Essays on Financial Returns' Distributions Modelling with Applications

by

Artur Semeyutin

Department of Accounting, Finance and Economics

University of Huddersfield

SUBMITTED IN PARTIAL FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

University of Huddersfield

November 25, 2019

This thesis is lovingly dedicated to my favourite superhero, my mum, Nadezhda Levonovna.

Acknowledgements

I am deeply grateful and indebted to my mother and my uncle for their unconditional love and support through and beyond this academic journey. I would like also to take this opportunity to express my gratitudes to my supervisory team, Dr Robert O'Neill and Prof Jill Johnes, for their advice and guidance on writing this thesis. Furthermore, I am sincerely grateful to my former supervisor Prof Caroline Elliott for her kindness, guidance and involvement in my personal and professional development during the first half of my PhD journey and to Dr Eric Njoya for his openness, support and advice within the years of my PhD studies.

Copyright Statement

Declaration

I certify that this thesis is presented for examination of PhD in Economics and Econometrics degree at the University of Huddersfield and it is solely my work other than where I clearly indicated that it is my joint work with others. No portion of this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

All computations for producing this thesis were scripted and performed in R version 3.5 by R Core Team (2018). All unknown parameters for estimations were obtained with the optimization algorithms within R standard optimization routine optim. All functions for all evaluations are scripted and developed solely by me unless explicitly stated otherwise and therefore any mistakes in either R code or related computations due to that are solely my responsibility.

This thesis contains approximately **47 250** words in total and does not violate maximum number of words requirement.

I declare that material from Chapters 2 and 3 was used for the below academic publication:

Semeyutin, A., & O'Neill, R. (2019). A brief survey on the choice of parameters for: "Kernel density estimation for time series data". *The North American Journal of Economics and Finance, 50*, 1-20.

Abstract

This thesis consists of three main chapters. Its first and second chapters are concerned with univariate distributions modelling of financial data, while the third chapter has more applied and pragmatic financial scope of the Value-at-Risk estimations. First chapter targets developing new parametric distribution models for financial applications and suggests six of such models on the basis of Student's t distribution. Second chapter shifts to the field of nonparametric statistics for financial data in the time series estimations context and is concerned with selection of parameters for estimation of densities and distributions of financial returns with dynamic kernel methods. It compares performances of the dynamic kernel estimators under the parameters chosen by maximum likelihood and several least-squares routines. Third chapter, enriched with results from the previous substantive part, aims to position performance of the dynamic kernel estimator for distributions of financial returns within relevant group of methods for Value-at-Risk modelling and forecasting. Main chapters of this thesis are preceded by the preface section to summarize main motivations behind, provide overarching theme of the thesis, encompass some of its limitations and future research paths, which may follow from the conducted work, while main contributions of each chapter are also covered in the section concluding this work.

Contents

List of Figures

List of Tables

Preface

This thesis consists of three chapters united under the broad banner of financial economics. It contributes to the field of financial economics by expanding our knowledge of how, among many other approaches at our disposal, we can model financial data. Events of the financial crisis, the collapse of Lehman Brothers in particular, matched the first semester of my academic journey as an undergraduate student in economics and motivated me for this work. Embrechts & Hofert (2014) looking back to the several financial and economic crises, highlight the emergence of a new and mostly applied research field of Quantitative Risk Management (QRM). QRM encompasses subjects in mathematics, statistics, finance and economics, leads to the specific regulatory practices and serves to enhance our understanding, awareness and hopefully, preparedness for the future negative economic and financial outcomes. For financial disciplines, Embrechts & Hofert (2014) note that the Value-at-Risk and Expected Shortfall are among the key numerical foundations of our risk relevant practices. However, due to its relative simplicity, when compared to the Expected Shortfall, Value-at-Risk is a more common quantification tool in the extensive relevant literature as conscientiously summarized by Nieto & Ruiz (2016). Abad et al. (2014) in another detailed Value-at-Risk study point out the importance of the distributional speicifications behind the data when quantifying risks, whether in the time varying context as thoroughly discussed in Andersen et al. (2006) or in the time invariant context as in Gencay & Selcuk (2004). In fact, these models, forming our assumptions behind the data, are often so key that QRM research results and analysis by the one group of researchers may be pushed further by another, if sets of their distributions are different. For example, Corlu & Corlu's (2015) exchange rates associated risks analysis is deepened further by the same financial data investigation of Nadarajah et al. (2015) just by considering more elaborated distribution models. On the other hand, research efforts to enhance some distributions with more modelling power and flexibility may lead to more complex functional forms as in Papastathopoulos & Tawn (2013b) and later argued to replicate properties and abilities of the models with less complex mathematical expressions as in Nadarajah et al. (2013). The list of such representative examples is by no means complete, but should be sufficient to highlight the importance of distribution choices, when we model financial data to quantify risks or have other relevant estimation tasks. It also highlights the empirical nature of the QRM research, where our knowledge and understanding of the financial data often progresses through the "trial and error" documented experiments and later academic debate.

Parametric distribution models are abundant and although there are some commonly used ones listed in the works of Abad et al. (2014) and Nieto & Ruiz (2016), researchers have not given up exploring new and alternative forms to be applied for modelling financial data. For example, Nadarajah (2012) offers a list consisting of the sixteen models designed to extend the functional form of the Gaussian distribution by "plugging" other distributions into this model. On the other hand, Afuecheta et al. (2018) focus on a flexible, very common and natural extension to the Gaussian distribution, Student's t distribution. Afuecheta et al.

(2018) following the methodological path of Nadarajah (2012) for the Student's t distribution model provide new combined models, but yield a notably smaller list of distributions than Nadarajah (2012). Their list consists of three models. Afuecheta et al. (2018) also provide general expressions to help interested researchers expanding their portfolio of distribution models and their small-scaled empirical experiment supports considering these models for applications to financial data.

The first chapter of this thesis looks into new distributions and contributes to the list of already existing methods by combining half normal, Fréchet, Lomax, Burr, inverse gamma and generalized gamma models with the functional form of the Student's t distribution similar to Afuecheta et al. (2018). It provides necessary expressions for modelling financial data and conducts a small empirical test of these models. Embrechts & Hofert (2014) tracing QRM research origins, point out a fundamental work of McNeil et al. (2005), which in turn, recommends Generalized Hyperbolic distribution for applied tasks in financial economics among other models. Modelling power of Generalized Hyperbolic distribution in financial applications is rarely outperformed due to its flexible functional form, allowing it to incorporate many of the observed empirical properties of financial data and several distributions as special cases. Therefore, it is often a pragmatic benchmark for positioning new models for financial data within relevant research or applied contexts. From the small applied experiment conducted in the first chapter, outlined distribution models demonstrate a good potential, when compared to the well-established Generalized Hyperbolic distribution, to be considered in further more expansive data or list of models investigations with more applied scope than in Chapter 1.

Empirical comparisons for models in Chapter 1 are performed using a set of the log-likelihood based evaluation criteria of Akaike (1974), Schwarz et al. (1978), Hannan & Quinn (1979), Bozdogan (1987) and Hurvich & Tsai (1989). Though, it is found that at least one of the outlined distribution models is able to outperform Generalized Hyperbolic distribution, it is noteworthy pointing out that only one model achieves this consistently in the samples considered. That is Student's t generalized gamma model, a generalization encompassing all three models supplied by Afuecheta et al. (2018) and likely benefiting from the functionality offered by its number of parameters in the adopted estimations setting. On the one hand, obtained results once again confirm the validity of McNeil et al.'s (2005) suggestion on the Generalized Hyperbolic model consideration, but on the other hand, these findigs may be a demonstration that outlined models with fewer number of parameters, may require expansion of their functinality to achieve better modelling outcomes more consistently. For example in future investigations, these models may allow accounting for, often observed in the financial data, assymetry as in Fernández & Steel (1998). Furthermore, similar to the Student's t gamma mixture model in Afuecheta et al. (2016) these models, alongside some of the others listed in Afuecheta et al. (2018), can be applied in the time series framework of GARCH models for Value-at-Risk forecasting. Also as in applications of Wichitaksorn et al. (2015) for similar distributions, investment decisions upon the parsimonious CAPM models' output, as listed in Gregoriou (2018) among others, where residuals are specified with the distributions provided in this chapter could be considered in the future investigations.

Chapter 2 of this thesis also looks into distributions modelling of financial returns, but focuses on the nonparametric techniques for financial data. Nonparametric estimators are well-known in the time invariant estimation contexts and unlike parametric specifications, may be argued as notably less often in the time evolving financial applications. For example, Nieto & Ruiz (2016) in their all-inclusive review devote a small discussion on the time evolving nonparameteric methods in the form of the several quantile regressions only, avoiding empirical examples of such methods due to their computational "heaviness" and complexity. Alternatively, Harvey & Oryshchenko (2012) show that nonparametric methods may not necessarily be restricted to the quantile regressions and can be used to estimate the entire time evolving density and distribution of financial data. Though Harvey & Oryshchenko's (2012) ideas may be similar to those employed to drive the quantile regression processes, the approach is more appropriate to be regarded as the most nonparametric version of the well-known special case Integrated GARCH model for volatility estimations or simply as the fully nonparametric version of RiskMetrics™ by J.P. Morgan (1996). Similar to Nieto & Ruiz (2016), Harvey & Oryshchenko (2012) recognize computational demands of the nonparametric methods and employ maximum likelihood to reduce overall computational burden using this type of estimators. Conducted empirical work and tests in this chapter provide evidence that methods as in Harvey & Oryshchenko (2012) are indeed valid and should be attractive for financial applications, however, computational short-cut of employing maximum likelihood is less preferable than other pragmatic alternatives considered in this part of the thesis. Presented ideas for estimations may be also similar to those employed in the nonparametric quantile regression contexts, but unlike Harvey & Oryshchenko (2012), this chapter contributes by arguing for nonparmetric estimations with least-squares routines. Empirical comparisons for the sets of parameters obtained in Chapter 2 using different techniques for estimations are performed using Kolmogorov-Smirnov and Cramer-Von Mises tests as well as Berkowitz (2001) compound forecasts evaluation criteria. It is found that least-squares based estimation techniques lead to parameters allowing estimators as in Harvey

& Oryshchenko (2012) to accurately forecast notably more in-sample evaluations than with parameters chosen by maximum likelihood. These findings remain unchanged if each financial series in the data set is adjusted for the volatility bursts by the ad-hoc ARMA-GARCH model specification. Also, given maximum likelihood's perk of the faster computations, it is found that rearranging Harvey & Oryshchenko's (2012) estimators to accommodate binned financial data may lead to the same computational time and better forecasts than with parameters by maximum likelihood. Based on the binned estimators, dynamic adaptive estimations are also considered and although these forecasts accuracy may be less appealing given obtained results and higher number of parameters involved, there may be valuable insights for further research with these bandwidths empirical experiments. For example, from the conducted work with parameters, it may be worthwhile combining kernels methods with parametric models for the tails as in MacDonald et al. (2011), but in the dynamic context of exponentially declining weights as in Harvey & Oryshchenko (2012). For such estimations, ad hoc adaptive strategy computations as in Chapter 2, suggest that least-squares may be rearranged to pick up parameters optimal for the body domains of densities and distributions of financial data, areas estimated nonparametrically by the dynamic kernel estimators in the semiparametric settings. Yet, given satisfactory diagnostic output for the financial data as obtained in this chapter for the parameters chosen with least-squares routines, these estimators are valid in the semiparametric contexts of bivariate copula QRM research agenda as described in Patton (2012, 2013) or higher copula dimensions modelling as in Nikoloulopoulos et al. (2012).

Chapters 1 and 2 of this work are empiric in their nature, similar to other QRM research, but may be too formal for the applied finance practitioners' discussion. Nieto & Ruiz (2016) outline that applied research questions are often narrowed down to the very pragmatic approximations of the "correct frequency of the losses" when holding financial assets, where losses are usually a financial asset returns' exceedances of the Value-at-Risk estimates at the predefined confidence level. Moreover, Nieto & Ruiz (2016) point out that for such estimations, nonparametric estimators avoid any misspecification of the assumptions behind the data formed by the parametric distribution models, as in Chapter 1 for example, and thus, should be appealing for the applied QRM research. Therefore, given Harvey & Oryshchenko's (2012) kernel estimator performance under the parameters obtained with least-squares routines, Chapter 3 conducts a small-scaled financial experiment to make applied comparisons of the Harvey & Oryshchenko's (2012) type nonparametric estimators with some of the direct and most up-to-date parametric competing methods presented in Lucas & Zhang (2016).

Empirical comparisons for the small Value-at-Risk estimation results in Chapter 3 are performed using applied tests of Kupiec (1995) and Christoffersen (1998) at some of the most common risk confidence levels as in Cheng & Hung (2011) among others. Obtained results do not provide an ultimate specification for the applied use under the weighting scheme as in Harvey & Oryshchenko (2012) and Lucas & Zhang (2016), however they demonstrate a modest and most importantly stable performance of the kernel estimator as in Harvey & Oryshchenko (2012) under the diligent choice of parameters. From the obtained results functionality of the kernels demonstrate notable gains over the simplified empirical time evolving distribution specification of the financial data, performs consistently well, when compared to other methods in the pool, at the lower risk confidence levels and is slightly inferior to the parametric specifications at the highest risk level. Moreover, given that at the highest Value-at-Risk confidence level, best out-of-sample performance is provided by the parsimonious parametric specification known for yielding "conservative" risk strategies, it may be valid arguing kernel method as in Harvey & Oryshchenko (2012) as a standard benchmark for assessing validity of the new methods for Value-at-Risk estimations in the relevant financial studies rather than, often employed, original RiskMetrics™ or more technically elaborated approaches under the same weighting scheme. Apart from modest and stable results, the approach posses endemic flexibility of the nonparametric methods and has arguably straightforward functional form, however, may still require techniques for faster computation of the appropriate parameters. This may be worthwhile addressing in the future studies. For now, if computational time for rolling daily re-estimations of the Value-at-Risk forecasts is considered burdensome, such parameters re-estimations may be less often as argued by Ardia & Hoogerheide (2014). Also, in-sample results in Chapters 2 and 3 indicate a good fit of the Harvey & Oryshchenko's (2012) estimators for financial assets time-evolving dependence investigations with the nonparameteric methods by Busetti & Harvey (2010), Harvey (2010) and Bücher et al. (2015), since estimations for such "in-sample exceedances count based" evaluations with the least-squares routines discussed in Chapter 2 are notably less time consuming.

Most of the conducted work in Chapters 1, 2 and 3 limitations may be unpacked from the future research paths outlined above. However, it is worthwhile emphasizing the largest obstacle encountered while working on this thesis. This consists of two components: initial intellectual capacity/knowledge of and time allowed to contribute to the chosen niche in the broad discipline of modern economics. Roughly, this thesis was first influenced by the collapse of Lehman Brothers, known and highlighted conservative pragamatism of practioners by Pérignon et al. (2008) and arguments of Embrechts & Hofert (2014) on the importance of risk modelling and the growing role of copula models for the task of minimizing "Lehman" precedents in the future. Therefore among abundant academic literature on applications of copula models in finance, this thesis was initialized with the parsimonious copula approach of Harvey (2010), which in turn, is explicitly linked to the work of Harvey & Oryshchenko (2012). This may explain the subjective component, if any, in locating the second chapter at the center of this work.

Analytical and critical assessment of the ideas expressed in Harvey & Oryshchenko (2012) as well as their replication supplied a necessary skills-set required for thorough understanding of the extensive technical material as presented by Nieto & Ruiz (2016) among others. Accumulated knowledge and related experiments lead to the second chapter, while growing interest in the approximations of data generating processes of financial returns lead to the developing parametric models in the first chapter. On the other hand, given complexity of the models presented in Chapter 1 and relative inconsistency of their majority, though in a small-scaled empirical experiment, in overcoming already existing models[1] third chapter took a turn towards more pragmatic research objectives. Applied results from Chapter 3 demonstrate that in QRM more elaborated approaches do not necessariluy imply better modelling outcomes. This may shape the view that in the modern abundancy of the methods for risk modelling, it is worthwhile performing more large scale empirical investigations with relatively simple and uncomplicated research targets similar to P. R. Hansen & Lunde (2005), Ardia & Hoogerheide (2014), Liu et al. (2015) and Ardia, Bluteau, et al. (2018) among others.

[1]Note that from McNeil et al. (2005) Generalized Hyperbolic distribution encompasses several popular distribution models in finance such as Normal Inverse Gaussian and generalized Student's t model of Aas & Haff (2006) for example.

This general conclusion does not necessesarily diminish the value of attempts in suggesting more elaborate or new and innovative approaches for risk modelling, but highlights that it may be worthwhile to begin collecting empirical facts on the various methods in practice of the data comprehesive (larger-scale) investigations. Such large scale empirical stress tests should have explicit empirical findings and allow to effectively summarize obtained methods' performance results. Similar had been implemented for empirical properties of financial returns by Cont (2001) among others and could be a prominent milestone to facilitate efficiency of further QRM research. Unfortunately, similar research path suggestions may come at the later stages of a long academic journey, after initial knowledge of the studied discipline is substantially upgraded with both documented and undocumented empirical experiments.

Chapter 1

Models for financial returns based on the Student's *t* distribution

1.1 Introduction and motivation

Student's t distribution of Gosset (1908) is the most common and parsimonious parametric choice for estimations in economics, insurance and finance (e.g. Abad et al. (2014); Nieto & Ruiz (2016)). It not only allows parametrizing leptokurtic features of financial data, but also can serve a foundation to accomodate other data properties (e.g. such as volatility clustering) as summarized for financial returns by Pagan (1996) and Cont (2001). Some notable modifications to its functional form are provided by B. E. Hansen (1994), Fernández & Steel (1998), Theodossiou (1998), Jones & Faddy (2003), Sahu et al. (2003), Bauwens & Laurent (2005), Aas & Haff (2006), Zhu & Galbraith (2010) and Papastathopoulos & Tawn (2013a) and are widely applied beyond the Bayesian finite and infinite variance (e.g. Tucker (1992)), Markov regime switching means, variance and mixing weights (e.g. Perez-Quiros &

Timmermann (2001)) as well as multivariate stochastic volatility (e.g. Wang et al. (2011)) models. A detailed review of the various Stundent's t modifications/generalizations for broad financial and other applied tasks is provided in R. Li & Nadarajah (2017), but the list is still by no means complete.

One of the Student's t popular generalizations, often recommended for risk quantification in finance (e.g. as in McNeil et al. (2005)), is Generalized Hyperbolic (GHYP) distribution. GHYP distribution model offers a flexible functional form allowing data asymmetry, is classified as a normal mean-variance mixture distribution and has Student's t as one of its special cases. Other normal mean-variance distributions are also widespread and not uncommon. For example, mixing of this type can be traced back to Press (1967) and Praetz (1972) and is followed by Andrews & Mallows (1974), Barndorff-Nielsen (1977), Barndorff-Nielsen et al. (1982), Kon (1984), West (1987), Madan & Seneta (1990), Madan et al. (1998), Tjetjep & Seneta (2006), Luciano & Semeraro (2010), Geweke & Amisano (2010) and Nadarajah (2012) among others. Typically, these compound models capture heterogeneous characteristics of financial data by randomizing one of the parameters of a "parent" distribution with the functional form of an appropriate "mixing" distribution (e.g. McDonald & Butler (1987); Hoogerheide et al. (2007); Ardia et al. (2009a,b)).

Unlike previous compositions and similar to Gencay & Selcuk (2004) among other researchers in finance, Afuecheta et al. (2018) focus on the leptokurtic empirical properties of financial data, take a step further in the compound distribution modelling and suggest a Student's t mean-variance distributions family. Moreover, Afuecheta et al. (2018) outline mixtures of the Student's t with exponential, Weibull and gamma distributions, conduct a small applied experiment and provide empirical evidence on the ability of these models to demonstrate better estimation outcomes than some normal mean-variance distribution mixtures. Therefore, this chapter targets to expand upon mixtures family of Afuecheta et al. (2018) and follows an illustrative example of Nadarajah (2012) for the set of normal mean-variance compositions. In particular, this chapter offers six new composition models, where financial returns are assumed to follow the Student's t distribution conditional on the randomized variance/ volatility by the: one parameter half normal, two parameter Fréchet, two parameter Lomax, two parameter Burr III, two parameter inverse gamma and three parameter generalized gamma distributions. For each new compound model following the guidelines of Afuecheta et al. (2018), its Probability Density Function (PDF), Cumulative Distribution Function (CDF), log-likelihood and basic properties/characteristics functions are provided. Moreover, Student's t generalized gamma mixture model provided in this chapter is a special contribution encompassing all three distributions previously supplied in Afuecheta et al. (2018) as its special cases.

Though the main aim of this chapter is portfolio of Student's t based models expansion, some basic applications are also conducted. Applications are performed over the financial returns for stock indices, energy commodities and cryptocurrencies. Estimations are performed using maximum likelihood (MLE) and for the samples considered, comparisons are made using a common set of the log-likelihood based evaluation criteria. It is found that at least one of the proposed models performs better than the GHYP distribution under the selected criteria framework and considered models may be worthwhile investigating in the more data inclusive investigations of the larger scale or studying their properties in more detail. Some Quantile-Quantile (Q-Q) and Probability Integral Transforms - Probability Integral Transforms (P-P) plots are also provided to deepen the log-likelihood based criteria analysis and lay down the directions for future research and investigations.

The chapter is organized as follows: Section 1.2 and corresponding subsections outline the general compound model development as in Afuecheta et al. (2018) as well as provides derivations and formulations for the proposed models, Section 1.3 describes data, conducts some exploratory analysis linked to the proposed models and outlines evaluation criteria, Section 1.4 provides and analyses estimation results as well as supplies estimated parameters for the best performing models and finally, Section 1.5 concludes and summarises the chapter.

1.2 Compound model development

From Nadarajah (2012), consider that $X \sim N(0, \sigma^2)$, where X is the true data generating process (DGP) of a continuous random variable, is normally distributed with zero mean and variance described by some density function $g(\sigma^2)$ always satisfying $\sigma^2 > 0$, then X can be outlined by the below mean-variance density composition

$$f_X(x) = \int_0^\infty \phi(x \mid \sigma^2)\, g(\sigma)\, d\sigma^2, \tag{1.1}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Now, PDF and CDF for the original Student's t distribution, given the degrees of freedom, ν, always satisfying $\nu > 0$ and $x \in \mathbb{R}$ are given by

$$f(x) = K(\nu)\left(1 + \frac{x^2}{\nu}\right)^{-\frac{1+\nu}{2}}$$

and

$$F(x)=\frac{1}{2}+x\Gamma\left(\frac{\nu+1}{2}\right)\frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)},$$

respectively, where

$$\Gamma(z)=\int_0^\infty t^{z-1}\exp(-t)dt$$

denotes the gamma function,

$${}_pF_q\left(a_1,\ldots,a_p;b_1,\ldots,b_q;z\right)=\sum_{j=0}^{\infty}\frac{\left(a_1\right)_j\left(a_2\right)_j\cdots\left(a_p\right)_j}{\left(b_1\right)_j\left(b_2\right)_j\cdots\left(b_q\right)_j}\frac{z^j}{j!}$$

generalized hypergeometric function, $(a)_k=a(a+1)\cdots(a+k-1)$ ascending factorial and $K(\nu)=\Gamma\left((\nu+1)/2\right)/\{\sqrt{\pi\nu}\Gamma\left(\nu/2\right)\}$. Provided that $n<\nu$ all moments of order n exist. In particular, $E(X)=0$ for $\nu>1$, $Var(X)=\nu/(\nu-2)$ for $\nu>2$ and the tail behaviour characterized by $f(x)\simeq cx^{-\nu-1}$, jointly allowing PDF of the Student's t distribution offer fatter/bigger tails than that of the normal distribution as pointed by Blattberg & Gonedes (1974) among others.

Considering the σ^2 conditional PDF of the Student's t of the below form

$$f\left(x\mid\sigma^2\right)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{2\sigma^2}\right)^{-\frac{\nu+1}{2}},$$

Afuecheta et al. (2018) proceed rewriting (1.1) as

$$f_X(x) = \int_0^\infty \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} g\left(\sigma^2\right) d\sigma^2$$

or for convenience by setting $\sigma^2 = \tau$ as

$$\begin{aligned} f_X(x) &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty \frac{1}{\sqrt{\tau}}\left(1+\frac{x^2}{\nu\tau}\right)^{-\frac{\nu+1}{2}} g(\tau)d\tau \\ &= \frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty \tau^{\frac{\nu}{2}}\left(1+\frac{\nu\tau}{x^2}\right)^{-\frac{\nu+1}{2}} g(\tau)d\tau \end{aligned} \tag{1.2}$$

and essentially bringing the tails of the Student's t into equation of the standard compound setting as in Nadarajah (2012). Further, by making use of the below series expansion

$$(1+z)^{-a} = \sum_{k=0}^{\infty}\binom{-a}{k} z^k,$$

(1.2) is then can be rewritten as

$$\begin{aligned} f_X(x) &= \frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\binom{-\frac{\nu+1}{2}}{k}\frac{\nu^k}{x^{2k}}\int_0^\infty \tau^{k+\frac{\nu}{2}} g(\tau)d\tau \\ &= \frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\left(\frac{\nu+1}{2}\right)_k \frac{(-\nu)^k}{k!\,x^{2k}}\int_0^\infty \tau^{k+\frac{\nu}{2}} g(\tau)d\tau, \end{aligned} \tag{1.3}$$

providing a general form of the Student's t based compound PDF. For (1.3), the corresponding general CDF form can be also outlined. From Afuecheta et al. (2018) that is,

$$F_X(x) = \frac{1}{2} + \frac{x\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty \frac{1}{\sqrt{\tau}}\,{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\tau\nu}\right) g(\tau)d\tau$$

or, by utilizing the below series expansion

$$
\begin{aligned}
{}_2F_1\left(a,b;c;z\right) &= \frac{\Gamma(b-a)\Gamma(c)(-z)^{-a}}{\Gamma(b)\Gamma(c-a)}\sum_{k=0}^{\infty}\frac{(a)_k(a-c+1)_k z^{-k}}{k!\,(a-b+1)_k} \\
&\quad +\frac{\Gamma(a-b)\Gamma(c)(-z)^{-b}}{\Gamma(a)\Gamma(c-b)}\sum_{k=0}^{\infty}\frac{(b)_k(b-c+1)_k z^{-k}}{k!\,(b-a+1)_k},
\end{aligned}
$$

rewritten to

$$
F_X(x)=\frac{1}{2}-\frac{\nu^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu+1}{2}\right)}{x^{\nu}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\frac{1+\nu}{2}\right)_k\left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k}\left(\frac{x^2}{\nu}\right)^{-k}\int_0^{\infty}\tau^{k+\frac{\nu}{2}}g(\tau)d\tau. \tag{1.4}
$$

Next, provided that $0<k<\nu$, the general form of the kth moment of X Afuecheta et al. (2018) express with

$$
E\left(X^k\right)=E\left[E\left(X^k\mid\tau\right)\right]=\left[\int_0^{\infty}\tau^{\frac{k}{2}}g(\tau)d\tau\right]\frac{1+(-1)^k}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)\nu^{\frac{k}{2}}. \tag{1.5}
$$

For the general form of the characteristic function of X Afuecheta et al. (2018) obtain

$$
E\left[\exp(\mathrm{i}tX)\right]=E\left\{E\left[\exp(\mathrm{i}tX)\mid\tau\right]\right\}=\left[\int_0^{\infty}K_{\frac{\nu}{2}}\left(\sqrt{\nu\tau}\mid t\mid\right)\tau^{\frac{\nu}{4}}g(\tau)d\tau\right]\frac{\left(\sqrt{\nu}\mid t\mid\right)^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu}{2}\right)},
$$

where $\mathrm{i}=\sqrt{-1}$ and $K_{\nu}(\cdot)$ denotes the modified Bessel function of the third kind defined by

$$
K_n(z)=\frac{x^n\Gamma(1/2)}{2^n\Gamma(n+1/2)}\int_1^{\infty}\exp(-zt)\left(t^2-1\right)^{n-1/2}dt,
$$

and by making use of the below series expansion

$$K_{\nu}(z) = \frac{\pi \csc(\pi\nu)}{2}\left[\sum_{k=0}^{\infty} \frac{z^{2k-\nu}}{2^{2k-\nu} k!\, \Gamma(k-\nu+1)} - \sum_{k=0}^{\infty} \frac{z^{2k+\nu}}{2^{2k+\nu} k!\, \Gamma(k+\nu+1)}\right]$$

further rewrite it to

$$\begin{aligned} E\left[\exp(\mathrm{i}tX)\right] &= \frac{\left(\sqrt{\nu} \mid t \mid\right)^{-\frac{\nu}{4}} \csc\left(\frac{\pi\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\sqrt{\nu} \mid t \mid\right)^{k}}{2^{2k} k!\, \Gamma\left(k - \frac{\nu}{2} + 1\right)} \left[\int_0^{\infty} \tau^{k} g(\tau) d\tau\right] \\ &\quad - \frac{\left(\sqrt{\nu} \mid t \mid\right)^{\frac{3\nu}{4}} \csc\left(\frac{\pi\nu}{2}\right)}{2^{\nu}\Gamma\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\sqrt{\nu} \mid t \mid\right)^{k}}{2^{2k} k!\, \Gamma\left(k + \frac{\nu}{2} + 1\right)} \left[\int_0^{\infty} \tau^{k+\frac{\nu}{2}} g(\tau) d\tau\right] \\ &= \frac{\left(\sqrt{\nu} \mid t \mid\right)^{-\frac{\nu}{4}} \csc\left(\frac{\pi\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\sqrt{\nu} \mid t \mid\right)^{k}}{2^{2k} k!\, \left(1 - \frac{\nu}{2}\right)_k} \left[\int_0^{\infty} \tau^{k} g(\tau) d\tau\right] \\ &\quad - \frac{\left(\sqrt{\nu} \mid t \mid\right)^{\frac{3\nu}{4}} \csc\left(\frac{\pi\nu}{2}\right)}{2^{\nu}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\sqrt{\nu} \mid t \mid\right)^{k}}{2^{2k} k!\, \left(\frac{\nu}{2} + 1\right)_k} \left[\int_0^{\infty} \tau^{k+\frac{\nu}{2}} g(\tau) d\tau\right]. \end{aligned} \tag{1.6}$$

Finally, for functions in (1.3), (1.4), (1.5) and (1.6), choosing appropriate functional form of $g(\cdot)$ closed form expressions for PDF, CDF, the kth moment and characteristic functions of new compound models for financial returns can be obtained. The choice of $g(\cdot)$ is abundant and may not be limited to the parametric forms provided in Afuecheta et al. (2018). Alternatively if $g(\cdot)$ takes form of a degenerate distribution, functions in (1.3), (1.4), (1.5) and (1.6) shall yield Student's t distribution of Gosset (1908) functions as follows from Afuecheta et al. (2018). Further, the half normal, Fréchet, Lomax and Burr distributions, previously listed in Nadarajah (2012) for the normal mean-variance compositions, but not covered in Afuecheta et al. (2018), as well as inverse and generalized gamma functional forms of the mixing distributions are used to expand the overall portfolio of the Student's t composition models for financial returns. Functional forms and some additional descriptions on the $g(\cdot)$

models considered, can be found in Nadarajah (2012) for the half normal, Fréchet, Lomax and Burr, and in Kilber & Kotz (2003) for the inverse and generalized gamma distributions respectively.

1.2.1 One parameter half normal

For one parameter half normal, functional form of g is given by

$$g(\tau) = \frac{2}{\sqrt{2\pi}\theta} \exp\left(-\frac{\tau^2}{2\theta^2}\right),$$

for $\tau > 0$ and $\theta > 0$. The half normal distribution is a form of the common normal distribution with zero mean and scale parameter θ bounded from below at zero. Its applications cut across production processes (e.g. Meeusen & van Den Broeck (1977); Chou & Liu (1998)), life data analysis (e.g. Lawless (2003); Cooray & Ananda (2008)), genetics (e.g. Dobzhansky & Wright (1947)) and other biological science applictions (e.g. Bland &

Altman (1999)). Here, for the one parameter half normal distribution it is obtained

$$
\begin{aligned}
\int_0^\infty \tau^\eta g(\tau)d\tau &= \int_0^\infty \tau^\eta \frac{2}{\sqrt{2\pi}\theta}\exp\left(-\frac{\tau^2}{2\theta^2}\right)d\tau, \\
&= \frac{2}{\sqrt{2\pi}\theta}\int_0^\infty \tau^\eta \exp\left(-\frac{\tau^2}{2\theta^2}\right)d\tau, \\
u &= \frac{\tau^2}{2\theta^2}; \\
\tau^2 &= 2u\theta^2; \\
\tau &= \sqrt{2u\theta^2}; \\
\frac{du}{d\tau} &= \tau/\theta^2; \\
d\tau &= \frac{\theta^2 du}{\tau} = \frac{\theta^2 du}{\sqrt{2u\theta^2}}; \\
\int_0^\infty \tau^\eta g(\tau)d\tau &= \frac{2}{\sqrt{2\pi}\theta}\int_0^\infty (\sqrt{2u\theta^2})^\eta \exp(-u)\frac{\theta^2}{\sqrt{2u\theta^2}}du, \\
&= \frac{2\cdot 2^{\eta/2}\theta^\eta}{\sqrt{2\pi}\theta}\int_0^\infty u^{\eta/2}\exp(-u)\frac{\theta^2}{\sqrt{2u\theta^2}}du, \\
&= \frac{2^{\eta/2+1}\theta^\eta\theta^2\theta^{-1}2^{-1/2}}{\sqrt{2\pi}\theta}\int_0^\infty u^{\eta/2}u^{-1/2}\exp(-u)du, \\
&= \frac{2^{\eta/2+1/2-1/2}\theta^\eta}{\sqrt{\pi}}\int_0^\infty u^{\eta/2-1/2+1-1}\exp(-u)du. \qquad (1.7)
\end{aligned}
$$

Note that (1.7) contains the Euler integral of the second kind, which can be approximated by the gamma function;

$$\Gamma(Z) = \int_0^\infty x^{z-1}\exp(-x)dx,$$

providing that

$$\int_0^\infty u^{(\frac{\eta+1}{2})-1}\exp(-u)du = \Gamma\left(\frac{\eta+1}{2}\right)$$

and yielding

$$\int_0^\infty \tau^\eta g(\tau)d\tau = \frac{[\sqrt{2}\theta]^\eta \Gamma\left(\dfrac{\eta+1}{2}\right)}{\sqrt{\pi}}.$$

Now, from (1.3) $\eta = k + \frac{\nu}{2}$; hence,

$$\begin{aligned}\int_0^\infty \tau^\eta g(\tau)d\tau &= \frac{[\sqrt{2}\theta]^k[\sqrt{2}\theta]^{\frac{\nu}{2}} \Gamma\left(\dfrac{k+\frac{\nu}{2}+1}{2}\right)}{\sqrt{\pi}} \\ &= \frac{[\sqrt{2}\theta]^k[\sqrt{2}\theta]^{\frac{\nu}{2}} \Gamma\left(\dfrac{2k+\nu+2}{4}\right)}{\sqrt{\pi}},\end{aligned}$$

while the closed form expression for Student's t half normal mixture PDF is then given by

$$f_X(x) = \frac{[\sqrt{2}\theta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{x^{\nu+1}\pi\Gamma\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \left(\frac{\nu+1}{2}\right)_k \frac{[-\sqrt{2}\theta\nu]^k \Gamma_k}{k!\, x^{2k}}, \tag{1.8}$$

where $\Gamma_k = \Gamma\left(\dfrac{2k+\nu+2}{4}\right)$. Likewise, from (1.4), (1.5) and (1.6) the closed form expressions for Student's t half normal mixture CDF

$$F_X(x) = \frac{1}{2} - \frac{[\sqrt{2}\theta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\nu x^\nu \pi\Gamma\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\frac{1+\nu}{2}\right)_k \left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k} \left(\frac{x^2}{\nu}\right)^{-k} [\sqrt{2}\theta]^k \Gamma_k,$$

moments

$$E\left(X^k\right) = E\left[E\left(X^k \mid \tau\right)\right] = \Gamma\left(\frac{k+2}{4}\right) \frac{1+(-1)^k}{2\pi\Gamma\left(\frac{\nu}{2}\right)} \Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{\nu-k}{2}\right) [\sqrt{2}\theta\nu]^{\frac{k}{2}},$$

and characteristic function

$$\begin{aligned} E\left[\exp(\mathrm{i}tX)\right] &= \frac{\left(\sqrt{\nu}\mid t\mid\right)^{-\frac{\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\sqrt{2}\theta\sqrt{\nu}\mid t\mid\right)^{k}}{2^{2k}k!\left(1-\frac{\nu}{2}\right)_{k}}\Gamma\left(\frac{k+1}{2}\right) \\ &\quad -\frac{\theta^{\frac{\nu}{2}}\left(\sqrt{\nu}\mid t\mid\right)^{\frac{3\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\sqrt{2\nu}\theta\mid t\mid\right)^{k}\Gamma_{k}}{2^{2k}k!\left(\frac{\nu}{2}+1\right)_{k}}, \end{aligned}$$

are derived. Further, at the estimation stage, all financial returns in the composed data set are location adjusted and therefore all further log-likelihood functions are given in accordance with their forms used in computations. For location variant of (1.8) that is

$$\begin{aligned} \log L\left(\mu,\nu,\theta\right) &= \frac{n\nu}{2}\log[\sqrt{2}\theta\nu]+n\log\Gamma\left(\frac{\nu+1}{2}\right)-n\log\Gamma\left(\frac{\nu}{2}\right)-n\log\pi \\ &+ \sum_{i=1}^{n}\log\left[\sum_{k=0}^{\infty}\binom{-\frac{1}{2}-\frac{\nu}{2}}{k}\frac{\Gamma\left(\frac{2k+\nu+2}{4}\right)[\sqrt{2}\theta\nu]^{k}}{(x_i-\mu)^{2k}}\right] \\ &- (\nu+1)\sum_{i=1}^{n}\log(x_i-\mu). \end{aligned} \tag{1.9}$$

1.2.2 Two parameter Fréchet

For two parameter Fréchet, functional form of g is given by

$$g(\tau)=\frac{\alpha\beta^{\alpha}\exp\left[-\left(\frac{\beta}{\tau}\right)^{\alpha}\right]}{\tau^{\alpha+1}},$$

for $\tau>0$, $\alpha>0$ and $\beta>0$. The latter two parameters can be classified as shape and location parameters of the Fréchet distribution. This distribution is skewed to the right, has a unique mode, and is also known as the inverse Weibull distribution. It is also a special case of the generalized extreme value distribution (GEV), which in turn, is widely used for the "tail risks" characterization in insurance and financial applications (e.g. McNeil et al. (2005)). Here, for the two parameter Fréchet distribution it is obtained

$$\begin{aligned}
\int_0^\infty \tau^\eta g(\tau)d\tau &= \int_0^\infty \frac{\alpha\beta^\alpha\tau^\eta \text{exp}(-[\beta/\tau]^\alpha)}{\tau^{\alpha+1}}d\tau, \\
&= \alpha\beta^\alpha \int_0^\infty \frac{\tau^\eta \text{exp}(-[\beta/\tau]^\alpha)}{\tau^{\alpha+1}}d\tau; \\
u &= (\beta/\tau)^\alpha; \\
u^{1/\alpha} &= \beta/\tau; \\
\tau &= \beta u^{-1/\alpha}; \\
\frac{du}{d\tau} &= \frac{\alpha(\beta/\tau)^\alpha}{\tau} = \alpha\beta^\alpha\tau^{-\alpha-1}; \\
d\tau &= \frac{du}{\alpha\beta^\alpha\tau^{-\alpha-1}}; \\
\int_0^\infty \tau^\eta g(\tau)d\tau &= \alpha\beta^\alpha \int_0^\infty \frac{\beta^\eta(u^{-1/\alpha})^\eta \text{exp}(-u)}{\beta^{\alpha+1}(u^{-1/\alpha})^{\alpha+1}} \frac{\beta^{\alpha+1}(u^{-1/\alpha})^{\alpha+1}}{\alpha\beta^\alpha}du; \\
&= \beta^\eta \int_0^\infty u^{-\eta/\alpha+1-1}\text{exp}(-u)du; \\
&= \beta^\eta\Gamma(1-\eta/\alpha),
\end{aligned}$$

allowing to derive from (1.3), (1.4), (1.5) and (1.6) the closed form expressions for Stundet's t Fréchet mixture PDF

$$f_X(x) = \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \left(\frac{\nu+1}{2}\right)_k \frac{(-\beta\nu)^k\Gamma_k}{k!\, x^{2k}}, \tag{1.10}$$

CDF

$$F_X(x) = \frac{1}{2} - \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\nu x^\nu\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \sum_{k=0}^{\infty} \frac{\left(\frac{1+\nu}{2}\right)_k \left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k} \left(\frac{x^2}{\nu}\right)^{-k} \beta^k\Gamma_k,$$

moments

$$E\left(X^k\right) = E\left[E\left(X^k \mid \tau\right)\right] = \Gamma\left(1-\frac{k}{2\alpha}\right)\frac{1+(-1)^k}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)[\beta\nu]^{\frac{k}{2}},$$

and characteristic function

$$\begin{aligned} E\left[\exp(\mathrm{i}tX)\right] &= \frac{\left(\sqrt{\nu}\mid t\mid\right)^{-\frac{\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^k}{2^{2k}k!\left(1-\frac{\nu}{2}\right)_k}\left[\Gamma\left(1-\frac{k}{\alpha}\right)\right] \\ &\quad -\frac{\beta^{\frac{\nu}{2}}\left(\sqrt{\nu}\mid t\mid\right)^{\frac{3\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{2^{\nu}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^k\Gamma_k}{2^{2k}k!\left(\frac{\nu}{2}+1\right)_k}, \end{aligned}$$

where $\Gamma_k = \Gamma\left(1-\frac{k}{\alpha}-\frac{\nu}{2\alpha}\right)$. For the location adjusted variant of (1.10) the log-likelihood function is then given by

$$\begin{aligned} \log L\left(\mu,\nu,\beta,\alpha\right) &= \frac{n\nu}{2}\log\beta\nu + n\log\Gamma\left(\frac{\nu+1}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) - \frac{n}{2}\log\pi \\ &+ \sum_{i=1}^{n}\log\left[\sum_{k=0}^{\infty}\binom{-\frac{1}{2}-\frac{\nu}{2}}{k}\frac{\Gamma\left(1-\frac{k}{\alpha}-\frac{\nu}{2\alpha}\right)[\beta\nu]^k}{(x_i-\mu)^{2k}}\right] \\ &- (\nu+1)\sum_{i=1}^{n}\log(x_i-\mu). \end{aligned} \tag{1.11}$$

