Risk Estimation with a Time Varying Probability of Zero Returns[*]

Genaro Sucarrat[1] and Steffen Grønneberg[1]

[1]*BI Norwegian Business School*

14 February 2020

Abstract:

The probability of an observed financial return being equal to zero is not necessarily zero, or constant. In ordinary models of financial return, however, e.g. ARCH, SV, GAS and continuous-time models, the zero-probability is zero, constant or both, thus frequently resulting in biased risk estimates (volatility, Value-at-Risk, Expected Shortfall, etc.). We propose a new class of models that allows for a time varying zero-probability that can either be stationary or non-stationary. The new class is the natural generalisation of ordinary models of financial return, so ordinary models are nested and obtained as special cases. The main properties (e.g. volatility, skewness, kurtosis, Value-at-Risk, Expected Shortfall) of the new model class are derived as functions of the assumed volatility and zero-probability specifications, and estimation methods are proposed and illustrated. In a comprehensive study of the stocks at New York Stock Exchange (NYSE) we find extensive evidence of time varying zero-probabilities in daily returns, and an out-of-sample experiment shows that corrected risk estimates can provide significantly better forecasts in a large number of instances.

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 liquidity issues (e.g. low trading volume), market closures, data issues (e.g. data imputation due to missing values), price discreteness and/or rounding error, characteristics specific to the market, and so on. Moreover, the zero-probability may change and depend on market conditions. In ordinary models of financial risk, however, the probability of a zero return is usually 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)), the Generalised Autoregressive Score (GAS) or Dynamic Conditional Score (DCS) model proposed by Creal et al. (2010) and Harvey (2013), respectively, and continuous-time models (e.g. Brownian motion).* A time varying zero-probability will generally lead to biased risk estimates in all of these model classes.

Several contributions relax the constancy assumption by specifying return as a discrete dynamic process. Hausman et al. (1992), for example, allow the zero-probability to depend on other conditioning variables (e.g. volume, duration and past returns) in a probit framework. This was extended in two different directions by Engle and Russell (1998), and Russell and Engle (2005), respectively. In the former, the durations between price increments are specified in terms of an Autoregressive Conditional Duration (ACD) model, whereas in the latter price-changes are specified in terms of an Autoregressive Conditional Multinomial (ACM) model in combination with an ACD model of the durations between trades. Liesenfeld et al. (2006) point to 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). Rydberg and Shephard (2003) propose a framework in which the price increment is decomposed multiplicatively into three components: Activity, direction and integer magnitude. Finally, Kümm and Küsters (2015) propose a zero-inflated model for German milk-based commodity returns with autoregressive persistence, where zeros occur either because there is no information available (i.e. a binary variable), or because of rounding.

Even though discrete models in many cases provide a more accurate characterisation of observed returns, the most common models used in risk analysis in empirical practice are continuous. Examples include ARCH, SV, GAS/DCS and continuous-time models. Arguably, the discreteness point that causes the biggest problem for continuous models is located at zero. This is because zero is usually 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 non-zero and/or time varying zero-probability may thus severely invalidate the parameter and risk estimates of continuous models, in particular if the zero process is non-stationary. 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 returns multiplicatively into a continuous part and a discrete part at zero that is appropriately scaled by the conditional zero-probability. The zero and volatility processes can be mutually dependent, and standard volatility models (e.g. ARCH, SV and continuous-time models) are nested and obtained as special cases when the conditional zero-probability is constant and equal to zero. Hautsch et al. (2013) propose a model for volume which uses a decomposition that is similar to ours. In their model the dynamics is governed by a logarithmic Multiplicative Error Model (MEM) with a Generalised F as conditional density, see Brownlees et al. (2012) for a survey of MEMs. Our model is much more general and nests the specification of Hautsch et al. (2013) as a special case: The dynamics need not be specified in logs,

---

*See Bauwens et al. (2012) for a survey of these models.

the density of the continuous part (squared) need not be Generalised F, our framework also applies to return models (not only MEMs), and the model class is not restricted to ARCH-type models. Another attraction of our model is that many return properties (e.g. conditional volatility, return skewness, Value-at-Risk and Expected Shortfall) are 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. Finally, a recent strand of the continuous-time literature introduces the idea of "stale" price increments, see e.g. Bandi et al. (2017), and Bandi et al. (2018). This can be viewed as a continuous-time analogue of our discrete-time framework.

Our results shed light on the effect and bias caused by zeros in several ways. First, for a given volatility level, our results imply that a higher zero-probability increases both the conditional skewness and conditional kurtosis of return, but reduces return variability when defined as conditional absolute return (see Proposition 2.1). Second, we derive general formulas for Value-at-Risk (VaR). They show that the bias induced by not correcting for zeros depends, in nonlinear ways, on the volatility bias caused by the misspecified model and/or estimator, and on the exact shape of the conditional density. In other words, whether the estimated risk is too low or too high will depend on a variety of factors that would vary from application to application. Nevertheless, for a given level of volatility, our results show that risk – when defined as VaR – will be biased downwards for rare loss events (5% or less) if zeros are not corrected for (see Section 2.3). Third, we derive general formulas for Expected Shortfall (ES). Since the formulas depend on the value of the VaR, also here the bias depends, in nonlinear ways, on the volatility bias caused by the misspecified model and/or estimator, and on the exact shape of the conditional density. Notwithstanding, for a given level of volatility, our results show that risk – when defined as ES – will be biased downwards (i.e. just as for VaR) for rare loss events (10% or less) if zeros are not corrected for (see Section 2.4). Fourth, since the models and/or estimators that are commonly used by practitioners can lead to severely biased risk estimates – in particular if the zero-probability is non-stationary, we outline an estimation and inference procedure that reduces the bias caused by a time varying zero-probability, and which can be combined with well-known models and estimators (see Section 2.5). Section 3 contains a detailed illustration of our results and methods applied to the daily returns of three stocks at the New York Stock Exchange (NYSE). The stocks have been carefully selected to illustrate three different types of zero-probability dynamics. Finally, in a comprehensive study of the stocks at the NYSE (see Section 4) we find that 24.4% of the daily returns we study are characterised by a time varying zero-probability. The actual proportion is likely to be higher, since the stocks we omit from our analysis – stocks with less than thousand observations in the in-sample – are likely to be characterised by a high zero-probability, and therefore also of a time varying zero-probability. Next, an out-of-sample experiment shows that corrected risk estimates can provide significantly better forecasts in a large number of instances.

The rest of the paper is organised as follows. Section 2 presents the new model class and derives some general properties, and the formulas for zero-corrected VaR and ES. The section ends by outlining situations

where volatility estimates are not biased even though the zero-probability is time varying (and stationary), and by outlining a general estimation and inference procedure that reduces the volatility bias caused by zeros when the zero-probability is non-stationary. A main attraction with the procedure is that it can be combined with common models and methods. Section 3 contains the detailed illustration of the results and methods of Section 2. Section 4 contains a comprehensive study of stocks at the NYSE, whereas Section 5 concludes. The Appendix contains the proofs and additional auxiliary material.

2   Financial return with time varying zero-probability

2.1   The ordinary model of return

The ordinary model of a financial return r_t is given by

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

where $\sigma_t > 0$ is a time varying scale or volatility (that does not need to equal the conditional standard deviation). The subscript $t-1$ is notational shorthand for conditioning on the past. Unless we state otherwise, the past will be the sigma-field generated by $\{r_u : u < t\}$, and when needed we will denote this sigma-field by $\mathcal{F}_{t-1}^r$. The w_t is an innovation and $P_{t-1}(w_t = 0)$ is the zero-probability of w_t conditional on the past. We refer to (1) as an "ordinary" model of return, since the zero-probability of return r_t is 0 for all t. An example of an ordinary model is the GARCH(1,1) of Bollerslev (1986), where

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

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

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

with v_i being independent of w_j for all pairs i, j. Other examples of σ_t include quadratic variation and other continuous-time notions of volatility, the Gaussian log-GARCH models proposed independently by Geweke (1986), Pantula (1986) and Milhøj (1987), the EGARCH model of Nelson (1991) with $w \sim GED$ (where GED stands for Generalised Error Distribution), 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 A model of return with time varying zero-probability

Let r_t denote a return governed by

$$r_t = \sigma_t z_t, \qquad \sigma_t > 0, \qquad t \in \mathbb{Z}, \tag{4}$$

$$z_t = w_t I_t \pi_{1t}^{-1/2}, \quad E_{t-1}(w_t) = 0, \quad E_{t-1}(w_t^2) = \sigma_w^2, \quad P_{t-1}(w_t = 0) = 0, \tag{5}$$

$$I_t \in \{0, 1\}, \qquad \pi_{1t} = P_{t-1}(I_t = 1), \qquad 0 < \pi_{1t} \le 1. \tag{6}$$

Again, the subscript $t - 1$ is shorthand notation for conditioning on the past, and again the past is given by the sigma-field generated by past returns, i.e. $\mathcal{F}_{t-1}^r$. The indicator variable I_t determines whether return r_t is zero or not: $r_t \ne 0$ if $I_t = 1$, and $r_t = 0$ if $I_t = 0$. This follows from $P_{t-1}(w_t = 0) = 0$, which is an assumption that is needed for identification (it ensures zeros do not originate from both w_t and I_t). The probability of a zero return conditional on the past is thus $\pi_{0t} = 1 - \pi_{1t}$. The motivation for letting π_{1t} enter the way it does in z_t is to ensure that $Var_{t-1}(z) = \sigma_w^2$ (see Proposition 2.1 below). In particular, if $\sigma_w^2 = 1$, then we can interpret σ_t and σ_t^2 as the conditional standard deviation and variance, respectively. Note that (4) – (6) do not exclude the possibility of I_t being contemporaneously dependent on the value of w_t, e.g. that small values of $|w_t|$ increase the probability of I_t being zero. A specific example is the situation where w_t conditional on the past is standard normal, and $I_t = 1$ if $|w_t| > 0.05$ and 0 otherwise (so that $\pi_{1t} = 0.96$ for all t). Note also that (4) – (6) do not exclude the possibility of σ_t being contemporaneously dependent on w_t or I_t, or both. Finally, we will refer to $\widetilde{r}_t = \sigma_t w_t$ as "zero-adjusted" or "zero-corrected" return, since $\widetilde{r}_t = r_t \pi_{1t}^{1/2}$ whenever $I_t \ne 0$.

An attractive feature of (4)–(6) is that many properties can be expressed as a function of the underlying models of volatility and the zero-probability. In deriving these properties we rely on suitable subsets of the following assumptions.

Assumption 1 (regularity of distribution) *Conditional on the past $\mathcal{F}_{t-1}^r$:*

(a) The joint probability distribution of w_t and I_t is regular.

(b) The joint probability distribution of $\widetilde{r}_t$ and I_t is regular.

Assumption 2 (identification) *For all t: $E_{t-1}(w_t|I_t = 1) = 0$ and $E_{t-1}(w_t^2|I_t = 1) = \sigma_w^2$ with $0 < \sigma_w^2 < \infty$.*

Assumption 1 is a technical condition ensuring that probabilities conditional on the past can be manipulated as usual, see Shiryaev (1996, pp. 226-227). In what follows, (a) will usually be needed when deriving properties involving z_t, whereas (b) will usually be needed when deriving properties involving r_t. Assumption 2 states that, conditional on both $\mathcal{F}_{t-1}^r$ and $I_t = 1$, the expectation of w_t is zero, and the expectation of w_t^2 exists and is equal to σ_w^2 for all t. The motivation behind this assumption is to ensure that z_t exhibits the first and second moment properties that are typically possessed by the scaled innovation in volatility models. In particular, if $\sigma_w^2 = 1$ (as in the ARCH class of models), then σ_t and σ_t^2 will usually correspond to the conditional standard

deviation and variance, respectively. The assumption can thus be viewed as an identification condition. The conditional zero-mean property will usually ensure that returns are Martingale Difference Sequences (MDSs). It should be noted, however, that Assumption 2 is used only once in the proofs of our results, namely in the proof of Proposition 2.1. In other words, Assumption 2 is not required for the other propositions. Proposition 2.1 collects some properties of z_t that follow straightforwardly.

Proposition 2.1 (properties of z_t) *Suppose* (4) – (6), *Assumption 1(a) and Assumption 2 hold. Then:*

(i) *If $E_{t-1}|z_t| < \infty$ for all t, then $\{z_t\}$ is a Martingale Difference Sequence (MDS).*

(ii) *If $E_{t-1}|z_t^2| < \infty$ for all t, then $Var_{t-1}(z_t) = \sigma_w^2$ for all t, and $\{z_t\}$ is covariance-stationary with $E(z_t) = 0$, $Var(z_t) = \sigma_w^2$ and $Cov(z_t, z_{t-j}) = 0$ when $j \neq 0$.*

(iii) *If $E_{t-1}|z_t^s| < \infty$ for some $s \geq 0$, then $E_{t-1}(z_t^s) = \pi_{1t}^{(2-s)/2} E_{t-1}(w_t^s | I_t = 1)$.*

(iv) *If $E_{t-1}|z_t^s| < \infty$ for some $s \geq 0$, then $E_{t-1}|z_t|^s = \pi_{1t}^{(2-s)/2} E_{t-1}(|w_t|^s | I_t = 1)$.*

Proof: See Appendix A.1.

Property (i) means that $\{z_t\}$ is a MDS even if π_{1t} is time varying. Indeed, it remains a MDS even if $\{I_t\}$ is non-stationary. Usually, (i) will imply that $\{r_t\}$ is also a MDS, e.g. in the ARCH class of models, since there $E_{t-1}(r_t) = \sigma_t E_{t-1}(z_t)$, see Assumption 4 and Proposition 2.4. Property (ii) means that σ_t^2 corresponds to the conditional variance in ARCH models, and that the unconditional second moment – if it exists – is not affected by the presence of time varying zero-probability. For example, in the semi-strong GARCH(1,1) of Lee and Hansen (1994), where z_t is strictly stationary and ergodic with $\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, we have $Var_{t-1}(r_t) = \sigma_t^2$ and $Var(r_t) = \alpha_0/(1 - \alpha_1 - \beta_1)$ regardless of whether π_{1t} is constant or time varying. Also, if the zero-probability is periodic (as is common in intraday returns) or downwards trending (as in some daily returns) so that I_t is non-stationary, then Property (ii) means that z_t will still be covariance stationary even though I_t and z_t are not stationary. The implications of I_t being non-stationary is discussed in Section 2.5. Property (iii) means higher order (i.e. $s > 2$) conditional moments (in absolute value) are scaled upwards by positive zero-probabilities, whereas the opposite is the case for lower order (i.e. $s < 2$) conditional moments. In particular, both conditional skewness ($s = 3$) and conditional kurtosis ($s = 4$) become more pronounced.[†] Similarly, property (iv) means that higher order (i.e. $s > 2$) conditional absolute moments are scaled upwards by positive zero-probabilities, whereas the opposite is the case for lower order (i.e. $s < 2$) conditional moments. In particular, for a given volatility level σ_t the conditional absolute return (i.e. $s = 1$) is scaled downwards.

