UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES
Departamento de Fundamentos del Análisis Económico II

TESIS DOCTORAL

Sample size, skewness and leverage effects in value at risk and expected shortfall estimation

Efectos del tamaño muestral, la asimetría y el apalancamiento en la estimación del valor en riesgo y de la pérdida esperada

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Laura García Jorcano

Director

Alfonso Novales Cinca

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

DEPARTAMENTO DE FUNDAMENTOS
DEL ANÁLISIS ECONÓMICO II
(ECONOMIA CUANTITATIVA)

Doctorado en Banca y Finanzas Cuantitativas

TESIS DOCTORAL

Sample size, skewness and leverage effects in Value at Risk and Expected Shortfall estimation

Efectos del tamaño muestral, la asimetría y el apalancamiento en la estimación del Valor en Riesgo y de la Pérdida Esperada

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
REALIZADA Y PRESENTADA POR

Laura García Jorcano

DIRECTOR
Alfonso Novales Cinca

Madrid, 2017

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTY OF ECONOMICS AND BUSINESS

DEPARTMENT OF FOUNDATIONS OF ECONOMIC ANALYSIS II

(QUANTITATIVE ECONOMICS)

Program in Banking and Quantitative Finance

DOCTORAL THESIS

Sample size, skewness and leverage effects in Value at Risk and Expected Shortfall estimation

AUTHOR
Laura García Jorcano

ADVISOR
Alfonso Novales Cinca

Madrid, 2017

Contents

List of Figures

List of Tables

Resumen

La estimación de las medidas de riesgo es un área de gran importancia en la industria financiera. Las medidas de riesgo juegan un papel principal en la gestión del riesgo y en el cálculo del capital requerido. El documento Basilea III [13] ha sugerido medir el riesgo en condiciones de tensión mediante la Pérdida Esperada (ES), en lugar del Valor en Riesgo (VaR), a un nivel de confianza del 97.5%. Este cambio viene motivado por las atractivas propiedades teóricas del ES como medida de riesgo y por las limitaciones del VaR. En particular, el VaR no captura el "riesgo de cola". En esta transición, el principal reto al que se enfrentan las instituciones financieras es la falta de disponibilidad de herramientas sencillas para la evaluación de las predicciones del ES, esto es, *backtesting* del ES.

El objetivo de la tesis es comparar la *performance* de una variedad de modelos para la estimación del VaR y del ES para un conjunto de activos de diferente naturaleza: índices de mercado, acciones, bonos, tipos de cambio y materias primas. A lo largo de la tesis, entendemos por "modelo" VaR o por "modelo" ES una especificación dada por un modelo de volatilidad condicional combinado con una distribución de probabilidad que suponemos siguen las innovaciones estandarizadas.

En concreto, el Capítulo 1 considera el concepto de insesgadez en la estimación del VaR. Francioni y Herzog (2012) [FH] [20] demuestran que existe una corrección analítica del sesgo del VaR cuando los rendimientos siguen una distribución Normal. En este capítulo, el análisis FH se extiende a la distribución t-Student así como a la Mixtura de dos Normales, utilizando el algoritmo bootstrapping propuesto por FH. El uso de medidas VaR insesgadas en probabilidad evita la infraestimación sistemática del riesgo como resultado del sesgo que presentan las medidas VaR estándar. La magnitud de la distorsión necesaria para pasar del cuantil del VaR estándar al del VaR insesgado en probabilidad depende del tamaño de la muestra y del supuesto de distribución de los rendimientos. Debido a que la distribución de los rendimientos financieros tiene colas más gruesas que la Normal, a menor tamaño muestral y menor grosor de las colas de la distribución supuesta en la estimación, mayor es la distorsión necesaria para obtener un estimador insesgado. Este ajuste del VaR nos permite trabajar con muestras pequeñas con la tranquilidad de que se obtendrá una buena *performance* del VaR. Además, los resultados obtenidos en la tesis demuestran que las muestras pequeñas permiten obtener estimaciones del VaR más exactas que las muestras grandes según la Probabilidad en Exceso y la Desviación Absoluta Observada por año (media de las diferencias absolutas entre el número esperado de excesos y el número observado de excesos respecto del VaR). La buena *performance* del VaR

insesgado en probabilidad viene del hecho de que las muestras cortas permiten capturar mejor los cambios estructurales que se producen a lo largo del tiempo en las rentabilidades financieras debido al comportamiento del mercado.

El Capítulo 2 analiza como la eficiencia del VaR depende de la especificación de volatilidad y del supuesto acerca de la distribución de probabilidad para las innovaciones de las rentabilidades. Esta cuestión es crucial para los gestores de riesgo debido a la existencia de un gran número de posibles elecciones de modelos de volatilidad y de distribuciones de probabilidad, siendo conveniente establecer algunas prioridades al modelar los rendimientos para estimar el riesgo. Consideramos diferentes modelos VaR condicionales para activos de diferente naturaleza, utilizando distribuciones simétricas y asimétricas para las innovaciones y modelos de volatilidad con y sin efecto apalancamiento. El VaR estimado se calcula siguiendo el enfoque paramétrico. La capacidad para explicar los momentos muestrales de las rentabilidades podría considerarse una condición natural para obtener una buena *performance* del VaR. Sin embargo, aunque hay normalmente un esfuerzo significativo para seleccionar una combinación apropiada de distribución de probabilidad y especificación de volatilidad en la estimación del VaR, la capacidad para explicar los momentos muestrales de las rentabilidades rara vez se examina. Tras la utilización de métodos de simulación para calcular los momentos de las rentabilidades implícitos a partir de los modelos estimados, comparamos los niveles implícitos de asimetría y curtosis de las rentabilidades con sus momentos muestrales análogos. Se observa que la capacidad para explicar los momentos muestrales está, de hecho, ligada a la *performance* del VaR. Dicha *performance* es examinada a través de contrastes estándar: el contraste de cobertura incondicional de Kupiec (1995) [78], el de cobertura condicional y de independencia de Christoffersen (1998) [27] y el Dynamic Quantile de Engle y Manganelli (2004) [39], así como, a través de funciones de pérdida propuestas por Lopez (1998, 1999) [85] [86] y Sarma et al. (2003) [113] y por Giacomini y Komunjer (2005) [47].

Respecto a una literatura cada vez más abundante, contribuimos de diferentes formas:

i) considerando un conjunto de distribuciones de probabilidad que recientemente han sido consideradas apropiadas para capturar la asimetría y la curtosis de datos financieros, pero cuya *performance* para las estimaciones del VaR no ha sido comparada previamente para una base de datos común: distribución t-Student asimétrica [41], Error Generalizada asimétrica [41], Johnson S_U [70], t-Generalizada asimétrica [123] y t-Student asimétrica Hiperbólica Generalizada [1], junto con las distribuciones Normal y t-Student como *benchmark*,

ii) considerando tres especificaciones de volatilidad con apalancamiento, GJR-GARCH, APARCH y FGARCH, así como el modelo estándar simétrico GARCH como *benchmark*. Los modelos FGARCH y APARCH son cada vez más apreciados al ser adecuados para las rentabilidades financieras ya que su especificación considera una potencia en la desviación estándar de las innovaciones lo que proporciona mayor flexibilidad a la dinámica de la volatilidad,

iii) evaluando explícitamente el ajuste de las rentabilidades y relacionando esa capacidad de ajuste con la *performance* del VaR e

iv) introduciendo un criterio de dominancia para establecer un ranking de modelos en base al comportamiento de los mismos en los contrastes estándar de validación del VaR y en las funciones de pérdida.

Obtenemos los siguientes resultados:

i) los modelos VaR que asumen distribuciones de probabilidad asimétrica para las innovaciones, como la t-Student asimétrica, la Error Generalizada asimétrica, la Johnson S_U y la t-Generalizada asimétrica proporcionan un mejor ajuste de los momentos muestrales de las rentabilidades que las distribuciones simétricas y logran una mejor *performance* del VaR,

ii) los modelos de volatilidad con apalancamiento, como el APARCH y el FGARCH, muestran una mejor *performance* del VaR que las especificaciones de volatilidad más estándar como el GARCH y el GJR-GARCH,

iii) los resultados de nuestras simulaciones fuera de muestra sugieren que el supuesto importante para la *performance* del VaR es el de la distribución de probabilidad de las innovaciones de las rentabilidades, jugando un papel secundario la elección del modelo de volatilidad,

iv) la consideración de la potencia de la desviación estándar condicional como un parámetro libre es una importante característica de las especificaciones de volatilidad APARCH/FGARCH, al sugerir nuestras estimaciones, para un número considerable de activos, que la especificación de la desviación condicional al cuadrado es inapropiada,

v) un buen ajuste de los momentos de las rentabilidades normalmente conlleva una buena *performance* del VaR. Los modelos APARCH o FGARCH con distribuciones Error Generalizada asimétrica, t-Generalizada asimétrica y Johnson S_U son preferidos a aquellos con otras distribuciones asimétricas, como la t-Student asimétrica y la t-Student asimétrica Hiperbólica Generalizada, o con distribuciones simétricas, como la t-Student y la Normal y

vi) modelos VaR alternativos parecen proporcionar distinta performance del VaR para las diferentes clases de activos.

En el Capítulo 3 estimamos el ES condicional basado en el enfoque de la Teoría de Valores Extremos (EVT) utilizando distribuciones de probabilidad asimétricas para las innovaciones de las rentabilidades y examinamos la exactitud de nuestras estimaciones antes y durante la crisis financiera de 2008 utilizando datos diarios a horizontes 1 día y 10 días. Tenemos en cuenta los efectos de los agrupamientos de volatilidad y de apalancamiento en la volatilidad de las rentabilidades al utilizar un modelo APARCH con diferentes distribuciones de probabilidad para las innovaciones estandarizadas: Gaussiana, t-Student, t-Student asimétrica, Error Generalizada asimétrica y Johnson S_U, así como, con el enfoque EVT siguiendo el procedimiento de dos pasos de McNeil y Frey (2000) [75]. Este procedimiento en dos pasos ajusta la distribución Pareto Generalizada a los valores extremos de los residuos estandarizados generados por los modelos APARCH. Posteriormente, comparamos la performance de las predicciones del ES fuera de muestra un día hacia adelante de todos estos modelos para diferentes niveles de significación (α). Previamente, los contrastes de backtesting existentes para ES han demostrado tener serias limitaciones, como son el contraste de McNeil y Frey (2000) [75], de Berkowitz (2001) [15], de Kerkhof y Melenberg (2004) [65] y de Wong (2008) [95]. Tales limitaciones son superadas por algunas de las propuestas más recientes de backtesting para ES las cuales utilizamos para la evaluación del ES: el contraste de Righi y Ceretta (2013), los dos contrastes de Acerbi y Szekely (2014) [4], que son sencillos pero requieren análisis de simulación (como el contraste de Righi y Ceretta), el contraste de Graham y Pál (2014) [57], que es una extensión del enfoque Lugannani-Rice de Wong, el contraste de cobertura incondicional de Costanzino y Curran (2015) [26] para la familia de medidas de riesgo espectrales, de la cual el ES es miembro y, finalmente, el contraste condicional de Du y Escanciano (2015) [36].

Este capítulo contribuye a la literatura de diferentes maneras:

i) considerando la especificación de volatilidad APARCH en modelos EVT utilizando Simulación Histórica Filtrada (FHS), lo que permite considerar los agrupamientos de volatilidad y la asimetría de los rendimientos,

ii) comparando modelos EVT condicionales que incorporan modelos condicionales con distribuciones de probabilidad asimétricas apenas utilizadas en la literatura financiera para la estimación del ES,

iii) analizando la *performance* de las estimaciones del VaR y del ES a horizonte 10 días, tal y como proponen los requerimientos de capital de Basilea,

iv) centrándonos en la exactitud de nuestros modelos de riesgo para estimar el VaR y el ES durante los periodos de pre-crisis y de crisis, así como, para diferentes niveles de significación (α) y

v) evaluando la *performance* del ES con las propuestas más recientes de backtesting para ES, considerando todas en el mismo estudio.

Obtenemos las siguientes conclusiones:

i) la Teoría de Valores Extremos produce una buena estimación del ES independientemente de la distribución de probabilidad supuesta en la estimación para las innovaciones de las rentabilidades. Esto es debido al hecho de que la cola, para todos estos casos, es modelizada con una distribución Pareto Generalizada,

ii) si consideramos modelos condicionales sin el enfoque EVT, observamos que la distribución Error Generalizada asimétrica y la Johnson S_U juegan un papel importante en la captura del "riesgo de cola" a horizontes 1 y 10 días. Esto se debe a que los hechos estilizados de las rentabilidades financieras tales como agrupamientos de volatilidad, colas pesadas y asimetría son capturados adecuadamente por las mismas,

iii) incluso durante el periodo de crisis, los modelos EVT condicionales son más adecuados y fiables para predecir las pérdidas generadas por riesgo de los activos que los modelos condicionales que no incorporan el enfoque EVT y

iv) los modelos EVT condicionales producen, en algunos casos, sobreestimación del ES.

Summary

The estimation of risk measures is an area of highest importance in the financial industry. Risk measures play a major role in the risk-management and in the computation of regulatory capital. The Basel III document [13] has suggested to shift from Value-at-Risk (VaR) into Expected Shortfall (ES) as a risk measure and to consider stressed scenarios at a new confidence level of 97.5%. This change is motivated by the appealing theoretical properties of ES as a measure of risk and the poor properties of VaR. In particular, VaR fails to control for "tail risk". In this transition, the major challenge faced by financial institutions is the unavailability of simple tools for evaluation of ES forecasts (i.e. backtesting ES)

The objective of this thesis is to compare the performance of a variety of models for VaR and ES estimation for a collection of assets of different nature: stock indexes, individual stocks, bonds, exchange rates, and commodities. Throughout the thesis, by a VaR or an ES "model" is meant a given specification for conditional volatility, combined with an assumption on the probability distribution of return innovations.

Specifically, Chapter 1 considers the concept of unbiasedness in VaR estimation. Franconi and Herzog (2012) (FH) [20] showed that there exists an analytical bias correction for VaR when returns are Normally distributed. In this chapter the FH analysis is extended to the Student-t distribution as well as to Mixtures of two Normal distributions, using a bootstrapping algorithm proposed by FH. The use of the probability-unbiased VaR avoids the systematic underestimation of risk implied by the bias of standard VaR measures. The magnitude of the distortion that needs to be exerted on the quantile to move from the standard VaR to the probability-unbiased VaR depends on the sample size and on the distribution assumption on returns. Since financial returns usually have thick tails, the smaller the sample size and the lower the heaviness of the tail of the assumed distribution in estimation, the higher will be the distortion to be applied to achieve unbiasedness. This VaR adjustment allows us to work with small samples knowing that the estimated VaR will generally display a good performance. Furthermore, the results in the thesis show that using a small sample may easily lead to more accurate VaR estimates than longer samples according to the Exceedance Probability and to the Observed Absolute Deviation per year (mean of the absolute differences between the expected number of exceedances and the number of observed exceedances). The good performance of the probability-unbiased VaR follows from the fact that a short sample size allows for capturing better the structural changes that arise over time in financial returns due to trading behaviour.

Chapter 2 analyzes how the efficiency of VaR depends on the volatility specification and the assumption on the probability distribution for return innovations. This question is crucial for risk managers, since there are so many potential choices for volatility model and probability distributions that it would be very convenient to establish some priorities in modelling returns for risk estimation. We consider different conditional VaR models for assets of different nature, using symmetric and asymmetric probability distributions for the innovations and volatility models with and without leverage. We calculate VaR estimates following the parametric approach. The ability to explain sample return moments might be considered a natural condition to obtain a good VaR performance. However, even though significant effort is usually placed in selecting an appropriate combination of probability distribution and volatility specification in VaR estimation, the ability to explain sample return moments is seldom examined. After using simulation methods to calculate implied return moments from estimated models, we compare the implied levels of skewness and kurtosis of returns with the analogue sample moments. We show that the ability to explain sample moments is in fact linked to performance in VaR estimation. Such performance is examined through standard tests: the unconditional coverage test of Kupiec (1995) [78], the independence and conditional coverage tests of Christoffersen (1998) [27], the Dynamic Quantile test of Engle and Manganelli (2004) [39], as well as the loss functions proposed by Lopez (1998, 1999) [85] [86] and Sarma et al. (2003) [113] and that of Giacomini and Komunjer (2005) [47].

Relative to an ever increasing literature, we contribute in different ways:

i) considering a set of probability distributions that have recently been rendered to be appropriate for capturing the skewness and kurtosis of financial data, but whose performance for VaR estimation has not been compared previously on a common dataset: Skewed Student-t [41], Skewed Generalized Error [41], Johnson S_U [70], Skewed Generalized-t [123] and Generalized Hyperbolic Skew Student-t [1] distributions, with the Normal and symmetric Student-t distributions as benchmark,

ii) considering three volatility specifications with leverage, GJR-GARCH, APARCH and FGARCH, as well as the standard symmetric GARCH model as benchmark. FGARCH and APARCH are increasingly being appreciated as being adequate for financial returns because they are specified for a power of the conditional standard deviation of the innovations, which provides more flexibility to the dynamics of volatility,

iii) explicitly evaluating the fit to return data, relating that fitting ability to VaR performance, and

iv) by introducing a dominance criterion to establish a ranking of models on the basis of their behavior under standard VaR validation tests and loss functions.

We obtain the following results:

i) VaR models that assume asymmetric probability distributions for the innovations, like the Skewed Student-t distribution, Skewed Generalized Error distribution, Johnson SU distribution, and Skewed Generalized-t distribution provide a better fit to sample return moments than symmetric distributions and achieve a better VaR performance,

ii) volatility models with leverage, like APARCH and FGARCH, show a better VaR performance than more standard GARCH and GJR-GARCH volatility specifications, *iii*) our out-of-sample simulation results suggest that the important assumption for VaR performance is that of the probability distribution of return innovations, with the choice of volatility model playing a secondary role,

iv) dealing with the power of the conditional standard deviation as a free parameter is an important feature of the APARCH/FGARCH volatility specifications because our estimates suggest that for a number of financial assets the squared conditional deviation specification is inappropriate,

v) a good fit to return moments usually leads to a good VaR performance. APARCH or FGARCH models with Skewed Generalized Error, Skewed Generalized-t and Johnson S_U distributions are preferred to other asymmetric distributions, like Skewed Student-t and Generalized Hyperbolic Skew Student-t, and symmetric distributions, like Student-t and Normal distributions, and

vi) alternative VaR models seem to provide a distinct performance for different classes of assets.

In Chapter 3 we estimate the conditional Expected Shortfall based on the Extreme Value Theory (EVT) approach using asymmetric probability distributions for return innovations, and we analyze the accuracy of our estimates before and during the 2008 financial crisis using daily data for 1- and 10-day horizons. We take into account volatility clustering and leverage effects in return volatility by using the APARCH model under different probability distributions assumed for the standardized innovations: Gaussian, Student-t, skewed Student-t, skewed generalized error and Johnson S_U and under EVT methods, following the two-step procedure of McNeil & Frey (2000) [75]. This two-step procedure fits a Generalized Pareto Distribution to the extreme values of the standardized residuals generated by APARCH models. Then, we compare the one-step-ahead out-of-sample ES forecast performance of all these models for different significance levels (α). Previously existing backtesting tests for ES have been shown have serious limitations [as McNeil & Frey (2000) [75] test, Berkowitz (2001) [15] test, Kerkhof and Melenberg (2004) [65] test and Wong (2008) [95] test]. Such limitations are overcome by some recent ES backtesting proposals that we use for ES evaluation: the Righi & Ceretta (2013) [83] test, two tests by Acerbi & Szekely (2014) [4] that are straightforward but require simulation analysis (like the Rigui & Ceretta test), the test of Graham & Pál (2014) [57], which is an extension of the Lugannani-Rice approach of Wong, the quantile-space unconditional coverage test of Costanzino & Curran (2015) [26] for the family of Spectral Risk Measures, of which ES is a member and, finally, the conditional test of Du & Escanciano (2015) [36].

This chapter contributes to the literature in different ways:

i) considering the APARCH volatility specification in an EVT model using Filtered Historical Simulation (FHS) [11] [12] to take into account volatility clustering and asymmetric returns,

ii) comparing conditional EVT models that incorporate conditional models with asymmetric probability distributions rarely used in the financial literature for ES estimation,

iii) by analyzing the performance of VaR and ES estimates over 10-day horizons for risk liquidity management, as proposed in Basel capital requirements [13],

iv) by focusing on the accuracy of our risk models for VaR and ES estimation during the pre-crisis and crisis periods as well as under different significance levels (α), and

v) by evaluating ES performance using the most recent ES backtesting proposals in the same study.

We obtain the following conclusions:

i) Extreme Value Theory produces a good ES performance regardless of the probability distribution assumed for return innovations in estimation. This is due to the fact that the tail is modeled with a Generalized Pareto Distribution not only with 1-day but also 10-day horizons,

ii) if we consider conditional models without the EVT approach, we observe that the Skewed Generalized Error distribution and the Johnson S_U distribution play an important role in capturing tail risk in 1-day and 10-day horizons. This is because the stylized facts of financial returns such as volatility clusters, heavy tails and asymmetry are suitably captured by these asymmetric distributions,

iii) even during the crisis period, conditional EVT models are more accurate and reliable for predicting asset risk losses than conditional models that do not incorporate the EVT approach, and

iv) sometimes conditional EVT models produce a ES overestimation.

Chapter 1

Probability-unbiased VaR estimator

Abstract

The Probability unbiasedness of a Value at Risk (VaR) estimator guarantees that the numerical VaR estimate obtained from a finite amount of data will be exceeded by the next observation drawn from the same distribution with an expected probability α. In the special case of a Normal distribution, closed-formed solutions for probability unbiased VaR estimators are known. For the general case, we use a bootstrapping algorithm to illustrate the outcomes obtained by estimating VaR from simulated random samples of different length generated from Normal, Student-t and Mixtures of two Normal distributions. Using the empirical distribution derived for the VaR estimate, we compute in short samples the probability-unbiased VaR as well as its confidence bands. Our results show that short samples may yield good VaR estimates. In fact, we show the probability unbiased VaR estimator to display a better performance than the standard VaR estimator obtained under different models, all of which are much more complex.

1.1 Introduction

The parametric or variance-covariance approach to the estimation of VaR in two steps: first, the distribution is estimated by statistical methods; in the second step, the estimated distribution is considered as the true distribution and VaR is computed. In the parametric case this is achieved by using the mathematical expression for VaR in each specific model and inserting the estimated parameters. This VaR estimator is called *plug-in* estimator. It is well-known that the highly nonlinear mapping from the model parameters to the risk-measure introduces biases and some statistical experiments show that this bias leads to a systematic underestimation of risk.

VaR performance is usually assessed by comparing the observed number of violations of the quantile estimator threshold with the theoretical frequency. Francioni and Herzog, "Probability-unbiased Value-at-Risk estimators" [20] suggest the use of probability unbiasedness as a criterion to judge the quality of VaR estimates. Probability unbiasedness means that the VaR estimator should be unbiased regarding the relative frequency of violations of the quantile. They show how the α-quantile may be modified so that the implied VaR estimate is unbiased.

We use the non-parametric method (bootstrapping) for the calculation of the unbiased VaR estimator introduced by Francioni and Herzog for the Normal distribution. We extend their approach to other distributions such as Student-t and mixture of Normals, while using the parametric approach to calculate probability-unbiased VaR in the Normal case. We show that the use of probability-unbiased VaR avoids the systematic underestimation of risk implied by the bias of standard VaR measures in small samples. The magnitude of the distortion that needs to be exerted on the quantile to move from the standard VaR to the probability-unbiased VaR depends on the sample size and on the distribution assumption on returns. Our results suggest that using a small sample may easily lead to more accurate VaR estimates than a historical estimator based on long samples according to the exceedence probability and to the Observed Absolute Deviation per year. Short samples are more robust to the structural changes that may arise over time in financial returns due to trading behavior.

The remainder of the chapter is organized as follows. In Section 1.2, we present a review of literature. In Section 1.3, we describe the concept of quantile or VaR estimator. In Section 1.4, we introduce the difference between parametric and non-parametric methods used to calculate probability-unbiased VaR. In Sections 1.5, 1.6 and 1.7, we explain and calculate probability-unbiased VaR with Normal, Student-t and Mixture of two Normal distributions, respectively. In Section 1.8, we report the results of an empirical application. Finally, Section 1.9 concludes the chapter.

1.2 A review of literature

The estimation of risk measures is an area of highest importance in the financial industry. Risk measures play a major role in the risk-management and in the computation of regulatory capital. [For an in-depth treatment of the topic, see textbooks of McNeil, Frey and Embrechts (2005) [75] and of Alexander (2009) [2]]. In particular, Embrechts and Hofert (2014) [37] highlight that a major part of quantitative risk management is actually of statistical nature, and the statistical aspects in the estimation of risk measures have recently raised a lot of attention [see Acerbi and Szekely (2007) [4], Davis (2014) [10], Emmer et al. (2015) [39], Du and Escanciano (2015) [36], Costanzino and Curran (2015) [26], Fissler et al. (2015) [45] and Ziegel (2016) [97]]. Surprisingly, it turns out that statistical properties of risk estimators have not yet been analyzed thoroughly. Some of the existing analysis show that standard risk estimators may be biased, and they systematically underestimate risk. Unfortunately, while the classical (statistical) definition of bias is always desirable from a theoretical point of view, it is not considered a priority by financial institutions or regulators, for whom the backtests are the main source of estimation accuracy.

Not surprisingly, the occurrence of biases in risk estimation plays an important role in practice. The Basel III document [13] has suggested to shift from Value-at-Risk into Expected Shortfall as a risk measure and to consider stressed scenarios at a new confidence level of 97.5%. In fact, such a correction may reduce the bias, but only in the right scenarios. Our goal is to obtain probability-unbiased estimators that pass the standard backtesting procedure proposed by Basel. That amounts to having an expected failure rate close to the theoretical VaR level α.

Francioni and Herzog (2012) [20] (FH) have shown how to distort the α quantile used in VaR estimation so that the implied VaR estimator is unbiased in probability, in a sense to be defined below. Our goal is to extend the concept of probability unbiased estimation introduced by FH to distributions different from Normal. In this line, Pitera and Schmidt (2016) [26] propose an unbiased bootstrapping estimator under Normality obtained by distorting the estimated parameters of the distribution instead of distorting the VaR confidence level as FH suggest. The FH strategy is the one we follow in this chapter.

1.3 Quantile or VaR estimator

In this section we describe the concepts associated to the probability unbiased estimation of Value at Risk.

Let us suppose that X is an absolutely continuous random variable with distribution function F_θ, where θ is a parameter vector. The α quantile Q_α of X is defined as

$$Q_\alpha = F_\theta^{-1}(\alpha)$$

By definition, the quantile has the property that

$$F_\theta(Q_\alpha) = \alpha$$

This equation represents the intuitive concept of the quantile as a threshold that is exceeded with probability α. The quantile Q_α of the distribution of returns of a given financial asset or portfolio is known as the Value-at-Risk (VaR) at the level α or at the confidence level $1 - \alpha$.

We assume that the parameter vector θ can be estimated by any method like Maximum Likelihood, Generalized Method of Moments or others in such a way that the observed data are well described. We will assume that estimator to be at least consistent.

In a general estimation setup, a *plug-in* estimator for a function $g(\theta)$ is an estimator obtained by replacing the parameter θ in the function by an estimate, that is

$$\widehat{g(\theta)} = g(\widehat{\theta})$$

The quantile Q_α can be seen as a function of the parameter vector and the significance level:

$$Q_\alpha = g(\theta, \alpha)$$

The *plug-in* VaR estimator is the only method to estimate VaR under a parametric approach:

$$\widehat{VaR}_\alpha = \widehat{Q}_\alpha = g(\widehat{\theta}, \alpha)$$

We aim at estimating the risk of the future position where $\theta \in \Theta$ are unknown. If θ were known, we could directly compute the corresponding VaR as a function of θ, $g(\theta)$, specifically with F_θ, and we would not need to consider the family of VaR, $(g(\theta))_{\theta \in \Theta}$.

Our aim is to estimate Q_α in such a way that the estimator satisfies this probabilistic 'threshold property' in the mean for a F_θ-distributed random variable X for all θ, i.e.

$$\mathbb{E}_\theta[F_\theta(\widehat{Q}_\alpha)] = \alpha$$

where $\mathbb{E}_\theta$ denotes the expectation operator under probability measure F_θ.

This is a standard unbiasedness condition on the probability of exceeding the VaR estimate $\widehat{Q}_\alpha$. That probability is usually checked by backtesting. Unbiasedness would imply that the VaR estimate $\widehat{Q}_\alpha$ will be exceeded with an expected probability equal to α.

Definition 1 *An estimator $\widehat{g(\theta)}$, obtained with sample observations $(X_1, ..., X_n) \sim F_\theta$ of $g(\theta)$, is said to be probability unbiased with respect to a random variable Z with distribution function F^Z, if*

$$F^Z(g(\theta)) = \mathbb{E}_\theta[F^Z(\widehat{g(\theta)})]$$

holds for all θ.

In the case of a quantile/VaR estimation where all $X_i \sim^{i.i.d} F_\theta$, $i = 1, ..., n$, $g(\theta)$ is the α-quantile Q_α, Z is the next sample observation $Z = X_{n+1}$, and F^Z is the probability distribution from which the sample has been obtained. Hence, a probability-unbiased VaR estimator with respect to $Z = X_{n+1}$ must satisfy:

$$\mathbb{E}_\theta[P(X_{n+1} < \widehat{Q}_\alpha)] = \alpha \tag{1.1}$$

Unfortunately, under nonlinear mappings of the parameter vector θ, as it is the case of the quantile, the *plug-in* procedure generally introduces a small sample bias: $\mathbb{E}(P(X_{new} <$

$\widehat{VaR}_\alpha) \neq \alpha$. The reason is that it treats the estimated parameter vector as deterministic, even though $\widehat{\theta}$ is a random variable, a fact that must be incorporated into the estimation procedure in order to obtain probability-unbiasedness. As a consequence, the equation:

$$\widehat{Q}_\alpha = F^{-1}_{\widehat{\theta}}(\alpha)$$

where $\widehat{\theta}$ is an estimator of the parameter θ, is only true asymptotically, i.e. as the number of observations goes to infinity, provided the *plug-in* estimator is consistent.

$$\widehat{VaR}_\alpha \equiv \widehat{Q}_\alpha \overset{n\to\infty}{\longrightarrow} VaR_\alpha \equiv Q_\alpha = F^{-1}_\theta(\alpha)$$

almost surely for each $\theta \in \Theta$, so that it is asymptotically unbiased.

To obtain a probability-unbiased estimator for the quantile there are two approaches,

1. Estimating a probability α_{pu} to modify the quantile from the estimated distribution for which the VaR is estimated. The VaR estimator will be

   $$\widehat{Q}_{\alpha_{pu}} = \widehat{VaR}_{\alpha_{pu}} = g(\widehat{\theta}, \alpha_{pu}) = F^{-1}_{\widehat{\theta}}(\alpha_{pu})$$

   where α_{pu} is chosen so that equation (1.1) is fulfilled.
   For example, if F is a Normal distribution, the VaR estimator can be written:

   $$\widehat{VaR}_\alpha = \widehat{\mu} + \widehat{\sigma} z_{\alpha_{pu}}$$

   where $\widehat{\mu}$ and $\widehat{\sigma}$ are the estimated mean and standard deviation, respectively, and $z_{\alpha_{pu}}$ is the inverse cumulative distribution function of the standard Normal(0,1) for α_{pu}.

2. Modifying the estimate of the parameter vector $\widehat{\theta}$ of the distribution F to $\widehat{\theta}_{pu}$ when computing the *plug-in* estimator

   $$\widehat{Q}_\alpha = F^{-1}_{\widehat{\theta}_{pu}}(\alpha)$$

   If F is a Normal distribution: $\widehat{\theta}_{pu} = (\widehat{\mu}_{pu}, \widehat{\sigma}_{pu})$, and the VaR estimator would then be written as follows:

   $$\widehat{VaR}_\alpha = \widehat{\mu}_{pu} + \widehat{\sigma}_{pu} z_\alpha$$

   On the other hand, the *plug-in* estimator, which has been used in the calculation of quantile / VaR is:

   $$\widehat{VaR}_\alpha = \widehat{\mu} + \widehat{\sigma} z_\alpha$$

In this chapter we follow the first of these two approaches to calculate the probability-unbiased VaR, and we use the second approach, whenever possible, to graph an approximation of the function F distorted by modifying the parameter $\widehat{\theta}$. Thus, we will be computing a probability-unbiased estimator of VaR, that is, an estimator:

$$\widehat{Q}_\alpha = F^{-1}_{\widehat{\theta}}(\alpha_{pu})$$

where α_{pu} is chosen so that equation (1.1) is fulfilled.

1.4 Parametric and Non-Parametric methods

We can use parametric or non-parametric methods to find α_{pu}. On the one hand, parametric methods, or classical statistical methods, have the basis for making inferences about the population in the theoretical sampling statistical distribution, whose parameters can be estimated from the observed statistical sample. On the other hand, there are different procedures based on non parametric methods. Those procedures generate samples from a set of observations constructing a sample distribution that can be used for parameter estimation and confidence intervals. Among them, probably the best known and most commonly used is the bootstrap method. The first mention of this method under this name is due to Efron (1979) [13], although the same basic ideas came handling for at least a decade ago (Simon, 1969 [29]). Efron conceived the bootstrap method as an extension of "jackknife techniques", which usually consist in extracting samples ever constructed by removing one element of the original sample to assess the effect on certain statistical (Quenouille, 1949 [28]; Tukey, 1958 [125] and Miller, 1974 [25]).

The bootstrap method, unlike classical estimation methods, does not make any distribution assumptions for the theoretical statistical. Instead, the distribution of the statistic is determined by simulating a large number of random samples constructed directly from the observed data. That is, the original sample is used to generate new samples from that as a basis for estimating inductively the sampling distribution of the statistic, rather than deriving it from a theoretical distribution assumed a priori. This method has an immediate predecessor in the techniques of Monte Carlo simulation, consisting in extracting a large number of random samples from a known population to calculate from them the value of the statistic whose sampling distribution is intended to be estimated. However, in practice the population is not known and the information we have is a sample drawn from it.

Definition 2 *The bootstrapping (bootstrap) is a resampling method or algorithm that consists in generating a large number of resamples using sampling with replacement from an original random sample of size n which represents the population from which it was extracted. Each resample is the same size as the original random sample. The resamples serve as population samples.*

According to the main idea of bootstrap, the procedure involves using the sample itself since we consider that it contains basic information about the population. Therefore, the suitability of this method will be greater when the sample contributes with more information about the population. A direct consequence is that the longer the sample size, the better the estimation about the sample distribution of a statistic. However, even with small samples, between ten and twenty observations, the bootstrap method can provide correct results (Bickel and Krieger, 1989 [6]) while being unsuitable for samples with less than five (Chernick, 1999 [8]).

1.5 Normal Distribution

The VaR calculated by the parametric approach for a Normal distribution is $Q_\alpha = \mu + \sigma z_\alpha$ where the parameter vector $\theta = (\mu, \sigma)$. To obtain the *plug-in* $\widehat{VaR}_\alpha$ the parameters of the Normal distribution are replaced by their Maximum Likelihood estimates $\widehat{Q}_\alpha = \bar{x} + z_\alpha s$ where z_α is the inverse cumulative distribution function of a Normal(0,1), $\bar{x}$ is the sample mean and s is the sample standard deviation. These statistics are independent since the sample comes from a distribution $N(\mu, \sigma^2)$

$$\bar{x} = \frac{1}{n}\Sigma X_i$$

$$s = \sqrt{\frac{1}{n-1}\Sigma(X_i - \bar{x})^2}$$

For the Normal distribution, the statistical distributions are known, where the distribution of sample mean is a Normal distribution,

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and the distribution of variance is calculates as follows,

If we start from a simple random sample with distribution $N(\mu, \sigma^2)$, then

$$\frac{n-1}{\sigma^2}s^2 \sim \chi^2_{n-1}$$

Using the previous proposition, we can obtain the distribution of s^2 through a transformation of random variable. It is obtained as

$$s^2 \sim \chi^2_{n-1}\left(\frac{n-1}{\sigma^2}s^2\right)\frac{n-1}{\sigma^2}$$

Note that the distributions of the two estimators depend on the size of the sample n. The estimator $\hat{x}$ is unbiased, and s^2 is consistent (Casella and Berger, 2002 [7]).

To obtain

$$\mathbb{E}(P(X_{n+1} < \widehat{VaR}_\alpha)) = \alpha$$

we have,

$$P\left(X_{n+1} < \bar{x} + z_\alpha s\right) = \int_\infty^{\bar{x}+z_\alpha s} N\left(x \mid \mu, \sigma^2\right) dx = P\left(\frac{X_{n+1} - \mu}{\sigma} < \frac{\bar{x} + z_\alpha s - \mu}{\sigma}\right) = \Phi\left(\frac{\bar{x} - \mu}{\sigma} + z_\alpha \frac{s}{\sigma}\right)$$

$$\mathbb{E}\left(\Phi\left(\frac{\bar{x} - \mu}{\sigma} + z_\alpha \frac{s}{\sigma}\right)\right) = \iint \Phi\left(\frac{\bar{x} - \mu}{\sigma} + z_\alpha \frac{\sqrt{s^2}}{\sigma}\right) N\left(\bar{x} \mid \mu, \frac{\sigma^2}{n}\right) \chi^2_{n-1}\left(\frac{n-1}{\sigma^2}s^2\right)\frac{n-1}{\sigma^2} d\bar{x} ds^2 \qquad (1.2)$$

where we have calculated the expectation of a two random variables continuous function:

$$\mathbb{E}(g(\bar{x}, s^2)) = \iint g(\bar{x}, s^2) f_{\bar{x}s^2} d\bar{x} ds^2$$

where $f_{\bar{x}s^2}$ is the joint density function of two random variables. As $\bar{x}$ and s^2 are independent, the joint density function is just the product of the density functions of each variable

$$N\left(\bar{x} \mid \mu, \frac{\sigma^2}{n}\right) \chi^2_{n-1}\left(\frac{n-1}{\sigma^2}s^2\right) \frac{n-1}{\sigma^2}$$

In (1.2) we do the following change of variable

$$X = \frac{\bar{x}-\mu}{\sigma} \qquad d\bar{x} = \sigma dX$$
$$Y = \frac{n-1}{\sigma^2}s^2 \qquad ds^2 = \frac{\sigma^2}{n-1}dY$$

The density function of the sample mean after the change of variable is

$$N\left(\bar{x} \mid \mu, \frac{\sigma^2}{n}\right) = \frac{\sqrt{n}}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} = \frac{\sqrt{n}}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{n}{2}X^2} = \frac{1}{\sigma}N\left(X \mid 0, \frac{1}{n}\right)$$

Therefore, the equation that defines α_{pu} with Maximum Likelihood estimator for the Normal distribution is

$$\iint \Phi\left(X + z_{\alpha_{pu}}\sqrt{\frac{Y}{n-1}}\right) N\left(X \mid 0, \frac{1}{n}\right) \chi^2_{n-1}(Y)dXdY = \alpha \tag{1.3}$$

Notice that equation (1.3) only depends on α and n, but it does not depend on θ, i.e. on μ and σ. Non-dependence on θ arises under the Normal distribution because of its strong invariance structure. Being a location-scale distribution, we can reduce it to a standard Normal distribution that does not depend on these parameters. This property is important because in estimating the parameter θ, the VaR estimator obtained is only an approximation to the *probability-unbiased* VaR.

In this case, the α_{pu} is unique for each sample size (n) and for each probability α and it does not vary from one sample to another of equal size because the function (1.3) does not depend on the distribution parameters. The VaR obtained with each of these α_{pu} will be probability-unbiased, that is, $\mathbb{E}(\widehat{VaR}) = VaR$.

To sum up, if the probability distribution from which we draw independent sample realizations belongs to the location-scale family, then we will be able to find an α_{pu} such that the VaR is unbiased.

Table 1.1 lists the probabilities α_{pu} obtained from equation (1.3), as a function of the sample size n and the value of α. We can see that $\alpha_{pu} \to \alpha$ when $n \to \infty$. For instance, under the estimated probability distribution for a sample size $n = 20$, the 3.82% percentile has a 5% probability of being exceeded by a future observation drawn from the full distribution of returns. As we can see, for small sample sizes the estimated distribution function from a Normal sample is much heavier tailed than the Normal distribution associated to the plug-in estimator. As a consequence, the *plug-in* VaR estimator underestimates risk.

	$\alpha(\%)$			
n	**0.5**	**1**	**5**	**10**
10	0.033	0.154	2.727	7.345
15	0.105	0.336	3.445	8.239
20	0.169	0.463	3.821	8.683
25	0.217	0.552	4.051	8.948
50	0.340	0.757	4.520	9.476
100	0.415	0.874	4.759	9.738
150	0.442	0.915	4.839	9.826
200	0.456	0.936	4.879	9.869

Table 1.1: Probabilities $\alpha_{pu}(\%)$ to be used to obtain the *probability-unbiased* VaR_α for different values of α and n in the i.i.d. Normal distribution case.

Table 1.2 presents the reverse question: What is the α associated to a given α_{pu}? Now, at 5% significance and $n = 20$, the *pu*-VaR estimate would have a 6.25% probability of being exceeded by a future sample observation from the full distribution of returns. We observe that the differences are greater when we have small sample sizes and we can also observe that $\alpha \to \alpha_{pu}$ when $n \to \infty$.

Figure 1.1 graphs the distortion function for different sample sizes (red line). It corroborates the fact that, as we have more observations, the correction in the probability level is smaller and the distortion function converges to the identity (black line). This distortion function describes how probabilities need to be changed in the plug-in quantile estimator such that the plug-in estimator becomes probability-unbiased. Figure 1.2 shows the distortion of the quantiles of the standard Normal distribution function which describes how the plug-in estimate of the cumulative density function has to be changed for a given sample size n such that the estimate becomes probability-unbiased. In both figures, we only represent the left extreme quantiles, but it would be possible to enlarge the graph to represent the entire distribution. In Figure 1.2. we observe that for more extreme quantiles the distortion is greater, i.e. the differences between α and α_{pu} are larger. Also, α_{pu} is always lower than α, in other words, *probability-unbiased* VaR is greater (in absolute value) than *plug-in* VaR. The latter underestimates the extreme events and, therefore, is not an appropriate method to estimate risk measure with small samples.

1.5.1 Parametric probability-unbiased VaR estimator for a Normal distribution.

We now turn to the estimation of VaR itself. We apply the first approach described in Section 1.3 to estimate the *probability-unbiased* VaR, which implies a modification of the quantile, replacing α by α_{pu}. Table 1.3 shows the *probability-unbiased* $\widehat{VaR}_\alpha$ (VaR_{pu}) and

	$\alpha_{pu}(\%)$			
n	**0.5**	**1**	**5**	**10**
10	1.820	2.686	7.563	12.639
15	1.288	2.043	6.678	11.752
20	1.056	1.751	6.247	11.312
25	0.928	1.585	5.992	11.048
50	0.697	1.277	5.490	10.523
100	0.594	1.134	5.243	10.261
150	0.562	1.089	5.162	10.174
200	0.546	1.066	5.121	10.130

Table 1.2: The shortfall probabilities $\alpha(\%)$ with which the next observation is lower than the plug-in VaR estimate $z_{\alpha_{pu}}$ for the Normal distribution case.

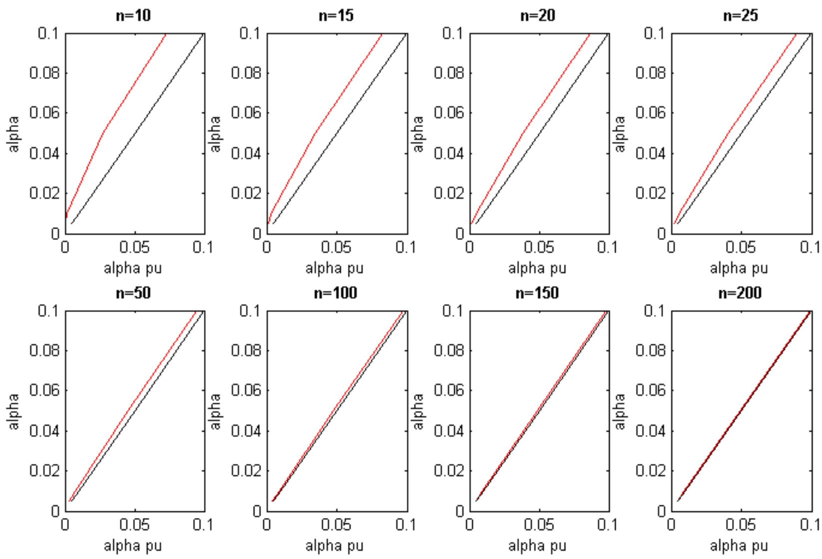


Figure 1.1: Distortion function for the Normal distribution calculated with the parametric method. The diagonal (black line) represents no distortion.

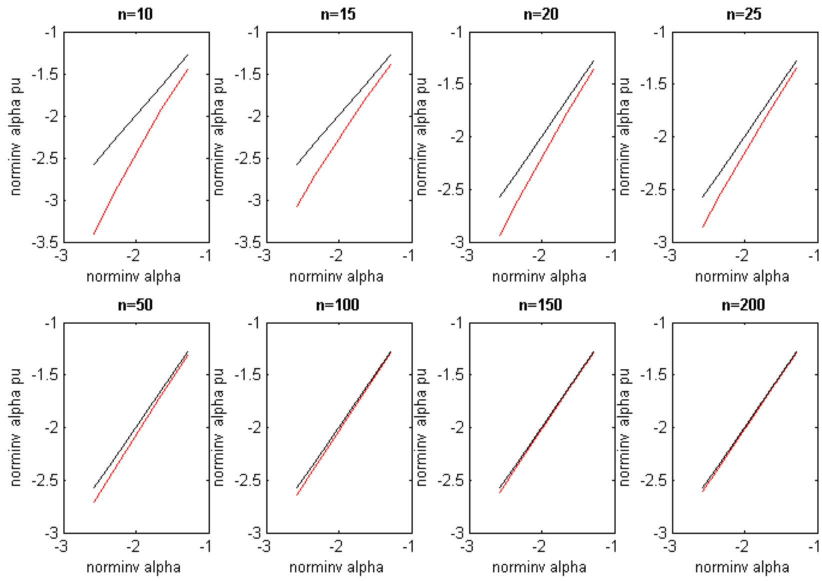


Figure 1.2: The quantiles of the Normal cdf versus the quantiles of the distorted Normal cdf calculated with the parametric method. The diagonal (black line) represents no distortion.

the *plug-in* $\widehat{VaR}_\alpha$ ($VaR_{plug-in}$) obtained for different sample sizes and α's. We can see that *plug-in* $\widehat{VaR}_\alpha$ underestimates risk, indicating smaller losses than we should really expect with α% probability. Thus, for instance, for a random sample of size 25, the maximum expected loss with 95% probability or, equivalently, the minimum loss with a 5% is not 1.844, but 1.949.

The calculation of *probability-unbiased* VaR is particularly relevant for small sample sizes, when the difference in the estimation of VaR is higher than for large samples, for which the *probability-unbiased* $\widehat{VaR}_\alpha$ and the *plug-in* $\widehat{VaR}_\alpha$ are very similar.

Now, we follow the second approach described in Section 1.3, to obtain the *probability-unbiased* VaR estimator by calculating the standard deviation $\widehat{\sigma}_{pu}$ of the distorted distribution function F.

If F is a Normal distribution, the *probability-unbiased* VaR estimator can be written in two alternative ways:

$$\widehat{VaR}_\alpha = \widehat{\mu}_{pu} + \widehat{\sigma}_{pu} z_\alpha = \widehat{\mu} + \widehat{\sigma} z_{\alpha_{pu}}$$

that illustrate the two equivalent approaches to *probability-unbiased* VaR estimation: either we distort the quantile and use the estimated parameters or we maintain the original quantile while distorting the estimated parameters. This equation also shows that once we have calculated α_{pu} we can obtain $\sigma z_{\alpha_{pu}}$ and viceversa.

	VaR_{pu}				$VaR_{plug-in}$			
n	**0.5**	**1**	**5**	**10**	**0.5**	**1**	**5**	**10**
10	-3.227	-2.786	-1.768	-1.305	-2.409	-2.165	-1.496	-1.139
15	-2.796	-2.440	-1.570	-1.150	-2.309	-2.065	-1.400	-1.045
20	-3.206	-2.866	-2.011	-1.587	-2.839	-2.582	-1.880	-1.506
25	-3.117	-2.789	-1.949	-1.527	-2.825	-2.562	-1.844	-1.461
50	-2.925	-2.624	-1.827	-1.415	-2.783	-2.513	-1.775	-1.382
100	-2.546	-2.307	-1.666	-1.329	-2.488	-2.262	-1.645	-1.316
150	-2.621	-2.374	-1.705	-1.352	-2.581	-2.342	-1.690	-1.343
200	-2.393	-2.165	-1.547	-1.220	-2.365	-2.143	-1.537	-1.213

Table 1.3: *Probability-unbiased* $\widehat{VaR}_\alpha$ versus *plug-in* $\widehat{VaR}_\alpha$ in the case of Normal(0,1).

Since the mean of the distribution is very low in high frequency returns and it is estimated with very low precision, we can consider it to be the same for the distorted distribution as for the original distribution, i.e. $\widehat{\mu}_{pu} = \widehat{\mu}$. Then, we will calculate the standard deviation $\widehat{\sigma}_{pu}$ of the distorted distribution function implicitly so that the previous equation holds. That standard deviation will be different for every α and for each sample size (n) because $\widehat{VaR}_\alpha$ also changes with α and with n.

Table 1.4 shows the $\widehat{\sigma}_{pu}$ values obtained for different α and n. Notice that $\widehat{\sigma}_{pu}$ is greater for small sample sizes suggesting the heavier tails of the distorted distribution. For a given sample size, we obtain larger differences between $\widehat{\sigma}$ and $\widehat{\sigma}_{pu}$ for the more extreme quantiles. For a given α, we obtain greater differences between $\widehat{\sigma}$ and $\widehat{\sigma}_{pu}$ for small sample sizes. Finally, for $n = 200$ we can see that the $\widehat{\sigma}_{pu}$'s are closer to the sample standard deviation ($\widehat{\sigma} = s$) for any α and, therefore, closer to the population standard deviation, 1.

Figure 1.3 shows the true density function of a random variable N(0,1) (blue line), the density function of the Normal distribution with the parameters estimated from a random sample of size 15 extracted from a N(0,1) (red line), and the density function of the distorted estimated distribution function using the $\widehat{\sigma}_{pu}$ estimate (green line). We can see that the distorted distribution function has heavier tails, which should allow for a better fit to a distribution of most asset returns. The *probability-unbiased* $\widehat{VaR}$ (green point) indicates higher losses than *plug-in* $\widehat{VaR}$ (red point). In other words, the *plug-in* estimator underestimates risk, particularly in small size samples. Furthemore, the smaller the sample the greater the correction or adjustment needed on the probability distribution.

	$\alpha(\%)$			
n	**0.5**	**1**	**5**	**10**
10	1.299	1.248	1.147	1.111
15	1.165	1.137	1.079	1.058
20	1.172	1.152	1.109	1.093
25	1.167	1.151	1.118	1.105
50	1.138	1.130	1.114	1.108
100	0.928	0.925	0.919	0.916
150	0.972	0.970	0.966	0.964
200	0.901	0.899	0.8965	0.895

Table 1.4: Estimated standard deviations for the distorted distribution function.

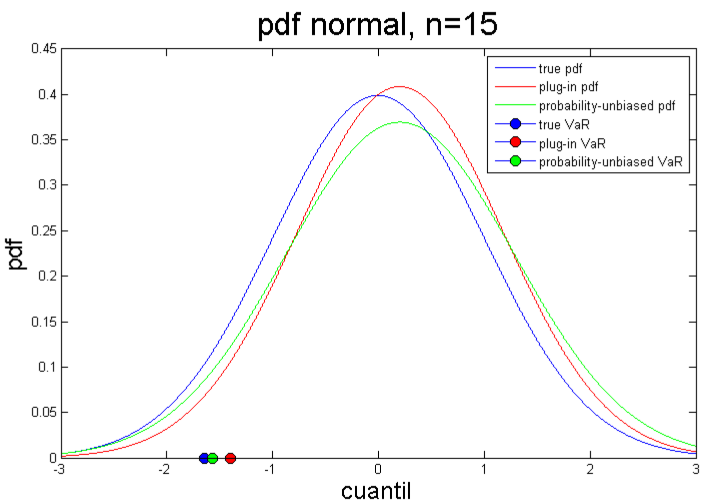


Figure 1.3: The true N(0,1) pdf (blue line), the *plug-in* pdf (red line) and the pdf of the *unbiased* cdf (green line). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *unbiased* $\widehat{VaR}_{5\%}$ (green point).

Figure 1.4 shows the cumulative distribution function of N(0,1) (blue line), the *plug-in* cumulative distribution function (red line) and the *probability-unbiased* cumulative distribution function (green line). It also displays VaR estimates at 5% significance level.

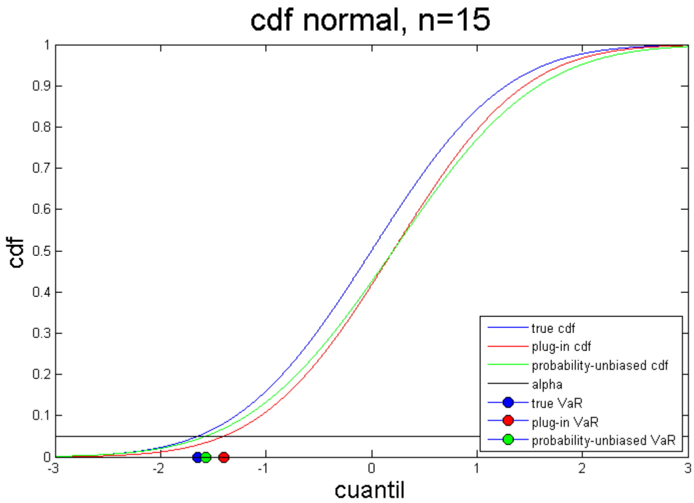


Figure 1.4: The true N(0,1) cdf (blue line), the *plug-in* cdf (red line) and the *unbiased* cdf (green line). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *unbiased* $\widehat{VaR}_{5\%}$ (green point) .

Figures 1.5 and 1.6 show pdf's and cdf's, respectively, for different sample sizes. We also show the *plug-in* $\widehat{VaR}_{5\%}$ and the *probability-unbiased* $\widehat{VaR}_{5\%}$. These Figures show the convergence of the *plug-in* distribution and the *probability-unbiased* distribution to the true distribution as the sample size increases. The pdf's and the cdf's have been represented based on a random sample of size 15, where the distortion can be easily appreciated, since the smaller the size sample the larger the distortion in the *plug-in* distribution function.

1.5.2 A comparison of probability-unbiased VaR and plug-in VaR under Normality

We want to compare the exceedance probabilities obtained from *probability-unbiased* VaR and *plug-in* VaR. According to the argument we used in the previous section to calculate an approximate $alpha_{pu}$ we expect to obtain a number of exceedances for *probability-unbiased* VaR to be very close to the theoretical α regardless of the sample size considered. For this, we simulate the estimation of the *plug-in* VaR estimator and the *probability-unbiased* VaR estimator and calculate the exceedance probabilities.

The Monte-Carlo exercise with S simulations is performed using the following steps:

1. Set the counter of the simulation $s = 0$.

2. Increment the counter of the simulation $s = s + 1$.

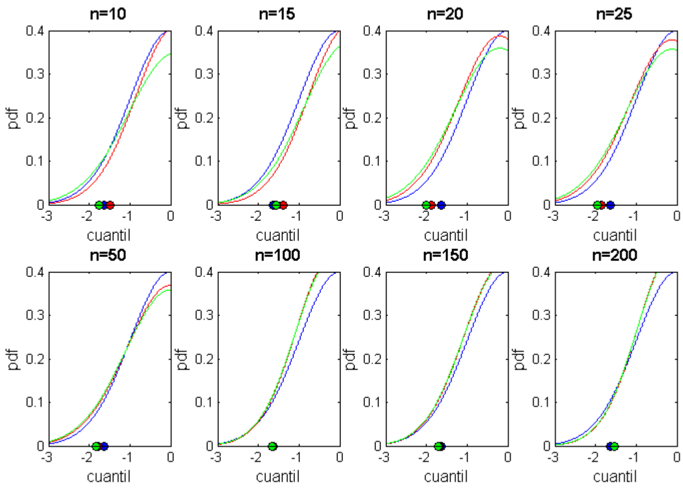


Figure 1.5: The true N(0,1) pdf (blue line), the *plug-in* pdf (red line) and the pdf of the *unbiased* cdf (green line) for different sample sizes (enlargement of the left tail). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point) .

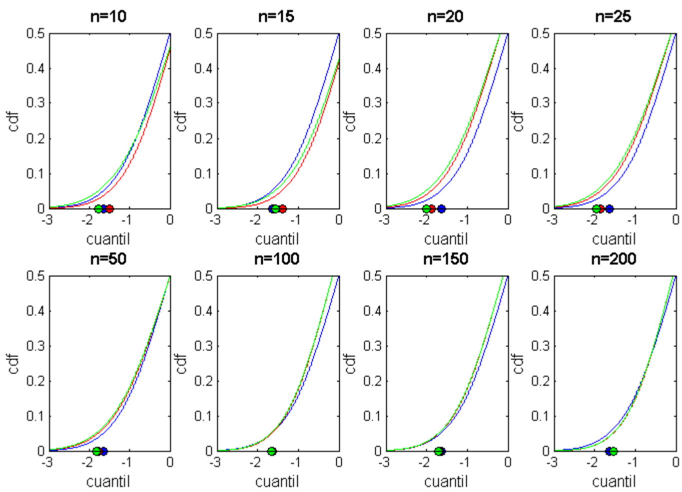


Figure 1.6: The true N(0,1) cdf (blue line), the *plug-in* cdf (red line) and the *unbiased* cdf (green line) for different sample sizes (enlargement of the left tail). Points on the horizontal axis the data points show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *unbiased* $\widehat{VaR}_{5\%}$ (green point) .

3. Simulate normally distributed data with $n+1$ observations and predefined values of μ and σ.

4. Calculate the mean and the standard deviation based on the first n data points.

5. Calculate the *plug-in* VaR estimator. CAlculate the *probability-unbiased* VaR estimator using the α_{pu} from Table 1.1 for the μ and σ estimates in step 4.

6. Check if the n+1 data point is smaller than the *plug-in* $\widehat{VaR}$ and the *probability-unbiased* estimator. When the data point is smaller (VaR exceedance) record a 1, otherwise a 0.

7. Return to step 2 while $s < S$.

8. Calculate the exceedance probability by summing the recorded values and dividing them by the number of simulations S.

The results in Table 1.5 for samples of size 10, 15, 20 and 25, for $\widehat{VaR}_{1\%}$ and $\widehat{VaR}_{5\%}$, with $\mu = 0$, $\sigma = 1$ and $S = 100000$, show that the probability of a VaR exceedence from the *probability-unbiased* estimator is close to the theoretical values of 1% and 5%. However, the probability of a VaR exceedence for the *plug-in* VaR differs than the theoretical probability. This confirms the results presented in Table 1.1. As the sample size increases, the probability of an excess from the *plug-in* VaR estimator calculated from the simulations approaches the theoretical value. For the *probability-unbiased* VaR estimator, that probability remains similar to the theoretical probability for all sample sizes.

n	$1\% plug-in$	$1\% pu$	$5\% plug-in$	$5\% pu$
10	2.674	0.985	7.567	5.015
15	2.063	0.953	6.748	5.047
20	1.825	1.028	6.268	5.038
25	1.608	0.999	5.978	4.930

Table 1.5: Shortfall probabilities $\alpha\%$ that the next observation will be lower than the *plug-in* $\widehat{VaR}$ and the *probability unbiased* $\widehat{VaR}$ in the Monte-Carlo simulation for the Normal distribution.

1.5.3 Bootstrapping estimation of probability-unbiased VaR

When sampling from a Normal distribution function (F_θ), FH [20] propose an algorithm to replace the level α by a suitably chosen level α_{pu} so as to minimize the average distance between the bootstrapped estimators and α. The α_{pu} obtained through resampling will change with the sample size (n), the confidence level α and the observed values in the sample. The algorithm approximates the α_{pu} level and achieves an approximation to the *probability-unbiased* VaR through a modification of the confidence level. The change from

α to α_{pu} corrects for the fact that we do not observe infinite realizations. For a large number of observations the *plug-in* estimator and the *probability-unbiased* estimator become very similar. The *plug-in* estimator has good properties only asymptotically, while the *probability-unbiased* estimator is a good estimator even in short samples.

Suppose we have a random sample of size n drawn from a distribution F_θ. Then, we generate B resamples of the same size n. These resamples are obtained by sampling with replacement. The steps to be performed are:

1. From observed values $X_1, ..., X_n \sim^{i.i.d.} F_\theta$
2. Calculate $\widehat{\theta} = \widehat{\theta}(X_1, ..., X_n)$
3. For i=1:B
   Samples $X_1^*, ..., X_n^*$ from $F_{\widehat{\theta}}$ [1]
   Calculate $\widehat{\theta}_i^*$
   Find the α_{pu} that minimizes the following objective function

$$\alpha_{pu} = argmin_\gamma \left| \frac{1}{B} \sum_{i=1}^{B} F_{\widehat{\theta}}(F^{-1}_{\widehat{\theta}_i^*}(\gamma)) - \alpha \right| \tag{1.4}$$

The level of α_{pu} is chosen so that equation (1.1) is satisfied. Substituting α for α_{pu} we obtain the *probability-unbiased* VaR estimator.

We start with a random sample of size n generated from a Normal distribution with mean 0 and standard deviation 1. From this original random sample we obtain 10,000 resamples of size n. As we increase the sample size, the Maximum Likelihood estimates of mean and standard deviation of the original random sample, μ_y and σ_y, tend to the population average ($\mu_x = 0$) and the population standard deviation ($\sigma_x = 1$). For each resample we estimate the mean and the standard deviation, obtaining 10000 means and 10000 standard deviations. These estimates are used to find the α_{pu} that minimizes the objective function (1.4). Table 1.6 shows the α_{pu} probabilities obtained under the bootstrap algorithm proposed by FH. They change with sample size (n) and with the value

[1] Even though this is unclear in FH, we understand that these samples are resamples of the original sample (our observed values) for three reasons: 1) if the samples were generated from the distribution function estimated with the original sample, we would obtain from each resample very different values of $\widehat{\theta}_i^*$, especially with small sample sizes, and we would have to draw many samples to obtain suitable results. Indeed, even performing 100000 samples from $F_{\widehat{\theta}}$ we have not obtained the expected results, and α_{pu} does not tend to α when n tends to ∞, 2) we have just one random sample, possibly of small size, and we cannot use classical statistical inference to find the sampling distribution because we do not know the parameters of the population distribution and we cannot take the estimated parameters as population parameters. Therefore, to find the sampling distribution, at least approximately, we create many resamples by repeatedly sampling with replacement from the original random sample. Each resample has the same size as the original random sample, and 3) a bootstrap algorithm as mentioned by FH is based on a large number of new samples obtained by sampling from the original sample, not by simulation.

of α [2]. Notice that $\alpha_{pu} \longrightarrow \alpha$ as $n \longrightarrow \infty$. Comparing Table 1.1 with Table 1.6, we can see that the α_{pu} values obtained by parametric approach and the bootstrap algorithm are very similar. In Table 1.6 these values change with the sample. The α_{pu} values from both tables differ from each other just for small sample sizes. As $n \longrightarrow \infty$ both α_{pu} values converge towards the theoretical α.

	$\alpha(\%)$			
n	**0.5**	**1**	**5**	**10**
10	0.021	0.091	2.167	6.402
15	0.049	0.122	2.536	7.010
20	0.154	0.425	3.613	8.362
25	0.169	0.453	3.652	8.365
50	0.319	0.715	4.358	9.245
100	0.405	0.854	4.686	9.635
150	0.429	0.892	4.775	9.740
200	0.457	0.936	4.864	9.839

Table 1.6: Probabilities α_{pu} to be used to obtain a *probability-unbiased* VaR_α for different values of α and n in the i.i.d. Normal distribution case using bootstrapping proposed by FH.

At the beginning of Section 1.5 we have shown that for a Normal distribution it is possible to obtain a closed-form solution for *probability-unbiased* VaR. But in this subsection we have reproduced the FH argument using a Monte Carlo simulation algorithm to calculate α_{pu} for a Normal distribution and we have compared the results obtained from the analytical and the numerical approaches. For other distributions for which the probability distributions of estimated parameters are either unknown or they are difficult to obtain, we suggest using this same bootstrap algorithm. In particular, we will use the above routine to calculate *probability-unbiased* VaR for Student-t distributions a well as for a mixture of two Normal distributions.

1.5.4 Approximate probability-unbiased confidence intervals for VaR under Normality

The numerical value of an estimator provides a point estimate of the statistical parameter under study. It is often preferred to estimate by a probability/confidence interval, since

[2] From the α_{pu} obtained by the bootstraping algorithm proposed by FH, we obtain *probability-unbiased* VaR and the distortion functions, and we can check by Monte-Carlo simulation the probability of an excess obtained with the *probability-unbiased* VaR and compare it with that obtained with *plug-in* VaR for different sample sizes. Since these analysis are just a reproduction of the simulation analysis in FH, we leave them to Appendix A.

the amplitude of the interval provides information about the possible estimation error (or about the level of uncertainty on the true value of the parameter, under a Bayesian interpretation).

In this section we construct a confidence interval for the VaR estimate. We want to estimate the confidence interval $(1-2\alpha_{CI})$ of the form $I = [CI_{low}, CI_{up}]$. The exact coverage probability indicates that

$$P(g(\theta) \in I) = 1 - 2\alpha_{CI}$$

FH propose again using the bootstrap method to calculate the confidence interval:

1. From observed values $X_1, ..., X_n \sim^{i.i.d} F_\theta$
2. Calculate $\widehat{\theta} = \widehat{\theta}(X_1, ..., X_n)$
3. For i=1:B
   Sample $X_1^*, ..., X_n^*$ of $F_{\widehat{\theta}}$
   Calculate $\widehat{\theta}_i^* = \widehat{\theta}_i^*(X_1^*, ..., X_n^*)$
   Calculate $g(\widehat{\theta}_i^*)$
4. Calculate an estimate $\widehat{G}^*$ of G, the distribution of $g(\widehat{\theta})$, with the help of the bootstrap samples $g(\widehat{\theta}_1^*), ...g(\widehat{\theta}_B^*)$
5. Calculate $\widehat{I}$ with the help of the distribution $\widehat{G}^*$

   $$\widehat{CI}_{low} = (\widehat{G}^*)^{-1}(\alpha_{CI}), \widehat{CI}_{up} = (\widehat{G}^*)^{-1}(1-\alpha_{CI})$$

   for a symmetric confidence interval.

The bootstrap distribution $(\widehat{G}^*)$ of the estimator $g(\widehat{\theta})$ based on many resamples approximates the sampling distribution of the estimator that we could obtain based on many samples, if they were available. The bootstrap distribution of VaR is obtained from VaR estimates from many resamples and provides information about the sampling distribution of VaR when it is unknown. The original sample represents the population from which it was drawn. So resamples from this sample represent what we would get if we took many samples from the population. As we increase the number of resamples B, interval estimation converges to the *plug-in* estimates rather than to the *probability-unbiased* estimates because the *plug-in* estimators were used in step 3.

Both bounds are estimated separately to obtain the probability-unbiased confidence interval. We calculate the upper bound CI_{up} as

$$P(CI < g(\theta)) = \alpha_{CI}$$

The respective lower bound CI_{low} can be found by replacing α_{CI} by $1-\alpha_{CI}$.

The aim is to calculate a probability-unbiased estimator for CI with respect to $g(\theta)$,

$$P(\widehat{CI} < g(\theta)) = \alpha_{CI}$$

We thus replace again α_{CI} by $\alpha_{CI,pu}$ and we obtain the estimator

$$\widehat{CI} = (\widehat{G}^*)^{-1}(\alpha_{CI,pu})$$

where $\widehat{G}$ is the estimate of G.

To estimate an approximation to $\alpha_{CI,pu}$ a bootstrap algorithm is used, following the steps:

1. Repeat B_2 times steps 2 and 3 of the bootstrap algorithm to obtain realizations $\widehat{G}_1^*, ..., \widehat{G}_{B_2}^*$ de $\widehat{G}$

2. Find $\alpha_{CI,pu}$ such that

$$\alpha_{CI,pu} = argmin_{\gamma} \left| \frac{1}{B_2} \sum_{i=1}^{B_2} 1_{(\widehat{G}_i^*)^{-1}(\gamma) < g(\widehat{\theta})} - \alpha_{CI} \right| \tag{1.5}$$

   where $1_{(\widehat{G}_i^*)^{-1}(\gamma) < g(\widehat{\theta})}$ is the indicator function taking the value 1 if the VaR falls outside the interval and it takes the value 0 if the VaR is inside the interval. In other words, $g(\widehat{\theta})$ is exceeded in mean with probability approximately $1 - \alpha_{CI}$ by the $\alpha_{CI,pu}$ quantile of $\widehat{G}_i$, $i = 1, ..., B$.

It is important to note that for the respective lower bound, we cannot just take the $1 - \alpha_{CI,pu}$ quantile of the estimated distribution $\widehat{G}^*$, because the distribution G is non-symmetric. We need to repeat the above algorithm for CI_{low} replacing α_{CI} for $1 - \alpha_{CI}$. That will yield an interval estimate with a confidence level of $1-2\alpha_{CI}$ where, for example, $\alpha_{CI} = 0.05$.

Using the bootstrap algorithm, we obtain the distribution of $\widehat{VaR}_{5\%}$ on which we calculate the confidence interval within which the true VaR will be find. Table 1.7 shows the *plug-in* interval range for a 90% confidence level. We observe that as the sample size increases, the confidence interval narrows. This table shows the distortions for *probability-unbiased* confidence intervals in the case where we use the plug-in VaR estimator in the objective function (1.5) of the bootstrap algorithm.

To calculate the *probability-unbiased* confidence interval, we need to obtain the $\alpha_{CI,pu}$. The $\alpha_{CI,pu}$ is the value that minimizes the objective function (1.5). This is the upper bound of the interval. For the lower bound we should perform the bootstrap algorithm again, in this case replacing α_{CI} for $1 - \alpha_{CI}$.

Table 1.8 shows the distortions for *probability-unbiased* confidence intervals in the case where we use the *plug-in* VaR estimator in step 3 of the bootstrap algorithm, i.e. $\alpha_{\widehat{VaR}} = \alpha$. For example, the lower bound of the 90% confidence level estimated from 20 data points is the 1.40% sample percentile. It will be exceeded with a probability of 5%. If this result is subtracted from the upper bound exceedence probability, we obtain a coverage probability for the interval of (62.33% – 1.40% = 60.93%). We can see that

n	α_{CI}	$1-\alpha_{CI}$	CI_{low}	CI_{up}
10	5	95	-3.1084	-0.5513
15	5	95	-1.8305	-0.7546
20	5	95	-1.5809	-0.6599
25	5	95	-1.8869	-1.0263
50	5	95	-1.6890	-1.0249
100	5	95	-2.0458	-1.3302
150	5	95	-1.6067	-1.2543
200	5	95	-1.6666	-1.3640

Table 1.7: Probability $\alpha_{CI}(\%)$, $1-\alpha_{CI}(\%)$ and quantiles corresponding to the lower and upper bounds of confidence interval 90% for a Normal distribution.

n	$\alpha_{CI,low,pu}$	$\alpha_{CI,up,pu}$	$CI_{low,pu}$	$CI_{up,pu}$
10	1.14	45.11	-3.5659	-2.1695
15	1.12	58.32	-2.0367	-1.2789
20	1.40	62.33	-1.6703	-1.0033
25	1.01	70.58	-2.0454	-1.3337
50	1.42	81.06	-1.7473	-1.1800
100	1.76	91.45	-2.0905	-1.4472
150	1.01	94.07	-1.6191	-1.2546
200	1.42	90.35	-1.7220	-1.4018

Table 1.8: Probabilities $\alpha_{CI,pu}(\%)$ such that $\alpha_{CI} = P(\widehat{CI} < VaR)$ with $\alpha_{\widehat{VaR}} = \alpha = 5\%$ in the $\widehat{VaR}$ and quantiles corresponding to the lower and upper bounds of confidence interval 90% in the case of the Normal distribution.

the confidence interval is shifted to the left relative to the standard (plug-in) confidence interval but not symmetrically because almost all exceedances occur at the lower bound.

As the sample size increases the distortion of the confidence interval is smaller and the range narrows.

1.6 Student-t Distribution

1.6.1 Probability-unbiased VaR estimator for a Student-t distribution

Let us assume that we have a finite-short sample from a Student-t distribution function (F_θ). Since we do not know the distribution of the number of degrees of freedom, we will use the bootstrap algorithm proposed by FH and defined in subsection 1.5.4.

Following the first approach described in section 1.3, the VaR estimator can be obtained using a modification on the α-quantile from the estimated distribution. If F is a Student-t distribution, the VaR estimator can be written

$$\widehat{VaR}_\alpha = t^{-1}(\alpha_{pu})$$

where α_{pu} is chosen so that the equation $E_\theta[P(X_{n+1}) < \widehat{Q}_\alpha)] = \alpha$ is satisfied. The α_{pu} approximation is obtained by a bootstrap algorithm. The change of α for α_{pu} corrects for the fact that we do not observe infinite realizations. The *probability-unbiased* estimator satisfies the unbiasedness condition by construction for any sample size, including small samples, while the *plug-in* estimator is *probability-unbiased* only when $n \longrightarrow \infty$

We start with a random sample of size n generated from a Student-t distribution with 2 degrees of freedom. This will be the original random sample from which 10,000 resamples of size n will be generated. The parameter to be estimated in this distribution is the number of degrees of freedom. We have used the method of moments estimator:

$$\nu = \frac{2\sigma^2}{\sigma^2 - 1}$$

because of its simplicity, although it requires that the distribution has a variance greater than 1. For each resample, we estimate the number of degrees of freedom, which is then used to find the $\alpha_{pu}$by solving the optimization problem:

$$\alpha_{pu} = argmin_\gamma \left| \frac{1}{B} \sum_{i=1}^{B} F_{\widehat{\theta}}(F^{-1}_{\widehat{\theta}^*_i}(\gamma)) - \alpha \right|$$

Our estimates in Table 1.9 show that $\alpha_{pu} \longrightarrow \alpha$ as $n \longrightarrow \infty$. The convergence to the *plug-in* VaR estimator under the Student-t distribution is faster than under the Normal, and for small sample sizes we obtain an α_{pu} closer to the theoretical α than under the Normal distribution. This is because the higher kurtosis of the Student-t distribution makes more likely the occurrence of extreme events, so that the correction needed on α is smaller.

	$\alpha(\%)$			
n	**0.5**	**1**	**5**	**10**
10	0.071	0.332	3.865	8.903
15	0.227	0.639	4.470	9.490
20	0.357	0.798	4.663	9.668
25	0.363	0.804	4.656	9.654
50	0.373	0.821	4.701	9.706
100	0.389	0.833	4.784	9.776
150	0.450	0.926	4.862	9.858
200	0.445	0.919	4.855	9.853

Table 1.9: Probabilities $\alpha_{pu}(\%)$ to be used to obtain a *probability-unbiased* VaR_α for different values of α and n in the i.i.d. Student-t distribution case.

Table 1.10 shows the *probability-unbiased* $\widehat{VaR}_\alpha$ (VaR_{pu}) and the *plug-in* $\widehat{VaR}_\alpha$ (the standard VaR estimator) for different sample sizes and α's. Notice that the *plug-in* $\widehat{VaR}_\alpha$ underestimates risk for any probability level $\alpha\%$. As with the Normal distribution, the calculation of *probability-unbiased* VaR is specially relevant for small sample sizes although in this case, differences between both VaR estimates are much greater than under a Normal distribution.

The distortion function in Figure 1.7 (red line) shows the decrease in the size of the VaR correction needed for unbiasedness as the number of sample observations increases, eventually converging to 1 (black line). The same evidence arises from the distortion of the quantiles of the Student-t distribution function in Figure 1.8.

Table 1.11 presents the reverse question of Table 1.10: What is the α associated to a given α_{pu}? Now, at 5% significance and $n = 20$, the *pu*-VaR estimate would have a 5.34% probability of being exceeded by a future sample observation from the full distribution of returns. We observe that differences are greater when we have small sample sizes and we can also observe that $\alpha \to \alpha_{pu}$ when $n \to \infty$.

We can now use the second approach described in Section 1.3 for the computation of the *probability-unbiased* VaR estimator, distorting the estimated parameter instead of the level of α. Since the *probability-unbiased* VaR estimator can be written:

$$\widehat{VaR}_\alpha = t^{-1}_{\widehat{\nu}_{pu}}(\alpha)$$

the, once we have the *probability-unbiased* VaR estimate, we can solve for $\widehat{\nu}_{pu}$. Estimates differ for every α and for each sample size (n), as shown Table 1.12. We observe that for

	VaR_{pu}				$VaR_{plug-in}$			
n	**0.5**	**1**	**5**	**10**	**0.5**	**1**	**5**	**10**
10	-17.102	-8.949	-2.975	-1.885	-7.515	-5.564	-2.609	-1.752
15	-14.275	-8.533	-3.078	-1.940	-9.650	-6.808	-2.887	-1.872
20	-7.928	-5.716	-2.612	-1.750	-6.919	-5.205	-2.522	-1.714
25	-9.530	-6.635	-2.815	-1.838	-8.247	-5.997	-2.709	-1.796
50	-7.437	-5.448	-2.556	-1.727	-6.634	-5.031	-2.479	-1.695
100	-8.666	-6.198	-2.736	-1.805	-7.762	-5.711	-2.643	-1.767
150	-8.372	-6.054	-2.715	-1.798	-7.990	-5.846	-2.674	-1.781
200	-8.512	-6.128	-2.728	-1.804	-8.078	-5.898	-2.686	-1.786

Table 1.10: *Probability-unbiased* $\widehat{VaR}_\alpha$ versus *plug-in* $\widehat{VaR}_\alpha$ in the case of Student-t distribution.

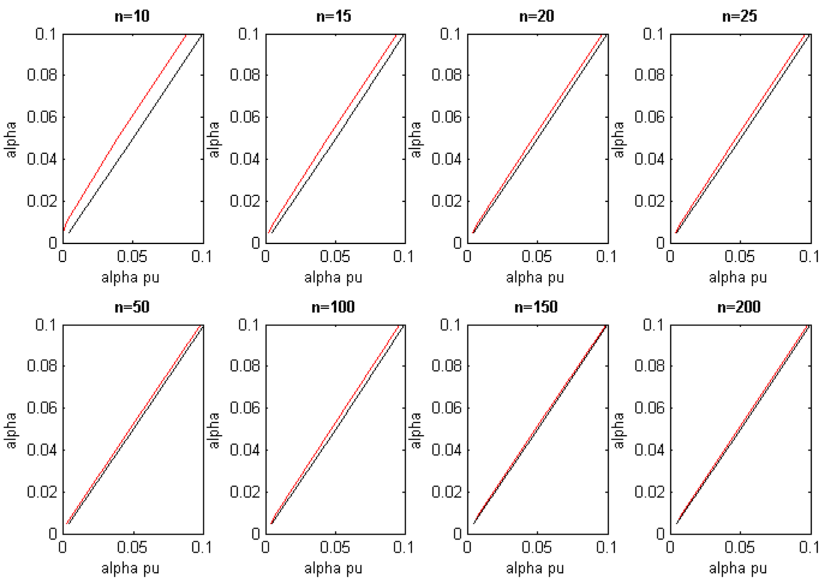


Figure 1.7: Distortion function for the Student-t distribution. The diagonal (black line) represents no distorsion.

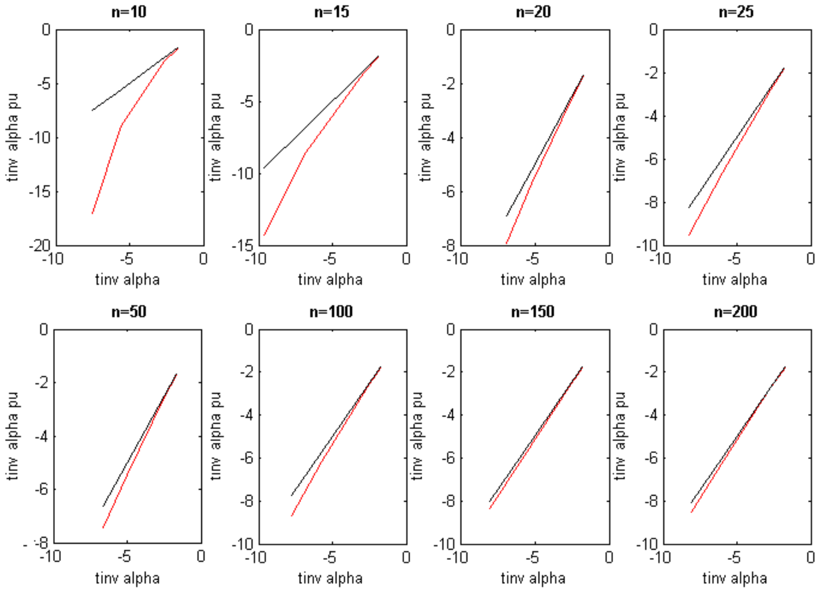


Figure 1.8: Quantiles of the Student-t cdf versus the quantiles of the distorted Student-t cdf. The diagonal (black line) represents no distortion.

	$\alpha_{pu}(\%)$			
n	**0.5**	**1**	**5**	**10**
10	1.248	1.895	6.149	11.068
15	0.838	1.404	5.532	10.504
20	0.666	1.221	5.339	10.329
25	0.658	1.215	5.347	10.344
50	0.645	1.193	5.301	10.294
100	0.628	1.183	5.319	10.322
150	0.553	1.077	5.139	10.141
200	0.558	1.083	5.145	10.146

Table 1.11: Shortfall probabilities $\alpha(\%)$ for the next observation being lower than the plug-in VaR estimate $t^{-1}(\alpha_{pu})$.

a given sample size, we obtain a lower number of degrees of freedom for the most extreme quantile, due to the fact that the estimated distribution function has thicker tails than the *plug-in* estimator.

	$\alpha(\%)$			
n	**0.5**	**1**	**5**	**10**
10	1.526	1.696	1.947	2.001
15	1.653	1.745	1.858	1.879
20	2.316	2.350	2.403	2.416
25	2.049	2.074	2.114	2.124
50	2.428	2.457	2.504	2.516
100	2.178	2.190	2.215	2.222
150	2.229	2.234	2.244	2.246
200	2.204	2.211	2.224	2.227

Table 1.12: Degrees of freedom estimated for the distorted distribution function.

Figure 1.9 shows the true density function of a random variable t(2) (blue line), the density function of the Student-t distribution estimated from a random sample of size 15 extracted from a t(2) distribution (red line), and the density function of the distorted estimated distribution function (green line). As with the Normal, the distorted distribution function has thicker tails than the density function of the random variable of size 15 (the *plug-in* pdf).

The *probability-unbiased* $\widehat{VaR}$ (green point) indicates higher losses than the *plug-in* $\widehat{VaR}$ (red point), a result that is consistent with the fact that *plug-in* VaR underestimates risk.

Figure 1.10 shows the distribution function of t(2) (blue line), the *plug-in* distribution function (red line) and the *probability-unbiased* distribution function (green line). We also show their 5% VaR estimates.

Figures 1.11 and 1.12 show pdf's and cdf's, respectively, for different sample sizes. We also present *plug-in* $\widehat{VaR}_{5\%}$ and the *probability-unbiased* $\widehat{VaR}_{5\%}$. These figures show the convergence of the *plug-in* distribution and the *probability-unbiased* distribution to the true distribution when increasing the size of the random sample. As discussed above, such convergence is faster than with the Normal, i.e. the distortion needed by the Student-t distribution for each n and for each α is smaller than for the Normal distribution.

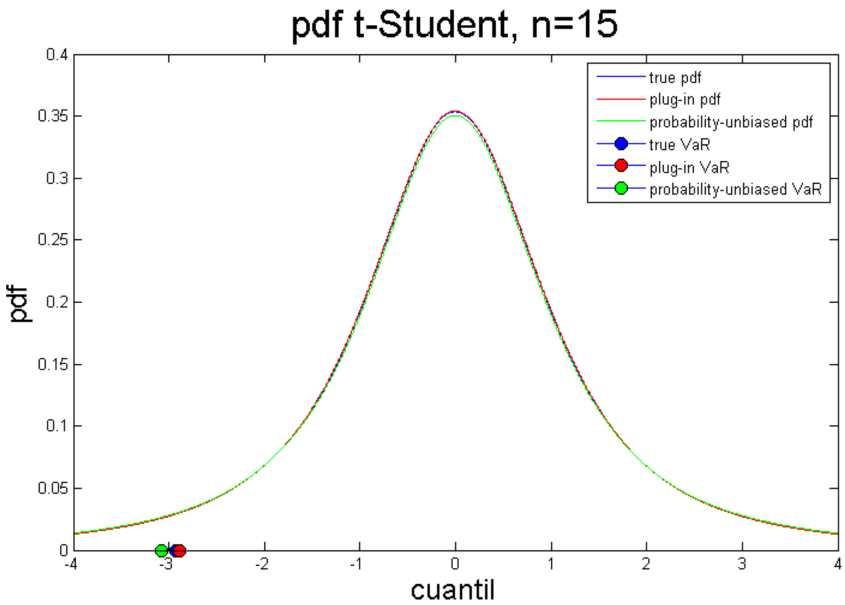


Figure 1.9: The true t(2) pdf (blue line), the *plug-in* pdf (red line) and the pdf of the *probabilty-unbiased* cdf (green line). On the horizontal axis the data points show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point) .

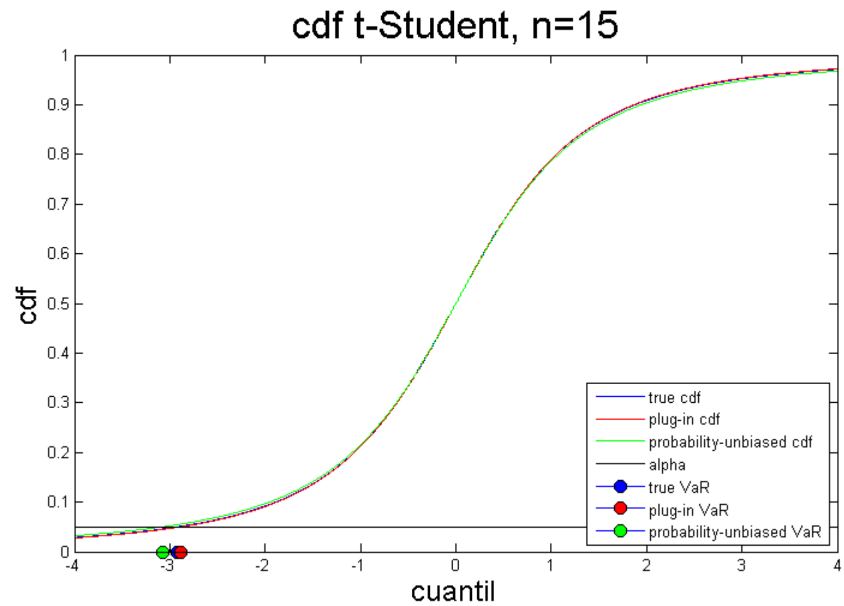


Figure 1.10: The true t(2) cdf (blue line), the *plug-in* cdf (red line) and the *probability-unbiased* cdf (green line). On the horizontal axis the data points show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point).

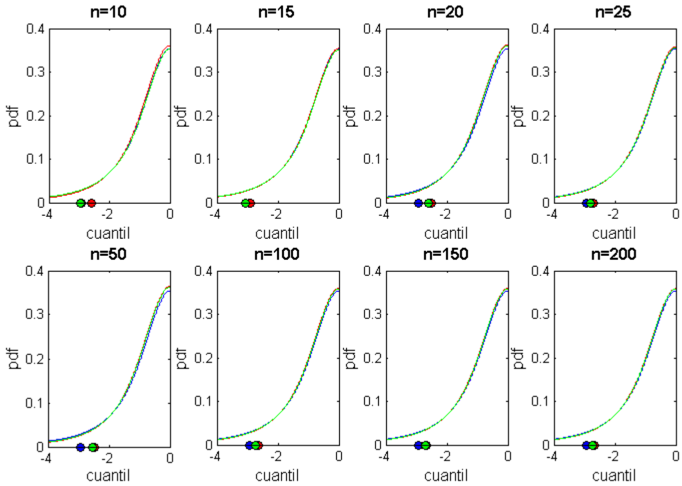


Figure 1.11: The true t(2) pdf (blue line), the *plug-in* pdf (red line) and the pdf of the *unbiased* cdf (green line) for different sample sizes (enlargement of the left tail). On the horizontal axis the data points show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *unbiased* $\widehat{VaR}_{5\%}$ (green point).

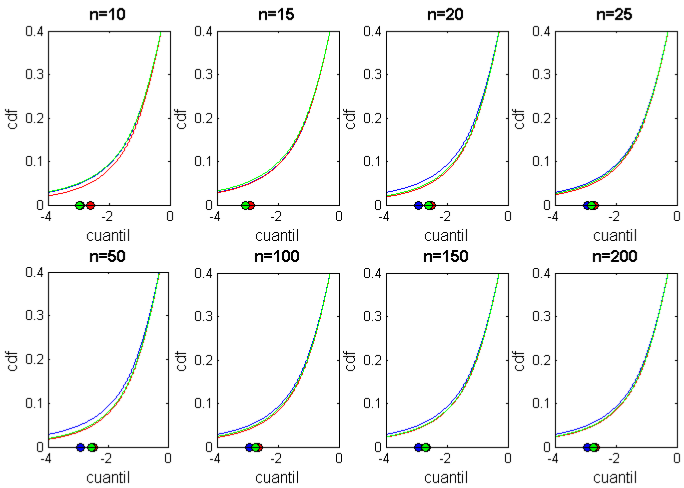


Figure 1.12: The true t(2) cdf (blue line), the *plug-in* cdf (red line) and the *unbiased* cdf (green line) for different sample sizes (enlargement of the left tail). On the horizontal axis the data points show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *unbiased* $\widehat{VaR}_{5\%}$ (green point).

1.6.2 A comparison of probability-unbiased VaR and plug-in VaR under Student-t

In order to test our theoretical calculations in a controlled simulation, we simulate the estimation of the *plug-in* VaR estimator and the *probability-unbiased* VaR estimator and calculate the exceedence probabilities. The results in Table 1.13, for samples of size 10, 15, 20 and 25, $\widehat{VaR}_{1\%}$ and $\widehat{VaR}_{5\%}$, with $gl = 2$ and $S = 100000$, show that the VaR exceedence of the *probability-unbiased* estimator is close to the theoretical values of 1% and 5%. However, the VaR exceedence of the *plug-in* VaR differs more than the theoretical probability.

