



Munich Personal RePEc Archive

Risk Estimation when the Zero Probability of Financial Return is Time-Varying

Grønneberg, Steffen and Sucarrat, Genaro

BI Norwegian Business School, BI Norwegian Business School

29 October 2014

Online at <https://mpra.ub.uni-muenchen.de/81882/>
MPRA Paper No. 81882, posted 12 Oct 2017 18:37 UTC

Risk Estimation when the Zero Probability of Financial Return is Time-Varying*

Steffen Grønneberg[†] and Genaro Sucarrat[‡]

First version: 29 October 2014

This version: 10th October 2017

Abstract

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 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 risk (e.g. volatility, Value-at-Risk, Expected Shortfall), however, the zero probability is zero, constant or both. We propose a new class of models that allows for a time-varying zero probability, and which nests ordinary models as special cases. The properties (e.g. volatility, skewness, kurtosis, Value-at-Risk, Expected Shortfall) of the new class are obtained as functions of the underlying volatility and zero probability models. For a given volatility level, our results imply that risk estimates can be severely biased if zeros are not accommodated: For rare loss events (i.e. 5% or less) we find that Conditional Value-at-Risk is biased downwards and that Conditional Expected Shortfall is biased upwards. An empirical application illustrates our results, and shows that zero-adjusted risk estimates can differ substantially from risk estimates that are not adjusted for the zero probability.

JEL Classification: C01, C22, C32, C51, C52, C58

Keywords: Financial return, volatility, zero-inflated return, Value-at-Risk, Expected Shortfall

*We are grateful to Christian Conrad, Christian Francq, participants at the HeiKaMEtrics Conference 2017 (Heidelberg), VieCo 2017 conference (Vienna), the CFE 2016 conference (Seville), the CEQURA 2016 conference (Munich), the CATE September 2016 workshop (Oslo), the CORE 50th. anniversary conference (Louvain-la-Neuve), the Maastricht econometrics seminar (May, 2016), the Uppsala statistics seminar (April 2016), the CREST econometrics seminar (February 2016), the SNDE Annual Symposium 2015 (Oslo) and the IAAE Conference 2015 (Thessaloniki) for useful comments, suggestions and questions.

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

[‡]Corresponding author. Department of Economics, BI Norwegian Business School. Email: genaro.sucarrat@bi.no. Webpage: <http://www.sucarrat.net/>

Contents:

1	Introduction	2
2	Financial return with time-varying zero probability	4
2.1	The ordinary model of return	4
2.2	A model of return with time-varying zero probability	5
2.3	Conditional VaR	6
2.4	Conditional ES	9
2.5	Estimation	10
3	Empirical application	12
3.1	Models	12
3.2	Volatility	13
3.3	Conditional VaR	13
3.4	Conditional ES	14
4	Conclusions	14
	References	17
A	Proofs	18
A.1	Proof of Proposition 1	18
A.2	Proof of Proposition 2	19
A.3	Proof of Proposition 3	19
A.4	Proof of Proposition 4	20
A.5	Proof of Proposition 5	21
A.6	Proof of Proposition 6	22
B	Missing values estimation algorithm	22

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.

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 then 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, therefore, 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 may 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 the Autoregressive Conditional Heteroscedasticity (ARCH) class of models of [Engle \(1982\)](#), the Stochastic Volatility (SV) class of models (see [Shephard \(2005\)](#)) and continuous time models (e.g. Brownian motion).¹ 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 return multiplicatively into a continuous part, which can be specified in terms of common volatility models, and a discrete part at zero that is appropriately scaled by the zero probability. Standard volatility models (e.g. ARCH, SV and continuous time models) are therefore nested and obtained as special cases. [Hautsch et al. \(2013\)](#) proposed a model for volume that uses a similar decomposition 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 their specification as a special case: The dynamics need not be specified in logs, 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 – in principle – for autoregressive conditional dynamics in both the zero probability and volatility specifications, and for a two-way feedback between the two.

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 skewness and kurtosis of return, but reduces return variability defined as absolute return (see [Proposition 1](#)). Second, if the model and/or estimator used by the practitioner does not accommodate zeros appropriately, then volatility estimates may be severely biased in unpredictable ways. This is particularly the case if the zero probability is non-stationary. To alleviate this problem we outline an estimation and inference procedure that reduces the bias caused by a time-varying zero probability (possibly non-stationary), and which can be combined with well-known models and estimators (see [Section 2.5](#)). Third, we derive general formulas for Conditional Value-

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

at-Risk (VaR). For a given level of volatility, we find that risk – when defined as Conditional VaR – will be biased downwards for rare loss events (5% or less) if zeros are not adjusted for (see Section 2.3). Fourth, we derive general formulas for Conditional Expected Shortfall (ES). For a given level of volatility, we find that risk – when defined as Conditional ES – will be biased upwards – i.e. the *opposite* of Conditional VaR – for rare loss events (10% or less) if zeros are not adjusted for (see Section 2.4). This may have implications for financial market supervision, due to the increased emphasis on Expected Shortfall in the Basel III regulatory framework. Finally, an empirical illustration shows that risk estimates can be substantially biased in practice if the time-varying zero probability is not accommodated appropriately (see Section 3).

The rest of the paper is organised as follows. Section 2 presents the new model class and derives some general properties, including the formulas for Conditional VaR and Conditional ES. The section ends by outlining a general estimation and inference procedure that reduces the biases caused by zeros, and which can be combined with common models and methods. Section 3 contains our empirical application, whereas Section 4 concludes. The Appendix contains the proofs and additional auxiliary material. Tables and Figures are located at the end.

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

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

where $\sigma_t > 0$ is a time-varying scale or volatility (that needs not 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

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

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

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

with v_i being independent of w_j for all pairs i, j . Other examples of σ_t include quadratic variation (e.g. Brownian motion) 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,² 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 process governed by

$$r_t = \sigma_t z_t, \quad \sigma_t > 0, \quad t \in \mathbb{Z}, \quad (4)$$

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

$$I_t \in \{0, 1\}, \quad \pi_{1t} = P_{t-1}(I_t = 1), \quad 0 < \pi_{1t} \leq 1. \quad (6)$$

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: $r_t \neq 0$ 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 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 1](#) below). It should be underlined 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 $|w_t|$ increases the probability of I_t being zero. A specific example is the situation where $I_t = 1$ if $|w_t| < 0.05$ and 0 otherwise, and where w_t conditional on the past is standard normal (in this specific example $\pi_{1t} = 0.96$). It should also be underlined 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 $\tilde{r}_t = \sigma_t w_t$ as “zero-adjusted” return, since $\tilde{r}_t = r_t \pi_{1t}^{1/2}$ whenever $I_t \neq 0$.

An attractive feature of (4)–(6) is that many properties can be expressed as a function of the underlying models of volatility and 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 \tilde{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 that ensures probabilities conditional on the past can be manipulated as usual, see [Shiryaev \(1996\)](#), pp. 226-227). (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

²The GED, which stands for Generalised Error Distribution, is also known as the Exponential Power distribution.

that z_t exhibits the first and second moment properties typically possessed by the scaled innovation in volatility models. The assumption can thus be viewed as an identification condition. The zero-mean property will usually ensure that returns are Martingale Difference Sequences (MDSs), and most commonly $\sigma_w^2 = 1$, as in the ARCH class of models. It should be noted, however, that Assumption 2 is only needed in Proposition 1. This proposition collects some properties that follow from (4) – (6) together with some additional moment assumptions.

