A Brief Survey on the Choice of Parameters for: "Kernel density estimation for time series data"

Artur Semeyutin[†] and Robert O'Neill[‡]

The Business School, University of Huddersfield, Huddersfield HD1 3DH, UK;

Department of Accountancy, Finance and Economics

[†]artur.semeyutin@hud.ac.uk,[1] [‡]r.o'neill@hud.ac,

---

[1]corresponding author

Abstract

This paper presents an overview of the empirical performance of some of the common methods for parameter selection in the area of enhanced dynamic kernel density and distribution estimation with exponentially declining weights. It is shown that exponential weighting delivers accurate nonparametric density forecasts, without common corrections for scale and/or location in most of the financial time series considered, provided that parameters are chosen appropriately. Computational simplifications for more efficient numerical optimizations and simple kernel adaptive estimation strategies for equally effective forecasting are also provided.

1 Introduction

Harvey & Oryshchenko (2012) formalize a method for dynamic kernel density (PDF) and cumulative distribution (CDF) estimation for time series data when employing exponentialy declining weights.[2] The attractiveness of the dynamic nonparametric CDF estimate for practitioners is driven by the fact that it allows for a full set of quantiles to be tracked simultaneously, thus allowing for exploratory dependence (e.g. as in Harvey, 2010) or risk exposure modelling (e.g. as in Taylor, 2007), without compromising the fact that quantile estimates may cross (e.g. Gouriéroux & Jasiak, 2008; Harvey, 2010). Overall, the method presented in Harvey & Oryshchenko (2012) is straightforward to implement and appealing for applied use, however in order to compete with nonparametric exponential weights quantile regression for Value-at-Risk (VaR) models of Chen & Tang (2005) or Taylor (2007), the method requires an appropriate choice of parameters, since Harvey & Oryshchenko (2012) fall into the computational short-cut of employing Maximum Likelihood (ML) for identifying the unknown values.[3]

In this paper we present explicit modifications to the nonparametric workhorse Least-Squares (LS) routine, accommodating exponential weights in the recursive form for PDF and CDF estimation.[4] We also demonstrate how binned kernel estimators may be employed reducing the computational effort involved in the recursive LS optimisation or efficiently varying the bandwidth parameter for adaptive time-varying PDF and CDF estimation, maintaining the accuracy of the density forecasts if necessary. We conduct our analysis on a data set consisting of ten samples of daily logarithmic financial returns of different lengths over in-sample performance evaluations with Probability of Integral Transforms (PITs) and lower tail domain time-varying quantiles (VaR) for the best PITs performing approach to the parameters estimation identified. Further, we perform in-sample VaR comparisons to the well-known, for various financial audiences, J.P. Morgan's (1996) RiskMetrics™ (RM) approach and to some of the RM's more advanced modifications presented by Lucas & Zhang

---

[2]The combination of exponential weights and kernel type estimators had been present in the relevant literature prior to this contribution (e.g. Yu & Jones, 1998; Chen & Tang, 2005; Taylor, 2007; Bessa et al., 2012; Zhang et al., 2014; Taylor & Jeon, 2015; Arora & Taylor, 2016), but Harvey & Oryshchenko (2012) were the first to provide a focused and comprehensive discussion of such estimators for tracking the entire probability density and distribution functions nonparametrically.

[3]In the nonparametric framework ML is computationally attractive, though it is well known to produce suboptimal estimates (e.g. Schuster & Gregory, 1981; Chow et al., 1983; Hall, 1987a,b; Härdle, 1990; Li & Racine, 2007).

[4]Optimal bandwidths for PDF and CDF estimation may be expected to have different convergence rates as discussed in Li & Racine (2007).

(2016) under the Generalized Autoregressive Score (GAS) time series framework. Finally, we discuss the method identified for choosing the parameters in Wang et al. (2018) and conclude that at its current development stage, it requires a more rigorous approach to the training constrains as our results for our LS parameters deliver a more appealing diagnostic output. For now, we suggest relying on the parsimonious empirical CDF (eCDF) and exponential weights based approach similar to kernel estimators of Harvey & Oryshchenko (2012) instead. From our results, this simple suggestion is appropriate for tracking time-varying quantiles and just marginally behind the kernel estimators of Harvey & Oryshchenko (2012).

2 Methodology

We begin our work by introducing two types of exponential weighting schemes and how they can be used to enhance kernel PDF and CDF estimation of time series data.

2.1 Dynamic kernel methods for time-varying density and distribution

A set of weights for exponential filtering can be defined by the following equation:

$$w_{t,i}(\omega) = (1 - \omega)(1 - \omega^t)^{-1}\omega^{t-i}, \tag{1}$$

where ω is the weights' learning rate restricted to be in the $(0, 1]$ interval so that $w_{t,i}$ satisfy $\sum_{i=1}^{t} w_{t,i} = 1$ for all $i = 1, ..., t$ and $t = 1, ..., T$. Hence t is the point at which an estimate is being made and i is an index of all time points leading up to this point. Given this definition a time-varying probability density of financial returns of value x can be estimated nonparametrically using:

$$\hat{f}_{t+1|t}(x) = h^{-1}\sum_{i=1}^{t} K\left([x - x_i]h^{-1}\right)w_{t,i}(\omega), \tag{2}$$

where $K(\cdot)$ is a symmetric kernel in the form of PDF (for examples of such standard second order type kernels see Wand & Jones (1995)) and h is an optimal bandwidth parameter. Hence, for the same weighting scheme employing a kernel $W(\cdot)$ in the form of CDF, as described in Li & Racine

(2007), a time-varying CDF of financial returns is given by:

$$\hat{F}_{t+1|t}(x) = \sum_{i=1}^{t} W\left([x - x_i]\beta^{-1}\right) w_{t,i}(\omega),$$ (3)

where β is a CDF optimal bandwidth parameter. In both of the cases above, we can see that densities and distributions alter depending on time at which the estimate is being made, while Figure 1 provides an illustration of such time-varying PDF estimate, where combination of the nonparametric estimator's flexibility (e.g. curvature in the PDF tail domains) with exponentially declining weights (e.g. time evolution of the curvature) is straightforward to note.

[Figure 1 around here.]

We can further expand this approach by employing two-sided weights:

$$w_{t,T,i}(\omega) = (1 - \omega)(1 + \omega - \omega^t - \omega^{T-t+1})^{-1}\omega^{|t-i|},$$ (4)

for an exponentially time conditional smoothed PDF estimate of financial returns, which is conditional on the choice of T [5] and is provided by:

$$\hat{f}_{t|T}(x) = h^{-1} \sum_{i=1}^{T} K\left([x - x_i]h^{-1}\right) w_{t,T,i}(\omega)$$ (5)

with their CDF estimate now defined by:

$$\hat{F}_{t|T}(x) = \sum_{i=1}^{T} W\left([x - x_i]\beta^{-1}\right) w_{t,T,i}(\omega).$$ (6)

---

[5]Note that similar to Harvey (1990) we refer to the estimations with exponential weights equipped with information up to t at the time of estimation as exponential filtering and to estimations equipped with information up to T at the time of estimation as exponential smoothing.

Adaptive dynamic nonparametric estimators as defined since Breiman et al. (1977) with weights in the form of (1) can be given by the below equations for the PDF and CDF respectively:

$$\hat{f}_{t+1|t}(x) \;=\; \sum_{i=1}^{t} K\left([x - x_i]h_i^{-1}\right) w_{t,i}(\omega)h_i^{-1}, \tag{7}$$

$$\hat{F}_{t+1|t}(x) \;=\; \sum_{i=1}^{t} W\left([x - x_i]\beta_i^{-1}\right) w_{t,i}(\omega). \tag{8}$$

Finally, for weights defined by (4) the time-varying adaptive PDF and CDF estimates can be defined as:

$$\hat{f}_{t|T}(x) \;=\; \sum_{i=1}^{T} K\left([x - x_i]h_i^{-1}\right) w_{t,T,i}(\omega)h_i^{-1}, \tag{9}$$

$$\hat{F}_{t|T}(x) \;=\; \sum_{i=1}^{T} W\left([x - x_i]\beta_i^{-1}\right) w_{t,T,i}(\omega). \tag{10}$$

2.2   Unknown parameters estimation

In this section we describe functions which can be used to obtain unknown parameters for time-varying PDF and CDF modelling with estimators in Section 2.1. We begin with ML as in Harvey & Oryshchenko (2012) and then provide reformulated canonical LS loss functions for dynamic kernel estimators. Li & Racine (2007) propagate loss functions in the form of LS for estimations. Such functional form often supply parameters leading to better estimation outputs, but are computationally heavier than ML, since they contain more blocks for evaluation. Moreover, these blocks increase in size with very iteration in the time series context. We suggest reducing time necessary to compute parameters with LS using binned simplifications of the dynamic kernel estimators next. Though not for the time series context, binned simplifications are available since Hall (1982), while their ability for approximating unknown parameters with minor loss in estimation accuracy, given sufficient number of bins, has been studied by Scott & Sheather (1985) and Hall & Wand (1996). Further, with binned estimators and LS for time series estimations it may be convenient allowing different bandwidths parameter for tail and body domains when performing time-varying adaptive PDF and CDF estimation. Scott (2015) highlights this for independent and identically distributed estimations framework and we just make necessary rearrangements to match

the time series context.

2.2.1 Maximum Likelihood

In the nonparametric framework ML implies optimization of the Kullback-Leibler loss function for kernels considered in their PDF form. For the weights of the form (1) the loss function takes the predictive recursive form of: [6]

$$l_f(\omega, h) = (T - m)^{-1} \sum_{t=m}^{T-1} \log \left[h^{-1} \sum_{i=1}^{t} K \left([x_{t+1} - x_i] h^{-1} \right) w_{t,i}(\omega) \right], \tag{11}$$

where m is the number of observations required for the initialisation of the procedure. For the weights of the form (4) the loss function effectively transforms into a cross-validation type problem, defined to maximise the following criterion function:

$$l_s(\omega, h) = T^{-1} \sum_{t=1}^{T} \log \left[h^{-1} \sum_{i=1, i \neq t}^{T} K \left([x_t - x_i] h^{-1} \right) w_{t,T,i}(\omega) \right]. \tag{12}$$

2.2.2 Canonical Least-Squares

Rewriting canonical LS in the recursive form for (1) provides:

$$
\begin{aligned}
ls_f(\omega, h) &= (T - m)^{-1} \sum_{t=m}^{T-1} \left[h^{-1} \sum_{j=1}^{t} \left(\sum_{i=1}^{t} \bar{K} \left([x_j - x_i] h^{-1} \right) w_{t,i}(\omega) \right) w_{t,j}(\omega) \right. \\
&\quad \left. - 2h^{-1} \sum_{i=1}^{t} K \left([x_{t+1} - x_i] h^{-1} \right) w_{t,i}(\omega) \right],
\end{aligned} \tag{13}
$$

where $\bar{K}$ is a convolution kernel.[7] Cross-validation for weights in (4) is thus given by:

$$
\begin{aligned}
ls_s(\omega, h) &= T^{-1} \sum_{t=1}^{T} \left[h^{-1} \sum_{j=1}^{T} \left(\sum_{i=1}^{T} \bar{K} \left([x_j - x_i] h^{-1} \right) w_{t,T,i}(\omega) \right) w_{t,T,j}(\omega) \right. \\
&\quad \left. - 2h^{-1} \sum_{i=1, i \neq t}^{T} K \left([x_t - x_i] h^{-1} \right) w_{t,T,i}(\omega) \right].
\end{aligned} \tag{14}
$$

---

[6] All provided ML expressions are outlined by Harvey & Oryshchenko (2012).

[7] For instance, from Li & Racine (2007), if $K(x) = \dfrac{1}{\sqrt{2\pi}} e^{(-x^2/2)}$, is widely used Gaussian kernel, then its convolution $\int_{-\infty}^{\infty} K(y)K(x-y)dy = \bar{K}(x) = \dfrac{1}{\sqrt{4\pi}} e^{(-x^2/4)}$.

Since the work of Bowman et al. (1998), LS has allowed effective evaluation of optimal parameters for CDF and in its recursive form for weights in (1) is given by:

$$LS_f(\omega, \beta) = (T-m)^{-1} \sum_{t=m}^{T-1}$$
$$\int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_{t+1}<x\}} - \sum_{i=1}^{t} W\left([x - x_i]\beta^{-1}\right) w_{t,i}(\omega) \right]^2 dx, \qquad (15)$$

where $\mathbb{I}_{\{x_{t+1}<x\}}$ is an indicator function taking value of 1 if the condition in the parentheses is satisfied. Similarly for weights in (4), LS for CDF takes the following form for optimizing:

$$LS_s(\omega, \beta) = T^{-1} \sum_{t=1}^{T}$$
$$\int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_t<x\}} - \sum_{i=1,\, i\neq t}^{T} W\left([x - x_i]\beta^{-1}\right) w_{t,T,i}(\omega) \right]^2 dx. \qquad (16)$$

2.2.3 Binned and adaptive Least-Squares

LS, in the recursive form for filtering or in the cross-validation format for smoothing may be computationally complex functions for optimization, especially if the sample size is large. Therefore, a simplification of (13) with:

$$ls_f^b(\omega, h) = (T-m)^{-1} \sum_{t=m}^{T-1} \left[h^{-1} \sum_{s=1}^{a} \left(\sum_{q=1}^{a} \bar{K}\left([c_s - c_q]h^{-1}\right) w_{t,q}(\omega) \right) w_{t,s}(\omega) \right.$$
$$\left. -2h^{-1} \sum_{i=1}^{t} K\left([x_{t+1} - c_s]h^{-1}\right) w_{t,s}(\omega) \right], \qquad (17)$$

where c_s is the centre of some bin B_s, a is the total number of bins and $w_{t,s}(\omega) = \sum_{i=1}^{t} \mathbb{I}_{\{x_i \in B_s\}} w_{t,i}(\omega)$, provides an alternative, where bins are used to simplify the estimation of the function over the range of returns.[8] Necessesary binning strategies are provided and well described by Hall & Wand (1996)

---

[8]The notation used here is largely adopted from Scott & Sheather (1985) and Sain & Scott (1996).