As the sample size increases, the probability of an excess for the *plug-in* VaR estimator calculated by simulation approaches its theoretical value. However, in the case of *probability-unbiased* VaR estimator, that probability remains broadly similar to the theoretical probability for all sample sizes. However, the number of exceedances is not as close to the theoretical number as in the case of the Normal distribution when the probability-unbiased VaR was calculated.

n	$1\% plug-in$	$1\% pu$	$5\% plug-in$	$5\% pu$
10	2.615	1.489	7.459	6.221
15	2.498	1.909	7.159	6.627
20	2.431	2.109	7.073	6.735
25	2.296	1.965	7.091	6.732

Table 1.13: Shortfall probabilities $\alpha\%$ that the next observation is always lower than the *plug-in* $\widehat{VaR}$ and the probability *probability-unbiased* $\widehat{VaR}$ in the Monte-Carlo simulation for Student-t distribution.

1.6.3 Approximate probability-unbiased confidence intervals for VaR under Studen-t

In this section we construct an interval estimate with a confidence level of $1 - 2\alpha_{CI}$ using the bootstrap methodology. We present results for $\alpha_{CI} = 0.05$. Table 1.14 shows the *plug-in* range of values at a confidence level of 90%. This is the standard confidence interval.

To calculate the *probability-unbiased* confidence interval we should obtain the $\alpha_{CI,pu}$ that minimizes the objective function

$$\alpha_{CI,pu} = argmin_{\gamma} \left| \frac{1}{B_2} \sum_{i=1}^{B_2} 1_{(\widehat{G}_i^*)^{-1}(\gamma) < g(\widehat{\theta})} - \alpha_{CI} \right|$$

where $1_{(\widehat{G}_i^*)^{-1}(\gamma) < g(\widehat{\theta})}$ is the indicator function that signals whether the VaR is out the confidence interval. This way, we obtain the upper bound of the range. For the lower

n	α_{CI}	$1-\alpha_{CI}$	CI_{low}	CI_{up}
50	5	95	-2.6304	-2.4079
75	5	95	-2.7098	-2.5021
80	5	95	-2.8457	-2.3540
90	5	95	-2.7791	-2.3202
100	5	95	-2.6606	-2.3685
200	5	95	-2.7659	-2.3314

Table 1.14: Probability $\alpha_{CI}(\%)$, $1-\alpha_{CI}(\%)$ and quantiles corresponding to the lower and upper bounds of a 90% *plug-in* confidence interval in the case of the Student-t distribution.

bound we must perform again in its entirety the bootstrap algorithm replacing α_{CI} by $1-\alpha_{CI}$.

Table 1.15 shows the 90% *probability-unbiased* confidence interval. The confidence interval is again shifted to the left relative to the *plug-in* confidence interval but not symmetrically because almost exceedances occur at lower bound. In this case the leftward shift of the confidence interval is smaller than for the Normal distribution. We can see again that the confidence interval narrows when we increase the sample size.

n	$\alpha_{CI,low,pu}$	$1-\alpha_{CI,up,pu}$	$CI_{low,pu}$	$CI_{up,pu}$
50	1.03	33.71	-2.6470	-2.5466
75	1.06	50.23	-2.7220	-2.6171
80	1.26	3.25	-2.8610	-2.8501
90	1.15	9.83	-2.7880	-2.7493
100	1.37	50.52	-2.6826	-2.5288
200	1.07	13.02	-2.8052	-2.7332

Table 1.15: Probabilities $\alpha_{CI,pu}(\%)$, $\alpha_{CI} = P(\widehat{CI} < VaR)$, with $\alpha_{\widehat{VaR}} = \alpha = 5\%$. Columns 4 and 5 show the quantiles corresponding to the lower and upper bounds of the 90% confidence interval in the case of the Student-t distribution.

1.7 Mixture of two Normal distributions

Definition 3 *A distribution is a mixture of two probability distributions when its distribution function can be written as:*

$$F(x) = pF_1(x) + (1-p)F_2(x)$$

where F_1 and F_2 are distribution functions and p is a value between zero and one representing the probability that the element x comes from the distribution F_1.

The mean of the mixture is

$$m_{mix} = pm_1 + (1-p)m_2$$

And the variance

- *If $m1 = m2$*

$$\sigma^2_{mix} = p\sigma^2_1 + (1-p)\sigma^2_2$$

- *If $m1 \neq m2$*

$$\sigma^2_{mix} = p\sigma^2_1 + (1-p)\sigma^2_2 + p(1-p)(m_1 - m_2)^2$$

Strictly speaking, any distribution observed in practice may be regarded as a mixed distribution. Unlike the linear combination of two Normal distributions, the mixture of two Normal distributions, does not follow a Normal distribution, since it is obtained by taking data at random from the first Normal distribution with probability p and from the second one with probability $1-p$.

A mixture could be understood as the modification produced on a distribution (the second distribution used in the mixture, in this chapter) due to the influence of another one (the first distribution). When we increase the value of the mixing parameter p, we increase the influence of the the first distribution on the second one [3].

1.7.1 Casuistry of Mixture of two Normal distribution

Different distributions may be obtained according to the Normal distributions used in the mixture and according to the mixing parameter used. We consider four mixture distribution from two Normal distributions with different means and different standard deviation. The four mixture distributions generated are: a mixture of two Normal with equal μ and similar σ, with equal μ and different σ, with different μ and similar σ and with different μ and different σ. We show here the last one, while the rest are described in Appendix B.

Mixture of two Normal distributions with different μ and different σ.

We take as an example a mixture of a Normal distribution with mean -5 and standard deviation of 10 and a Normal distribution with mean 0 and standard deviation 1. When

[3]We provide more description about mixtures in Appendix B.

the mixing parameter is $p = 0.1$, this mixture has more extreme values in the left tail than other mixtures that we consider in Appendix B. This is because the first Normal distribution has a smaller mean and a higher standard deviation.

As we increase the mixing parameter we raise the influence of the second Normal distribution on the first one. When we reach $p = 0.9$ we obtain a mixture distribution that is similar to the first Normal distribution modified by a few values taken from the second Normal distribution. These few values will mainly affect the right side of the distribution because the mean of the second Normal distribution is to the right of the mean of the first Normal distribution (see Figure 1.13).

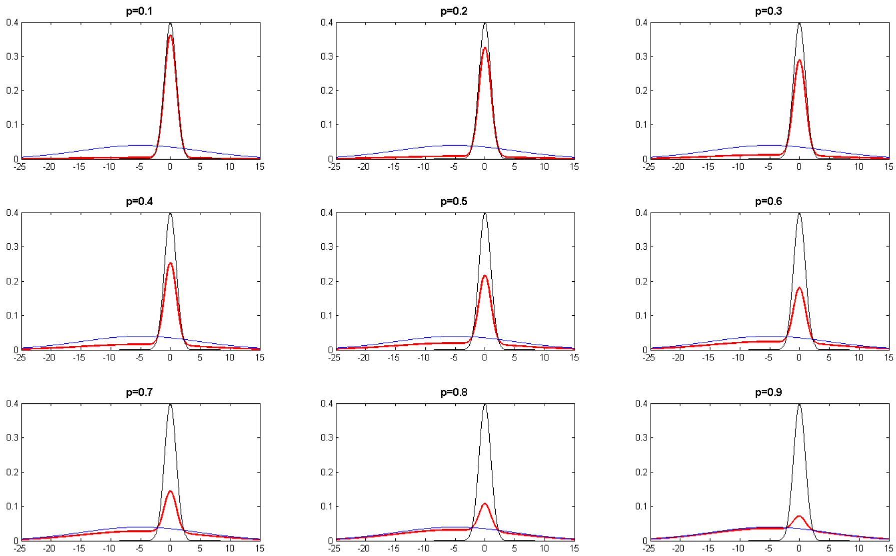


Figure 1.13: Density function of the mixture of N1 (-5,10) and N2 (0,1) (red line), density function for N1 (-5,10) (blue line) and density function for N2 (0,1) (black line) for different values of the mixing parameter.

Figure 1.14 clearly shows that the mixture distribution is not a Normal distribution. The mixture has less central values than the Normal distributions that are mixed, especially for intermediate p values, when most values fall at the extremes.

1.7.2 Probability-unbiased VaR estimator for Mixtures of Normal distributions

In this section we estimate the *probability-unbiased* VaR estimator for four different mixtures of Normal distributions introduced in the previous section.

We begin with a mixture of two Normal distributions with equal mean and similar standard deviation, specifically a mixture of a Normal(0,2) and a Normal(0,1) with a mixing parameter $p = 0.1$. We use a small p to obtain a mixture distribution not too different from a Normal distribution, although the mixture will incorporate some atypical data from

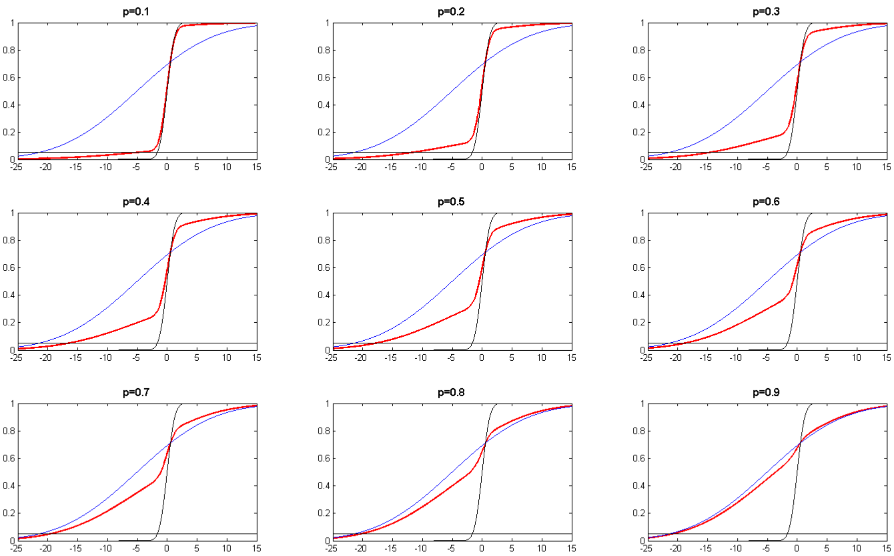


Figure 1.14: Distribution function of the mixture of N1(-5,10) and N2(0,1) (red line), distribution function for N1(-5,10) (blue line) and distribution function for N2 (0,1) (black line) for different values of the mixing parameter. The black horizontal line indicates $\alpha = 0.05$.

the first Normal distribution. Sample sizes are 100, 200, 300 and 400 observations. The quantiles of a mixture distribution do not accept a closed form solution but rather, they require solving an implicit equation. Therefore, to calculate the VaR we cannot use the parametric approach. We then need to work with samples larger than in the case of Normal and t-Student distributions because both *plug-in* VaR and *probability-unbiased* VaR are now calculated as a sample percentile. For example, if we want to compute the 1% percentile, we must compute it from a sample of considerable size to avoid that it might fall outside the data range. For instance, the *prctile* function of MatLab would return the first value of the sample, in spite of the fact that the first value might be significantly larger than the 1% percentile.

Table 1.16 shows the main sample moments from this mixture distribution. Notice that the standard deviation and the kurtosis of the mixture is higher than those of the second Normal distribution used in the mixture because of the influence of the first Normal distribution, specially when larger sample sizes allow for higher precision in estimation. However, the mixture has a virtually Normal distribution because the two Normal distributions that have been used for mixing are very similar.

Once again, to obtain the *probability-unbiased* VaR estimator, for each given α we calculate the quantile corresponding to the probability level α_{pu}. The mixture of two Normal distributions is a parametric distribution, with parameters depending on the parameters of the two mixing Normal and the mixing probability. Table 1.17 shows the *probability-unbiased* VaR, α_{pu}, that satisfies equation (1.1). Notice that the 1% and 5% *plug-in* VaR underestimate the level of risk. We can also see that in a larger sample size

MIXTURE OF N(0,2) and N(0,1)				
n	μ_{mix}	σ_{mix}	$skewness_{mix}$	$kurtosis_{mix}$
100	-0.0111	1.0427	-0.2318	2.5156
200	0.1246	1.1602	0.0827	3.2897
300	-0.0229	1.2119	0.4906	4.8465
400	0.0946	1.1553	-0.1329	4.8632

Table 1.16: First four moments for samples of size 100, 200, 300 and 400 of a mixture distribution of N(0,2) and N(0,1) with a mixing parameter $p = 0.1$.

the *probability-unbiased* VaR approaches the *plug-in* VaR.

	$\alpha = 1\%$			$\alpha = 5\%$		
n	α_{pu}	VaR_{pu}	$VaR_{plug-in}$	α_{pu}	VaR_{pu}	$VaR_{plug-in}$
100	0.6431	-2.6599	-2.4817	4.5353	-2.0092	-1.9267
200	0.7106	-2.6183	-2.5894	4.6221	1.8305	1.8256
300	0.8947	-2.7030	-2.6819	4.9033	-1.9933	-1.9875
400	0.8774	-2.5118	-2.5005	4.9113	-1.7869	-1.7762

Table 1.17: Probabilities $\alpha_{pu}(\%)$ needed to obtain probability-unbiased $\widehat{VaR}_{\alpha}$ for samples of size 100, 200, 300 y 400 of a mixture distribution of a Normal(0,2) and of a Normal(0,1) with mixing parameter $p = 0.1$ and respective *probability-unbiased* VaR and *plug-in* VaR for $\alpha = 1\%$ and $\alpha = 5\%$.

Figures 1.15 and 1.16 show density and distribution functions, respectively, for the mixture and for the two mixed Normal distributions with a mixing parameter $p = 0.1$ and sample size 100. Also shown on the horizontal axis are the 5% *plug-in* VaR and the *unbiased in probability* VaR.

The second mixture considers two Normal distributions with equal mean and different standard deviations, Normal(0,10) and Normal(0,1), with a mixing parameter $p = 0.1$, and random samples of size 100, 200, 300 and 400. This mixture distribution is more similar to a Student's t distribution than the previous one, because of the different standard deviations of the two mixing Normal distributions. So, this case is more relevant than the previous one. Table 1.18 shows the first four sample moments of the mixture for different sample sizes. The mixture distribution can be interpreted as the distribution obtained by substituting 10% of the data from the second Normal distribution with data from the first distribution. The mixture has higher standard deviation and higher kurtosis than

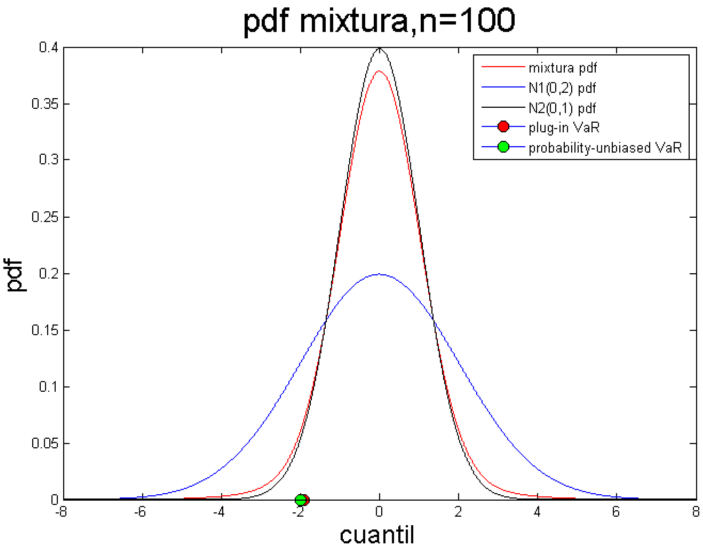


Figure 1.15: The mixture density function (red line), the density function of N(0,2) (blue line) and the density function of N(0,1) (black line). Points on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

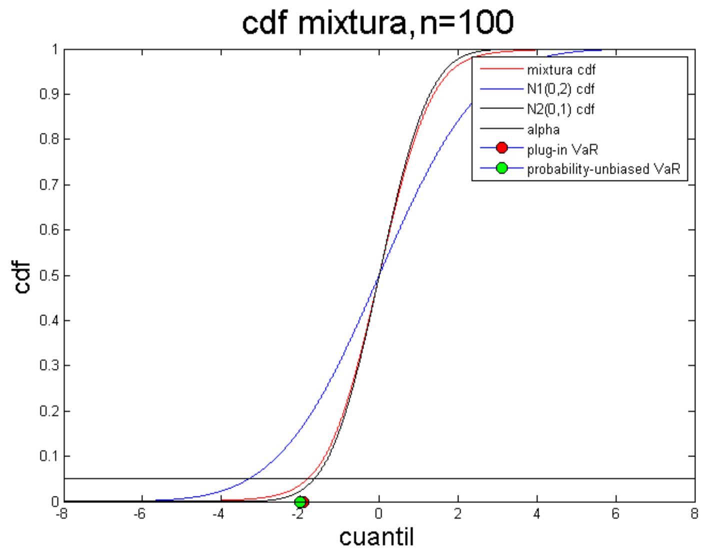


Figure 1.16: The mixture distribution function (red line), the distribution function of N(0,2) (blue line) and the distribution function of N(0,1) (black line). Points on the horizontal axis indicate the points *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

the N(0,1) distribution, both to a greater degree than in the previous case, because of the ratio of the standard deviations of the two Normal distributions. Furthermore, the level of kurtosis obtained in the mixture for different sample sizes suggests a distribution closer to an asymmetric Student-t than to a Normal.

MIXTURE of N(0,10) and N(0,1)				
n	μ_{mix}	σ_{mix}	$skewness_{mix}$	$kurtosis_{mix}$
100	-0.1630	1.5950	1.8959	10.7701
200	0.1145	3.0345	-1.6851	28.7797
300	0.2841	3.2635	1.2468	18.2806
400	-0.0734	3.7079	-0.3584	22.2558

Table 1.18: First four moments for samples of size 100, 200, 300 and 400 of a mixture distribution of N(0,10) and N(0,1) with mixing parameter $p = 0.1$.

Table 1.19 displays the *probability-unbiased* α_{pu} obtained by bootstrapping. It is less than α for all sample sizes, indicating that *plug-in* VaR again underestimates risk. As the sample size increases, α_{pu} tends to α. We also present the comparison between the *probability-unbiased* and the *plug-in* VaR estimates.

	$\alpha = 1\%$			$\alpha = 5\%$		
n	α_{pu}	VaR_{pu}	$VaR_{plug-in}$	α_{pu}	VaR_{pu}	$VaR_{plug-in}$
100	0.5000	-8.2684	-7.3471	4.7367	-2.3926	-2.3618
200	0.8304	-10.5555	-10.0159	4.8514	2.0401	-1.9588
300	0.8330	-13.2550	-12.8525	4.9213	-1.9506	-1.9471
400	0.8744	-14.5035	-14.3003	4.9522	-2.3772	-2.3749

Table 1.19: Probabilities $\alpha_{pu}(\%)$ needed to obtain the probability-unbiased $\widehat{VaR}_\alpha$ for samples of size 100, 200, 300 y 400 of a mixture distribution of a Normal(0,10) and of a Normal(0,1) with mixing parameter $p = 0.1$. The table also presents the associated *probability-unbiased* VaR and *plug-in* VaR estimates for $\alpha = 1\%$ and $\alpha = 5\%$.

Figure 1.17 shows the density functions for the mixture, the N(0,10) and N(0,1) distributions, while Figure 1.18 presents the respective cumulative distribution functions. The *plug-in* $VaR_{5\%}$ and *probability-unbiased* $VaR_{5\%}$ are shown in both Figures.

The third mixture is made up by Normal distributions with different mean and sim-

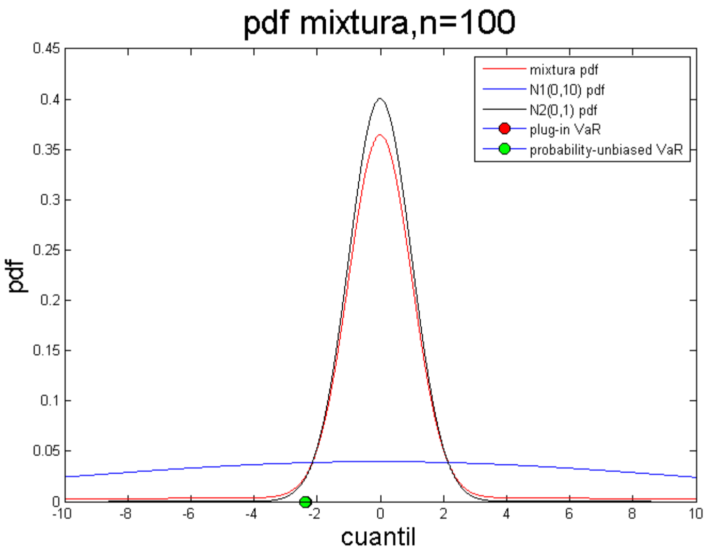


Figure 1.17: Mixture density function (red line), density function of N(0,10) (blue line) and density function of N(0,1) (black line). poionts on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

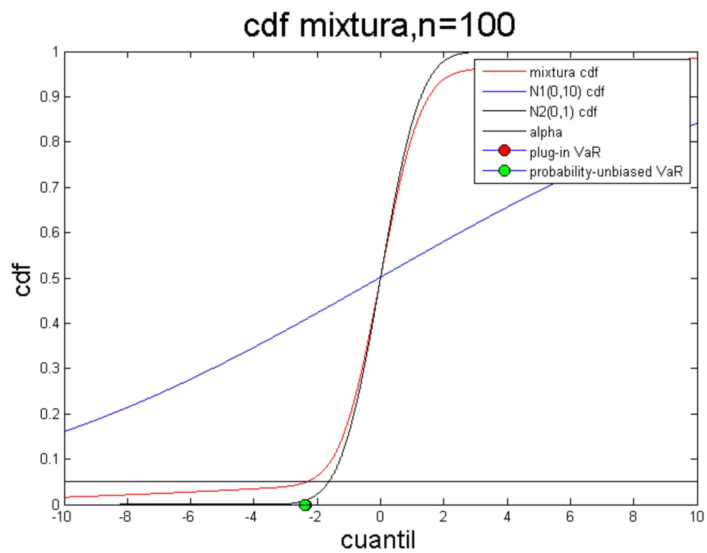


Figure 1.18: Mixture distribution function (red line), N(0,10) distribution function (blue line) and N(0,1) distribution function (black line). Points on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

ilar standard deviation, N(-5,2) and N(0,1) and a mixing parameter $p = 0.1$ for random samples of size 100, 200, 300 and 400. The left tail (tail relevant for the VaR) is now more distorted than in the previous cases.

In practice, it is important to detect that a time series from a mixture distribution rather than from a more standard distribution, in order not to distort the estimates. When mixing two Normal distributions, such detection will be easier if both distributions are very different and p is small. Then, the values generated by the first distribution Normal will be atypical compared to the second one, which might produce some data with entirely different conditions from the rest. That might produces some asymmetry and rather high kurtosis, but it will generally not be too simple to distinguish that from other asymmetric probability distributions with thick tails.

The mixture distribution has a mean which is shifted to the left: $m_{mix} = 0.1(-5) + (1-0.1)(0) = -0.5$. Sample moments are shown in Table 1.20, with a sample mean around the theoretical value. The standard deviation of the mixture is greater than that exhibited by the N(0,1), because the distortion from the N(-5,2) distribution. The mixture distribution also presents negative skewness due to the mixing of two Normal distributions with a different mean. The kurtosis of the mixture is well above 3 due to the different mean and standard deviation of the mixing distributions. This mixture presents a left tail thicker than the right tail because the mean of the distorting Normal distribution falls to the left of the mean of the second Normal distribution and we use a small mixing parameter p.

MIXTURE of N(-5,2) and N(0,1)				
n	μ_{mix}	σ_{mix}	$skewness_{mix}$	$kurtosis_{mix}$
100	-0.4173	1.9107	-2.0661	7.9924
200	-0.2348	1.5175	-1.1762	6.2227
300	-0.5152	1.8328	-2.0430	8.1173
400	-0.5945	2.0769	-1.7753	6.4910

Table 1.20: First four moments for samples of size 100, 200, 300 and 400 of a mixture distribution of N(-5,2) and N(0,1) with a mixing parameter $p = 0.1$.

The *probability-unbiased* VaR estimator is obtained as in the previous cases. Table 1.21 shows the α_{pu}, the VaR_{pu} and the $VaR_{plug-in}$ for $\alpha = 1\%$ and $\alpha = 5\%$, where we could notice similar observations to those in the previous mixtures. Figures 1.19 and 1.20 display again the density functions and cumulative distribution functions, respectively, for this case.

Finally, we consider a mixture of Normal distributions with different mean and different standard deviation: N(-5,10) and N(0,1) with mixing parameter $p = 0.1$. With so different Normal distributions and a small p, the resulting mixture can capture potential atypical

	$\alpha = 1\%$			$\alpha = 5\%$		
n	α_{pu}	VaR_{pu}	$VaR_{plug-in}$	α_{pu}	VaR_{pu}	$VaR_{plug-in}$
100	0.7931	-8.0283	-7.7839	4.4389	-5.1368	-4.8630
200	0.8565	-5.4241	-5.3490	4.9500	-2.9854	-2.9849
300	0.8269	-7.3096	-7.0326	4.8311	-5.2021	-5.1947
400	0.9459	-7.9207	-7.8929	4.9980	-5.8046	-5.8040

Table 1.21: Probabilities $\alpha_{pu}(\%)$ needed to obtain probabily-unbiased $\widehat{VaR}_\alpha$ for samples of size 100, 200, 300 y 400 of a mixture distribution of a Normal(-5,2) and of a Normal(0,1) with mixing parameter $p = 0.1$. We also show the associated *probability-unbiased* VaR and *plug-in* VaR for $\alpha = 1\%$ and $\alpha = 5\%$.

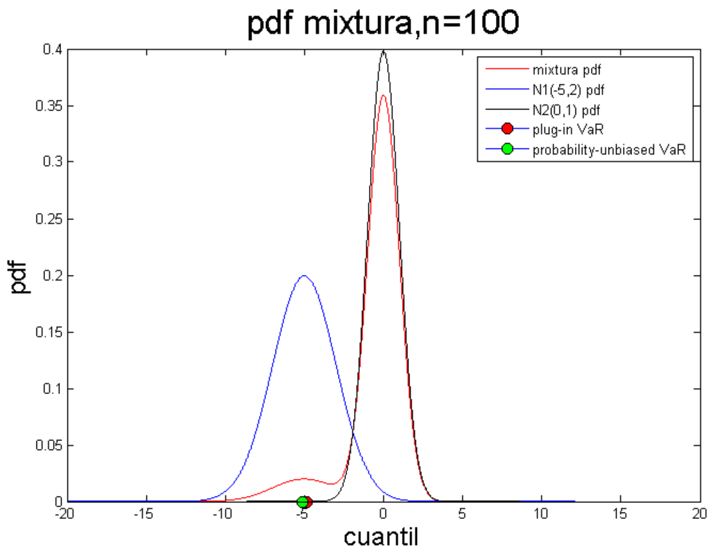


Figure 1.19: Mixture density function (red line), N(-5,2) the density function (blue line) and N(0,1) density function (black line). Points on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

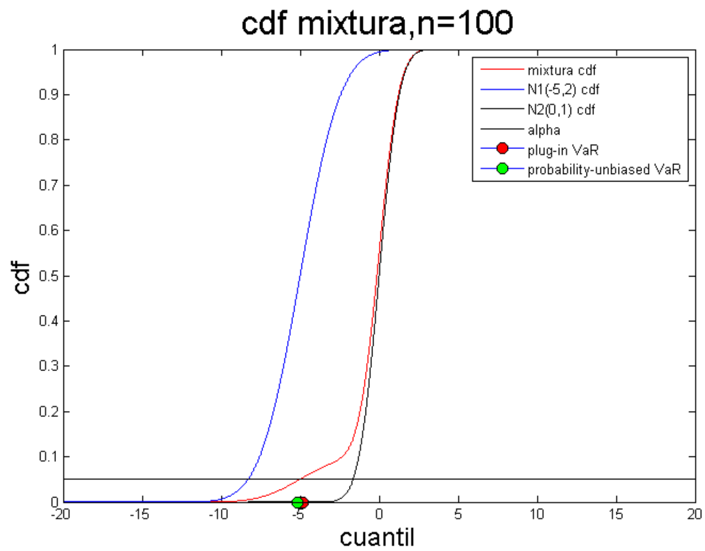


Figure 1.20: Mixture distribution function (red line), N(-5,2) distribution function (blue line) and N(0,1) distribution function (black line). Points on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

data much better than a Normal distribution, which may provide a better fit to some of the statistical characteristics observed in asset returns. Table 1.22 shows moments for sample sizes 100, 200, 300 and 400. As a result of this mixing, we obtain a distribution having a smaller mean, greater deviation and largest kurtosis than the second Normal distribution. This difference between the mixture and the second Normal distribution is larger than in previous mixtures.

MIXTURE of N(-5,10) and N(0,1)				
n	μ_{mix}	σ_{mix}	$skewness_{mix}$	$kurtosis_{mix}$
100	-0.5274	3.8400	-3.0260	18.8137
200	-0.5415	3.4229	-4.9035	34.4905
300	-0.2653	2.7198	-2.1651	29.9570
400	-0.2644	3.4468	-2.6797	27.9013

Table 1.22: First four moments for samples of size 100, 200, 300 and 400 of a mixture distribution of N(-5,10) and N(0,1) with a mixing parameter $p = 0.1$.

Table 1.23 shows the α_{pu} calculated from the bootstrap method to estimate the *probability-unbiased* VaR. We again see that the 1% and 5% *plug-in* VaR underestimate risk. Figures 1.21 and 1.22 show the density and cumulative distribution functions, respectively.

	$\alpha = 1\%$			$\alpha = 5\%$		
n	α_{pu}	VaR_{pu}	$VaR_{plug-in}$	α_{pu}	VaR_{pu}	$VaR_{plug-in}$
100	0.5806	-23.2413	-19.0515	4.6457	-6.6337	-5.5756
200	0.7963	-18.4699	-17.7663	4.6833	-2.2543	-1.8864
300	0.8333	-12.6246	-12.4064	4.8733	-2.1380	-2.1126
400	0.8677	-20.5952	-18.8131	4.8821	-2.7855	-2.6125

Table 1.23: Probabilities $\alpha_{pu}(\%)$ needed to obtain probability-unbiased $\widehat{VaR}_\alpha$ for samples of size 100, 200, 300 y 400 of a mixture distribution of a Normal(-5,10) and of a Normal(0,1) with mixing parameter $p = 0.1$. We also show the associated *probability-unbiased* VaR and *plug-in* VaR for $\alpha = 1\%$ and $\alpha = 5\%$.

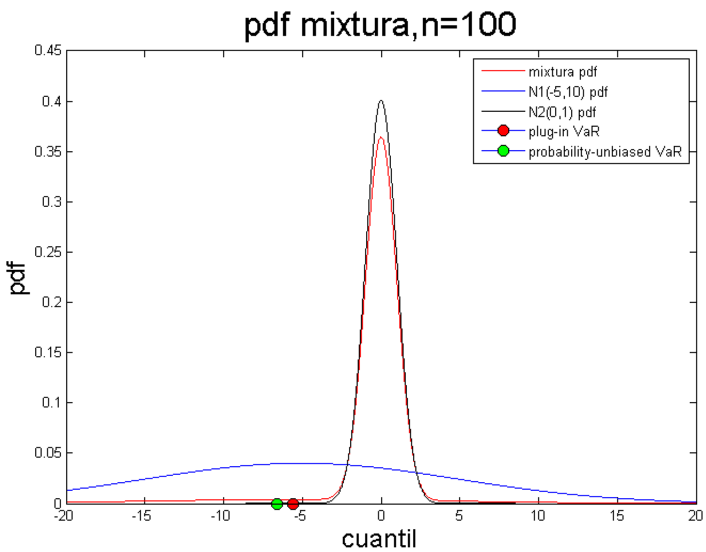


Figure 1.21: Mixture density function (red line), N(-5,10) density function (blue line) and N(0,1) density function (black line). Points on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

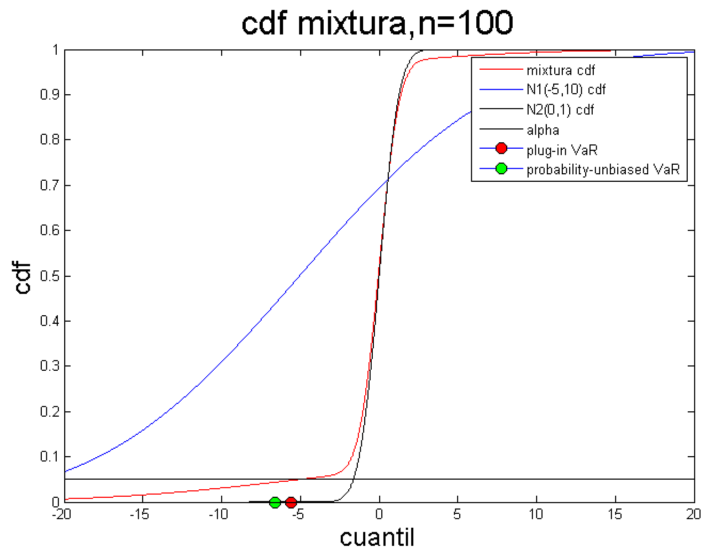


Figure 1.22: Mixture distribution function (red line), N(-5,10) distribution function (blue line) and N(0,1) distribution function (black line). Points on the horizontal axis indicate *plug-in* $\widehat{VaR}_{5\%}$ (red dot) and *probability-unbiased* $\widehat{VaR}_{5\%}$ (green dot).

1.7.3 A comparison of probability-unbiased VaR and plug-in VaR under Mixtures of Normal distributions

The theoretical results obtained for the *plug-in* VaR and *probability-unbiased* VaR estimators are now tested through simulations, and we calculate the probability of exceeding the VaR estimate.

We perform S = 10000 simulations for each of the four mixtures. The results shown in Table 1.24 for different sample sizes and $\widehat{VaR}_{1\%}$ and $\widehat{VaR}_{5\%}$ significance indicate that the probability of an excess from the *probability-unbiased* VaR estimator for different sample sizes is very close to the theoretical probability of 1% and 5%, respectively. The exceedence probability of the *plug-in* VaR estimator departs from the theoretical probability more than the *probability-unbiased* VaR estimator.

1.7.4 Approximate probability-unbiased confidence intervals for VaR under Mixtures of Normal distributions

We compute an estimation interval with a confidence level of $1 - 2\alpha_{CI}$, where $\alpha_{CI} = 0.05$, using bootstrapping. Table 1.25 shows the *plug-in* range for a confidence level of 90%. As the sample size increases, the confidence intervals narrow.

To calculate the upper bound of the *probability-unbiased* confidence interval we must obtain the $\alpha_{CI,pu}$ that minimizes the following objective function:

$$\alpha_{CI,pu} = argmin_\gamma \left| \frac{1}{B_2} \sum_{i=1}^{B_2} 1_{(\widehat{G}_i^*)^{-1}(\gamma) < g(\widehat{\theta})} - \alpha_{CI} \right|$$

	Mixture of N(0,2) and N(0,1)				Mixture of N(0,10) and N(0,1)			
n	$1\%plug-in$	$1\%pu$	$5\%plug-in$	$5\%pu$	$1\%plug-in$	$1\%pu$	$5\%plug-in$	$5\%pu$
100	1.20	0.93	4.92	4.56	1.47	1.10	5.47	5.27
200	1.25	0.91	5.12	4.75	1.18	1.02	5.18	5.05
300	1.16	1.09	4.96	4.89	1.03	0.91	5.32	5.24
400	1.09	1.01	5.37	5.3	1.12	0.99	5.44	5.38
	Mixture of N(-5,2) and N(0,1)				Mixture of N(-5,10) and N(0,1)			
n	$1\%plug-in$	$1\%pu$	$5\%plug-in$	$5\%pu$	$1\%plug-in$	$1\%pu$	$5\%plug-in$	$5\%pu$
100	1.24	1.07	5.29	4.74	1.53	1.20	5.41	5.13
200	1.23	1.12	5.05	4.99	1.29	0.94	5.47	5.08
300	1.19	1.01	5.53	5.38	1.22	1.05	5.11	4.98
400	1.14	1.08	5.11	5.11	1.05	0.98	4.97	4.97

Table 1.24: Shortfall probabilities $\alpha\%$ that the next observation is lower than the *plug-in* $\widehat{VaR}$ or lower that the probability *probability-unbiased* $\widehat{VaR}$ in the Monte-Carlo simulation, for different mixture distributions.

	Mixture of N(0,2) and N(0,1)				Mixture of N(0,10) and N(0,1)			
n	α_{CI}	$1-\alpha_{CI}$	CI_{low}	CI_{up}	α_{CI}	$1-\alpha_{CI}$	CI_{low}	CI_{up}
100	5	95	-2.4752	-1.4887	5	95	-5.8543	-3.9735
200	5	95	-2.3401	-1.2894	5	95	-5.1978	-2.2713
300	5	95	-2.0025	-1.6501	5	95	-6.1352	-4.6217
400	5	95	-1.9453	-1.5704	5	95	-6.5160	-4.3535
	Mixture of N(-5,2) and N(0,1)				Mixture of N(-5,10) and N(0,1)			
n	α_{CI}	$1-\alpha_{CI}$	CI_{low}	CI_{up}	α_{CI}	$1-\alpha_{CI}$	CI_{low}	CI_{up}
100	5	95	-8.2934	-4.8051	5	95	-17.6472	-2.7137
200	5	95	-6.9048	-4.7741	5	95	-17.3019	-4.2875
300	5	95	-6.5033	-4.8190	5	95	-8.3347	-1.9905
400	5	95	-6.1024	-4.4760	5	95	-10.9462	-3.0907

Table 1.25: Probability $\alpha_{CI}(\%)$, $1-\alpha_{CI}(\%)$ and quantiles corresponding to the lower and upper bounds of confidence interval 90% for different mixture distributions.

To obtain the lower bound of the confidence interval the bootstrap algorithm needs to be performed again in its entirety replacing α_{CI} by $1 - \alpha_{CI}$. Table 1.26 shows the 90% *probability-unbiased* confidence interval. The range is shifted to the left and is not symmetrical.

	Mixture of N(0,2) and N(0,1)				Mixture of N(0,10) and N(0,1)			
n	$\alpha_{CI,low,pu}$	$\alpha_{CI,up,pu}$	$CI_{low,pu}$	$CI_{up,pu}$	$\alpha_{CI,low,pu}$	$\alpha_{CI,up,pu}$	$CI_{low,pu}$	$CI_{up,pu}$
100	1.03	15.15	-2.5607	-2.3087	1.62	87.14	-6.1361	-4.4374
200	1.02	72.37	-2.4649	-1.5942	1.43	90.54	-5.2439	-2.4456
300	1.07	34.57	-2.1279	-1.8756	4.14	98.11	-6.1778	-4.5701
400	1.05	63.46	-1.9509	-1.6326	1.16	19.16	-6.8459	-6.1403
	Mixture of N(-5,2) and N(0,1)				Mixture of N(-5,10) and N(0,1)			
n	$\alpha_{CI,low,pu}$	$\alpha_{CI,up,pu}$	$CI_{low,pu}$	$CI_{up,pu}$	$\alpha_{CI,low,pu}$	$\alpha_{CI,up,pu}$	$CI_{low,pu}$	$CI_{up,pu}$
100	1.15	52.56	-8.8757	-6.1944	1.43	65.91	-17.7105	-9.3074
200	4.44	97.64	-6.9072	-4.1446	2.87	95.61	-18.4403	-4.1731
300	1.04	64.07	-6.5355	-5.5148	1.32	89.35	-9.2881	-2.4894
400	2.66	87.45	-6.1756	-4.6320	9.36	98.67	-9.4781	-2.0642

Table 1.26: Probabilities $\alpha_{CI,pu}(\%)$ such that $\alpha_{CI} = P(\widehat{CI} < VaR)$ with $\alpha_{\widehat{VaR}} = \alpha = 5\%$ in the $\widehat{VaR}$ and quantiles corresponding to the lower and upper bounds of confidence interval 90% for the different cases of the mixture distribution.

1.8 Empirical application and comparison with other VaR models

We follow McNeil et al. (2005, chapter 2.3.6) to test different VaR estimation methods using 1000 data observations from a portfolio that invests 30% in the FTSE 100 index, 40% in the S&P 500 index and 30% in SMI index between 1992 and 2003. The methods considered are,

- VC: The standard unconditional variance-covariance method assuming multivariate Gaussian risk-factor changes.
- HS: The standard unconditional historical simulation method.
- VC-t: The standard unconditional variance-covariance method assuming multivariate Student-t risk-factor changes.
- HS-GARCH: A conditional version of the historical simulation method in which GARCH(1,1) models with a constant conditional mean term and Gaussian innovations are fitted to the historically simulated losses to estimate the volatility of the next day's loss.

- VC-MGARCH: A conditional version of the variance-covariance method in which a multivariate GARCH model (a first-order constant conditional correlation model) with multivariate Normal innovations is used to estimate the conditional covariance matrix of the next day's risk-factor changes.
- HS-EWMA: A conditional method, similar to HS-GARCH, in which the EWMA method is used to estimate the conditional covariance matrix of the next day's risk-factor changes.
- VC-EWMA: A similar method to VC-MGARCH but a multivariate version of the EWMA method is used to estimate the conditional covariance matrix of the next day's risk-factor changes.
- HS-GARCH-t: A similar method to HS-GARCH but Student-t innovations are assumed in the GARCH model.
- VC-MGARCH-t: A similar method to VC-MGARCH but multivariate Student-t innovations are used in the MGARCH model.
- HS-CONDEVT: A conditional method using a combination of GARCH modeling and EVT (extreme value theory).

The return series show little serial correlation, to the point that it is safe to treat returns as being i.i.d.. The characteristics of the distortion of α allows for efficiently estimating the VaR quantile from a short amount of data to capture the clusters in the data. This is relevant because extreme returns appear in clusters (McNeil et al., 2005) and if we use the i.i.d. model with long windows we will be likely to underestimate risk. VaR estimates would then change very slowly, being unable to capture changes that may occur in the market as soon as they happen.

Calculation of the probability-unbiased VaR estimator from short rolling windows under the i.i.d. approach has some advantages [(Francioni and Herzog, 2010): i) only a few data points are needed to obtain a very good VaR estimate and ii) this approach outperforms other alternatives that need many data points to calibrate the model, e.g. EWMA, GARCH, ... (at least 1000 data are necessary to calibrate the models, McNeil et al., 2005). Francioni and Herzog showed how to distort the significance level α so that the resulting VaR estimate is unbiased. They also showed that the standard plug-in VaR estimates of a Normal population is biased.

In this section, we use the α_{pu} values to calculate the probability-unbiased VaR estimator in a rolling window of size t. We start with the simpler case of the Normal distribution, a member of the location-scale family. The Student-t has the number of degrees of freedom as an additional parameter, that we estimate first from portfolio return data, to subsequently calculate α_{pu} values. A similar procedure is followed for mixtures of two Normals, starting with the GMM estimation of the five additional parameters, μ_1, μ_2, σ_1, σ_2 and p, using portfolio returns. We propose mixtures of Normals as a more realistic distributions to further improve the results obtained under the Normal distribution.

Table 1.27 shows the number of annual VaR exceedences for the i.i.d. rolling window 5% VaR estimator under Normal and Student-t distributions, for different sample sizes. Every year, that number is relatively close to its expected value of 13 (aprox. $5\% \cdot 260$ days) [4]. The results in Table 19 suggest that under a Student-t distribution, windows with $n = 20$, 25 and 50 data points are generally outperformed by the shorter $n = 15$ window. In particular, for 1993, 1994, 2000 and 2001 VaR is poorly estimated under the Student-t distribution, with too many violations of the 95% VaR estimates. In general, 2000, 2001 and 2002 were the most difficult years to use in prediction for most models, since returns became very volatile, with many extreme losses. In the case of the Normal distribution, $n = 15$ and $n = 25$ perform better than $n = 50$.

For the mixture of two Normal distributions, the window with $n = 100$ outperforms longer window sizes especially during 1996, 1997 and 1998. We again work with samples larger than in the case of Normal and Student-t distributions for reasons explained above.

Normal						**Year**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
n = 10	5 (8)	5 (8)	8 (5)	8 (5)	8 (5)	6 (7)	5 (8)	4 (9)	9 (4)	3 (10)	3 (10)	3 (10)
n= 15	7 (6)	7 (6)	8 (5)	11 (2)	14 (1)	9 (4)	11 (2)	7 (6)	12 (1)	7 (6)	8 (5)	6 (7)
n = 20	7 (6)	14 (1)	11 (2)	8 (5)	12 (1)	10 (3)	15 (2)	9 (4)	14 (1)	14 (1)	6 (7)	8 (5)
n = 25	7 (6)	18 (5)	12 (1)	10 (3)	13 (0)	10 (3)	13 (0)	12 (1)	13 (0)	13 (0)	9 (4)	6 (7)
n = 50	7 (6)	13 (0)	17 (4)	11 (2)	15 (2)	12 (1)	14 (1)	11 (2)	15 (2)	18 (5)	11 (2)	9 (4)
Student-t						**Year**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
n = 10	8 (5)	12 (1)	10 (3)	13 (0)	13 (0)	9 (4)	10 (3)	14 (1)	17 (4)	10 (3)	11 (2)	10 (3)
n = 15	8 (5)	15 (2)	12 (1)	12 (1)	14 (2)	10 (3)	15 (2)	12 (1)	15 (2)	16 (3)	13 (0)	10 (3)
n = 20	9 (4)	19 (6)	16 (3)	10 (3)	15 (2)	14 (1)	17 (4)	12 (1)	17 (4)	18 (5)	14 (1)	8 (5)
n = 25	10 (3)	19 (6)	15 (2)	11 (2)	16 (3)	13 (0)	18 (5)	13 (0)	17 (4)	18 (5)	12 (1)	9 (4)
n = 50	9 (4)	13 (0)	18 (5)	11 (2)	15 (2)	13 (0)	16 (3)	12 (1)	17 (4)	19 (6)	12 (1)	9 (4)
Mixture						**Year**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
n = 100	7 (6)	13 (0)	15 (2)	9 (4)	15 (2)	14 (1)	14 (1)	8 (5)	19 (6)	15 (2)	19 (6)	8 (5)
n = 200	1 (12)	12 (1)	15 (2)	5 (8)	17 (4)	16 (3)	19 (6)	6 (7)	14 (1)	15 (2)	16 (3)	2 (11)
n = 300	0 (13)	7 (6)	20 (7)	4 (9)	22 (9)	23 (10)	19 (6)	4 (9)	17 (4)	16 (3)	20 (7)	3 (10)
n = 400	0 (13)	4 (9)	19 (6)	6 (7)	21 (8)	26 (13)	20 (7)	5 (8)	13 (0)	19 (6)	22 (9)	4 (9)

Table 1.27: Number of 5% VaR exceedences per year for the i.i.d. mixture rolling window model with window length n. Absolute differences between the expected number of exceedences (13 per year) and the number of observed exceedences are reported in parentheses.

When comparing the performance of the different models for VaR estimation we look at estimates from 1996 through 2003, because that is the time period considered by McNeil et al. (2005) with whom we want to compare our results. Two different quantities are calculated: i) the overall exceedence probability, defined as the number of observed exceedences in the period divided by the number of data points, ii) the Observed Absolute Deviation per year (OAD), used by McNeil et al. (2005), which was introduced by Francioni and Herzog as the mean of the absolute difference between the expected num-

[4] Except for 1992, when we lost n data observations due to the rolling window

ber of exceedences (i.e. 3 for 1% VaR and 13 for 5% VaR) and the number of observed exceedences.

Table 1.28 clearly shows that the i.i.d. model with rolling window outperforms the other models with respect to the overall exceedence probability and OAD. For the 1% probability-unbiased VaR, the Normal model with window of length 25, the Student-t model with window of length 15 and the mixture with window of length 100 show the best overall probabilistic properties. At 5% significance, Table 1.29 shows that the i.i.d. Student-t model with rolling window of size 15 and the Normal model with rolling window of size 50 outperform the other models with respect to OAD and to the overall exceedence probability, respectively. The mixture with a window of length 100 is the best within models with mixture distribution and outperforms in VaR estimation many of the methods proposed by McNeil et al.

Model	**Exc. Prob. (%)**	**OAD**
N i.i.d. n = 15	0.16	2.50
N i.i.d. n = 25	0.87	1.33
N i.i.d. n = 50	1.43	1.42
ST i.i.d. n = 15	0.83	1.25
ST i.i.d. n = 25	1.32	1.33
ST i.i.d. n = 50	1.65	1.50
NM i.i.d. n = 100	1.02	1.33
NM i.i.d. n = 200	1.29	2.00
NM i.i.d. n = 300	1.38	2.42
NM i.i.d. n = 400	1.43	2.33
VC	3.03	5.55
HS	2.02	3.00
VC-t	2.35	3.87
HS-GARCH	2.26	2.87
VC-MGARCH	2.31	2.87
HS-EWMA	2.07	2.62
VC-EWMA	2.02	2.62
HS-GARCH-t	1.68	1.62
VC-MGARCH-t	1.44	3.12
HS-CONDEVT	1.35	1.25

Table 1.28: Historical 1% VaR exceedence probabilities of the various models and historical Observed Absolute Deviation (OAD) per year.

In general, the i.i.d. approach with short windows incorporates too little information to produce good VaR estimates, whereas with large windows VaR estimates are too static and they do not adapt to new information fast enough. However, the i.i.d. approach to compute the 1% and 5% probability-unbiased VaR outperform the other models proposed by McNeil et al., which are more complex and use the *plug-in* VaR estimator. Figures 1.23 and 1.24 show a plot of the probability-unbiased VaR estimates of the i.i.d. Student-t

Model	Exc. Prob. (%)	OAD
N i.i.d. n = 15	3.43	4.16
N i.i.d. n = 25	4.38	2.41
N i.i.d. n = 50	4.97	2.42
ST i.i.d. n = 15	4.87	1.92
ST i.i.d. n = 25	5.5	2.83
ST i.i.d. n = 50	5.32	2.50
NM i.i.d. n = 100	5.15	2.92
NM i.i.d. n = 200	4.71	4.17
NM i.i.d. n = 300	5.48	6.50
NM i.i.d. n = 400	5.82	6.25
VC	7.36	7.88
HS	7.65	8.12
VC-t	8.46	10.25
HS-GARCH	6.11	3.38
VC-MGARCH	6.64	4.50
HS-EWMA	6.2	3.62
VC-EWMA	5.92	3.38
HS-GARCH-t	6.34	3.75
VC-MGARCH-t	6.93	5.50
HS-CONDEVT	5.77	2.75

Table 1.29: Historical 5% VaR exceedence probabilities of the various models and historical Observed Absolute Deviation (OAD) per year.

model with a rolling window size of 20 and 200 data points, respectively. The rolling window with 20 data points clearly reacts extremely fast to new data, with occasional large jumps in the VaR estimate.

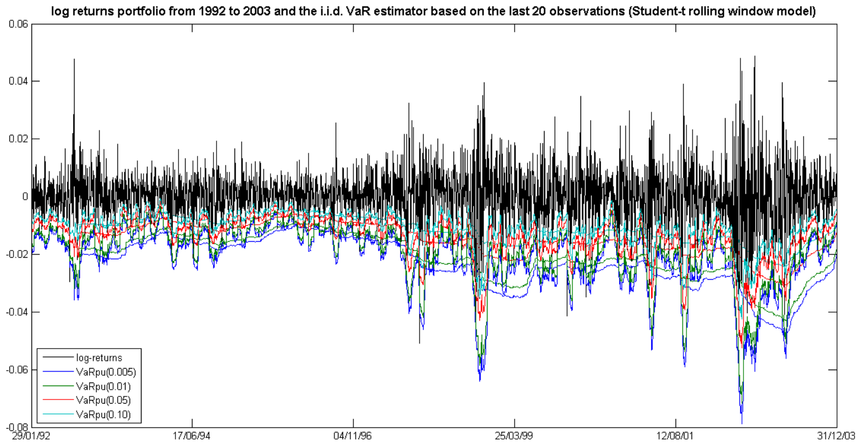


Figure 1.23: Portfolio log-returns from 1992 to 2003 and i.i.d. VaR estimates for different α based on a Student-t rolling window model with a window length of 20 observations.

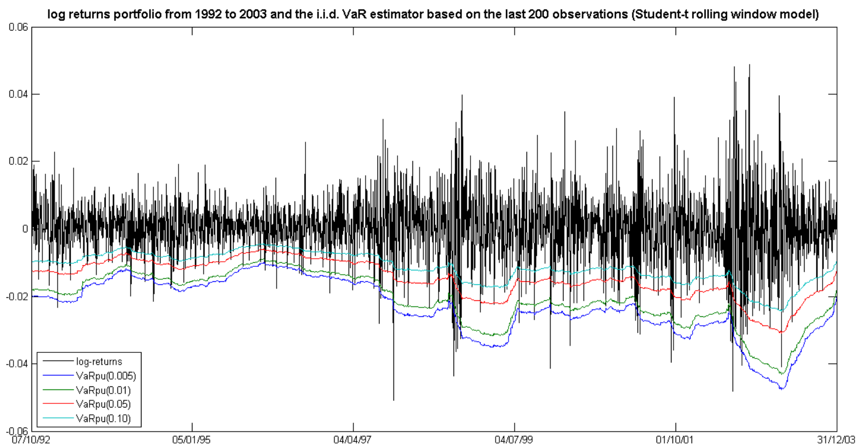


Figure 1.24: Portfolio log-returns from 1992 to 2003 and i.i.d. VaR estimates for different α based on a Student-t rolling window model with window length of 200 observations.

1.9 Conclusions

Francioni and Herzog (2010) (FH) proposed a standard resampling bootstrap algorithm to estimate a probability unbiased VaR in the case of Normal returns. The main idea is to replace the level α by a suitable chosen level α_{pu} which minimizes the averaged distance of the bootstrapped estimators to α. In other words, their strategy consisted on modifying the desired significance level α to obtain α_{pu} in such a way that we obtain an unbiased estimate of VaR at the original significance level. Their analysis suggested that VaR estimates based on short samples may have a good performance for Normal distributions, often beating standard VaR estimates based on long samples.

We explore the properties of the *probability-unbiased* VaR proposed by FH as an interesting alternative to *plug-in* VaR when working with short samples and a small significance level α. It is then when the *probability-unbiased* VaR differs more from *plug-in* VaR. We extend work by FH to Student-t distributions and mixtures of Normal distributions. Our results suggest that for a variety of distributions the *plug-in* VaR estimator underestimates risk for a given range of probabilities (α) when estimated from short samples. The smaller the sample size, the greater the underestimation of risk by the *plug-in* VaR estimator. The range of probabilities for which *plug-in* VaR underestimates risk depends on the sample size and on the assumed probability distribution for returns.

In the Gaussian case we can use the parametric approach to estimate VaR in closed form. For other cases we use an appropriate bootstrapping algorithm suggested by FH. We show that the performance of the *probability-unbiased* estimators for small sample sizes is surprisingly good also for Student-t distributions as well as for mixtures of Normals. The reason is that the shorter the period, the more uniform will be the sample. Besides, the conditional volatility will not change much over a short sample, making the sample almost i.i.d.. The difference between α_{pu} and α is larger for a Normal sample than for a Student-t distribution. For a Mixture of Normals, the difference depends on the mixing parameters.

We also estimate *probability-unbiased* confidence intervals for the VaR estimator. For the three distributions (Normal, Student-t and mixture of Normals) the *probability-unbiased* confidence interval is shifted to the left, relative to the standard confidence interval calculated using the *plug-in* VaR estimator. The leftward shift of the *probability-unbiased* confidence interval is due to the fact that most exceedences occur at the lower bound. Hence, a symmetric confidence interval would not be appropriate. The findings in this chapter suggest that the unbiased VaR estimator is a valuable tool for estimation of risk in practice.

Bibliography

[1] Acerbi, C. and Szekely, B., 2014. *Backtesting Expected Shortfall*. Publication of MSCI.

[2] Alexander, C., 2009. *Market Risk Analysis, Value at Risk Models*. John Wiley & Sons, Vol. 4.

[3] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., 1998. *Coherent measures of risk*, Math. Finance, Vol. 9, pp. 203-228.

[4] Basel Committee on Banking Supervision, 2016. Standards: Minimum Capital requirements for market risk. Bank for International Settlements.

[5] Berkowitz, J. and O'Brien, J., 2002. *How accurate are Value-at-Risk models at commercial banks?*. Journal of Finance, pp. 1093-1111.

[6] Bickel, P.J. and Krieger, A. M., 1989. *Confidence bands for a distribution function using the bootstrap*. Journal of the American Statistical Association, Vol. 84, pp. 95-100.

[7] Casella, G. and Berger, R., 2002. *Statistical Inference*. Duxbury Advanced Series.

[8] Chernick, M.R., 1999. *Bootstrap Methods: A Practitioner's Guide*. John Wiley & Sons.

[9] Costanzino, N. and Curran, M., 2015. *Backtesting General Spectral Risk Measures with application to Expected Shortfall*. Available at SSRN 2514403

[10] Davis, M., 2013. *Consistency of risk measure estimates*. Available at SSRN: *https://ssrn.com/abstract*=2342279 or *http://dx.doi.org/10.2139/ssrn.2342279*

[11] Du, Z. and Escanciano, J.C., 2015. *Backtesting Expected Shortfall: Accounting for Tail Risk*. Available at SSRN 2548544.

[12] Duffie, D and Pan, J., 1997. *An overview of value at risk*. Journal of Derivatives, Vol. 4, pp. 7-49.

[13] Efron, B., 1979. *Bootstrap methods: Another look at the jackknife*. Annals of Statistics, Vol. 7, pp. 1-26

[14] Efron, B. and Tibshirani, R.J., 1993. *An introduction to the Bootstrap*. Chapman & Hall: New York.

[15] Embrechts, P. and Hofert, M., 2014. *Statistics and quantitative risk management for banking and insurance.* Anual Review of Statistics and Its Applications, Vol. 1, pp. 493-514.

[16] Emmer, S., Kratz, M. and Tasche, D., 2015. *What is the best risk measure in practice? A comparison of standard measures.* Journal of Risk, Vol. 18, No. 2.

[17] Fissler, T., Ziegel, J.F. and Gneiting, T., 2015. *Expected Shortfall is jointly elicitable with value at risk-implications for backtesting.* arXiv preprint arXiv:1507.00244.

[18] Föllmer, H. and Knispel, T., 2013. *Convex risk measures: Basic facts, law-invariance and beyond, asymptotics for large portfolios.* Handbook of the fundamentals of Financial Decision Making, Part II, pp. 507-554.

[19] Föllmer, H. and Schied, A., 2002. *Stochastic Finance: An introduction in discrete time.* Gruyters Studies in Mathematics, Vol. 27.

[20] Francioni, I. and Herzog, F., 2010. *Probability-unbiased Value-at-Risk estimators*, Quantitative Finance, Vol. 12, pp. 755-768.

[21] Hull, J., 2015. *Risk Management and Financial Institutions.* Pearson, Prentice Hall.

[22] Jackel, P., 2003. *Monte Carlo methods in finance.* Wiley Finance.

[23] Jorion, P., 2007. *Value at Risk: The New Benchmark for Controlling Market Risk.* McGraw-Hill.

[24] McNeil, A., Frey, A. and Enbrechts, P., 2005. *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press.

[25] Miller, R., 1974. *The jackknife: a review.* Biometrika, Vol. 61, pp. 1-15.

[26] Pitera, M. and Schmidt, T., 2016. *Unbiased estimation of risk.* arXiv preprint arXiv:1603.02615.

[27] Pearson, N.D. and Smithson, C., 2002. *VaR The State of Play.* Journal of Financial Econometrics, Vol. 11, pp. 175-189.

[28] Quenouille, M. H., 1949.*Approximate tests of correlation in time series.* Journal of the Royal Statistical Society, Series B, Vol. 11, pp. 68-84.

[29] Simon, J.L., 1969. *Basic Research Methods in Social Science.* Random House.

[30] Tukey, J., 1958. *Bias and confidence in not quite so large samples (Abstract).* Annals of Mathematical Statistics, Vol. 29, pp. 614.

[31] Vose, D., 1996. *Quantitative risk analysis: a guide to MC simulation modelling.* John Wiley & Sons.

[32] Ziegel, J.F., 2016. *Coherence and elicitability.* Mathematical Finance, Vol. 26, No. 4, pp. 901-918.

Appendices

A VaR$_{pu}$ calculated with the bootstrap algorithm proposed by FH for a Normal distribution

In this appendix we study the case where explicit unbiased estimators for the VaR risk measure are not available. We use the bootstrap algorithm proposed by FH in [20] and consider the Normal distribution. We compare the results obtained here with those obtained previously in Section 1.5.

Table 30 shows the *probability-unbiased* $\widehat{VaR}_\alpha$, (VaR_{pu}), and the *plug-in* $\widehat{VaR}_\alpha$ (the standard VaR estimator) obtained for different samples sizes and α's. Notice that the *plug-in* $\widehat{VaR}_\alpha$ underestimates the risk, indicating fewer losses than they really arise with a probability α%.

As with the parametric method in the main text, calculating the probability-unbiased VaR is relevant especially for small sample sizes, when the difference between the alternative VaR estimates is larger, since for large samples the *probability-unbiased* $\widehat{VaR}_\alpha$ is similar to *plug-in* $\widehat{VaR}_\alpha$.

	VaR_{pu}				$VaR_{plug-in}$			
n	**0.5**	**1**	**5**	**10**	**0.5**	**1**	**5**	**10**
10	-3.233	-2.942	-1.864	-1.375	-2.409	-2.165	-1.496	-1.139
15	-3.111	-2.751	-1.701	-1.234	-2.309	-2.065	-1.400	-1.045
20	-3.229	-2.884	-2.005	-1.566	-2.839	-2.582	-1.880	-1.506
25	-3.205	-2.874	-2.033	-1.609	-2.825	-2.562	-1.844	-1.461
50	-2.947	-2.646	-1.846	-1.429	-2.783	-2.513	-1.775	-1.382
100	-2.553	-2.315	-1.673	-1.335	-2.488	-2.262	-1.645	-1.316
150	-2.632	-2.383	-1.711	-1.357	-2.581	-2.342	-1.690	-1.343
200	-2.393	-2.165	-1.548	-1.221	-2.365	-2.143	-1.537	-1.213

Table 30: *Probability-unbiased* $\widehat{VaR}_\alpha$ versus *plug-in* $\widehat{VaR}_\alpha$ in the case of Normal distribution (0,1) using bootstrapping proposed by FH.

Figure 25 shows the graphs of the distortion function for different sample sizes (red line). It corroborates that the correction decreases as we get more sample observations, and the distortion function converges to the identity (black line) as the sample size grows. Figure 26 shows the distortion in the quantiles of the standard Normal distribution function.

Table 31 presents the probability α% that the next observation will exceed $\widehat{VaR}$ plug-in

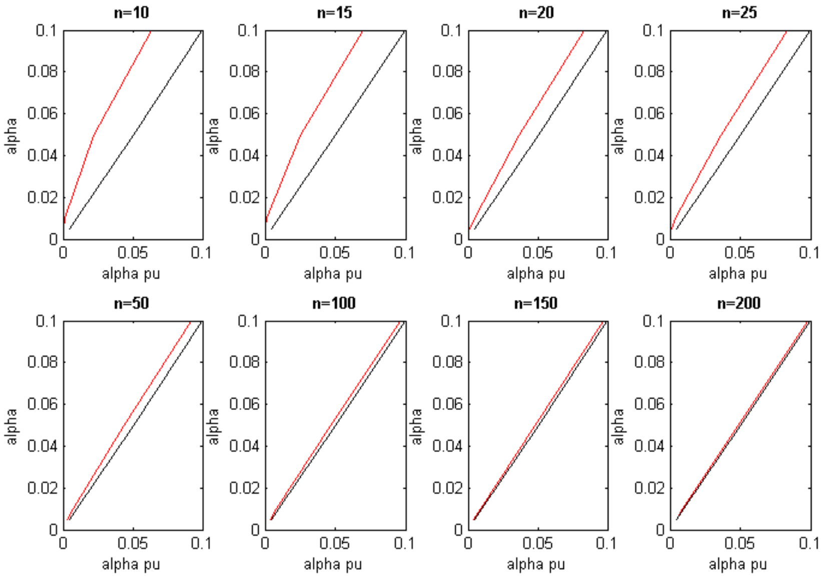


Figure 25: Distorsion function for the Normal distribution using bootstrapping proposed by FH. The diagonal (black line) represents no distorsion.

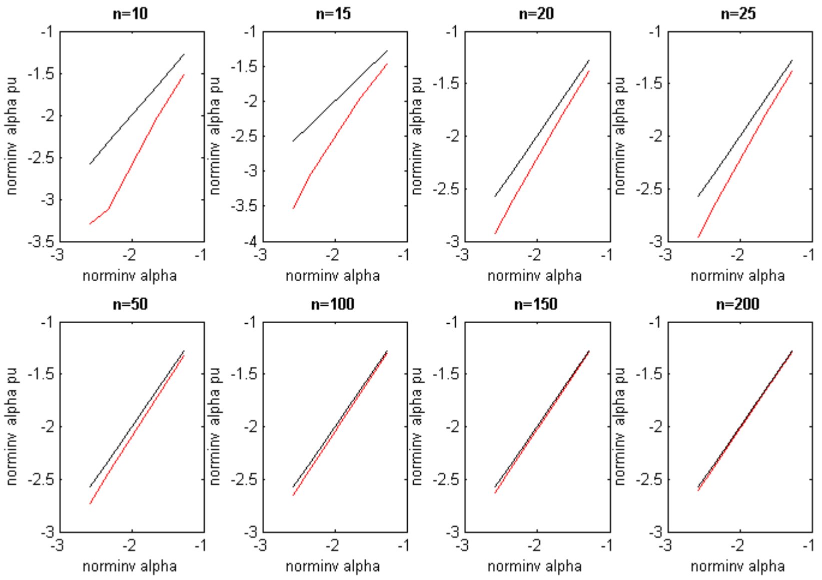


Figure 26: The quantiles of the Normal cdf versus the quantiles of the distorted Normal cdf using bootstrapping proposed by FH. The diagonal (black line) represents no distortion.

for a given α_{pu} and for different sample sizes. For example, if we estimate the population parameters with a sample of 25 observations, the probability that the next observation falls below $\widehat{VaR}_{5\%}$ *plug-in* is 6.47%.

	$\alpha_{pu}(\%)$			
n	**0.5**	**1**	**5**	**10**
10	2.225	3.210	8.466	13.684
15	1.958	2.853	7.864	13.016
20	1.121	1.854	6.510	11.659
25	1.075	1.801	6.478	11.650
50	0.734	1.340	5.668	10.759
100	0.608	1.159	5.320	10.366
150	0.578	1.115	5.227	10.260
200	0.545	1.067	5.137	10.160

Table 31: The shortfall probabilities $\alpha(\%)$ with which the next observation is lower than the plug-in VaR estimate $z_{\alpha_{pu}}$ sing bootstrapping proposed by FH to calculate α_{pu}.

Tables 30 and 31 will be different each time the bootstrap algorithm is performed, as this is a non-parametric method that depends on the values obtained on the sample and those obtained in the resamples. Instead, the parametric method used in subsection 1.5.1 yields unique Tables 1.3 and 1.2, providing a specific α_{pu} for each α and n. We obtain an unbiased estimator $\alpha_{pu}$of α and therefore an unbiased VaR.

Now, we use the second approach described in Section 1.3 to calculate the probability-unbiased VaR estimator by calculating the standard deviation $\widehat{\sigma}_{pu}$ of the distorted distribution function F. If F is a Normal distribution, the *probability-unbiased* VaR estimator can be written:

$$\widehat{VaR}_{\alpha} = \widehat{\mu}_{pu} + \widehat{\sigma}_{pu} z_{\alpha}$$

so that once we have the probability unbiased VaR, we can solve for $\sigma_{pu} z_{\alpha}$ We maintain the same mean: $\widehat{\mu}_{pu} = \widehat{\mu}$, calculating only the standard deviation $\widehat{\sigma}_{pu}$ of the distorted distribution function. This deviation will be different for every α and for each sample size (n), as in Table 32.

Figure 27 shows the true density function of a random variable N(0,1) (blue line), the density function of a random sample of size 15 from a random variable following a distribution N(0,1) (red line), and the density function of the distorted distribution function of the random sample of size 15 of the random variable that follows a N(0,1) (green line). Notice that the distorted distribution function has heavy tails that would allow for

	α(%)			
n	**0.5**	**1**	**5**	**10**
10	1.335	1.316	1.205	1.165
15	1.254	1.271	1.159	1.123
20	1.210	1.192	1.151	1.136
25	1.172	1.155	1.122	1.110
50	1.146	1.140	1.126	1.120
100	0.931	0.928	0.923	0.921
150	0.976	0.974	0.969	0.968
200	0.901	0.899	0.897	0.896

Table 32: Estimated standard deviations for the distorted distribution function using bootstrapping proposed by FH to calculate α_{pu}.

a better fit of the empirical evidence.

The *probability-unbiased* $\widehat{VaR}$ (green point) indicates higher losses than *plug-in* $\widehat{VaR}$ (red point). In other words, the it plug-in estimator underestimates risk particularly in small size samples, i.e. the smaller the sample, the greater the correction or adjustment needed for the distribution.

Figure 28 shows the cumulative distribution function of N(0,1) (blue line), the *plug-in* cumulative distribution function (red line) and the *probability-unbiased* cumulative distribution function (green line). We also show their 5% VaR estimates.

Figures 29 and 30 show pdf's and cdf's, respectively, for different sample sizes. They also show the respective point estimates of the 5% VaR. These figures show the convergence of the *plug-in* distribution and *probability-unbiased* distribution to the true distribution as the size of the random sample increases.

In order to test our theoretical calculations in a controlled simulation, we simulate the estimation of the *plug-in* VaR estimator and the *probability-unbiased* VaR estimator and calculate their exceedence probabilities.

The results, Table 33, for samples of size 10, 15, 20 and 25, for $\widehat{VaR}_{1\%}$ and for $\widehat{VaR}_{5\%}$, with $\mu = 0$, $\sigma = 1$ and $S = 100000$, show that the VaR exceedence of the *probability-unbiased* estimator is close to the theoretical values of 1% and 5%. The VaR exceedence of the *plug-in* VaR from the theoretical probability is clearly larger.

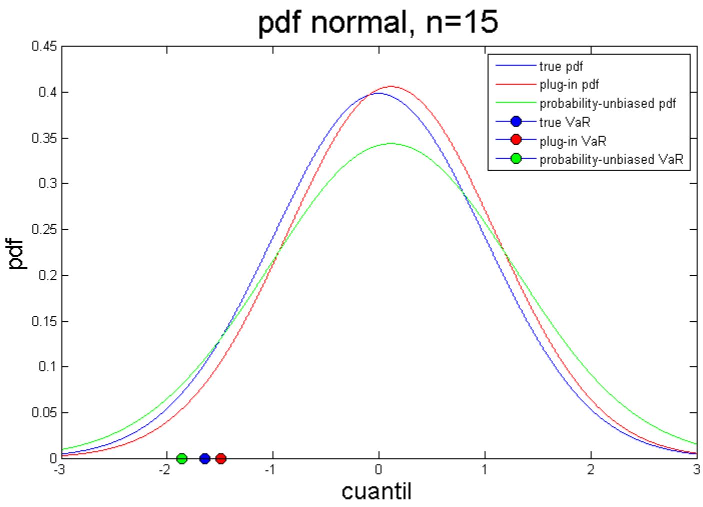


Figure 27: The true N(0,1) pdf (blue line), the *plug-in* pdf (red line) and the pdf of the *probabilty-unbiased* cdf (green line). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point) . Here we use bootstrapping proposed by FH to calculate α_{pu}.

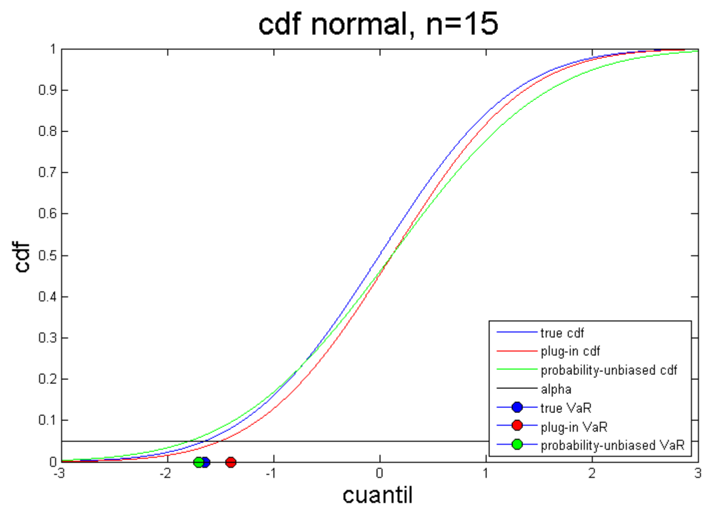


Figure 28: The true N(0,1) cdf (blue line), the *plug-in* cdf (red line) and the *probability-unbiased* cdf (green line). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point) . Here we use bootstrapping proposed by FH to calculate α_{pu}.

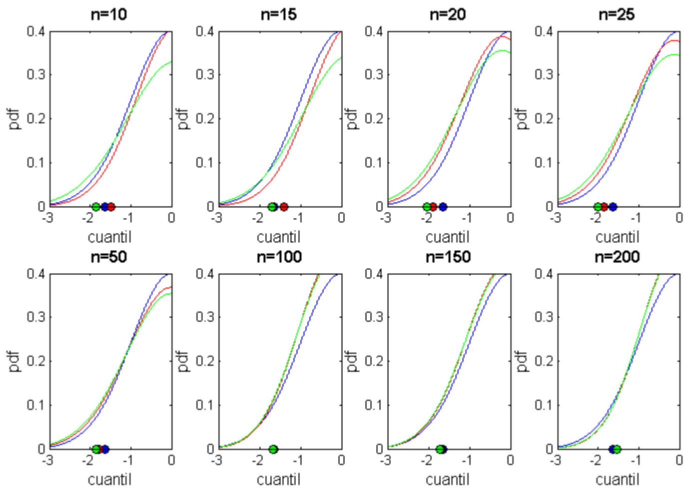


Figure 29: The true N(0,1) pdf (blue line), the *plug-in* pdf (red line) and the pdf of the *probability-unbiased* cdf (green line) for different sample sizes (enlargement of the left tail). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point). Here we use the bootstrapping algorithm proposed by FH to calculate α_{pu}

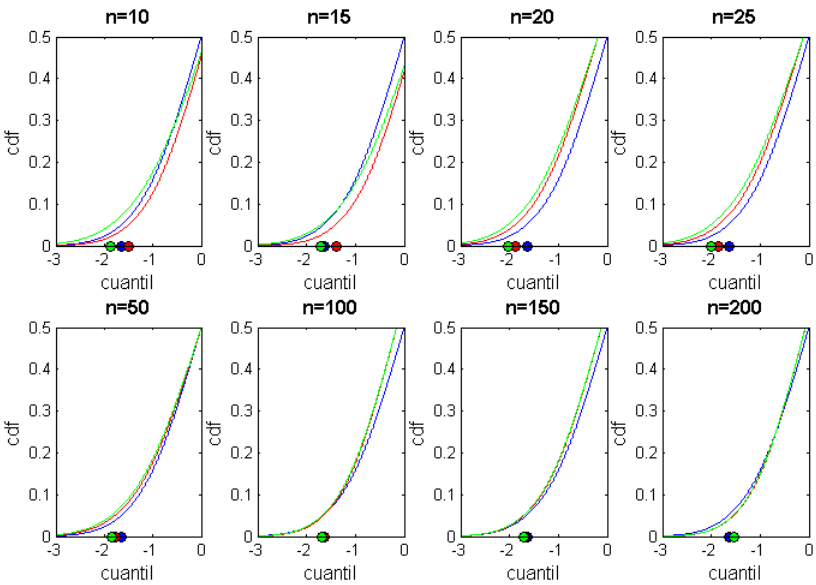


Figure 30: The true N(0,1) pdf (blue line), the *plug-in* pdf (red line) and the *probability-unbiased* cdf (green line) for different sample sizes (enlargement of the left tail). Points on the horizontal axis show the true $\widehat{VaR}_{5\%}$ (blue point), the *plug-in* $\widehat{VaR}_{5\%}$ (red point) and the *probability-unbiased* $\widehat{VaR}_{5\%}$ (green point) . Here we use the bootstrapping algorithm proposed by FH to calculate α_{pu}

As the sample size increases, the probability of an excess for the *plug-in* VaR estimator calculated from the simulation approaches the theoretical value. In the case of *probability-unbiased* VaR estimator that probability remains broadly similar to the theoretical probability for all sample sizes. However, the results are not as good as those obtained parametrically since in that case the value of α_{pu} was not just an approximation.

n	$1\%plug-in$	$1\%pu$	$5\%plug-in$	$5\%pu$
10	2.674	0.803	7.567	4.289
15	2.063	0.553	6.748	3.907
20	1.825	0.958	6.268	4.782
25	1.608	0.772	5.978	4.635

Table 33: Shortfall probabilities $\alpha\%$ that the next observation will be lower than the *plug-in* $\widehat{VaR}$ and the probability *probability-unbiased* $\widehat{VaR}$ in the Monte-Carlo simulation for Normal distribution using bootstrapping proposed by FH.

B Mixture of two distributions

The density function of a mixture of two distributions is

$$f(x) = pf_1(x) + (1-p)f_2(x)$$

donde f_1 and f_2 are density functions.

When we mix two Normal distributions, both density functions intersect at two points, for which the density function of the mixture also intersects at those two points, except in the following cases

- The two Normal distributions have equal standard deviation (σ) and different mean (μ). For all values of mixture parameter (p), the two Normal density functions intersect at a point.
- The two Normal distributions have equal mean (μ) and equal standard deviation (σ). For all values of mixture parameter (p), the two Normal density functions intersect at infinite points. The mixture distribution is a Normal.
- With mixture parameter 0 or 1, independently of the mean and standard deviation of the Normal distributions mixed, the density function of the mixture intersects at an infinite number of points with one of the two Normal, i.e. the mixture distribution is a Normal distribution.

The mixtures obtained in the last two points are symmetrical, being Normal. It is also symmetric density function of mixture in the following cases:

- When the means of the Normal distributions are equal.
- When the standard deviations of the Normal distributions are equal and mixing probability is $p = 0.5$.

Mixture of two Normal distributions with equal μ and similar σ

In this case, the mixture distribution obtained is similar to Normal. The average of the mixture will be the same as the mean of Normal distributions and the variance of the mixture will be the weighting of the respective variances of the mixed Normal distributions.

Figure 31 shows the distribution mixture obtained by mixing a Normal distribution with mean 0 and standard deviation 2, and a Normal distribution with mean 0 and standard deviation 1.

With $p = 0.1$ mixed distribution is similar to the second Normal distribution. If this mixing parameter is increased, the mixture will increasingly be alike to the first distribution, because it is weighted the first one more than the second one, arriving in $p = 0.9$ the

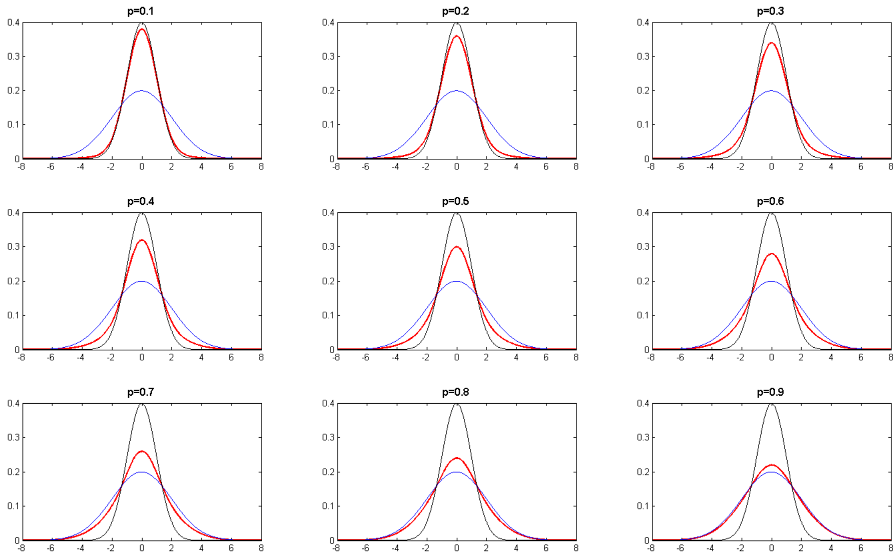


Figure 31: Density function of the mixture of N1(0,2) and N2(0,1) (red line), the density function N1(0,2) (blue line) and density function N2(0,1) (black line) for different values of the mixing parameter.

mixture being very similar to the first distribution.

Figure 32 shows the respective cumulative distribution functions. It has drawn a horizontal line (black line) to indicate the value $\alpha = 0.05$. The *plug-in* VaR to a 5% is the quantile where this line cuts the distribution function of the mixture (red line). As you increase the mixing parameter, VaR is shifted to the left, indicating a higher expected loss to a 95% confidence level. This is because as the mixing parameter is greater, the distortion produces by the first Normal distribution is larger than the second one, and if this first Normal distribution has greater variance than the second one, that introduces into this one more extreme values, increasing the tails and shifting the VaR to the left, i.e. VaR is more negative.

Mixture of two Normal distributions with equal μ and different σ

In this second case, we have a Normal distribution with mean 0 and standard deviation 10 and a Normal distribution with mean 0 and standard deviation 1. When distributions are mixed with very different deviations, the mixture distribution obtained is more similar to Student-t than a Normal, because there is higher probability in the tails due to the extreme values introduced by the first distribution Normal.

Figure 33 shows density functions of the mixture and of the two Normal distributions mixed.

Figure 34 represents the corresponding cumulative distribution functions. The mixture

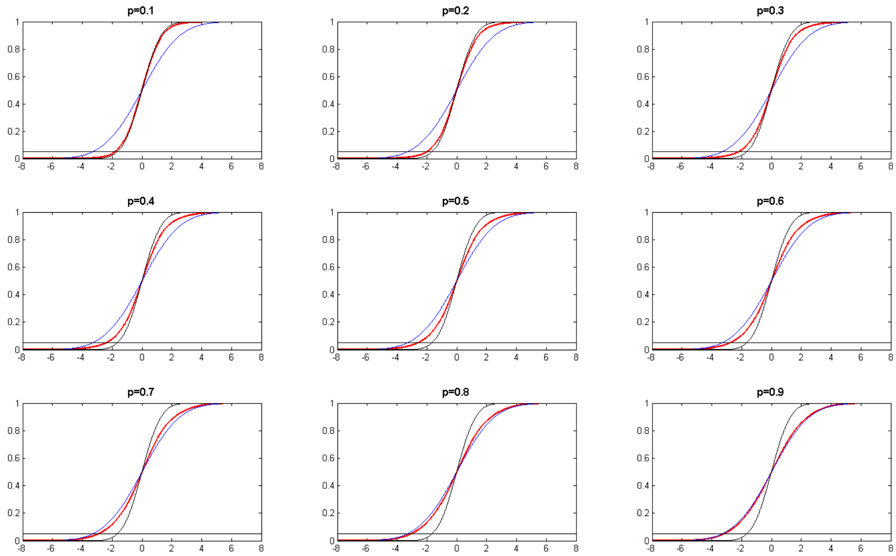


Figure 32: Distribution function of the mixture of N1(0,2) and N2(0,1) (red line), the distribution function N1(0,2) (blue line) and distribution function N2(0,1) (black line) for different values of the mixing parameter. The black horizontal line indicates the $\alpha = 0.05$.

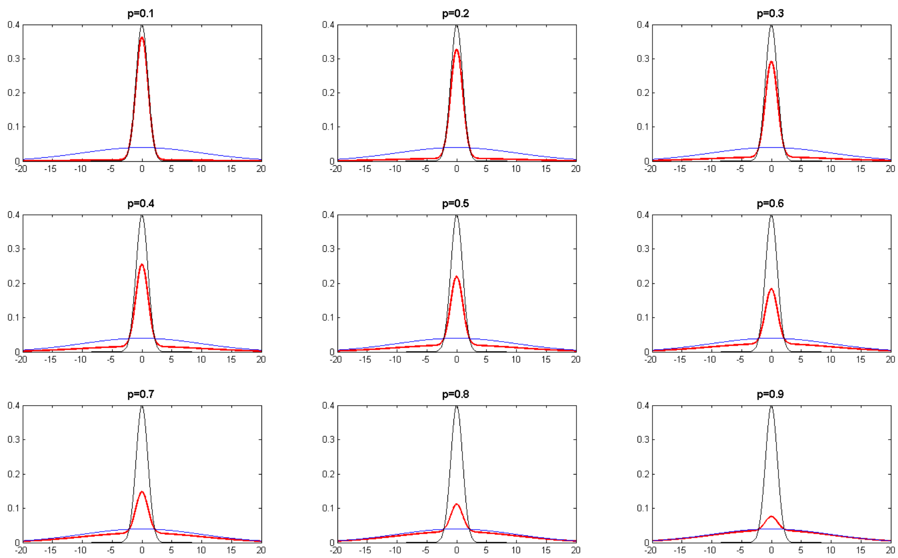


Figure 33: Density function of the mixture of N1(0,10) and N2(0,1) (red line), the density function N1(0,10) (blue line) and density function N2(0,1) (black line) for different values of the mixing parameter.

is mixing the two Normal, beginning to weigh very little the second distribution and very much the first one to reach that the second distribution weights much and the first one only a bit.

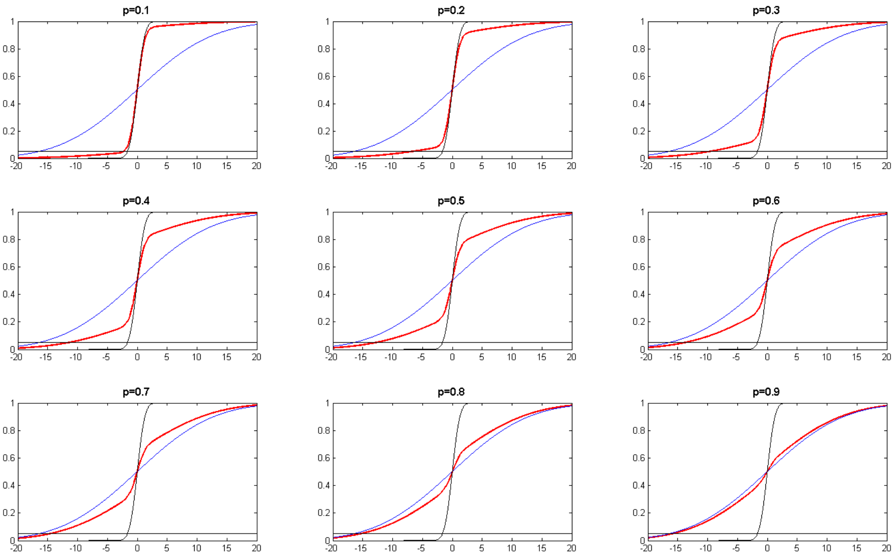


Figure 34: Distribution function of the mixture of N1(0.10) and N2(0.1) (red line), the distribution function N1(0.10) (blue line) and the distribution function N2(0, 1) (black line) for different values of the mixing parameter. The black horizontal line indicates the $\alpha = 0.05$.

Mixture of two Normal distributions with different μ and similar σ

Now, mixture distribution is created from a Normal distribution with mean -5 and standard deviation 2 and a Normal distribution with mean 0 and standard deviation 1.

It is observed that from Normal distributions with means very different, the left tail is heavier than a Normal when mixing parameter weights little the first Normal distribution and much the second one, and the right tail is heavier than a Normal when mixing parameter weights much the first distribution and little the second one.

In this case, unlike the previous two, we do not have Normal distributions with equal mean, the two tails are not thicken equally but thickens one more than the other depending on whether the mean of the first Normal is located to the left or to the right on the mean of the second Normal and the weighting given by the parameter mixing p. This mixture is suitable when you want to get a mixed distribution with skewness.

In this example, the left tail becomes heavier (when $p = 0.1$) than the right tail because the first Normal has its mean to the left of the mean of the second Normal.

Figure 35 shows that there is a p for which the distribution is bimodal mixture (between $p = 0.6$ and $p = 0.7$).

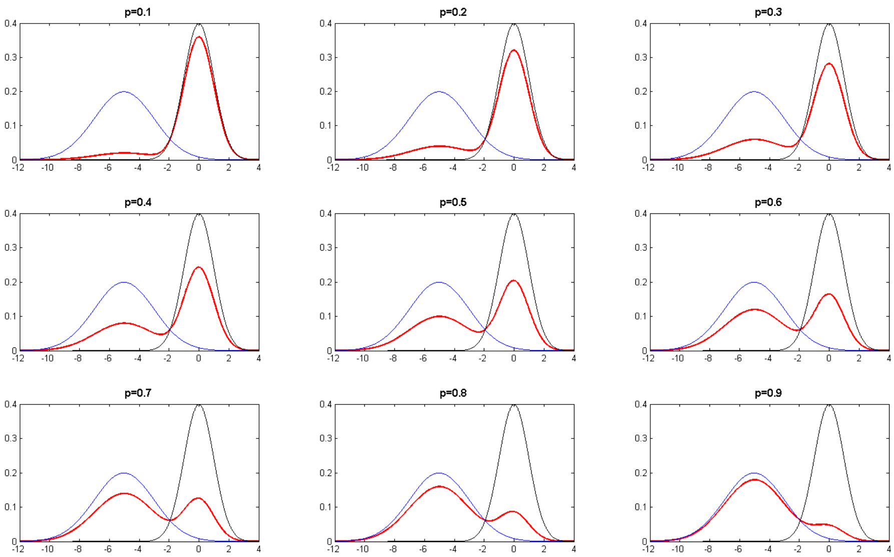


Figure 35: Density function of the mixture of N1(-5,2) and N2(0,1) (red line), the density function N1(-5,2) (blue line) and density function N2(0,1) (black line) for different values of the mixing parameter.

Figure 36 represents the cumulative distribution functions.

Unlike what happened in the first two cases, the VaR is shifted to the left when we increase the value of the mixing parameter because it is generated from Normal distributions with means very different. In the first subgraph, when $p = 0.1$, the mixed distribution has the left tail heavier than the second Normal distribution due to the small influence of the first Normal distribution. In contrast, in the last subplot when $p = 0.9$, we observe that the mixture distribution has the right tail heavier but less so than the left in the first subgraph, because the second Normal distribution has lower standard deviation than the first Normal distribution.

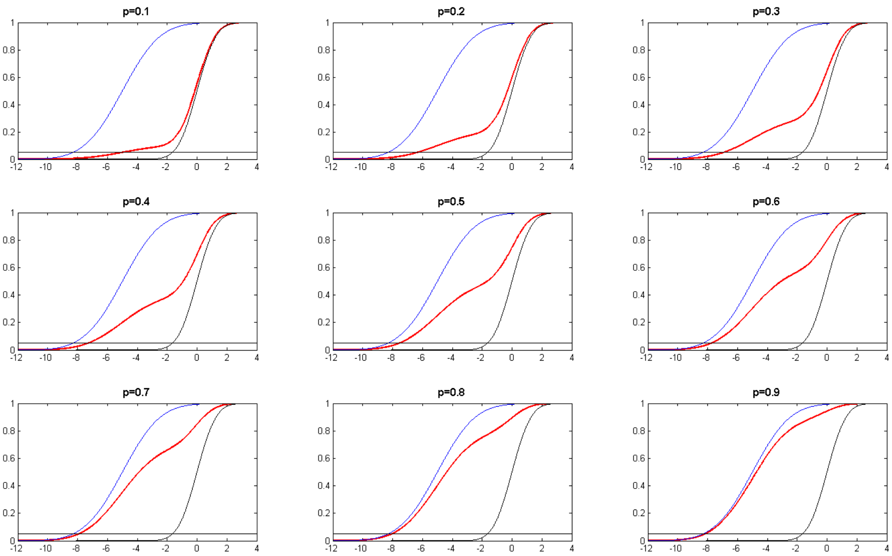


Figure 36: Distribution function of the mixture of N1(-5,2) and N2(0,1) (red line), the distribution function N1(-5,2) (blue line) and the distribution function N2(0,1) (black line) for different values of the mixing parameter. The black horizontal line indicates the $\alpha = 0.05$.