[†]Whether this implies that higher order conditional moments of return r_t become more pronounced or not depends on the specification of σ_t and π_{1t}, and on the nature of their inter-dependence.

2.3 Value-at-Risk (VaR)

For notational simplicity we will, henceforth, denote the cumulative distribution function (cdf) of a random variable X_t conditional on $\mathcal{F}^r_{t-1}$ as $F_{X_t}(x)$, hence omitting the subscript $t-1$. Conditional on both $\mathcal{F}^r_{t-1}$ and $I_t = 1$, we will use the notation $F_{X_t|1}(x)$.

Proposition 2.2 (cdfs of z_t and r_t) *Suppose* (4) – (6) *hold, and let* $1_{\{x \geq 0\}}$ *denote an indicator function equal to 1 if* $x \geq 0$ *and 0 otherwise:*

(i) If also Assumption 1(a) holds, then the cdf of z_t at t conditional on $\mathcal{F}^r_{t-1}$ is

$$F_{z_t}(x) \;\; = \;\; F_{w_t|1}(x\pi_{1t}^{1/2})\pi_{1t} + 1_{\{x \geq 0\}}(1 - \pi_{1t}). \tag{7}$$

(ii) If also Assumption 1(b) holds, then the cdf of r_t at t conditional on $\mathcal{F}^r_{t-1}$ is

$$F_{r_t}(x) \;\; = \;\; F_{\widetilde{r}_t|1}(x\pi_{1t}^{1/2})\pi_{1t} + 1_{\{x \geq 0\}}(1 - \pi_{1t}). \tag{8}$$

Proof: See Appendix A.2.

Natural examples of $F_{w_t|1}$ and $F_{\widetilde{r}_t|1}$ are, respectively, $N(0,1)$ and $N(0, \sigma_t^2)$.

If $F_{X_t}(x)$ denotes the cdf of a random variable X_t conditional on the past $\mathcal{F}^r_{t-1}$, then its lower c-quantile with $c \in (0,1)$ is given by

$$X_{c,t} = \inf\{x \in \mathbb{R} : F_{X_t}(x) \geq c\}. \tag{9}$$

We will write $F^{-1}_{X_t}(c) = X_{c,t}$ even though the inverse of F_X does not exist, and we will refer to $F^{-1}_{X_t}(c)$ as the generalised inverse of $F_{X_t}(x)$, see e.g. Embrechts and Hofert (2013). In order to derive general formulas for quantiles and VaRs, we introduce an additional, technical assumption on the distributions of w_t and $\widetilde{r}_t$. The assumption can be relaxed, but at the cost of more complicated formulas.

Assumption 3 *Conditional on the past $\mathcal{F}^r_{t-1}$ and $I_t = 1$:*

(a) The cdf of w_t, denoted $F_{w_t|1}$, is strictly increasing.

(b) The cdf of $\widetilde{r}_t$, denoted $F_{\widetilde{r}_t|1}$, is strictly increasing.

The assumption is fairly mild, since it holds for most of the conditional densities that have been used in the literature, including the standard normal, the Student's t and the GED, and also for many skewed versions. In particular, the assumption does not require smoothness or continuity. A consequence of (a) and (b) is that F_{z_t} and F_{r_t} are both increasing. Accordingly, their lower and upper c-quantiles – as defined in Acerbi and Tasche (2002, Definition 2.1, p. 1489) – coincide. This simplifies the expressions for the quantile, VaR and ES.

Proposition 2.3 (quantiles and VaRs) *Suppose* (4) – (6) *hold and that* $c \in (0,1)$:

(a) If also Assumptions 1(a) and 3(a) hold, then the cth. quantile of z_t conditional on the past $\mathcal{F}^r_{t-1}$ is

$$
\begin{aligned}
z_{c,t} &= F_z^{-1}(c) \\
&= \begin{cases}
\pi_{1t}^{-1/2} F_{w_t|1}^{-1}(c/\pi_{1t}) & \text{if } c < F_{w_t|1}(0)\pi_{1t} \\
0 & \text{if } F_{w_t|1}(0)\pi_{1t} \leq c < F_{w_t|1}(0)\pi_{1t} + \pi_{0t} \\
\pi_{1t}^{-1/2} F_{w_t|1}^{-1}\left[\frac{(c-\pi_{0t})}{\pi_{1t}}\right] & \text{if } c \geq F_{w_t|1}(0)\pi_{1t} + \pi_{0t},
\end{cases}
\end{aligned}
\tag{10}
$$

and the $100 \cdot (1-c)\%$ Value-at-Risk (VaR_c) of z_t conditional on the past $\mathcal{F}^r_{t-1}$ is $-z_{c,t}$.

(b) If also Assumptions 1(b) and 3(b) hold, then the cth. quantile of r_t conditional on the past $\mathcal{F}^r_{t-1}$ is

$$
\begin{aligned}
r_{c,t} &= F_r^{-1}(c) \\
&= \begin{cases}
\pi_{1t}^{-1/2} F_{\widetilde{r}_t|1}^{-1}(c/\pi_{1t}) & \text{if } c < F_{\widetilde{r}_t|1}(0)\pi_{1t} \\
0 & \text{if } F_{\widetilde{r}_t|1}(0)\pi_{1t} \leq c < F_{\widetilde{r}_t|1}(0)\pi_{1t} + \pi_{0t} \\
\pi_{1t}^{-1/2} F_{\widetilde{r}_t|1}^{-1}\left[\frac{(c-\pi_{0t})}{\pi_{1t}}\right] & \text{if } c \geq F_{\widetilde{r}_t|1}(0)\pi_{1t} + \pi_{0t},
\end{cases}
\end{aligned}
\tag{11}
$$

and the $100 \cdot (1-c)\%$ Value-at-Risk (VaR_c) of r_t conditional on the past $\mathcal{F}^r_{t-1}$ is $-r_{c,t}$.

Proof: See Appendix A.3.

The expression for $r_{c,t}$ is not necessarily the most convenient from a practitioner's point of view. Indeed, in some situations it is desirable to be able to write $r_{c,t} = \sigma_t z_{c,t}$, so that the estimation of σ_t and $z_{c,t}$ may be separated into two different steps. The following assumption ensures that $r_{c,t}$ can indeed be written as $\sigma_t z_{c,t}$.

Assumption 4 *σ_t is measurable with respect to $\mathcal{F}^r_{t-1}$.*

The assumption is fulfilled by most ARCH models, but not necessarily by SV models. The assumption is only needed to prove Propositions 2.4 and 2.6.

Proposition 2.4 *Suppose* (4) – (6) *and Assumptions 1, 3 and 4 hold. If $c \in (0,1)$, then $r_{c,t} = \sigma_t z_{c,t}$, where $z_{c,t}$ is given by* (10).

Proof: See Appendix A.4

Note that we need both the (a) and (b) parts of Assumptions 1 and 3 for the proposition to hold.

Figures 1 and 2 provide an insight into the effect of zeros on VaR for a fixed value of volatility σ_t. Figure 1 plots VaR (i.e. $-z_{c,t}$) for different values of c and π_{0t}, and for three different densities of w_t: The standard normal, the standardised Student's t with five degrees of freedom, and the standardised skew Student's t with five

degrees of freedom.[‡] When $c \in \{0.05, 0.01\}$, then VaR always increases when the zero-probability π_{0t} increases. By contrast, when $c = 0.10$ then VaR generally falls, with the exception being when $w_t \sim N(0, 1)$. There, VaR first falls and then increases in π_{0t}. In summary, therefore, the main implication of Figure 1 is that the effect of zeros on VaR, for a given level of volatility, is highly non-linear and dependent on the density of w_t. Nevertheless, if c is sufficiently small, then the figure suggests VaR usually increases when the zero-probability increases. In other words, if VaR is not corrected for the zero-probability, then risk – defined in terms of VaR – will be biased downwards. Figure 2 provides an insight into the relative size of the bias. The figure contains the ratio of the incorrect VaR (numerator) divided by the correct VaR (denominator). That is, $w_{c,t}/z_{c,t}$, where $w_{c,t}$ is the cth. quantile of w_t. Of course, $w_{c,t} = z_{c,t}$ when $\pi_{1t} = 1$. The plot reveals that, in relative terms, the effect depends, in non-linear ways, on c, π_{0t} and the density of w_t. Nevertheless, one general characteristic is that when $c \in \{0.05, 0.01\}$, then the largest effect on VaR occurs when w_t is normal. That is, the most commonly used density assumption.

2.4 Expected Shortfall (ES)

Let $F_X(x)$ and x_c denote the cdf and c-quantile of a random variable X, and let $1_{\{X < x_c\}}$ denote an indicator function equal to 1 if $X < x_c$ and 0 otherwise. Following Acerbi and Tasche (2002, Definition 2.6, p. 1491), we define the Expected Shortfall at level $c \in (0, 1)$ for a random variable X as

$$ES_c = -\frac{1}{c} \left[E(X 1_{\{X < x_c\}}) + x_c \left(c - F_X(x_c) \right) \right]. \tag{12}$$

The last term in the definition, i.e. $x_c \left(c - F_X(x_c) \right)$, is needed if F_X is discontinuous. This may complicate the expressions for ES_c considerably. As a mild simplifying assumption, therefore, we introduce a continuity assumption on $F_{w_t|1}$ and $F_{\tilde{r}_t|1}$, which ensures that the term is zero for F_{z_t} and F_{r_t}.

Assumption 5 *Conditional on the past $\mathcal{F}^r_{t-1}$ and $I_t = 1$:*

(a) The cdf of w_t, denoted by $F_{w|1}$, is continuous and has density with respect to the Lebesgue measure.

(b) The cdf of $\tilde{r}_t$, denoted by $F_{\tilde{r}_t|1}$, is continuous and has density with respect to the Lebesgue measure.

The assumption is mild in the sense that it is assumed in most of the empirical applications that compute VaR and ES. That the assumption indeed ensures that $x_c \left(c - F_X(x_c) \right)$ is zero for both z_t and r_t, is shown in Appendix A.5 (see Lemma A.2).

Proposition 2.5 (ES) *Suppose (4) – (6) hold and that $c \in (0, 1)$:*

[‡]The skewing method used is that of Fernández and Steel (1998), and it is implemented by means of the corresponding functions in the R package fGarch, see Wuertz et al. (2016).

(a) If Assumptions 1(a), 3(a) and 5(a) also hold, then the $100 \cdot (1-c)\%$ Expected Shortfall (ES_c) of z_t conditional on the past $\mathcal{F}^r_{t-1}$ is $-c^{-1}E_{t-1}(z_t|z_t \leq z_{c,t})$, where

$$
E_{t-1}(z_t|z_t \leq z_{c,t})
= \begin{cases}
\pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w|1}^{-1}(c/\pi_{1t})\}}\right) & \text{if } c < F_{w|1}(0)\pi_{1t}, \\
\pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq 0\}}\right) & \text{if } F_{w|1}(0)\pi_{1t} \leq c < F_{w|1}(0)\pi_{1t} + \pi_{0t}, \\
\pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w|1}^{-1}[(c-\pi_{0t})/\pi_{1t}]\}}\right) & \text{if } c \geq F_{w|1}(0)\pi_{1t} + \pi_{0t},
\end{cases}
\tag{13}
$$

(b) If Assumptions 1(b), 3(b) and 5(b) also hold, then the $100 \cdot (1-c)\%$ Expected Shortfall (ES_c) of r_t conditional on the past $\mathcal{F}^r_{t-1}$ is $-c^{-1}E_{t-1}(r_t|r_t \leq r_{c,t})$, where

$$
E_{t-1}(r_t|r_t \leq r_{c,t})
= \begin{cases}
\pi_{1t}^{1/2} E_{t-1}\left(\widetilde{r}_t 1_{\{\widetilde{r}_t \leq F_{\widetilde{r}|1}^{-1}(c/\pi_{1t})\}}\right) & \text{if } c < F_{\widetilde{r}|1}(0)\pi_{1t}, \\
\pi_{1t}^{1/2} E_{t-1}\left(\widetilde{r}_t 1_{\{\widetilde{r}_t \leq 0\}}\right) & \text{if } F_{\widetilde{r}|1}(0)\pi_{1t} \leq c < F_{\widetilde{r}|1}(0)\pi_{1t} + \pi_{0t}, \\
\pi_{1t}^{1/2} E_{t-1}\left(\widetilde{r}_t 1_{\{\widetilde{r}_t \leq F_{\widetilde{r}|1}^{-1}[(c-\pi_{0t})/\pi_{1t}]\}}\right) & \text{if } c \geq F_{\widetilde{r}|1}(0)\pi_{1t} + \pi_{0t},
\end{cases}
\tag{14}
$$

Proof: See Appendix A.5.

Just as with the expression for the quantile $r_{c,t}$ in Proposition 2.3, the expression for $E_{t-1}(r_t|r_t \leq r_{c,t})$ is not necessarily the most convenient from a practitioner's point of view. Indeed, in many situations it would be desirable if we could write $E_{t-1}(r_t|r_t \leq r_{c,t})$ as $\sigma_t E_{t-1}(z_t|z_t \leq z_{c,t})$, so that the estimation of σ_t and $E_{t-1}(z_t|z_t \leq z_{c,t})$ may be separated into two different steps. If we rely on all of the assumptions stated so far, apart from Assumption 2, then we can indeed write the expression in this way.