Proposition 1. *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 $\text{Var}_{t-1}(z_t) = \sigma_w^2$ for all t , and $\{z_t\}$ is covariance-stationary with $E(z_t) = 0$, $\text{Var}(z_t) = \sigma_w^2$ and $\text{Cov}(z_t, z_{t-j}) = 0$ 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 $\{z_t\}$ is an MDS even if π_{1t} is time-varying. Indeed, it remains an MDS even if $\{I_t\}$ is non-stationary. Usually, (i) will imply that $\{r_t\}$ is also an MDS, e.g. in the ARCH class of models, since there $E_{t-1}(r_t) = \sigma_t E_{t-1}(z_t)$. Property (ii) means 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 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 $\text{Var}_{t-1}(r_t) = \sigma_t^2$ and $\text{Var}(r_t) = \alpha_0 / (1 - \alpha_1 - \beta_1)$ regardless of whether π_{1t} is constant or time-varying. If z_t is not strictly stationary, e.g. because the zero probability is periodic (this is common in intraday returns), then Property (ii) means z_t will still be covariance stationary. 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 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. For a given volatility level σ_t , this means the conditional absolute return (i.e. $s = 1$) is scaled downwards, since $E|x|^s < E|x|^2$ for $0 < s < 2$ due to the Lyapounov inequality.

2.3 Conditional VaR

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

Proposition 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}_{t-1}^r is

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

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

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

Proof: See Appendix A.2.

Natural examples of $F_{w_t|1}$ and $F_{\tilde{r}_t|1}$, respectively, are $N(0, 1)$ 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}_{t-1}^r , then its lower c -quantile with $c \in (0, 1)$ is given by

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

We will write $F_{X_t}^{-1}(c) = X_{c,t}$ even though the inverse of F_X does not exist, and we will refer to $F_{X_t}^{-1}(c)$ 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 conditional VaRs, we introduce an additional, technical assumption on the distributions of w_t and \tilde{r} . The assumption can be relaxed, but at the cost of more complicated formulas.

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

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

(b) *The cdf of \tilde{r}_t , denoted $F_{\tilde{r}_t|1}$, 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 nor 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 will simplify the conditional quantile, VaR and ES expressions.

Proposition 3 (conditional 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 c th. quantile of z_t conditional on the past \mathcal{F}_{t-1}^r is*

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

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

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

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

and the $(100 \cdot c)\%$ Value-at-Risk (VaR_c) of r_t conditional on the past \mathcal{F}_{t-1}^r 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 estimation of σ_t and $z_{c,t}$ may be separated into two different steps. The following assumption, which is fulfilled by most ARCH models but not necessarily by SV models, ensures $r_{c,t}$ can indeed be written as $\sigma_t z_{c,t}$.

Assumption 4. σ_t is measurable with respect to \mathcal{F}_{t-1}^r .

Proposition 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

It should be noted 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 Conditional VaR for a fixed value of volatility σ_t . Figure 1 plots Conditional VaR (i.e. $-z_{c,t}$) for different values of c and π_{0t} , and for four different densities of w_t : The standard normal, the standardised skew normal, the standardised Student's t with five degrees of freedom, and the standardised skew Student's t with five degrees of freedom.³ When $c \in \{0.05, 0.01\}$, then Conditional VaR always increases when the zero probability π_{0t} increases. By contrast, when $c = 0.10$ then Conditional VaR generally falls, the exception being when $w_t \sim N(0, 1)$. There, Conditional 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 conditional 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 Conditional VaR usually increases when the zero probability increases. In other words, if the estimation of Conditional VaR is not adjusted for the zero probability, then the estimate of risk – defined in terms of Conditional VaR – will be biased downwards. Figure 2 provides an insight into the relative size of the bias. The Figure contains plots of the ratio of the unadjusted Conditional VaR (numerator) versus the zero-adjusted Conditional VaR (denominator). The unadjusted Conditional VaRs are those of w_t (i.e. those of z_t under the assumption that $\pi_{1t} = 1$), whereas the zero-adjusted Conditional VaRs are those of Figure 1. The plot reveals

³The 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.

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 $c = 0.01$, then the largest effect on $z_{c,t}$ (and hence conditional VaR) occurs when w_t is normal and skew normal. That is, the most commonly used density assumption.

2.4 Conditional 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

$$ES_c = -\frac{1}{c} [E(X1_{\{X < X_c\}}) + X_c(c - F_X(X_c))]. \quad (12)$$

The last term in the definition, i.e. $X_c(c - F_X(X_c))$, 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}_{t-1}^r and $I_t = 1$:*

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

The assumption is mild in the sense that it is fulfilled in most of the empirical applications that compute VaR and ES. That the assumption indeed ensures that $X_c(c - F_X(X_c))$ is zero for both z_t and r_t , is showed in [Appendix A.5](#) (see [Lemma 2](#)).

Proposition 5. *Suppose (4) – (6) hold and that $c \in (0, 1)$:*

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

$$E_{t-1}(z_t|z_t \leq z_{c,t}) = \begin{cases} \pi_{1t}E_{t-1}\left(w_t1_{\{w_t \leq F_{w|1}^{-1}(c/\pi_{1t})\}}\right) & \text{if } c < F_{w|1}(0)\pi_{1t}, \\ \pi_{1t}E_{t-1}\left(w_t1_{\{w_t \leq 0\}}\right) & \text{if } F_{w|1}(0)\pi_{1t} \leq c < F_{w|1}(0)\pi_{1t} + \pi_{0t}, \\ \pi_{1t}E_{t-1}\left(w_t1_{\{w_t \leq F_{w|1}^{-1}[(c-\pi_{0t})/\pi_{1t}]\}}\right) & \text{if } c \geq F_{w|1}(0)\pi_{1t} + \pi_{0t}, \end{cases} \quad (13)$$

- (b) *If also Assumptions 1(b), 3(b) and 5(b) hold, then the $(100 \cdot c)\%$ Expected Shortfall*

(ES_c) of r_t conditional on the past \mathcal{F}_{t-1}^r is $-c^{-1}E_{t-1}(r_t|r_t \leq r_{c,t})$, where

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

Proof: See Appendix A.5.

Just as for the expression for the quantile $r_{c,t}$ in Proposition 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 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 the assumptions stated so far, apart from Assumption 2, then we can indeed write the expression in this way.

Proposition 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 , Conditional ES is determined by $-\pi_{1t}c^{-1}E_{t-1}(z_t|z_t \leq z_{c,t})$. Figure 3 plots this expression for different values of c and π_{0t} , and for different conditional densities of w_t (the same as those for Conditional VaR above). Contrary to the Conditional VaR case, here the effect is always monotonous for $c \in \{0.10, 0.05, 0.01\}$ albeit opposite to that of Conditional VaR: Conditional ES falls as the zero probability increases. In other words, risk – defined as Conditional ES – will be biased upwards if not adjusted for the zero probability. Figure 4 provides insight into the magnitude of bias in relative terms. The plots contain the ratio of unadjusted Conditional ES (numerator) versus the zero-adjusted Conditional ES (denominator). The unadjusted ones are computed under the assumption that $\pi_{1t} = 1$, whereas the zero-adjusted ones are those of Figure 3. The plot reveals that, in relative terms, the largest effect occurs when $c = 0.10$ and w_t is skew t . Also, contrary to the Conditional VaR case, the normal and skew normal densities produce the smallest biases in relative terms.