Chapter 2

Volatility specifications versus probability distributions in VaR estimation

Abstract

We provide evidence suggesting that the assumption on the probability distribution for return innovations is more influential for VaR performance than the volatility specification. Even though the more recent Basel requirements center around the Expected Shortfall (ES), a precise estimate of VaR is needed to have a good estimate of ES. We compare 28 alternative combinations of probability distributions on return innovations and volatility specifications on the basis of the results of a variety of VaR tests. To summarize the results of each model on the tests we use a criterion of dominance introduced in this paper. Our results also show that the combination of generalized skewed t distributions and APARCH and FGARCH volatility specifications beat more standard alternatives and should be preferred when estimating tail risk. Clear-cut conclusions on comparisons of this type are clearly helpful for risk management.

2.1 Introduction

A traditional discussion in risk measurement analysis has been whether volatility models that incorporate a leverage effect, with negative innovations having a larger impact on volatility than positive innovations of same size, lead to better Value-at-Risk (VaR) estimates than symmetric volatility models. A second modeling issue refers to whether asymmetric probability distributions for the return innovations lead to an improved VaR model.

The contribution of this paper is to examine the relative importance of the two issues, the volatility specification and the assumption on the probability distribution of return innovations, for the efficiency of VaR estimates. To that end, we have performed an extensive analysis of VaR estimates in assets of different nature, using symmetric and asymmetric probability distributions for the innovations on volatility models with and without leverage. The question is crucial for risk managers, since there are so many potential choices for volatility model and probability distributions that it would be very convenient to establish some priorities in modeling returns for risk estimation.

We consider three general volatility models with leverage, GJR-GARCH, APARCH and FGARCH as well as the standard symmetric GARCH model as benchmark. We like the FGARCH model because it includes as special cases many other volatility specifications, like the symmetric GARCH, GJR-GARCH and APARCH, i.e., it is in fact a nested family of asymmetric GARCH models, thereby allowing for testing how simpler models fit the data. Besides, the model is specified for a power of the conditional standard deviation and the innovations, like the APARCH model, which provides more flexibility to the dynamics of volatility and it allows shifts and rotations in the news impact curve. In principle, these two types of asymmetry are distinct, and they should not be treated as substitutes for each other. In fact, the shift is the dominant source of asymmetry for small shocks, while the rotation is more important for large shocks. As probability distributions for the innovations we compare the performance of the Skewed Student-t distribution and Skewed Generalized Error distribution as introduced in Fernandez and Steel (1998) [41], the Johnson S_U distribution [70], Skewed Generalized-t distribution (Theodossiou, 1998) [123] and Generalized Hyperbolic Skew Student-t distribution (Aas and Haff, 2006) [1], with the Normal and symmetric Student-t distributions as benchmark. In all models we jointly estimate by maximum likelihood the parameters in the equation for the mean return, the equation for its conditional variance and the probability distribution for the return innovations, except with Skewed Generalized-t distribution where we use a two-step estimation [1]. Estimated models are simulated to obtain the implied levels of skewness and kurtosis of returns, which are compared with the analogue sample moments. An interesting feature of our work is the consideration of a variety of assets of different nature: stock

[1]The conditional mean-volatility models are estimated first, and the parameters of the Skewed Generalized-t distribution (SGT) in a second stage because of the difficulty of estimating all parameters jointly. We estimate the parameters for the SGT using the standardized returns obtained from the GED-AR(1)-GARCH specifications estimated in the first step. We use the Generalized Error distribution (GED) in the fist step estimation following Bali and Theodossiou (2007) [9].

market indices, individual stocks, interest rates, commodity prices and exchange rates.

A novel approach of our analysis is to use standard statistical tests to examine the extent to which the estimated probability distributions fit the distribution of empirical return innovations. Additionally, each estimated combination of volatility specification and probability distribution for return innovations determines the distribution of returns themselves. We use simulation methods to analyze whether our estimated models fit the main characteristics of return distributions. These should be expected to be two natural conditions for the good VaR performance of a model. But, in spite of the fact that significant effort is generally placed in selecting an appropriate probability distribution and volatility model, the ability of estimates to explain sample moments is seldom examined.

We calculate VaR estimates following the parametric approach. The performance of VaR estimates is examined through standard tests: the unconditional coverage test of Kupiec (1995) [78], the independence and conditional coverage tests of Christoffersen (1998) [27], the Dynamic Quantile test of Engle and Manganelli (2004) [39], as well as the loss functions proposed by Lopez (1998, 1999) [85, 86] and Sarma et al. (2003) [113] and that of Giacomini and Komunjer (2005) [46].

Our out-of-sample simulation results suggest that the important assumption for VaR performance is that of the probability distribution of the innovations, with the choice of volatility model playing a secondary role. Indeed, validation tests for VaR estimates yield very similar results for a given probability distribution as we change the volatility model. On the contrary, test results drastically change for a given volatility model when we change the assumption on the probability distribution of the innovations. In fact, the main difference arises when we move from symmetric to asymmetric probability distributions for the innovations. This result is relatively consistent with Gerlach et al. (2011) [44].

Relative to the ability to reproduce sample moments, different volatility models with the same probability distribution for the innovations fit sample moments similarly. On the other hand, while it is obviously true that asymmetric distributions are needed to explain the skewness in returns, symmetric and asymmetric probability distributions, imposed on the same volatility model lead to minor differences in kurtosis. The ability of estimated models to fit the empirical distributions of returns and returns innovations seems in fact, a necessary condition for a good VaR performance.

The remainder of the paper is organized as follows. In Section 2.2, we present the review of literature. In Section 2.3, we introduce the volatility models and distributions used in our analysis. In Section 2.4, we present preliminary statistics for our dataset. In Section 2.5, we report the results of the empirical investigation of the estimated models and in Section 2.6 we include the fit to sample moments of the probability distributions and of the returns. In Section 2.7, we provide a description of statistical test and loss function used and we asses VaR performance. Finally, Section 2.8 concludes the paper.

2.2 A review of literature

In VaR estimation we can distinguish between semiparametric/nonparametric and parametric methods. The first group includes: historical simulation, CAViaR, nonparametric estimators of the return distribution and extreme value theory. The historical simulation is by far the most popular procedure for forecasting VaR among commercial banks [see for example Berkowitz, Christoffersen and Pelletier (2011) [16], Pérignon and Smith (2008, 2010b) [103] [104] and Pritsker (2006) [81]]. Pérignon and Smith (2010b) [104] document that almost three quarters of the banks that disclose their VaR methods report using historical simulation. The popularity of the historical simulation method emerges from its simplicity and smoothness, but it also has serious limitations: it is based on the assumption of independent and identically distributed (i.i.d.) returns, which is generally not an adequate assumption. Besides, VaR estimates can exhibit predictable jumps when large negative returns either enter into or drop out of the window used to obtain them. Furthermore, empirical quantiles are not efficient estimators of extreme population quantiles. An alternative VaR forecast procedure is CAViaR, which represents the dynamic evolution of quantiles directly [Engle and Manganelli (2004) [39]]. However, according to Taylor (2008a) [119], estimating expectiles is computationally more attractive than estimating quantiles. Many CAViaR extensions have been proposed; see Gourieroux and Jasiak (2008) [57], De Rossi and Harvey (2009) [30], Gerlach, Chen and Chan (2011) [43], Yu, Li and Jin (2010) [128], Chen, Gerlach, Hwang and McAleer (2012) [24], Huang et al. (2010) [67], Jeon and Taylor (2013) [69], Rubia and Sanchis-Marco (2013) [112], Drakos, Kouretas and Zarangas (2015) [36], Kim and Lee (2016) [75] and Nagai (2016) [96].

Alternatively, several authors have proposed nonparametric estimators of the return distribution, to avoid the effect of potential misspecifications. These nonparametric methods are more complicated computationally, but they can result in inferential gains when the assumptions of the parametric models are wrong. A variety of approaches have been proposed. Chen and Tang (2005) [25] suggest implementing kernel smoothing on the empirical distribution of returns. The goal is to forecast quantiles using a double kernel smoothing estimator of the cumulative distribution (Taylor, 2008b) [120], an approach that provides a greater accuracy for tail quantiles that are changing relatively quickly over time. Xu (2013) [127] suggests using a fully nonparametric quantile regression model based on a double-smoothing local polynomial estimation of the conditional distribution function. Finally, the extreme value theory (EVT), which is a semiparametric method, consists on modeling the tails of the distribution of returns without making any specific assumption concerning the center of the distribution. The tail index parameter in EVT can be estimated nonparametrically, without assuming any particular model for the tail. There are many estimators that can be used to accomplish this task, such as Hill estimator (Hill, 1975) [60] and Pickands estimator (Pickands, 1975) [80]. In practice, the number of data points in the tails are limited, leading to small sample biases. To address this problem, many solutions have been proposed by Huisman, Koedijk, Kool and Palm (2001) [68], Gomes, de Haan and Rodrigues (2008) [51], Gomes, Figueiredo, Rodrigues and Miranda (2012) [52], Gomes, Matins and Neves (2007) [53] and Gomes and Pestana (2007) [54]. Gourieroux and Jasiak (2010a) [56] point out that the accuracy of these non-parametric estimators is rather poor, due to the difficulty of estimating the probability of infrequent events. Another problem is that these estimators depend on the number of observations in a very erratic way.

The second group includes parametric methods. Their distinguishing feature is that they require an explicit specification of the statistical distribution from which the data observations are drawn. To apply any parametric approach it is necessary to estimate the parameters involved and it is important to choose an estimation method that is suitable for the distribution we are dealing with. There are various approaches to parameter estimation, including least squares, maximum likelihood, semi-parametric methods, robust estimation methods and the generalized method of moments.

For the estimation of the tail index parameter in EVT there are also two parametric approaches. The parameter of the distribution of extremes, including tail index, are directly estimated by classical methods such as the maximum likelihood. The first parametric approach is the Block Maxima, which divides the sample into m subsamples of n observations each and picks the maximum of each subsample; see for example Longin (2000) [70], Diebold, Schuermann and Stroughair (2000) [34]. The second EVT parametric approach is the Peak Over Threshold method, according to which any observations that exceed a given high threshold, u, are modeled separately from non-extreme observations. McNeil and Frey (2000) [75] show that the EVT method based on General Pareto distribution gives quantile estimates that are more stable than those from the Hill estimator. Any EVT approach entails choosing an adequate cut-off between the central part of the distribution and the tails. When working with threshold exceedances, the cut-off is induced by the number of observations in the tail, while in the block maxima procedure, it is implied by the choice of the number of blocks. The choice of the cut-off may have severe consequences for the risk estimates. Different authors have proposed methods for optimal threshold selection.

Among parametric methods for VaR estimation, many authors have analyzed the improvement on VaR estimation provided by asymmetric volatility models. Giot and Laurent (2003a) [48] estimated daily VaR for stock indices using different volatility models. They stated that more complex models like APARCH performed better than RiskMetrics or GARCH specifications (for a comparison of volatility models in VaR estimation see also El Babsiri and Zakoian, 2001 [37]). Angelidis, Benos and Degiannakis (2004) [10] show that asymmetric volatility models fare better than symmetric ones, as they capture more efficiently the characteristics of the underlying series and provide better estimators since they perform better in the low probability regions that VaR tries to measure (see also Ane, 2006) [9]. McMillan and Kambouroudis (2009) [91] provide evidence on the performance of alternative VaR models for a large number of individual stocks and exchange rates. They conclude that the APARCH model should be preferred for more extreme VaR estimates, while the RiskMetrics model seems to be adequate at more moderate significance levels. In their work, RiskMetrics seems adequate in providing volatility forecasts for most Asian markets; however, the APARCH model is superior in obtaining forecasts for the G7 markets, as well as other European markets and the larger Asia markets.

Given the widespread evidence on the skewness of the distribution of asset returns, analyzing whether the assumption of an asymmetric distribution of return innovations leads to more efficient VaR estimates is a second methodological issue of interest. Based on the influence of leverage effects on the accuracy of VaR estimates, Brooks and Persand (2003) [21] concluded that models which do not allow for asymmetries either in the unconditional distribution of returns or in the volatility specification underestimate the true VaR. Giot and Laurent (2003a) [48] used daily data for stock market indices and individual stocks, showing that models which rely on a symmetric density for the innovation underperform with respect to skewed density models that require modeling both the left and right tails of the distribution of returns. Lee & Su (2015) [81] estimate the VaR through the accuracy evaluation for the eight stock indices in Europe and Asia stock markets by a parametric approach (only GARCH model) and by the semi-parametric approach of Hull and White. As measures of accuracy these authors use the unconditional coverage test by Kupiec as well as two loss functions, quadratic loss function, and the unexpected loss. They only consider the asymmetric distribution skewed generalized Student-t. They conclude that the skewed generalized Student-t has the best VaR forecasting performance followed by the Student-t, while the Normal has the worst performance in VaR forecasting. Corlu, Meterelliyoz and Tiniç (2016) [29] investigate the ability of five alternative probability distributions to represent the behavior of daily equity index returns over the period 1979-2014: the skewed Student-t distribution, the generalized lambda distribution, the Johnson system of distributions, the normal inverse Gaussian distribution, and the g-and-h distribution. The explanatory power of the distributions is tested using in-sample Value-at-Risk (VaR) failure rates. Their focus is on the unconditional distribution of equity returns, not in conditional distributions.

We consider that financial returns exhibit temporal dependence in mean and volatility, calculate out-of-sample VaR, perform a battery of VaR statistical tests and loss functions, we consider the fit of return innovations moments and return moments and contemplate a broad set of assets of different nature.

More recently, some papers have jointly examined the performance of both, the variance specification and the probability distribution of return innovations in VaR estimation. Gerlach et al. (2011) [44] examine the performance of a wide class of volatility models: RiskMetrics, asymmetric GARCH, IGARCH, GJR-GARCH and EGARCH, under four alternative probability distributions: Gaussian, Student-t, Generalized Error Distribution and Skewed Student-t in VaR estimation at 1% and 5% significance in different time periods (pre-crisis, crisis-GFC and post-crisis) incorporating parameter uncertainty through a Bayesian approach. In such a global analysis, results are varied and hard to summarize, but their evidence suggests a preference for asymmetric probability distributions for the innovations of the return process. Giot and Laurent (2003b) [49] use skewed density models for daily returns on commodities. Bubak (2008) [22], Tu, Wong and Chang (2008) [124], Kang and Yoon (2009) [73] and Diamandis et al. (2011) [31], analyze Eastern and Central European stock markets, Asian stock markets, Asian emerging markets and developed and emerging markets, respectively. Comparing a wide range of univariate conditional variance models, they show that models that incorporate an asymmetric distribution of the error term tend to perform better than models with a symmetric distribution to produce both in-sample and out-of-sample (one-day-ahead) VaR forecasts. Tang and Shieh (2006) [118] and Mabrouk and Saadi (2012) [88] include Fractionally Integrated time varying GARCH models designed to capture not only volatility clustering, but also long memory in assets return volatility. They find that FIAPARCH model, under a skewed Student-t distribution, outperforms the alternative models they consider, including some widely used ones such as GARCH and HYGARCH. However, given that the VaR forecasts required by the Basel accords are short run, the inclusion of long-memory is expected not to make any fundamental differences; for support of this result, see for example So and Yu (2006) [117].

We also examined the performance of both, the variance specification and the probability distribution of return innovations in VaR estimation. We consider a complex and flexible volatility model, FGARCH model proposed by Hentschel (1995 [63]), which is an omnibus model which subsumes some of the most popular GARCH models. To the best of our knowledge, there are no papers examining the performance of this model for VaR estimation. Besides, we introduce distributions rarely used in the literature on VaR performance, such as Skewed Generalized Error Distribution [Fernandez and Steel (1998) [41]], Johnson S_U distribution [Johnson (1949) [70]], Skewed Generalized-t [Theodossiou (1998) [123]] and Generalized Hyperbolic Skew Student-t distribution (GHST) [Aas and Haff (2016) [1]]. In the VaR literature, Johnson distributions are suggested in Zangari (1996) [129], Mina and Ulmer (1999) [94], in RiskMetrics Technical Document (1996) [110] and Choi and Nam (2008) [26]. The last one examines empirically a GARCH model with Johnson innovations. Simonato (2011) [115] documents the performance of the Johnson system relative to closely competing approaches, such as the Gram-Charlier and Cornish-Fisher approximations. He considers the case of Expected Shortfall computation without performing a backtesting analysis, just comparing the moments of the distributions and root-mean-squared errors. The GHST distribution has hardly been employed in financial applications because its estimation is computationally demanding. Nakajima and Omori (2012) [97] use GHST distribution but they perform a Bayesian analysis of a stochastic volatility model. Hu (2005) [66] estimates Multivariate Generalized Hyperbolic Distribution using the EM algorithm. Paolella and Polak (2015) [101] also use Generalized Hyperbolic distribution in a context of multivariate time series.

Relative to this ever increasing literature, we contribute in different ways: i) considering a set of probability distributions that have recently been rendered to be appropriate to capture the skewness and kurtosis of financial data, but whose performance for VaR estimation has not been compared previously on a common dataset, ii) considering volatility models with leverage, APARCH and FGARCH, that again, have been appreciated as being adequate for financial returns, iii) by explicitly evaluating their fit to return data, relating that fitting ability to their VaR performance, iv) by introducing a dominance criterion to establish a ranking of models on the basis of their behavior under standard VaR validation tests and loss functions.

Tables 2.1-2.2 show an overview of the recent literature.

	APARCH	Skewed distribution	Variety of assets	Significance Level VaR
Aas & Aaf (2006)		NIG, AC, GHST	2 indices,, 1 interest rate, 1 exchange rate	0.5%, 1%, 5%, 95%, 99%, 99.5%
Abad & Benito (2010)			8 indices	1%
Abad, Benito & Lopez (2015)			1 commodity, 1 index	0.1%, 0.5%, 1%, 5%
Abad, Benito, Sanchez-Granero & Lopez (2013)		H, SGT, SGED, JSU	9 indices	0.25%, 1%
Andersen, Bollerslev, Diebold & Labys (2003)			2 exchange rates	1%, 5%, 10%, 90%, 95%, 99%
Angelidis, Benos & Degiannakis (2007)	X	LL	2 indices	1%, 2.5%, 97.5%, 99%
Bali & Theodossiou (2007)		SGT	1 index	0.5%, 1%, 1.5%, 2%, 2.5%, 5%
Bubak (2008)	X	LL	4 indices, 14 stocks	0.5%, 1%, 2.5%, 5%, 95%, 97.5%, 99%, 99.5%
Corlu, Meterelliyoz and Tiniç (2016)		GLD, JSU, AC, gh, NIG	20 indices	0.5%, 1%, 5%, 95%, 99%, 99.5%
Diamandis, Drakos, Kouretas & Zarangas (2011)	X	LL	21 indices	0.25%, 0.5%,1%, 2.5%, 5%
Ergün & Jun (2010)		H	1 indices	0.1%, 1%, 5%, 95%, 99%, 99%
Gerlach, Chen, Lin & Lee (2011)		H	4 stocks	1%, 5%
Giot & Laurent (2003a)	X	LL	3 indices, 3 stocks	0.25%, 0.5%, 1%, 2.5%, 5%, 95% 97.5%, 99%, 99.5%, 99.25%
Giot & Laurent (2003b)	X	LL	3 commodities	0.25%, 0.5%, 1%, 2.5%, 5%, 95% 97.5%, 99%, 99.5%, 99.25%
Kang & Yong (2009)	X	LL	5 indices	0.25%, 0.5%, 1%, 2.5%, 5%, 95% 97.5%, 99%, 99.5%, 99.25%
Kuester, Mittnik & Paolella (2006)		P	1 index	1%, 2.5%, 5%
Lee & Su (2015)		SGT	8 indices	0.5%, 1%, 5%
Louzis, Xanthopoulus-Sisinis & Refenes (2014)		LL	1 index, 1 exchange rate, 1 future, 1 commodity	1%
Mabrouk & Saadi (2012)	FIAPARCH	LL	7 indices	0.25%, 0.5%, 1%, 2.5%, 5%, 95% 97.5%, 99%, 99.5%, 99.25%
McMillan & Kambouroudis (2009)	X		31 indices	1%, 5%
Ozun, Cifter & Yilmazer (2010)	X	LL	1 index	1%, 5%
Polanski & Stoja (2010)		SGT, EGB2	3 indices	0.5%, 1%, 1.5%, 2%, 5%
So & Yu (2006)			12 indices, 4 exchange rates	1%, 2.5%, 5%, 95%, 97.5%, 99%
Tang & Shieh (2006)		LL	3 indices	0.25%, 0.5%, 1%, 2.5%, 5%, 95% 97.5%, 99%, 99.5%, 99.25%
Tu, Wong & Chang (2008)	X	LL	10 indices	0.5%, 1%, 2.5%, 5%, 95% 97.5%, 99%, 99.5%

Table 2.1: Overview of papers that compare VaR models.

Note: The table summarizes some characteristics of empirical papers involving comparison of VaR models and the variety of assets. The volatility model and the VaR methodology are marked with a cross when they are included in a paper. The skewed distributions included in these table are the following: the skewed Student-t distribution of Hansen (H) [59], the skewed Student-t distribution of Lambert & Laurent (LL) [80], the skewed Studen-t proposed by Azzalini and Capitanio (AC) [7], the skewed generalised-t distribution of Theodossiou (SGT) [123], the skewed error generalised distribution of Theodossiou (SGED) [122], the Johnson S_U distribution (JSU) [70], the generalized asymmetric t distribution of Paolella [100] and Mittnik and Paolella [95] (P), the exponential generalized beta of the second kind (EGB2) [90], the generalized lambda distribution (GLD) [42] [71] [109], g-and-h distribution (gh) [125], the Normal Inverse Gaussian of Barndorff-Nielsen (NIG) [15] and the Generalized Skew Student-t of Aas and Haff (GHST) [1].

	Tests VaR	Loss Functions
Aas & Aaf (2006)	FR	
Abad & Benito (2010)	FR, LR_{uc}, LR_{ind}, LR_{cc}, BTC, DQT	RQ
Abad, Benito & Lopez (2015)	FR, LR_{uc}, LR_{ind}, LR_{cc}, BTC	RQL, RL, RQ, RC_1, RC_2, RC_3, FS, FC_1, FC_2, FC_3, FABL
Abad, Benito, Sanchez-Granero & Lopez (2013)	FR, LR_{uc}, LR_{ind}, LR_{cc}, BTC, DQT	RQ, FS
Andersen, Bollerslev, Diebold & Labys (2003)	FR	
Angelidis, Benos & Degiannakis (2007)	FR, LR_{uc}	
Bali & Theodossiou (2007)	FR, LR_{uc}, LR_{ind}, LR_{cc}	
Bubak (2008)	FR, LR_{uc}	
Corlu, Meterelliyoz and Tiniç (2016)	FR, LR_{uc}	
Diamandis, Drakos, Kouretas & Zarangas (2011)	FR, LR_{uc}, LR_{ind}, LR_{cc}, DQT	
Ergün & Jun (2010)	FR, LR_{uc}, LR_{ind}, LR_{cc}	
Gerlach, Chen, Lin & Lee (2011)	FR, MRC, LR_{uc}, LR_{cc}	RC_3
Giot & Laurent (2003a)	LR_{uc}	
Giot & Laurent (2003b)	LR_{uc}	
Kang & Yong (2009)	LR_{uc}	
Kuester, Mittnik & Paolella (2006)	FR, LR_{uc}, LR_{ind}, LR_{cc}, DQT	
Lee & Su (2015)	FR, LR_{uc}	RQL, UL
Louzis, Xanthopoulus-Sisinis & Refenes (2014)	FR, MRC, LR_{uc}, LR_{ind}, LR_{cc}, DQT	RQL, FS
Mabrouk & Saadi (2012)	FR, LR_{uc}, DQT	
McMillan & Kambouroudis (2009)	FR, LR_{uc}, DQT	
Ozun, Cifter & Yilmazer (2010)	LR_{uc}, LR_{cc}	RQL
Polanski & Stoja (2010)	FR, LR_{uc}, LR_{cc}	
So & Yu (2006)	FR	
Tang & Shieh (2006)	FR, LR_{uc}	
Tu, Wong & Chang (2008)	FR, LR_{uc}	

Table 2.2: Overview of papers that compare VaR models (continued).

Note: The table summarizes the tests used to evaluate the accuracy of VaR models and the loss functions used in the comparative exercise. We indicate the test to evaluate the accuracy of VaR models and/or the loss function used in the comparative exercise. FR is the failure rate, MRC is the Market Risk Charge, LR_{uc} is the unconditional coverage test [78], LR_{ind} is the statistic for the serial independence, LR_{cc} is the conditional coverage test [27], BTC is the Back-Testing Criterion, DQT is the Dynamic Quantile Test [39], RQL is the Regulator's Quadratic loss function of Lopez [85, 86], RL is Regulator's Lineal loss function of Lopez [85, 86], RQ is the Regulator's Quadratic Loss Function of Sarma et al. [113] , RC_1, RC_2 and RC_3 are the Regulator's loss functions of Caporin [23], FS is the Firm's loss function of Sarma et al. [113], FC_1, FC_2 and FC_3 are the Firm's loss functions of Caporin [23], FABL is the Firm's loss functions of Abad, Benito and Lopez [3] and UL is unexpected loss of Lee and Su [81].

2.3 Models and probability distributions

Let x_t, for $t = 1, ..., T$, be a time series of asset returns. It is convenient to break down the complete characterization of x_t into three components: (i) the conditional mean, μ_t (ii) the conditional variance, which contains the scale parameter that measures the dispersion of the distribution, σ_t^2 and (iii) the shape parameters (e.g., skewness, kurtosis) that determine the form of a conditional distribution within the general family of distributions. Thus, we may write

$$x_t = \mu_t(\theta) + \varepsilon_t \qquad \mu_t(\theta) = \mathbb{E}[x_t|\mathcal{F}_{t-1}] = \mu(\theta, \mathcal{F}_{t-1}) \qquad \varepsilon_t = \sigma_t(\theta) z_t$$

$$\sigma_t^2(\theta) = \mathbb{E}[(x_t - \mu_t)^2|\mathcal{F}_{t-1}] = \sigma^2(\theta, \mathcal{F}_{t-1}) \qquad z_t \sim f(z_t|\theta)$$

The standardized innovation, $z_t = (x_t - \mu_t(\theta))/\sigma_t(\theta)$ has zero mean and unit variance and it follows a conditional distribution f. Vector θ contains all the parameters associated with the conditional mean, the conditional variance and conditional distribution. In the last, we have shape parameters which capture asymmetry and fat-tailedness of the distribution, except if the conditional distribution is assumed to be $\mathcal{N}(0,1)$, we do not have shape parameters. We estimate all parameters jointly by Maximum Likelihood.

An AR(1) model was specified for the conditional mean return in all cases, which is sufficient to produce serially uncorrelated innovations. We consider three general volatility models with leverage, GJR-GARCH, APARCH and FGARCH as well as the standard symmetric GARCH model as benchmark. As probability distributions for the innovations we compare the performance of Skewed Student-t, Skewed Generalized Error, Johnson S_U, Skewed Generalized-t and Generalized Hyperbolic Skew Student-t distributions, with the Normal and symmetric Student-t distributions as benchmark. We provide a description of volatility models and probability distributions in Appendices A.1 - A.9.

2.4 The data

We work with daily percentage returns on five groups of assets of different nature over the sample period 01/04/2000 - 12/31/2015 (4173 observations). Daily returns are computed as 100 times the first difference of log prices, i.e. $100[\ln(P_{t+1}) - \ln(P_t)]\%$. The financial assets considered are: stock market indices: IBEX 35 (€), NASDAQ 100 ($), FTSE 100 (£) and NIKKEI 225 (¥); individual stocks: International Business Machines [IBM] ($), Banco Santander [SAN] (€), AXA (€) and BP (£); interest rates: Interest Rate Swap 5Y [IRS 5Y] (€), interest rate of GERMAN BOND 10Y (€) and interest rate of US BOND 10Y($); commodity prices CRUDE OIL BRENT ($ per barrel), NATURAL GAS ($ per Million British Thermal Units), GOLD ($ per Troy Ounce) and SILVER (Cents $ per Troy Ounce) and exchange rates EUR/USD (€), GBP/USD (£), JPY/USD (¥) and AUD/USD (Australian $). The data were extracted from Datastream.

Table 2.3 reports descriptive statistics for daily returns. All of them have a mean close to zero. Median returns are zero, except for IBEX, NASDAQ, IRS, interest rate of GERMAN BOND, interest rate of US BOND, GOLD, JPY/USD and AUD/USD. GAS is the one with a wider total range ($max - min$) followed by AXA and SAN. In general, interest rates have a narrow range. The unconditional standard deviation is specially high for GAS (4.19), AXA (2.67), OIL BRENT (2.28) and SAN (2.19) and very low for interest rates and exchange rates. For the rest of assets, the standard deviation moves between 1.13 GOLD and 1.85 NASDAQ. According to the skewness statistic, AUD/USD, SILVER, GOLD and NIKKEI have relatively high negative skewness, while GAS, AXA, JPY/USD and NASDAQ have relatively high positive skewness. GBP/USD, IBM and the interest rate of GERMAN BOND display a small negative skewness. For all assets considered, the kurtosis statistic is large, implying that the distributions of those returns have much thicker tails than Normal distribution. For instance, AUD/USD, GAS, IBM and AXA have a high degree of kurtosis while EUR/USD, interest rate of GERMAN BOND and JPY/USD have a small kurtosis. Similarly, the Jarque-Bera statistic is statistically significant, rejecting the assumption of normality in all cases.

	Mean (%)	Median (%)	Max	Min	S.D.	Skewness	Kurtosis	J-B
IBEX	-0.47	2.89	13.48	-9.58	1.49	0.08	7.93	4234.84
NASDAQ	0.46	3.68	17.20	-11.11	1.85	0.19	9.62	7652.53
FTSE	-0.25	0	9.38	-9.26	1.21	-0.16	9.36	7042.80
NIKKEI	0.01	0	13.23	-12.11	1.50	-0.41	9.72	7979.58
IBM	0.42	0	12.26	-16.89	1.66	-0.07	11.63	12947.74
SAN	1.01	0	20.87	-15.19	2.19	0.15	9.11	6515.50
AXA	0.55	0	19.78	-20.35	2.67	0.27	10.09	8790.79
BP	-1.35	0	10.58	-14.04	1.71	-0.13	7.81	4041.28
IRS	0.55	0.48	1.92	-1.86	0.21	-0.28	8.53	5367.17
GER BOND	1.11	0.97	3.39	-2.33	0.41	-0.09	5.97	1536.83
US BOND	0.98	0.96	4.53	-5.57	0.59	-0.22	7.96	4307.77
BRENT	0.98	0	17.97	-18.72	2.28	-0.19	8.26	4831.81
GAS	0.01	0	37.81	-28.90	4.19	0.56	12.81	16946.14
GOLD	3.10	0.01	6.86	-10.16	1.13	-0.41	8.81	5991.49
SILVER	2.26	0	13.66	-12.98	1.93	-0.57	8.62	5724.23
EUR/USD	0.16	0	4.62	-3.84	0.63	0.14	5.48	1091.11
GBP/USD	-0.20	0	4.43	-3.88	0.57	-0.04	7.27	3170.80
JPY/USD	-0.41	-0.99	4.61	-3.71	0.63	0.27	6.96	2779.74
AUD/USD	0.23	1.86	6.70	-8.83	0.83	-0.82	15.13	26058.43

Table 2.3: Descriptive statistics for the daily percentage returns.

Figure 2.1 displays daily percentage returns of each stock market indices. It is clear from the graph that large price changes tend to also be followed by large changes, and small changes tend to follow small changes. Such volatility clustering is a property of asset prices that each index seems to exhibit. This graphical evidence is an indication of the presence of ARCH effect in our daily returns series that should be accounted for when estimating Value at Risk. Figure 2.1 also displays QQ-plot of each index against the Normal distribution. These QQ-plot show that all returns distributions exhibit fat tails and also fat tails are not symmetric.

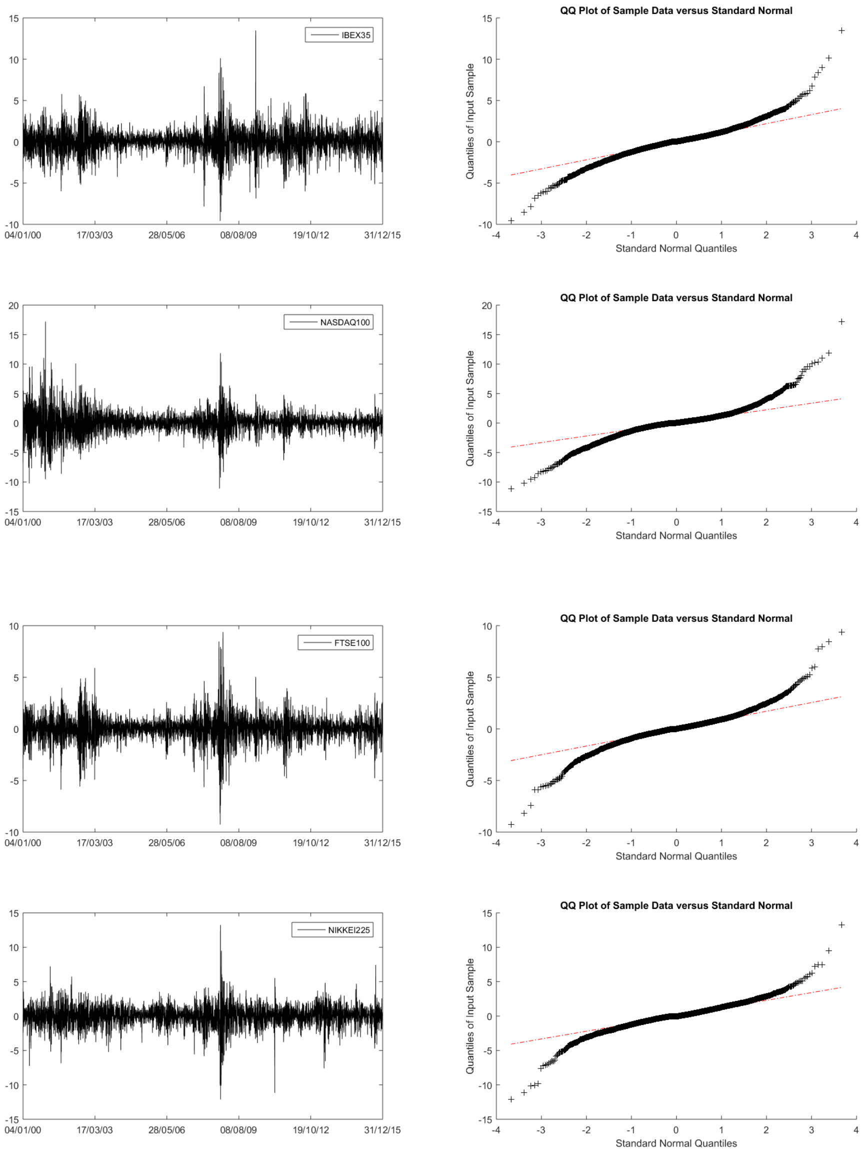


Figure 2.1: Stock market indices daily percentage returns and QQ-plot against the Normal distribution.

2.5 Parameter estimates

To perform a VaR analysis we estimate four volatility models: GARCH, GJR-GARCH, APARCH and FGARCH under each of the different probability distributions assumed for the innovations: Gaussian, Student-t, skewed Student-t, skewed generalized error, unbiased Johnson S_U, skewed generalized-t and generalized hyperbolic skewed Student-t distributions. To save space, we only report estimation results of the APARCH model under the different probability distributions in Tables 2.4 - 2.8 for the different assets according to their nature. Results for alternative models are available from the authors upon request [2].

In Tables 2.4 - 2.8 we observe that the APARCH model is particularly successful in capturing the heteroscedasticity exhibited by the data. We show the parameter estimates under the seven probability distributions. The Ljung-Box Q statistic for five lags computed on the standardized residuals does not show evidence of autocorrelation at 1% significance level except for GAS. But for one lag, GAS does not show autocorrelation at 1%, inasmuch as the p-values of the Q statistics are 0.0899, 0.2621, 0.2440, 0.0452, 0.2288, 0.0447 and 0.4053 for N-, ST-, SKST-, SGED-, JSU-, SGT- and GHST-APARCH models, respectively. The same statistic computed with nine lags on the squared standardized residuals is not significant at 1% except for IBEX, SAN, IRS, GERMAN BOND, OIL, GOLD and SILVER. If we consider one lag, we obtain a Q statistic not significant at 1% significance level for IBEX and SAN but it remains significant for the remaining assets. A significant statistic indicates a possible problem with this model. In the lower panels of these tables we present the log-likelihood values of the four volatility models (GARCH, GJR-GARCH, APARCH and FGARCH). Their similarity suggests that the implied volatility specifications are very similar. The autoregressive effect in volatility is strong, with a β_1-parameter generally above 0.90, suggesting strong memory effects. The range of β_1 is [0.88, 0.97] where the minimum is obtained for GAS and the maximum is obtained for EUR/USD. The coefficient γ_1 is positive and statistically significant for most series, indicating the existence of a leverage effect for negative returns in the conditional variance specification. Estimates of γ_1 are close to 1 for IBEX, NASDAQ and FTSE (in the GJR-GARCH model we also obtain an α_1-parameter close to 0). Compared to estimates for other assets these values are very high, suggesting that only negative shocks contribute to volatility. We also obtain a γ_1 estimate close to 1 in the APARCH model (equivalently α_1 close to 0 in GJR-GARCH) with other indices not considered in this paper such as CAC 40, DAX 30 and S&P 500 for this same sample period. We obtain the same parameter estimates for these models using MatLab, R, Eviews and Gretl. The coefficient γ_1 is negative and statistically significant for interest rates with some models, GOLD, SILVER and JPY/USD, indicating that a positive shock generates greater volatility than a negative shock of equal size.

[2] All computations were performed with the *rugarch* package (version 1.3-4) of R software (version 3.1.1) designed for the estimation and forecast of various univariate ARCH-type models, except in the estimation of models under the Skewed Generalized-t and Generalized Hyperbolic Skew Student-t distributions that we used the *sgt* package (version 2.0) and the *SkewHyperbolic* package (version 0.3-2), respectively.

It is also important that estimates of ξ in the skewed Student-t and Skewed Generalized Error are less than 1 for most assets, suggesting the convenience of incorporating negative asymmetric features in the probability distribution in order to model innovations appropriately. A similar consideration applies to the skewness parameter γ of the Johnson S_U, λ of the Skewed Generalized-t and β of Generalized Hyperbolic Skew Student-t, which in these cases the skewness parameters have negative sign. We obtain positive skewness with GAS and GOLD with some models, EUR/USD and JPY/USD. According to kurtosis, the estimates of ν (Student-t and Skewed Student-t) and δ(Johnson S_U) are between 1.35 and 12.50, capturing the heavy tails of the distribution. The smallest values are obtained with Johnson S_U. The kurtosis parameters η and p of Skewed Generalized Error and Skewed Generalized-t, respectively, measure the peakness of the distribution. For most assets and with most models, we obtain values lower than 2 indicating that the distribution is leptokurtic. Note that Skewed Generalized-t have two parameters related to kurtosis, p and q. The parameters p and q control the peak and the tails of density, respectively. And the parameter q only has the degrees of freedom interpretation in case $\lambda = 0$ and $p = 2$. We obtain high q values accompanied with low p values for some assets, indicating in these cases that the kurtosis is mainly due to higher peak, rather than thicker tails of the distribution. Finally, δ takes value between 0.95 and 2.33, being significantly different from 2 in most cases. These results suggest that, contrary to standard practice, we should model the conditional standard deviation, rather than the conditional variance [3]. Our estimates for the different asset classes suggest that we should model the conditional standard deviation for stock market indices, individual stocks and commodities (metals), the conditional variance ($\delta = 2$) for interest rates, and something between conditional standard deviation and variance ($\delta = 1.5$) for commodities (energy) and exchange rates. In summary, these results indicate the need for a model featuring a negative leverage effect in the conditional variance equation (conditional asymmetry) combined with an asymmetric distribution for the underlying error term (unconditional asymmetry) when representing market data.

[3] This result is in line with those of Taylor (1986) [121], Schwert (1990) [114] and Ding et al. (1993) [35] who indicate that there is substantially more correlation among absolute returns than among squared returns, a reflection of the 'long memory' of high-frequency financial return.

	N-APARCH				ST-APARCH				SKST-APARCH				SGED-APARCH				JSU-APARCH				SGT-APARCH				GHST-APARCH			
	IBEX	NASDAQ	FTSE	NIKKEI	IBEX	NASDAQ	FTSE	NIKKEI	IBEX	NASDAQ	FTSE	NIKKEI	IBEX	NASDAQ	FTSE	NIKKEI	IBEX	NASDAQ	FTSE	NIKKEI	IBEX	NASDAQ	FTSE	NIKKEI	IBEX	NASDAQ	FTSE	NIKKEI
μ	0.00346	0.02519	-0.00911	0.01048	0.02506	0.05402	0.00792	0.01254	0.00752	0.03093	-0.00812	0.01254	0.00358	0.02680	-0.01119	-0.00216	0.00753	0.03217	-0.00822	0.01139	0.02655	0.05178	0.00817	0.01254	-0.00697	0.02261	-0.00967	-0.00887
ϕ_1	0.00916	-0.03500	-0.03540	-0.01187	0.00775	-0.03276	-0.03241	-0.02012	-0.00101	-0.04474	-0.04171	-0.02389	-0.00143	-0.03828	-0.03983	-0.02278	-0.00015	-0.04282	-0.04123	-0.02483	0.00678	-0.02447	-0.02882	-0.01669	-0.00323	-0.04478	-0.03889	-0.02592
ω	0.02370	0.01977	0.01996	0.04707	0.01914	0.01594	0.01901	0.04177	0.01759	0.0160	0.01954	0.04171	0.02219	0.01894	0.02025	0.04446	0.02018	0.01725	0.01929	0.04149	0.02098	0.01759	0.01945	0.04380	0.02104	0.01771	0.02019	0.04264
α_1	0.05989	0.04828	0.06423	0.08889	0.06028	0.05462	0.07168	0.07584	0.05699	0.04988	0.07104	0.07620	0.06037	0.05633	0.06934	0.08175	0.06005	0.05786	0.07008	0.07674	0.06048	0.05421	0.06885	0.08178	0.06204	0.06124	0.07458	0.07771
γ_1	0.99999	0.99996	0.99999	0.41495	0.99999	0.99961	0.99999	0.65828	0.99999	1	0.99999	0.63868	0.99999	0.99997	0.99999	0.58538	1.00000	0.99999	1.00000	0.61973	0.99999	0.99998	0.99999	0.58551	0.99999	1.00000	0.99999	0.62190
β_1	0.93056	0.93293	0.92071	0.89362	0.93380	0.93442	0.92100	0.90911	0.93375	0.93512	0.92252	0.90881	0.93169	0.93193	0.92151	0.90379	0.93389	0.93402	0.92298	0.90868	0.93163	0.93167	0.91973	0.90379	0.93454	0.93473	0.92209	0.91044
δ	1.15445	1.44813	1.23332	1.45041	1.13205	1.31798	1.10371	1.22593	1.27901	1.42125	1.19407	1.22654	1.14450	1.31073	1.12985	1.25059	1.13403	1.26183	1.10837	1.22986	1.14118	1.35943	1.16286	1.25555	1.12816	1.22402	1.07694	1.22180
ξ skewness									0.905903	0.89431	0.87372	0.93325	0.91335	0.91073	0.88128	0.97689												
ν/η kurtosis					9.34167	8.69135	11.02488	7.78825	9.43469	9.50618	12.50485	8.17265	1.51441	1.44312	1.58483	1.28553												
γ skewness																	-0.43680	-0.48479	-0.77439	-0.32977								
δ kurtosis																	2.35587	2.25426	2.69699	2.11659								
λ skewness																					-0.0851	-0.0798	-0.1166	-0.0135				
p peakedness																					1.6616	1.4318	1.5822	1.2791				
q tail-thickness																					14.6977	44514.4041	485.3757	46575.4178				
β skewness																									-1.05492	-1.09668	-0.89900	-1.07513
ν kurtosis																									8.19653	8.19639	8.19704	8.19646
Ljung-Box Test on Standardized Residuals Lag[5]																												
statistic	2.1724	3.0733	2.9535	0.8080	2.3193	3.5274	3.0940	1.2543	3.123	3.4560	3.4607	1.7140	3.1750	3.0911	3.2666	1.7840	3.0140	3.3590	3.3990	1.8810	2.4113	3.8954	3.0402	0.9848	3.4950	3.3718	3.2045	2.3160
p-value	0.6613	0.4130	0.4431	0.9714	0.6184	0.3110	0.4080	0.9018	0.4009	0.3257	0.3248	0.7915	0.3884	0.4087	0.3673	0.7724	0.4278	0.3466	0.3378	0.7453	0.5918	0.2426	0.4212	0.9491	0.3176	0.3438	0.3816	0.6195
Ljung-Box Test on Standardized Squared Residuals Lag[9]																												
statistic	14.3540	12.6100	5.0950	5.0970	13.1550	11.8330	4.3310	7.0743	13.1260	12.3900	4.4250	6.8504	13.6500	12.8600	4.7710	6.2090	13.1100	12.3300	4.436	6.6131	13.6600	12.3200	4.6660	6.1380	12.8430	12.3600	4.3320	6.8393
p-value	0.0051	0.0131	0.4156	0.4153	0.0098	0.0198	0.5336	0.1930	0.0099	0.0148	0.5184	0.2120	0.0075	0.0115	0.4637	0.2748	0.0101	0.0153	0.5165	0.2338	0.0075	0.0153	0.4800	0.2826	0.0116	0.0150	0.5336	0.2130
Log-Likelihoods																												
FGARCH	-6857.293	-7173.316	-5734.019	-7043.726	-6823.375	-7149.125	-5709.985	-6981.778	-6811.683	-7129.065	-5689.814	-6974.403	-6809.176	-7121.970	-5685.727	-6956.129	-6810.914	-7137.549	-5686.932	-6970.236	-6820.942	-7129.161	-5705.586	-6958.155	-6826.554	-7143.381	-5694.074	-6972.850
APARCH	-6868.201	-7199.368	-5752.936	-7067.510	-6827.756	-7162.948	-5726.196	-7004.187	-6817.215	-7148.359	-5706.765	-6997.693	-6815.463	-7131.517	-5702.196	-6976.295	-6816.443	-7145.028	-5703.821	-6993.738	-6826.387	-7142.862	-5721.051	-6977.464	-6823.141	-7150.344	-5709.698	-6998.739
GJRGARCH	-6883.275	-7205.335	-5760.899	-7072.092	-6840.630	-7170.654	-5735.475	-7012.864	-6830.409	-7157.616	-5716.723	-7006.270	-6827.821	-7139.414	-5711.072	-6983.184	-6829.479	-7154.870	-5713.324	-7001.968	-6838.401	-7149.12	-5728.725	-6984.185	-6840.208	-7180.600	-5719.975	-7027.461
GARCH	-6970.794	-7268.542	-5850.492	-7106.658	-6907.600	-7222.490	-5850.492	-7043.798	-6900.903	-7215.762	-5799.776	-7036.438	-6897.008	-7193.125	-5789.600	-7009.636	-6899.894	-7213.013	-5796.144	-7031.025	-6903.445	-7197.259	-5799.959	-7010.538	-6907.563	-7224.668	-5799.967	-7034.622

Table 2.4: Parameter estimates of APARCH model for stock market indices under different probability distributions.

	N-APARCH				ST-APARCH				SKST-APARCH				SGED-APARCH				JSU-APARCH				SGT-APARCH				GHST-APARCH			
	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP	IBM	SAN	AXA	BP
μ	-0.00030	-0.01224	-0.01056	-0.01921	0.01072	0.01082	0.00887	-0.01080	0.00208	-0.00925	-0.00727	-0.02150	-0.00701	-0.01321	-0.00889	-0.02536	-0.00183	-0.00943	-0.00808	-0.02175	0.00370	0.01119	0.01047	-0.00631	-0.03240	-0.02694	-0.04348	-0.06164
ϕ_1	-0.02063	-0.01502	0.01785	-0.01874	-0.02694	-0.01279	0.02072	-0.01652	-0.02738	-0.01714	0.02013	-0.01599	-0.02087	-0.01750	0.01879	-0.01185	-0.02746	-0.01694	0.02013	-0.01595	-0.01727	-0.01177	-0.01818	-0.01113	-0.02347	-0.01894	0.02030	-0.01512
ω	0.02742	0.03310	0.03234	0.04707	0.01528	0.02621	0.03072	0.02445	0.01554	0.02742	0.03114	0.02452	0.02015	0.02987	0.03295	0.02739	0.01594	0.02766	0.02487	0.04140	0.01998	0.02853	0.03254	0.02700	0.01525	0.02886	0.03231	0.02696
α_1	0.07871	0.06751	0.05334	0.08889	0.07102	0.06809	0.06451	0.06749	0.07061	0.06851	0.06362	0.06635	0.07398	0.06916	0.06902	0.06111	0.06992	0.06872	0.06402	0.06569	0.07449	0.06922	0.06981	0.06195	0.05875	0.06932	0.06274	0.06218
γ_1	0.53084	0.81436	0.69978	0.67314	0.53185	0.80632	0.76565	0.58662	0.53969	0.81606	0.79219	0.60617	0.53969	0.81606	0.53388	0.81747	0.74415	0.63103	0.78804	0.60735	0.52443	0.80467	0.72367	0.60845	0.60721	0.81937	0.85278	0.67477
β_1	0.92645	0.93293	0.92738	0.93458	0.93768	0.93181	0.93439	0.93173	0.93803	0.93179	0.93530	0.93291	0.93199	0.92903	0.93101	0.93419	0.93797	0.93138	0.93508	0.93308	0.93144	0.92867	0.93011	0.93321	0.94642	0.93192	0.93815	0.93755
δ	0.97548	1.26759	1.07037	1.34272	0.99638	1.14563	1.12595	1.15755	0.99353	1.12851	1.11760	1.14768	0.98185	1.17336	1.07988	1.20524	0.98521	1.13434	1.11377	1.15349	0.98857	1.19012	1.54067	1.21690	1.01156	1.12808	1.10735	1.13736
ξ skewness									0.97781	0.93722	0.95295	0.96364	0.98772	0.95371	0.95970	0.97914												
ν/η kurtosis					4.82633	7.89389	9.824196	7.538588	4.83464	8.02733	9.94087	7.54868	1.18036	1.40931	1.54665	1.38413												
γ skewness																	-0.06871	-0.24757	-0.21555	-0.13045								
δ kurtosis																	1.55408	2.08511	2.39059	2.01485								
λ skewness																					-0.0204	-0.0446	-0.0443	-0.0230				
p peakedness																					1.9354	1.5397	1.9928	1.6274				
q tail-thickness																					2.6644	16.9412	5.0522	9.3767				
β skewness																									-1.07510	-0.83842	0.89812	-1.14018
ν kurtosis																									8.19648	8.19724	8.19704	8.19625
Ljung-Box Test on Standardized Residuals Lag[5]																												
statistic	1.6299	2.2824	6.1028	5.3090	1.8911	2.0989	6.0561	4.8190	1.9666	2.3960	5.9941	4.7612	1.7920	2.4656	6.0493	4.8480	1.9822	2.3797	5.9844	4.7851	1.7881	2.0745	6.2108	4.9456	1.9830	2.5705	5.7598	4.7089
p-value	0.8138	0.6292	0.04263	0.0829	0.7425	0.6827	0.0444	0.1226	0.7209	0.5962	0.0468	0.7915	0.7703	0.5762	0.0446	0.1198	0.7165	0.6009	0.0472	0.1258	0.7713	0.6898	0.0388	0.1110	0.7162	0.5464	0.0571	0.1335
Ljung-Box Test on Standardized Squared Residuals Lag[9]																												
statistic	0.7930	12.7591	7.6448	3.3190	0.7345	13.2923	9.6050	1.9707	0.7471	13.2453	10.1310	2.0960	0.7037	12.9138	8.5296	2.4953	0.7612	13.1766	9.9960	2.1371	0.6958	12.9515	8.1854	2.2980	1.1983	13.1809	11.7490	2.9528
p-value	0.9933	0.01215	0.1509	0.7053	0.9947	0.0091	0.0608	0.9085	0.9944	0.0093	0.0470	0.8932	0.9954	0.0112	0.1013	0.8384	0.9941	0.0097	0.0503	0.8880	0.9955	0.0109	0.1186	0.8666	0.6765	0.0097	0.0206	0.7667
Log-Likelihoods																												
FGARCH	-7255.489	-8356.278	-8949.090	-7547.478	-7031.218	-8295.946	-8910.258	-7475.460	-7030.560	-8290.907	-8907.609	-7473.707	-7053.236	-8286.609	-8913.446	-7474.457	-7032.462	-8289.825	-8907.877	-7473.873	-7053.612	-8290.177	-8915.944	-7475.776	-7069.200	-8305.243	-8910.200	-7493.013
APARCH	-7258.499	-8359.171	-8960.933	-7550.978	-7031.730	-8297.918	-8918.848	-7478.854	-7031.149	-8293.079	-8916.456	-7477.268	-7054.174	-8288.537	-8922.595	-7476.979	-7033.070	-8291.966	-8916.729	-7477.180	-7054.508	-8292.016	-8924.736	-7478.075	-7072.793	-8296.995	-8926.393	-7499.351
GJRGARCH	-7291.113	-8371.481	-8974.059	-7556.869	-7052.943	-8311.735	-8927.718	-7487.892	-7052.613	-8307.714	-8925.753	-7486.571	-7075.011	-8300.844	-8932.716	-7484.128	-7054.587	-8306.28	-8926.076	-7486.198	-7075.137	-8303.569	-8934.517	-7484.956	-7086.052	-8315.64	-8941.190	-7499.048
GARCH	-7319.913	-8439.541	-9024.659	-7590.371	-7067.567	-8355.674	-8966.350	-7510.826	-7067.336	-8353.324	-8965.638	-7510.162	-7089.816	-8347.629	-8972.366	-7506.898	-7068.898	-8352.110	-8965.944	-7510.162	-7089.817	-8348.581	-8973.074	-7507.040	-7113.547	-8374.881	-8996.552	-7534.774

Table 2.5: Parameter estimates of the APARCH model for individual stocks under different probability distributions.

	N-APARCH			ST-APARCH			SKST-APARCH			SGED-APARCH			JSU-APARCH			SGT-APARCH			GHST-APARCH		
	IRS	GERMAN	US BOND	IRS	GERMAN	US BOND	IRS	GERMAN	US BOND	IRS	GERMAN	US BOND	IRS	GERMAN	US BOND	IRS	GERMAN	US BOND	IRS	GERMAN	US BOND
μ	0.00728	0.00918	0.00881	0.00931	0.01601	0.01579	0.00710	0.01069	0.00868	0.00577	0.00818	0.00767	0.00668	0.00998	0.00851	0.00695	0.01444	0.01542	0.00535	0.00940	0.00622
ϕ_1	-0.01559	0.04811	-0.02571	-0.01846	0.04596	-0.02239	-0.01937	0.04461	-0.02313	-0.02121	0.04161	-0.02002	-0.01958	0.04421	-0.02336	-0.01951	0.04007	-0.01627	-0.00979	0.04448	-0.02369
ω	0.00108	0.00122	0.00164	0.00090	0.00148	0.00207	0.00086	0.00155	0.00202	0.00110	0.00148	0.00184	0.00092	0.00158	0.00198	0.00110	0.00142	0.00192	0.00015	0.00162	0.00196
α_1	0.07354	0.03196	0.03324	0.05590	0.03433	0.03368	0.05510	0.03496	0.03336	0.06774	0.03527	0.03329	0.05723	0.03550	0.03310	0.06822	0.03446	0.03398	0.01298	0.03609	0.03237
γ_1	0.01042	0.02083	-0.06909	-0.00811	-0.03203	-0.12339	-0.00490	-0.02853	-0.11455	0.00508	0.00446	-0.08668	-0.00369	-0.02385	-0.11202	0.00216	-0.00286	-0.09320	1.00000	-0.02663	-0.11545
β_1	0.92579	0.95278	0.95893	0.95391	0.95880	0.96195	0.95444	0.95819	0.96206	0.93989	0.95389	0.96070	0.95210	0.95718	0.96207	0.93960	0.95506	0.96035	0.98595	0.95726	0.96222
δ	1.58070	2.33188	2.18176	1.02333	1.90899	1.90483	1.04306	1.88864	1.91893	1.25605	2.07945	2.03121	1.07272	1.90389	1.93816	1.25298	2.08664	2.01211	1.28134	1.87989	1.95515
ξ skewness							0.95454	0.93246	0.93951	0.98351	0.94622	0.96157									
ν/η kurtosis				5.62811	8.06565	7.01447	5.69630	8.25788	7.05796	1.25233	1.44811	1.36696									
γ skewness													-0.14007	-0.28885	-0.20925						
δ kurtosis													1.70711	2.14801	1.94649						
λ skewness																-0.04000	-0.06330	-0.04640			
p peakedness																1.76990	1.88900	1.67210			
q tail-thickness																4.16170	5.06920	7.10430			
β skewness																			-0.31304	-0.53511	-0.55720
ν kurtosis																			8.75182	8.19824	8.19704
Ljung-Box Test on Standardized Residuals Lag[5]																					
statistic	2.6808	2.7171	4.9978	3.6203	2.7965	4.9318	3.5380	2.9051	4.9457	2.9796	3.2213	4.9931	3.3815	2.9527	4.9431	2.9788	3.3469	5.1129	13.4100	2.9359	4.9446
p-value	0.5156	0.5057	0.1065	0.2925	0.4842	0.1122	0.3088	0.4555	0.1110	0.4365	0.3777	0.1069	0.3417	0.4433	0.1112	0.4366	0.3492	0.0972	0.0000	0.4476	0.1111
Ljung-Box Test on Standardized Squared Residuals Lag[9]																					
statistic	21.7300	32.3500	9.2740	126.8000	52.1300	11.2010	125.9000	51.9400	10.9940	51.4200	39.2200	10.2790	109.6000	49.8400	10.9310	51.0300	40.5200	10.5400	796.1000	50.8100	10.9750
p-value	0.0001	0.0000	0.0713	0.0000	0.0000	0.0274	0.0000	0.0000	0.0305	0.0000	0.0000	0.0437	0.0000	0.0000	0.0315	0.0000	0.0000	0.0384	0.0000	0.0000	0.0308
Log-Likelihoods																					
FGARCH	926.259	-2003.963	-3441.672	1103.253	-1937.591	-3359.877	1106.286	-1932.359	-3355.629	1084.776	-1938.523	-3356.930	1104.802	-1931.928	-3355.412	1084.181	-1942.813	-3359.233	1106.706	-1931.335	-3357.194
APARCH	924.602	-2006.190	-3442.768	1097.711	-1938.100	-3360.789	1100.172	-1932.823	-3356.336	1081.777	-1939.186	-3357.536	1099.238	-1932.407	-3356.103	1081.310	-1943.573	-3359.936	1036.118	-1931.799	-3358.235
GJRGARCH	919.917	-2007.073	-3443.090	1079.145	-1938.145	-3360.854	1081.789	-1932.892	-3356.384	1071.212	-1939.221	-3357.543	1082.006	-1932.458	-3356.131	1070.746	1943.614	-3359.937	1081.661	-1931.941	-3375.987
GARCH	919.886	-2007.235	-3444.361	1079.037	-1938.265	-3362.703	1081.688	-1932.988	-3358.003	1071.204	-1939.225	-3358.628	1081.922	-1932.528	-3357.709	1070.735	-1943.614	-3361.164	1072.947	-1931.939	-3364.000

Table 2.6: Parameter estimates of the APARCH model for interest rates under different probability distributions.

	N-APARCH				ST-APARCH				SKST-APARCH				SGED-APARCH			
	OIL	GAS	GOLD	SILVER	OIL	GAS	GOLD	SILVER	OIL	GAS	GOLD	SILVER	OIL	GAS	GOLD	SILVER
μ	0.00534	-0.01040	0.03610	0.00552	0.02705	-0.01040	0.04435	0.04440	0.00523	-0.01038	0.03480	0.01432	0.00118	0.00110	0.02619	-0.01000
ϕ_1	-0.00576	0.04516	-0.00458	-0.00559	0.00245	0.05621	-0.03070	-0.03864	0.00131	0.05539	-0.03153	-0.04065	0.00071	0.03889	-0.03653	-0.02486
ω	0.01006	0.08723	0.01549	0.01171	0.00810	0.12421	0.00622	0.00936	0.00740	0.12346	0.00606	0.00807	0.00799	0.11940	0.00762	0.00750
α_1	0.03776	0.13290	0.01549	0.04907	0.03701	0.12705	0.05057	0.04947	0.03695	0.12692	0.05017	0.04862	0.03804	0.12946	0.05114	0.04862
γ_1	0.20107	0.02672	-0.12890	-0.19197	0.29254	-0.04129	-0.38067	-0.27443	0.30731	-0.04205	-0.37435	-0.26535	0.26152	-0.00851	-0.28531	-0.23375
β_1	0.95919	0.88435	0.93681	0.95480	0.96424	0.88073	0.95861	0.95863	0.96503	0.88093	0.95885	0.95961	0.96275	0.88032	0.95402	0.95857
δ	0.95919	1.46520	1.84700	1.61294	1.68710	1.31254	1.16249	1.54822	1.65137	1.31113	1.15811	1.55272	1.76565	1.36941	1.33060	1.52007
ξ skewness									0.95114	1.01349	0.97337	0.94765	0.96882	0.99931	1.00536	0.98476
ν/η kurtosis					6.51509	5.80898	4.07112	3.93669	6.59810	5.82746	4.09096	3.94872	1.32548	1.23600	1.04792	0.99962
γ skewness																
δ kurtosis																
λ skewness																
p peakedness																
q tail-thickness																
β skewness																
ν kurtosis																
Ljung-Box Test on Standardized Residuals Lag[5]																
statistic	2.3145	37.0380	0.7856	0.6398	2.4215	37.5520	4.7270	5.7400	2.3990	37.6490	4.8870	6.0830	2.3344	39.3720	6.5510	2.4990
p-value	0.6198	0.0000	0.9737	0.9860	0.5889	0.0000	0.1316	0.0580	0.5953	0.0000	0.1162	0.0434	0.6140	0.0000	0.0288	0.5668
Ljung-Box Test on Standardized Squared Residuals Lag[9]																
statistic	11.6040	4.4670	29.8800	37.8400	20.9500	5.7750	71.4400	42.7100	21.6700	5.7870	71.4700	43.5200	17.2900	5.1653	57.6300	42.9300
p-value	0.0223	0.5116	0.0000	0.0000	0.0001	0.3249	0.0000	0.0000	0.0001	0.3234	0.0000	0.0000	0.0010	0.4055	0.0000	0.0000
Log-Likelihoods																
FGARCH	-8890.499	-11068.200	-6027.327	-8152.271	-8797.377	-10953.830	-5790.620	-7894.708	-8794.636	-10953.520	-5789.507	-7890.226	-8793.402	-10944.710	-5771.547	-7858.002
APARCH	-8896.963	-11074.470	-6043.597	-8157.958	-8798.970	-10956.900	-5795.265	-7895.974	-8796.022	-10956.640	-5794.283	-7891.792	-8795.537	-10948.340	-5776.379	-7859.400
GJRGARCH	-8896.970	-11080.690	-6043.961	-8159.962	-8799.577	-10963.820	-5805.213	-7899.261	-8796.797	-10963.530	-5804.435	-7904.482	-8795.799	-10953.840	-5781.740	-7861.838
GARCH	-8908.400	-11081.450	-6048.792	-8172.483	-8806.808	-10963.820	-5812.830	-7909.014	-8804.449	-10963.610	-5811.778	-7891.792	-8803.374	-10953.860	-5787.113	-7869.136

	JSU-APARCH				SGT-APARCH				GHST-APARCH			
	OIL	GAS	GOLD	SILVER	OIL	GAS	GOLD	SILVER	OIL	GAS	GOLD	SILVER
μ	0.00311	-0.01040	0.02986	0.00430	0.02290	0.00297	0.02368	0.00622	0.06650	-0.01040	0.01607	-0.01776
ϕ_1	0.00102	0.05513	-0.03478	-0.04145	0.00375	0.03880	-0.03717	-0.01516	0.00376	0.06142	-0.02707	-0.03811
ω	0.00752	0.12163	0.00619	0.00780	0.00832	0.11927	0.00758	0.00807	0.01035	0.12103	0.00688	0.00705
α_1	0.03699	0.12662	0.04964	0.04749	0.03799	0.12940	0.05102	0.04923	0.03812	0.13183	0.04422	0.04202
γ_1	0.30224	-0.03921	-0.35909	-0.25787	0.24608	-0.00871	-0.28296	-0.23586	0.24812	-0.00417	-0.33332	-0.24331
β_1	0.96475	0.88109	0.95828	0.95989	0.96199	0.88038	0.95409	0.95808	0.96195	0.88040	0.95811	0.96091
δ	1.66451	1.31143	1.17609	1.52168	1.82435	1.36917	1.32880	1.52749	1.76914	1.37259	1.28408	1.61472
ξ skewness												
ν/η kurtosis												
γ skewness	-0.17648	0.06274	-0.07894	-0.14137								
δ kurtosis	1.87374	1.73191	1.37965	1.34976								
λ skewness					-0.03460	0.00010	-0.00460	-0.00320				
p peakedness					1.59440	1.35470	1.19960	0.99130				
q tail-thickness					7.92790	18.88350	13.51120	64003.76400				
β skewness									0.72538	-1.08763	-0.89820	-1.19491
ν kurtosis									8.19685	8.19642	8.19704	8.19607
Ljung-Box Test on Standardized Residuals Lag[5]												
statistic	2.3874	37.6310	5.7760	6.2240	2.3903	39.3900	6.7310	1.2348	2.4590	36.2946	3.7870	5.4480
p-value	0.5987	0.0000	0.0563	0.0384	0.5979	0.0000	0.0246	0.9058	0.5781	0.0000	0.2615	0.0740
Ljung-Box Test on Standardized Squared Residuals Lag[9]												
statistic	21.1200	5.7530	68.9000	44.2300	16.4300	5.1738	57.5800	42.6400	18.2500	5.0631	64.2600	43.7000
p-value	0.0001	0.3276	0.0000	0.0000	0.0016	0.4043	0.0000	0.0000	0.0006	0.4202	0.0000	0.0000
Log-Likelihoods												
FGARCH	-8793.578	-10952.000	-5784.585	-7883.179	-8794.686	-10944.720	-5771.568	-7858.712	-8801.163	-11008.400	-5830.550	-7930.550
APARCH	-8795.110	-10955.330	-5789.511	-7884.831	-8797.112	-10948.340	-5776.426	-7859.984	-8832.886	-11014.960	-5836.129	-7933.716
GJRGARCH	-8795.773	-10962.160	-5798.928	-7888.069	-8797.276	-10953.850	5781.842	-7862.298	-8809.132	-11038.850	-5851.344	-7928.186
GARCH	-8803.463	-10962.220	-5805.687	-7896.624	-8804.413	-10953.870	-5787.142	-7869.240	-8830.847	-11000.680	-5853.270	-7946.992

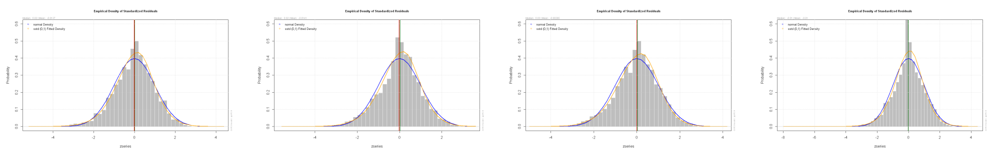
Table 2.7: Parameter estimates of the APARCH model for commodities under different probability distributions.

	N-APARCH				ST-APARCH				SKST-APARCH				SGED-APARCH			
	EUR/USD	GBP/USD	JPY/USD	AUD/USD	EUR/USD	GBP/USD	JPY/USD	AUD/USD	EUR/USD	GBP/USD	JPY/USD	AUD/USD	EUR/USD	GBP/USD	JPY/USD	AUD/USD
μ	0.00485	0.00055	-0.00710	0.00028	0.00528	0.00402	-0.01276	0.01455	0.00541	0.00071	-0.00758	0.00286	0.00620	0.00045	-0.00608	-0.00069
ϕ_1	0.00005	0.01289	-0.02004	0.01178	-0.00414	0.01185	-0.03709	0.00604	-0.00412	0.01037	-0.03708	0.00060	-0.00720	0.00999	-0.04123	0.00020
ω	0.00110	0.00184	0.00874	0.00692	0.00099	0.00122	0.00541	0.00549	0.00099	0.00122	0.00559	0.00511	0.00104	0.00139	0.00685	0.00567
α_1	0.03052	0.03254	0.05525	0.05571	0.03253	0.03140	0.04635	0.05027	0.03254	0.03129	0.04664	0.04844	0.03145	0.03202	0.05023	0.05048
γ_1	0.13145	0.21562	-0.17963	0.27995	0.10438	0.18375	-0.06856	0.28024	0.10436	0.18566	-0.07087	0.27574	0.11615	0.19242	-0.11524	0.27546
β_1	0.97024	0.95711	0.94116	0.94203	0.96902	0.95983	0.95113	0.94910	0.96900	0.96027	0.95040	0.95085	0.96985	0.95843	0.94630	0.94801
δ	1.71404	2.12363	1.26742	1.46472	1.71932	2.20836	1.50093	1.41978	1.71989	2.19741	1.50774	1.46261	1.69878	2.20529	1.39941	1.47868
ξ skewness									1.00110	0.96782	1.03474	0.90846	1.00731	0.98164	1.03042	0.90699
ν/η kurtosis					8.61483	9.35785	5.68722	8.13601	8.61507	9.34700	5.72134	8.57177	1.43454	1.41824	1.28043	1.47874
γ skewness																
δ kurtosis																
λ skewness																
p peakedness																
q tail-thickness																
β skewness																
ν kurtosis																
Ljung-Box Test on Standardized Residuals Lag[5]																
statistic	1.4585	0.7123	0.8316	1.9019	1.8611	0.7714	2.2510	2.2613	1.8592	0.8394	2.2380	2.8180	2.2200	0.8442	2.9750	2.8550
p-value	0.8567	0.9805	0.9688	0.7394	0.7509	0.9751	0.6382	0.6353	0.7515	0.9679	0.6421	0.4783	0.6473	0.9674	0.4375	0.4687
Ljung-Box Test on Standardized Squared Residuals Lag[9]																
statistic	12.1700	2.1377	8.6750	0.7394	12.9600	1.7830	8.3070	1.2866	12.9600	1.8380	8.1600	1.4780	12.3900	2.0462	8.3140	1.3265
p-value	0.0166	0.8879	0.0947	0.9624	0.0109	0.9292	0.1122	0.9712	0.0109	0.9234	0.1199	0.9573	0.0148	0.8994	0.1118	0.9685
Log-Likelihoods																
FGARCH	-3726.654	-3212.542	-3821.917	-4486.170	-3684.254	-3179.633	-3685.445	-4420.022	-3684.249	-3178.347	-3684.105	-4410.368	-3673.048	-3160.342	-3693.424	-4415.481
APARCH	-3726.955	-3213.268	-3825.044	-4486.320	-3685.922	-3180.539	-3685.445	-4420.023	-3685.921	-3179.208	-3684.106	-4410.379	-3673.874	-3161.226	-3693.649	-4415.498
GJRGARCH	-3727.443	-3213.355	-3832.559	-4489.558	-3686.289	-3180.716	-3687.313	-4422.863	-3686.286	-3179.371	-3685.914	-4412.971	-3674.282	-3161.387	-3696.432	-4417.949
GARCH	-3730.746	-3222.614	-3835.508	-4500.710	-3688.007	-3186.048	-3687.822	-4430.385	-3688.003	-3184.822	-3686.464	-4420.525	-3676.230	-3166.980	-3697.513	-4425.803

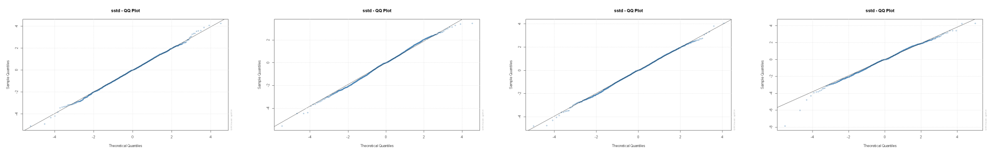
	JSU-APARCH				SGT-APARCH				GHST-APARCH			
	EUR/USD	GBP/USD	JPY/USD	AUD/USD	EUR/USD	GBP/USD	JPY/USD	AUD/USD	EUR/USD	GBP/USD	JPY/USD	AUD/USD
μ	0.00566	0.00131	-0.00679	0.00122	0.00450	0.00410	-0.01252	0.01334	0.01865	-0.00391	-0.02296	0.00048
ϕ_1	-0.00458	0.01035	-0.03799	-0.00009	-0.00747	0.01036	-0.03970	0.00716	-0.00222	0.00825	-0.03219	-0.00011
ω	0.00099	0.00121	0.00574	0.00502	0.00103	0.00138	-0.03970	0.00618	0.00120	0.00116	0.00534	0.00495
α_1	0.03237	0.03131	0.04699	0.04780	0.03140	0.03216	0.05004	0.05350	0.03435	0.03149	0.04657	0.04731
γ_1	0.10627	0.18422	-0.07697	0.27712	0.11724	0.18905	-0.10925	0.27483	0.09224	0.18447	-0.07449	0.28330
β_1	0.96916	0.96007	0.94980	0.95140	0.96997	0.95808	0.94700	0.94511	0.96727	0.96096	0.95138	0.95252
δ	1.71488	2.21070	1.49499	1.47118	1.69446	2.21266	1.39307	1.43540	1.76964	2.22515	1.42114	1.46576
ξ skewness												
ν/η kurtosis												
γ skewness	0.00812	-0.12791	0.10432	-0.42106								
δ kurtosis	2.15574	2.24144	1.71643	2.21822								
λ skewness					0.00570	-0.0151	0.02710	-0.08700				
p peakedness					1.43470	1.4178	1.75190	1.84960				
q tail-thickness					37075.337	48971.258	4.30350	5.78010				
β skewness									0.74763	-0.53192	-0.60067	-0.70704
ν kurtosis									8.19612	8.19825	8.19802	8.19767
Ljung-Box Test on Standardized Residuals Lag[5]												
statistic	1.9078	0.8394	2.3800	2.9050	2.2630	0.8173	2.7140	2.1721	1.6360	0.97570	1.6581	2.9150
p-value	0.7377	0.9679	0.6009	0.4556	0.6347	0.9704	0.5065	0.6614	0.8122	0.95040	0.8064	0.4530
Ljung-Box Test on Standardized Squared Residuals Lag[9]												
statistic	12.8600	1.8401	8.1520	1.5430	12.4300	1.9763	8.4190	1.1156	12.8100	1.91290	8.7390	1.7123
p-value	0.0115	0.9232	0.1204	0.9520	0.0145	0.9078	0.1066	0.9810	0.0118	0.91520	0.0919	0.9364
Log-Likelihoods												
FGARCH	-3682.975	-3177.066	-3684.075	-4408.619	-3673.165	-3160.890	-3694.745	-4427.420	-3685.677	-3181.968	-3711.697	-4407.483
APARCH	-3684.512	-3177.975	-3684.077	-4408.636	-3673.976	-3161.815	-3694.975	-4427.423	-3699.408	-3183.140	-3713.161	-4407.483
GJRGARCH	-3684.885	-3183.478	-3685.969	-4411.185	-3674.397	-3161.987	-3697.855	-4430.104	-3699.702	-3184.392	-3720.262	-4410.242
GARCH	-3686.625	-3178.151	-3686.595	-4418.804	-3676.380	-3167.394	-3698.814	-4437.724	-3710.757	-3194.772	-3709.002	-4420.237

Table 2.8: Parameter estimates of the APARCH model for exchange rates under different probability distributions.

Figure 2.2 displays, for each stock market index, histograms and QQ-plots against theoretical quantiles for estimated standardized residuals ($\hat{z}_t$) of the SKST-APARCH model. We can observe that standardized innovations show, indeed, fat tails and negative skewness.



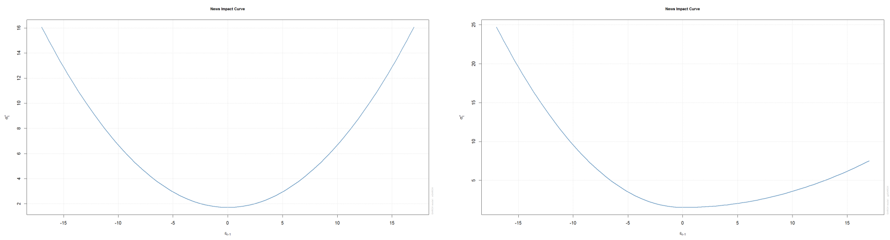
(a) IBEX 35 (b) NASDAQ 100 (c) FTSE 100 (d) NIKKEI 225



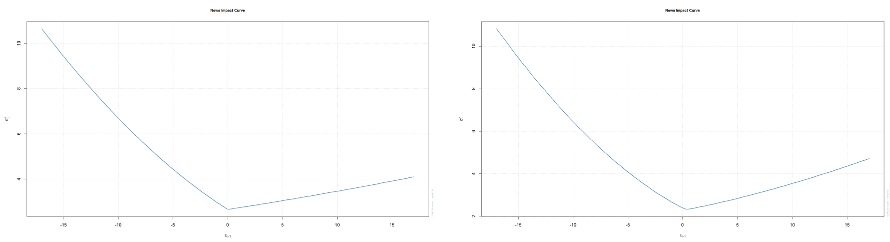
(e) IBEX 35 (f) NASDAQ 100 (g) FTSE 100 (h) NIKKEI 225

Figure 2.2: Histograms and QQ-plots of standardized innovations from SKST-APARCH model for stock market indices against the skewed Student-t distribution.

Figure 2.3 displays the news impact curves of different volatility models for IBM. We can observe that GARCH and GJRGARCH models are based on the variance equation, while APARCH and FGARCH models introduce the Box-Cox transformation in the conditional standard deviation, and the free parameter (δ in APARCH and λ in FGARCH) determines the shape of the transformation. For IBM the value of this parameter is $\delta = 1.01$ and $\lambda = 1.10$ for the APARCH and FGARCH model, respectively. This parameters are significantly different from zero and two, but not from one. Furthermore, FGARCH model permit not only rotations, like APARCH model, but also shifts of the news impact curve. As can be seen from Figure 2.3 d), the asymmetry caused by the shift $\eta_2 = 0.20$ is most pronounced for small shocks. For extremely large shocks, the asymmetric effect becomes a negligible part of the total response. On the other hand, the rotated news impact curve of Figure 2.3 c), $\gamma = 0.61$ maintains the hypothesis that a zero shock results in the smallest increase of conditional variance. Additionally, the size of the asymmetric effect of small shocks is very small in absolute terms. The estimates of γ in APARCH model imply that negative shocks result in higher volatility than equally large positive shocks, which is in accordance with the "leverage effect". In Figure 2.3 d) the shift $\eta_2 = 0.20$ and rotation $\eta_1 = 0.42$ are combined in one news impact curve. Both parameters are significant. By appropriately shifting and rotating the news impact curve, it is possible to have asymmetry for small shocks, a roughly symmetric response for moderate shocks, and asymmetry for very large shocks.



(a) GHST-GARCH (b) GHST-GJRGARCH



(c) GHST-APARCH (d) GHST-FGARCH

Figure 2.3: News impact curves of different volatility model for IBM.

2.6 Fitting the data

VaR models are usually evaluated according to the performance of their VaR estimates using appropriate testing procedures. However, the ability of a VaR model to reproduce the main characteristics of return data is hardly ever examined. A possible justification for such inattention is the argument that good VaR estimates have to do just with the quality of the fit to the tails of the distribution of returns. A good overall fit might not be all that interesting because it might be obtained at the expense of not fitting so well the distribution tails. However, the fit to the tail of return distribution is usually not examined either. The fact is that it is unclear whether a good overall fit of the return distribution helps to produce good VaR estimates or whether it should be enough to care about the fit to the tail of the distribution, and we want to throw some light into that question. In particular, if fitting the tail distribution is what matters, that might explain why the type of models considered in extreme value theory tend to beat other alternatives in VaR estimation.

We examine in this section the extent to which each model fits the return data, and we will later check whether the models with a better overall fit lead to better VaR estimates. We start by checking the extent to which each model fits the likelihood of return data. After that, we examine the ability of each model to fit the main sample moments of returns. To evaluate the fit to the distribution of returns Monte Carlo simulation is needed, as explained below.

2.6.1 Likelihood ratio tests

The lower panels of the Tables 2.4 - 2.8 show that models with FGARCH volatility, combined with JSU and SGED distributions for stock market indices, with SKST and SGED distributions for individual stocks, with JSU and GHST distributions for interest rates, with SGED for commodities and with SGED and JSU for exchange rates, often achieve the highest log-likelihood. Likelihood ratio tests in Table 2.9 show a superiority of the FGARCH specification over the APARCH, GJR-GARCH and the symmetric GARCH specifications for stock market indices. In all comparisons in the table, the more restricted model appears to the left. At 5% significance, the test clearly favors the APARCH model against the GJRGARCH model and the FGARCH model against the APARCH. Indeed, for stock market indices, individual stocks and commodities the FGARCH model is preferred to the APARCH model whereas for interest rates and exchange rates the APARCH model is preferred. Overall, FGARCH and APARCH are the best models according to this criterion.

Test statistic	IBEX	NASDAQ	FTSE	NIKKEI
N-GARCH vs N-APARCH	205.186	138.348	195.112	78.296
N-GJRGARCH vs N-APARCH	30.148	11.934	15.908	9.164
N-APARCH vs N-FGARCH	21.816	52.104	37.834	47.568
ST-GARCH vs ST-APARCH	159.688	119.084	248.592	79.222
ST-GJRGARCH vs ST-APARCH	25.748	15.412	18.558	17.354
ST-APARCH vs ST-FGARCH	8.762	27.646	32.422	44.818
SKST-GARCH vs SKST-APARCH	167.376	134.806	186.022	77.49
SKST-GJRGARCH vs SKST-APARCH	26.388	18.518	19.916	17.154
SKST-APARCH vs SKST-FGARCH	11.064	38.588	33.902	46.58
SGED-GARCH vs SGED-APARCH	163.090	123.216	137.098	66.682
SGED-GJRGARCH vs SGED-APARCH	24.716	15.794	**-19.958**	13.778
SGED-APARCH vs SGED-FGARCH	12.574	19.094	30.93	40.332
JSU-GARCH vs JSU-APARCH	166.902	135.970	184.646	74.574
JSU-GJRGARCH vs JSU-APARCH	25.992	19.584	19.006	16.460
JSU-APARCH vs JSU-FGARCH	**1.216**	14.958	33.778	47.004
SGT-GARCH vs SGT-APARCH	154.116	108.794	157.816	66.148
SGT-GJRGARCH vs SGT-APARCH	24.028	12.516	15.348	13.442
SGT-APARCH vs SGT-FGARCH	10.89	27.402	30.93	38.618
GHST-GARCH vs GHST-APARCH	168.844	148.648	180.538	71.766
GHST-GJRGARCH vs GHST-APARCH	34.134	60.512	19.006	16.460
GHST-APARCH vs GHST-FGARCH	**-6.826**	13.926	20.554	57.444

Table 2.9: Likelihood ratio tests of volatility specifications for stock market indices. Note: The null hypothesis is rejected, except where indicated by boldface.

2.6.2 Fitting standardized innovations

2.6.2.1 Fitting the empirical distribution of return innovations

Table 2.10 reports the results obtained when comparing the empirical distribution of estimated innovations to the theoretical distribution used in estimation for the four stock market indices [4]. We use the Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933, Smirnov, 1939 and Massey, 1951) [76] [116] [89], which quantifies the distance between the empirical distribution function of standardized innovation and the cumulative distribution function of the reference distribution, and the Chi-square (Chi2) test (Pearson, 1900) [102] applied to a partition of the return data range into 10 bins [5]. The null distribution of these statistics is calculated under the null hypothesis that the sample is drawn from the reference distribution. These tests suggest that models with an asymmetric distribution for the innovations are to be preferred. Test statistics also tend to be smaller for the APARCH and FGARCH volatility specifications.

	IBEX35		NASDAQ100		FTSE100		NIKKEI225	
	KS	Chi2	KS	Chi2	KS	Chi2	KS	Chi2
N-GARCH	0.039 (0.000)	243110 (0.000)	0.043 (0.000)	86.329 (0.000)	0.032 (0.000)	107.345 (0.000)	0.051 (0.000)	243267 (0.000)
ST-GARCH	0.022 (0.038)	21.348 (0.011)	0.028 (0.002)	17.048 (0.048)	0.020 (0.083)	37.191 (0.000)	0.039 (0.000)	21.667 (0.010)
SKST-GARCH	0.027 (0.005)	9.892 (0.359)	0.030 (0.001)	7.344 (0.601)	0.031 (0.001)	15.078 (0.089)	0.038 (0.000)	13.744 (0.132)
SGED-GARCH	0.022 (0.035)	50.690 (0.000)	0.023 (0.027)	5.633 (0.776)	0.027 (0.005)	17.152 (0.046)	0.026 (0.006)	39.174 (0.000)
JSU-GARCH	0.026 (0.006)	10.010 (0.349)	0.030 (0.001)	6.336 (0.706)	0.031 (0.001)	13.381 (0.146)	0.041 (0.000)	14.011 (0.122)
SGT-GARCH	0.029 (0.001)	23.392 (0.005)	0.028 (0.003)	6.001 (0.740)	0.034 (0.000)	17.480 (0.042)	0.028 (0.003)	39.408 (0.000)
GHST-GARCH	0.021 (0.056)	8.799 (0.456)	0.028 (0.003)	7.665 (0.568)	0.019 (0.099)	22.349 (0.008)	0.037 (0.000)	10.983 (0.277)
N-GJRGARCH	0.034 (0.000)	228.860 (0.000)	0.047 (0.000)	39.537 (0.000)	0.039 (0.000)	117.613 (0.000)	0.048 (0.000)	971626 (0.000)
ST-GJRGARCH	0.024 (0.020)	49.752 (0.000)	0.031 (0.001)	40.861 (0.000)	0.029 (0.002)	57.236 (0.000)	0.038 (0.000)	73.794 (0.000)
SKST-GJRGARCH	0.018 (0.129)	14.610 (0.102)	0.022 (0.041)	12.553 (0.184)	0.016 (0.222)	8.206 (0.514)	0.036 (0.000)	43.630 (0.000)
SGED-GJRGARCH	0.015 (0.338)	21.714 (0.010)	0.017 (0.166)	17.948 (0.036)	0.016 (0.262)	9.942 (0.355)	0.030 (0.001)	127.937 (0.000)
JSU-GJRGARCH	0.017 (0.155)	14.262 (0.113)	0.020 (0.072)	12.947 (0.165)	0.016 (0.248)	5.923 (0.748)	0.039 (0.000)	38.568 (0.000)
SGT-GJRGARCH	0.021 (0.046)	20.689 (0.014)	0.025 (0.012)	20.488 (0.015)	0.026 (0.008)	11.848 (0.222)	0.029 (0.002)	133.807 (0.003)
GHST-GJRGARCH	0.027 (0.004)	18.724 (0.028)	0.033 (0.000)	12.947 (0.165)	0.028 (0.003)	28.488 (0.001)	0.035 (0.000)	24.784 (0.000)
N-APARCH	0.036 (0.000)	248.980 (0.000)	0.047 (0.000)	141.086 (0.000)	0.042 (0.000)	111.868 (0.000)	0.048 (0.000)	243023 (0.000)
ST-APARCH	0.026 (0.006)	48.544 (0.000)	0.030 (0.001)	29.656 (0.001)	0.031 (0.001)	44.470 (0.000)	0.038 (0.000)	47.912 (0.000)
SKST-APARCH	0.019 (0.093)	15.015 (0.091)	0.020 (0.062)	2.589 (0.978)	0.019 (0.110)	3.208 (0.956)	0.037 (0.000)	20.405 (0.016)
SGED-APARCH	0.019 (0.114)	24.748 (0.003)	0.017 (0.162)	3.676 (0.931)	0.015 (0.275)	5.960 (0.744)	0.031 (0.001)	54.437 (0.000)
JSU-APARCH	0.020 (0.081)	16.025 (0.066)	0.020 (0.064)	1.759 (0.995)	0.018 (0.120)	3.576 (0.937)	0.038 (0.000)	15.731 (0.073)
SGT-APARCH	0.019 (0.094)	21.066 (0.012)	0.026 (0.008)	5.237 (0.813)	0.025 (0.013)	6.753 (0.663)	0.029 (0.002)	60.521 (0.000)
GHST-APARCH	0.030 (0.001)	22.105 (0.009)	0.031 (0.001)	8.823 (0.454)	0.030 (0.001)	21.973 (0.009)	0.037 (0.000)	20.293 (0.016)
N-FGARCH	0.037 (0.000)	132.260 (0.000)	0.047 (0.000)	97.099 (0.000)	0.042 (0.000)	106.330 (0.000)	0.051 (0.000)	971730 (0.000)
ST-FGARCH	0.027 (0.004)	14.633 (0.102)	0.025 (0.011)	32.188 (0.000)	0.033 (0.000)	44.745 (0.000)	0.037 (0.000)	60.110 (0.000)
SKST-FGARCH	0.019 (0.102)	4.929 (0.840)	0.022 (0.035)	15.5601(0.077)	0.019 (0.088)	2.967 (0.966)	0.034 (0.000)	27.787 (0.001)
SGED-FGARCH	0.018 (0.142)	10.078 (0.344)	0.023 (0.029)	14.149 (0.117)	0.017 (0.158)	2.206 (0.988)	0.032 (0.000)	110.135 (0.000)
JSU-FGARCH	0.019 (0.097)	4.467(0.878)	0.026 (0.007)	12.654 (0.179)	0.020 (0.082)	2.393 (0.984)	0.035 (0.000)	22.448 (0.008)
SGT-FGARCH	0.018 (0.123)	7.566 (0.578)	0.028 (0.002)	14.121 (0.118)	0.023 (0.022)	3.179 (0.957)	0.027 (0.005)	121.760 (0.000)
GHST-FGARCH	0.026 (0.008)	15.856 (0.070)	0.032 (0.000)	33.125 (0.000)	0.032 (0.000)	25.915 (0.002)	0.037 (0.000)	25.226 (0.003)

Table 2.10: Goodness-of-fit tests for standardized innovations of stock market indices. Figures in brackets denote p-value.