Proposition 2.6 *Suppose* $(4) - (6)$, *and Assumptions 1 and $3 - 5$ hold. If $c \in (0,1)$, then $E_{t-1}(r_t|r_t \leq r_{c,t}) = \sigma_t E_{t-1}(z_t|z_t \leq z_{c,t})$, where $E_{t-1}(z_t|z_t \leq z_{c,t})$ is given by* (13).

Proof: See Appendix A.6.

For a given volatility level σ_t, ES is determined by the ES of z_t, i.e. $-c^{-1}E_{t-1}(z_t|z_t \leq z_{c,t})$ from Proposition 2.5(a). Figure 3 plots this expression for different values of c and π_{0t}, and for different densities of w_t (the same as those for VaR above). Contrary to the VaR case, here the effect is always monotonous for $c \in \{0.10, 0.05, 0.01\}$: ES increases as the zero-probability increases. In other words, risk – defined as ES – will be biased downwards if it is not corrected for the zero-probability. Figure 4 provides an insight into the magnitude of the bias in relative terms. The plots contain the ratios of ES of z_t: The numerator contains ES under the assumption that $\pi_{1t} = 1$, i.e. $-c^{-1}E_{t-1}(w_t|w_t \leq w_{c,t})$, whereas the denominator contains ES of z_t adjusted for zeros, i.e. $-c^{-1}E_{t-1}(z_t|z_t \leq z_{c,t})$. Of course, the expressions are equal when $\pi_{1t} = 1$. The plots reveal that, in relative

terms, the smaller the c, the larger the effect. The largest effect occurs when $c = 0.01$ and w_t is normal, just as in the VaR case.

2.5   Estimation of volatility

The σ_t can be specified in terms of a wide range of volatility models. If $\{z_t\}$ is a MDS that is strictly stationary and ergodic, for example, then the result by Lee and Hansen (1994) means that σ_t can be specified as a GARCH(1,1) in the usual way, i.e.

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{15}$$

since Gaussian QML then provides strongly consistent and asymptotically normal estimates of α_0, α_1 and β_1. Of course, this holds even if z_t is non-normal and skewed in unknown ways (in fact, the conditional third and fourth moments of z_t can even be time varying). Escanciano (2009) and Francq and Thieu (2018) extend this result to the GARCH(p,q) and GARCH(p,q)-X specifications, respectively. In particular, the latter accommodates asymmetry (i.e. "leverage") and stationary covariates ('X'), including past values of I_t, as conditioning variables. Another example of σ_t with z_t stationary is a log-GARCH(1,1) that "skips" the zeros, i.e.

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 I_{t-1} \ln r_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2, \tag{16}$$

where $I_t \ln r_t^2 = \ln r_t^2$ if $I_t = 1$ and 0 otherwise. A MEM version of this specification was proposed by Hautsch et al. (2013) for volume, and according to Francq and Zakoïan (2019) an extended version of the specification is strictly stationary and ergodic.

If the zero process $\{I_t\}$ is not stationary, however, then z_t is not strictly stationary. The zero process can be non-stationary if, say, the zero-probability is periodic (as in intraday returns), or if it is trending upwards or downwards over time because of general market developments (e.g. the influx of high-frequency algorithmic trading, increased trading volume, increased quoting frequency, lower tick-size, etc.). In this case, an alternative approach to the specification of σ_t is to formulate it in terms of zero-corrected return $\widetilde{r}_t = \sigma_t w_t$. For example, the GARCH(1,1) model in terms of zero-corrected return is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{17}$$

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

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \widetilde{r}_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2. \tag{18}$$

If $\widetilde{r}_t$ were observed, then estimation could proceed as usual by, say, maximising $\sum_{t=1}^n \ln f_{\widetilde{r}_t}(\widetilde{r}_t)$, where $f_{\widetilde{r}_t}$ is a suitably chosen density. In practice, however, $\widetilde{r}_t$ is not observed. Instead, therefore, we propose an approximate

estimation and inference procedure that consists of first replacing $\widetilde{r}_t$ with its estimate $r_t \widehat{\pi}_{1t}^{1/2}$, and then to treat zeros as "missing":

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

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-corrected return $\widetilde{r}_t$ is unobserved or "missing".

3. Use an estimation procedure that handles missing values to estimate the volatility model.

Sucarrat and Escribano (2017) propose an algorithm of this type for the log-GARCH model, where missing values are replaced by estimates of the conditional expectation (see also Francq and Sucarrat (2018)). If Gaussian (Q)ML is used for estimation, then this can be viewed as a dynamic variant of the Expectation Maximisation (EM) algorithm. A similar algorithm can be devised for many additional volatility models, including the GARCH model, subject to suitable assumptions. Appendix B contains the details of the algorithm together with a small simulation study, whereas Section 3 illustrates the usage of the algorithm. It should be noted that the algorithm does not necessarily provide consistent parameter estimates – in particular if the zero-probability is large. The reason for this is that the missing values induce a repeated irrelevance of initial value problem, see the discussion in Sucarrat and Escribano (2017).

3 An illustration

The aim of this section is to provide a detailed illustration of the results and methods of the previous section. To this end, we use the daily returns of three stocks listed at the NYSE. The stocks have been carefully selected to illustrate three different types of zero-probability dynamics. The first stock, General Electric (GE), is a high-volume stock, since its trading volume averages about 68 million USD per day over the sample. The second stock, Vonage Holdings Corporation (VG), a cloud communication services company, is a medium-volume stock, since its traded volume on average is about 2.2 million USD per day over the sample. The third stock, The Bank of New York Mellon Corporation (BKT), a financial products and services firm, is a low-volume stock, since its trading volume averages about 0.18 million USD per day over the sample. The daily returns are computed as $(\ln S_t - \ln S_{t-1}) \cdot 100$, where S_t is the stock price at the end of day t. Saturdays, Sundays and other non-trading days are excluded from the sample, and the sample period is 3 January 2007 – 31 December 2014. The sample period thus coincides with the in-sample analysis in Section 4. The source of the data is Bloomberg, and the data was obtained with the R package `Rblpapi` (Armstrong et al. (2018)) on a Bloomberg terminal. Descriptive statistics of the returns are contained in the upper part of Table 1. The statistics confirm that the returns exhibit the usual properties of excess kurtosis when compared with the normal distribution, and ARCH as measured by first order autocorrelation in the squared return. The fraction of zeros over the sample is 1.5% for GE, 7.4% for VG and 12.8% for BKT.

3.1 Models

The middle part of Table 1 contains estimates of three logit models for each return:

$$
\begin{array}{rlll}
\text{Constant:} & h_t & = & \rho_0, \\
\text{ACL(1,1):} & h_t & = & \rho_0 + \rho_1 s_{t-1} + \zeta_1 h_{t-1}, \qquad s_t = (I_t - \pi_{1t})/\sqrt{\pi_{1t}\pi_{0t}}, \\
\text{Trend:} & h_t & = & \rho_0 + \lambda t^*, \qquad t^* = t/T, \qquad t^* \in (0,1].
\end{array}
$$

In all three the conditional zero-probability π_{0t} is given by $(1 - \pi_{1t})$ with $\pi_{1t} = 1/(1 + \exp(-h_t))$. In the first model the zero-probability is constant, whereas in the second it is driven by a first order Autoregressive Conditional Logit (ACL) specification. The ACL is the binomial version of the Autoregressive Conditional Multinomial (ACM) of Russell and Engle (2005). In the third model the conditional zero-probability is governed by a deterministic trend (t^* is "relative time"). To select the specification that best characterises the zero-probability, we use the Schwarz (1978) information criterion (SIC), whose values are contained in the second-to-last column of the middle part in Table 1. For GE returns it is the first specification that fits the data best, for VG it is the second, and for BKT it is the third. In other words, according to the SIC, the conditional zero-probability of GE returns is constant, the conditional zero-probability of VG returns is time varying and stationary, whereas the conditional zero-probability of BKT returns is time varying and non-stationary. The first row of graphs in Figure 5 contains the fitted conditional zero-probability $\widehat{\pi}_{0t}$ of the selected models. For GE returns it is constant at 1.5%. For VG returns it varies between 5.6% and 25.9%, and the dynamics is characterised by clustering. That is, a high $\widehat{\pi}_{0t}$ tends to be followed by another high one, and a low $\widehat{\pi}_{0t}$ tends to be followed by another low one. The fitted conditional zero-probability of BKT returns exhibits a clear upwards trend. It starts at a minimum of 8.4% in the beginning of the sample, and increases gradually to a maximum of 18.4% at the end of the sample.

The bottom part of Table 1 contains GARCH(1,1) estimates of the return series. We fit an Ordinary GARCH-specification to all three return series, whereas to BKT returns we also fit a zero-corrected GARCH-specification. The Ordinary specification is given by

$$
\text{Ordinary:} \qquad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{19}
$$

If z_t is strictly stationary and ergodic, then the results by Escanciano (2009), and Francq and Thieu (2018) imply that Gaussian QML provides consistent parameter estimates (subject to additional regularity conditions) even if π_{0t} is time varying. As noted above, however, I_t is non-stationary for BKT. This means that z_t is not strictly stationary, and so the results by Escanciano (2009), and Francq and Thieu (2018) are not applicable. To accommodate the non-stationarity of I_t in the BKT case, we also fit a zero-corrected GARCH(1,1) specification

to its returns:

$$\text{0-corrected:} \qquad \sigma_t^2 = \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \qquad\qquad (20)$$

The parameters are estimated by Gaussian QML in combination with the missing values algorithm outlined in Section 2.5. The algorithm proceeds by replacing $\widetilde{r}_t$ with its estimate $\widehat{\pi}_{1t}^{1/2} r_t$ whenever $r_t \neq 0$, while treating zeros as missing observations. The $\widehat{\pi}_{1t}$'s are those of the Trend model. Next, the missing values are replaced by estimates of their conditional expectations, i.e. $\widehat{E}_{t-1}(\widetilde{r}_t^2) = \widehat{\sigma}_t^2$. Since Gaussian QML is used in the estimation, the algorithm can be viewed as a dynamic variant of the Expectation-Maximisation (EM) algorithm (see Appendix B for more details). The nominal differences between the parameter estimates of the Ordinary and 0-corrected specifications may appear small. However, as we will see, these nominal differences – together with the different treatment of zeros – can lead to substantially different risk estimates and risk dynamics.

3.2 Volatility

For GE and VG, estimates of σ_t^2 are unaffected by zeros (subject to the assumption that z_t is strictly stationary and ergodic). For BKT, the difference between the estimates is $x_t = \widehat{\sigma}_{t,\text{0-adj}}^2 - \widehat{\sigma}_t^2$, where $\widehat{\sigma}_{t,\text{0-adj}}^2$ is the estimate produced by the zero-corrected GARCH, and $\widehat{\sigma}_t^2$ is the estimate obtained under the erroneous statistical assumption that the zero-probability is stationary. So x_t can be interpreted as an estimate of the error incurred by the ordinary GARCH. The second row in Figure 5 contains graphs of the errors. For GE and VG, the errors are all 0 over the sample, since estimates of σ_t^2 are unaffected by zeros. The Mean Error (ME) provides a measure of the overall or unconditional error, whereas the Mean Absolut Error (MAE) provides a measure of the day-to-day or conditional error. For BKT, the ME and MAE are computed as $T^{-1} \sum_{t=1}^{T} x_t$ and $T^{-1} \sum_{t=1}^{T} |x_t|$, respectively. Accordingly, a negative value on ME means the incorrect risk estimate is, on average, higher than the zero-corrected one. In the graphs, the values in square brackets are p-values associated with tests of ME and MAE. For both ME and MAE the tests are implemented via the OLS estimated regression $x_t = \mu + u_t$ with standard error of the Newey and West (1987) type. For ME, $H_0 : \mu = 0$ and $H_A : \mu \neq 0$. For MAE, to avoid non-standard inference, we specify the null as $H_0 : \mu = 0.01$, i.e. away from the lower bound 0 of the permissible parameter space, and the alternative as $H_A : \mu > 0.01$. The ME is -0.013 and significantly different from zero at the most common significance levels. The value of -0.013 means the risk, as measured by the conditional variance, is estimated to be too high by 0.013 points on average if the zeros are not corrected for. However, the graph shows that, on a day-to-day basis, the differences can be much larger in absolute value: The maximum difference is 0.37 points, whereas the minimum is -1.33 points. In other words, on a day-to-day basis, the difference can be very large with substantial implications for risk analysis. The MAE, which provides an overall measure of the day-to-day differences, is 0.04 and significantly greater than 0.01 at all the usual significance levels.

3.3 Value-at-Risk (VaR)

To illustrate the effect of time varying zero-probability on VaR, we choose $c = 0.025$. This corresponds to the 97.5% VaR. The differences between the estimated VaRs are contained in the third row of graphs in Figure 5. The difference or error at t is given by $x_t = \hat{r}_{c,t} - \hat{r}_{c,t,0-adj}$, which is equivalent to $x_t = -\hat{r}_{c,t,0-adj} - (-\hat{r}_{c,t})$. That is, zero-corrected VaR minus incorrect VaR. Since return r_t is expressed in percent, the difference x_t can be interpreted as the percentage point difference between the VaRs, and $100 \cdot x_t$ can be interpreted as the basis point difference. For GE, VG and BKT, $\hat{r}_{c,t}$ is computed as $\hat{\sigma}_t \hat{z}_c$, where $\hat{\sigma}_t$ is the fitted value of (19), and $\hat{z}_c$ is the empirical c-quantile of the residuals $\hat{z}_t$. Subject to suitable regularity assumptions, $\hat{z}_t$ provides a consistent estimate, see e.g. Francq and Zakoïan (2015), and Ghourabi et al. (2016). For GE and VG, $\hat{r}_{c,t,0-adj}$ is computed as $\hat{\sigma}_t \hat{z}_{c,t}$, where $\hat{z}_{c,t}$ is obtained using the relevant formula in (10), i.e. $\pi_{1t}^{-1/2} F_{w|1}^{-1}(c/\pi_{1t})$. To estimate $F_{w|1}^{-1}(c/\pi_{1t})$ at t we use the empirical $c/\hat{\pi}_{1t}$-quantile of the zero-corrected residuals $\hat{w}_t$ (zeros excluded). For BKT, $\hat{r}_{c,t,0-adj}$ is computed as $\hat{\sigma}_{t,0-adj} \hat{z}_{c,t}$, where $\hat{\sigma}_{t,0-adj}$ is the fitted value of (20), and $\hat{z}_{c,t}$ is computed in the same way as for GE and VG. Again we use the ME as an overall or unconditional measure of the errors, and MAE as an average measure of the day-to-day differences. We also implement tests of ME and MAE in the same way as above (Section 3.2).