among others. The same binning strategy can then be employed to simplify (14) into:

$$
\begin{aligned}
ls_s^b(\omega, h) \;=\; & T^{-1} \sum_{t=1}^{T} \left[h^{-1} \sum_{s=1}^{a} \left(\sum_{q=1}^{a} \bar{K}\left([c_s - c_q]h^{-1}\right) w_{t,T,q}(\omega) \right) w_{t,T,s}(\omega) \right. \\
& \left. - 2h^{-1} \sum_{s=1}^{a} K\left([x_t - c_s]h^{-1}\right) w^*_{t,T,s}(\omega) \right],
\end{aligned}
\tag{18}
$$

where $w_{t,T,s}(\omega) = \sum_{i=1}^{t} \mathbb{I}_{\{x_i \in B_s\}} w_{t,T,i}(\omega)$ and $w^*_{t,T,s}(\omega) = \sum_{\substack{i=1 \\ t \neq i}}^{t} \mathbb{I}_{\{x_i \in B_s\}} w_{t,T,i}(\omega)$. Next, the same approach can be extended to the loss functions designed for obtaining optimal CDF parameters. For filtering the function (15) becomes:

$$
\begin{aligned}
LS_f^b(\omega, \beta) = \quad & (T-m)^{-1} \sum_{t=m}^{T-1} \\
& \int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_{t+1} < x\}} - \sum_{s=1}^{a} W\left([x - c_s]\beta^{-1}\right) w_{t,s}(\omega) \right]^2 dx,
\end{aligned}
\tag{19}
$$

and for smoothing from (16) we obtain:

$$
\begin{aligned}
LS_s^b(\omega, \beta) = \quad & T^{-1} \sum_{t=1}^{T} \\
& \int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_t < x\}} - \sum_{s=1}^{a} W\left([x - c_s]\beta^{-1}\right) w^*_{t,T,s}(\omega) \right]^2 dx.
\end{aligned}
\tag{20}
$$

Binned simplifications can be extended to allow varying the bandwidth parameter for estimators in (7), (8), (9) and (10) so that $h_i = h_s(x_i \in B_s)$ and $\beta_i = \beta_s(x_i \in B_s)$ respectively, thus allowing the bandwidth parameters to change according to the bin they are being used in.[9] For PDF optimal parameters for filtering this can be obtained using:

$$
\begin{aligned}
ls_f^a(\omega, h_s) \;=\; & (T-m)^{-1} \sum_{t=m}^{T-1} \left[\sum_{s=1}^{a} \left(\sum_{q=1}^{a} \bar{K}\left([c_s - c_q]h_q^{-1}\right) w_{t,q}(\omega) \right) w_{t,s}(\omega) \cdot h_s^{-1} \right. \\
& \left. - 2 \sum_{i=1}^{t} K\left([x_{t+1} - c_s]h_s^{-1}\right) w_{t,s}(\omega) \cdot h_s^{-1} \right]
\end{aligned}
\tag{21}
$$

---

[9]Such estimators and their properties have been available since Sain & Scott (1996).

and for smoothing:

$$ls_s^a(\omega, h_s) \;=\; T^{-1} \sum_{t=1}^{T} \left[\sum_{s=1}^{a} \left(\sum_{q=1}^{a} \bar{K}\left([c_s - c_q]h_q^{-1}\right) w_{t,T,q}(\omega) \right) w_{t,T,s}(\omega) \cdot h_s^{-1} \right.$$
$$\left. -2 \sum_{s=1}^{a} K\left([x_t - c_s]h_s^{-1}\right) w_{t,T,s}^*(\omega) \cdot h_s^{-1} \right], \tag{22}$$

while for CDF for filtering:

$$LS_f^a(\omega, \beta_s) = \quad (T-m)^{-1} \sum_{t=m}^{T-1}$$
$$\int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_{t+1} < x\}} - \sum_{s=1}^{a} W\left([x - c_s]\beta_s^{-1}\right) w_{t,s}(\omega) \right]^2 dx \tag{23}$$

and finally, for exponentially smoothing for CDF we obtain:

$$LS_s^a(\omega, \beta_s) = \quad T^{-1} \sum_{t=1}^{T}$$
$$\int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_t < x\}} - \sum_{s=1}^{a} W\left([x - c_s]\beta_s^{-1}\right) w_{t,T,s}^*(\omega) \right]^2 dx. \tag{24}$$

2.3   Nonparametric recursions and direct parametric alternatives

In this section we rewrite nonparametric estimators given in (2), (3), (5) and (6) with appropriate recursions, which are more common for their direct alternatives, relying on the exponential weights, for parametric estimation of the time-evolving densities and distributions of financial returns. Since Harvey & Oryshchenko's (2012) dynamic kernel estimators allow modelling entire dynamic density and distribution of returns, we also describe several parametric alternatives, which follow the same estimation concept.

For the Gaussian PDF, J.P. Morgan's (1996) RM is often outlined by:

$$\sigma_{t+1}^2 = \omega \cdot \sigma_t^2 + (1 - \omega) \cdot x_t^2 \tag{25}$$

9

and equivalently expressed with:

$$\sigma_{t+1}^2 = \frac{1 - \omega}{1 - \omega^t} \sum_{i=1}^{t} \omega^{t-i} \cdot x_{t-i}^2$$

by employing exponential weights given in (1). RM in (25) is a well-known special case of Bollerslev's (1986) Integrated GARCH (1,1) model, given by:

$$\sigma_{t+1}^2 = c + A \cdot x_t^2 + B \cdot \sigma_t^2 = c + A \cdot (x_t^2 - \sigma_t^2) + (A + B) \cdot \sigma_t^2,$$

where $c = 0$, $B = \omega$ and $A = 1 - B$, and under the GAS framework of Creal et al. (2013) can be outlined with:

$$\mathbf{f}_{t+1} = \mathbf{f}_t + A \cdot s_t$$

if $A = 1 - \omega$, where $s_t = \mathbf{S}_t \cdot \dfrac{\partial \mathcal{L}_t}{\partial \mathbf{f}_t}$ for $\mathbf{S}_t = \mathbf{S}(\mathbf{f}_t, \mathbf{F}_{t-1}; \theta)$ and $\mathcal{L}_t = \log f(x_t | \mathbf{F}_{t-1}; \mathbf{f}_t, \theta)$ with $\mathcal{L}_t(\cdot)$ denoting the logarithm of the conditional PDF and $\mathbf{S}_t(\cdot)$ a scaling function, which as in Lucas & Zhang (2016) is the inverse diagonal of the Fisher information matrix and $\mathbf{F}_{t-1}$ is the information set available at time $t - 1$, $\mathbf{f}_t$ and θ are vectors of the time-varying and static parameters respectively. From Harvey & Oryshchenko (2012), we may also rewrite kernel estimator in (2) with the RM recursive form for PDF forecasts:

$$\hat{f}_{t+1}(x) = \omega \cdot \hat{f}_t(x) + (1 - \omega) \cdot h^{-1} \cdot K\left([x - x_t]h^{-1}\right). \tag{26}$$

Respectively, CDF estimator in (3) in the form of the one-step-ahead recursive CDF forecasts is:

$$\hat{F}_{t+1}(x) = \omega \cdot \hat{F}_t(x) + (1 - \omega) \cdot W\left([x - x_t]\beta^{-1}\right). \tag{27}$$

From (26), recursive form similar to GAS dynamics in (26) is then given as follows:

$$\hat{f}_{t+1}(x) = \hat{f}_t(x) + (1 - \omega) \cdot \mathbb{e}_t(x), \tag{28}$$

where $\mathbb{e}_t(x) = h^{-1} \cdot K\left(\dfrac{x - x_t}{h}\right) - \hat{f}_t(x)$, for PDF and from (27):

$$\hat{F}_{t+1}(x) = \hat{F}_t(x) + (1 - \omega) \cdot \mathbb{E}_t(x), \tag{29}$$

where $\mathbb{E}_t(x) = W\left(\dfrac{x - x_t}{\beta}\right) - \hat{F}_t(x)$, for CDF respectively.

On the other hand, recursions for estimators in (5) and (6) can be implemented in several steps if necessesary. First, obtain $\hat{f}_t(x)$ and $\hat{F}_t(x)$ from recursions in (26) and (27). Then, calculate backward recursions:

$$r_{t-1}(x) = \omega\left[r_t(x) + h^{-1}K\left([x - x_t]h^{-1}\right) - \hat{f}_t(x)\right]; \ \ t = T, \ldots, 2$$

for PDF,

$$R_{t-1}(x) = \omega\left[R_t(x) + W\left([x - x_t]\beta^{-1}\right) - \hat{F}_t(x)\right]; \ \ t = T, \ldots, 2$$

and for CDF. Finally, forward recursions are:

$$\hat{f}_{t|T}(x) = \omega \cdot \hat{f}_t(x) + (1 - \omega) \cdot \left[r_t(x) + h^{-1}K\left([x - x_t]h^{-1}\right)\right]; \ \ t = 1, \ldots, T$$

for PDF,

$$\hat{F}_{t|T}(x) = \omega \cdot \hat{F}_t(x) + (1 - \omega) \cdot \left[R_t(x) + W\left([x - x_t]\beta^{-1}\right)\right]; \ \ t = 1, \ldots, T$$

and for exponentially smoothed CDF estimate.

Recursions in (26), (27), (28) and (29) explicitly position estimators as in Harvey & Oryshchenko (2012) alongside exponential weights parametric specifications family, however given criticism of the Gaussian parametrization in (25), it is worthwhile highlighting its robust fully parametric alternatives.

First, to account for heavy tails property of financial returns with exponential weights, Guermat & Harris (2002) consider Laplace distribution with PDF for estimations given by:

$$f(x_t \mid \mathbf{F}_{t-1}; \mathbf{f}_t, \theta) = \frac{1}{\sqrt{2}\sigma_t} e^{-\dfrac{\sqrt{2}|x_t|}{\sigma_t}}, \tag{30}$$

while Lucas & Zhang (2016) extend it to the GAS (L-GAS) framework and parametrise its scale dynamics with:

$$\sigma_{t+1}^2 = \omega \cdot \sigma_t^2 + (1 - \omega) \cdot \sqrt{2}|x_t|\sigma_t. \tag{31}$$

Next, for the Student's t PDF described by:

$$f(x_t \mid \mathbf{F}_{t-1}; \mathbf{f}_t, \theta) = \frac{\Gamma\left(\dfrac{\nu_t + 1}{2}\right)}{\Gamma\left(\dfrac{\nu_t}{2}\right)\sqrt{(\nu_t - 2)\pi\sigma_t^2}}\left(1 + \frac{x_t^2}{(\nu_t - 2)\sigma_t^2}\right)^{-\frac{\nu_t + 1}{2}}, \tag{32}$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$, and with $\sigma_t^2 = \mathbf{f}_{1,t}$ and $\nu_t = 2 + \exp(\mathbf{f}_{2,t})$ Lucas & Zhang (2016) provide closed form recursions:

$$\mathbf{f}_{1,t+1} = \mathbf{f}_{1,t} + A_{\sigma_t^2} \cdot (1 + 3\nu_t^{-1}) \cdot \left(\frac{\nu_t + 1}{\nu_t - 2 + x_t^2/\mathbf{f}_{1,t}} \cdot x_t^2 - \mathbf{f}_{1,t}\right) \tag{33}$$

for σ_{t+1}^2 and: [10]

$$
\begin{aligned}
\mathbf{f}_{2,t+1} = \quad & \mathbf{f}_{2,t} - A_{\nu_t} \cdot \frac{2}{\nu_t - 2} \cdot \left(\gamma''\left(\frac{\nu_t + 1}{2}\right) - \gamma''\left(\frac{\nu_t}{2}\right) + \right. \\
& \left. \frac{2(\nu_t + 4)(\nu_t - 3)}{(\nu_t + 1)(\nu_t + 3)(\nu_t - 2)^2}\right)^{-1} \cdot \left(\gamma'\left(\frac{\nu_t + 1}{2}\right) - \right. \\
& \gamma'\left(\frac{\nu_t}{2}\right) - \frac{1}{\nu_t - 2} - \log\left(1 + \frac{x_t^2}{(\nu_t - 2)\mathbf{f}_{1,t}}\right) + \\
& \left. \frac{\nu_t + 1}{\nu_t - 2} \cdot \frac{x_t^2}{(\nu_t - 2)\mathbf{f}_{1,t} + x_t^2}\right),
\end{aligned} \tag{34}
$$

where $\gamma'(\cdot)$ and $\gamma''(\cdot)$ are the first and second order derivatives of $\gamma(\cdot) = \log\Gamma(\cdot)$, for ν_{t+1} under $A_\iota > 0$ restriction for both (33) and (34). The recursions in (33) and (34) (t-GAS) are notably more involved at the implementation stage than the standard RM specification, but they allow for the modelling of time-variation in the tails of financial returns within the functionality of the popular Student's t distribution as well as represent a good benchmark to position performance of Harvey & Oryshchenko's (2012) kernel estimators within the group of methods driven by the exponential weights.

---

[10] Note that setting $\omega = A \cdot (1 + 3\nu_t^{-1})$ in (33) provides a recursive form similar to (25), see Lucas & Zhang (2016) for details.

3 Data, estimated parameters and evaluation framework

To investigate the empirical performance of the approaches to parameters estimation outlined in Section 2.2, we constructed a data set of daily logarithmic returns for ten financial series. There were no specific criteria established for this data set composition. However, we aimed obtaining returns which cover at least one commonly recognised event of financial turbulence and posses common characteristics for logarithmic returns as consolidated by Cont (2001), to keep our modelling competition challenging for the methods considered. Some of the characteristics of these returns are given in Table 1. Overall, returns under investigation can be described as stationary, non-normal, heavy-tailed and with evident volatility clustering. For example, our data set offers high volatility periods of the "Dot-com" bubble collapse for NASDAQ composite, GBP/USD exchange rate "BREXIT" referendum outcome correction and BTC/USD exchange rate volatility in the cryptocurrencies markets.

There are also log-returns with very contrasting characteristics which may be worth investigating in more detail on their own. For example, NOK/USD exchange rate depicts lowest kurtosis value among all series. For a hydrocarbons exporting country this can be a notable characteristic given the last decade evolution of oil prices.[11] On the contrary, RUB/USD have the highest kurtosis value and is the most heavy-tailed among all series.[12] Tracking dynamic dependences of these two may be interesting to researchers in the contagion, central bank policy evaluation or oil exporting countries related disciplines. Therefore, we believe that our data set is sufficiently large to represent a fruitful field for empirical stress testing of the competing unknown parameters estimation methods and this class of estimators in general.

[Table 1 and Figure 2 around here.]

First, to obtain parameters with LS variations or ML in Section 2.2 a choice of the initialisation constant m is necessary. For example, Harvey & Oryshchenko (2012) initialise with $m = 100$,

---

[11]Note how the same characteristics are also different for BRENT series.

[12]Though the sample for RUB/USD is larger than for example for BTC/USD, it has heavier tails characteristics. This shows that increasing sample size not necessary implies "aggregate normality" features of financial returns in practice, keeping large samples potentially problematic for ML.

recommended minimum observations by Markovich (2008). On the other hand, Pérez (2012) discusses the choice of the initialising constant in the ML estimation setting and concludes that its choice has little impact on the CDF estimates. We make a more conservative choice in our empirical experiments and set $m = 250$. This has very pragmatic rationale and simple explanation, which is also discussed and supported by Markovich (2008). Nonparametric estimators are naturally data driven and a good precision of estimation, especially for weakly dependent data as financial returns, can be achieved by supplying estimators with sufficient observations for initialisation. However, it is also worthwhile highlighting that different research objectives or data constraints may lead to different choices of the initialising constant and therefore, may rely on the conclusions of Pérez (2012).