1.2.3 Two parameter Lomax

For two parameters Lomax, functional form of g is given by

$$g(\tau) = \frac{\alpha\beta^\alpha}{(\beta+\tau)^{\alpha+1}},$$

for $\tau > 0$, $\beta > 0$ and $\alpha > 0$. The latter two parameters can also be classified as scale and shape parameters of the Lomax distribution. This PDF has a unique mode at zero and is typically used for characterizing business failures. Lomax distribution is also known as type II Pareto distribution and is a special case of the generalized Pareto distribution (GPD). GPD is commonly utilized in risk management applications and typically serves as an alternative to the GEV distribution (e.g. McNeil et al. (2005)). Overall, Lomax distribution belongs to the Pareto distribution family and is extensively used for lifetime data analysis alongside other applications (e.g. Benckert & Jung (1974); Revankar et al. (1974); Arnold (1983); Hogg & Klugman (1983); Nair & Hitha (1990)). Here, for the two parameter Lomax distribution it is first obtained

$$\begin{aligned}
\int_0^\infty \tau^\eta g(\tau)d\tau &= \int_0^\infty \frac{\alpha\beta^\alpha\tau^\eta}{(\beta+\tau)^{\alpha+1}}d\tau, \\
&= \alpha\beta^\alpha \int_0^\infty \frac{\tau^\eta}{(\beta+\tau)^{\alpha+1}}d\tau, \\
&= \alpha\beta^\alpha \int_0^\infty \frac{\tau^{\eta+1-1}}{[\beta(1+\frac{1}{\beta}\tau)]^{\alpha+1}}d\tau, \\
&= \alpha\beta^{\alpha-\alpha-1} \int_0^\infty \frac{\tau^{\eta+1-1}}{(1+\frac{1}{\beta}\tau)^{\alpha+1}}d\tau. \qquad (1.12)
\end{aligned}$$

Second, from Gradshteyn & Ryzhik (2014) it can be shown that

$$\int_0^\infty \frac{x^{z-1}}{(1+bx)^v} = b^{-z}\mathbf{B}(z, v-z). \qquad (1.13)$$

Therefore, for $z = \eta + 1$, $b = 1/\beta$ and $v = z + 1$ with (1.13) it can be demonstrated that (1.12) takes the following form and further simplified to:

$$\begin{aligned}
\int_0^\infty \tau^\eta g(\tau) d\tau &= \alpha\beta^{\alpha-\alpha-1}[1/\beta]^{-\eta-1}\mathbf{B}(\eta+1, \alpha+1-\eta-1), \\
&= \alpha\beta^{-1}\beta^{\eta+1}\mathbf{B}(\eta+1, \alpha-\eta), \\
&= \alpha\beta^{\eta}\mathbf{B}(\eta+1, \alpha-\eta), \\
&= \alpha\beta^{\eta}\frac{\Gamma(\eta+1)\Gamma(\alpha-\eta)}{\Gamma(\eta+1+\alpha-\eta)}, \\
&= \alpha\beta^{\eta}\frac{\Gamma(\eta+1)\Gamma(\alpha-\eta)}{\Gamma(1+\alpha)}, \\
&= \alpha\beta^{\eta}\frac{\Gamma(\eta+1)\Gamma(\alpha-\eta)}{\Gamma(\alpha)\alpha}, \\
&= \beta^{\eta}\frac{\Gamma(\eta+1)\Gamma(\alpha-\eta)}{\Gamma(\alpha)},
\end{aligned}$$

allowing to derive from (1.3), (1.4), (1.5) and (1.6) the closed form expressions for Student's t Lomax mixture PDF

$$f_X(x) = \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\alpha)x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\left(\frac{\nu+1}{2}\right)_k\frac{(-\beta\nu)^k}{k!\,x^{2k}}\Gamma\Big(\alpha-k-\frac{\nu}{2}\Big)\Gamma\Big(1+k+\frac{\nu}{2}\Big), \quad (1.14)$$

CDF

$$F_X(x) = \frac{1}{2} - \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\alpha)\nu x^{\nu}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\frac{1+\nu}{2}\right)_k\left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k}\left(\frac{x^2}{\nu}\right)^{-k}\beta^k\Gamma\Big(\alpha-k-\frac{\nu}{2}\Big)\Gamma\Big(1+k+\frac{\nu}{2}\Big),$$

moments

$$E\left(X^k\right) = E\left[E\left(X^k \mid \tau\right)\right] = \frac{\Gamma\left(\frac{k}{2}+1\right)\Gamma\left(\alpha-\frac{k}{2}\right)}{\Gamma(\alpha)}\frac{1+(-1)^k}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)[\beta\nu]^{\frac{k}{2}},$$

and characteristic function

$$
\begin{aligned}
E\left[\exp(\mathrm{i}tX)\right] &= \frac{\left(\sqrt{\nu}\mid t\mid\right)^{-\frac{\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\alpha)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^{k}\Gamma(1+k)\Gamma(\alpha-k)}{2^{2k}k!\left(1-\frac{\nu}{2}\right)_{k}} \\
&\quad -\frac{\beta^{\frac{\nu}{2}}\left(\sqrt{\nu}\mid t\mid\right)^{\frac{3\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\alpha)2^{\nu}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^{k}\Gamma\left(\alpha-k-\frac{\nu}{2}\right)\Gamma\left(1+k+\frac{\nu}{2}\right)}{2^{2k}k!\left(\frac{\nu}{2}+1\right)_{k}}.
\end{aligned}
$$

For the location variant of (1.14) the log-likelihood function is then given by

$$
\begin{aligned}
\log L\left(\mu,\nu,\beta,\alpha\right) &= \frac{n\nu}{2}\log\beta\nu+n\log\Gamma\left(\frac{\nu+1}{2}\right)-n\log\Gamma\left(\frac{\nu}{2}\right)-n\log\Gamma(\alpha)-\frac{n}{2}\log\pi \\
&+ \sum_{i=1}^{n}\log\left[\sum_{k=0}^{\infty}\binom{-\frac{1}{2}-\frac{\nu}{2}}{k}\frac{\Gamma\left(\alpha-k-\frac{\nu}{2}\right)\Gamma\left(1+k+\frac{\nu}{2}\right)[\beta\nu]^{k}}{(x_i-\mu)^{2k}}\right] \\
&- (\nu+1)\sum_{i=1}^{n}\log(x_i-\mu).
\end{aligned} \tag{1.15}
$$

1.2.4 Two parameter Burr III distribution

For two parameter Burr, functional form of g is given by

$$
g(\tau)=\frac{c\lambda\tau^{c-1}}{\left(\tau^{c}+1\right)^{\lambda+1}},
$$

for $\tau > 0$, $c > 0$ and $\lambda > 0$. The later two parameters are both commonly referred to as the shape parameters of the Burr III (Burr) distribution. This PDF has a unique mode and moderately skewed to the right. The Burr distribution is often used in reliability analysis, forestry and meteorology among other applications. Here, for the two parameter

Burr distribution it is first obtained

$$\begin{aligned}
\int_0^\infty \tau^\eta g(\tau)d\tau &= \int_0^\infty \tau^\eta \frac{c\lambda\tau^{c-1}}{(\tau^c+1)^{\lambda+1}}d\tau, \\
&= c\lambda\int_0^\infty \frac{\tau^\eta\tau^{c-1}}{(\tau^c+1)^{\lambda+1}}d\tau; \\
u &= \tau^c; \\
u^{1/c} &= \tau; \\
\frac{du}{d\tau} &= c\tau^{c-1}; \\
d\tau &= \frac{du}{c\tau^{c-1}}; \\
\int_0^\infty \tau^\eta g(\tau)d\tau &= c\lambda\int_0^\infty \frac{(u^{1/c})^\eta\tau^{c-1}}{(u+1)^{\lambda+1}c\tau^{c-1}}du, \\
&= \lambda\int_0^\infty \frac{u^{\eta/c+1-1}}{(u+1)^{\lambda+1}}du. \qquad (1.16)
\end{aligned}$$

Now, from Beals & Wong (2010) it can be shown that

$$\mathbf{B}(a,b) = \int_0^\infty u^a \left(\frac{1}{1+u}\right)^{a+b} \frac{du}{u}. \qquad (1.17)$$

With (1.17) it can be demonstrated that (1.16) is decomposed and further simplified to

$$\begin{aligned}
\int_0^\infty \tau^\eta g(\tau)d\tau &= \lambda\mathbf{B}(\eta/c+1, \lambda+1-[\eta/c+1]), \\
&= \lambda\mathbf{B}(\eta/c+1, \lambda-\eta/c), \\
&= \lambda\frac{\Gamma(\eta/c+1)\Gamma(\lambda-\eta/c)}{\Gamma(\eta/c+1+\lambda-\eta/c)}, \\
&= \lambda\frac{\Gamma(\eta/c+1)\Gamma(\lambda-\eta/c)}{\Gamma(\lambda)\lambda}, \\
&= \frac{\Gamma(\eta/c+1)\Gamma(\lambda-\eta/c)}{\Gamma(\lambda)}, \qquad (1.18)
\end{aligned}$$

allowing to derive from (1.3), (1.4), (1.5) and (1.6) the closed form expressions for Student's t Burr mixture PDF

$$\begin{aligned} f_X(x) = & \frac{\nu^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\lambda)x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\left(\frac{\nu+1}{2}\right)_k\frac{(-\nu)^k}{k!\,x^{2k}}\Gamma\left(1+\frac{k}{c}+\frac{\nu}{2c}\right)\times \\ & \times\Gamma\left(\lambda-\frac{k}{c}-\frac{\nu}{2c}\right), \end{aligned} \tag{1.19}$$

CDF

$$F_X(x) = \frac{1}{2} - \frac{\nu^{\frac{\nu}{2}-1}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\lambda)x^{\nu}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\frac{1+\nu}{2}\right)_k\left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k}\left(\frac{x^2}{\nu}\right)^{-k}\Gamma\left(1+\frac{k}{c}+\frac{\nu}{2c}\right)\Gamma\left(\lambda-\frac{k}{c}-\frac{\nu}{2c}\right),$$

moments

$$E\left(X^k\right) = E\left[E\left(X^k \mid \tau\right)\right] = \frac{\Gamma\left(\frac{k}{2c}+1\right)\Gamma\left(\lambda-\frac{k}{2c}\right)}{\Gamma(\lambda)}\frac{1+(-1)^k}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)\nu^{\frac{k}{2}}$$

and characteristic function

$$\begin{aligned} E\left[\exp(\mathrm{i}tX)\right] = & \frac{\left(\sqrt{\nu}\mid t\mid\right)^{-\frac{\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\lambda)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\sqrt{\nu}\mid t\mid\right)^k}{2^{2k}k!\left(1-\frac{\nu}{2}\right)_k}\Gamma\left(\frac{k}{c}+1\right)\Gamma\left(\lambda-\frac{k}{c}\right) \\ & -\frac{\left(\sqrt{\nu}\mid t\mid\right)^{\frac{3\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\lambda)2^{\nu}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\sqrt{\nu}\mid t\mid\right)^k\Gamma_k}{2^{2k}k!\left(\frac{\nu}{2}+1\right)_k}, \end{aligned}$$

where $\Gamma_k = \Gamma\left(1+\frac{k}{c}+\frac{\nu}{2c}\right)\Gamma\left(\lambda-\frac{k}{c}-\frac{\nu}{2c}\right)$. For the location adjusted variant of (1.19) the log-likelihood function is then given by

$$\begin{aligned}\log L\left(\mu,\nu,c,\lambda\right) &= \frac{n\nu}{2}\log\nu + n\log\Gamma\left(\frac{\nu+1}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) - n\log\Gamma(\lambda) - \frac{n}{2}\log\pi \\ &+ \sum_{i=1}^{n}\log\left[\sum_{k=0}^{\infty}\binom{-\frac{1}{2}-\frac{\nu}{2}}{k}\frac{\Gamma\left(1+\frac{k}{c}+\frac{\nu}{2c}\right)\Gamma\left(\lambda-\frac{k}{c}-\frac{\nu}{2c}\right)\nu^k}{(x_i-\mu)^{2k}}\right] \\ &- (\nu+1)\sum_{i=1}^{n}\log(x_i-\mu). \end{aligned} \tag{1.20}$$

1.2.5 Two parameter inverse gamma

For two parameter inverse gamma, functional form of g is given by

$$g(\tau) = \frac{\beta^{\alpha}\tau^{-\alpha-1}\exp\left(-\frac{\beta}{\tau}\right)}{\Gamma(\alpha)},$$

for $\tau > 0$, $\alpha > 0$ and $\beta > 0$. The latter two parameters are commonly known as shape and scale parameters of the inverse gamma distribution. This PDF has a unique mode defined by $\tau = \beta/(\alpha+1)$ and is moderately skewed to the right. It is used in modelling a wide range of physical phenomena in climatology, survival analysis as well as option pricing in finance (e.g. Bouchaud & Potters (2003)). Here, for the two parameter inverse gamma distribution it is obtained

$$
\begin{aligned}
\int_0^\infty \tau^\eta g(\tau)d\tau &= \int_0^\infty \frac{\beta^\alpha \tau^\eta \tau^{-\alpha-1}\exp(-\beta/\tau)}{\Gamma(\alpha)}d\tau,\\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \tau^\eta \tau^{-\alpha-1}\exp(-\beta/\tau)d\tau,\\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \tau^{\eta-\alpha-1}\exp(-\beta/\tau)d\tau;\\
S &= \beta/\tau;\\
\tau &= \beta/S;\\
d\tau/dS &= -\beta S^{-2};\\
d\tau &= -\beta S^{-2}dS;\\
\int_0^\infty \tau^\eta g(\tau)d\tau &= \frac{\beta^\alpha}{\Gamma(\alpha)}(-)\int_\infty^0 (\beta/S)^{\eta-\alpha-1}\exp(-S)(-\beta S^{-2})dS,\\
&= \frac{\beta^{\alpha+\eta-\alpha-1+1}}{\Gamma(\alpha)}\int_0^\infty S^{-\eta+\alpha+1}S^{-2}\exp(-S)dS,\\
&= \frac{\beta^\eta}{\Gamma(\alpha)}\int_0^\infty S^{(\alpha-\eta)-1}\exp(-S)dS,\\
&= \frac{\beta^\eta}{\Gamma(\alpha)}\Gamma\left(\alpha-\eta\right),
\end{aligned}
$$

allowing to derive from (1.3), (1.4), (1.5) and (1.6) the closed form expressions for Student's t inverse gamma mixture PDF

$$
f_X(x) = \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\alpha\right)x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\left(\frac{\nu+1}{2}\right)_k\frac{(-\beta\nu)^k}{k!\,x^{2k}}\Gamma\left(\alpha-k-\frac{\nu}{2}\right), \tag{1.21}
$$

CDF

$$
F_X(x) = \frac{1}{2} - \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\alpha)\nu x^\nu\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\frac{1+\nu}{2}\right)_k\left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k}\left(\frac{x^2}{\nu}\right)^{-k}\beta^k\Gamma\left(\alpha-k-\frac{\nu}{2}\right),
$$

moments

$$E\left(X^k\right)=E\left[E\left(X^k \mid \tau\right)\right]=\frac{\Gamma\left(\alpha-\frac{k}{2}\right)}{\Gamma(\alpha)}\frac{1+(-1)^k}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)[\beta\nu]^{\frac{k}{2}},$$

and characteristic function

$$\begin{aligned} E\left[\exp(\mathrm{i}tX)\right] &= \frac{\left(\sqrt{\nu}\mid t\mid\right)^{-\frac{\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\alpha)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^k\Gamma(\alpha-k)}{2^{2k}k!\left(1-\frac{\nu}{2}\right)_k} \\ &\quad -\frac{\beta^{\frac{\nu}{2}}\left(\sqrt{\nu}\mid t\mid\right)^{\frac{3\nu}{4}}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\alpha)2^{\nu}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^k}{2^{2k}k!\left(\frac{\nu}{2}+1\right)_k}\Gamma\Big(\alpha-k-\frac{\nu}{2}\Big). \end{aligned}$$

For the location adjusted variant of (1.21) the log-likelihood function is then given by

$$\begin{aligned} \log L\left(\mu,\nu,\beta,\alpha\right) &= \frac{n\nu}{2}\log\beta\nu+n\log\Gamma\left(\frac{\nu+1}{2}\right)-n\log\Gamma\left(\frac{\nu}{2}\right)-n\log\Gamma(\alpha)-\frac{n}{2}\log\pi \\ &+ \sum_{i=1}^{n}\log\left[\sum_{k=0}^{\infty}\binom{-\frac{1}{2}-\frac{\nu}{2}}{k}\frac{[\beta\nu]^k\Gamma\left(\alpha-k-\frac{\nu}{2}\right)}{(x_i-\mu)^{2k}}\right] \\ &- (\nu+1)\sum_{i=1}^{n}\log(x_i-\mu). \end{aligned} \tag{1.22}$$

1.2.6 Three parameter generalized gamma

For three parameter generalized gamma, functional form of g is given by

$$g(\tau)=\frac{\lambda\tau^{\alpha\lambda-1}\exp\left[-\left(\frac{\tau}{\beta}\right)^{\lambda}\right]}{\Gamma(\alpha)\beta^{\alpha\lambda}},$$

for $\tau>0$, $\beta>0$, $\lambda>0$ and $\alpha>0$. Its scale, first and second shape parameters are described by β, λ and α respectively. This PDF has a unique mode and is skewed to the right. The generalized gamma distribution has extensive applications in hydrology, biology, economics and finance (e.g. Madan & Seneta (1990); Tjetjep & Seneta (2006)). It encompasses a number of other distributions often used in survival analysis. For instance, if $\lambda = \alpha = 1$ then the generalized gamma distribution takes the form of the exponential distribution, when $\lambda = 1$ the generalized gamma distribution is identical to the gamma distribution, and if $\alpha = 1$ the generalized gamma is identical to the Weilbull distribution. Here, for the three parameter generalized gamma distribution it is obtained

$$
\begin{aligned}
\int_0^\infty \tau^\eta g(\tau) d\tau &= \int_0^\infty \tau^\eta \frac{\lambda \tau^{\alpha\lambda-1} \exp\left[-(\tau/\beta)^\lambda\right]}{\Gamma(\alpha)\beta^{\alpha\lambda}} d\tau, \\
&= \frac{\lambda}{\Gamma(\alpha)\beta^{\alpha\lambda}} \int_0^\infty \tau^\eta \tau^{\alpha\lambda-1} \exp\left[-(\tau/\beta)^\lambda\right] d\tau, \\
u &= (\tau/\beta)^\lambda; \\
u^{1/\lambda} &= \tau/\beta; \\
\tau &= u^{1/\lambda}\beta; \\
\frac{du}{d\tau} &= \frac{\lambda(\tau/\beta)^\lambda}{\tau}; \\
d\tau &= \frac{\tau du}{\lambda(\tau/\beta)^\lambda}; \\
\int_0^\infty \tau^\eta g(\tau) d\tau &= \frac{\lambda}{\Gamma(\alpha)\beta^{\alpha\lambda}} \int_0^\infty \beta^\eta u^{\eta/\lambda} \beta^{\alpha\lambda-1} \left(u^{1/\lambda}\right)^{\alpha\lambda-1} \exp(-u) \frac{\beta u^{1/\lambda}}{\lambda\left(\frac{\beta u^{1/\lambda}}{\beta}\right)^\lambda} du, \\
&= \frac{\beta^\eta \beta \beta^{\alpha\lambda-1}}{\Gamma(\alpha)\beta^{\alpha\lambda}} \int_0^\infty u^{\eta/\lambda} u^\alpha u^{-1/\lambda} \exp(-u) \frac{u^{1/\lambda}}{u} du, \\
&= \frac{\beta^\eta}{\Gamma(\alpha)} \int_0^\infty u^{\eta/\lambda} u^\alpha u^{-1} \exp(-u) du, \\
&= \frac{\beta^\eta}{\Gamma(\alpha)} \int_0^\infty u^{\eta/\lambda+\alpha-1} \exp(-u) du, \\
&= \frac{\beta^\eta \Gamma\left(\alpha + \frac{\eta}{\lambda}\right)}{\Gamma(\alpha)}.
\end{aligned}
$$

allowing to derive from (1.3), (1.4), (1.5) and (1.6) the closed form expressions for Student's t generalized gamma mixture PDF

$$f_X(x) = \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\alpha)x^{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\left(\frac{\nu+1}{2}\right)_k\frac{(-\beta\nu)^k}{k!\,x^{2k}}\Gamma\left(\alpha+\frac{k}{\lambda}+\frac{\nu}{2\lambda}\right) \tag{1.23}$$

CDF

$$F_X(x) = \frac{1}{2} - \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\alpha)\nu x^{\nu}\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\frac{1+\nu}{2}\right)_k\left(\frac{\nu}{2}\right)_k}{k!\left(1+\frac{\nu}{2}\right)_k}\left(\frac{x^2}{\nu}\right)^{-k}\beta^k\Gamma\left(\alpha+\frac{k}{\lambda}+\frac{\nu}{2\lambda}\right)$$

moments

$$E\left(X^k\right) = E\left[E\left(X^k \mid \tau\right)\right] = \frac{\Gamma\left(\frac{k}{2\lambda}+\alpha\right)}{\Gamma(\alpha)}\frac{1+(-1)^k}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right)[\beta\nu]^{\frac{k}{2}}$$

characteristic function

$$\begin{aligned} E\left[\exp(\mathrm{i}tX)\right] &= \frac{\left(\sqrt{\nu}\mid t\mid\right)^{-\frac{\nu}{4}}\beta^{-\alpha\lambda}\csc\left(\frac{\pi\nu}{2}\right)}{\Gamma(\alpha)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^k}{2^{2k}k!\left(1-\frac{\nu}{2}\right)_k}\Gamma\left(\alpha+\frac{k}{\lambda}\right) \\ &\quad -\frac{\left(\sqrt{\nu}\mid t\mid\right)^{\frac{3\nu}{4}}\beta^{\frac{\nu}{2}}\csc\left(\frac{\pi\nu}{2}\right)}{2^{\nu}\Gamma(\alpha)\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(1+\frac{\nu}{2}\right)}\sum_{k=0}^{\infty}\frac{\left(\beta\sqrt{\nu}\mid t\mid\right)^k}{2^{2k}k!\left(\frac{\nu}{2}+1\right)_k}\Gamma\left(\alpha+\frac{k}{\lambda}+\frac{\nu}{2\lambda}\right). \end{aligned}$$

For the location adjusted variant of (1.23) the log-likelihood function is then given by

$$\begin{aligned} \log L\left(\mu,\nu,\beta,\alpha,\lambda\right) &= \frac{n\nu}{2}\log\beta\nu + n\log\Gamma\left(\frac{\nu+1}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) - n\log\Gamma(\alpha) - \frac{n}{2}\log\pi \\ &+ \sum_{i=1}^{n}\log\left[\sum_{k=0}^{\infty}\binom{-\frac{1}{2}-\frac{\nu}{2}}{k}\frac{[\beta\nu]^k\Gamma\left(\alpha+\frac{k}{\lambda}+\frac{\nu}{2\lambda}\right)}{(x_i-\mu)^{2k}}\right] \\ &- (\nu+1)\sum_{i=1}^{n}\log(x_i-\mu). \end{aligned} \tag{1.24}$$

1.3 Data

To investigate empirical performance and validity of the suggested models a data set consisting of the six financial series has been constructed. Composed data set includes the daily log-returns for two financial stock indices, two fuel commodities prices and two crypto-currencies exchange rates. Stock indices are Standard & Poor's 500 (S&P500) and Dow Jones Industrial Average (DJI) for the period starting on the 28th of April 2003 and ending on the 15th of June 2018 as provided by Bloomberg. Fuel commodities are spot prices for the Los Angeles Ultra-Low-Sulfur Diesel (Diesel) and Mont Belvieu, Texas Propane (Propane) in USDs per gallon for the period beginning on the 2nd of January 1997 and ending on the 15th of June 2018 as provided by the United States Energy Information Administration. Cryptocurrencies are Bitcoin (BTC) for the period starting on the 18th of July 2010 and ending on the 16th of June 2018 and Litecoin (LTC) for the period beginning on the 24th of October 2013 and ending on the 16th of June 2018. Both cryptocurrencies are denominated in USD with their sample sizes representing their entire life cycle on the moment of the data download from Kraken cryptocurrencies exchange database. Since this chapter is partially motivated by the heavy tail potential of the parent distribution in the compound models considered, alongside some very common stock indices (S&P500 and DJI) the target was to include some financial series with notable tail (excess kurtosis) characteristics (e.g. Propane and LTC). In general, Diesel and Propane series are not uncommon in the energy relevant investigations as in Laporta et al. (2018), while BTC and LTC are usually present in the cryptocurrencies context as in Chan et al. (2017) among others. Therefore, this forms a small and representative data set consisting of stock indices, energy commodities and

Statistics	S&P500	DJI	Diesel	Propane	BTC	LTC
n	3811	3811	5388	5388	2890	1696
Minimum	-0.09469	-0.08201	-0.22716	-0.49913	-0.49152	-0.93452
Q1	-0.03903	-0.00388	-0.01149	-0.01025	-0.01273	-0.02003
Median	0.00069	0.00053	0.00000	0.00000	0.00141	0.00000
Mean	0.00029	0.00028	0.00019	0.00009	0.00386	0.00204
Q3	0.00526	0.00507	0.01139	0.01121	0.02276	0.01810
Maximum	0.10957	0.10508	0.26826	0.19979	0.42457	0.89035
Skewness	-0.37707	-0.15289	0.10768	-2.02435	-0.34656	0.63046
Kurtosis	15.2352	14.59654	13.8692	45.13415	14.70389	36.08584
SD	0.01146	0.01064	0.02336	0.02572	0.05840	0.07981
Variance	0.00013	0.00011	0.00054	0.00066	0.00341	0.00637
Range	0.20426	0.18708	0.49542	0.69892	0.91611	1.82488
IQR	0.00916	0.00895	0.02288	0.02146	0.03550	0.03814

Table 1.1: Summary descriptive statistics for the daily log-returns of S&P500, DJI, Diesel, Propane, BTC and LTC.

cryptocurrencies for empirical investigations.

For the above discussed financial series, log-returns have been computed with

$$R_{i,t} = \log\left(\frac{P_{i,t}}{P_{i,t-1}}\right),$$

where $R_{i,t}$ is the return on the index/commodity i for the period t, $P_{i,t}$ is the closing rate/price of the index/commodity at the end of period t and $P_{i,t-1}$ is the price of the index/commodity at the end of the period $t-1$. Obtained log-returns and their histogram density evaluations are illustrated in Figure 1.1. Some of their characteristics are described in Table 1.1 including: minimum, first quartile (Q1), median, mean, third quartile (Q3), maximum, skewness, kurtosis, standard deviation (SD), variance, range and inter quartile range (IQR).

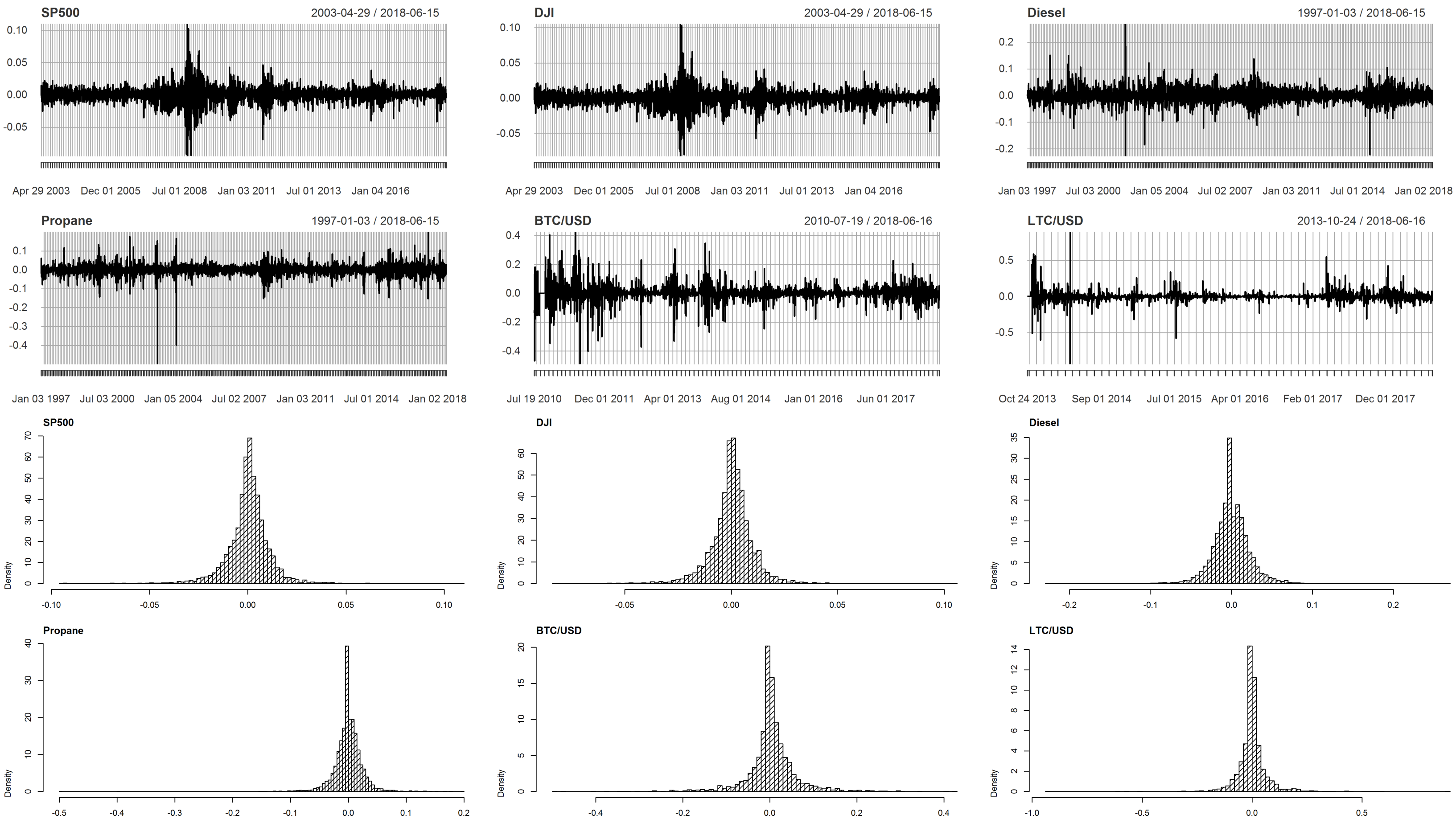

Figure 1.1: Time series plots for the daily log-returns of S&P500, DJI, Diesel, Propane, BTC and LTC with their histogram based density evaluations.

From Table 1.1 the highest range is obsevred for the cryptocurrencies returns, then for commodities and is the smallest for stock indices. From Figure 1.1 sharp market corrections for Diesel and Propane in the early 2000s can be articulated by the changes in fundamentals of hydrocarbons, while lack of the clearly defined fundamentals best explains the highest range for the cryptocurrencies. Stock indices highest turbulence is arguably attributed to the events associated with the financial crisis and their, relatively other returns in Table 1.1, low range may be explained by the composite nature of the indices themselves. The highest kurtosis values are depicted by the Propane and LTC series. Kurtosis values for S&P500, DJI, Deisel and BTC are notably excess normal and have similar value to each other. Diesel and LTC are the only two positively skewed series. Overall, inspecting histograms in Figure 1.1 it may be noted that participating returns are mostly positioned around zero and are arguably symmetric (apart from Propane returns). Next, from the time series plots of returns it can be observed that periods of high turbulence are typically followed by the periods of low turbulence and vice versa. This can be often expected for financial returns and is commonly known as volatility clustering (e.g. Mandelbrot (1963); Pagan (1996); Cont (2001)). Therefore, for volatility, as measured by the standard deviation for non-overlapping window of the 50 observations for log-returns, Figure 1.2 is produced. It may be noted that all histograms of the standard deviation in Figure 1.2 are unimodal and exhibit gradual decay over the horizontal axis, suggesting overall appropriateness of the choice for the functional forms of $g(\cdot)$ in Section 1.2. Moreover, using parameters obtained by MLE, the best fitting $g(\cdot)$ forms from the pool described in Section 1.2 are illustrated alongside every histogram of the computed standard deviations. For simplicity, the best performing $g(\cdot)$ model is chosen on the basis of the lowest value for the negative log-likelihood function for each volatility

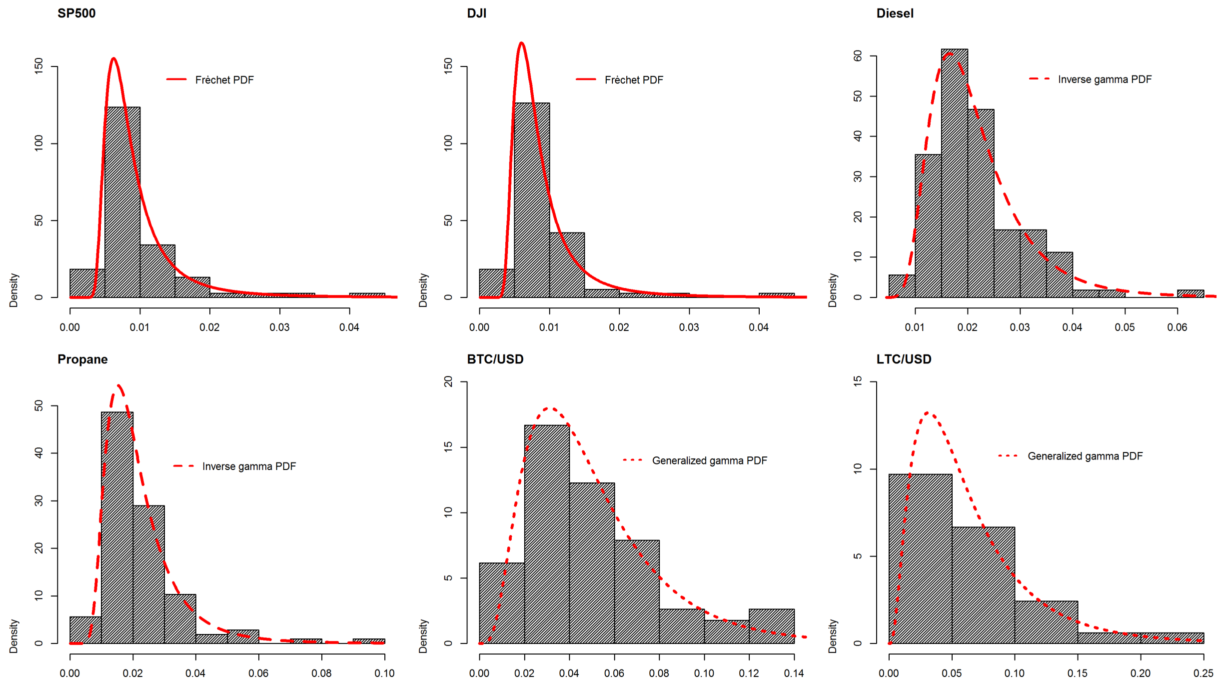

Figure 1.2: Histogram densities of the non-overlapping window standard deviations for the specified daily log-returns.

measure. Each fitted PDF seems capturing volatility histograms well, though it is interesting highlighting that volatility for stock indices is the best described by the Fréchet PDF, for fuel commodities by the inverse gamma PDF and for cryptocurrencies by the generalized gamma PDF, overall pointing out that in the general financial setting "one model fits all" may not always be an optimal strategy.

Finally, after conducting some preliminary data analysis and roughly rationalizing new models for financial returns provision, for compound distributions in Section 1.2, empirical performance evaluations and estimations are performed using unknown parameters obtained with MLE. That formally is: for observations $x_1, x_2, \cdots, x_n$ from DGP X, optimal parameters of the estimated models are the values maximizing the likelihood

$$L(\Theta) = \prod_{i=1}^{n} f(x_i; \Theta),$$

or by exploiting the logarithmic properties to simplify estimations, maximizing the log-likelihood

$$\log L(\Theta) = \sum_{i=1}^{n} \log(x_i; \Theta),$$

where $\Theta = (\theta_1 \cdots \theta_k)'$ is the parameter vector that specifies the distribution density of X, $f(\cdot)$. Consequently, the optimal estimates for Θ are $\widehat{\Theta} = \left(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k\right)'$. Since considered distributions are not nested, discrimination among them is performed using the following standard criteria:

- the Akaike information criterion (AIC) due to Akaike (1974), defined by

$$\text{AIC} = 2k - 2\log L(\widehat{\Theta});$$

- the Bayesian information criterion (BIC) due to Schwarz et al. (1978), defined by

$$\text{BIC} = k \log n - 2\log L(\widehat{\Theta});$$

- the corrected Akaike information criterion (AICc) due to Hurvich & Tsai (1989), defined by

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1};$$

- the Hannan-Quinn criterion (HQC) due to Hannan & Quinn (1979), defined by

$$\mathrm{HQC} = -2\log L(\widehat{\Theta}) + 2k\log\log n;$$

- the consistent Akaike information criterion (CAIC) due to Bozdogan (1987), defined by

$$\mathrm{CAIC} = -2\log L(\widehat{\Theta}) + k\left(\log n + 1\right).$$

More extensive discussions on these commonly used criteria are provided by Burnham & Anderson (2004) and Fang (2011). Roughly, for all participating models given particular data sample, the smallest computed criteria values outline the best performing model. Models in Section 1.2 compete against GHYP distribution as provided by the ghyp package of Luethi & Breymann (2016). This specification matches GHYP model outlined in McNeil et al. (2005). Function values and criteria for GHYP distribution are computed for the parameters obtained with MLE as provided by the fit.ghypuv command of the ghyp package.

1.4 Some estimation results and discussion

The log-likelihood and values for the five models selection criteria (AIC, BIC, AICc, HQC and CAIC) for all participating distributions are provided in Table 1.2. From Table 1.2, for stock indices returns, it can be observed that the main competitors are GHYP and generalized gamma mixture distributions, though performance of the inverse gamma may

Returns	Mixing distribution	$-\log L(\cdot)$	AIC	BIC	CAIC	AICc	HQC
S&P500	Half normal	-12241.35	-24476.70	-24457.96	-24454.96	-24476.69	-24470.04
	Frechet	-12326.20	-24644.40	-24619.41	-24615.41	-24644.39	-24635.52
	Lomax	-12123.29	-24238.58	-24213.60	-24209.60	-24238.57	-24229.70
	Burr	-11662.53	-23317.06	-23292.08	-23288.08	-23317.05	-23308.19
	Inverse gamma	-12331.33	-24654.66	-24629.67	-24625.67	-24654.65	-24645.78
	Gen. gamma	-12336.21	-24662.42	-24631.20	-24626.20	-24662.41	-24651.33
	GHYP	-12334.85	-24659.70	-24628.47	-24623.47	-24659.68	-24648.60
DJI	Half normal	-12440.16	-24874.32	-24855.59	-24852.59	-24874.32	-24867.67
	Frechet	-12526.14	-25044.27	-25019.29	-25015.29	-25044.26	-25035.40
	Lomax	-12276.93	-24545.86	-24520.88	-24516.88	-24545.85	-24536.98
	Burr	-12179.49	-24350.98	-24326.00	-24322.00	-24350.97	-24342.11
	Inverse gamma	-12530.78	-25053.56	-25028.58	-25024.58	-25053.55	-25044.68
	Gen. gamma	-12536.95	-25063.89	-25032.67	-25027.67	-25063.88	-25052.80
	GHYP	-12533.64	-25057.28	-25026.05	-25021.05	-25057.26	-25046.18
Diesel	Half normal	-14163.66	-28321.33	-28301.55	-28298.55	-28321.32	-28314.42
	Frechet	-13147.06	-26286.11	-26259.74	-26255.74	-26286.10	-26276.90
	Lomax	-13769.83	-27531.65	-27505.29	-27501.29	-27531.65	-27522.45
	Burr	-9656.51	-19305.02	-19278.65	-19274.65	-19305.01	-19295.81
	Inverse gamma	-13149.67	-26291.34	-26264.97	-26260.97	-26291.33	-26282.14
	Gen. gamma	-18401.46	-36792.91	-36759.95	-36754.95	-36792.90	-36781.40
	GHYP	-13145.98	-26281.96	-26249.00	-26244.00	-26281.95	-26270.45
Propane	Half normal	-14343.46	-28680.91	-28661.14	-28658.14	-28680.91	-28674.01
	Frechet	-13172.12	-26336.23	-26309.87	-26305.87	-26336.23	-26327.03
	Lomax	-14071.09	-28134.18	-28107.81	-28103.81	-28134.17	-28124.97
	Burr	-18170.40	-36332.80	-36306.43	-36302.43	-36332.79	-36323.59
	Inverse gamma	-13180.95	-26353.91	-26327.54	-26323.54	-26353.90	-26344.70
	Gen. gamma	-17550.68	-35091.36	-35058.40	-35053.40	-35091.35	-35079.85
	GHYP	-15338.40	-30666.80	-30633.84	-30628.84	-30666.79	-30655.29
BTC	Half normal	-5358.07	-10710.14	-10692.23	-10689.23	-10710.13	-10703.69
	Frechet	-5037.48	-10066.96	-10043.09	-10039.09	-10066.95	-10058.36
	Lomax	-5326.56	-10645.12	-10621.24	-10617.24	-10645.11	-10636.51
	Burr	-5341.80	-10675.61	-10651.73	-10647.73	-10675.60	-10667.00
	Inverse gamma	-5049.81	-10091.62	-10067.74	-10063.74	-10091.61	-10083.02
	Gen. gamma	-6328.61	-12647.22	-12617.37	-12612.37	-12647.20	-12636.46
	GHYP	-5246.81	-10483.62	-10453.77	-10448.77	-10483.60	-10472.86
LTC	Half normal	-2945.52	-5885.05	-5868.74	-5865.74	-5885.03	-5879.01
	Frechet	-2707.95	-5407.89	-5386.15	-5382.15	-5407.87	-5399.84
	Lomax	-2938.51	-5869.02	-5847.27	-5843.27	-5868.99	-5860.97
	Burr	-2916.45	-5824.90	-5803.16	-5799.16	-5824.88	-5816.85
	Inverse gamma	-2714.11	-5420.23	-5398.49	-5394.49	-5420.21	-5412.18
	Gen. gamma	-4047.36	-8084.72	-8057.54	-8052.54	-8084.68	-8074.65
	GHYP	-2722.79	-5435.58	-5408.40	-5403.40	-5435.54	-5425.51

Table 1.2: Log-likelihood values and the AIC, BIC, CAIC, AICc and HQC criteria for the six proposed models and GHYP distribution.

be also worth highlighting. In the current setting, generalized gamma mixture has five parameters, the same as the GHYP distribution, and likely benefits from the flexibility offered by its extra parameter to beat the inverse gamma composition and yield the best performing model for both stock indices among proposed compositions in Section 1.2. Moving over to fuel commodities in Table 1.2 an interesting trend may be observed in the best performing models. For diesel: Fréchet, inverse gamma, Lomax, half normal and generalized gamma mixtures outperform GHYP as per all evaluation criteria considered. It is also noteworthy highlighting that the half normal, three parameter mixture, is the "first runner up" behind the best performing five parameter generalized gamma compound distribution. For propane, Burr distribution takes the lead leaving both the generalized gamma and GHYP behind. For both cryptocurrencies the best performing model is generalized gamma with Lomax, Burr and half normal for BTC and with Burr, Lomax and half normal for LTC performing better than the GHYP distribution. Again, it is wroth highlighting remarkable performance of the half normal compound model given its low number of parameters in the setting. Therefore, it can be observed that portfolio of the proposed distribution models may be viable for different application subjects, since apart from the stock indices, fuels and cryptocurrencies log-returns applications provide good perspectives for the most of the Student's t parametrizations here, when compared to the GHYP distribution. Finally, for the best fitting models from Table 1.2 estimated parameters are provided in Table 1.3. Q-Q and P-P plots for these parameters are provided in Figures 1.3 and 1.4 respectively.