According to the KS test, models with N distributions fits the data well 11 out of 76 cases (4 volatility models by 19 assets), ST fits the data well in 53 cases, SKST in 54, SGED in 59, JSU in 47, SGT in 62 and GHST in 47 cases. Regarding volatility models, distributions with GARCH model fit the data well 79 out of 133 cases (7 probability distributions by 19 assets), GJRGARCH and APARCH fit the data well in 85 cases and FGARCH in 84 cases. According to the Chi2 test, models with N distributions fits the data well 1 out of 76 cases, ST fits the data well in 20 cases, SKST and SGED in 32, JSU

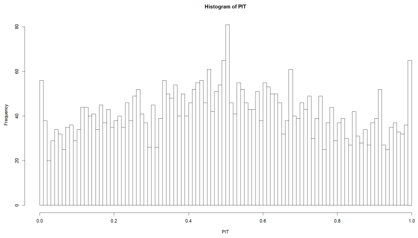
[4] Results for another assets are available on request.

[5] The number of bins affects the results of the Pearson test.

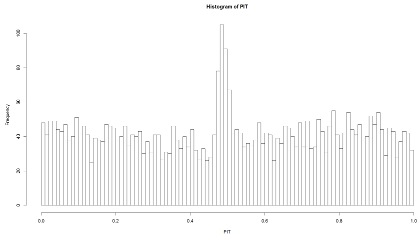
and SGT in 30 and GHST in 18 cases. Respect to volatility models, distributions with GARCH, APARCH and FGARCH models fit the data well 40 out of 133 cases and GJR-GARCH in 43 cases. To sum up, the SGED and SGT are preferred to fit the innovations and GJRGARCH and APARCH to model the volatility.

To compare the adequacy of the different distributions we can also employ out-of-sample density forecasts, as proposed by Diebold, Gunther and Tay (1998) [34] (DGT). Let $f_i(y_i|\Omega_i)_{i=1}^m$ be a sequence of m one-step-ahead density forecasts produce by a given model, where Ω_i is the conditioning information set, and $p_i(y_i|\Omega_i)_{i=1}^m$ the sequence of densities defining the Data Generating Process y_i (which is never observed). The null hypothesis is $H_0 : f_i(y_i|\Omega_i)_{i=1}^m = p_i(y_i|\Omega_i)_{i=1}^m$. DGT use the fact that under the null hypothesis, the probability integral transform $\zeta_i = \int_{-\infty}^{y_i} f_i(t)dt$ is i.i.d. with a Uniform(0,1) distribution. To check H_0, they propose to use independence test for i.i.d. U(0,1). The i.i.d.-ness property of ζ_i can be evaluated by plotting an histogram of ζ_i. A humped shape of the ζ-histogram would indicate that the issued forecasts are too narrow and that the tails of the true density are not accounted for. On the other hand, a U-shape of the histogram would suggest that the model issues forecasts that either under- or overestimate too frequently [Bauwens, Giot, Grammig and Veredas (2000) [16]].

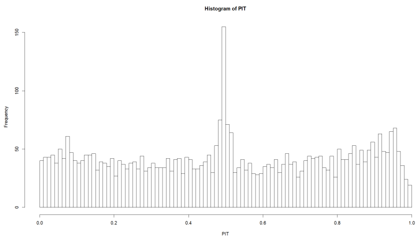
Figures 2.4 a) and 2.4 b) show a sample of such histograms for the assets in our data set. The humped shape of the histograms shows that symmetrical distributions are not suitable to model the OIL and US BOND 10Y returns. Figures 2.4 c) and 2.4 d) show that the Skewed Generalized Error distribution is not suitable for NIKKEI 225. It is appropriate for JPY/USD because its probability integral transform is Uniformly distributed. In 4 e) the Johnson S_U distribution is also appropriate for AUD/USD. Figure 2.4 f) shows that the assumption of a Generalized Hyperbolic Skew Student-t for the innovation is not appropriate for SAN. These results are consistent with the goodness-of-fit tests previously carried out. For the rest of assets, the results are similar, the symmetrical distributions and the Generalized Hyperbolic Skew Student-t for the innovations are not appropriate for most of the assets whereas Skewed Student-t, Skewed Generalized Error and Johnson S_U are suitable.



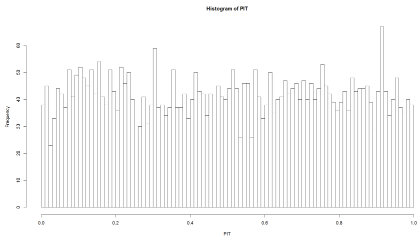
(a) OIL - N



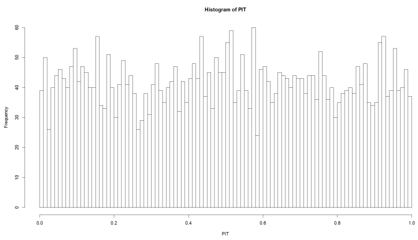
(b) US BOND 10Y - ST



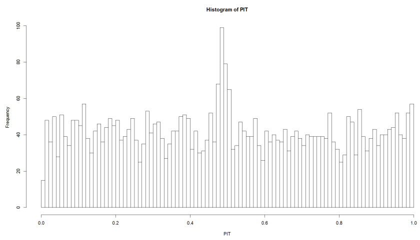
(c) NIKKEI 225 - SGED



(d) JPY/USD - SGED



(e) AUD/USD - JSU



(f) SAN - GHST

Figure 2.4: ζ-histograms (100 cells) for 4173 one-step-ahead forecasts. We assume different distributions with AR(1)-FGARCH(1,1) model for different assets.

2.6.2.2 Fitting the sample moments of return innovations

For a given asset, the innovations change with the estimated model, so we compare the theoretical moments of a given probability distribution with the sample moments for the standardized innovations for that model. In fact, however, sample moments for innovations are similar across models, showing a near zero mean and a unit variance in all models, as expected. But that is also the case under all the estimated probability distributions, so it makes sense to focus on the ability of each estimated distribution to fit the sample skewness and kurtosis of standardized innovations. Tables 2.11 - 2.15 compare the theoretical value of skewness and kurtosis from the estimated probability distribution with the similar sample moments of the standardized innovations calculating the absolute differences between these both values. Obviously, the Normal and the symmetric Student distribution do not produce any skewness. This is a limitation of these distributions since skewness and kurtosis are present in standardized innovations. For most assets, the skewed t-Student distribution produces negative skewness, although not as much as it is observed in the data. The unbounded Johnson distribution achieves a higher level of negative skewness, often being close to that observed in the data. The GHST distribution does not fit innovation moments very well, especially overestimating the degree of negative skewness. Indeed, the GHST distribution usually produces the maximum absolute difference between the theoretical and the sample skewness in most stock market indices, individual stocks and exchange rates. The GHST distribution has been proposed as being suitable for assets with high skewness and heavy-tailed (Aas & Haff, 2006 [1]) and the assets we consider do not have high skewness. In fact, only the standardized innovations in SILVER and AUD/USD have a negative high skewness and in these two cases models with GHST produce the best fit to sample skewness. Additionally, asymmetric probability distributions are unable to reproduce the positive skewness shown by a few return innovations, such as those in IRS5Y and GAS.

On the other hand, the GHST distribution can explain the high kurtosis often observed in our standardized innovations, except when used with a GJRGARCH volatility for stock market indices or when used with APARCH and FGARCH specifications for IRS5Y. The symmetric and the skewed Student-t distributions explain the level of kurtosis observed in the data [6], while the Johnson distribution generally implies higher kurtosis than it is observed in the data [7].

In fact, for skewness the results are concentrated in the SKST distribution, it fits skewness best in 8 of the 19 cases. For kurtosis results are not so concentrated: ST (for 5 assets), SKST (4), SGT (6) y SGED (4) fit kurtosis best. Pulling together the fit of both moments, the SKST distribution performs best in 12 out of the 38 cases, followed by

[6] The theoretical kurtosis for the Student-t distribution has been calculated as $K = \frac{6}{\nu - 4} + 3$ (The same result is obtained by replacing $\xi = 1$ and ν in (2)-(3); see Appendices). For GOLD and SILVER the Student-t distribution for some volatility models produces negative kurtosis because we have obtained in the estimation a number of degrees of freedom (ν) less than 4.

[7] For GOLD and SILVER, as well as for IBM and BP, the unbounded Johnson distribution produces extremely high kurtosis because the estimated kurtosis parameter (δ) of the Johnson distribution is close to 1. As $\delta \to \infty$ the distribution approaches the Normal density function and we obtain a kurtosis = 3.

SGT and SGED with 7 cases. The FGARCH specification fits skewness best in 7 assets, while the GARCH specification fits kurtosis best in 8 assets. Overall, the SGT and SKST distributions with GARCH, GJR-GARCH and FGARCH do better in capturing skewness and kurtosis of the standardized innovations than other combinations.

	IBEX35		**NASDAQ100**		**FTSE100**		**NIKKEI225**	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
N-GARCH	0.263	1.327	0.235	0.900	0.318	0.748	0.377	1.334
ST-GARCH	0.284	0.149	0.245	0.651	0.317	0.488	0.398	0.278
SKST-GARCH	0.079	0.045	0.048	0.555	0.065	0.313	0.197	0.127
SGED-GARCH	0.104	0.420	0.101	0.206	0.102	0.063	0.341	**0.074**
JSU-GARCH	0.070	2.529	0.145	3.869	0.062	1.020	0.076	3.273
SGT-GARCH	0.114	0.254	0.123	0.222	0.118	0.082	0.351	0.105
GHST-GARCH	0.345	1.303	0.416	2.406	0.210	0.333	0.249	1.558
N-GJRGARCH	0.260	0.952	0.275	0.734	0.332	0.633	0.366	1.499
ST-GJRGARCH	0.269	0.159	0.288	0.476	0.332	0.196	0.384	0.267
SKST-GJRGARCH	0.032	0.119	0.030	0.421	**0.047**	0.115	0.198	0.339
SGED-GJRGARCH	0.055	0.187	0.070	0.188	0.066	0.011	0.316	0.268
JSU-GJRGARCH	0.025	0.952	0.069	2.094	0.171	0.165	0.032	1.718
SGT-GJRGARCH	0.063	0.084	0.096	0.203	0.081	0.010	0.336	0.264
GHST-GJRGARCH	0.501	4.861	0.843	18.402	0.304	2.140	0.731	16.204
N-APARCH	0.251	0.916	0.307	0.773	0.338	0.685	0.365	1.577
ST-APARCH	0.258	0.159	0.332	0.420	0.346	0.085	0.395	0.451
SKST-APARCH	0.019	0.130	0.061	0.323	0.052	0.033	0.206	0.511
SGED-APARCH	0.043	0.172	0.094	**0.108**	0.067	0.088	0.313	0.461
JSU-APARCH	0.027	0.852	**0.023**	1.778	0.176	0.097	**0.001**	2.072
SGT-APARCH	0.049	0.058	0.118	0.127	0.083	0.089	0.343	0.458
GHST-APARCH	0.391	2.119	0.334	2.779	0.199	0.498	0.277	1.444
N-FGARCH	0.232	0.763	0.299	0.677	0.338	0.649	0.385	1.632
ST-FGARCH	0.250	0.197	0.297	0.204	0.349	0.038	0.415	0.513
SKST-FGARCH	**0.003**	0.184	0.053	0.352	0.055	**0.002**	0.216	0.596
SGED-FGARCH	0.027	0.082	0.122	0.131	0.065	0.099	0.318	0.588
JSU-FGARCH	0.035	0.745	0.185	0.140	0.201	0.021	0.051	1.326
SGT-FGARCH	0.035	**0.011**	0.106	0.168	0.080	0.096	0.353	0.582
GHST-FGARCH	0.291	2.414	0.313	2.303	0.080	1.123	0.218	0.818

Table 2.11: Absolute differences between standardized innovation moments and theoretical moments for stock market indices.

	IBM		SAN		AXA		BP	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
N-GARCH	0.339	4.930	0.218	1.784	0.147	1.323	0.097	1.718
ST-GARCH	0.510	4.606	0.225	0.196	0.152	0.066	0.084	0.032
SKST-GARCH	0.440	4.448	0.092	0.120	0.085	0.093	0.014	0.021
SGED-GARCH	0.455	3.691	0.157	0.642	0.094	0.554	0.070	0.664
JSU-GARCH	0.916	149.379	0.171	6.433	0.056	1.217	0.113	5.795
SGT-GARCH	0.384	**0.033**	0.150	0.477	0.087	0.117	0.057	0.485
GHST-GARCH	0.121	3.295	0.543	4.010	0.678	6.267	0.596	2.078
N-GJRGARCH	0.286	4.536	0.221	1.365	0.109	1.125	0.080	1.536
ST-GJRGARCH	0.416	3.042	0.222	0.170	0.111	0.082	0.063	0.025
SKST-GJRGARCH	0.335	2.956	0.062	0.135	**0.004**	0.098	0.032	**0.017**
SGED-GJRGARCH	0.327	3.035	1.113	0.356	0.024	0.455	0.022	0.560
JSU-GJRGARCH	0.897	110.441	0.084	2.988	0.008	0.507	0.139	3.744
SGT-GJRGARCH	0.276	0.441	0.117	0.290	0.009	**0.051**	0.019	0.404
GHST-GJRGARCH	**0.072**	4.745	0.380	0.810	0.543	1.923	0.469	0.456
N-APARCH	0.225	4.237	0.233	1.508	0.102	1.052	0.073	1.646
ST-APARCH	0.277	2.664	0.235	0.037	0.106	0.074	0.045	0.274
SKST-APARCH	0.173	2.612	0.058	0.050	0.011	0.085	0.060	0.254
SGED-APARCH	0.213	2.592	0.111	0.538	0.009	0.407	**0.003**	0.748
JSU-APARCH	1.092	88.537	0.092	2.782	0.006	0.364	0.177	3.646
SGT-APARCH	0.165	1.135	0.115	0.430	0.005	0.057	0.005	0.554
GHST-APARCH	0.343	1.368	0.270	0.835	0.438	0.140	0.652	2.511
N-FGARCH	0.234	4.161	0.220	1.459	0.104	1.014	0.064	1.686
ST-FGARCH	0.280	2.559	0.226	0.011	0.106	0.086	0.015	0.345
SKST-FGARCH	0.170	2.518	**0.045**	**0.003**	0.016	0.092	0.099	0.613
SGED-FGARCH	0.212	2.553	0.100	0.500	0.004	0.393	0.019	0.896
JSU-FGARCH	1.099	86.192	0.107	2.849	0.013	0.222	0.210	2.980
SGT-FGARCH	0.163	1.065	0.102	0.380	0.009	0.070	0.027	0.694
GHST-FGARCH	0.314	1.905	0.531	3.974	0.194	2.066	0.650	1.498

Table 2.12: Absolute differences between standardized innovation moments and theoretical moments for individual stocks.

	IRS 5Y		GERMANBOND 10Y		USBOND 10Y	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
N-GARCH	0.136	6.020	0.259	1.625	0.173	1.644
ST-GARCH	0.187	**3.704**	0.253	**0.176**	0.176	0.322
SKST-GARCH	0.364	3.837	0.066	0.201	0.033	0.353
SGED-GARCH	0.215	5.141	0.120	0.721	0.066	0.534
JSU-GARCH	1.171	19.504	0.045	1.793	0.329	6.449
SGT-GARCH	0.289	4.371	0.088	0.287	0.040	**0.190**
GHST-GARCH	0.311	8.732	0.052	2.600	0.314	1.122
N-GJRGARCH	0.315	6.002	0.260	1.603	0.178	1.655
ST-GJRGARCH	0.193	3.807	0.250	0.206	0.183	0.290
SKST-GJRGARCH	0.369	3.945	0.064	0.230	**0.004**	0.302
SGED-GJRGARCH	0.216	5.161	0.121	0.716	0.074	0.556
JSU-GJRGARCH	1.174	19.380	0.047	1.769	0.303	6.189
SGT-GJRGARCH	0.291	4.394	0.088	0.288	0.046	0.196
GHST-GJRGARCH	0.452	9.239	0.074	1.519	0.600	4.455
N-APARCH	0.088	6.168	0.263	1.547	0.181	1.639
ST-APARCH	0.080	6.300	0.249	0.225	0.182	0.284
SKST-APARCH	0.266	6.463	0.062	0.253	0.005	0.298
SGED-APARCH	0.129	5.963	0.121	0.700	0.074	0.551
JSU-APARCH	1.160	20.361	0.051	1.776	0.305	6.211
SGT-APARCH	0.216	4.963	0.090	0.277	0.047	0.196
GHST-APARCH	0.098	26.225	0.062	2.621	0.144	1.558
N-FGARCH	0.067	6.154	0.257	1.500	0.191	1.646
ST-FGARCH	**0.030**	8.461	0.246	0.208	0.189	0.261
SKST-FGARCH	0.226	8.610	0.061	0.237	0.006	0.272
SGED-FGARCH	0.097	6.436	0.119	0.675	0.084	0.561
JSU-FGARCH	1.228	19.090	**0.013**	0.443	0.279	5.958
SGT-FGARCH	0.183	5.372	0.087	0.261	0.057	0.206
GHST-FGARCH	0.337	13.856	0.051	2.659	0.122	2.636

Table 2.13: Absolute differences between standardized innovation moments and theoretical moments for interest rates.

	OIL BRENT		GAS		GOLD		SILVER	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
N-GARCH	0.248	1.897	0.296	2.590	0.213	**4.389**	0.621	**4.960**
ST-GARCH	0.251	0.668	0.310	0.470	0.214	1027.771	0.635	330.553
SKST-GARCH	0.107	0.609	0.266	0.443	**0.021**	380.951	**0.253**	183.017
SGED-GARCH	0.177	0.632	0.310	1.129	0.235	2.210	0.587	1.852
JSU-GARCH	0.343	11.883	**0.079**	21.572	7.085	669.264	15.594	923.526
SGT-GARCH	0.158	**0.251**	0.306	1.009	0.212	**2.003**	0.625	1.629
GHST-GARCH	0.569	8.280	0.785	2.355	0.284	3.881	0.100	0.283
N-GJRGARCH	0.256	1.887	0.288	2.528	0.244	4.768	0.650	5.124
ST-GJRGARCH	0.263	0.462	0.315	0.451	0.309	48.982	0.681	24.536
SKST-GJRGARCH	0.110	0.413	0.271	**0.422**	0.155	40.928	0.379	18.715
SGED-GJRGARCH	0.174	0.666	0.309	1.112	0.323	3.066	0.608	2.330
JSU-GJRGARCH	0.305	9.572	0.079	21.738	5.608	564.599	9.670	529.738
SGT-GJRGARCH	0.159	0.317	0.305	0.997	0.291	2.791	0.652	2.220
GHST-GJRGARCH	0.377	0.975	1.128	4.548	0.303	3.085	0.099	3.690
N-APARCH	0.255	1.883	0.284	2.515	0.394	8.649	0.645	5.004
ST-APARCH	0.274	0.371	0.313	0.505	0.395	75.737	0.665	100.057
SKST-APARCH	0.118	0.315	0.265	0.475	0.215	57.680	0.270	Inf
SGED-APARCH	0.178	0.710	0.305	1.103	0.373	4.778	0.585	2.169
JSU-APARCH	0.294	9.220	0.102	21.777	5.964	560.019	14.586	870.605
SGT-APARCH	0.161	0.338	0.302	0.965	0.333	4.343	0.632	2.056
GHST-APARCH	0.698	2.044	0.975	0.847	0.149	6.751	0.087	**0.018**
N-FGARCH	0.230	1.773	0.257	2.314	0.117	3.662	0.618	4.929
ST-FGARCH	0.252	0.396	0.284	0.724	0.326	82.896	0.651	85.487
SKST-FGARCH	**0.106**	0.341	0.230	0.698	0.137	64.476	0.242	Inf
SGED-FGARCH	0.166	0.635	0.272	0.895	0.297	4.077	0.567	2.109
JSU-FGARCH	0.259	8.370	0.167	21.559	6.264	564.523	14.912	863.214
SGT-FGARCH	0.150	0.321	0.271	0.772	0.265	3.656	0.619	1.986
GHST-FGARCH	0.248	1.120	0.918	0.452	0.173	6.510	**0.068**	0.745

Table 2.14: Absolute differences between standardized innovation moments and theoretical moments for commodities.

	EUR/USD		**GBP/USD**		**JPY/USD**		**AUD/USD**	
	Skewness	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**	**Skewness**	**Kurtosis**
N-GARCH	0.042	0.910	0.034	0.800	0.170	2.894	0.410	1.681
ST-GARCH	0.039	0.405	0.035	0.399	0.173	0.779	0.423	0.208
SKST-GARCH	0.035	0.407	0.045	0.402	0.045	0.718	0.172	0.318
SGED-GARCH	0.018	0.021	0.006	**0.192**	0.085	1.478	0.185	0.852
JSU-GARCH	0.026	2.323	0.069	1.942	0.634	26.765	0.047	1.432
SGT-GARCH	0.022	0.020	**0.001**	0.194	0.079	0.304	0.186	0.410
GHST-GARCH	0.583	0.267	0.431	0.523	0.428	4.092	0.204	1.054
N-GJRGARCH	0.037	0.868	0.033	0.743	0.179	2.838	0.394	1.575
ST-GJRGARCH	0.036	0.422	0.034	0.347	0.176	0.768	0.411	**0.162**
SKST-GJRGARCH	0.032	0.421	0.048	0.356	**0.003**	1.340	0.162	0.259
SGED-GJRGARCH	0.016	0.049	0.014	0.209	0.086	1.441	0.019	2.046
JSU-GJRGARCH	0.026	2.244	0.060	1.525	0.631	26.106	0.055	1.308
SGT-GJRGARCH	0.020	0.049	0.006	0.209	0.083	0.318	0.172	0.349
GHST-GJRGARCH	0.441	1.196	0.311	1.512	0.577	3.216	0.061	1.223
N-APARCH	0.036	0.865	0.034	0.743	0.154	2.676	0.388	1.577
ST-APARCH	0.035	0.421	0.035	0.350	0.159	0.770	0.407	0.195
SKST-APARCH	0.032	0.421	0.047	0.358	0.033	0.720	0.162	0.285
SGED-APARCH	0.016	0.054	0.013	0.210	0.069	1.330	0.169	0.778
JSU-APARCH	0.027	2.225	0.059	1.548	0.594	24.778	0.063	1.216
SGT-APARCH	0.020	0.052	0.004	0.209	0.066	0.298	0.170	0.367
GHST-APARCH	0.406	0.804	0.276	1.711	0.510	3.450	**0.003**	1.856
N-FGARCH	0.037	0.887	0.034	0.739	0.161	2.568	0.388	1.575
ST-FGARCH	0.037	0.348	0.036	0.370	0.159	0.770	0.407	0.196
SKST-FGARCH	0.031	0.349	0.045	0.377	0.033	0.719	0.162	0.287
SGED-FGARCH	**0.016**	**0.009**	0.011	0.215	0.076	1.304	0.169	0.781
JSU-FGARCH	0.024	2.176	0.059	1.616	0.007	**0.217**	0.064	1.200
SGT-FGARCH	0.019	0.009	0.003	0.216	0.074	0.303	0.170	0.367
GHST-FGARCH	0.115	2.419	0.262	1.769	0.499	3.512	0.015	2.012

Table 2.15: Absolute differences between standardized innovation moments and theoretical moments for exchange rates.

2.6.3 Fitting observed returns

2.6.3.1 Fitting the empirical distribution of asset returns

How about the ability of each estimated model to fit sample return moments? Unfortunately, except in cases when returns do not show any stochastic structure, it is not easy to derive the moments of asset returns from the estimated probability distribution for return innovations. Hence, we characterize the implied probability distribution for returns by simulation. Taking random draws for the estimated probability distribution for innovations, we generate 1000 time series for returns with the same length as our data set. For each simulation we apply the two-sample KS test (Kolmogorov, 1933, Smirnov, 1939 and Massey, 1951) [76] [116] [89] and the Chi2 test (Pearson, 1900) [102] to compute the failure rates of the respective null hypotheses.

The KS test quantifies the distance between the empirical distribution function of observed returns and the one obtained from each simulated time series. The KS test statistic is:

$$D = sup_x |F_{1,n}(x) - F_{2,n'}(x)|$$

where sup_x is the supremum of the set of distances between the two empirical distributions, $F_{1,n}$ and $F_{2,n'}$. The null hypothesis is rejected at level α if $D > c(\alpha)\sqrt{\frac{n+n'}{nn'}}$ where n and n' are the sizes of first and second sample respectively. The value of $c(\alpha)$ is given in the table below for each level of α,

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

Table 2.16 reports the failure rates of the KS and Chi2 null hypothesis at confidence level 99%. The models with lower failure rate in either the KS and the Chi2 tests are the SGED distribution with GJRGARCH, APARCH or FGARCH volatility specifications, and the SKST and JSU distributions with APARCH and FGARCH specifications, respectively. Hence, we observe again the preference for asymmetric distributions and volatility models with leverage.

confidence level = 0.99	IBEX35		NASDAQ100		FTSE100		NIKKEI225	
fail rate	**KS**	**Chi2**	**KS**	**Chi2**	**KS**	**Chi2**	**KS**	**Chi2**
N-GARCH	0.821	0.985	0.993	1.000	0.367	0.846	1.000	0.999
ST-GARCH	0.069	0.978	0.391	0.999	0.100	0.676	0.798	0.997
SKST-GARCH	0.202	0.714	0.575	0.994	0.551	0.423	0.836	0.773
SGED-GARCH	0.110	0.594	0.155	0.991	0.519	0.461	0.460	0.979
JSU-GARCH	0.208	0.713	0.549	0.993	0.555	0.436	0.855	0.542
SGT-GARCH	0.324	0.669	0.332	0.992	0.800	0.428	0.465	0.994
GHST-GARCH	0.067	0.889	0.416	0.998	0.067	0.447	0.845	0.811
N-GJRGARCH	0.593	0.997	0.993	1.000	0.721	0.970	0.997	0.999
ST-GJRGARCH	0.149	0.995	0.420	1.000	0.221	0.887	0.806	0.997
SKST-GJRGARCH	0.029	0.775	0.185	0.995	0.054	0.433	0.789	0.818
SGED-GJRGARCH	0.009	0.668	0.051	0.993	0.056	0.450	0.521	0.993
JSU-GJRGARCH	0.025	0.803	0.164	0.996	0.050	0.459	0.841	0.637
SGT-GJRGARCH	0.074	0.648	0.150	0.990	0.270	0.421	0.461	0.991
GHST-GJRGARCH	0.223	0.974	0.617	0.996	0.218	0.705	0.923	0.804
N-APARCH	0.672	0.999	0.992	1.000	0.819	0.993	0.999	0.998
ST-APARCH	0.250	0.994	0.407	1.000	0.313	0.943	0.791	1.000
SKST-APARCH	0.032	0.812	0.177	0.993	0.040	0.527	0.747	0.826
SGED-APARCH	0.024	0.706	0.058	0.989	0.034	0.583	0.565	0.989
JSU-APARCH	0.037	0.818	0.154	0.996	0.037	0.551	0.762	0.708
SGT-APARCH	0.045	0.657	0.170	0.983	0.219	0.515	0.473	0.996
GHST-APARCH	0.318	0.984	0.463	0.999	0.302	0.789	0.808	0.938
N-FGARCH	0.735	0.998	0.984	1.000	0.858	0.983	1.000	1.000
ST-FGARCH	0.305	0.999	0.285	1.000	0.423	0.939	0.768	0.999
SKST-FGARCH	0.033	0.831	0.158	0.998	0.046	0.453	0.714	0.825
SGED-FGARCH	0.030	0.746	0.142	0.999	0.025	0.452	0.607	0.991
JSU-FGARCH	0.033	0.864	0.338	1.000	0.047	0.500	0.700	0.700
SGT-FGARCH	0.040	0.670	0.336	0.998	0.173	0.405	0.441	0.989
GHST-FGARCH	0.257	0.998	0.455	1.000	0.344	0.861	0.802	0.959

Table 2.16: Goodness-of-fit tests for observed returns of stock market indices. Figures denote the fail rates for each model.

2.6.3.2 Fitting the sample moments of asset returns

In addition to the fit to the whole distribution, we now examine the ability of each combination of volatility specification and probability distribution to fit the main moments of observed returns: sample mean, standard deviation, skewness, kurtosis, maximum, minimum and the observed range. To that end, we assign to each model the average value for each of these moments over the set of 1000 simulations, to be compared with their sample return analogues. Tables 2.17 - 2.21 present sample return moments for each asset together with a summary of the average simulated return moments over probability distributions and volatility specifications. Column 1 in Tables 2.17 - 2.21 displays sample moments, while column 2 shows the median value of the average simulated moments across all models (28 in total). The remaining columns show median values of moments across subsets of models[8]. The first panel, from third to ninth column, considers median values of moments across alternative volatility specifications, for a given probability distribution for return innovations. The second panel, from tenth to thirteenth column, presents median values of simulated moments across probability distributions, for a given volatility specification. We also compute the mean absolute difference between the average moments obtained by simulation and the analogue sample moments (mean, standard deviation, skewness, kurtosis, maximum, minimum and the observed range). The last row displays the median value of these absolute differences. Finally, we take the range [9] of MAE values across the set of volatility specifications or across the set of probability distributions, as shown in the last two columns.

The first panel shows that for most assets all probability distributions explain the standard deviations of return data similarly, with the Normal and Student-t distributions doing somewhat better than the rest. The Johnson S_U distribution approximates very well the level of skewness in returns and Skewed Generalized Error distribution does better than other distributions to approximate the level of kurtosis. We conclude that the Normal distribution performs well on this account for stock market indices because it fits very well the second moment but not because it fits well the higher order moments, i.e. the third and fourth moment. In the second panel, the differences between volatility specifications are small compared with differences between probability distributions but APARCH and FGARH models fit standard deviation better than another volatility models, GJRGARCH and FGARCH volatilities seem to fit skewness best, while APARCH and FGARCH fit kurtosis best.

Summarizing, all the probability distributions other than the Normal produce levels of kurtosis as high as those in the return data, but they fall short of explaining the negative skewness observed in some market returns. They also fall a bit short of reproducing the maximum returns. However, they tend to produce a minimum that is higher in absolute value than the one for observed returns. Consequently, the range of values implied by the estimated models is just a bit narrower than that observed in return data for all assets.

According to the median MAE, the Normal distribution is the preferred one for 2 assets, the symmetric Student-t is the best for 4 assets, the Skewed Student-t for 3, the Skewed Generalized Error for 2, the Johnson S_U for 4, the Skewed Generalized-t for 1 and Generalized Hyperbolic Skew Student-t for 3 assets. In terms of volatility models, the standard GARCH is the preferred volatility specifications for 4 assets, the GJR-GARCH model for 1, the APARCH model for 6 and the FGARCH model is the best for 8 assets. So, from this point of view, it looks as if the FGARCH and APARCH volatility specifications and the symmetric Student-t and the Johnson S_U probability distribution should be preferred.

If we exclude from consideration the ability to reproduce the maximum and minimum observed returns the Normal distribution is the preferred one for 2 assets, the symmetric

[8]Remember that for each model we take the average value of each moment over 1000 simulations.

[9]The difference between the highest and the lowest MAE values.

Student-t is the best for 2 assets, the Skewed Student-t for 2, the Skewed Generalized Error for 5, the Johnson S_U for 3, the Skewed Generalized-t for 2 and Generalized Hyperbolic Skew Student-t for 3 assets. In terms of volatility models, the standard GARCH is the preferred volatility specifications for 5 assets, the GJR-GARCH model for 1, the APARCH model for 7 and the FGARCH model is the best for 6 assets. Again, the APARCH and FGARCH volatility models perform better than GARCH and GJR-GARCH, but now the Skewed Generalized Error distribution is the preferred one.

Interestingly enough, the last two columns show that median values of the simulated statistics for different volatility specifications are more similar among them than median values for the alternative probability distributions. This suggests again that the assumption we can make on the probability distribution of return innovations may be more important to fit return data than the assumption on the volatility specification.

		Median over all models	Median over probability distributions							Median over volatility models				Ranges over distributions (left) and over volatility models (right)	
	Sample		N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH		
IBEX35															
Mean	-0.005	0.018	0.007	0.028	0.011	0.006	0.012	0.029	0.009	0.046	0.015	0.007	0.004	0.023	0.042
Standard deviation	1.495	1.743	1.592	1.474	1.874	1.685	1.743	2.004	2.400	2.049	1.750	1.516	1.674	0.926	0.532
Skewness	0.083	-0.193	0.042	0.064	-0.272	-0.224	-0.318	-0.078	-0.740	-0.244	-0.182	-0.235	-0.079	0.804	0.165
Kurtosis	7.932	12.364	6.875	12.419	14.043	12.206	13.732	10.226	16.545	12.860	13.755	9.423	13.217	9.670	4.333
Maximum	13.484	13.675	10.356	12.234	15.004	13.348	14.102	14.953	16.584	15.095	13.989	10.735	14.316	6.228	4.360
Minimum	-9.586	-13.746	-9.976	-11.114	-15.613	-13.746	-15.138	-14.938	-21.414	-16.437	-13.701	-11.861	-13.792	11.438	4.576
Range	23.070	27.584	20.332	23.348	30.866	27.208	29.557	29.891	37.448	31.533	27.690	22.596	27.791	17.115	8.936
		Median MAE	**0.853**	**1.237**	**2.483**	**1.580**	**2.174**	**1.630**	**4.130**	**2.459**	**1.820**	**0.953**	**2.001**	**3.277**	**1.505**
NASDAQ100															
Mean	0.005	0.052	0.027	0.060	0.032	0.042	0.049	0.059	0.037	0.070	0.033	0.031	0.055	0.034	0.039
Standard deviation	1.848	1.759	1.591	1.420	1.828	1.714	1.789	1.955	2.676	1.876	1.939	1.744	1.245	1.257	0.694
Skewness	0.192	-0.097	0.062	0.066	-0.227	-0.177	-0.297	-0.043	-0.788	-0.199	-0.140	-0.219	0.044	0.854	0.263
Kurtosis	9.623	11.838	6.833	11.691	15.384	11.343	12.643	9.933	33.514	11.669	15.312	12.087	7.035	26.681	8.277
Maximum	17.203	13.741	9.866	11.939	15.234	12.849	13.616	15.024	19.485	13.598	15.853	13.885	10.031	9.618	5.822
Minimum	-11.115	-14.403	-9.292	-10.803	-15.034	-13.247	-15.093	-14.524	-24.307	-14.978	-15.248	-14.469	-7.301	15.015	7.946
Range	28.318	28.319	19.158	22.742	30.079	26.096	28.709	29.548	42.950	28.326	31.101	29.092	19.883	23.792	11.218
		Median MAE	**2.134**	**1.507**	**2.296**	**1.739**	**2.431**	**1.508**	**7.888**	**1.783**	**2.534**	**1.845**	**2.896**	**6.380**	**1.113**
FTSE100															
Mean	-0.003	-0.003	-0.006	0.010	-0.005	-0.009	-0.005	0.010	-0.008	0.028	-0.003	-0.008	-0.011	0.019	0.039
Standard deviation	1.210	1.346	1.221	1.237	1.321	1.244	1.279	1.665	1.762	1.287	1.326	1.231	1.551	0.541	0.320
Skewness	-0.161	-0.277	0.065	0.067	-0.325	-0.277	-0.404	-0.094	-0.596	-0.290	-0.261	-0.305	-0.265	0.663	0.044
Kurtosis	9.356	11.624	7.223	11.786	13.472	12.302	13.579	10.420	21.043	11.269	14.362	9.968	17.074	13.820	7.106
Maximum	9.384	10.687	8.027	9.726	10.406	9.813	10.133	12.133	14.531	9.083	10.786	8.984	13.586	6.504	4.602
Minimum	-9.266	-10.675	-7.673	-9.437	-10.967	-10.049	-11.161	-12.544	-16.806	-10.260	-10.751	-10.189	-14.170	9.132	3.981
Range	18.650	21.293	15.700	19.164	21.337	19.863	21.294	24.677	31.337	19.343	21.537	19.229	27.293	15.637	8.064
		Median MAE	**0.888**	**0.926**	**1.213**	**0.856**	**1.294**	**1.349**	**4.219**	**0.868**	**1.384**	**0.539**	**2.941**	**3.363**	**2.402**
NIKKEI225															
Mean	0.000	0.012	0.012	0.012	0.012	-0.001	0.012	0.012	-0.011	0.012	0.012	0.012	0.002	0.023	0.010
Standard deviation	1.499	1.561	1.533	1.502	1.506	1.529	1.500	2.182	1.693	1.545	1.482	1.484	1.678	0.681	0.195
Skewness	-0.410	-0.064	0.014	0.024	-0.218	-0.064	-0.336	0.000	-0.751	-0.066	-0.037	-0.061	-0.092	0.775	0.055
Kurtosis	9.725	8.711	4.825	8.581	8.987	9.273	8.898	8.565	13.131	8.672	9.301	7.444	11.667	8.306	4.223
Maximum	13.235	10.939	8.372	10.785	10.597	11.194	10.177	15.646	11.342	10.992	10.886	10.076	12.773	7.275	2.697
Minimum	-12.120	-11.439	-8.158	-10.532	-11.305	-11.072	-11.567	-15.496	-15.983	-11.422	-11.189	-10.693	-14.054	7.825	3.361
Range	25.354	22.376	16.530	21.318	21.902	22.266	21.744	31.142	27.325	22.309	22.074	20.357	27.491	14.612	7.134
		Median MAE	**2.367**	**0.947**	**0.798**	**0.991**	**0.821**	**1.341**	**1.618**	**0.983**	**1.076**	**1.323**	**1.259**	**1.570**	**0.340**

Table 2.17: Empirical moments vs sample moments for stock market indices.

		Median over	Median over probability distributions							Median over volatility models				Ranges over distributions (left) and over volatility models (right)	
	Sample	all models	N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH		
IBM															
Mean	0.004	0.004	0.006	0.017	0.006	-0.004	0.004	0.005	-0.024	0.017	0.009	0.000	-0.002	0.041	0.019
Standard deviation	1.660	1.674	1.961	1.724	1.674	1.598	1.602	2.231	1.645	1.711	1.630	1.724	1.670	0.633	0.095
Skewness	-0.071	-0.023	0.026	0.109	-0.072	-0.020	-0.127	0.006	-0.697	-0.025	0.015	-0.029	-0.034	0.806	0.049
Kurtosis	11.628	11.097	5.551	17.933	16.961	9.479	12.434	9.300	10.583	9.604	11.133	11.120	11.074	12.382	1.529
Maximum	12.260	13.474	11.472	16.512	15.725	11.803	13.418	16.904	10.513	12.852	13.942	13.451	13.498	6.391	1.090
Minimum	-16.892	-14.357	-11.070	-15.174	-15.534	-11.680	-13.784	-16.714	-14.417	-13.488	-13.680	-15.174	-14.802	5.645	1.686
Range	29.152	27.493	22.489	31.701	31.455	23.482	27.318	33.619	24.930	26.341	27.622	29.332	27.273	11.129	2.992
		Median MAE	**2.382**	**1.954**	**1.629**	**1.346**	**0.834**	**1.347**	**0.996**	**1.756**	**1.426**	**1.424**	**1.404**	**1.548**	**0.351**
SAN															
Mean	-0.010	0.006	-0.004	0.017	-0.002	-0.009	-0.001	0.021	-0.025	0.046	0.006	-0.010	-0.011	0.046	0.057
Standard deviation	2.190	2.762	2.609	2.273	2.608	2.582	2.510	3.698	3.883	3.496	2.799	2.387	2.387	1.611	1.109
Skewness	0.147	-0.068	0.083	0.078	-0.165	-0.092	-0.228	-0.131	-0.743	-0.113	-0.006	-0.098	-0.086	0.827	0.106
Kurtosis	9.114	13.557	9.549	13.501	15.442	13.908	14.606	14.459	19.032	16.694	17.199	11.117	12.291	9.483	6.081
Maximum	20.877	23.153	19.093	19.729	22.595	21.992	21.659	29.758	26.997	27.046	24.751	18.690	19.496	10.665	8.356
Minimum	-15.186	-23.581	-17.330	-17.981	-22.189	-21.065	-21.839	-30.572	-36.777	-28.494	-24.015	-19.281	-19.664	19.447	9.214
Range	36.063	47.416	36.423	37.710	44.785	43.057	43.498	60.330	63.774	55.541	48.624	37.601	38.373	27.351	17.940
		Median MAE	**0.838**	**1.412**	**2.821**	**2.477**	**2.728**	**5.599**	**6.740**	**4.908**	**3.696**	**1.607**	**1.759**	**5.902**	**3.301**
AXA															
Mean	-0.005	-0.001	-0.004	0.011	-0.003	-0.006	-0.002	0.013	-0.037	0.055	0.001	-0.007	-0.019	0.050	0.074
Standard deviation	2.673	2.599	2.626	2.432	2.572	2.536	2.499	3.437	4.113	2.504	2.584	2.514	2.638	1.681	0.134
Skewness	0.266	-0.080	0.034	0.034	-0.099	-0.086	-0.132	-0.029	-0.644	-0.079	-0.062	-0.101	-0.092	0.678	0.039
Kurtosis	10.090	9.303	6.865	9.626	10.368	9.393	9.729	8.772	16.564	8.946	10.625	8.529	10.313	9.699	2.097
Maximum	19.778	19.666	17.039	18.115	19.501	18.454	18.564	24.632	29.093	18.022	19.840	18.247	19.768	12.054	1.817
Minimum	-20.350	-18.910	-16.344	-17.476	-19.567	-18.274	-18.803	-24.561	-38.232	-18.022	-19.481	-18.124	-20.143	21.888	2.121
Range	40.128	38.617	33.352	35.591	39.170	36.707	37.417	49.193	66.634	36.044	39.234	35.725	39.794	33.282	4.070
		Median MAE	**1.720**	**0.926**	**0.653**	**0.828**	**0.734**	**1.916**	**5.797**	**1.273**	**0.956**	**1.339**	**0.521**	**5.144**	**0.817**
BP															
Mean	-0.014	-0.016	-0.017	-0.008	-0.019	-0.024	-0.018	-0.005	-0.053	0.007	-0.015	-0.021	-0.027	0.048	0.034
Standard deviation	1.715	1.757	1.670	1.719	1.780	1.681	1.740	2.295	2.139	1.815	1.721	1.731	1.757	0.625	0.094
Skewness	-0.127	-0.039	0.018	0.031	-0.089	-0.039	-0.124	-0.008	-0.786	-0.029	-0.028	-0.048	-0.050	0.817	0.023
Kurtosis	7.814	8.232	4.342	8.851	9.900	7.323	8.937	6.501	12.452	9.444	9.215	7.910	8.431	8.110	1.533
Maximum	10.583	12.818	8.578	12.744	13.466	11.179	12.535	14.673	13.687	13.448	12.978	12.278	12.869	6.096	1.169
Minimum	-14.037	-12.996	-8.246	-12.158	-13.580	-11.055	-12.947	-14.394	-20.002	-13.976	-12.937	-12.420	-12.956	11.756	1.556
Range	24.619	25.859	16.731	24.902	27.098	22.389	25.593	29.067	33.689	27.679	25.915	24.297	25.270	16.958	3.382
		Median MAE	**1.936**	**0.882**	**0.887**	**0.709**	**0.659**	**1.078**	**2.458**	**1.033**	**1.056**	**0.789**	**0.802**	**1.799**	**0.267**

Table 2.18: Empirical moments vs sample moments for individual stocks.

		Median over	Median over probability distributions							Median over volatility models				Ranges over distributions (left) and over volatility models (right)	
	Sample	all models	N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH		
IRS 5Y															
Mean	0.006	0.006	0.007	0.008	0.006	0.005	0.006	0.006	0.006	0.006	0.006	0.007	0.006	0.003	0.001
Standard deviation	0.208	0.254	0.241	0.276	0.281	0.235	0.251	0.341	0.199	0.286	0.345	0.223	0.213	0.142	0.132
Skewness	-0.282	-0.069	0.004	0.014	-0.195	-0.069	-0.257	-0.053	-0.234	-0.067	-0.068	-0.070	-0.083	0.271	0.016
Kurtosis	8.527	9.488	4.508	12.239	12.260	8.673	10.632	8.073	9.034	9.744	12.415	7.601	8.818	7.752	4.814
Maximum	1.921	1.763	1.232	2.216	2.107	1.610	1.820	2.323	1.314	2.142	2.539	1.497	1.385	1.092	1.154
Minimum	-1.862	-1.793	-1.217	-2.157	-2.238	-1.641	-1.995	-2.379	-1.565	-2.333	-2.715	-1.658	-1.655	1.161	1.060
Range	3.783	3.495	2.449	4.373	4.345	3.248	3.815	4.702	2.879	4.476	5.289	3.155	3.155	2.253	2.134
		Median MAE	**0.946**	**0.826**	**0.817**	**0.309**	**0.488**	**0.399**	**0.634**	**0.773**	**0.975**	**0.384**	**0.445**	**0.636**	**0.591**
GERMAN BOND 10Y															
Mean	0.011	0.010	0.009	0.016	0.011	0.008	0.010	0.014	0.009	0.010	0.010	0.010	0.011	0.008	0.001
Standard deviation	0.413	0.407	0.402	0.407	0.407	0.407	0.404	0.579	0.413	0.407	0.409	0.407	0.403	0.177	0.006
Skewness	-0.096	-0.150	0.001	0.000	-0.197	-0.150	-0.244	-0.062	-0.321	-0.150	-0.150	-0.149	-0.147	0.321	0.003
Kurtosis	5.967	5.087	3.379	5.261	5.296	4.657	5.087	4.292	5.709	5.106	5.090	5.084	5.062	2.330	0.044
Maximum	3.392	2.150	1.678	2.391	2.238	2.001	2.095	2.864	2.150	2.172	2.150	2.137	2.150	1.186	0.034
Minimum	-2.317	-2.434	-1.647	-2.364	-2.477	-2.165	-2.434	-2.988	-2.792	-2.441	-2.435	-2.434	-2.417	1.341	0.024
Range	5.709	4.695	3.325	4.755	4.716	4.166	4.529	5.852	4.947	4.726	4.717	4.715	4.652	2.527	0.074
		Median MAE	**0.847**	**0.310**	**0.349**	**0.486**	**0.409**	**0.513**	**0.368**	**0.405**	**0.408**	**0.409**	**0.409**	**0.537**	**0.005**
US BOND 10Y															
Mean	0.010	0.009	0.009	0.016	0.008	0.007	0.009	0.015	0.003	0.007	0.009	0.009	0.009	0.013	0.003
Standard deviation	0.594	0.585	0.593	0.595	0.582	0.585	0.576	0.847	0.583	0.588	0.607	0.585	0.575	0.270	0.033
Skewness	-0.217	-0.126	-0.002	-0.009	-0.198	-0.126	-0.235	-0.054	-0.424	-0.124	-0.128	-0.128	-0.124	0.422	0.005
Kurtosis	7.958	5.858	3.675	6.345	6.339	5.350	5.917	4.864	6.263	5.977	5.924	5.844	5.857	2.671	0.133
Maximum	4.531	3.274	2.610	3.773	3.521	3.119	3.274	4.523	3.001	3.331	3.274	3.264	3.273	1.913	0.067
Minimum	-5.572	-3.767	-2.600	-3.827	-3.861	-3.326	-3.751	-4.645	-4.177	-3.778	-3.935	-3.745	-3.689	2.046	0.246
Range	10.104	7.070	5.210	7.600	7.360	6.447	7.020	9.168	7.200	7.421	7.472	6.740	6.962	3.959	0.732
		Median MAE	**1.568**	**0.724**	**0.735**	**1.061**	**0.860**	**0.756**	**0.804**	**0.784**	**0.728**	**0.940**	**0.878**	**0.844**	**0.213**

Table 2.19: Empirical moments vs sample moments for interest rates.

		Median over	Median over probability distributions							Median over volatility models				Ranges over distributions (left)	
	Sample	all models	N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH	and over volatility models (right)	
OIL BRENT															
Mean	0.010	0.011	0.009	0.030	0.009	0.003	0.009	0.026	-0.027	0.019	0.007	0.005	0.010	0.057	0.013
Standard deviation	2.281	3.464	3.092	2.963	4.080	3.480	3.447	4.389	3.507	3.450	4.061	3.741	3.197	1.426	0.864
Skewness	-0.195	-0.070	0.023	0.055	-0.123	-0.079	-0.185	-0.013	-0.590	-0.108	-0.068	-0.011	-0.084	0.644	0.097
Kurtosis	8.257	8.999	5.259	9.597	10.206	8.518	9.322	7.587	10.504	9.120	9.524	8.720	8.877	5.246	0.803
Maximum	17.969	21.604	15.471	20.771	24.479	20.762	21.497	26.023	19.315	19.782	23.212	22.216	20.345	10.553	3.430
Minimum	-18.725	-22.363	-15.052	-19.929	-25.125	-21.239	-23.092	-25.993	-28.254	-22.390	-24.648	-22.336	-21.147	13.202	3.501
Range	36.694	44.573	30.523	40.650	49.604	42.001	44.588	52.007	47.569	45.117	47.860	44.552	42.612	21.484	5.248
		Median MAE	**1.701**	**1.061**	**2.815**	**1.152**	**1.712**	**3.055**	**2.483**	**1.875**	**2.374**	**1.887**	**1.755**	**1.995**	**0.620**
GAS															
Mean	0.000	-0.011	-0.008	-0.012	-0.013	-0.002	-0.011	0.003	-0.011	-0.011	-0.011	-0.010	-0.009	0.016	0.002
Standard deviation	4.196	4.536	13.261	4.424	4.425	4.619	4.366	6.570	5.583	4.574	4.809	4.428	4.332	8.896	0.477
Skewness	0.559	0.019	0.015	0.028	0.030	-0.004	0.074	0.014	-0.867	0.014	0.020	0.018	0.027	0.941	0.013
Kurtosis	12.809	15.878	16.181	15.632	15.727	14.372	13.991	13.873	16.636	16.278	16.365	13.549	13.352	2.763	3.013
Maximum	37.814	41.562	101.791	39.990	39.636	38.399	37.802	55.771	40.469	41.650	42.040	37.688	37.809	63.989	4.352
Minimum	-28.899	-41.957	-103.351	-39.504	-39.912	-38.870	-37.449	-55.403	-57.221	-42.193	-42.033	-37.791	-36.755	65.902	5.437
Range	66.714	83.591	205.142	79.511	79.549	77.269	75.252	111.177	97.690	83.732	83.618	75.479	74.565	129.891	9.168
		Median MAE	**25.553**	**2.722**	**2.779**	**2.661**	**2.125**	**8.301**	**6.320**	**3.844**	**3.813**	**1.773**	**1.735**	**23.428**	**2.109**
GOLD															
Mean	0.031	0.029	0.036	0.044	0.034	0.025	0.030	0.023	0.014	0.027	0.032	0.030	0.029	0.030	0.005
Standard deviation	1.129	1.247	1.147	1.605	1.579	1.186	1.327	1.678	0.963	1.392	1.605	1.261	1.149	0.715	0.457
Skewness	-0.415	-0.052	-0.002	-0.052	-0.210	-0.006	-0.275	-0.007	-0.569	-0.006	-0.180	-0.077	-0.027	0.567	0.174
Kurtosis	8.811	9.739	3.842	20.339	20.383	9.288	14.431	7.524	6.895	9.312	11.569	9.961	8.556	16.541	3.013
Maximum	6.865	10.430	5.172	15.426	14.787	8.583	11.132	11.280	4.799	11.492	13.887	10.357	9.352	10.627	4.535
Minimum	-10.162	-10.984	-5.143	-15.543	-15.425	-8.593	-12.180	-11.260	-7.387	-11.346	-13.995	-11.174	-9.915	10.400	4.081
Range	17.028	21.580	10.315	30.969	30.212	17.177	23.312	22.540	12.118	22.839	27.882	21.861	19.499	20.653	8.383
		Median MAE	**2.042**	**4.388**	**4.240**	**0.708**	**2.041**	**1.295**	**1.230**	**2.142**	**2.127**	**1.636**	**1.189**	**3.680**	**0.953**
SILVER															
Mean	0.023	0.005	0.005	0.044	0.014	-0.011	0.004	0.005	-0.017	-0.001	0.006	0.005	0.006	0.061	0.007
Standard deviation	1.926	2.884	3.547	3.108	2.778	2.861	2.491	4.598	1.656	3.086	2.884	4.748	2.429	2.942	2.320
Skewness	-0.568	-0.167	-0.023	-0.134	-0.400	-0.132	-0.506	-0.032	-0.756	-0.063	-0.176	-0.252	-0.108	0.733	0.189
Kurtosis	8.624	13.237	5.800	21.952	20.713	13.237	16.437	11.294	8.466	13.138	14.129	13.336	10.181	16.152	3.948
Maximum	13.665	21.732	18.751	27.384	24.203	21.034	20.098	32.277	8.124	23.170	20.994	28.401	18.516	24.153	9.886
Minimum	-12.982	-23.125	-16.153	-28.334	-26.412	-22.499	-23.751	-32.756	-13.927	-26.339	-22.548	-26.806	-19.338	18.828	7.468
Range	26.646	44.302	31.928	55.701	50.616	43.533	44.083	65.032	22.288	49.350	43.575	49.229	37.077	42.745	12.273
		Median MAE	**2.729**	**7.235**	**6.187**	**3.895**	**4.338**	**7.381**	**1.256**	**5.317**	**3.972**	**7.681**	**2.512**	**6.125**	**5.170**

Table 2.20: Empirical moments vs sample moments for commodities.

		Median over	Median over probability distributions							Median over volatility models				Ranges over distributions (left)	
	Sample	all models	N	ST	SKST	SGED	JSU	SGT	GHST	GARCH	GJRGARCH	APARCH	FGARCH	and over volatility models (right)	
EUR/USD															
Mean	0.002	0.006	0.005	0.005	0.006	0.006	0.006	0.005	0.008	0.007	0.006	0.005	0.006	0.004	0.001
Standard deviation	0.630	0.789	0.716	0.813	0.823	0.725	0.773	1.043	0.911	0.801	0.825	0.751	0.797	0.328	0.074
Skewness	0.141	0.017	0.007	0.013	0.024	0.026	0.022	0.014	-0.107	0.011	0.016	0.018	0.035	0.133	0.025
Kurtosis	5.489	6.193	3.918	6.504	6.486	5.700	6.241	5.182	7.662	6.257	6.456	5.901	6.535	3.744	0.634
Maximum	4.617	4.835	3.219	4.892	4.975	4.016	4.637	5.542	5.683	4.810	4.973	4.500	5.080	2.464	0.580
Minimum	-3.844	-4.537	-3.169	-4.869	-4.843	-3.963	-4.537	-5.452	-5.951	-4.827	-4.863	-4.418	-4.604	2.782	0.445
Range	8.462	9.482	6.388	9.772	9.820	7.980	9.173	10.994	11.634	9.671	9.873	8.977	9.501	5.247	0.897
		Median MAE	**0.644**	**0.441**	**0.446**	**0.202**	**0.293**	**0.619**	**1.026**	**0.435**	**0.459**	**0.279**	**0.514**	**0.824**	**0.235**
GBP/USD															
Mean	-0.002	0.001	0.001	0.004	0.001	0.000	0.002	0.004	-0.005	0.005	0.000	0.001	0.001	0.009	0.004
Standard deviation	0.575	0.594	0.538	0.571	0.599	0.583	0.591	0.803	0.862	0.594	0.583	0.603	0.593	0.323	0.020
Skewness	-0.047	-0.046	0.006	0.015	-0.073	-0.046	-0.087	-0.007	-0.332	-0.053	-0.047	-0.044	-0.041	0.347	0.012
Kurtosis	7.269	5.907	3.653	5.900	6.083	5.773	6.035	5.126	8.555	5.887	5.860	5.914	5.985	4.902	0.125
Maximum	4.431	3.554	2.401	3.470	3.644	3.353	3.547	4.444	5.804	3.548	3.449	3.644	3.650	3.403	0.201
Minimum	-3.887	-3.615	-2.331	-3.365	-3.683	-3.363	-3.650	-4.363	-6.291	-3.622	-3.546	-3.717	-3.678	3.960	0.171
Range	8.318	7.155	4.732	6.830	7.319	6.706	7.192	8.809	12.474	7.137	6.995	7.361	7.328	7.742	0.366
		Median MAE	**1.215**	**0.487**	**0.369**	**0.518**	**0.403**	**0.490**	**1.065**	**0.531**	**0.540**	**0.479**	**0.475**	**0.846**	**0.065**
JPY/USD															
Mean	-0.004	-0.008	-0.008	-0.013	-0.008	-0.007	-0.007	-0.013	-0.023	-0.011	-0.008	-0.008	-0.008	0.016	0.003
Standard deviation	0.635	0.632	0.688	0.632	0.637	0.626	0.628	0.886	0.605	0.644	0.633	0.630	0.630	0.281	0.014
Skewness	0.270	0.040	-0.002	0.007	0.126	0.085	0.153	0.040	-0.361	0.041	0.040	0.039	0.037	0.513	0.004
Kurtosis	6.962	5.766	3.515	7.788	7.954	5.399	6.765	4.923	5.766	5.512	5.832	5.746	5.787	4.439	0.320
Maximum	4.610	4.294	2.949	4.527	4.683	3.557	4.294	4.988	3.083	4.350	4.318	4.270	4.271	2.039	0.080
Minimum	-3.710	-4.259	-2.975	-4.503	-4.477	-3.455	-4.029	-4.847	-4.259	-4.064	-4.209	-4.308	-4.395	1.872	0.331
Range	8.320	8.323	5.924	9.030	9.159	7.012	8.323	9.835	7.271	8.414	8.382	8.264	8.264	3.911	0.150
		Median MAE	**1.029**	**0.329**	**0.330**	**0.511**	**0.160**	**0.674**	**0.660**	**0.507**	**0.519**	**0.515**	**0.505**	**0.869**	**0.014**
AUD/USD															
Mean	0.002	0.003	0.001	0.015	0.003	-0.001	0.002	0.013	0.000	0.007	0.002	0.002	0.001	0.015	0.006
Standard deviation	0.831	0.788	0.828	0.727	0.786	0.786	0.781	1.036	0.862	0.799	0.788	0.785	0.784	0.309	0.014
Skewness	-0.818	-0.260	0.008	0.010	-0.274	-0.260	-0.339	-0.085	-0.477	-0.273	-0.084	-0.260	-0.260	0.487	0.189
Kurtosis	15.132	6.143	4.175	6.170	6.671	5.802	6.395	5.037	8.324	6.575	6.266	6.072	6.074	4.148	0.503
Maximum	6.701	4.615	4.068	4.604	4.715	4.277	4.416	5.581	5.126	4.776	4.624	4.601	4.606	1.513	0.175
Minimum	-8.828	-5.256	-3.991	-4.493	-5.327	-4.794	-5.302	-5.851	-6.766	-5.365	-5.376	-5.237	-5.238	2.775	0.138
Range	15.529	9.866	8.059	9.096	10.067	9.083	9.717	11.432	11.923	10.247	10.017	9.628	9.625	3.865	0.622
		Median MAE	**3.203**	**2.733**	**2.419**	**2.761**	**2.512**	**2.526**	**1.801**	**2.366**	**2.521**	**2.559**	**2.559**	**1.402**	**0.193**

Table 2.21: Empirical moments vs sample moments for exchange rates.

2.7 VaR Performance

We now analyze VaR performance of our estimated models restricting our attention to the left tail of the distribution and the 1% significance level. Results for alternative significance levels are available from the authors upon request. The choice of the 1% level is a compromise between trying to capture extreme events and trying to avoid a too low number of exceptions. Considering the left tail is not a trivial choice, since results for both tails may differ significantly for asymmetric return distributions. In all cases we show out-of-sample VaR estimates over the last five years in the sample: 2011-2015 (1260 data). Every day we compute 1-day ahead 1% VaR, reestimating each model every 50 days. The latter choice tries to reduce the computational cost as well as avoiding frequent parameter variation that might be due in part to pure noise.

The one-step ahead $VaR_{\alpha,t} = \mu_t(\theta) + \sigma_t(\theta)F^{-1}(\alpha|\theta)$, where $\mu_t(\theta)$ represents the conditional mean, $\sigma_t(\theta)$ is the conditional standard deviation and $F^{-1}(\alpha|\theta)$ denotes the corresponding quantile of the distribution of the standardized innovations z_t at a given α% significance.

The performance of VaR is examined through standard tests: the unconditional coverage test of Kupiec (1995) [78], the independence and conditional coverage tests of Christoffersen (1998) [27], the Dynamic Quantile test of Engle and Manganelli (2004) [39], as well as the loss functions proposed by Lopez (1998, 1999) [85, 86] and Sarma et al. (2003) [113] and that of Giacomini and Komunjer (2005) [46].

2.7.1 Backtesting VaR

The unconditional coverage test introduced by Kupiec (1995) [78] is based on the number of violations, i.e. the number of times returns exceed the predicted VaR (T_1) over a period of time T for a given significance level. If the VaR model is correctly specified, the failure rate ($\hat{\pi} = \frac{T_1}{T}$) should be equal to the pre-specified VaR level (α). The null hypothesis $H0 : \pi = \alpha$ is evaluated through a likelihood ratio test:

$$LR_{uc} = -2\ln\left(\frac{L(\Pi_\alpha)}{L(\widehat{\Pi})}\right) = -2\ln\left(\frac{(1-\alpha)^{T_0}\alpha^{T_1}}{(1-\hat{\pi})^{T_0}\hat{\pi}^{T_1}}\right) \quad \overset{T\to\infty}{\longrightarrow} \chi^2_1$$

where $T_0 = T - T_1$.

Christoffersen (1998) [27] developed a conditional coverage test from the unconditional coverage test (LR_{uc}) and the independence test (LR_{ind}).

The LR_{ind} statistic: $LR_{ind} = -2\ln(L(\widehat{\Pi})/L(\widehat{\Pi}_1))$ is the likelihood ratio statistic for the hypothesis of serial independence against first-order Markov dependence. It has an asymptotic χ^2_1 distribution. The likelihood function under the null hypothesis ($\pi_{01} = \pi_{11} = \pi = (T_{11}+T_{01})/T$) is $L(\widehat{\Pi}) = (1-\hat{\pi})^{T_0}\hat{\pi}^{T_1}$ where $T_0 = T_{00}+T_{10}$ and $T_1 = T_{11}+T_{01}$. The likelihood function under the alternative hypothesis is $L(\widehat{\Pi}_1) = (1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}$ where T_{ij} denotes the number of observations in state j after having been in state i in the previous period, $\hat{\pi}_{01} = T_{01}/(T_{00} + T_{01})$ and $\hat{\pi}_{11} = T_{11}/(T_{10} + T_{11})$.

He assumes that, under the alternative hypothesis of VaR inefficiency, the process of violations $I_t(\alpha)$, where $I_t(\alpha) = 1$ if $r_t < VaR(\alpha)$ and $I_t(\alpha) = 0$ otherwise, can be modeled as a Markov chain with $\pi_{ij} = Pr[I_t(\alpha) = j|I_{t-1}(\alpha) = i]$. This leads to a test of the null hypothesis of conditional coverage using a simple likelihood ratio statistic, $LR_{cc} = -2\ln(L(\Pi_\alpha)/L(\widehat{\Pi}_1)) = LR_{uc} + LR_{ind}$, which is asymptotically distributed χ^2_2.

While this test is easy to use, it is rather limited for two main reasons, i) The independence is tested against a very particular form of alternative dependence structure that does not take into account a dependence of order higher than one, ii) The use of a Markov chain only considers the influence of past violations $I_t(\alpha)$ and not the influence of any other exogenous variable.

The Dynamic Quantile Test proposed by Engle and Manganelli (2004) [39] overcomes these two drawbacks of the conditional coverage test. These authors suggest using a linear regression model that links current violations to past violations. Let us define the auxiliary variable: $Hit_t(\alpha) = I_t(\alpha) - \alpha$ so that $Hit_t(\alpha) = 1 - \alpha$ if $r_t < VaR_{t|t-1}(\alpha)$ and $Hit_t(\alpha) = -\alpha$ otherwise. The null hypothesis of this test is that the sequence of hits (Hit_t) is uncorrelated with any variable that belongs to the information set Ω_{t-1} available when the VaR was calculated and it has a mean value of zero, which implies that the hits are not autocorrelated.

The Dynamic Quantile test is a Wald test of the null hypothesis that all slopes in the regression model,

$$Hit_t(\alpha) = \delta_0 + \sum_{i=1}^{p} \delta_i Hit_{t-i} + \sum_{j=p+1}^{q} \delta_j X_j + \epsilon_t$$

are zero, where X_j are explanatory variables contained in Ω_{t-1}. The test statistic has an asymptotic χ^2_{p+q+1} distribution. In our implementation of the test, we use $p = 5$ and $q = 1$ (where $X_1 = VaR(\alpha)$) as proposed by Engle and Manganelli (2004). By doing so, we are testing whether the probability of an exception depends on the level of the VaR.

Lopez (1998, 1999) [85, 86] introduced loss functions in VaR evaluation to take into account the magnitude of the excesses that occur with respect to the VaR. Using that insight, Sarma et al. (2003) [113] introduced the Quadratic Loss Function (QLF) that uses squared distances between the observed returns and the $VaR(\alpha)$ predicted when a violation occurs, to ensure a greater penalty on large excesses:

$$lf_{t+1} = \begin{cases} (r_{t+1} - Var(\alpha))^2 & \text{if } r_{t+1} < VaR(\alpha) \\ 0 & \text{if } r_{t+1} \geq VaR(\alpha) \end{cases}$$

A VaR model should be preferable to another if it has a lower average value of the loss function, $\left(\sum_{t=1}^{T} \frac{lf_t}{T}\right)$.

Later on, Giacomini and Komunjer (2005) [46] suggested the use of the Asymmetric Linear Tick Loss Function (AlTick) that takes into account the magnitude of the implicit cost associated with VaR forecasting errors. Hence, it takes into account not only the returns that exceed the VaR, but also the opportunity cost produced by an overestimation of VaR. When there are not exceptions, the loss function also penalizes due to the excess capital retained:

$$L_\alpha(e_{t+1}) = \begin{cases} (\alpha - 1)e_{t+1} & \text{if } e_{t+1} < 0 \\ \alpha e_{t+1} & \text{if } e_{t+1} \geq 0 \end{cases}$$

where $e_{t+1} = r_{t+1} - VaR_{t+1}$. Giacomini and Komunjer use the asymmetric linear loss function of order α because the object of interest is the conditional α-quantile of the distribution of returns. If a quadratic loss function is used, the optimal forecast is the conditional mean of the distribution of returns and if, on the other hand, an absolute value loss function is used, the optimal forecast corresponds to the conditional median of the distribution of returns. For this reason, the AlTick is the implicit loss function whenever the object of interest is a forecast of a particular quantile of the conditional distribution of returns. A VaR model is preferable if it has a lower average value of the loss function.

2.7.2 VaR Analysis

The different combinations of probability distributions and volatility specifications, applied to each of the 19 assets considered, yield a large number of VaR tests and it is hard to summarize so much information in order to achieve some clear-cut conclusion on the adequacy of each model.

Some authors compare the VaR methodologies and VaR models using a two-stage selection process. This approach, proposed by Sarma et al. (2003) [113], consists in removing in a first stage those methods or models that fail to pass statistical accuracy tests (backtesting), like those described above. The VaR models selected in this stage are then compared in a second stage on the basis of loss functions. Even though this two-stage selection approach helps in selecting a smaller set of competing models, it may fail to identify suitable models because they have been removed in the first stage. Indeed, a model may be rejected in the first stage because it is not statistically appropriate for a given test at a specific confidence level, in spite of having a smaller loss than another one that has been judged to be statistically appropriate in the first stage. Under that approach the VaR accuracy tests resemble more a decision-making process than an evaluation using loss functions. In the extreme case when we identify a single model as appropriate in the first stage, we would be making a decision based on statistical accuracy tests without taking into account the size of the losses beyond the VaR.

Instead, we will proceed in the next section along four lines: i) the frequency of rejections of the model according to each test when applied to each asset, ii) how often a given test statistic increases in value when switching between two models differing in either the probability distribution or the volatility specification, iii) by a concept of Dominance among VaR models that we introduce below, iv) by implementing a Model Confidence Set approach to select the preferred VaR models for each asset.

2.7.3 Frequency of violations

Tables 2.25 - 2.28 present the the number of violations, the test statistics and p-values of each test for each combination of volatility model and probability distribution for the innovations. Naturally, violation rates close to $\alpha = 0.01$ (13 violations) are desirable. Further, under the Basel Accord, models that over-estimate risk are preferable to those that under-estimate risk levels. We can see that most models with Normal innovations under-estimate the level of risk. In fact, such models are not very acceptable for many assets with more than 20 violations out of 1260 observations, so that the models fall in the yellow zone.[10]

Models with a Student-t distribution for the innovations are not suitable for stock market indices, falling again in the yellow zone, although they show an exact coverage, 13 violations, for IBM and JPY/USD. Models with a skewed Student-t and the unbounded Johnson distributions for the innovations are good models, with less than 20 violations (green zone) for all assets. Models with skewed generalized error, skewed generalized-t and generalized hyperbolic skew Student-t distributions for the innovations are also good models, except for GOLD and SILVER, for which they get more than 20 violations. Median frequency of violations over volatility specifications is 1.75% for models with Normal innovations, 1.27% for Student-t innovations, 1.19% for skewed Student-t, skewed generalized error and skewed generalized-t innovations, 1.11% for Johnson S_U innovations and 0.79% for generalized hyperbolic skew Student-t innovations. According to the frequency of violations, Johnson S_U distribution shows the best behavior among the asymmetric probability distributions.

On the other hand, the frequency of violations for all volatility specifications is relatively similar: 1.19% for GARCH, 1.11% for GJR-GARCH, 1.27% for APARCH and FGARCH models. This observation already suggests the need to be careful when choosing an appropriate probability distribution for return innovations. Needless to say, selecting an appropriate volatility specification is also important, although differences across volatility models might not be so crucial.

2.7.4 Switching between models

For each of the four tests described above (Kupiec, independence, conditional coverage and Dynamic Quantile tests) we compare in this section the values of the test statistics for models that differ in either the probability distribution for the innovations or the volatility

[10] In terms of Basel Accord, based on a sample of 250 observations, if the number of exception is less than, or equal to 4 (the green zone), the test results are consistent with an accurate model, and the possibility of erroneously accepting an inaccurate model is low. At the other extreme, if there are 10 or more exceptions (the red zone), the test results are extremely unlikely to have a resulted from an accurate model, and the probability of erroneously rejecting an accurate model on this basis is remote. In between these two cases, however, is the yellow zone where the backtesting results could be consistent with either accurate or inaccurate models, and the supervisor should encourage a bank to present additional information about its model before taking action.

specification. To summarize the results of this analysis, Table 2.22 displays the number of cases in which the numerical value of the test statistic decreases or increases by a change in the probability distribution or by a change in the specification of the volatility model.

If we consider all the possible specifications sharing the same probability distribution for return innovations, we see that switching from a Normal to a Student-t distribution for return innovations reduces the value of the statistic in 160 out of a total of 216 comparisons, leading in those cases to a more accurate VaR model.[11] Even though such statistics are obviously subject to sampling error, that frequency of reductions in value suggests that, as expected, the Student-t distribution is generally more appropriate than the Normal to represent financial returns. Switching from the symmetric to the skewed Student-t distribution achieves a further reduction in 114 comparisons, while increasing in 75 cases. Moving from the asymmetric Student-t to other asymmetric distributions the unbounded Johnson achieves a reduction in 91 cases while increasing in 55 cases. Switching from the asymmetric Student-t (SKST) to other asymmetric distributions (SGT, JSU, SGED), the numerical value of the statistic decreases more often than otherwise. On the contrary, if we switch from the SKST, SGED, JSU or SGT distributions to the GHST distribution, the opposite happens, with the test statistic usually increasing in value. Hence, we consider the SKST, SGED, JSU and SGT distributions to be preferable to GHST. Between these asymmetric distributions, switching to JSU or SGT lead to a reduction in the test statistic in a greater number of cases.

Among volatility models, switching from the symmetric GARCH to GJR-GARCH reduces the value of the statistic in 176 out of 378 comparisons. The value of the test statistic decreases in 131 cases when switching from GJR-GARCH to APARCH, but it increases in 167 cases. On the other hand, if we move from the APARCH to the FGARCH model, the statistic decreases in 151 out of 378 cases, increasing in 128 cases. Overall, the FGARCH model seems the preferable volatility specification.

Percent differences between the number of cases in which the value of the test statistic increases or decreases when switching between volatility models are not as large as the ones obtained when switching between two probability distributions. That suggests again that, according to the performance of the models for VaR estimation, the specification of the probability distribution of the innovation in returns seems to be more important than the specification of the volatility dynamics. This is consistent with our results in subsection 2.6.3.

[11] The number of comparisons arises from applying all the VaR tests to all the assets. When comparing two probability distribution or two volatility models, the difference between the total number of comparisons considered and the sum of decreases and increases in the test statistic is the number of cases in which the numerical value of the test statistic does not change.

	LR_{uc}		LR_{ind}		LR_{cc}		DQT		**TOTAL**	
Total number statistics	**76**		**32**		**32**		**76**		**216**	
Decreases/Increases	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$
N→ST	64	12	8	24	30	2	58	18	160	56
ST→SKST	45	11	9	20	21	7	39	37	114	75
SKST→JSU	25	6	4	15	16	4	46	30	91	55
SKST→SGT	14	15	6	12	16	2	49	27	85	56
SKST→GHST	33	37	9	19	11	16	32	44	85	116
SKST→SGED	17	16	5	15	14	6	47	29	83	66
SGED→JSU	28	13	8	9	9	8	41	35	86	65
SGED→SGT	6	11	7	2	6	3	52	24	71	40
SGED→GHST	29	38	9	16	8	17	29	47	75	118
JSU→SGT	10	30	12	6	9	9	43	33	74	78
JSU→GHST	22	43	8	17	8	18	25	51	63	129
SGT→GHST	29	29	6	16	2	20	29	47	66	112
Total number statistics	**133**		**56**		**56**		**133**		**378**	
Decreases/Increases	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$	$\Downarrow$	$\Uparrow$
GARCH→GJRGARCH	46	56	36	19	35	21	59	74	176	170
GJRGARCH→APARCH	32	50	25	16	16	26	58	75	131	167
APARCH→FGARCH	34	44	21	13	18	16	78	55	151	128

Table 2.22: Number of cases in which the numerical value of the test statistics decreases or increases when changing probability distribution or changing volatility model for all assets.

2.7.5 Dominance among VaR models

In the previous sections we have used four backtesting tests for VaR performance: the unconditional likelihood-ratio test, the independence test, the conditional coverage test, and the dynamic quantile test, and each test has been run for a variety of models [12] and assets. In this section we evaluate the adequacy of the different models considered by comparing the specific situations in which each model has been rejected by each one of these tests.

We introduce now the concept of dominance: we say that model M1 is dominated by model M2 if i) M1 has been rejected in at least as many cases as M2, and ii) whenever M2 is rejected by a test, M1 is also rejected. This introduces a transitive relationship among VaR models but it is too strong to be satisfied in practice. So, we also consider the concept of p-dominance: Given a confidence level between 0 and 1, we say that model M1 is p-dominated by model M2 if i) M1 has been rejected in at least as many cases

[12] As in the rest of the paper, a model is a combination of a volatility specification and a probability distribution for return innovations.

as M2, and ii) in a percentage of at least p of the cases when M2 is rejected by a test, M1 is also rejected. It is interesting to consider the special case p=1: Two models with exactly the same set of rejections p-dominate each other. They also dominate each other. Unfortunately, p-dominance is not a transitive relationship.

The upper panel of Table 2.23 compares the rejections of models using probability distributions D1 (left) and D2 (right) when combined with all the volatility specifications. The lower panel compares the rejections of models made up with volatility specifications M1 (left) and M2 (right) when combined with all the probability distributions. The first two columns of each panel ($C1$ and $C2$) in Table 2.23 show the number of cases when the two probability distributions D1 and D2 (or volatility models M1 and M2) listed in the first column have been rejected by the data. The third column (p) displays the percentage of rejections of D2 that were also rejections of D1. For instance, the independence test rejected 7 models made up with either the Normal or the Student-t distributions. In 5 of the 7 cases (0.714) when a model with a Student-t distribution was rejected, it was also rejected with a Normal distribution for return innovations. The number of pairwise comparisons is very high because they could be made in both directions, but we show in Table 2.23 the more interesting ones. For instance, we do not explicitly show the comparisons between the Normal distribution and asymmetric distributions because the latter always dominate. Similarly, we do not show pairwise comparisons between Student-t and any asymmetric distribution other than the skewed Student-t because the skewed Student-t tend to p-dominate the standard Student-t, and the majority of asymmetric distributions p-dominate the skewed Student-t distribution[13].

At $\alpha = 95\%$ we can summarize the comparisons over the set of tests as:[14]

$$JSU \succ SGT \succ SGED \succ SKST \succ GHST \succ ST$$

No matter whether we take $\alpha = 99\%$ or $\alpha = 95\%$ the Student-t, SKST and SGED distributions are dominated by other alternatives, specially JSU and SGT. According to this dominance criterion the GHST distribution is judged again not to be appropriate for VaR estimation, since it is dominated by the rest of asymmetric distributions. The Normal distribution is dominated by all other distributions.

At $\alpha = 99\%$ there is not a clear dominance ordering between volatility specifications. For $\alpha = 95\%$ the FGARCH specification seems to dominate.

[13]Even though p-dominance is not transitive it seems safe to focus on the models that tend to be p-dominant.

[14]Looking at specific tests we would have similar orderings characterized with somewhat lower precision:

$$LR_{uc} : JSU \succ SGT = SKST \succ SGED \succ ST$$

$$LR_{ind} : SGT \succ JSU = SGED$$

$$LR_{cc} : JSU = SGT \succ SGED = SKST$$

$$DQT : JSU \succ SGT = SKST \succ SGED \succ ST$$

Confidence level 99%	LR_{uc}			LR_{ind}			LR_{cc}			DQT			TOTAL		
Total number statistics	**76**			**32**			**32**			**76**			**216**		
D1 → D2	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p
N → ST	36	7	1	7	7	0.714	25	13	1	44	29	1	112	56	0.964
ST → SKST	7	0	1	7	6	0.833	13	7	1	29	21	1	56	34	0.971
SKST → JSU	0	0	1	6	4	1	7	4	1	21	21	0.952	34	29	0.966
SKST → SGT	0	1	0	6	5	1	7	6	1	21	21	0.952	34	33	0.939
SKST → SGED	0	1	0	6	6	0.833	7	7	0.857	21	22	0.955	34	36	0.889
SGED → JSU	1	0	1	6	4	1	7	4	1	22	21	1	36	29	1
SGED → SGT	1	1	1	6	5	1	7	6	1	22	21	1	36	33	1
SGT → JSU	1	0	1	5	4	1	6	4	1	21	21	1	33	29	1
GHST → SKST	9	0	1	7	6	0.667	9	7	1	24	21	0.762	49	34	0.794
GHST → SGED	9	1	1	7	6	1	9	7	1	24	22	0.727	49	36	0.833
GHST → JSU	9	0	1	7	4	1	9	4	1	24	21	0.714	49	29	0.793
GHST → SGT	9	1	1	7	5	1	9	6	1	24	21	0.714	49	33	0.818
Total number statistics	**133**			**56**			**56**			**133**			**378**		
M1 → M2	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p
GARCH → GJRGARCH	10	12	0.833	9	9	0.778	16	12	0.917	48	45	0.844	83	78	0.846
GJRGARCH → APARCH	12	14	0.714	9	13	0.615	12	23	0.609	45	46	0.739	78	96	0.688
APARCH → FGARCH	14	18	0.722	13	11	0.818	23	20	0.800	46	43	0.930	96	92	0.848
Confidence level 95%	LR_{uc}			LR_{ind}			LR_{cc}			DQT			TOTAL		
Total number statistics	**76**			**32**			**32**			**76**			**216**		
D1 → D2	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p
N → ST	50	23	0.826	13	13	0.769	32	23	1	52	35	1	147	94	0.926
ST → SKST	23	6	1	13	16	0.813	23	18	1	35	27	0.963	94	67	0.940
SKST → JSU	6	3	1	16	17	0.941	18	17	1	27	25	0.960	67	62	0.968
SKST → SGT	6	6	1	16	16	0.875	18	17	1	27	28	0.964	67	67	0.955
SKST → SGED	6	8	0.750	16	17	0.882	18	18	1	27	28	0.964	67	71	0.930
SGED → JSU	8	3	1	17	17	0.941	18	17	1	28	25	1	71	62	0.984
SGED → SGT	8	6	1	17	16	1	18	17	1	28	28	0.964	71	67	0.985
SGT → JSU	6	3	1	16	17	0.882	17	17	0.941	28	25	1	67	62	0.866
GHST → SKST	21	6	0.833	17	16	0.938	19	18	0.944	30	27	0.926	87	67	0.925
GHST → SGED	21	8	0.875	17	17	0.882	19	18	0.944	30	28	0.929	87	71	0.901
GHST → JSU	21	3	1	17	17	0.882	19	17	0.941	30	25	1	87	62	0.952
GHST → SGT	21	6	1	17	16	0.875	19	17	0.941	30	28	0.929	87	67	0.925
Total number statistics	**133**			**56**			**56**			**133**			**378**		
M1 → M2	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p	C1	C2	p
GARCH → GJRGARCH	23	30	0.700	32	30	0.867	40	37	0.865	56	51	0.863	151	148	0.831
GJRGARCH → APARCH	30	33	0.758	30	25	0.960	37	36	0.972	51	59	0.797	148	153	0.856
APARCH → FGARCH	33	31	0.968	25	22	0.955	36	32	1	59	59	0.915	153	144	0.951

Table 2.23: Dominance among VaR models. $C1$ is the set consisting of the number of times H0 is rejected with D1/M1 for the different assets, $C2$ is the set consisting of the number of times H0 is rejected with D2/M2 for the different assets and p is the proportion of times that H0 is rejected with D2/M2 and and also rejected with D1/M1.

A preference for APARCH and FGARCH models against standard GARCH and GJR-GARCH has been a constant throughout our analysis. So, a robust conclusion is the need to incorporate a leverage effect in volatility and, possibly more important, the convenience to model standard deviations, rather than variances. The preference for asymmetric probability distributions in Table 2.23 is also consistent with results in Table 2.22 when comparing the numerical values of the test statistics. Both analysis are based on the same information, but they use it in a very different fashion[15]. Nothing guarantees that the con-

[15]Switching between models we compared the values of the test statistics under a continuous criterion,

clusions on the preferred probability distributions should be the same in both analysis. On the contrary, this coincidence should be seen as a proof of the robustness of such preference.

The interesting feature of this dominance criterion is that it compares any two model specifications across all the statistical tests and assets. The criterion could accommodate different weights to each asset and test depending on the relevance we want to assign them. The dominance criterion would then be applied to such weighted sums, as representing the number of rejections, weighted by relevance. An interesting possibility would consist of assigning a larger weight to tests having a larger ability to discriminate among models.

A further variation of the dominance criterion would choose weights as a bounded function of the size of the test rejection, either in terms of the test statistic or the p-value of the test.

2.7.6 Loss functions

Tables 2.29 - 2.33 present the values of the QLF and AlTick loss functions for different models and assets. Including a leverage effect in volatility reduces the AlTick loss function with independence of the assumption on the probability distribution of innovations, except for interest rates and GOLD and SILVER. Indeed, the accuracy of the GARCH specification falls well below that of GJR-GARCH, APARCH and FGARCH models.

For these three volatility specifications there is some reduction in the loss function when switching from a Normal to a Student-t distribution. The reduction achieved by the skewed t-Student distribution over the symmetric Student-t distribution is again not so evident [16]. On the contrary, there is a noticeable improvement when we move from symmetric to asymmetric distributions in terms of formal tests and also in terms of loss functions. VaR models under an unbounded Johnson, skewed Generalized Error, Skewed Generalized-t and Generalized Hyperbolic Skew Student-t distributions significantly reduce the value of the loss function relative to the Skewed Student-t with independence of the assumption on the volatility model.

Hence, our results suggest that it is the explicit consideration of the skewness in the probability distribution of innovations that is truly important for VaR performance according to the loss function criterion.

2.7.7 Model Confidence Sets

The availability of several model specifications being able to adequately describe the unobserved data generating process (DGP) opens the question of selecting the 'best fitting model' according to a given optimality criterion. Recently, significant effort has been placed on developing testing procedures being able to deliver the 'best fitting' models

while in this section we compare the results of the tests under alternative models in a discrete way.

[16] See OIL, GAS and exchange rates.

among a set of alternatives. One of the first proposals was Diebold & Mariano (1995) [33], but it is not applicable when the forecasts come from nested models or when they depend on semiparametric or non parametric estimators (Giacomini & Komunjer, 2005 [46]). This has been overcome by the Reality Check (RC) approach of White (2000) [126], the Stepwise Multiple Testing procedure of Romano and Wolf (2005) [111], the Superior Predictive Ability (SPA) test of Hansen and Lunde (2005) [60], the Conditional test of Giacomini and White (2006) [47], and the Model Confidence Set (MCS) procedure developed by Hansen, Lunde and Nason (2011) [62]. All these tests are relevant from an empirical point of view, especially when the set of competing alternatives is large.

We implement the Model Confidence Set (MCS) procedure developed by Hansen, Lunde and Nason (2011) [62] to discriminate among models. The MCS procedure is a general approach to model selection that neither assumes knowledge of the correct specification, nor does it require that the "true" model is available as one of the competing models. Another advantage is that MCS does not discard a model unless it is found to be significantly inferior relative to other models [17]. The MCS has an interpretation similar to a confidence interval for a parameter in the sense that, with a given level of confidence, a MCS contains the best model. It is an appealing method to use when comparing a set of forecasting models because in practice it often cannot be ruled out that two or more competing models are equally good. In this sense, the MCS approach may be preferred over methods that require a single model to be selected as "best model".

The MCS procedure consists on a sequence of tests to construct the 'Set of Superior Models' (SSM) where the null hypothesis of Equal Predictive Ability (EPA) is not rejected at a certain confidence level, while such set is characterized as having better predictive ability than models not in the set. The EPA test statistic is evaluated under a given loss function, which essentially means that it is possible to test models on various aspects depending on the chosen loss function. The possibility of user supplied loss functions provides enough flexibility to the procedure that can be used to test competing models with respect to different dimensions. This is in common with Diebold & Mariano (1995) [33], although we are here not so much interested on whether VaR forecasts are significantly different, but rather on whether the number and size of VaR violations are different across models.

Formally, the loss function $\ell_{i,t}$ associated to the i-th model $\ell_{i,t} = \ell(Y_t, \hat{Y}_{i,t})$ measures the cost produced by the difference between the observation at time t, Y_t, and $\hat{Y}_{i,t}$ the output of model i at time t. The MCS procedure starts from an initial set of models $\hat{M}^0$ of dimension m made up by all combinations of probability distribution and volatility specification considered in previous sections. Then, for a given confidence level $1-\alpha$, we obtain a smaller set, the superior set of models, SSM, $\hat{M}^*_{1-\alpha}$ of dimension $m^* \leq m$. Let us denote by d_{ij} the loss differential between models i and j,

$$d_{ij,t} = \ell_{i,t} - \ell_{j,t} \qquad i,j = 1,...,m, \qquad t = 1,...,n,$$

[17]In this respect it is clearly different from the two-stage approach to model selection we described in subsection 2.7.2.

The EPA hypothesis for a given set of models M can be formulated:[18]

$$H_{0,M} : c_{ij} = 0, \qquad \text{for all} \qquad i, j = 1, ..., m$$

$$H_{1,M} : c_{ij} \neq 0, \qquad \text{for some} \qquad i, j = 1, ..., m$$

where $c_{ij} = \mathbb{E}(d_{ij})$ is assumed to be finite and not time dependent. This hypothesis can be tested using the test statistic [Hansen et al. (2011)],

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \qquad \text{for} \qquad i, j \in M$$

where $\bar{d}_{ij} = n^{-1} \sum_{t=1}^{n} d_{ij,t}$ measures the relative sample loss between the i-th and j-th models, while $\widehat{var}(\bar{d}_{ij})$ is a bootstrapped estimate of $var(\bar{d}_{ij})$. According to Hansen et al. (2011), to calculate the bootstrapped variances, we perform a block-bootstrap procedure [19] of 10000 resamples, where the block length p is the maximum number of significant parameters obtained by fitting an AR(p) process on all the d_{ij} terms, in our case $p = 1$.

As discussed in Hansen et al. (2011) [62] the EPA null hypothesis maps naturally into the statistic,

$$T_{R,M} = \max_{i,j \in M} |t_{ij}|$$

Since the asymptotic distributions of this test statistic is nonstandard, the relevant distribution under the null hypothesis needs to be estimated using a bootstrap procedure

[18]An alternative test is: Let $d_{i\cdot,t} = (m-1)^{-1} \sum_{j \in M} d_{ij,t} \quad i = 1, ..., m$ the simple average loss of model i relative to any other model j at time t, the EPA hypothesis for a given set of models M is

$$H_{0,M} : c_{i\cdot} = 0, \qquad \text{for all} \qquad i, j = 1, ..., m$$

$$H_{1,M} : c_{i\cdot} \neq 0, \qquad \text{for some} \qquad i, j = 1, ..., m$$

where $c_{i\cdot} = \mathbb{E}(d_{i\cdot})$ is assumed to be finite and not time dependent. The statistic is,

$$t_{i\cdot} = \frac{\bar{d}_{i\cdot}}{\sqrt{\widehat{var}(\bar{d}_{i\cdot})}} \qquad \text{for} \qquad i, j \in M$$

where $\bar{d}_{i\cdot} = (m-1)^{-1} \sum_{j \in M} \bar{d}_{ij}$ is the simple loss of the i-th model relative to the average losses across models in the set M, while $\widehat{var}(\bar{d}_{i\cdot})$ is bootstrapped estimates of $var(\bar{d}_{i\cdot})$.
The EPA null hypothesis is $T_{max,M} = \max_{i \in M} t_{i\cdot}$ and the relevant distribution under the null hypothesis need to be estimated using a bootstrap procedure similar to that used to estimate $var(\bar{d}_{i\cdot})$ since the asymptotic distribution of the test statistic is nonstandard. The choice of the worst model to be eliminated is coherent with the test statistic:

$$e_{max,M} = arg \max_{i \in M} \frac{\bar{d}_{i\cdot}}{\widehat{var}(\bar{d}_{i\cdot})}$$

[19]The block-bootstrap is the most general method to improve the accuracy of bootstrap for time series data. By dividing the data into several blocks, it can preserve the original time series structure within a block. However the accuracy of the block-bootstrap is sensitive to the choice of block length, and the optimal block length depends on the sample size, the data generating process, and the statistic considered (see Goncalves and White, 2004 [55], 2005 [56], Künsch, 1989 [77], Liu and Singh, 1992 [82] [83] and Politis and Romano, 1994 [107]). Details about the implemented bootstrap procedure can be found in White (2000) [126], Kilian (1999) [74], Clark and McCracken (2001) [28], Hansen et al. (2003) [61], Hansen and Lunde (2005) [60], Hansen et al. (2011) [62] and Bernardi et al. (2016) [18].

similar to that used to estimate $var(\bar{d}_{ij})$.

