial Returns. https://mpra.ub.uni-muenchen.de/67140/.

Francq, C. and L. Q. Thieu (2015). Qml inference for volatility models with covariates. http://mpra.ub.uni-muenchen.de/63198/.

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.

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

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131*, 59–95.

Gottlieb, G. and A. Kalay (1985). Implications of the Discreteness of Observed Stock prices. *Journal of Finance 40*, 135–153.

Han, H. and D. Kristensen (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates. *Journal of Business and Economic Statistics*, ??–??

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

Hausman, J., A. Lo, and A. MacKinlay (1992). An ordered probit analysis of transaction stock prices. *Journal of financial economics 31*, 319–379.

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.

Li, Y. and P. A. Mykland (2014). Rounding Errors and Volatility Estimation. *Journal of Financial Econometrics*. Forthcoming. doi:10.1093/jjfinec/nbu005.

Liesenfeld, R., I. Nolte, and W. Pohlmeier (2006). Modelling Financial Transaction Price Movements: A Dynamic Integer Count Data Model. *Empirical Economics 30*, 795–825.

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.

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

Russell, J. R. and R. F. Engle (2005). A Discrete-State Continuous-Time Model of Financial Transaction Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model. *Journal of Business and Economic Statistics 23*, 166–180.

Rydberg, T. H. and N. Shephard (2003). Dynamics of Trade-by-Trade Price Movements: Decomposition and Models. *Journal of Financial Econometrics 1*, 2–25.

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

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

Sucarrat, G. and Á. Escribano (2013). Unbiased QML Estimation of Log-GARCH Models in the Presence of Zero Returns. MPRA Paper No. 50699. Online at http://mpra.ub.uni-muenchen.de/50699/.

Sucarrat, G., S. Grønneberg, and Á. Escribano (2013). Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown. MPRA Paper No. 49344. Online at http://mpra.ub.uni-muenchen.de/49344/.

Sucarrat, G., S. Grønneberg, and Á. Escribano (2015). Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown. *Computational Statistics and Data Analysis*. Forthcoming. DOI: http://dx.doi.org/10.1016/j.csda.2015.12.005. Working Paper version: http://mpra.ub.uni-muenchen.de/62352/.

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

A Derivation of properties 7 to 10

A.1 Pdfs and cdfs

Let $X_t = w_t I_t \pi_{1t}^{-1/2}$, and let $P_{t-1}(X_t \leq x_t)$ denote the cdf of X_t conditional on the past. Then

$$\begin{aligned} P_{t-1}(X_t \leq x_t) &= P_{t-1}(w_t I_t \pi_{1t}^{-1/2} \leq x_t) && (37)\\ &\overset{(a)}{=} P_{t-1}(w_t I_t \pi_{1t}^{-1/2} \leq x_t, I_t = 1) + P_{t-1}(w_t I_t \pi_{1t}^{-1/2} \leq x_t, I_t = 0) && (38)\\ &\overset{(b)}{=} P_{t-1}(w_t \pi_{1t}^{-1/2} \leq x_t, I_t = 1) + P_{t-1}(0 \leq x, I_t = 0) && (39)\\ &\overset{(c)}{=} P_{t-1}(w_t \leq x_t\sqrt{\pi_{1t}})P_{t-1}(I_t = 1) + P_{t-1}(0 \leq x)P_{t-1}(I_t = 0) && (40)\\ &\overset{(d)}{=} P_{t-1}(w_t \leq x_t\sqrt{\pi_{1t}})\pi_{1t} + 1_{0\leq x}\pi_{0t}, && (41) \end{aligned}$$

where we use: (a) $P(A) = P(A \cap B) + P(A \cap B^c)$, (b) $I_t = 1$ in $w_t I_t \pi_{1t}^{-1/2}$ in the first term and $I_t = 0$ in the second, (c) w_t and I_t are independent conditional on the past and (d) $P_{t-1}(0 \leq x)$ reduces to the indicator function for non-random events. Replacing X_t with z_t in the derivation gives (14), whereas replacing X_t with r_t and w_t with $\widetilde{r}_t$ gives (16). Next, the pdfs in (13) and (15) are obtained by differentiating the cdfs.