Analysing Q-Q plots in Figure 1.3, quantiles fit provided by the best performing models is overall satisfactory. Both estimated tail quantiles for diesel, propane and BTC lie on the vector of the sample quantiles suggesting accurate quantile approximations. LTC quantile

Returns	Estimated parameters				
S&P500	$\hat{\mu} = 0.00072$	$\hat{\nu} = 3.75036$	$\hat{\lambda} = 1.68896$	$\hat{\alpha} = 1.69071$	$\hat{\beta} = 0.00564$
DJI	$\hat{\mu} = 0.00056$	$\hat{\nu} = 3.58858$	$\hat{\lambda} = 6.52243$	$\hat{\alpha} = 0.31171$	$\hat{\beta} = 0.01012$
Diesel	$\hat{\mu} = -0.171 \cdot 10^7$	$\hat{\nu} = 11.26778$	$\hat{\lambda} = 1.41116$	$\hat{\alpha} = 0.70864$	$\hat{\beta} = 0.03025$
Propane	$\hat{\mu} = -0.319 \cdot 10^7$	$\hat{\nu} = 99.9901$	$\hat{\lambda} = 49.04078$	$\hat{c} = 1.00003$	-
BTC	$\hat{\mu} = -0.499 \cdot 10^8$	$\hat{\nu} = 10.11002$	$\hat{\lambda} = 0.89034$	$\hat{\alpha} = 1.12321$	$\hat{\beta} = 0.03133$
LTC	$\hat{\mu} = -0.109 \cdot 10^7$	$\hat{\nu} = 10.55378$	$\hat{\lambda} = 1.07568$	$\hat{\alpha} = 0.92964$	$\hat{\beta} = 0.04607$

Table 1.3: Estimated parameters for the best fitting distribution models as per results in Table 1.2.

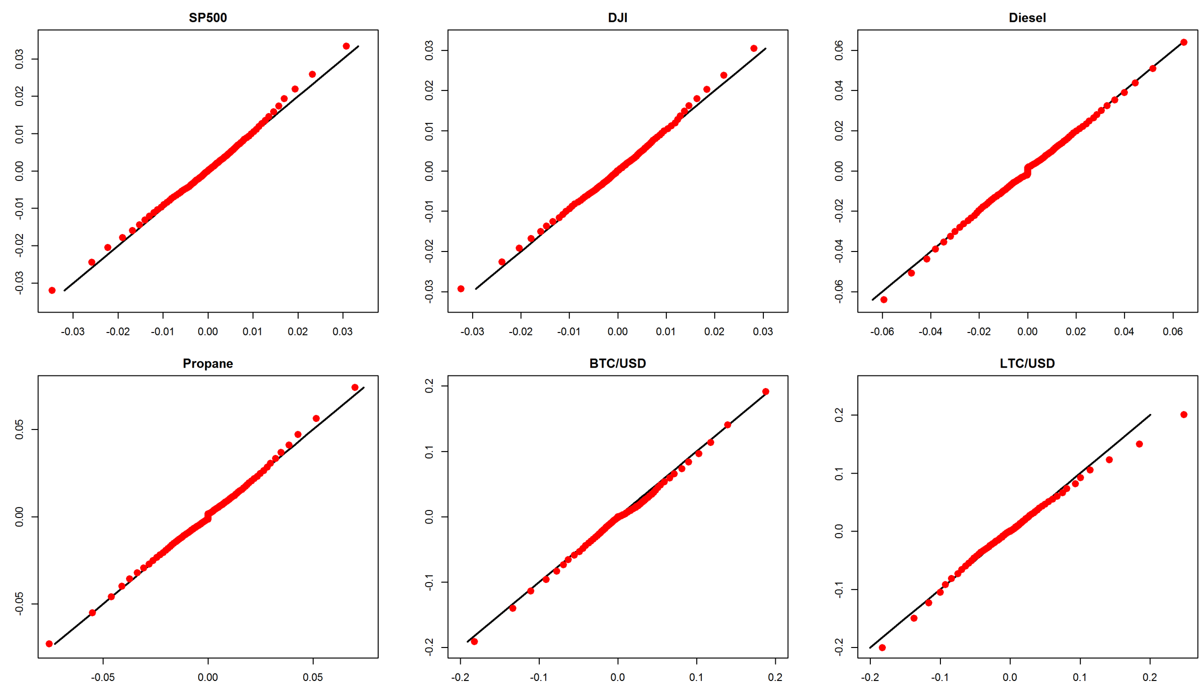

Figure 1.3: Q-Q plots for the best fitting distribution models.

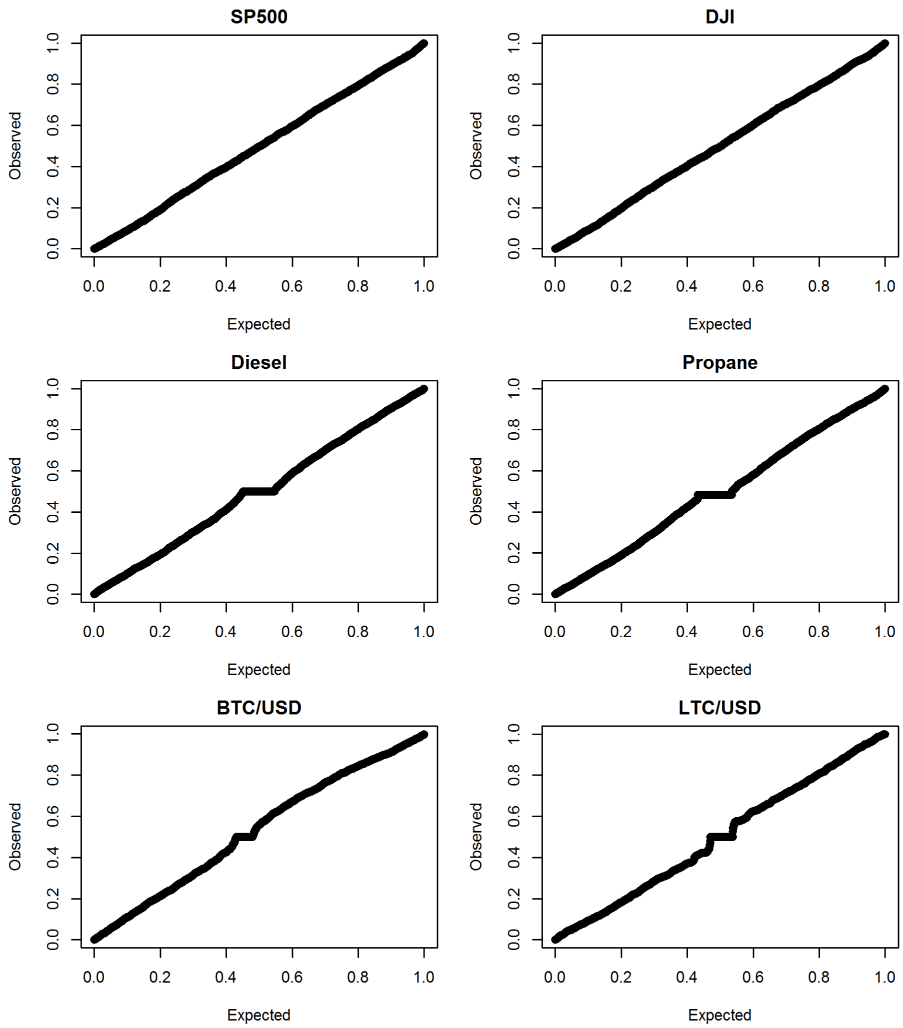

Figure 1.4: P-P plots of the best fitting distribution models.

estimates for the lower tail are overall close to matching the vector of sample quantiles, while for the upper tail quantile estimates are rather "conservative" and deviate from the vector of observed sample quantiles. S&P500 and DJI quantile outlook is very similar to each other and is arguably adequate with small deviations from sample quantiles observed in the tails. However, there some "curvature" may be noted in the centres of the Q-Q plots for diesel, propane, BTC and LTC. This may be an indication that centres of the distributions for these returns are not captured adequately. Indeed, Figure 1.4 confirms this. For all four, it is clear that estimates of the centres are far from satisfactory. Returning to the histogram density evaluations in Figure 1.1 it may be spoted that all four series are arguably more dense towards the centre than nonparametric evaluations for S&P500 and DJI. Given that the P-P evaluations for S&P500 and DJI are notably uniform for both tails and body domains it may be rationalized that there is not enough flexibility in the centres of the compound models offered to the series of diesel, propane, BTC and LTC. To overcome this, it may be worthwhile introducing functionality of asymmetry in the parent distribution or considering any other functional forms allowing more contrast in the characteristics of the tails and body domains for the returns under investigation. It also may be worth introducing time-variation in the parameters of the proposed distributions using appropriate autoregression dynamics. It is interesting highlighting that, though the GHYP distribution allows for asymmetries and has several distributions as its special cases (e.g. as shown by Aas & Haff (2006)), it provides inferior results for these four returns in Table 1.2. Therefore, it is likely that optimal P-P projections for both body and tail domains may be achieved by the certain nonparametric (e.g. as in Sain & Scott (1996)) or semiparametric compositions only (e.g. as MacDonald et al. (2011)) in the time invariant context as here.

1.5 Concluding remarks

In this chapter, based on the scale mixing of the Student's t distribution and following the guidelines developed by Afuecheta et al. (2018) six new compound distribution models were developed. PDF, CDF, moments and characteristic functions for every suggested model alongside some empirical applications to the small set of financial data were provided. From the results obtained, considered compositions demonstrate a good potential for both future applied financial investigations and further functional developments of the discussed models. From the P-P and Q-Q plots provided it is worthwhile introducing functionality of assymetry to enhance modelling power of the considered models. It may be also worth considering the functional form of the Laplace distribution for compositions of new, but similar models. This should lead to more parsimonious closed form expressions for calculations than for the models considered here, allow introducing skewness for asymmetric parametrizations and keep the heavy tails property in the parent distribution. Though, Laplace based mixing may be less innovative than the Student's t models here, portfolio of these combined distribution models is still underdeveloped and largely limited to the gamma and beta distributions mixing in the relevant applications (e.g. Wichitaksorn et al. (2015); L. Chen et al. (2018)). For example, Laplace combination with the generalized gamma distribution has been suggested since Choy & Chan (2008), but its closed forms for estimations are yet to be derived and reported. Given early performance of the Student's t generalized gamma mixture, its Laplace alternative may be attractive due to the smaller set of parameters under the scaling and functional power of the generalized gamma distribution.

Less burdensome introduction of asymmetry to the proposed distribution models may be implemented using approach of the inverse scale factors as in Fernández & Steel (1998). On the other hand, more applied investigations may be interested in including proposed models in the Value-at-Risk forecasting investigations under the autoregressive time series context of GARCH models. Apart from the Student's t gamma combined distribution model (e.g. Afuecheta et al. (2016)), such investigations for the similar and models provided here are yet to be conducted. Time series framework of the Generalized Autoregressive Score (GAS) models of Creal et al. (2013) and Harvey (2013) may be also appropriate for the applied Value-at-Risk forecasting studies. GAS allows relatively straightforward introduction of the time-varying dynamics for any desired parameter of the distribution models, but closed form recursions for such estimations may be more challenging to provide.

Time series context extensions may also allow considering these distribution models in the multivariate fully parametric copula applications as discussed in Nikoloulopoulos et al. (2012) and Patton (2012, 2013), though more "traditional" extensions for the multivariate analysis of financial data as discussed in Kotz & Nadarajah (2004) may be also worth investigating in the future. Future studies may be also more inclusive and make relevant comparisons to the wider pool of the popular distribution models in finance as well as go beyond the likelihood based comparisons here. May be it is worth comparing relatively good performing half normal and inverse gamma combined distributions with models governed by the same number of parameters to thoroughly realize these distributions modelling potential.

Chapter 2

On the choice of parameters for dynamic kernel density and distribution estimation of financial returns

2.1 Introduction and motivation

Harvey & Oryshchenko (2012) formalize dynamic kernel PDF and CDF estimators for time series data employing exponentially declining weights. Exponential weights are well-known and widely applied, see for example comprehesive overviews of Gardner (1985) and Gardner (2006) for this weighting family in general, while nonparametric estimators in combination with exponentially declining weights are considered by Yu & Jones (1998), Gijbels et al.

(1999), S. X. Chen & Tang (2005), Taylor (2007), Bessa et al. (2012), Zhang et al. (2014), Taylor & Jeon (2015) and Arora & Taylor (2016) among others. However, Harvey & Oryshchenko (2012) are the first to provide a thorough discussion focused on the dynamic kernel estimators for the time-varying PDF, CDF and quantile modelling with weighting schemes derived from volatility modelling (e.g. as thoroughtly discussed in Andersen et al. (2006)) dedicating their empirical applications to the exponential weights. Attractiveness of Harvey & Oryshchenko's (2012) nonparametric approach, for example over its alternative, nonparametric quantile regression as in Taylor (2007), is driven by the fact that it allows for a full set of time-varying quantiles to be tracked simultaneously either for exploratory dependence (e.g. as in Busetti & Harvey (2010), Harvey (2010) and Bücher et al. (2015)) or risk exposure modelling (e.g. if considered as a variant of the well-known J.P. Morgan's (1996) RiskMetrics™ methodology for Value-at-Risk estimation) without compromising the fact that these quantiles may cross.[1] This prominent empirical property may be less crucial for the risk modelling context, however for the dependence modelling, dynamic estimators as in Harvey & Oryshchenko (2012) can be a more useful tool in the semiparametric copula frameworks (e.g. see comprehensive copula discussions by Patton (2012, 2013)) among other applications.

Overall, methods described in Harvey & Oryshchenko (2012) seem to be extending common perks of the nonparametric methods to the time series data context, have some of their properties broadly discussed by Robinson (1983), Wu et al. (2010) and Aït-Sahalia & Park (2016) with some empirical attempts traced back to Hall & Patil (1994) and should be

[1] For quantile regressions Gouriéroux & Jasiak (2008) show that quantiles may cross on the example of the widespread Conditional Autoregressive VAR, CAViaR, model of Engle & Manganelli (2004). Harvey (2010) and Harvey & Oryshchenko (2012) argue that their quantile algorithms, based on the empirical estimations of PDFs and CDFs, insure that this does not happen.

appealing for practical use in the context of financial returns, but Harvey & Oryshchenko's (2012) computational short-cut of employing MLE motivates, often essential, choice of the parameters dilemma for the nonparametric specifications of the DGP. To specify, MLE in nonparametric estimations has a strong reputation of providing parameters often leading to unsatisfactory or suboptimal estimation outcomes, see wide-ranging discussions in Schuster & Gregory (1981), Chow et al. (1983), Hall (1987a,b) and Q. Li & Racine (2007) among others. MLE pitfalls motivate this chapter to focus on its rigorous, yet parsimonious and simple alternative routine of least-squares estimations (LSE).

LSE is also common in the general exponential weights context (e.g. Hyndman et al. (2008)) and unlike MLE can be rewritten for parameters optimal for PDF and CDF in the nonparametric estimations context (e.g. Q. Li & Racine (2007)). It, however, comes at the cost of more demanding and often longer running estimations than MLE, especially in the time series context, when following the guidelines of Harvey & Oryshchenko (2012), LSE function "blocks" in the recursive optimizations get heavier with every iteration. Therefore, this chapter first outlines LSE for the nonparametric estimators enhaced with exponential weights as used in applications of Harvey & Oryshchenko (2012) both in the form of PDF and CDF. Second, realizing the high computational demands of LSE and aiming to simplify this burden, this chapter next describes simplifications using binned estimators similar to the traditional estimators as in Hall (1982), Scott & Sheather (1985) and Hall & Wand (1996), but enhanced to accommodate exponentially declining weights both for PDF and CDF LSE optimizations. Third, given binned estimators for exponential weights this chapters considers, perhaps the most basic, dynamic adaptive/varying bandwidths estimators similar to those described in Sain & Scott (1996) and Scott (2015), but on the basis of binned estimators with exponentially declining weights. Fourth, for a small data set of financial returns, consisting of ten daily series of different length, under the composite density forecast evaluation criteria as in Berkowitz (2001), this chapter demonstrates that all parameters chosen with LSE, both for the full-scale estimators as in Harvey & Oryshchenko's (2012) and their binned computational simplifications and bandwidth variation based adaptations, provide appropriate estimations for the most of the series without corrections for location/ scale as previously achieved in Harvey & Oryshchenko (2012) for parameters obtained with MLE. Fifth, it is pointed out that even after GARCH pre-filtering for location and scale, MLE struggles suggesting good combinations of the parameters as per evaluation criteria chosen in this chapter. Overall, similar to the investigation of Q. Li & Racine (2008), but for the dynamic nonparameteric context as in Harvey & Oryshchenko (2012), it is empirically shown that the parameters optimal either for PDF or CDF should lead to the good estimation outcomes using LSE routine and, most importantly, that Harvey & Oryshchenko's (2012) estimators are in fact viable for financial returns applications as long as their parameters for estimations are chosen accordingly.

This chapter does not go beyond the weighting schemes used in applications of Harvey & Oryshchenko (2012), but following their work it continues to Section 2.2 by introducing kernel estimators in the recursive forms for the time series data context. Then it proceeds to Section 2.3 by describing related, but different quantile mining approaches that briefly outline how approaches similar to Harvey & Oryshchenko (2012) stand-out from the similar techniques available prior to their work. Next, Section 2.4 describes considered approaches to parameters estimation. Section 2.5 provides chosen criteria/tests for evaluation of different parameters choices provided by LSE variations and MLE. Section 2.6 describes data set of financial returns used to obtain empirical findings for the selected parameters. Section 2.7 conducts diagnostics and describes obtained empirical results. Finally, Section 2.8 concludes this chapter and outlines some further research paths.

2.2 Dynamic kernel density and distribution estimation

Time-varying PDF of financial returns can be estimated using a kernel (KDE) and an appropriate weighting scheme used for volatility modelling as in Andersen et al. (2006) among others. That is,

$$\hat{f}_t(x) = \frac{1}{h}\sum_{i=1}^{t} K\left(\frac{x - x_i}{h}\right) w_{t,i} \text{ for } i = 1, ..., t \text{ and } t = 1, ..., T, \tag{2.1}$$

where $K(\cdot)$ is a kernel in the form of PDF as provided in Silverman (1986), Wand & Jones (1995), Q. Li & Racine (2007) and Tsybakov (2009) among others, h is a PDF optimal bandwidth parameter and $x \in \mathbb{R}$ as in Chapter 1. Similarly, dynamic CDF of financial returns can be estimated using an appropriate kernel functional form (KCDE) and the weighting scheme as in (2.1) or by integrating obtained PDF estimate with (2.1). That is,

$$\hat{F}_t(x) = \sum_{i=1}^{t} W\left(\frac{x - x_i}{\beta}\right) w_{t,i} \text{ for } i = 1, ..., t \text{ and } t = 1, ..., T, \tag{2.2}$$

where $W(\cdot)$ is a kernel in the form of CDF and β is the CDF optimal bandwidth parameter as described in Q. Li & Racine (2007). $w_{t,i}$ in (2.1) and (2.2) may be some non-negative weights always satisfying $\sum_{i=1}^{t} w_{t,i} = 1$ and strategically allocating more weight to recent observations. To achieve this and obtain one step ahead time conditional nonparametric PDF forecasts, Harvey & Oryshchenko (2012) employ recursive format of the simple exponential weights as described in Gardner (1985) among others. The forecast is outlined by

$$\hat{f}_{t+1|t}(x) = \omega \cdot \hat{f}_{t|t-1}(x) + \frac{1-\omega}{h} \cdot K\left(\frac{x-x_t}{h}\right), \tag{2.3}$$

where ω is the "learning rate" or more formally, parameter governing the dynamics of exponential weights. By recursive substitution from (2.3) it can be shown that

$$\begin{aligned}
\hat{f}_{t+1|t} &= \omega\left[\omega \cdot \hat{f}_{t-1|t-2} + \frac{1-\omega}{h} \cdot K\left(\frac{x-x_{t-1}}{h}\right)\right] + \frac{1-\omega}{h} \cdot K\left(\frac{x-x_t}{h}\right), \\
&= \omega^2 \cdot \hat{f}_{t-1|t-2}(x) + \omega^1 \cdot \frac{1-\omega}{h} \cdot K\left(\frac{x-x_{t-1}}{h}\right) + \omega^0 \cdot \frac{1-\omega}{h} \cdot K\left(\frac{x-x_t}{h}\right), \\
&= \omega^1 \hat{f}_{t-1}(x) + \left[(1-\omega)\sum_{i=1}^{t} \omega^{t-i}\right] \cdot K\left(\frac{x-x_i}{h}\right) h^{-1}; \qquad (2.4)
\end{aligned}$$

hence for estimators in (2.1) and (2.2) $w_{t,i}(\omega) = (1-\omega) \cdot \omega^{t-i}$ as given in Andersen et al. (2006) can be employed for estimations. However, Harvey & Oryshchenko (2012) suggest relying on

$$w_{t,i}(\omega) = \begin{cases} \dfrac{1-\omega}{1-\omega^t} \cdot \omega^{t-i} & \text{if } \omega \in (0,1), \\ \\ 1/t & \text{if } \omega = 1. \end{cases} \tag{2.5}$$

Weights in (2.5) are identical to weights used in Andersen et al. (2006) and obtained by recursive substitution in (2.4) as $t \longrightarrow \infty$ if $\omega \in (0,1)$ and simply take the form of equally weighted expanding window if $\omega = 1$. Figure 2.1 provides illustrations of the weights from (2.5) for different ω values. Since this weighting scheme employs information only up to t, following Harvey (1990), it is further referred to as the weighting scheme for exponential filtering. Dynamic KCDE in (2.2) in the form of the one step ahead forecast of the time

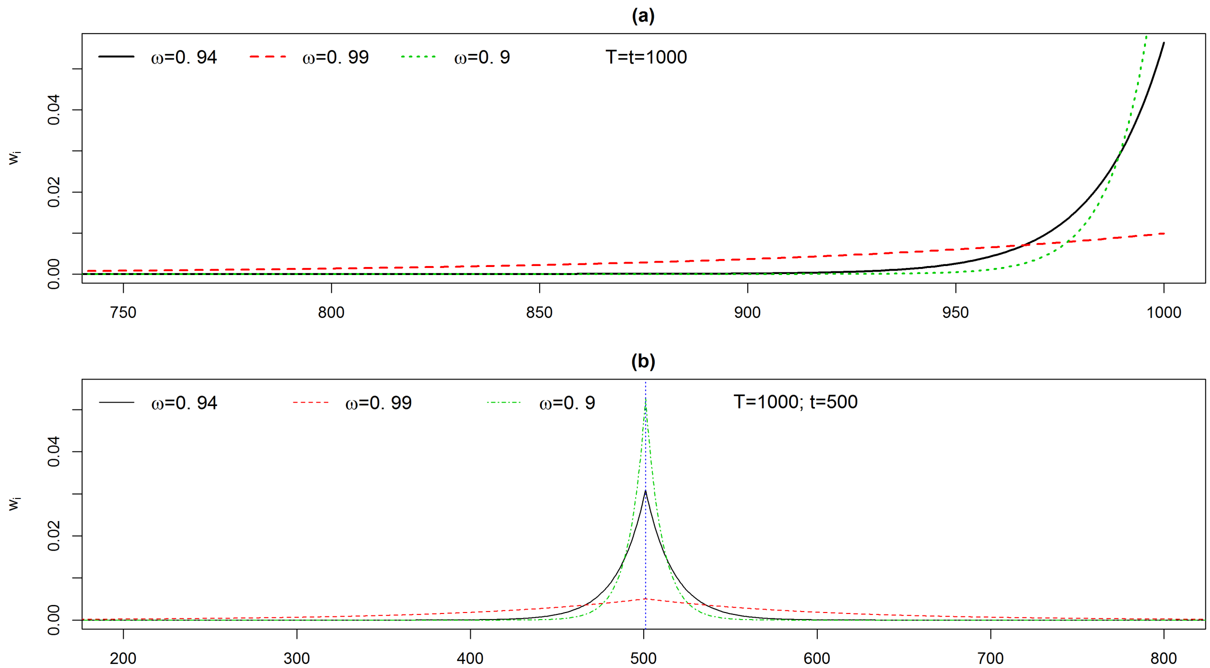

Figure 2.1: Exponentially declining weights for (a) filtering at a time $t = T = 1000$ and (b) smoothing at a time $t = 500$; $T = 1000$ for different values of ω.

conditional CDF is then given by

$$\hat{F}_{t+1|t}(x) = \omega \cdot \hat{F}_{t|t-1}(x) + (1-\omega) \cdot W\left(\frac{x - x_t}{\beta}\right). \tag{2.6}$$

Following Gardner (1985) forecasts in (2.3) and (2.6) can be also rewritten in the error correction form. That is,

$$\hat{f}_{t+1|t}(x) = \omega \cdot \hat{f}_{t|t-1}(x) + (1-\omega) \cdot \mathbb{e}_t(x), \tag{2.7}$$

where $\mathbb{e}_t(x) = h^{-1} \cdot K\left(\frac{x - x_t}{h}\right) - \hat{f}_{t|t-1}(x)$, for PDF and

$$\hat{F}_{t+1|t}(x) = \omega \cdot \hat{F}_{t|t-1}(x) + (1-\omega) \cdot \mathbb{E}_t(x), \tag{2.8}$$

where $\mathbb{E}_t(x) = W\left(\frac{x - x_t}{\beta}\right) - \hat{F}_{t|t-1}(x)$, for CDF respectively. Further recursive upgrades may include a "stable" form to offer a slightly different dynamics. To be specific, following variance targeting scheme as in Kristensen & Linton (2004), Shephard & Sheppard (2010) and Francq et al. (2011) among others, "stable" PDF forecasts are given by

$$\hat{f}_{t+1|t}(x) = (1 - \omega^* - \omega) \cdot \bar{f}(x) + \frac{\omega^*}{h} K\left(\frac{x - x_t}{h}\right) + \omega \cdot \hat{f}_{t|t-1}(x) \tag{2.9}$$

and "stable" CDF forecasts are respectively outlined by

$$\hat{F}_{t+1|t}(x) = (1 - \omega^* - \omega) \cdot \bar{F}(x) + \omega^* \cdot W\left(\frac{x - x_t}{\beta}\right) + \omega \cdot \hat{F}_{t|t-1}(x), \tag{2.10}$$

where $\bar{f}(x)$ in (2.9) is the unconditional nonparametric PDF estimate and $\bar{F}(x)$ in (2.10) is the unconditional nonparameteric CDF estimate at the time T respectively. Recursions in (2.9) and (2.10) are stable with ω^* adjusting to account for unconditional PDF and CDF estimates if $\omega^* + \omega < 1$ and reverse to the recursions of the simple exponential form in (2.3) and (2.6) if $\omega^* + \omega = 1$.

A two-sided exponentially smoothed conditional PDF ($\hat{f}_{t|T}(x)$) and CDF ($\hat{F}_{t|T}(x)$) estimates may also be constructed. To illustrate, an algorithm may be designed storing $\hat{f}_{t|t-1}(x)$ and $\hat{F}_{t|t-1}(x)$ estimates from recursions (2.3) and (2.6) and then calculating backward recursions

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

for PDF and

$$\mathbb{R}_{t-1}(x) = \omega \left[\mathbb{R}_t(x) + W\left(\frac{x - x_t}{\beta}\right) - \hat{F}_{t|t-1}(x) \right];\ \ t = T, \ldots, 2$$

for CDF, where $\mathbb{r}_T = 0$ and $\mathbb{R}_T = 0$ insuring that estimates at the time $t = T$ are identical with (2.3) and (2.6). Finally, a forward recursion

$$\hat{f}_{t|T}(x) = \omega \cdot \hat{f}_{t|t-1}(x) + (1 - \omega) \cdot \left[\mathbb{r}_t(x) + h^{-1}K\left(\frac{x - x_t}{h}\right) \right];\ \ t = 1, \ldots, T \tag{2.11}$$

outlines a two-sided conditional PDF estimate, while

$$\hat{F}_{t|T}(x) = \omega \cdot \hat{F}_{t|t-1}(x) + (1 - \omega) \cdot \left[\mathbb{R}_t(x) + W\left(\frac{x - x_t}{\beta}\right) \right];\ \ t = 1, \ldots, T \tag{2.12}$$

outlines a two-sided conditional CDF estimate. All of the above recursions may require a suitable initialization procedures (e.g. if the entire conditional PDF and CDF estimates at each time t are necessary). For (2.11) and (2.12) $\hat{f}_{1|T}(x) = \left[\mathbb{r}_1(x) + h^{-1}K\left(\frac{x - x_1}{h}\right) \right]$ and $\hat{F}_{1|T}(x) = \left[\mathbb{R}_1(x) + W\left(\frac{x - x_1}{\beta}\right) \right]$ while for (2.3), (2.6), (2.7) and (2.8) $\hat{f}_{1|0}(x) = 0$ and $\hat{F}_{1|0}(x) = 0$ may be appropriate respectively. Most importantly, weights used in computations are constrained to sum to unity. Computation algorithms should also run over some predefined grid range $[x_{min}; x_{max}] \in \mathbb{R}$ and store matrices of $\mathbb{e}_t(x)$, $\mathbb{E}_t(x)$ or $\mathbb{r}_t(x)$, $\mathbb{R}_t(x)$ depending on their recursive forms. Estimations may be simplified by precomputing weights for time t such that $t = m + 1, \ldots, T$, where m is some number of observations initializing the procedure, with algorithms as in Koopman & Harvey (2003) among others and plugging them into the estimators given in (2.1) and (2.2). To accommodate precomputed two-sided exponential weights, simple rearrangement of (2.1) provides

$$\hat{f}_{t|T}(x) = \frac{1}{h}\sum_{i=1}^{T} K\left(\frac{x-x_i}{h}\right) w_{t,T,i}(\omega) \;\text{ for }\; i=1,...,t \;\text{ and }\; t=1,...,T, \tag{2.13}$$

for PDF estimate and similar rearrangement of (2.2) produces

$$\hat{F}_{t|T}(x) = \sum_{i=1}^{T} W\left(\frac{x-x_i}{\beta}\right) w_{t,T,i}(\omega) \;\text{ for }\; i=1,...,t \;\text{ and }\; t=1,...,T, \tag{2.14}$$

for CDF estimate. In a large sample two-sided exponential weights can be given by

$$w_{t,i}(\omega) = \begin{cases} \dfrac{1-\omega}{1+\omega-\omega^{t}-\omega^{T-t+1}} \cdot \omega^{|t-i|} & \text{if } \omega \in (0,1), \\ 1/T & \text{if } \omega = 1, \end{cases} \tag{2.15}$$

where identical to the weights in (2.5) ω is the parameter governing weights' dynamics. From Harvey (1990) such weighting scheme may be referred to as exponential smoothing, since it employs information on time T, commonly the entire sample at hand. Also note that the weights in (2.15) for the middle of a large sample may take a reduced form of $\dfrac{1-\omega}{1+\omega} \cdot \omega^{|t-i|}$ and are illustrated in Figure 2.1 for different values of ω.

Dynamic kernel estimators may be also allowed to have varying bandwidths parameters similar to adaptive forms as first defined by Breiman et al. (1977). For weights in (2.5) dynamic adaptive KDE (AKDE) is given by

$$\hat{f}_t(x) = \sum_{i=1}^{t} K\left(\frac{x-x_i}{h_i}\right) w_{t,i}(\omega) \cdot h_i^{-1} \tag{2.16}$$

and for weights in (2.15) it is

$$\hat{f}_{t|T}(x) = \sum_{i=1}^{T} K\left(\frac{x - x_i}{h_i}\right) w_{t,T,i}(\omega) \cdot h_i^{-1}. \tag{2.17}$$

Similarly, adaptive KCDE (AKCDE) for (2.5) is given by

$$\hat{F}_t(x) = \sum_{i=1}^{t} W\left(\frac{x - x_i}{\beta_i}\right) w_{t,i}(\omega) \tag{2.18}$$

and for (2.15) by

$$\hat{F}_{t|T}(x) = \sum_{i=1}^{T} W\left(\frac{x - x_i}{\beta_i}\right) w_{t,T,i}(\omega) \tag{2.19}$$

respectively. Recursive form of (2.16) for the one step ahead PDF forecasts is then

$$\hat{f}_{t+1|t}(x) = \omega \cdot \hat{f}_{t|t-1}(x) + \frac{1-\omega}{h_t} \cdot K\left(\frac{x - x_t}{h_t}\right) \tag{2.20}$$

and of (2.18) for the one step ahead CDF forecasts respectively is

$$\hat{F}_{t+1|t}(x) = \omega \cdot \hat{F}_{t|t-1}(x) + (1-\omega) \cdot W\left(\frac{x - x_t}{\beta_t}\right). \tag{2.21}$$

Note that bandwidths h_t and β_t in the above recursions alter with each observation at the estimation time t. This also allows imposing assumptions on the DGPs of bandwidths, allocating them a specific time-varying structure (e.g. from Silverman's (1986) rule-of-thumb, bandwidths can be approximated with a scale/volatility parameter, hence a time-varying dynamics as discussesd in Lucas & Zhang (2016) among others may be adopted for evolving bandwidths), and developing data and time depended plug-in bandwidths choice similar to Altman & Leger (1995) respectively. Next, error correction form of (2.20) is then simply

$$\hat{f}_{t+1|t}(x) = \omega \cdot \hat{f}_{t|t-1}(x) + (1-\omega) \cdot \mathbb{e}_t(x), \tag{2.22}$$

where $\mathbb{e}_t(x) = h_t \cdot K\left(\frac{x-x_t}{h_t}\right) - \hat{f}_{t|t-1}(x)$, and of (2.21)

$$\hat{F}_{t+1|t}(x) = \omega \cdot \hat{F}_{t|t-1}(x) + (1-\omega) \cdot \mathbb{E}_t(x), \tag{2.23}$$

where $\mathbb{E}_t(x) = W\left(\frac{x-x_t}{\beta_t}\right) - \hat{F}_{t|t-1}$. "Stable" form PDF forecast for (2.16) is consequently specified as

$$\hat{f}_{t+1|t}(x) = (1-\omega^* - \omega) \cdot \bar{f}(x) + \frac{\omega^*}{h_t} K\left(\frac{x-x_t}{h_t}\right) + \omega \cdot \hat{f}_{t|t-1}(x), \tag{2.24}$$

where $\bar{f}(x)$ is now can be given by the unconditional PDF estimate as described in Abramson (1982) for example, and CDF forecast for (2.18) is then

$$\hat{f}_{t+1|t}(x) = (1-\omega^* - \omega) \cdot \bar{F}(x) + \omega^* \cdot W\left(\frac{x-x_t}{\beta_t}\right) + \omega \cdot \hat{F}_{t|t-1}(x). \tag{2.25}$$

Recursion for (2.17) can be constructed upon $\hat{f}_{t|t-1}(x)$ output from (2.20) by first computing

$$\mathbb{r}_{t-1} = \omega\left[\mathbb{r}_t(x) + h_t^{-1} K\left(\frac{x-x_t}{h_t}\right) - \hat{f}_{t|t-1}(x)\right]; \quad t = T, \cdots, 2,$$

where $\mathbb{r}_T = 0$, and is given by

$$\hat{f}_{t|T}(x) = \omega \cdot \hat{f}_{t|t-1}(x) + (1-\omega) \cdot \left[\mathbb{r}_t(x) + h_t^{-1} K\left(\frac{x - x_t}{h_t}\right)\right]; \;\; t = 1, \cdots, T. \tag{2.26}$$

Similarly for (2.19) recursive form is produced calculating

$$\mathbb{R}_{t-1} = \omega \left[\mathbb{R}_t(x) + W\left(\frac{x - x_t}{\beta_t}\right) - \hat{F}_{t|t-1}(x)\right]; \;\; t = T, \cdots, 2,$$

where $\mathbb{R}_T = 0$, and plugging it in to

$$\hat{F}_{t|T}(x) = \omega \cdot \hat{F}_{t|t-1}(x) + (1-\omega) \cdot \left[\mathbb{R}_t(x) + W\left(\frac{x - x_t}{\beta_t}\right)\right]; \;\; t = 1, \cdots, T. \tag{2.27}$$

If financial returns need to be accounted for trends, seasonality or serial correlation, Harvey & Oryshchenko (2012) suggest overcoming imperfections in the above discussed recursive estimators and their weighting schemes respectively, using parameteric pre-filtering for location and/or scale. This pragmatic suggestion formally transforms $h^{-1}K(\cdot)$, $h_t^{-1}K(\cdot)$ and $W(\cdot)$ (e.g. in (2.3), (2.16), (2.6) and (2.18) respectively) to

$$\begin{aligned}
(h\hat{\sigma}_{t|t-1})^{-1} K\left(\frac{x - (x_t - \hat{y}_{t|t-1})}{h\hat{\sigma}_{t|t-1}}\right) &= (h\hat{\sigma}_{t|t-1})^{-1} K\left(\frac{x - x_t + \hat{y}_{t|t-1}}{h\hat{\sigma}_{t|t-1}}\right); \\
W\left(\frac{x - (x_t - \hat{y}_{t|t-1})}{\beta\hat{\sigma}_{t|t-1}}\right) &= W\left(\frac{x - x_t + \hat{y}_{t|t-1}}{\beta\hat{\sigma}_{t|t-1}}\right); \\
(h_t\hat{\sigma}_{t|t-1})^{-1} K\left(\frac{x - (x_t - \hat{y}_{t|t-1})}{h_t\hat{\sigma}_{t|t-1}}\right) &= (h_t\hat{\sigma}_{t|t-1})^{-1} K\left(\frac{x - x_t + \hat{y}_{t|t-1}}{h_t\hat{\sigma}_{t|t-1}}\right); \\
W\left(\frac{x - (x_t - \hat{y}_{t|t-1})}{\beta_t\hat{\sigma}_{t|t-1}}\right) &= W\left(\frac{x - x_t + \hat{y}_{t|t-1}}{\beta_t\hat{\sigma}_{t|t-1}}\right),
\end{aligned}$$

where $\hat{y}_{t|t-1}$ GARCH/GAS conditional location/mean equation and $\hat{\sigma}_{t|t-1}$ GARCH/GAS conditional location/variance equation outputs. These rough adjustments may be regarded as necessary location and scale corrections, though the target here: is to attempt providing accurate forecasts without pre-filtering.

2.3 Time-varying quantiles

2.3.1 Mining quantiles from time-varying kernel CDF estimate

Time-varying CDF estimate may be useful for obtaining time-varying quantiles. From Nadaraya (1964) for dynamic nonparametric estimators it directly follows that

$$\hat{F}_{t|t-1}(x) = \tau$$

and consequently dynamic τ-quantile may be obtained by solving

$$\hat{F}^{-1}_{t|t-1}(\tau) = \hat{\xi}_{t|t-1}(\tau),$$

where $\hat{F}_{t|t-1}(x)$ is obtained with the appropriate nonparametric estimators in Section 2.2. Harvey (2010) and Harvey & Oryshchenko (2012) suggest simplifying quantile extraction by employing linear interpolation between

$$\xi_{low} = \max_i(\xi_i : \hat{F}_t(\xi_i) \leq \tau) \text{ and } \xi_{high} = \min_i(\xi_i : \hat{F}_t(\xi_i) \geq \tau)$$

and then using this point for initialization of optimization for the quantile of interest at a time t. Obviously, to implement this, CDF estimate for each t on a fine grid is necessary. In this work, numerical algorithm is designed to store CDF estimates for each t at the grid of at least 10 000 points. For such grid scale, interpolation and then numerical solution may not be even necessary. In fact, either $\xi_{low} = \max_i(\xi_i : \hat{F}_t(\xi_i) \leq \tau)$ or $\xi_{high} = \min_i(\xi_i : \hat{F}_t(\xi_i) \geq \tau)$ is typically sufficient and is not very different from the final solution of the τ-quantile.

2.3.2 Nonparametric quantile regression

Yu & Jones (1998) show that quantiles can be modelled directly by the means of the nonparameteric quantile regression, which in its simplest form can be given by

$$\hat{F}_t(x) = \frac{\sum_t K_h(y - y_i) * \mathbb{I}_{\{x_i < x\}}}{\sum_t K_h(y - y_i)}, \tag{2.28}$$

where $K_h(y - y_i)$ is a conventional PDF kernel function exploited as a weighting function for observations and $\mathbb{I}_{\{x_i < x\}}$ is an indicator function taking the value of 1 if x_i observation is less than estimation point x as described in Taylor (2007). Estimator in (2.28) can take more elaborate form of the double kernel quantile estimator outlined by

$$\hat{F}_t(x) = \frac{\sum_t K_h(y - y_i) * W[(x - x_i)\beta^{-1}]}{\sum_t K_h(y - y_i)}, \tag{2.29}$$

combining kernels for PDF and CDF estimation to approximate kernel weighted (from the PDF component) and quantiles (from the CDF component) respectively. Gijbels et al. (1999) provides proofs and specific technical conditions to obtain a set of weights for exponential filtering from $\sum_t K_h(y-y_i)$ estimator; however from and similar to Taylor (2007), this can be briefly described by first setting $y = t$ and $y_i = i$ components of the $\sum_t K_h(y-y_i)$ weighting estimator. If $K(\cdot)$ is a one-sided exponential kernel allocating weights on data to the left of the location t Taylor (2007) points out that $K_h(y-y_i) = \omega^{t-i}$ or for more elaborate form of the exponential weights obtained by the recursive substitution in (2.4) $K_h(y-y_i) = (1-\omega)\omega^{t-i}$ and redefines (2.28) as

$$\hat{F}_t(x) = \frac{\sum_t (1-\omega)\omega^{t-i} * \mathbb{I}_{\{x_i<x\}}}{\sum_t (1-\omega)\omega^{t-i}} \tag{2.30}$$

and consequently redefines the estimator in (2.29) as

$$\hat{F}_t(x) = \frac{\sum_t (1-\omega)\omega^{t-i} * W[(x-x_i)\beta^{-1}]}{\sum_t (1-\omega)\omega^{t-i}} \tag{2.31}$$

for the large samples of financial returns. This also allows rewriting (2.30) and (2.31) in the recursive formats of

$$\hat{F}_t(x) = \omega \cdot \hat{F}_{t-1}(x) + (1-\omega) \cdot \mathbb{I}_{\{x_t<x\}}$$

and

$$\hat{F}_t(x) = \omega \cdot \hat{F}_{t-1}(x) + (1-\omega) \cdot W\left(\frac{x-x_t}{\beta}\right)$$

respectively. With minor misuse of notation, the latter is essentially identical to the estimator in (2.6), however approach to estimations is different. Right-hand side of (2.30) and (2.31) is usually solved iteratively for different values of x until the estimation target of $\hat{F}_t(x) = \tau$ is achieved with specific technical estimation details described in Gijbels et al. (1999), Taylor

(2007) and Taylor & Jeon (2015) among others. This allows modelling each quantile with individual ω parameter, providing different quantiles with different dynamics (e.g. as also empirically identified in Taylor (2007)) and differentiates approaches discussed in Section 2.2 from other direct time-varying quantile modelling tools (e.g. as in Engle & Manganelli (2004) and De Rossi & Harvey (2009) among others).