The MCS is a sequential testing procedure that eliminates at each step the worst model, until the hypothesis of equal predictive ability (EPA) is not rejected for any of the models in the current SSM. The choice of the worst model to be eliminated has been made using an elimination rule that is coherent with the statistic test which is

$$e_{R,M} = arg \max_i \left\{ sup_{j \in M} \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \right\}$$

Table 2.24 reports the frequency by which each probability distribution and each volatility specification enter into the Superior Set of Models for each asset using the AlTick loss function [20]. Tests are performed at the 90% confidence level, using a block-bootstrap procedure of 10000 resamples with a block length of 1. The table shows that for some assets the SSM with AlTick function is quite homogeneous with respect to the volatility and probability distribution assumptions, specially for NASDAQ 100, FTSE 100 and US BOND in distributions and for EUR/USD in volatility models and probability distributions. In those cases the one step ahead 1% VaR forecasting performance of the competing combinations are quite similar, suggesting that for those series the use of complicated non-linear combinations is not entirely justified. Among the volatility models FGARCH seems to describe very well the financial time series behaviour. Concerning the distribution specifications, we observe that the MCS confirms the common finding that the Gaussian distribution provides a poor description of the behavior of financial time series. Under the AlTick loss, the Skewed Generalized-t and Skewed Error Distributions perform better than the Generalized Hyperbolic Skew Student-t. Definitely, Gaussian, Student-t and Skewed Student-t distributions do not seem to be appropriate for the wide set of financial assets considered in this paper. Regarding QLF loss function, the results obtained with SSM are not homogeneous. Surprisingly there is a clear preference for GHST distribution and for GJR-GARCH volatility model. This is due to that QLF only considers the underestimation of risk, and GHST tends to overestimate it. For this reason, we obtain lower QLF loss functions with this distribution. We observe that the conclusions are different and depend on we focus only on the magnitude of failure or also on the opportunity cost of capital. The QLF is preferred by regulators which are concerned about the underestimation of the risk and AlTick is preferred by firms which have a conflict between the goal of safety and the goal of profit maximization.

[20] Care must be exerted when choosing the loss function, and it might be worthwhile to explore other functions that might focus on different characteristics of VaR estimates. We believe that the opportunity cost of overestimating VaR is non trivial, and therefore, the AlTick loss function is to be preferred over the QLF loss.

QLF	IBEX	NASDAQ	FTSE	NIKKEI	IBM	SAN	AXA	BP	IRS	GER BOND	US BOND	BRENT	GAS	GOLD	SILVER	EUR/USD	GBP/USD	JPY/USD	AUD/USD	TOTAL
											Models									
GARCH	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	**4**
GJRGARCH	1	1	1	1	0	1	0	1	0	1	1	1	1	0	0	1	1	1	1	**14**
APARCH	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**3**
FGARCH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
											Distributions									
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
ST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
SKST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
SGED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
JSU	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	**3**
SGT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
GHST	3	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	**18**
Total Number	**3**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	
AlTick	**IBEX**	**NASDAQ**	**FTSE**	**NIKKEI**	**IBM**	**SAN**	**AXA**	**BP**	**IRS**	**GER BOND**	**US BOND**	**BRENT**	**GAS**	**GOLD**	**SILVER**	**EUR/USD**	**GBP/USD**	**JPY/USD**	**AUD/USD**	**TOTAL**
											Models									
GARCH	0	0	0	0	1	0	0	0	1	2	6	0	0	1	4	2	0	3	0	**20**
GJRGARCH	0	2	2	1	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	**9**
APARCH	2	5	5	0	0	0	0	2	0	0	0	0	0	0	0	2	0	4	2	**22**
FGARCH	3	4	3	2	0	1	1	0	0	3	1	1	1	0	1	6	0	4	1	**32**
											Distributions									
N	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	2	**6**
ST	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	3	1	**8**
SKST	0	2	2	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	**7**
SGED	2	2	2	2	1	0	1	1	0	1	1	0	0	0	0	1	0	3	0	**17**
JSU	0	2	3	0	0	1	0	0	0	2	1	0	0	0	2	1	0	0	0	**12**
SGT	2	3	3	1	0	0	0	0	1	1	1	0	0	0	1	4	0	3	0	**20**
GHST	1	1	0	0	0	0	0	0	0	1	2	0	0	1	2	3	0	2	0	**13**
Total Number	**5**	**11**	**10**	**3**	**1**	**1**	**1**	**2**	**1**	**5**	**7**	**1**	**1**	**1**	**6**	**12**	**1**	**11**	**3**	

Table 2.24: Composition of remaining probability distribution-volatility model combinations in the Superior Set for each asset discriminated by model and distribution according QLF loss function (top panel) and AlTick loss function (bottom panel).

2.8 Conclusions

This paper completes previous work on the forecasting performance of alternative VaR models by considering four volatility specifications: GARCH, GJR-GARCH, APARCH and FGARCH and a set of distributions including skewed Student-t, skewed generalized error, unbounded Johnson, skewed generalized-t and generalized hyperbolic skew Student-t distributions, some of them not widely used yet in the literature. Standard symmetric distributions and GARCH models without leverage are also used as a benchmark. We employed data covering the recent financial crisis of 2007-2009 for assets of different nature.

Two clear results refer to issues that have been analyzed in previous research by a number of authors: *i*) VaR models that assume asymmetric probability distributions for the innovations, like the Skewed Student-t distribution, Skewed Generalized Error distribution, Johnson S_U distribution, and Skewed Generalized-t distribution provide a better fit of the sample return moments than symmetric distributions and achieve better VaR performance, *ii*) volatility models with leverage, like APARCH and FGARCH, show a better VaR performance than more standard GARCH and GJR-GARCH volatility specifications.

Our results highlight other important issues. A third result is that the shape and the skew of the assumed probability distribution for innovations are more important than including a leverage effect in volatility for the performance of a Value-at-Risk model. This corroborates results by other authors (Lopez and Walter, 2000 [87], Angelidis and Degiannakis, 2006 [11], and Braione and Scholtes, 2016 [20]) suggesting that the assumption on the probability distribution is more important than the chosen volatility specification. We provide a thorough analysis of that issue by showing that for the wide set of assets considered: *i*) different volatility models with the same probability distribution for the innovations fit sample return moments similarly, *ii*) the frequency of rejections of VaR tests in models that differ in their volatility specification are similar, while rejection frequencies among models with the same volatility specification but different probability distribution for the innovations can differ very significantly, *iii*) changing the probability distribution in a VaR model affects the numerical value of the statistic for VaR tests much more than changing the volatility specification, and *iv*) the dominance criterion establishes a clear ranking between models differing in their probability distribution.

A fourth result deals with the fact that if the true, unobserved volatility dynamics is not in terms of squared conditional standard deviations, then models specified for the conditional variance are prone to produce biased results. We believe that by dealing with the power of the conditional standard deviation as a free parameter is an important feature of the APARCH/FGARCH volatility specifications. In fact, our estimates suggest that for a number of financial assets the squared conditional deviation specification is inappropriate.

Fifth, our analysis suggests that, as expected, a good fit of the moments of the distribution of returns usually leads to a good VaR performance. The MAE calculated over estimates for the four first moments selects the combination of a Skewed Generalized Error distribution and an APARCH/FGARCH volatility specification as the best model to reproduce the skewness and kurtosis in asset returns. According to VaR performance, the results obtained are similar. The use of a dominance criterion introduced in this paper on the results of backtesting tests suggests that Johnson S_U, Skewed Generalized-t and Skewed Generalized Error distributions dominate over other asymmetric distributions, like Skewed Student-t and Generalized Hyperbolic Skew Student-t, and symmetric distributions, like Student-t and Normal distributions. FGARCH seems the preferable volatility model. If we consider the AlTick loss function, the Skewed Generalized-t and Skewed Error distributions perform better than the other distributions in terms of the Model Confidence Sets procedure. Among the volatility models FGARCH seems again to describe well the financial time series behavior.

Finally, we have examined in the paper whether alternative VaR models provide different evidence in VaR performance for assets of different nature. FGARCH volatility model with Skewed Generalized Error, Skewed Generalized-t and Johnson S_U distributions are the most suitable for stock market indices and individual stocks. The Generalized Hyperbolic Skew Student-t seems to perform well for interest rate and exchange rates, and the models combining APARCH or FGARCH volatility specifications and symmetric distributions or simple volatility models with asymmetric distributions are good combinations for commodities.

Bibliography

[1] Aas, K. and Haff, I.H., 2006 *The Generalized Hyperbolic Skew Student's t-Distribution.* Journal of Financial Econometrics, Vol.4, No. 2, pp. 275-309.

[2] Abad, P., Benito, S., 2012. *A detail comparison of value at risk estimates.* Mathematics and Computers in Simulation, Vol. 94, pp. 258-276.

[3] Abad, P., Benito, S. and Lopez, C., 2015. *Role of the loss function in the VaR comparison.* Journal of Risk Model Validation, Vol. 9, No. 1, pp. 1-19.

[4] Abad, P., Benito, S., Sanchez-Granero M.A. and Lopez, C., 2013. *Evaluating the Performance of the Skewed distributions to forecast Value at Risk in the Global Financial Crisis.* Available at E-prints: http://eprints.ucm.es/23999/1/1340

[5] Abad, P., Benito, S. and Lopez, C., 2012. *A comprehensive review of Value-at-Risk methodologies.* The Spanish Review of Financial Economics, Vol. 12, pp. 15-32.

[6] Alberg, D., Shalit, H. and Yosef, R., 2008. *Estimating Stock Market Volatility using Assymetric Garch Models.* Applied Financial Economics, Vol. 18, pp. 1201-1208.

[7] Alexander, C. and Lazard, E., 2006.*Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling.* Journal of Applied Econometrics, Vol. 21, pp. 307-336.

[8] Andersen, T., Bollerslev, T., Diebold, F. and Labys, P., 2003. *Modelling and forecasting realized volatility.* Econometrica, Vol. 71, pp. 529-629.

[9] Ane, T., 2006. *An analysis of the flexibility of Asymmetric Power GARCH models.* Computational Statistics and Data Analysis, Vol. 51, pp. 1293-1311.

[10] Angelidis, T., Benos, A. and Degiannakis, S., 2007. *A robust VaR model under different time periods and weighting schemes.* Review of Quantitative Finance and Accounting, Vol. 28, pp. 187-201.

[11] Angelidis, T and Degiannakis, S., 2006. *Backtesting VaR Models: An Expected Shortfall Approach.* Available at SSRN 898473

[12] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., 1998. *Coherent measures of risk.* Mathematical Finance, Vol. 9, pp. 203-228.

[13] Azzalini, A., and Capitanio, A., 2003. *An asymmetric Generalization of Gaussian and Laplace Laws.* Journal of the Royal Statistical Society B, Vol. 65, pp. 579-602.

[14] Bali, T. and Theodossiou, P., 2007. *A conditional-SGT-VaR approach with alternative GARCH model.* Annals of Operations Research, Springer.

[15] Barndorff-Nielsen, O.E., 1997. *Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling.* Scandinavian Journal of Statistics, Vol. 24, pp. 1-14.

[16] Bauwens, L., Giot, P., Grammig, J. and Veredas, D., 2004. *A comparison of financial duration models via density forecasts.* International Journal of Forecasting, Vol. 20, pp. 589-609.

[17] Berkowitz, J., Christoffersen, P. and Pelletier, D, 2011. *Evaluating Value-at-Risk models with desk-level data.* Management Science, Vol. 57, No. 2. pp. 2213-2227.

[18] Bernardi, M., Catania, L. and Petrella, L., 2016. *Are News Important to Predict the Value-at-Risk?.* The European Journal of Finance. URL http://dx.doi.org/10.1080/1351847X.2015.1106959

[19] Bollerslev, T., 1986. *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics, Vol. 31, pp. 307-327.

[20] Braione, M. and Scholtes, N.K., 2016. *Forecasting Value-at-Risk under Different Distributional Assumptions.* Econometrics, Vol. 4, No. 3.

[21] Brooks, C. and Persand, G., 2003. *Volatility Forecasting for Risk Management.* Journal of Forecasting, Vol. 22, pp. 1-22.

[22] Bubak, V., 2008. *Value-at-Risk on Central and Eastern European stock markets: An empirical investigation using GARCH models.* Institute of Economics Studies, Charles University, No. 18.

[23] Caporin, M., 2008. *Evaluating value-at-risk measures in the presence of long memory conditional volatility.* The journal of Risk, Vol. 10, No. 3, pp. 79-110.

[24] Chen, C.W.S., Gerlach, R. Hwang, B.B.K. and McAleer, M., 2012. *Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range.* International Journal of Forecasting, Vol. 28, No. 3, pp. 557-574.

[25] Chen, S.X. and Tang, C.Y., 2005. *Nonparametric inference of value-at-risk for dependent financial returns.* Journal of Financial Econometrics, Vol. 3, No. 2, pp. 227-255.

[26] Choi, P., and Nam, K., 2008. *Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the SU-normal distribution.* Journal of Empirical finance, Vol. 15, No.1, pp. 41-63.

[27] Christoffersen, P., 1998. *Evaluating internal Forecasting.* International Economic Review. Vol. 39, pp. 841-862.

[28] Clark, T.E. and McCraken, M.W., 2001. *Test of equal forecast accuracy and encompassing for nested models.* Journal of Econometrics, Vol. 105, No. 1, pp. 85-110.

[29] Corlu, C.G., Meterelliyoz, M. and Tiniç, M, 2016. *Emirical distributions of daily equity index returns: A comparison.* Expert System with Applications, Vol. 54, pp. 170-192.

[30] De Rossi, G. and Harvey, A.C., 2009. *Quantiles, expectiles and splines.* Journal of Econometrics, Vol. 152, pp. 179-185.

[31] Diamandis, P.F., Drakos, A.A., Kouretas, G.P. and Zarangas, L., 2011. *Value-at-Risk for long and short trading positions: Evidence from developed and emerging equity markets.* International Review of Financial Analysis, Vol. 20, pp. 165-176.

[32] Diebold, F.X., Gunther, T.A. and Tay, A.S., 1998. *Evaluating Density Forecasts, with Applications to Financial Risk Management.* International Economic Review, Vol. 39, pp. 863-883.

[33] Diebold, F.X. and Mariano, R. S., 1995. *Comparing predictive accuracy.* Journal of Business and Economic Statistics, Vol. 13, No. 3, pp. 253-263.

[34] Diebold, F.X., Schuermann, T. and Stroughair, J.D., 2000. *Pitfalls and opportunities in the use of extreme value theory in risk management.* The Journal of Risk Finance, Vol. 50, pp. 264-272.

[35] Ding, Z., Granger, C.W.J. and Engle, R.F., 1993. *A long memory property of stock market returns and a new model.* Journal of Empirical Finance, Vol.1, pp. 83-106.

[36] Drakos, A.A., Kouretas, G.P. and Zarangas, L., 2015.*Predicting Conditional Autoregressive Value-at-Risk for Stock Markets during Tranquil and Turbulent Periods.* Journal of Financial Risk Management, 2015, Vol. 4, pp. 168-186.

[37] El Babsiri, M. and Zakoian, J. M., 2001. *Contemporaneous asymmetry in GARCH processes.* Journal of Econometrics, Vol. 101, pp. 257-294.

[38] Engle R.F. and Bollerslev, T., 1986. *Modelling the Persistence of Conditional Variances.* Econometric Reviews, Vol. 5, pp. 1-50.

[39] Engle R.F. and Manganelli, S., 2004. *CAViaR: conditional autoregressive value at risk by regression quantiles.* Journal of Business & Economic Statistics, Vol. 22, pp. 367-381.

[40] Ergün, A., Jun, J., 2010. *Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall.* Quarterly Review of Economics and Finance, Vol. 50, pp. 264-272.

[41] Fernandez, C. and Steel, M., 1998. *On Bayesian Modelling of Fat Tails and Skewness.* Journal of the American Statistical Association, Vol. 93, No. 441, pp. 359-371.

[42] Filliben, J.J., 1975. *The probability plot correlation coefficient test for normality.* Technometrics, Vol. 17, No. 1, pp. 111-117.

[43] Gerlach, R.H., Chen, C.W.S. and Chan, N.Y.C, 2011. *Bayesian time varying quantile forecasting for value-at-risk in financial markets.* Journal of Business and Economic Statistics, Vol. 29, No. 4, pp. 481-492.

[44] Gerlach, R., Chen, C.W.S., Lin, E.M.H. and Lee, W.C.W., 2011. *Bayesian Forecasting for Financial Risk Management, Pre and Post the Global Financial Crisis.* Journal of Forecasting, Vol. 31, No. 8, pp. 661-687.

[45] Geweke, J., 1986. *Modeling the persistence of conditional variances: a comment.* Econometric Review, Vol. 5, pp. 57-61.

[46] Giacomini, R. and Komunjer, I., 2005. *Evaluation and Combination of Conditional Quantile Forecasts.* Journal of Business and Economic Statistics, Vol. 23, No. 4, pp. 416-431.

[47] Giacomini, R. and White, H., 2006. *Tests of conditional predictive ability.* Econometrica, Vol. 74, No. 6, pp. 1545-1578.

[48] Giot, P. and Laurent, S., 2003a. *Value-at-Risk for Long and Short Trading Positions.* Journal of Applied Econometrics, Vol. 18, pp. 641-664.

[49] Giot, P. and Laurent, S., 2003b. *Market Risk in commodity markets: a VaR approach.* Energy Economics, Vol. 25, pp. 435-457.

[50] Glosten, L., Jagannathan R. and Runkle, D., 1993. *On the relation between the expected value and the volatility of the nominal excess return on stocks.* Journal of Finance, Vol. 48, pp. 1779-1801.

[51] Gomes, M.I., de Haan, L. and Rodrigues, L.H., 2008. *Tail index estimation for heavy-tailed models: Accomodation of bias in weighted log-excesses.* Journal of the Royal Statistical Society, Series B, Vol. 70, No. 1, pp. 31-52.

[52] Gomes, M.I., Figueiredo, F., Rodrigues, L.H. and Miranda, M.C., 2012. *A computational sturdy of a quasi-PORT methodology for VaR based on second-order reduced-bias estimation.* Journal of Statistical Computation and Simulation, Vol. 82, No. 4, pp. 587-602.

[53] Gomes, M.I., Matins, M.J. and Neves, M.M., 2007. *Improving second order reduced-bias tail index estimator.* Review of Statistics, Vol. 5, No. 2, pp. 177-207.

[54] Gomes, M.I. and Pestana. D., 2007. *A sturdy reduced bias extreme quantile (VaR) estimator.* Journal of the american Statistical Association, Vol. 102, pp. 477, pp. 280-292.

[55] Goncalves, S. and White, H., 2004. *Maximum likelihood and the bootstrap for nonlinear dynamics models.* Journal of Econometrics, Vol. 119, pp. 199-219.

[56] Goncalves, S. and White, H., 2005. *Bootstrap standard error estimates for linear regressions.* Journal of the American Statistical Association, Vol. 100, No. 471, pp. 970-979.

[57] Gourieroux, C. and Jasiak, J., 2008. *Dynamic quantile models*. Journal of Econometrics, Vol. 147, pp. 198-205.

[58] Gourieroux, C. and Jasiak, J., 2010. *Value-at-risk*. Handbook of financial econometrics, Vol. 1, pp. 553-615.

[59] Hansen, B., 1994. *Autorregressive conditional density estimation*. International Economic Review, Vol. 35, pp. 705-730.

[60] Hansen, P.R. and Lunde, A., 2005. *A forecast comparison of volatility models: does anything beat a GARCH(1,1)?*. Journal of Applied Econometrics, Vol. 20, No. 7, pp. 873-889.

[61] Hansen, P.R., Lunde, A. and Nason J.M., 2003. *Choosing the best volatility models: The model confidence set approach*. Oxford Bulletin of Economics and Statistics, Vol. 65, No. s1, pp. 839-861.

[62] Hansen P.R., Lunde A. and Nason, J.M., 2011. *The model confidence set*. Econometrica, Vol. 79, No. 2, pp. 453-497.

[63] Hentschel, L., 1995. *All in the family Nesting symmetric and asymmetric GARCH models*. Journal of Financial Economics, Vol. 39, pp. 71-104.

[64] Higgins, M.L. and Bera, A.K., 1992. *A class of Non Linear Arch Models*. International Economic Review, Vol. 33, No. 1, pp. 137-158.

[65] Hill, B.M., 1975. *A simple general approach to inference about the tail of a distribution*. The Annals of Statistic, Vol. 3, No. 5, pp. 1163-1174.

[66] Hu, W., 2017. *Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk*. Dissertation in the Florida State University.

[67] Huang, D., Yu, B, Luz, Z, Fabozzi, F., Focardi, S. and Fukushima, M., 2010. *Index-exciting CAViaR: a new empirical time-varying risk model*. Studies in Nonlinear Dynamics and Econometrics, Vol. 14, No. 1.

[68] Huisman, R., Koedijk, K.G., Kool, C.J. and Palm, F., 2001. *Tail index estimates in small samples*. Journal of Business and Economic Statistics, Vol. 19, pp. 208-216.

[69] Jeon, J. and Taylor, J.W, 2013. *Using CAViaR models with implied volatility for value-at-risk estimation*. Journal of Forecasting, Vol. 32, pp. 62-74.

[70] Johnson, N.L., 1949. *Systems of frequency curves generated by methods of translations*. Biometrika, Vol. 36, pp. 149-176.

[71] Joiner, B.L. and Rosenblatt, J.R., 1971. *Some properties of the range in samples from Tukey's symmetric lambda distributions*. Journal of the American Statistical Association, Vol. 66, No. 334, pp. 394-399.

[72] Jondeau, E. and Rockinger, M., 2003. *Conditional volatility, skewness and kurtosis: existence, persistence and comovements*. Journal of Economic Dynamics and Control, Vol. 27, pp. 1699-1737.

[73] Kang, S.H. and Yong S-M., 2009. *Value-at-Risk analysis for Asian emerging markets: asymmetry and fat tails in returns innovation*. The Korean Economic Review, Vol. 25, No. 387-411.

[74] Kilian, L., 1999. *Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?*. Journal of Applied Econometrics, Vol. 14, No. 5, pp. 491-510.

[75] Kim, M. and Lee, S., 2015. *Comparison of semiparametric methods to estimate VaR and ES*. Korean Journal of Applied Statistics, Vol. 29, No.1, pp. 171-180

[76] Kolmogorov, A., 1933. *Sulla Determinazione Empirica di una Leggi di Distribucione*. Giornalle dell' Istitute Italiano degli Attuari, Vol. 4, pp. 1-11

[77] Künsch, H.R., 1989. *The jackknife and the bootstrap for general stationary observations*. Annals of Statistics, Vol. 17, pp. 1217-1241.

[78] Kupiec, P., 1995. *Techniques for Verifying the Accuracy of Risk Measurement models*. Journal of Derivatives, Vol. 2, pp. 174-184.

[79] Lambert, P. and Laurent, S., 2002. *Modelling Skewness Dynamics in Series of Financial Data using Skewed location-scale distribution*. Discussion Paper. Institut de Statistique, Louvain-la-Neuve.

[80] Lambert, P. and Laurent, S., 2001. *Modelling Financial Time Series using GARCH-type models with a skewed student distribution for the innovations*. Mimeo, Université de Liege.

[81] Lee, C.F. and Su, J.B., 2015. *Value-at-Risk Estimation via a Semi-Parametric Approach: Evidence from the Stock Markets*. Handbook of Financial Econometrics and Statistics, Springer Science Business Media New York.

[82] Liu, R.Y. and Singh, K., 1992a. *Moving blocks jackknife and bootstrap capture weak dependence*. LePage and Billiard (Eds), *Exploring the limits of the bootstrap*, Wiley, New York.

[83] Liu, R.Y. and Singh, K., 1992b. *Efficiency and robustness in resampling*. Annals of Statistics, Vol. 20, pp. 370-384.

[84] Longin, F., 2005. *The choice of the distribution of asset returns: How extreme value theory can help?* Journal of Banking and Finance, Vol. 29, pp. 1017-1035.

[85] Lopez, J.A., 1998. *Testing your risk tests*. Financial Survey (May-Jun), pp. 18-20.

[86] Lopez, J.A., 1999. *Methods for evaluating Value-at-Risk estimates*. Federal Reserve Bank of San Francisco Economic Review, Vol. 2, pp. 3-17.

[87] Lopez, J.A. and Walter, C.A., 2000. *Evaluating Covariance Matrix Forecasts in a Value-at-Risk Framework*. Available at SSRN 305279

[88] Mabrouk, S. and Saadi, S., 2012. *Parametric Value-at-Risk analysis: Evidence from stock indices*. The Quarterly Review of Economics and Finance, Vol. 52, pp. 305-321.

[89] Massey, F. J., 1951. *The Kolmogorov-Smirnov test for goodness of fit*. Journal of the American Statistical Association, Vol. 46, pp. 68-78.

[90] McDonald, J. and Xu, Y., 1995. *A generalization of the Beta distribution with applications*. Journal of Econometrics, Vol. 66, pp. 133-152.

[91] McMillan, D.G. and Kambourodis, D., 2009. *Are RiskMetrics forecasts good enough? Evidence from 31 stock markets*. International Review of Financial Analysis, Vol. 18, pp. 117-124.

[92] McMillan, D.G. and Speight, A.E.H., 2007. *Value-at-Rik in emerging equity markets: Comparative evidence for simmetric, asymmetric and long-memory GARCH models*. International Review of Finance, Vol. 7, pp. 1-19.

[93] McNeil, A. and Frey, R., 2000. *Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach*. Journal of Empirical Finance, Vol. 7, pp. 271-300.

[94] Mina, J. and Ulmer, A., 1999. *Delta-Gamma four ways*. Technical report, RiskMetrics Group, pp. 1-17.

[95] Mittnik, S. and Paolella, M.S., 2000. *Prediction of Financial Downside-Risk with Heavy-Tailed Conditional Distributions*. Available at SSRN 391261.

[96] Nagai, M., 2016. *Estimation of Extreme Value at Risks Using CAViaR models*. Available at: http://www.econ.hit-u.ac.jp/ finmodel/student/15Nagai.pdf

[97] Nakajima, J. and Omori, Y., 2012. *Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution*. Computational Statistics & Data Analysis, Vol. 56, No.11, pp. 3690-3704.

[98] Nelson, D.B., 1991. *Conditional Heteroskedasticity in Asset Returns: A New Approach*. Econometrica, Vol. 59, No. 2, pp. 347-370.

[99] Pantula, S., 1986. *Modeling the persistence in conditional variances: a comment*. Econometric Review, Vol. 5, pp. 71-74

[100] Paolella, M., 1997. *Tail Estimation and Conditional Modeling of Heteroskedstic Time-Series*. Ph.D Thesis, Institute of Statistics and Econometrics, Christian Albrechts University of Kiel.

[101] Paolella, M. S. and Polak, P., 2015. *COMFORT: A common market factor non-Gaussian returns model*. Journal of Econometrics, Vol. 187, No. 2, pp. 593-605.

[102] Pearson, K., 1990. *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.* Philosophy Magazine Series (5), Vol. 50, pp. 157-172.

[103] Pérignon, C. and Smith, D.R., 2008. *A new approach to comparing VaR estimation method.* The Journal of Derivatives, Vol. 16, No. 2, pp. 54-66.

[104] Pérignon, C. and Smith, D.R., 2010b. *The level and quality of Value-at-Risk disclosure by commercial banks.* Journal of Banking and Finance, Vol. 34, pp. 362-377.

[105] Pickands, J., 1975. *Statistical inference using extreme order statistics.* Annals of Statistics, Vol. 3, pp. 119-131.

[106] Polanski, A. and Stoja, E., 2010. *Incorporating higher moments into value-at-risk forecasting.* Journal of Forecasting, No. 29, pp. 523-535.

[107] Politis, D. and Romano, J., 1994. *The stationary bootstrap.* Journal of American Statistics Association, Vol. 89, pp. 1303-1313.

[108] Pritsker, M., 2006. *The hidden dangers of historical simulation.* Journal of Banking and Finance, Vol. 30, No. 2, pp. 561-582.

[109] Ramberg, J.S. and Schmeiser, B. W., 1974. *An approximate method for generating asymmetric random variables* Communications of the ACM, Vol. 17, No. 2, pp. 78-82.

[110] Riskmetrics, T. M., 1996. JP Morgan Technical Document.

[111] Romano, J.P. and Wolf, M., 2005. *Stepwise multiple testing as formalized data snooping.* Econometrica, Vol. 73, No. 4, pp. 1237-1282.

[112] Rubia, A. and Sanchis-Marco, L., 2013. *On downside risk predictability through liquidity and trading activity: A dynamic quantile approach.* International Journal of Forecasting, Vol. 29, No. 1, pp. 202-219.

[113] Sarma, M., Thomas, S., and Shah, A., 2003. *Selection of value at risk models.* Journal of Forecasting, Vol. 22, pp. 337-358.

[114] Schwert, W., 1990. *Stock volatility and the crash of '87.* Review of Financial Studies, Vol. 3, pp. 77-102.

[115] Simonato, J. G., 2011. *The performance of Johnson distributions for computing value at risk and expected shortfall.* The Journal of Derivatives, Vol. 19, No. 1, pp. 7-24.

[116] Smirnov, H., 1939. *On the deviation of the empirical distribution function.* Recueil Mathematique (Matematiceskii Sbornik), N. S., Vol. 6, pp. 3-26.

[117] So, M.K.P. and Yu, P.L.H., 2006. *Empirical analysis of GARCH models in value at risk estimation.* Journal of International Financial Markets, Institutions and Money, Vol. 16, pp. 180-197.

[118] Tang, T.L. and Shieh, S.J., 2006. *Long Memory in Stock Index Futures Markets: A value-at-risk approach.* Physica A, Vol. 366, pp. 437-448.

[119] Taylor, J.W., 2008a. *Estimating value at risk and expected shortfall using expectiles.* Journal of Financial Econometrics, Vol. 6, No. 2, pp. 231-252.

[120] Taylor, J.W., 2008b. *Using exponentially weighted quantile regression to estimate value at risk and expected shortfall.* Journal of Financial Econometrics, Vol. 6, No. 3, pp. 382-406.

[121] Taylor, S.J., 1986. *Modelling Financial Time Series.* John Wiley and Sons, Inc.

[122] Theodossiou, P., 2001. *Skewness and kurtosis in financial data and the pricing of options.* Working Paper. Rutgers University.

[123] Theodossiou, P., 1998. *Financial data and skewed generalized t distribution..* Management Science, Vol. 44, pp. 1650-1661.

[124] Tu, A.H., Wong, W.K. and Chang, M.C., 2008. *Value-at-Risk Long and Short Positions of Asian Stock Markets.* International Research Journal of Finance and Economics, Vol. 22, pp. 135-143.

[125] Tukey, J., 1977. *Exploratory data analysis.* Addison-Wesley series in behavioral science. Adison-Wesley Publishing Company.

[126] White, H., 2000. *A reality check for data snooping.* Econometrica, Vol. 68, No. 5, pp. 1097-1126.

[127] Xu, K.-L., 2013. *Nonparametric inference for conditional quantiles of time series.* Econometric Theory, Vol. 19, No. 4, pp. 673-698.

[128] Yu, P.L.H., Li, W.K. and Jin, S., 2010. *On some models for value-at-risk.* Econometrics Reviews, Vol. 29, No. 5-6, pp. 622-641.

[129] Zangari, P., 1996. *An Improved Methodology for Measuring VaR.* RiskMetrics Monitor, 2nd quarter, pp. 7-25.

[130] Zakoian, J.M., 1994. *Threshold heteroskedastic models.* Journal of Economic Dynamics and Control, Vol. 18, pp. 931-955.

Appendices

A Models and Probability distributions

A.1 GARCH volatility model

The GARCH model is based on an infinite order ARCH specification. It improves upon the ARCH model by reducing the number of estimated parameters from infinity to two. The conditional variance of GARCH(p,q) model (Bollerslev, 1986) [19] is supposed to be not only a linear function of lagged squared residuals but also a linear function of lagged conditional variance. The standard GARCH model captures the existence of volatility clustering but it is unable to express the leverage effect, since it assumes that positive and negative error terms have the same effect on volatility,

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

where $\omega > 0$, $\alpha_i, \beta_j \geq 0$, $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$.

A.2 GJR-GARCH volatility model

To incorporate asymmetric effects on volatility from positive and negative surprises, Glosten, Jagannathan and Runkle (1993) [50] proposed a GJR-GARCH(p,q) model, adding the negative impact of leverage in the conditional variance equation. This model incorporates positive and negative shocks on the conditional variance asymmetrically via the use of the indicator function $I(\varepsilon_{t-i} \leq 0)$, so that the variance equation becomes,

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \left[\alpha_i \varepsilon_{t-i}^2 + \gamma_i I(\varepsilon_{t-i} \leq 0) \varepsilon_{t-i}^2\right] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

The volatility effect of a unit negative shock is $\alpha_i + \gamma_i$ while the effect of a unit positive shock is α_i. A positive value of γ_i indicates that a negative innovation generates greater volatility than a positive innovation of equal size, and on the contrary for a negative value of γ_i.

A.3 APARCH volatility model

The APARCH model (Asymmetric Power ARCH model) was proposed by Ding, Granger and Engle (1993) [35]. This model can well express volatility clustering, fat tailed, excess kurtosis, leverage effect and Taylor effect. The latter effect named after Taylor (1986) [121] who observed that the sample autocorrelation of absolute returns was usually larger than that of squared returns. The variance equation is now,

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} \beta_j (\sigma_{t-j})^\delta$$

where ω, α_i, γ_i, β_j and δ are additional parameters to be estimated. The parameter γ_i reflects the leverage effect ($-1 < \gamma_i < 1$). A positive (resp. negative) value of γ_i means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive (resp. negative) shocks. The parameter δ plays the role of a Box-Cox transformation of $\sigma_t(\delta > 0)$.

The APARCH equation is supposed to satisfy the following conditions,

1. $\omega > 0$,$\alpha_i \geq 0$, $i = 1, 2, ..., q$, $\beta_j \geq 0$, $j = 1, 2, ..., p$, when $\alpha_i = 0$,$i = 1, 2, ..., q$, $\beta_j = 0$, $j = 1, 2, ..., p$, then $\sigma_t^2 = \omega$. Due to the variance is positive, so $\omega > 0$.
2. $0 \leq \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \leq 1$

The APARCH model is a general model because it has great flexibility, having as special cases: i) The simple ARCH model of Engle (1982) when $\delta = 2$, $\beta_j = 0(j = 1, ..., p)$ and $\gamma_i = 0(i = 1, ..., q)$, ii) The simple GARCH model of Bollerslev (1986) [19] when $\delta = 2$ and $\gamma_i = 0(i = 1, ..., q)$, iii) The Absolute Value GARCH (AVGARCH) model of Taylor (1986) [121] and Schwert (1990) [114] when $\delta = 1$ and $\gamma_i = 0(i = 1, ..., q)$, iv) The GJR-GARCH model of Glosten et al. (1993) [50] when $\delta = 2$, v) The Threshold GARCH (TGARCH) model of Zakoian (1994) [130] when $\delta = 1$, vi) The Non Linear ARCH model of Higgins et al. (1992) [64] when $\beta_j = 0(j = 1, ..., p)$ and $\gamma_i = 0(i = 1, ..., q)$, vii) The Log-ARCH model of Geweke (1986) [45] and Pantula (1986) [99] when $\delta \to 0$.

A.4 FGARCH volatility model

The FGARCH model (Family GARCH) of Hentschel (1995) [63] is an omnibus model which subsumes some of the most popular GARCH models. It is similar to the APARCH model, but more general since it allows the decomposition of the residuals in the conditional variance equation to be driven by different powers for z_t and σ_t and also allowing for both shifts and rotations in the news impact curve, where the shift is the main source of asymmetry for small shocks while rotation drives large shocks.

$$\sigma_t^\lambda = \omega + \sum_{i=1}^{q} \alpha_i \sigma_{t-i}^\lambda f^\delta(z_{t-i}) + \sum_{j=1}^{p} \beta_j (\sigma_{t-j})^\lambda$$

where $f^\delta(z_{t-i}) = (|z_{t-i} - \eta_{2i}| - \eta_{1i}(z_{t-i} - \eta_{2i}))^\delta$.

Positivity of $f^\delta(z_{t-i})$ is guaranteed when $|\eta_1| \leq 1$, which ensures that neither arm of the rotated absolute value function crosses the abscissa. The parameter η_2, however, is unrestricted in size and sign.

In the FGARCH model, the magnitude and direction of a shift in the news impact curve are controlled by the parameter η_2; a positive value of η_2 causes a rightward shift of the news impact curve. When the news impact curve is shifted to the right by the distance η_2, one obtains an asymmetric model that matches the stylized facts of stock return volatility, with a negative shock rising volatility more than an equally large but positive shock. The magnitude and direction of a rotation in the news impact curve are controlled by the parameter η_1. By allowing slopes of different magnitudes on either side of the origin, the news impact curves of this type also produce asymmetric variance responses. A positive value of η_1 corresponds to a clockwise rotation. If the news impact curve is rotated clockwise, negative shocks increase volatility more than positive shocks. Notice that η_1 does not cause a pure rotation of the absolute value function. Rather, η_1 controls the slopes of the news impact curve, which are different on either side of the minimum at $\varepsilon = \eta_2$. To achieve a pure rotation, one must also pick an appropriate value of α. Other GARCH models only permit either a shift or a rotation, but not both. In principle, these two types of asymmetry are distinct, and they should not be treated as substitutes for each other.

Various submodels arise as special cases: i) The simple GARCH model of Bollerslev (1986) [19] when $\lambda = \delta = 2$ and $\eta_{1i} = \eta_{2i} = 0 (i = 1, ..., q)$, ii) The Absolute Value GARCH (AVGARCH) model of Taylor (1986) [121] and Schwert [122] when $\lambda = \delta = 1$ and $|\eta_{1i}| \leq 1 (i = 1, ..., q)$, iii) The GJR-GARCH model of Glosten et al. (1993) [70] when $\lambda = \delta = 2$ and $\eta_{2i} = 0 (i = 1, ..., q)$, iv) The Threshold GARCH (TGARCH) model of Zakoian (1994) [130] $\lambda = \delta = 1$, $\eta_{2i} = 0 (i = 1, ..., q)$ and $|\eta_{1i}| \leq 1 (i = 1, ..., q)$, v) The Nonlinear ARCH model of Higgins et al.(1992) [64] when $\delta = \lambda$ and $\eta_{1i} = \eta_{2i} = 0 (i = 1, ..., q)$, vi) The Nonlinear Asymmetric GARCH model of Engle and Ng (1993) when $\delta = \lambda$ and $\eta_{1i} = 0 (i = 1, ..., q)$, vii) The Asymmetric Power ARCH model of Ding et al. (1993) [35] when $\delta = \lambda$, $\eta_{2i} = 0 (i = 1, ..., q)$ and $|\eta_{1i}| \leq 1 (i = 1, ..., q)$, viii) The Exponential GARCH model of Nelson (1991) when $\delta = 1$, $\lambda = 0$ and $\eta_{2i} = 0 (i = 1, ..., q)$.

A.5 Skewed Student-t distribution

To account for the excess skewness and kurtosis typical of financial data, Fernandez and Steel [41] proposed to extend the Student-t distribution by adding a skewness parameter. Their procedure allows for the introduction of skewness in any continuous unimodal and symmetric (about zero) distribution $g(\cdot)$ by changing the scale at each side of the mode. It is helpful to express it as a mixture of two truncated densities. The main drawback of this density is that it is expressed in terms of the mode and the dispersion. To keep it in the ARCH tradition, Lambert and Laurent (2001) [80] expressed the skewed Student-t density in terms of mean and variance.

According to Lambert and Laurent the innovation process z_t is said to follow a (standardized) skewed Student-t distribution, $SKST(0, 1, \xi, \nu)$, if

$$f(z|\xi, \nu) = \frac{2}{\xi + \frac{1}{\xi}} s\{g[\xi(sz + m)|\nu]I_{(-\infty,0)}(z + m/s) + g[(sz + m)/\xi|\nu]I_{[0,\infty)}(z + m/s)\} \quad (1)$$

where $g(\cdot|\nu)$ is the symmetric (unit variance) Student-t density and ξ is the skewness parameter [21]; m and s^2 are, respectively the mean and the variance of the non-standardized

[21] The skewness parameter $\xi > 0$ is defined such that the ratio of probability masses above and below the mean is

$$\frac{Prob(z \geq 0|\xi)}{Prob(z < 0|\xi)} = \xi^2$$

skewed Student-t and are defined as,

$$E(\varepsilon|\xi,\nu) = \frac{\Gamma(\frac{(\nu-1)}{2})\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}\left(\xi - \frac{1}{\xi}\right) \equiv m$$

$$V(\varepsilon|\xi,\nu) = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2 \equiv s^2$$

where

$$z_t = \frac{\varepsilon_t - m}{s}$$

is the standardized random variable with mean 0 and variance 1.

For a standardized skewed Student-t, the log-likelihood is

$$L_{SKST}(\theta) = ln\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \frac{1}{2}ln[\pi(\nu-2)] + ln\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + ln(s) -$$

$$-\frac{1}{2}\sum_{t=1}^{T}\left[ln(\sigma_t^2) + (1+\nu)ln\left(1 + \frac{sz_t+m}{\nu-2}\xi^{-I_t}\right)\right]$$

where $I_t = \begin{cases} 1 & \text{si } z_t \geq -\frac{m}{s} \\ -1 & \text{si } z_t < -\frac{m}{s} \end{cases}$

Notice also that the density $f(z_t|1/\xi,\nu)$ is the mirror of $f(z_t|\xi,\nu)$ with respect to the (zero) mean, i.e., $f(z_t|1/\xi,\nu) = f(-z_t|\xi,\nu)$. Therefore, the sign of $ln(\xi)$ indicates the direction of the skewness: the third moment is positive (negative), and the density is skewed to the right (left), if $ln(\xi) > 0 (< 0)$.

Provided the positive real values are finite, we can easily obtain from the estimated standardized residuals $\hat{z}_t$ the empirical skewness and kurtosis coefficients. If $\hat{z}_t$ is normally distributed, $Sk(\hat{z}_t)$ and $K(\hat{z}_t)$ should not be significantly different from 0 and 3, respectively [22]. Accordingly and following Lambert and Laurent (2001) [80], if $\hat{z}_t \sim SKST(0,1,\xi,\nu)$,

$$Sk(\hat{z}_t|\xi,\nu) = \frac{E(\hat{z}^3|\xi,\nu) - 3E(\hat{z}|\xi,\nu)E(\hat{z}^2|\xi,\nu) + 2E(\hat{z}|\xi,\nu)^3}{V(\hat{z}|\xi,\nu)^{3/2}} \tag{2}$$

$$K(\hat{z}_t|\xi,\nu) = \frac{E(\hat{z}^4|\xi,\nu) - 4E(\hat{z}|\xi,\nu)E(\hat{z}^3|\xi,\nu) + 6E(\hat{z}^2|\xi,\nu)E(\hat{z}|\xi,\nu)^2 - 3E(\hat{z}|\xi,\nu)^4}{V(\hat{z}|\xi,\nu)^2} \tag{3}$$

The skewed Student-t distribution leads to a finite r^{th} order moment (r$\in \Re$) if and only if the corresponding moment of $g(\cdot)$ exists (i.e., for $\xi = 1$). In our case, $g(\cdot)$ is the

[22] Lambert and Laurent (2001) [80] and Giot and Laurent (2003a) [48] have shown that for various financial daily returns, it is realistic to assume that $\hat{z}_t$ is skewed Student-t distribution.

(standardized) Student-t probability density function, with number of degrees of freedom $\nu > 2$. In particular,

$$E(z^r|\xi,\nu) = M_r \frac{\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}}{\xi + \frac{1}{\xi}}$$

where

$$M_r = \int_0^\infty 2s^r g(s)ds$$

Of course, $E(z^r|\xi)$ will be real-valued only for an integer r.

M_r is the r^{th} order moment of $g(\cdot)$ truncated to the positive real line. For $r \leq -1$, the unimodality of $g(\cdot)$ implies that $M_r = \infty$. Thus, we concentrate on positive integer order moments. From M_r, the following properties can be shown to hold for noncentered moments: For odd r, the r^{th} order moment retains the same absolute value but changes sign if we invert ξ, takes the value 0 only for $\xi = 1$, and it is an increasing function of ξ with $lim_{\xi\to\infty} E(z^r|\xi) = \infty$. Even moments, on the other hand, are entirely unaffected by inverting ξ and again increase without bounds in ξ for $\xi > 1$. Consequently, $min_\xi E(z^r|\xi) = E(z^r|\xi = 1)$ for even r. Expressions for centered moments are readily available from M_r. In particular, the variance possesses all of the properties just mentioned for even noncentered moments.

As shown in Equations (2) and (3), ξ and ν characterize skewness and kurtosis. Furthermore, careful scrutiny of the algebra yielding (2) shows that skewness exist if $\nu > 3$. Last, kurtosis in (3) is well defined if $\nu > 4$. Given these restrictions on the underlying parameters, it is clear that the range of skewness and kurtosis will also be restricted to a certain domain. The dominating feature of skewness is the ξ parameter while kurtosis is mainly governed by ν. The Student-t distribution $g(\cdot)$ depends on degrees of freedom. Note that when $\xi = 1$ and $\nu = +\infty$, we get the skewness and the kurtosis of the Gaussian density and when $\xi = 1$ but $\nu > 2$, we have the skewness and the kurtosis of the (standardized) Student-t distribution.

Lambert and Laurent (2001) [80] show that the quantile function $skst^*_{\alpha,\nu,\xi}$ of a non-standardized skewed Student-t density is

$$skst^*_{\alpha,\nu,\xi} = \begin{cases} \frac{1}{\xi} st_{\alpha,\nu} \left[\frac{\alpha}{2}(1+\xi^2)\right], & \text{if } \alpha < \frac{1}{1+\xi^2}, \\ -\xi st_{\alpha,\nu} \left[\frac{1-\alpha}{2}(1+\xi^{-2})\right], & \text{if } \alpha \geq \frac{1}{1+\xi^2}. \end{cases}$$

where $st_{\alpha,\nu}$ is the quantile function of the (unit variance) Student-t density. It is straightforward to obtain the quantile function of the standardized skewed Student-t $skst_{\alpha,\nu,\xi} = (skst^*_{\alpha,\nu,\xi} - m)/s$

A.6 Skewed Generalized Error distribution

The Generalized Error Distribution (GED) by Nelson (1991) [98] is a three parameter distribution belonging to the exponential family with conditional density given by,

$$f(x) = \frac{\kappa e^{-\frac{1}{2}\left|\frac{x-\alpha}{\beta}\right|^{\kappa}}}{2^{1+\kappa^{-1}}\beta\Gamma(\kappa^{-1})}$$

with α, β and κ representing the location, scale and shape parameters respectively and $\Gamma(\cdot)$ is the gamma function. Since the distribution is symmetric and unimodal the location parameter is also mode, median and mean of the distribution (i.e., μ). By symmetry, all odd moments beyond the mean are zero. The variance and kurtosis are given by,

$$Var(X) = \beta^2 2^{2/\kappa} \frac{\Gamma(3\kappa^{-1})}{\Gamma(\kappa^{-1})}$$

$$K(X) = \frac{\Gamma(5\kappa^{-1})\Gamma(\kappa^{-1})}{\Gamma^2(3\kappa^{-1})}$$

As κ increases the density gets flatter and flatter. In the limit as $\kappa \to \infty$ the distribution tends toward the uniform. Special cases are the Normal when $\kappa = 2$ and the Laplace when $\kappa = 1$. For $\kappa > 2$ the distribution is platykurtic and for $\kappa < 2$ it is leptokurtic.

Standardization is simple and involves rescaling the density to have unit standard deviation,

$$Var(X) = 1 \Rightarrow \beta = \sqrt{2^{-2/\kappa} \frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}}$$

Finally substituting into conditional density,

$$f(z|\kappa) = \frac{\kappa e^{-\frac{1}{2}\left|\sqrt{2^{-2/\kappa}\frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}}z\right|^{\kappa}}}{\sqrt{2^{-2/\kappa}\frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}}2^{1+\kappa^{-1}}\Gamma(\kappa^{-1})}$$

The skewed version proposed by Fernandez and Steel is obtained from GED probability density function. We replace $g(\cdot)$ in equation (1) by GED standardized density.

A.7 Johnson S_U distribution

The Johnson S_U distribution was one of the distributions derived by Johnson (1949) [70] based on transformations of the Normal distribution by certain functions. Letting $Z \sim N(0,1)$, the standard Normal distribution, the random variable Y has the Johnson system of frequency curves from this method of transformation $Z = \gamma + \delta g((Y - \xi)/\lambda)$. The form of the resulting distribution depends on the choice of g-function. When $g(u) = sinh^{-1}(u)$, the distribution is unbounded, called the Johnson S_U distribution. The parameters of the distribution are ξ, $\lambda > 0$, γ, $\delta > 0$.

We use a parametrization [23] of the original Johnson S_U distribution, so that the parameters ξ and λ are the mean and the standard deviation of the distribution. The parameter γ determines the skewness of the distribution with $\gamma > 0$ indicating positive

[23] This parametrization is used by the R rugarch package, which we used for estimating the parameters of our models.

skewness and $\gamma < 0$ negative. The parameter δ determines the kurtosis of the distribution. δ should be positive and most likely in the region above 1.

The pdf of the Johnson's S_U, denoted here as $JSU(\xi, \lambda, \gamma, \delta)$, is defined by

$$f_Y(y) = \frac{\delta}{c\lambda}\frac{1}{\sqrt{(r^2+1)}}\frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}z^2\right]$$

where

$$z = -\gamma + \delta sinh^{-1}(r) = -\gamma + \delta \log\left[r + (r^2+1)^{1/2}\right]$$

$$r = \frac{y - (\xi + c\lambda\omega^{1/2} sinh\Omega)}{c\lambda}$$

$$c = \left\{\frac{1}{2}(\omega - 1)[\omega cosh 2\Omega + 1]\right\}^{-1/2}$$

$\omega = exp(\delta^{-2})$ and $\Omega = -\gamma/\delta$. Note that $Z \sim N(0,1)$. Here $E(Y) = \xi$ and $Var(Y) = \lambda^2$.

The pdf of the original Johnson's S_U denoted as $JSU_o(\xi, \lambda, \gamma, \delta)$ is

$$f_Y(y) = \frac{\delta}{\lambda}\frac{1}{\sqrt{(r^2+1)}}\frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}z^2\right]$$

where

$$z = \gamma + \delta sinh^{-1}(r) = \gamma + \delta \log\left[r + (r^2+1)^{1/2}\right]$$

$$r = (y - \xi)/\lambda$$

Note that $Z \sim N(0,1)$, $y \in \mathbb{R}$, ϕ is the probability density function of the standard Normal distribution. The parameters of the JS_U are $(\xi, \lambda, \gamma, \delta)'$ each of them affecting the location, scale, skewness and kurtosis of the distribution. The distribution is positively or negatively skewed according as γ is negative or positive. For a given γ , increasing δ reduces the kurtosis. As $\delta \to \infty$ the distribution approaches the Normal density function. The parameters are not the discreet raw moments of the distribution. We give the first four moments of Johnson S_U distribution. The mean and variance are :

$$E(Y) = \mu = \xi + \lambda\omega^{1/2} sinh(\Omega)$$

$$Var(Y) = \sigma^2 = \frac{\lambda^2}{2}(\omega - 1)(\omega cosh 2\Omega + 1)$$

where $\omega = exp(\delta^{-2})$ and $\Omega = \gamma/\delta$. Since there is not much simplification in the expressions for skewness and kurtosis, we give the third and fourth central moments μ_3 and μ_4, respectively

$$\mu_3 = -\frac{1}{4}\omega^2(\omega^2 - 1)^2[\omega^2(\omega^2+2) sinh 3\Omega + 3 sinh\Omega]$$

$$\mu_4 = \frac{1}{8}(\omega^2 - 1)^2[\omega^4(\omega^8 + 2\omega^6 + 3\omega^4 - 3) cosh 4\Omega + 4\omega^4(\omega^2+2) cosh 2\Omega + 3(2\omega^2 + 1)]$$

From the transformation of the Normal distribution, the cumulative distribution function of the JS_U distribution is shown below. If $Y \sim JS_U(\xi, \lambda, \gamma, \delta)$, $F_Y(y) = \Phi(\gamma + \delta sinh^{-1}[(y-\xi)/\lambda])$ where the function $\Phi(u)$ is the cumulative distribution function of the standard Normal distribution.

From the equation above, the quantile function F_Y^{-1} can be directly derived as $F_Y^{-1} = \xi + \lambda sinh[(\Phi^{-1}(p) - \gamma)/\delta]$ where the quantile function simply depends on the quantiles of the standard Normal distribution $\Phi^{-1}(p)$.

Figure 2.5 shows the distribution's authorized domain, i.e. the region of values of skewness and kurtosis for which a density exists. This is known as the Hamburger moment problem, which characterizes the maximum attainable skewness given a level of kurtosis (see Widder, 1946). From the plot, it is clear that the skewed Student-t has the widest possible combination of skewness and kurtosis for values of kurtosis less than ~ 9, whilst the Johnson S_U distribution has the widest combination for values greater than ~ 9.

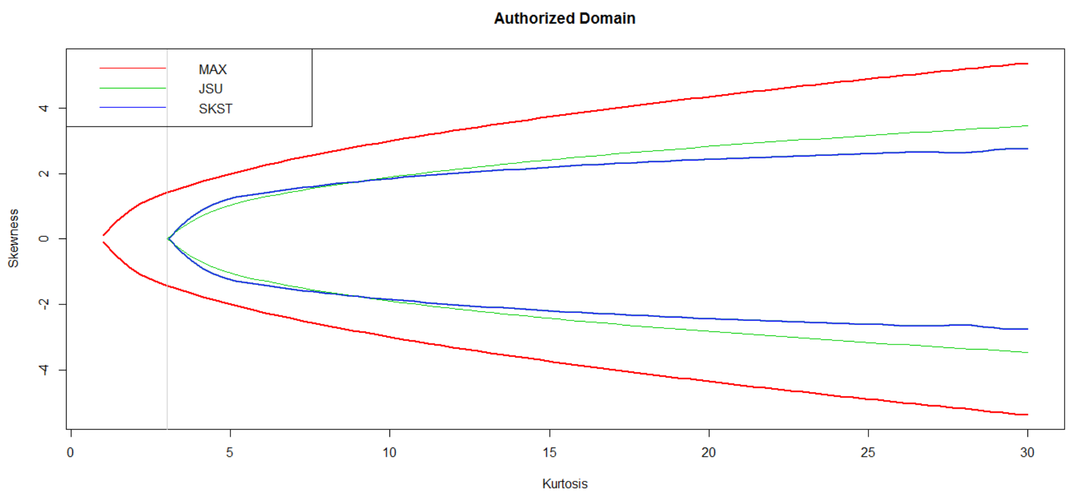


Figure 5: Region for Skewness-Kurtosis for which Skewed Student-t and Johnson S_U distributions exist.

A.8 Skewed Generalized-t distribution

Theodossiou (1998) [123] [9] developed a skewed version of the Generalized t (GT) distribution introduced by McDonald and Newey (1988). The skewed GT is a flexible distribution accommodating the skewness and excess kurtosis often present in financial data.

The skewed GT distribution has the probability density function

$$f(x|\mu, \sigma, \lambda, p, q) = \frac{p}{2\nu\sigma q^{1/p} B(\frac{1}{p}, q) \left(\frac{|x-\mu+m|^p}{q(\nu\sigma)^p (\lambda sign(x-\mu+m)+1)^p} + 1 \right)^{\frac{1}{p}+q}}$$

where

$$m = \frac{2\nu\sigma\lambda q^{\frac{1}{p}} B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)}$$

$$\nu = q^{-\frac{1}{p}} \left[(3\lambda^2 + 1) \left(\frac{B\left(\frac{3}{p}, q - \frac{2}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right) - 4\lambda^2 \left(\frac{B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right)^2 \right]^{-\frac{1}{2}}$$

$B(\cdot)$ is the beta function, and μ, σ, λ, p and q are the location, scale, skewness, peakedness and tail-thickness parameters, respectively. Notice that the parameters have the following restrictions $\sigma > 0$, $-1 < \lambda < 1$, $p > 0$ and $q > 0$. The skewness parameter λ controls the rate of descent of the density around $x = 0$. The parameters p and q control the height and tails of the density, respectively. The parameter q has the degrees of freedom interpretation in case $\lambda = 0$ and $p = 2$.

The skewed GT distribution generates for $\lambda = 0$ McDonald's and Newey's GT distribution; for $p = 2$, Hansen's skewed Student-t distribution; for $\lambda = 0$ and $q = \infty$, the Subbotin's power exponential distribution; for $\lambda = 0$, $p = 1$ and $q = \infty$, the Laplace distribution; for $\lambda = 0$, $p = 2$ and $q = 1$, the Cauchy distribution ; for $\lambda = 0$, $p = 2$ and $q = \infty$, the normal distribution; and for $\lambda = 0$, $p = \infty$ and $q = \infty$, the uniform distribution.

The mean, for $pq > 1$, is

$$E(X) = \mu + \frac{2\nu\sigma\lambda q^{\frac{1}{p}} B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)} - m$$

The variance, for $pq > 2$, is

$$Var(X) = (\nu\sigma)^2 q^{\frac{2}{p}} \left[(3\lambda^2 + 1) \left(\frac{B\left(\frac{3}{p}, q - \frac{2}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right) - 4\lambda^2 \left(\frac{B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right)^2 \right]$$

The skewness, for $pq > 3$, is,

$$\begin{aligned} Sk(X) = & \frac{2q^{\frac{3}{p}}\lambda(\nu\sigma)^3}{B\left(\frac{1}{p}, q\right)^3} \left[8\lambda^2 B\left(\frac{2}{p}, q - \frac{1}{p}\right)^3 - 3(1+3\lambda^2) B\left(\frac{1}{p}, q\right) \right. \\ & \left. B\left(\frac{2}{p}, q - \frac{1}{p}\right) B\left(\frac{3}{p}, q - \frac{2}{p}\right) + 2(1+\lambda^2) B\left(\frac{1}{p}, q\right)^2 B\left(\frac{4}{p}, q - \frac{3}{p}\right) \right] \end{aligned}$$

The kurtosis, for $pq > 4$, is,

$$K(X) = \frac{q^{\frac{4}{p}}(\nu\sigma)^4}{B\left(\frac{1}{p},q\right)^4}\Bigg[-48\lambda^4 B\left(\frac{2}{p},q-\frac{1}{p}\right)^4 + 24\lambda^2(1+3\lambda^2)B\left(\frac{1}{p},q\right)B\left(\frac{2}{p},q-\frac{1}{p}\right)^2$$
$$B\left(\frac{3}{p},q-\frac{2}{p}\right) - 32\lambda^2(1+\lambda^2)B\left(\frac{1}{p},q\right)^2 B\left(\frac{2}{p},q-\frac{1}{p}\right)B\left(\frac{4}{p},q-\frac{3}{p}\right)$$
$$+(1+10\lambda^2+5\lambda^4)B\left(\frac{1}{p},q\right)^3 B\left(\frac{5}{p},q-\frac{4}{p}\right)\Bigg]$$

A.9 Generalized Hyperbolic Skew Student-t distribution

Aas and Haff (2006) [1] proposed a special case of the generalized hyperbolic (GH) family that they denote as the GH skew Student-t distribution. This distribution has the important property that one tail has a polynomial and the other an exponential behavior. Further, it is the only subclass of the GH family of distribution having this property. This is an alternative for modeling the empirical distribution of financial returns. It is often skewed, having one heavy and one semiheavy or more gaussian-like tail. The skew extensions to the Student-t distribution, like that of Fernandez and Steel, have two tails behaving as polynomials. This means that they fit heavy-tailed data well, but they do not handle substantial skewness. Substantial skewness is reached by combining one heavy tail and one nonheavy tail.

The probability density function of the GH Skew Student-t is given by

$$f_X(x) = \frac{2^{\frac{1-\nu}{2}}\delta^\nu|\beta|^{\frac{\nu+1}{2}}K_{\frac{\nu+1}{2}}\left(\sqrt{\beta^2(\delta^2+(x-\mu)^2)}\right)exp(\beta(x-\mu))}{\Gamma(\frac{\nu}{2})\sqrt{\pi}\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{\frac{\nu+1}{2}}} \qquad \beta \neq 0$$

and

$$f_X(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\delta\Gamma(\frac{\nu}{2})}\left[1+\frac{(x-\mu)^2}{\delta^2}\right]^{-(\nu+1)/2} \qquad \beta = 0$$

where $K_\nu(x) \sim \sqrt{\frac{\pi}{2x}}exp(-x)$ for $x \to \pm\infty$ is the modified Bessel function (Abramowitz and Stegun, 1972), μ, δ, β and ν determine the location, scale, skew and shape parameters, respectively.

The density $f_X(x)$ when $\beta = 0$ can be recognized as that of noncentral Student-t distribution with ν degrees of freedom, expectation μ and variance $\delta^2/(\nu-2)$.

The mean and variance of a GH skew Studen-t distributed random variate X are

$$E(X) = \mu + \frac{\beta\delta^2}{\nu-2}$$

and

$$Var(X) = \frac{2\beta^2\delta^4}{(\nu-2)^2(\nu-4)} + \frac{\delta^2}{\nu-2}$$

The variance is only finite when $\nu > 4$, as opposed to the symmetric Student-t distribution, which only requires $\nu > 2$. The derivation of the skewness and kurtosis is relative straightforward (but cumbersome) due to the normal mixture structure of the distribution. These are given by

$$Sk(X) = \frac{2(\nu-4)^{1/2}\beta\delta}{[2\beta^2\delta^2 + (\nu-2)(\nu-4)]^{3/2}}\left[3(\nu-2) + \frac{8\beta^2\delta^2}{\nu-6}\right]$$

and

$$K(X) = \frac{6}{[2\beta^2\delta^2 + (\nu-2)(\nu-4)]^2}\left[(\nu-2)^2(\nu-4) + \frac{16\beta^2\delta^2(\nu-2)(\nu-4)}{\nu-6} + \frac{8\beta^4\delta^4(5\nu-22)}{(\nu-6)(\nu-8)}\right]$$

The skewness and kurtosis do not exist when $\nu \leq 6$ and $\nu \leq 8$, respectively.

Utilizing the property of the modified Bessel function, it can be shown that in the tails, the skew Student-t density behaves as

$$f_X(x) \sim const|x|^{-\nu/2-1} exp(-|\beta x| + \beta x) \qquad x \to \pm\infty$$

Hence the heaviest tail decays as

$$f_X(x) \sim const|x|^{-\nu/2-1} \text{ when } \begin{cases} \beta < 0 & \text{and } x \to -\infty \\ \beta > 0 & \text{and } x \to +\infty \end{cases}$$

and the lightest as

$$f_X(x) \sim const|x|^{-\nu/2-1} exp(-2|\beta x|) \text{ when } \begin{cases} \beta < 0 & \text{and } x \to +\infty \\ \beta > 0 & \text{and } x \to -\infty \end{cases}$$

B Tables

	IBEX	NASDAQ	FTSE	NIKKEI	IBM	SAN	AXA	BP	IRS	GER BOND	US BOND	BRENT	GAS	GOLD	SILVER	EUR/USD	GBP/USD	JPY/USD	AUD/USD
N-GARCH (/1260)	**27**	**30**	**29**	**25**	**19**	**23**	**22**	**19**	**17**	**21**	**21**	**21**	**9**	**34**	**28**	**24**	**17**	**19**	**14**
LR_{uc}	12.522	17.494	15.765	9.583	2.841	6.970	5.794	2.841	1.399	4.711	4.711	4.711	1.154	25.070	14.107	8.234	1.399	2.841	0.152
	(0.000)	(0.000)	(0.000)	(0.002)	(0.092)	(0.008)	(0.016)	(0.092)	(0.237)	(0.029)	(0.029)	(0.029)	(0.000)	(0.000)	(0.000)	(0.004)	(0.237)	(0.092)	(0.697)
LR_{ind}	0.308	1.709	4.795	2.797	-	-	0.752	1.158	10.244	3.976	-	-	-	1.078	5.154	-	1.502	1.158	-
	(0.579)	(0.191)	(0.028)	(0.094)	(-)	(-)	(0.386)	(0.282)	(0.001)	(0.046)	(-)	(-)	(-)	(0.299)	(0.023)	(-)	(0.220)	(0.282)	(-)
LR_{cc}	12.830	19.203	20.560	12.380	-	-	6.546	3.999	11.644	8.687	-	-	-	26.148	19.261	-	2.901	3.999	-
	(0.002)	(0.000)	(0.000)	(0.002)	(-)	(-)	(0.038)	(0.135)	(0.003)	(0.013)	(-)	(-)	(-)	(0.000)	(0.000)	(-)	(0.234)	(0.135)	(-)
DQT	36.155	75.778	55.047	29.016	22.656	27.486	12.802	16.123	32.081	26.574	16.635	13.364	1.308	57.167	46.024	18.852	9.533	8.039	6.701
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.077)	(0.024)	(0.000)	(0.000)	(0.019)	(0.064)	(0.988)	(0.000)	(0.000)	(0.009)	(0.217)	(0.329)	(0.461)
ST-GARCH (/1260)	**19**	**22**	**23**	**19**	**13**	**18**	**15**	**14**	**15**	**16**	**16**	**12**	**6**	**20**	**22**	**18**	**10**	**12**	**11**
LR_{uc}	2.841	5.794	6.970	2.841	0.013	2.064	0.435	0.152	0.435	0.854	0.854	0.029	4.332	3.725	5.794	2.064	0.583	0.029	0.214
	(0.092)	(0.016)	(0.008)	(0.092)	(0.910)	(0.151)	(0.509)	(0.697)	(0.509)	(0.355)	(0.355)	(0.864)	(0.037)	(0.053)	(0.016)	(0.151)	(0.455)	(0.864)	(0.643)
LR_{ind}	-	0.752	3.347	4.699	-	-	-	2.159	10.944	7.064	-	-	-	1.009	3.651	-	-	2.720	-
	(-)	(0.386)	(0.067)	(0.030)	(-)	(-)	(-)	(0.147)	(0.001)	(0.008)	(-)	(-)	(-)	(0.315)	(0.056)	(-)	(-)	(0.099)	(-)
LR_{cc}	-	6.546	10.316	7.540	-	-	-	2.311	11.379	6.858	-	-	-	4.734	9.445	-	-	2.750	-
	(-)	(0.038)	(0.006)	(0.023)	(-)	(-)	(-)	(0.315)	(0.003)	(0.032)	(-)	(-)	(-)	(0.094)	(0.009)	(-)	(-)	(0.253)	(-)
DQT	15.651	19.940	38.373	22.471	2.046	23.716	3.763	10.078	37.290	24.959	7.280	0.870	3.533	11.699	22.533	7.967	2.603	7.682	8.221
	(0.028)	(0.006)	(0.000)	(0.002)	(0.957)	(0.001)	(0.807)	(0.184)	(0.000)	(0.000)	(0.400)	(0.997)	(0.832)	(0.111)	(0.002)	(0.335)	(0.919)	(0.361)	(0.313)
SKST-GARCH (/1260)	**17**	**20**	**16**	**18**	**13**	**17**	**14**	**13**	**13**	**16**	**15**	**10**	**6**	**17**	**18**	**18**	**10**	**13**	**9**
LR_{uc}	1.399	3.725	0.854	2.064	0.013	1.399	0.152	0.013	0.013	0.854	0.435	0.583	4.332	1.399	2.064	2.064	0.583	0.013	1.154
	(0.237)	(0.053)	(0.355)	(0.151)	(0.910)	(0.237)	(0.697)	(0.910)	(0.910)	0.355	(0.509)	(0.445)	(0.037)	(0.237)	(0.151)	(0.151)	(0.445)	(0.910)	(0.283)
LR_{ind}	-	1.009	6.004	5.102	-	-	-	2.425	11.778	6.004	-	-	-	1.502	5.102	-	-	2.425	-
	(-)	(0.315)	(0.014)	(0.024)	(-)	(-)	(-)	(0.119)	(0.000)	(0.014)	(-)	(-)	(-)	(0.220)	(0.024)	(-)	(-)	(0.119)	(-)
LR_{cc}	-	4.734	6.858	7.166	-	-	-	2.438	11.791	6.858	-	-	-	2.901	7.166	-	-	2.438	-
	(-)	(0.094)	(0.032)	(0.027)	(-)	(-)	(-)	(0.295)	(0.003)	(0.032)	(-)	(-)	(-)	(0.234)	(0.028)	(-)	(-)	(0.295)	(-)
DQT	12.788	18.156	30.500	22.598	2.058	24.311	3.193	10.234	24.855	24.988	6.818	0.943	3.534	8.786	21.847	8.012	2.605	6.843	11.417
	(0.077)	(0.011)	(0.000)	(0.002)	(0.956)	(0.001)	(0.866)	(0.176)	(0.001)	(0.001)	(0.448)	(0.996)	(0.831)	(0.268)	(0.003)	(0.331)	(0.919)	(0.445)	(0.121)
SGED-GARCH (/1260)	**17**	**16**	**16**	**16**	**13**	**17**	**15**	**13**	**13**	**16**	**15**	**11**	**5**	**20**	**18**	**18**	**8**	**13**	**9**
LR_{uc}	1.399	0.854	0.854	0.854	0.013	1.399	0.435	0.013	0.013	0.854	0.435	0.214	6.004	3.725	2.064	2.064	1.949	0.013	1.154
	(0.237)	(0.355)	(0.355)	(0.355)	(0.910)	(0.237)	(0.509)	(0.910)	(0.910)	0.355	(0.509)	(0.643)	(0.014)	(0.053)	(0.151)	(0.151)	(0.163)	(0.910)	(0.283)
LR_{ind}	-	1.700	6.004	6.004	-	-	-	2.425	11.778	6.004	-	-	-	1.009	5.102	-	-	2.425	-
	(-)	(0.192)	(0.014)	(0.014)	(-)	(-)	(-)	(0.119)	(0.000)	(0.014)	(-)	(-)	(-)	(0.315)	(0.024)	(-)	(-)	(0.119)	(-)
LR_{cc}	-	2.554	6.858	6.858	-	-	-	2.425	11.791	6.858	-	-	-	4.734	7.166	-	-	2.438	-
	(-)	(0.279)	(0.032)	(0.032)	(-)	(-)	(-)	(0.295)	(0.003)	(0.032)	(-)	(-)	(-)	(0.094)	(0.028)	(-)	(-)	(0.295)	(-)
DQT	12.768	16.329	30.499	26.322	1.848	24.414	3.792	10.277	24.779	25.129	6.806	0.643	4.638	11.870	21.928	7.958	3.157	6.860	11.422
	(0.078)	(0.022)	(0.000)	(0.000)	(0.968)	(0.001)	(0.803)	(0.173)	(0.001)	(0.001)	(0.449)	(0.996)	(0.704)	(0.105)	(0.003)	(0.336)	(0.870)	(0.444)	(0.121)
JSU-GARCH (/1260)	**16**	**18**	**16**	**16**	**13**	**17**	**14**	**13**	**13**	**16**	**13**	**10**	**6**	**14**	**16**	**18**	**10**	**13**	**8**
LR_{uc}	0.854	2.064	0.854	0.854	0.013	1.399	0.152	0.013	0.013	0.854	0.013	0.583	4.332	0.152	0.854	2.064	0.583	0.013	1.949
	(0.355)	(0.151)	(0.355)	(0.355)	(0.910)	(0.237)	(0.697)	(0.910)	(0.910)	(0.355)	(0.910)	(0.455)	(0.037)	(0.697)	(0.355)	(0.151)	(0.445)	(0.910)	(0.163)
LR_{ind}	-	1.322	6.004	6.004	-	-	-	2.425	11.778	6.004	-	-	-	2.159	6.004	-	-	2.425	-
	(-)	(0.250)	(0.014)	(0.014)	(-)	(-)	(-)	(0.119)	(0.000)	(0.014)	(-)	(-)	(-)	(0.142)	(0.014)	(-)	(-)	(0.119)	(-)
LR_{cc}	-	3.385	6.858	6.858	-	-	-	2.438	11.791	6.858	-	-	-	2.311	6.858	-	-	2.438	-
	(-)	(0.184)	(0.032)	(0.032)	(-)	(-)	(-)	(0.295)	(0.003)	(0.032)	(-)	(-)	(-)	(0.315)	(0.032)	(-)	(-)	(0.295)	(-)
DQT	11.188	16.793	30.495	26.444	2.061	24.335	3.194	10.208	24.860	24.988	7.738	0.943	3.534	7.419	23.437	8.011	2.589	6.831	13.673
	(0.131)	(0.019)	(0.000)	(0.000)	(0.956)	(0.001)	(0.866)	(0.178	(0.001)	(0.001)	(0.356)	(0.996)	(0.831)	(0.387)	(0.001)	(0.332)	(0.920)	(0.447)	(0.057)
SGT-GARCH (/1260)	**17**	**18**	**16**	**16**	**13**	**17**	**15**	**13**	**13**	**16**	**15**	**10**	**5**	**20**	**18**	**18**	**9**	**13**	**9**
LR_{uc}	1.399	2.064	0.854	0.854	0.013	1.399	0.435	0.013	0.013	0.854	0.435	0.583	6.004	3.725	2.064	2.064	1.154	0.013	1.154
	(0.237)	(0.151)	(0.355)	(0.355)	(0.910)	(0.237)	(0.509)	(0.910)	(0.910)	(0.355)	(0.509)	(0.455)	(0.014)	(0.053)	(0.150)	(0.150)	(0.283)	(0.910)	(0.283)
LR_{ind}	-	1.322	6.004	6.004	-	-	-	2.425	11.778	6.004	-	-	-	1.009	5.102	-	-	2.438	-
	(-)	(0.250)	(0.014)	(0.014)	(-)	(-)	(-)	(0.119)	(0.000)	(0.014)	(-)	(-)	(-)	(0.315)	(0.024)	(-)	(-)	(0.119)	(-)
LR_{cc}	-	3.385	6.858	6.858	-	-	-	2.438	11.791	6.858	-	-	-	4.734	7.166	-	-	2.438	-
	(-)	(0.184)	(0.032)	(0.032)	(-)	(-)	(-)	(0.295)	(0.003)	(0.032)	(-)	(-)	(-)	(0.094)	(0.028)	(-)	(-)	(0.295)	(-)
DQT	12.612	15.839	30.404	26.249	1.822	24.392	3.805	10.298	24.738	25.100	6.777	0.933	4.638	11.824	21.897	7.939	2.436	6.850	13.395
	(0.082)	(0.026)	(0.000)	(0.000)	(0.966)	(0.001)	(0.802)	(0.172	(0.001)	(0.001)	(0.452)	(0.996)	(0.704)	(0.106)	(0.003)	(0.338)	(0.932)	(0.445)	(0.122)
GHST-GARCH (/1260)	**12**	**9**	**9**	**10**	**13**	**11**	**8**	**7**	**13**	**15**	**13**	**7**	**5**	**21**	**18**	**11**	**3**	**8**	**8**
LR_{uc}	0.029	1.154	1.154	0.583	0.013	0.214	1.949	2.996	0.013	0.435	0.013	2.996	6.004	4.711	2.064	0.214	10.663	1.949	1.949
	(0.864)	(0.283)	(0.283)	(0.445)	(0.910)	(0.643)	(0.163)	(0.083)	(0.910)	(0.509)	(0.910)	(0.083)	(0.014)	(0.029)	(0.151)	(0.643)	(0.001)	(0.163)	(0.163)
LR_{ind}	-	3.834	3.834	9.879	-	-	-	-	11.778	6.512	-	-	-	3.976	5.102	-	-	-	-
	(-)	(0.050)	(0.050)	(0.002)	(-)	(-)	(-)	(-)	(0.000)	(0.011)	(-)	(-)	(-)	(0.046)	(0.024)	(-)	(-)	(-)	(-)
LR_{cc}	-	4.988	4.988	10.462	-	-	-	-	11.791	6.947	-	-	-	8.687	7.166	-	-	-	-
	(-)	(0.082)	(0.082)	(0.005)	(-)	(-)	(-)	(-)	(0.003)	(0.031)	(-)	(-)	(-)	(0.013)	(0.028)	(-)	(-)	(-)	(-)
DQT	9.321	20.271	11.937	42.347	2.000	16.949	2.120	3.191	24.974	26.040	7.749	2.634	4.658	25.940	21.674	1.936	7.604	3.112	13.740
	(0.231)	(0.005)	(0.103)	(0.000)	(0.959)	(0.018)	(0.953)	(0.867	(0.001)	(0.000)	(0.355)	(0.917)	(0.702)	(0.000)	(0.003)	(0.963)	(0.369)	(0.874)	(0.056)

Table 2.2: $VaR_{1\%}$ results with GARCH volatility model for all assets: unconditional coverage test, independence and conditional coverage tests and Dynamic Quantile Test. P-values are reported in parentheses. For each model, the first line shows the number of VaR excesses produced with the different assets (boldface).

	IBEX	NASDAQ	FTSE	NIKKEI	IBM	SAN	AXA	BP	IRS	GER BOND	US BOND	BRENT	GAS	GOLD	SILVER	EUR/USD	GBP/USD	JPY/USD	AUD/USD
N-GJRGARCH (/1260)	**3**	**24**	**26**	**27**	**17**	**24**	**22**	**16**	**16**	**21**	**22**	**16**	**9**	**36**	**31**	**23**	**17**	**21**	**12**
LR_{uc}	11.013	8.234	11.013	12.522	1.399	8.234	5.794	0.854	0.854	4.711	5.794	0.854	1.154	29.229	19.291	6.970	1.399	4.711	0.029
	(0.001)	(0.004)	(0.001)	(0.000)	(0.237)	(0.004)	(0.016)	(0.355)	(0.355)	(0.029)	(0.016)	(0.355)	(0.283)	(0.000)	(0.000)	(0.008)	(0.237)	(0.029)	(0.864)
LR_{ind}	-	0.543	2.549	2.317	-	-	-	-	10.579	3.976	-	-	-	0.829	7.555	-	1.502	0.874	-
	(-)	(0.461)	(0.110)	(0.128)	(-)	(-)	(-)	(-)	(0.001)	(0.046)	(-)	(-)	(-)	(0.363)	(0.006)	(-)	(0.220)	(0.349)	(-)
LR_{cc}	-	8.777	13.562	14.839	-	-	-	-	11.433	8.687	-	-	-	30.058	26.846	-	2.901	5.585	-
	(-)	(0.012)	(0.001)	(0.000)	(-)	(-)	(-)	(-)	(0.000)	(0.013)	(-)	(-)	(-)	(0.000)	(0.000)	(-)	(0.234)	(0.061)	(-)
DQT	26.920	19.355	30.116	30.541	7.779	28.274	12.287	5.086	34.714	26.581	21.127	2.930	1.310	65.991	67.379	16.046	7.020	10.135	7.284
	(0.000)	(0.007)	(0.000)	(0.000)	(0.352)	(0.000)	(0.091)	(0.649)	(0.000)	(0.000)	(0.003)	(0.891)	(0.988)	(0.000)	(0.000)	(0.024)	(0.427)	(0.181)	(0.399)
ST-GJRGARCH (/1260)	**21**	**21**	**23**	**20**	**14**	**20**	**16**	**12**	**14**	**18**	**15**	**8**	**6**	**23**	**21**	**17**	**11**	**13**	**10**
LR_{uc}	4.711	4.711	6.970	3.725	0.152	3.725	0.854	0.029	0.152	2.064	0.435	1.949	4.332	6.970	4.711	1.399	0.214	0.013	0.583
	(0.030)	(0.030)	(0.009)	(0.053)	(0.697)	(0.053)	(0.355)	(0.864)	(0.697)	(0.151)	(0.509)	(0.163)	(0.037)	(0.008)	(0.029)	(0.237)	(0.643)	(0.910)	(0.445)
LR_{ind}	-	0.874	0.642	4.325	-	-	-	-	11.342	5.102	-	-	-	3.347	3.976	-	-	2.425	-
	(-)	(0.349)	(0.423)	(0.037)	(-)	(-)	(-)	(-)	(0.001)	(0.024)	(-)	(-)	(-)	(0.067)	(0.046)	(-)	(-)	(0.119)	(-)
LR_{cc}	-	5.585	7.611	8.050	-	-	-	-	11.494	7.166	-	-	-	10.316	8.687	-	-	2.438	-
	(-)	(0.061)	(0.022)	(0.018)	(-)	(-)	(-)	(-)	(0.003)	(0.028)	(-)	(-)	(-)	(0.006)	(0.013)	(-)	(-)	(0.295)	(-)
DQT	22.016	16.745	14.692	22.933	3.279	22.511	2.808	2.326	41.604	27.881	7.205	1.859	3.535	32.509	21.961	4.566	1.497	6.931	9.439
	(0.002)	(0.019)	(0.040)	(0.002)	(0.858)	(0.002)	(0.902)	(0.939)	(0.000)	(0.000)	(0.408)	(0.967)	(0.831)	(0.000)	(0.002)	(0.713)	(0.982)	(0.436)	(0.223)
SKST-GJRGARCH (/1260)	**16**	**16**	**14**	**17**	**13**	**19**	**14**	**12**	**13**	**16**	**14**	**7**	**6**	**20**	**21**	**18**	**10**	**14**	**10**
LR_{uc}	0.854	0.854	0.152	1.399	0.013	2.841	0.152	0.029	0.013	0.854	0.152	2.996	4.332	3.725	4.711	2.064	0.583	0.152	0.583
	(0.355)	(0.355)	(0.697)	(0.237)	(0.910)	(0.092)	(0.697)	(0.864)	(0.910)	(0.355)	(0.697)	(0.083)	(0.037)	(0.053)	(0.030)	(0.151)	(0.445)	(0.697)	(0.445)
LR_{ind}	-	1.700	-	5.536	-	-	-	-	11.778	6.004	-	-	-	0.096	3.976	-	-	2.159	-
	(-)	(0.192)	(-)	(0.019)	(-)	(-)	(-)	(-)	(0.000)	(0.014)	(-)	(-)	(-)	(0.757)	(0.046)	(-)	(-)	(0.142)	(-)
LR_{cc}	-	2.554	-	6.935	-	-	-	-	11.791	6.858	-	-	-	8.050	8.687	-	-	2.311	-
	(-)	(0.279)	(-)	(0.031)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.178)	(0.013)	(-)	(-)	(0.315)	(-)
DQT	13.971	11.379	1.104	25.619	2.578	21.315	1.244	2.280	25.067	25.028	8.942	2.594	3.536	29.175	21.958	8.100	2.190	6.386	9.449
	(0.052)	(0.123)	(0.993)	(0.000)	(0.921)	(0.003)	(0.989)	(0.943)	(0.001)	(0.001)	(0.257)	(0.919)	(0.831)	(0.000)	(0.002)	(0.324)	(0.948)	(0.495)	(0.222)
SGED-GJRGARCH (/1260)	**16**	**14**	**14**	**15**	**13**	**19**	**14**	**12**	**14**	**16**	**14**	**8**	**5**	**21**	**20**	**18**	**8**	**14**	**10**
LR_{uc}	0.854	0.152	0.152	0.435	0.013	2.841	0.152	0.029	0.152	0.854	0.152	1.949	6.004	4.711	3.725	2.064	1.949	0.152	0.583
	(0.355)	(0.697)	(0.697)	(0.509)	(0.910)	(0.092)	(0.697)	(0.864)	(0.697)	(0.355)	(0.697)	(0.163)	(0.014)	(0.030)	(0.054)	(0.151)	(0.163)	(0.697)	(0.445)
LR_{ind}	-	2.159	-	6.512	-	-	-	-	11.342	6.004	-	-	-	3.976	4.325	-	-	2.159	-
	(-)	(0.142)	(-)	(0.011)	(-)	(-)	(-)	(-)	(0.000)	(0.014)	(-)	(-)	(-)	(0.046)	(0.037)	(-)	(-)	(0.142)	(-)
LR_{cc}	-	2.311	-	6.947	-	-	-	-	11.494	6.858	-	-	-	8.687	8.050	-	-	2.311	-
	(-)	(0.314)	(-)	(0.031)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.013)	(0.018)	(-)	(-)	(0.315)	(-)
DQT	13.887	11.948	1.103	26.744	2.378	21.307	1.237	2.265	41.603	25.143	8.808	1.861	4.638	25.977	21.672	8.044	2.359	6.407	9.444
	(0.053)	(0.102)	(0.993)	(0.000)	(0.936)	(0.003)	(0.990)	(0.944)	(0.000)	(0.001)	(0.267)	(0.967)	(0.704)	(0.000)	(0.003)	(0.329)	(2.359)	(0.493)	(0.222)
JSU-GJRGARCH (/1260)	**15**	**14**	**14**	**15**	**13**	**19**	**14**	**12**	**13**	**16**	**14**	**7**	**7**	**17**	**19**	**18**	**10**	**15**	**9**
LR_{uc}	0.435	0.152	0.152	0.435	0.013	2.841	0.152	0.029	0.013	0.854	0.152	2.996	2.996	1.399	2.841	2.064	0.583	0.435	1.154
	(0.509)	(0.697)	(0.697)	(0.509)	(0.910)	(0.092)	(0.697)	(0.864)	(0.910)	(0.355)	(0.696)	(0.083)	(0.083)	(0.237)	(0.092)	(0.151)	(0.445)	(0.509)	(0.283)
LR_{ind}	-	2.159	-	6.512	-	-	-	-	11.778	6.004	-	-	-	1.502	4.699	-	-	1.919	-
	(-)	(0.142)	(-)	(0.011)	(-)	(-)	(-)	(-)	(0.000)	(0.014)	(-)	(-)	(-)	(0.220)	(0.030)	(-)	(-)	(0.166)	(-)
LR_{cc}	-	2.311	-	6.947	-	-	-	-	11.778	6.858	-	-	-	2.901	7.540	-	-	2.354	-
	(-)	(0.315)	(-)	(0.031)	(-)	(-)	(-)	(-)	(0.000)	(0.032)	(-)	(-)	(-)	(0.234)	(0.023)	(-)	(-)	(0.308)	(-)
DQT	12.589	11.978	1.104	26.821	2.587	21.315	1.244	2.266	25.071	25.019	8.952	2.594	2.611	10.392	21.448	8.097	2.178	6.671	11.208
	(0.083)	(0.101)	(0.993)	(0.000)	(0.920)	(0.003)	(0.989)	(0.944)	(0.001)	(0.001)	(0.256)	(0.919)	(0.918)	(0.167)	(0.003)	(0.324)	(0.949)	(0.464)	(0.130)
SGT-GJRGARCH (/1260)	**17**	**15**	**14**	**16**	**13**	**19**	**14**	**12**	**14**	**16**	**14**	**8**	**5**	**18**	**13**	**18**	**8**	**14**	**10**
LR_{uc}	1.399	0.435	0.152	0.854	0.013	2.841	0.152	0.029	0.152	0.854	0.152	1.949	6.004	2.064	0.123	2.064	1.949	0.152	0.583
	(0.237)	(0.509)	(0.697)	(0.355)	(0.910)	(0.092)	(0.697)	(0.864)	(0.697)	(0.355)	(0.697)	(0.163)	(0.014)	(0.151)	(0.726)	(0.151)	(0.163)	(0.697)	(0.445)
LR_{ind}	-	1.919	-	6.004	-	-	-	-	11.342	6.004	-	-	-	5.102	2.903	-	-	2.159	-
	(-)	(0.166)	(-)	(0.014)	(-)	(-)	(-)	(-)	(0.001)	(0.014)	(-)	(-)	(-)	(0.239)	(0.088)	(-)	(-)	(0.142)	(-)
LR_{cc}	-	2.354	-	6.858	-	-	-	-	11.494	6.858	-	-	-	7.166	3.026	-	-	2.311	-
	(-)	(0.308)	(-)	(0.032)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.028)	(0.220)	(-)	(-)	(0.315)	(-)
DQT	15.621	11.585	1.101	25.871	2.389	21.288	1.239	2.301	41.550	25.124	8.798	1.861	4.638	26.865	7.602	8.029	2.356	6.390	9.434
	(0.029)	(0.115)	(0.993)	(0.000)	(0.935)	(0.003)	(0.990)	(0.941)	(0.000)	(0.001)	(0.267)	(0.967)	(0.704)	(0.000)	(0.369)	(0.330)	(0.937)	(0.495)	(0.223)
GHST-GJRGARCH (/1260)	**10**	**7**	**6**	**8**	**12**	**7**	**7**	**6**	**13**	**13**	**11**	**5**	**4**	**20**	**19**	**9**	**1**	**12**	**5**
LR_{uc}	0.583	4.861	4.332	1.949	0.029	2.996	2.996	4.332	0.013	0.013	0.214	6.004	8.080	3.725	2.841	1.154	18.240	0.029	6.004
	(0.445)	(0.027)	(0.037)	(0.163)	(0.864)	(0.083)	(0.083)	(0.037)	(0.910)	(0.910)	(0.643)	(0.014)	(0.004)	(0.053)	(0.092)	(0.283)	(0.000)	(0.864)	(0.014)
LR_{ind}	-	4.861	-	4.310	-	-	-	-	11.778	7.667	-	-	-	4.325	4.699	-	-	2.720	-
	(-)	(0.027)	(-)	(0.037)	(-)	(-)	(-)	(-)	(0.000)	(0.005)	(-)	(-)	(-)	(0.037)	(0.030)	(-)	(-)	(0.099)	(-)
LR_{cc}	-	7.857	-	6.259	-	-	-	-	11.791	7.679	-	-	-	8.050	7.540	-	-	2.750	-
	(-)	(0.019)	(-)	(0.043)	(-)	(-)	(-)	(-)	(0.003)	(0.021)	(-)	(-)	(-)	(0.018)	(0.023)	(-)	(-)	(0.253)	(-)
DQT	14.140	27.404	4.210	13.725	3.514	17.124	3.065	3.604	25.191	30.054	9.480	4.639	5.909	24.876	21.802	3.434	10.822	7.664	24.143
	(0.049)	(0.000)	(0.755)	(0.056)	(0.834)	(0.017)	(0.879)	(0.824)	(0.001)	(0.000)	(0.220)	(0.704)	(0.550)	(0.001)	(0.003)	(0.842)	(0.146)	(0.363)	(0.001)

Table 2.3: $VaR_{1\%}$ results with GJRGARCH volatility model for all assets: unconditional coverage test, independence and conditional coverage tests and Dynamic Quantile Test. P-values are reported in parentheses. For each model, the first line shows the number of VaR excesses produced with the different assets (boldface).