A.2 Quantiles

Let $X_t = w_t I_t \pi_{1t}^{-1/2}$ with cdf $F_X(x_t) = F_w(x_t\sqrt{\pi_{1t}})\pi_{1t} + 1_{\{x_t\geq 0\}}\pi_{0t}$. We wish to find the generalised inverse of F_X, given by $F_X^{-1}(c) = \inf\{x \in \mathbb{R} : F_X(x) \geq c\}$. Suppose first that $c < F_w(0)\pi_{1t}$, so that $x_t < 0$. Then

$$F_w(x_t\sqrt{\pi_{1t}})\pi_{1t} = c \quad \Leftrightarrow \quad x_t\sqrt{\pi_{1t}} = F_w^{-1}(c/\pi_{1t}) \quad \Leftrightarrow \quad x_t = \pi_{1t}^{-1/2}F_w^{-1}(c/\pi_{1t}). \qquad (42)$$

Suppose now that $c \geq F_w(0)\pi_{1t} + \pi_{0t}$. Then

$$F_w(x_t\sqrt{\pi_{1t}})\pi_{1t} + \pi_{0t} = c \quad \Leftrightarrow \quad F_w(x_t\sqrt{\pi_{1t}}) = (c - \pi_{0t})/\pi_{1t}, \qquad (43)$$

so that $x_t = \pi_{1t}^{-1/2}F_w^{-1}[(c - \pi_{0t})/\pi_{1t}]$. Finally, if $F_w(0)\pi_{1t} \leq c < F_w(0)\pi_{1t} + \pi_{0t}$, then $x_t = 0$ by the definition of the generalised inverse. Replacing X_t with z_t in the derivation gives (17), whereas replacing X_t with r_t and w_t with $\widetilde{r}_t$ gives (18).

A.3 Tail expectations

Let x_t denote a realisation of the random variable $X_t = w_t I_t \pi_{1t}^{-1/2}$ at t, so that $F_X(x_t) = P_{t-1}(X_t \leq x_t)$. If $0 < c < F_X(0)\pi_{1t}$, then the c-level (lower) tail-expectation of X_t conditional on the past is given by

$$\frac{1}{c}E_{t-1}\left(X_t 1_{\{X_t \leq F_X^{-1}(c)\}}\right) = \frac{1}{c}\int_A y_t \, dF_X(y_t), \qquad (44)$$

where $A = (-\infty, \pi_{1,t}^{-1/2} F_w^{-1}(c/\pi_{1,t}))$. Because $c < F_X(0)\pi_{1t}$, we have that $F_X^{-1}(c) < 0$, so that the area we integrate over only includes negative numbers. In this region $F_X(x_t) = \pi_{1t} F_w(x_t \sqrt{\pi_{1t}})$ with derivative $dF_X(x_t)/dx_t$ equal to $\pi_{1t}^{3/2} f_w(x_t \sqrt{\pi_{1t}})$. Hence,

$$E_{t-1}\left(X_t 1_{\{X_t \leq F_X^{-1}(c)\}}\right) = \pi_{1t}^{3/2} \int_A y_t f_w(y_t \sqrt{\pi_{1t}})\, dy_t. \tag{45}$$

Letting $u_t = y_t \sqrt{\pi_{1t}}$ gives $dy_t = du_t / \sqrt{\pi_{1t}}$, so that the area of integration is changed to $(-\infty, F_w^{-1}(c/\pi_{1t}))$. This gives

$$\begin{aligned} E_{t-1}\left(X_t 1_{\{X_t \leq F_X^{-1}(c)\}}\right) &= \pi_{1t} \int_{-\infty}^{F_w^{-1}(c/\pi_{1t})} u_t f_w(u_t)\, du_t & (46)\\ &= \pi_{1t} E_{t-1}\left(w_t 1_{\{w_t \leq F_w^{-1}(c/\pi_{1t})\}}\right), & (47) \end{aligned}$$

so that

$$\frac{1}{c} E_{t-1}\left(X_t 1_{\{X_t \leq F_X^{-1}(c)\}}\right) = \frac{\pi_{1t}}{c} E_{t-1}\left(w_t 1_{\{w \leq F_w^{-1}[c/\pi_{1t}]\}}\right). \tag{48}$$

Replacing X_t with z_t gives (20), whereas replacing X_t with r_t and w_t with $\widetilde{r}_t$ gives (21).