2.4 Unknown parameters estimation

2.4.1 Maximum likelihood

For exponential filtering, Harvey & Oryshchenko (2012) suggest maximizing a log-likelihood function in the predictive recursive form given by

$$\begin{aligned} l_f(\omega, h) &= \frac{1}{T-m}\sum_{t=m}^{T-1} \log[\hat{f}_{t+1|t}(x_{t+1})] \\ &= \frac{1}{T-m}\sum_{t=m}^{T-1} \log\left[\frac{1}{h}\sum_{i=1}^{t} K\left(\frac{x_{t+1}-x_i}{h}\right) w_{t,i}(\omega)\right], \end{aligned} \tag{2.32}$$

where m is the number of observations initializing the estimation procedure and $w_{t,i}(\omega)$ are one-sided exponential weights. There is no firm theoretical foundation on the value of m, but following Markovich's (2008) guidelines for estimations in the weekly dependent data enviroment, such as financial returns described in Cont (2001), good estimations with recursive kernels may be data demanding. Therefore, Harvey & Oryshchenko (2012) follow the recommended minimum by Markovich (2008) and set $m = 100$ in their applied examples. Further, in this thesis for dynamic nonparametric estimators, $m = 250$ unless stated otherwise.

This choice of m approximately equates to one year of trading data for daily financial returns once holidays and weekends are excluded.

For exponential smoothing, Harvey & Oryshchenko (2012) obtain unknown parameters by maximizing a modified log-likelihood cross-validation (CV) criterion given by

$$\begin{aligned} l_s(\omega, h) =& \frac{1}{T}\sum_{t=1}^{T}\log[\hat{f}_{-t|T}(x_t)] \\ =& \frac{1}{T}\sum_{t=1}^{T}\log\left[\frac{1}{h}\sum_{i=1,\, i\neq t}^{T} K\left(\frac{x_t - x_i}{h}\right) w_{t,T,i}(\omega)\right], \end{aligned} \tag{2.33}$$

where $w_{t,T,i}(\omega)$ are two-sided exponential weights.

2.4.2 Least-squares

From the LSE routines derived in Q. Li & Racine's (2007), but for the context of the exponentially weighted time-series data, it can be shown that

$$E_X[\hat{f}_{t+1}(X)] = \int \hat{f}_{t+1|t}(x) f_{t+1}(x) dx$$

and [2]

$$E_X[\hat{f}_t(X)] = \int \hat{f}_{t|T}(x) f_t(x) dx,$$

where E_X denotes expectations with respect to the true DGP X and not with respect to the time-series observations $\{x_i\}_{i=1}^{t}$ and $\{x_i\}_{\substack{i=1 \\ i\neq t}}^{T}$ allocated to estimate $\hat{f}(\cdot)$ respectively. Similar to likelihoods in (2.32) and (2.33) of Harvey & Oryshchenko (2012) and standard LSE practises described in Q. Li & Racine (2007), expectations over the DGP can be fulfilled by replacing

[2] $\int_{-\infty}^{\infty} = \int$ unless stated otherwise.

E_X with appropriately chosen exponentially weighted sample average. When estimations target is exponential filtering, expectations are given by

$$E_X[\hat{f}_{t+1}(X)] \approx \frac{1}{h}\sum_{i=1}^{t} K\left(\frac{x_{t+1}-x_i}{h}\right) w_{t,i}(\omega), \tag{2.34}$$

and when estimations target is exponential smoothing, expectations are outlined by

$$E_X[\hat{f}_t(X)] \approx \frac{1}{h}\sum_{i=1,i\neq t}^{T} K\left(\frac{x_t-x_i}{h}\right) w_{t,T,i}(\omega) \tag{2.35}$$

respectively. Equation in (2.34) approximates DGP with respect to the predictive observation x_{t+1} using exponentially weighted observations up to t and weighting scheme for exponential filtering. On the other hand, for (2.35) there are less information restrictions and thus, equation in (2.35) constitutes a different expectation, which is fulfilled by the exponentially smoothing PDF with respect to the omitted x_t observation. The latter is an explicit modification of the common leave-one-out estimator as in Q. Li & Racine (2007) designed to accommodate two-sided exponential weights. Forming expectations set by (2.34) and (2.35), it is relatively straightforward to provide least-squares loss functions for unknown parameters estimation.

LSE for filtering in the form of PDF begins with

$$\mathbb{ls}_f(\omega, h) = \frac{1}{T-m}\sum_{t=m}^{T-1}\int \left[f_{t+1}(x) - \hat{f}_{t+1|t}(x)\right]^2 dx, \tag{2.36}$$

where the main difference from (2.32) is that the objective is minimization of the squared integrated distance between estimated and true PDFs. (2.36) decomposes to

$$\mathbb{ls}_f(\omega, h) = \frac{1}{T-m}\sum_{t=m}^{T-1}\left[\int f_{t+1}(x)^2 dx - 2\int f_{t+1}(x)\hat{f}_{t+1|t}(x)dx + \int \hat{f}_{t+1|t}(x)^2 dx\right],$$

where the first squared integrand is irrelevant for optimization and can be omitted providing

$$\mathbb{ls}_f(\omega, h) = \frac{1}{T-m}\sum_{t=m}^{T-1}\left[\int \hat{f}_{t+1|t}(x)^2 dx - 2\int f_{t+1}(x)\hat{f}_{t+1|t}(x)dx\right]. \tag{2.37}$$

In (2.37) with (2.1) the first squared integrand is straightforward, while double integrated cross product of the true PDF and its estimate now follow from (2.34) providing

$$\begin{aligned}\mathbb{ls}_f(\omega, h) = \frac{1}{T-m}\sum_{t=m}^{T-1}\Bigg[\frac{1}{h^2}\sum_{j=1}^{t}\sum_{i=1}^{t}\int K\left(\frac{x-x_i}{h}\right)w_{t,i}(\omega)K\left(\frac{x-x_j}{h}\right)w_{t,j}(\omega)dx \\ -\frac{2}{h}\sum_{i=1}^{t}K\left(\frac{x_{t+1}-x_i}{h}\right)w_{t,i}(\omega)\Bigg]\end{aligned}$$

or with a twofold convolution kernel

$$\begin{aligned}\mathbb{ls}_f(\omega, h) = \frac{1}{T-m}\sum_{t=m}^{T-1}\Bigg[\frac{1}{h}\sum_{j=1}^{t}\left(\sum_{i=1}^{t}\bar{K}\left(\frac{x_j-x_i}{h}\right)w_{t,i}(\omega)\right)w_{t,j}(\omega) \\ -\frac{2}{h}\sum_{i=1}^{t}K\left(\frac{x_{t+1}-x_i}{h}\right)w_{t,i}(\omega)\Bigg],\end{aligned} \tag{2.38}$$

where $w_{t,i}(\omega)$ are one-sided exponential weights, m is the constant initializing the procedure as in (2.32) and $\bar{K}(\cdot)$ is a twofold convolution kernel. LSE for smoothing in the form of PDF requires decomposition steps relying on (2.35) and is similar to (2.33) respectively. That is,

$$
\begin{aligned}
\mathbb{ls}_s(\omega, h) =& \frac{1}{T}\sum_{t=1}^{T}\int\left[f_t(x) - \hat{f}_{-t|T}(x)\right]^2 dx,\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\int f_t(x)^2 dx - 2\int f_t(x)\hat{f}_{t|T}(x)dx + \int \hat{f}_{t|T}(x)^2 dx\right];\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\int \hat{f}_{t|T}(x)^2 dx - 2\int f_t(x)\hat{f}_{t|T}(x)dx\right];\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{h^2}\sum_{j=1}^{T}\sum_{i=1}^{T}\int K\left(\frac{x-x_i}{h}\right)w_{t,T,i}(\omega)K\left(\frac{x-x_j}{h}\right)w_{t,T,j}(\omega)dx\right.\\
&\left.-\frac{2}{h}\sum_{i=1,\,i\neq t}^{T}K\left(\frac{x_t-x_i}{h}\right)w_{t,T,i}(\omega)\right];\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{h}\sum_{j=1}^{T}\left(\sum_{i=1}^{T}\bar{K}\left(\frac{x_j-x_i}{h}\right)w_{t,T,i}(\omega)\right)w_{t,T,j}(\omega)\right.\\
&\left.-\frac{2}{h}\sum_{i=1,i\neq t}^{T}K\left(\frac{x_t-x_i}{h}\right)w_{t,T,i}(\omega)\right].
\end{aligned}
\tag{2.39}
$$

Rewriting (2.36) for CDF provides

$$
\mathbb{LS}_f(\omega, \beta) = \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[F_{t+1}(x) - \hat{F}_{t+1|t}(x)\right]^2 dx,
$$

which for the approximation of the DGP as in Bowman et al. (1998) and for CDF estimate outlined by (2.2) is given by

$$
\mathbb{LS}_f(\omega, \beta) = \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[\mathbb{I}_{\{x_{t+1}<x\}} - \sum_{i=1}^{t}W\left(\frac{x-x_i}{\beta}\right)w_{t,i}(\omega)\right]^2 dx, \tag{2.40}
$$

where $\mathbb{I}_{\{x_{t+1}<x\}}$ is already familiar indicator function taking the value of 1 if predictive observation x_{t+1} is less than estimation point x. Although approximation of the DGP as in Bowman et al. (1998) provides a wide ground for (2.40) to be modified for the entire range of recursive estimators in Section 2.2 and "stable" form in (2.10) in particular, aims of this work do not go beyond the weighting schemes used in empirical applications of Harvey & Oryshchenko (2012). Therefore, all loss function expressions' style matches the original work.

LSE for smoothing for CDF is then provided by

$$\begin{aligned}\mathbb{LS}_s(\omega,\beta) =& \frac{1}{T}\sum_{t=1}^{T}\int\left[F_t(x)-\hat{F}_{t|T}(x)\right]^2 dx,\\ =& \frac{1}{T}\sum_{t=1}^{T}\int\left[\mathbb{I}_{\{x_t<x\}}-\sum_{i=1,\,i\neq t}^{T}W\left(\frac{x-x_i}{\beta}\right)w_{t,T,i}(\omega)\right]^2 dx.\end{aligned} \tag{2.41}$$

For all parameters estimation functions a choice of the kernel functional form is necessary. For stationary or close to being stationary data as the most of financial returns (e.g. Cont (2001)) with dynamic kernel estimators, selection of the kernel functional form is less crucial than the choice of the bandwidth parameter (e.g. Robinson (1983)) similar to the "static" kernel estimators (e.g. Wand & Jones (1995)). A pragmatic and popular choice is Gaussian kernel. From Q. Li & Racine (2007), its PDF form is given by $K(x)=\dfrac{1}{\sqrt{2\pi}}e^{(-x^2/2)}$, its convolution form by $\int K(y)K(x-y)dy=\bar{K}(x)=\dfrac{1}{\sqrt{4\pi}}e^{(-x^2/4)}$ and finally, its CDF form is outlined by $W(x)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{(-z^2/2)}dz$ respectively. Technically, convolution (e.g. as in (2.37)) for LSE squared integrand with Gaussian kernel is derived under the data independence assumptions, but Hart & Vieu (1990) experiments for the LSE efficiency under the dependent data, indicate that LSE seems to have an ability of tolerating moderate dependence levels in the data under investigation (e.g. below 0.5 in the absolute values as given by the autocorrelation functions (ACFs)). Therefore in Section 2.6, to meet prior estimation formalities before the empirical comparisons of MLE and its LSE alternatives, alongside some descriptive statistics, ACFs for the raw log-returns of the data under investigation are also provided.

In all further nonparametric computations Gaussian kernels are employed unless stated otherwise. Unbounded support of the Gaussian kernels theoretically insures that the density is not zero at any estimation point and potentially benefits estimations with MLE, however, problem of the different tail decay factors for DGP and employed kernel, remains unresolved for such estimations. To be specific, Hall (1987b) points out that if the DGP exhibit tails decay slower than is specified by the kernel functional form, MLE offers $h \to \infty$. It is worthwhile highlighting that ussually for filtering estimations are performed under some information restrictions and therefore, sudden market corrections may be more problematic for MLE for such estimations than under the CV conditions discussed in Hall (1987b).

2.4.3 Dynamic binned estimators and least-squares

Unlike MLE, LSE typically demands more computing power to perform identification of the optimal unknowns, since it has more components to evalute in each iteration. Problem of faster evaluation with LSE for parameters optimization in indepedent and identically distributed (i.i.d.) data framework mainly remains for the multivariate data settings (e.g. Sain (2002)), however recursive forms for the time series data, where size of the estimation blocks' increase with every iteration, bring the problem of the time efficient estimations to the univariate time series setting. Binned estimators are well-known to provide good evaluations of the parameters and reduce computational burden in the i.i.d. univariate set up (e.g. Scott & Sheather (1985), Wand & Jones (1995) and Hall & Wand (1996) among others) and can also be rewritten to accommodate weighting schemes appropriate for the time series modelling and obtain dynamic expression forms.

For weights in (2.5) a dynamic binned KDE (BKDE) for exponential filtering can be expressed as

$$\dot{f}_{t+1|t}(x) = \frac{1}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}} K\left(\frac{x - c_{\mathbf{s}}}{h}\right)\dot{w}_{t,\mathbf{s}}(\omega), \tag{2.42}$$

where $\dot{w}_{t,\mathbf{s}}(\omega) = \sum_{i=1}^{t} \mathbb{I}_{\{x_i \in B_{\mathbf{s}}\}} \cdot w_{t,i}(\omega)$, which essentially is a total dynamic weights count function for observations in each bin $B_{\mathbf{s}}$ up to the estimation time t for the total number of non-empty bins $\mathbf{a}$ with bin centres $c_{\mathbf{s}}$. Employing a similar strategy but weights in (2.15) a dynamic BKDE for exponential smoothing of PDF is then can be given by

$$\dot{f}_{t|T}(x) = \frac{1}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}} K\left(\frac{x - c_{\mathbf{s}}}{h}\right)\dot{w}_{t,T,\mathbf{s}}(\omega), \tag{2.43}$$

where $\dot{w}_{t,T,\mathbf{s}}(\omega) = \sum_{i=1}^{T} \mathbb{I}_{\{x_i \in B_{\mathbf{s}}\}} \cdot w_{t,T,i}(\omega)$. Now, for dynamic binned KCDE (BKCDE), for exponential filtering of CDF, it is obtained

$$\dot{F}_{t+1|t}(x) = \sum_{\mathbf{s}=1}^{\mathbf{a}} W\left(\frac{x - c_{\mathbf{s}}}{\beta}\right)\dot{w}_{t,\mathbf{s}}(\omega) \tag{2.44}$$

and for exponential smoothing respectively

$$\dot{F}_{t|T}(x) = \sum_{\mathbf{s}=1}^{\mathbf{a}} W\left(\frac{x - c_{\mathbf{s}}}{\beta}\right)\dot{w}_{t,T,\mathbf{s}}(\omega). \tag{2.45}$$

Estimators in (2.42), (2.43), (2.44) and (2.45) can also be transformed to provide bandwidths for dynamic adaptive PDF and CDF estimation, as outlined in Section 2.2, mimicking the approach of Sain & Scott (1996) and Hazelton (2003) for i.i.d. framework and yielding $h_i = h_{\mathbf{s}}(x_i \in B_{\mathbf{s}})$ and $\beta_i = \beta_{\mathbf{s}}(x_i \in B_{\mathbf{s}})$ for the dynamic adaptive estimations respectively. For exponential filtering of PDF that is,

$$\tilde{f}_{t+1|t}(x) = \sum_{\mathbf{s}=1}^{\mathbf{a}} K\left(\frac{x - c_{\mathbf{s}}}{h_{\mathbf{s}}}\right) \dot{w}_{t,\mathbf{s}}(\omega) \cdot h_{\mathbf{s}}^{-1}, \tag{2.46}$$

for exponential smoothing of PDF

$$\tilde{f}_{t|T}(x) = \sum_{\mathbf{s}=1}^{\mathbf{a}} K\left(\frac{x - c_{\mathbf{s}}}{h_{\mathbf{s}}}\right) \dot{w}_{t,T,\mathbf{s}}(\omega) \cdot h_{\mathbf{s}}^{-1}, \tag{2.47}$$

for exponential filtering of CDF

$$\tilde{F}_{t+1|t}(x) = \sum_{\mathbf{s}=1}^{\mathbf{a}} W\left(\frac{x - c_{\mathbf{s}}}{\beta_{\mathbf{s}}}\right) \dot{w}_{t,\mathbf{s}}(\omega) \tag{2.48}$$

and finally, for exponential smoothing of CDF

$$\tilde{F}_{t|T}(x) = \sum_{\mathbf{s}=1}^{\mathbf{a}} W\left(\frac{x - c_{\mathbf{s}}}{\beta_{\mathbf{s}}}\right) \dot{w}_{t,T,\mathbf{s}}(\omega). \tag{2.49}$$

All of the binned time conditional PDF and CDF expressions can also be rewritten in the recursive format as in Section 2.2 for the full-scale estimators, however, the main target of the binned estimators is to achieve less time consuming, hopefully accurate evaluations, with LSE and vary the bandwidth parameter over the range of returns within their functionality.

Therefore, for clearness and simplicity reasons, only estimators which are actually used for computations are provided.

To achieve outlined estimation targets with binned estimators, a binning strategy can be decisive. There is sufficiently known on binning details and strategies for i.i.d. estimation framework (e.g. Hall & Wand (1996)), however specific binning guidelines for the time series recursive estimations are yet to be reported. Hypothetically, time series framework implies that bins and their respective bin centres should be time-varying for exponential filtering. That is: $B_{t,\mathbf{s}}$ and $c_{t,\mathbf{s}}$, which can be defined on the basis of the parsimonious quantile based binning recommendation of Hazelton (2003) among others, but for the time-varying quantiles specifications to match information assumptions behind the dynamic estimations. To be more specific, time-varying histograms as provided in Harvey (2010), or any other approach to dynamic quantiles from Section 2.3, can be used to provide dynamic bins and their bin centres for estimations. This can be viewed as a necessary dynamic pilot estimation to perform binned related evaluations. While this might be justified for the dynamic adaptive estimations, it may be more problematic to explain an additional layer of complexity, if the estimation aim is to merely speed up unknown parameters estimation with LSE.[3] Though keeping bins constant is less ground breaking, it is simple and may provide relative stability at the estimation stage. For the constant bins specifications, binned estimators' dynamics are driven only by the time-varying bin weights, therefore, the key challenge for not evolving bins may be approximation of the parameters governing the dynamics of the weighting scheme. On the other hand, for exponential smoothing, constant bins do not violate information

[3]Dynamic bins should require a matrix of dynamic quantiles from the pilot estimate upon which vectors/matrices, depending on the computation algorithm design, of dynamic bin centres and bin weights to be computed.

assumptions and parameters obtained with this type of binned estimators may be exploited for filtering. This is the most appropriate when the dynamic kernel estimators serve an exploratory purpose, as it ussually is for this division of statistical/econometrics methods in general, and also discussed by Harvey & Oryshchenko (2012), though not in the context of dynamic binned estimators.

As for the dynamic adaptive estimations, it may be also interesting pre-computing weights for the ω parameter obtained by other methods to speed up estimations. Present chapter does not alter weighting schemes used in Harvey & Oryshchenko's (2012) applications and though, higher flexibility in the bandwidth parameters through adaptive estimations should hypothetically improve overall density and distributions specifications, it cannot go beyond the forecasts evaluation characteristics dependent on the weighting scheme as discussed in the next section. ω obtained with MLE may be of a particular interest, since estimation performance of the exponential learning rate and bandwidth by MLE can be separated. However, for the binned reliant bandwidths variation strategy, it may be more informative also experimenting with the ω parameter obtained for the binned simplifications as in (2.42), (2.43), (2.44) and (2.45).

Now, keeping bin specifications less burdensome, LSE for the time series estimations can be rearranged for faster evaluations and for multiple bandwidths provision with previously outlined dynamic binned estimators for PDF and CDF. First, rewriting (2.36) for (2.42) provides

$$
\begin{aligned}
\mathbb{ls}_f^b(\omega,h) =& \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[f_{t+1}(x)-\dot{f}_{t+1|t}(x)\right]^2dx,\\
=&\frac{1}{T-m}\sum_{t=m}^{T-1}\left[\int f_{t+1}(x)^2dx-2\int f_{t+1}(x)\dot{f}_{t+1|t}(x)dx+\int \dot{f}_{t+1|t}(x)^2dx\right],\\
=&\frac{1}{T-m}\sum_{t=m}^{T-1}\left[\int \dot{f}_{t+1|t}(x)^2dx-2\int f_{t+1}(x)\dot{f}_{t+1|t}(x)dx\right],\\
=&\frac{1}{T-m}\sum_{t=m}^{T-1}\left[\frac{1}{h^2}\sum_{\mathbf{s}=1}^{\mathbf{a}}\sum_{\mathbf{q}=1}^{\mathbf{a}}\int K\left(\frac{x-c_{\mathbf{s}}}{h}\right)\dot{w}_{t,\mathbf{s}}(\omega)K\left(\frac{x-c_{\mathbf{q}}}{h}\right)\dot{w}_{t,\mathbf{q}}(\omega)dx\right.\\
&\left.-\frac{2}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}}K\left(\frac{x_{t+1}-c_{\mathbf{s}}}{h}\right)\dot{w}_{t,\mathbf{s}}(\omega)\right],\\
=&\frac{1}{T-m}\sum_{t=m}^{T-1}\left[\frac{1}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}}\left(\sum_{\mathbf{q}=1}^{\mathbf{a}}\bar{K}\left(\frac{c_{\mathbf{s}}-c_{\mathbf{q}}}{h}\right)\dot{w}_{t,\mathbf{q}}(\omega)\right)\dot{w}_{t,\mathbf{s}}(\omega)\right.\\
&\left.-\frac{2}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}}K\left(\frac{x_{t+1}-c_{\mathbf{s}}}{h}\right)\dot{w}_{t,\mathbf{s}}(\omega)\right],
\end{aligned}
\tag{2.50}
$$

where $\dot{w}_{t,\mathbf{s}}(\omega)$ is the one-sided non-empty bins exponential weights count function as in (2.42), m is the constant initializing the procedure and $\bar{K}(\cdot)$ is a convolution kernel similar to the full-scale LSE for PDF loss function in (2.38). Then, (2.39) with (2.43) is now outlined by

$$
\begin{aligned}
\mathbb{ls}_s^b(\omega, h) =& \frac{1}{T}\sum_{t=1}^{T}\int\left[f_t(x) - \dot{f}_{-t|T}(x)\right]^2 dx,\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\int f_t(x)^2 dx - 2\int f_t(x)\dot{f}_{-t|T}(x)dx + \int \dot{f}_{-t|T}(x)^2 dx\right],\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\int \dot{f}_{-t|T}(x)^2 dx - 2\int f_t(x)\dot{f}_{-t|T}(x)dx\right],\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{h^2}\sum_{\mathbf{s}=1}^{\mathbf{a}}\sum_{\mathbf{q}=1}^{\mathbf{a}}\int K\left(\frac{x - c_{\mathbf{s}}}{h}\right)\dot{w}_{t,T,\mathbf{s}}(\omega)K\left(\frac{x - c_{\mathbf{q}}}{h}\right)\dot{w}_{t,T,\mathbf{q}}(\omega)dx\right.\\
&\left. - \frac{2}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}} K\left(\frac{x_t - c_{\mathbf{s}}}{h}\right)\dot{w}_{t,T,\mathbf{s}}(\omega)\right],\\
=& \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}}\left(\sum_{\mathbf{q}=1}^{\mathbf{a}}\bar{K}\left(\frac{c_{\mathbf{s}} - c_{\mathbf{q}}}{h}\right)\dot{w}_{t,T,\mathbf{q}}(\omega)\right)\dot{w}_{t,T,\mathbf{s}}(\omega)\right.\\
&\left. - \frac{2}{h}\sum_{\mathbf{s}=1}^{\mathbf{a}} K\left(\frac{x_t - c_{\mathbf{s}}}{h}\right)\dot{w}^*_{t,T,\mathbf{s}}(\omega)\right],
\end{aligned} \tag{2.51}
$$

where $\dot{w}^*_{t,T,\mathbf{s}}(\omega) = \sum_{\substack{i=1\\ t\neq i}}^{t} \mathbb{I}_{\{x_i \in B_{\mathbf{s}}\}} w_{t,T,i}(\omega)$. Consequently, rewriting (2.40) for (2.44) provides

$$
\mathbb{LS}_f^b(\omega, \beta) = \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[F_{t+1}(x) - \dot{F}_{t+1|t}(x)\right]^2, \tag{2.52}
$$

$$
= \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[\mathbb{I}_{\{x_{t+1}<x\}} - \sum_{\mathbf{s}=1}^{\mathbf{a}} W\left(\frac{x - c_{\mathbf{s}}}{\beta}\right)\dot{w}_{t,\mathbf{s}}(\omega)\right]^2 \tag{2.53}
$$

and (2.41) for (2.45) is rearranged to

$$
\mathbb{LS}_s^b(\omega, \beta) = \frac{1}{T}\sum_{t=1}^{T}\int\left[F_t(x) - \dot{F}_{-t|T}(x)\right]^2, \tag{2.54}
$$

$$
= \frac{1}{T}\sum_{t=1}^{T}\int\left[\mathbb{I}_{\{x_t<x\}} - \sum_{\mathbf{s}=1}^{\mathbf{a}} W\left(\frac{x - c_{\mathbf{s}}}{\beta}\right)\dot{w}^*_{t,T,\mathbf{s}}(\omega)\right]^2. \tag{2.55}
$$

Finally, rewriting (2.36) for (2.46) provides

$$\begin{aligned}
\mathbb{ls}^a_f(\omega, h_{\mathbf{s}}) =& \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[f_{t+1}(x) - \tilde{f}_{t+1|t}(x)\right]^2 dx,\\
=& \frac{1}{T-m}\sum_{t=m}^{T-1}\left[\int f_{t+1}(x)^2 dx - 2\int f_{t+1}(x)\tilde{f}_{t+1|t}(x)dx + \int \tilde{f}_{t+1|t}(x)^2 dx\right],\\
=& \frac{1}{T-m}\sum_{t=m}^{T-1}\left[\int \tilde{f}_{t+1|t}(x)^2 dx - 2\int f_{t+1}(x)\tilde{f}_{t+1|t}(x)dx\right],\\
=& \frac{1}{T-m}\sum_{t=m}^{T-1}\left[\sum_{\mathbf{s}=1}^{\mathbf{a}}\sum_{\mathbf{q}=1}^{\mathbf{a}}\int K\left(\frac{x-c_{\mathbf{s}}}{h_{\mathbf{s}}}\right)\dot{w}_{t,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\cdot K\left(\frac{x-c_{\mathbf{q}}}{h_{\mathbf{q}}}\right)\dot{w}_{t,\mathbf{q}}(\omega)h_{\mathbf{q}}^{-1}dx\right.\\
&\left.-2\sum_{\mathbf{s}=1}^{\mathbf{a}}K\left(\frac{x_{t+1}-c_{\mathbf{s}}}{h_{\mathbf{s}}}\right)\dot{w}_{t,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\right],\\
=& \frac{1}{T-m}\sum_{t=m}^{T-1}\left[\sum_{\mathbf{s}=1}^{\mathbf{a}}\left(\sum_{\mathbf{q}=1}^{\mathbf{a}}\bar{K}\left(\frac{c_{\mathbf{s}}-c_{\mathbf{q}}}{h_{\mathbf{q}}}\right)\dot{w}_{t,\mathbf{q}}(\omega)\right)\dot{w}_{t,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\right.\\
&\left.-2\sum_{\mathbf{s}=1}^{\mathbf{a}}K\left(\frac{x_{t+1}-c_{\mathbf{s}}}{h_{\mathbf{s}}}\right)\dot{w}_{t,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\right]. \qquad (2.56)
\end{aligned}$$

Likewise, (2.39) for (2.47) is

$$
\begin{aligned}
\mathbb{Is}^a_s(\omega, h_{\mathbf{s}}) =& \frac{1}{T}\sum_{t=1}^{T}\int\left[f_t(x)-\tilde{f}_{-t|T}(x)\right]^2 dx,\\
=&\frac{1}{T}\sum_{t=1}^{T}\left[\int f_t(x)^2dx - 2\int f_t(x)\tilde{f}_{-t|T}(x)dx + \int \tilde{f}_{-t|T}(x)^2dx\right],\\
=&\frac{1}{T}\sum_{t=1}^{T}\left[\int \tilde{f}_{-t|T}(x)^2dx - 2\int f_t(x)\tilde{f}_{-t|T}(x)dx\right],\\
=&\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{\mathbf{s}=1}^{\mathbf{a}}\sum_{\mathbf{q}=1}^{\mathbf{a}}\int K\left(\frac{x-c_{\mathbf{s}}}{h_{\mathbf{s}}}\right)\dot{w}_{t,T,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\cdot K\left(\frac{x-c_{\mathbf{q}}}{h_{\mathbf{q}}}\right)\dot{w}_{t,T,\mathbf{q}}(\omega)h_{\mathbf{q}}^{-1}dx\right.\\
&\left.-2\sum_{\mathbf{s}=1}^{\mathbf{a}}K\left(\frac{x_t-c_{\mathbf{s}}}{h_{\mathbf{s}}}\right)\dot{w}_{t,T,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\right],\\
=&\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{\mathbf{s}=1}^{\mathbf{a}}\left(\sum_{\mathbf{q}=1}^{\mathbf{a}}\bar{K}\left(\frac{c_{\mathbf{s}}-c_{\mathbf{q}}}{h_{\mathbf{q}}}\right)\dot{w}_{t,T,\mathbf{q}}(\omega)\right)\dot{w}_{t,T,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\right.\\
&\left.-2\sum_{\mathbf{s}=1}^{\mathbf{a}}K\left(\frac{x_t-c_{\mathbf{s}}}{h_{\mathbf{s}}}\right)\dot{w}^*_{t,T,\mathbf{s}}(\omega)h_{\mathbf{s}}^{-1}\right],
\end{aligned}
\tag{2.57}
$$

while (2.41) for (2.48)

$$
\mathbb{LS}^a_f(\omega, \beta_{\mathbf{s}}) = \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[F_{t+1}(x)-\tilde{F}_{t+1|t}(x)\right]^2, \tag{2.58}
$$

$$
= \frac{1}{T-m}\sum_{t=m}^{T-1}\int\left[\mathbb{I}_{\{x_{t+1}<x\}}-\sum_{\mathbf{s}=1}^{\mathbf{a}}W\left(\frac{x-c_{\mathbf{s}}}{\beta_{\mathbf{s}}}\right)\dot{w}_{t,\mathbf{s}}(\omega)\right]^2 \tag{2.59}
$$

and (2.41) for (2.49)

$$
\mathbb{LS}^a_s(\omega, \beta_{\mathbf{s}}) = \frac{1}{T}\sum_{t=1}^{T}\int\left[F_t(x)-\tilde{F}_{-t|T}(x)\right]^2, \tag{2.60}
$$

$$
= \frac{1}{T}\sum_{t=1}^{T}\int\left[\mathbb{I}_{\{x_t<x\}}-\sum_{\mathbf{s}=1}^{\mathbf{a}}W\left(\frac{x-c_{\mathbf{s}}}{\beta_{\mathbf{s}}}\right)\dot{w}^*_{t,T,\mathbf{s}}(\omega)\right]^2. \tag{2.61}
$$

2.5 Forecasts evaluation criteria

Diebold et al. (1998) point out that good PDF forecasts are described by the independently and uniformly distributed Probability Integral Transforms (PITs). PITs for such forecasts evaluations can be obtained with

$$\hat{F}_{t|t-1}(x_t) = \sum_{t=m}^{T} \left[\sum_{i=1}^{t-1} W\left(\frac{x_t - x_i}{\beta}\right) w_{t,i}(\omega) \right], \tag{2.62}$$

if the density forecasts are performed with (2.1) and

$$\hat{F}_{t|t-1}(x_t) = \sum_{t=m}^{T} \left[\sum_{i=1}^{t-1} W\left(\frac{x_t - x_i}{\beta_i}\right) w_{t,i}(\omega) \right], \tag{2.63}$$

if the forecasts are obtained with (2.16) and evaluations are restricted for exponential filtering as in Harvey & Oryshchenko (2012). It is worthwhile pointing out that both time-varying quantiles and PITs are CDF optimal outputs. Therefore, parameters optimal for CDF may have higher preference in the further discussion than the estimated parameters optimal for PDF.

Uniformity of the PITs, obtained by (2.62) or by (2.63), can be then assessed with Kolmogorov-Smirnov (K-S) and/or Cramer-von Mises (CvM) nonparametric tests. To specify, K-S test quantifies departures from the uniformity with

$$K - S_d = sup|\tilde{F}_k(\beta, \omega) - U_k|, \tag{2.64}$$

where U_k denotes the true uniform CDF at some grid estimation point k and $\tilde{F}_k(\beta,\omega)$ is empirical CDF (eCDF) estimate of the PITs at k with *sup* indicating maximum value of the absolute distances between PITs eCDF and the expected true uniform CDF. On the other hand, CvM test has the same components as in (2.64), but quantifies departures from the uniformity differently. That is,

$$CvM_d = \sum_k \left[\tilde{F}_k(\beta,\omega) - U_k\right]^2. \tag{2.65}$$

Patton (2013) highlights that for both (2.64) and (2.65) asymptotic properties are well-known and thus, their respective p-values ($K-S_p$ and CvM_p) are often available within relevant computing software enviroments. $K-S_p$ and CvM_p values are likelihoods to observe a sample of the same size with the same maximum absolute for K-S and total squared for CvM distances from the target distribution. Hence roughly, the higher are the $K-S_p$ or CvM_p, the greater are the chances that PITs are uniformly distributed. Patton (2013) sets PITs uniformity decision cut-off points for both tests at the 5% significance levels. Diebold et al. (1998), however, argue that K-S and CvM tests on thier own are not sufficient for forecasts evaluations, since they only evaluate uniformity of the PITs and neglect their independence property. Diebold et al. (1998) then proceed suggesting an "expert judgement" based visual assesment criteria, which has been quantified into a convinient likelihood ratio (LR) test by Berkowitz (2001).

If PITs are transformed with $\Phi^{-1}\left(\hat{F}_{t|t-1}(x_t)\right)$, where $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard normal, they may be expected to be independent and normally distributed with $\mu=0$, $\rho=0$ and $\sigma^2=0$ 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 for PITs normality and independence is then formally defined as

$$\mathrm{LR}_d = -2(\log L(0,1,0) - \log L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})) \tag{2.66}$$

and is expected to be $\chi^2(3)$ distributed with the rejection threshold set up at the 5% significance level similar to the K-S and CvM tests. Further statistic computed with (2.64), (2.65) and (2.66) is used to evaluate forecasts for the data described in the next section. Also note that further evaluations with (2.62) and (2.63) are also made for the PDF by MLE and LSE optimal parameters, but following notation and estimators convergence formalities, they are provided in this form. Generally, such practice is not uncommon, see Q. Li & Racine (2008) for example, and is actually used in Harvey & Oryshchenko (2012).

2.6 Data

2.6.1 Data and descriptive statistics

To survey performance of parameters provided by the estimation approaches described in Section 2.4 a data set consisting of ten daily logarithmic financial returns has been constructed. It includes log-returns for stock index prices, commodities and currency exchange rates obtained such as:

$$x_{\iota,t} = \log\left(\frac{P_{\iota,t}}{P_{\iota,t-1}}\right) \cdot 100,$$

Date	Returns	T	MIN	MAX	Mean	St.d.	Skew.	Kurt.	LB(12)	LB(12)^2	JB	AH(12)
30.03.01-31.08.17	BRENT	4172	-19.89	18.13	0.019	2.23	-0.09	8.59	0.07	≈ 0.00	≈ 0.00	≈ 0.00
30.03.01-31.08.17	NOK/USD	4136	-5.92	4.42	-0.003	0.77	0.11	5.89	0.47	≈ 0.00	≈ 0.00	≈ 0.00
30.03.01-31.08.17	RUB/USD	4080	-12.86	10.94	0.017	0.83	0.76	41.62	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
28.03.13-27.03.18	GBP/USD	1303	-8.39	3.001	-0.005	0.58	-2.26	36.89	0.85	≈ 0.00	≈ 0.00	≈ 0.00
10.12.98-28.06.02	NASDAQ	890	-10.16	13.25	-0.036	2.47	0.16	4.73	0.06	≈ 0.00	≈ 0.00	≈ 0.00
03.01.06-01.03.10	S&P 500	1045	-9.46	10.95	-0.012	1.64	-0.22	11.35	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
03.01.06-01.03.10	FTSE 100	1052	-9.26	9.38	0.004	1.52	-0.09	6.92	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
01.09.88-28.02.92	NIKKEI 225	857	-6.82	12.43	-0.02	1.40	0.50	12.34	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
01.01.10-30.01.15	GOLD/USD	1325	-9.51	3.98	0.011	1.10	-0.84	9.04	0.96	≈ 0.00	≈ 0.00	≈ 0.00
28.03.13-27.03.18	BTC/USD	1301	-60.1	51.70	0.349	6.14	-0.62	25.79	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00

Table 2.1: Descriptive statistics for the specified returns series, where T denotes sample size, MIN minimum value, MAX maximum value, St.d. standard deviation, Skew. skewness, Kurt. kurtosis, LB(12) and LB(12)^2 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 the Jarque-Bera probabilities for normality and AH(12) the Lagrange Multiplier test for autoregressive conditional heteroscedasticity probabilities of order 12 for no autocorrelation, normality and homoscedasticity.

where ι indexes the financial time series price/exchange rate and corresponding returns on the trading day t.[4] Obtained log-returns are described in Table 2.1 and are illustrated in Figure 2.2. There were no specific guidelines adopted for composition of this data set. However, a crucial criteria for inclusion was: to obtain log-returns which cover at least one commonly recognized event causing or directly connected to the financial turbulence and posses common characteristics for log-returns as consolidated by Cont (2001). Therefore, composed data set consists of returns of different length, ranging from medium to large sample sizes, and covers such events as "BREXIT" referendum outcome correction for GBP/USD exchange rate, "Dot-com" bubble collapse for NASDAQ composite index, cryptocurrencies "hype" for BTC/USD, Japanese "asset price" bubble collapse for NIKKEI225 and unfulfilled "quantitative easing" expectations correction for GOLD/USD, while S&P500 and FTSE100 exemplify, usually very representative in financial research, financial crisis associated volatility. Specific

[4]Further all log-returns for computations are obtained using this expression unless stated otherwise. In Chapter 1, log-returns are not scaled by 100, merely to insure that all evaluated likelihoods have negative values and are convenient for the log-likelihood based model comparisons.

dates and time-periods for the obtained log-returns are provided in Table 2.1 and/or could be found on the time series plots of these observations in Figure 2.2.

Considered samples of the larger size are also interesting for estimations and comparisons. For example, obtained BRENT log-returns cover volatility in the hydrocarbon markets over the periods of both substantial price increases and meltdowns linked to the financial crisis events and concerns on the oil market oversupplies. Therefore, inclusion of NOK/USD and RUB/USD may be also interesting from the applied research perspective of the currencies of oil exporting countries. For example, despite sharp oil market corrections, considered NOK/USD returns depict the lowest kurtosis value among series reported in Table 2.1. On the other hand, RUB/USD depicts the highest kurtosis value. It is interesting to investigate how considered methods for parameters estimation perform for each of these series, since NOK/USD can be concluded to be relatively light-tailed, BRENT moderate-tailed and RUB/USD heavy-tailed respectively. Oil and oil exporting countries contagion researchers may also be interested for considering these series for investigation, since central banks of these two countries had opposite monetary policy responses for the oil market lows in the late 2015 - early 2016 in addition to the contrasting characteristics of these series and BRENT log-returns. Therefore, considered time series are interesting not only from an applied econometric perspective, but also from an economy policy perspective.