Unsurprisingly, both ME and MAE are essentially 0 for GE, although the latter is statistically significant at the usual significance levels. For VG, the tests of ME and MAE are both significant at the usual levels, and both are equal to 0.09. That is, on average the incorrect VaR is 0.09%-points lower than the zero-corrected VaR, both overall and on a day-to-day basis. The reason they are identical is that the zero-corrected VaR is always higher than the incorrect VaR over the sample. The maximum difference over the sample is 1.21%-points. For BKT, the tests of ME and MAE are also significant at the usual levels, and their values are both negative and equal to -0.25 when rounded to two decimals. On a day-to-day basis the discrepancy can be as large as -1.91. The negative sign on ME is opposite to that of VG. In other words, the presence of a time varying zero-probability may bias VaR either upwards or downwards.

3.4 Expected Shortfall (ES)

To illustrate the effect of zeros on ES, we choose $c = 0.025$. This corresponds to the 97.5% ES. The differences between the estimated ESs are contained in the bottom row of graphs in Figure 5. The difference at t is given by $x_t = \widehat{ES}_{c,t,0-adj} - \widehat{ES}_{c,t}$, where $\widehat{ES}_{c,t,0-adj}$ is the zero-corrected ES and $\widehat{ES}_{c,t}$ is incorrect ES. Also here can x_t and $100 \cdot x_t$ be interpreted as the percentage point and basis point difference, respectively. For GE, VG and BKT, $\widehat{ES}_{c,t}$ is computed as $-c^{-1}\hat{\sigma}_t \hat{E}_{t-1}(z_t|z_t \leq z_{c,t})$, where $\hat{\sigma}_t$ is the estimate from (19), and $\hat{E}_{t-1}(z_t|z_t \leq z_c)$ is computed as the sample average of the residuals $\hat{z}_t$ that are equal to or lower than $\hat{z}_c$ as defined above (i.e. the empirical c-quantile of the residuals $\hat{z}_t$). Subject to suitable regularity assumptions, $\hat{E}_{t-1}(z_t|z_t \leq z_c)$ provides a consistent estimate, see e.g. Francq and Zakoïan (2015). For GE and VG, the zero-corrected estimate $\widehat{ES}_{c,t,0-adj}$ is computed as $-c^{-1}\hat{\sigma}_t \hat{E}_{t-1}(z_t|z_t \leq z_{c,t})$, where $\hat{E}_{t-1}(z_t|z_t \leq z_{c,t})$ now is obtained via the relevant

formula in (13), i.e. $\pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w_t|1}^{-1}(c/\pi_{1t})\}}\right)$. To estimate $F_{w_t|1}^{-1}(c/\pi_{1t})$ at t we use the empirical $c/\widehat{\pi}_{1t}$-quantile of the zero-corrected residuals $\widehat{w}_t$ (zeros excluded). Next, we estimate $E_{t-1}\left(w_t 1_{\{w_t \leq F_{w_t|1}^{-1}(c/\pi_{1t})\}}\right)$ at t by forming an average made up of the non-zero residuals $\widehat{w}$: $T_1^{-1} \sum_{I_t=1} \widehat{w}_t 1_{\{\widehat{w}_t \leq \widehat{F}_{w_t|1}^{-1}(c/\widehat{\pi}_{1t})\}}$, where T_1 is the number of non-zero observations (i.e. $T_1 = \sum_{t=1}^n I_t$), $\widehat{F}_{w_t|1}^{-1}(c/\widehat{\pi}_{1t})$ is the estimate of $F_{w_t|1}^{-1}(c/\pi_{1t})$, and the symbolism $\sum_{I_t=1}$ means the summation is over non-zero values only. For BKT, the zero-corrected estimate $\widehat{ES}_{c,t,0-adj}$ is computed as $-c^{-1}\widehat{\sigma}_{t,0-adj}\widehat{E}_{t-1}(z_t|z_t \leq z_{c,t})$, where $\widehat{\sigma}_{t,0-adj}$ is the estimate from (20), and where $\widehat{E}_{t-1}(z_t|z_t \leq z_{c,t})$ is computed in the same way as for GE and VG. Again we use the ME as an overall measure, and MAE as an average measure of the day-to-day differences. Tests of ME and MAE are implemented in the same way as above.

As indicated by the bottom row of graphs in Figure 5, for GE the ME and MAE are both essentially 0. The test of ME, however, reject the null at the usual significance levels. Note that, here, the difference is not due to a time varying zero-probability, but the discreteness in the cumulative density function. For VG, the ME and MAE are -0.06 and 0.08, respectively, and the null is rejected at the usual significance levels in both tests. The negative sign on ME means the incorrect ES is biased upwards by about 0.06%-points on average. However, as the graph show, on a day-to-day basis it can be about 1.1%-points in absolute value. Interestingly, the negative sign of the overall bias is opposite to its VaR case, since there the sign of the overall bias is positive. For BKT, the ME and MAE are both 0.73, and also here is the null rejected at the usual significance levels in both tests. The positive sign on ME means the incorrect ES is, on average, 0.73%-points lower. On a day-to-day basis, however, the graph reveals the difference can be as large as 4.3%-points in absolute value. The positive sign on ME is opposite to that of VG. So just as for VaR, the presence of a time varying zero-probability may bias ES either upwards or downwards. Finally, the positive sign of the overall bias on ME for BTK is opposite to its VaR case, since there the sign of the overall bias is negative.

4   The importance of time varying zero-probabilities at the NYSE

The New York Stock Exchange (NYSE) is one of the largest stock exchanges in the world measured by market capitalisation. The period we study is 3 January 2007 – 4 February 2019, i.e. a maximum of 3043 daily observations before lagging and differencing. Weekends and non-trading days are excluded from the sample. We split the sample period in two. The first part, the in-sample period, goes from the start of 2007 until the end of 2014 (up to 2014 observations before lagging and differencing). This part is used to identify the zero-probability dynamics that characterises each stock return. The remaining part (up to 1029 observations) is used for the out-of-sample comparison. To ensure that a sufficient number of observations is used for the in-sample identification, we exclude all stocks with less than 1000 observations in the in-sample period. This leaves us with 1665 stocks out of the about 2300 stocks listed at NYSE in February 2019. It is reasonable to conjecture that this induces a selection bias: The stocks that are left out are more likely to be characterised by a time varying zero-probability. To identify the type of zero-probability dynamics exhibited by each stock, we use the strategy

of Section 3.1. That is, we fit three logit-models to each return (Constant, ACL(1,1) and Trend), and compare their fit by means of the Schwarz (1978) information criterion. The source of the data is Bloomberg, and the data were downloaded with the R package Rblpapi (Armstrong et al. (2018)) on a Bloomberg terminal.

Table 2 contains the identification results. Out of the 1665 stock return series, 1259 are found to have a constant zero-probability, 228 are found to have a time varying zero-probability of the ACL(1,1) type, and 178 are found to have a trend-like time varying zero-probability. That means 24.4% of the stocks we study at NYSE are characterised by a time varying zero-probability. As noted above, the actual proportion is likely to be higher, since the stocks we omit from our analysis are likely to be characterised by a high zero-probability, and therefore also by a time varying zero-probability. This conjecture is supported by Table 2: The average of the zero-proportions is higher among the stocks characterised by ACL and trend-like dynamics (2.6% and 3.2% in comparison to 1.9%). As expected, the average daily trading volume is lower among the stocks with a time varying zero-probability. However, the relationship between zero-proportions and daily average volumes is maybe not as strong as expected. Across all stocks, the sample correlation is -0.14. Among the stocks with a constant zero-probability, the correlation is -0.13. Among the stocks with time varying zero-probability, the correlation is -0.21 for the stocks with ACL-like dynamics, and -0.28 for the stocks with trend-like dynamics.

4.1 Out-of-sample forecasting of volatility

To shed light on the importance of a time varying zero-probability in out-of-sample volatility forecasting, we compare the 1-step ahead volatility forecasts of an ordinary GARCH(1,1) with that of a zero-corrected GARCH(1,1). We use the same approach as in Section 3.2. Recall that the QMLE estimates of an ordinary GARCH(1,1) are valid when the zero process is stationary, even if the the zero-probability is time varying. Accordingly, we restrict the comparison to the 178 stock returns that are characterised by a non-stationary zero process. The ordinary GARCH(1,1) is thus estimated under the erroneous statistical assumption that the zero process is stationary, whereas the zero-corrected GARCH(1,1) accommodates non-stationarity by means of the method proposed in Section 2.5.

Let $\widehat{\sigma}_{it,0\text{-adj}}^2$ denote the fitted zero-corrected volatility of stock i, and let $\widehat{\sigma}_{it}^2$ denote the fitted ordinary volatility of stock i, $t = 1, 2, \ldots, T_i$, where T_i is the number of out-of-sample observations for stock i. Note that T_i varies slightly across the 178 stocks, but is usually 1029 (the minimum T_i across the stocks is 988). For each out-of-sample day $t = 1, 2, \ldots, T_i$, we fit an ordinary and a zero-corrected GARCH(1,1) model to each stock return, and then generate 1-step forecasts of volatility. The sample used for estimation and forecasting consists of the observations preceding t. So the sample size increases with t as more observations become available. It is unclear whether and to what extent standard volatility proxies made up of high-frequency intraday data provide accurate estimates of volatility in the presence of time varying and non-stationary zero-probabilities. So the best measure of volatility at hand is probably the estimate provided by the zero-corrected model. Let $x_{it} = \widehat{\sigma}_{it,0\text{-adj}}^2 - \widehat{\sigma}_{it}^2$ denote the 1-step forecast error at t. The ME and MAE are computed as $\sum_{t=1}^{T_i} x_{it}/T_i$ and $\sum_{t=1}^{T_i} |x_{it}|/T_i$, respectively. The former provides a measure of the overall or unconditional error, whereas the

latter provides a measure of the day-to-day or conditional error. Tests of ME and MAE are implemented as in Section 3.2.

The results are contained in the upper part of Table 3. The average of the MEs is -0.059, the maximum ME is 2.832 and the minimum is -1.686. In other words, although the average of the MEs is negative, the results do not suggest that there is a clear tendency in the sign of the bias. Out of the 178 tests with $H_0 : \mu_i = 0$ and $H_A : \mu_i \neq 0$, the null is rejected 149 times at the 10% significance level, 140 times at 5% and 127 times at 1%. This is substantially more than what is expected by chance: If $\mu_i = 0$ for all i, then one should on average expect 17.8 false rejections at the 10% significance level, 8.9 false rejections at 5% and 1.78 false rejections at 1%. Accordingly, the large number of rejections provide comprehensive evidence of an overall or unconditional effect of time varying zero-probability. As for a day-to-day effect, the average of the MAEs is 0.302, the maximum MAE is 4.092 and the minimum is 0.008. Out of the 178 tests with $H_0 : \mu_i = 0.01$ and $H_A : \mu_i > 0.01$, the null is rejected 175 times at the 10% and 5% significance levels, and 173 times at 1%. By chance, one would on average expect the same number of false rejection as in the ME tests. So the results provide even more comprehensive evidence of a day-to-day discrepancy than in the unconditional case.

4.2 Out-of-sample VaR forecasting

To shed light on the importance of a time varying zero-probability in the out-of-sample forecasting of VaR, we compare the incorrect 1-step ahead VaR forecasts with the zero-corrected ones. The comparison is made for all the $n = 406$ stocks with a time varying zero-probability. As in Section 3.3, we choose $c = 0.025$, which corresponds to the 97.5% VaR. Let $-\widehat{r}_{c,it,0-adj}$ denote the zero-corrected 97.5% VaR of stock i at t, and let $-\widehat{r}_{c,it}$ denote the incorrect 97.5% VaR of stock i at t. The ME and MAE are computed as $\sum_{t=1}^{T_i} x_{it}/T_i$ and $\sum_{t=1}^{T_i} |x_{it}|/T_i$, respectively, where $x_{it} = -\widehat{r}_{c,it,0-adj} - (-\widehat{r}_{c,it}) = \widehat{r}_{c,it} - \widehat{r}_{c,it,0-adj}$ is the error at t. Tests of ME and MAE are implemented as above. For each out-of-sample day $t = 1, 2, \ldots, T_i$, forecasts are obtained as described in Section 3.3. The sample used for estimation consists of the observations preceding t, so the sample size increases with t as more observations become available, just as in the out-of-sample forecasting of volatility above.

The middle part of Table 3 contains the results. The average of the MEs is 0.004, and they range from -0.241 (minimum) to 0.434 (maximum). As for volatility, the results do not suggest a clear tendency in the sign of the bias across stocks. Out of the 406 tests of ME, the null is rejected 328, 307 and 269 times at the 10%, 5% and 1% significance levels, respectively. Again, this is substantially more rejections than what is expected by chance: If $\mu_i = 0$ for all i, then one should on average expect 40.6, 20.3 and 4.06 false rejections, respectively. The average of the MAEs is 0.050, and they range from 0.000 (minimum) to 0.692 (maximum). Out of the 406 tests of MAE, the null is rejected 255, 254 and 248 times at the 10%, 5% and 1% levels, respectively. Just as for ME, this is substantially more than what is expected by chance. All-in-all, therefore, the large number of rejections – both for ME and MAE – provide comprehensive support of the hypothesis that an appropriate zero-correction can improve out-of-sample VaR forecasts significantly.

Table 4 provides some diagnostics on the VaR forecasts. The table contains the results of two tests proposed by Christoffersen (1998): The unconditional coverage test and an independence test. In both tests one should – on average – expect 40.6, 20.3 and 4.06 false rejections, respectively, at the 10%, 5% and 1% significance levels, respectively. In the first test there are 62, 36 and 14 rejections, respectively, for the unadjusted model. For the zero-corrected model there are 67, 44 and 13 rejections, respectively. The number of rejections is thus slightly higher for the zero-corrected model at 10% and 5%, and slightly lower at 1%. All-in-all, the number of rejections are not substantially higher than what one should on average expect by chance. This means both methods produce, in general, good VaR forecasts in the unconditional coverage sense. For the independence test, the number of rejections is identical for the two models, and substantially higher than one should expect by chance. However, it should be noted that independence may not be required by either method. The large number of rejections nevertheless suggests there is room for improved risk estimates, e.g. by adding lagged covariates in the volatility and/or zero-probability specifications.