B Missing values estimation algorithm

In Section 3.2 we propose an estimation procedure where volatility models are estimated in a second step conditional on estimates of $\widetilde{r}_t$ from a first step, treating zeros as missing values. Here, we provide the details of how our missing values algorithm is implemented for the GARCH(1,1) and log-GARCH(1,1) models. Extensions of the algorithm to specifications of higher orders and/or with covariates is straightforward and self-explanatory.

Let $\widehat{\alpha}_0^{(k)}, \widehat{\alpha}_1^{(k)}$ and $\widehat{\beta}_1^{(k)}$ denote the parameter estimates of a GARCH(1,1) model after k iterations with some numerical method (*e.g.* Newton-Raphson). The initial values are at $k = 0$. If there are no zeros so that $r_t = \widetilde{r}_t$ for all t, then the kth. iteration of the numerical method proceeds in the usual way:

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

$$\widehat{\sigma}_t^2 = \widehat{\alpha}_0^{(k-1)} + \widehat{\alpha}_1^{(k-1)} \widetilde{r}_{t-1}^2 + \widehat{\beta}_1^{(k-1)} \widehat{\sigma}_{t-1}^2. \tag{49}$$

2. Compute the log-likelihood $\sum_{t=1}^n \ln f_{\widetilde{r}}(\widetilde{r}_t, \widehat{\sigma}_t)$ and other quantities (*e.g.* the gradient and/or Hessian) needed by the numerical method to generate $\widehat{\alpha}_0^{(k)}, \widehat{\alpha}_1^{(k)}$ and $\widehat{\beta}_1^{(k)}$.

Usually, $f_{\widetilde{r}}$ is the Gaussian density, so that the estimator may be interpreted as a Gaussian QML estimator. The algorithm we propose modifies the kth. iteration in several ways. Let G denote the set that contains the locations of the non-zeros, and let n^* denote the number of non-zero returns. The kth. iteration now proceeds as follows:

1. Compute, recursively, for $t = 1, \dots, n$:

$$\text{a) } \overline{r^2}_t = \begin{cases} \widetilde{r}_t^2 & \text{if } t \in G \\ \widehat{\sigma}_t^2 & \text{if } t \notin G, \text{ where } \widehat{\sigma}_t^2 = \widehat{\alpha}_0^{(k-1)} + \widehat{\alpha}_1^{(k-1)}\overline{r^2}_{t-1} + \widehat{\beta}_1^{(k-1)}\widehat{\sigma}_{t-1}^2, \end{cases} \tag{50}$$

$$\text{b) } \widehat{\sigma}_t^2 = \widehat{\alpha}_0^{(k-1)} + \widehat{\alpha}_1^{(k-1)}\overline{r^2}_{t-1} + \widehat{\beta}_1^{(k-1)}\widehat{\sigma}_{t-1}^2. \tag{51}$$

2. Compute the log-likelihood $\sum_{t \in G} \ln f_{\widetilde{r}}(\widetilde{r}_t, \widehat{\sigma}_t)$ and other quantities (*e.g.* the gradient and/or Hessian) needed by the numerical method to generate $\widehat{\alpha}_0^{(k)}, \widehat{\alpha}_1^{(k)}$ and $\widehat{\beta}_1^{(k)}$.

Step 1.a) means $\overline{r^2}_t$ is equal to an estimate of its conditional expectation at the locations of the zero-values. In Step 2 the symbolism $t \in G$ means the log-likelihood only includes contributions from non-zero locations. A practical implication of this is that any likelihood comparison (*e.g.* via information criteria) with other models should be in terms of the average log-likelihood, *i.e.* division by n^* rather than n.