To sum up, returns under investigation can be described as non-normal, heavy-tailed, with evident volatility clustering and should be fruitful for empirical investigation of the unknown parameters' role in estimations with the dynamic kernel methods. Sharp market corrections that are linked to the events briefly described above are not hard to visually distinguish on the corresponding time series plots for these returns in Figure 2.2, while their ACFs plots

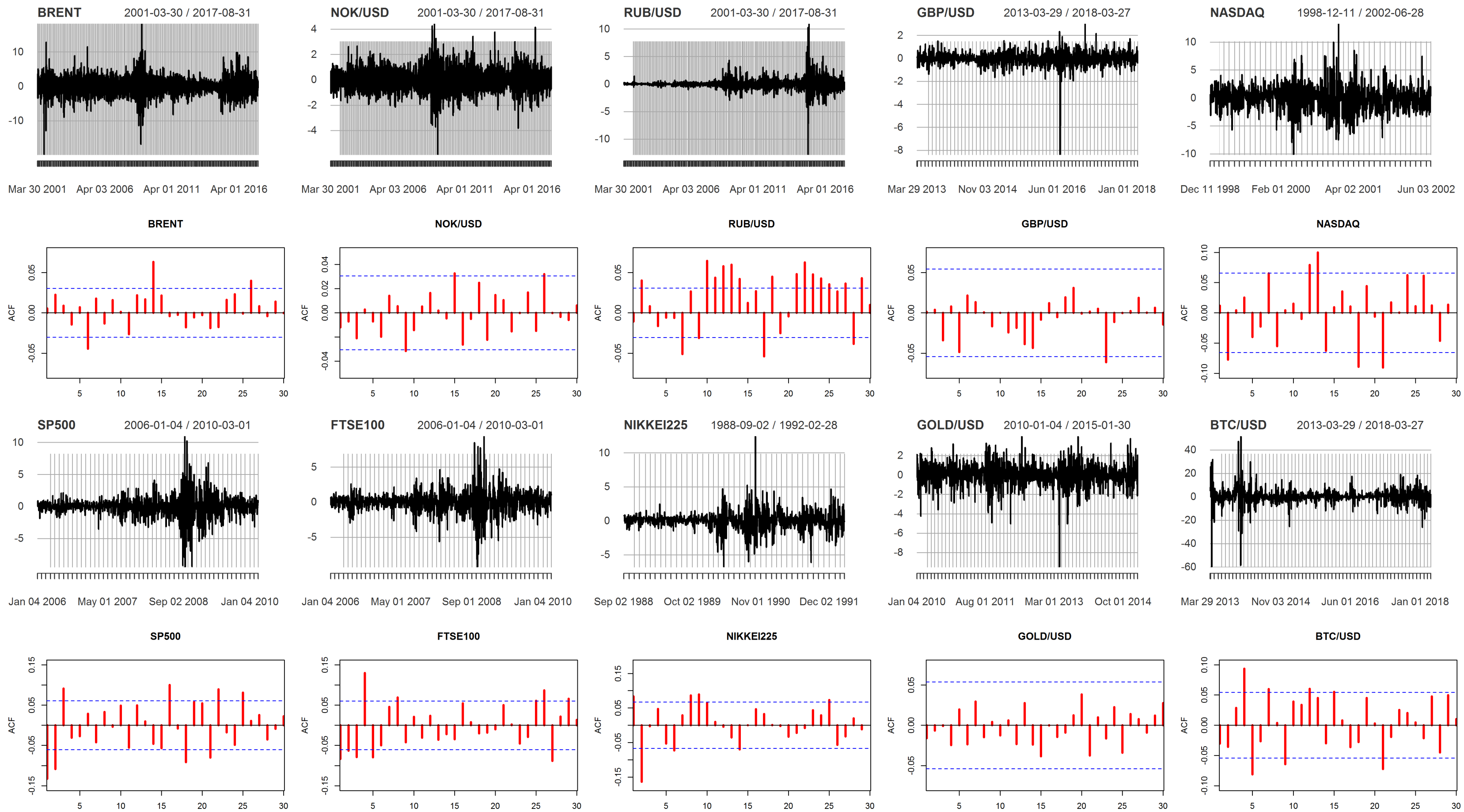

Figure 2.2: Time series plots for the daily log-returns of BRENT, NOK/USD, RUB/USD, GBP/USD, NASDAQ, SP500, FTSE100, NIKKEI225, GOLD/USD and BTC/USD with their autocorrelation functions.

parameters	BRENT	NOK/USD	RUB/USD	GBP/USD	NASDAQ	S&P 500	FTSE 100	NIKKEI 225	GOLD/USD	BTC/USD
θ_0	0.0382	-0.0156	0.0033	0.0028	0.0656	0.0709^{***}	0.0624^{***}	0.0709^{**}	0.0443^{**}	0.2529
θ_1	0.8371^{***}	0.7430^{*}	–	-0.887^{***}	-0.670	0.6994^{***}	0.6445^{***}	-0.404^{**}	-0.295	0.9735^{***}
θ_2	-0.8227^{***}	-0.7541^{***}	0.0694^{***}	0.8700^{***}	0.7028	-0.786^{***}	-0.7154^{***}	0.5017^{***}	0.2444	-0.957^{***}
α_0	0.0119^{***}	0.0032^{***}	0.00003	0.0016	0.1595^{**}	0.0087^{**}	0.0177^{**}	0.0177^{**}	0.0148^{**}	0.6869
α_1	0.0391^{***}	0.0336^{***}	0.1055^{***}	0.0369^{***}	0.0857^{***}	0.0909^{***}	0.1176^{***}	0.1561^{***}	0.0428^{***}	0.2106^{***}
α_2	0.9593^{***}	0.9609^{***}	0.8935^{***}	0.9586^{***}	0.8873^{***}	0.9080^{***}	0.8794^{***}	0.8428^{***}	0.9481^{***}	0.7884^{***}
ν	7.1556	9.7599	5.6439	5.1013	29.97	5.6279	9.5503	7.5402	4.0332	3.0516

Table 2.2: ARMA & GARCH pre-filtering parameters for the specified series. *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels.

illustrate that these series are weekly dependent and could be expected to have a small-scaled impact on the parameters estimated with LSE routines from Section 2.4. All data on the currency rates were obtained from the central bank's databases of the corresponding countries, all other stock indices and exchange rates were supplied by Bloomberg.

2.6.2 Estimated parameters

Parameters optimizing loss functions described in Section 2.4 are estimated over two sets of returns illustrated in Figure 2.2. The first set includes raw returns, while the second set encompasses these returns after ARMA (1,1) & GARCH (1,1) pre-filtering similar to Harvey & Oryshchenko (2012). The ARMA & GARCH setting employed can be broadly outlined by

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 \epsilon_{t-1} + \epsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \sigma_{t-1}^2;$$

$$\sqrt{\frac{\nu}{\sigma_t^2(\nu-2)}} \cdot \epsilon_t \sim \text{i.i.d.}\, t_\nu$$

and is performed using `rugarch` package of Ghalanos (2018). Note that the above setting did not converge for RUB/USD series and therefore ARMA (1,0) & GARCH (1,1) under

#	Returns	LSE PDF - filtering	LSE PDF - smoothing	MLE - filtering	MLE - smoothing	LSE CDF - filtering	LSE CDF smoothing
1	BRENT	$h = 0.5418$, $\omega = 0.9868$	$h = 0.3780$, $\omega = 0.9919$	$h = 0.9835$, $\omega = 0.9785$	$h = 0.9675$, $\omega = 0.9620$	$\beta = 0.4261$, $\omega = 0.9841$	$\beta = 0.4289$, $\omega = 0.9788$
2	NOK/USD	$h = 0.2582$, $\omega = 0.9837$	$h = 0.1913$, $\omega = 0.9908$	$h = 0.3528$, $\omega = 0.9847$	$h = 0.2929$, $\omega = 0.9813$	$\beta = 0.1879$, $\omega = 0.9866$	$\beta = 0.1312$, $\omega = 0.9829$
3	RUB/USD	$h = 0.0681$, $\omega = 0.9769$	$h = 0.0655$, $\omega = 0.9873$	$h = 0.3359$, $\omega = 0.9845$	$h = 0.2439$, $\omega = 0.9620$	$\beta = 0.0586$, $\omega = 0.9703$	$\beta = 0.0552$, $\omega = 0.9452$
4	GBP/USD	$h = 0.1435$, $\omega = 0.9855$	$h = 0.1236$, $\omega = 0.9836$	$h = 0.4761$, $\omega = 0.9999$	$h = 0.3811$, $\omega = 0.9769$	$\beta = 0.1249$, $\omega = 0.9896$	$\beta = 0.0695$, $\omega = 0.9882$
5	NASDAQ	$h = 1.0241$, $\omega = 0.9656$	$h = 0.7807$, $\omega = 0.9660$	$h = 1.2796$, $\omega = 0.9690$	$h = 1.0535$, $\omega = 0.9474$	$\beta = 0.8237$, $\omega = 0.9757$	$\beta = 0.5143$, $\omega = 0.9707$
6	S&P 500	$h = 0.3026$, $\omega = 0.9799$	$h = 0.1586$, $\omega = 0.9836$	$h = 0.8356$, $\omega = 0.9565$	$h = 0.3671$, $\omega = 0.9669$	$\beta = 0.3664$, $\omega = 0.9708$	$\beta = 0.1751$, $\omega = 0.9715$
7	FTSE 100	$h = 0.4734$, $\omega = 0.9799$	$h = 0.3168$, $\omega = 0.9752$	$h = 0.7989$, $\omega = 0.9569$	$h = 0.4640$, $\omega = 0.9578$	$\beta = 0.2611$, $\omega = 0.9726$	$\beta = 0.2445$, $\omega = 0.9675$
8	NIKKEI 225	$h = 0.4379$, $\omega = 0.9591$	$h = 0.2654$, $\omega = 0.9594$	$h = 0.9755$, $\omega = 0.9266$	$h = 0.6991$, $\omega = 0.9176$	$\beta = 0.4513$, $\omega = 0.9562$	$\beta = 0.2283$, $\omega = 0.9444$
9	GOLD	$h = 0.2846$, $\omega = 0.9775$	$h = 0.2246$, $\omega = 0.9777$	$h = 0.5688$, $\omega = 0.9783$	$h = 0.5291$, $\omega = 0.9731$	$\beta = 0.2402$, $\omega = 0.9811$	$\beta = 0.1640$, $\omega = 0.9785$
10	BTC/USD	$h = 0.6260$, $\omega = 0.9773$	$h = 0.6118$, $\omega = 0.9634$	$h = 0.8913$, $\omega = 0.9921$	$h = 1.2381$, $\omega = 0.9891$	$\beta = 0.6112$, $\omega = 0.9748$	$\beta = 0.6659$, $\omega = 0.9257$
11	†BRENT	$h = 0.0984$, $\omega = 0.9994$	$h = 0.1372$, $\omega = 0.9986$	$h = 0.3051$, $\omega = 0.9942$	$h = 0.3575$, $\omega = 0.9985$	$\beta = 0.0811$, $\omega = 0.9990$	$\beta = 0.1087$, $\omega = 0.9986$
12	†NOK/USD	$h = 0.2344$, $\omega = 0.9985$	$h = 0.1542$, $\omega = 0.9985$	$h = 0.3857$, $\omega = 1$	$h = 0.3044$, $\omega = 0.9982$	$\beta = 0.1639$, $\omega = 0.9985$	$\beta = 0.1014$, $\omega = 0.9854$
13	†RUB/USD	$h = 0.0870$, $\omega = 0.9969$	$h = 0.0571$, $\omega = 0.9968$	$h = 0.5902$, $\omega = 1$	$h = 0.5407$, $\omega = 0.9962$	$\beta = 0.1894$, $\omega = 0.9934$	$\beta = 0.1012$, $\omega = 0.9834$
14	†GBP/USD	$h = 0.2536$, $\omega = 1$	$h = 0.2347$, $\omega = 0.9969$	$h = 0.6768$, $\omega = 1$	$h = 0.6427$, $\omega = 0.9947$	$\beta = 0.1984$, $\omega = 1$	$\beta = 0.1514$, $\omega = 0.9958$
15	†NASDAQ	$h = 0.3907$, $\omega = 0.9937$	$h = 0.3735$, $\omega = 0.9844$	$h = 0.3389$, $\omega = 0.9968$	$h = 0.2975$, $\omega = 0.9931$	$\beta = 0.2799$, $\omega = 0.9919$	$\beta = 0.1986$, $\omega = 0.9911$
16	†S&P 500	$h = 0.2269$, $\omega = 1$	$h = 0.1783$, $\omega = 0.9963$	$h = 0.5913$, $\omega = 1$	$h = 0.4944$, $\omega = 0.9934$	$\beta = 0.321$, $\omega = 1$	$\beta = 0.1507$, $\omega = 0.9925$
17	†FTSE 100	$h = 0.3459$, $\omega = 1$	$h = 0.2548$, $\omega = 0.9951$	$h = 0.4315$, $\omega = 1$	$h = 0.3135$, $\omega = 0.9943$	$\beta = 0.3277$, $\omega = 1$	$\beta = 0.1511$, $\omega = 0.9949$
18	†NIKKEI 225	$h = 0.3604$, $\omega = 0.9968$	$h = 0.2451$, $\omega = 0.9899$	$h = 0.4992$, $\omega = 1$	$h = 0.3361$, $\omega = 0.9941$	$\beta = 0.3378$, $\omega = 0.997$	$\beta = 0.1509$, $\omega = 0.9939$
19	†GOLD	$h = 0.2$, $\omega = 0.9941$	$h = 0.1758$, $\omega = 0.9935$	$h = 0.3757$, $\omega = 0.9983$	$h = 0.3366$, $\omega = 0.9957$	$\beta = 0.1289$, $\omega = 0.9962$	$\beta = 0.1499$, $\omega = 0.9958$
20	†BTC/USD	$h = 0.1933$, $\omega = 0.9884$	$h = 0.1613$, $\omega = 0.9852$	$h = 0.2808$, $\omega = 0.9969$	$h = 0.3124$, $\omega = 0.9954$	$\beta = 0.1584$, $\omega = 0.9909$	$\beta = 0.1133$, $\omega = 0.9845$

Table 2.3: Filtering and smoothing unknown parameters for the specified returns series estimated with MLE, full-scale LSE for PDF and CDF loss functions; † denotes estimated parameters for ARMA & GARCH filtered data.

#	Returns	binned LSE PDF-filtering	binned LSE PDF-smoothing	binned LSE CDF-filtering	binned LSE CDF-smoothing
1	BRENT	$h = 0.6252$, $\omega = 0.9865$	$h = 0.5693$, $\omega = 0.9913$	$\beta = 0.4246$, $\omega = 0.9874$	$\beta = 0.3534$, $\omega = 0.9883$
2	NOK/USD	$h = 0.2610$, $\omega = 0.9851$	$h = 0.2178$, $\omega = 0.9914$	$\beta = 0.1635$, $\omega = 0.9887$	$\beta = 0.1310$, $\omega = 0.9881$
3	RUB/USD	$h = 0.0694$, $\omega = 0.9833$	$h = 0.0614$, $\omega = 0.9851$	$\beta = 0.0797$, $\omega = 0.9875$	$\beta = 0.0587$, $\omega = 0.9896$
4	GBP/USD	$h = 0.1462$, $\omega = 0.9866$	$h = 0.1331$, $\omega = 0.9894$	$\beta = 0.1135$, $\omega = 0.9925$	$\beta = 0.0819$, $\omega = 0.9908$
5	NASDAQ	$h = 1.0022$, $\omega = 0.9672$	$h = 0.7036$, $\omega = 0.9853$	$\beta = 0.7559$, $\omega = 0.9782$	$\beta = 0.3815$, $\omega = 0.9801$
6	S&P 500	$h = 0.4222$, $\omega = 0.9808$	$h = 0.3304$, $\omega = 0.9868$	$\beta = 0.3664$, $\omega = 0.9725$	$\beta = 0.2191$, $\omega = 0.9796$
7	FTSE 100	$h = 0.5271$, $\omega = 0.9834$	$h = 0.3725$, $\omega = 0.9884$	$\beta = 0.3957$, $\omega = 0.9811$	$\beta = 0.2242$, $\omega = 0.9817$
8	NIKKEI 225	$h = 0.4604$, $\omega = 0.9629$	$h = 0.3229$, $\omega = 0.9821$	$\beta = 0.4081$, $\omega = 0.9625$	$\beta = 0.2024$, $\omega = 0.9749$
9	GOLD	$h = 0.3135$, $\omega = 0.9755$	$h = 0.2739$, $\omega = 0.9857$	$\beta = 0.2158$, $\omega = 0.9826$	$\beta = 0.1674$, $\omega = 0.9845$
10	BTC/USD	$h = 0.8676$, $\omega = 0.9734$	$h = 0.8771$, $\omega = 0.9784$	$\beta = 0.6903$, $\omega = 0.9767$	$\beta = 0.6179$, $\omega = 0.9690$
11	†BRENT	$h = 0.2667$, $\omega = 0.9987$	$h = 0.2563$, $\omega = 0.9990$	$\beta = 0.1578$, $\omega = 0.9988$	$\beta = 0.1557$, $\omega = 0.9985$
12	†NOK/USD	$h = 0.2801$, $\omega = 0.9988$	$h = 0.2791$, $\omega = 0.9988$	$\beta = 0.1660$, $\omega = 0.9991$	$\beta = 0.1115$, $\omega = 0.9983$
13	†RUB/USD	$h = 0.2089$, $\omega = 0.9965$	$h = 0.1974$, $\omega = 0.9956$	$\beta = 0.1873$, $\omega = 0.9965$	$\beta = 0.1129$, $\omega = 0.9938$
14	†GBP/USD	$h = 0.2544$, $\omega = 1$	$h = 0.2478$, $\omega = 0.9974$	$\beta = 0.1863$, $\omega = 1$	$\beta = 0.1563$, $\omega = 0.9959$
15	†NASDAQ	$h = 0.4146$, $\omega = 0.9914$	$h = 0.3501$, $\omega = 0.9956$	$\beta = 0.2599$, $\omega = 0.9925$	$\beta = 0.1969$, $\omega = 0.9924$
16	†S&P 500	$h = 0.2779$, $\omega = 1$	$h = 0.2611$, $\omega = 0.9964$	$\beta = 0.2634$, $\omega = 1$	$\beta = 0.1686$, $\omega = 0.9931$
17	†FTSE 100	$h = 0.3903$, $\omega = 1$	$h = 0.3408$, $\omega = 0.9966$	$\beta = 0.2847$, $\omega = 1$	$\beta = 0.1759$, $\omega = 0.9950$
18	†NIKKEI 225	$h = 0.3901$, $\omega = 0.9987$	$h = 0.3039$, $\omega = 0.9952$	$\beta = 0.3064$, $\omega = 0.9974$	$\beta = 0.1615$, $\omega = 0.9937$
19	†GOLD	$h = 0.2531$, $\omega = 0.9928$	$h = 0.2336$, $\omega = 0.9965$	$\beta = 0.1641$, $\omega = 0.9961$	$\beta = 0.1499$, $\omega = 0.9960$
20	†BTC/USD	$h = 0.2283$, $\omega = 0.9866$	$h = 0.2179$, $\omega = 0.9892$	$\beta = 0.1628$, $\omega = 0.9912$	$\beta = 0.1401$, $\omega = 0.9906$

Table 2.4: Binned filtering and smoothing unknown parameters for the specified returns series estimated with LSE for PDF and CDF loss functions; † denotes estimated parameters for ARMA & GARCH filtered data.

	BRENT	NOK/USD	RUB/USD	GBP/USD	NASDAQ	S&P500	FTSE100	NIKKEI225	GOLD/USD	BTC/USD
ω'	0.9785	0.9847	0.9845	0.9999	0.9690	0.9565	0.9569	0.9266	0.9783	0.9921
β'_1	0.4607	0.1336	0.1357	0.1093	0.4091	0.3790	0.3441	0.3562	0.2491	1.2210
β'_2	0.3224	0.1115	0.0732	0.0771	0.5626	0.3701	0.2965	0.3265	0.1618	0.5093
β'_3	0.2486	0.1067	0.0341	0.0736	0.6965	0.2066	0.2883	0.2779	0.1399	0.3424
β'_4	0.2362	0.1009	0.0283	0.0710	0.4222	0.2480	0.2759	0.3094	0.1467	0.3403
β'_5	0.2989	0.1131	0.0751	0.0788	0.7339	0.2225	0.2253	0.2424	0.2048	0.6459
β'_6	0.4062	0.1517	0.1524	0.1144	0.5411	0.3054	0.2867	0.3824	0.2102	1.1843
ω'	0.9865	0.9851	0.9833	0.9866	0.9672	0.9808	0.9834	0.9629	0.9755	0.9734
β'_1	0.4541	0.1358	0.1272	0.1093	0.4419	0.3579	0.3288	0.3457	0.2634	1.2210
β'_2	0.3105	0.1142	0.0675	0.0771	0.6118	0.3533	0.2856	0.3213	0.1807	0.5093
β'_3	0.2373	0.1104	0.0334	0.0735	0.7190	0.2127	0.2905	0.2713	0.1520	0.3424
β'_4	0.2266	0.1043	0.0271	0.0710	0.4326	0.2401	0.2691	0.3181	0.1599	0.3403
β'_5	0.2884	0.1160	0.0705	0.0788	0.7511	0.2113	0.2233	0.2369	0.2165	0.6459
β'_6	0.4002	0.1538	0.1431	0.1144	0.5535	0.2881	0.2667	0.3714	0.2191	1.1843
ω'	0.9874	0.9887	0.9875	0.9925	0.9782	0.9725	0.9811	0.9625	0.9826	0.9767
β'_1	0.4491	0.1313	0.1265	0.1090	0.3988	0.3741	0.3290	0.3422	0.2446	1.2209
β'_2	0.3053	0.1070	0.0680	0.0771	0.5478	0.3623	0.2748	0.3199	0.1564	0.5093
β'_3	0.2334	0.1013	0.0339	0.0738	0.6850	0.2041	0.2971	0.2404	0.1363	0.3423
β'_4	0.2231	0.0957	0.0280	0.0712	0.4208	0.2446	0.2856	0.3223	0.1427	0.3402
β'_5	0.2847	0.1093	0.0727	0.0787	0.7211	0.2186	0.2197	0.2357	0.2014	0.6458
β'_6	0.3947	0.1491	0.1424	0.1171	0.5381	0.3029	0.2707	0.3697	0.2082	1.1842
ω'	0.9841	0.9866	0.9703	0.9896	0.9757	0.9708	0.9726	0.9562	0.9811	0.9748
β'_1	0.4607	0.1336	0.1357	0.1093	0.4091	0.3790	0.3441	0.3562	0.2491	1.2210
β'_2	0.3224	0.1115	0.0732	0.0771	0.5626	0.3701	0.2965	0.3265	0.1618	0.5093
β'_3	0.2486	0.1067	0.0341	0.0736	0.6965	0.2066	0.2883	0.2779	0.1399	0.3424
β'_4	0.2362	0.1009	0.0283	0.0710	0.4222	0.2480	0.2759	0.3094	0.1467	0.3403
β'_5	0.2989	0.1131	0.0751	0.0788	0.7339	0.2225	0.2253	0.2424	0.2048	0.6459
β'_6	0.4062	0.1517	0.1524	0.1144	0.5411	0.3054	0.2867	0.3824	0.2102	1.1843
$\omega^{\backprime}$	0.9883	0.9881	0.9896	0.9908	0.9801	0.9796	0.9817	0.9749	0.9835	0.9690
$\beta^{\backprime}_1$	0.4105	0.1232	0.1196	0.1041	0.3877	0.3081	0.2712	0.2679	0.2329	1.0926
$\beta^{\backprime}_2$	0.2631	0.0901	0.0615	0.0601	0.3894	0.1527	0.1656	0.1371	0.1271	0.4101
$\beta^{\backprime}_3$	0.2072	0.0771	0.0304	0.0504	0.3147	0.0981	0.1325	0.0938	0.0998	0.2490
$\beta^{\backprime}_4$	0.2003	0.0776	0.0234	0.0487	0.2653	0.0816	0.1339	0.1099	0.0951	0.2678
$\beta^{\backprime}_5$	0.2576	0.0965	0.0593	0.0639	0.2931	0.1338	0.1379	0.1122	0.1443	0.5507
$\beta^{\backprime}_6$	0.3701	0.1402	0.1341	0.1042	0.3845	0.2604	0.2401	0.2538	0.1709	1.1169
$\omega^{\backprime}$	0.9788	0.9829	0.9452	0.9882	0.9707	0.9715	0.9675	0.9444	0.9785	0.9257
$\beta^{\backprime}_1$	0.4327	0.1271	0.1330	0.1061	0.4165	0.3254	0.2929	0.3035	0.2430	1.2731
$\beta^{\backprime}_2$	0.2907	0.0960	0.0726	0.0625	0.4448	0.1671	0.1928	0.1798	0.1394	0.5657
$\beta^{\backprime}_3$	0.2316	0.0844	0.0334	0.0525	0.3798	0.1103	0.1590	0.1301	0.1110	0.3450
$\beta^{\backprime}_4$	0.2222	0.0851	0.0269	0.0513	0.3181	0.0920	0.1629	0.1512	0.1048	0.3567
$\beta^{\backprime}_5$	0.2833	0.1035	0.0709	0.0666	0.3386	0.1486	0.1594	0.1399	0.1537	0.7136
$\beta^{\backprime}_6$	0.3875	0.1445	0.1486	0.1064	0.4135	0.2742	0.2605	0.2996	0.1778	1.2585

Table 2.5: Adaptive bandwidths for the specified returns series, where $'$ denotes parameters for filtering and $^{\backprime}$ parameters for smoothing. Note that the bandwidths are obtained for the weights precomputed for ω, from top to bottom: by MLE, binned LSE PDF, binned LSE CDF, LSE CDF for filtering and binned LSE CDF and LSE CDF for smoothing.

symmetric Student's t distribution filter was used for this sample. The latter specification is identical to what is used in Harvey & Oryshchenko (2012). Estimated parameters used for pre-filtering are provided in Table 2.2. It is worthwhile highlighting that Harvey & Oryshchenko (2012) select their ARMA & GARCH specification under the formal likelihood based evaluation criteria framework similar to the one used in Chapter 1, but here the choice of the ARMA & GARCH specification is ad hoc. This has straightforward and applied motivation. First, considered dynamic estimators are foremost hypothesised to achieve accurate in-sample PITs forecasts without pre-filtering at least in some of the raw returns series. Second, if they fail achieving that, they are expected to pick up any remaining time-variation after rough pre-filtering of the series to empirically support their potential for further financial applications.

Estimated parameters for raw and filtered returns with different estimation methods are reported in Tables 2.3, 2.4 and 2.5. All binning procedures used to obtain adaptive and fixed estimators' parameters are based on the sample quantiles as is instructed and commonly used for spline fitting in Harrell (2010) and also applied in the kernel density estimation context by Hazelton (2003).

Following simple linear binning as in Hall & Wand (1996), for the dynamic binned fixed bandwidth estimators 15 "static" bins with 15 bin centres for each sample were employed. For the sample quantile $\xi_s(\tau)$, these bins and bin centres are outlined by the intervals in $[\mathbf{min}(x_\iota) - 0.01; \xi_s(0.015, 0.03, 0.05, 0.15, \cdots, 0.85, 0.95, 0.97, 0.985); \mathbf{max}(x_\iota) + 0.01]$. For adaptive estimations, the same linear binning was used, but samples were binned over the intervals in $[\mathbf{min}(x_\iota) - 0.01; \xi_s(0.1, 0.23, 0.37, 0.5, 0.67, 0.76, 0.9); \mathbf{max}(x_\iota) + 0.01]$ and bandwidths were computed only for the bins in $[\xi_s(0.1, 0.23, 0.37, 0.5, 0.67, 0.76, 0.9)]$. This suggests 6 bandwidths for each sample with bandwidths for $[\mathbf{min}(x_\iota) - 0.01; \xi_s(0.1))$ and $(\xi_s(0.9); \mathbf{max}(x_\iota) + 0.01]$ being the same as the bandwidths for $[\xi_s(0.1), \xi_s(0.23))$ and $(\xi_s(0.76), \xi_s(0.9)]$ respectively. All adaptive bandwidths are obtained with LSE loss functions optimal for CDF, where weights are precomputed using ω parameters for dynamic binned and full-scale estimators in LSE optimizations as well as those ω parameters obtained with MLE. Since the weighting scheme remains unchanged in all estimations, pre-computing weights is justified by more time-efficient applied adaptive experiments. Such estimation still keeps its distinctive feature of varying the bandwidth parameter over the range of returns and also reserves time for experimenting with different values of ω. This also points out that the main gains, when evaluating PITs for adaptive estimations, could be expected in their uniformity characteristics outlined by the K-S and CvM tests output. On the other hand, the fixed bandwidth dynamic binned estimators may be expected not only to provide parameters similar to those delivered by the full-scale estimators under the LSE routine, and therefore match the diagnostic evaluations for the full-scale estimators in the LSE experiments, but also reduce their computational burden.

First, reviewing Table 2.3, it may be observed that optimal bandwidth parameters offered for the LSE for PDF, MLE and LSE for CDF are notably different in value for every returns series without pre-filtering, while exponential "learning rates" offered by these methods are typically similar in value. LSE for CDF usually offers the smallest value of the bandwidth parameter, then follows its PDF counterpart, while MLE is observed suggesting the highest bandwidths. After pre-filtering the pattern of the parameter values is similar, but can be argued as less explicit. MLE is still observed mainly offering the highest bandwidths and LSE for CDF offering the smallest values of this parameter after pre-filtering. Now, from

Table 2.4 it may be noted that the binned estimators deliver parameters which are arguably comparable in value to the parameters reported for the full-scale estimators they simplify in the LSE routines, though binned LSE for CDF parameters are typically closer to what is reported for its full-scale alternative. Differences in the parameters may be commanded by the "one binning fits all" computational strategy and therefore, it may also be hypothesised that individual and perhaprs more intensive approach to binning of each sample may lead to the parameters closer to the full-scale estimations. It is, however, worthwhile reporting that function evaluation time for the LSE for PDF under the chosen binning rule and the number of bins is comparable and is very close (hardly differentiable) to the function evaluation time for MLE routine, if both estimations are scripted and run in R. Given what is observed for the obtained parameters, this can be considered as a satisfactory result for speeding up LSE optimization for PDF. For LSE optimization for CDF, the necessity of integrating the loss function over the grid of estimation points leads to the "a priori" longer evaluations, though simple integration algorithms as "Simpson's" procedure within pracma package of Borchers (2018) seem to help speeding up estimations for CDF optimal parameters. Also, it is important noting that, though under quite basic non-evolving binning, dynamic binned estimators seem to be able of picking up the exponential learning rate parameter relatively well. However, a fair evaluation of the parameters by the binned estimators shall be conducted in the next section under the selected K-S, CvM and Berkowitz's (2001) criteria. Noting that LSE for CDF has a tendency of being less sensitive to the number of bins used for evaluations, dynamic adaptive estimations were restricted to the experiments with CDF optimal computaions for more stable evaluations of the bandwidths. The bandwidths strategy for the tails is experimental and motivated by faster evaluations, since it is typically expected (e.g. see Sain & Scott (1996)) that a good approximation of the bandwidths for the tails may require more extensive and carefull binning, resulting in a notably higher number of parameters, which may be less rationalized for one dimensional data estimations. Moreover, under the simple non-evolving binning as here, extensive binning in the tails may imply empty bins at some initial iterations, which are typically avoided by Sain & Scott (1996) and Hazelton (2003) in the previous applied examples. LSE estimations have a history of experiments when observations are omitted in optimizations (e.g. see Feluch & Koronacki (1992); Stute (1992)), while here however, all observations are used, but not all bandwidths are computed and employed. After all, computed bandwidths are body domains oriented and such bandwidths experiment should be very attractive for future dynamic semiparametric models, where bandwidths may be preferred being restricted from approximating the tails defined by the weighted Hill's (1975) tail index estimator. That formally is: if z_t represents $x_{1,t} \leq, \cdots, \leq x_{i,t}$ for $i = 1, \cdots, t$ and $t = 1, \cdots, T$ as previously defined in this chapter, then rewriting Hill's (1975) estimator, defined by

$$\widehat{\Theta}_{\mathbf{Hill}} = \frac{\kappa}{\sum_{j=1}^{\kappa} \log z_{t-\kappa+j} - \kappa \log z_{t-\kappa}},$$

where κ outlines the total number of observations dedicated to the tail approximation, for the exponentially weighted with (2.5) ordered observations provides

$$\widehat{\Theta}_{\mathbf{Hill}}^{\omega} = \frac{\sum_{j=1}^{\kappa} w_{t-\kappa+j}(\omega)}{\sum_{j=1}^{\kappa} w_{t-\kappa+j}(\omega)\left[\log z_{t-\kappa+j} - \log z_{t-\kappa}\right]}$$

for approximation of the simple parametric Pareto type tail described in Kilber & Kotz

(2003) among others. Note that the above outlines the upper/right Pareto tail and for the lower/left Pareto tail it is necessary considering returns multiplied by minus one, though there are investigations where absolute values of returns are considered for regular Hill's (1975) tail estimations (e.g. Dupuis et al. (2014)). It must also be highlighted that niether applications of the weighted Hill's (1975) estimator to financial data, nor its combination with the exponential weights have been defined on the date of producing this work and though, quantiles are common thresholds for the tail initializtion points (e.g. Gencay & Selcuk (2004); Chavez-Demoulin et al. (2014)), a formal discussion outlining that the obtained time-varying PDFs always integrate to one and thus, are always true densities (e.g. see Markovich (2008)) as well as considerations of the known sensitivities of the regular tail estimator and any potential negative estimation consequences due to that as discussed in Hill (2010), Embrechts et al. (2013) and Hill (2015) are taken into account before applications. Extreme value theory and semiparametric methods as in MacDonald et al. (2011), but for the dynamic, exponentially weights setting as here, are beyond the scope of this work and are left out for future investigations.

Overall, bandwidths reported for the adaptive strategy match the expected pattern for the heavy-tailed financial data as described by Markovich (2008) and Scott (2015). From Table 2.5 it can be observed that typically higher bandwidths are allocated closer to the tails, where observations are sparse, and smaller bandwidths are allocated closer to the center, where observations are abundant, with the notable exception for the bandwidth allocations for exponential filtering of the considered NASDAQ series. This can be articulated by the non-evolving bins used for estimations, which may not always reflect the time-variation in data effectively. Time-variation in such bins is strictly driven by the chosen exponential learning rate parameter, while bin edges and their centres reflect the entire sample at hand and may not always approximate/represent data falling into the bins well at some iterations. In the fixed bandwidth estimations, such binning imperfections are averaged out into a single bandwidth parameter, while in the Cross-Validation for the exponential smoothing, there are less restrictions and simple non-evolving bins reflect the data, note that also time-conditional and exponentially weighted, at all iterations proportionately, yielding expected pattern for the parameters in Table 2.5. Therefore, it may be also worthwhile employing exponential smoothing parameters in the exponential filtering applications, since Cross-Validation is a generally accepted computational compromise in the time series nonparametric applications (e.g. Harvey & Oryshchenko (2012); Taylor & Jeon (2015)). Most importantly the loss functions of this type for the LSE routine are available, provided in Section 2.4.3, seem delivering expected vectors of the bandwidths and are notably different from what is provided by MLE. Though, all further evaluations are restricted to the exponential filtering parameters following Harvey & Oryshchenko (2012), all adaptive parameters are still valuable and informative, allowing thorough investigation of the LSE bandwidth allocation procedure, confirming its satisfactory modest reaction to outliers as well as pointing out that for NASDAQ series the bandwidth variation may not be even necessary, as there are no notable changes in the per bin optimal bandwidths identified.

2.7 Diagnostics output and discussion

All in-sample forecast evaluations are provided in Tables 2.6, 2.7 and 2.8. First, from the diagnostic output provided in Table 2.6 it is explicit that LSE for CDF provides overall the

#	Returns	LSE PDF						MLE						LSE 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	0.1858	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 2.6: Diagnostic output for parameters reported in Table 2.3. $K-S_d$, $K-S_d$, CvM_d, CvM_p, LR$_d$ and LR$_p$ denote Kolmogorov-Smirnov, Cramer-von Mises and Berkowitz tests' distances and corresponding p-values, while tests' statistic exceeding 5% significance threshold is highlighted in grey respectively.

		binned LSE PDF						binned LSE 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 2.7: Diagnostic output for parameters reported in Table 2.4. $K-S_d$, $K-S_d$, CvM_d, CvM_p, LR_d and LR_p denote Kolmogorov-Smirnov, Cramer-von Mises and Berkowitz tests' distances and corresponding p-values, while tests' statistic exceeding 5% significance threshold is highlighted in grey respectively.

		ω as per MLE						ω as per LSE 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	0.0128	0.5438	0.0792	0.6964	10.9686	0.0119	0.0112	0.7093	0.0854	0.6614	4.3208	0.2288
2	NOK/USD	0.0079	0.9693	0.0248	0.9900	9.9893	0.0187	0.0074	0.9833	0.0243	0.9911	7.7350	0.0518
3	RUB/USD	0.0119	0.6546	0.0690	0.7578	24.1977	0.0000	0.0107	0.7748	0.0600	0.8140	31.3471	0.0000
4	GBP/USD	0.0470	0.0189	0.5456	0.0305	28.3862	0.0000	0.0159	0.9531	0.0329	0.9658	6.3839	0.0944
5	NASDAQ	0.0271	0.7340	0.0775	0.7069	10.7047	0.0134	0.0256	0.7961	0.0772	0.7082	8.3789	0.0388
6	S&P 500	0.0237	0.7611	0.0701	0.7513	34.2816	0.0000	0.0228	0.8016	0.0571	0.8323	30.3964	0.0000
7	FTSE 100	0.0261	0.6455	0.0691	0.7572	31.7502	0.0000	0.0225	0.8113	0.0539	0.8525	22.8836	0.0000
8	NIKKEI 225	0.0300	0.6459	0.0892	0.6407	39.7170	0.0000	0.0314	0.5859	0.0779	0.7043	26.2552	0.0000
9	GOLD/USD	0.0129	0.9939	0.0171	0.9990	7.9772	0.0465	0.0112	0.9993	0.0149	0.9996	6.2971	0.0980
10	BTC/USD	0.0361	0.1280	0.1829	0.3032	1.2473	0.7417	0.0198	0.8061	0.0948	0.6108	5.8338	0.1200

		ω as per binned LSE PDF						ω as per binned LSE 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	0.0121	0.6104	0.0880	0.6471	2.7359	0.4342	0.0122	0.6018	0.0893	0.6397	2.2836	0.5157
2	NOK/USD	0.0073	0.9864	0.0245	0.9906	9.0776	0.0283	0.0084	0.9470	0.0248	0.9899	5.9622	0.1135
3	RUB/USD	0.0110	0.7392	0.0575	0.8300	26.4875	0.0000	0.0124	0.5944	0.0735	0.7304	27.1667	0.0000
4	GBP/USD	0.0145	0.9793	0.0272	0.9845	7.2138	0.0654	0.0197	0.8086	0.0474	0.8921	6.2671	0.0993
5	NASDAQ	0.0291	0.6487	0.0808	0.6872	10.8568	0.0125	0.0264	0.7633	0.0758	0.7166	7.8522	0.0492
6	S&P 500	0.0183	0.9518	0.0504	0.8738	31.6466	0.0000	0.0205	0.8930	0.0554	0.8431	30.5157	0.0000
7	FTSE 100	0.0176	0.9655	0.0452	0.9051	21.0887	0.0001	0.0174	0.9686	0.0442	0.9105	21.0832	0.0001
8	NIKKEI 225	0.0306	0.6207	0.0768	0.7106	24.0430	0.0000	0.0271	0.7620	0.0716	0.7417	24.8138	0.0000
9	GOLD/USD	0.0129	0.9937	0.0164	0.9992	8.8647	0.0311	0.0112	0.9992	0.0152	0.9996	5.7021	0.1270
10	BTC/USD	0.0199	0.7976	0.0926	0.6225	6.6021	0.0857	0.0207	0.7606	0.0983	0.5936	4.8930	0.1798

Table 2.8: Diagnostic output for parameters reported in Table 2.5. $K-S_d$, $K-S_d$, CvM_d, CvM_p, LR_d and LR_p denote Kolmogorov-Smirnov, Cramer-von Mises and Berkowitz tests' distances and corresponding p-values, while tests' statistic exceeding 5% significance threshold is highlighted in grey respectively.

best choice of parameters for estimations in the considered samples without pre-filtering. PITs for all of these samples are uniform as reported by the K-S and CvM outputs, but are likely not always independent as reported by Berkowitz's (2001) compound evaluation criteria. Such output may be an indication that LSE for CDF suggests a good choice of the bandwidth parameter, but weighting under exponential filtering may not always be sufficient for the considered financial returns or simply the choice of the exponential learning rate is not always ideal. Without pre-filtering, LSE for PDF also performs well, but it fails insuring uniformity of the PITs for two samples less than the LSE for CDF. Accurate evaluations ratio under Berkowitz's (2001) test is the same for both LSE for PDF and CDF choice of parameters. Unfortunately, MLE, if compared to its direct PDF optimal LSE competitor, delivers quite unsatisfactory results. Uniform PITs under both K-S and CvM tests are obtained only for one sample, however a strong outlook under Berkowitz's (2001) test for BTC/USD suggests that the compound ratio of correct evaluations may be admitted at the two out of ten level. Poor PITs uniformity characteristics points out the poor choice of the bandwidth by the MLE routine.

Moving to the pre-filtered returns both LSE based approaches demonstrate a very strong diagnostic output, picking up both uniformity and independence properties of the PITs. After employed parametric pre-filtering, it may be argued that there is a little variation left in the series, as can be noted from the exponential learning rate parameters in Table 2.3 for example, and therefore, the choice of the bandwidth becomes more important than the choice of the learning rate for accurate evaluations. As it is observed from the results without pre-filtering, LSE based estimations do not exhibit difficulties choosing the appropriate bandwidth and after some additional help from the corrections for scale and location they seem accomplishing this task comfortably. MLE, however, still struggles providing a good evaluation outlook. Strong PITs uniformity characteristics are provided only for two series now, while Berkowitz's (2001) criteria suggest only four correct evaluations. The output is still notably inferior to what is provided for all LSE parameters discussed so far even without pre-filtering. It may be advocated that with MLE, an appropriate and thorough choice of the filtering specification, insuring absence of outliers in innovations, should be adopted to achieve accurate evaluations, however from the applied viewpoint, this may not always be very well justified given the results reported in Table 2.7 for binned computational simplifications using LSE.

Comparing diagnostic output for the dynamic binned estimators, evaluations by the CDF optimal parameters again provide the most appealing outlook. Without pre-filtering number of correct evaluations for Berkowitz's (2001) test is the same as for the full-scale CDF optimal optimizations, but PITs for RUB/USD are no longer uniform. On the other hand, for the binned LSE for PDF estimations, PITs with confirmed uniformity characteristics match its full-scale estimations alternative, but under Berkowitz's (2001) evaluation NOK/USD forecast is no longer accurate. After ARMA & GARCH filter, LSE binned outlook is overall strong and very similar to the full-scale PDF and CDF estimations with an exception for the BRENT series, where uniformity of the PITs is not confirmed by the CvM test for binned LSE for PDF. Anyway, given quite experimental and simplified approach to binning, very motivated by achieving computational time comparable to the MLE routine, gains for the binned LSE for PDF estimations over MLE are notable and easy to argue with the achieved applied results, though it is hard to deny the role of binning and higher sensitivity of the PDF optimal estimations to this procedure.