2.5 Estimation

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

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

since Gaussian QML then provides strongly consistent and asymptotically normal estimates of α_0 , α_1 and β_1 . Escanciano (2009) and Francq and Thieu (2015) 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 co-variates (‘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.

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

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 [Francq and Zakoïan \(2017\)](#) show that 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 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-adjusted return $\tilde{r}_t = \sigma_t w_t$. For example, the GARCH(1,1) model in terms of zero-adjusted return is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \tilde{r}_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (17)$$

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

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \tilde{r}_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2. \quad (18)$$

If \tilde{r}_t were observed, then estimation could proceed as usual by, say, maximising $\sum_{t=1}^n \ln f_{\tilde{r}_t}(\tilde{r}_t)$, where $f_{\tilde{r}_t}$ is a suitably chosen density. The approximate estimation and inference procedure we propose consists of first replacing \tilde{r}_t with its estimate $r_t \hat{\pi}_{1t}^{1/2}$, 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 \tilde{r}_t by multiplying r_t with $\hat{\pi}_{1t}^{1/2}$, where $\hat{\pi}_{1t}$ is the fitted value of π_{1t} from Step 1. At zero locations the zero-adjusted return \tilde{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. 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 applies the algorithm to the daily Apple return series. 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 induces a repeated invertibility or irrelevance of initial values issue, see the discussion in [Sucarrat and Escribano \(2017\)](#).

3 Empirical application

In order to shed light on how returns with time-varying zero probabilities affect volatility dynamics, Value-at-Risk and Expected Shortfall in practice, we revisit three of the return series in [Sucarrat and Escribano \(2017\)](#). These series are of interest, since they exhibit a variety of zero-dynamics characteristics. The three series are the daily Standard and Poor’s 500 stock market index (SP500) return, the daily Ekornes stock price return and the daily Apple stock price return. The first and third return series are well-known, whereas the second is a leading Nordic furniture manufacturer listed on the Oslo Stock Exchange. Ekornes is a medium-sized company in international terms, since its market value is approximately 300 million euros (at the end of the series). Our interest in Ekornes is mainly due to its relatively large – for daily returns – proportion of zeros over the sample (about 19%). The source of the data is Yahoo Finance (<http://finance.yahoo.com>). All three returns are computed as $(\ln S_t - \ln S_{t-1}) \cdot 100$, where S_t is the index level or stock price at day t . Saturdays and Sundays, where returns are usually 0, are not included in our sample. Descriptive statistics are contained in the upper part of [Table 1](#). The statistics confirm that the returns exhibit the usual properties of excess kurtosis compared with the normal, and ARCH as measured by first order serial correlation in the squared return. The number of zeros varies from only 2 observations (about 0.1% of the sample) for SP500 to 667 observations (about 19% of the sample) for Ekornes.

3.1 Models

The middle part of [Table 1](#) contains estimates of three dynamic logit models for each return:

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

The conditional zero probability π_{0t} is thus given by $(1 - \pi_{1t})$ with $\pi_{1t} = 1/(1 + \exp(-h_t))$. In the first model the zero probability is constant, in the second it is governed by a deterministic trend (t^* is “relative time”) and in the third is a first order Autoregressive Conditional Logit (ACL). The ACL is the binomial version of the Autoregressive Conditional Multinomial (ACM) of [Russell and Engle \(2005\)](#). For SP500 returns, it is the first logit specification that fits the data best according to the [Schwarz \(1978\)](#) information criterion (SIC), whereas for Apple and Ekornes returns the best model according to SIC is the ACL(1,1). The first row of [Figure 5](#) contains the fitted zero probabilities. For SP500 and Ekornes the series appear to be stationary, whereas for Apple the series appear to be downward trending over the sample and hence non-stationary.

The bottom part of [Table 1](#) contains GARCH(1,1) estimates of the return series. In the SP500 and Ekornes cases we fit a single specification, namely

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

If z_t is (strictly) stationary (and ergodic), then the results by [Escanciano \(2009\)](#), and [Francq and Thieu \(2015\)](#) imply that Gaussian QML provides consistent parameter estimates (subject to additional conditions). In the Apple case we fit two different GARCH(1,1) specifications, namely (19) and a zero-adjusted specification:

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

The zero-adjusted specification is estimated by Gaussian QML in combination with the missing values algorithm proposed in [Sucarrat and Escribano \(2017\)](#), since I_t appears to be non-stationary with a downwards trend over the sample in the zero probability. To recall (see Section 2.5), the algorithm proceeds by replacing \tilde{r}_t with its estimate $\hat{\pi}_{1t}^{1/2} r_t$ whenever $r_t \neq 0$, while treating zeros as missing observations. Next, the missing values are replaced by estimates of their conditional expectations, i.e. $\hat{E}_{t-1}(\tilde{r}_t^2) = \hat{\sigma}_t^2$. Since Gaussian QML is used for 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-adj specifications appear small. However, as we will see shortly, these small nominal differences – together with the different treatment of zeros – can lead to substantially different risk estimates and risk dynamics.

3.2 Volatility

The second row in Figure 5 contains graphs of the ratios of the fitted conditional standard deviations. For SP500 and Ekornes estimates of σ_t are unaffected by zeros (subject to the assumption that z_t is strictly stationary and ergodic), so the ratios are 1 over the whole sample. For Apple the ratio is computed as $\hat{\sigma}_t / \hat{\sigma}_{t,0\text{-adj}}$, i.e. the fitted values of (19) over those from the 0-adjusted specification, i.e. (20). The Mean Percentage Error (MPE), computed as $n^{-1} \sum_{t=1}^n (x_t - 1) \cdot 100$ where $x_t = \hat{\sigma}_t / \hat{\sigma}_{t,0\text{-adj}}$ is the ratio at t , provides an overall measure of relative difference. Of course, the MPE is by construction 0% for SP500 and Ekornes. For Apple the MPE is 0.32%, which suggests volatility on average is only 0.32% higher than zero-adjusted volatility. A closer inspection, however, reveals that volatility is biased downwards in the beginning of the sample, and biased upwards towards the end. Also, the day-to-day difference between the two measures vary much more in the beginning, when there are more zeros, than towards the end. Finally, the difference in percentage terms is relatively low towards the end of the sample.