Second, for all binned estimations a binning strategy is necessary and may be important. We employ a binning strategy based on the sample quantiles. This is also common for spline fitting (e.g. Harrell, 2010) and have been successfully employed for adaptive kernel density estimation with binned estimators by Hazelton (2003). To be specific, for the fixed bandwidth binned estimators and $\xi_s(\tau)$ denoting sample quantiles, where $\tau \in (0, 1)$, we binned each sample in Table 1 over the interval $[\mathbf{min}(x_i) - 0.01; \xi_s(0.015, 0.03, 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.950, 0.970, 0.985); \mathbf{max}(x_i) + 0.01]$ producing a total of 15 bins with 15 bin centres under the simple linear binning as in Hall & Wand (1996).

A further worthwhile consideration for estimations is the choice of the appropriate kernel functional form. Due its wide availability in various forms for estimations with different software packages and since kernel's form is typically far from the most crucial component of the forecasting scheme as discussed by Robinson (1983) or Harvey & Oryshchenko (2012) among others, an applied choice is often the Gaussian kernel. However, given the discussion of Hall (1987b) on ML performance under different kernels, unbounded support of Gaussian kernel can theoretically benefit ML estimation routine, if the true data generating process exhibit tails decaying to zero at the same rate or faster as the kernel's tails. Moreover, unbounded support insures that estimated time-varying PDF is not zero at any estimation point and therefore, estimation with (11) and (12) should not collapse due to the fact that natural logarithm of zero is undefined. For kernels with unbounded support with ML routine, Harvey & Oryshchenko (2012) suggest ensuring that at all estimation points the density is not zero by setting it equal to the smallest number of machine tolerance. However,

from Hall (1987b) if the tails of the data generating process are heavier than the tails of the kernel for computations, the ML inevitably offers $h \longrightarrow \infty$. Moreover, for filtering, ML chooses unknown parameters under constrained information of future observations. This implies that at some estimation points of the sudden market corrections, the tails of the data generating processes can be expected to be heavier, since density gaps are not yet filled.

[Table 2 around here.]

Further in this work, we mainly concentrate on the estimations with Gaussian kernel, however also provide appropraite diagnostics and discussion with uniform and biweight [13] kernels for a more comprehensive view on the parameters for estimations. We choose uniform kernel as it is the most simple and less efficient kernel among the common set of kernels for estimations (e.g. see Härdle & Linton, 1994), to support the theoretical argument that kernel functional form is not the key modelling component for time series data as well. On the other hand, biweight kernel is just a marginally inferior form to the Epanechnikov kernel used in Harvey & Oryshchenko (2012) and may be valuable to perform comparisons to the Gaussian's average efficiency as discussed in Härdle & Linton (1994) or Wand & Jones (1995) among others for Gaussian, Epanechnikov and biweight kernels.

We illustrate obtained parameters for LS for PDF and CDF and their corresponding binned simplifications as well as for ML in Figure 2, while computing time required to obtain this parameters with R 3.5 in Table 2 using Gaussian kernels.[14] We employ convolution kernels from the kedd package of Guidoum (2015), while perform integration with "Simpson" integration algorithm from pracma package of Borchers (2018) over ten intervals. It is probably the simplest and the fastest approach to the integration available. From our experience, estimated parameters are almost identical to those obtained by more complex integration algorithms, however time required for evaluation is dramatically different. Perhaps, "Simpson" procedure delivers inferior results as data dimensions rise.

---

[13] Also often referred as the quartic kernel as in Härdle & Linton (1994) for example.

[14] All parameters used in our estimations, including those used for producing Figure 2, are available upon request from the corresponding author.

From Table 2 the fastest evaluations are typically performed with ML, then with binned LS for PDF, binned LS for CDF, LS for PDF and finally with LS for CDF. Estimations for filtering are less time consuming than estimations for smoothing due to the structural forms of the objective functions for filtering and smoothing. Binned LS for CDF is typically more demanding than LS for PDF, since it requires numeric integration. Overall, it is notable how binned estimators reduce the computational burden of the functions they simplify. However, further improvements in the computing time with binned estimators can be also achieved. First, more elaborate and computationally efficient binning procedure than linear binning we employ can be adopted (e.g. see Härdle & Linton, 1994; Hall & Wand, 1996). Second, necessary functions can be programmed in non-interpretive language such as, `C++` and later compiled with `Rccp` package of Eddelbuettel & François (2011) to perform computations in `R`. Finally, individual approach to binning can be employed for more efficient estimations. However, our goal here is to demonstrate that binned estimators can reduce computational time of our LS functions and keep evaluations accurate. We hope compiling estimators and estimations in our work into a full-scale `R` time-efficient package in the future, making it available for interested researchers and their future investigations.

Overall, from the illustrated parameters in Figure 2, optimal bandwidths parameters as per ML have the highest contrasts with the bandwidths reported for all LS functions and are higher in value. The only exception is the bandwidth for BTC/USD sample for binned LS for PDF when filtering. It is close to the obtained ML parameter. When inspecting parameters for smoothing, the outlook for BTC/USD is different and therefore, binned LS for PDF higher bandwidth for filtering this sample may be explained by the imperfections of our uniform binning rule imposed on all samples. Cryptocurrencies are known for their high volatility and may require more extensive and precise binning. On the other hand, binning can be also performed using dynamic quantiles obtained with a pilot CDF estimate. However, this adds another layer of complexity to the estimations. An alternative and pragmatic solution may be employing bandwidths obtained for smoothing. Since weights for exponential smoothing have less information restrictions, dynamic bins and bin centres are not necessary, shall effect the expected pattern of parameters less and from (4) forecasts with estimators for smoothing at $T = t$ should be identical to the forecasts for filtering. We, however, keep our evaluations to the exponential filtering as discussed in the Section 3.1 and provide parameters for smoothing just to highlight that ML has a notable tendency providing

higher bandwidths.

We do not note a clear difference in the reported patterns of ω parameter in Figure 2 by the considered approaches. However, it is worthwhile highlighting that binned estimators pick up ω relatively well given non-evolving bins we employ, though parameters can be noted as different from their respective full-scale LS estimations. Similar can be also observed for the bandwidths inspecting the output of the binned estimations. We evaluate how these differences in reported parameters impact the performance of Harvey & Oryshchenko's (2012) estimators in Section 4 with the criteria we describe in the next section.

3.1 Performance evaluation and PITs

Harvey & Oryshchenko (2012) employ a set of visual diagnostic procedures for PITs as prescribed by Diebold et al. (1998), restricted with exponentially filtered density forecasts. The procedure is standardised to the likelihood ratio (LR) test by Berkowitz (2001) and is convenient if the number of samples under investigation is not small.[15] For evaluation of the forecasts produced by estimators in (2) and (3) or any of their recursive forms in (26), (27), (28) or (29), first obtain PITs with:

$$\hat{F}_{t|t-1}(x_t) = \sum_{t=m}^{T} \left[\sum_{i=1}^{t-1} W[x_t - x_i]\beta^{-1}w_{t,i}(\omega) \right], \tag{35}$$

and then[16] if the PITs forecasts are adequate, $\Phi^{-1}\left(\hat{F}_{t|t-1}(x_t)\right)$ transforms, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal, are independent and normally distributed with $\mu = 0$ and $\rho = 0$ and $\sigma^2(\epsilon_t) = 1$ for $z_t - \mu = \rho(z_{t-1} - \mu) + \epsilon_t$, where $z_t = \Phi^{-1}\left(\hat{F}_{t|t-1}(x_t)\right)$. The test statistics is $\chi^2(3)$ distributed and may have a rejection cut off point set at the 5% significance level. Further we denote actual computed likelihood ratio statistic of Berkowitz (2001) for obtained PITs with LR_d, while corresponding p-values with LR_p.

---

[15]Evaluation as per Diebold et al. (1998) criteria typically requires analysing 4 to 5 diagnostic plots for each forecast produced. Hence, for example, for a setting of 10 samples and several approaches to parameters estimation for forecasting, a notably larger set of figures would be required, making Diebold et al.'s (1998) procedure less appropriate for reporting and then describing forecasting outcomes.

[16]If the parameters are obtained by the loss function in the form of PDF, (e.g. by ML as in Harvey & Oryshchenko (2012)) then $\beta = h$ is used for PITs computation.

3.1.1 Nonparametric goodness of fit tests and on the optimization with Cramer-Von Misses

From Diebold et al. (1998), if the forecasts are accurate, PITs obtained with (35) are expected to be independent and uniformly distributed in the $[0, 1]$ domain, implying that nonparmetric tests may be appropriate for testing their property of uniformity, such as Kolmogorov-Smirnov (K-S) and Cramer-Von Mises (CvM) tests.

The K-S test allows straightforward quantification of the departures from the uniformity of the PITs with:

$$K - S_d = sup|\dot{F}_k(\beta, \omega) - U_k|,$$

where sup denotes maximum value of the distances between PITs eCDF and their true uniform CDF, for which asymptotic critical values are available since Massey (1951) and allow straightforward $K - S_p$ values evaluation.[17] Wang et al. (2018) suggest exploiting uniformity of the PITs and select the unknowns for estimation by minimising:

$$CvM_d = \sum_k \left[\dot{F}_k(\beta, \omega) - U_k\right]^2, \tag{36}$$

where U_k denotes the true uniform CDF at some grid estimation point k and $\dot{F}_i(\beta, \omega)$ is the eCDF estimate of the PITs at the same estimation point, employing Artificial Neural Networks (ANN) algorithms. The goodness of fit of parameter is then inspected using eCDFs of the PITs and K-S test outputs. It is essential to recognise that (36) is the well-known Cramer-Von Mises statistics, though with simplified notation, for the goodness of fit (e.g. see Patton, 2013) and that Wang et al. (2018) suggest employing one nonparametric test for choosing the parameters for estimation and validating their appropriateness by another. We leave justifications of such practices to Wang et al. (2018), though some attempts in this direction are supplied by Oryshchenko (2016).

[Figure 3 and Table 3 around here.]

---

[17]$K - S_p$ values are likelihoods to observe a sample of the same size with the same maximum absolute distance from the target distribution. Hence roughly, the higher is the $K - S_p$, the greater are the chances that PITs are uniformly distributed. Similar pattern of significance is true for the other p-values of tests we employ in our work.

Moreover, let's consider the sample of NASDAQ log-returns as in Wang et al. (2018). It is illustrated in Figure 3 (a). For this sample, ANN trained by Wang et al. (2018) reports parameters: $\beta = 0.0087$ and $\omega = 0.9486$, when employing Gaussian kernel.[18] Now, skipping ANN step and under the same estimation settings, we optimise (36) directly with very standard Nelder-Mead optimisation routine and obtain the following parameters for this sample: $\beta = 0.00083$ and $\omega = 0.9917$. It is important highlighting that computation time of the direct CvM optimisation for this sample on our system took only ≈ 27.5 seconds. For comparison, evaluation of ML in (11) for $m = 50$ on the same system, takes ≈ 26.8 seconds. Though, we posses higher computing power, the improvement over the computational time reported in Wang et al. (2018) is positively and substantially different. Next, for the same sample, optimal parameters by the LS for CDF in (15) are $\beta = 0.2547$ and $\omega = 0.9778$ and by the binned LS for CDF in (19) under the same binning as described in Section 3 we obtain: $\beta = 0.2530$ and $\omega = 0.9823$. Now, for an ad hoc adaptive estimations, we binned Wang et al.'s (2018) NASDAQ data over the interval $[\mathbf{min}(x_i) - 0.01, \xi_s(0.1, 0.23, 0.37, 0.5, 0.67, 0.76, 0.9), \mathbf{max}(x_i) + 0.01]$ and computed bandwidths only for the bin centres defined for $[\xi_s(\tau = 0.1, 0.23, 0.37, 0.5, 0.67, 0.76, 0.9)]$ with LS for CDF in (23). This implies a total of six bandwidths parameter for the sample. Bandwidths for $[\min(x_i) - 0.01, \xi_s(0.1))$ and $(\xi_s(0.9), \max(x_i) + 0.01]$ are the same as the bandwidths for $[\xi_s(0.1), \xi_s(0.23))$ and $(\xi_s(0.76), \xi_s(0.9)]$. This way we effectively keep outliers involved in the bandwidths computations and avoid employing high bandwidths typically produced for the tails as for financial time series.[19] For faster optimisation of our adaptive estimators we pre-computed weights using ω for the binned LS for CDF since these parameters were available. Obtained adaptive set of bandwidths contains: $\beta_1 = 0.2938$, $\beta_2 = 0.1735$, $\beta_3 = 0.1366$, $\beta_4 = 0.1334$, $\beta_5 = 0.1544$, $\beta_6 = 0.2286$. Note how obtained set of the adaptive badnwidths meets a general expectation that β_1 and β_6 should take higher values than bandwidths for the middle of financial returns, since more smoothing is necessary in the tail domains, where data is typically scarce, unlike associated smaller values for the body domains, where data points are often abundant (e.g. Scott, 2015).[20]

---

[18]Note that these are essentially the parameters reported by Wang et al. (2018).

[19]Feluch & Koronacki (1992) as well as Stute (1992) experiment with parameters chosen by LS when certain data points are purposely omitted for kernel density estimation. Here, however, all data points are used, but certain bandwidths are omitted for estimations.

[20]We also experimented with this adaptive strategy on the data in Table 1. We found no substantial gains over the LS results we report and discuss in Section 4. These results are available upon request.

Adopting very similar in-sample PITs forecast evaluation setting as in Wang et al. (2018), just substituting visual eCDF of the PITs analysis with simple histograms, we provide diagnostic outlook in Figure 3 for the parameters reported in the previous paragraph. To clarify, Diebold et al. (1998) insist on histograms, since they are more informative than eCDFs and are simple for PITs uniformity analysis. Histogram evaluations and $K - S_p$ values for these parameters are provided in Figure 3 (b), (c), (d) and (e).[21] PITs histogram in Figure 3 (c) has "horns" indicating that parameters reported by Wang et al. (2018) do not capture tail behaviour properly, depicting severe departures from uniformity.[22] This is a clear sign that this forecast is not accurate. On the other hand, optimising CvM directly delivers the highest value for the K-S statistic. It also provides very appealing uniform shaped histogram of the PITs. PITs in histograms for the LS for CDF and adaptive estimation are also arguably uniform with the most of evaluations lying within 95% uniformity confidence intervals. K-S test outputs are higher than for the Wang et al.'s (2018) parameters and inferior to the direct optimisation of CvM in (36). Overall, we can observe that under the evaluation framework chosen by Wang et al. (2018), we can achieve better PITs diagnostic output by either LS routines or by simply skipping the ANN step.