4.3   Out-of-sample ES forecasting

In this subsection we shed light on whether a correction for the time varying zero-probability improves the out-of-sample forecasting of ES. We use the same approach as for VaR: The incorrect 1-step ahead forecasts are compared out-of-sample with the zero-corrected ones. The comparison is made for all the $n = 406$ stocks return with time varying zero-probability. Again we choose $c = 0.025$, which corresponds to the 97.5% ES. Let $\widehat{ES}_{c,it,0-adj}$ denote the zero-corrected 97.5% ES forecast of stock i at t, and let $\widehat{ES}_{c,it}$ denote the incorrect 97.5% ES forecast of stock i at t. The forecasts are computed as in Section 3.4, so the difference or error is given by $x_{it} = \widehat{ES}_{c,it,0-adj} - \widehat{ES}_{c,it}$. The ME and MAE, and their associated tests, are defined in the same way as earlier. Finally, as for volatility and VaR, the sample used for estimation consists of the observations preceding t. So the sample size increases with t as more observations become available.

The bottom part of Table 3 contains the results. The average of the MEs is 0.004, and the MEs range from -0.340 (minimum) to 0.691 (maximum). So yet again there is no clear tendency with respect to the sign of the bias across stocks. Out of the 406 tests of ME, the null is rejected 255, 232 and 205 times at the 10%, 5% and 1% levels, respectively. Again, this is substantially more than what is expected on average by chance (40.6, 20.3 and 4.06 false rejections, respectively, under the null). The average of the MAEs is 0.074, and they range from 0.001 (minimum) to 0.926 (maximum). Out of the 406 tests of MAE, the null is rejected 328, 245 and 70 times at the 10%, 5% and 1% significance levels, respectively. Albeit this is substantially more than what is expected by chance, the number of rejections is notably smaller than for ME at the 1% level. This may suggest that the improvement induced by zero-correcting is – in general – small in nominal terms. Nevertheless, all-in-all, the results provide comprehensive support of the hypothesis that an appropriate zero-correction can improve out-of-sample ES forecasts significantly.

5   Conclusions

We propose a new class of financial return models that allows for a time varying zero-probability that can either be stationary or non-stationary. Standard volatility models (e.g. ARCH, SV and continuous-time models) are nested and obtained as special cases when the zero-probability is zero or constant, the zero and volatility processes are allowed to be mutually dependent, and properties of the new class (e.g. conditional volatility, skewness, kurtosis, VaR, ES, etc.) are obtained as functions of the underlying volatility model. Analytically, our results imply that, for a given volatility level, a higher conditional zero-probability increases the conditional skewness and kurtosis of return, but reduces return variability when defined as conditional absolute return. Moreover, for a given level of volatility and sufficiently rare loss events (5% or less), risk defined as VaR or ES will be biased downwards if zeros are not corrected for. Empirically, the sign and size of the bias will depend on a number of additional circumstances and how they interact: The magnitude of the zero-proportion, the stationarity properties of the zero process, the exact type of the zero-probability dynamics, the exact volatility model and/or estimator, and on the conditional density of return. To alleviate the unpredictable biases caused by non-stationary zero processes, we outline an approximate estimation and inference procedure that can be combined with standard volatility models and estimators. Finally, we undertake a comprehensive study of the stocks listed at the New York Stock Exchange (NYSE). We identify 24.4% of the daily returns that we study to be characterised by a time varying zero-probability. However, the actual proportion is likely to be higher, since we restrict our analysis to stocks with more than 1000 observations in the in-sample. Next, we conduct an out-of-sample forecast evaluation of our results and methods. Our results show that zero-corrected risk estimates provide an improvement in a large number of cases.

Our results have several empirical, theoretical and practical implications. First, we found a widespread presence of time varying zero-probabilities in daily stock returns at NYSE, which is one of the most liquid markets in the world. In less liquid markets, in other asset-classes, and at higher frequencies (i.e. intradaily), the proportion of zeros is likely to be substantially higher, and the zero-probability dynamics is likely to be much more pronounced. Accordingly, our results are likely to be of even greater importance in markets that are not as liquid as the NYSE. Second, the widespread presence of a non-stationary zero process prompts the need for new theoretical results. This is because most models, estimators and methods are derived under the assumption of a stationary zero process. Finally, at a practical level our results suggests more attention should be paid to how market quotes and transaction prices are aggregated in order to obtain the asset prices reported by data providers, Central Banks and others. In particular, if a non-stationary zero process is the result of specific data practices, then it may be worthwhile to re-consider these.

References

Acerbi, C. and D. Tasche (2002). On the coherence of expected shortfall. *Journal of Banking and Finance 26*, 1487–1503.

Armstrong, W., D. Eddelbuettel, and J. Laing (2018). *Rblpapi: R Interface to 'Bloomberg'*. R package version 0.3.8.

Bandi, F. M., D. Pirino, and R. Reno (2017). Excess idle time. *Econometrica 85*, 1793–1846.

Bandi, F. M., D. Pirino, and R. Reno (2018). Systematic staleness. Working paper available at SSRN: http://dx.doi.org/10.2139/ssrn.3208204.

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.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review 39*, 841–862.

Creal, D., S. J. Koopmans, and A. Lucas (2010). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*.

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

Embrechts, P. and M. Hofert (2013). A note on generalized inverses. *Mathematical Methods of Operations Research 77*(3), 423–432.

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.

Escanciano, J. C. (2009). Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory 25*, 561–570.

Fernández, C. and M. Steel (1998). On Bayesian Modelling of Fat Tails and Skewness. *Journal of the American Statistical Association 93*, 359–371.

Francq, C. and G. Sucarrat (2017). An Equation-by-Equation Estimator of a Multivariate Log-GARCH-X Model of Financial Returns. *Journal of Multivariate Analysis 153*, 16–32.

Francq, C. and G. Sucarrat (2018). An Exponential Chi-Squared QMLE for Log-GARCH Models Via the ARMA Representation. *Journal of Financial Econometrics 16*, 129–154. Working Paper version: http://mpra.ub.uni-muenchen.de/51783/.

Francq, C. and L. Q. Thieu (2018). Qml inference for volatility models with covariates. *Econometric Theory*. https://doi.org/10.1017/S0266466617000512.

Francq, C., O. Wintenberger, and J.-M. Zakoïan (2013). GARCH Models Without Positivity Constraints: Exponential or Log-GARCH? *Journal of Econometrics 177*, 34–36.

Francq, C. and J.-M. Zakoïan (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics 184*, 158–173.

Francq, C. and J.-M. Zakoïan (2019). *GARCH Models*. New York: Wiley. 2nd. Edition.

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

Ghourabi, M. E., C. Francq, and F. Telmoudi (2016). Consistent estimation of the value at risk when the error distribution of the volatility model is misspecified. *Journal of Time Series Analysis 37*, 46–76.

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.

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 12*, 89–121.

Kümm, H. and U. Küsters (2015). Forecasting zero-inflated price changes with a markov switching mixture model for autoregressive and heteroscedastic time series. *International Journal of Forecasting 31*, 598–608.

Lee, S. and B. Hansen (1994). Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory 10*, 29–52.

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.

Newey, W. and K. West (1987). A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703–708.

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

R Core Team (2018). *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.

Shiryaev, A. N. (1996). *Probability*. Springer-Verlag, New York,.

Shorack, G. and J. Wellner (1986). *Empirical Processes with Applications to Statistics*. Wiley.

Sucarrat, G. and Á. Escribano (2017). Estimation of Log-GARCH Models in the Presence of Zero Returns. *European Journal of Finance*. http://dx.doi.org/10.1080/1351847X.2017.1336452.

Sucarrat, G., S. Grønneberg, and Á. Escribano (2016). Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown. *Computational Statistics and Data Analysis 100*, 582–594.

Wuertz, D., Y. C. with contribution from Michal Miklovic, C. Boudt, P. Chausse, and others (2016). *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. R package version 3010.82.1.

A   Proofs

A.1   Proof of Proposition 2.1

Throughout, $E_{t-1}(w_t^s \cdot 0 | I_t = 0)\pi_{0t}$ with $s \geq 0$ stands for $E_{t-1}(w_t^s \cdot 0)$ whenever $\pi_{0t} = 0$.

(i) Assumption 2 and $E_{t-1}|z_t| < \infty$ imply that

$$
\begin{aligned}
E_{t-1}(z_t) &= \pi_{1t}^{-1/2} E_{t-1}(w_t I_t) \\
&= \pi_{1t}^{-1/2} \Big(E_{t-1}(w_t \cdot 1|I_t = 1)\pi_{1t} + E_{t-1}(w_t \cdot 0|I_t = 0)\pi_{0t} \Big) \\
&= 0
\end{aligned}
$$

for all t. Accordingly, $\{z_t\}$ is a MDS.

(ii) Assumption 2 and $E_{t-1}|z_t^2| < \infty$ imply that

$$
\begin{aligned}
E_{t-1}(z_t^2) &= \pi_{1t}^{-1} E_{t-1}(w_t^2 I_t^2) \\
&= \pi_{1t}^{-1} \Big(E_{t-1}(w_t^2 \cdot 1|I_t = 1)\pi_{1t} + E_{t-1}(w_t^2 \cdot 0|I_t = 0)\pi_{0t} \Big) \\
&= \pi_{1t}^{-1} \Big(\sigma_w^2 \pi_{1t} \Big) \\
&= \sigma_w^2.
\end{aligned}
$$

for all t. Next, since $\{z_t\}$ is a MDS and $Var_{t-1}(z_t) = \sigma_w^2$ for all t, we have (for all t) that $E(z_t) = 0$, $E(z_t^2) = \sigma_w^2$ and $Cov(z_{t-i}, z_{t-j}) = 0$ for all $i \neq j$. So $\{z_t\}$ is covariance-stationary.

(iii) Since $E_{t-1}|z_t^s| < \infty$, we have that

$$
\begin{aligned}
E_{t-1}(z_t^s) &= \pi_{1t}^{-s/2} E_{t-1}(w_t^s I_t) \\
&= \pi_{1t}^{-s/2} \Big(E_{t-1}(w_t^s \cdot 1|I_t = 1)\pi_{1t} + E_{t-1}(w_t^s \cdot 0|I_t = 0)\pi_{0t} \Big) \\
&= \pi_{1t}^{(2-s)/2} E_{t-1}(w_t^s|I_t = 1)
\end{aligned}
$$

for all t.

(iv) If $E_{t-1}|z_t^s| < \infty$, we have that

$$
\begin{aligned}
E_{t-1}|z_t|^s &= \pi_{1t}^{-s/2} E_{t-1}(|w_t|^s I_t^s) \\
&= \pi_{1t}^{-s/2} \Big(E_{t-1}(|w_t|^s \cdot 1|I_t = 1)\pi_{1t} + E_{t-1}(|w_t|^s \cdot 0|I_t = 0)\pi_{0t} \Big) \\
&= \pi_{1t}^{(2-s)/2} E_{t-1}(|w_t|^s|I_t = 1)
\end{aligned}
$$

for all t. The notation $E_{t-1}(|w_t|^s \cdot 0|I_t = 0)\pi_{0t}$ stands for $E_{t-1}(|w_t|^s \cdot 0)$ whenever $\pi_{0t} = 0$.

A.2 Proof of Proposition 2.2

Let $X_t = w_t I_t \pi_{1t}^{-1/2}$, and let $P_{t-1}(X_t \leq x)$ denote the cdf of X_t at t conditional on $\mathcal{F}_{t-1}^r$. By Assumption 1(a) this conditional probability is regular. Hence:

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

where we have used (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 term, (c) for $0 > x$ we have $P_{t-1}(0 \leq x \cap I_t = 0) = P_{t-1}(\emptyset \cap I_t = 0) = 0$, and for $0 \leq x$ we have $P_{t-1}(0 \leq x, I_t = 0) = P_{t-1}(\Omega \cap \{I_t = 0\}) = P_{t-1}(I_t = 0) = \pi_{0t}$, where Ω is the whole outcome set of the underlying probability space, (d) the assumption $\pi_{1t} = P_{t-1}(I_t = 1)$ in (6) implies that π_{1t} is measurable with respect to $\mathcal{F}_{t-1}^r$.

Replacing w_t with $\widetilde{r}_t$ so that $X_t = r_t$, and assuming Assumption 1(b) instead of Assumption 1(a), gives (8).

A.3 Proof of Proposition 2.3

Let f, g denote two functions, and let $f \circ g$ denote function composition so that $f \circ g(x) = f(g(x))$. The statements in the following Lemma will be used in the proofs of Propositions 2.3 and 2.5.

Lemma A.1 *Let $\xi \sim U[0,1]$, let F be a cdf, and let F^{-1} be the generalised inverse of F as defined in (9).*

(a) *We have that $X := F^{-1}(\xi) \sim F$, that is, X is distributed according to F.*

(b) *We have $\{F^{-1}(\xi) \leq x\} = \{\xi \leq F(x)\}$ as events, for any x.*

(c) *We have that $F \circ F^{-1}(c) \geq c$ for all $0 \leq c \leq 1$ with equality failing if and only if c is not in the range of F on $[-\infty, \infty]$.*

(d) *We have that $F^{-1} \circ F(x) \leq x$ for all $-\infty < x < \infty$ with equality failing if and only if $F(x - \varepsilon) = F(x)$ for some $\varepsilon > 0$.*

All four statements are contained and proved in Shorack and Wellner (1986): (a) and (b) are in Theorem 1 on p. 3, (c) is Proposition 1 on p. 5, and (d) is Proposition 3 on p. 6.