QML Estimation of the log-GARCH model is via its ARMA-representation, since the standard Gaussian ML estimator must be interpreted as *exact* ML in the presence of missing values, see Sucarrat and Escribano (2013). If $|E(\ln w_t^2)| < \infty$, then the ARMA(1,1) representation exists and is given by

$$\ln \widetilde{r}_t^2 = \phi_0 + \phi_1 \ln \widetilde{r}_{t-1}^2 + \theta_1 u_{t-1} + u_t, \quad u_t = \ln w_t^2 - E(\ln w_t^2), \tag{52}$$

where $\phi_0 = \alpha_0 + (1-\beta_1)E(\ln w_t^2)$, $\phi_1 = \alpha_1 + \beta_1$, $\theta_1 = -\beta_1$ and u_t is zero-mean and IID conditionally on the past. Accordingly, the usual ARMA-methods can be used to estimate ϕ_0, ϕ_1 and θ_1, and hence the log-GARCH parameters α_1 and β_1. To identify α_0, however, an estimate of $E(\ln w_t^2)$ is needed. Sucarrat et al. (2015) show that, under very general assumptions, the formula $-\ln\left[n^{-1}\sum_{t=1}^n \exp(\widehat{u}_t)\right]$ provides a consistent estimate (see also Francq and Sucarrat (2015)). To accommodate the missing values, this formula is modified to $-\ln\left[n^{*-1}\sum_{t \in G} \exp(\widehat{u}_t)\right]$.

C Estimation of the joint model of π_{1t} and σ_t with two-way feedback

For notational economy we illustrate the first order case only, since the extension to higher order specifications is straightforward. The procedure starts by casting the $\ln \sigma_t^2$ equation into its "ARMA-X" form and then changing the h_t equation accordingly. Assuming that $|E(\ln w_t^2)| < \infty$ and denoting $\tilde{y}_t = (\ln \widetilde{r}_t^2, h_t)'$, $\tau = E(\ln w_t^2)$, $u_t = (\ln w_t^2 - \tau, s_t)'$ and $u_t^* = (\ln w_t^2 - \tau, 0)'$, the VARMA representation is given by

$$\begin{aligned} \ln \widetilde{r}_t^2 &= \alpha_0^* + \phi_1 \ln \widetilde{r}_{t-1}^2 + \rho_{11} s_{t-1} + \beta_{11}^* u_{t-1} + \zeta_{11} h_{t-1} + u_t, & (53) \\ h_t &= \rho_0^* + \phi_2 \ln \widetilde{r}_{t-1}^2 + \rho_{21} s_{t-1} + \beta_{21}^* u_{t-1} + \zeta_{21} h_{t-1}, & (54) \end{aligned}$$

where $\alpha_0^* = \alpha_0 + (1-\beta_{11})E(\ln w_t^2)$, $\rho_0^* = \rho_0 - \beta_{21}E(\ln w_t^2)$, $\phi_1 = (\alpha_{11} + \beta_{11})$, $\phi_2 = (\alpha_{21} + \beta_{21})$, $\beta_{11}^* = -\beta_{11}$, $\beta_{21}^* = -\beta_{21}$, and $u_t = \ln w_t^2 - E(\ln w_t^2)$ with u_t conditional on the past being $IID(0, \sigma_u^2)$. The stability condition for this representation is identical to the untransformed model, *i.e.* the log-GARCH-ACL specification, but the unconditional expectations are now given by

$$E(\ln \widetilde{r}_t^2) = \frac{(1-\zeta_{21})\alpha_0^* + \zeta_{11}\rho_0^*}{(1-\phi_1)(1-\zeta_{21}) - \phi_2\zeta_{11}}, \quad (55)$$

$$E(h_t) = \frac{\phi_2\alpha_0^* + (1-\phi_1)\rho_0^*}{(1-\phi_1)(1-\zeta_{21}) - \phi_2\zeta_{11}}. \quad (56)$$

More compactly, the VARMA representation can be written as

$$\tilde{y}_t = \phi_0 + \phi_1\tilde{y}_{t-1} + \theta_1 u_{t-1} + u_t^*, \quad (57)$$

where $\phi_0 = \omega + (1_2 \circ D - B_1 \circ D)\tau$, $\phi_1 = (A_1 \circ D + B_1)$ and

$$\theta_1 = \begin{pmatrix} -\beta_{11} & \rho_{11} \\ -\beta_{21} & \rho_{21} \end{pmatrix}. \quad (58)$$