3.3 Conditional VaR

To illustrate the effect of zeros on Conditional VaR, we choose $c = 0.01$. Ratios of the estimated Conditional VaRs are contained in the third row of graphs in Figure 5, and the ratio at t is given by $x_t = \hat{r}_{c,t} / \hat{r}_{c,t,0\text{-adj}}$. For SP500, Ekornes and Apple, $\hat{r}_{c,t}$ 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 . For SP500 and Ekornes, $\hat{r}_{c,t,0\text{-adj}}$ 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-adjusted residuals \hat{w}_t (zeros excluded). For Apple, $\hat{r}_{c,t,0\text{-adj}}$ is computed as $\hat{\sigma}_{t,0\text{-adj}} \hat{z}_{c,t}$, where $\hat{\sigma}_{t,0\text{-adj}}$ is the

fitted value of (20), and $\widehat{z}_{c,t}$ is computed in the same way as for SP500 and Ekornes. Again we use the MPE ($= n^{-1} \sum_{t=1}^n (x_t - 1) \cdot 100$) as an overall measure of relative difference. Unsurprisingly, the MPE is essentially 0% for SP500. For Ekornes, by contrast, the MPE is about -10% . This means risk defined as Conditional VaR, on average, is biased downwards by 10% if zeros are not adjusted for. For Apple the MPE is only -0.40% , which suggests the downward bias is low. However, recalling that the zero probability has been gradually declining over the sample, a closer look reveals that the Conditional VaR is biased downwards in the first part of the sample – at times more than 10%, and biased upwards towards the end – always below 5%. The graph also shows that the ratio is more variable when there are more zeros, i.e. in the beginning of the sample.

3.4 Conditional ES

To illustrate the effect of zeros on Conditional ES, we choose $c = 0.01$ also here. Ratios of the estimated Conditional ESs are contained in the bottom row of graphs in Figure 5, and the ratio at t is given by $x_t = \widehat{ES}_{c,t} / \widehat{ES}_{c,t,0-adj}$. For SP500, Ekornes and Apple, $\widehat{ES}_{c,t}$ is computed as $-c^{-1} \widehat{\sigma}_t \widehat{E}_{t-1}(z_t | z_t \leq z_{c,t})$, where $\widehat{\sigma}_t$ is the estimate from (19), and $\widehat{E}_{t-1}(z_t | z_t \leq z_c)$ is computed as the sample average of the residuals \widehat{z}_t that are equal to or lower than \widehat{z}_c as defined above (i.e. the empirical c -quantile of the residuals \widehat{z}_t). For SP500 and Ekornes, the zero-adjusted estimate $\widehat{ES}_{c,t,0-adj}$ is computed as $-c^{-1} \widehat{\sigma}_t \widehat{E}_{t-1}(z_t | z_t \leq z_{c,t})$, where $\widehat{E}_{t-1}(z_t | z_t \leq z_{c,t})$ is equal to the relevant formula in (13), i.e. $\pi_{1t} E_{t-1} \left(w_t 1_{\{w_t \leq F_{w_t|1}^{-1}(c/\pi_{1t})\}} \right)$. As for Conditional VaR, to estimate $F_{w_t|1}^{-1}(c/\pi_{1t})$ at t we use the empirical $c/\widehat{\pi}_{1t}$ -quantile of the zero-adjusted 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} : $n_1^{-1} \sum_{I_t=1} \widehat{w}_t 1_{\{\widehat{w}_t \leq \widehat{F}_{w_t|1}^{-1}(c/\widehat{\pi}_{1t})\}}$, where n_1 is the number of non-zero observations (i.e. $n_1 = \sum_{t=1}^n I_t$), $\widehat{F}_{w_t|1}^{-1}(c/\widehat{\pi}_{1t})$ 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 Apple, $\widehat{ES}_{c,t,0-adj}$ 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 SP500 and Ekornes. The graphs of x_t and the MPEs show that Conditional ES is biased upwards in the Ekornes and Apple cases, and much more so than for VaR. Indeed, for Ekornes Conditional ES is on average biased upwards by about 54%, which is huge. Usually, daily stock returns will not exhibit a zero probability of about 19%, as in the Ekornes case. Usually, they will be below 5%. Nevertheless, the results do suggest the effect of zeros is much higher on Conditional ES than on Conditional VaR. Finally, also here does the graph of x_t for Apple exhibit a trend over the sample due to the downwards trend in the zero probability.

4 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. Key features of the

new class is that standard volatility models (e.g. ARCH, SV and continuous time models) are nested and obtained as special cases, and that the properties of the new class (e.g. conditional volatility, skewness, kurtosis, Value-at-Risk, Expected Shortfall, etc.) are obtained as functions of the underlying volatility model. Our results imply that, for a given volatility level, more zeros increases the skewness and kurtosis of return, but reduces return variability when defined as absolute return. Moreover, for a given level of volatility and sufficiently rare loss events (5% or less), risk defined as Conditional VaR will be biased downwards if zeros are not adjusted for, and risk defined as Conditional ES will be biased upwards if zeros are not adjusted for. The effect of zeros on volatility estimates will depend on the exact volatility model, the conditional density of return and on whether the zero probability is stationary or not. 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, our empirical illustration shows that risk estimates can be substantially biased in practice if the time-varying zero probability is not accommodated appropriately.

Our results have several practical and theoretical implications. First, our results suggests more attention should be paid to how market quotes and transaction prices are aggregated in order to compute 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. Second, for rare loss events we find that Conditional ES is biased upwards – sometimes substantially – if the zero probability is not adjusted for. This may have implications for the supervision of financial institutions, since recent regulatory changes emphasise the importance of ES (rather than VaR).

References

- Acerbi, C. and D. Tasche (2002). On the coherence of expected shortfall. *Journal of Banking and Finance* 26, 1487–1503.
- 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.
- 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 Inflation. *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 L. Q. Thieu (2015). Qml inference for volatility models with covariates. <http://mpira.ub.uni-muenchen.de/63198/>.
- 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 (2017). *GARCH Models*. New York: Marcel Dekker. 2nd. Edition, in press.
- Geweke, J. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews* 5, 57–61.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131, 59–95.
- 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.

- Liesenfeld, R., I. Nolte, and W. Pohlmeier (2006). Modelling Financial Transaction Price Movements: A Dynamic Integer Count Data Model. *Empirical Economics* 30, 795–825.
- Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika* 66, 265–270.
- Milhøj, A. (1987). A Multiplicative Parametrization of ARCH Models. Research Report 101, University of Copenhagen: Institute of Statistics.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347–370.
- Pantula, S. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews* 5, 71–73.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Russell, J. R. and R. F. Engle (2005). A Discrete-State Continuous-Time Model of Financial Transaction Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model. *Journal of Business and Economic Statistics* 23, 166–180.
- Rydberg, T. H. and N. Shephard (2003). Dynamics of Trade-by-Trade Price Movements: Decomposition and Models. *Journal of Financial Econometrics* 1, 2–25.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Shiryayev, 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. Escibano (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. Escibano (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.

A Proofs

A.1 Proof of Proposition 1

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

$$E_{t-1}(z_t) = \pi_{1t}^{-1/2} E_{t-1}(w_t I_t) \quad (21)$$

$$= \pi_{1t}^{-1/2} \left(E_{t-1}(w_t \cdot 1 | I_t = 1) \pi_{1t} + E_{t-1}(w_t \cdot 0 | I_t = 0) \pi_{0t} \right) \quad (22)$$

$$= 0 \quad (23)$$

for all t . Note that the notation $E_{t-1}(w_t \cdot 0 | I_t = 0) \pi_{0t}$ stands for $E_{t-1}(w_t \cdot 0)$ whenever $\pi_{0t} = 0$. Accordingly, $\{z_t\}$ is an MDS.