	IBEX	NASDAQ	FTSE	NIKKEI	IBM	SAN	AXA	BP	IRS	GER BOND	US BOND	BRENT	GAS	GOLD	SILVER	EUR/USD	GBP/USD	JPY/USD	AUD/USD
N-APARCH (/1260)	**26**	**22**	**25**	**27**	**23**	**25**	**25**	**16**	**16**	**21**	**22**	**16**	**11**	**22**	**32**	**24**	**16**	**26**	**11**
LR_{uc}	11.013	5.794	9.583	12.522	6.970	9.583	9.583	0.854	0.854	4.711	5.794	0.854	0.214	5.794	21.154	8.234	0.854	11.013	0.214
	(0.001)	(0.016)	(0.002)	(0.000)	(0.008)	(0.002)	(0.002)	(0.355)	(0.355)	(0.030)	(0.016)	(0.355)	(0.643)	(0.016)	(0.000)	(0.004)	(0.355)	(0.001)	(0.643)
LR_{ind}	-	0.752	0.455	2.317	-	-	0.455	-	10.579	3.976	-	-	-	3.651	7.104	-	-	0.376	-
	(-)	(0.386)	(0.500)	(0.128)	(-)	(-)	(0.500)	(-)	(0.001)	(0.046)	(-)	(-)	(-)	(0.0561)	(0.008)	(-)	(-)	(0.539)	(-)
LR_{cc}	-	6.546	10.037	14.839	-	-	10.037	-	11.433	8.687	-	-	-	9.445	28.258	-	-	11.390	-
	(-)	(0.038)	(0.007)	(0.000)	(-)	(-)	(0.007)	(-)	(0.003)	(0.013)	(-)	(-)	(-)	(0.009)	(0.000)	(-)	(-)	(0.003)	(-)
DQT	26.812	16.790	17.953	31.881	25.773	28.302	19.629	4.727	34.679	24.328	21.050	2.943	8.057	38.869	68.698	18.995	2.698	19.616	8.433
	(0.000)	(0.019)	(0.012)	(0.000)	(0.000)	(0.000)	(0.006)	(0.693)	(0.000)	(0.001)	(0.004)	(0.890)	(0.328)	(0.000)	(0.000)	(0.008)	(0.911)	(0.006)	(0.296)
ST-APARCH (/1260)	**22**	**20**	**23**	**22**	**16**	**19**	**17**	**14**	**15**	**18**	**15**	**8**	**6**	**27**	**22**	**18**	**11**	**14**	**10**
LR_{uc}	5.794	3.725	6.970	5.794	0.854	2.841	1.399	0.152	0.435	2.064	0.435	1.949	4.332	12.522	5.794	2.064	0.214	0.152	0.583
	(0.016)	(0.053)	(0.008)	(0.016)	(0.355)	(0.918)	(0.237)	(0.697)	(0.509)	(0.151)	(0.509)	(0.163)	(0.037)	(0.000)	(0.016)	(0.151)	(0.643)	(0.697)	(0.445)
LR_{ind}	-	1.009	0.642	3.651	-	-	-	-	10.944	5.102	-	-	-	2.317	7.798	-	-	2.159	-
	(-)	(0.315)	(0.423)	(0.056)	(-)	(-)	(-)	(-)	(0.001)	(0.024)	(-)	(-)	(-)	(0.128)	(0.005)	(-)	(-)	(0.142)	(-)
LR_{cc}	-	4.734	7.611	9.445	-	-	-	-	11.379	7.166	-	-	-	14.839	13.592	-	-	2.311	-
	(-)	(0.094)	(0.022)	(0.008)	(-)	(-)	(-)	(-)	(0.003)	(0.028)	(-)	(-)	(-)	(0.000)	(0.001)	(-)	(-)	(0.315)	(-)
DQT	23.056	15.381	16.037	26.160	27.979	15.812	3.636	4.704	38.969	27.961	7.282	1.862	3.524	48.341	41.604	8.038	1.633	6.569	9.373
	(0.002)	(0.031)	(0.025)	(0.000)	(0.000)	(0.027)	(0.820)	(0.696)	(0.000)	(0.000)	(0.400)	(0.967)	(0.833)	(0.000)	(0.000)	(0.329)	(0.977)	(0.475)	(0.227)
SKST-APARCH (/1260)	**17**	**15**	**14**	**17**	**16**	**18**	**16**	**13**	**14**	**17**	**14**	**7**	**6**	**22**	**18**	**18**	**10**	**17**	**10**
LR_{uc}	1.399	0.435	0.152	1.399	0.854	2.064	0.854	0.013	0.152	1.399	0.152	2.996	4.332	5.794	2.064	2.064	0.583	1.399	0.583
	(0.237)	(0.509)	(0.697)	(0.237)	(0.355)	(0.151)	(0.355)	(0.910)	(0.697)	(0.237)	(0.697)	(0.083)	(0.037)	(0.016)	(0.151)	(0.151)	(0.445)	(0.237)	(0.445)
LR_{ind}	-	1.919	1.078	1.502	-	-	-	-	11.342	5.536	-	-	-	3.651	10.181	-	-	1.502	-
	(-)	(0.166)	(0.299)	(0.220)	(-)	(-)	(-)	(-)	(0.001)	(0.019)	(-)	(-)	(-)	(0.056)	(0.001)	(-)	(-)	(0.220)	(-)
LR_{cc}	-	2.354	1.276	2.901	-	-	-	-	11.494	6.935	-	-	-	9.445	12.244	-	-	2.901	-
	(-)	(0.308)	(0.528)	(0.234)	(-)	(-)	(-)	(-)	(0.003)	(0.031)	(-)	(-)	(-)	(0.009)	(0.002)	(-)	(-)	(0.234)	(-)
DQT	14.518	11.365	7.357	7.049	28.015	15.371	2.600	3.031	41.966	29.511	9.082	2.593	3.524	38.869	47.187	8.083	2.374	6.885	9.381
	(0.042)	(0.123)	(0.393)	(0.424)	(0.000)	(0.031)	(0.919)	(0.882)	(0.000)	(0.000)	(0.247)	(0.919)	(0.833)	(0.000)	(0.000)	(0.325)	(0.936)	(0.441)	(0.226)
SGED-APARCH (/1260)	**17**	**15**	**14**	**15**	**15**	**18**	**16**	**12**	**14**	**16**	**14**	**7**	**5**	**21**	**21**	**18**	**9**	**16**	**13**
LR_{uc}	1.399	0.435	0.152	0.435	0.435	2.064	0.854	0.029	0.152	0.854	0.152	2.996	6.004	4.711	4.711	2.064	1.154	0.854	0.039
	(0.237)	(0.509)	(0.697)	(0.509)	(0.509)	(0.151)	(0.355)	(0.864)	(0.697)	(0.355)	(0.697)	(0.083)	(0.014)	(0.030)	(0.030)	(0.151)	(0.283)	(0.355)	(0.844)
LR_{ind}	-	1.919	-	1.919	-	-	-	-	11.342	6.004	-	-	-	3.976	8.337	-	-	1.700	0.549
	(-)	(0.166)	(-)	(0.166)	(-)	(-)	(-)	(-)	(0.001)	(0.014)	(-)	(-)	(-)	(0.046)	(0.004)	(-)	(-)	(0.192)	(0.458)
LR_{cc}	-	2.354	-	2.354	-	-	-	-	11.494	6.858	-	-	-	8.687	13.048	-	-	2.554	0.588
	(-)	(0.308)	(-)	(0.308)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.013)	(0.001)	(-)	(-)	(0.279)	(0.745)
DQT	14.433	11.339	7.337	6.308	28.183	15.354	2.583	1.971	41.796	25.135	8.861	2.592	4.630	27.205	42.468	8.020	2.093	7.024	3.766
	(0.044)	(0.124)	(0.395)	(0.504)	(0.000)	(0.032)	(0.921)	(0.961)	(0.000)	(0.001)	(0.263)	(0.920)	(0.705)	(0.000)	(0.000)	(0.331)	(0.954)	(0.426)	(0.806)
JSU-APARCH (/1260)	**17**	**15**	**12**	**15**	**16**	**18**	**16**	**13**	**13**	**17**	**14**	**7**	**6**	**20**	**17**	**18**	**10**	**17**	**9**
LR_{uc}	1.399	0.435	0.029	0.435	0.854	2.064	0.854	0.013	0.013	1.399	0.152	2.996	4.332	3.725	1.399	2.064	0.583	1.399	1.154
	(0.237)	(0.509)	(0.864)	(0.509)	(0.355)	(0.151)	(0.355)	(0.910)	(0.910)	(0.237)	(0.697)	(0.083)	(0.037)	(0.053)	(0.236)	(0.151)	(0.455)	(0.237)	(0.283)
LR_{ind}	-	1.919	-	1.919	-	-	-	-	11.778	5.536	-	-	-	4.325	5.536	-	-	1.502	-
	(-)	(0.166)	(-)	(0.166)	(-)	(-)	(-)	(-)	(0.000)	(0.019)	(-)	(-)	(-)	(0.037)	(0.019)	(-)	(-)	(0.220)	(-)
LR_{cc}	-	2.354	-	2.354	-	-	-	-	11.791	6.935	-	-	-	8.050	6.935	-	-	2.901	-
	(-)	(0.308)	(-)	(0.308)	(-)	(-)	(-)	(-)	(0.003)	(0.031)	(-)	(-)	(-)	(0.018)	(0.031)	(-)	(-)	(0.234)	(-)
DQT	10.242	11.330	8.142	6.363	28.061	15.353	2.601	3.012	25.350	29.492	9.076	2.593	3.525	26.137	22.408	8.080	2.367	6.861	11.108
	(0.175)	(0.125)	(0.320)	(0.498)	(0.000)	(0.032)	(0.919)	(0.884)	(0.001)	(0.000)	(0.247)	(0.919	(0.833)	(0.000)	(0.002)	(0.326)	(0.937)	(0.443)	(0.134)
SGT-APARCH (/1260)	**18**	**15**	**14**	**15**	**16**	**18**	**16**	**12**	**14**	**16**	**14**	**7**	**5**	**21**	**19**	**18**	**9**	**16**	**10**
LR_{uc}	2.064	0.435	0.152	0.435	0.854	2.064	0.854	0.029	0.152	0.854	0.152	2.996	6.004	4.711	2.841	2.064	1.154	0.854	0.583
	(0.151)	(0.509)	(0.697)	(0.509)	(0.355)	(0.151)	(0.355)	(0.864)	(0.697)	(0.355)	(0.697)	(0.083)	(0.014)	(0.030)	(0.092)	(0.151)	(0.283)	(0.355)	(0.445)
LR_{ind}	-	1.919	-	1.919	-	-	-	-	11.342	6.004	-	-	-	3.976	9.525	-	-	1.700	-
	(-)	(0.166)	(-)	(0.166)	(-)	(-)	(-)	(-)	(0.001)	(0.014)	(-)	(-)	(-)	(0.046)	(0.002)	(-)	(-)	(0.192)	(-)
LR_{cc}	-	2.354	-	2.354	-	-	-	-	11.494	6.858	-	-	-	8.687	12.366	-	-	2.554	-
	(-)	(0.308)	(-)	(0.308)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.013)	(0.002)	(-)	(-)	(0.279)	(-)
DQT	17.110	11.333	7.314	6.292	27.868	15.341	2.588	2.021	41.755	25.116	8.867	2.593	4.630	27.028	45.222	8.003	2.090	6.237	9.375
	(0.017)	(0.125)	(0.397)	(0.506)	(0.000)	(0.032)	(0.920)	(0.959)	(0.000)	(0.001)	(0.262)	(0.920	(0.705)	(0.000)	(0.000)	(0.332)	(0.955)	(0.512)	(0.227)
GHST-APARCH (/1260)	**12**	**6**	**9**	**10**	**14**	**10**	**12**	**7**	**13**	**16**	**14**	**5**	**4**	**27**	**21**	**16**	**3**	**12**	**7**
LR_{uc}	0.029	4.332	1.154	0.583	0.152	0.583	0.029	2.996	0.013	0.854	0.152	6.004	8.080	12.522	4.711	0.854	10.663	0.029	2.996
	(0.864)	(0.037)	(0.283)	(0.445)	(0.697)	(0.445)	(0.864)	(0.083)	(0.910)	(0.355)	(0.697)	(0.014)	(0.000)	(0.030)	(0.355)	(0.001)	(0.864)	(0.083)	(0.283)
LR_{ind}	-	5.511	-	3.418	-	-	-	-	11.778	6.004	-	-	-	2.317	8.337	-	-	2.720	-
	(-)	(0.019)	(-)	(0.064)	(-)	(-)	(-)	(-)	(0.000)	(0.014)	(-)	(-)	(-)	(0.128)	(0.004)	(-)	(-)	(0.099)	(-)
LR_{cc}	-	9.843	-	4.001	-	-	-	-	11.791	6.858	-	-	-	14.839	13.048	-	-	2.750	-
	(-)	(0.007)	(-)	(0.135)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.000)	(0.001)	(-)	(-)	(0.253)	(-)
DQT	11.979	32.801	2.710	10.378	31.600	19.424	0.664	2.878	25.536	25.218	8.817	4.684	5.905	48.514	42.397	7.255	7.477	8.069	15.980
	(0.101)	(0.000)	(0.910)	(0.168)	(0.000)	(0.007)	(0.007)	(0.998)	(0.897)	(0.001)	(0.266)	(0.698)	(0.551	(0.000)	(0.000)	(0.403)	(0.384)	(0.326)	(0.025)

Table 2.4: $VaR_{1\%}$ results with APARCH volatility model for all assets: unconditional coverage test, independence and conditional coverage tests and Dynamic Quantile Test. P-values are reported in parentheses. For each model, the first line shows the number of VaR excesses produced with the different assets (boldface).

	IBEX	NASDAQ	FTSE	NIKKEI	IBM	SAN	AXA	BP	IRS	GER BOND	US BOND	BRENT	GAS	GOLD	SILVER	EUR/USD	GBP/USD	JPY/USD	AUD/USD
N-FGARCH (/1260)	**29**	**42**	**23**	**27**	**23**	**29**	**23**	**18**	**17**	**22**	**21**	**11**	**12**	**38**	**39**	**24**	**15**	**24**	**11**
LR_{uc}	15.760	43.032	6.970	12.522	6.970	15.765	6.970	2.064	1.399	5.794	4.711	0.214	0.029	33.616	35.892	8.234	0.435	8.234	0.214
	(0.000)	(0.000)	(0.008)	(0.000)	(0.008)	(0.000)	(0.008)	(0.151)	(0.237)	(0.016)	(0.030)	(0.643)	(0.864)	(0.000)	(0.000)	(0.004)	(0.509)	(0.004)	(0.643)
LR_{ind}	-	31.228	0.642	2.317	-	-	-	-	10.244	3.651	-	-	-	0.620	4.487	-	-	0.543	-
	(-)	(0.000)	(0.423)	(0.128)	(-)	(-)	(-)	(-)	(0.001)	(0.056)	(-)	(-)	(-)	(0.431)	(0.034)	(-)	(-)	(0.461)	(-)
LR_{cc}	-	74.260	7.611	14.839	-	-	-	-	11.644	9.445	-	-	-	34.236	40.379	-	-	8.777	-
	(-)	(0.000)	(0.022)	(0.000)	(-)	(-)	(-)	(-)	(0.003)	(0.010)	(-)	(-)	(-)	(0.000)	(0.000)	(-)	(-)	(0.012)	(-)
DQT	35.266	574.922	14.936	29.199	25.773	34.730	14.284	7.155	32.477	24.642	18.543	0.777	0.945	73.664	85.997	18.521	2.445	15.993	8.436
	(0.000)	(0.000)	(0.037)	(0.000)	(0.000)	(0.000)	(0.046)	(0.413)	(0.000)	(0.001)	(0.010)	(0.998)	(0.996)	(0.000)	(0.000)	(0.010)	(0.931)	(0.025)	(0.296)
ST-FGARCH (/1260)	**22**	**19**	**21**	**23**	**16**	**21**	**16**	**15**	**15**	**18**	**15**	**7**	**5**	**27**	**22**	**16**	**11**	**14**	**10**
LR_{uc}	5.794	2.841	4.711	6.970	0.854	4.711	0.854	0.435	0.435	2.064	0.435	2.996	6.004	12.522	5.794	0.854	0.214	0.152	0.583
	(0.016)	(0.092)	(0.030)	(0.009)	(0.355)	(0.030)	(0.355)	(0.509)	(0.509)	(0.151)	(0.509)	(0.083)	(0.014)	(0.000)	(0.016)	(0.355)	(0.643)	(0.697)	(0.445)
LR_{ind}	-	1.158	-	3.347	-	-	-	-	10.944	5.102	-	-	-	2.317	7.798	-	-	2.159	-
	(-)	(0.282)	(-)	(0.067)	(-)	(-)	(-)	(-)	(0.001)	(0.024)	(-)	(-)	(-)	(0.128)	(0.005)	(-)	(-)	(0.142)	(-)
LR_{cc}	-	3.999	-	10.316	-	-	-	-	11.379	7.166	-	-	-	14.839	13.592	-	-	2.311	-
	(-)	(0.135)	(-)	(0.006)	(-)	(-)	(-)	(-)	(0.003)	(0.028)	(-)	(-)	(-)	(0.000)	(0.001)	(-)	(-)	(0.315)	(-)
DQT	21.867	13.427	10.154	24.461	27.918	19.582	2.389	6.244	38.790	27.746	7.628	2.592	4.631	39.923	41.665	4.116	1.761	6.500	9.378
	(0.002)	(0.062)	(0.180)	(0.001)	(0.000)	(0.006)	(0.935)	(0.511)	(0.000)	(0.000)	(0.366)	(0.920)	(0.705)	(0.000)	(0.000)	(0.766)	(0.972)	(0.483)	(0.227)
SKST-FGARCH (/1260)	**17**	**17**	**17**	**19**	**16**	**19**	**15**	**12**	**14**	**18**	**15**	**7**	**5**	**12**	**19**	**16**	**10**	**16**	**10**
LR_{uc}	1.399	1.399	1.399	2.841	0.854	2.841	0.435	0.029	0.152	2.064	0.435	2.996	6.004	0.013	2.841	0.854	0.583	0.854	0.583
	(0.237)	(0.237)	(0.237)	(0.092)	(0.355)	(0.092)	(0.509)	(0.864)	(0.697)	(0.151)	(0.509)	(0.083)	(0.014)	(0.909)	(0.092)	(0.355)	(0.445)	(0.355)	(0.445)
LR_{ind}	-	1.502	-	1.158	-	-	-	-	11.342	5.102	-	-	-	12.695	4.699	-	-	1.700	-
	(-)	(0.220)	(-)	(0.282)	(-)	(-)	(-)	(-)	(0.001)	(0.024)	(-)	(-)	(-)	(0.000)	(0.030)	(-)	(-)	(0.192)	(-)
LR_{cc}	-	2.901	-	3.999	-	-	-	-	11.494	7.166	-	-	-	12.708	7.540	-	-	2.554	-
	(-)	(0.234)	(-)	(0.135)	(-)	(-)	(-)	(-)	(0.003)	(0.027)	(-)	(-)	(-)	(0.002)	(0.023)	(-)	(-)	(0.278)	(-)
DQT	14.310	14.800	6.830	10.378	27.990	14.679	1.669	3.255	41.743	27.808	7.688	2.593	4.630	8.288	21.504	4.138	2.515	6.147	9.389
	(0.046)	(0.039)	(0.447)	(0.168)	(0.000)	(0.040)	(0.976)	(0.860)	(0.000)	(0.000)	(0.361)	(0.920)	(0.705)	(0.308)	(0.003)	(0.764)	(0.926)	(0.523)	(0.226)
SGED-FGARCH (/1260)	**18**	**15**	**18**	**15**	**15**	**19**	**15**	**12**	**14**	**16**	**15**	**7**	**5**	**25**	**20**	**16**	**10**	**16**	**10**
LR_{uc}	2.064	0.435	2.064	0.435	0.435	2.841	0.435	0.029	0.152	0.854	0.435	2.996	6.004	9.583	3.725	0.854	0.583	0.854	0.583
	(0.151)	(0.509)	(0.151)	(0.509)	(0.509)	(0.092)	(0.509)	(0.864)	(0.697)	(0.355)	(0.509)	(0.083)	(0.014)	(0.002)	(0.053)	(0.355)	(0.445)	(0.355)	(0.445)
LR_{ind}	-	1.919	-	1.919	-	-	-	-	11.342	6.004	-	-	-	2.797	8.911	-	-	1.700	-
	(-)	(0.166)	(-)	(0.166)	(-)	(-)	(-)	(-)	(0.001)	(0.014)	(-)	(-)	(-)	(0.094)	(0.003)	(-)	(-)	(0.192)	(-)
LR_{cc}	-	2.354	-	2.354	-	-	-	-	11.494	6.858	-	-	-	12.380	12.637	-	-	2.554	-
	(-)	(0.308)	(-)	(0.308)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.002)	(0.001)	(-)	(-)	(0.279)	(0.745)
DQT	14.496	16.116	7.208	7.555	28.103	14.744	1.654	3.248	41.597	24.875	7.532	2.592	4.629	36.749	43.533	4.137	2.481	6.202	9.388
	(0.044)	(0.024)	(0.407)	(0.373)	(0.000)	(0.039)	(0.976)	(0.861)	(0.000)	(0.001)	(0.377)	(0.920)	(0.705)	(0.000)	(0.000)	(0.764)	(0.928)	(0.516)	(0.226)
JSU-FGARCH (/1260)	**16**	**17**	**11**	**14**	**16**	**19**	**15**	**12**	**13**	**16**	**14**	**7**	**6**	**18**	**16**	**16**	**10**	**16**	**9**
LR_{uc}	0.854	1.399	0.214	0.152	0.854	2.841	0.435	0.029	0.013	0.854	0.152	2.996	4.332	2.064	0.854	0.854	0.583	0.854	1.154
	(0.355)	(0.237)	(0.643)	(0.697)	(0.355)	(0.092)	(0.509)	(0.864)	(0.910)	(0.355)	(0.697)	(0.083)	(0.037)	(0.151)	(0.355)	(0.355)	(0.445)	(0.355)	(0.283)
LR_{ind}	-	1.502	-	2.159	-	-	-	-	11.778	6.004	-	-	-	5.102	6.004	-	-	1.700	-
	(-)	(0.220)	(-)	(0.142)	(-)	(-)	(-)	(-)	(0.000)	(0.014)	(-)	(-)	(-)	(0.023)	(0.014)	(-)	(-)	(0.192)	(-)
LR_{cc}	-	2.901	-	2.311	-	-	-	-	11.791	6.858	-	-	-	7.166	6.858	-	-	2.554	-
	(-)	(0.234)	(-)	(0.315)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.027)	(0.032)	(-)	(-)	(0.279)	(-)
DQT	13.951	14.787	1.002	7.681	28.045	14.679	1.668	3.244	25.192	24.805	6.996	2.592	3.526	26.539	23.641	4.122	2.512	6.139	11.130
	(0.052)	(0.039)	(0.994)	(0.361)	(0.000)	(0.040)	(0.976)	(0.861)	(0.001)	(0.001)	(0.429)	(0.919)	(0.832)	(0.000)	(0.001)	(0.765)	(0.926)	(0.523)	(0.133)
SGT-FGARCH (/1260)	**19**	**15**	**17**	**17**	**16**	**19**	**15**	**12**	**14**	**16**	**15**	**7**	**5**	**25**	**19**	**16**	**10**	**16**	**10**
LR_{uc}	2.841	0.435	1.399	1.399	0.854	2.841	0.435	0.029	0.152	0.854	0.435	2.996	6.004	9.583	2.841	0.854	0.583	0.854	0.583
	(0.091)	(0.509)	(0.237)	(0.237)	(0.355)	(0.092)	(0.509)	(0.864)	(0.697)	(0.355)	(0.509)	(0.083)	(0.014)	(0.002)	(0.092)	(0.355)	(0.445)	(0.355)	(0.445)
LR_{ind}	-	1.919	-	1.502	-	-	-	-	11.342	6.004	-	-	-	2.797	4.699	-	-	1.700	-
	(-)	(0.166)	(-)	(0.220)	(-)	(-)	(-)	(-)	(0.001)	(0.014)	(-)	(-)	(-)	(0.094)	(0.030)	(-)	(-)	(0.192)	(-)
LR_{cc}	-	2.354	-	2.901	-	-	-	-	11.494	6.858	-	-	-	12.380	7.540	-	-	2.554	-
	(-)	(0.308)	(-)	(0.234)	(-)	(-)	(-)	(-)	(0.003)	(0.032)	(-)	(-)	(-)	(0.002)	(0.023)	(-)	(-)	(0.279)	(-)
DQT	16.693	15.724	6.773	8.392	27.992	14.762	1.656	3.335	41.571	24.843	7.527	2.592	4.629	36.517	21.492	4.121	2.466	6.188	9.380
	(0.019)	(0.028)	(0.453)	(0.299)	(0.000)	(0.039)	(0.977)	(0.852)	(0.000)	(0.001)	(0.376)	(0.920	(0.705)	(0.000)	(0.003)	(0.766)	(0.930)	(0.518)	(0.226)
GHST-FGARCH (/1260)	**16**	**15**	**9**	**9**	**14**	**9**	**11**	**8**	**12**	**16**	**14**	**5**	**4**	**29**	**21**	**14**	**3**	**12**	**8**
LR_{uc}	0.854	0.435	1.154	1.154	0.152	1.154	0.214	1.949	0.029	0.854	0.152	6.004	8.080	15.765	4.711	0.152	10.663	0.029	1.949
	(0.355)	(0.509)	(0.283)	(0.283)	(0.697)	(0.283)	(0.643)	(0.163)	(0.864)	(0.355)	(0.697)	(0.014)	(0.004)	(0.000)	(0.029)	(0.697)	(0.001)	(0.864)	(0.163)
LR_{ind}	-	1.919	-	3.834	-	-	-	-	-	6.004	-	-	-	1.898	8.337	-	-	2.720	-
	(-)	(0.166)	(-)	(0.050)	(-)	(-)	(-)	(-)	(-)	(0.014)	(-)	(-)	(-)	(0.168)	(0.004)	(-)	(-)	(0.099)	(-)
LR_{cc}	-	2.354	-	4.988	-	-	-	-	-	6.858	-	-	-	17.663	13.048	-	-	2.750	-
	(-)	(0.308)	(-)	(0.082)	(-)	(-)	(-)	(-)	(-)	(0.032)	(-)	(-)	(-)	(0.000)	(0.001)	(-)	(-)	(0.253)	(-)
DQT	13.677	16.272	3.968	77.936	7.052	20.919	1.018	2.404	17.322	25.084	6.544	4.633	5.905	48.544	42.315	2.644	7.529	8.029	13.333
	(0.057)	(0.023)	(0.783)	(0.000)	(0.423)	(0.004)	(0.994)	(0.934)	(0.154)	(0.001)	(0.477)	(0.705)	(0.551	(0.000)	(0.000)	(0.916)	(0.376)	(0.330)	(0.064)

Table 2.5: $VaR_{1\%}$ results with FGARCH volatility model for all assets: unconditional coverage test, independence and conditional coverage tests and Dynamic Quantile Test. P-values are reported in parentheses. For each model, the first line shows the number of VaR excesses produced with the different assets (boldface).

	IBEX35		NASDAQ100		FTSE100		NIKKEI225	
	QLF	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**
N-GARCH	1.680	4.668	1.263	3.544	0.848	3.145	6.141	5.095
ST-GARCH	1.115	4.519	0.966	3.432	0.609	3.024	5.730	5.029
SKST-GARCH	0.873	4.504	0.818	3.397	0.479	2.985	5.365	4.995
SGED-GARCH	0.842	4.499	0.733	3.368	0.461	2.980	4.939	4.956
JSU-GARCH	0.799	4.499	0.763	3.384	0.430	2.977	5.082	4.974
SGT-GARCH	0.847	4.494	0.732	3.364	0.442	2.965	4.955	4.951
GHST-GARCH	0.478	4.520	0.418	3.400	0.260	3.003	4.230	5.002
N-GJRGARCH	1.517	4.523	0.747	3.356	0.484	2.932	5.388	4.896
ST-GJRGARCH	1.140	4.429	0.536	3.302	0.339	4.958	4.958	4.804
SKST-GJRGARCH	0.890	4.403	0.348	3.256	0.210	2.804	4.557	4.769
SGED-GJRGARCH	0.876	4.394	0.292	3.242	0.201	2.800	**4.228**	4.746
JSU-GJRGARCH	0.829	4.403	0.296	3.249	0.179	2.801	4.311	4.760
SGT-GJRGARCH	0.888	4.395	0.311	3.238	0.208	2.792	4.296	4.747
GHST-GJRGARCH	**0.457**	4.602	**0.176**	3.334	**0.141**	3.030	4.340	4.785
N-APARCH	1.428	4.485	0.782	3.322	0.500	2.911	5.994	4.960
ST-APARCH	1.052	4.387	0.586	3.258	0.366	2.836	5.587	4.863
SKST-APARCH	0.793	4.338	0.398	3.228	0.239	2.788	5.169	4.809
SGED-APARCH	0.784	4.327	0.338	3.216	0.230	2.783	4.845	4.777
JSU-APARCH	0.735	4.335	0.347	3.219	0.210	2.789	4.924	4.792
SGT-APARCH	0.801	4.328	0.353	3.216	0.241	2.783	4.915	4.778
GHST-APARCH	**0.457**	4.340	**0.176**	3.255	**0.141**	2.833	4.340	4.807
N-FGARCH	1.299	4.453	1.761	4.075	0.511	2.908	6.234	4.921
ST-FGARCH	1.047	4.376	0.749	3.257	0.370	2.840	5.928	4.830
SKST-FGARCH	0.785	4.326	0.574	3.235	0.230	2.780	5.531	4.759
SGED-FGARCH	0.754	**4.309**	0.543	3.253	0.225	**2.776**	5.181	**4.713**
JSU-FGARCH	0.738	4.330	0.596	3.242	0.201	2.778	5.307	4.736
SGT-FGARCH	0.767	4.312	0.447	**3.211**	0.237	2.779	7.403	5.412
GHST-FGARCH	0.772	4.318	0.585	3.273	0.198	2.861	4.845	4.796

Table 2.6: $VaR_{1\%}$ loss functions for stock market indices: Quadratic Loss Function (QLF) and Asymmetric Linear Tick Loss Function (AlTick).

	IBM		SAN		AXA		BP	
	QLF	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**
N-GARCH	9.127	5.417	11.834	7.181	4.090	6.694	3.390	4.591
ST-GARCH	8.025	5.370	10.158	7.042	3.053	6.562	2.805	4.577
SKST-GARCH	7.966	5.372	9.663	7.038	2.975	6.565	2.716	4.574
SGED-GARCH	7.785	5.358	9.564	7.043	2.917	6.576	2.672	4.567
JSU-GARCH	7.830	5.373	9.409	7.036	2.882	6.570	2.665	4.574
SGT-GARCH	7.854	5.535	8.912	6.863	2.655	6.672	2.234	4.536
GHST-GARCH	6.970	5.570	7.274	6.941	0.402	6.916	1.377	4.741
N-GJRGARCH	9.146	5.527	10.893	7.019	3.848	6.779	2.858	4.513
ST-GJRGARCH	8.120	5.542	9.523	6.905	2.829	6.666	2.379	4.538
SKST-GJRGARCH	8.020	5.543	8.904	6.862	2.603	6.668	2.239	4.542
SGED-GJRGARCH	7.804	5.536	8.817	6.857	2.585	6.670	2.203	4.533
JSU-GJRGARCH	7.879	5.546	8.668	6.848	2.508	6.667	2.192	4.544
SGT-GJRGARCH	7.854	5.535	8.912	6.863	2.655	6.672	2.234	4.536
GHST-GJRGARCH	6.970	5.570	7.274	6.941	**0.402**	6.916	**1.377**	4.741
N-APARCH	8.849	5.542	11.265	6.969	3.269	6.660	2.822	**4.482**
ST-APARCH	7.623	5.450	9.977	6.813	2.330	6.511	2.374	4.506
SKST-APARCH	7.485	5.446	9.322	6.771	2.084	6.498	2.213	4.500
SGED-APARCH	7.348	5.443	9.102	6.758	2.003	6.494	2.165	4.499
JSU-APARCH	7.348	5.443	9.102	6.758	2.003	6.494	2.165	4.499
SGT-APARCH	7.333	5.422	9.283	6.766	2.062	6.493	2.172	4.489
GHST-APARCH	6.970	5.451	7.274	6.773	**0.402**	6.471	**1.377**	4.667
N-FGARCH	8.814	5.529	10.917	6.918	3.004	6.535	2.823	4.517
ST-FGARCH	7.607	5.429	9.622	6.733	2.239	6.445	2.425	4.522
SKST-FGARCH	7.444	5.421	8.970	6.667	1.974	6.435	2.240	4.516
SGED-FGARCH	7.311	5.397	8.962	6.672	1.919	**6.412**	2.194	4.505
JSU-FGARCH	7.307	5.417	8.764	**6.650**	1.892	6.429	2.190	4.518
SGT-FGARCH	**5.230**	**4.716**	9.076	6.687	1.985	6.417	2.250	4.512
GHST-FGARCH	7.048	5.414	**7.137**	6.697	0.454	6.454	1.403	4.664

Table 2.7: $VaR_{1\%}$ loss functions for individual stocks: Quadratic Loss Function (QLF) and Asymmetric Linear Tick Loss Function (AlTick).

	IRS 5Y		GERMANBOND 10Y		USBOND 10Y	
	QLF	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**
N-GARCH	0.101	0.652	0.331	1.520	0.222	1.679
ST-GARCH	0.083	0.647	0.268	1.489	0.139	1.632
SKST-GARCH	0.076	0.649	0.234	1.481	0.110	1.626
SGED-GARCH	0.075	0.646	0.237	1.485	0.114	1.629
JSU-GARCH	**0.072**	0.649	0.225	1.479	0.102	1.626
SGT-GARCH	0.074	**0.644**	0.241	1.486	0.114	1.628
GHST-GARCH	0.076	0.654	0.203	1.480	**0.088**	**1.625**
N-GJRGARCH	0.104	0.660	0.330	1.520	0.245	1.698
ST-GJRGARCH	0.085	0.658	0.270	1.496	0.165	1.652
SKST-GJRGARCH	0.078	0.660	0.236	1.485	0.134	1.648
SGED-GJRGARCH	0.078	0.657	0.238	1.487	0.134	1.648
JSU-GJRGARCH	0.075	0.660	0.227	1.483	0.125	1.648
SGT-GJRGARCH	0.077	0.655	0.241	1.488	0.135	1.649
GHST-GJRGARCH	0.108	0.663	0.209	1.493	0.130	1.657
N-APARCH	0.106	0.660	0.324	1.513	0.242	1.697
ST-APARCH	0.119	0.665	0.274	1.502	0.169	1.654
SKST-APARCH	0.110	0.666	0.240	1.490	0.137	1.651
SGED-APARCH	0.099	0.661	0.237	1.487	0.135	1.649
JSU-APARCH	0.106	0.666	0.231	1.486	0.127	1.651
SGT-APARCH	0.098	0.660	0.241	1.488	0.135	1.650
GHST-APARCH	0.108	0.670	0.209	1.484	0.130	1.656
N-FGARCH	0.103	0.658	0.307	1.498	0.241	1.693
ST-FGARCH	0.121	0.665	0.265	1.493	0.163	1.658
SKST-FGARCH	0.112	0.665	0.231	1.481	0.132	1.652
SGED-FGARCH	0.096	0.659	0.227	**1.476**	0.131	1.652
JSU-FGARCH	0.106	0.665	0.222	1.476	0.122	1.651
SGT-FGARCH	0.096	0.658	0.229	1.476	0.131	1.652
GHST-FGARCH	0.180	0.730	**0.203**	1.483	0.093	1.629

Table 2.8: $VaR_{1\%}$ loss functions for interest rates: Quadratic Loss Function (QLF) and Asymmetric Linear Tick Loss Function (AlTick).

	OIL BRENT		GAS		GOLD		SILVER	
	QLF	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**
N-GARCH	4.887	5.500	3.188	8.480	7.298	4.748	15.631	8.630
ST-GARCH	3.716	5.426	1.483	8.872	5.609	4.394	11.606	8.152
SKST-GARCH	3.447	5.453	1.621	8.838	5.422	4.376	10.066	8.004
SGED-GARCH	3.498	5.450	1.438	8.942	5.675	4.394	10.310	8.034
JSU-GARCH	3.319	5.467	1.681	8.821	5.263	4.369	9.075	7.948
SGT-GARCH	3.456	5.447	1.432	8.946	5.644	4.385	9.960	8.005
GHST-GARCH	2.681	5.599	0.703	9.743	5.536	4.353	10.035	7.960
N-GJRGARCH	4.108	5.306	3.215	8.485	8.361	4.896	18.011	8.772
ST-GJRGARCH	3.004	5.327	1.503	8.878	6.582	4.484	14.520	8.291
SKST-GJRGARCH	2.703	5.382	1.651	8.843	6.412	4.465	13.025	8.147
SGED-GJRGARCH	2.783	5.373	1.459	8.938	6.583	4.471	12.983	8.154
JSU-GJRGARCH	2.574	5.410	1.715	8.826	6.190	4.451	12.036	8.064
SGT-GJRGARCH	2.779	5.373	1.450	8.943	6.545	4.461	**8.823**	**7.097**
GHST-GJRGARCH	2.888	5.785	0.450	9.766	7.295	4.459	13.582	8.020
N-APARCH	4.128	5.310	3.128	8.352	7.012	4.432	19.479	9.022
ST-APARCH	2.984	5.335	1.396	8.662	7.208	4.481	14.305	8.299
SKST-APARCH	2.670	5.399	1.553	8.621	7.012	4.432	12.629	8.128
SGED-APARCH	2.771	5.381	1.407	8.745	7.198	4.462	13.272	8.211
JSU-APARCH	2.546	5.426	1.628	8.603	6.862	4.406	11.872	8.086
SGT-APARCH	2.774	5.382	1.394	8.754	7.151	4.448	12.921	8.167
GHST-APARCH	2.888	5.498	0.450	9.232	7.295	4.493	13.582	8.226
N-FGARCH	3.597	**5.288**	2.920	**8.288**	8.315	4.935	18.720	9.139
ST-FGARCH	2.791	5.401	1.282	8.580	6.958	4.470	14.409	8.261
SKST-FGARCH	8.527	5.466	1.445	8.525	**2.433**	**3.875**	12.668	8.092
SGED-FGARCH	2.582	5.464	1.293	8.642	6.918	4.474	13.290	8.147
JSU-FGARCH	2.408	5.495	1.502	8.500	6.615	4.408	11.902	8.032
SGT-FGARCH	2.564	5.464	1.274	8.653	6.882	4.457	12.969	8.110
GHST-FGARCH	**1.764**	5.680	**0.450**	9.217	7.114	4.503	13.717	8.192

Table 2.9: $VaR_{1\%}$ loss functions for commodities: Quadratic Loss Function (QLF) and Asymmetric Linear Tick Loss Function (AlTick).

	EUR/USD		GBP/USD		JPY/USD		AUD/USD	
	QLF	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**	**QLF**	**AlTick**
N-GARCH	0.209	1.814	0.050	1.271	1.283	2.221	0.784	2.172
ST-GARCH	0.137	1.761	0.032	1.273	1.028	2.198	0.658	2.184
SKST-GARCH	0.145	1.767	0.028	1.278	1.117	2.200	0.556	2.217
SGED-GARCH	0.136	1.762	0.025	1.281	1.100	2.198	0.546	2.218
JSU-GARCH	0.143	1.767	0.027	1.280	1.136	2.201	0.526	2.229
SGT-GARCH	0.135	1.759	0.026	1.280	1.084	2.198	0.548	2.210
GHST-GARCH	**0.051**	**1.745**	0.015	1.347	**0.882**	2.212	0.493	2.257
N-GJRGARCH	0.198	1.805	0.037	**1.259**	1.281	2.240	0.701	2.167
ST-GJRGARCH	0.130	1.759	0.023	1.265	1.022	2.200	0.586	2.174
SKST-GJRGARCH	0.138	1.766	0.020	1.269	1.114	2.205	0.487	2.200
SGED-GJRGARCH	0.132	1.761	0.017	1.274	1.097	2.203	0.474	2.199
JSU-GJRGARCH	0.136	1.765	0.019	1.272	1.135	2.209	0.459	2.210
SGT-GJRGARCH	0.128	1.757	0.018	1.273	1.080	2.203	0.477	2.193
GHST-GJRGARCH	0.097	1.771	**0.013**	1.410	0.938	2.206	**0.383**	2.283
N-APARCH	0.199	1.807	0.038	1.261	1.266	2.243	0.648	**2.151**
ST-APARCH	0.131	1.761	0.025	1.268	1.014	2.186	0.524	2.160
SKST-APARCH	0.139	1.767	0.021	1.272	1.105	2.196	0.433	2.185
SGED-APARCH	0.133	1.762	0.018	1.277	1.082	2.189	1.215	2.667
JSU-APARCH	0.137	1.766	0.020	1.275	1.124	2.201	0.407	2.195
SGT-APARCH	0.129	1.758	0.019	1.276	1.063	2.186	0.424	2.180
GHST-APARCH	0.097	1.762	**0.013**	1.328	0.938	**2.181**	**0.383**	2.213
N-FGARCH	0.194	1.793	0.039	1.267	1.220	2.236	0.651	**2.151**
ST-FGARCH	0.128	1.746	0.026	1.273	0.995	2.188	0.529	2.162
SKST-FGARCH	0.136	1.749	0.023	1.278	1.085	2.198	0.437	2.189
SGED-FGARCH	0.131	1.748	0.019	1.282	1.063	2.190	0.429	2.189
JSU-FGARCH	0.134	1.748	0.021	1.280	1.105	2.202	0.425	2.202
SGT-FGARCH	0.126	1.746	0.020	1.281	1.044	2.189	0.429	2.182
GHST-FGARCH	0.107	1.755	0.014	1.339	0.922	2.181	0.394	2.233

Table 2.10: $VaR_{1\%}$ loss functions for exchange rates: Quadratic Loss Function (QLF) and Asymmetric Linear Tick Loss Function (AlTick).

Chapter 3

Testing ES estimation models: An extreme value theory approach

Abstract

We investigate whether there is a pattern regarding the quality of several models and methods in Expected Shortfall (ES) estimation for a set of assets, considering a variety of significance levels. We use conditional models applied to the full distribution and also models that focus on extreme events through the extreme value theory (EVT) approach, following the two-step procedure of McNeil & Frey (2000). We assess the performance of the models using different ES backtests recently proposed in the literature. Our results suggest that the conditional EVT-based models produce a better 1-day ES performance compared with conditional models with asymmetric probability distributions for return innovations. We find that these results are also valid for the recent crisis period. They are also robust to considering 10-day ES, where the conditional EVT models are again more accurate and reliable for predicting asset risk losses.

3.1 Introduction

The Basel Committee on Banking Supervision (BIS) has recently sanctioned Expected Shortfall (ES) as the market risk measure to be used for banking regulation purposes, replacing the well-known Value-at-Risk (VaR). This change is motivated by the appealing theoretical properties of ES as a measure of risk and the poor properties of VaR. In particular, VaR fails to control for "tail risk". In this transition, the major challenge faced by financial institutions is the unavailability of simple tools for evaluation of ES forecasts (i.e. backtesting ES). In fact, the Basel Committee backed down on requiring the backtesting of ES. A debate, started by Gneiting [49] led many to believe that ES could not be backtested because it was not "elicitable". That point was settled recently by Fissler, Ziegel and Gneiting [45] and by Acerbi [4]; they demonstrate that lack of elicitability is not an impediment to backtesting ES. The latest Basel consultative document of January 2016 [13], however, proposed to calculate risk and capital using ES, but to conduct backtesting only on VaR. The backtests are applied comparing whether the observed percentage of outcomes covered by the risk measure is consistent with the intended level of coverage. However, it is important that the capital reserve indicated by the VaR calculation could be tested, and the hypothesis that the level of reserves is adequate could be subject to a valid statistical test.

There is not much work evaluating and comparing the performance of ES estimation models using recently introduced ES backtesting. Alexander and Sheedy (2008) [2] develop a methodology for conducting stress tests in the context of a risk model, proposing a two-stage approach whereby an initial shock event is linked to the probability of its occurrence. The risk model is used to model the consequences of that shock event in simulation. They implemented this stress testing procedure for three major currency pairs and found that results compared favorably with the traditional historical scenario stress testing approach in all but one extraordinary case. Jalal and Rockinger (2008) [62] use the circular block bootstrap, consisting in wrapping the data around a circle. This ensure that each of the original observations has an equal chance of appearing in a simulated series and that all the blocks have the same length. This method is adequate to take into account the possible dependency among the exceedances. Focusing on ES forecasts, obtained for the two step procedure of McNeil and Frey (2000) [75], they find that one cannot reject the assumption that the forecasted ES measure captures actual shortfalls in a satisfactory manner. Ergün and Jun (2010) [42], apply the Autoregressive Conditional Density (ARCD) model of Hansen (1994), which allows time-variation in higher-order conditional moments, to five minute stock index futures returns and examine their out-of-sample one-step-ahead VaR and ES forecast performance for long and short positions. They also estimate other GARCH-based models and a model based on the extreme value theory (EVT), following the two-step approach of McNeil and Frey (2000). Estimation results show that the ARCD model with a time-varying conditional skewness parameter seems to provide more accurate ES forecasts. Other studies have VaR as their primary measure of interest, leaving ES to a second level, such as Venter and Jongh (2004) [92], Marinelli, D'Addona and Rachev (2007) [74], Zhou (2012) [96], Degiannakis, Floros and Dent (2013) [32] and Tolikas (2014) [91], where no extensive focus is placed on ES forecasting patters.

Wong, Fan and Zeng (2012) [93] compare ES estimation models considering only the saddlepoint backtest proposed by Wong (2008) [95]. Righi and Ceretta (2015) [83] evaluate unconditional, conditional and quantile/expectile regression-based models for ES estimation using only the ES backtest proposed by McNeil and Frey (2000) [75] and a proposed test based on the standard deviation of returns beyond VaR. Clift, Costanzino and Curran (2016) [23] apply three approaches recently proposed in the literature for backtesting ES, i.e. Wong (2008), Acerbi & Szekely (2014) and Costanzino & Curran (2015) approaches, but they only use GARCH volatility and a Normal distribution for ES estimation.

We estimate the conditional Expected Shortfall at 1-day and 10-day horizons based on the EVT approach using asymmetric probability distributions for return innovations, and we analyze the accuracy of our estimates before and during the 2008 financial crisis using daily data. We take into account volatility clustering and leverage effects in return volatility by using the APARCH model (Ding, Granger and Engle, 1993 [35]) under different probability distributions assumed for the standardized innovations: Gaussian, Student-t, skewed Student-t [Fernandez and Steel (1998) [43]], skewed generalized error [Fernandez and Steel (1998) [43]] and Johnson S_U [Johnson (1949) [63]] and following the two-step procedure of McNeil & Frey (2000) [75]. This two-step procedure fits a generalized Pareto distribution to the extreme values of the standardized residuals generated by an APARCH model. Then, we compare the out-of-sample one-step-ahead ES forecast performance of all these models. For ES evaluation, we use the most recent ES backtesting proposals, which overcome the limitations of previous tests [McNeil & Frey (2000) [75], Berkowitz (2001) [15], Kerkhof and Melenberg (2004) [65] and Wong (2008) [95]]. These are the test of Righi & Ceretta (2013) [83], the first two tests of Acerbi & Szekely (2014) [4], which are straightforward but require simulation analysis (like the Rigui & Ceretta test), the test of Graham & Pál (2014) [57] which is an extension of the Lugannani-Rice approach of Wong (2008) [95], the quantile-space unconditional coverage test of Costanzino & Curran (2015) [26] for the family of Spectral Risk Measures, of which ES is a member and, finally, the conditional test of Du & Escanciano (2015) [36]. The last two tests can be thought of as the continuous limit of the Emmer, Kratz & Tasche (2013) [39] idea in that it is a joint test of a continuum of VaR levels.

Implementation of the EVT for ES estimation has rarely been applied beyond a one-day horizon when estimating the ES of financial assets, even though there are several economic and practical reasons for computing long-term risk measures. Risk horizons longer than one day are particularly important for risk liquidity management, for long term strategic asset allocation and for capital requirement and the Basel Committee obliges banks to compute their level of risk over a ten-day horizon. The difficulty is that it is hard enough data on 10-day returns over non-overlapping periods. With this motivation we use Filtered Historical Simulation (FHS) to obtain time series of 10-day returns and we estimate the 10-day ES by applying the same methodology we have used to estimate the 1-day ahead ES. That way we avoid the limitations of the scaling law.

To sum up, this work contributes to the literature in four ways. First, we use the APARCH volatility specification in an EVT model and in Filtered Historical Simulation

(FHS) to take into account volatility clustering and asymmetric returns. Second, we compare conditional EVT models that incorporate conditional models with asymmetric probability distributions rarely used in the financial literature to calculate ES. Third, we calculate VaR and ES over 10-day horizons for risk liquidity management and Basel capital requirements. Finally, we focus on the accuracy of our risk models for VaR and ES estimation during pre-crisis and crisis periods as well as under different significance levels (α).

The remainder of the paper is organized as follows. In Section 3.2 we present a review of the literature. In Section 3.3 we describe the mathematical properties of the standard risk measures and explain the backtesting approaches used in our analysis. In Section 3.4 we present the procedure to estimate ES models, including McNeil & Frey procedure for EVT. In Section 3.5 we present preliminary statistics for our daily data and parameter estimates. In Section 3.6 we report the results of the empirical investigation of the estimated 1-day ES models and 1-day ES backtesting and in Section 3.7 we report the results obtained after dividing the sample in pre-crisis and crisis periods. In Section 3.8 we provide a description of the different methods to calculate ES for risk horizons longer than one-day and we assess 10-day ES performance. Finally, Section 3.9 concludes the paper.

3.2 Review of Literature

The quantiles of the distribution of returns (VaR) can be estimated by extreme value theory (EVT), which models the tails of the distribution of returns without making any specific assumption concerning the center of the distribution (Rocco, 2014 [84]). The tail index parameter in EVT can be estimated nonparametrically without assuming any particular model for the tail. There are many estimators that can be used to accomplish this task, such as Hill estimator (Hill, 1975) [60] and Pickands estimator (Pickands, 1975) [80]. In practice, the number of data points in the tails are limited, leading to small sample biases. To address this problem, many solutions have been proposed by Huisman, Koedijk, Kool and Palm (2001) [61], Gomes, de Haan and Rodrigues (2008) [51], Gomes, Figueiredo, Rodrigues and Miranda (2012) [52], Gomes, Matins and Neves (2007) [53] and Gomes and Pestana (2007) [54]. Gourieroux and Jasiak (2010a) [56] point out that the accuracy of these non-parametric estimators is rather poor, due to the difficulty of estimating the probability of infrequent events. Another problem is that these estimators depend on the number of observations in a very erratic way.

For the estimation of the tail index parameter in EVT there are also two parametric approaches. The parameter of the distribution of extremes, including tail index, are directly estimated by classical methods such as maximum likelihood. The first parametric approach is Block Maxima (BM) based on the Generalized Extreme Value (GEV), which divides the sample into m subsamples of n observations each, and picks the maximum of each subsample; see for example Longin (2000) [70], Diebold, Schuermann and Stroughair (2000) [34]. The second EVT parametric approach is the Peak Over Threshold (POT) based on the Generalized Pareto Distribution, according to which any observations that exceeds a given high threshold, u, are modelled separately from non-extreme observations. McNeil and Frey (2000) [75] show that the EVT method based on the General Pareto distribution yields quantile estimates that are more stable than those from the Hill estimator. Any EVT approach entails choosing an adequate cut-off between the central part of the distribution and the tails. When working with threshold exceedances the cut-off is induced by the number of observations in the tail while in the block maxima procedure it is implied by the choice of the number of blocks. The choice of the cut-off may have severe consequences for the risk estimates. If it is too low, the VaR forecasts will be biased and the asymptotic limit theorems do not apply. Conversely, if the threshold is too large, the VaR forecasts will have large standard deviations due to the limited number of observations over the threshold. Danielsson, de Haan, Peng and de Vries (2001) [29] and Ferreira, de Haan, and Peng (2003) [44] develop bootstrap methods for optimal threshold selection in the context of the Hill and GPD estimators, respectively. The former authors choose the threshold by minimizing the asymptotic MSE of the Hill estimator. However, the selection of the threshold using bootstrap procedures is very time consuming. Alternatively, Gonzalo and Olmo (2004) [55] propose a single-step approach to threshold selection. Gençay and Selçuk (2004) [47] determine the threshold using a combination of the mean excess function and the Hill plots. Chavez-Demoulin, Embrechts and Sardy (2014) [19] propose the inclusion of a sensitivity analysis across several threshold values for a full POT application. As another alternative, Li, Peng and Yang (2010) [69] propose choosing the threshold in such a way as to reduce the bias of the tail index estimator.

Alternatively, it is possible to calculate the conditional quantile. Based on parametric methods, the most popular option to calculate the conditional quantile is assuming a particular distribution for return innovations. The most popular parametric distribution for standardized returns are Gaussian and Student-t distributions, the Skewed Student-t distribution of Hansen (1994) [59] [see Ardia and Hoogerheide (2014) [6], Bali and Theodossiou (2007) [9], Giot and Laurent (2003) [48], Halbleib and Pohlmeier (2012) [58], Kuester et al. (2006) [67], Louzis et al. (2013) [71], Pérignon and Smith (2010a) [79] and Sajjad et al. (2008) [87] for applications using these distributions]. An alternative leptokurtic and asymmetric distribution that has been considered in this context is the Skewed-Generalized-t (SGT) distribution proposed by Theodossiou (1998) [90] [see Bali, Mo and Tang (2008) [8], Bali and Theodossiou (2007) [9] and Cheng and Hung (2011) [21] for applications of the SGT distribution to VaR forecasting]. The SGT distribution has the attractive feature of encompassing most of the distributions that are usually assumed for standardized returns, such as Gaussian, Generalized Error Distribution (GED), Student-t and Skewed Student-t distributions, for example. Recently, Ergen (2015) [41] has considered the Skewed-t distribution proposed by Azzalini and Capitanio (2003) [7] and Aas and Haff (2016) [1] propose to use the Generalized Hyperbolic Skew Student-t distribution for unconditional and conditional VaR forecasting.

Another option is to calculate the conditional quantile using the EVT approach. Danielsson and de Vries (2000) [30] and McNeil and Frey (2000) [75] propose to estimate the quantiles of the innovations by applying EVT to the standardized returns, which are i.i.d. if the conditional mean and variance are specified correctly. Chan and Gray (2006)

[18] introduce a description of the conditional EVT and its application to the forecasting of the VaR of daily electricity prices. In particular, McNeil and Frey (2000) [75] propose first filtering the returns from estimating a GARCH model, then applying EVT to the tails of the innovations while bootstrapping to the central part of the distribution. They verify that the General Pareto distribution of EVT results in better estimates for ES than the Gaussian model. Jalal and Rockinger (2008) [62] show that this procedure appears to perform a remarkable job when combined with a well-chosen threshold estimation, such as that in Gonzalo and Olmo (2004) [55].

Based on non-parametric methods, the quantiles of the innovations can be also estimated using bootstrap methods that do not assume any particular distribution (Ruiz and Pascual, 2002 [86]). In particular, Barone-Adesi, Giannopoulus and Vosper (1999, 2002) [11] [12] propose a bootstrap method known as filtered historical simulation (FHS), which is based on the idea of using random draws with replacement from the standardized residuals and does not incorporate parameter uncertainty; see Engle (2003) [40] and Pritsker (2006) [81] for implementations. Pascual, Ruiz and Romo (2006) [78] propose a bootstrap procedure that allows for the incorporation of parameter uncertainty. Bootstrap procedures have the advantage that they allow for the construction of confidence intervals for VaR estimates. Kourema et al. (2011) [66] compare unconditional and conditional historical simulation and EVT in VaR and ES estimation. They conclude that conditional EVT model is more accurate and reliable for VaR forecasting, according to the rate of violations and Wald, Kupiec and Christoffersen tests, and for ES forecasting, according to an ES test proposed by them based on average difference between the realized returns and the forecast ES.

With respect to ES backtesting Berkowitz (2001) [15], Kerkhof & Melenberg (2004) [65] and Wong (2008) [95] proposed backtesting of risk measures based on size tail losses. While Berkowitz's censored Gaussian approach and Kerkhof & Melenberg's functional delta method rely on large samples for convergence to the required limiting distributions, the saddlepoint techniques proposed by Wong are accurate and have reasonable test power even if the sample size is small. The saddlepoint technique makes use of a small sample asymptotic method that involves higher order moments of the underlying distribution and is able to approximate to a very high degree of accuracy the required tail probability even for very small sample sizes. But this test has a few disadvantages, such as the Gaussian distribution assumption and the full distribution conditional standard deviation that is used as the dispersion measure. To overcome these limitations, Emmer, Kratz & Tasche (2013) [39] propose a new ES backtest based on a simple approximative approach to the backtesting of ES from a representation of ES based on several VaR levels, by Righi & Ceretta (2013) [82], which verifies whether the average deviation from the ES estimate is zero but they consider the dispersion only for the exceptions rather than for the full sample. Later, Acerbi & Szekely (2014) [4] introduce three model-free, non-parametric backtest methodologies for ES that are shown to be more powerful than the Basel VaR test. Graham & Pál (2014) [57] generalize Wong's result in a tractable and intutive manner to allow for any VaR modeling, and therefore distributional, approach. Costanzino & Curran (2015) [26] developed a methodology that can be used to backtest any spectral risk measure, including ES. It is based on the idea that ES is an average of a continuum of VaR levels. They explain an unconditional ES backtest. Later, Du & Escanciano (2015) [36] propose backtest for ES based on cumulative violations, which are the natural analogue of the commonly used conditional backtest for VaR, extending the results obtained by Costanzino & Curran (2015) [26].

3.3 Background

3.3.1 The Mathematical Properties of Risk Measures

(The mathematical properties of risk measures have been extracted from Emmer, Kratz and Tasche (2015) [39] and Ziegel (2016) [97].)

3.3.1.1 Coherence and Related Properties

Artzner et al. [3] state four axioms which any risk measure used for effective risk regulation and management should satisfy. Such risk measures are then said to be coherent. Coherence is a fundamental concept related to the acceptability of a risk measure.

A risk measure ρ is called **coherent** if it satisfies the following conditions,

- Homogeneity. ρ is **homogeneous** if for all variables Y and $h \geq 0$ it holds that $\rho(hY) = h\rho(Y)$.
- Subadditivity. ρ is **subadditive** if for all variables Y_1 and Y_2 it holds that $\rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$.
- Monotonicity. ρ is **monotonic** if for all variables Y_1 and Y_2 it holds that $Y_1 \leq Y_2 \Rightarrow \rho(Y_1) \leq \rho(Y_2)$
- Translation invariance. ρ is **translation invariant** if for all variables Y and $a \in \mathbb{R}$ it holds that $\rho(Y - a) = \rho(Y) - a$.

The concept of **convex** measure of risk is an extension of that of coherent risk measure. A risk measure is convex if it satisfies the condition of monotonicity, translation and convexity defined as

$$\rho(\lambda Y_1 + (1-\lambda)Y_2) \leq \lambda\rho(Y_1) + (1-\lambda)\rho(Y_2)$$

for any $\lambda \in (0,1)$. It is also assumed that $\rho(0) = 0$ as a convenient normalization.

Here, we can notice the direct link with the notion of diversification: the risk of a diversed portfolio, which in this case is $\lambda Y_1 + (1-\lambda)Y_2$, is less or equal to the weighted average of individual risk. Accordingly, it is not surprising that any positive homogeneous and subadditive risk measure is also convex.

Comonotonic additivity is another property of risk measures that is mainly of interest as a complementary property to subadditivity.

Two real-valued random variables Y_1 and Y_2 are said comonotonic if there exist a real-valued random variable X (the common risk factor) and monotonic non-decreasing functions f_1 and f_2 such that

$$Y_1 = f_1(X) \qquad \text{and} \qquad Y_2 = f_2(X)$$

This simply means that, for instance, two risky positions Y_1 and Y_2 are perfectely and also positively dependent on the same source of risk X. Comonotonicity may be considered the strongest possible dependence of random variables (Embrechts et al. [37]).

A risk measure ρ is **comonotonically additive** if for any comonotonic random variables Y_1 and Y_2 it holds that

$$\rho(Y_1 + Y_2) = \rho(Y_1) + \rho(Y_2)$$

The reason why this property extremely matters, is intuitive and related to the notion of diversification.

As a matter of fact, if two different risky positions perfectly depend on the same risk factor, they should not benefit from diversification effects.

Thus, in risk management, we should always use comonotonic additive risk measures.

A risk measure ρ is law-invariant if it depends entirely on the distribution of the random variable associated to it. More precisely, let two random variables Y_1 and Y_2 and their corresponding distribution functions F_{Y_1}, F_{Y_2}, a risk measure $\rho(\cdot)$ is a **law-invariant** risk measure if

$$F_{Y_1} = F_{Y_2} \Rightarrow \rho(Y_1) = \rho(Y_2)$$

A direct consequence of this fact is that, whenever we deal with risk measures that are not law-invariant, we could not evaluate the riskiness of a position through the loss distribution.

3.3.1.2 Elicitability

An interesting criterion when estimating a risk measure is elicitability, introduced by Osband [77] and Lambert et al. [68], then by Gneiting [49]. We briefly recall its definition, which is linked to the one of scoring function. It is also relevant a recent review on probabilistic forecasting, including the notion of elicitability, by Gneiting and Katzfuss [50]. For the definition of elicitability we first introduce the concept of strictly consistent scoring functions.

A scoring function aims at assigning a numerical score to a single-valued point forecast based on the predictive point and realization. A **scoring function** is a function

$$\begin{array}{rccl} s: & \mathbb{R}\times\mathbb{R} & \to & [0,\infty) \\ & (x,y) & \to & s(x,y) \end{array}$$

where x and y are the point forecasts and observations respectively.

Let ν be a functional on a class of probability measures $\mathcal{P}$ on $\mathbb{R}$,

$$\begin{array}{rccl} \nu: & \mathcal{P} & \to & 2^{\mathbb{R}} \\ & P & \mapsto & \nu(P) \subset \mathbb{R} \end{array}$$

A scoring function $s: \quad \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is **consistent** for the functional ν relative to the class $\mathcal{P}$ if and only if, for all $P \in \mathcal{P}$, $t \in \nu(P)$ and $x \in \mathbb{R}$,

$$\mathbb{E}_P[s(t, Y)] \leq \mathbb{E}_P[s(x, Y)]$$

Y being the random variable defined on $(\Omega, \mathcal{F}, P)$.

The function s is **strictly consistent** if it is consistent and

$$\mathbb{E}_P[s(t, Y)] = \mathbb{E}_P[s(x, Y)] \Rightarrow x \in \nu(P)$$

The functional ν is **elicitable** relative to $\mathcal{P}$ if and only if there is a scoring function s which is strictly consistent for ν relative to $\mathcal{P}$.

Elicitability is a helpful criterion for the determination of optimal point forecasts: the class of (strictly) consistent scoring functions for a functional is identical to the class of functions under which (only) the functional is an optimal point forecast. Hence, if we have found a strictly consistent scoring function for a functional ν, we can determine the optimal forecast $\hat{x}$ for $\nu(P)$ by

$$\hat{x} = arg \min_x \mathbb{E}_P[s(x, Y)]$$

Hence elicitability of a functional of probability distributions may be interpreted as the property that the functional can be estimated by generalized regression. Another property, that makes elicitability an important concept, is that it can be used for comparing the performance of different forecast methods.

3.3.1.3 Conditional Elicitability

So far we have only distinguished between elicitable and non-elicitable functionals. However, it turns out that some useful risk measures are not elicitable but "2nd order" elicitable in the following sense.

A functional ν of $\mathcal{P}$ is called **conditionally elicitable** if there exist functionals $\widetilde{\gamma}$ and $\gamma : \mathcal{D} \to 2^{\mathbb{R}}$ with $\mathcal{D} \subset \mathcal{P} \times 2^{\mathbb{R}}$ such that

(i) $\widetilde{\gamma}$ is elicitable relative to $\mathcal{P}$.

(ii) $(P, \widetilde{\gamma}(P)) \in \mathcal{D}$ for all $P \in \mathcal{P}$

(iii) for all $c \in \widetilde{\gamma}(\mathcal{P})$ the functional $\gamma_c : \mathcal{P}_c \to 2^{\mathbb{R}}$, $P \mapsto \gamma(P, c) \subset \mathbb{R}$ is elicitable relative to $\mathcal{P}_c = \{P \in \mathcal{P} : (P, c) \in \mathcal{D}\}$, and

(iv) $\nu(P) = \gamma(P, \widetilde{\gamma}(P))$ for all $P \in \mathcal{P}$.

Sometimes, c and $\gamma(P, c)$ respectively are single-valued. In this case we identify the one-point sets c and $\gamma(P, c)$ respectively with their unique elements.

Conditional elicitability is a helpful concept for the forecasting of some risk measures which are not elicitable. ES is an example of a risk measure whose conditional elicitability provides the possibility to forecast it in two steps. Indeed, due to the elicitability of $\widetilde{\gamma}(P)$ as fix and forecast $\gamma(P,c)$ due to the elicitability of γ_c. Note that every elicitable functional is conditionally elicitable.

3.3.1.4 Robustness

Another important issue when estimating risk measures is robustness. Without robustness, results may not meaningful, since then small measurement errors in the loss distribution can have a huge impact on the estimate of the risk measure. This is why we investigate robustness in terms of continuity. Since most of the relevant risk measures are not continuous with respect to the weak topology, we need stronger notion of convergence. Therefore, and due to some scaling properties which are convenient in risk management, it is useful to consider the Wasserstein distance when investigating the robustness of risk measures (see Bellini et al. [14]).

The Wasserstein distance between two probability measures P and Q is defined as follows,

$$d_W(P,Q) = inf\mathbb{E}(|X-Y|) : X \sim P, Y \sim Q$$

When we call a risk measure robust with respect to the Wasserstein distance , we mean continuity with respect to the Wasserstein distance in the following sense,

Let P_n, $n \geq 1$, and P be probability measures, and $X_n \sim P_n$, $n \geq 1$ and $P \sim X$. A risk measure ρ is called continuous at X with respect to the Wasserstein distance if

$$\lim_{n\to\infty} d_W(X_n, X) = 0 \Rightarrow \lim_{n\to\infty} |\rho(X_n) - \rho(X)| = 0$$

Cont et al. [25] use a different, potentially more intuitive concept of robustness which takes the estimation procedure into account. They investigate robustness as the sensitivity of the risk measure estimate to the addition of a new data point to the data set which is used as basis for estimation. It turns out that for the same risk measure the estimation method can have a significant impact on the sensitivity. For instance, the risk measure estimate can react in a completely different way on an additional data point if we fit a parametric model instead of using the empirical loss distribution. Thus, robustness in the sense of Cont et al. relates more to sensitivity to outliers in the data sample than to mere measurement errors. Cont et al. also show that there is a conflict between the subadditivity and robustness of a risk measure.

In contrast to robustness based on continuity with respect to weak topology or Wasserstein distance, the concept of Cont et al. allows to distinguish between different degrees of robustness. This concept may make it hard to decide whether or not a risk measure is still reasonably risk sensitive or no longer robust with respect to data outliers in the estimation sample. However, in finance and insurance, large values do occur and are not outliers or measurement errors, but facts that are parts of the observed process itself. In particular, in (re)insurance, one could argue that large claims are actually more accurately monitored than small ones, and their values better estimated. Thus the question of robustness in the sense of Cont et al. may not be so relevant in this context.

3.3.2 Standard Risk Measures

Value-at-Risk (VaR) is a simple risk measure that tells us what loss will be exceeded only a small percentage of time in the next K trading days ($\alpha \cdot 100\%$). Thus, VaR for some significance level α is implicitly defined from the probability of getting an even larger loss as in $Pr(r_{t+k} < VaR^{\alpha}_{t+k}) = \alpha$, where r_{t+k} is the log-return of an asset in period $t+k$. In short, VaR^{α}_{t+k} is defined as the number so that we would get a worse log-return only with probability α. It is easy to obtain an analytical formulation for VaR. Suppose that we are predicting the VaR for some α for 1-day ahead return, $Pr(r_{t+1} < VaR^{\alpha}_{t+1}) = \alpha$, where $r_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1}$. We subtract both terms by μ_t, and divide the resulting by σ_{t+1} returns and we obtain $Pr(z_{t+1} < (VaR^{\alpha}_{t+1} - \mu_{t+1})/\sigma_{t+1}) = \alpha$, where μ_{t+1} is the conditional mean of an asset in period $t+1$, σ^2_{t+1} is the conditional variance of an asset in period $t+1$ and z_{t+1} represents the white noise time series of return innovations which will follow a given probability distribution. This expression can be rewritten as, $F((VaR^{\alpha}_{t+1} - \mu_{t+1})/\sigma_{t+1}) = \alpha$, isolating VaR^{α}_{t+1} in this expression gives us,

$$VaR^{\alpha}_{t+1} = \mu_{t+1} + \sigma_{t+1} F^{-1}(\alpha) \tag{3.1}$$

where F denotes the probability distribution function of the return innovations z_t. In practical financial analysis the α-th quantile of the distribution of log-returns will be negative for low α values. Given the drawbacks of VaR as a risk measure, it is convenient to compute the ES, which accounts for the magnitude of large losses as well as their occurring probability. The ES is defined from VaR as $ES^{\alpha}_{t+k} = \mathbb{E}_{t+k}[r_{t+k}|r_{t+k} < VaR^{\alpha}_{t+k}]$ and tells us the expected value of day k loss, conditional on it being worse than the VaR. As VaR is frequently negative, the expectation below its value is also negative. Extending the deduction of (3.1) to 1-day ahead, ES gives us the following, $ES^{\alpha}_{t+1} = \mathbb{E}_{t+1}[r_{t+1}|r_{t+1} < VaR^{\alpha}_{t+1}]$, from the fact that $r_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1}$ and using the properties of the expectation operator, $ES^{\alpha}_{t+1} = \mu_{t+1} + \sigma_{t+1}\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < (VaR^{\alpha}_{t+1} - \mu_{t+1})/\sigma_{t+1}]$, recalling from (3.1) that $VaR^{\alpha}_{t+1} = \mu_{t+1} + \sigma_{t+1}\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < F^{-1}(\alpha)]$, we get [1],

$$ES^{\alpha}_{t+1} = \mu_{t+1} + \sigma_{t+1}\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < F^{-1}(\alpha)] \tag{3.2}$$

If we assume the existence of an absolutely continuous cdf F, ES is defined as

$$\mathbb{E}_{t+1}[z_{t+1}|z_{t+1} < F^{-1}(\alpha)] = \frac{1}{\alpha}\int_0^{\alpha} F^{-1}(s)ds = \frac{1}{\alpha}\int_{-\infty}^{F^{-1}(\alpha)} rf(r)dr$$

3.3.3 Properties of the Standard Risk Measures

The subadditivite property fails to hold for VaR in general, so VaR is not a coherent measure. Indeed, examples (see e.g. Embrechts et al. [38]) can be given where it is superadditive, i.e.

$$VaR_{\alpha}(\sum_{i=1}^{n} Y_i) < \sum_{i=1}^{n} VaR_{\alpha}(Y_i)$$

Whether or not VaR is subadditive depends on the properties of the joint loss distributions. Three standard cases of subadditivity of VaR,

[1] We assume that VaR_{α} and ES_{α} are < 0.

(i) The random variables are independent and identically distributed (*iid*) as well as positively regularly varying.

(ii) The random variables have an elliptical distribution.

(iii) The random variables have an Archimedean survival dependence structure.

The lack of subadditivity contradicts the notion that there should be a diversification benefit associated with merging portfolios. As a consequence, a decentralization of risk management using VaR is difficult since we cannot be sure that by aggregating VaR numbers for different portfolios or business units we will obtain a bound for the overall risk of the enterprise. Moreover, VaR at level α gives no information about the severity of tail losses which occur with a probability less than $1-\alpha$, in contrast to ES at the same confidence level. When looking at aggregated risks $\sum_{i=1}^{n} Y_i$, it is well known (Acerbi and Tasche [5]) that the ES risk measure is coherent. In particular, in contrast to VaR, it is generally subadditive.

With respect to the weak topology most of the common risk measures are discontinuous. Therefore and due to some convenient scaling properties detailed in Proposition 2.1 of Stahl et al. [89], it is standard in risk management to consider robustness as continuity with respect to the Wasserstein distance. According to them, ES is discontinuous with respect to the weak topology whereas VaR at the level α is robust at F_0 if F_0^{-1} is continuous at α. Stahl et al. observe that ES is continuous with respect to the Wasserstein distance with constant $C = max\{\frac{\alpha}{1-\alpha}; \frac{1-\alpha}{\alpha}\}$, which implies continuity with respect to the Wasserstein distance.

With regard to robustness in the sense given in Cont et al. [25], these authors demonstrate that historical ES is much more sensitive to the addition of a data point than VaR. Moreover, in contrast to VaR, ES is sensitive to the data point's size.

Finally, although ES is a coherent risk measure and, in contrast to VaR, is sensitive to the severity of losses beyond VaR, a potential deficiency arises compared to VaR, when it comes to forecasting and backtesting ES. Gneiting [49] showed that ES is not elicitable. He proved that the existence of convex level sets is a necessary condition for the elicitability of a risk measure and disproved the existence of convex level sets for the ES. This means that it is not a possible to find a scoring function $s(x, y)$ such that ES is defined as the forecast x given a distribution Y that minimizes the scoring function $s(x, y)$.

What Gneiting showed was that this was not possible to do for ES since the scoring function does not exist. Following his findings, many others have interpreted this as evidence that it is not possible to backtest ES at all. This can be seen in for example Carver (2013) [17]. The paper by Gneiting changed the discussion of ES from how it could be backtest to a question of whether it was even possible to do so.

Not all people have interpreted Gneiting's findings as evidence that ES is not backtestable. One of the outstanding issues after his findings was that successful attempts of backtesting ES had been made before 2011. For example Kerkhof and Melenberg (2004)

[65] found methods that performed better than comparable VaR backtests. Following Gneiting's findings, Emmer *et al.* (2013) [39], showed that ES is conditionally elicitable for continuous distributions with finite means. ES consists of two elicitable components. Backtesting can then be done by testing the two components separately. We let Y denote a random variable with a parametric or empirical distribution from which the estimates are drawn. They proposed using the following algorithm,

- Calculate the quantile as

$$VaR_\alpha(Y) = arg \min_{x\in\mathbb{R}} \mathbb{E}[(\alpha - \mathbb{1}_{\{Y<x\}})(Y - x)]$$

- Calculate $ES_\alpha(Y) = \mathbb{E}[Y|Y < VaR_\alpha]$ using the scoring function $\mathbb{E}_P[(Y-x)^2]$, with probabilities $P(A) = P(A|Y < VaR_\alpha)$. This gives

$$ES_\alpha(Y) = arg \min_{x\in\mathbb{R}} \mathbb{E}_P[(Y - x)^2]$$

We know that VaR is elicitable (see Appendix A). If we first confirm this, then what is left is simply a conditional expectation and expectations are always elicitable. In the same paper, Emmer *et al.* (2013) [39] made a careful comparison of different measures and their mathematical properties. They concluded that ES is the most appropriate risk measure even though it is not elicitable. A similar discussion of the implications of different risk measures and its effect on regulation can be found in Chen (2014) [20].

Acerbi and Szekely (2014) [4] argued in a recent article that even without the conditional elicitability, ES is still backtestable. Elicitability is mainly a way to rank the forecasting performance of different models. While VaR is elicitable, this property is not exploited in a normal VaR backtest. In fact, VaR backtests are still based on counting exceptions. If these tests are simple and entail the recording of just one number, it is not because VaR is elicitable, but because quantiles define a Bernoulli random variable. Any other elicitable statistic simply does not. This means that ES cannot be backtested through any scoring function but there is no reason why this could not be done using another method. This means that if we can find a backtest that does not exploit the property of elicitability, there is no reason why backtest would not work.

Elicitability allows to compare in a natural way different models that forecast a statistics in the exact same sequence of events, while recording only point predictions. Acerbi and Szekely (2014) [4] put the following example: "if a bank A has multiple VaR models in place for its $P\&L$, the mean score can be used to select the best in class due to VaR is elicitable. But this is model selection, not model testing. It's a relative ranking not an absolute validation".

Regulators on the contrary need to validate individual models from different banks on an absolute scale. To this purpose, elicitability would still require either the collection of the predictive distributions or strong distributional assumptions, with no guarantee of better power a priori.

It is interesting to note that other important risk measures like the variance are not elicitable either (Lambert et al. [68]). Emmer *et al.* (2015) [39] showed that variance is conditionally elicitable for continuous distributions with finite second moments.

Property	**variance**	**VaR**	**ES**
Coherence			X
Comonotonic additivity		X	X
Robustness, weak topology		X	
Robustness, Wasserstein distance	X	X	X
Elicitability		X	
Conditional Elicitability	X	X	X

Table 3.1: Properties of standard risk measures (Emmer *et al.*, 2015 [39])

3.3.4 Estimating risk: Conditional models for the full distribution

We now define the VaR and ES under conditional models. For that purpose, consider that R is a stationary process with a fully parametric location-scale specification based on the expectation, dispersion and random component; conforming to $r_t = \mu_t + \sigma_t z_t$, where for period t, r_t is the returns of an asset, μ_t is the conditional mean (location), σ_t is the conditional standard deviation (scale) and z_t represents a zero location and unit scale innovations white noise series, which can assume many probability distribution functions F. Under this specification, the risk measures become,

$$VaR_t^\alpha = \mu_t + \sigma_t F^{-1}(\alpha)$$

$$ES_t^\alpha = \mu_t + \sigma_t \left(\frac{1}{\alpha} \int_0^\alpha F^{-1}(s)ds \right)$$

$$SD_t^\alpha = \left[\sigma_t^2 \frac{1}{\alpha} \int_0^\alpha \left(F^{-1}(s) - \left(\frac{1}{\alpha} \int_0^\alpha F^{-1}(s)ds \right) \right)^2 ds \right]^{1/2}$$

The last one is the dispersion around the expected value truncated by the VaR, the SD measure. This will be considered for ES backtesting of Righi & Ceretta.

3.3.5 Estimating risk: Conditional models for extreme events

The other approach is to consider only extreme events, precisely those captured by risk measures. In this regard, there is the Extreme Value Theory (EVT), which is concerned with the distribution of the smallest order statistics and focuses only on the tail of the returns distribution. For further reference on the EVT for modeling the distribution of asset returns, see Longin (2005) [70]. Although the EVT is interesting in risk modeling, the stylized facts make the *iid* assumption inappropriate for most financial data. To solve this issue, one should apply the EVT analysis to the filtered residuals z_t, as proposed by Diebold, Schuermann and Stroughair (2000) [34] and McNeil and Frey (2000) [75]. This is possible because under a correct model specification, the filtered residuals will be approximately *iid*, an assumption for EVT modeling. Under *iid* assumption, consider the distribution function of excess $Y = u - Z$ [2] over a high, fixed threshold u [3], that is $F_u(y) = P(Y = u - Z \leq y | Z < u) = [F(u) - F(u-y)]/[F(u)]$, $y \geq 0$ [4]. For excesses over the threshold, Pickands (1975) [80] elucidates that the generalized Pareto distribution (GPD) arises naturally as the limit distribution of the scaled excesses of identical and independently distributed (*iid*) random variables over high thresholds. We say that excess $Y = u - Z \sim GPD(\xi, \beta)$ if

$$F_u(y) \approx GPD_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\beta}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - exp\left(-\dfrac{y}{\beta}\right), & \xi = 0. \end{cases}$$

$GPD_{\xi,\beta}(y)$ has support $y \geq 0$ if $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ if $\xi < 0$ where $\beta > 0$ is scale parameter and ξ is the tail shape parameter, which is crucial because it governs the tail behavior of $GPD_{\xi,\beta}(y)$. The case $\xi > 0$ corresponds to heavy-tailed distributions whose tails decay like power functions, such as Pareto, Student-t, Cauchy, Burr, loggamma and Fréchet distributions. In this case, the tail index parameter equal to $1/\xi$ corresponds to, for example, the degrees of freedom of the Student-t distribution. The case $\xi = 0$ corresponds to distributions like normal, exponential, gamma and lognormal, whose tails essentially decay exponentially. The final group of distributions are short-tailed distributions ($\xi < 0$) with a finite right endpoint, such as the uniform and beta distributions.

We assume the tail of the underlying distribution begins at the threshold u. From our sample of T data a random number T_u will exceed this threshold. If we assume that the T_u excesses over the threshold are *iid* with exact GPD distribution, Smith (1987) [88] has shown that maximum likelihood estimates $\hat{\xi} = \hat{\xi}_N$ and $\hat{\beta} = \hat{\beta}_N$ of the GPD parameters ξ and β are consistent and asymptotically normal as $T_u \to \infty$, provided $\xi > -1/2$. Under the weaker assumption that the excesses are *iid* from $F_u(y)$ which is only approximately GPD he also obtains asymptotic normality results for ξ and β.

Consider now the following equality for points $z < u$ in the left tail of F

$$F(z) = F(u) - F_u(u-z)F(u) = F(u)(1 - F_u(u-z))$$

[2] Notice that we focus on the lower tail of the data. We have adapted all formulations to consider this issue.

[3] The choice of u has a trade-off: very high u leads to an estimator with large variance, while low u induces bias. The choice of u is the most important implementation issue in EVT.

[4] VaR and ES will be modeled using EVT concept of threshold exceedance. The peaks-over-threshold (POT) models developed around this concept center on the analysis of the generalized Pareto distribution, which may be understood as a limiting tail distribution for a wide variety of commonly studied continuous distributions. The POT is the typical approach used in finance.

If we estimate the first term on the right hand side of the equation using random proportion of the data in the tail T_u/T, and if we estimate the second term by approximating the excess distribution with a generalized Pareto distribution fitted by maximum likelihood, we get the tail estimator

$$\widehat{F}_Z(z) = \frac{T_u}{T}\left(1 + \hat{\xi}\frac{u-z}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

It is very important to note that the distribution F of the conditional model and the distribution $GPD_{\xi,\nu}$ for $\{z\}$ over threshold u are not linked. Thus, it is possible to use distinct conditional models to filter data before applying the EVT to z_T. In this chapter, we assume that F can be different asymmetric distributions to maintain a pattern with the distributions functions we used in the previous chapter. Thus, we have different conditional EVT approaches. The risk measures are,

$$VaR_t^\alpha = \mu_t + \sigma_t F_{z,u}^{-1}(\alpha) = \mu_t + \sigma_t\left(u + \frac{\beta}{\xi}\left[1 - \left(\frac{\alpha}{T_u/T}\right)^{-\xi}\right]\right)$$

$$ES_t^\alpha = \mu_t + \sigma_t\left(\frac{1}{\alpha}\int_0^\alpha F_{z,u}^{-1}(s)ds\right) = \frac{VaR_t^\alpha}{1-\xi} - \left(\frac{\beta+\xi u}{1-\xi}\right)$$

$$SD_t^\alpha = \left[\sigma_t^2\frac{1}{\alpha}\int_0^\alpha\left(F_{z,u}^{-1}(s) - \left(\frac{1}{\alpha}\int_0^\alpha F_{z,u}^{-1}(s)ds\right)\right)^2 ds\right]^{1/2}$$

To sum up, for this purpose, McNeil & Frey proceed as follows: In the first step, they filter the dependence in the returns series by computing the residual of a GARCH-type model, which should be *iid* if the GARCH-type model correctly fits the data. In the second step, they model the extreme behavior of the residual using the tail approach explained previously. Finally, in order to produce a VaR estimate of the original return, they trace back the steps by first producing the α-quantile estimate for the GARCH-type filtered residuals and convert the α-quantile estimate to the original return using the conditional forecast for the required horizon.

It is worth emphasizing that the GARCH-EVT approach incorporates the two ingredients required for an accurate evaluation of the conditional VaR, i.e. a model for the dynamics of the first and second moments, and an appropriate model for the conditional distribution. An obvious improvement of this approach as compared to the unconditional EVT is that incorporates in the VaR changes in expected return and in volatility. For instance, if we assume a change in volatility over the recent period, the GARCH-EVT is able to incorporate this new feature in its VaR evaluation, whereas the unconditional EVT remains stuck at the average level of volatility over the estimation sample.

McNeil & Frey (2000) [75] also provides a backtesting experiment, in which they compare the performances of various methods to correctly reproduce the quantiles of several asset returns. They show that the GARCH-EVT performs much better than unconditional EVT, suggesting that the ability to capture changes in volatility is crucial for VaR computation.

3.3.6 Estimating risk: Filtered Historical Simulation

Filtered Historical Simulation (FHS) method of Barone-Adesi et al. (1998, 1999) [10] [11] applies statistical bootstrap on a parametric, dynamic model for return distributions, such as GARCH-type model. This filtering allows h-day return distributions to be generated from overlapping samples and we can also increase the number of observations used for building the h-day portfolio return distribution through the use of the bootstrap. FHS is in fact a hybrid method combining some attractive features of both historical and Monte Carlo VaR models. We use this method to generate 1-day and 10-day return distributions.

Suppose that at a time s, we want to simulate returns for the next h days. We select $\{z^*_{s+1}, z^*_{s+2}, ..., z^*_{s+h}\}$ at random with replacement from the set of standardized innovations of our model $\{\hat{z}_1, \hat{z}_2, ..., \hat{z}_s\}$ (statistical bootstrap). The data for time s is known. Using the APARCH model to calculate future returns in h days for the dates $t = s+1, s+2, ..., s+h$ is as below:

$$\sigma^*_t = (\hat{\omega} + \hat{\alpha}_1(|\varepsilon^*_{t-1}| - \hat{\gamma}_1\varepsilon^*_{t-1})^{\delta} + \hat{\beta}_1(\sigma^*_{t-1})^{\hat{\delta}})^{1/\hat{\delta}} \tag{3.3}$$

$$\varepsilon^*_t = z^*_t\sigma^*_t \tag{3.4}$$

$$r^*_t = \hat{\phi}_0 + \hat{\phi}_1 r^*_{t-1} + \varepsilon^*_t \tag{3.5}$$

The algorithm contains the following steps,

(i) Select a set $\{z^*_{s+1}, z^*_{s+2}, ..., z^*_{s+h}\}$ which has h elements and is chosen randomly with replacement from $\{\hat{z}_1, \hat{z}_2, ..., \hat{z}_s\}$.

(ii) Assumed as initial values the last estimates: $\sigma^*_s = \hat{\sigma}_s$ and $\varepsilon^*_s = \hat{\varepsilon}_s$.

(iii) Set up for $t = s+1, s+2, ..., s+h$,

- Plug σ^*_{t-1} and ε^*_{t-1} in equation (3.3) and get σ^*_t.
- Plug z^*_t (obtained in step 1) and σ^*_t (obtained previously) in equation (3.4) and get ε^*_t.
- Plug r_{t-1} (last return of the historical sample, i.e. r_s) and ε^*_t (obtained previously) in equation (3.5) and get r^*_t.
- Then the simulated log return over a risk horizon of h days ($r^*_{s:h}$) is the sum $r^*_{s+1} + r^*_{s+2} + ... + r^*_{s+h}$

(iv) Repeating this procedure N times, in our case 5000, produces 5000 simulated h-day returns.

We will have in $s = 1$, where we have used 2915 observations previously for estimating the model and generate from in-sample standardized innovations (filtering out APARCH and AR models) 5000 paths with length h periods using bootstrap and recover the h-day density by cumulating returns, h-day VaR, h-day ES and h-day SD calculated as

$$VaR^{\alpha}_{t+h} = Percentile\left\{ r^*_{i,s=1,s:h}{}^{N}_{i=1}, 100\alpha \right\} \quad i = 1, 2, ..., N$$

where $r^*_{i,s=1,s:h}{}_{i=1}^{N}$ is the 5000 by 1 vector than contains the 10-day log-returns of the asset;

$$ES^{\alpha}_{t+h} = (N\alpha)^{-1}\sum_{i=1}^{N}(r^*_{i,s=1,s:h}\mathbb{1}_{\{r^*_{s=1,s:h}<VaR^{\alpha}_{t+h}\}})$$

where $\mathbb{1}$ is the indicator function that assumes value 1 if the h-day returns are lower than VaR and 0 otherwise. Thus, the ES is just the mean of the values below VaR;

$$SD^{\alpha}_{t+h} = \left\{(N\alpha)^{-1}\sum_{i=1}^{N}[(r^*_{i,s=1,s:h}\mathbb{1}_{\{r^*_{i,s=1,s:h}<VaR^{\alpha}_{t+h}\}}) - ES^{\alpha}_{t+h}]\right\}^{1/2}$$

and, thus SD is just the standard deviation around the ES, considering only the values below VaR.

We repeat all steps using recursive forecasts (expanding window) to estimate the model, and h-day ahead out-of-sample forecasts are produced, where the sample is increasing by one until $s = 1260$, the model are re-estimated each day, and h-day ahead forecasts are produced. Therefore we are going to 1260 (the last five years of the sample) forecast of h-day ahead VaR, ES and SD risk measures.

Under EVT approach, following the McNeil & Frey (2000) [75] approach, we fit the left tail of the standardized innovations with the Generalized Pareto distribution (GPD) and we estimate the parameters ξ and β. We generate patterns (5000 simulations) with length h periods of returns from the fitted models and sampling innovations from in-sample residuals (FHS) using a combination of bootstrap and GPD simulation according to the following algorithm which was also proposed independently by Danielsson and de Vries (2000) [31],

(i) Randomly select standardized innovations from the samples generated ($N = 5000$) in $s = 1$.

(ii) If standardized innovations are less than threshold (u), sample a GPD($\hat{\xi}, \hat{\beta}$) distributed excess z^* from the left tail and return $u - z^*$.

(iii) Otherwise return standardized innovations themselves.

(iv) Finally, the procedure is the same as previously described, i.e. we trace back from simulated standardized innovations to recover the returns and we end up with N sequences of hypothetical daily returns for day $s + 1$ through day $s + h$. From these hypothetical daily returns, we calculate the hypothetical h-day returns as $r^*_{s:h} = \sum_{h=1}^{H} r_{i,s+h}$ for $i = 1, 2, ..., N$. If we collect the N hypothetical h-day returns in a set $r^*_{i,s=1,s:h}{}_{i=1}^{N}$, then we calculate the h-day VaR, 10-day ES and h-day SD, as we have defined previously.

(v) We repeat this procedure for $s + 1, s + 2, s + 3, ..., s + 1259$ (out-of-sample period).

The advantages of FHS approach are 1) we capture current market conditions by means of the volatility dynamics, 2) no assumptions need to be made on the distribution of the return shocks and 3) the method allows for the compution of any risk measure for any investment horizon of interest because we can generate as many h-day returns as we like.

3.4 Data and Estimation Models

We work with daily percentage returns on assets over the sample period 10/02/2000 - 09/30/2016 (number of observations 4175). Daily returns are computed as 100 times the difference of the log prices, i.e. $100[\ln(P_{t+1}) - \ln(P_t)]\%$. The financial assets considered are: International Business Machines [IBM] ($), Banco Santander [SAN] (€), AXA [AXA] (€) and BP [BP] (£). The data were extracted from Datastream.

Table 3.2 reports descriptive statistics for the daily percentage returns series. All of them have a mean close to zero. Median returns are zero. SAN is the one with a wider total range ($max - min$) and BP has a narrower range. The unconditional standard deviation (S.D.) is around 2, being the highest one for AXA and the lowest one for IBM. According to the skewness statistic, all assets have negative skewness, except AXA. For all assets considered, the kurtosis statistic is large, implying that the distributions of those returns have much thicker tails than Normal distribution. Similarly, the Jarque-Bera statistic (J-B) is statistically significant, rejecting the assumption of normality in all cases.

	Mean (%)	Median (%)	Max	Min	S.D.	Skewness	Kurtosis	J-B
IBM	0.83	0	11.35	-16.89	1.58	-0.22	12.33	15194.87
SAN	1.56	0	20.88	-22.17	2.26	-0.07	10.50	9793.17
AXA	1.47	0	19.78	-20.35	2.69	0.19	10.24	9155.81
BP	-0.69	0	10.58	-14.04	1.69	-0.19	8.01	4390.88

Table 3.2: Descriptive statistics for the daily percentage returns.

In order to perform a ES analysis we estimate volatility model APARCH (Ding, Granger and Engle, 1993 [35]) under the different probability distributions assumed for the innovations: Gaussian, Student-t, skewed Student-t, skewed generalized error and Johnson S_U [5]. An AR(1) model was considered for the conditional mean return, which is sufficient to produce serially uncorrelated innovations [6].