To estimate the VARMA parameters ϕ_0, ϕ_1 and θ_1 we replace the density $f_{\widetilde{r}}$ in the log-likelihood at t, which is given by (see the beginning of Section 3)

$$\ln f_r(r_t) = I_t \ln f_{\widetilde{r}}(\widetilde{r}_t) + I_t \ln \pi_{1t} + (1 - I_t)\ln(1 - \pi_{1t}), \quad (59)$$

with an instrumental QML density (*e.g.* the normal) $f_{u_{1t}}$ in the IID error u_{1t}. The joint log-likelihood at t conditional on the past thus becomes

$$I_t \ln f_{u_{1t}}(u_t) + I_t \ln \pi_{1t} + (1 - I_t)\ln(1 - \pi_{1t}). \quad (60)$$

Maximisation of the total log-likelihood provides estimates of the VARMA parameters. All the parameters of interest, apart from α_0 and ρ_0, can be identified from the VARMA estimates. In order to identify α_0 and ρ_0 an estimate of τ is needed, since the VARMA intercepts are given by $\phi_{10} = \alpha_0 + (1 - \beta_{11})\tau$ and $\phi_{20} = \rho_0 - \beta_{21}\tau$, respectively. To this end we use the same formula as in the univariate case to estimate τ, *i.e.* $-\ln\left[n^{*-1}\sum_{t\in G}\exp(\widehat{u}_{1t})\right]$, where $\widehat{u}_{1t}$ are the residuals from the first equation in the estimated VARMA model. Next, from the definition of ϕ_0 we can solve for α_0 and ρ_0, respectively, in order to obtain plug-in estimators of α_{10} and ρ_0. Table 1 provides a small Monte Carlo study of this estimation procedure. The simulations show that estimates are close to their population counterparts in finite samples for different densities of w_t, even when as much as 38% of the returns r_t are zero.

Table 1: QML estimation of the log-GARCH-ACL model via the VARMA representation

	DGP	T	$\widehat{\alpha}_0$	$\widehat{\alpha}_{11}$	$\widehat{\rho}_{11}$	$\widehat{\beta}_{11}$	$\widehat{\zeta}_{11}$	$\widehat{\rho}_0$	$\widehat{\alpha}_{21}$	$\widehat{\rho}_{21}$	$\widehat{\beta}_{21}$	$\widehat{\zeta}_{21}$	$\widehat{E}(\ln w_t^2)$	$\widehat{E}(\pi_{0t})$
$w_t \sim N(0,1)$:	A	5000	0.04	0.301	0.053	0.082	0.059	0.152	0.053	0.094	0.051	0.949	-1.318	0.375
		10000	0.05	0.302	0.054	0.094	0.053	0.152	0.055	0.094	0.042	0.950	-1.321	0.376
	B	5000	0.02	0.101	0.006	0.796	-0.006	0.189	0.001	0.099	-0.003	0.937	-1.272	0.050
		10000	0.00	0.101	0.001	0.797	-0.002	0.158	0.001	0.100	-0.004	0.947	-1.273	0.050
	C	5000	0.00	0.101	0.055	0.794	0.052	0.214	0.051	0.101	0.006	0.945	-1.277	0.076
		10000	0.00	0.101	0.053	0.795	0.051	0.209	0.052	0.105	0.001	0.948	-1.275	0.074
$w_t \sim t(10)$:	A	5000	0.06	0.302	0.062	0.089	0.053	0.148	0.050	0.096	0.055	0.950	-1.457	0.440
		10000	0.06	0.303	0.059	0.086	0.051	0.148	0.050	0.100	0.054	0.949	-1.455	0.440
	B	5000	0.00	0.101	0.000	0.794	-0.001	0.210	0.001	0.104	-0.003	0.930	-1.391	0.049
		10000	0.00	0.101	0.000	0.796	0.001	0.168	0.000	0.101	0.002	0.945	-1.394	0.050
	C	5000	0.00	0.104	0.055	0.791	0.052	0.215	0.051	0.105	0.005	0.945	-1.397	0.112
		10000	0.01	0.103	0.056	0.797	0.048	0.209	0.051	0.107	0.004	0.948	-1.398	0.112