If binned estimators are employed to vary the bandwidth parameter for adaptive estimations, typically a good outlook in the PITs uniformity characteristics may be achieved as shown for computed distances for K-S and CvM tests in Table 2.5. However, losses in the Berkowitz's (2001) distances are also notable, suggesting an overall mixed performance of the adopted variable bandwidth adaptive strategy, especially considering additional effort it requires for estimations. From the conducted experiments for the different values of the exponential learning rates, it can be concluded that the adaptive estimations as here, are more sensitive to abrupt changes in the series and to the good choice of the parameter governing the weights' dynamics. It is, however, worth highlighting that the MLE based ω yields the worst results for PITs independence, diminishing value of this estimation procedure further. LSE seems to yield improvements upon this routine, providing a better combination of the bandwidths than with MLE previously, but since the weights are dictated by the MLE optimal parameters, the notable losses in the PITs independence point out additional problems when choosing parameters with MLE. On the other hand, reviewing output for weights as per LSE for CDF ω parameter, it can be noted that the NASDAQ evaluation does not path independence assessment, unlike for the simple binned and full-scale LSE previous evaluations. This outcome is well matched by the preliminary analysis of the parameters conducted for this sample in the preceding section. Time evolving bins, more careful and sample oriented binning, or simply parameters optimal for the CV procedure may provide uniform and independent evaluations for this sample.

2.8 Concluding remarks

This chapter empirically demonstrated that exponentially weighted kernel estimators as in Harvey & Oryshchenko (2012) are valid for forecasting the densities and distributions of financial returns. Previously, to achieve accurate and appropriate in-sample forecast evaluations with these estimators, Harvey & Oryshchenko (2012) had to correct for scale and location applying an ARMA & GARCH filter. Here, it is found that such pre-filtering may not be even necessary, if parameters are chosen by credible techniques for unknown parameters estimation. In this chapter, it is achieved with LSE functions modified to incorporate exponentially declining weights enhanced kernel estimators. Moreover, from the small conducted experiment, it can also be reported that MLE may remain unable providing an appropriate combination of parameters for dynamic kernel estimations after pre-filtering. If estimation samples are large and computing power is limited, obtained results suggest that reducing computational burden with dynamic binned estimators and employing accordingly modified LSE loss functions can be among valid estimation strategies. Empirically, these estimators appear to have an ability of picking up both the bandwidth and parameter governing the dynamics of exponential weights, though binning strategies may be additional and important component of the accurate forecasting. Here, perhaps the most elementary approach to the time series binning has been adopted to deliver relatively satisfactory to and more time efficient results than the full-scale LSE estimations. Future research may adopt less empirically oriented agenda on the binning type estimators discussed in this chapter. For example, it may be interesting investigating in more detail how number of bins employed in estimations impacts "time efficiency - estimation accuracy" trade-off in the time series conditions as here. Since there can be more problematic financial series than those considered in this chapter, it may also be worthwhile considering time evolving bins and estimation benefits from their characteristics. On the other hand, exploiting the cross-validation approaches with the two-sided exponential weights, as provided in this chapter for dynamic binned estimators, could also be considered a valid pragmatic suggestion. Most importantly, binned estimators based evaluations remain time efficient and parsimonious for applied use.

Binned estimators may also serve a foundation for dynamic adaptive estimations. In this chapter, the ad hoc approach to such bandwidth variation strategies has been adopted. It demonstrated to provide improvements in the PITs uniformity characteristics, but gains in the dependency component of the evaluation criteria for these pseudo-observations appeared to be smaller. Exponential weights enhaced estimators as in Harvey & Oryshchenko (2012) seem to fit slowly evolving series well, but adaptive strategies as here, may be more sensitive to volatility bursts in financial data. Therefore, it may be worthwhile exploring other weighting schemes to accurately forecast faster evolving series or alternative strategies for bandwidths variations.

This chapter does not go beyond DGP assumptions free methods and thus, approaches with a reference to some parametric distributions as in Sheather & Jones (1991), Altman & Leger (1995) or Polansky & Baker (2000) may also be worth modifying to accomodate exponentially weighted kernel estimators. This may lead to more time-efficient estimations and more inclusive reviews of the approaches to parameters estimations for Harvey & Oryshchenko's (2012) estimators than here. Overall, obtained results demonstrate empirical evidence for applications of these estimators within semiparametric or nonparameteric multivariate copula frameworks. On the other hand, for univariate Valua-at-Risk applied experiments, it may be poorly justified comparing these estimators with the full-scale GAS models of Creal et al. (2013) and Harvey (2013), but it is interesting considering kernel methods for exponential filtering alongside similar methods under the GAS framework of Lucas & Zhang (2016). This is addressed in the next chapter of this thesis.

Chapter 3

A RiskMetrics competition for the dynamic kernel density estimator

3.1 Introduction and motivation

J.P. Morgan's (1996) RiskMetrics™ is the most recognized and basic benchmark model in financial research for Value-at-Risk (VAR) estimation (e.g. Wilhelmsson (2009); Boucher et al. (2014); Danielsson et al. (2016); Nieto & Ruiz (2016)). Its original form of the exponentially weighted (EWMA) conditional variance for Gaussian distributed financial returns has been criticised (e.g. Guermat & Harris (2002)), but its intrinsic simplicity and pragmatism still attract practitioners (e.g. Zumbach (2007)) as well as academics (e.g. Gerlach et al. (2013); Lucas & Zhang (2016)) to introduce necessary upgrades and keep it valid for applied financial practices such as VAR estimations (e.g. Pafka & Kondor (2001); Taylor (2007); McMillan & Kambouroudis (2009); Boucher et al. (2014)). Therefore, present chapter aims to review some of the up-to-date EWMA VAR methods, which fall into the same category as the original RiskMetrics™ to question whether this weighting scheme can be still valid for risk measurement under the basic VAR testing statistic and, perhaps similar to the work of P. R. Hansen & Lunde (2005) for GARCH (1,1) model, also questions whether parsimonious variations of RiskMetrics, including its nonparametric versions as in Chapter 2 for example, can outperform or be equally effective as more technically involved methods. The choice for the most advanced EWMA VAR benchmark scheme here falls on the Student's t Generalized Autoregressive Score (GAS) version of Lucas & Zhang (2016). This belongs to the group of approaches which have been gaining popularity within the GAS time-varying framework of Creal et al. (2013) and Harvey (2013). The approach of Lucas & Zhang (2016) allows exponential decay type dynamics in the tails of returns distribution, offers leptokurtic properties, falling in line with up-to-date regulatory demands (see Rossignolo et al. (2012, 2013); Danielsson et al. (2016) for an interesting discussion on the models allowing heavy tails, Basel accords' capital requirements and VAR), and can incorporate dynamic skewness parametrisation for assymetry modelling if necessesary. The latter feature may be less justified practically, given its restriction to the exponential weighting and overall rising level of complexity, which is higher than or comparable to the full-scale and more flexible GAS models (see Ardia, Boudt, & Catania (2018) and the closed form expressions of some full-scale GAS models therein as well as Ardia et al.'s (2016) notes on the practical pitfalls for using assymetric GAS parametrizations for VAR modelling). Roughly, this chapter supports the view that RiskMetrics type VAR models must be parsimonious and easy to implement, since the more technically involved this scheme becomes, the less reasonable it is to restrict one's portfolio of methods to the exponential decay weighting given the wide availability of other effective methods (e.g. see Laporta et al. (2018) for a recent full-scale GAS based

VAR performance competition under the less restricted portfolio of methods or Nieto & Ruiz (2016) for a comprehensive review of VAR methods in general). RiskMetrics type models may be very useful to compare new or existing advanced methods to, but they may not always be relied upon for risk quantification. Therefore, this chapter also hopes that obtained empirical output may be helpful to identify a good benchmark model in the form a very simple and widely familiar weighting scheme for future studies.

Further in this chapter t-GAS RiskMetrics is compared to its parsimonious, but heavy tails "friendly", "robust" Laplace alternative introduced in Guermat & Harris (2002) under the GAS (L-GAS) parametrization of Lucas & Zhang (2016), to the kernel (kCDF) nonparamtric EWMA VAR presented in Harvey & Oryshchenko (2012), Harvey (2013) and throughtly discussed in Chapter 2 as well as to, perhaps the most parsimonious, nonparametric alternative in the form of the dynamic empirical Cumulative Distribution Function (eCDF). Nonparametric RiskMetrics variations are naturally free from distributional assumptions, account for asymmetries and have arguably simpler closed functional forms than the t-GAS RiskMetrics of Lucas & Zhang (2016), however they may struggle the most when approximating high quantiles for VAR evaluations (e.g. Jones & Signorini (1997); Markovich (2008)).

Originall RiskMetrics™, t-GAS and L-GAS RiskMetrics as well as two of their nonparametric alternatives are set to compete over the daily Brazil, Russia, India, China, Turkey and South Africa (BRICTS) - USD exchange rates returns and the returns data set used in Chapter 2 under the commonly used VAR estimation criteria of Kupiec (1995) and Christoffersen (1998) over the VAR confidence levels as in applied work of Cheng & Hung (2011) among others. There are several interesting findings from the conducted analysis: first, for the 99% VAR confidence level, it is identified that for the BRICTS in-sample and out-of-sample settings

L-GAS RiskMetrics provides the best VAR projections, followed by the estimations for kCDF, t-GAS and eCDF variations respectively. Though, the t-GAS RiskMetrics performs worse than some of its parsimonious alternatives in the BRICTS settings, it provides the strongest outlook for the slightly more expansive data set from Chapter 2. For this data set, the second best VAR estimations are again provided by the kCDF RiskMetrics of Harvey & Oryshchenko (2012) with eCDF alongside L-GAS demonstrating a similar VAR accuracy ratio and sharing the third best performing approach position. RiskMetrics™ mostly performs the worst in the considered data and testing settings. Secondly, obtained findings can be considered valuable for the nonparametric class of estimators used in the RiskMetrics competition. The kCDF and eCDF EWMA VAR type estimators have been succesfully applied in electricity/power generation time series modelling (e.g. Taylor & Jeon (2015); Arora & Taylor (2016)), but have yet to be acknowledged by finance practitioners for financial returns applications. From the obtained results kCDF RiskMetrics appears to be very comfortable at the 95% and lower VAR confidence level and its performance is slightly inferior to the parametric heavy tail RiskMetrics specifications at the 99% risk level. The method has a good potential as a benchmark for other VAR competitions due to its relative simplicity and flexibility, properties which are endemic to the nonparametric estimators in general.

Obtained results also suggest that more complex methods do not necessesarily "a priori" imply better estimation outcomes in practice and such results are often reflected in the pragmatism of the most finance practicioners according to Pérignon & Smith (2010). It is important to highlight that the test statistic employed here focuses on the frequency of the losses and not on their magnitude or associated capital opportunity costs. Different data, backtesting or methods may provide a different as well as better outlook and perspectives on the VAR estimations than the results for RiskMetrics variations here. Moreover, obtained good L-GAS results for the BRICTS settings may very well be an example of the "conservative strategy compensating imperfections in the forecasting scheme" case. This, on the other hand, does not necessesary imply that such strategies are not valid for practical use. For example, Pérignon et al. (2008) point out that the six largest commeracial banks in Canada prefer to overestimate their exposure to avoid additional financial penalties (see McAleer (2009); McAleer et al. (2013) for examples of the number of violations and capital penalties under Basel II and III standards). Therefore, mostly aiming to provide an interesting discussion within the methods of RiskMetrics™ relevance, this chapter is organized as follows: Section 3.2 introduces employed RiksMetrics variations, Section 3.3 provides VAR evaluation criteria, Section 3.4 describes the data, Section 3.5 reports some of the estimated parameters for VAR evaluations, Section 3.6 discusses the results of the VAR projections and Section 3.7 summarizes the discussion.

3.2 RiskMetrics, GAS and nonparametric approaches

For the PDF in

$$f(x_t \mid \mathbf{F}_{t-1}; \mathbf{f}_t, \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{x_t^2}{2\sigma_t^2}}, \tag{3.1}$$

where $\mathbf{F}_{t-1}$ is the information set available at time $t-1$, $\mathbf{f}_t$ and θ are vectors of time-varying and static parameters respectively; setting $\mathbf{f}_t = \sigma_t^2$ produces J.P. Morgan's (1996) RiskMetrics™ which parametrizes volatility as the weighted sum of the past squared observations given by the following recursive form

$$\sigma_{t+1}^2 = \omega \cdot \sigma_t^2 + (1-\omega) \cdot x_t^2, \;\; 0 < \omega < 1; \tag{3.2}$$

equivalently expressed through appropriate steps of the recursive substitution similar to estimators as in Chapter 2 with

$$\sigma_{t+1}^2 = (1-\omega) \sum_{i=1}^{t} \omega^i x_{t-i}^2 \tag{3.3}$$

or by

$$\sigma_{t+1}^2 = \frac{1-\omega}{1-\omega^t} \sum_{i=1}^{t} \omega^i x_{t-i}^2 \tag{3.4}$$

which insures that weights always sum to 1 over $i = 1, \cdots, t$ and is a zero intercept special case of Bollerslev's (1986) Integrated GARCH (1,1) (IGARCH) model. The more general form of the IGARCH model for conditional volatility under DGP assumption as in (3.1) is

$$\sigma_{t+1}^2 = c + A \cdot x_t^2 + B \cdot \sigma_t^2 = c + A \cdot (x_t^2 - \sigma_t^2) + (A+B) \cdot \sigma_t^2 \tag{3.5}$$

and the special case outlined in (3.2) occurs when $c = 0$, $B = \omega$ and $A = 1 - B$ with $c + A + B = 1$ (see Nelson (1990) and Bollerslev et al. (1994) for specific conditions and details). Also note that error correction form similar to kernel estimators described in Chapter 2 is outlined on the right-hand side of the IGARCH volatility parametrization in (3.5).

For the Gaussian PDF in (3.1) and under the GAS time-varying parameters framework outlined by

$$\mathbf{f}_{t+1} = c + A \cdot s_t + B \cdot \mathbf{f}_t, \tag{3.6}$$

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 as in (3.1) and $\mathbf{S}_t(\cdot)$ a scaling function, which as in Lucas & Zhang (2016) is the inverse diagonal of the Fisher information matrix (see Creal et al. (2013); Harvey (2013) for more details or other scaling options), setting $c = 0$ and $B = 1$ Creal et al. (2013) show that the Integrated GAS (IGAS) reduces to

$$\mathbf{f}_{t+1} = \mathbf{f}_t + A \cdot s_t \tag{3.7}$$

and is identical to the IGARCH in (3.5) if $A = 1 - \omega$.

For the Student's t PDF described by

$$f(x_t \mid \mathbf{F}_{t-1}; \mathbf{f}_t, \theta) = \frac{\Gamma\left(\dfrac{\nu_t + 1}{2}\right)}{\Gamma\left(\dfrac{\nu_t}{2}\right)\sqrt{(\nu_t - 2)\pi\sigma_t^2}} \left(1 + \frac{x_t^2}{(\nu_t - 2)\sigma_t^2}\right)^{-\dfrac{\nu_t + 1}{2}}, \tag{3.8}$$

where $\Gamma(\cdot)$ denotes a gamma function as in Chapter 1, 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 for the t-GAS form of RiskMetrics. The recursions are outlined by

$$\mathbf{f}_{1,t+1} = \mathbf{f}_{1,t} + A_{\sigma_t^2} \cdot (1 + 3\nu_t^{-1}) \cdot \left(\frac{\nu_t + 1}{\nu_t - 2 + x_t^2/\mathbf{f}_{1,t}} \cdot x_t^2 - \mathbf{f}_{1,t}\right) \tag{3.9}$$

for σ^2_{t+1} and [1]

$$
\begin{aligned}
\mathbf{f}_{2,t+1} = \quad & \mathbf{f}_{2,t} - A_{\nu_t} \cdot \frac{2}{\nu_t - 2} \cdot \left(\gamma'' \left(\frac{\nu_t + 1}{2} \right) - \gamma'' \left(\frac{\nu_t}{2} \right) + \right. \\
& \left. \frac{2(\nu_t + 4)(\nu_t - 3)}{(\nu_t + 1)(\nu_t + 3)(\nu_t - 2)^2} \right)^{-1} \cdot \left(\gamma' \left(\frac{\nu_t + 1}{2} \right) - \right. \\
& \gamma' \left(\frac{\nu_t}{2} \right) - \frac{1}{\nu_t - 2} - \log \left(1 + \frac{x_t^2}{(\nu_t - 2)\mathbf{f}_{1,t}} \right) + \\
& \left. \frac{\nu_t + 1}{\nu_t - 2} \cdot \frac{x_t^2}{(\nu_t - 2)\mathbf{f}_{1,t} + x_t^2} \right), \qquad (3.10)
\end{aligned}
$$

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 (3.9) and (3.10). Although Lucas & Zhang (2016) show that setting $\omega = A \cdot (1 + 3\nu_t^{-1})$ in (3.9) provides a recursive form similar to (3.2), the recursions in (3.9) and (3.10) are notably more involved at the implementation stage than the standard RiskMetrics™ specification. However, they allow for the modelling of time-variation in the tails of financial returns within the functionality of the popular Student's t distribution.

An alternative parsimonious parametric choice, which also allows for heavy tails, though with more restrictions in the tails' shape, can be found in Laplace distribution. Its PDF for estimations is given by

$$f(x_t \mid \mathbf{F}_{t-1}; \mathbf{f}_t, \theta) = \frac{1}{\sqrt{2}\sigma_t} e^{-\frac{\sqrt{2}|x_t|}{\sigma_t}}, \qquad (3.11)$$

while its IGAS dynamics are specified as

$$\mathbf{f}_{3,t+1} = c + 2A \cdot \sqrt{2}|x_t|\sigma_t + (B - 2A) \cdot \mathbf{f}_{3,t}, \qquad (3.12)$$

[1]Note that Gaussian and Student's t PDFs used here have minor notation and parametrization differences from the forms previously given in Chapter 1.

which under $c = 0$, $A = \dfrac{1-\omega}{2}$ and $B = 1$ takes the "robust" form of Guermat & Harris (2002), given by

$$\sigma_{t+1}^2 = \omega \cdot \sigma_t^2 + (1-\omega) \cdot \sqrt{2}|x_t|\sigma_t, \tag{3.13}$$

as shown by Lucas & Zhang (2016) for L-GAS RiskMetrics parametrization. Even more pragmatic, but still a valid strategy, under RiskMetrics type weightings, may result from removing any particular form of distributional parametrzizations as in Harvey & Oryshchenko (2012) for dynamic kernel CDF estimator with weights for exponential filtering. From Chapter 2 and for the notation format adopted here, it is given by the recursion

$$\mathbf{F}_{1,t+1}(x) = \omega \cdot \mathbf{F}_{1,t}(x) + (1-\omega) \cdot W\left(\frac{x - x_t}{\beta}\right). \tag{3.14}$$

The recursion in (3.14) can be equivalently expressed as

$$\mathbf{F}_{1,t+1}(x) = (1-\omega)\sum_{i=1}^{t} W\left(\frac{x - x_i}{\beta}\right)\omega^i \tag{3.15}$$

or as

$$\mathbf{F}_{1,t+1}(x) = \frac{1-\omega}{1-\omega^t}\sum_{i=1}^{t} W\left(\frac{x - x_i}{\beta}\right)\omega^i. \tag{3.16}$$

The latter form insures non-negative weights summing to unity at all estimation stages and is the particular form used in all of the kernel affiliated computations considered further. Next, if the nonparametric setting offered by (3.16) is still regarded as overly complex, it can be simplified to the dynamic empirical CDF approach driven by the below already familiar recursion

$$\mathbf{F}_{2,t+1}(x) = \omega \cdot \mathbf{F}_{2,t}(x) + (1-\omega) \cdot \mathbb{I}_{\{x_t \leq x\}}, \tag{3.17}$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator function taking a value of 1 if the condition in the parenthesis is satisfied, (3.17) can be equivalently rewritten as

$$\mathbf{F}_{2,t+1}(x) = (1-\omega) \sum_{i=1}^{t} \mathbb{I}_{\{x_i \leq x\}} \omega^i, \tag{3.18}$$

and therefore takes the below form for computations

$$\mathbf{F}_{2,t+1}(x) = \frac{1-\omega}{1-\omega^t} \sum_{i=1}^{t} \mathbb{I}_{\{x_i \leq x\}} \omega^i. \tag{3.19}$$

For (3.17), (3.18) and (3.19) as covered in Section 2.3.2, similar forms are shown by Taylor (2007) in the form of the nonparametric quantile regression. Similar, to the kernel approach of Harvey & Oryshchenko (2012), dynamic eCDF and the empirical approach utilised in Taylor (2007) should be identical. However since in practice, quantile regressions may offer different dynamics for different quantiles, while here and as in Harvey & Oryshchenko (2012), exponentially declining weights are used to drive the dynamics of the entire time-varying eCDF. Estimated time-varying eCDF is then used for time-varying quantile extraction. Hence, as noted in Harvey (2010), quantiles should not cross. Direct evaluation of the dynamic quantiles with both parametric and nonparametric quantile regressions are beyond the scope of this thesis.

3.3 Framework for RiskMetrics competition

For simplicity, let VAR be outlined as

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

where Ω^{-1} is the inverse CDF of (3.1), (3.8), (3.11), (3.16) or (3.19) for the risk confidence level $\alpha \in (0,1)$. For the nonparametric versions of EWMA VAR the quantile algorithm described in Harvey & Oryshchenko (2012) and Section 2.3.1 is employed.

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

$$\mathrm{LRuc}_d(\alpha) = 2\left(\log\left[\left(\frac{N}{T}\right)^N \cdot \left(1-\frac{N}{T}\right)^{T-N}\right] - \log\left[(1-\alpha)^{T-N}\cdot\alpha^N\right]\right). \tag{3.20}$$

The statistic assumes that $\dfrac{N}{T} = 1-\alpha$ 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

$$\mathrm{LRcc}_d(\alpha) = \mathrm{LRuc}_d(\alpha) - \mathrm{LRin}_d(\alpha), \tag{3.21}$$

where LRuc_d is the unconditional coverage in (3.20) and LRin_d is independence LR test outlined by

$$\text{LRin}_d(\alpha) = 2\left(\log\left[\pi_{00}^{T_{00}}\pi_{01}^{T_{01}}\pi_{10}^{T_{10}}\pi_{11}^{T_{11}}\right] - \log\left[(1-\alpha)^{T_{01}+T_{11}} \cdot \alpha^{T_{00}+T_{10}}\right]\right), \tag{3.22}$$

where $\pi_{ij} = P(\mathbb{I}_t = j \mid \mathbb{I}_{t-1} = i) = \dfrac{T_{ij}}{T_{i0}+T_{i1}}$ 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 (3.21) follows $\chi^2(2)$ for $T \to \infty$. Essentially, the statistic in (3.22) tests the serial independence of VAR violations against a Markov first order dependence hypothesis, which in combination with (3.20) provides a joint test of independence and conditional coverage in (3.21).

3.4 Data

Applications of the outlined RiskMetrics variations are performed for the daily log-returns described in Tables 2.1 and 3.1. The first group is a data set from Chapter 2 used to demonstrate that nonparametric estimators as in Harvey & Oryshchenko (2012) can provide accurate density evaluations under the compound Berkowitz's (2001) criteria if the parameters are chosen accordingly. Since Nieto & Ruiz (2016) point out that accurate density forecasts are too formal for practical use and that this chapter is partially dedicated to investigating the fitness of the dynamic nonparametric methods for more common and generally applied tasks, one of the targets here is to support previous findings with the results for more pragmatic evaluation techniques and make comparisons to other up-to-date relevant methods in the literature.

Though, using parametric tail evalution criteria, Lucas & Zhang (2016), similar to what is outlined for kCDF RiskMetrics with nonparametric criteria in Chapter 2 by the general uniformity of the PITs, show that t-GAS RiskMetrics can correctly outline tail domains of financial returns. Therefore, in-sample first data group VAR comparisons of the dynamic kCDF method with its most technically advanced alternative on the date of estimations, should be of a particular value for the estimator of Harvey & Oryshchenko (2012) and its general potential future use in financial applications, since applied investigations for this type of estimator is scarce and still in the initial stages.

The second data group in Table 3.1 consists of the log-returns for BRICTS currencies exchange rates against USD for the period from 18.08.2010 to 14.09.2018. This yields a slightly more than 8 years of trading data observations for these currencies. Data for this group has been obtained from the daily spot exchange rates as reported by Bloomberg and is illustrated in Figure 3.1. Note that small differences in the sample length for each series in this group are explained by the differences in country specific trading calendars. Motivation for composing the BRICTS data group is based on the desire to obtain a group of the financial series demonstrating turbulence at the moment of producing this part of the thesis and insure an up-to-date empirical and valid investigation of the considered methods. For example, since the beginning of 2018: CNY experienced devaluations against USD on the agenda of the emerging trade barriers, RUB devalued against USD, though under "bullish" sentiment in the oil markets, on the agenda of sanctions, while INR, BRL and ZAR devaluations towards

Training	Returns	T	MIN	MAX	Mean	St.d.	Skew.	Kurt.	LB(12)	$LB(12)^2$	JB	AH(12)
18.08.10-15.01.16	BRL/USD	1356	-4.89	5.97	-0.061	0.95	0.03	5.71	0.02	≈ 0.00	≈ 0.00	≈ 0.00
18.08.10-15.01.16	RUB/USD	1363	-17.01	17.35	-0.069	1.21	-0.24	66.32	0.13	≈ 0.00	≈ 0.00	≈ 0.00
18.08.10-15.01.16	INR/USD	1310	-3.93	3.32	-0.028	0.53	-0.25	8.96	0.25	≈ 0.00	≈ 0.00	≈ 0.00
18.08.10-15.01.16	CNY/USD	1341	-1.83	0.61	0.002	0.13	-2.36	36.23	0.03	≈ 0.00	≈ 0.00	≈ 0.00
18.08.10-15.01.16	TRY/USD	1413	-3.35	3.16	-0.05	0.68	-0.29	4.65	0.87	0.02	≈ 0.00	≈ 0.00
18.08.10-15.01.16	ZAR/USD	1413	-6.44	5.08	-0.059	0.91	-0.29	6.07	0.61	≈ 0.00	≈ 0.00	≈ 0.00
Testing	Returns	T	MIN	MAX	Mean	St.d.	Skew.	Kurt.	LB(12)	$LB(12)^2$	JB	AH(12)
18.01.16-14.09.18	BRL/USD	668	-7.11	5.13	0.005	1.02	-0.31	8.22	0.03	≈ 0.00	≈ 0.00	0.07
18.01.16-14.09.18	RUB/USD	672	-4.19	5.76	0.023	0.94	0.10	6.81	0.04	≈ 0.00	≈ 0.00	≈ 0.00
18.01.16-14.09.18	INR/USD	644	-1.58	1.19	-0.006	0.29	-0.49	5.08	0.81	0.63	≈ 0.00	0.72
18.01.16-14.09.18	CNY/USD	652	-0.83	1.19	-0.007	0.24	0.29	5.63	0.91	≈ 0.00	≈ 0.00	0.02
18.01.16-14.09.18	TRY/USD	695	-14.74	8.07	-0.101	1.20	-2.51	40.31	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
18.01.16-14.09.18	ZAR/USD	694	-4.91	3.18	0.020	1.12	-0.34	3.98	0.42	0.01	≈ 0.00	0.09

Table 3.1: Descriptive statistics for the specified 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, 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.

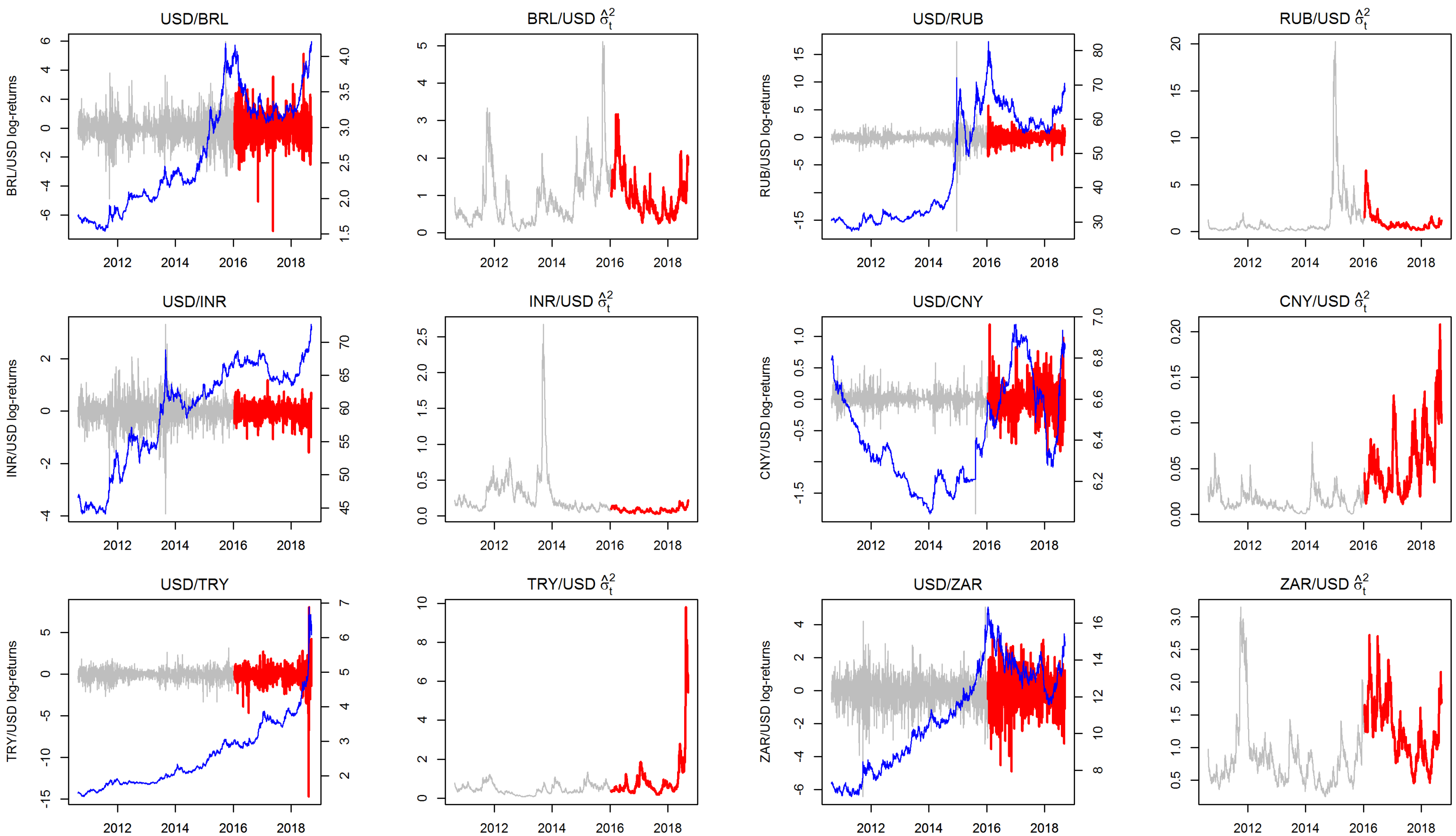

Figure 3.1: BRICTS log-returns series and their t-GAS RM $\hat{\sigma}_t^2$ for the parameters in Table 3.2. Testing sample log-returns and their $\hat{\sigma}_t^2$ are highlighted in red.

USD were likely due to the international capital outflow from the emerging markets to the USA on the attractive "quantitative tightening" policy yields and finally, TRY devaluation against USD could be explained by "liberal" monetary policy rates and emerging trade barriers. In fact, Turkey may be argued to be one of the key factors underpinning emerging markets volatility in August-September 2018 and therefore the cut-off point for the BRICTS sample is set on the 14.09.2018, which is the end of the trading week for Turkish monetary policy adjustment and start of the relatively calm sentiment in the emerging markets in September 2018.

The choice of the backtesting procedure for the second data group is conventional and straightforward. BRICTS data is split into two sub-samples: training (in-sample) and testing (out-of-sample) parts. The data is partitioned on the 18.01.2016 producing 5+ years of the data for training and 2+ years of trading data for testing. The training data part is used to obtain parameters for estimations which are then used for VAR estimations in the testing part of the BRICTS data. For the daily out-of-sample VAR projections a rolling window of 1000 observations parameter updating backtests similar to Laporta et al. (2018) is also performed. Both backtesting frameworks for the out-of-samples evaluations are common and are also employed in Lucas & Zhang (2016) for introducing t and other GAS RiskMetrics specifications as well as validating their appropriateness and should aid analytical discussion of the obtained results.

3.5 Estimated parameters

To estimate parameters for VAR evaluations of financial returns, described in Tables 2.1 and 3.1, MLE is utilized as outlined in Creal et al. (2013),

$$\Upsilon = \arg\max_{\mathbf{f_t},\theta} \sum_{i=1}^{t} \log f(x_t \mid \mathbf{F}_{t-1}; \mathbf{f}_t, \theta) \tag{3.23}$$

for (3.1), (3.8) and (3.11), while recursive version of the Bowman et al.'s (1998) LSE

$$\Lambda = \frac{1}{T-m} \sum_{t=m}^{T-1} \int_{-\infty}^{\infty} \left[\mathbb{I}_{\{x_{t+1} \leq x\}} - \mathbf{F}_{\iota,t+1}(x) \right]^2 dx \tag{3.24}$$

is used for (3.16) and (3.19), since results in Chapter 2 indicate its superiority over MLE, as was originally suggested in Harvey & Oryshchenko (2012). Estimated parameters for all of the methods described in Section 3.2 are provided in Table 3.2. Since for the nonparmetric methods $m = 250$, all VAR evaluations for log-returns as in Table 2.1 and for BRICTS training sub-samples, are performed excluding the initial 250 observations. These observations are used in the estimation of unknown parameters for all methods, but are excluded from VAR in-sample evaluations.

For the rolling window VAR projections of the nonparametric methods estimations are simplified with the rolling window parameters obtained for RiskMetrics™. For the kCDF RiskMetricks the bandwidth parameter is then approximated as follows

$$\beta = 0.93 \cdot \mathbf{min}\left(\sigma_s; \mathrm{IQR}_s/1.34\right) \cdot \left(\frac{1}{1-\omega}\right)^{-1/3}, \tag{3.25}$$

Returns	eCDF RM	kCDF RM	RM$^{\text{TM}}$	L-GAS RM	$t-$GAS RM		†
BRENT	$\omega = 0.9864$	$\beta = 0.4261$; $\omega = 0.9841$	$\omega = 0.9566$	$A_{\sigma_t^2} = 0.0371$	$A_{\sigma_t^2} = 0.0371$	-	$\nu = 7.7383$
NOK/USD	$\omega = 0.9886$	$\beta = 0.1879$; $\omega = 0.9866$	$\omega = 0.9688$	$A_{\sigma_t^2} = 0.0329$	$A_{\sigma_t^2} = 0.0332$	$A_{\nu_t} = 0.0008$	$\nu = 13.8645$
RUB/USD	$\omega = 0.9716$	$\beta = 0.0586$; $\omega = 0.9703$	$\omega = 0.9399$	$A_{\sigma_t^2} = 0.0683$	$A_{\sigma_t^2} = 0.0718$	$A_{\nu_t} = 0.0148$	$\nu = 2.0426$
GBP/USD	$\omega = 0.9906$	$\beta = 0.1249$; $\omega = 0.9896$	$\omega = 0.9398$	$A_{\sigma_t^2} = 0.0373$	$A_{\sigma_t^2} = 0.0335$	-	$\nu = 5.1987$
NASDAQ	$\omega = 0.9811$	$\beta = 0.8237$; $\omega = 0.9757$	$\omega = 0.9337$	$A_{\sigma_t^2} = 0.0560$	$A_{\sigma_t^2} = 0.0635$	-	$\nu = 22.5702$
SP500	$\omega = 0.9731$	$\beta = 0.3664$; $\omega = 0.9708$	$\omega = 0.9317$	$A_{\sigma_t^2} = 0.0640$	$A_{\sigma_t^2} = 0.0676$	$A_{\nu_t} = 0.0047$	$\nu = 3.7728$
FTSE100	$\omega = 0.9800$	$\beta = 0.2611$; $\omega = 0.9726$	$\omega = 0.9079$	$A_{\sigma_t^2} = 0.0793$	$A_{\sigma_t^2} = 0.0869$	-	$\nu = 10.1051$
NIKKEI225	$\omega = 0.9639$	$\beta = 0.4513$; $\omega = 0.9562$	$\omega = 0.8804$	$A_{\sigma_t^2} = 0.0943$	$A_{\sigma_t^2} = 0.0997$	-	$\nu = 7.3343$
GOLD	$\omega = 0.9847$	$\beta = 0.2402$; $\omega = 0.9811$	$\omega = 0.9559$	$A_{\sigma_t^2} = 0.0423$	$A_{\sigma_t^2} = 0.0421$	-	$\nu = 4.1925$
BTC/USD	$\omega = 0.9774$	$\beta = 0.6112$; $\omega = 0.9748$	$\omega = 0.9035$	$A_{\sigma_t^2} = 0.1018$	$A_{\sigma_t^2} = 0.0465$	$A_{\nu_t} = 0.1021$	$\nu = 3.4921$
BRL/USD	$\omega = 0.9795$	$\beta = 0.2389$; $\omega = 0.9765$	$\omega = 0.9213$	$A_{\sigma_t^2} = 0.0696$	$A_{\sigma_t^2} = 0.0743$	$A_{\nu_t} = 0.0026$	$\nu = 9.5772$
RUB/USD	$\omega = 0.9632$	$\beta = 0.2319$; $\omega = 0.9569$	$\omega = 0.9194$	$A_{\sigma_t^2} = 0.0721$	$A_{\sigma_t^2} = 0.0727$	-	$\nu = 5.3637$
INR/USD	$\omega = 0.9838$	$\beta = 0.1130$; $\omega = 0.9794$	$\omega = 0.9422$	$A_{\sigma_t^2} = 0.0544$	$A_{\sigma_t^2} = 0.0514$	-	$\nu = 5.5282$
CNY/USD	$\omega = 0.9665$	$\beta = 0.0206$; $\omega = 0.9631$	$\omega = 0.9939$	$A_{\sigma_t^2} = 0.0224$	$A_{\sigma_t^2} = 0.0701$	$A_{\nu_t} = 0.0335$	$\nu = 3.0226$
TRY/USD	$\omega = 0.9869$	$\beta = 0.1561$; $\omega = 0.9844$	$\omega = 0.9541$	$A_{\sigma_t^2} = 0.0481$	$A_{\sigma_t^2} = 0.0498$	$A_{\nu_t} = 0.0017$	$\nu = 14.6825$
ZAR/USD	$\omega = 0.9927$	$\beta = 0.2244$; $\omega = 0.9906$	$\omega = 0.9543$	$A_{\sigma_t^2} = 0.0401$	$A_{\sigma_t^2} = 0.0435$	-	$\nu = 10.0561$

Table 3.2: RM estimated parameters for the specified returns series. † for t-GAS RM with $A_{\nu_t} > 0$, ν is the estimated initialization point for the d.f. dynamics.

parameters for (3.25) may be worthwhile investigating in the greater detail in the future.

3.6 Competition results and discussion

Output for all test statistics of the VAR projections for the corresponding samples are provided in Tables 3.3 and 3.4. For tests, the decision significance rule is set at the standard 5% level and obtained values exceeding this threshold are not differentiated among each other. The statistics of Christoffersen (1998) is preffered over Kupiec's (1995) test output, though in fruther interpretations the actual over expected exceedances (AE) ratio is also considered important and essentially serves as a simplified version of Kupiec's (1995) statistic. Further discussion mainly focuses on the common 99% and 95% risk confidence levels similar to Laporta et al. (2018), but also provides results for the 90% level to check the nonparametric estimators ability of picking up evolving quantiles more into the "body" domain of financial returns and thus their potential appropriateness for studies interested in their broader range (e.g. as in Reboredo & Ugolini (2016)). Illustration of differing types quantile dynamics provided by the different variations of RiskMetrics is exemplified on the BRENT log-returns in Figure 3.2. From Figure 3.2, differences in the quantiles evolution are most visually distinguished at the 99% level. It is easy to see that dynamic eCDF quantiles for this risk confidence level are the least smooth with little time variation at certain periods of estimation. This may be expected from the functional form of the dynamic eCDF in (3.17), (3.18) and (3.19) or from what is generally known of its static version (e.g. Markovich (2008)). Enhancing the capabilities of eCDF with kernels in the form of CDF adds smoothness in the evolution of the nonparametric quantiles, but the changes are arguably slower than those for the parametric quantiles. Faster time adjustments in the parametric quantiles may be explained by the differences in the estimated parameters between nonparametric and parametric methods reported in Table 3.2 and discussed in the previous section as well as by the "robustifying volatility responses" of GAS models. For example, Lucas & Zhang (2016), similar to Creal et al. (2013) and Harvey (2013), argue that under the GAS framework, the closer the latest available observation is to the tail domain of the Student's t distribution, the sharper should be the response in the projected volatility estimate and quantiles respectively. If the observation is closer to the middle, it should have a smaller impact on the respective projections. On the other hand, for the illustrated parametric quantiles in Figure 3.2, the events of the financial crisis and associated turbulence in the oil markets highlight the most notable differences in the estimates based on the Laplace and Student's t distributions. Laplace quantiles are higher in absolute values, and are more conservative, than the estimated quantiles for the Student's t distribution. In fact, Laplace quantiles may be argued to be visually closer to the values of the financial crisis quantiles for the kernel method than that of the Student's t distribution in Figure 3.2.