The APARCH model is particularly successful in capturing the heteroscedasticity exhibited by the data due to the power of the conditional standard deviation is a free parameter, which provides more flexibility to the dynamics of volatility.

For a given return series $r_1, ..., r_T$, the model adopted is

$$r_t = \phi_0 + \phi_1 r_{t-1} + \varepsilon_t \qquad \varepsilon_t = \sigma_t z_t \qquad t = 1, 2, ...T$$

$$\sigma_t^\delta = \omega + \alpha_1(|\varepsilon_{t-i}| - \gamma_1 \varepsilon_{t-i})^\delta + \beta_1(\sigma_{t-j})^\delta$$

[5] We provided a description of probability distributions in Appendices A.5 - A.9 of Chapter 2.

[6] All computations were performed with the *rugarch* package (version 1.3-4) of R software (version 3.1.1) designed for the estimation and forecast of various univariate ARCH-type models. In the estimation of EVT models, we use *ismev* (version 1.41) and *evir* (version 1.7-3) packages.

where ω, α_i, γ_i, β_j and δ are additional parameters to be estimated. The parameter γ_i reflects the leverage effect $(-1 < \gamma_i < 1)$. A positive (resp. negative) value of γ_i means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive (resp. negative) shocks. The parameter δ plays the role of a Box-Cox transformation of $\sigma_t(\delta > 0)$.

For the EVT, we use 10% of the data as the threshold excess. For the conditional models, a filter is necessary to model the conditional mean and the variance of data. Thus, we estimate the AR(1)-APARCH(1,1) model, conforming to formulations. The z represents a F distributed white noise series. As explained previously, we set F to be Gaussian, Student-t, skewed Student-t, skewed generalized error and Johnson S_U distributions. In all models we jointly estimate by maximum likelihood the parameters in the equation for the mean return, the equation for its conditional standard deviation and the probability distribution for the return innovations. In addition, through the usual diagnostics performed in linear and quadratic standardized residuals, we assess that the information is properly filtered. Based on this filtering, the conditional models are estimated as described in subsection 3.3.5

Table 3.3 presents the results of the estimation by the maximum likelihood method of the generalized Pareto distribution parameters jointly with the respective parameters of the distribution of innovation and of the model AR(1)-APARCH(1,1), for a given threshold u, for each asset. For all asset returns, the estimated tail index ξ of generalized Pareto distribution is positive. Left tails of these return distributions are fat, i.e. the probability of occurrence of extreme loss is higher than what the Normal distribution predicts (tails decrease polynomially). We observe that the tail indexes of IBM and SAN are higher than ones of AXA and BP, i.e. the left tail of those returns is higher than of these ones.

As an example, we observe, in Table 3.4 the estimated parameters of AR(1)-APARCH(1,1) model with JSU distribution for IBM [7]. We observe that the autoregressive effect in the volatility specification is strong, β_1 is around 0.93, suggesting strong memory effects. The coefficient γ_1 was found to be positive and statistically significant, indicating the existence of a leverage effect for negative returns in the conditional volatility specification. It is also important that skewness parameter in the Johnson S_U is less than 0, suggesting the convenience of incorporating negative asymmetric to model innovations appropriately, although this parameter is not significant at 5%, and the shape parameter is low, implying high kurtosis. Finally, δ takes value 1.07, being significantly different from 2. This result suggests that, instead of modeling the conditional variance, it might be better to model the conditional standard deviation.

The maximum likelihood estimates of the generalized Pareto distribution parameters are $(\hat{\xi}, \hat{\beta}) = (0.39, 0.51)$, with standard errors of 0.12 and 0.07 respectively. In Figure 3.1, we observe that for $\hat{\xi} = 0.39$ the maximum log-likelihood is reached (-91.877). Thus, the model we have fitted is essentially a very heavy-tailed, infinite-variance model.

[7] To save space, we only report estimation results of the EVT-JSU-AR(1)-APARCH(1,1) model for IBM. Results for alternative models and for another assets are available from the authors upon request.

Daily		u	ξ	β	(ξ)	(β)
N	**IBM**	-1.041	0.392	0.493	0.121	0.072
	SAN	-1.239	0.240	0.522	0.103	0.070
	AXA	-1.139	0.048	0.712	0.067	0.079
	BP	-1.131	0.055	0.642	0.086	0.079
ST	**IBM**	-1.061	0.391	0.514	0.120	0.075
	SAN	-1.249	0.235	0.534	0.101	0.071
	AXA	-1.175	0.059	0.697	0.070	0.079
	BP	-1.158	0.072	0.635	0.089	0.080
SKST	**IBM**	-1.051	0.390	0.514	0.120	0.075
	SAN	-1.235	0.229	0.539	0.100	0.071
	AXA	-1.159	0.057	0.702	0.069	0.079
	BP	-1.154	0.078	0.627	0.090	0.079
SGED	**IBM**	-1.037	0.376	0.524	0.118	0.076
	SAN	-1.233	0.225	0.542	0.100	0.072
	AXA	-1.152	0.055	0.705	0.069	0.079
	BP	-1.145	0.072	0.628	0.089	0.079
JSU	**IBM**	-1.053	0.392	0.516	0.121	0.075
	SAN	-1.236	0.230	0.539	0.100	0.071
	AXA	-1.157	0.057	0.703	0.069	0.079
	BP	-1.152	0.074	0.631	0.089	0.079

Table 3.3: Estimated parameters of GPD with daily returns. u is the threshold, ξ is the shape parameter, β is the scale parameter and (ξ) and (β) respectively correspond to the standard error of shape parameter and scale parameter.

	Estimate	**Std. Error**	**t value**	$Pr(>\|t\|)$
ϕ_0	-0.00049	0.01760	-0.02818	0.97752
ϕ_1	-0.02048	0.01442	-1.41984	0.15565
ω	0.02108	0.02664	0.79118	0.42884
α_1	0.07614	0.05961	1.27726	0.20151
β_1	0.92882	0.06486	14.31942	0
γ_1	0.50884	0.29529	1.72320	0.08485
δ	1.07512	0.27157	3.95886	0.00007
skew	-0.09240	0.04808	-1.92187	0.05462
shape	1.52764	0.06332	24.12484	0
ξ	0.39168	0.12073	3.2442	0.00120
β	0.51558	0.07523	6.8532	0

Table 3.4: Parameter estimates of AR(1)-APARCH(1,1) model with JSU distribution for IBM and with GPD for IBM residuals of the previous model.

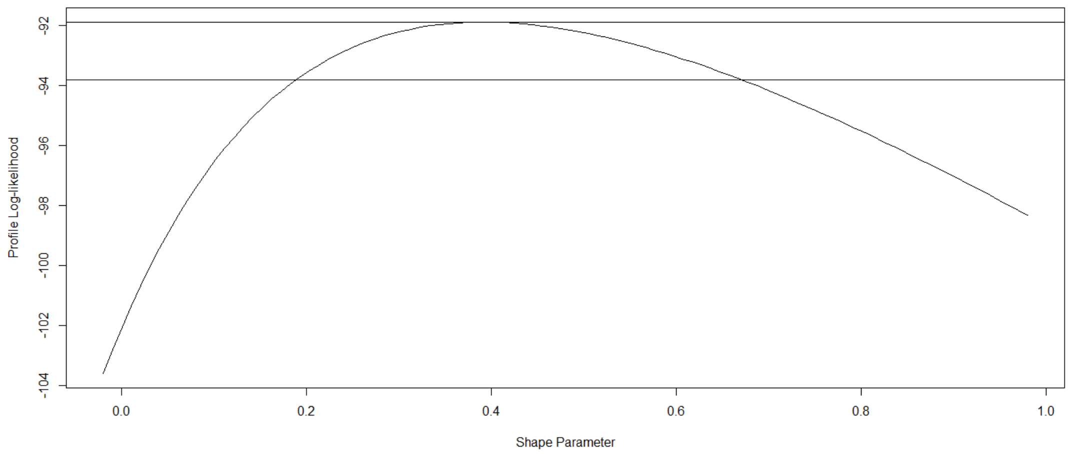


Figure 3.1: Profile likelihood for ξ in threshold excess model of percentage daily loss filtered residuals of IBM with JSU-EVT model.

We consider tail from threshold u=1.0533 [8]. In this case, we have 126 exceedances (10% of 1260 data). Figure 3.2 shows the fitted GPD model for the excess distribution, $F_u(y)$ where $y = z - u$, superimposed on points plotted at empirical estimates of excess probabilities for each loss (126 losses); note the good correspondence between the empirical estimates and the GPD curve [9] .

Figure 3.3 shows the estimation tail probabilities on logarithmic axes. The points on the graph are the 126 threshold exceedances and are plotted at y-values corresponding to the tail of the empirical distribution function; the smooth curve running through the points is the tail estimator (defined in right tail)

$$1 - \widehat{F}(z) = \frac{T_u}{T}\left(1 + \hat{\xi}\frac{z-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

3.5 Evaluating 1-day ES

3.5.1 A Review of Backtesting Approaches

Despite the ES advantages, it is still less used than VaR. The principal reason is that backtesting ES is much harder than backtesting VaR. Recently, some ES backtesting procedures have been developed, like the residual approach introduced by McNeil and Frey (2000) [75], the censored Gaussian approach proposed by Berkowitz (2001) [15], the functional delta approach of Kerkhof and Melenberg (2004) [65], and the saddlepoint technique introduced by Wong (2008) [95].

[8] In the following Figures 3.2 and 3.3, we plot the right tail. Losses are positives.

[9] All considered models whose filtered residuals have been fitted by GPD (EVT approach) show a very similar fit of the GPD curve, specially when the filtered residuals come from asymmetric distributions.

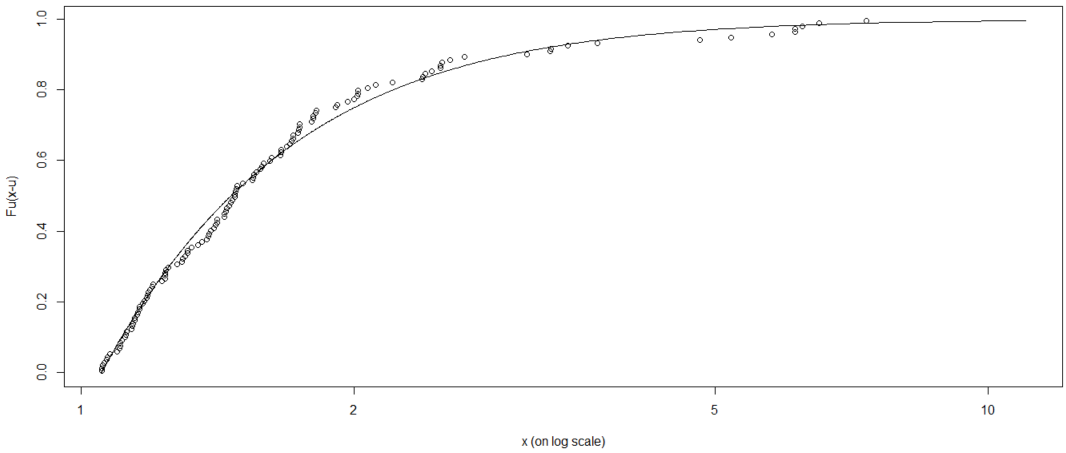


Figure 3.2: Empirical distribution of excess of IBM filtered residuals of AR(1)-APARCH(1,1)-JSU and its fitted GPD.

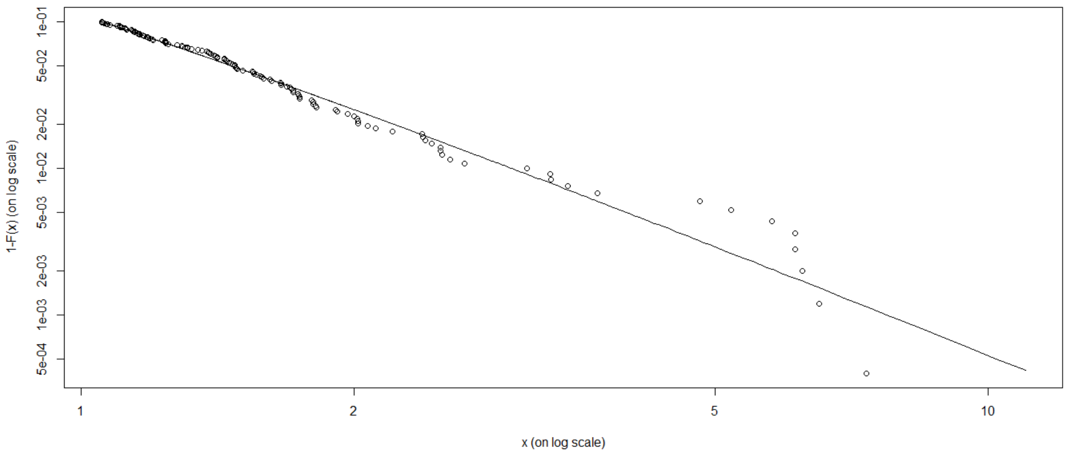


Figure 3.3: The smooth curve through the points shows the estimated tail of the IBM percentage loss filtered residuals of AR(1)-APARCH(1,1)-JSU using tail estimator. Points are plotted at empirical tail probabilities calculated from empirical distribution function.

However, these approaches present some drawbacks. The backtest of McNeil and Frey (2000) [75], Berkowitz (2001) [15] and Kerkhof and Melenberg (2004) [65] rely on asymptotic test statistics that might be inaccurate when the sample size is small, and this could penalize financial institutions because of an incorrect estimation of ES. Further, these tests compute the required p-value based on the full sample size rather than conditional on the number of exceptions. The test proposed by Wong (2008) is robust to these questions, making it possible to detect failure of a risk model based on just one or two exceptions before any more data is observed. Nonetheless, the Wong (2008) backtest [95] has some disadvantages, such as the Gaussian distribution assumption, and the use of the full distribution conditional standard deviation as a dispersion measure. All the mentioned procedures are limited to backtesting the ES estimation considering the whole sample period.

Backtesting ES	**Definition**	**Statistic**	**Distribution**		**Parametric**
McNeil & Frey (2000)	Residual Approach	t	Unknown	Simulations	**X**
Berkowitz (2001)	Censored Gaussian Approach	LR_{tail}	χ^2_2		**X**
Kerkhof & Melenberg (2004)	Functional Delta Approach	S_T	$N(0,1)$		**X**
Wong (2008)	VaR tail losses (Normal)	TR	Unknown	Lugananni & Rice	**X**
Emmer, Kratz & Tasche (2013)	Quantile Approximation	LR_{uc}	χ^2_1		
Righi & Ceretta (2013)	Truncated tail distribution	BT_T	Unknown	Simulations	**X**
Acerbi & Szekely (2014)	Testing ES after VaR Testing ES directly Estimating ES from realized ranks	Z_1 Z_2 Z_3	Unknown Unknown Unknown	Simulations Simulations Simulations	
Graham & Pál (2014)	VaR tail losses (Exponential)	TR	Unknown	Lugananni & Rice	**X**
Costanzino & Curran (2015)	Unconditional ES Spectral Risk Measure Test	U_{ES}	$N(0,1)$		
Du & Escanciano (2015)	Conditional ES Spectral Risk Measure Test	C_{ES}	χ^2_m		

Table 3.5: Overview of ES backtesting procedures in literature.

In this chapter we use five approaches for evaluation of ES estimates, with six methods overall. The test of Righi & Ceretta [83] and the first two tests of Acerbi & Szekely are straightforward, but require simulations, the test of Graham & Pál [57], which is an extension of the Lugannani-Rice approach of Wong [95], the quantile-space unconditional coverage test of Costanzino & Curran [26] for the family of Spectral Risk Measures, of which ES is a member, and, finally, the conditional test of Du & Escanciano [36]. The last two tests can be thought of as the continuous limit of the Emmer, Kratz & Tasche [39] idea in that it is a joint test of a continuum of VaR levels.

3.5.1.1 The Righi & Ceretta Approach

(The Righi & Ceretta approach has been extracted from Righi & Ceretta (2013) [83].)

The ES backtest approach of Righi & Ceretta (2013) [83] extends and improves those previously introduced in the literature in three main ways. First, they use the dispersion of the truncated distribution by the estimated VaR upper limit, instead of the whole probability function. Second, they do not limit the approach to the Gaussian case. They permit other probability distribution functions and even an empirical distribution, making this approach more flexible. Finally, their approach allows to separately test if each VaR violation is significantly different from the ES, allowing a faster model of error verification, which is extremely useful since prompt action is often required in order to avert extreme financial losses due to market risk.

The Expectation term in (3.2) is the expectation of a truncated distribution of z_{t+1} with upper limit $F^{-1}(\alpha)$. Based on this idea, they propose that the dispersion around this expected value, truncated by the VaR should be considered for ES backtesting. They refer to this dispersion as Shortfall Deviation (SD). The SD is the square root of the truncated variance for some quantile conditional to the probability α, i.e. $SD^{\alpha}_{t+k} = (VAR[r_{t+k}|r_{t+k} < VaR^{\alpha}_{t+k}])^{1/2}$. With a similar deduction of the ES, we obtain the 1-day ahead SD as, $SD^{\alpha}_{t+k} = (VAR_{t+k}[r_{t+k}|r_{t+k} < VaR^{\alpha}_{t+k}])^{1/2}$, and since $r_{t+1} = \mu_{t+1} + \sigma_{t+1}z_{t+1}$ then by standardization we get, $SD^{\alpha}_{t+1} = (\sigma^2_{t+1}VAR_{t+1}[z_{t+1}|z_{t+1} < (VaR^{\alpha}_{t+1} - \mu_{t+1})/\sigma_{t+1}])^{1/2}$. Again, substituting the expression (3.1) in this formulation gives us,

$$SD^{\alpha}_{t+1} = (\sigma^2_{t+1}VAR_{t+1}[z_{t+1}|z_{t+1} < F^{-1}(\alpha)])^{1/2} \tag{3.6}$$

The SD is a better estimate than the whole sample standard deviation because when extreme negative returns occur, it is the risk in the left tail that risk managers and financial institutions are concerned about. Further, to precise how much a loss was far from its expected value, one needs to use some dispersion measure intrinsic to this expectation rather than linked with the absolute distribution expectation.

For each day in the forecast period which a violation in the predicted VaR occurs, it is possible to estimate the expected loss, as well as its dispersion through the ES and SD, respectively. In that sense, Righi & Ceretta propose to backtest if the day k violation is significantly worse from that expected for certain α VaR quantile. To that end, they use a backtest ratio as equation (3.7),

$$BT_{t+k} = \frac{r_{t+k} - ES^{\alpha}_{t+k}}{SD^{\alpha}_{t+k}} \tag{3.7}$$

In (3.7) we compute by how many units of the dispersion measure the occurred loss is far from its expected value. This test has the null hypothesis $BT_{t+k} = 0$ against the alternative that $BT_{t+k} < 0$. We use the test of Righi & Ceretta (2015) [82] and we focus on a single test for all of the out-sample observations, i.e. $H_0 : \mathbb{E}(BT_t) = 0$ against $H_1 : \mathbb{E}(BT_t) < 0$. This is in contrast to Righi & Ceretta (2013) [83] who test for each day of the out-of-sample period.

As $r_{t+k} = \mu_{t+k} + \sigma_{t+k}z_{t+k}$ then, substituting (3.2) and (3.6) in the expression (3.7) we get,

$$BT_{t+k} = \frac{\mu_{t+k} + \sigma_{t+k}z_{t+k} - (\mu_{t+k} + \sigma_{t+k}E_{t+k}[z_{t+k}|z_{t+k} < F^{-1}(\alpha)])}{(\sigma^2_{t+k}VAR_{t+k}[z_{t+k}|z_{t+k} < F^{-1}(\alpha)])^{1/2}}$$

and simplifying the expression we obtain a simpler form for our backtest, as represented by (3.8),

$$BT_{t+k} = \frac{z_{t+k} - E_{t+k}[z_{t+k}|z_{t+k} < F^{-1}(\alpha)]}{(VAR_{t+k}[z_{t+k}|z_{t+k} < F^{-1}(\alpha)])^{1/2}} \tag{3.8}$$

As we know from the GARCH-type model that $z_{t+k} \sim i.i.d.\ F$ by definition, we conclude that expression (3.8) has more tractable properties than (3.7) because instead of z_{t+k}, we do not know the real probability distribution function of r_{t+k}. The proposed backtest is a one-tailed test with the alternative hypothesis that the occurred loss is worse than the expected one. To robustly obtain the statistical probability linked with the calculated value of BT_{t+k}, i.e. without the need to rely in any distribution assumption about the ratio, Righi & Ceretta (2013) [83] use Monte Carlo simulations as the following algorithm,

1. Generate N times n random variables $u_{ij} \sim iid\ F$, $i = 1, ..., n$; $j = 1, ..., N$.
2. Calculate for each sample N, $E[u_{ij}|u_{ij} < F^{-1}(\alpha)]$ and $VAR[u_{ij}|u_{ij} < F^{-1}(\alpha)]$, where $F^{-1}(\alpha)$ denotes the α-quantile of u_{ij}.
3. For every $u_{ij} < F^{-1}(\alpha)$, calculate $BT_{ij} = \frac{u_{ij} - E[u_{ij}|u_{ij} < F^{-1}(\alpha)]}{(VAR[u_{ij}|u_{ij} < F^{-1}(\alpha)])^{1/2}}$.
4. Given a significance level α', determine the critical value as the median of all BT_{ij} series α'-th quantile.
5. Given the actual BT_{t+k}, determine the test p-value as the median of $Pr(BT_{ij} < BT_{t+k})$.

3.5.1.2 The Acerbi & Szekely Approaches

(The Acerbi & Szekely approaches have been extracted from Acerbi & Szekely (2014) [4].)

Acerbi & Szekely (2014) [4] define three test statistics, each with slightly different assumptions. Each test involves somewhat different null and alternative hypotheses. We use two of the three tests proposed by them. In all cases, the general test procedure for the test statistics Z_i, $i \in 1, 2$ is the same. These tests are non-parametric and free from distributional assumptions other than continuity, necessary conditions for any application in banking regulation.

Starting with realized $Z_i(\overrightarrow{R})$ score from T returns, where $\overrightarrow{R} \in \mathbb{R}^T$ is the vector $\overrightarrow{R} = \{r_t\}$ of all T observations and $i \in 1, 2$,

1. Simulate a set of *iid* trials following the distributions $r_t^j \sim F_t$, $\forall t = 1, ..., T$, $j = 1, ..., M$ for a suitably large M [10].
2. Compute the test statistics of interest from the simulation r_t^j, $Z_i^j = Z_i(r_t^j)$, $i \in 1, 2$.

[10] In Acerbi & Szekely and Righi & Ceretta tests, we simulate 10000 process of length 1000.

3. Estimate the quantile, or p-value, of the observed score $p = \frac{1}{M}\sum_{k=1}^{M} \mathbb{1}_{\{Z_i^k < Z(\overrightarrow{R})\}}$

The two test statistics are as follows [11]

Statistic Z_1: Testing ES after VaR.

$$Z_1(\overrightarrow{R}) = \frac{1}{N_T}\sum_{t=1}^{T}\frac{I_t r_t}{ES_t^\alpha} - 1$$

if $N_T = \sum_{t=1}^{T} I_t > 0$ where $I_t = \mathbb{1}_{\{r_t < VaR_t^\alpha\}}$ is the indicator of VaR breaches. The null hypothesis is $H_0 : P_t^\alpha = F_t^\alpha \quad \forall t$ where F_t^α is the tail of the forecast cumulative distributions for each day when $r_t < VaR_t^{\alpha,F}$ and P_t^α represents the tail of (unknown) real distributions from which the realized events r_t are drawn. The alternative hypothesis is

$$\begin{aligned} H_1 : \quad & ES_t^{\alpha,P} \le ES_t^{\alpha,F} && \forall t \text{ and } < \text{ for some } t \\ & VaR_t^{\alpha,P} = VaR_t^{\alpha,F} && \forall t \end{aligned}$$

We see that the predicted VaR_α is still correct under H_1, in line with the idea that this test is subordinated to a preliminary VaR test. This test is in fact completely insensitive to an excessive number of exceptions as it is an average taken over exceptions themselves.

Under these conditions $\mathbb{E}_{H_0}[Z_1|N_T > 0] = 0$ and $\mathbb{E}_{H_1}[Z_1|N_T > 0] > 0$. So, the realized value $Z_1(\overrightarrow{R})$ is expected to be zero, and it signals a problem when it is positive.

Statistic Z_2: Testing ES directly.

$$Z_2(\overrightarrow{R}) = \frac{1}{T\alpha}\sum_{t=1}^{T}\frac{I_t r_t}{ES_t^\alpha} - 1$$

if $N_T > 0$. H_0 is as in the previous test and the alternative hypothesis is

$$\begin{aligned} H_1 : \quad & ES_t^{\alpha,P} \le ES_t^{\alpha,F} && \forall t \text{ and } < \text{ for some } t \\ & VaR_t^{\alpha,P} \le VaR_t^{\alpha,F} && \forall t \end{aligned}$$

We note that $\mathbb{E}_{H_0}[N_T] = T\alpha$. We have again $\mathbb{E}_{H_0}[Z_2] = 0$ and $\mathbb{E}_{H_1}[Z_2] > 0$.

Unlike the Z_1 statistic, the sum of the VaR breach event returns is divided by the expected value. The Z_2 statistic will tend to reject a large number of VaR breach events of small magnitude. This leads to the difference in H_1 between the two statistics; rejecting the H_0 of $Z_2(\overrightarrow{R})$ includes rejecting $VaR_t^{\alpha,F}$ as correctly specified.

[11] Here, the test statistics are defined to work with negative values of VaR_α and ES_α, but in Acerbi & Szekely (2014), these tests were defined with positive ES values.

3.5.1.3 The Graham & Pál Approach

(The Graham & Pál approach has been extracted from Graham & Pál (2014) [57] and Wong (2010) [94].)

Graham & Pál (2014) [57] generalize Wong's approach (2010) [94]. Wong's backtest is well-defined and used advanced statistical techniques to deal with the small number of VaR violations usually found within a year's worth of VaR forecasts. Unfortunately, though Wong acknowledges the possibility of assuming otherwise, his backtest is not model- or distribution-agnostic, as it assumes normal distributions. Graham & Pál generalize Wong's result in a tractable and intuitive manner to allow for any VaR modeling, and therefore distributional, approach.

Their goal is to quantify how extreme each VaR violation is in relation to its forecast distribution. The approach can be intuitively described as an extension of the "hit" time series concept, wherein each of the "1" values (when VaR violation occurs) are modified so as to measure the distance between each violation and its corresponding VaR threshold. In other words, the nonzero time series values in our context are negative values consisting of the difference between each percentile smaller than the VaR threshold percentile and the VaR threshold percentile itself. If there is no VaR violation on a specific observation date, then a value of zero will still be recorded.

Now that a time series of transformed tail losses has been formed, it remains to determine how to quantify its behavior. We would expect this series to be uniformly distributed within the tail region if the series of forecast distributions accurately modeled the portfolio's $P\&L$.

The remainder of the backtest consists of a number of steps. The first is to transform the percentile time series into a suitable context under which we can easily compare its average value to the one we would expect; that is, to the one that would occur if the VaR forecast distributions accurately (or at least, conservatively) modeled their corresponding realized $P\&L$ values. The second is to then derive an appropriate methodology to compare the average of the sample time series of tail-loss values with the expectation of the mean tail loss. By tail-loss we understand the difference of the losses beyond VaR minus the VaR value itself. Finally, as there is a small expected number of VaR violations, a small-sample asymptotic technique will be used to determine to what extent we can be certain that the sample average VaR violation is not too extreme. This is, in fact, the basis of the hypothesis test that forms the core of the backtest.

Tail Risk concept

The central risk concept that they employ within the backtest is that of tail risk, as defined by Wong (2010) [94]. Tail Risk (TR), is related to VaR and ES in the following way,

$$TR_\alpha = \int_{-\infty}^{q(\alpha)} (r - q(\alpha))f(r)dr = \alpha(ES_\alpha - VaR_\alpha)$$

where $VaR_\alpha = q(\alpha) = F^{-1}(\alpha)$.

The tail risk will always be a negative quantity. We can consider $\alpha^{-1}TR$ as the difference between the ES and the VaR.

Given a sample of N returns $r_1, r_2, ..., r_N$ of R, the sample estimator for tail risk at confidence level $(1 - \alpha)$ may be calculated as

$$\widehat{TR}_\alpha = \frac{1}{N}\sum_{i=1}^{N}(r_i - q(\alpha))\mathbb{1}_{\{r_i<q(\alpha)\}} \tag{3.9}$$

Proceeding from (3.9), we define the random variable X as

$$X = (R - q(\alpha))\mathbb{1}_{\{R<q(\alpha)\}}$$

and observe that the sample estimator for the mean of X is equal to the sample estimator for the tail risk,

$$\widehat{TR}_\alpha = \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

We note that the range of X is range$(X) = (-\infty, 0]$. It is also important to note that the expected value of X is the theoretical tail risk of the portfolio, which is

$$\mathbb{E}[X] = \mathbb{E}[\widehat{TR}_\alpha] = TR_\alpha = \alpha(ES_\alpha - VaR_\alpha) \tag{3.10}$$

Wong (2010) [94] incorporates TR as his test statistic by making the assumption of normally distributed portfolios (though he notes that the small-sample asymptotic technique may be derived under nonnormal conditions). However, though perhaps still widely modeled as such, return distributions may and do certainly deviate quite greatly from assumptions of normality. Graham & Pál generalize Wong's result by demonstrating how the TR test statistic may be implemented even for dynamically updating asset/portfolio, with forecast distribution the change from day to day.

Transformation of the random variable X

To understand how to proceed, we simply transform the realized losses and forecast distributions through the probability integral transform (PIT) to ensure that our sample estimator for the average tail risk is created through identically distributed sample values.[12] The exact transformation is simple; it is necessary only to locate each VaR violation

[12] Rosenblatt (1952) [85], Crnkovic and Drachman (1996) [27], Diebold *et al.* (1998) [33] and Berkowitz (2001) [15] are often credited with introducing PIT into the financial risk management backtesting literature.

as a percentile value within its forecast distribution,

$$p_t = \int_{-\infty}^{y_t} f_t(u)du = F_t(y_t) \tag{3.11}$$

Using (3.11), we may now redefine the random variable X as follows

$$X_t = (F_t(y_t) - F_t(q_t(\alpha)))\mathbb{1}_{\{y_t<q_t(\alpha)\}} = (F_t(y_t) - \alpha)\mathbb{1}_{\{F_t(y_t)<\alpha\}} = (p_t - \alpha)\mathbb{1}_{\{p_t<\alpha\}} \tag{3.12}$$

It is well-known that if the forecast CDFs $F_t(\cdot)$ are correct estimates of the real (though unobservable, as they are dynamically changing) $P\&L$ distributions, then the series p_t is distributed uniformly $\mathbb{U}(0,1)$. Moreover, if the sequence of forecast CDFs is correctly conditionally calibrated, then the corresponding p_t sequence is independent and identically distributed (*iid*) $\mathbb{U}(0,1)$ [13]. Clearly, once X has been redefined as per (3.12) and under the assumption of correctly conditionally calibrated forecast CDFs, $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ is then an average of *iid* random variables.

Though it is possible to state the hypothesis test for the revised random variable X as defined (3.12), they choose to make a further transformation to the exponential context. The reason is that the hypothetical distribution of X_t as defined in (3.9) has a noncontinuous CDF with large mass at zero, and the form of the moment- and cumulant generating functions necessitates the use of numerical techniques to obtain the saddle point. In contrast, making an additional transformation to the exponential context allows us to solve for the saddle point analytically; this solution is, moreover, well-defined over the complete interval of interest for tail losses.

The more common approach in such cases might be to perform a transformation to the normal context. This appears, for example, in Berkowitz (2001) [15], and Wong (2010) [94] intentionally uses normal distributions as his starting point. However, a normal transformation such as $X_t^{\mathbb{N}} = (\Phi^{-1}(p_t) - \Phi^{-1}(\alpha))\mathbb{1}_{\{p_t<\alpha\}}$ where $\Phi(\cdot)$ is the standard normal CDF, would also require a numerical saddle-point solution. In fact, some experimentation has shown that, for sample mean losses very close to the threshold, this is the VaR, standard numerical techniques are unable to solve for the saddle point under the normal transformation.

In sum, the exponential approach renders much more tractable the calculations underlying the small-sample asymptotic technique used to derive the hypothesis test's p-value, and provides results consistent with those obtained under the normal transformation.

Redefining the random variable X, we obtain

$$X_t = (\ln F_t(y_t) - \ln F_t(q_t(\alpha)))\mathbb{1}_{\{y_t<q_t(\alpha)\}} = (\ln p_t - \ln\alpha)\mathbb{1}_{\{F_t(y_t)<\alpha\}} = (\ln p_t - \ln\alpha)\mathbb{1}_{\{\ln p_t<\ln\alpha\}} \tag{3.13}$$

where, as with (3.12), the transformations within the indicator function are possible due to the monotonicity of the functions uses. Under this series of transformations, if the series

[13] Another necessary condition for the series to be *iid* is that the $P\&L$ time horizons do not overlap; otherwise, serial interdependencies may occur within the data.

$\{p_t\}$ is *iid* $\mathbb{U}(0,1)$, then the series $\{\ln p_t\}$ is *iid* $Exp(-\infty,0)$ and $\{X_t\}$ is then also clearly *iid*.

Distribution of X_t

Now we need to determine the exact distribution of $\{X_t\}$; knowing its CDF and PDF allows us to determine its moments and cumulants, and these will be used both in determining its theoretical average value and in calculating its test statistic sample value using the small-sample asymptotic technique described below. Under the assumption that the sequence of forecast $P\&L$ distributions is correctly conditionally calibrated, we know that $\{r_t\} := \{\ln p_t\}$ is *iid* $Exp(-\infty,0)$ with PDF $\phi_E(\cdot)$ and CDF $\Phi_E(\cdot)$ as follows

$$\phi_E(r) = \begin{cases} e^r, & -\infty < r < 0, \\ 0, & r \geq 0. \end{cases}$$

and

$$\Phi_E(r) = \begin{cases} e^r, & -\infty < r < 0, \\ 1, & r \geq 0. \end{cases}$$

VaR and ES for the $\{r_t\}$ distribution are

$$\left. \begin{aligned} VaR^0 &= q_E(\alpha) = \ln\alpha \\ ES^0 &= \frac{1}{\alpha}\int_{-\infty}^{\ln\alpha} r\phi_E(r)dr = \frac{1}{\alpha}\int_{-\infty}^{\ln\alpha} re^r dr = \frac{1}{\alpha}(re^r - e^r)|_{-\infty}^{\ln\alpha} = \ln\alpha - 1 \end{aligned} \right\} \tag{3.14}$$

With X defined as per (3.13), we derive its CDF,

$$\begin{aligned} F_X(x) &= \mathbb{P}^X[X \leq x] \\ &= \mathbb{P}^{\phi_E}[r \in \mathbb{R} | (r - ln\alpha)\mathbb{1}_{\{r<\ln\alpha\}} \leq x] \\ &= \begin{cases} \mathbb{P}^{\phi_E}[r \leq \ln\alpha + x], & x < 0, \\ \mathbb{P}^{\phi_E}[r \leq 0], & x \geq 0, \end{cases} \\ &= \begin{cases} \Phi_E(\ln\alpha + x), & x < 0, \\ \Phi_E(0), & x \geq 0, \end{cases} \\ &= \begin{cases} e^{\ln\alpha + x}, & x < 0, \\ 1, & x \geq 0, \end{cases} \\ &= \begin{cases} \alpha e^x, & x < 0, \\ 1, & x \geq 0. \end{cases} \end{aligned}$$

This implies that the PDF of X may be defined in terms of ϕ_E and δ_0, the Dirac delta function with mass concentrates at zero,

$$f_X(x) = \mathbb{P}^X[X = x] = \begin{cases} \alpha e^x, & x < 0, \\ (1-\alpha)\delta_0, & x = 0, \\ 0, & x > 0. \end{cases}$$

Lugannani-Rice formula

We now derive the moment-generating and cumulant-generating functions of X, as these will be needed in the calculation of the Lugannani-Rice formula. We state the relevant results, when $t > -1$, noting only the care that must be taken around the origin in the case of the moment-generating function,

$$\left.\begin{aligned} M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{0} e^{tx} dF_X(x) = \frac{\alpha}{t+1} + 1 - \alpha, \\ M'(t) = -\frac{\alpha}{(t+1)^2}, \quad M''(t) = \frac{2\alpha}{(t+1)^3}, \quad M'''(t) = -\frac{6\alpha}{(t+1)^4}, \\ K(t) = \ln M(t), \quad K'(t) = \frac{M'(t)}{M(t)} = -\frac{\alpha}{(t+1)[t(1-\alpha)+1]}, \\ K''(t) = \frac{\alpha[2t(1-\alpha)+(2-\alpha)]}{(t+1)^2[t(1-\alpha)+1]^2}, \\ K'''(t) = 2\alpha\frac{(1-\alpha)(t+1)[t(1-\alpha)+1] - [2t(1-\alpha)+(2-\alpha)]^2}{(t+1)^3[t(1-\alpha)+1]^3}. \end{aligned}\right\} \tag{3.15}$$

We are now easily able to derive the mean of X as

$$\mu_X = \mathbb{E}[X] = M'(0) = -\alpha \tag{3.16}$$

To highlight the consistency between (3.16), (3.14) and (3.10), we simply note that,

$$TR_0 = \alpha(ES^0 - VaR^0) = \mu_X = -\alpha \tag{3.17}$$

It is also quite simple to calculate the variance of X as

$$\sigma_X^2 = var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = M''(0) - (M'(0))^2 = 2\alpha - \alpha^2 \tag{3.18}$$

It is now straightforward to analyze the sample mean of X in the context of (3.16) and (3.18), using standard statistical asymptotic results. Indeed, given a sample $\{X_t\}$ of size T, the sample mean $\bar{X}$ converge in distribution to standard normality as T tends to ∞ by the Central Limit Theorem. In other words, given population mean μ_X and variance σ_X as per (3.16) and (3.18),

$$\sqrt{T}\left(\frac{\bar{X}-\mu_X}{\sigma_X}\right) \xrightarrow{d} \mathbb{N}(0,1) \tag{3.19}$$

However, as Wong (2010) [94] notes, the above result (3.19) will generally never be valid for sample sizes encountered in practice, due to the inherently nature of the test statistic.

Instead, we now work towards a statement of the Lugannani-Rice formula [72]. As a general approach using the saddle-point technique, Lugannani and Rice (1980) provide a closed-form solution that can be used to accurately approximate the tail probability or CDF of the sample mean of the *iid* random sample $\{X_1, X_2, ..., X_T\}$ drawn from any distribution having well-defined moments and cumulants to the third order, such as the above-defined distribution X.

A general formula may be derived for the PDF of the sample mean $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ by considering the inversion formula when $y < 0$

$$f_{\bar{X}}(y) = \frac{T}{2\pi}\int_{-\infty}^{+\infty} e^{T[K(it)-ity]}dy$$

where $f_{\bar{X}}(\cdot)$ denotes the PDF of the sample mean and $K(t) = \ln M(t)$ is the cumulant-generating function of $f_X(\cdot)$. Then the tail probability can be written as

$$\mathbb{P}[\bar{X} > y] = \int_{y}^{0} f_{\bar{X}}(u)du = \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty} e^{T[K(z)-zy]}z^{-1}dz$$

where s is the saddlepoint, chosen to satisfy

$$K'(s) = \frac{M'(s)}{M(s)} = y$$

for $y < 0$.

Omitting the remainder details [14], we now simply state Lugannani & Rice's result, first defining, for readability,

$$\eta = s\sqrt{TK''(s)} \quad \text{and} \quad \varsigma = sgn(s)\sqrt{2T(s\bar{x} - K(s))}, \qquad K'(s) = \bar{x}$$

where $sgn(s)$ equals zero when $s = 0$, and takes the sign of s when $s \neq 0$. According to Lugannani & Rice, then, the tail probability of exceeding the sample mean $\bar{x} \neq \mu_X$ is given by

$$\mathbb{P}[\bar{X} > \bar{x}] = 1 - \Phi(\varsigma) + \phi(\varsigma)\left(\frac{1}{\eta} - \frac{1}{\varsigma} + \mathcal{O}(T^{-3/2})\right)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent the standard normal CDF and PDF, respectively. The case $\bar{x} = \mu_X$ is rarely encountered in practice, but in this case

$$\mathbb{P}[\bar{X} > \mu_x] = \frac{1}{2} - \frac{K'''(0)}{6\sqrt{2\pi T[K''(0)]^3}} + \mathcal{O}(T^{-3/2})$$

Backtest implementation

They now explicitly formulate the hypothesis test. Though they could use the tail loss $TR_\alpha = \alpha(ES_\alpha - VaR_\alpha)$ as the test statistic, they follow Wong (2010) [94] and define as the test statistic the standardized variable

$$z = \alpha^{-1}TR_\alpha = ES_\alpha - VaR_\alpha$$

[14] For more details, Lugannani and Rice (1980) [72], Daniels (1987) [28] and Wong (2010) [94].

Therefore, under the exponential tail distribution null hypothesis obtained after transforming to (3.13), and with $\alpha = 0.01$, we have that

$$z_0 = \alpha^{-1} TR^0 = ES^0 - VaR^0 = -1, \qquad \text{ie.} \quad TR^0 = -0.01$$

as we observed (3.17). If the theoretical tail risk TR^0 predicted by the model is more negative than the sample tail risk $\widehat{TR}_\alpha$, then the risk model is said to " 'capture' the tail risk ", or to "provide sufficient risk coverage". Otherwise, we may describe the risk model as "having failed to capture the tail risk", based on the empirical tail risk backtest.

Accordingly, a one-tailed regulatory backtest to check whether the risk model provides sufficient risk coverage may be formulated in terms of z, with the null and alternative hypotheses defined as follows:

$$H_0: \quad z \geq z_0, \qquad \text{ie.} \quad TR_\alpha \geq TR^0$$

$$H_1: \quad z < z_0, \qquad \text{ie.} \quad TR_\alpha < TR^0$$

- We wish to asses the evidence in the observed statistical value $\hat{z}$ for H_0.
- The test should measure how surprising the observed value $\hat{z}$ (or, equivalently $\bar{x} = \alpha\hat{z}$) is when H_0 is assumed to be true. It is accepted that $\hat{z}$ is surprising whenever $\hat{z}$ lies in a region of low probability for each distribution in the model for which H_0 is true.
- In terms of the distribution of $\bar{X}$, it means that we should examine the left tail of the distribution.
- Explicitly, the null hypothesis is rejected if the realized value of the sample statistic $\hat{z} = \alpha^{-1}\widehat{TR} = \alpha^{-1}\bar{x}_T$ is significantly less than $z_0 = -1$ under the exponential tail assumption. This is equivalent to $\bar{x}_T = \alpha\hat{z}$ being significantly less than -0.01,

  $$\hat{z} = \alpha^{-1}\widehat{TR} << -1 = z_0$$

  or equivalently,

  $$\bar{x}_T = \alpha\hat{z} << -0.01 \Rightarrow \text{we reject} \quad H_0$$

  In this case, we will tend to see at least one of two occurrences in the sample data:

  1. A large number of (not necessarily extreme) negative x_t sample values, each representing one of the (not necessarily extreme) VaR violations.
  2. a (not necessarily large) number of extreme negative x_t sample values, ie. extreme VaR violations.
- The p-value of the hypothesis test is simply given by the Lugannani-Rice formula as

  $$p - value = \mathbb{P}[\bar{X} \leq \bar{x}_T] = 1 - \mathbb{P}[\bar{X} > \bar{x}_T]$$

As Wong (2010) [94] notes, a two-tailed hypothesis test can be formulated in terms of z, with the null and alternative hypothesis as follows

$$H_0: \quad z = z_0$$

$$H_1: \quad z \neq z_0$$

Strictly speaking, the two-tailed hypothesis identifies whether a given sample of VaR violations is consistent with the tail distribution under the null hypothesis that the risk model is correct. The null hypothesis is rejected if the realized value of the sample statistic $\hat{z}$ is significantly less than or significantly more than $z_0 = -1$.

Summary

Based on the above methodology, it is possible to construct a conceptually flexible, theoretically robust and operationally feasible backtest implementation. The step-by-step procedure is

1. Transform the realized asset/portfolio $P\&L$ values to the exponential context

$$x_t = \begin{cases} \ln F_t(y_t) - \ln \alpha, & y_t < VaR_{t,\alpha} = F_t^{-1}(\alpha), \\ 0, & y_t \geq VaR_{t,\alpha} = F_t^{-1}(\alpha). \end{cases}$$

2. Having constructed the sample $\{x_t\}$ drawn from X, we now construct the sample mean $\bar{x}$ as

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$$

3. We solve the saddle-point equation, recalling the moments and cumulants previously determined in (3.15),

$$K'(s) = \frac{M'(s)}{M(s)} = -\frac{\alpha}{(s+1)[s(1-\alpha)+1]} = \bar{x} \quad \text{for} \quad \bar{x} < 0$$

The unique solution s in the interval $(-1, \infty)$ is then given by,

$$s = \frac{(\alpha - 2) + \sqrt{\Delta}}{2(1-\alpha)} \quad \text{for} \quad \Delta = \alpha^2 + \frac{4\alpha}{\bar{x}}(\alpha - 1) > 0, \quad \bar{x} < 0$$

We now calculate the required components of the hypothesis test,

$$\eta = s\sqrt{TK''(s)} = -s\bar{x}\sqrt[4]{\Delta}\frac{\sqrt{\alpha T}}{\alpha} \quad \text{and} \quad \varsigma = sgn(s)\sqrt{2T(s\bar{x} - K(s))}$$

where $sgn(s)$ takes the sign of s if $s \neq 0$, and is 0 otherwise.

4. We now calculate the Lugannani-Rice formula. The theoretical mean is $\mu_X = M'(0) = -\alpha$. In the case that $\bar{x} \neq \mu_X$, the Lugannani-Rice p-value is

$$p - value = \mathbb{P}[\bar{X} \leq \bar{x}] = \Phi(\varsigma) - \phi(\varsigma)\left(\frac{1}{\eta} - \frac{1}{\varsigma}\right)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent the standard normal CDF and PDF, respectively. The case $\bar{x} = \mu_X$ is hardly ever encountered in practice, though, as previously mentioned, it should be included in any implementation,

$$p - value = \mathbb{P}[\bar{X} \leq \mu_X] = \frac{1}{2} + \frac{K'''(0)}{6\sqrt{2\pi T[K''(0)]^3}}$$

5. We now perform the hypothesis analysis (one-tailed),

$$H_0 : TR_\alpha \geq TR^0$$

$$H_1 : TR_\alpha < TR^0$$

The null hypothesis is that the risk model provides sufficient tail risk coverage, whereas the alternative hypothesis is that the risk model provides insufficient tail risk coverage.

3.5.1.4 The Costanzino & Curran and Du & Escanciano Approaches

(These approaches have been extracted from Costanzino & Curran (2015) [26] and Du & Escanciano (2015) [36].)

Du & Escanciano (2015) [36] propose backtest for ES based on cumulative violations, which are the natural analogue of the commonly used backtest for VaR. It is well-known that for each coverage level, violations should be unpredictable if the risk model is appropriate, i.e. centered violations should be a martingale difference sequence (*mds*). Indeed, rather than just one *mds*, centered violations form a class of *mds* indexed by the coverage level. The integral of the violations over the coverage level in the left tail, which they refer to as cumulative violations, also form a *mds*. The cumulative violation process accumulates all violations in the left tail, just like the ES accumulates the VaR in the left tail. They suggest a Box-Pierce test to check for the *mds* property. Their Box-Pierce test is analogue for ES of the conditional backtest proposed by Christoffersen (1998) [22] and Berkowitz, Christoffersen and Pelletier (2011) [16] for VaR.

This approach is developed from Costanzino & Curran (2015) [26]. Their methodology can be used to backtest any spectral risk measure [15], including ES. It is based on

[15] Spectral risk measures are a class of coherent measures, which can be thought of weighing VaR by a spectrum $\phi \in L^1([0,1])$, which is an admissible risk spectrum if, *i)* ϕ is non-negative, *ii)* ϕ is non-increasing, and *iii)* $||\phi||_1 = 1$. Let X a random variable with cdf F_X and ϕ an admissible risk spectrum, we say that $\mathcal{M}_\phi$ defined by $\mathcal{M}_\phi = \int_0^1 \phi(p)VaR(p)dp$ is a spectral risk measure with risk spectrum ϕ. If $\phi(p) = Dirac_\alpha(p)$, we obtain $\mathcal{M}_{Dirac_\alpha} = VaR_\alpha$ but $Dirac_\alpha(p)$ is not an admissible risk spectrum for violation of properties *ii)* and *iii)*. For this reason, VaR is not a spectral risk measure. For more details, Costanzino & Curran (2015) [26].

the idea that ES is an average of a continuum of VaR levels. Indeed, Emmer, Kratz & Tasche (2015) [39] were the first to suggest that, despite not being directly elicitable, ES is indirectly elicitable since it can be approximated by several VaR estimates. They suggest approximating ES by

$$ES_t^\alpha = \frac{1}{\alpha}\int_0^\alpha VaR_t^\nu d\nu \approxeq \frac{1}{4}(VaR_t^{\alpha/4} + VaR_t^{\alpha/2} + VaR_t^{3\alpha/4} + VaR_t^\alpha)$$

and then backtesting ES by backtesting the individual VaRs at the four different confidence levels. However, this leads to a possible ambiguous decision making framework. The Costanzino & Curran Coverage Test can be thought of as the continuous limit of the Emmer, Kratz & Tasche idea in that it is a joint test of a continuum of VaR levels. Furthermore, the general methodology is perfectly compatible with the VaR coverage test in the sense that if the spectrum ϕ converges to the point mass at α, ie. $\phi_{SR} \to \delta_\alpha$, the test converges to the VaR coverage test despite the fact that δ_α is not a spectral risk measure (for more details [26]).

Unlike the test proposed by Du & Escanciano [36], the test proposed by Costanzino & Curran [26] does not test independence, but they are the first to propose a coverage test for Spectral Risk Measures essentially amounts to a joint test of a continuum of weighted VaR quantiles and gives a single decision at a fixed confidence level. The key of the method is to show that the Spectral Measure Failure Rate is asymptotically Normal under the null hypothesis and therefore admits a formal Z-test.

The cumulative violation process

The cumulative violation process is defined as

$$H_t(\alpha) = \frac{1}{\alpha}\int_0^\alpha h_t(u)du$$

where $h_t(u) = \mathbb{1}_{(r_t \leq VaR_t(u))}$ is the u-violation or hit at time t.

Since $h_t(u)$ has mean u, by Fubini's Theorem $H_t(\alpha)$ has mean $1/\alpha \int_0^\alpha udu = \alpha/2$. Moreover, again by Fubini's Theorem, the *mds* property of the class $\{h_t(\alpha) - \alpha : \alpha \in [0,1]\}_{t=1}^\infty$ is preserved by integration, which means that $\{H_t(\alpha) - \alpha/2\}_{t=1}^\infty$ is also *mds*.

For computational purposes, it is convenient to define $u_t = F(r_t, \Omega_{t-1})$ where $F(\cdot, \Omega_{t-1})$ denote the conditional cumulative distribution function (cdf) of r_t given Ω_{t-1} . Using that $h_t(u) = \mathbb{1}_{(r_t \leq VaR_t(u))} = \mathbb{1}_{(u_t \leq u)}$, we obtain [16],

$$H_t(\alpha) = \frac{1}{\alpha}\int_0^\alpha \mathbb{1}_{(u_t \leq u)}du = \frac{1}{\alpha}(\alpha - u_t)\mathbb{1}_{(u_t \leq \alpha)} \tag{3.20}$$

Like violations, cumulative violations are distribution-free, since $\{u_t\}_{t=1}^\infty$ comprises a sample of independent and identically distributed (*iid*) $\mathbb{U}(0,1)$ variables. Acerbi &

[16] $H_t(\alpha) = \frac{1}{\alpha}\int_0^\alpha \mathbb{1}_{(u_t \leq u)}du = \frac{1}{\alpha}\mathbb{1}_{(u_t \leq \alpha)}\int_0^\alpha \mathbb{1}_{(u_t \leq u)}du = \frac{1}{\alpha}\mathbb{1}_{(u_t \leq \alpha)}\int_{u_t}^\alpha 1du = \frac{1}{\alpha}\mathbb{1}_{(u_t \leq \alpha)}(\alpha - u_t)$.

Tasche (2002) [4] and Emmer, Kratz and Tasche (2014) [39] used the representation $ES_t(\alpha) = 1/\alpha \int_0^\alpha VaR_t(u)du$ to approximate the integral with a Riemann sum with four terms. Working with violations avoids approximations, as the integral in (3.20) can be computed exactly. Unlike violations, cumulative violations are also zero, but when a violation occurs, the cumulative violation measures how far is the actual value of r_t from its quantile, through the term $\alpha - u_t = F(F^{-1}(\alpha, \Omega_{t-1}), \Omega_{t-1}) - F(r_t, \Omega_{t-1})$.

The variables $\{u_t\}_{t=1}^{\infty}$ necessary to construct $\{H_t(\alpha)\}_{t=1}^{\infty}$ are generally unknown, since the distribution of the data F is unknown. In practice, researchers and risk managers specify a parametric conditional distribution $F(\cdot, \Omega_{t-1}, \theta_0)$, where θ_0 is some unknown parameter in $\Theta \subset \mathbb{R}^p$, and proceed to estimate θ_0 before producing VaR and ES forecasts.

With the parametric model, we can define the "generalized errors",

$$u_t(\theta_0) = F(r_t, \Omega_{t-1}, \theta_0)$$

and the associated cumulative violations,

$$H_t(\alpha, \theta_0) = \frac{1}{\alpha}(\alpha - u_t(\theta_0))\mathbb{1}_{(u_t \leq \alpha)}$$

Very much like for VaRs, the arguments above provide a theoretical justification for backtesting ES by checking whether $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^{\infty}$ have zero mean (unconditional ES backtest) and whether $\{H_t(\alpha, \theta_0) - \alpha/2\}_{t=1}^{\infty}$ are uncorrelated (conditional ES backtest).

The unconditional backtest

The unconditional backtest for ES is a standard t-test for the null hypothesis

$$H_{0u} = \mathbb{E}(H_t(\alpha, \theta_0)) = \alpha/2$$

Note that a simple calculations show that $\mathbb{E}[H_t^2(\alpha, \theta_0)] = \alpha/3$, and hence, $Var(H_t(\alpha)) = \alpha(1/3 - \alpha/4)$. Therefore, a simple t-test statistic is as follows

$$U_{ES} = \frac{\sqrt{n}(\bar{H}(\alpha) - \alpha/2)}{\sqrt{\alpha(1/3 - \alpha/4)}}$$

where n is the size of the out-of-sample period which is used to evaluate (backtest) the ES model and $\bar{H}(\alpha)$ denotes the sample mean of $\{\hat{H}_t(\alpha)\}_{t=1}^{n}$, i.e.

$$\bar{H}(\alpha) = \frac{1}{n}\sum_{t=1}^{n}\hat{H}_t(\alpha)$$

The U_{ES} statistic has a standard normal limit distribution when the estimation period is much larger than the evaluation period, $U_{ES} \xrightarrow{d} N(0,1)$.

The conditional backtest

Next, the conditional backtest has the following null hypothesis,

$$H_{0c} : \mathbb{E}[H_t(\alpha, \theta_0) - \alpha/2|\Omega_{t-1}] = 0$$

which is the analogue of the null hypothesis of conditional backtest for VaR. Define the lag-j autocovariance and autocorrelation of $H_t(\alpha)$ for $j \geq 0$ by

$$\gamma_j = Cov(H_t(\alpha), H_{t-j}(\alpha)) \quad \text{and} \quad \rho_j = \frac{\gamma_j}{\gamma_0}$$

respectively. We drop the dependence of γ_j and other related quantities on α for simplicity of notation. The sample counterparts of γ_j and ρ_j based on a sample $\{H_t(\alpha)\}_{t=1}^{n}$ are [17]

$$\gamma_{nj} = \frac{1}{n-j} \sum_{t=1+j}^{n} (H_t(\alpha) - \alpha/2)(H_{t-j}(\alpha) - \alpha/2) \quad \text{and} \quad \rho_{nj} = \frac{\gamma_{nj}}{\gamma_{n0}}$$

Notice that $\rho_j = 0$ for $j \geq 1$ under H_{0c}. Simple conditional tests can be constructed using $\hat{\rho}_{nj}$, for example the Box-Pierce test statistic

$$C_{ES}(m) = n \sum_{j=1}^{m} \hat{\rho}_{nj}^2$$

The C_{ES} statistic has a chi-square distribution with m degrees of freedom when the estimation period is much larger than the evaluation period, $C_{ES}(m) \xrightarrow{d} \chi_m^2$.

3.5.2 Full sample analysis

First, we calculate VaR and ES estimates by the parametric approach. We restrict our attention to the left tail of the distribution and the 1%, 2.5% and 5% significance levels. We choose to work with these α's because these values are the significance levels most commonly used in the literature. In all cases we show out-of-sample VaR and ES estimates over the last five years in the sample: 2012-2016 (1260 data). Every day we compute 1-day ahead VaR and ES, estimating each model every 50 days. The latter choice tries to reduce the computational cost while avoiding frequent parameter variation due in part to pure noise.

Table 3.6 displays the descriptive statistics for the in-sample (10/02/2000-12/02/2011) and out-of-sample (12/05/2011-09/30/2016) periods. It gives an overview of the behavior

[17] Notice that we use the conditional mean restriction in the definition of autocorrelations. As a result, test bases on γ_{nj} are expected to have power against deviations from H_{0c}, where $H_t(\alpha)$ are uncorrelated but have mean different from $\alpha/2$. It is also possible consider a test that desegregates power against deviations from zero autocorrelations of $H_t(\alpha)$ and power against deviations from H_{0u}, analogue for ES of the conditional backtests proposed by Christoffersen (1998) [22]. One could also consider tests that do not use the unconditional mean restriction in the definition of autocorrelations, for example, using the sample mean of cumulative violations.

of asset distributions for time horizons. Skewness is negative, except for SAN and AXA in the in-sample period. Likewise, kurtosis is higher than 3 for all assets in both periods of time. We are thus confronted with leptokurtic distributions, i.e. with fat tails in comparison with the Normal distribution. These distributions do not follow Normal distribution as shown by the Jarque-Bera statistic (JB stat). These statistics suggest that, the study of the behavior of the extremes in the left tails of these leptokurtic distributions seems justified as it should allow better estimation of extreme variations of financial returns. VaR and ES estimates based on the assumption of a Normal distribution of returns must then be rejected and we need to compute them under a non-Normal framework using EVT framework, for example.

	In-Sample				**Out-of-Sample**			
Daily	**IBM**	**SAN**	**AXA**	**BP**	**IBM**	**SAN**	**AXA**	**BP**
No. Obs.	2915	2915	2915	2915	1260	1260	1260	1260
Mean (%)	1.79	-2.18	-3.94	-0.89	-1.41	-0.13	4.26	-0.26
Median (%)	0.00	0.00	0.00	0.00	0.00	1.94	11.04	0.00
St. Dev	1.74	2.31	2.96	1.80	1.18	2.13	1.93	1.43
Skewness	-0.12	0.27	0.31	-0.19	-1.00	-1.09	-0.72	-0.16
Kurtosis	11.47	9.09	9.33	7.87	9.77	14.79	9.10	6.84
Maximum	11.35	20.88	19.78	10.58	4.91	10.14	7.28	6.93
10 percentile	-1.73	-2.63	-3.07	-1.96	-1.24	-2.43	-2.09	-1.60
5 percentile	-2.66	-3.67	-4.60	-2.73	-1.73	-3.38	-3.22	-2.31
1 percentile	-5.13	-6.68	-8.46	-5.32	-3.59	-4.97	-4.96	-3.59
Minimum	-16.89	-12.72	-20.35	-14.04	-8.64	-22.17	-16.82	-9.08
JB stat	8724.16	4548.12	4918.84	2902.54	2614.85	7549.73	2063.42	780.63

Table 3.6: Descriptive statistics for the log-returns (%) of four indexes for in-sample and out-of-sample period. JB stat is the statistic of Jarque-Bera test.

We estimate both risk measures, not only with the full distribution but also using only extreme events. The estimation exercise was explained in subsections 3.3.4 and 3.3.5.

Figure 3.4 shows IBM daily percentage returns (1260 data) and the out-of-sample $VaR_{1\%}$ and $VaR_{5\%}$ calculated with AR(1) model with JSU-APARCH(1,1) as well as with JSU-APARCH(1,1) innovation, fitting a GPD density in the tail of the distribution (EVT). The differences in VaR calculated with the two models are not significant for the 5% quantile but they become more important for the 1% quantile. In the latter case, the VaR with EVT indicate higher losses than VaR without EVT.

Figure 3.5 shows the $ES_{1\%}$ and $ES_{5\%}$ estimates with EVT and without EVT. In this case, differences are significant, especially for the more extreme quantiles. ES measures the average losses exceeding the VaR and we observe that the GPD distribution considers greater losses than JSU for the tail of the distribution.

It is interesting to look at how the ratio of the two risk measures estimated under EVT

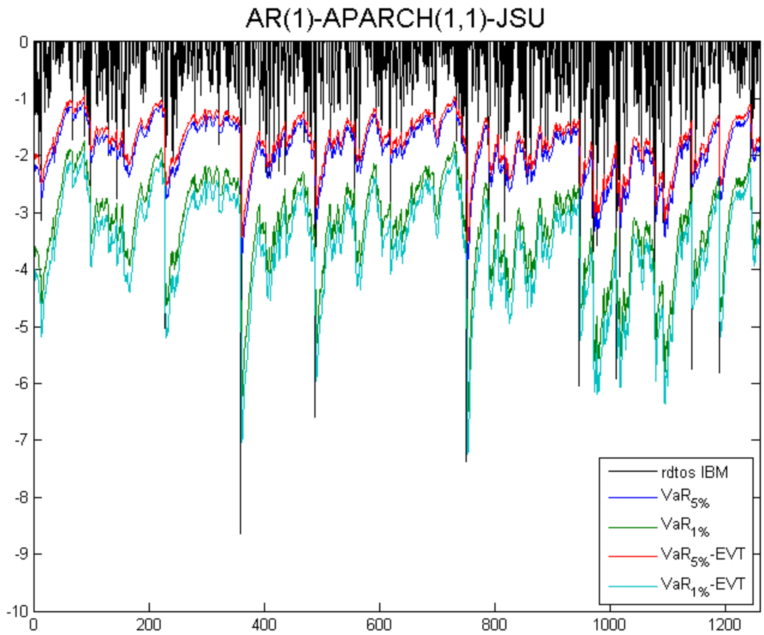


Figure 3.4: IBM daily percentage returns and $VaR_{1\%}$ and $VaR_{5\%}$ calculated with all sample as well as using only extreme values.

approach behaves for large values of the quantile probability α.

$$\lim_{\alpha \to 0} = \frac{ES_\alpha}{VaR_\alpha} = \begin{cases} (1-\xi)^{-1}, & \xi \geq 0, \\ 1, & \xi < 0. \end{cases}$$

so that the shape parameter ξ of the GPD effectively determines the ratio when we go far enough out into the tail.

Figure 3.6 shows the evolution of the ratio ES_α and VaR_α calculated with EVT from model JSU-APARCH for IBM, with parameter $\xi = 0.392$. When $\alpha \to 0$, this ratio tends to $(1-\xi)^{-1} = 1.644$.

Tables 3.8 - 3.11 present the average value of the estimated ES [18], the violations ratio (Viol.) of the underlying VaR and the backtesting results for the distinct models.

The results across assets indicate that there are too many variations regarding the ES estimation models in all of the dimensions of our analysis. Although there are certain exceptions, our discussion here is focused on the general patterns that appear in the results. In what concerns parsimony of ES_t [19], the conditional EVT-based models present "more negative" values than do conditional models not based on EVT, as indicated by

[18] We have obtained out-of-sample 1-day ES series whose average value is shown in the first column of these tables.

[19] We are talking about parsimony, when we want to refer to the behavior of the ES series, since this

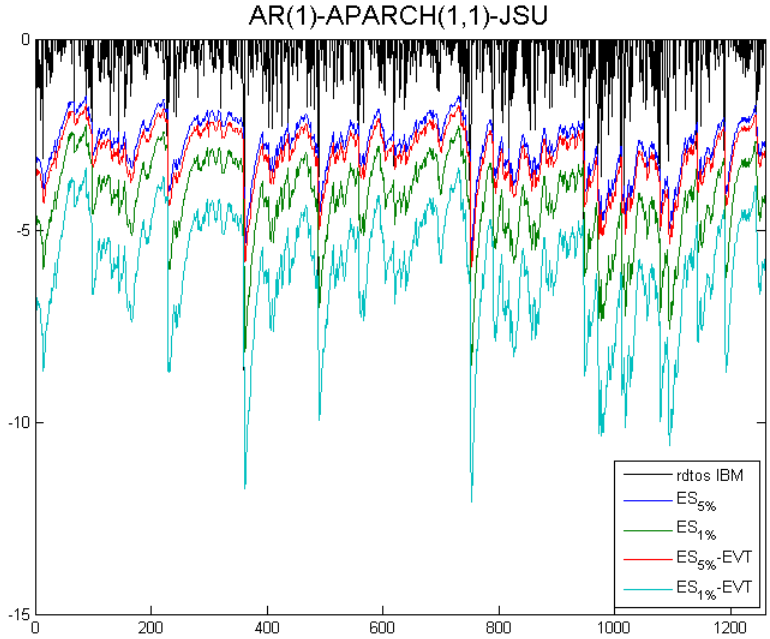


Figure 3.5: IBM daily percentage returns and $ES_{1\%}$ and $ES_{5\%}$ calculated with all sample as well as using only extreme values.

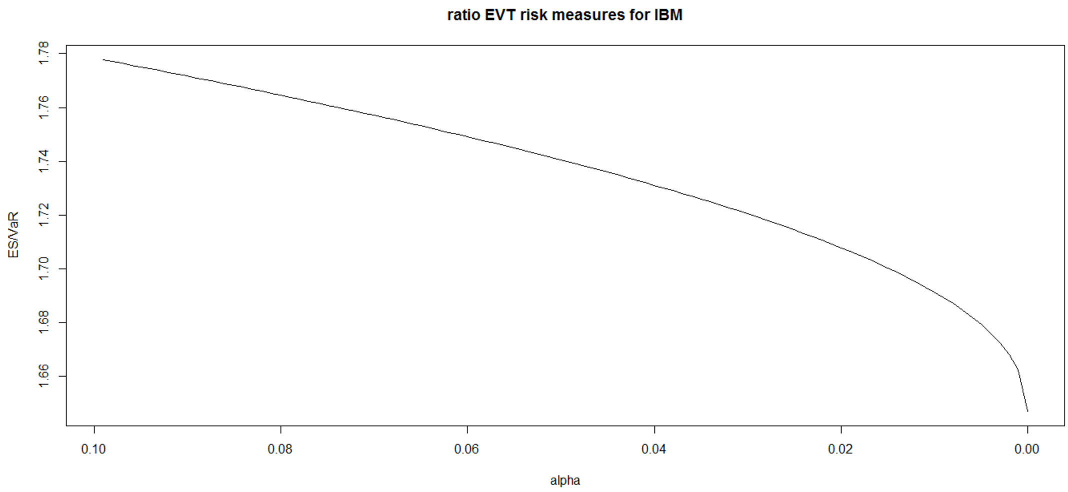


Figure 3.6: Evolution of the ES_α/VaR_α ratio calculated with EVT from model JSU-APARCH for IBM.

the averages for the ES estimates. In addition, the 1% significance level presents more differences in ES estimates than does the 5% significance level. This analysis of the parsimony is related to the complete out-of-sample period (5 years, 1260 data). Furthermore, an ES model should be precise when violations occur, but it is not always like this. Regarding VaR violation rates, we observe that the conditional EVT-based models present more violations than what is theoretically expected. Most of these models obtain correct ES estimates, as we observe in backtest results, but their VaR predictions, in some cases, underestimate risk, for example AXA and SAN at the 2.5% and 5% significance levels, but with all models we obtain a lower violation rate than this obtained with the homologous model not based on EVT. In general, we can say that conditional EVT-based models estimate the quantile correctly, corroborating Kuester, Mittnik and Paolella (2006) [67], which attest to the superiority of this type of model. On the other hand, conditional models with heavy-tailed distributions also perform very well, corroborating Mabrouk and Saadi (2012) [73]. It is very important to note that these tests are one-sided by nature and focus on risk underestimation, except the tests of Costanzino & Curran and Du & Escanciano. Therefore, in those tests risk overestimation does not lead to a rejection of the null hypothesis.

In general, we observe that conditional EVT-based models obtain the best results in ES estimation according to the different ES backtests. In many cases, we obtain p-values close to 1 with EVT-based models. The success of the EVT models for the ES estimation corroborates Marinelli *et al.* (2007) [74], Jalal and Rockinger (2008) [62] and Wong *et al.* (2012) [93]. If we focus on the conditional models not based on EVT, we observe that JSU and SKST are the best probability distributions for ES estimation according to Righi & Ceretta test [83] and Graham & Pál test [57] and SGED and JSU according to Acerbi & Szekely tests [4] and Costanzino & Curran and Du & Escanciano tests. Besides, in the Acerbi & Szekely tests we obtain p-values for $Z2$ greater than the p-values for $Z1$ at the 5% significance level. The second test of Acerbi & Szekely ($Z2$) tests ES directly and jointly evaluates frequency and magnitude of α-tail events as shown by the relationship $Z_2 = (1 + Z_1)N_T/T\alpha - 1$ if $N_T = \sum_{t=1}^{T} I_t > 0$ where $I_t = \mathbb{1}_{\{r_t < VaR_t^{\alpha}\}}$ is the indicator of VaR breaches. Remember that $\mathbb{E}_{H_0}[N_T] = T\alpha$. The difference between both is that $Z1$ is insensitive to an excessive number of exceptions as it is an average taken over excesses themselves, whereas null hypothesis of $Z2$ test is not rejecting if not only the magnitude but also the frequency of the excesses is statistically equal to the expected one [20].

At the 1% significance level, p-values of Acerbi & Szekely tests for the conditional models are very close to 0. In these cases, we obtain positive realized value $Z_1(\overrightarrow{R})$ and $Z_2(\overrightarrow{R})$, instead of being equal to zero. In short, we reject H_0 because of risk underestimation.

The p-values obtained in Graham & Pál test show, like Acerbi & Szekely, large differ-

series changes over time thanks to the parsimonious model used for the volatility, in our case, APARCH model.

[20] Acerbi & Szekely (2014) show that Z_2 test is the most powerful in the case of alternative hypothesis with different volatility, while Z_1 is the most powerful in the case of the different tail index.

ences between conditional models based on EVT and not based on EVT, in favor of the first ones.

The Righi & Ceretta, Acerbi & Szekely and Graham & Pál tests are one-tail tests, i.e. only risk underestimation lead to a rejection of the null hypothesis. We indicate in bold face the p-values of the test in which we have obtained statistics with opposite sign to the collected in the alternative hypothesis. In the Righi & Ceretta test we have $H_0 : \mathbb{E}(BT_t) = 0$ against $H_1 : \mathbb{E}(BT_t) < 0$ but with some models we obtain $\mathbb{E}(BT_t) > 0$, i.e. most excesses are between VaR and ES, not beyond ES, especially, under EVT approach. In the first test of Acerbi & Szekely we have $H_0 : \mathbb{E}(Z_1) = 0$ against $H_1 : \mathbb{E}(Z_1) > 0$ and in the second one, $H_0 : \mathbb{E}(Z_2) = 0$ against $H_1 : \mathbb{E}(Z_2) > 0$, but with some models, especially models based on EVT, we obtain $\mathbb{E}(Z_1) < 0$ and $\mathbb{E}(Z_2) < 0$, respectively, i.e. in the first test, the average of the realized excesses is lower than forecast ES and in the second one, not only the average taken over excesses but also the number of excesses is lower than expected according to forecast VaR and ES. Finally, in the Graham & Pál test we have $H_0 : TR_\alpha = TR^0$ against $H_1 : TR_\alpha < TR^0$ where TR^0 is equal to $-\alpha$ under exponential assumption and the null hypothesis is rejected if the realized value of the sample statistic $\widehat{TR}_\alpha$ is significantly less than the theoretical tail risk TR^0. If we obtain $TR_\alpha > TR^0$, we will say that the risk model captures the tail risk or provide sufficient risk coverage, but sometimes it is possible a significant risk overestimation not identified in this contrast. This happens when the logarithmic difference between the probabilities of excess and the VaR probability (α) does not follow an exponential distribution (this series follows an exponential distribution if the forecast CDF is correct estimates of the real and unobservable P&L distribution) but another distribution with thicker tails than the exponential is obtained. In most cases, the distance from a statistic equal to zero is not very large, so it would not lead to a significant overestimation of the risk, but it is important to keep in mind that sometimes the EVT approach, and to a lesser extent non-EVT-based models, can produce risk overestimation does not identified with these tests.

The tests of Costanzino & Curran and Du & Escanciano are two-tail tests, i.e. both, risk underestimation and overestimation lead to a rejection of the null hypothesis. These tests are based on cumulative violation process. Unlike violations, cumulative violations H_t [21] contain information on the tail risk and, therefore, provide a more complete description of the risk involved. Besides, its main advantage is that the distribution of the test statistic is available for finite out-of-sample size which leads to better size and power properties compared to another tests. Table 3.7 reports the expected value of violations ($n\alpha$), the number of violations $\left(V(\alpha) = \sum_{t=1}^{n} \hat{h}_t(\alpha)\right)$ and the cumulative violations $\left(CV(\alpha) = \sum_{t=1}^{n} \hat{H}_t(\alpha)\right)$. Comparing $V(\alpha)$ and $CV(\alpha)$ with AR(1)-APARCH(1,1)-JSU model based on EVT and not based on EVT for IBM, we observe that with EVT approach the number of $VaR_{5\%}$ violations and of $VaR_{2.5\%}$ violations increase, but for 2.5%, the losses are less than with model not based on EVT. At 1% significance level, we obtain the same number of violations (13) but we get smaller losses with EVT approach.

[21] In subsection 3.3.4, in "The Costanzino & Curran and Du & Escanciano Approaches" is the explanation of how to calculate H_t.

Comparing IBM and SAN, for SAN we obtain in general a greater number of violations and larger losses, except at 1% significance level. For this significance level, for IBM we obtain a greater number of violations and larger losses with JSU-EVT, and we obtain 17 violations for SAN versus 13 for IBM but 7.06 versus 10.41 of losses with JSU probability distribution. Table 3.7 shows significant discrepancies between violations and cumulative violations at the different coverage levels.

In Tables 3.8 - 3.11, according to p-values obtained in Costanzino & Curran and Du & Escanciano tests, we do not reject the null hypothesis of conditional test with all considered models, except for SAN with models not based on EVT, i.e. the cumulative violations calculated for SAN is having autocorrelation for the first five lags when those are calculated at 1% significance level. Regarding the unconditional test, in general we do not reject the null hypothesis with all considered models, except for IBM with models not based on EVT for $ES_{1\%}$ considering at 5% significance level for the hypothesis contrast, for SAN, AXA and BP with Normal distribution for $ES_{1\%}$, for SAN and AXA with Normal and Student-t distributions for $ES_{2.5\%}$ and for AXA with Normal and Student-t distributions for $ES_{5\%}$.

	IBM		**SAN**	
	JSU	**JSU-EVT**	**JSU**	**JSU-EVT**
$n\alpha$	63	63	63	63
V(0.05)	51	64	63	66
CV(0.05)	26.56	30.03	32.32	31.48
$n\alpha$	31.5	31.5	31.5	31.5
V(0.025)	24	29	34	33
CV(0.025)	16.05	14.72	17.63	15.04
$n\alpha$	12.6	12.6	12.6	12.6
V(0.01)	13	13	17	11
CV(0.01)	10.41	7.63	7.06	4.73

Table 3.7: Descriptive Analysis of Violations for IBM and SAN with AR(1)-APARCH(1,1)-JSU and with AR(1)-APARCH(1,1)-EVT-JSU.

Figures 3.7 and 3.8 plot the cumulative violations $\{\hat{H}_t(0.05)\}$, $\{\hat{H}_t(0.025)\}$ and $\{\hat{H}_t(0.01)\}$ of the individual stocks IBM and SAN respectively in the out-of-sample period with JSU-APARCH and JSU-EVT-APARCH models. We do not observe large values of $\{\hat{H}_t(\alpha)\}$, but we observe some clusters of cumulative violations, which suggest deviations from the martingale difference sequence *mds* hypothesis that would be implied by an appropriate ES forecast. In fact, we reject null hypothesis $\mathbb{E}(\hat{H}_t(0.01)) \neq \alpha/2$ at 5% significance level, only with JSU-APARCH for IBM, because U_{ES} p-value is equal to 0.02. We observe, in general, for SAN and with models not based on EVT more cumulative violations than for IBM and with models based on EVT.

IBM				**1% significance level**					
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.249	0.014	0.00	0.01	0.01	0.00	0.00	0.65	0.97
ST	-4.205	0.010	0.07	0.00	0.00	0.00	0.01	0.72	0.99
SKST	-4.266	0.010	0.07	0.00	0.00	0.00	0.02	0.73	0.99
SGED	-3.921	0.010	0.02	0.00	0.00	0.00	0.01	0.72	0.99
JSU	-4.206	0.010	0.06	0.00	0.00	0.00	0.02	0.73	0.99
N-EVT	-5.931	0.010	**0.38**	**0.96**	**1.00**	0.30	0.26	0.76	0.99
ST-EVT	-6.059	0.010	**0.35**	**0.92**	**0.99**	0.30	0.26	0.77	0.99
SKST-EVT	-6.050	0.010	**0.35**	**0.93**	**1.00**	0.30	0.26	0.77	0.99
SGED-EVT	-5.923	0.010	**0.34**	**0.89**	**0.98**	0.29	0.25	0.77	0.99
JSU-EVT	-6.052	0.010	**0.34**	**0.90**	**1.00**	0.30	0.26	0.76	0.99
IBM				**2.5% significance level**					
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.847	0.022	0.01	0.01	0.03	0.00	0.13	0.79	0.84
ST	-3.305	0.024	0.13	0.03	0.05	0.01	0.29	0.86	0.89
SKST	-3.350	0.021	0.12	0.01	0.08	0.03	0.39	1.00	0.91
SGED	-3.255	0.018	0.05	0.01	0.32	0.00	0.48	0.63	0.91
JSU	-3.352	0.019	0.10	0.01	0.32	0.02	0.46	0.81	0.92
N-EVT	-4.042	0.022	**0.36**	**0.79**	**0.95**	0.49	0.35	0.70	0.91
ST-EVT	-4.122	0.022	0.31	0.68	0.94	0.47	0.38	1.00	0.93
SKST-EVT	-4.116	0.024	**0.34**	**0.73**	**0.96**	0.47	0.38	0.99	0.93
SGED-EVT	-4.056	0.022	0.32	0.66	0.92	0.48	0.35	0.69	0.91
JSU-EVT	-4.116	0.023	0.32	0.68	**0.99**	0.47	0.38	1.00	0.93
IBM				**5% significance level**					
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.510	0.037	0.05	0.04	0.49	0.00	0.17	0.37	0.55
ST	-2.699	0.044	0.18	0.13	0.40	0.09	0.33	0.11	0.41
SKST	-2.734	0.044	0.20	0.11	0.39	0.16	0.22	0.11	0.40
SGED	-2.738	0.037	0.15	0.07	0.75	0.02	0.06	0.36	0.58
JSU	-2.752	0.040	0.18	0.11	0.67	0.15	0.14	0.13	0.42
N-EVT	-3.003	0.053	**0.48**	**0.69**	**0.90**	0.52	0.41	0.41	0.67
ST-EVT	-3.055	0.049	**0.40**	**0.70**	**0.97**	0.54	0.37	0.13	0.46
SKST-EVT	-3.050	0.050	**0.41**	**0.76**	**0.96**	0.54	0.37	0.13	0.46
SGED-EVT	-3.015	0.053	**0.44**	**0.69**	**0.96**	0.54	0.36	0.23	0.59
JSU-EVT	-3.050	0.051	**0.41**	**0.63**	**0.93**	0.54	0.37	0.13	0.45

Table 3.8: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for IBM. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

SAN	1% significance level								
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-5.319	0.021	0.05	0.01	0.00	0.00	0.00	0.58	0.06
ST	-6.143	0.015	0.14	0.01	0.00	0.01	0.06	0.70	0.00
SKST	-6.400	0.014	0.16	0.01	0.00	0.03	0.28	0.77	0.00
SGED	-6.174	0.014	0.11	0.01	0.01	0.00	0.22	0.77	0.00
JSU	-6.473	0.013	0.17	0.01	0.00	0.02	0.35	0.79	0.00
N-EVT	-7.757	0.009	**0.96**	**0.99**	**1.00**	0.52	0.24	0.84	0.09
ST-EVT	-7.811	0.009	**0.96**	**0.99**	**1.00**	0.52	0.22	0.85	0.07
SKST-EVT	-7.766	0.009	**0.96**	**0.97**	**0.99**	0.52	0.22	0.85	0.06
SGED-EVT	-7.724	0.009	**0.97**	**1.00**	**1.00**	0.52	0.22	0.85	0.08
JSU-EVT	-7.771	0.009	**0.96**	**0.99**	**0.99**	0.52	0.22	0.85	0.07
SAN	**2.5% significance level**								
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-4.667	0.033	0.08	0.02	0.02	0.00	0.00	0.44	0.52
ST	-5.096	0.033	0.20	0.06	0.01	0.01	0.04	0.43	0.35
SKST	-5.294	0.028	0.19	0.09	0.07	0.04	0.20	0.44	0.24
SGED	-5.232	0.027	0.17	0.04	0.03	0.00	0.27	0.46	0.20
JSU	-5.354	0.027	0.20	0.03	0.03	0.05	0.28	0.45	0.20
N-EVT	-5.805	0.028	**0.96**	**0.96**	**0.99**	0.53	0.43	0.48	0.17
ST-EVT	-5.842	0.026	**0.94**	**0.95**	**0.99**	0.53	0.42	0.48	0.14
SKST-EVT	-5.814	0.026	**0.94**	**0.93**	**0.95**	0.54	0.41	0.48	0.14
SGED-EVT	-5.788	0.026	**0.95**	**0.94**	**0.97**	0.54	0.41	0.48	0.15
JSU-EVT	-5.816	0.026	**0.94**	**0.96**	**0.99**	0.54	0.41	0.48	0.14
SAN	**5% significance level**								
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-4.120	0.052	0.14	0.06	0.06	0.00	0.07	0.70	0.52
ST	-4.322	0.056	0.23	0.15	0.06	0.01	0.09	0.99	0.48
SKST	-4.477	0.050	0.23	0.06	0.05	0.09	0.34	0.82	0.52
SGED	-4.479	0.050	0.24	0.12	0.16	0.01	0.47	0.60	0.52
JSU	-4.518	0.050	0.24	0.12	0.13	0.11	0.43	0.77	0.53
N-EVT	-4.589	0.053	**0.93**	**0.95**	**0.97**	0.50	0.49	0.72	0.60
ST-EVT	-4.609	0.052	**0.91**	**0.92**	**0.97**	0.51	0.50	0.82	0.53
SKST-EVT	-4.586	0.052	**0.92**	**0.90**	**0.94**	0.51	0.50	0.86	0.55
SGED-EVT	-4.567	0.052	**0.92**	**0.98**	**0.99**	0.51	0.49	0.77	0.57
JSU-EVT	-4.588	0.052	**0.91**	**0.93**	**0.96**	0.51	0.50	0.84	0.55

Table 3.9: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for SAN. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano.The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

AXA			1% significance level						
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-5.042	0.021	0.13	0.03	0.00	0.00	0.00	0.68	0.84
ST	-5.623	0.017	0.25	0.06	0.01	0.03	0.08	0.70	0.99
SKST	-5.774	0.013	0.19	0.01	0.00	0.09	0.25	0.75	0.99
SGED	-5.606	0.014	0.18	0.00	0.00	0.00	0.21	0.76	0.99
JSU	-5.807	0.012	0.20	0.02	0.01	0.09	0.32	0.76	0.99
N-EVT	-6.399	0.009	**1.00**	**1.00**	**1.00**	**0.66**	0.12	0.84	1.00
ST-EVT	-6.460	0.009	**1.00**	**1.00**	**1.00**	**0.66**	0.13	0.84	1.00
SKST-EVT	-6.448	0.009	**1.00**	**0.99**	**0.99**	**0.66**	0.13	0.84	1.00
SGED-EVT	-6.429	0.009	**1.00**	**0.99**	**1.00**	**0.66**	0.12	0.84	1.00
JSU-EVT	-6.447	0.009	**1.00**	**1.00**	**1.00**	**0.66**	0.13	0.84	1.00
AXA			**2.5% significance level**						
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-4.425	0.034	0.15	0.05	0.03	0.00	0.00	0.41	0.55
ST	-4.727	0.033	0.24	0.08	0.03	0.01	0.01	0.33	0.55
SKST	-4.846	0.032	0.26	0.09	0.02	0.06	0.06	0.29	0.58
SGED	-4.801	0.032	0.26	0.04	0.99	0.01	0.11	0.35	0.64
JSU	-4.876	0.032	0.28	0.11	0.00	0.07	0.10	0.29	0.60
N-EVT	-4.989	0.027	**1.00**	**0.99**	**1.00**	**0.70**	0.27	0.57	0.82
ST-EVT	-5.035	0.027	**1.00**	**1.00**	**1.00**	**0.68**	0.31	0.31	0.73
SKST-EVT	-5.023	0.027	**1.00**	**0.97**	**0.99**	**0.69**	0.30	0.29	0.72
SGED-EVT	-5.009	0.026	**1.00**	**0.97**	**1.00**	**0.69**	0.29	0.37	0.75
JSU-EVT	-5.022	0.027	**1.00**	**0.95**	**1.00**	**0.69**	0.30	0.30	0.72
AXA			**5% significance level**						
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.908	0.056	0.20	0.13	0.07	0.00	0.01	0.66	0.52
ST	-4.046	0.057	0.26	0.24	0.10	0.01	0.02	0.70	0.59
SKST	-4.141	0.056	0.29	0.12	0.07	0.07	0.13	0.61	0.56
SGED	-4.148	0.052	0.27	0.08	0.04	0.02	0.24	0.55	0.55
JSU	-4.164	0.056	0.29	0.17	0.08	0.09	0.17	0.59	0.56
N-EVT	-3.962	0.056	**1.00**	**0.96**	**1.00**	**0.57**	0.45	0.58	0.51
ST-EVT	-4.008	0.055	**1.00**	**0.95**	**0.98**	**0.56**	0.44	0.55	0.55
SKST-EVT	-3.994	0.056	**1.00**	**0.95**	**0.98**	**0.56**	0.45	0.55	0.55
SGED-EVT	-3.981	0.056	**1.00**	**0.96**	**0.99**	**0.56**	0.45	0.55	0.52
JSU-EVT	-3.992	0.056	**1.00**	**0.95**	**0.95**	**0.57**	0.45	0.55	0.55

Table 3.10: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for AXA. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

BP				**1% significance level**					
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.696	0.014	0.06	0.03	0.02	0.00	0.00	0.69	0.98
ST	-4.283	0.012	0.20	0.01	0.00	0.06	0.12	0.74	0.99
SKST	-4.372	0.012	0.22	0.00	0.01	0.11	0.20	0.76	0.99
SGED	-4.213	0.011	0.16	0.02	0.04	0.02	0.19	0.78	1.00
JSU	-4.384	0.012	0.24	0.02	0.01	0.10	0.24	0.77	0.99
N-EVT	-4.585	0.007	**0.64**	**0.99**	**1.00**	0.39	0.45	0.82	1.00
ST-EVT	-4.696	0.007	**0.55**	**0.98**	**0.99**	0.41	0.47	0.81	1.00
SKST-EVT	-4.698	0.007	**0.56**	**0.99**	**1.00**	0.41	0.47	0.81	1.00
SGED-EVT	-4.651	0.007	**0.58**	**0.98**	**1.00**	0.41	0.45	0.82	1.00
JSU-EVT	-4.689	0.007	**0.56**	**0.99**	**1.00**	0.41	0.46	0.81	1.00
BP				**2.5% significance level**					
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.243	0.031	0.18	0.07	0.03	0.00	0.05	1.00	0.89
ST	-3.550	0.029	0.29	0.13	0.04	0.10	0.23	0.83	0.92
SKST	-3.619	0.028	0.30	0.18	0.06	0.20	0.42	0.78	0.94
SGED	-3.577	0.025	0.26	0.07	0.25	0.09	0.44	0.77	0.94
JSU	-3.637	0.028	0.31	0.20	0.11	0.22	0.50	0.77	0.94
N-EVT	-3.631	0.024	**0.90**	**0.97**	**0.99**	**0.59**	0.25	0.88	0.94
ST-EVT	-3.704	0.025	**0.80**	**0.96**	**0.99**	0.58	0.27	0.69	0.95
SKST-EVT	-3.702	0.026	**0.81**	**0.95**	**1.00**	0.58	0.28	0.70	0.94
SGED-EVT	-3.671	0.025	**0.84**	**0.98**	**0.99**	0.58	0.27	0.80	0.95
JSU-EVT	-3.698	0.025	**0.81**	**0.99**	**1.00**	0.58	0.28	0.71	0.95
BP				**5% significance level**					
1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.862	0.044	0.17	0.11	0.49	0.00	0.24	0.69	0.66
ST	-3.010	0.047	0.26	0.12	0.39	0.13	0.28	0.79	0.60
SKST	-3.064	0.044	0.26	0.11	**0.63**	0.24	0.43	0.79	0.64
SGED	-3.068	0.043	0.27	0.18	**0.81**	0.18	0.40	0.75	0.68
JSU	-3.079	0.044	0.27	0.13	**0.74**	0.27	0.49	0.79	0.65
N-EVT	-2.940	0.049	**0.81**	**0.90**	**0.96**	**0.55**	0.43	0.74	0.64
ST-EVT	-2.996	0.048	**0.70**	**0.90**	**0.97**	0.53	0.44	0.82	0.63
SKST-EVT	-2.995	0.048	**0.71**	**0.90**	**0.99**	0.52	0.45	0.81	0.63
SGED-EVT	-2.972	0.048	**0.73**	**0.91**	**0.97**	0.53	0.45	0.79	0.64
JSU-EVT	-2.991	0.047	**0.70**	**0.93**	**0.98**	0.53	0.45	0.81	0.63

Table 3.11: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for BP. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano.The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

Figures 3.9 and 3.10 show insignificant autocorrelations from $\{\hat{H}_t(0.01)\}$ for IBM serie but some significant autocorrelations for SAN with both models, in fact the p-value of $C_{ES}(5)$ statistic with JSU-APARCH is equal to 0. In this case, the number of extreme losses and the average losses is not very large but these are highly correlated. In general, for IBM with both models, autocorrelations are close to zero for the first twelve lags, i.e. $Cov(H_t(\alpha), H_{t-j}(\alpha)) = 0$.