The table contains the average parameter estimates from 100 simulations. All computations in the programming language R (R Core Team (2014)) using the same initial values across DGPs. $w_t \sim N(0,1)$, w_t is standard normal with $E(\ln w_t^2) = -1.27$. $w_t \sim t(10)$, w_t is standardised t with 10 degrees of freedom and $E(\ln w_t^2) = 1.39$. T, sample size. DGP A, $(\alpha_0, \alpha_{11}, \rho_{11}, \beta_{11}, \zeta_{11}) = (0, 0.3, 0.05, 0.1, 0.05)$ and $(\rho_0, \alpha_{21}, \rho_{21}, \beta_{21}, \zeta_{21}) = (0.15, 0.05, 0.05, 0.05, 0.95)$. DGP B, $(\alpha_0, \alpha_{11}, \rho_{11}, \beta_{11}, \zeta_{11}) = (0, 0.1, 0, 0.8, 0)$ and $(\rho_0, \alpha_{21}, \rho_{21}, \beta_{21}, \zeta_{21}) = (0.15, 0, 0.1, 0, 0.95)$. DGP C, $(\alpha_0, \alpha_{11}, \rho_{11}, \beta_{11}, \zeta_{11}) = (0, 0.1, 0.05, 0.8, 0.05)$ and $(\rho_0, \alpha_{21}, \rho_{21}, \beta_{21}, \zeta_{21}) = (0.2, 0.05, 0.1, 0, 0.95)$. $E(\pi_{0t})$, unconditional zero probability (estimated by the fraction of zeros, *i.e.* $1 - T^{-1}\sum_{t=1}^{T} I_t$). All computations in R (R Core Team (2014)).

Table 2: Descriptive statistics, dynamic logit models and GARCH-models of SP500, Apple and Ekornes returns (see Section 5)

Descriptive statistics:

	s^2	s^4	ARCH $[p-val]$	n	$0s$	$\widehat{\pi}_0$
SP500	1.73	10.30	143.10 [0.00]	3684	2	0.001
Apple	9.25	55.03	7.12 [0.01]	7303	294	0.040
Ekornes	5.70	10.32	54.01 [0.00]	3546	667	0.189

Dynamic logit-models:

		$\widehat{\rho}_0$ $(s.e.)$	$\widehat{\rho}_1$ $(s.e.)$	$\widehat{\zeta}_1$ $(s.e.)$	$\widehat{\lambda}$ $(s.e.)$	SIC	$Logl$
SP500	Constant	7.158 (0.707)				0.0115	−17.04
	Trend	7.200 (1.309)			0.673 (2.478)	0.0137	−17.00
	ACL(1,1)	0.032 ($4e-05$)	−1.147 ($4e-05$)	0.997 ($4e-05$)		0.0116	−9.022
Apple	Constant	3.171 (0.060)				0.3387	−1232.5
	Trend	1.870 (0.094)			3.437 (0.263)	0.3102	−1123.9
	ACL(1,1)	$1e-09$ ($4e-05$)	0.024 (0.011)	0.999 ($9e-05$)		0.3095	−1116.9
Ekornes	Constant	1.462 (0.043)				0.9692	−1714.3
	Trend	1.183 (0.083)			0.576 (0.150)	0.9673	−1706.9
	ACL(1,1)	0.445 (0.130)	0.207 (0.036)	0.701 (0.087)		0.9599	−1689.6

GARCH-models:

		$\widehat{\alpha}_0$ $(s.e.)$	$\widehat{\alpha}_1$ $(s.e.)$	$\widehat{\beta}_1$ $(s.e.)$
SP500	Ordinary	0.015 (0.003)	0.083 (0.008)	0.908 (0.009)
	0-adj.	0.015 (0.003)	0.083 (0.008)	0.908 (0.009)
Apple	Ordinary	0.168 (0.033)	0.087 (0.008)	0.901 (0.010)
	0-adj.	0.175 (0.037)	0.093 (0.010)	0.894 (0.012)
Ekornes	Ordinary	0.036 (0.009)	0.019 (0.002)	0.974 (0.004)
	0-adj.	0.039 (0.011)	0.025 (0.004)	0.968 (0.005)

s^2, sample variance. s^4, sample kurtosis. ARCH, Ljung and Box (1979) test statistic of first-order serial correlation in the squared return. $p - val$, the p-value of the test-statistic. n, number of returns. $0s$, number of zero returns. $\widehat{\pi}_0$, proportion of zero returns. $s.e.$, approximate standard errors (obtained via the numerically estimated Hessian). k, the number of estimated model coefficients. $LogL$, log-likelihood. SIC, the Schwarz (1978) information criterion. All computations in R (R Core Team (2014)).