Now given the fact that skipping ANN step notably reduces computational time necessary to obtain the unknown parameters, we would like to test how these parameters perform under more comprehensive evaluation criteria of Berkowitz (2001) with our data set in Table 1. These results are provided in Table 3. It is straightforward to see that reported parameters reported in Table 3 deliver very good K-S test output and fail forecasting all series under Berkowitz's (2001) LR test. It is important to recognise that we observe one of the key arguments by Diebold et al. (1998): good uniformity characteristics do not necessary imply good density forecasts. We can also note that typically values of the bandwidths parameters in Table 3 are substantially lower than those for the canonical methods in Figure 2. This may be problematic in the context of PDF evaluations. The bandwidths are drivers of the kernel estimators' bias-variance trade-off, which for stationary or close to stationary data in the time series framework, as logarithmic financial returns in Table 1,

---

[21]We omitted histograms for PITs by the parameters optimal for the binned LS for CDF as these parameters are close to the parameters by the full-scale LS for CDF, providing very similar histograms and K-S test outputs and are not very informative.

[22]The same can also be spotted, though under very carefull inspection, in the PITs eCDFs reported by Wang et al. (2018). Histograms just expose this limitation.

is expected to hold as argued by Robinson (1983).[23] To clarify, from Wand & Jones (1995) among others, the small values of the bandwidth constitute higher variance of the kernel estimators and downgrade estimates to histograms with large number of bins.

Therefore, upon the reported bandwidths parameters and test results in Table 3, it is valid suggesting that instead of the estimation path by Wang et al. (2018), one can employ an alternative nonparametric estimator based on the time-varying eCDF estimate, as per guidelines of Taylor (2007) for nonparametric quantile regressions and estimation framework of Harvey & Oryshchenko (2012) in general. That is:

$$\hat{F}_{2,t+1}(x) = \frac{1-\omega}{1-\omega^t} \sum_{i=1}^{t} \mathbb{I}_{\{x_i \leq x\}} \omega^{t-i}, \tag{37}$$

where $\mathbb{I}_{\{.\}}$ is an indicator function taking a value of 1 if the condition in the parenthesis is satisfied. The approach is arguably straightforward [24] at the estimation stage and can be directly used with the loss function in (15) for estimation of ω parameter. Since (37) is based upon the weights for exponential filtering, we can also rewrite it in the RM recursive form outlined by:

$$\hat{F}_{2,t+1}(x) = \omega \cdot \hat{F}_{2,t}(x) + (1-\omega) \cdot \mathbb{I}_{\{x_t \leq x\}}. \tag{38}$$

Similar, to the kernel approach of Harvey & Oryshchenko (2012), dynamic eCDFs in (37) or (38) and the empirical approach described in Taylor (2007) should be identical, but since in practice, quantile regressions may offer different dynamics for different quantiles, while here, as in Harvey & Oryshchenko (2012), exponentially declining weights are used to drive the dynamics of the entire time-varying eCDF, which is then used for time-varying quantile extraction. Hence, as noted in Harvey (2010), quantiles should not cross.

Estimator in (37) can also be rewritten to accommodate weights for exponential smoothing in (4) and take corresponding recursive forms as described in Section 2.3, however since, similar to Harvey & Oryshchenko (2012), our diagnostics is focused on the weights for exponential filtering and we regard approach in (37) as mostly appropriate for time-varying quantiles extraction, similar to RM and its variations in Section 2.3, only recursive forms used in the quantile evaluation phase of our empirical experiment are provided for brevity reasons.

---

[23]These expectations may be different for nonstationary data as shown by Aït-Sahalia & Park (2016).

[24]In fact, alongside Gaussian RM is the most straightforward among presented in our work.

Overall, before proceeding into the description of the performance evaluation techniques based on the estimated quantiles in the next section, we highlight that appropriately trained ANN maybe a worthwhile consideration for a future research, however it has to be a subject of more training constrains than used in Wang et al. (2018) to deliver accurate density forecasts. We now proceed to the next section, where we describe quantile based tests of Kupiec (1995) and Christoffersen (1998) employed in our empirical experiments.

3.2 Performance evaluation techniques based on the estimated quantiles

For simplicity, let VaR be outlined as:

$$\text{VAR}_{\iota,t+1,\alpha} = \Omega^{-1}_{\iota,t+1}(1-\alpha),$$

where Ω^{-1} is the inverse CDF, quantile function, of (30), (32), (27) or (38) for the risk confidence level $\alpha \in (0,1)$. For the nonparametric versions of RM we employ the quantile algorithm described in Harvey & Oryshchenko (2012), while quantile functions for PDFs in (30) and (32) are typically available within the standard set of `R` packages employed in time series econometrics, such as `rugarch` by Ghalanos (2018) for example. In Figure 4, we illustrate time-varying quantiles estimated with (30), (32), (27) and (38) for $\alpha = 0.9$, 0.95 and 0.99, similar to the applied work of Cheng & Hung (2011) among others, once again relying on BRENT returns from Table 1. When inspecting time-varying quantiles in the lower tail domain in Figure 4, it is straightforward to observe the role of kernels and their widths for estimations. eCDF based estimates are less smoothed than quantiles provided by the kernel enhanced estimator, since eCDF quantiles are only data reliant and tail areas are relatively data scares domains. Similar may be expected for the kernel quantiles, if the very small value of the bandwidth parameter is used for estimations as in Table 3 for BRENT returns when optimising (36) directly. On the other hand, inspecting parametric quantile dynamics, typical conservatism of the Laplace based estimations due to its fixed tail shape can also noted.

Now, for the number of VaR violations, $N = \sum_{t=1}^{T} \mathbb{I}_t$, with $\mathbb{I}_t$ denoting an indicator function taking a value of 1 everytime there is a larger loss than the VaR projection for the period of trading days

T, Kupiec (1995) suggests employing the following statistic:

$$\text{LRuc}_d(\alpha) = 2\left(\log\left[\left(\frac{N}{T}\right)^N \cdot \left(1 - \frac{N}{T}\right)^{T-N}\right] - \log\left[(1-\alpha)^{T-N} \cdot \alpha^N\right]\right). \tag{39}$$

The statistic assumes that $\frac{N}{T} = 1 - \alpha$ under a $\chi^2(1)$ distribution for $T \to \infty$ and is commonly known as the unconditional coverage LR test. Broadly, the statistic quantifies how well the VaR failure rate matches expectations. Christoffersen (1998) suggests a more inclusive procedure outlined by using:

$$\text{LRcc}_d(\alpha) = \text{LRuc}_d(\alpha) - \text{LRin}_d(\alpha), \tag{40}$$

where LRuc_d is the unconditional coverage in (39) and LRin_d is independence LR test outlined by:

$$\text{LRin}_d(\alpha) = 2\left(\log\left[\pi_{00}^{T_{00}}\pi_{01}^{T_{01}}\pi_{10}^{T_{10}}\pi_{11}^{T_{11}}\right] - \log\left[(1-\alpha)^{T_{01}+T_{11}} \cdot \alpha^{T_{00}+T_{10}}\right]\right), \tag{41}$$

where $\pi_{ij} = P(\mathbb{I}_t = j \mid \mathbb{I}_{t-1} = i) = \frac{T_{ij}}{T_{i0} + T_{i1}}$ for the first-order Markov chain transition matrix:

$$\nabla = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}$$

with T_{ij} accounting for transition times from states i and j for $i, j \in \{0, 1\}$. Since $\text{LRin}_d \sim \chi^2(1)$ for $T \to \infty$ the complete conditional coverage LR test in (40) follows $\chi^2(2)$ for $T \to \infty$. Essentially, the statistic in (41) tests the serial independence of VaR violations against a Markov first order dependence hypothesis, which in combination with (39) provides a joint test of independence and conditional coverage in (40).

[Figure 4 around here.]

Overall, our testing framework with K-S, CvM and LR tests for PITs and with LRuc and LRcc for VaR can be described as the most standard and widely recognised for densities and distribution (e.g. Patton, 2013) and quantiles (e.g. Nieto & Ruiz, 2016) accuracy evaluation respectively. K-S

and CvM both test uniformity, while LRuc tests expectations for the number of violations. LR and LRcc build upon the concepts of the K-S, CvM and LRuc tests, adding constraints of the first order independence. In the next section, we first evaluate in-sample characteristics of PITs using binned LS, full-scale LS and ML and then proceed evaluating in-sample lower-tail quantiles for the best performing approach for PITs estimation. We also evaluate quantiles performance by the RM and its eCDF, L-GAS as well as t-GAS parametrisations for completeness of our analysis.

4   Empirical performance and results discussion

We begin our empirical evaluations with the results of full-scale LS for PDF, ML and LS for CDF parameters and corresponding PITs in-sample K-S, CvM and LR diagnostics in Table 4. In Table 4 we present results for the returns in Table 1 and their transformations with ARMA & GARCH filter. Similar to Harvey & Oryshchenko (2012) we employ ARMA & GARCH filter outlined by:

$$y_{J,t} = \hat{\theta}_{J,0} + \hat{\theta}_{J,1} y_{t-1} + \hat{\theta}_{J,2} \epsilon_{J,t-1} + \epsilon_{J,t},$$
$$\sigma_{J,t}^2 = \hat{\theta}_{J,3} + \hat{\theta}_{J,4} \epsilon_{J,t-1}^2 + \hat{\theta}_{J,5} \sigma_{J,t-1}^2, \tag{42}$$

under the Student's t in (32) errors specification. Pre-filtering as in (42) formally transforms $K_h(\cdot)$ component of the estimators in (2) and (5) into:

$$(h\hat{\sigma}_{t|t-1})^{-1} K \left([x - x_t + \hat{y}_{t|t-1}] h \hat{\sigma}_{t|t-1}^{-1} \right)$$

and $W_\beta(\cdot)$ component of the estimators in (3) and (6) respectively to:

$$W \left([x - x_t + \hat{y}_{t|t-1}] \beta \hat{\sigma}_{t|t-1}^{-1} \right).$$

Harvey & Oryshchenko (2012) employ pre-filtering to overcome imperfections in the forecasting scheme of the employed nonparametric estimators. We, however, aim to obtain accurate PITs and quantiles evaluations without pre-filtering as exponential weights in similar research settings have repeatedly shown modest empirical performance on their own (e.g. Taylor, 2007; Boucher et al., 2014). Moreover, there is little time-variation to be picked up by the dynamic estimators

after pre-filtering. For example, from the LS for CDF estimated parameters, returns for NASDAQ composite have $\hat{\omega} = 0.9919$. This indicates an effective sample size of ≈ 566 observations out of 890 in total for the series after ARMA & GARCH correction. Therefore, pre-filtering may be more appropriate for the samples of a very large size as in Harvey & Oryshchenko (2012), with dynamic estimators employed for empirical justification of the time-varying parametric tails of Chavez-Demoulin et al. (2014)'s application to the ARMA & GARCH innovations, which is beyond the scope of our goals here.

[Tables 4, 5 and 6 around here.]

From Table 4 without pre-filtering, LS for CDF parameters deliver the highest number of forecasts, which can be classified as accurate at the 5% significance level. All PITs are uniformly distributed as shown by the K-S and CvM tests. Moreover, six out of the ten considered samples satisfy PITs independence property of the LR test without pre-filtering. These results highlight that LS for CDF picks up the right choice of the bandwidth parameter very well. On the other hand, results for the LR test before and after ARMA & GARCH filtering point out not sufficiency of the simple exponential filtering for evaluation of four samples. After pre-filtering all PITs are uniform and independent if the parameters are chosen by the LS for CDF. On the hand, a similar parameters as per LS for PDF also demonstrate similar PITs performance after pre-filtering and falling behind the LS for CDF in the number of appropriate uniformity characteristics of the raw returns at the 5% level. From (35), PITs are CDF based transformations and parameters optimal for PDF evaluations may be expected to have inferior performance. However, overall performance of the LS for PDF parameters is notably superior to the PITs outlook evaluated with the parameters provided by ML. ML provides only two correct samples evaluation under LR test. ML PITs outlook improves with pre-filtering, however, the number of correctly evaluated samples is still notably behind both full-scale LS functions without pre-filtering. Generally and as also highlighted by Li & Racine (2008), we observe that parameters optimal for PDF loss functions are tolerable for estimating CDF optimal output such as time-varying quantiles or PITs, but the technique must be robust to outliers and consistent in estimation. After pre-filtering the residual time-variation is typically

small and the choice of the bandwidth becomes the main mechanism for improving the accuracy of PITs evaluations. In our setting, this is achieved by the LS consistently, while ML fails insuring an appropriate choice of the bandwidth. This is also reflected in the K-S and CvM tests output without pre-filtering for ML.

Simplifying LS approaches with binned estimators also delivers good PITs characteristics as reported in Table 5. Number of the samples passing LR statistic at the 5% level is almost the same as for the full-scale LS functions with only NOK/USD binned LS for PDF evaluation falling behind the appointed significance level. We also may note that CvM statistic is behind the selected threshold with binned LS for PDF for pre-filtered BRENT sample, while LS for CDF struggles supplying PITs uniformity for RUB/USD without ARMA & GARCH filter as shown by the K-S and CvM output. We motivate loss in the PITs evaluation accuracy of the binned estimators with our strategy "one binning rule fits all". If each data set is binned individually with additional bins added where they are necessary, an accurate forecast ratio can be expected to be identical with the full scale LS functions. Given PITs characteristics by ML, computational time gains and basic approach to binning these can be argued as are more than satisfactory results. For example, given reduced computational burden binned estimators provide, in Table 6 we illustrate performance of ML and LS with binned estimators PITs performance for uniform and biweight kernels.

From Table 6, overall performance of parameters for different kernels by LS or ML has very similar pattern as in previously reported results in Tables 4 and 5. It is important highlighting that binned LS for CDF number of accurate PITs evaluations with both uniform and biweight kernels at the 5% confidence level is identical to the Gaussian kernel results reported in Table 5. This supports the argument that the choice of the optimal bandwidth is the main component of the accurate nonparametric evaluation. PITs are CDF optimal transformations and therefore, LS for CDF insures the name number of accurate evaluations with uniform, Gaussian and biweight kernels. On the other hand, with LS for PDF we may observe small estimation gains with biweight kernel over the results reported for Gaussian kernel in Table 5 binned and in Table 4 with full-scale LS for PDF parameters. Employing uniform kernel marginally reduces the number of correct binned evaluations in Table 5. Finally, ML again provides sub-optimal PITs and does not gain from employing a more efficient kernel for estimations. Generally, we can conclude that suggestion of Harvey & Oryshchenko (2012) for estimations with ML is an estimation trick preventing the

26

optimisation from collapsing with unbounded support kernels.