From Assumption 3(a) and the expression for $F_{z_t}(x)$ in Proposition 2.2, it follows that $F_{z_t}(x)$ is strictly increasing for $x \in (-\infty, 0) \cup (0, \infty)$. So in these regions the inverse function exists, and solves the equation $F_{z_t}(x) = c$ for c. We first deal with the intervals $(-\infty, 0)$ and $(0, \infty)$, and then the case corresponding to $x = 0$:

1. For $x \in (-\infty, 0)$ it follows from Proposition 2.2 that $F_{z_t}(x) = F_{w_t|1}(x\pi_{1t}^{1/2})\pi_{1t}$, and hence that $c < F_{w_t|1}(0)\pi_{1t}$. Next: $F_{z_t}(x) = c \Leftrightarrow F_{w_t|1}(x\sqrt{\pi_{1t}})\pi_{1t} = c \Leftrightarrow F_{w_t|1}^{-1} \circ F_{w_t|1}(x\sqrt{\pi_{1t}}) = F_{w_t|1}^{-1}(c/\pi_{1t})$. Since $F_{w_t|1}$ is assumed to be strictly increasing, we have $F_{w_t|1}^{-1} \circ F_{w_t|1}(x) = x$ by Lemma A.1 (d). So $x = \pi_{1t}^{-1/2}F_{w_t|1}^{-1}(c/\pi_{1t})$.

2. For $x \in (0, \infty)$, then it follows from the expression of $F_{z_t}(x)$ in Proposition 2.2 that $c \geq F_{w_t|1}(0)\pi_{1,t} + \pi_{0,t}$. We search for the solution x to $F_{z_t}(x) = F_{w_t|1}(c)\pi_{1,t} + \pi_{0,t} \Leftrightarrow F_{w_t|1}(x\sqrt{\pi_{1,t}}) = (c - \pi_{0,t})/\pi_{1,t} \Leftrightarrow F_{w_t|1}^{-1}F_{w_t|1}(x\sqrt{\pi_{1,t}}) = F_{w_t|1}^{-1}[(c - \pi_{0,t})/\pi_{1,t}]$. Since $F_{w_t|1}$ is assumed to be strictly increasing, we have $F_{w_t|1}^{-1} \circ F_{w_t|1}(x) = x$ by Lemma A.1 (d). So $x = \pi_{1,t}^{-1/2}F_{w_t|1}^{-1}[(c - \pi_{0,t})/\pi_{1,t}]$.

3. For $F_{w_t|1}(0)\pi_{1,t} \leq c < F_{w_t|1}(0)\pi_{1,t} + \pi_{0,t}$, then there is no solution x to $F_{z_t}(x) = c$. In this region, the generalised inverse is by definition equal to the smallest value x such that $F_{z_t}(x)$ is more than or equal to c, see equation (9). Since $F_{z_t}(x)$ makes this jump at $x = 0$ and is therefore never equal to c, we get that $F_{z_t}^{-1}(c) = 0$ which is the smallest possible choice of x so that $F_{z_t}(x) \geq c$.

Relying on Assumption 3(b) instead of Assumption 3(a), and replacing w_t with $\widetilde{r}_t$ and z_t with r_t, gives (11).

A.4   Proof of Proposition 2.4

Due to Assumptions 1 and 4 we have

$$
\begin{aligned}
F_{\widetilde{r}_t|1}(x) &= P_{t-1}(\widetilde{r}_t \leq x | I_t = 1) \\
&= P_{t-1}(\sigma_t w_t \leq x | I_t = 1) \\
&= P_{t-1}(w_t \leq x\sigma_t^{-1} | I_t = 1) \\
&\overset{(4)}{=} F_{w_t|1}(x\sigma_t^{-1}),
\end{aligned}
$$

where (4) indicates where we have used Assumption 4. Both $F_{w_t|1}$ and $F_{\widetilde{r}_t|1}$ are assumed strictly increasing in Assumption 3, so both $F_{w_t|1}$ and $F_{\widetilde{r}_t|1}$ are invertible. Denote $y = F_{\widetilde{r}_t|1}(x)$, so that $F_{\widetilde{r}_t|1}^{-1}(y) = x$. Since $F_{\widetilde{r}_t|1}(x) = F_{w_t|1}(x\sigma_t^{-1})$, this means $y = F_{w_t|1}(x\sigma_t^{-1})$, and hence $F_{w_t|1}^{-1}(y) = x\sigma_t^{-1}$. Substituting for x (we have that $x = F_{\widetilde{r}_t|1}^{-1}(y)$) in this expression and re-arranging, gives

$$
F_{\widetilde{r}_t|1}^{-1}(y) = \sigma_t F_{w_t|1}^{-1}(y).
$$

From this it follows that (11) can be re-written as

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

That is, $r_{c,t} = \sigma_t z_{c,t}$.

A.5 Proof of Proposition 2.5

In deriving the expression for $E_{t-1}(z_t | z_t \leq z_{c,t})$ we start by showing that $x_c(c - F_X(x_c))$ in (12) is indeed equal to zero for z_t:

Lemma A.2 *If Assumptions 1(a), 3(a) and 5(a) hold, then $z_{c,t}(c - F_{z_t}(z_{c,t})) = 0$.*

Proof. (a) and (b) in Lemma A.1 imply that $P_{t-1}(z_t \leq F_{z_t}^{-1}(c)) = P_{t-1}(F_{z_t}^{-1}(\xi) \leq F_{z_t}^{-1}(c)) = P_{t-1}(\xi \leq F_{z_t} \circ F_{z_t}^{-1}(c))$. Next, since $\xi \sim U[0,1]$, we have that $P_{t-1}(\xi \leq x) = x 1_{\{0 \leq x \leq 1\}} + 1_{\{x > 1\}}$. Since $0 \leq F_{z_t} \leq 1$ we get $P_{t-1}(\xi \leq F_{z_t} \circ F_{z_t}^{-1}(c)) = F_{z_t} \circ F_{z_t}^{-1}(c)$. Hence we are left with computing $F_{z_t} \circ F_{z_t}^{-1}(c)$:

Case 1. If $c \in [0, F_{w_t|1}(0)\pi_{1t}) \cup [F_{w_t|1}(0)\pi_{1t} + \pi_{0t}, \infty)$, which is the range of F_{z_t} by Proposition 2.2 and Assumption 5, then $F_{z_t} \circ F_{z_t}^{-1}(c) = c$ by (c) in Lemma A.1. So $F_{z_t}^{-1}(c)[c - P_{t-1}(z_t \leq F_{z_t}^{-1}(c))] = 0$.

Case 2. If on the other hand $F_{w_t|1}(0)\pi_{1t} \leq c < F_{w_t|1}(0)\pi_{1t} + \pi_{0t}$, then $F_{z_t}^{-1}(c) = 0$ by Proposition 2.2, so $F_{z_t}^{-1}(c)[c - P_{t-1}(z_t \leq F_{z_t}^{-1}(c))] = 0$.

We now turn to the three cases in (13):

Case 1: $c < F_{w_t|1}(0)\pi_{1t}$. In this case $F_{z_t}^{-1}(c) = \pi_{1t}^{-1/2} F_{w_t|1}^{-1}(c/\pi_{1t})$ according to Proposition 2.3, and so

$$E(z_t 1_{\{z_t \leq F_{z_t}^{-1}(c)\}}) = \int_A x \, dF_{z_t}(x), \qquad A = (-\infty, \pi_{1t}^{-1/2} F_{w_t|1}^{-1}[c/\pi_{1t}]).$$

Because $c < F_{w_t|1}(0)\pi_{1t}$ and $F_{z_t}^{-1}$ is a non-decreasing function, we have that $F_{z_t}^{-1}(c) < F_{z_t}^{-1}[F_{w_t|1}(0)\pi_{1t}] = 0$. Hence, the area we integrate over only includes negative numbers. In this region

$$F_{z_t}(x) = \pi_{1t} F_{w_t|1}(x\sqrt{\pi_{1t}}) + 1_{\{0 \leq x\}} \pi_{0t} = \pi_{1t} F_{w_t|1}(x\sqrt{\pi_{1t}})$$

with derivative equal to $\pi_{1t}^{3/2} f_{w_t|1}(x\sqrt{\pi_{1t}})$ by Assumption 5. So

$$E(z_t 1_{\{z_t \leq F_{z_t}^{-1}(c)\}}) = \pi_{1t}^{3/2} \int_A x f_{w_t|1}(x\sqrt{\pi_{1t}}) \, dx.$$

Letting $u = x\sqrt{\pi_{1t}}$ so that $x = u/\sqrt{\pi_{1t}}$ gives $dx = du/\sqrt{\pi_{1t}}$, and the area of integration is changed to $(-\infty, F_{w_t|1}^{-1}[c/\pi_{1t}])$ because, for the function $u(x) = x\sqrt{\pi_{1t}}$, we have $u(-\infty) = -\infty$ and $u(\pi_{1t}^{-1/2} F_{w_t|1}^{-1}[c/\pi_{1t}]) =$

$\pi_{1t}^{-1/2}F_{w_t|1}^{-1}[c/\pi_{1t}\sqrt{pi_{1t}}] = F_{w_t|1}^{-1}[c/\pi_{1t}]$. This gives

$$E(z_t 1_{\{z_t \le F_{z_t}^{-1}(c)\}}) = \pi_{1t}^{3/2} \int_{-\infty}^{F_{w_t|1}^{-1}[c/\pi_{1t}]} (u/\sqrt{\pi_{1t}}) f_{w_t|1}(u)\, du/\sqrt{\pi_{1t}}$$

$$= \sqrt{\pi_{1t}} \int_{-\infty}^{F_{w_t|1}^{-1}(c/\pi_{1t})} u f_{w_t|1}(u)\, du$$

$$= \sqrt{\pi_{1t}} E(w_t 1_{\{w_t \le F_{w_t|1}^{-1}(c/\pi_{1t})\}}).$$

Case 2: $F_{w_t|1}(0)\pi_{1t} \le c < F_{w_t|1}(0)\pi_{1t} + \pi_{0t}$. In this case $E(z_t 1_{\{z_t \le F_{z_t}^{-1}(c)\}}) = E(z_t 1_{\{z_t \le 0\}})$ according to Proposition 2.3, and so

$$E(z_t 1_{\{z_t \le 0\}}) = \int_{-\infty}^0 x\, dF_{z_t}(x) = \int_{-\infty}^0 x\, d[\pi_{1t} F_{w_t}(x\sqrt{\pi_{1t}})] + \int_{-\infty}^0 x\, d[\pi_{0t} 1_{\{0 \le x\}}].$$

We have $\int_{-\infty}^0 x\, d[\pi_{0t} 1_{\{0 \le x\}}] = \pi_{0t} \int_{\mathbb{R}} 1_{\{x \le 0\}} x\, d1_{\{0 \le x\}} = \pi_{0t} 1_{\{x \le 0\}} x|_{x=0} = 0$, since $1_{\{0 \le x\}}$ is the cumulative distribution function of a (degenerate) random variable Z with $P(Z = 0) = 1$. We therefore get that $E(z_t 1_{\{z_t \le 0\}}) = \int_{-\infty}^0 x\, d[\pi_{1t} F_{w_t}(x\sqrt{\pi_{1t}})]$, which equals $\sqrt{\pi_{1t}} E(w_t 1_{\{w_t \le 0\}})$ by means of the same sort of calculations as in case 1.

Case 3: $c \ge F_{w_t|1}(0)\pi_{1t} + \pi_{0t}$. In this case $E(z_t 1_{\{z_t \le F_{z_t}^{-1}(c)\}}) = E(z_t 1_{\{z_t \le \pi_{1t}^{-1/2} F_{w_t|1}^{-1}[(c-\pi_{0t})/\pi_{1t}]\}})$ according to Proposition 2.3. Let $B := (-\infty, \pi_{1t}^{-1/2} F_{w_t|1}^{-1}[(c-\pi_{0t})/\pi_{1t}])$. As in case 2, we use the linearity of the Lebesgue-Stieltjes integral in terms of its measure to see that

$$E(z_t 1_{\{z_t \le F_{z_t}^{-1}(c)\}}) = \int_B x\, dF_{z_t}(x) = \int_B x\, d[\pi_{1t} F_{w_t|1}(x\sqrt{\pi_{1t}})] + \int_B x\, d[\pi_{0t} 1_{\{0 \le x\}}].$$

The integral from the discrete component is computed as in case 2, and we see that

$$\int_B x\, d[\pi_{0t} 1_{\{0 \le x\}}] = \pi_{0t} \int_{\mathbb{R}} 1_{\{x \in B\}} x\, d1_{\{0 \le x\}} = \pi_{0t} 1_{\{x \in B\}} x|_{x=0} = 0.$$

As in case 1 we see that

$$\int_B x\, d[\pi_{1t} F_{z_t}(x\sqrt{\pi_{1t}})] = \pi_{1t}^{3/2} \int_B x f_{w_t|1}(x\sqrt{\pi_{1t}})\, dx = \sqrt{\pi_{1t}} E\left(w_t 1_{\{w_t \le F_{w_t|1}^{-1}[(c-\pi_{0t})/\pi_{1t}]\}}\right).$$

Relying on Assumptions 1(b), 3(b) and 5(b) instead of 1(a), 3(a) and 5(a), and replacing w_t with $\widetilde{r}_t$ and z_t with r_t, gives (14).

A.6   Proof of Proposition 2.6

From the measurability of σ_t with respect to $\mathcal{F}_{t-1}^r$ (i.e. Assumption 4) it follows that $E_{t-1}(\widetilde{r}_t 1_A) = \sigma_t E_{t-1}(w_t 1_A)$, where A denotes an event. Denote $y = F_{\widetilde{r}_t|1}(x)$, so that $F_{\widetilde{r}_t|1}^{-1}(y) = x$. From the proof of Proposition 2.4 in

Appendix A.4 it follows that $F_{\widetilde{r}_t|1}(x) = F_{w_t|1}(x\sigma_t^{-1})$ and $F_{\widetilde{r}_t|1}^{-1}(y) = \sigma_t F_{w_t|1}^{-1}(y)$. Accordingly, we can re-write (14) as

$$
\begin{aligned}
&E_{t-1}(r_t|r_t \leq r_{c,t}) \\
&= \sigma_t \begin{cases} \pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w|1}^{-1}(c/\pi_{1t})\}}\right) & \text{if } c < F_{w_t|1}(0)\pi_{1t}, \\[2mm] \pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq 0\}}\right) & \text{if } F_{w|1}(0)\pi_{1t} \leq c < F_{w_t|1}(0)\pi_{1t} + \pi_{0t}, \\[2mm] \pi_{1t}^{1/2} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w|1}^{-1}[(c-\pi_{0t})/\pi_{1t}]\}}\right) & \text{if } c \geq F_{w_t|1}(0)\pi_{1t} + \pi_{0t}. \end{cases}
\end{aligned}
$$

That is, $E_{t-1}(r_t|r_t \leq r_{c,t}) = \sigma_t E_{t-1}(z_t|z_t \leq z_{c,t})$.