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

$$E_{t-1}(z_t^2) = \pi_{1t}^{-1} E_{t-1}(w_t^2 I_t^2) \quad (24)$$

$$= \pi_{1t}^{-1} \left(E_{t-1}(w_t^2 \cdot 1 | I_t = 1) \pi_{1t} + E_{t-1}(w_t^2 \cdot 0 | I_t = 0) \pi_{0t} \right) \quad (25)$$

$$= \pi_{1t}^{-1} \left(\sigma_w^2 \pi_{1t} \right) \quad (26)$$

$$= \sigma_w^2. \quad (27)$$

for all t . Note that also here the notation $E_{t-1}(w_t^2 \cdot 0 | I_t = 0) \pi_{0t}$ stands for $E_{t-1}(w_t^2 \cdot 0)$ whenever $\pi_{0t} = 0$. Next, since $\{z_t\}$ is an 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

$$E_{t-1}(z_t^s) = \pi_{1t}^{-s/2} E_{t-1}(w_t^s I_t) \quad (28)$$

$$= \pi_{1t}^{-s/2} \left(E_{t-1}(w_t^s \cdot 1 | I_t = 1) \pi_{1t} + E_{t-1}(w_t^s \cdot 0 | I_t = 0) \pi_{0t} \right) \quad (29)$$

$$= \pi_{1t}^{(2-s)/2} E_{t-1}(w_t^s | I_t = 1) \quad (30)$$

for all t . Again, 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 $\pi_{0t} = 0$.

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

$$E_{t-1}|z_t|^s = \pi_{1t}^{-s/2} E_{t-1}(|w_t|^s I_t^s) \quad (31)$$

$$= \pi_{1t}^{-s/2} \left(E_{t-1}(|w_t|^s \cdot 1 | I_t = 1) \pi_{1t} + E_{t-1}(|w_t|^s \cdot 0 | I_t = 0) \pi_{0t} \right) \quad (32)$$

$$= \pi_{1t}^{(2-s)/2} E_{t-1}(|w_t|^s | I_t = 1) \quad (33)$$

for all t . Again, 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

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:

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

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, (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 \tilde{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 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 3 and 5.

Lemma 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 1 on p. 6.

From Assumption 3(a) and the expression for $F_{z_t}(x)$ in Proposition 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 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 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 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 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 \tilde{r}_t and z_t with r_t , gives (11).

A.4 Proof of Proposition 4

Due to Assumptions 1 and 4 we have

$$F_{\tilde{r}_t|1}(x) = P_{t-1}(\tilde{r}_t \leq x | I_t = 1) \quad (34)$$

$$= P_{t-1}(\sigma_t w_t \leq x | I_t = 1) \quad (35)$$

$$= P_{t-1}(w_t \leq x\sigma_t^{-1} | I_t = 1) \quad (36)$$

$$\stackrel{(4)}{=} F_{w_t|1}(x\sigma_t^{-1}), \quad (37)$$

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

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

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

$$r_{c,t} = F_r^{-1}(c) = \sigma_t \begin{cases} \pi_{1t}^{-1/2} F_{\tilde{w}_t|1}^{-1}(c/\pi_{1t}) & \text{if } c < F_{\tilde{w}_t|1}(0)\pi_{1t} \\ 0 & \text{if } F_{\tilde{w}_t|1}(0)\pi_{1t} \leq c < F_{\tilde{w}_t|1}(0)\pi_{1t} + \pi_{0t} \\ \pi_{1t}^{-1/2} F_{\tilde{w}_t|1}^{-1} \left[\frac{(c - \pi_{0t})}{\pi_{1t}} \right] & \text{if } c \geq F_{\tilde{w}_t|1}(0)\pi_{1t} + \pi_{0t}. \end{cases} \quad (38)$$

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

A.5 Proof of Proposition 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 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 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 $P_{t-1}(\xi \leq x) = x1_{\{0 \leq x \leq 1\}} + 1_{\{x > 1\}}$. 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 and Assumption 5, then $F_{z_t} \circ F_{z_t}^{-1}(c) = c$ by (c) in Lemma 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, so $F_{z_t}^{-1}(c)[c - P_{t-1}(z_t \leq F_{z_t}^{-1}(c))] = 0$. \square

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 3, and so

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

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 $u = x\sqrt{\pi_{1t}}$ gives $dx = du/\sqrt{\pi_{1t}}$, and the area of integration is changed to $(-\infty, F_{w_t|1}^{-1}(c/\pi_{1t})]$. This gives

$$E(z_t 1_{\{z_t \leq F_{z_t}^{-1}(c)\}}) = \pi_{1t} \int_{-\infty}^{F_{w_t|1}^{-1}(c/\pi_{1t})} u f_{w_t|1}(u) du = \pi_{1t} E(w_t 1_{\{w_t \leq F_{w_t|1}^{-1}(c/\pi_{1t})\}}).$$

Case 2: $F_{w_t|1}(0)\pi_{1t} \leq c < F_{w_t|1}(0)\pi_{1t} + \pi_{0t}$. In this case $E(z_t 1_{\{z_t \leq F_{z_t}^{-1}(c)\}}) = E(z_t 1_{\{z_t \leq 0\}})$ according to Proposition 3, and so

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

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

Case 3: $c \geq F_{w_t|1}(0)\pi_{1t} + \pi_{0t}$. In this case $E(z_t 1_{\{z_t \leq F_{z_t}^{-1}(c)\}}) = E(z_t 1_{\{z_t \leq \pi_{1t}^{-1/2} F_{w_t|1}^{-1}[(c - \pi_{0t})/\pi_{1t}]\}})$ according to Proposition 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

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

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

$$\int_A x d[\pi_{0t} 1_{\{0 \leq x\}}] = \pi_{0t} \int_{\mathbb{R}} 1_{\{x \in A\}} x d1_{\{0 \leq x\}} = \pi_{0t} 1_{\{x \in A\}} x|_{x=0} = 0.$$

As in case 1 we see that

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

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

A.6 Proof of Proposition 6

From the measurability of σ_t with respect to \mathcal{F}_{t-1}^r (i.e. Assumption 4) it follows that $E_{t-1}(\tilde{r}_t 1_A) = \sigma_t E_{t-1}(w_t 1_A)$, where A denotes an event. Denote $y = F_{\tilde{r}_t|1}(x)$, so that $F_{\tilde{r}_t|1}^{-1}(y) = x$. From the proof of Proposition 4 in Appendix A.4 it follows that $F_{\tilde{r}_t|1}(x) = F_{w_t|1}(x \sigma_t^{-1})$ and $F_{\tilde{r}_t|1}^{-1}(y) = \sigma_t F_{w_t|1}^{-1}(y)$. Accordingly, we can re-write (14) as

$$E_{t-1}(r_t | r_t \leq r_{c,t}) = \sigma_t \begin{cases} \pi_{1t} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w_t|1}^{-1}(c/\pi_{1t})\}}\right) & \text{if } c < F_{w_t|1}(0)\pi_{1t}, \\ \pi_{1t} E_{t-1}\left(w_t 1_{\{w_t \leq 0\}}\right) & \text{if } F_{w_t|1}(0)\pi_{1t} \leq c < F_{w_t|1}(0)\pi_{1t} + \pi_{0t}, \\ \pi_{1t} E_{t-1}\left(w_t 1_{\{w_t \leq F_{w_t|1}^{-1}[(c - \pi_{0t})/\pi_{1t}]\}}\right) & \text{if } c \geq F_{w_t|1}(0)\pi_{1t} + \pi_{0t}. \end{cases} \quad (39)$$