[Table 7 around here.]

From results based upon PITs in Tables 4, 5 and 6, LS for CDF can be concluded as the optimal technique for parameters estimations with kernel based estimations. Therefore, we report diagnostics for Gaussian kernel in-sample time-varying quantiles in Table 7 estimated with LS for CDF. Moreover, in Table 7 we provide similar evaluations for eCDF nonparametric, Gaussian, Laplace and Student's t based parametric evaluations driven by the (38), (25), (31), (33) and (34) respectively. First, evaluating 99% in-sample VaR confidence level in Table 7 we may observe that Gaussian RM performs the worst providing the highest number of violations. On the other hand, we may note another expected results for this VaR level. Laplace evaluations tend to overestimate quantiles, leading to the smaller number of violations than expected. The next highest number of matching expectations and first order independent quantiles are provided by the eCDF RM variation. Kernel based estimations provide a small gain over its nonparametric counterpart, arriving second at this VaR level. It is interesting pointing out that though the gains are small for the kernel based estimations, Actual over Expected violations ratio is typically closer to one for kernel than eCDF based evaluations. Yet dropping the bandwidth and kernel from the functional form for quantiles evaluation demonstrates an appealing performance given its simplicity for estimations. Student's t based estimations take their expected leading performance at this estimation level, given the higher flexibility and time-variation in its tails shape.

As we reduce our VaR confidence to 95% and 90% levels respectively, we may observe growing dominance of the nonparametric based estimations and kernel enhanced approach demonstrating the highest number of expected and independent violations. For 95% level, quantiles for all samples are adequately captured at the common 5% significance level, while only for RUB/USD violations do not pass the first order independence threshold. eCDF evaluations have similar output, however struggle with NIKKEI 225 at these levels. All parametric RM variations in Table 7 provide a very similar overall performance at the lower VaR confidence levels to each other, though RM may be worthwhile highlighting due to its improved number of correct evaluations when compared

27

to its performance at the 99% risk level. Overall, similar to Jones & Signorini (1997), we point out that in the tail domains at the high risk confidence levels, parametric assumptions may be necessary to achieve accurate quantiles evaluation consistently. On the other hand, if the target is to mine quantiles for nonparametric estimations as in Busetti & Harvey (2010), Harvey (2010) or Bücher et al. (2015) additional layer of kernels complexity or estimations of their parameters as in Wang et al. (2018) may not be appealing for practitioners as gains over more simple eCDF based evaluations are small. May be it is worthwhile considering extending semiparametric mixtures as in MacDonald et al. (2011) to the time series context by combining kernels of Harvey & Oryshchenko (2012) with parametric tails as in Chavez-Demoulin et al. (2014) for equally effective evaluations in the extreme and lower tail domains of the density and distribution of financial data. Finally, we note an interesting pattern, if the LS for CDF parameters pass the 5% significance level in the reported Berkowitz (2001) statistic in Table 4, quantiles for these parameters also tend to pass this significance threshold in Table 7 at the 99% level. This observation highlights the inefficiency of ML's inflated bandwidths parameters.

5   Concluding remarks

We have empirically demonstrated that exponentially weighted kernel estimators are valid for forecasting the densities and distributions of financial time series. Previously, to achieve accurate forecasts with these estimators Harvey & Oryshchenko (2012) had to correct for scale and location by applying ARMA & GARCH filter. We show that such pre-filtering may not be necessary, if the parameters are chosen by credible techniques for the unknown parameters estimation. Moreover, empirical evidence is given on the parameters obtained by ML after pre-filtering failing to insure accurate forecasts. We suggest modifying LS to the recursive forms appropriate for accommodating exponentially declining weights. If the sample size is large and computing power is limited, our results indicate that reducing computational burden with binned estimators and employing accordingly modified LS loss functions can be preferable. Such estimators empirically appear having a good ability identifying both the bandwidths and parameters governing the dynamics of exponential weights, though binning strategies may be an additional important component of accurate forecasting. If necessary binned estimators may also be used to vary bandwidths for the

tails and body domains of the modelled time series. Using binned estimators, we also demonstrate that our results and conclusions hold for the different choices of the kernel functional forms. Finally, we find that under the optimal choice of parameters, estimators as in Harvey & Oryshchenko (2012) provide accurate quantile evaluations, which are comparable to and may be preferred over, at certain confidence levels, the estimated quantiles by the more advanced, exponential weights driven and GAS based parametric competitors as in Lucas & Zhang (2016). On the other hand, we highlight that more simple exponential weights driven eCDF estimations also pick up in-sample quantiles well.

Overall, we provide a comprehensive empirical analysis, pointing out that the estimators as in Harvey & Oryshchenko (2012) are valuable addition to the semiparametric copula dependence family as in Patton (2012) or worthwhile consideration in the fully nonparametric copula dependence setting of Harvey (2010) if parameters are obtained with LS loss functions. Our comparisons to the relevant GAS models and highlights of Pérignon & Smith (2010) on the fact that practitioners may favour nonparametric methods for historical simulation of VaR measurement, indicate that the these methods have a good potential in the portfolios of models for risk exposure evaluation. Further research, may concentrate on the out-of-sample VaR performance of the Harvey & Oryshchenko's (2012) models, also performing relevant comparisons with quantile regressions of Taylor's (2007) and using more comprehensive VaR backtesting framework with other tests described in Nieto & Ruiz (2016). For such investigation, it may be also promising to consider modifying the methods of Sheather & Jones (1991), Altman & Léger (1995) or Polansky & Baker (2000) to accommodate exponential weights for equally fast and accurate parameters evaluation, since recursive daily updates of the estimated parameters for the out-of-sample VaR may be demanding even with our binned estimators. However, an alternative pragmatic, but valid strategy as highlighted by Ardia & Hoogerheide (2014) for ARMA & GARCH settings, may be performing parameters updates for VaR forecasts less often with the binned estimators. It may be also worthwhile improving upon our basic binning strategy to achieve accurate results more efficiently.

Another direction for further research may be consideration of other weighting schemes. For example, Harvey & Oryshchenko (2012) suggest investigating exponential filtering in the "stable" form similar to the variance targeting incorporated in GARCH models as in Francq & Zakoïan (2011). Employing simple indicator functions, as in the LS for CDF here, this is straightforward to

implement, however kernel estimators should also be in the appropriate recursive form as provided in Harvey & Oryshchenko (2012). Estimation of the parameters for these weighting schemes may be expected to be computationally demanding and it may be worth restricting the grid of estimation points to reflect modelling objectives more specifically. For example, for VaR estimations, grid may be restricted to reflect the domain of interest only, or the loss function may be further transformed to the specific quantile regression level as in Taylor (2007). This computational necessity implies abandoning the key feature of the entire density and distribution modelling, however to the best of our knowledge, performance of the "stable" filter nonparametric quantile regression for VaR modelling is still to be investigated. On the other hand, modelling entire distribution may benefit exploring tails asymmetries, medians and interquartile range changes over time as in Harvey (2010).

References

Aït-Sahalia, Y., & Park, J. Y. (2016). Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models. *Journal of Econometrics*, *192*(1), 119–138.

Altman, N., & Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, *46*(2), 195–214.

Ardia, D., & Hoogerheide, L. F. (2014). GARCH models for daily stock returns: impact of estimation frequency on Value-at-Risk and Expected Shortfall forecasts. *Economics Letters*, *123*(2), 187–190.

Arora, S., & Taylor, J. W. (2016). Forecasting electricity smart meter data using conditional kernel density estimation. *Omega*, *59*, 47–59.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.

Bessa, R. J., Miranda, V., Botterud, A., Wang, J., & Constantinescu, E. M. (2012). Time adaptive conditional kernel density estimation for wind power forecasting. *IEEE Transactions on Sustainable Energy*, *3*(4), 660–669.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Borchers, H. W. (2018). pracma: Practical numerical math functions. [Computer software manual]. (R package version 2.1.4)

Boucher, C. M., Daníelsson, J., Kouontchou, P. S., & Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, *44*, 72–92.

Bowman, A., Hall, P., & Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, 799–808.

Breiman, L., Meisel, W., & Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, *19*(2), 135–144.

Bücher, A., Jäschke, S., & Wied, D. (2015). Nonparametric tests for constant tail dependence with an application to energy and finance. *Journal of Econometrics*, *187*(1), 154–168.

Busetti, F., & Harvey, A. (2010). When is a copula constant? A test for changing relationships. *Journal of Financial Econometrics*, *9*(1), 106–131.

Chavez-Demoulin, V., Embrechts, P., & Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, *181*(1), 44–52.

Chen, S. X., & Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of financial econometrics*, *3*(2), 227–255.

Cheng, W.-H., & Hung, J.-C. (2011). Skewness and leptokurtosis in GARCH-typed VaR estimation of petroleum and metal asset returns. *Journal of Empirical Finance*, *18*(1), 160–173.

Chow, Y.-S., Geman, S., & Wu, L.-D. (1983). Consistent cross-validated density estimation. *The Annals of Statistics*, 25–38.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 841–862.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 863–883.

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. Retrieved from `http://www.jstatsoft.org/v40/i08/` doi: 10.18637/jss.v040.i08

Feluch, W., & Koronacki, J. (1992). A note on modified cross-validation in density estimation. *Computational statistics & data analysis*, *13*(2), 143–151.

Francq, C., & Zakoïan, J.-M. (2011). Merits and drawbacks of variance targeting in garch models. *Journal of Financial Econometrics*, *9*(4), 619–656.

Ghalanos, A. (2018). rugarch: Univariate garch models. [Computer software manual]. (R package version 1.4-0.)

Gouriéroux, C., & Jasiak, J. (2008). Dynamic quantile models. *Journal of econometrics*, *147*(1), 198–205.

Guermat, C., & Harris, R. D. (2002). Robust conditional variance estimation and Value-at-Risk. *Journal of Risk*, *4*, 25–42.

Guidoum, A. (2015). kedd: Kernel estimator and bandwidth selection for density and its derivatives. [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=kedd` (R package version 1.0.3)

Hall, P. (1982). The influence of rounding errors on some nonparametric estimators of a density and its derivatives. *SIAM Journal on Applied Mathematics*, *42*(2), 390–399.

Hall, P. (1987a). On kullback-leibler loss and density estimation. *The Annals of Statistics*, 1491–1519.

Hall, P. (1987b). On the use of compactly supported density estimates in problems of discrimination. *Journal of multivariate analysis*, *23*(1), 131–158.

Hall, P., & Wand, M. (1996). On the accuracy of binned kernel density estimators. *Journal of Multivariate Analysis*, *56*(2), 165–184.

Härdle, W. (1990). *Applied nonparametric regression* (No. 19). Cambridge university press.

Härdle, W., & Linton, O. (1994). Applied nonparametric methods. *Handbook of econometrics*, *4*, 2295–2339.

Harrell, F. E. (2010). *Regression modeling strategies.* New York: Springer.

Harvey, A. (1990). *Forecasting, structural time series models and the kalman filter.* Cambridge university press.

Harvey, A. (2010). Tracking a changing copula. *Journal of Empirical Finance*, *17*(3), 485–500.

Harvey, A., & Oryshchenko, V. (2012). Kernel density estimation for time series data. *International Journal of Forecasting*, *28*(1), 3–14.

Hazelton, M. L. (2003). Variable kernel density estimation. *Australian & New Zealand Journal of Statistics*, *45*(3), 271–284.

Jones, M. C., & Signorini, D. (1997). A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association*, *92*(439), 1063–1073.

J.P. Morgan. (1996). *RiskMetrics technical document 4th edition.* J.P. Morgan, New York.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*(2).

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: theory and practice.* Princeton University Press.

Li, Q., & Racine, J. S. (2008). Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, *26*(4), 423–434.

Lucas, A., & Zhang, X. (2016). Score-driven exponentially weighted moving averages and Value-at-Risk forecasting. *International Journal of Forecasting*, *32*(2), 293–302.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., & Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, *55*(6), 2137–2157.

Markovich, N. (2008). *Nonparametric analysis of univariate heavy-tailed data: research and practice* (Vol. 753). John Wiley & Sons.

Massey, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68–78.

Nieto, M. R., & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, *32*(2), 475–501.

Oryshchenko, V. (2016). Indirect maximum entropy bandwidth. *arXiv preprint arXiv:1607.03698*.

Patton, A. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, *110*, 4–18.

Patton, A. (2013). Copula methods for forecasting multivariate time series. In *Handbook of economic forecasting* (Vol. 2, pp. 899–960). Elsevier.

Pérez, A. (2012). Comments on kernel density estimation for time series data. *International Journal of Forecasting*, *28*(1), 15–19.

Pérignon, C., & Smith, D. R. (2010). The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking & Finance*, *34*(2), 362–377.

Polansky, A. M., & Baker, E. R. (2000). Multistage plugin bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, *65*(1-4), 63–80.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, *4*(3), 185–207.

Sain, S. R., & Scott, D. W. (1996). On locally adaptive density estimation. *Journal of the American Statistical Association*, *91*(436), 1525–1534.

Schuster, E. F., & Gregory, G. G. (1981). On the nonconsistency of maximum likelihood nonparametric density estimators. In *Computer science and statistics: Proceedings of the 13th symposium on the interface* (pp. 295–298).

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization.* John Wiley & Sons.

Scott, D. W., & Sheather, S. J. (1985). Kernel density estimation with binned data. *Communications in Statistics-Theory and Methods*, *14*(6), 1353–1359.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 683–690.

Stute, W. (1992). Modified cross-validation in density estimation. *Journal of Statistical Planning and Inference*, *30*(3), 293–305.

Taylor, J. W. (2007). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, *6*(3), 382–406.

Taylor, J. W., & Jeon, J. (2015). Forecasting wind power quantiles using conditional kernel estimation. *Renewable Energy*, *80*, 370–379.

Wand, M., & Jones, M. (1995). *Kernel smoothing* (Vol. 60). CRC Press.

Wang, X., Tsokos, C. P., & Saghafi, A. (2018). Improved parameter estimation of time dependent kernel density by using artificial neural networks. *The Journal of Finance and Data Science*.

Yu, K., & Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, *93*(441), 228–237.