B   Missing values estimation algorithm

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 $t = 1, \ldots, T$:

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

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 non-zero locations, and let T^* denote the number of non-zero returns. The kth. iteration now proceeds as follows:

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

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

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

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 that $\overline{r^2}_t$ 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 that 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 T^* rather than T.

QML Estimation of the log-GARCH model is via its ARMA-representation, see Sucarrat and Escribano (2017). If $|E(\ln w_t^2)| < \infty$, then the ARMA(1,1) representation is given by

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

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. Accordingly, subject to suitable assumptions, 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 an estimate of $E(\ln w_t^2)$ is needed. Sucarrat et al. (2016) show that, under very general assumptions, the formula $-\ln\left[T^{-1}\sum_{t=1}^n \exp(\widehat{u}_t)\right]$ provides a consistent estimate (see also Francq and Sucarrat (2017)). To accommodate the missing values, this formula is modified to $-\ln\left[T^{*-1}\sum_{t\in G} \exp(\widehat{u}_t)\right]$.

In order to study the finite sample bias of the algorithm, we undertake a simulation study. In the simulations the Data Generating Process (DGP) of return is given by

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

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

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

The term $t^* = t/T$ 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

$$
\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 \widetilde{r}_{t-1}^2 + \sigma_{t-1}^2, \\
\ln \sigma_t^2 &= \alpha_0 + \alpha_1 \ln \widetilde{r}_{t-1}^2 + \ln \sigma_{t-1}^2,
\end{aligned}
$$

with $(\alpha_0, \alpha_1, \beta_1) = (0.02, 0.1, 0.8)$ in each. We compare two estimation approaches. In the first, which we label "Ordinary", $\widetilde{r}_t^2$ is replaced by 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. (2016). The second estimation approach, which we label "Algorithm", uses the missing value algorithm as described in the above. Figure 6 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 (i.e. the second estimation approach), whereas a dotted red line stands for the bias of ordinary Gaussian QML

estimation *without* zero-adjustment (i.e. the first estimation approach). The figure confirms that the algorithm provides approximately unbiased estimates in finite samples in the presence of missing values, and that the bias is increasing in the zero-probability. 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.

Table 1: Descriptive statistics, logit models and GARCH-models of the daily returns of three NYSE-listed stocks (see Section 3)

Descriptive statistics:

	Sample	Volume	s^2	s^4	ARCH $[p-val]$	T	$0s$	$\widehat{\pi}_0$
GE	3/1/2007–31/12/2014	67.75	4.55	12.57	154.1 [0.00]	2013	30	0.015
VG	3/1/2007–31/12/2014	2.20	32.00	75.21	40.05 [0.00]	2013	148	0.074
BKT	3/1/2007–31/12/2014	0.176	0.621	21.45	31.81 [0.00]	2013	258	0.128

Logit models:

		$\widehat{\rho}_0$ (s.e.)	$\widehat{\rho}_1$ (s.e.)	$\widehat{\zeta}_1$ (s.e.)	$\widehat{\lambda}_1$ (s.e.)	SIC	$Logl$
GE	Constant:	4.191 (0.184)				0.1587	-155.961
	ACL(1,1):	3.315 (2.418)	-2.624 (6.331)	0.278 (0.404)		0.1649	-154.574
	Trend:	4.736 (0.421)			-1.008 (0.653)	0.1613	-154.739
VG	Constant:	2.534 (0.085)				0.5291	-528.726
	ACL(1,1):	0.756 (0.275)	0.270 (0.054)	0.710 (0.106)		0.5222	-514.163
	Trend:	2.585 (0.173)			-0.102 (0.296)	0.5328	-528.667
BKT	Constant:	1.917 (0.067)				0.7696	-770.752
	ACL(1,1):	0.127 (0.120)	0.070 (0.040)	0.934 (0.062)		0.7729	-766.476
	Trend:	2.393 (0.147)			-0.901 (0.235)	0.7659	-763.281

GARCH models:

		$\widehat{\alpha}_0$ (s.e.)	$\widehat{\alpha}_1$ (s.e.)	$\widehat{\beta}_1$ (s.e.)
GE	Ordinary:	0.024 (0.012)	0.066 (0.016)	0.925 (0.017)
VG	Ordinary:	1.031 (0.563)	0.190 (0.071)	0.795 (0.071)
BKT	Ordinary:	0.029 (0.011)	0.144 (0.035)	0.798 (0.049)
	0-adjusted:	0.024 (0.008)	0.148 (0.030)	0.804 (0.041)

GE, the ticker of General Electric. VG, the ticker of Vonage Holdings Corporation. BKT, the ticker of The Bank of New York Mellon Corporation. Volume, average daily trading volume in millions of USD over the sample. s^2, sample variance of return. s^4, sample kurtosis of return. 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. T, number of observations before differencing and lagging. $0s$, 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. Datasource: Bloomberg. All computations in R (R Core Team, 2018).

Table 2: In-sample descriptives of logit-models (see Section 4)

	n	$\mathrm{avg}(\widehat{\pi}_{0i})$	$\max \widehat{\pi}_{0i}$	$\min \widehat{\pi}_{0i}$	$\mathrm{avg}(vol_i)$	$\rho(\widehat{\pi}_{0i}, vol_i)$
All	1665	0.0211	0.1931	0.0000	1.822	-0.14
Constant	1259	0.0188	0.1311	0.0000	1.907	-0.13
ACL(1,1)	228	0.0259	0.1931	0.0015	1.580	-0.21
Trend	178	0.0317	0.1282	0.0030	1.533	-0.28

n, number of stocks. $\widehat{\pi}_{0i}$, stock i's proportion of zero returns. $\mathrm{avg}(\widehat{\pi}_{0i})$, average of the $\widehat{\pi}_{0i}$'s. $\max \widehat{\pi}_{0i}$, the largest zero-proportion across stocks. $\min \widehat{\pi}_{0i}$, the smallest zero-proportion across stocks. vol_i, stock i's daily average volume in million USD. $\mathrm{avg}(vol_i)$, average of the vol_i's. $\rho(\widehat{\pi}_{0i}, vol_i)$, sample correlation between $\widehat{\pi}_{0i}$ and vol_i.

Table 3: Out-of-sample ME and MAE results (see Section 4)

Volatility:

	n	$avg.$	$max.$	$min.$	$n(0.10)$	$n(0.05)$	$n(0.01)$
ME	178	-0.059	2.832	-1.686	149	140	127
MAE	178	0.302	4.092	0.008	175	175	173

97.5% VaR:

	n	$avg.$	$max.$	$min.$	$n(0.10)$	$n(0.05)$	$n(0.01)$
ME	406	0.004	0.434	-0.241	328	307	269
MAE	406	0.050	0.692	0.000	255	254	248

97.5% ES:

	n	$avg.$	$max.$	$min.$	$n(0.10)$	$n(0.05)$	$n(0.01)$
ME	406	0.004	0.691	-0.340	255	232	205
MAE	406	0.074	0.926	0.001	328	245	70

n, number of stocks. $avg.$, the average of the MEs or MAEs across stocks. $max.$, the maximum ME or MAE across the stocks. $min.$, the minimum ME or MAE across the stocks. $n(\alpha)$, the number of rejections of H_0 at significance level α. The tests are implemented via OLS estimated regressions with Newey and West (1987) standard error. For ME, $x_{it} = \mu_i + u_{it}$ with $H_0 : \mu_i = 0$ and $H_A : \mu_i \neq 0$. For MAE, $|x_{it}| = \mu_i + u_{it}$ with $H_0 : \mu_i = 0.01$ and $H_A : \mu_i > 0.01$.

Table 4: Coverage and independence tests of out-of-sample VaR forecasts (see Section 4)

97.5% VaR:

	n	Unconditional coverage			Independence		
		$n(0.10)$	$n(0.05)$	$n(0.01)$	$n(0.10)$	$n(0.05)$	$n(0.01)$
Ordinary	406	62	36	14	398	397	397
0-adjusted	406	67	44	13	398	397	397

The tests are those of Christoffersen (1998). n, number of stocks. $n(\alpha)$, the number of rejections of H_0 at significance level α.
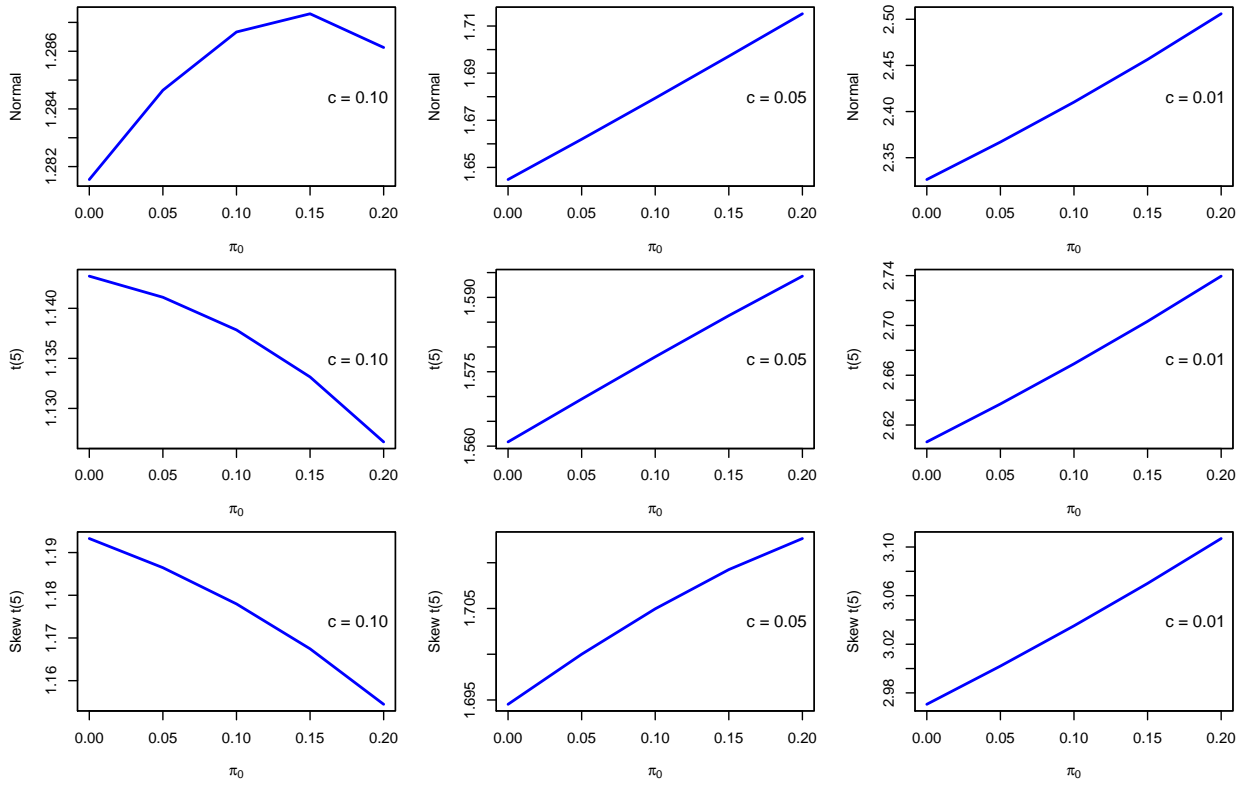
Figure 1: VaR of z_t, i.e. $-z_{c,t}$, where $z_{c,t}$ is given by (10), for different values of π_{0t} and c, and for different densities of w_t, see Section 2.3
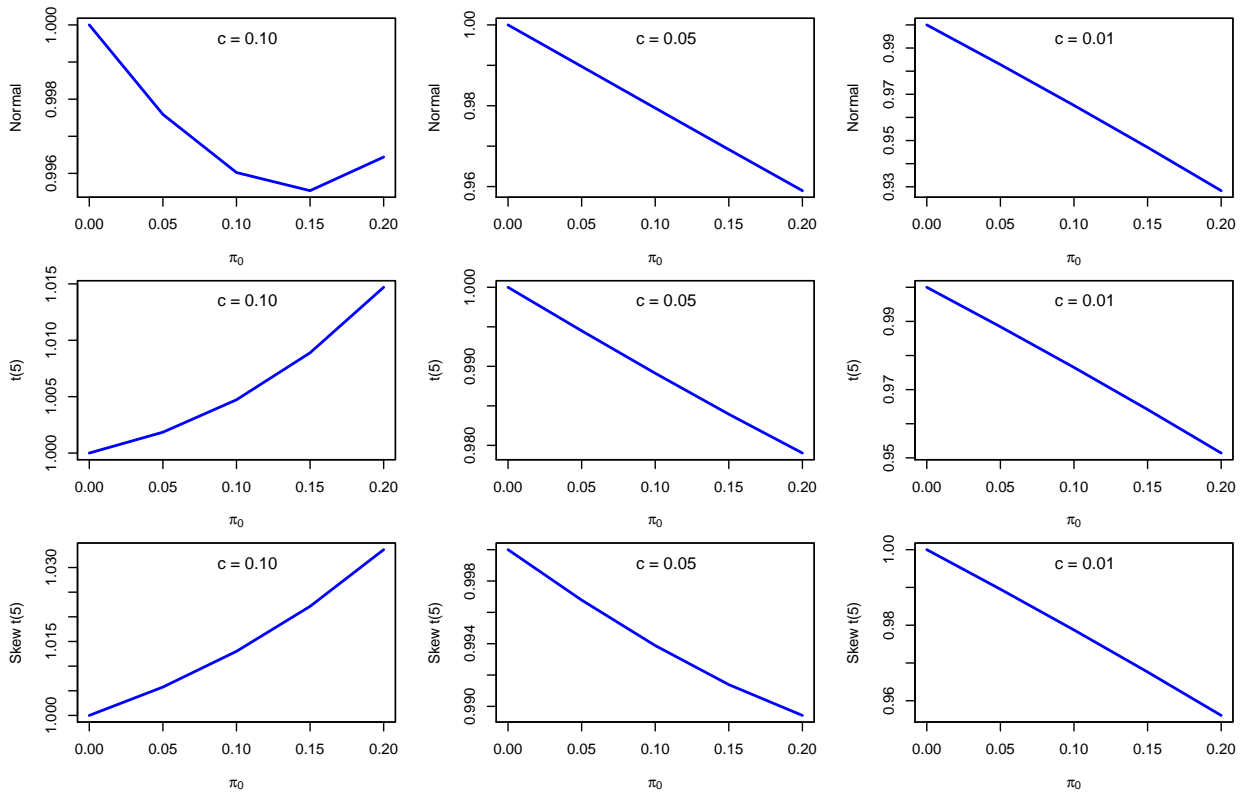