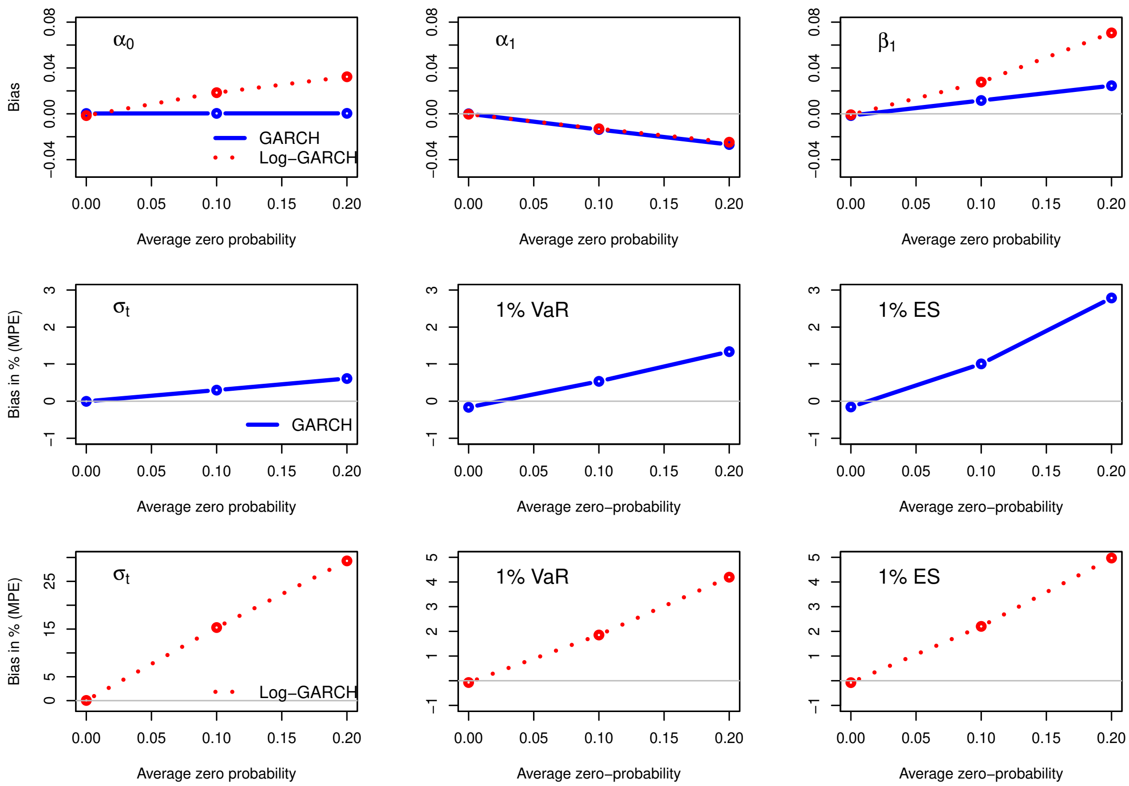

Figure 1: Simulated parameter and risk estimation biases in GARCH(1,1) and log-GARCH(1,1) models (see Section 4)