Zhang, Y., Wang, J., & Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, *32*, 255–270.
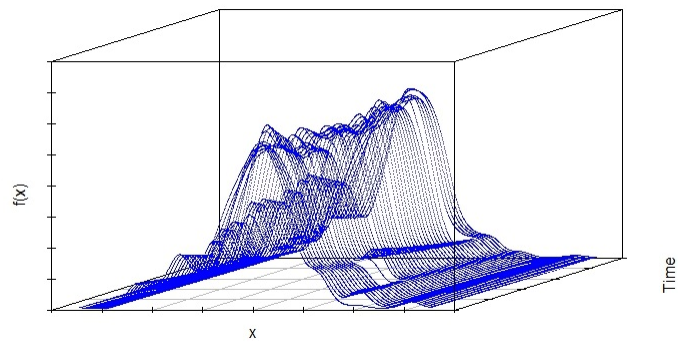
Figure 1: Illustration of the nonparametric time-varying PDF estimate based on the 27th of January 2009 - 7th of April 2009 window of BRENT log-returns from Table 1 with parameters obtained by LS for PDF as in Section 2.2.
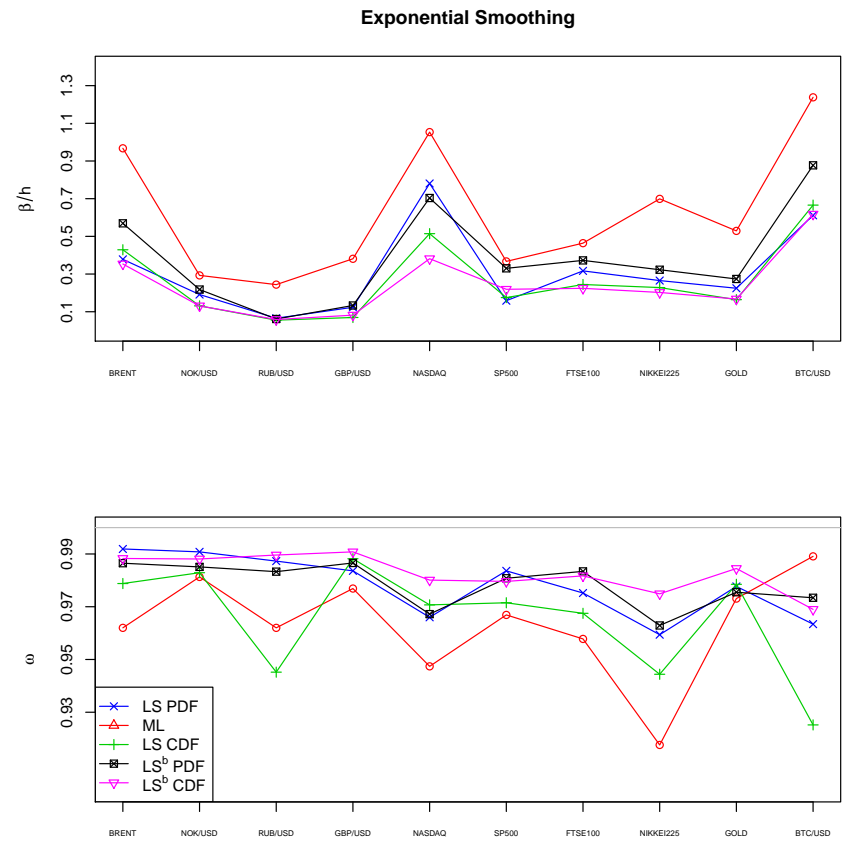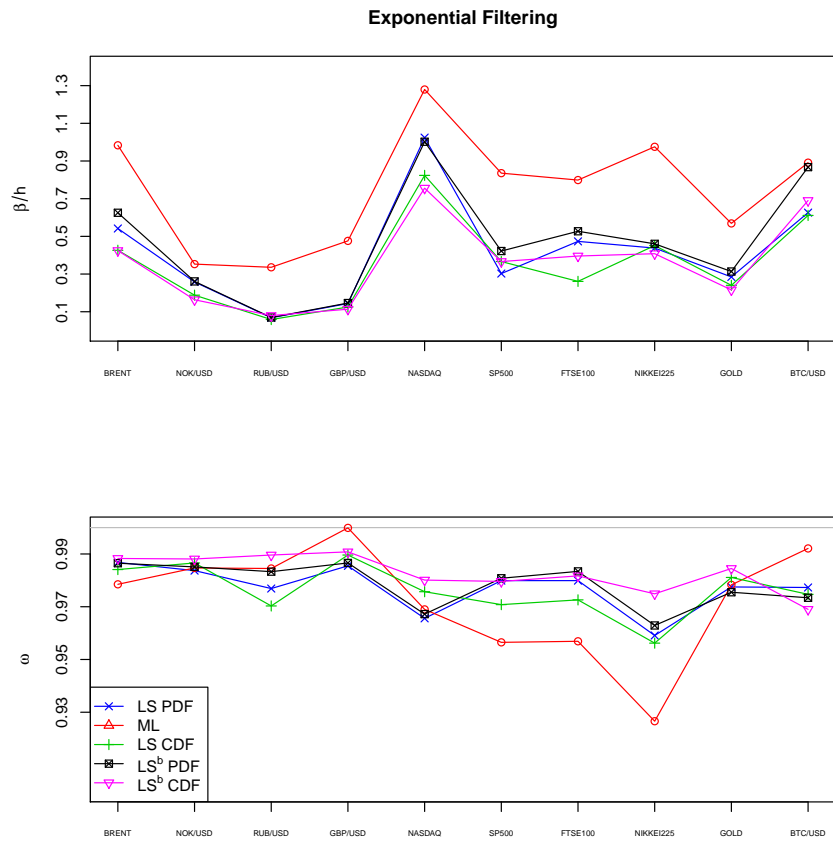
Figure 2: Unknown parameters for the log-returns in Table 1.
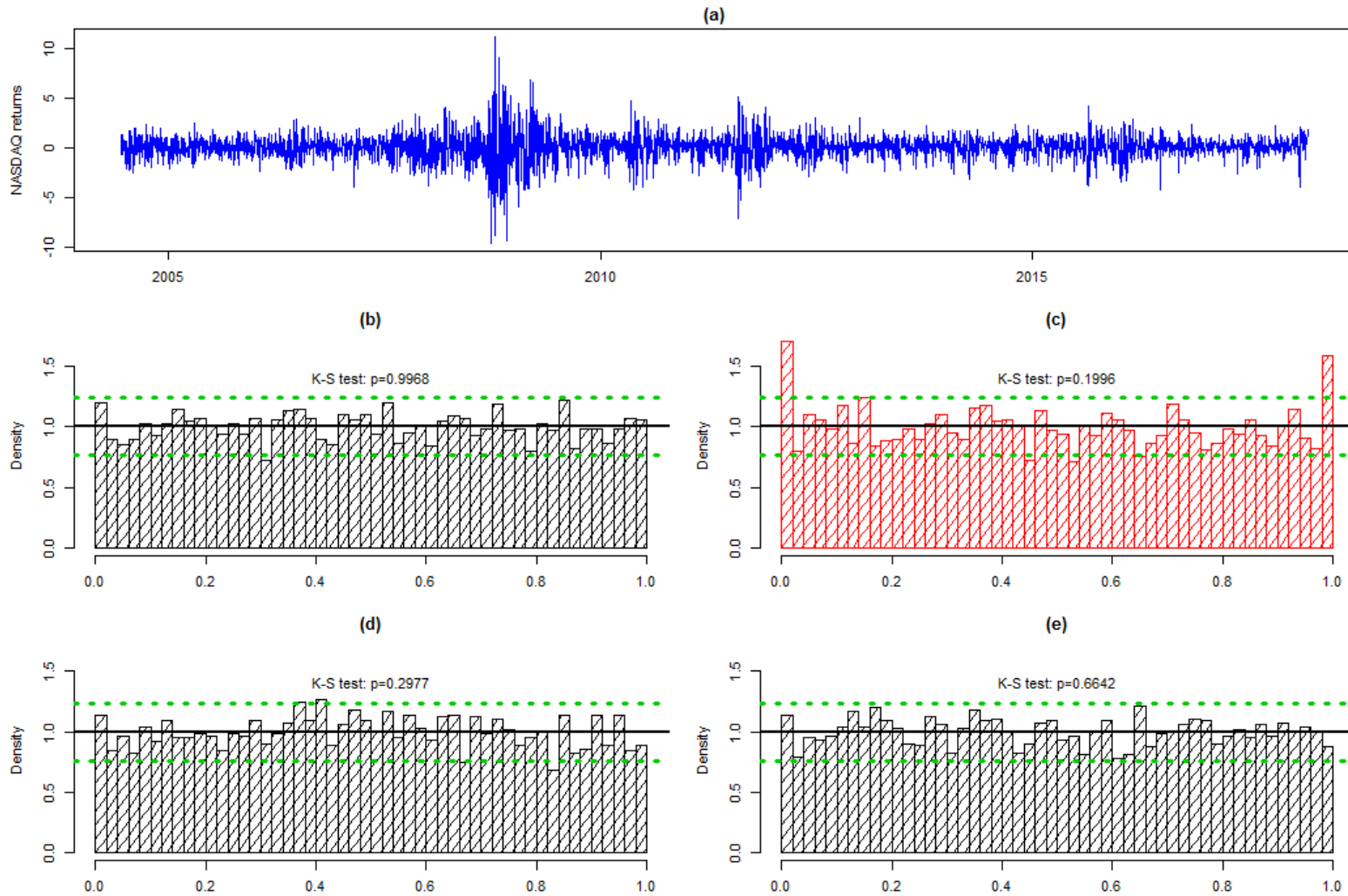
Figure 3: (a) NASDAQ daily logarithmic financial returns as in Wang et al. (2018). Histograms of the PITs for the NASDAQ returns and parameters (b) directly minimising CvM; $\beta = 0.00083$ and $\omega = 0.9917$, (c) ANN CvM as reported by Wang et al. (2018); $\beta = 0.0087$ and $\omega = 0.9486$, (c) LS for CDF; $\beta = 0.2547$ and $\omega = 0.9778$ and adaptive binned estimator for CDF; $\beta_1 = 0.2938$, $\beta_2 = 0.1735$, $\beta_3 = 0.1366$, $\beta_4 = 0.1344$, $\beta_5 = 0.1544$ and $\beta_6 = 0.2286$. Adaptive bandwidths are computed over the fixed binned LS for CDF ω parameter; $\beta = 0.2520$ and $\omega = 0.9823$. Lines parallel to the horizontal axis show ± 2 standard deviations confidence interval for PITs uniformity; $\pm 2\sqrt{(\kappa - 1)T^{-1}}$ where κ is the number of bins for histogram evaluation and is set at $\kappa = 50$ as in Diebold et al. (1998).
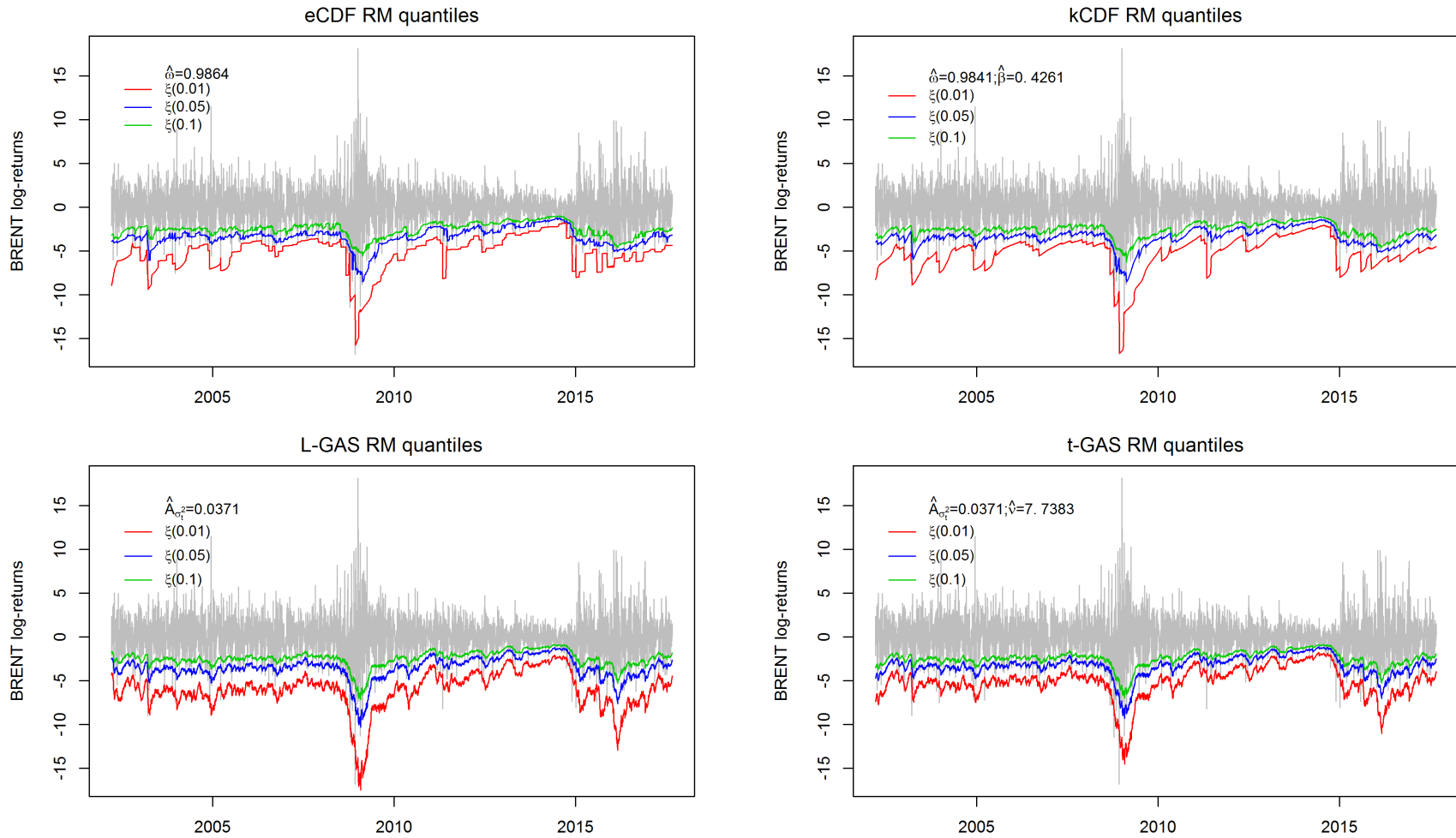
Figure 4: BRENT time-varying quantiles for the outlined confidence level and estimation method. For both nonparametric specifications, parameters for estimations are obtained employing LS for CDF in (15), while for parametric RM variations ML is employed as prescribed by Creal et al. (2013) and Lucas & Zhang (2016).