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 $\hat{\alpha}_0^{(k)}$, $\hat{\alpha}_1^{(k)}$ and $\hat{\beta}_1^{(k)}$ 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 $r_t = \tilde{r}_t$ for all t , then the k th. iteration of the numerical method proceeds in the usual way:

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

$$\hat{\sigma}_t^2 = \hat{\alpha}_0^{(k-1)} + \hat{\alpha}_1^{(k-1)} \tilde{r}_{t-1}^2 + \hat{\beta}_1^{(k-1)} \hat{\sigma}_{t-1}^2. \quad (40)$$

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

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

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

$$\text{a) } \overline{r}_t^2 = \begin{cases} \tilde{r}_t^2 & \text{if } t \in G \\ \hat{\sigma}_t^2 & \text{if } t \notin G, \text{ where } \hat{\sigma}_t^2 = \hat{\alpha}_0^{(k-1)} + \hat{\alpha}_1^{(k-1)} \overline{r}_{t-1}^2 + \hat{\beta}_1^{(k-1)} \hat{\sigma}_{t-1}^2, \end{cases} \quad (41)$$

$$\text{b) } \hat{\sigma}_t^2 = \hat{\alpha}_0^{(k-1)} + \hat{\alpha}_1^{(k-1)} \overline{r}_{t-1}^2 + \hat{\beta}_1^{(k-1)} \hat{\sigma}_{t-1}^2. \quad (42)$$

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

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

QML Estimation of the log-GARCH model is via its ARMA-representation, since the standard Gaussian ML estimator must be interpreted as *exact* ML in the presence of missing values, see [Sucarrat and Escibano \(2017\)](#). If $|E(\ln w_t^2)| < \infty$, then the ARMA(1,1) representation is given by

$$\ln \tilde{r}_t^2 = \phi_0 + \phi_1 \ln \tilde{r}_{t-1}^2 + \theta_1 u_{t-1} + u_t, \quad u_t = \ln w_t^2 - E(\ln w_t^2), \quad (43)$$

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 $-\ln [n^{-1} \sum_{t=1}^n \exp(\hat{u}_t)]$ provides a consistent estimate (see also [Francq and Sucarrat \(2017\)](#)). To accommodate the missing values, this formula is modified to $-\ln [n^{*-1} \sum_{t \in G} \exp(\hat{u}_t)]$.

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

$$r_t = \sigma_t I_t w_t \pi_{1t}^{-1/2}, \quad w_t \sim N(0, 1), \quad t = 1, \dots, n = 10000, \quad (44)$$

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

$$\pi_{1t} = 1/(1 + \exp(-h_t)), \quad h_t = \rho_0 + \lambda t^*, \quad t^* = t/n. \quad (45)$$

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

$$\sigma_t^2 = \alpha_0 + \alpha_1 \tilde{r}_{t-1}^2 + \sigma_{t-1}^2, \quad (46)$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \tilde{r}_{t-1}^2 + \ln \sigma_{t-1}^2, \quad (47)$$

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”, \tilde{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 sketched 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 produce large differences in the dynamics.

Table 1: Descriptive statistics, dynamic logit models and GARCH-models of SP500, Apple and Ekornes returns (see Section 3)

Descriptive statistics:							
	s^2	s^4	ARCH [<i>p-val</i>]	n	$0s$	$\hat{\pi}_0$	
SP500	1.73	10.30	143.10 [0.00]	3684	2	0.001	
Ekornes	5.70	10.32	54.01 [0.00]	3546	667	0.189	
Apple	9.25	55.03	7.12 [0.01]	7303	294	0.040	
Dynamic logit-models:							
		$\hat{\rho}_0$ (<i>s.e.</i>)	$\hat{\rho}_1$ (<i>s.e.</i>)	$\hat{\zeta}_1$ (<i>s.e.</i>)	$\hat{\lambda}$ (<i>s.e.</i>)	<i>SIC</i>	<i>Logl</i>
SP500	Constant	7.158 (0.707)				0.0115	-17.04
	Trend	7.200 (1.309)			0.673 (2.478)	0.0137	-17.00
	ACL(1,1)	0.032 (4e-05)	-1.147 (4e-05)	0.997 (4e-05)		0.0116	-9.022
Ekornes	Constant	1.462 (0.043)				0.9692	-1714.3
	Trend	1.183 (0.083)			0.576 (0.150)	0.9673	-1706.9
	ACL(1,1)	0.445 (0.130)	0.207 (0.036)	0.701 (0.087)		0.9599	-1689.6
Apple	Constant	3.171 (0.060)				0.3387	-1232.5
	Trend	1.870 (0.094)			3.437 (0.263)	0.3102	-1123.9
	ACL(1,1)	1e-09 (4e-05)	0.024 (0.011)	0.999 (9e-05)		0.3095	-1116.9
GARCH-models:							
		$\hat{\alpha}_0$ (<i>s.e.</i>)	$\hat{\alpha}_1$ (<i>s.e.</i>)	$\hat{\beta}_1$ (<i>s.e.</i>)			
SP500	Ordinary:	0.015 (0.003)	0.083 (0.008)	0.908 (0.009)			
Ekornes	Ordinary:	0.036 (0.009)	0.019 (0.002)	0.974 (0.004)			
Apple	Ordinary:	0.168 (0.033)	0.087 (0.008)	0.901 (0.010)			
	0-adjusted:	0.175 (0.037)	0.093 (0.010)	0.894 (0.012)			