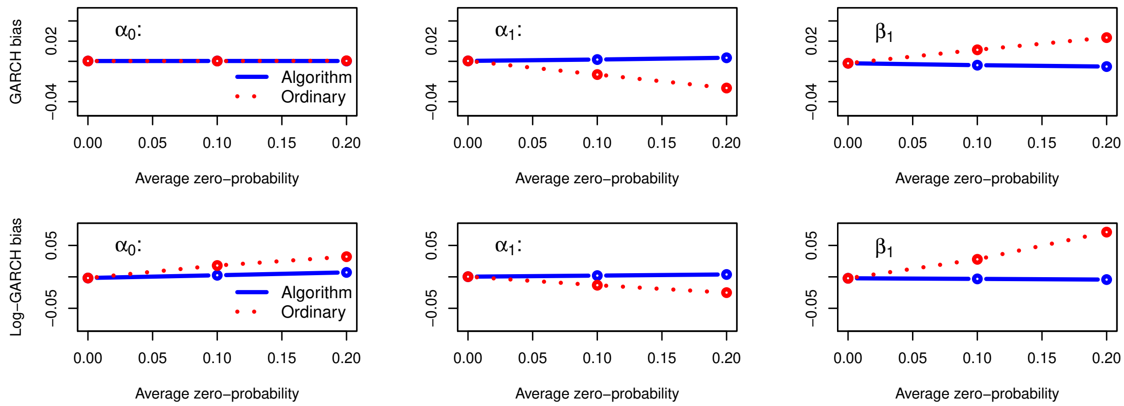

Figure 2: Simulated parameter biases in GARCH(1,1) and log-GARCH(1,1) models for the missing values algorithm in comparison with ordinary methods (see Section 4.2)

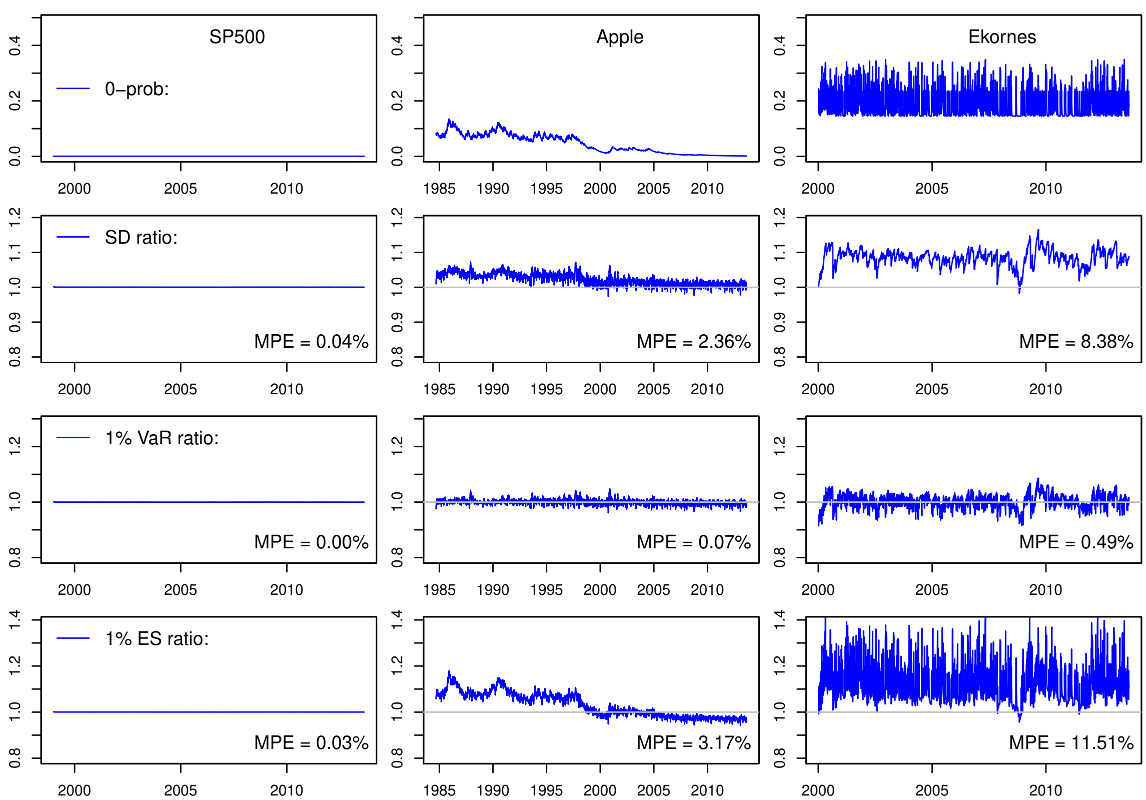

Figure 3: Fitted 0-probabilities, and the ratios of fitted σ_t, 1% VaR and 1% ES (see Section 5).