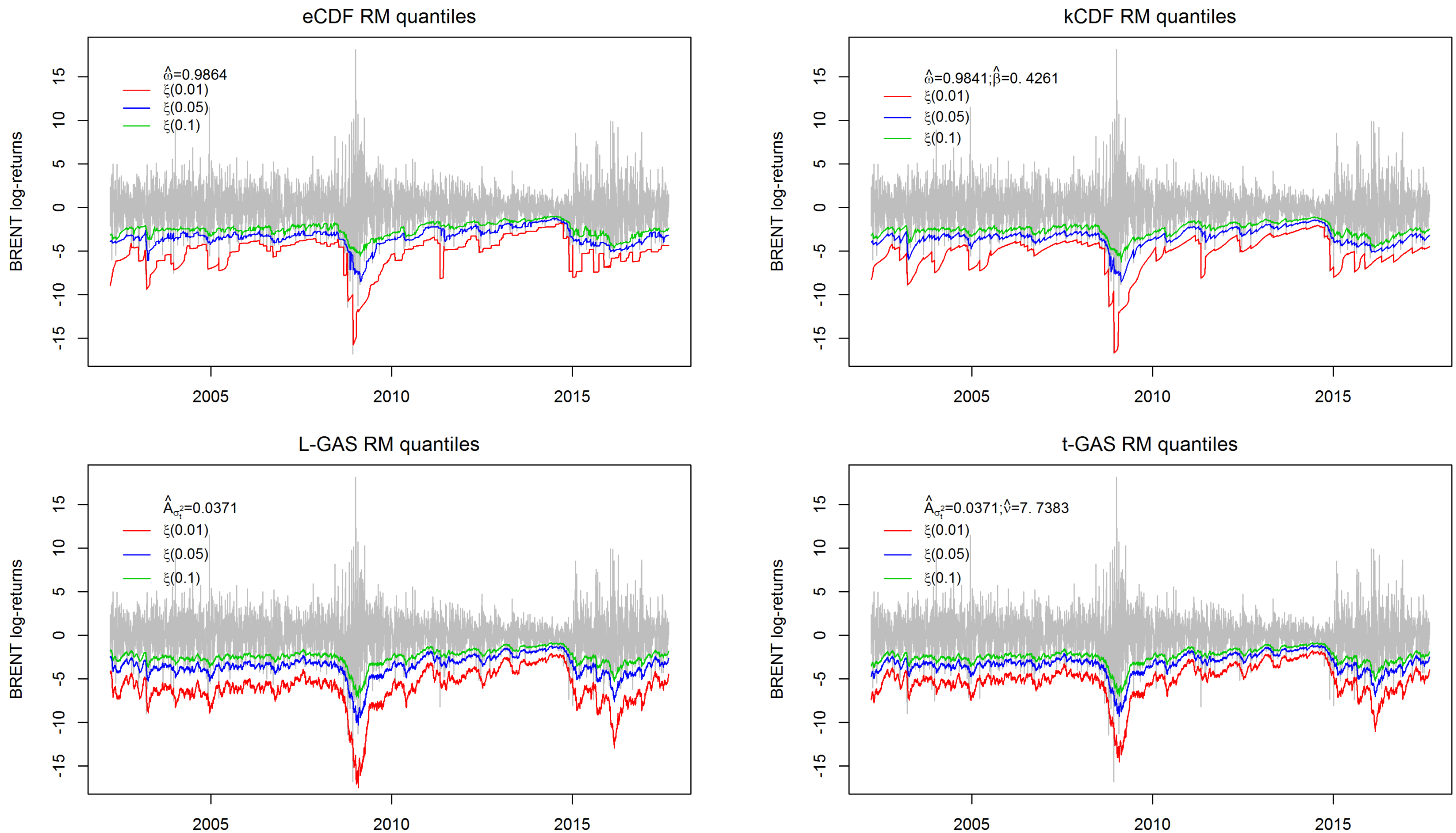

Figure 3.2: eCDF, kCDF, L-GAS and t-GAS RM lower tail time-varying quantiles for BRENT log-returns.

		eCDF RM					kCDF RM					RM^{TM}					L-GAS RM					t-GAS RM				
VAR_α	Returns	$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	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	0.0040	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	1.0503	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	0.6064	0.4361	0.9190	0.6316	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 3.3: VAR estimation results for the specified methods and VAR levels for data set used in Chapter 2, where $LRuc_p$ and $LRcc_p$ denote p-values for Kupiec (1995) and Christoffersen (1998) statistics, AE is actual over expected exceedances ratio, while statistic exceeding 5% significance threshold is highlighted in grey.

First, the VAR estimation results in Table 3.3 are discussed and analysed. At the 99% level the poor performance of RiskMetrics™ is the most notable result, then with observed improvements, follow the eCDF and L-GAS results, which, in turn, are slightly outperformed by the kCDF RiskMetrics. The leading position is taken by the t-GAS RiskMetrics. Going to the 95% level the outlook most notably changes for RiskMetrics™. For this level RiskMetrics™ results are, in fact, more satisfactory than that of its t-GAS alternative, but results for L-GAS parameterization are also worth highlighting. Mainly due to the initial expectation that if the Laplace specification works well that would be at the 99% level domains due to the greater shape adaptability characteristics of the Student's t distribution (or also see previous results of Gerlach et al. (2013) or Lucas & Zhang (2016)). Most likely different decays in speed of the distributions' tails and body domains (e.g. Cont (2001); Markovich (2008)) are observed for these groups of returns. This may be quite problematic when capturing with a specific tail oriented parametrization, even for in-sample VAR evaluations as here. On the other hand, the functionality of the nonparametric methods are data driven and as long as the estimation domains possess enough information in the observed data, these estimators should perform well under good choices of parameters as empirically shown by the density forecasts in the previous chapter and well outlined by the good results at the 95% level in Table 3.3. Though, the 95% confidence level still interests researchers a lot (e.g. Ji et al. (2018) among others for a recent application at this particular level) nonparametric in-sample results for this and 90% levels together, in particular what is observed for the AE ratio at these levels, add more empirical evidence for applications of these methods in the full-scale nonparametric changing dependence testing frameworks. For example, the obtained quantiles may be directly used in applications similar to Harvey (2010) or pseudo-observations may be obtained with these estimators specifically without parametric pre-filtering for the tail dependency tests as in Bücher et al. (2015) given that the diagnotic output for the obtained pseudo-observations is satisfactory (e.g. using Berkowitz's (2001) criteria). It is also interesting to note that the kCDF results even at the 99% level are in-line with the previous density forecasts, though noted as not very practical by Nieto & Ruiz (2016), and accuracy ratio for this estimator in Chapter 2 indicating the robustness of both the choice of parameters and adopted LSE approach to parameters estimation in general.

The results presented in Table 3.4 allow providing a discussion more focused on VAR than the above, since the setting is more appropriate. Reviewing the 99% level results for in-sample evaluations a different outlook on the best performing specification is observed. L-GAS comfortably outperforms its technically more elaborate Student's t alternative. In fact the t-GAS performance is not very encouraging and is comparable to the most parsimonious eCDF specification. The kCDF RiskMetrics again takes the "first runner-up" position for this level of in-sample evaluations, while RikMetrics™ performs the worst. Given a good pool of in-sample evaluations it can be now noted that Laplace based estimations tend to overestimating risk exposure when compared to the other methods considered. On the basis of its slowly evolving quantile, kCDF may be expected to be the second most conservative method, but this is not consistently observed in obtained results at the 99% level when compared to the t-GAS estimation results. This reflects an intermittent vulnerability of the strictly data driven approach and benefits of the parametric assumptions in the tails. Moving to the 95% and 90% levels the dominance of the nonparametric estimators is difficult to debate, though results for the Laplace based specification for the 95% level may be again worth highlighting.

		eCDF RM - training					kCDF RM - training					RM - training					L-GAS RM - training					t-GAS RM - training				
VAR_α	Returns	$LRuc_d$	$LRuc_p$	$LRcc_d$	$LRcc_p$	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	BRL/USD	5.8891	0.0152	6.6265	0.0364	1.8083	5.8891	0.0152	6.6265	0.0364	1.8083	7.1222	0.0076	7.9351	0.0189	1.8970	0.4177	0.5181	0.5653	0.7538	0.8130	1.2685	0.2601	1.6809	0.4315	1.3550
	RUB/USD	14.5811	0.0001	23.8295	0.0000	2.3360	4.6385	0.0313	13.4424	0.0012	1.7071	22.0437	0.0000	29.1850	0.0000	2.6930	0.0018	0.9663	0.2214	0.8952	0.9874	4.6242	0.0315	13.4330	0.0012	1.7056
	INR/USD	1.6344	0.2011	3.2433	0.1976	1.4151	1.0007	0.3171	2.8435	0.2413	1.3208	26.0942	0.0000	29.5420	0.0000	2.9218	2.3930	0.1219	7.7430	0.0208	1.5080	9.4308	0.0021	16.2713	0.0003	2.0735
	CNY/USD	7.4179	0.0065	8.2431	0.0162	1.9248	6.1388	0.0132	6.8865	0.0320	1.8332	11.7969	0.0006	27.1777	0.0000	2.1978	2.0877	0.1485	7.5422	0.0230	1.4652	4.9466	0.0261	19.1065	0.0001	1.7399
	TRY/USD	0.0118	0.9136	0.2622	0.8771	1.0318	0.0118	0.9136	0.2622	0.8771	1.0318	10.1462	0.0014	11.1576	0.0038	2.0619	1.2914	0.2558	1.4022	0.4960	0.6873	3.0084	0.0828	3.5744	0.1674	1.5464
	ZAR/USD	0.9037	0.3418	7.0935	0.0288	1.2898	0.1570	0.6919	2.4208	0.2981	1.1178	16.6678	0.0000	16.8126	0.0002	2.4055	0.2451	0.6206	0.4185	0.8112	0.8591	10.1462	0.0014	10.5641	0.0051	2.0619
0.95	BRL/USD	1.6997	0.1923	1.7088	0.4255	1.1754	0.0545	0.8154	0.0558	0.9725	1.0307	4.8355	0.0279	5.9913	0.0500	1.3008	0.1065	0.7442	0.1939	0.9076	0.9575	6.6412	0.0100	7.3947	0.0248	1.3550
	RUB/USD	1.9172	0.1662	12.4587	0.0020	1.1860	0.0342	0.8533	2.8793	0.2370	1.0243	3.1173	0.0775	6.2680	0.0435	1.2388	1.2448	0.2645	2.6570	0.2649	1.1490	11.4843	0.0007	17.2132	0.0002	1.4722
	INR/USD	0.6908	0.4059	6.3404	0.0420	1.1132	0.1756	0.6751	4.7907	0.0911	1.0566	7.1043	0.0077	7.9778	0.0185	1.3761	4.6300	0.0314	5.1615	0.0757	1.3007	13.4430	0.0002	17.2762	0.0002	1.5269
	CNY/USD	0.7745	0.3788	4.1248	0.1271	1.1182	0.1142	0.7354	1.4054	0.4952	1.0449	1.8851	0.1698	18.5740	0.0001	0.8242	0.8739	0.3499	19.1428	0.0001	0.8791	1.2989	0.2544	7.9082	0.0192	1.1538
	TRY/USD	0.0130	0.9092	0.0130	0.9935	1.0146	0.7088	0.3998	0.7615	0.6833	0.8942	8.4247	0.0037	9.0305	0.0109	1.3918	0.8072	0.3690	0.9378	0.6257	1.1168	9.1407	0.0025	9.8426	0.0073	1.4089
	ZAR/USD	1.3544	0.2445	1.3654	0.5052	1.1522	0.2629	0.6081	0.4415	0.8019	1.0662	2.7776	0.0956	2.8075	0.2457	1.2199	0.0115	0.9145	0.3550	0.8374	1.0137	5.2490	0.0220	5.5143	0.0635	1.3058
0.9	BRL/USD	2.2907	0.1302	2.3027	0.3162	1.1392	0.0684	0.7937	0.3029	0.8594	0.9765	9.1107	0.0025	11.3548	0.0034	1.2827	13.4341	0.0002	13.9857	0.0009	1.3460	18.5177	0.0000	18.7582	0.0001	1.4092
	RUB/USD	3.6876	0.0548	18.8223	0.0001	1.1770	0.0169	0.8965	8.1658	0.0169	0.9883	3.6480	0.0561	20.7965	0.0000	1.1759	10.3847	0.0013	26.2572	0.0000	1.3016	14.9357	0.0001	31.8286	0.0000	1.3645
	INR/USD	0.0417	0.8382	2.6887	0.2607	1.0189	0.0951	0.7577	1.8665	0.3933	0.9717	0.6397	0.4238	3.9487	0.1388	1.0745	5.1747	0.0229	7.4401	0.0242	1.2158	10.4323	0.0012	12.9261	0.0016	1.3101
	CNY/USD	0.3490	0.5547	7.3203	0.0257	1.0541	0.2687	0.6042	7.0547	0.0294	0.9533	22.9155	0.0000	42.3085	0.0000	0.5952	3.1631	0.0753	31.8829	0.0000	0.8425	0.2314	0.6305	16.9502	0.0002	1.0440
	TRY/USD	0.0047	0.9455	0.0698	0.9657	1.0060	0.0047	0.9455	0.0698	0.9657	1.0060	5.4494	0.0196	5.5285	0.0630	1.2113	8.8561	0.0029	8.8801	0.0118	1.2715	9.4055	0.0022	9.4545	0.0089	1.2801
	ZAR/USD	3.1936	0.0739	3.6618	0.1603	1.1608	0.0009	0.9766	0.6472	0.7235	0.9974	2.2366	0.1348	2.3021	0.3163	1.1340	7.8040	0.0052	7.8491	0.0198	1.2543	11.7559	0.0006	11.7559	0.0028	1.3144
		eCDF RM - testing					kCDF RM - testing					RM - testing					L-GAS RM - testing					t-GAS RM - testing				
0.99	BRL/USD	0.7341	0.3916	0.9803	0.6125	1.3473	0.0724	0.7879	0.1813	0.9133	0.8982	1.4460	0.2292	3.6934	0.1578	1.4970	1.2683	0.2601	1.3165	0.5178	0.5988	0.4675	0.4942	0.5430	0.7622	0.7485
	RUB/USD	6.0711	0.0137	21.8283	0.0000	2.0833	3.3977	0.0653	15.0989	0.0005	1.7857	1.4061	0.2357	15.5673	0.0004	1.4881	0.4879	0.4848	5.5440	0.0625	0.7440	0.0808	0.7762	12.1091	0.0023	0.8929
	INR/USD	1.7010	0.1922	3.8844	0.1434	1.5528	0.9148	0.3388	3.4895	0.1747	1.3975	10.1460	0.0014	10.9627	0.0042	2.4845	0.9148	0.3388	1.1704	0.5570	1.3975	1.7010	0.1922	2.0170	0.3648	1.5528
	CNY/USD	5.0472	0.0247	6.3523	0.0417	1.9939	5.0472	0.0247	6.3523	0.0417	1.9939	48.6454	0.0000	50.0166	0.0000	4.7546	30.7743	0.0000	36.9296	0.0000	3.8344	103.2521	0.0000	114.5631	0.0000	7.0552
	TRY/USD	12.3371	0.0004	19.2126	0.0001	2.5899	8.7027	0.0032	12.5591	0.0019	2.3022	90.3313	0.0000	101.0310	0.0000	6.3309	23.3280	0.0000	40.3412	0.0000	3.3094	54.9357	0.0000	70.0998	0.0000	4.8921
	ZAR/USD	2.0371	0.1535	8.7983	0.0123	1.5850	0.0005	0.9818	0.1434	0.9308	1.0086	16.4671	0.0000	26.4105	0.0000	2.8818	2.0371	0.1535	4.0070	0.1349	1.5850	14.3642	0.0002	25.1330	0.0000	2.7378
0.95	BRL/USD	0.9400	0.3323	1.1757	0.5555	1.1677	0.0795	0.7780	0.7817	0.6765	1.0479	1.8621	0.1724	1.8623	0.3941	0.7784	5.5513	0.0185	5.7127	0.0575	0.6287	0.3767	0.5394	0.4848	0.7847	0.8982
	RUB/USD	0.0113	0.9152	8.5830	0.0137	0.9821	0.6937	0.4049	8.4298	0.0148	0.8631	9.1086	0.0025	26.0415	0.0000	0.5357	7.8681	0.0050	24.2109	0.0000	0.5655	9.1086	0.0025	26.0415	0.0000	0.5357
	INR/USD	0.1041	0.7470	0.8654	0.6488	1.0559	0.9331	0.3341	1.5310	0.4651	0.8385	7.1381	0.0075	7.7313	0.0209	1.4907	0.7202	0.3961	2.2337	0.3273	1.1491	14.1508	0.0002	16.5163	0.0003	1.7081
	CNY/USD	0.3616	0.5476	0.3616	0.8346	1.1043	0.3616	0.5476	0.3616	0.8346	1.1043	26.6349	0.0000	27.1051	0.0000	1.9939	31.2494	0.0000	32.5930	0.0000	2.0859	133.8295	0.0000	145.1031	0.0000	3.4969
	TRY/USD	4.1137	0.0425	8.2222	0.0164	1.3525	4.1137	0.0425	10.3642	0.0056	1.3525	27.9662	0.0000	45.3287	0.0000	1.9856	3.4953	0.0615	31.2340	0.0000	1.3237	39.0807	0.0000	55.7513	0.0000	2.1871
	ZAR/USD	1.1406	0.2855	3.5876	0.1663	1.1816	0.0507	0.8219	4.2334	0.1204	1.0375	2.4274	0.1192	7.7932	0.0203	1.2680	0.3209	0.5711	6.0913	0.0476	1.0951	1.9485	0.1627	7.7848	0.0204	1.2392
0.9	BRL/USD	0.0797	0.7777	0.3157	0.8540	1.0329	0.5746	0.4484	0.6124	0.7362	0.9132	3.3848	0.0658	3.5507	0.1694	0.7934	1.3420	0.2467	1.5493	0.4609	0.8683	0.3917	0.5314	2.3404	0.3103	0.9281
	RUB/USD	0.0007	0.9795	1.9382	0.3794	0.9970	0.4578	0.4987	3.9320	0.1400	0.9226	16.4610	0.0000	27.8519	0.0000	0.5655	11.9554	0.0005	27.4229	0.0000	0.6250	15.2542	0.0001	25.9110	0.0000	0.5804
	INR/USD	0.0438	0.8341	0.0531	0.9738	1.0248	0.5162	0.4725	1.0409	0.5943	0.9161	2.2079	0.1373	2.7611	0.2514	1.1801	4.9633	0.0259	5.2523	0.0724	1.2733	8.0312	0.0046	8.5724	0.0138	1.3509
	CNY/USD	3.9816	0.0460	5.0181	0.0813	1.2423	3.5072	0.0611	4.7445	0.0933	1.2270	13.4590	0.0002	16.0381	0.0003	1.4571	31.4287	0.0000	37.0024	0.0000	1.7178	91.2746	0.0000	97.1334	0.0000	2.2853
	TRY/USD	3.1731	0.0749	8.3775	0.0152	1.2086	2.3759	0.1232	5.7180	0.0573	1.1799	8.7424	0.0031	26.9650	0.0000	1.3525	0.8720	0.3504	31.3853	0.0000	1.1079	18.6604	0.0000	44.7524	0.0000	1.5252
	ZAR/USD	0.1071	0.7435	3.0957	0.2127	1.0375	0.0316	0.8590	1.8559	0.3954	0.9798	0.0932	0.7602	5.1536	0.0760	0.9654	0.3323	0.5643	5.5983	0.0609	1.0663	0.4905	0.4837	5.3254	0.0698	1.0807
		eCDF RM - roll testing					kCDF RM - roll testing					RM - roll testing					L-GAS RM - roll testing					t-GAS RM - roll testing				
0.99	BRL/USD	19.7793	0.0000	19.9407	0.0000	3.1437	1.4460	0.2292	3.6934	0.1578	1.4970	0.7341	0.3916	0.9803	0.6125	1.3473	2.5774	0.1084	2.6045	0.2719	0.4491	1.2683	0.2601	1.3165	0.5178	0.5988
	RUB/USD	19.6050	0.0000	24.5998	0.0000	3.1250	3.3977	0.0653	3.8347	0.1470	1.7857	9.3301	0.0023	10.0705	0.0065	2.3810	0.0808	0.7762	4.3586	0.1131	0.8929	0.7063	0.4007	3.3577	0.1866	1.3393
	INR/USD	8.3611	0.0038	9.0777	0.0107	2.3292	5.2107	0.0224	5.7472	0.0565	2.0186	8.3611	0.0038	9.2203	0.0100	2.3292	0.3544	0.5516	0.5560	0.7573	1.2422	6.7128	0.0096	7.7711	0.0205	2.1739
	CNY/USD	1.6130	0.2041	1.9250	0.3819	1.5337	1.6130	0.2041	1.9250	0.3819	1.5337	20.4925	0.0000	22.3567	0.0000	3.2209	0.0430	0.8357	0.1547	0.9256	0.9202	2.4016	0.1212	2.4294	0.2968	0.4601
	TRY/USD	18.6317	0.0000	23.7950	0.0000	3.0216	8.7027	0.0032	12.5591	0.0019	2.3022	25.8130	0.0000	27.1133	0.0000	3.4532	5.5813	0.0182	6.7521	0.0342	2.0144	10.4596	0.0012	18.0008	0.0001	2.4460
	ZAR/USD	12.3695	0.0004	13.3296	0.0013	2.5937	4.2524	0.0392	4.7495	0.0930	1.8732	3.0599	0.0802	3.4828	0.1753	1.7291	1.4845	0.2231	1.5309	0.4651	0.5764	1.1993	0.2735	1.4922	0.4742	1.4409
0.95	BRL/USD	1.2948	0.2552	1.4542	0.4833	1.1976	0.0795	0.7780	0.7817	0.6765	1.0479	0.2080	0.6483	0.7684	0.6810	1.0778	0.3767	0.5394	0.4848	0.7847	0.8982	0.9400	0.3323	1.1757	0.5555	1.1677
	RUB/USD	5.7681	0.0163	6.5887	0.0371	1.4286	1.0390	0.3081	1.0667	0.5866	0.8333	0.1765	0.6744	2.2629	0.3226	1.0714	1.4588	0.2271	10.1683	0.0062	0.8036	0.0606	0.8055	7.1163	0.0285	1.0417
	INR/USD	2.3388	0.1262	2.5163	0.2842	1.2733	0.0208	0.8854	0.0793	0.9611	1.0248	1.0423	0.3073	2.3323	0.3116	1.1801	0.0013	0.9711	2.9558	0.2281	0.9938	7.1381	0.0075	7.1930	0.0274	1.4907
	CNY/USD	13.5523	0.0002	13.5837	0.0011	1.6871	14.6918	0.0001	14.7000	0.0006	1.7178	0.8960	0.3438	2.2353	0.3270	1.1656	0.0117	0.9139	1.1710	0.5568	0.9816	0.0624	0.8027	0.0929	0.9546	1.0429
	TRY/USD	1.9233	0.1655	2.5951	0.2732	1.2374	1.1215	0.2896	2.1138	0.3475	1.1799	10.6344	0.0011	12.2763	0.0022	1.5827	7.8733	0.0050	10.2940	0.0058	1.4964	16.0180	0.0001	20.4309	0.0000	1.7266
	ZAR/USD	0.8139	0.3670	3.5641	0.1683	1.1527	0.0027	0.9584	2.3859	0.3033	1.0086	1.5192	0.2177	5.5424	0.0626	1.2104	0.0891	0.7654	5.6020	0.0608	0.9510	1.9485	0.1627	5.5939	0.0610	1.2392
0.9	BRL/USD	0.8362	0.3605	4.0649	0.1310	1.1078	1.3420	0.2467	1.6174	0.4454	0.8683	0.0007	0.9794	0.5933	0.7433	1.0030	1.3541	0.2446	1.7811	0.4104	1.1377	5.1178	0.0237	6.1622	0.0459	1.2725
	RUB/USD	5.9883	0.0144	6.1489	0.0462	1.2946	0.4578	0.4987	0.4857	0.7844	0.9226	0.2972	0.5856	0.5732	0.7508	0.9375	0.0007	0.9795	1.9382	0.3794	0.9970	0.7427	0.3888	1.3051	0.5207	1.1012
	INR/USD	1.2287	0.2677	2.0982	0.3502	1.1335	0.2200	0.6390	0.2264	0.8930	1.0559	1.5247	0.2169	1.5593	0.4586	1.1491	4.4324	0.0353	4.4377	0.1087	1.2578	6.7224	0.0095	6.9081	0.0316	1.3199
	CNY/USD	20.0504	0.0000	20.0531	0.0000	1.5644	20.0504	0.0000	20.0531	0.0000	1.5644	1.2703	0.2597	1.2840	0.5262	1.1350	0.2420	0.6228	0.3207	0.8518	1.0583	0.1319	0.7164	0.7250	0.6959	1.0429
	TRY/USD	2.7608	0.0966	3.2933	0.1927	1.1942	0.6576	0.4174	2.5166	0.2841	1.0935	1.1156	0.2909	4.6178	0.0994	1.1223	5.0904	0.0241	7.6073	0.0223	1.2662	6.8026	0.0091	8.5115	0.0142	1.3094
	ZAR/USD	0.4905	0.4837	1.6785	0.4320	1.0807	0.0026	0.9596	0.7594	0.6841	0.9942	0.8963	0.3438	1.7240	0.4223	1.1095	0.8963	0.3438	3.5111	0.1728	1.1095	1.7233	0.1893	3.5177	0.1722	1.1527

Table 3.4: VAR estimation results for the specified methods and VAR levels for BRICTS training, testing and roll testing data sets.

For the testing period with no parameter updating estimation strategy in Table 3.4 the outlook for the best performing models is very similar to the in-sample BRICTS evaluations. Again, Laplace distribution captures the 99% level better than all other models. It is then followed by the kCDF EWMA performance. eCDF and t-GAS RiskMetrics accuracy ratios are similar again, while RiskMetrics™ performance is still the worst. It is worth pointing out that none of the specifications capture Turkish Lira out-of-sample volatility at the 99% level with quite worrying levels for the number of exceedances ratio for every method. The lowest exceedances are depicted for the kCDF. This may be explained by the strategy of keeping parameters unchanged within the out-of-sample evaluations. Since the parameters are fixed, growing excess volatility, which is not observed in the in-sample part, may be captured better by the more data reliant methods such as the kCDF based RiskMetrics. Also for this setting additional gains are realized by the kernel functional form over the eCDF based estimations with the exceedances ratio are becoming more notable. At the 95% level, it can be observed that nonparametric estimators are again claiming more accurate evaluations, however, estimations for Turkey are still not adequately captured. L-GAS is the closest to claiming accurate evaluation of Turkish currency risk at the 95% risk level, but it struggles in providing independent exceedances as can be concluded from the gap between Kupiec (1995) and Christoffersen (1998) reported LR distances for this sample. From the observed distances for Turkey, nonparameteric estimators seem to be dealing with the repeating exceedances better than the L-GAS, but fail to meet the 5% significance threshold level for the overall number of expected violations. At the 90% level nonparametric estimators provide the most satisfactory results, generally supporting previously stated argument for these estimators appropriateness for the nonparametric dependence tests, but now under a more relaxed choice of parameters.

For the recursive parameters, updated with the last 1000 observations preceding each out-of-sample observation, in Table 3.4 the most notable VAR estimation gains can be noted by the parametric specifications. L-GAS, though still being the most conservative, provides the highest number of acceptable evaluations. RiskMetrics™ also improves its previous performance by achieving now two acceptable evaluations. t-GAS picks up its far from the best previous BRICTS performances with four correct VAR specifications. However, nonparametric estimators, for the parameters based on the RiskMetrics™ estimations, provide a mixed performance. eCDF EWMA performs the worst, providing only one correct specification, while kCDF clearly demonstrates improvement upon the number of accurate RiskMetrics™ VAR evaluations. Moreover, kCDF has a higher number of acceptable conditional coverage statistics than the t-GAS RiskMetrics, but most of the reported AE ratios are alarmingly higher than one and are typically higher than those for the fixed parameters in out-of-sample evaluations for this specification. In comparing output for both of the nonparametric approaches the benefit of the additional smoothing from the kernel functional form can now be argued to be clear. Also, for the "effective sample size" RiskMetrics™ choice of parameters and from the conditional coverage statistic, kernel smoothing prevents repeated violations occurring, but at the same time seems lacking information in the tails and therefore permits higher number of violations in general. This likely occurs due to the faster "learning rates" of the window RiskMetrics™ estimations. Though this interpration is obtained exploring reported unconditional coverage, conditional coverage and AE ratio results for the nonparametric methods, Taylor (2007) also reports similar conclusions experimenting with the exponential weighting nonparametric quantile regressions. Generally, similar to Taylor

(2007), it can be observed that for the adequate high quantile approximation, nonparametric methods require more information contained in, though old, but rare and thus still valuable, tail observations. Given relatively good results for these methods reported in Table 3.4 for the 90% and 95% levels and similar to Taylor (2007), it is observed that for these domains, observations are more often, and thus, there should be sufficient information under the faster discounting of older observations for adequate estimations. This also explains,the previously noted, higher values for the parameter governing the dynamics of the nonparametric methods in Table 3.2.

Overall, for the two out-of-sample BRICTS settings it is identified that the TRY/USD associated risks quantification is the most problematic for all methods in the pool. Certainly this series may be linked to the events which are "too uncertain" for any technical risk quantification tool to model, but it may also serve as an indication for the insufficiency of RiskMetrics weighting for similar risk exposure problems. It does not cancel out the other satisfactory results achieved for the L-GAS specification at the 99% and 95% levels as well as other good performances of the t-GAS and kCDF RiskMetrics variations, suggesting that RiskMetrics scheme may well be a valid benchmark to make comparisons with. From the observed success of Laplace distribution in the emerging markets samples a valid suggestion for this data may be a full-scale GAS specification under the Generalized Student's t distribution as in Harvey & Lange (2017). This specification allows the Generalized Error Distribution as its special case and therefore can take the forms of and close to the well performing Laplace distribution if necessary. This GAS parametrization is too technical for the parsimonious scope of this chapter, but it should address the observed Laplace conservatism concerns in VAR estimations, which in turn, is a likely factor for its success here.

Moreover, RiskMetrics tradition assumes zero mean of returns for simplicity of estimation and accounting for simple time-variation in the location/mean parameter may help achieving accurate risk projections for the discussed Turkish out-of-sample volatility.

3.7 Concluding remarks

In this chapter a small-scale stress-test for the up-to-date EWMA VAR type models was conducted. Firstly, from the obtained results it can be observed that parsimonious exponential decay type weighting may be still valid for VAR estimations, serving as a simple, yet competitive scheme for comparision with other more involved methods, however distributional specifications behind the estimation remain important. Considering the VAR estimation statistic output for all of the data samples and settings it may be problematic to pick up an ultimate RiskMetrics specification, since for the more narrow test setting but slightly more expansive data set used in Chapter 2 at the 99% confidence level t-GAS RiskMetrics provides the most accurate evaluations, while for the same level and under more involved backtesting procedure of the BRICTS data, L-GAS evaluations seem to be the most solid and overall appealing. Secondly, comparing all considered models' performance for each data set and test setting it is valuable to highlight the modest results for the kernel based RiskMetrics approach. Due to its nonparametric flexibility and parsimonious nature, as well as given obtained moderate results, this approach can be suggested as "the first threshold to pass" for other relevant methods in comprehesive and inclusive discussions as in Boucher et al. (2014) or for more explicintly linked to the financial regulation investigations as in McAleer et al. (2013) than RiskMetrics choices employed there or may be even than some of the more involved methods as in Lucas & Zhang (2016).

Generally this chapter aimed to avoid giving a specific and ultimate recommendation for such delicate field as risk modelling, however from the results obtained and if the RiskMetrics scheme is among the models for consideration, a valid strategy may be adopting Laplace and kernel specification for the 99% and 95% risk confidence levels respectively. The Laplace form may lead to higher capital reservations, but this can be viewed as the modelling scheme's simplicity costs. On the other hand, kernel specification may be less comfortable at the higher confidence levels due to the lack of the data in these domains, but at the 95% threshold, it seems benefiting from its data driven functionality providing arguably consistent and good results.

Overall, this chapter unpacks the main practical limitation of the kernel EWMA VAR scheme. The approach is sensitive to the optimal combination of parameters, which for the daily rolling forecasts, can theoretically be identified using robust and simple recursive least squares appoach, but is poorly justified from the applied perspective even for the computional time with the binned estimators as in Chapter 2. Here, time efficiency is acheived with the ad hoc procedure which references to the RiskMetrics™ parameters. This certainly simplified estimation as well as demonstrated that the kernel functional form is able to provide an additional "cushion" for improvement over the dynamic eCDF and RiskMetrics™ results, but obtained number of violations was not attractive for the risk modelling context. Therefore, future research may look into the fast but robust settings of the parameters estimation for this method in more detail. For now, in the upcoming VAR studies involving dynamic kernel approach, it may be worthwhile to perform parameter updating less often, since for example, Ardia & Hoogerheide (2014) argue that results for quarterly parameters updating should be close or equally identicall to the daily updates used here. As for the contexts beyond VAR applications, obtained results can serve as additional evidence for these dynamic nonparameteric methods validity for the fully nonparametric changing dependence tests and explorations similar to Busetti & Harvey (2010) and Harvey (2010) and for the full-scale dynamic distribution/density semiparametric estimations similar to the MacDonald et al. (2011) for the independent and identically distributed modelling framework.

Chapter 4

Conclusions and further work

4.1 Concluding summary

This thesis contributes to the field of financial economics by: considering six new parametric distribution models for financial returns on the basis of the Student's t distribution in Chapter 1; evaluating, analysing and comparing in-sample forecasting performance of the nonparametric methods for financial returns' distributions modelling in the time series context of the exponentially declining weights under the maximum likelihood and several least-squares routines for estimation of the unknown parameters in Chapter 2; and empirically testing a forecasting ability of the Value-at-Risk measure by the methods falling in the parsimonious "exponential weighted moving average" modelling category in Chapter 3.

Chapter 1 extends portfolio of distributions models which can be utilized to model financial data in the relevant applied tasks in finance and economics. Models in Chapter 1 are fitted to the data set of daily financial returns for stock indices, energy commodities and cryptocurrencies and challenge Generalized Hyperbolic distribution as in McNeil et al. (2005)

among others. Empirical results based on the log-likelihood criteria analysis suggest overall appropriateness of the distribution models developed in the first chapter, when compared to the Generalized Hyperbolic model. For the models yielding the best likelihood criteria output for each individual sample in the considered data set basic quantiles and Probability Integral Transformations plots are provided and analysed. From the conducted empirical analysis it follows that best performing distribution models fit financial returns in the data set relatively well, though there is evidence that these models still, similar to other parametric distribution models in general, may struggle approximating domains in the center of financial returns.

Chapter 2 investigates in-sample forecasting performance of the exponentially weighted kernel estimators for density and distributions of financial returns under different choices of estimation parameters. It challenges suggestion of Harvey & Oryshchenko (2012) for employing maximum-likelihood for estimation of the unknown parameters for nonparametric density forecasts. Chapter 2 advocates for the computationally more conservative than maximum likelihood, least-square routines, which can be also re-written to accommodate dynamic binned kernel estimators for faster optimizations or varying bandwidths over the range of financial returns. Parameters for the estimation techniques in Chapter 2 are obtained for daily financial returns of stock indices, commodities, currencies and cryptocurrencies exchange rates. Empirical results based on the nonparametric and likelihood ratio tests suggest that parameters by least-squares estimations lead to notably more accurate in-sample forecasts than estimations relying on the parameters chosen by maximum likelihood, including binned variations with similar evaluation time as for maximum likelihood, though considered adaptive strategy may struggle with fast evolving series of financial returns.

Chapter 3 positions kernel based exponential weighted moving average technique within the group of relevant methods for Value-at-Risk estimations. Value-at-Risk techniques in Chapter 3 are fitted to the data set of daily financial returns of stock indices, commodities, currencies and cryptocurrencies exchange rates. Time evolving nonparametric estimators in Chapter 3 challenge approaches presented in Lucas & Zhang (2016). Empirical results based on the likelihood ratio tests for in-sample and out-of-sample estimations suggest a notably modest and stable performance of the kernel technique under the diligent choice of parameters, when compared to other competing methods. It provides estimation gains over its simplified empirical alternative and is found to be a good competitor to the considered parametric models, including the domains in the tails of financial returns, though in-sample estimation gains over the empirical approach may be smaller, while out-of-sample estimations may be still computationally demanding and therefore less appealing than its parametric alternatives for applied tasks.

4.2 Suggestions for further work

There are several directions to expand work conducted in Chapters 1, 2 and 3. On the one hand, portfolio of distributions in Chapter 1 may be expanded to fully match the list of compound distributions provided in Nadarajah (2012). On the other hand, models in Chapter 1 may be more intriguing to test in the time-series GARCH and GAS contexts for Value-at-Risk and/or Expected Shortfall forecasting experiments similar to the competition conducted in Chapter 3 for the dynamic nonparametric estimator as in Harvey & Oryshchenko (2012). In addition, it may be valuable introducing asymmetry to the compound distribution models in Chapter 1 to achieve better modelling outcomes. Asymmetric extensions may be then also considered in the time-series context and included in the empiric Value-at-Risk testing.

Chapter 2 can be extended by considering methods for automatic choice of the bandwidth parameters. Similar to the least-squares techniques, approaches provided in Sheather & Jones (1991), Altman & Leger (1995) and Polansky & Baker (2000) can be modified to accommodate exponentially weighted kernel density and distribution estimators with a reference to the parametric models driven by the similar exponential weights outlined in Lucas & Zhang (2016). For example, a robust Laplace scheme of Guermat & Harris (2002) may be adopted for approximation of the DGP to vary the bandwidths parameter over the range of financial returns and provide more time-effective estimations than with the least-squares routines and binned estimators in Chapter 2. This research extension path shall make estimators in Harvey & Oryshchenko (2012) more attractive for practitioners given that automatic approaches deliver robust parameters for estimations. An alternative extension can combine parametric Extreme-Value-Theory tails of Zhao et al. (2018) with estimators in Harvey & Oryshchenko (2012). Model of Zhao et al. (2018) may be modified to match the dynamics of exponential weights and yield a mixed frequency (e.g. intra-day and daily frequencies combination) semiparametric model for financial returns. On the other hand, estimators in Harvey & Oryshchenko (2012) may accommodate weights commonly used for intra-day financial data as in Mixed Data Sampling models of Gorgi et al. (2019) among others to yield a semiparametric density and distribution model for intra-day data. This also highlights the potential of further empirical applications to the financial data of different frequencies and alternative weighting schemes for Harvey & Oryshchenko's (2012) estimators.

Chapter 3 can be extended by considering a wider pool of the competing models for Value-at-Risk forecasting. There is a set of parametric and nonparametric quantile regressions driven by the exponential weights as in Taylor (2007) and Laporta et al. (2018) among others to be considered in the further work. In addition, Expected Shortfall estimations and forecasting with the list of approaches from Chapter 3 may be worthwhile considering as well as more elaborate and inclusive testing for Value-at-Risk forecasting accuracy can be adopted in future studies. For example, Model Confidence Set inclusive testing procedure of P. R. Hansen et al. (2011) is a good candidate for consideration. High-frequency data contexts are also research stimulating and given satisfactory backtesting results further high-frequency risk spillover investigations can be conducted to upgrade daily frequency risk spillover investigations of Corsi et al. (2018) and Peng et al. (2018) among others.

Appendix A

PDF and CDF functions in Chapter 1 may take more elegant forms for estimations with special hypergeometric functions. For example, from Prudnikov et al. (1986), PDF in (1.14) can be given by

$$f(x) = \frac{\alpha\nu[\beta\nu]^{\frac{\nu}{2}}\Gamma(\frac{\nu+1}{2})\Gamma(\alpha-\frac{\nu}{2})}{2\Gamma(1+\alpha)x^{\nu+1}\sqrt{\pi}} {}_3F_0\left(\frac{\nu+1}{2}, \alpha-\frac{\nu}{2}; 1+\frac{\nu}{2};; \frac{\beta\nu}{x^2}\right), \tag{4.1}$$

while a corresponding CDF outlined by

$$F(x) = \frac{1}{2} - \frac{\alpha[\nu\beta]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\alpha-\frac{\nu}{2}\right)}{\Gamma(\alpha+1)2x^{\nu}\sqrt{\pi}} {}_3F_0\left(\frac{1+\nu}{2}, \frac{\nu}{2}, \frac{\nu}{2}-\alpha+1;; -\frac{\nu\beta}{x^2}\right),$$

and PDF in (1.21) expressed with

$$f(x) = \frac{[\beta\nu]^{\frac{\nu}{2}}\Gamma(\frac{\nu+1}{2})\Gamma(\alpha-\frac{\nu}{2})}{\Gamma(\alpha)x^{\nu+1}\sqrt{\pi}\Gamma(\frac{\nu}{2})} {}_2F_0\left(\frac{\nu+1}{2}, \frac{\nu}{2}-\alpha+1;; \frac{\beta\nu}{x^2}\right), \tag{4.2}$$

while its CDF outlined by

$$F(x) = \frac{1}{2} - \frac{[\nu\beta]^{\frac{\nu}{2}}\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\alpha-\frac{\nu}{2}\right)}{\nu\Gamma(\alpha)x^{\nu}\sqrt{\pi}\Gamma(\frac{\nu}{2})} {}_3F_1\left(\frac{1+\nu}{2}, \frac{\nu}{2}, \frac{\nu}{2}-\alpha+1; \left(1+\frac{\nu}{2}\right)_k; -\frac{\nu\beta}{x^2}\right).$$

Expressions in (1.21) and (1.22) may be more straightforward at the implementation stage within R computing enviroment under the hypergeo package functionality of Hankin (2016). PDFs in (4.1) and(4.2) lead to the following log-likelihoods

$$\begin{aligned}\log L(\nu,\alpha,\beta) = \quad & n\log(\alpha\nu) + \frac{n\nu}{2}\log(\nu\beta) + n\log\Gamma\left(\frac{v+1}{2}\right) + n\log\Gamma(\alpha-\nu/2) - \\ & -n\log(2\sqrt{\pi}) - (\nu+1)\sum_{i=1}^{n}\log x_i - n\log\Gamma(1+\alpha) + \\ & + \sum_{i=1}^{n}\log\ {}_3F_0\left(\frac{\nu+1}{2};\alpha-\frac{\nu}{2},1+\frac{\nu}{2};\frac{\beta\nu}{x_i^2}\right)\end{aligned} \tag{4.3}$$

and

$$\begin{aligned}\log L(\nu,\alpha,\beta) = \quad & \frac{n\nu}{2}\log(\nu\beta) + n\log\Gamma\left(\frac{v+1}{2}\right) + n\log\Gamma(\alpha-\nu/2) - \frac{n}{2}\log\pi - \\ & -(\nu+1)\sum_{i=1}^{n}\log x_i - n\log\Gamma(\alpha) - n\log\Gamma\left(\frac{\nu}{2}\right) + \\ & + \sum_{i=1}^{n}\log\ {}_2F_0\left(\frac{\nu+1}{2};\frac{\nu}{2}-\alpha+1;\frac{\beta\nu}{x_i^2}\right)\end{aligned} \tag{4.4}$$

for estimations respectively.

Appendix B

Here, Chapter 2 raw and ARMA & GARCH filtered log-returns PITs histograms density evaluations for different sets of parameters reported in Tables 2.3, 2.4 and 2.5 are provided. Figures below allow visually assessing uniformity characteristics of the obtained PITs. Plots allowing assessment of the PITs' dependency characteristics are available upon request.