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

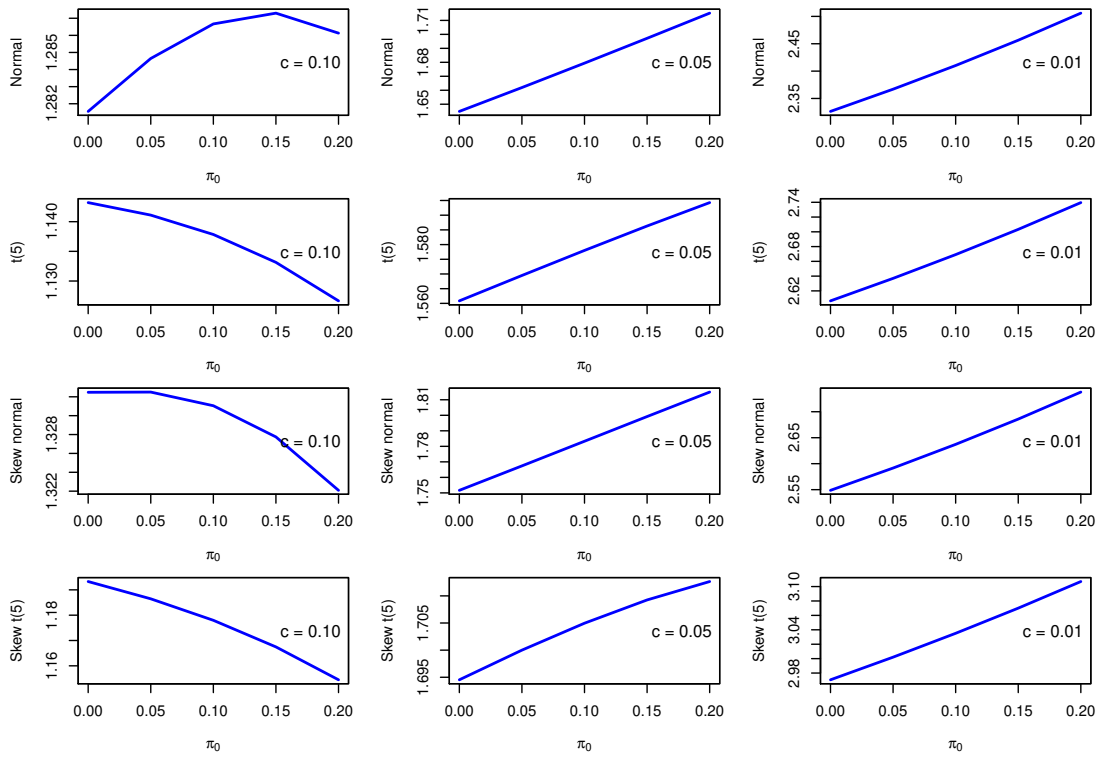


Figure 1: The effect of zeros on the Conditional VaR of z_t , see Section 2.3

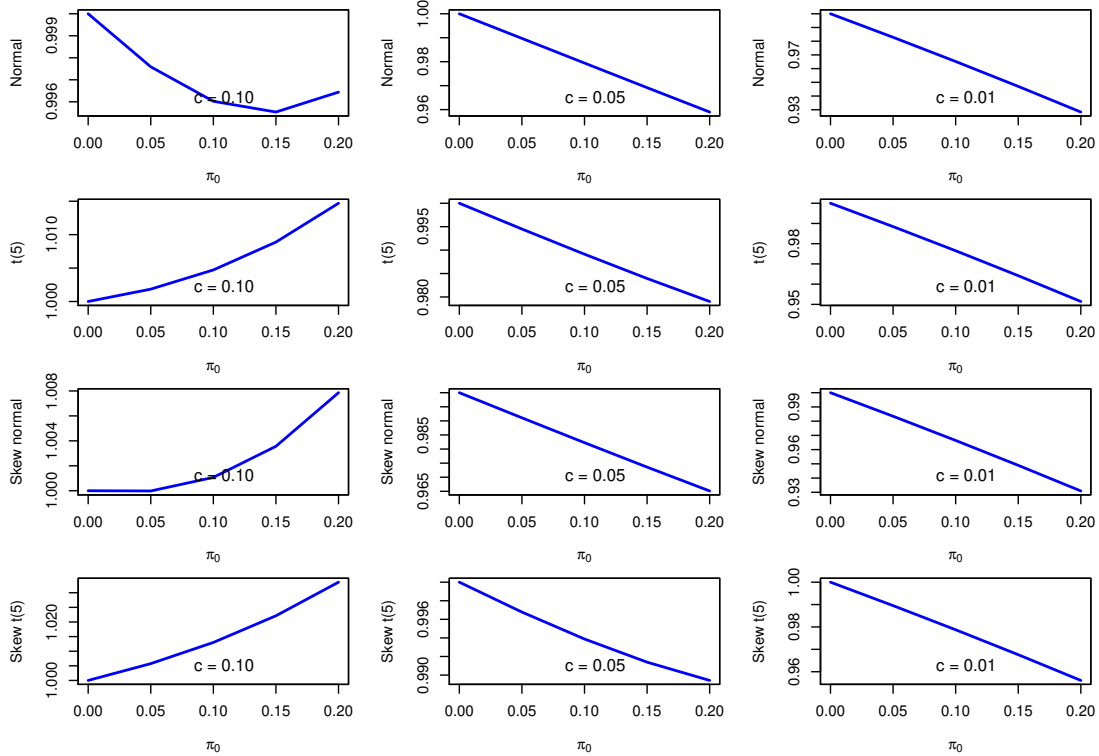


Figure 2: The effect of zeros on Conditional VaR ratios (unadjusted Conditional VaR in numerator, zero-adjusted Conditional VaR in denominator), see Section 2.3