Figure 2: Ratios of VaRs (computed as $-w_{c,t}/-z_{c,t}$ where $w_{c,t}$ is the cth. quantile of w_t) for different values of π_{0t} and c, and for different densities of w_t, see Section 2.3
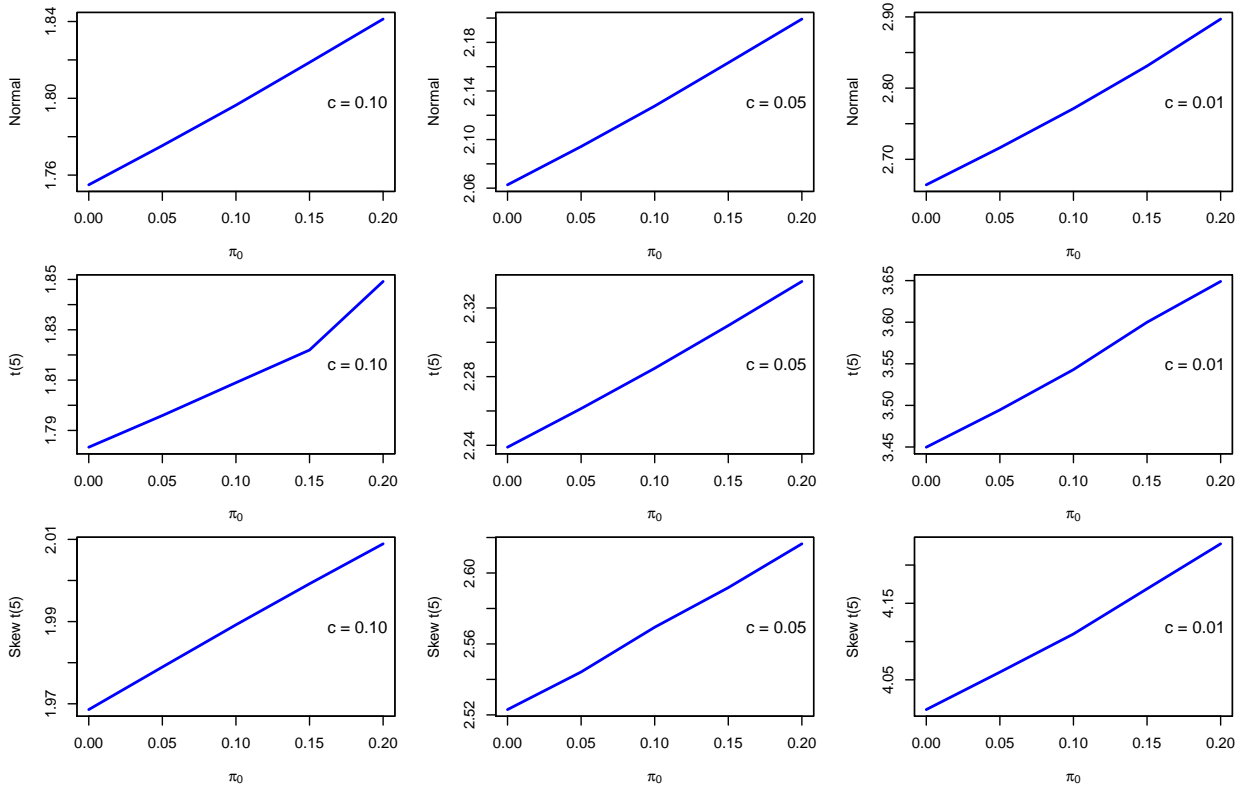
Figure 3: ES of z_t, i.e. $-c^{-1}E_{t-1}(z_t|z_t \leq z_{c,t})$, for different values of π_{0t} and c, and for different densities of w_t, see Section 2.4
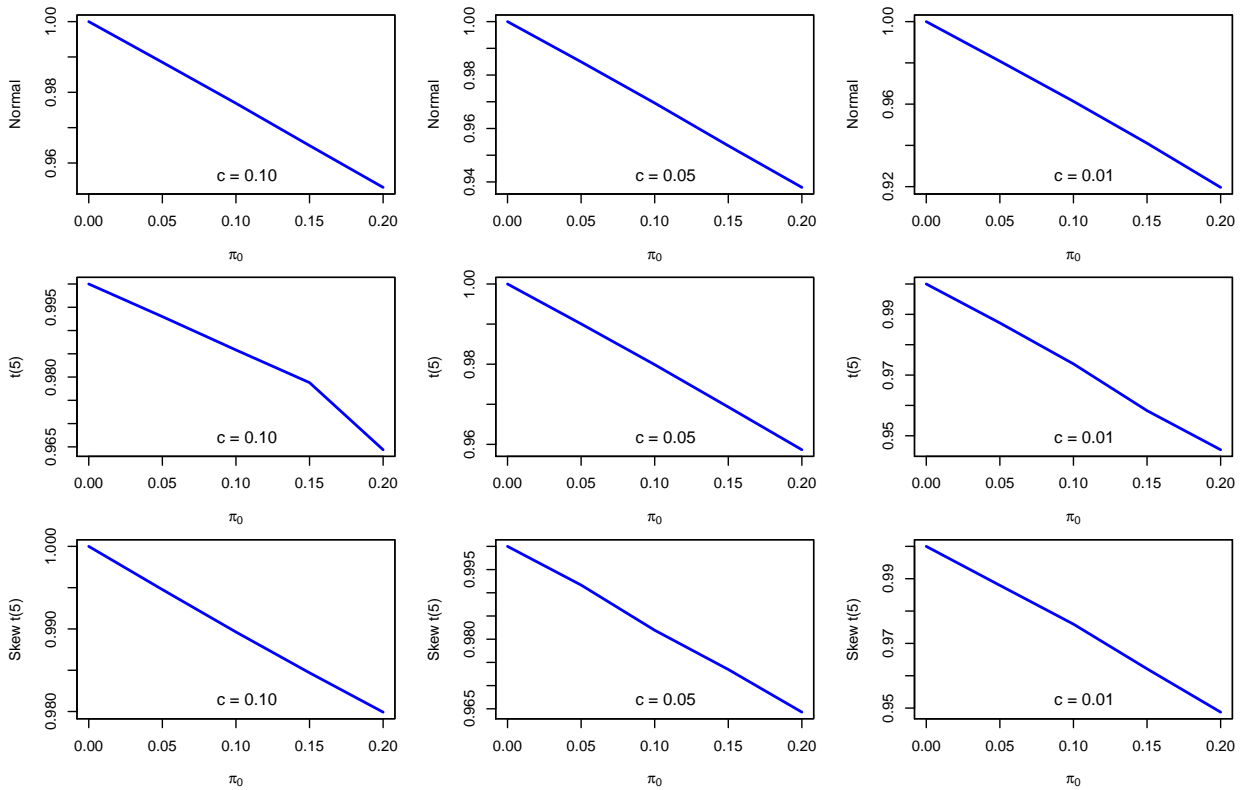

Figure 4: Ratios of ESs ($-c^{-1}E_{t-1}(w_t|w_t \leq w_{c,t})$ in the numerator, $-c^{-1}E_{t-1}(z_t|z_t \leq z_{c,t})$ in the denominator) for different values of π_{0t} and c, and for different densities of w_t, see Section 2.4
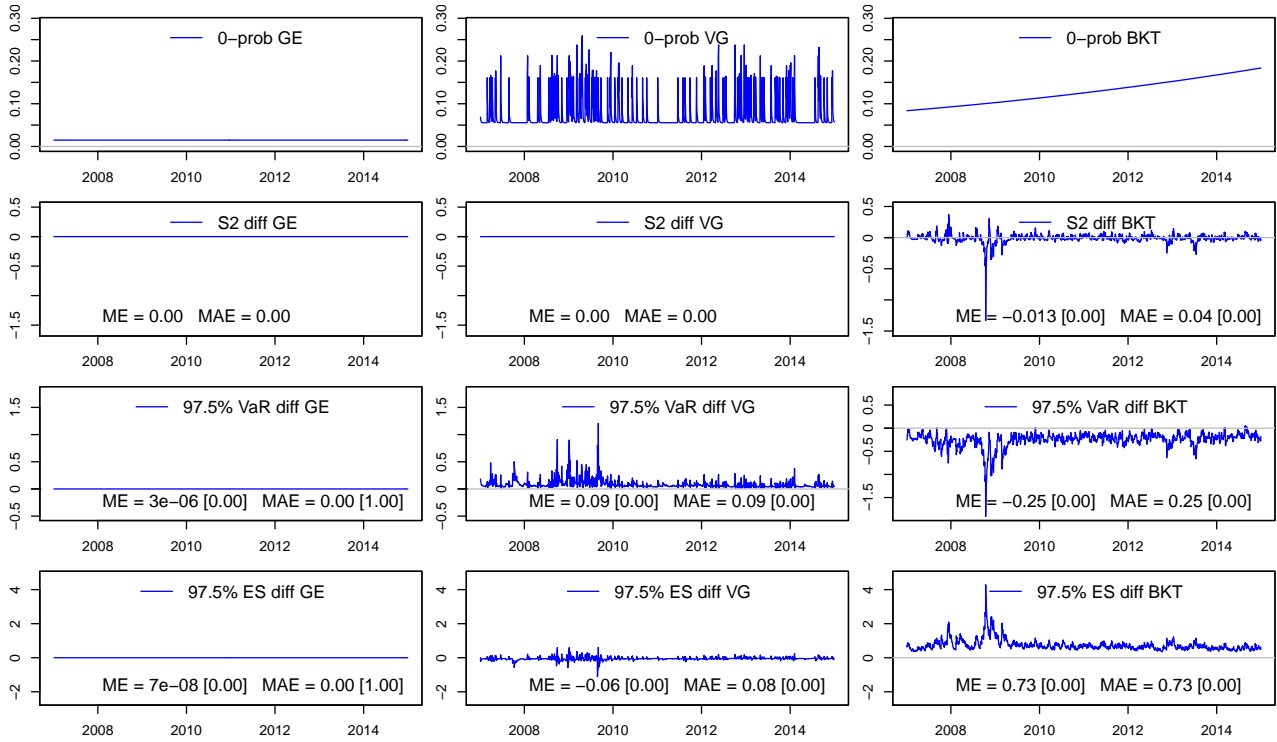
Figure 5: Fitted zero-probabilities (0-prob), and the differences between fitted σ_t^2, 97.5% VaR and 97.5% ES (see Section 3). The difference or error x_t at t is computed as the zero-corrected risk-estimate minus the incorrect one. The Mean Error (ME) is computed as $T^{-1} \sum_{t=1}^{T} x_t$, and the Mean Absolute Error (MAE) is computed as $T^{-1} \sum_{t=1}^{T} |x_t|$. For ME, the p-value in square brackets is from a test implemented via the OLS estimated regression $x_t = \mu + u_t$ with $H_0 : \mu = 0$ and $H_A : \mu \neq 0$. The t-distributed test statistic is $\widehat{\mu}/se(\widehat{\mu})$, where $se(\widehat{\mu})$ is the standard error of Newey and West (1987). For MAE, the p-value in square brackets is from a test implemented via the OLS estimated regression $|x_t| = \mu + u_t$ with $H_0 : \mu = 0.01$ and $H_A : \mu > 0.01$. The t-distributed test statistic is $(\widehat{\mu} - 0.01)/se(\widehat{\mu})$, where $se(\widehat{\mu})$ is the standard error of Newey and West (1987).
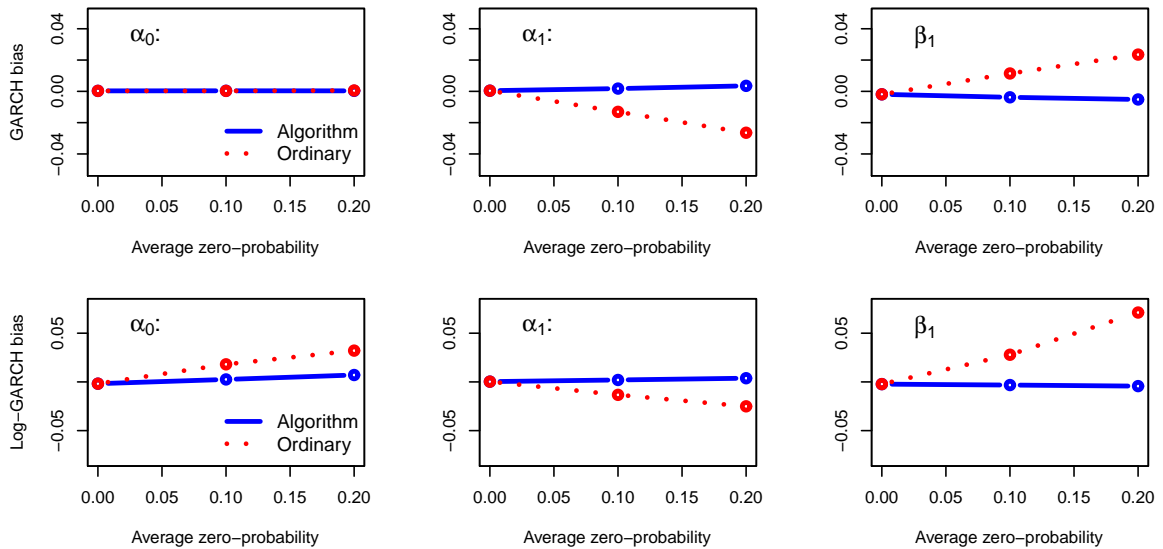

Figure 6: Simulated parameter biases in GARCH(1,1) and log-GARCH(1,1) models for the missing values algorithm in comparison with ordinary methods (see Appendix B)