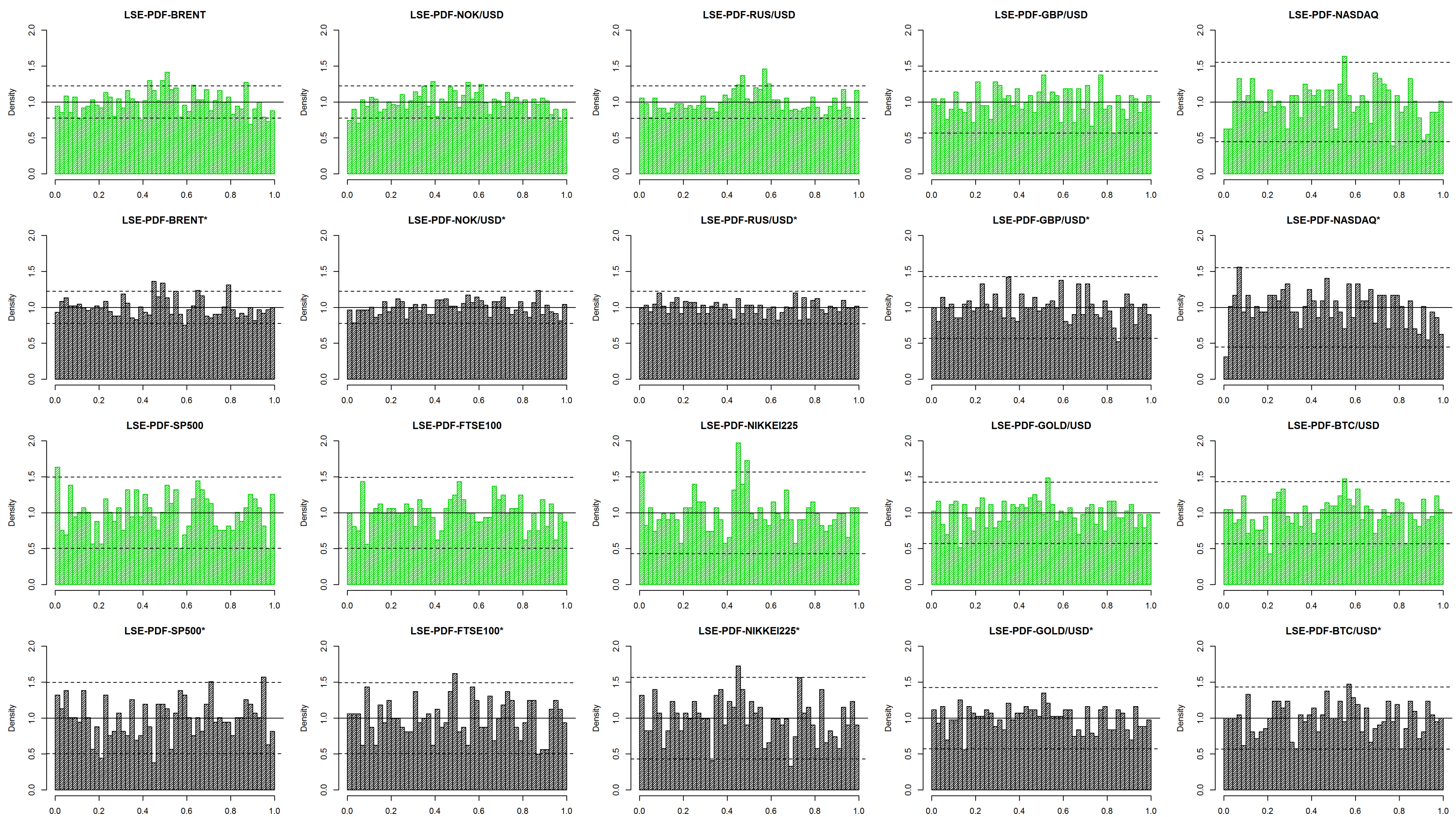

Figure 5.1: Chapter 2 log-returns data set histograms of PITs for parameters by LSE for PDF. * indicates PITs for ARMA & GARCH filtered returns. Lines parallel to the horizontal axis show $+/-$ 2 standard deviations confidence interval for PITs uniformity; $+/-2\sqrt{(\kappa-1)T^{-1}}$, where $\kappa = 50$ and is the number of bins used in histogram density evaluation as per Diebold et al. (1998).

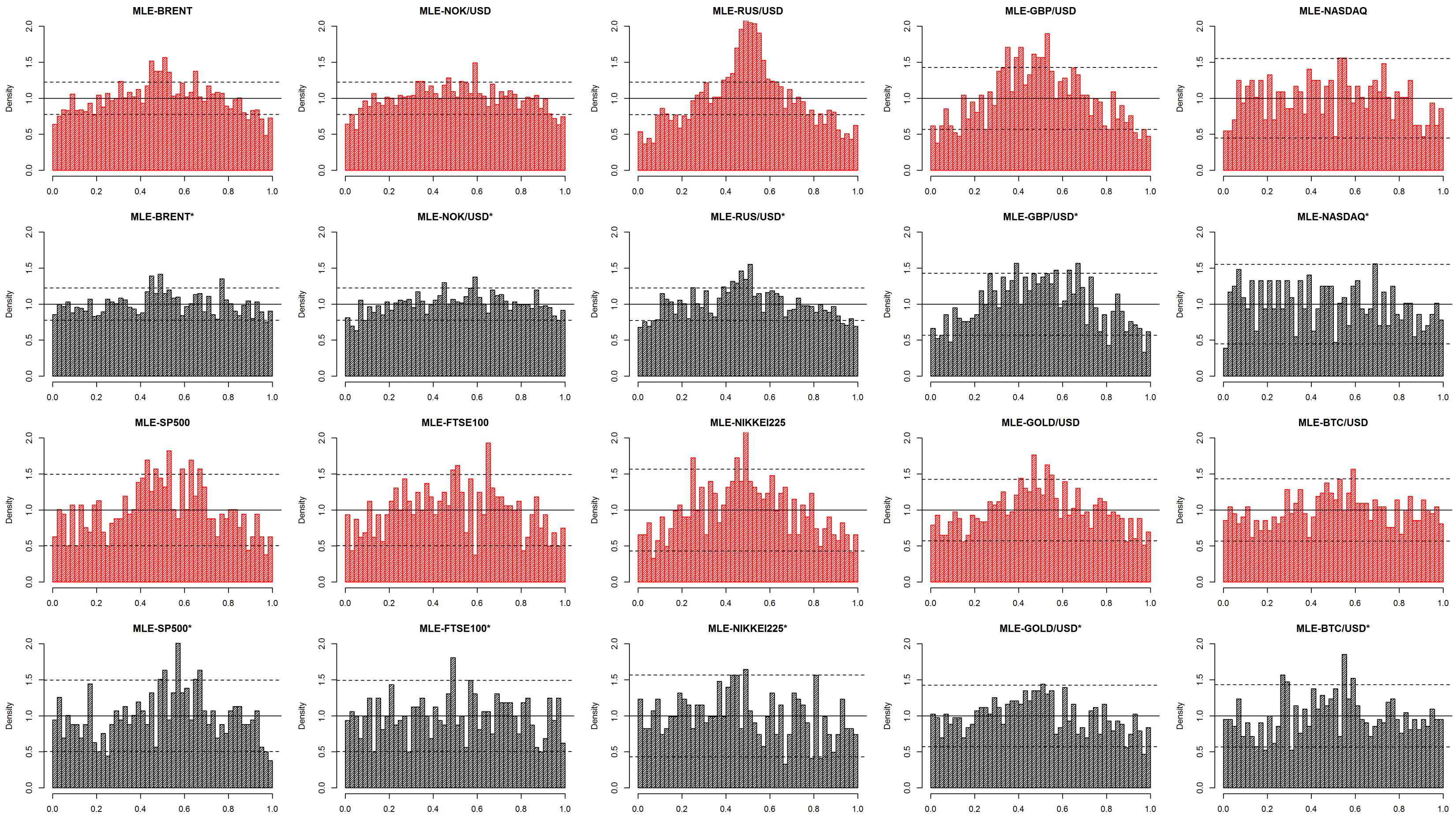

Figure 5.2: Chapter 2 log-returns data set histograms of PITs for parameters by MLE.

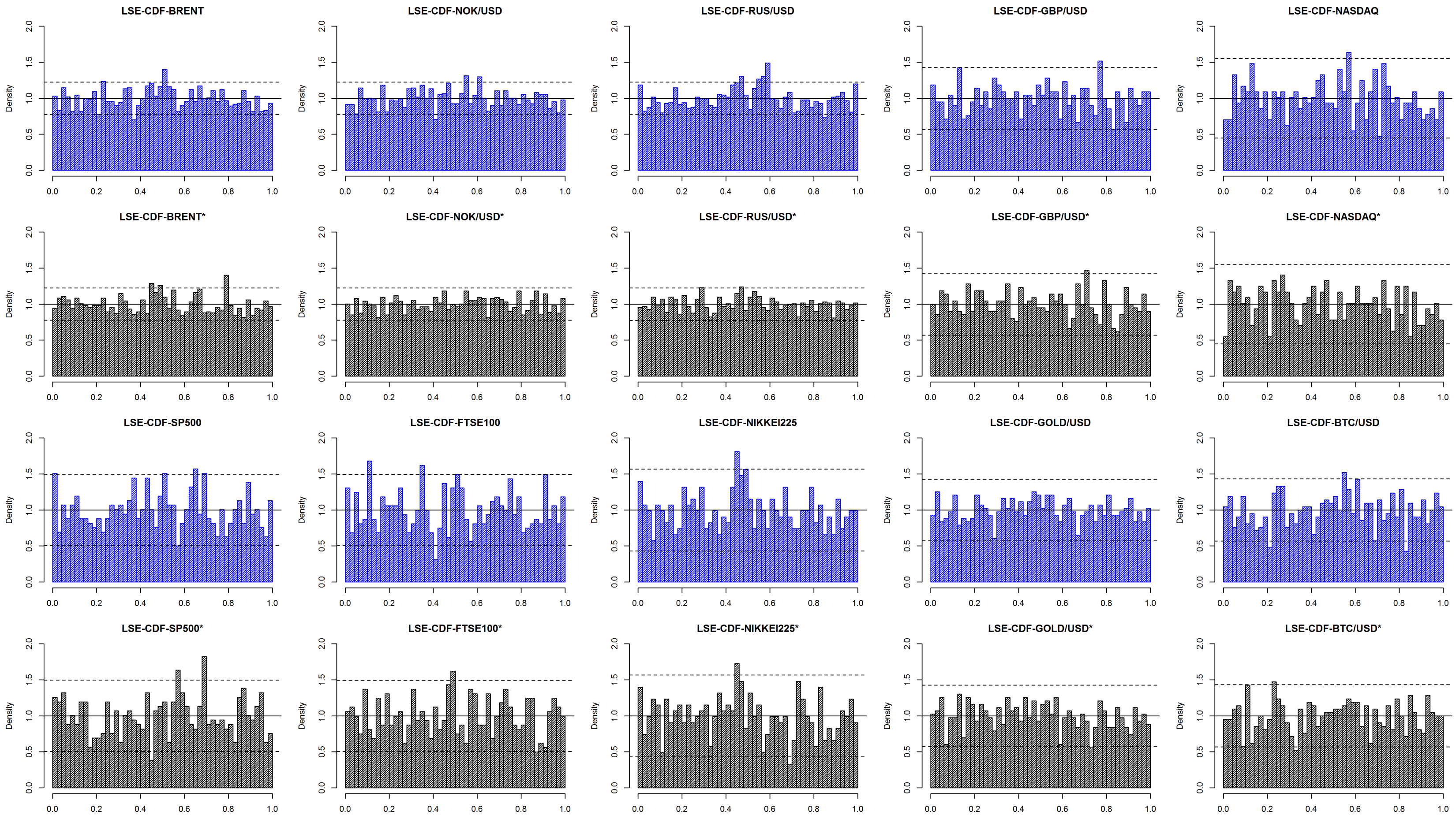

Figure 5.3: Chapter 2 log-returns data set histograms of PITs for parameters by LSE for CDF.

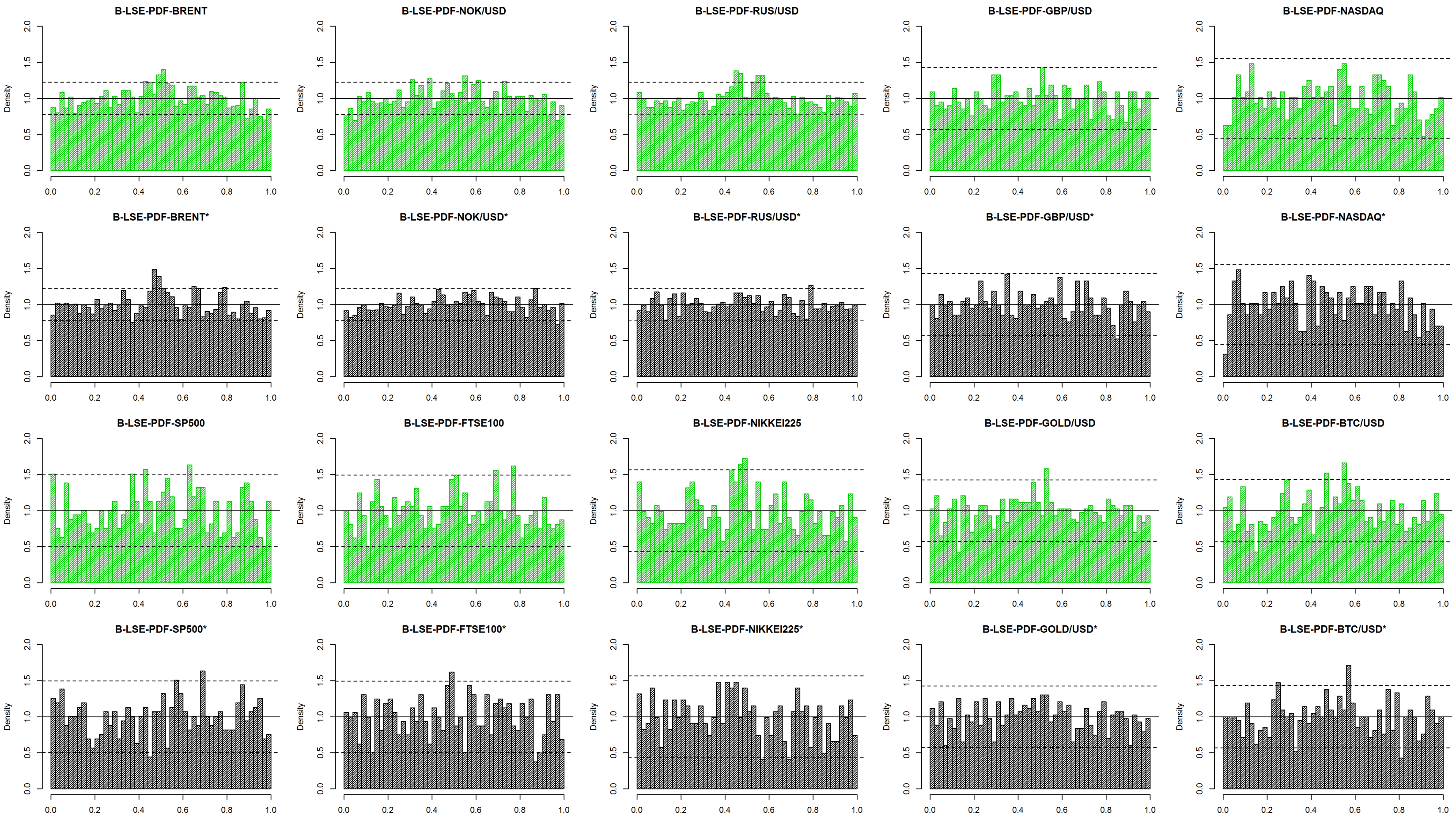

Figure 5.4: Chapter 2 log-returns data set histograms of PITs for parameters by binned LSE for PDF.

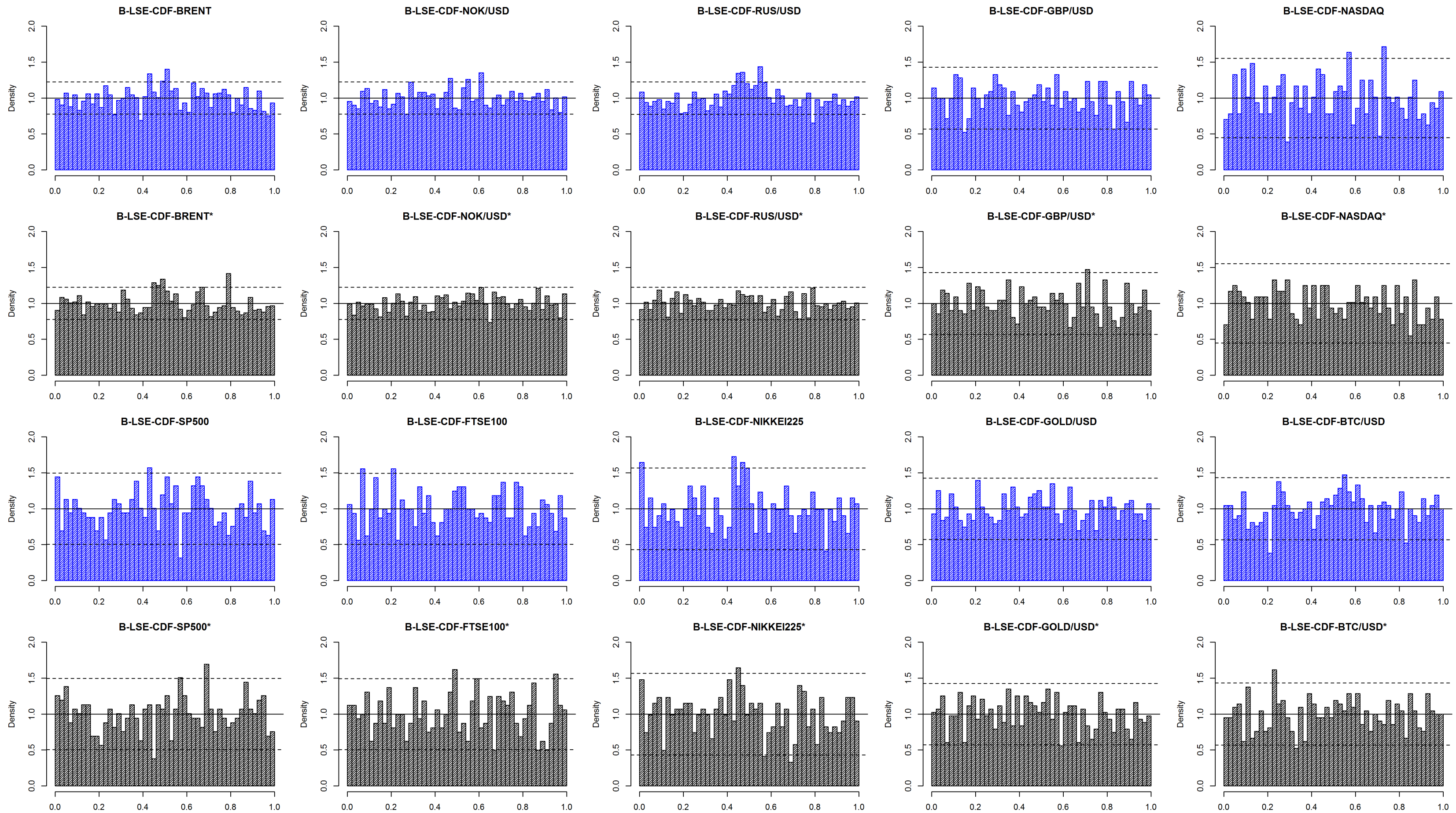

Figure 5.5: Chapter 2 log-returns data set histograms of PITs for parameters by binned LSE for CDF.

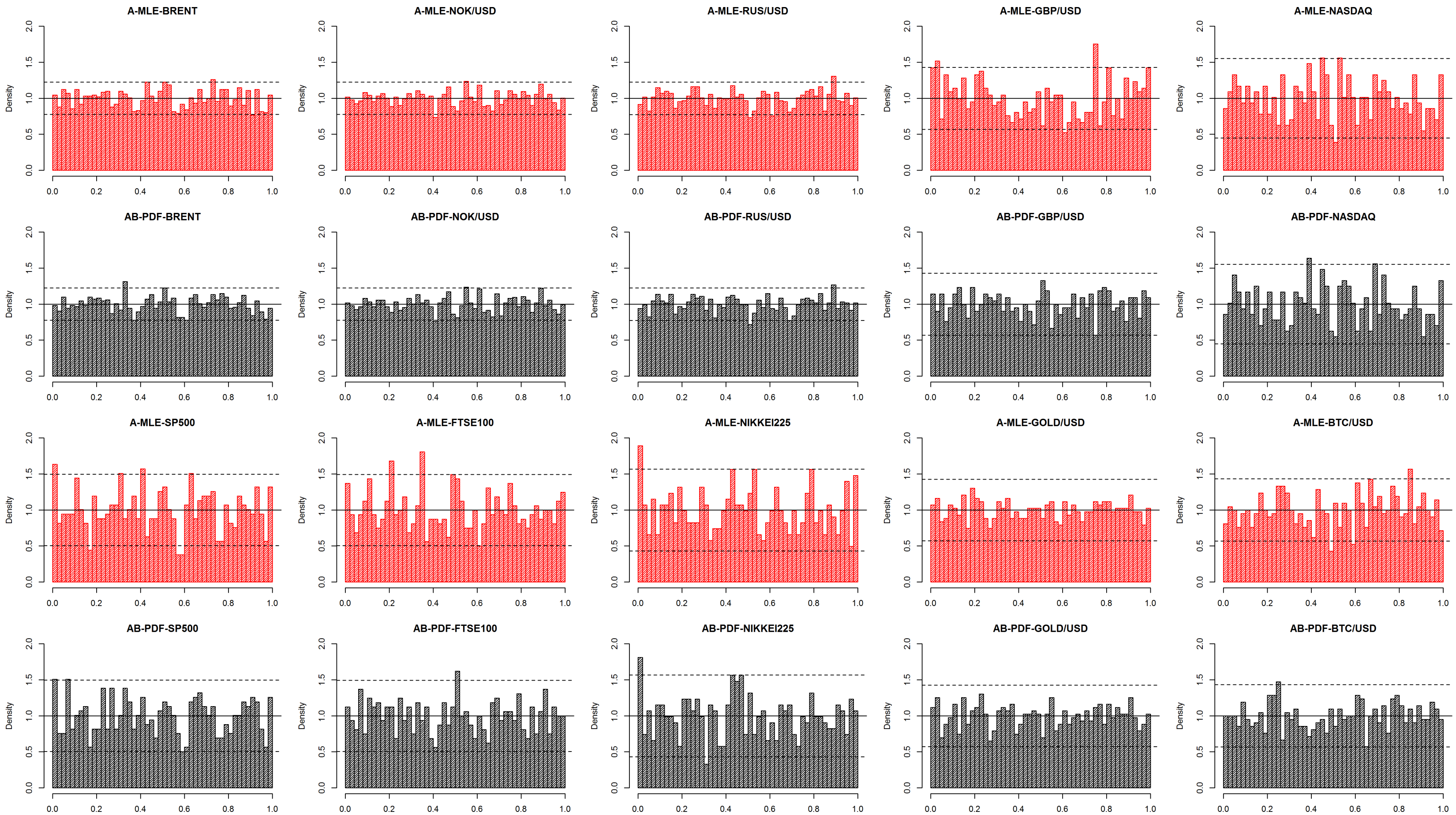

Figure 5.6: Chapter 2 log-returns data set adaptive histograms of PITs for MLE and binned LSE for CDF ω parameters.

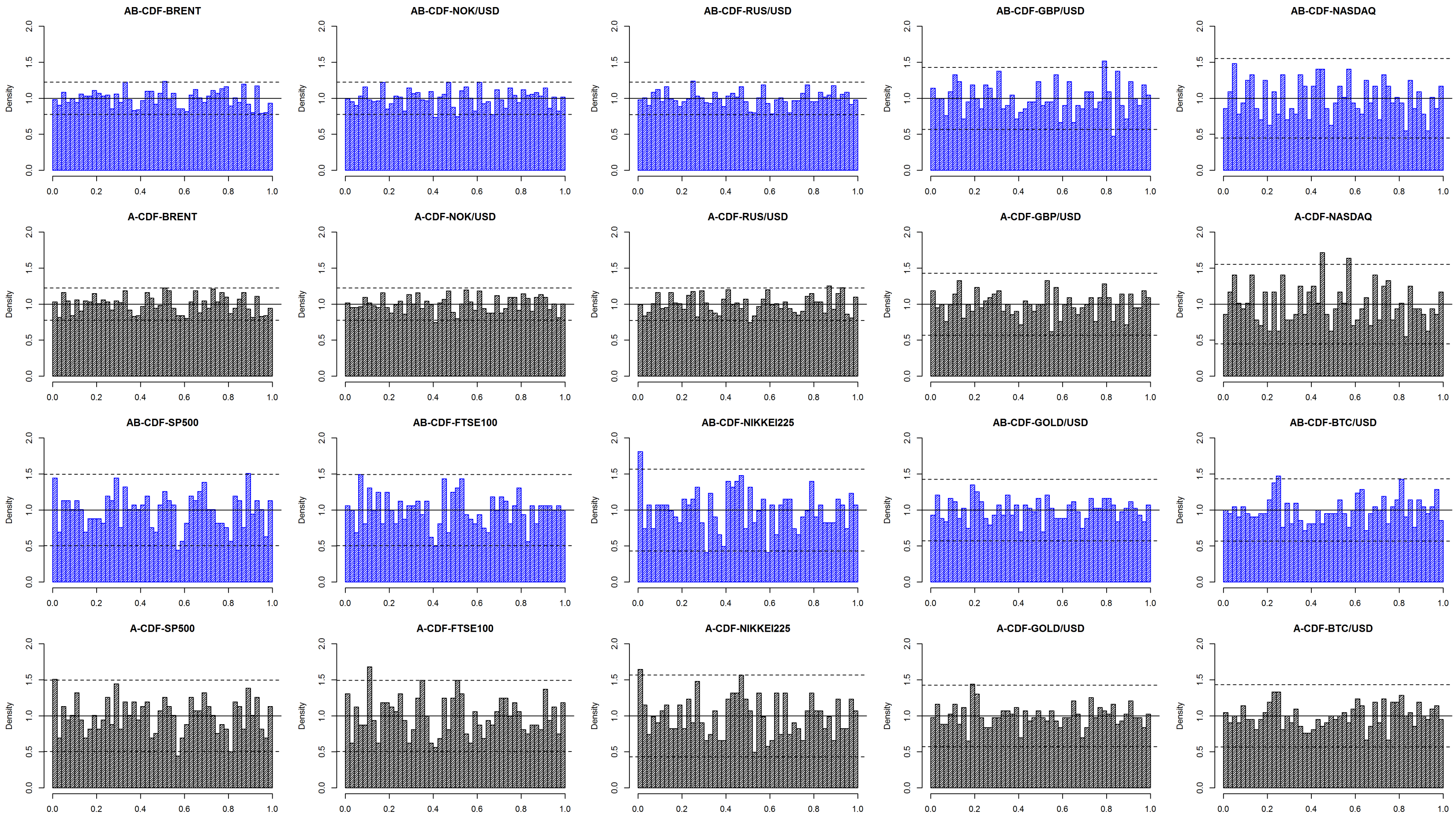

Figure 5.7: Chapter 2 log-returns data set adaptive histograms of PITs for binned and full-scale LSE for CDF ω parameters.

References

Aas, K., & Haff, I. H. (2006). The generalized hyperbolic skew Student's t distribution. *Journal of Financial Econometrics*, *4*(2), 275–309.

Abad, P., Benito, S., & López, C. (2014). A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*, *12*(1), 15–32.

Abramson, I. S. (1982). On bandwidth variation in kernel estimates - a square root law. *The Annals of Statistics*, *10*(4), 1217–1223.

Afuecheta, E., Chan, S., & Nadarajah, S. (2018). Flexible models for stock returns based on Student's t distribution. *The Manchester School, forthcoming*.

Afuecheta, E., Pérez Ruiz, D. A., Utazi, C., & Nwosu, C. (2016). On the flexibility of GARCH-family models with an application to the BRICS stock indices. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *2*(1-2), 44–77.

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.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trransactions on Automatic Control*, *19*(6), 716–723.

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

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of Economic Forecasting*, *1*, 777–878.

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(1), 99–102.

Ardia, D., Bluteau, K., Boudt, K., & Catania, L. (2018). Forecasting risk with Markov-switching GARCH models: a large-scale performance study. *International Journal of Forecasting*, *34*, 733-747.

Ardia, D., Boudt, K., & Catania, L. (2016). Value-at-Risk prediction in R with the GAS package. *R Journal*, *forthcoming*.

Ardia, D., Boudt, K., & Catania, L. (2018). Generalized autoregressive score models in R: the GAS package. *Journal of Statistical Software*, *forthcoming*.

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.

Ardia, D., Hoogerheide, L. F., & van Dijk, H. K. (2009a). Adaptive mixture of Student's t distribution as a flexible candidate distribution for efficient simulation: the R package AdMit. *Journal of Statistical Software*, *29*(3), 1–32.

Ardia, D., Hoogerheide, L. F., & van Dijk, H. K. (2009b). AdMit: adaptive mixtures of Student's t distributions. *The R Journal*, *1*(EI-1516), 25–30.

Arnold, B. C. (1983). *Pareto Distributions.* International Cooperative Publishing House, Fairland, MD.

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

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *353*(1674), 401–419.

Barndorff-Nielsen, O., Kent, J., & Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review/Revue Internationale de Statistique*, 145–159.

Bauwens, L., & Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics*, *23*(3), 346–354.

Beals, R., & Wong, R. (2010). *Special functions: a graduate text* (Vol. 126). Cambridge University Press.

Benckert, L.-G., & Jung, J. (1974). Statistical models of claim distributions in fire insurance. *ASTIN Bulletin: The Journal of the IAA*, *8*(1), 1–25.

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.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160.

Blattberg, R. C., & Gonedes, N. J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *The Journal of Business*, *47*, 244–280.

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

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. *Handbook of Econometrics*, *4*, 2959–3038.

Borchers, H. W. (2018). pracma: Practical Numerical Math Functions. [Computer software manual]. (R package version 2.1.4)

Bouchaud, J.-P., & Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management.* Cambridge University Press.

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*, *85*(4), 799–808.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370.

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.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

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

Chan, S., Chu, J., Nadarajah, S., & Osterrieder, J. (2017). A statistical analysis of cryptocurrencies. *Journal of Risk and Financial Management*, *10*(2), 12.

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

Chen, L., Zerilli, P., & Baum, C. F. (2018). Leverage effects and stochastic volatility in spot oil returns: a Bayesian approach with VaR and CVaR applications. *Energy Economics, forthcoming*.

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.

Chou, C.-Y., & Liu, H.-R. (1998). Properties of the half-normal distribution and its application to quality control. *Journal of Industrial Technology*, *14*(3), 4–7.

Chow, Y.-S., Geman, S., & Wu, L.-D. (1983). Consistent cross-validated density estimation. *The Annals of Statistics*, *11*(1), 25–38.

Choy, S. B., & Chan, J. S. (2008). Scale mixtures distributions in statistical modelling. *Australian & New Zealand Journal of Statistics*, *50*(2), 135–146.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841–862.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223-226.

Cooray, K., & Ananda, M. M. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics—Theory and Methods*, *37*(9), 1323–1337.

Corlu, C. G., & Corlu, A. (2015). Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance*, *15*(11), 1851–1864.

Corsi, F., Lillo, F., Pirino, D., & Trapin, L. (2018). Measuring the propagation of financial distress with granger-causality tail risk networks. *Journal of Financial Stability*, *38*, 18–36.

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

Danielsson, J., James, K. R., Valenzuela, M., & Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, *23*, 79–91.

De Rossi, G., & Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, *152*(2), 179–185.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

Dobzhansky, T., & Wright, S. (1947). Genetics of natural populations. XV. Rate of diffusion of a mutant gene through a population of Drosophila pseudoobscura. *Genetics*, *32*(3), 303.

Dupuis, D. J., Papageorgiou, N., & Rémillard, B. (2014). Robust conditional variance and Value-at-Risk estimation. *Journal of Financial Econometrics*, *13*(4), 896–921.

Embrechts, P., & Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, *1*, 493–514.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling Extremal Events: For Insurance and Finance* (Vol. 33). Springer Science & Business Media.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Fang, Y. (2011). Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. *Journal of Data Science*, *9*(1), 15–21.

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

Fernández, C., & Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, *93*(441), 359–371.

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

Gardner, E. S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, *4*(1), 1–28.

Gardner, E. S. (2006). Exponential smoothing: the state of the art — Part II. *International Journal of Forecasting*, *22*(4), 637–666.

Gencay, R., & Selcuk, F. (2004). Extreme value theory and Value-at-Risk: relative performance in emerging markets. *International Journal of Forecasting*, *20*(2), 287–303.

Gerlach, R., Lu, Z., & Huang, H. (2013). Exponentially smoothing the skewed Laplace distribution for Value-at-Risk forecasting. *Journal of Forecasting*, *32*(6), 534–550.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.

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

Gijbels, I., Pope, A., & Wand, M. P. (1999). Understanding exponential smoothing via kernel regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(1), 39–50.

Gorgi, P., Koopman, S. J., & Li, M. (2019). Forecasting economic time series using score-driven dynamic models with mixed-data sampling. *International Journal of Forecasting*, *35*(4), 1735–1747.

Gosset, W. S. (1908). The probable error of a mean. *Biometrika*, *6*, 1–25.

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

Gradshteyn, I. S., & Ryzhik, I. M. (2014). *Table of integrals, series, and products.* Academic Press.

Gregoriou, A. (2018). Cryptocurrencies and asset pricing. *Applied Economics Letters*, *26*(12), 1–4.

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

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., & Patil, P. (1994). On the efficiency of on-line density estimators. *IEEE Transactions on Information Theory*, *40*(5), 1504–1512.

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

Hankin, K. S., Robin. (2016). hypergeo: the Gauss hypergeometric function. [Computer software manual]. (R package version 1.2-13)

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*(2), 190–195.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*(3), 705–730.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Harrell, F. E. (2010). *Regression Modeling Strategies.* New York: Springer.

Hart, J. D., & Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *The Annals of Statistics*, *18*(2), 873–890.

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. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series* (Vol. 52). Cambridge University Press.

Harvey, A., & Lange, R.-J. (2017). Volatility modeling with a Generalized t distribution. *Journal of Time Series Analysis*, *38*(2), 175–190.

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.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163–1174.

Hill, B. (2010). On tail index estimation for dependent, heterogeneous data. *Econometric Theory*, *26*(5), 1398–1436.

Hill, B. (2015). Robust estimation and inference for heavy tailed GARCH. *Bernoulli*, *21*(3), 1629–1669.

Hogg, R. V., & Klugman, S. A. (1983). On the estimation of long tailed skewed distributions with actuarial applications. *Journal of Econometrics*, *23*(1), 91–102.

Hoogerheide, L. F., Kaashoek, J. F., & Van Dijk, H. K. (2007). On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks. *Journal of Econometrics*, *139*(1), 154–180.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Science & Business Media.

Ji, Q., Bouri, E., Roubaud, D., & Shahzad, S. J. H. (2018). Risk spillover between energy and agricultural commodity markets: a dependence-switching CoVaR-copula model. *Energy Economics*, *75*, 14–27.

Jones, M., & Faddy, M. (2003). A skew extension of the t distribution with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 159–174.

Jones, M., & 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.

Kilber, C., & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Science*. Willey, New York.

Kon, S. J. (1984). Models of stock returns — a comparison. *The Journal of Finance*, *39*(1), 147–165.

Koopman, S. J., & Harvey, A. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, *27*(7), 1317–1333.

Kotz, S., & Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press.

Kristensen, D., & Linton, O. (2004). Consistent standard errors for target variance approach to GARCH estimation. *Econometric Theory*, *20*(5), 990–993.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*(2), 73–84.

Laporta, A. G., Merlo, L., & Petrella, L. (2018). Selection of Value at Risk models for energy commodities. *Energy Economics*, *74*, 628–643.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons, New York.

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.

Li, R., & Nadarajah, S. (2017). A review of Student's t distribution and its generalizations. *Empirical Economics*, 1–30. doi: 10.1007/s00181-018-1570-0

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293–311.

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

Luciano, E., & Semeraro, P. (2010). A generalized normal mean-variance mixture for return processes in finance. *International Journal of Theoretical and Applied Finance*, *13*(03), 415–440.

Luethi, D., & Breymann, W. (2016). ghyp: a package on Generalized Hyperbolic distribution and its special cases [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=ghyp` (R package version 1.5.7)

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.

Madan, D. B., Carr, P. P., & Chang, E. C. (1998). The variance gamma process and option pricing. *Review of Finance*, *2*(1), 79–105.

Madan, D. B., & Seneta, E. (1990). The variance gamma (VG) model for share market returns. *Journal of Business*, *63*(4), 511–524.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

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

McAleer, M. (2009). The Ten commandments for optimizing Value-at-Risk and daily capital charges. *Journal of Economic Surveys*, *23*(5), 831–849.

McAleer, M., Jiménez-Martín, J.-Á., & Pérez-Amaral, T. (2013). Has the Basel accord improved risk management during the global financial crisis? *The North American Journal of Economics and Finance*, *26*, 250–265.

McDonald, J. B., & Butler, R. J. (1987). Some generalized mixture distributions with an application to unemployment duration. *The Review of Economics and Statistics*, *69*(2), 232–240.

McMillan, D. G., & Kambouroudis, D. (2009). Are RiskMetrics forecasts good enough? Evidence from 31 stock markets. *International Review of Financial Analysis*, *18*(3), 117–124.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton university press.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, *18*(2), 435–444.

Nadarajah, S. (2012). Models for stock returns. *Quantitative Finance*, *12*(3), 411–424.

Nadarajah, S., Afuecheta, E., & Chan, S. (2013). A double generalized Pareto distribution. *Statistics & Probability Letters*, *83*(12), 2656–2663.

Nadarajah, S., Afuecheta, E., & Chan, S. (2015). A note on "Modelling exchange rate returns: which flexible distribution to use?". *Quantitative Finance*, *15*(11), 1777–1785.

Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory of Probability & Its Applications*, *9*(3), 497–500.

Nair, N., & Hitha, N. (1990). Characterizations of Pareto and related distributions. *Journal of the Indian Statistical Association*, *28*, 75–79.

Nelson, D. B. (1990). Stationarity and persistence in the garch (1, 1) model. *Econometric theory*, *6*(3), 318–334.

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

Nikoloulopoulos, A. K., Joe, H., & Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, *56*(11), 3659–3673.

Pafka, S., & Kondor, I. (2001). Evaluating the RiskMetrics methodology in measuring volatility and Value-at-Risk in financial markets. *Physica A: Statistical Mechanics and Its Applications*, *299*(1-2), 305–310.

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, *3*(1), 15–102.

Papastathopoulos, I., & Tawn, J. A. (2013a). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, *143*(1), 131–143.

Papastathopoulos, I., & Tawn, J. A. (2013b). A generalised Student's *t* distribution. *Statistics & Probability Letters*, *83*(1), 70–77.

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

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

Peng, C., Zhu, H., Guo, Y., & Chen, X. (2018). Risk spillover of international crude oil to China's firms: evidence from Granger causality across quantile. *Energy Economics*, *72*, 188–199.

Perez-Quiros, G., & Timmermann, A. (2001). Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities. *Journal of Econometrics*, *103*(1-2), 259–306.

Pérignon, C., Deng, Z. Y., & Wang, Z. J. (2008). Do banks overstate their Value-at-Risk? *Journal of Banking & Finance*, *32*(5), 783–794.

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 plug—in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, *65*(1-4), 63–80.

Praetz, P. D. (1972). The distribution of share price changes. *Journal of Business*, *45*(1), 49–55.

Press, S. J. (1967). A compound events model for security prices. *Journal of Business*, *40*(3), 317–335.

Prudnikov, P., A, Brychkov, A., Y, & Marichev, I., O. (1986). *Integrals and Series* (Vol. 1). Gordon and Breach Science Publishers, Amsterdam.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org`

Reboredo, J. C., & Ugolini, A. (2016). Quantile dependence of oil price movements and stock returns. *Energy Economics*, *54*, 33–49.

Revankar, N. S., Hartley, M. J., Pagano, M., et al. (1974). A characterization of the Pareto distribution. *The Annals of Statistics*, *2*(3), 599–601.

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

Rossignolo, A. F., Fethi, M. D., & Shaban, M. (2012). Value-at-Risk models and Basel capital charges: evidence from emerging and frontier stock markets. *Journal of Financial Stability*, *8*(4), 303–319.

Rossignolo, A. F., Fethi, M. D., & Shaban, M. (2013). Market crises and Basel capital requirements: could Basel III have been different? Evidence from Portugal, Ireland, Greece and Spain (PIGS). *Journal of Banking & Finance*, *37*(5), 1323–1339.

Sahu, S. K., Dey, D. K., & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, *31*(2), 129–150.

Sain, S. R. (2002). Multivariate locally adaptive density estimation. *Computational Statistics & Data Analysis*, *39*(2), 165–186.

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).

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

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., & Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, *53*(3), 683–690.

Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, *25*(2), 197–231.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis* (Vol. 26). CRC Press.

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.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, *44*(12-part-1), 1650–1661.

Tjetjep, A., & Seneta, E. (2006). Skewed normal variance-mean models for asset pricing and the method of moments. *International Statistical Review*, *74*(1), 109–126.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer Series in Statistics. Springer, New York.

Tucker, A. L. (1992). A reexamination of finite- and infinite-variance distributions as models of daily stock returns. *Journal of Business & Economic Statistics*, *10*(1), 73–81.

Varma, J. R. (2008). *Derivatives and Risk Management.* New Delhi, India: Tata McGraw-Hill.

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

Wang, J. J., Chan, J. S., & Choy, S. B. (2011). Stochastic volatility models with leverage and heavy-tailed distributions: a Bayesian approach using scale mixtures. *Computational Statistics & Data Analysis*, *55*(1), 852–862.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, *74*(3), 646–648.

Wichitaksorn, N., Wang, J. J., Choy, S. B., & Gerlach, R. (2015). Analyzing return asymmetry and quantiles through stochastic volatility models using asymmetric Laplace error via uniform scale mixtures. *Applied Stochastic Models in Business and Industry*, *31*(5), 584–608.

Wilhelmsson, A. (2009). Value at Risk with time varying variance, skewness and kurtosis — the NIG-ACD model. *The Econometrics Journal*, *12*(1), 82–104.

Wu, W. B., Huang, Y., & Huang, Y. (2010). Kernel estimation for time series: an asymptotic theory. *Stochastic Processes and Their Applications*, *120*(12), 2412–2431.

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.

Zhao, Z., Zhang, Z., & Chen, R. (2018). Modeling maxima with autoregressive conditional Fréchet model. *Journal of Econometrics*, *207*(2), 325–351.

Zhu, D., & Galbraith, J. W. (2010). A generalized asymmetric Student's t distribution with application to financial econometrics. *Journal of Econometrics*, *157*(2), 297–305.

Zumbach, G. (2007). *The RiskMetricks 2006 Methodology.* RiskMetricks Group.