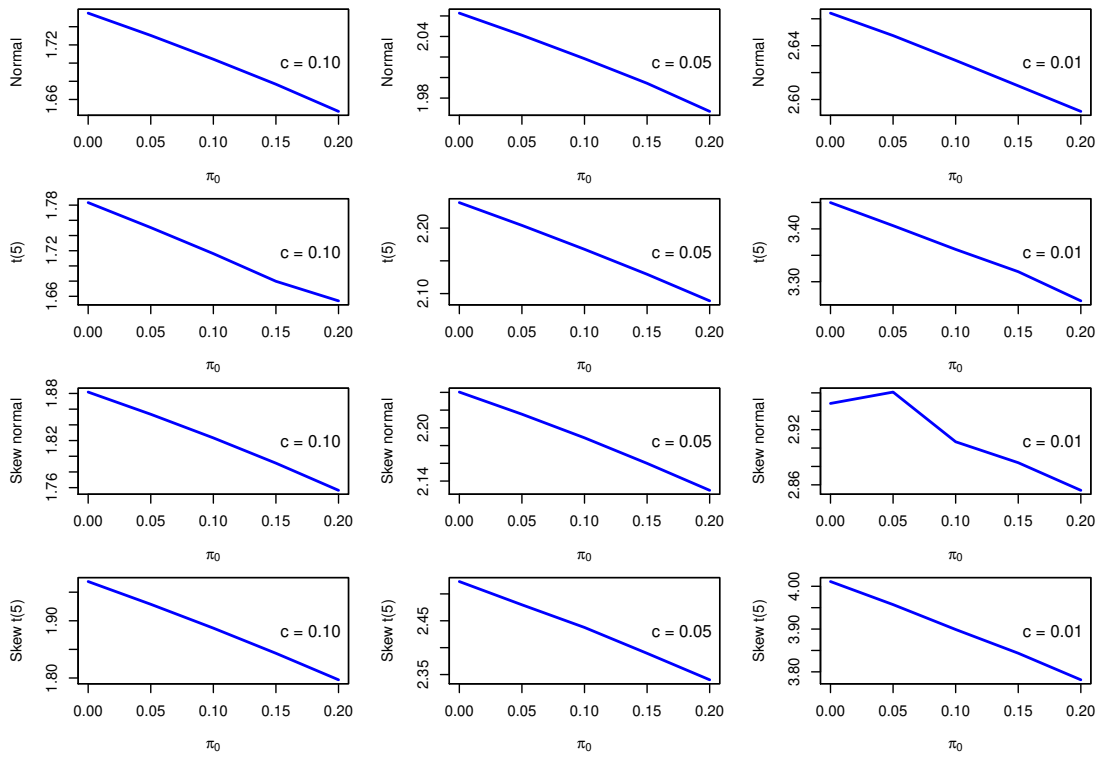


Figure 3: The effect of zeros on Conditional ES, see Section 2.4

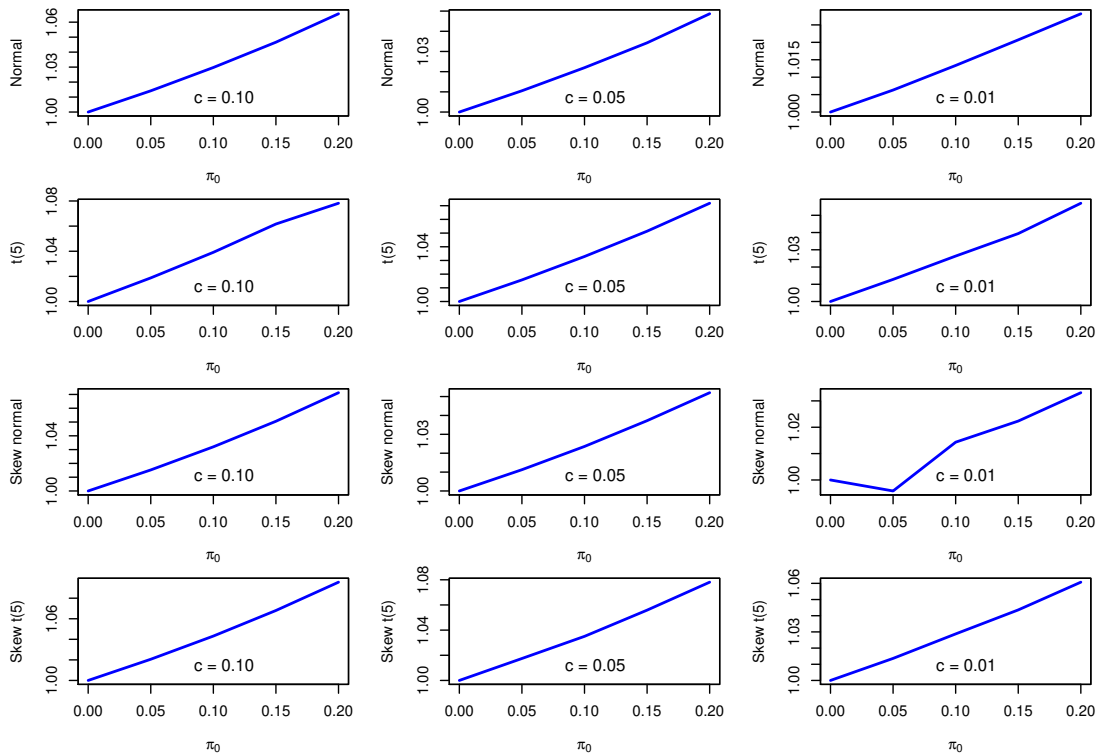


Figure 4: The effect of zeros on Conditional ES ratios (unadjusted Conditional ES in numerator, zero-adjusted Conditional ES in denominator), see Section 2.4

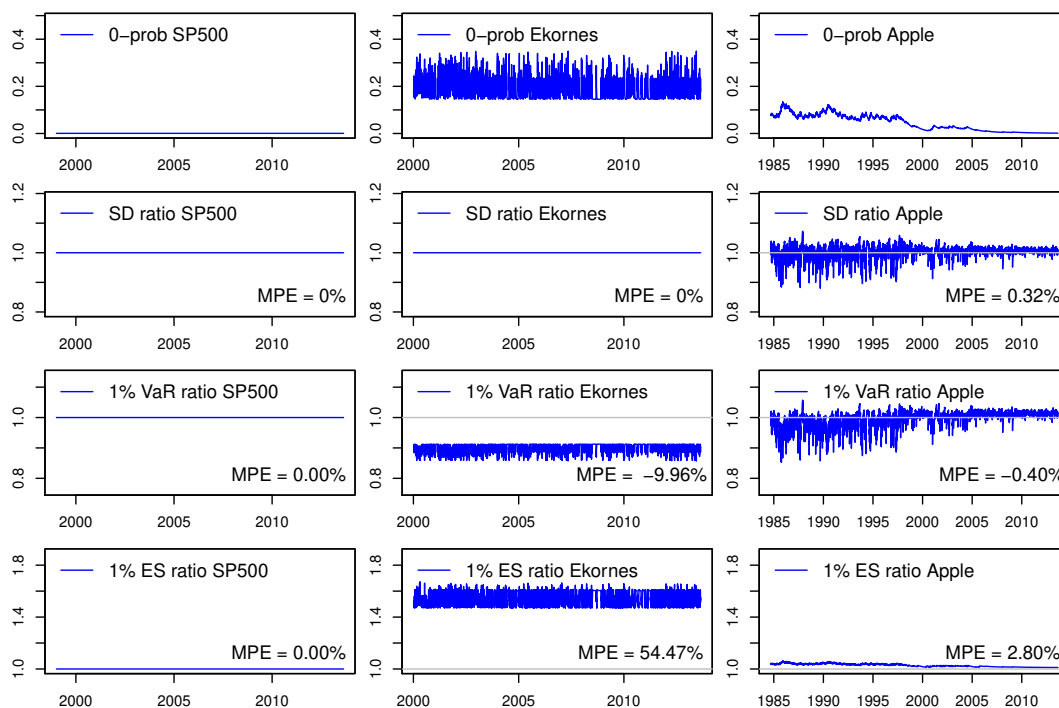


Figure 5: Fitted zero probabilities, and the ratios of fitted σ_t , 1% VaR and 1% ES (see Section 3).

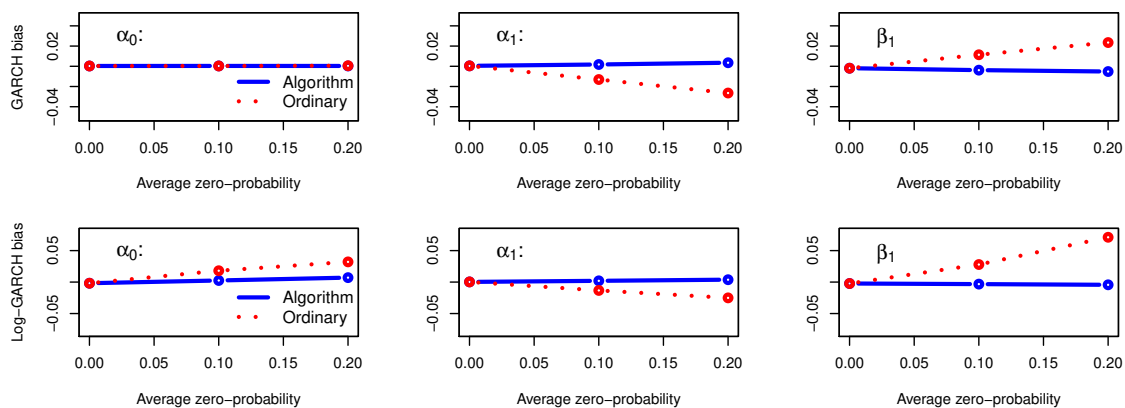


Figure 6: Simulated parameter biases in GARCH(1,1) and log-GARCH(1,1) models for the missing values algorithm in comparison with ordinary methods (see Appendix B)