Date	Returns	T	MIN	MAX	Mean	St.d.	Skew.	Kurt.	ADF	LB(12)	LB(12)^2	JB	AH(12)
03.30.01-08.31.17	BRENT	4172	-19.89	18.13	0.019	2.23	-0.09	8.59	≈ 0.99	0.07	≈ 0.00	≈ 0.00	≈ 0.00
03.30.01-08.31.17	NOK/USD	4136	-5.92	4.42	-0.003	0.77	0.11	5.89	≈ 0.99	0.47	≈ 0.00	≈ 0.00	≈ 0.00
03.30.01-08.31.17	RUB/USD	4080	-12.86	10.94	0.017	0.83	0.76	41.62	≈ 0.99	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
03.28.13-03.27.18	GBP/USD	1303	-8.39	3.001	-0.005	0.58	-2.26	36.89	≈ 0.99	0.85	≈ 0.00	≈ 0.00	≈ 0.00
12.10.98-06.28.02	NASDAQ	890	-10.16	13.25	-0.036	2.47	0.16	4.73	≈ 0.99	0.06	≈ 0.00	≈ 0.00	≈ 0.00
01.03.06-03.01.10	S&P 500	1045	-9.46	10.95	-0.012	1.64	-0.22	11.35	≈ 0.99	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
01.03.06-03.01.10	FTSE 100	1052	-9.26	9.38	0.004	1.52	-0.09	6.92	≈ 0.99	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
09.01.88-02.28.92	NIKKEI 225	857	-6.82	12.43	-0.020	1.40	0.50	12.34	≈ 0.99	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
01.01.10-01.30.15	GOLD	1325	-9.51	3.98	0.011	1.10	-0.84	9.04	≈ 0.99	0.96	≈ 0.00	≈ 0.00	≈ 0.00
03.28.13-03.27.18	BTC/USD	1301	-60.1	51.70	0.349	6.14	-0.62	25.79	≈ 0.99	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00

Table 1: Descriptive statistics for the returns series, where T is the sample size, MIN is the minimum value, MAX is the maximum value, St.d. is standard deviation, Skew. is skewness, Kurt. is kurtosis, ADF is the augmented Dickey-Fuller probabilities for returns stationarity, LB(12) and LB(12)^2 are the Ljung-Box probabilities for no serial correlation of order 12 in returns level, x_t and squared demeaned returns, $(x_t - \bar{x})^2$ respectively, JB are the Jarque-Bera probabilities for normality and AH(12) are the Lagrange Multiplier test for autoregressive conditional heteroscedasticity probabilities of order 12 for no autocorrelation, normality and homoscedasticity. Data on the currency rates were obtained from the Central Banks' databases of the corresponding countries, all other stock indices and exchange rates were supplied by Bloomberg.

	Exponential Filtering					Exponential Smoothing				
	ML	LS^b PDF	LS PDF	LS^b CDF	LS CDF	ML	LS^b PDF	LS PDF	LS^b CDF	LS CDF
BRENT	39.98	899.69	139001.81	5733.33	19649.04	125.45	2454.13	166006.60	6033.23	55654.20
NOK/USD	19.11	765.54	79274.88	2272.83	22963.92	141.31	2120.46	164570.12	3886.09	39208.79
RUB/USD	145.51	614.78	232219.34	1756.73	32954.21	59.68	2284.51	310848.12	3334.89	169536.71
GBP/USD	2.91	26.89	1148.31	373.36	936.86	9.35	774.86	5559.79	1522.39	5543.89
NASDAQ	2.95	22.78	722.95	361.65	724.68	15.69	110.37	2966.01	392.14	2595.72
S&P 500	3.18	46.03	1663.22	259.33	1921.91	5.22	290.45	3854.78	700.57	5730.03
FTSE 100	1.93	49.87	1690.99	206.02	2732.8	4.72	152.43	2514.91	1202.57	5629.61
NIKKEI 225	1.52	13.29	392.67	204.05	542.71	5.64	175.69	866.43	321.94	2883.74
GOLD	3.45	35.01	1500.63	469.09	2040.91	20.08	153.77	2880.24	381.52	8125.35
BTC/USD	8.44	53.34	4356.53	1036.01	9483.17	19.40	460.05	8806.28	1193.89	18354.17

Table 2: Computing time for the specified estimation method in seconds. Computations were performed at the starting values for $h = \beta = 0.2$ and $\omega = 0.98$, under the same R, version 3.5, settings for parallel computing with 8-core 3.4 GHz Intel i7-6700 CPU, 16 GB RAM.

	BRENT	NOK/USD	RUB/USD	GBP/USD	NASDAQ	S&P500	FTSE 100	NIKKEI 225	GOLD	BTC/USD
CvM ω	0.9878	0.9956	0.9839	0.9793	0.9680	0.9879	0.9864	0.9379	0.9747	0.9757
CvM β	$1 \cdot 10^{-5}$	0.0239	0.0156	0.0677	0.0035	0.1920	0.2722	0.2668	0.1048	0.0896
$K - S_d$	0.0086	0.0480	0.0107	0.0114	0.0172	0.0143	0.0147	0.0278	0.0123	0.0158
$K - S_p$	0.9152	0.9999	0.7304	0.9957	0.9546	0.9827	0.9767	0.9519	0.9878	0.8984
LR_d	33.82	19.50	92.37	29.09	72.73	44.49	27.16	48.85	31.18	36.57
$\chi^2(3)_{0.95}$	7.81	7.81	7.81	7.81	7.81	7.81	7.81	7.81	7.81	7.81

Table 3: Parameters directly optimizing CvM statistic for the specified returns series. $K - S_p$ and $K - S_d$ are K-S test distances and probabilities for the PITs uniformity for these parameters. LR_d is Berkowitz (2001) computed statistics for the obtained CvM optimisation parameters and $\chi^2(3)_{0.95}$ is the cut-off point for the null hypothesis of normality and independence of the PITs at the 5% significance level.

#	Returns	LS PDF						ML						LS CDF					
		$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p	$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p	$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p
1	BRENT	0.0233	0.0281	0.6142	0.0206	2.1900	0.5339	0.0443	0.0000	3.1694	0.0000	57.7727	0.0000	0.0178	0.1651	0.3120	0.1251	1.6163	0.6557
2	NOK/USD	0.0194	0.1085	0.5297	0.0335	7.1148	0.0683	0.0316	0.0009	1.7247	0.0001	44.9830	0.0000	0.0140	0.4307	0.1553	0.3739	2.1934	0.5332
3	RUB/USD	0.0243	0.0215	0.6019	0.0221	30.6972	0.0000	0.0891	0.0000	15.9535	0.0000	294.5671	0.0000	0.0212	0.0631	0.4512	0.0531	53.8257	0.0000
4	GBP/USD	0.0178	0.8922	0.0645	0.7858	3.3947	0.3347	0.0824	0.0000	3.0255	0.0000	68.6819	0.0000	0.0162	0.9442	0.0371	0.9477	3.8881	0.2738
5	NASDAQ	0.0401	0.2549	0.1787	0.3128	0.5752	0.9021	0.0494	0.0874	0.3347	0.1082	5.5962	0.1330	0.0307	0.5817	0.1154	0.5145	0.8539	0.8365
6	S&P 500	0.0224	0.8200	0.0540	0.8520	30.2058	0.0000	0.0694	0.0009	1.3027	0.0005	34.9723	0.0000	0.0303	0.4564	0.1094	0.5409	24.7393	0.0000
7	FTSE 100	0.0289	0.5151	0.0821	0.6801	9.3721	0.0247	0.0573	0.0102	0.7678	0.0087	20.0786	0.0002	0.0191	0.9319	0.0441	0.9112	27.5035	0.0000
8	NIKKEI 225	0.0335	0.5039	0.1315	0.4514	13.2020	0.0042	0.0754	0.0020	1.3014	0.0005	16.7355	0.0008	0.0319	0.5673	0.1413	0.4173	12.8009	0.0051
9	GOLD/USD	0.0206	0.7506	0.1022	0.5744	2.6624	0.4466	0.0521	0.0058	1.2315	0.0007	15.2172	0.0016	0.0169	0.9183	0.0568	0.8341	3.7648	0.2880
10	BTC/USD	0.0290	0.3382	0.1609	0.3582	4.3831	0.2230	0.0430	0.0406	0.4424	0.0559	2.7624	0.4297	0.0286	0.3553	0.1494	0.3914	5.9063	0.1163
11	†BRENT	0.0177	0.1692	0.2531	0.1844	1.5100	0.6800	0.0202	0.0809	0.5290	0.0336	5.1859	0.1587	0.0149	0.3446	0.1649	0.3472	0.8867	0.8286
12	†NOK/USD	0.0144	0.3982	0.2037	0.2605	1.9522	0.5824	0.0295	0.0023	1.0989	0.0014	29.7748	0.0000	0.0127	0.5572	0.1191	0.4992	0.6944	0.8745
13	†RUB/USD	0.0104	0.8056	0.0890	0.6415	6.4206	0.0928	0.0391	0.0000	2.2931	0.0000	75.5268	0.0000	0.0113	0.7118	0.0785	0.7007	4.7126	0.1941
14	†GBP/USD	0.0230	0.6334	0.1012	0.5792	0.7165	0.8693	0.0718	0.0000	1.8893	0.0000	46.1361	0.0000	0.0188	0.8524	0.0775	0.7066	1.1620	0.7621
15	†NASDAQ	0.0412	0.2275	0.2742	0.1601	4.3581	0.2253	0.0445	0.1571	0.4331	0.0592	3.6784	0.2984	0.0304	0.5958	0.1693	0.3359	1.2531	0.7403
16	†S&P 500	0.0272	0.5959	0.1196	0.4971	2.8326	0.4182	0.0483	0.0488	0.7011	0.0126	13.8634	0.0031	0.0326	0.3661	0.1219	0.4880	1.4325	0.6979
17	†FTSE 100	0.0221	0.8257	0.0461	0.8997	0.2248	0.9735	0.0310	0.4213	0.1115	0.5313	2.2530	0.5216	0.0204	0.8909	0.0397	0.9346	1.1858	0.9798
18	†NIKKEI 225	0.0464	0.1452	0.1970	0.2733	1.4922	0.6841	0.0614	0.0204	0.5302	0.0334	5.1429	0.1616	0.0465	0.1436	0.1988	0.2699	1.9352	0.5860
19	†GOLD/USD	0.0300	0.2857	0.1529	0.3809	3.1701	0.3661	0.0564	0.0021	0.9174	0.0038	10.1915	0.0170	0.0340	0.1666	0.2204	0.2313	5.7729	0.1232
20	†BTC/USD	0.0207	0.7602	0.0836	0.6715	3.7965	0.2843	0.0456	0.0254	0.4409	0.0565	8.1782	0.0425	0.0179	0.8908	0.0507	0.8722	3.7895	0.2851

Table 4: PITs diagnostic output for LS loss functions and ML, where † denotes GARCH filtered returns.

		LS^b PDF						LS^b CDF					
#	Returns	$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p	$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p
1	BRENT	0.0265	0.0080	0.9061	0.0041	6.3049	0.0977	0.0180	0.1559	0.3367	0.1069	0.8236	0.8438
2	NOK/USD	0.0195	0.1038	0.5602	0.0281	8.1734	0.0426	0.0121	0.6179	0.0930	0.6203	2.3553	0.5020
3	RUB/USD	0.0227	0.0382	0.5747	0.0259	24.2755	0.0000	0.0251	0.0163	0.7039	0.0125	19.1349	0.0003
4	GBP/USD	0.0198	0.8030	0.0707	0.7471	3.0654	0.3817	0.0152	0.9680	0.0286	0.9806	4.6363	0.2004
5	NASDAQ	0.0392	0.2781	0.1701	0.3337	0.4938	0.9203	0.0300	0.6094	0.1025	0.5730	1.4859	0.6855
6	S&P 500	0.0340	0.3171	0.1485	0.3943	21.8548	0.0001	0.0303	0.4584	0.1075	0.5497	24.6007	0.0000
7	FTSE 100	0.0312	0.4160	0.1219	0.4881	8.2791	0.0406	0.0231	0.7828	0.0406	0.9303	12.8325	0.0050
8	NIKKEI 225	0.0351	0.4409	0.1467	0.3998	10.5487	0.0144	0.0325	0.5412	0.1119	0.5300	14.9636	0.0018
9	GOLD/USD	0.0239	0.5698	0.1464	0.4007	1.7859	0.6180	0.0158	0.9506	0.0404	0.9314	4.1936	0.2413
10	BTC/USD	0.0372	0.1082	0.3282	0.1128	2.9118	0.4054	0.0304	0.2871	0.1937	0.2801	3.8329	0.2801
11	†BRENT	0.0215	0.0530	0.4942	0.0412	7.5218	0.0570	0.0170	0.2057	0.2188	0.2339	1.2793	0.7341
12	†NOK/USD	0.0171	0.2037	0.3465	0.1005	5.4342	0.1426	0.0171	0.2062	0.2022	0.2633	1.4222	0.7003
13	†RUB/USD	0.0136	0.4747	0.1047	0.5625	3.4145	0.3320	0.0128	0.5580	0.0810	0.6863	3.7823	0.2860
14	†GBP/USD	0.0231	0.6299	0.1017	0.5765	0.7138	0.8700	0.0179	0.8881	0.0754	0.7192	1.3087	0.7271
15	†NASDAQ	0.0410	0.2309	0.2361	0.2072	4.9364	0.1765	0.0295	0.6320	0.1797	0.3106	1.2403	0.7434
16	†S&P 500	0.0302	0.4630	0.1119	0.5296	1.9034	0.5927	0.0293	0.5035	0.1122	0.5284	2.1362	0.5446
17	†FTSE 100	0.0265	0.6235	0.0720	0.7395	0.8707	0.8325	0.0165	0.9808	0.0329	0.9660	0.5545	0.9068
18	†NIKKEI 225	0.0530	0.0656	0.3222	0.1172	2.4390	0.4864	0.0471	0.1351	0.2158	0.2388	2.8964	0.4079
19	†GOLD/USD	0.0345	0.1539	0.1824	0.3043	2.4106	0.4917	0.0355	0.1334	0.2293	0.2173	4.3398	0.2270
20	†BTC/USD	0.0273	0.4142	0.1533	0.3798	3.5861	0.3098	0.0190	0.8402	0.0550	0.8456	3.7223	0.2930

Table 5: PITs diagnostic output for LS loss functions with binned estimators.

#	Returns	LS^b PDF						ML						LS^b CDF					
		$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p	$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p	$K-S_d$	$K-S_p$	CvM_d	CvM_p	LR_d	LR_p
1	†BRENT	0.0164	0.2433	0.2639	0.1715	1.2396	0.7435	0.1871	0.0000	63.6274	0.0013	1947.9432	0.0000	0.0166	0.2270	0.2940	0.1405	1.6137	0.6563
2	†NOK/USD	0.0259	0.0109	1.1317	0.0012	21.1928	0.0001	0.2044	0.0000	73.0365	0.0022	2476.5496	0.0000	0.0123	0.5968	0.0845	0.6664	3.0068	0.3906
3	RUB/USD	0.0297	0.0023	1.1211	0.0013	18.3081	0.0004	0.3897	0.0000	245.5043	0.0498	10807.3308	0.0000	0.0262	0.0102	0.7276	0.0109	24.6676	0.0000
4	†GBP/USD	0.0380	0.0946	0.4877	0.0428	3.4940	0.3215	0.3950	0.0000	66.0643	0.0015	2493.6417	0.0000	0.0159	0.9517	0.0297	0.9772	5.8598	0.1186
5	†NASDAQ	0.0294	0.6370	0.1051	0.5606	5.0687	0.1668	0.0880	0.0001	1.4188	0.0003	42.1589	0.0000	0.0291	0.6495	0.1050	0.5610	4.5820	0.2051
6	†S&P 500	0.0280	0.5619	0.0798	0.6934	30.0902	0.0000	0.2481	0.0000	22.0566	0.0000	704.7850	0.0000	0.0319	0.3944	0.1119	0.5298	29.0852	0.0000
7	†FTSE 100	0.0330	0.3457	0.1541	0.3774	9.9175	0.0193	0.2008	0.0000	13.6674	0.0000	401.5413	0.0000	0.0236	0.7603	0.0460	0.9002	15.5282	0.0014
8	†NIKKEI 225	0.0331	0.5187	0.1275	0.4661	15.3379	0.0015	0.2858	0.0000	22.4737	0.0000	736.0580	0.0000	0.0318	0.5687	0.1142	0.5199	18.9060	0.0003
9	†GOLD/USD	0.0276	0.3848	0.2045	0.2589	2.0246	0.5673	0.2701	0.0000	34.8524	0.0000	1156.4072	0.0000	0.0181	0.8738	0.0413	0.9264	4.5025	0.2121
10	†BTC/USD	0.0422	0.0470	0.4469	0.0545	2.3694	0.4993	0.0481	0.0153	0.5441	0.0308	3.1917	0.3630	0.0307	0.2747	0.2063	0.2556	4.4968	0.2126
11	‡BRENT	0.0243	0.0193	0.6962	0.0130	2.6188	0.4542	0.1154	0.0000	24.2924	0.0000	770.6095	0.0000	0.0182	0.1487	0.3364	0.1071	0.9293	0.8183
12	‡NOK/USD	0.0188	0.1300	0.4734	0.0466	5.7329	0.1254	0.1152	0.0000	26.0998	0.0000	980.7074	0.0000	0.0122	0.6118	0.0897	0.6380	2.5811	0.4608
13	‡RUB/USD	0.0198	0.0991	0.3871	0.0782	34.4131	0.0000	0.3388	0.0000	198.6414	0.0348	7806.0238	0.0000	0.0258	0.0122	0.7265	0.0110	20.6105	0.0001
14	‡GBP/USD	0.0161	0.9487	0.0461	0.8998	4.0479	0.2563	0.3398	0.0000	51.9919	0.0005	2014.2356	0.0000	0.0156	0.9607	0.0291	0.9790	5.5443	0.1360
15	‡NASDAQ	0.0310	0.5674	0.1117	0.5305	4.8091	0.1863	0.0543	0.0459	0.4222	0.0632	7.1817	0.0663	0.0300	0.6095	0.1025	0.5728	4.4784	0.2142
16	‡S&P 500	0.0329	0.3541	0.1240	0.4794	25.9269	0.0000	0.1742	0.0000	10.7487	0.0000	342.7140	0.0000	0.0315	0.4101	0.1119	0.5298	27.5953	0.0000
17	‡FTSE 100	0.0298	0.4738	0.1000	0.5852	9.5964	0.0223	0.1210	0.0000	4.9615	0.0000	151.5443	0.0000	0.0237	0.7570	0.0422	0.9218	15.2215	0.0016
18	‡NIKKEI 225	0.0334	0.5081	0.1257	0.4730	16.7086	0.0008	0.2048	0.0000	11.1018	0.0000	317.6945	0.0000	0.0337	0.4958	0.1114	0.5317	18.1261	0.0004
19	‡GOLD/USD	0.0228	0.6326	0.1169	0.5085	3.5607	0.3130	0.1805	0.0000	16.2434	0.0000	501.4592	0.0000	0.0166	0.9277	0.0418	0.9236	4.2839	0.2324
20	‡BTC/USD	0.0358	0.1350	0.2916	0.1428	3.5892	0.3094	0.0417	0.0514	0.4171	0.0651	2.6008	0.4573	0.0305	0.2818	0.1978	0.2717	3.9972	0.2618

Table 6: PITs diagnostic output for LS loss functions with binned estimators and ML employing uniform † and biweight ‡ kernels.

VAR_α	Returns	eCDF RM					kCDF RM					RM^{TM}					L-GAS RM					t-GAS RM				
		$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	AE	$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	AE	$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	AE	$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	AE	$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	AE
0.99	BRENT	0.1945	0.6592	1.1041	0.5758	1.0709	0.0012	0.9718	0.7849	0.6754	0.9944	14.2743	0.0002	14.9307	0.0006	1.6569	7.9661	0.0048	8.2375	0.0163	0.5863	2.2777	0.1312	2.4898	0.2880	1.2490
	NOK/USD	3.4881	0.0618	3.6339	0.1625	1.3124	0.0335	0.8549	0.6490	0.7229	1.0293	4.0510	0.0441	4.1716	0.1242	1.3378	25.7196	0.0000	25.7939	0.0000	0.3087	0.0004	0.9833	0.7912	0.6733	1.0033
	RUB/USD	20.0839	0.0000	20.4890	0.0000	1.8016	8.8417	0.0029	9.9340	0.0070	1.5144	17.7411	0.0000	17.7686	0.0001	1.7489	0.3481	0.5552	6.8314	0.0329	1.0963	2.2918	0.1301	4.3843	0.1117	1.2529
	GBP/USD	0.5446	0.4605	0.8699	0.6473	1.2346	0.0209	0.8851	0.2534	0.8810	1.0446	9.5844	0.0020	10.5233	0.0052	2.0873	0.5399	0.4625	2.6328	0.2681	1.2334	1.0401	0.3078	2.8734	0.2377	1.3283
	NASDAQ	0.0551	0.8144	0.2102	0.9002	1.0938	0.0258	0.8724	0.1395	0.9326	0.9375	1.0566	0.3040	1.1069	0.5750	0.6240	4.1917	0.0406	4.2043	0.1222	0.3120	1.0566	0.3040	1.1069	0.5750	0.6240
	SP500	11.1640	0.0008	12.0957	0.0024	2.3899	2.7187	0.0992	3.1515	0.2069	1.6352	21.2227	0.0000	21.3229	0.0000	3.0151	0.0002	0.9886	0.1628	0.9218	1.0050	2.7058	0.1000	3.1381	0.2082	1.6332
	FTSE100	3.6845	0.0549	4.1826	0.1235	1.7456	3.6845	0.0549	4.1826	0.1235	1.7456	18.7499	0.0000	20.1083	0.0000	2.8643	0.1139	0.7357	0.3182	0.8529	1.1208	7.6626	0.0056	8.4750	0.0144	2.1171
	NIKKEI225	15.5109	0.0001	18.1142	0.0001	2.9654	6.0216	0.0141	6.5916	0.0370	2.1417	11.2871	0.0008	12.1535	0.0023	2.6316	0.8175	0.3659	0.8706	0.6471	0.6579	3.2439	0.0717	4.9909	0.0825	1.8092
	GOLD	0.4459	0.5043	7.4895	0.0236	1.2093	0.4459	0.5043	7.4895	0.0236	1.2093	21.3388	0.0000	22.7661	0.0000	2.6952	0.4417	0.5063	2.5702	0.2766	1.2082	2.2419	0.1343	3.6634	0.1601	1.4870
	BTC/USD	1.0602	0.3032	1.4386	0.4871	1.3321	0.5541	0.4566	0.8801	0.6440	1.2369	16.3218	0.0001	18.2957	0.0001	2.4715	6.8247	0.0090	10.5002	0.0052	1.9011	0.6647	0.4149	4.6161	0.0995	0.7605
0.95	BRENT	1.0147	0.3138	1.7182	0.4236	1.0709	0.0065	0.9357	0.0168	0.9916	0.9944	3.9880	0.0458	6.9835	0.0304	1.1420	0.0183	0.8924	4.6077	0.0999	1.0094	7.6381	0.0057	12.0266	0.0024	1.1981
	NOK/USD	0.5019	0.4787	1.3330	0.5135	1.0499	3.6204	0.0571	3.9088	0.1417	0.8698	3.6341	0.0566	4.5508	0.1028	0.8696	23.6162	0.0000	25.4523	0.0000	0.6792	0.9870	0.3205	1.8366	0.3992	0.9313
	RUB/USD	0.9802	0.3221	3.2627	0.1957	1.0705	0.0014	0.9704	3.0144	0.2215	0.9974	0.0132	0.9084	2.1818	0.3359	0.9919	0.8840	0.3471	10.3352	0.0057	0.9345	2.9118	0.0879	18.8482	0.0001	1.1224
	GBP/USD	0.5548	0.4564	0.7654	0.6820	1.1016	0.1089	0.7414	0.1148	0.9442	1.0446	0.5441	0.4607	0.5576	0.7567	1.1006	0.1480	0.7004	0.3171	0.8534	0.9488	1.0208	0.3123	1.0808	0.5825	1.1385
	NASDAQ	0.5068	0.4765	0.5072	0.7760	1.1250	1.2618	0.2613	1.9986	0.3681	0.8125	1.9308	0.1647	2.0520	0.3584	1.2480	1.7602	0.1846	1.7608	0.4146	0.7800	2.4267	0.1193	2.6108	0.2711	1.2793
	SP500	1.6952	0.1929	2.0470	0.3593	1.2075	0.9868	0.3205	1.1895	0.5517	1.1572	8.5448	0.0035	12.7749	0.0017	1.4824	5.4889	0.0191	8.7397	0.0127	1.3819	12.1869	0.0005	13.2572	0.0013	1.5829
	FTSE100	0.0211	0.8845	1.5724	0.4556	1.0224	0.0211	0.8845	0.4056	0.8165	1.0224	2.3678	0.1239	2.6316	0.2683	1.2453	0.5945	0.4407	1.4317	0.4888	1.1208	5.2092	0.0225	5.2248	0.0734	1.3699
	NIKKEI225	4.2263	0.0398	7.2475	0.0267	1.3839	0.0928	0.7606	5.0919	0.0784	1.0544	7.3315	0.0068	12.6281	0.0018	1.5132	1.4149	0.2342	4.3465	0.1138	1.2171	6.4720	0.0110	10.2097	0.0061	1.4803
	GOLD	0.3452	0.5568	0.3516	0.8388	1.0791	0.0979	0.7544	0.4566	0.7959	1.0419	1.2567	0.2623	1.3665	0.5050	1.1524	0.5136	0.4736	0.5337	0.7658	1.0967	5.2828	0.0215	5.3059	0.0704	1.3197
	BTC/USD	0.0032	0.9493	3.5082	0.1731	1.0086	0.0418	0.8381	5.2739	0.0716	1.0276	0.0032	0.9549	0.6467	0.7237	1.0076	2.0413	0.1531	3.3244	0.1897	1.1977	1.1503	0.3054	2.8728	0.2378	1.1407
0.9	BRENT	0.4597	0.4977	3.9398	0.1395	1.0326	0.4256	0.5141	2.5315	0.2820	0.9689	0.0082	0.9280	4.4806	0.1064	1.0043	5.7223	0.0168	9.2945	0.0096	1.1165	7.8575	0.0051	9.9284	0.0070	1.1369
	NOK/USD	0.6703	0.4129	5.2508	0.0724	1.0396	1.1153	0.2909	5.7967	0.0551	0.9496	1.7779	0.1824	5.6929	0.0580	0.9365	0.2710	0.6027	8.7285	0.0127	0.9750	0.5006	0.4792	7.4935	0.0236	1.0342
	RUB/USD	1.5082	0.2194	10.0378	0.0066	1.0601	0.4217	0.5161	6.5723	0.0374	0.9687	4.3408	0.0372	8.6024	0.0136	0.9005	2.5352	0.1113	8.0136	0.0182	1.0780	10.9590	0.0009	20.9771	0.0000	1.1642
	GBP/USD	0.1430	0.7053	0.3333	0.8465	1.0351	0.1160	0.7334	0.2166	0.8974	0.9687	0.1230	0.7258	0.6519	0.7218	0.9677	0.0707	0.7903	0.1643	0.9212	1.0247	1.3747	0.2410	1.4740	0.4785	1.1101
	NASDAQ	0.4244	0.5148	0.4592	0.7949	1.0781	0.0701	0.7912	0.0701	0.9655	0.9688	6.9417	0.0084	7.0003	0.0302	1.3261	8.9772	0.0027	8.9783	0.0112	1.3729	11.2498	0.0008	11.2502	0.0036	1.4197
	SP500	3.1810	0.0745	5.6066	0.0606	1.1950	0.1690	0.6810	1.2973	0.5227	1.0440	2.0544	0.1518	6.6072	0.0368	1.1558	6.4914	0.0108	10.7818	0.0046	1.2814	7.6550	0.0057	12.5771	0.0019	1.3065
	FTSE100	2.1705	0.1407	2.1753	0.3370	1.1596	0.0200	0.8874	0.7400	0.6907	0.9850	9.9190	0.0016	9.9191	0.0096	1.0834	2.4761	0.1156	2.5861	0.2744	1.1706	3.2307	0.0723	3.2594	0.1960	1.1955
	NIKKEI225	4.0006	0.0455	7.6571	0.0217	1.2521	0.0307	0.8608	3.6746	0.1592	1.0214	2.1778	0.1400	9.2919	0.0096	1.1842	1.8141	0.1780	11.4713	0.0032	1.1678	2.9994	0.0833	10.7651	0.0046	1.2171
	GOLD	0.1254	0.7232	4.2109	0.1218	1.0326	0.0650	0.7987	1.5980	0.4498	0.9767	1.1959	0.2741	1.8503	0.3965	0.9015	0.2967	0.5859	1.2596	0.5327	1.0502	1.7898	0.1810	3.4530	0.1779	1.1245
	BTC/USD	0.2504	0.6168	0.8809	0.6438	1.0466	0.0380	0.8455	3.7792	0.1511	1.0181	1.9134	0.1666	3.0992	0.2123	0.8745	0.4793	0.4887	1.4073	0.4948	1.0646	0.7985	0.3716	2.0542	0.3581	1.0837

Table 7: VaR estimation results for the specified methods and VaR levels, where $LRuc_p$ and $LRcc_p$ denote p-values for Kupiec (1995) and Christoffersen (1998) statistics and AE is actual over expected exceedances ratio.