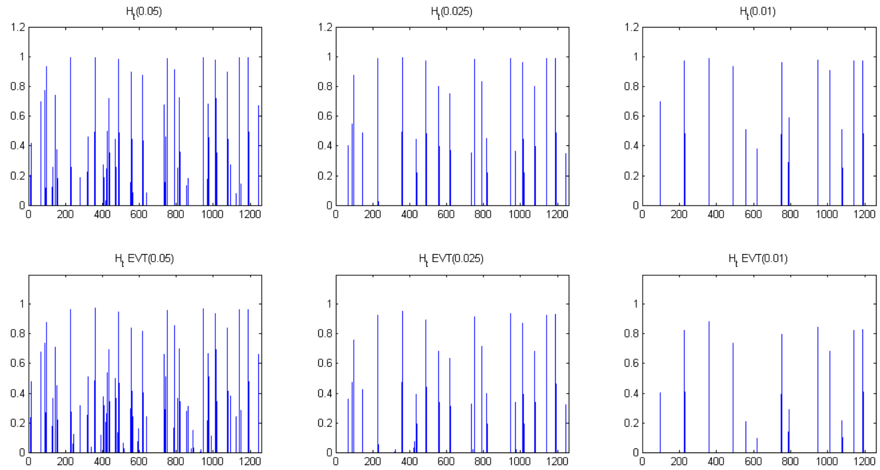


Figure 3.7: Cumulative hits (violations) of IBM with model JSU-APARCH and JSU-EVT-APARCH for different α.

Figure 3.11 shows the tail-distributions with their respective estimated parameters for IBM. We observe that the Normal distribution tends to underestimate the weight of the extreme returns contained in the distribution tails. The GPD is suitable to capture tail risk, and it avoids underestimating extreme risks.

Now, we calculate VaR, ES and SD estimates by FHS approach as described in subsection 3.3.6. Then, we evaluate the performance results of one-day out-of-sample ES forecasts using the test of Righi & Ceretta and the two tests of Acerbi & Szekely because these are more suitable for non-parametric VaR and ES forecasts. In FHS we obtain the VaR as the percentile of the return distribution and the ES as the mean of values below the VaR, we do not assume any volatility specification and/or probability distribution for the calculation of VaR and ES. The other tests, the Graham & Pál test and Costanzino & Curran and Du & Escanciano tests are suitable if we estimate VaR and ES by parametric approach.

Tables 3.12 - 3.15 show the average value of the estimated ES, the violations ratio (Viol.) of the underlying VaR and the backtesting results for the distinct models. If we compare these tables with Tables 3.8 - 3.11, we observe that (i) in what concerns parsi-

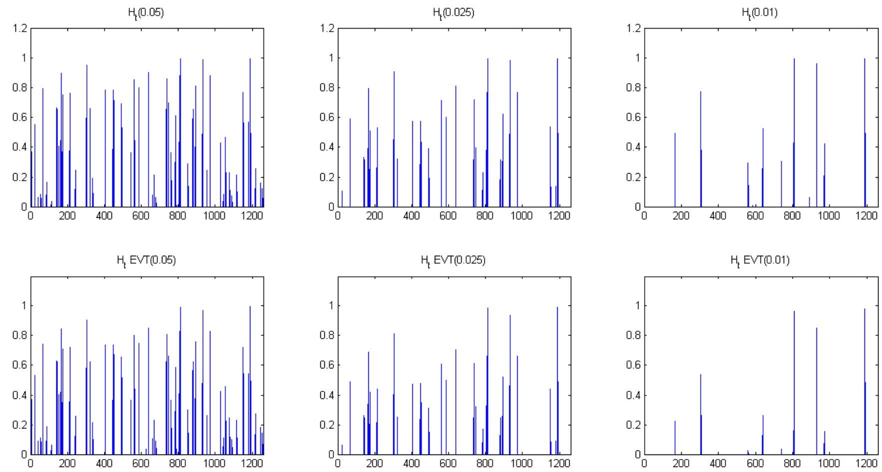


Figure 3.8: Cumulative hits (violations) of SAN with model JSU-APARCH and JSU-EVT-APARCH for different α.

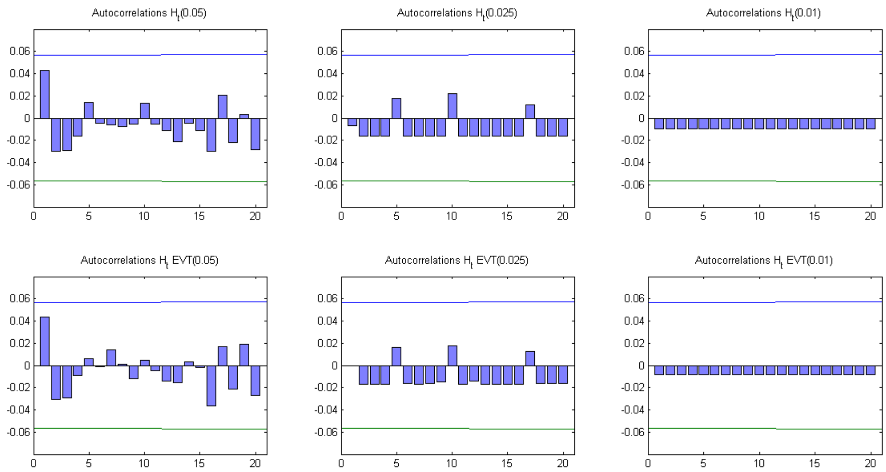


Figure 3.9: Sample autocorrelations of cumulative hits (violations) of IBM under model JSU-APARCH and JSU-EVT-APARCH for different values of α.

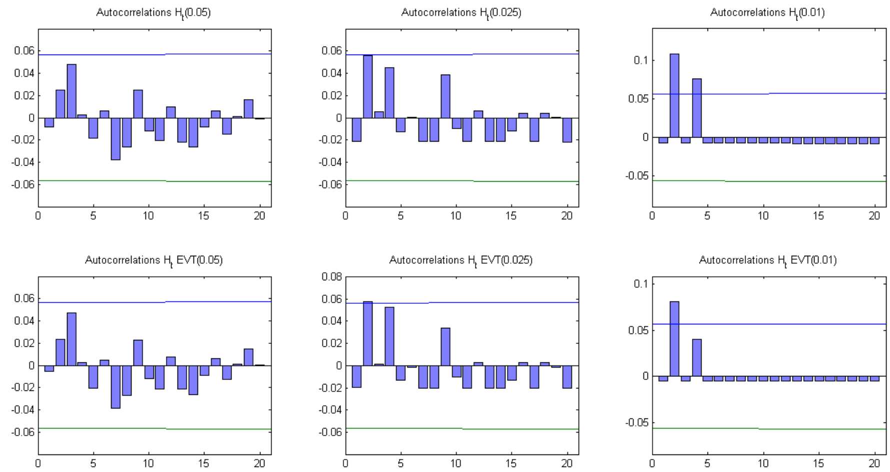


Figure 3.10: Sample autocorrelations of cumulative hits (violations) of SAN under model JSU-APARCH and JSU-EVT-APARCH for different values of α.

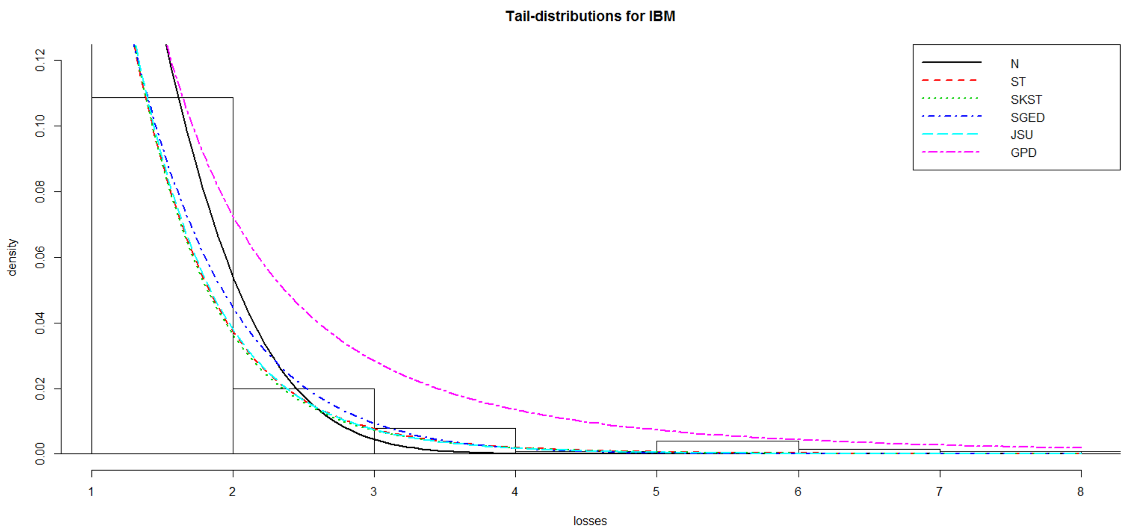


Figure 3.11: Tail-distributions for IBM. N is the Normal distribution, ST is the Student-t(4.67), SKST is the Skewed Student-t(0.97, 4.69), SGED is the Skewed Generalized Error(0.99, 1.15) and JSU is the Johnson SU (-0.092, 1.53) distribution.

mony of ES_t, the conditional EVT-based models do not always present "more negatives" values than do conditional models not based on EVT, as indicated by the averages for the ES estimates. Besides, the differences in average value of the ES are more similar among models than when we forecast ES by parametric approach (Tables 3.8 - 3.11) for the out-of-sample period (5 years, 1260 data), (ii) regarding VaR violation rates, in some models we observe an underestimation of risk, for example SAN at the 2.5% significance level and AXA at the 2.5% and 5% significance levels for all models, and unlike Tables 3.8 - 3.11, we do not obtain in all models based on EVT a lower violation rate than this obtained with the homologous model not based on EVT, and (iii) according to p-values of tests, models based on EVT, especially SKST, SGED and JSU are preferred in terms of ES backtesting for the three significance levels, although this model is possible that the risk is overestimated. Models not based on EVT are not suitable in terms of ES forecasts because we reject the null hypothesis of Z_1 and Z_2 test at 5% significance level for $ES_{1\%}$ and $ES_{2.5\%}$ calculated for IBM and SAN returns and we reject the null hypothesis of Z_2 test at significance level for $ES_{1\%}$ and $ES_{2.5\%}$ calculated for AXA returns. For BP, we obtain a good ES performance at 5% significance level with all models, based or not based on EVT, although ES is possibly overestimated regarding Z_2 test. But, the best models are N-EVT and SGED-EVT for this asset.

If we compare parametric and FHS approaches, we observe that the conclusions obtained on ES performance are similar, although the differences between conditional models based on EVT and not based on EVT are more significant with the parametric approach due to that FHS is a semi-parametric approach and the power and flexibility of conditional volatility models is diluted by historical simulation. The dilution depends on the number of realizations or paths generated by the standardized residuals. A considerable number of realizations are necessary in order to obtain robust results.

3.5.3 Pre-crisis and crisis periods

In this section we divide the full sample in two sub-samples. The pre-crisis period is defined so as to have the same number of observations as the crisis period (1239 data points). And we leave the 2007-2009 period for the out-of-sample crisis period. In Table 3.16, we show the time intervals for the pre-crisis and crisis periods for in-sample and out-of-sample evaluation.

Tables 3.17 - 3.20 and 3.21 - 3.24 show the average value of the estimated ES, the violations ratio of the underlying VaR and the backtesting results for the distinct models for the pre-crisis and crisis periods, respectively, using parametric approach.

For all assets and models, we observe in the pre-crisis period an average ES estimate lower than during the crisis and we also observe ratios of violations closer to the respective significance level (α) compared to those in crisis period. For instance, for IBM in the pre-crisis period, for 1% significance level the violations ratio of the different models is between $1.2\% - 1.5\%$, for 2.5% significance level the violations ratio is between $1.9\% - 2.7\%$, and for 5% significance level the violations ratio is between $3.3\% - 5\%$, while in the crisis period

IBM	**1% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-4.553	0.011	0.15	0.02	0.02
ST	-4.629	0.011	0.11	0.00	0.00
SKST	-4.627	0.010	0.10	0.02	0.02
SGED	-4.577	0.010	0.13	0.02	0.02
JSU	-4.619	0.011	0.15	0.00	0.12
N-EVT	-4.490	0.010	**1.00**	**0.90**	**0.97**
ST-EVT	-4.592	0.011	0.12	0.14	0.97
SKST-EVT	-4.575	0.011	0.12	0.10	0.97
SGED-EVT	-4.519	0.011	0.09	0.08	1.00
JSU-EVT	-4.573	0.011	0.12	0.00	0.97
IBM	**2.5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-3.474	0.021	0.19	0.02	0.29
ST	-3.486	0.021	0.14	0.03	0.26
SKST	-3.485	0.022	0.16	0.01	0.13
SGED	-3.471	0.021	0.19	0.03	0.30
JSU	-3.481	0.022	0.18	0.03	0.45
N-EVT	-3.471	0.021	**1.00**	**0.80**	**0.87**
ST-EVT	-3.492	0.021	0.14	0.37	0.94
SKST-EVT	-3.481	0.022	0.17	0.42	0.88
SGED-EVT	-3.466	0.021	0.14	0.31	0.93
JSU-EVT	-3.482	0.022	0.17	0.16	0.88
IBM	**5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-2.791	0.042	0.27	0.14	**0.68**
ST	-2.784	0.044	0.21	0.13	0.50
SKST	-2.782	0.043	0.22	0.05	0.43
SGED	-2.780	0.043	0.26	0.10	0.51
JSU	-2.781	0.044	0.24	0.12	0.45
N-EVT	-2.792	0.042	**1.00**	**0.75**	**0.75**
ST-EVT	-2.792	0.044	0.25	0.44	0.74
SKST-EVT	-2.784	0.045	0.26	0.42	0.80
SGED-EVT	-2.784	0.044	0.25	0.39	0.82
JSU-EVT	-2.786	0.044	0.25	0.52	0.79

Table 3.12: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for IBM and for 1-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

SAN	**1% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-6.147	0.013	0.12	0.01	0.00
ST	-6.159	0.013	0.08	0.02	0.01
SKST	-6.149	0.013	0.05	0.00	0.00
SGED	-6.144	0.013	0.06	0.01	0.00
JSU	-6.150	0.013	0.06	0.00	0.00
N-EVT	-6.203	0.014	0.14	0.49	0.93
ST-EVT	-6.206	0.014	0.14	0.60	0.89
SKST-EVT	-6.215	0.014	0.11	0.53	0.94
SGED-EVT	-6.214	0.015	0.13	0.58	0.94
JSU-EVT	-6.216	0.014	0.11	0.47	0.89
SAN	**2.5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-5.254	0.028	0.19	0.05	0.03
ST	-5.255	0.028	0.17	0.01	0.00
SKST	-5.250	0.026	0.13	0.04	0.03
SGED	-5.254	0.026	0.15	0.02	0.02
JSU	-5.251	0.026	0.14	0.03	0.03
N-EVT	-5.264	0.026	0.22	0.45	0.77
ST-EVT	-5.263	0.025	0.19	0.43	0.80
SKST-EVT	-5.266	0.027	0.21	0.52	0.82
SGED-EVT	-5.268	0.028	0.22	0.58	0.85
JSU-EVT	-5.267	0.027	0.21	0.49	0.86
SAN	**5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-4.512	0.052	0.28	0.17	0.16
ST	-4.507	0.051	0.23	0.09	0.08
SKST	-4.503	0.048	0.20	0.15	0.21
SGED	-4.508	0.048	0.23	0.06	0.17
JSU	-4.505	0.048	0.20	0.10	0.17
N-EVT	-4.506	0.049	0.35	0.38	0.71
ST-EVT	-4.503	0.049	0.31	0.55	0.81
SKST-EVT	-4.503	0.050	0.32	0.37	0.68
SGED-EVT	-4.507	0.049	0.31	0.39	0.77
JSU-EVT	-4.504	0.050	0.32	0.43	0.76

Table 3.13: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for SAN and for 1-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

AXA	**1% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-5.954	0.014	0.26	0.08	0.00
ST	-5.978	0.015	0.27	0.22	0.00
SKST	-5.986	0.017	0.31	0.24	0.00
SGED	-5.965	0.015	0.29	0.05	0.00
JSU	-5.985	0.017	0.34	0.29	0.00
N-EVT	-5.867	0.012	0.25	0.70	0.94
ST-EVT	-5.879	0.012	0.27	0.58	0.87
SKST-EVT	-5.900	0.010	0.19	0.73	0.94
SGED-EVT	-5.888	0.010	0.18	0.66	0.91
JSU-EVT	-5.900	0.010	0.19	0.63	0.93
AXA	**2.5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-4.889	0.034	0.32	0.13	0.03
ST	-4.883	0.034	0.29	0.07	0.00
SKST	-4.889	0.033	0.28	0.13	0.00
SGED	-4.887	0.034	0.32	0.16	0.03
JSU	-4.889	0.033	0.29	0.11	0.00
N-EVT	-4.923	0.034	0.50	0.59	0.80
ST-EVT	-4.919	0.034	0.48	0.61	0.82
SKST-EVT	-4.931	0.033	0.47	0.62	0.75
SGED-EVT	-4.929	0.033	0.46	0.58	0.80
JSU-EVT	-4.932	0.033	0.47	0.58	0.81
AXA	**5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-4.166	0.055	0.31	0.19	0.10
ST	-4.159	0.056	0.29	0.20	0.07
SKST	-4.162	0.057	0.29	0.21	0.10
SGED	-4.162	0.057	0.31	0.15	0.04
JSU	-4.162	0.057	0.29	0.18	0.07
N-EVT	-4.186	0.055	0.47	0.57	0.83
ST-EVT	-4.177	0.056	0.45	0.48	0.78
SKST-EVT	-4.184	0.056	0.47	0.42	0.76
SGED-EVT	-4.186	0.056	0.46	0.56	0.81
JSU-EVT	-4.185	0.057	0.48	0.44	0.74

Table 3.14: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for AXA and for 1-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

BP	**1% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-4.234	0.010	0.17	0.03	0.09
ST	-4.241	0.013	0.18	0.03	0.02
SKST	-4.254	0.012	0.20	0.00	0.00
SGED	-4.244	0.011	0.23	0.05	0.07
JSU	-4.253	0.012	0.22	0.02	0.00
N-EVT	-4.315	0.010	0.20	0.50	0.92
ST-EVT	-4.314	0.012	0.24	0.54	0.92
SKST-EVT	-4.316	0.011	0.18	0.58	0.92
SGED-EVT	-4.313	0.011	0.18	0.47	0.94
JSU-EVT	-4.315	0.011	0.19	0.54	0.90
BP	**2.5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-3.579	0.022	0.23	0.08	**0.66**
ST	-3.572	0.023	0.20	0.07	0.40
SKST	-3.583	0.027	0.27	0.13	0.04
SGED	-3.581	0.024	0.26	0.06	0.41
JSU	-3.583	0.026	0.26	0.14	0.14
N-EVT	-3.622	0.025	0.38	0.51	0.81
ST-EVT	-3.619	0.028	0.39	0.55	0.87
SKST-EVT	-3.619	0.027	0.37	0.61	0.86
SGED-EVT	-3.616	0.026	0.35	0.58	0.88
JSU-EVT	-4.315	0.027	0.37	0.52	0.78
BP	**5% significance level**				
FHS 1-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-3.067	0.044	0.29	0.21	**0.76**
ST	-3.062	0.044	0.24	0.17	**0.82**
SKST	-3.070	0.044	0.26	0.14	**0.61**
SGED	-3.068	0.045	0.29	0.12	**0.64**
JSU	-3.069	0.044	0.26	0.13	**0.73**
N-EVT	-3.079	0.044	0.40	0.58	**0.84**
ST-EVT	-3.076	0.044	0.35	0.50	**0.77**
SKST-EVT	-3.074	0.044	0.34	0.46	**0.76**
SGED-EVT	-3.072	0.043	0.33	0.53	**0.85**
JSU-EVT	-3.074	0.044	0.34	0.44	**0.74**

Table 3.15: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for BP and for 1-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

Daily	In-Sample			Out-of-Sample			
	Start	**End**	**T**	**Start**	**End**	**T**	**N**
pre-crisis	10/02/2000	06/30/2005	1239	07/01/2005	06/29/2007	521	1760
crisis	10/01/2002	06/29/2007	1239	07/02/2007	06/29/2009	521	1760
full sample	*10/02/2000*	*12/02/2011*	*2915*	*12/05/2011*	*09/30/2016*	*1260*	*4175*

Table 3.16: Sample information about the different considered periods.

those figures are between 1.2% − 1.7%, between 3.3% − 5%, and between 6.1% − 8.6%, respectively. We observe two facts: (i) As we increase the level of significance the differences between the violations ratio in pre-crisis and crisis periods increase, and (ii) in the pre-crisis period some models achieve a ratio of violations equal to the expected ratio (α): for BP data at $\alpha = 0.01$ significance, that is the case for N and ST models and for all models based on EVT except JSU-EVT; for the SGED-EVT model when analyzing IBM data at $\alpha = 0.025$, and for all models based on EVT for AXA data. At $\alpha = 0.05$ significance, that obtains for all models based on EVT for IBM data, for all models based on EVT except N-EVT for SAN data, for the SGED-EVT model for AXA data, and for the JSU-EVT model for BP data. In general, during the pre-crisis period, models based on EVT, especially with SGED and JSU distributions, perform the best in terms of the violations ratio.

If we compare the results obtained in the ES tests in pre-crisis and crisis periods, we observe (i) asymmetric distributions are preferred to symmetric distributions (with or without EVT) in pre-crisis and crisis period to obtain a better ES performance, (ii) among conditional models do not based on EVT are preferred SGED and JSU for the three significance levels both pre-crisis and crisis, except in pre-crisis for IBM at 1% significance level that SKST is preferred and in crisis for BP at 1% that ST is preferred, (iii) among conditional EVT models are also preferred SGED and JSU for the three significance levels and both pre-crisis and crisis, except for IBM that N-EVT model is preferred for the three significance levels, especially in crisis period, (iv) in general, in pre-crisis period the p-values obtained in all tests are higher than those obtained in crisis period, which implies a greater probability of assuming that the null hypothesis is not false, and (v) the differences between p-values obtained in pre-crisis and crisis period for ES based on the conditional EVT models are lower than those for ES do not based on conditional EVT models. In general, both pre-crisis and crisis, models based on EVT, especially with SGED and JSU are preferred in terms of ES backtesting.

If we compare the results obtained in each ES test individually, we observe: (i) according to the Righi & Ceretta test, the N-EVT and SGED-EVT models are suitable in terms of ES performance in both periods, and in pre-crisis JSU-EVT is also suitable, (ii) according to the first test of Acerbi & Szekely, that it is insensitive to an excessive number of exceptions, the N-EVT and SGED-EVT models play an important role in ES performance and it is possible to not reject H0 if there are a large number of exceptions of small magnitude; according to the second test of Acerbi & Szekely, where rejecting the H0 includes rejecting VaR as correctly specified, the SGED-EVT model has predominance in the ES performance, (iii) according to Graham & Pál test, JSU model performs the best at capturing tail risk as measured by ES and (iv) according to Du & Escanciano test, JSU and SKST-EVT models generally lead to not rejecting H0, implying that the cumulative violation process has mean equal to zero, with more probability than other EVT-based models; if we consider models not based on EVT, SGED and JSU are appropriate, because the cumulative violations do not display significant autocorrelation. Note that for AXA the cumulative violations have significant autocorrelation (in 5 lags) at 1% and 5% significance level in the pre-crisis but not in the crisis period and for IBM and BP, cumulative violations have significant autocorrelation for the three significance levels at both 1 and 5 lags in the crisis period but not in the pre-crisis period. In this sense, the number of extreme losses may not be large, but the average loss can be large and highly correlated. Therefore, this last test is able to better detect the problems of the commonly used risk models during the 2008 financial crisis. In addition, at a difference of conditional VaR backtests (Christoffersen, 1998) [22], we consider not only the clusters of tail events but also their magnitude. To sum up, we conclude that models with SGED and JSU, either based or not based on EVT, are preferred because they have more flexibility to capture the risk in periods of crisis and not crisis. We should also point out that the N-EVT model also yields good results, especially with the Rigui & Ceretta and Acerbi & Szekely tests.

IBM				**1% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.582	0.015	0.09	0.03	0.00	0.00	0.00	0.74	0.99
ST	-3.170	0.013	**0.35**	**0.57**	0.00	0.33	0.27	0.82	1.00
SKST	-3.224	0.012	0.31	0.37	0.02	0.35	0.28	0.82	1.00
SGED	-3.103	0.012	0.30	0.04	0.01	0.27	0.26	0.83	1.00
JSU	-3.218	0.012	0.34	0.34	0.02	0.36	0.31	0.82	1.00
N-EVT	-3.175	0.013	0.38	0.88	1.00	0.37	0.31	0.82	1.00
ST-EVT	-3.253	0.012	0.18	0.02	1.00	0.35	0.32	0.83	1.00
SKST-EVT	-3.253	0.012	0.18	0.03	1.00	0.35	0.32	0.83	1.00
SGED-EVT	-3.220	0.012	0.23	0.03	1.00	0.36	0.31	0.83	1.00
JSU-EVT	-3.254	0.012	0.19	0.02	0.99	0.35	0.31	0.83	1.00

IBM				**2.5% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.268	0.027	0.14	0.09	0.07	0.01	0.17	0.65	0.96
ST	-2.532	0.023	0.26	0.07	**0.53**	0.39	0.46	0.68	0.97
SKST	-2.573	0.019	0.22	0.07	**0.96**	0.40	0.48	0.69	0.98
SGED	-2.591	0.019	0.23	0.07	**0.96**	0.38	0.46	0.71	0.98
JSU	-2.595	0.019	0.23	0.07	**0.94**	0.43	0.48	0.70	0.98
N-EVT	-2.590	0.027	0.36	0.59	0.98	0.40	0.45	0.67	0.97
ST-EVT	-2.616	0.023	0.21	0.06	1.00	0.42	0.50	0.69	0.98
SKST-EVT	-2.618	0.023	0.22	0.08	0.99	0.42	0.50	0.69	0.98
SGED-EVT	-2.612	0.025	0.28	0.09	0.99	0.41	0.50	0.69	0.98
JSU-EVT	-2.623	0.023	0.22	0.06	1.00	0.42	0.50	0.69	0.98

IBM				**5% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.004	0.038	0.16	0.10	**0.90**	0.05	0.40	0.49	0.89
ST	-2.094	0.040	0.26	0.15	**0.85**	0.53	0.31	0.46	0.86
SKST	-2.125	0.040	0.27	0.21	**0.84**	0.56	0.29	0.47	0.87
SGED	-2.190	0.033	0.24	0.10	**0.92**	0.61	0.17	0.52	0.90
JSU	-2.152	0.040	0.28	0.29	**0.87**	0.61	0.23	0.49	0.88
N-EVT	-2.134	0.050	0.42	0.80	0.97	0.46	0.45	0.40	0.90
ST-EVT	-2.137	0.050	0.30	0.23	0.98	0.50	0.39	0.41	0.86
SKST-EVT	-2.140	0.050	0.31	0.35	0.98	0.50	0.39	0.41	0.86
SGED-EVT	-2.144	0.050	0.35	0.54	0.98	0.49	0.40	0.41	0.87
JSU-EVT	-2.145	0.050	0.32	0.34	0.96	0.50	0.39	0.41	0.86

Table 3.17: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for IBM and for pre-crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

SAN				**1% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.964	0.021	0.22	0.08	0.00	0.01	0.00	0.73	0.25
ST	-3.350	0.017	**0.40**	**0.78**	0.00	0.25	0.06	0.71	0.17
SKST	-3.515	0.017	**0.52**	**0.94**	0.00	0.52	0.30	0.76	0.21
SGED	-3.418	0.017	**0.50**	**0.95**	0.01	0.44	0.23	0.76	0.20
JSU	-3.581	0.017	**0.57**	**1.00**	0.00	**0.63**	0.45	0.77	0.24
N-EVT	-3.250	0.015	**0.78**	**0.97**	0.99	0.34	0.15	0.76	0.24
ST-EVT	-3.305	0.017	**0.64**	**0.99**	1.00	0.37	0.18	0.76	0.16
SKST-EVT	-3.270	0.017	**0.71**	**0.98**	1.00	0.36	0.17	0.75	0.19
SGED-EVT	-3.278	0.017	**0.71**	**1.00**	1.00	0.37	0.17	0.75	0.19
JSU-EVT	-3.266	0.017	**0.71**	**0.98**	1.00	0.36	0.17	0.75	0.18
SAN				**2.5% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.599	0.027	0.14	0.01	0.01	0.01	0.02	0.54	0.09
ST	-2.791	0.029	0.27	0.09	0.02	0.15	0.07	0.66	0.16
SKST	-2.926	0.027	0.32	0.19	0.13	0.33	0.20	0.78	0.24
SGED	-2.905	0.027	0.32	0.24	0.09	0.31	0.20	0.84	0.23
JSU	-2.976	0.027	0.35	0.36	0.13	0.42	0.26	0.84	0.26
N-EVT	-2.793	0.027	0.61	0.97	1.00	0.27	0.20	0.85	0.16
ST-EVT	-2.822	0.027	0.47	0.80	0.99	0.28	0.20	0.87	0.28
SKST-EVT	-2.806	0.027	0.54	0.92	0.99	0.27	0.19	0.86	0.29
SGED-EVT	-2.805	0.027	0.53	0.9	0.99	0.27	0.19	0.86	0.23
JSU-EVT	-2.804	0.027	0.54	0.97	1.00	0.27	0.19	0.87	0.28
SAN				**5% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.292	0.052	0.25	0.17	0.12	0.04	0.29	0.55	0.08
ST	-2.374	0.056	0.31	0.30	0.10	0.20	0.24	0.72	0.11
SKST	-2.484	0.050	0.34	0.36	0.46	0.45	0.48	0.61	0.11
SGED	-2.491	0.050	0.36	0.39	0.40	0.47	0.40	0.58	0.11
JSU	-2.519	0.050	0.36	0.43	0.54	**0.53**	0.42	0.61	0.11
N-EVT	-2.392	0.048	0.61	0.91	**0.99**	0.43	0.39	0.58	0.08
ST-EVT	-2.411	0.050	0.49	0.79	0.97	0.42	0.44	0.62	0.13
SKST-EVT	-2.403	0.050	0.56	0.89	1.00	0.42	0.42	0.60	0.12
SGED-EVT	-2.398	0.050	0.55	0.81	1.00	0.42	0.43	0.60	0.10
JSU-EVT	-2.401	0.050	0.56	0.84	0.99	0.42	0.42	0.61	0.12

Table 3.18: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for SAN and for pre-crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

AXA			**1% significance level**						
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.884	0.013	0.15	0.02	0.01	0.01	0.06	0.79	0.00
ST	-4.184	0.013	0.29	0.10	0.00	0.17	0.24	0.83	0.00
SKST	-4.286	0.012	0.26	0.04	0.00	0.23	0.34	0.85	0.00
SGED	-4.255	0.012	0.26	0.03	0.01	0.18	0.36	0.86	0.00
JSU	-4.328	0.010	0.23	0.03	0.19	0.26	0.39	0.86	0.00
N-EVT	-4.330	0.012	**1.00**	**0.97**	**0.99**	0.41	0.44	0.86	0.01
ST-EVT	-4.350	0.012	**1.00**	**0.99**	**0.99**	0.41	0.44	0.86	0.00
SKST-EVT	-4.340	0.012	**1.00**	**0.98**	**0.99**	0.41	0.44	0.86	0.01
SGED-EVT	-4.330	0.012	**1.00**	**0.98**	**1.00**	0.41	0.44	0.86	0.01
JSU-EVT	-4.334	0.012	**1.00**	**0.99**	**1.00**	0.40	0.44	0.86	0.01
AXA			**2.5% significance level**						
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.408	0.021	0.16	0.05	**0.63**	0.05	0.37	0.68	0.06
ST	-3.546	0.023	0.25	0.05	**0.51**	0.27	0.48	0.69	0.03
SKST	-3.629	0.021	0.26	0.06	**0.88**	0.36	0.43	0.71	0.02
SGED	-3.657	0.021	0.28	0.08	**0.89**	0.33	0.36	0.73	0.01
JSU	-3.664	0.021	0.28	0.09	**0.90**	0.40	0.38	0.72	0.01
N-EVT	-3.461	0.025	**0.97**	**0.94**	**0.97**	0.44	0.49	0.67	0.05
ST-EVT	-3.472	0.025	**0.93**	**0.99**	**1.00**	0.44	0.50	0.67	0.04
SKST-EVT	-3.468	0.025	**0.94**	**0.94**	**0.99**	0.44	0.50	0.67	0.04
SGED-EVT	-3.459	0.025	**0.95**	**0.93**	**0.99**	0.44	0.49	0.67	0.04
JSU-EVT	-3.464	0.025	**0.94**	**0.98**	**1.00**	0.44	0.50	0.67	0.04
AXA			**5% significance level**						
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.009	0.036	0.22	0.14	**0.92**	0.17	0.28	0.50	0.37
ST	-3.052	0.040	0.28	0.20	**0.87**	0.43	0.30	0.46	0.38
SKST	-3.121	0.038	0.30	0.17	**0.93**	0.54	0.22	0.49	0.31
SGED	-3.169	0.033	0.26	0.13	**0.98**	0.58	0.14	0.53	0.25
JSU	-3.147	0.038	0.31	0.20	**0.94**	0.59	0.19	0.50	0.28
N-EVT	-2.832	0.048	**0.95**	**0.89**	**0.94**	0.47	0.47	0.63	0.55
ST-EVT	-2.842	0.048	**0.90**	**0.88**	**0.95**	0.47	0.49	0.71	0.56
SKST-EVT	-2.839	0.048	**0.91**	**0.94**	**0.99**	0.47	0.48	0.68	0.55
SGED-EVT	-2.829	0.050	**0.94**	**0.92**	**0.99**	0.47	0.48	0.62	0.54
JSU-EVT	-2.835	0.048	**0.91**	**0.86**	**0.95**	0.47	0.48	0.67	0.55

Table 3.19: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for AXA and for pre-crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

BP				**1% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.361	0.010	0.36	0.40	0.46	0.41	0.42	0.73	1.00
ST	-3.621	0.010	**0.55**	**0.97**	**0.71**	**0.72**	0.30	0.89	1.00
SKST	-3.717	0.006	**0.46**	**0.86**	**0.99**	**0.79**	0.23	0.90	1.00
SGED	-3.668	0.006	**0.45**	**0.84**	**1.00**	**0.77**	0.24	0.90	1.00
JSU	-3.218	0.012	**0.51**	**0.93**	**1.00**	**0.83**	0.19	0.91	1.00
N-EVT	-3.155	0.010	**0.80**	**0.97**	**1.00**	0.37	0.39	0.68	1.00
ST-EVT	-3.144	0.010	**0.80**	**0.95**	**1.00**	0.37	0.40	0.76	1.00
SKST-EVT	-3.146	0.010	**0.82**	**0.98**	**1.00**	0.37	0.39	0.73	1.00
SGED-EVT	-3.150	0.010	**0.82**	**0.99**	**0.99**	0.38	0.39	0.72	1.00
JSU-EVT	-3.254	0.012	**0.83**	**0.97**	**1.00**	0.37	0.39	0.72	1.00
BP				**2.5% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.946	0.021	0.34	0.25	**0.92**	0.47	0.45	0.03	0.41
ST	-3.086	0.019	**0.39**	**0.56**	**0.97**	**0.73**	0.29	0.05	0.49
SKST	-3.167	0.019	**0.45**	**0.78**	**0.97**	**0.80**	0.22	0.06	0.54
SGED	-3.160	0.017	**0.42**	**0.72**	**0.95**	**0.81**	0.20	0.06	0.57
JSU	-2.595	0.019	**0.40**	**0.67**	**1.00**	**0.84**	0.18	0.07	0.58
N-EVT	-2.741	0.021	**0.85**	**0.93**	**1.00**	0.43	0.50	0.04	0.42
ST-EVT	-2.737	0.021	**0.84**	**0.96**	**0.98**	0.43	0.48	0.04	0.46
SKST-EVT	-2.738	0.021	**0.85**	**0.94**	**0.99**	0.43	0.50	0.04	0.46
SGED-EVT	-2.738	0.021	**0.85**	**0.97**	**0.99**	0.42	0.50	0.04	0.45
JSU-EVT	-2.623	0.023	**0.85**	**0.95**	**0.96**	0.42	0.50	0.04	0.46
BP				**5% significance level**					
pre-crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-2.598	0.044	0.38	0.48	**0.83**	0.56	0.34	0.20	0.58
ST	-2.667	0.046	**0.43**	**0.73**	**0.86**	**0.73**	0.30	0.22	0.60
SKST	-2.734	0.042	**0.45**	**0.80**	**0.91**	**0.82**	0.21	0.19	0.59
SGED	-2.743	0.040	**0.44**	**0.78**	**0.97**	**0.84**	0.16	0.16	0.58
JSU	-2.152	0.040	**0.47**	**0.79**	**0.91**	**0.86**	0.17	0.17	0.59
N-EVT	-2.379	0.048	**0.80**	**0.96**	**0.99**	0.46	0.47	0.27	0.61
ST-EVT	-2.375	0.048	**0.79**	**0.89**	**1.00**	0.47	0.45	0.26	0.62
SKST-EVT	-2.378	0.046	**0.79**	**0.96**	**1.00**	0.46	0.47	0.27	0.62
SGED-EVT	-2.378	0.048	**0.80**	**0.96**	**1.00**	0.46	0.47	0.27	0.62
JSU-EVT	-2.145	0.050	**0.79**	**0.89**	**0.98**	0.46	0.47	0.27	0.62

Table 3.20: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for BP and for pre-crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

IBM	**1% significance level**								
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-4.620	0.017	0.26	0.09	0.00	0.05	0.04	0.00	0.04
ST	-5.707	0.013	0.27	0.03	0.01	**0.62**	0.45	0.00	0.00
SKST	-5.664	0.015	0.27	0.04	0.00	**0.61**	0.47	0.00	0.00
SGED	-5.433	0.012	0.15	0.02	0.02	**0.56**	0.46	0.00	0.02
JSU	-5.591	0.013	0.22	0.01	0.00	**0.61**	0.45	0.00	0.00
N-EVT	-4.633	0.013	**1.00**	**1.00**	1.00	0.40	0.43	0.02	0.34
ST-EVT	-4.524	0.015	**0.91**	**1.00**	1.00	0.33	0.34	0.00	0.00
SKST-EVT	-4.519	0.015	**0.90**	**1.00**	1.00	0.33	0.34	0.00	0.00
SGED-EVT	-4.570	0.012	**0.74**	**0.97**	1.00	0.36	0.39	0.00	0.01
JSU-EVT	-4.511	0.015	**0.90**	**0.99**	1.00	0.33	0.35	0.00	0.00

IBM	**2.5% significance level**								
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-4.049	0.038	0.28	0.06	0.01	0.05	0.08	0.02	0.24
ST	-4.544	0.048	**0.37**	**0.61**	0.01	**0.56**	0.46	0.00	0.04
SKST	-4.512	0.050	0.36	**0.61**	0.00	**0.54**	0.43	0.00	0.04
SGED	-4.532	0.038	0.35	0.26	0.02	**0.56**	0.46	0.00	0.04
JSU	-4.508	0.048	**0.36**	0.49	0.02	**0.57**	0.48	0.00	0.03
N-EVT	-3.937	0.033	**1.00**	**0.97**	1.00	0.45	0.48	0.00	0.10
ST-EVT	-3.833	0.046	**0.99**	**0.99**	0.99	0.47	0.39	0.00	0.01
SKST-EVT	-3.830	0.046	**0.98**	**1.00**	1.00	0.47	0.39	0.00	0.01
SGED-EVT	-3.881	0.036	**0.96**	**0.98**	0.98	0.47	0.42	0.00	0.02
JSU-EVT	-3.825	0.046	**0.98**	**0.98**	0.99	0.47	0.39	0.00	0.01

IBM	**5% significance level**								
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.570	0.071	0.32	0.28	0.04	0.06	0.13	0.37	0.09
ST	-3.745	0.083	0.32	0.34	0.04	0.39	0.21	0.08	0.01
SKST	-3.721	0.086	0.32	0.31	0.02	0.37	0.20	0.09	0.01
SGED	-3.827	0.071	0.34	0.36	0.03	0.48	0.39	0.11	0.03
JSU	-3.735	0.083	0.32	0.35	0.03	0.41	0.25	0.08	0.01
N-EVT	-3.363	0.061	**0.99**	**0.97**	**0.98**	0.46	0.50	0.19	0.15
ST-EVT	-3.248	0.065	**0.90**	**0.99**	**0.99**	0.50	0.40	0.03	0.02
SKST-EVT	-3.245	0.069	**0.91**	**0.96**	0.98	0.51	0.40	0.03	0.02
SGED-EVT	-3.301	0.065	**0.94**	**0.97**	**0.99**	0.49	0.43	0.09	0.07
JSU-EVT	-3.243	0.067	**0.89**	**0.96**	0.98	0.51	0.40	0.03	0.02

Table 3.21: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for IBM and for crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

SAN				**1% significance level**					
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-6.654	0.019	0.15	0.01	0.00	0.01	0.02	0.75	0.70
ST	-8.090	0.013	**0.42**	**0.88**	0.01	**0.51**	0.47	0.86	1.00
SKST	-8.293	0.013	**0.47**	**0.96**	0.01	**0.58**	0.39	0.88	1.00
SGED	-7.829	0.015	**0.45**	**0.86**	0.00	0.44	0.46	0.87	1.00
JSU	-8.353	0.010	**0.40**	**0.72**	0.59	**0.58**	0.34	0.89	1.00
N-EVT	-7.128	0.012	**1.00**	**0.99**	**0.99**	0.40	0.48	0.87	1.00
ST-EVT	-7.018	0.013	**1.00**	**0.99**	**0.99**	0.40	0.48	0.87	1.00
SKST-EVT	-7.018	0.013	**1.00**	**0.99**	**0.99**	0.40	0.48	0.87	1.00
SGED-EVT	-6.918	0.010	**1.00**	**1.00**	**1.00**	0.40	0.47	0.87	1.00
JSU-EVT	-6.996	0.013	**1.00**	**0.99**	**0.99**	0.40	0.48	0.87	1.00
SAN				**2.5% significance level**					
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-5.831	0.050	0.30	0.12	0.02	0.01	0.04	0.52	0.10
ST	-6.610	0.040	**0.44**	**0.80**	0.00	0.42	0.36	0.61	0.32
SKST	-6.767	0.036	**0.45**	**0.84**	0.02	0.52	0.48	0.64	0.46
SGED	-6.596	0.035	**0.41**	**0.61**	0.00	0.46	0.49	0.66	0.58
JSU	-6.838	0.035	**0.45**	**0.85**	0.03	**0.58**	0.45	0.66	0.53
N-EVT	-5.969	0.025	**1.00**	**0.99**	**0.99**	0.48	0.41	0.69	0.40
ST-EVT	-5.904	0.025	**1.00**	**0.99**	**0.99**	0.48	0.41	0.69	0.62
SKST-EVT	-5.904	0.027	**1.00**	**1.00**	**1.00**	0.48	0.41	0.69	0.63
SGED-EVT	-5.830	0.025	**1.00**	**1.00**	**1.00**	0.49	0.40	0.69	0.69
JSU-EVT	-5.889	0.025	**1.00**	**0.99**	**1.00**	0.48	0.41	0.69	0.64
SAN				**5% significance level**					
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-5.139	0.086	0.32	0.27	0.01	0.01	0.01	0.33	0.01
ST	-5.545	0.088	**0.42**	**0.64**	0.02	0.15	0.02	0.33	0.02
SKST	-5.667	0.083	**0.43**	**0.84**	0.02	0.24	0.07	0.35	0.02
SGED	-5.616	0.081	**0.43**	**0.73**	0.01	0.26	0.12	0.40	0.02
JSU	-5.726	0.083	**0.45**	**0.79**	0.01	0.30	0.10	0.36	0.02
N-EVT	-5.106	0.058	**1.00**	**0.97**	**0.98**	0.51	0.44	0.37	0.03
ST-EVT	-5.063	0.060	**1.00**	**0.99**	**1.00**	0.52	0.42	0.38	0.05
SKST-EVT	-5.063	0.060	**1.00**	**0.97**	**0.98**	0.52	0.43	0.38	0.04
SGED-EVT	-4.989	0.060	**1.00**	**0.99**	**0.99**	0.54	0.40	0.39	0.06
JSU-EVT	-5.047	0.060	**1.00**	**0.97**	**0.99**	0.52	0.42	0.38	0.05

Table 3.22: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for SAN and for crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

AXA				**1% significance level**					
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-9.217	0.021	0.09	0.02	0.00	0.00	0.06	0.79	1.00
ST	-10.337	0.017	0.20	0.01	0.00	0.10	0.26	0.84	1.00
SKST	-10.372	0.015	0.18	0.03	0.00	0.09	0.25	0.84	1.00
SGED	-10.175	0.012	0.07	0.02	0.01	0.04	0.28	0.85	1.00
JSU	-10.412	0.013	0.16	0.02	0.00	0.09	0.27	0.84	0.88
N-EVT	-10.697	0.012	**1.00**	**0.99**	**1.00**	0.40	0.46	0.88	1.00
ST-EVT	-11.258	0.012	**1.00**	**1.00**	**1.00**	0.42	0.44	0.88	1.00
SKST-EVT	-11.276	0.012	**1.00**	**0.99**	**0.99**	0.42	0.44	0.88	1.00
SGED-EVT	-10.887	0.012	**1.00**	**1.00**	**1.00**	0.41	0.45	0.88	1.00
JSU-EVT	-11.201	0.012	**1.00**	**1.00**	**1.00**	0.42	0.44	1.00	1.00
AXA				**2.5% significance level**					
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-8.080	0.040	0.16	0.03	0.00	0.00	0.08	0.05	0.43
ST	-8.672	0.038	0.25	0.11	0.05	0.14	0.28	0.08	0.54
SKST	-8.701	0.038	0.27	0.15	0.03	0.12	0.25	0.06	0.49
SGED	-8.697	0.038	0.28	0.11	0.00	0.08	0.35	0.15	0.74
JSU	-8.750	0.038	0.29	0.17	0.02	0.13	0.29	0.08	0.40
N-EVT	-8.511	0.036	**1.00**	**1.00**	**1.00**	0.53	0.36	0.52	0.97
ST-EVT	-8.876	0.033	**1.00**	**0.98**	**0.98**	0.50	0.41	0.38	0.93
SKST-EVT	-8.890	0.033	**1.00**	**0.97**	**0.97**	0.50	0.41	0.37	0.93
SGED-EVT	-8.635	0.033	**1.00**	**0.98**	**0.99**	0.52	0.38	0.49	0.96
JSU-EVT	-8.840	0.033	**1.00**	**0.99**	**1.00**	0.51	0.40	0.57	0.94
AXA				**5% significance level**					
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-7.125	0.065	0.21	0.09	0.01	0.01	0.14	0.08	0.33
ST	-7.411	0.071	0.30	0.26	0.02	0.14	0.22	0.05	0.28
SKST	-7.435	0.069	0.30	0.26	0.02	0.12	0.20	0.05	0.28
SGED	-7.500	0.063	0.29	0.22	0.06	0.11	0.30	0.06	0.30
JSU	-7.478	0.069	0.31	0.28	0.06	0.13	0.24	0.05	0.03
N-EVT	-6.996	0.060	**1.00**	**0.96**	**0.98**	0.49	0.50	0.04	0.29
ST-EVT	-7.301	0.058	**1.00**	**0.96**	**0.98**	0.49	0.49	0.03	0.24
SKST-EVT	-7.314	0.058	**1.00**	**0.96**	**0.97**	0.49	0.49	0.03	0.24
SGED-EVT	-7.104	0.058	**1.00**	**0.96**	**0.96**	0.49	0.50	0.04	0.27
JSU-EVT	-7.273	0.058	**1.00**	**0.95**	**0.99**	0.49	0.49	0.28	0.25

Table 3.23: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for AXA and for crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags = 1$ and $lags = 5$) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

BP	**1% significance level**								
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-5.048	0.031	0.26	0.08	0.00	0.01	0.00	0.65	0.00
ST	-5.498	0.025	**0.38**	**0.57**	0.00	0.20	0.09	0.75	0.01
SKST	-5.419	0.025	0.34	0.37	0.00	0.17	0.06	0.74	0.00
SGED	-5.303	0.029	0.36	0.35	0.00	0.11	0.05	0.75	0.00
JSU	-5.439	0.025	0.37	0.43	0.00	0.18	0.07	0.75	0.00
N-EVT	-5.159	0.012	**1.00**	**1.00**	**1.00**	0.42	0.49	0.87	1.00
ST-EVT	-5.019	0.017	**1.00**	**1.00**	1.00	0.40	0.48	0.88	1.00
SKST-EVT	-5.013	0.017	**1.00**	**0.98**	0.99	0.41	0.48	0.88	1.00
SGED-EVT	-5.052	0.015	**1.00**	**1.00**	1.00	0.41	0.49	0.88	1.00
JSU-EVT	-5.022	0.017	**1.00**	**1.00**	1.00	0.40	0.48	0.88	1.00
BP	**2.5% significance level**								
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-4.425	0.048	0.22	0.07	0.01	0.00	0.00	0.45	0.01
ST	-4.656	0.046	0.30	0.24	0.01	0.07	0.01	0.46	0.00
SKST	-4.593	0.048	0.30	0.18	0.01	0.05	0.01	0.45	0.00
SGED	-4.569	0.050	0.30	0.28	0.01	0.04	0.01	0.48	0.00
JSU	-4.619	0.048	0.31	0.29	0.00	0.06	0.01	0.46	0.00
N-EVT	-4.418	0.031	**1.00**	**0.99**	**1.00**	0.40	0.39	0.62	0.00
ST-EVT	-4.330	0.031	**1.00**	**1.00**	**1.00**	0.40	0.41	0.63	0.00
SKST-EVT	-4.325	0.031	**1.00**	**0.99**	**0.99**	0.40	0.41	0.63	0.00
SGED-EVT	-4.346	0.031	**1.00**	**1.00**	**1.00**	0.40	0.40	0.63	0.00
JSU-EVT	-4.332	0.031	**1.00**	**1.00**	**1.00**	0.40	0.41	0.63	0.00
BP	**5% significance level**								
crisis	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-3.902	0.077	0.25	0.11	0.01	0.00	0.00	0.61	0.01
ST	-4.005	0.075	0.28	0.24	0.00	0.03	0.01	0.48	0.01
SKST	-3.955	0.077	0.28	0.14	0.00	0.02	0.01	0.50	0.01
SGED	-3.969	0.077	0.28	0.20	0.02	0.02	0.01	0.48	0.00
JSU	-3.980	0.075	0.28	0.16	0.02	0.03	0.01	0.49	0.01
N-EVT	-3.777	0.052	**0.99**	**0.98**	**1.00**	0.42	0.46	0.37	0.00
ST-EVT	-3.710	0.054	**0.98**	**0.98**	**1.00**	0.42	0.46	0.37	0.00
SKST-EVT	-3.706	0.054	**0.98**	**0.99**	**1.00**	0.42	0.45	0.36	0.00
SGED-EVT	-3.720	0.054	**0.99**	**0.98**	**1.00**	0.43	0.46	0.37	0.00
JSU-EVT	-3.712	0.054	**0.98**	**0.97**	**0.99**	0.42	0.46	0.37	0.00

Table 3.24: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for BP and for crisis period. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

3.6 Evaluating 10-day ES

Not much work has been done on implementation of extreme value theory for ES estimation beyond a one-day horizon, although risk horizons longer than one day are particularly important for risk liquidity management, for long term strategic asset allocation and for capital requirements, since the Basel Committee [13] obliges banks to compute their level of risk over a 10-day horizon. This motivates our analysis in this section on different methods to estimate ES for risk horizons longer than 1-day.

3.6.1 Scaling law

It is well-known that the variance of a Gaussian variable follows a simple scaling law. Indeed, the Basel Committee, in its 1996 Amendment (Basel II), states that it will accept a simple $\sqrt{h}$ scaling of 1-day VaR for deriving the 10-day VaR required in calculating market risk and the related risk capital, and the Basel Committee in 2016 (Basel III) proposes using the square root of time scaling assumption to calculate ES for risk horizons longer than one-day. However, the stylized facts on financial market volatility and research findings have repeatedly shown that the 10-day VaR is not likely to be the same as $\sqrt{10}\cdot$1-day VaR. First, the dynamics of a stationary volatility process suggests that if the current level of volatility is higher than unconditional volatility, the subsequent daily volatility forecasts will decline and convergence to unconditional volatility, and the contrary will happen when the initial volatility is lower than the unconditional one, with the rate of convergence being a function of the degree of volatility persistence. When initial volatility is higher than unconditional volatility, the scaling factor should be less than $\sqrt{10}$. In practice, due to volatility asymmetry and other predictive variables that might be included in the volatility model, it is always better to calculate $\hat{\sigma}^2_{t+1}$, $\hat{\sigma}^2_{t+2}$,..., $\hat{\sigma}^2_{t+10}$ separately. The 10-day VaR and 10-day ES is then produced using the 10-day volatility estimate computed from the sum $\sum_{i=1}^{10} \hat{\sigma}^2_{t+i}$.

Second, financial asset returns are not normally distributed. Danielsson and de Vries (1997)[30] show that the scaling parameter for quantiles derived using the EVT method increases at the approximate rate of h^{ξ}, which is typically less than the square-root-of-time adjustment. In stable distributions like GPD, the whole probability distribution, including the quantiles, and not just the standard deviation, scales as h^{ξ} [22]. The Feller convolution theorem explains this different scaling factor for VaR and ES estimates based on EVT [23].

In view of the conflicting empirical findings, one possible solution is to build models using 10-day return data. This again highlights the difficulty that arises from the inconsistency between the rule use to estimate VaR for calculating risk capital and the one applied to VaR estimation for backtesting.

[22] A random variable X has a stable distribution if the sum of n independent copies of X is a random variable that has the same type of distribution. Only the Normal, Cauchy and Lévy distributions are stable.

[23] If X and Y are both Pareto distributed (i.i.d.), then for a sufficiently large a one has $P(X+Y < a) \approx 2ca^{-1/\xi} = P(X < a) + P(Y < a)$ where $\hat{c} = \frac{T_u}{T} u^{1/\xi}$. This theorem established the additivity rule for Pareto-distributed random variables.

To sum up, scaling corresponds to an i.i.d. stable distribution assumption for returns and its use is incorrect for other distributions. Furthermore, this type of scaling is theoretically incorrect for volatility adjustment when we use a volatility model with mean reversion because it assumes that volatility remains constant or fluctuates around a local mean over the risk horizon and does not mean-revert at all. Obviously, under the scaling rule, the longer the risk horizon, the higher the error in VaR estimation.

3.6.2 Filtered Historical Simulation

An alternative for VaR and ES estimations at risk horizons longer than one-day is Filtered Historical Simulation (FHS). Barone-Adesi et al. (1998, 1999) [10] [11] extend the idea of volatility adjustment to multi-step historical simulation, using overlapping data in a way that does not create blunt tails for the h-day portfolio return distribution. Their idea is to use a parametric dynamic model of return volatility to simulate log returns on each day over the risk horizon [24].

Filtered Historical Simulation applies a statistical bootstrap on a parametric, dynamic model for return distributions. This filtering allows h-day return distributions to be generated from overlapping samples, since the bootstrap allows us to increasing the number of observations used for building the h-day return distribution. Whilst the standard historical approach is sometimes limited to the 1-day horizon because we simply do not have enough relevant historical data to use non-overlapping h-day returns. On the other hand, if we used overlapping h-day returns, that would distort the tail behavior of the return distributions, leading to significant error in VaR and ES estimates at extreme quantiles.

Tables 3.25 - 3.28 show the average value of the estimated ES, the violations ratio of the underlying VaR and the backtesting results for the distinct models for 10-day ES calculated by FHS. In this case, the entire out-of-sample is 1260 observations, but we have 1250 10-day ES which we can compare to the realized 10-day returns (we loose 10 observations).

If we compare these tables with Tables 3.8 - 3.11 and Tables 3.12 - 3.15 calculated for 1-day ES by parametric and FHS approaches, respectively, we observe: (i) as expected, the average ES estimates at 10-day horizon are greater than 1-day estimates, (ii) according to parsimony, i.e. ES series obtained with the different models used, which in our case are parsimonious models, the conditional EVT-based models present more negative values than the conditional models not based on EVT and at the 1% significance level differences in parsimony are larger than at the 5% signification but not as large as those presented for 1-day ES, (iii) regarding VaR violation rates, all models have a number of violations closer to the theoretical number even lower than this, (iv) if we focus on the conditional models not based on EVT, we observe that models with asymmetric distributions are better for 10-day ES estimation under all tests, except for $ES_{1\%}$ calculated for IBM according to Z_2 test, (v) if we focus on the conditional models based on EVT, we observe that the

[24] We provide a description about this approach in subsection 3.3.6

differences in performance between models that differ in the probability distribution for returns are not significant; in some cases as with Rigui & Ceretta and Acerbi & Szekely tests we obtain p-values close to 1; and (vi) for AXA and BP the differences between models, either based on or not on EVT, are not so huge. In this sense, these all models are suitable for 10-day ES estimation although it is possible a risk overestimation.

3.6.3 10-day historical returns

Under the historical return approach, 10-day returns are obtained as the sum of daily logarithmic returns over non-overlapping 10-day time intervals. The logarithmic transformation allows us to obtain continuously compound returns by sums and we can estimate ES over 10-day horizon without multiplying by a scaling factor.

The drawback of this method is that we might not have enough data to obtain precise parameter estimates and to perform ES backtesting. This is specially a problem under the EVT approach.

In our analysis, if we maintained the out-of-sample period (12/05/2011-09/30/2016) we would lack enough data for backtesting the 10-day ES, since we would only have 64 non-overlapping 10-day returns. To solve this, we enlarge the out-of-sample period from 10/03/2008 to 09/30/2016 (105 data) even though that will shorten the in-sample period. Therefore, the implied results will not be comparable with those of Tables 3.25 - 3.28.

In Tables 3.29 - 3.32, we observe that (i) in what concerns parsimony, conditional EVT-based models produce less negative forecasted ES values than the conditional models not based on EVT as indicated by the average ES estimates, (ii) violation rates are not so good in some cases because many models underestimate risk, with their VaR estimates producing more excesses than expected, especially at 2.5% significance level; notice that some models produce no violations at 1% significance level, (iii) conditional models based on EVT are better suited for risk measurement than those not based on EVT in terms of their ES estimates, (iv) ES models based on EVT with asymmetric distributions are better suited for risk measurement in terms of ES estimates, except at 1% significance level, where N-EVT is also a good performing model and (v) according to Costanzino & Curran [26] and Du & Escanciano [36] tests, the cumulative violations have autocorrelation with all models for IBM (at 2.5% and 5% significance levels); in this sense none of the considered models is able to capture the volatility clusters that are prevalent in financial assets, neither in quantity nor in magnitude.

IBM	**1% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-14.157	0.016	**0.55**	**0.97**	0.01
ST	-13.595	0.021	**0.45**	**0.88**	0.00
SKST	-13.598	0.020	**0.45**	**0.94**	0.01
SGED	-13.831	0.011	**0.48**	**0.97**	0.00
JSU	-13.546	0.021	**0.46**	**0.96**	0.00
N-EVT	-14.364	0.016	**0.85**	**0.87**	0.94
ST-EVT	-13.907	0.018	**0.56**	**0.89**	0.98
SKST-EVT	-13.937	0.018	**0.52**	**0.85**	0.97
SGED-EVT	-14.059	0.018	**0.66**	**0.93**	0.97
JSU-EVT	-13.802	0.018	**0.56**	**0.91**	0.96
IBM	**2.5% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-11.334	0.026	0.36	0.27	0.14
ST	-10.978	0.027	0.22	0.09	0.05
SKST	-10.977	0.028	0.25	0.06	0.01
SGED	-11.134	0.017	0.31	0.10	0.06
JSU	-10.940	0.027	0.24	0.07	0.03
N-EVT	-11.402	0.025	0.39	0.68	0.88
ST-EVT	-11.098	0.028	0.30	0.54	0.85
SKST-EVT	-11.111	0.026	0.26	0.63	0.92
SGED-EVT	-11.211	0.026	0.29	0.60	0.88
JSU-EVT	-11.025	0.027	0.28	0.60	0.91
IBM	**5% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-9.335	0.047	0.36	0.35	0.64
ST	-9.093	0.046	0.23	0.16	0.49
SKST	-9.083	0.047	0.24	0.16	0.27
SGED	-9.205	0.030	0.31	0.24	0.45
JSU	-9.063	0.046	0.25	0.16	0.41
N-EVT	-9.139	0.046	0.56	0.52	0.80
ST-EVT	-9.142	0.046	0.29	0.50	0.80
SKST-EVT	-9.148	0.046	0.28	0.50	0.79
SGED-EVT	-9.227	0.047	0.32	0.45	0.80
JSU-EVT	-9.089	0.046	0.29	0.53	0.80

Table 3.25: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for IBM and for 10-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

SAN			**1% significance level**		
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-23.026	0.010	**0.40**	**0.73**	**0.53**
ST	-23.074	0.010	0.33	0.50	0.53
SKST	-22.838	0.010	0.32	0.32	0.44
SGED	-22.973	0.010	0.37	**0.61**	**0.56**
JSU	-22.897	0.010	0.35	0.56	**0.54**
N-EVT	-22.919	0.010	**0.69**	**0.73**	**0.88**
ST-EVT	-22.919	0.009	0.42	0.80	**0.94**
SKST-EVT	-22.689	0.009	0.41	0.70	**0.92**
SGED-EVT	-22.845	0.010	0.53	**0.71**	**0.92**
JSU-EVT	-22.757	0.009	0.50	0.80	**0.94**
SAN			**2.5% significance level**		
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-18.846	0.022	0.38	0.47	**0.88**
ST	-18.926	0.021	0.32	0.32	**0.92**
SKST	-18.722	0.022	0.32	0.24	**0.91**
SGED	-18.805	0.022	0.36	0.37	**0.92**
JSU	-18.767	0.022	0.33	0.38	**0.94**
N-EVT	-18.784	0.022	**0.63**	**0.46**	**0.75**
ST-EVT	-18.840	0.021	0.48	0.52	**0.83**
SKST-EVT	-18.650	0.021	0.47	0.56	**0.82**
SGED-EVT	-18.734	0.022	0.53	0.63	**0.87**
JSU-EVT	-18.696	0.021	0.50	0.57	**0.80**
SAN			**5% significance level**		
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-15.757	0.046	**0.40**	**0.65**	**0.77**
ST	-15.847	0.045	0.36	**0.55**	**0.86**
SKST	-15.667	0.046	**0.36**	**0.57**	**0.81**
SGED	-15.720	0.046	0.38	**0.55**	**0.88**
JSU	-15.702	0.045	0.35	0.51	**0.91**
N-EVT	-15.715	0.047	**0.66**	**0.49**	**0.76**
ST-EVT	-15.794	0.045	0.55	**0.49**	**0.69**
SKST-EVT	-15.627	0.046	**0.57**	**0.42**	**0.71**
SGED-EVT	-15.676	0.045	0.55	**0.54**	**0.78**
JSU-EVT	-15.656	0.046	0.58	**0.55**	**0.82**

Table 3.26: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for SAN and for 10-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

AXA	**1% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-21.784	0.006	**0.43**	**0.70**	**1.00**
ST	-21.546	0.006	**0.41**	**0.64**	**0.99**
SKST	-21.542	0.007	**0.36**	**0.70**	**1.00**
SGED	-21.669	0.006	**0.42**	**0.72**	**1.00**
JSU	-21.527	0.006	0.36	0.46	**0.99**
N-EVT	-21.757	0.006	**0.72**	**0.79**	**0.95**
ST-EVT	-21.503	0.006	**0.68**	**0.77**	**0.93**
SKST-EVT	-21.487	0.008	**0.84**	**0.83**	**0.95**
SGED-EVT	-21.643	0.006	**0.71**	**0.74**	**0.92**
JSU-EVT	-21.517	0.006	**0.64**	**0.77**	**0.94**
AXA	**2.5% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-18.015	0.015	**0.40**	**0.66**	**0.96**
ST	-17.868	0.016	**0.38**	**0.66**	**0.95**
SKST	-17.851	0.014	0.33	0.30	**1.00**
SGED	-17.941	0.015	**0.39**	**0.67**	**0.98**
JSU	-17.853	0.016	**0.39**	**0.74**	**0.95**
N-EVT	-18.005	0.016	**0.71**	**0.63**	**0.91**
ST-EVT	-17.839	0.015	**0.62**	**0.72**	**0.91**
SKST-EVT	-17.824	0.012	0.50	0.68	**0.95**
SGED-EVT	-17.932	0.017	**0.72**	**0.60**	**0.87**
JSU-EVT	-17.845	0.016	**0.66**	**0.76**	**0.88**
AXA	**5% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-15.177	0.034	**0.42**	**0.64**	**0.98**
ST	-15.085	0.035	**0.41**	**0.60**	**0.90**
SKST	-15.067	0.032	0.37	**0.61**	**0.98**
SGED	-15.128	0.034	**0.41**	**0.66**	**0.99**
JSU	-15.074	0.034	**0.40**	0.78	**0.95**
N-EVT	-15.174	0.030	**0.65**	**0.63**	**0.85**
ST-EVT	-15.069	0.034	**0.65**	**0.64**	**0.80**
SKST-EVT	-15.053	0.033	**0.64**	**0.55**	**0.80**
SGED-EVT	-15.127	0.031	**0.63**	**0.43**	**0.82**
JSU-EVT	-15.070	0.031	**0.61**	**0.45**	**0.77**

Table 3.27: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for AXA and for 10-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

BP	**1% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-14.368	0.005	**0.56**	**0.94**	**1.00**
ST	-14.579	0.004	**0.59**	**0.97**	**1.00**
SKST	-14.559	0.005	**0.64**	**0.95**	**1.00**
SGED	-14.522	0.005	**0.63**	**0.95**	**1.00**
JSU	-14.528	0.004	**0.57**	**0.99**	**1.00**
N-EVT	-14.531	0.006	**0.82**	**0.81**	**0.98**
ST-EVT	-14.590	0.006	**0.91**	**0.84**	**0.94**
SKST-EVT	-14.507	0.006	**0.88**	**0.82**	**0.97**
SGED-EVT	-14.543	0.006	**0.86**	**0.80**	**0.96**
JSU-EVT	-14.540	0.004	**0.73**	**0.85**	**0.96**
BP	**2.5% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-11.967	0.015	**0.53**	**0.91**	**0.98**
ST	-12.115	0.014	**0.56**	**0.97**	**0.99**
SKST	-12.095	0.014	**0.56**	**0.91**	**1.00**
SGED	-12.086	0.013	**0.50**	**0.92**	**0.99**
JSU	-12.080	0.015	**0.56**	**0.93**	**1.00**
N-EVT	-11.992	0.014	**0.72**	**0.58**	**0.92**
ST-EVT	-12.126	0.014	**0.71**	**0.63**	**0.90**
SKST-EVT	-12.070	0.014	**0.74**	**0.64**	**0.85**
SGED-EVT	-12.104	0.014	**0.69**	**0.63**	**0.88**
JSU-EVT	-12.090	0.014	**0.70**	**0.67**	**0.86**
BP	**5% significance level**				
FHS 10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2
N	-10.144	0.038	**0.50**	**0.81**	**0.95**
ST	-10.250	0.041	**0.56**	**0.94**	**0.97**
SKST	-10.233	0.039	**0.50**	**0.79**	**0.95**
SGED	-10.239	0.038	**0.50**	**0.88**	**0.99**
JSU	-10.225	0.042	**0.53**	**0.87**	**0.92**
N-EVT	-10.164	0.042	**0.73**	**0.45**	**0.77**
ST-EVT	-10.261	0.042	**0.71**	**0.61**	**0.86**
SKST-EVT	-10.219	0.042	**0.71**	**0.58**	**0.77**
SGED-EVT	-10.253	0.038	**0.68**	**0.60**	**0.83**
JSU-EVT	-10.234	0.040	**0.68**	**0.56**	**0.83**

Table 3.28: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for BP and for 10-day returns calculated by FHS. BT_T is the test of Righi & Ceretta and Z_1 and Z_2 are the tests of Acerbi & Szekely. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

IBM				**1% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-11.239	0.029	0.13	0.02	0.00	0.02	0.04	0.85	1.00
ST	-13.983	0.019	**0.35**	**0.49**	0.00	0.23	0.30	0.92	1.00
SKST	-15.061	0.010	0.20	0.02	0.16	0.28	0.34	0.92	1.00
SGED	-13.791	0.010	0.09	0.00	0.00	0.20	0.28	0.92	1.00
JSU	-15.099	0.010	0.21	0.03	0.10	0.28	0.34	0.92	1.00
N-EVT	-17.144	0.010	**1.00**	**1.00**	**1.00**	**0.44**	0.49	0.92	1.00
ST-EVT	-15.390	0.010	**1.00**	**1.00**	**1.00**	0.37	0.43	0.92	1.00
SKST-EVT	-16.246	0.010	**1.00**	**0.99**	**0.99**	**0.39**	0.44	0.92	1.00
SGED-EVT	-17.710	0.010	**1.00**	**0.99**	**0.99**	**0.42**	0.48	0.92	1.00
JSU-EVT	-16.495	0.010	**1.00**	**0.99**	**0.99**	**0.39**	0.45	0.92	1.00
IBM				**2.5% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-9.846	0.038	0.10	0.08	0.03	0.02	0.06	0.71	0.99
ST	-11.018	0.038	0.31	0.22	0.00	0.23	0.26	0.78	1.00
SKST	-11.859	0.029	0.30	0.16	0.30	0.31	0.38	0.82	1.00
SGED	-11.504	0.029	0.25	0.05	0.13	0.26	0.39	0.83	1.00
JSU	-12.019	0.029	0.31	0.08	0.32	0.34	0.43	0.83	1.00
N-EVT	-12.344	0.038	**1.00**	**0.95**	**0.95**	0.44	0.42	0.80	1.00
ST-EVT	-11.179	0.029	**1.00**	**0.99**	**0.99**	0.42	0.48	0.84	1.00
SKST-EVT	-11.676	0.029	**1.00**	**0.96**	**0.98**	0.43	0.47	0.83	1.00
SGED-EVT	-12.557	0.029	**1.00**	**0.97**	**0.98**	0.44	0.45	0.82	1.00
JSU-EVT	-11.820	0.029	**1.00**	**1.00**	**1.00**	0.43	0.46	0.83	1.00
IBM				**5% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-8.676	0.048	0.09	0.03	0.03	0.04	0.18	0.63	0.97
ST	-9.000	0.048	0.21	0.07	0.09	0.24	0.31	0.64	0.97
SKST	-9.670	0.048	0.26	0.14	0.14	0.33	0.40	0.67	0.98
SGED	-9.704	0.048	0.25	0.12	0.13	0.30	0.43	0.68	0.98
JSU	-9.834	0.048	0.28	0.13	0.17	0.36	0.44	0.68	0.98
N-EVT	-9.399	0.048	**1.00**	**0.90**	**0.95**	0.39	0.36	0.64	0.97
ST-EVT	-8.446	0.048	**1.00**	**0.91**	**0.95**	0.42	0.44	0.67	0.98
SKST-EVT	-8.806	0.048	**1.00**	**0.94**	**0.94**	0.41	0.43	0.66	0.98
SGED-EVT	-9.473	0.048	**0.99**	**0.97**	**0.97**	0.40	0.39	0.65	0.97
JSU-EVT	-8.910	0.048	**1.00**	**0.95**	**0.95**	0.41	0.42	0.66	0.98

Table 3.29: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for IBM and for 10-day returns calculated as the sum of 10 daily logarithmic returns not overlapped. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

SAN				**1% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-17.410	0.019	**0.43**	**0.74**	0.00	0.30	0.30	0.89	1.00
ST	-19.553	0.019	**0.70**	**0.98**	0.02	**0.48**	0.44	0.92	1.00
SKST	-22.571	0.000	-	-	-	-	-	-	-
SGED	-22.462	0.000	-	-	-	-	-	-	-
JSU	-23.318	0.000	-	-	-	-	-	-	-
N-EVT	-15.567	0.019	**1.00**	**0.99**	**0.99**	**0.43**	0.49	0.92	1.00
ST-EVT	-18.570	0.019	**1.00**	**0.99**	**0.99**	**0.50**	0.43	0.92	1.00
SKST-EVT	-14.438	0.019	**1.00**	**1.00**	**1.00**	**0.40**	0.46	0.92	1.00
SGED-EVT	-14.060	0.019	**1.00**	**1.00**	**1.00**	0.39	0.44	0.92	1.00
JSU-EVT	-14.877	0.010	**1.00**	**1.00**	**1.00**	**0.41**	0.46	0.92	1.00
SAN				**2.5% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-15.288	0.029	0.27	0.08	0.28	0.25	0.23	0.78	0.99
ST	-15.974	0.029	**0.36**	**0.57**	**0.54**	0.37	0.31	0.77	0.99
SKST	-18.497	0.019	**0.57**	**0.97**	**1.00**	0.70	0.29	0.87	1.00
SGED	-18.881	0.019	**0.65**	**0.99**	**1.00**	**0.74**	0.29	0.89	1.00
JSU	-19.062	0.019	**0.62**	**0.99**	**1.00**	**0.71**	0.27	0.88	1.00
N-EVT	-12.752	0.029	**1.00**	**0.99**	**0.99**	0.33	0.28	0.78	0.99
ST-EVT	-14.660	0.029	**1.00**	**0.95**	**0.99**	0.36	0.29	0.77	0.99
SKST-EVT	-11.940	0.029	**1.00**	**0.99**	**0.99**	0.33	0.30	0.78	1.00
SGED-EVT	-11.699	0.029	**1.00**	**0.99**	**0.99**	0.33	0.29	0.78	1.00
JSU-EVT	-12.246	0.029	**1.00**	**0.98**	**0.98**	0.34	0.30	0.78	1.00
SAN				**5% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-13.506	0.029	0.14	0.07	0.96	0.37	0.46	0.76	0.99
ST	-13.400	0.048	0.34	0.32	0.40	0.45	0.49	0.72	0.99
SKST	-15.497	0.029	0.37	0.41	**1.00**	0.71	0.29	0.77	0.99
SGED	-16.024	0.029	**0.42**	**0.86**	**0.99**	**0.75**	0.25	0.77	0.99
JSU	-15.894	0.029	**0.40**	**0.84**	**1.00**	**0.72**	0.28	0.77	0.99
N-EVT	-10.556	0.048	**1.00**	**0.93**	**0.95**	0.42	0.48	0.74	0.99
ST-EVT	-12.098	0.048	**1.00**	**0.93**	**0.93**	0.43	0.48	0.71	0.98
SKST-EVT	-9.912	0.038	**1.00**	**0.99**	**1.00**	0.43	0.46	0.75	0.99
SGED-EVT	-9.704	0.038	**1.00**	**0.97**	**0.97**	0.43	0.18	0.75	0.99
JSU-EVT	-10.152	0.038	**1.00**	**0.98**	**1.00**	0.43	0.46	0.75	0.99

Table 3.30: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for SAN and for 10-day returns calculated as the sum of 10 daily logarithmic returns not overlapped. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The hyphen means that there are no violations. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

AXA			**1% significance level**						
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-14.922	0.029	**0.57**	**0.88**	0.00	0.13	0.04	0.81	0.95
ST	-18.896	0.010	**0.61**	**0.98**	**0.89**	**0.52**	0.40	0.92	1.00
SKST	-21.686	0.000	-	-	-	-	-	-	-
SGED	-20.187	0.000	-	-	-	-	-	-	-
JSU	-21.777	0.000	-	-	-	-	-	-	-
N-EVT	-15.926	0.010	**1.00**	**1.00**	**1.00**	**0.62**	0.28	0.92	1.00
ST-EVT	-21.074	0.019	**1.00**	**0.98**	**0.98**	**0.49**	0.44	0.92	1.00
SKST-EVT	-21.874	0.019	**1.00**	**1.00**	**1.00**	**0.48**	0.44	0.92	1.00
SGED-EVT	-12.745	0.010	**1.00**	**0.99**	**0.99**	0.34	0.39	0.92	1.00
JSU-EVT	-18.746	0.019	**1.00**	**1.00**	**1.00**	**0.45**	0.48	0.92	1.00
AXA			**2.5% significance level**						
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-13.084	0.048	0.39	0.47	0.00	0.10	0.05	0.70	0.18
ST	-15.122	0.029	**0.42**	**0.84**	**0.74**	**0.46**	0.43	0.80	0.00
SKST	-17.456	0.019	**0.62**	**1.00**	**1.00**	**0.73**	0.26	0.88	0.19
SGED	-16.992	0.019	**0.61**	**0.99**	**1.00**	**0.72**	0.27	0.88	0.36
JSU	-17.684	0.019	**0.65**	**1.00**	**1.00**	**0.76**	0.23	0.89	0.48
N-EVT	-13.139	0.029	**1.00**	**1.00**	**1.00**	**0.47**	0.41	0.79	0.56
ST-EVT	-16.505	0.029	**1.00**	**0.97**	**0.97**	0.39	0.35	0.79	0.00
SKST-EVT	-17.095	0.029	**0.99**	**0.97**	**0.97**	0.40	0.36	0.79	0.00
SGED-EVT	-10.826	0.029	**1.00**	**0.98**	**0.99**	0.36	0.41	0.83	0.00
JSU-EVT	-14.950	0.029	**1.00**	**0.98**	**0.98**	0.38	0.36	0.80	0.00
AXA			**5% significance level**						
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-11.539	0.057	0.24	0.18	0.05	0.09	0.08	0.55	0.38
ST	-12.500	0.057	**0.38**	**0.66**	0.11	**0.45**	0.43	0.66	0.03
SKST	-14.443	0.038	**0.49**	**0.98**	**1.00**	**0.77**	0.23	0.79	0.00
SGED	-14.441	0.038	**0.48**	**0.95**	**0.98**	**0.76**	0.24	0.78	0.00
JSU	-14.673	0.038	**0.51**	**0.97**	**0.99**	**0.79**	0.22	0.79	0.00
N-EVT	-11.055	0.048	**1.00**	**0.95**	**0.97**	**0.46**	0.43	0.66	0.23
ST-EVT	-13.899	0.057	**0.99**	**0.97**	**0.99**	0.43	0.45	0.68	0.02
SKST-EVT	-14.431	0.057	**0.98**	**0.91**	**0.91**	0.43	0.44	0.69	0.02
SGED-EVT	-9.134	0.048	**1.00**	**0.98**	**0.99**	0.46	0.40	0.77	0.01
JSU-EVT	-12.625	0.057	**1.00**	**0.94**	**0.95**	0.43	0.48	0.70	0.02

Table 3.31: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for AXA and for 10-day returns calculated as the sum of 10 daily logarithmic returns not overlapped. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The hyphen means that there are no violations. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

BP				**1% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-11.594	0.010	**0.67**	**0.99**	**0.80**	**0.61**	0.28	0.92	1.00
ST	-12.068	0.010	**0.53**	**0.93**	**0.67**	**0.65**	0.23	0.92	0.92
SKST	-14.786	0.000	-	-	-	-	-	-	-
SGED	-14.408	0.000	-	-	-	-	-	-	-
JSU	-15.314	0.000	-	-	-	-	-	-	-
N-EVT	-7.773	0.010	**1.00**	**0.99**	**0.99**	0.28	0.34	0.92	1.00
ST-EVT	-8.448	0.010	**1.00**	**1.00**	**1.00**	0.28	0.34	1.00	1.00
SKST-EVT	-7.312	0.010	**1.00**	**0.99**	**0.99**	0.28	0.34	0.92	1.00
SGED-EVT	-6.860	0.010	**1.00**	**1.00**	**1.00**	0.25	0.32	0.92	1.00
JSU-EVT	-7.551	0.010	**1.00**	**1.00**	**1.00**	0.28	0.34	0.92	1.00

BP				**2.5% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-10.193	0.010	0.27	0.12	**0.99**	**0.70**	0.25	0.92	1.00
ST	-10.388	0.010	0.20	0.06	**1.00**	**0.73**	0.23	0.92	0.90
SKST	-12.501	0.010	**0.85**	**1.00**	**1.00**	**0.90**	0.11	0.92	1.00
SGED	-12.360	0.010	**0.87**	**0.99**	**1.00**	**0.91**	0.10	0.92	1.00
JSU	-12.795	0.010	**0.85**	**0.98**	**1.00**	**0.89**	0.12	0.92	1.00
N-EVT	-6.748	0.029	**1.00**	**0.97**	**0.97**	**0.44**	0.36	0.91	1.00
ST-EVT	-7.221	0.029	**1.00**	**0.97**	**0.99**	**0.43**	0.39	1.00	1.00
SKST-EVT	-6.410	0.029	**1.00**	**1.00**	**1.00**	0.40	0.44	0.88	1.00
SGED-EVT	-6.115	0.029	**1.00**	**1.00**	**1.00**	0.39	0.41	0.89	1.00
JSU-EVT	-6.589	0.029	**1.00**	**1.00**	**1.00**	0.41	0.44	0.88	1.00

BP				**5% significance level**					
10-day	$\overline{ES}$	Viol	BT_T	Z_1	Z_2	TR	U_{ES}	$C_{ES}(1)$	$C_{ES}(5)$
N	-9.017	0.029	0.39	0.52	**0.97**	**0.82**	0.15	0.84	1.00
ST	-9.046	0.029	0.35	0.48	**0.98**	**0.84**	0.14	0.85	0.77
SKST	-10.702	0.029	**0.69**	**1.00**	**1.00**	**0.95**	0.07	0.88	1.00
SGED	-10.677	0.029	**0.72**	**0.98**	**1.00**	**0.95**	0.07	0.88	1.00
JSU	-10.851	0.029	**0.69**	**0.96**	**0.99**	**0.94**	0.08	0.87	1.00
N-EVT	-5.795	0.029	**1.00**	**0.95**	**0.98**	**0.50**	0.37	0.91	0.99
ST-EVT	-6.204	0.038	**1.00**	**0.97**	**0.98**	**0.53**	0.33	1.00	0.99
SKST-EVT	-5.475	0.057	**1.00**	**0.95**	**0.96**	**0.48**	0.40	0.73	0.99
SGED-EVT	-5.244	0.038	**1.00**	**0.98**	**0.98**	**0.46**	0.42	0.70	0.98
JSU-EVT	-5.619	0.057	**1.00**	**0.97**	**0.98**	**0.48**	0.39	0.73	0.99

Table 3.32: Mean estimates, violations ratio and backtesting results (p-values) for ES estimates from all the models for BP and for 10-day returns calculated as the sum of 10 daily logarithmic returns not overlapped. BT_T is the test of Righi & Ceretta, Z_1 and Z_2 are the tests of Acerbi & Szekely, TR is the test of Graham & Pál, and U_{ES}, $C_{ES}(1)$ and $C_{ES}(5)$ are the unconditional and the conditional ($lags$ = 1 and $lags$ = 5) tests of Costanzino & Curran and Du & Escanciano. The hyphen means that there are no violations. The p-values in bold indicate that the statistics obtained in these tests have an opposite sign to that specified in the alternative hypothesis.

3.7 Conclusions

In spite of the substantial theoretical evidence documenting the superiority of ES over VaR as a measure of risk, financial institutions and regulators have only recently embraced ES as an alternative to VaR for financial risk management. One of the major obstacles in this transition has been the unavailability of simple tools for the evaluation of ES forecasts. While the Basel rules for VaR tests are based on counting the number of exceptions, assessing the adequacy of a ES model requires the consideration of the size of tail losses beyond the VaR boundary. Different approaches have been proposed in the literature for ES backtesting in the last few years but, to the best of our knowledge, this is the first extensive comparison of a variety of ES backtesting procedures.

We use daily market closing prices for 10/02/2000 to 09/30/2016 on IBM, Santander, AXA and BP, and we consider some flexible families of asymmetric distributions for asset returns that include more standard probability distributions as special cases. Normal and Student's t distributions are considered as a benchmark for comparison. Given the evidence in Garcia-Jorcano and Novales (2017) [46] we use an APARCH volatility specification for all assets. We start by exploring which probability distribution seems to be more appropriate to model asset returns in order to get good ES estimates. Following the standard risk management methodology, once we estimate the dynamics of returns and the parameters of the probability distribution for the innovations, we forecast returns and volatility and apply a parametric approach to estimate 1- and 10-day ahead VaR and ES. After that, we use a variety of tests recently proposed for ES model validation.

As the true temporal dependency of financial returns is a complex issue, the standard approach to risk management can be improved by considering a two-step procedure that applies Extreme Value Theory (EVT): First, filtering the returns through a more or less complex GARCH model and second, estimating an extreme value theory type of density for the tail of the distribution of return innovations, using their assumed *iid* structure. This two-step procedure was proposed by McNeil & Frey (2000)[75] and it leads to a significant improvement, since VaR and ES estimates then incorporate changes in expected returns and volatility over time. So, we substitute the Generalized Pareto Distribution for the type of probability distributions mentioned above for return innovations. As in the standard approach, we then forecast VaR and ES at different significance levels and 1- and 10-day horizons and compare the results with those obtained under the standard approach.

In standard conditional models fitted to the full distribution of return innovations we observe that the Skewed Generalized Error distribution and the Johnson SU distribution play an important role in capturing tail risk. This is because some stylized facts of financial returns, such as volatility clusters, heavy tails and asymmetry are collected suitably by these asymmetric distributions. When we apply EVT to return innovations by modeling the tail with a GPD we obtain good ES forecasts regardless of the probability distribution used for returns. So, it looks as if considering just the return innovations in the tail of the distribution is more important than discriminating among probability distributions when estimating ES. Besides, each combination of APARCH volatility and probability distribution under the EVT approach dominates the similar specification under the standard approach that fits the full distribution. Furthermore, the conditional EVT models are more accurate and reliable than standard conditional models for predicting losses beyond the VaR during both pre-crisis and crisis periods.

All the results mentioned, in terms of the preference for SGED and Johnson SU distributions, or for the dominance of conditional EVT models over more standard conditional models for VaR and ES estimation, are valid not only for a 1-day horizon but also over a 10-day horizon. Since the standard historical approach is often limited to the 1-day horizon because of the lack of enough historical data, Filtered Historical Simulation based on bootstrapping allows us to use overlapping samples. That way, we increase the number of available observations for building the 10-day return distribution and estimate the 10-day ES. This avoids the systematic underestimation of risk, increasing with the time horizon, that arises when we apply the scaling law.

The ES tests we consider focus on a possible underestimation of risk, except for Costanzino & Curran and Du & Escanciano tests which are two-tailed tests. We note, however, that in some cases backtesting does not reject the model specification because the sample evidence is against both the null and the alternative hypothesis. In other words, some ES models are not rejected because they overestimate risk. When using ES to build the institution's reserves to cover potential losses in times of crisis, the underestimation may be fatal, but overestimation will lead to inefficient use of capital. This is a relevant consideration that should be taken into account for ES model validation.

Bibliography

[1] Aas, K. and Haff, I.H., 2006. *The Generalized Hyperbolic Skew Student's t-Distribution.* Journal of Financial Econometrics, Vol.4, No. 2, pp. 275-309.

[2] Alexander, C. and Sheedy, E., 2008. *Developing a stress testing framework based on market risk models.* Journal of banking and Finance, Vol. 32, pp. 2220-2236.

[3] Artzner, P., Delbaen, F., Eber, J.M. and Heath, M., 1999. *Coherent measures of risks.* Mathematical Finance, Vol. 9, pp. 203-228.

[4] Acerbi, C. and Szekely, B., 2014. *Backtesting Expected Shortfall.* Publication of MSCI.

[5] Acerbi, C. and Tasche, D., 2002. *On the Coherence of expected Shortfall.* Journal of Banking & Finance, Vol 26, pp. 1487-1503.

[6] Ardia, D. and Hoogerheide, L. F., 2014. *GARCH models for daily stock returns: Impact of estimation frequency on Value-at-Risk and Expected Shortfall forecasts.* Economics Letters, Vol. 123, No. 2, pp. 187-190.

[7] Azzalini, A. and Capitanio, A., 2003. *Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 65, No. 2, pp. 367-389.

[8] Bali, T. G., Mo, H. and Tang, Y., 2008. *The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR.* Journal of Banking and Finance, Vol. 32, No. 2, pp. 269-282.

[9] Bali, T. and Theodossiou, P., 2007. *A conditional-SGT-VaR approach with alternative GARCH model.* Annals of Operations Research, Springer.

[10] Barone-Adesi, G., Bourgoin, F. and Giannopoulus, K., 1998. *Don't look back.* Risk, Vol. 11, pp. 100-103.

[11] Barone-Adesi, Giannopoulus, K. and Vosper, L., 1999. *VaR without correlations for portfolios of derivative securities.* Journal of Futures Markets, Vol. 19, pp. 583-602.

[12] Barone-Adesi, G., Giannopoulos, K. and Vosper, L., 2002. *Backtesting derivative portfolios with filtered historical simulation (FHS).* European Financial Management, Vol. 8, No. 1, pp. 31-58.

[13] Basel Committee on Banking Supervision, 2016. Standards: Minimum Capital requirements for market risk. Bank for International Settlements.

[14] Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E., 2014. *Generalized quantiles as risk measures*. Insurance: Mathematics and Economics, Vol. 54, pp. 41-48.

[15] Berkowitz, J., 2001. *Testing density forecasts, with applications to risk management*. Review of Financial Studies, Vol. 14, pp. 371-405.

[16] Berkowitz, J., Christoffersen, P.F. and Pelletier, D., 2011. *Evaluating Value-at-Risk models with desk-level data*. Management Science, Vol. 57, No. 12, pp. 2213-2227.

[17] Carver, L. , 2013. *Mooted var substitute cannot be backtested, says top quant*. Risk.net

[18] Chan, K. F. and Gray, P., 2006. *Using extreme value theory to measure value-at-risk for daily electricity spot prices*. International Journal of forecasting, Vol. 22, No.2, pp. 283-300.

[19] Chavez-Demoulin, V., Embrechts, P. and Sardy, S., 2014. *Extreme-quantile tracking for financial time series*. Journal of Econometrics, Vol. 181, No. 1, pp. 44-52.

[20] Chen, J. M., 2014. *Measuring market risk under the basel accords: VaR, stressed VaR, and Expected Shortfall*. Aestimatio, The IEB International Journal of Finance, Vol. 8, pp. 184-201.

[21] Cheng, W. H. and Hung, J. C., 2011. *Skewness and leptokurtosis in GARCH-typed VaR estimation of petroleum and metal asset returns*. Journal of Empirical Finance, Vol. 18, No. 1, pp. 160-173.

[22] Christoffersen, P.F., 1998. *Evaluating interval forecasts*. International Economic Review, Vol. 39, pp. 841-862

[23] Clift, S.S., Costanzino, N. and Curran, M, 2016. *Empirical Performance of Backtesting Methods for Expected Shortfall*. Available at SSRN 2618345

[24] Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Events*. Springer.

[25] Cont, R., Deguest, R. and Scandolo, G., 2010. *Robustness and sensitivity analysis of risk measurement procedures*. Quantitative Finance, Vol. 16, No. 6, pp. 593-291.

[26] Costanzino, N. and Curran, M., 2015. *Backtesting General Spectral Risk Measures with application to Expected Shortfall*. Available at SSRN 2514403

[27] Crnkovic, C. and Drachman, J., 1996. *Quality control*. Risk, Vol. 9, pp. 139-143.

[28] Daniels, H.E., 1987. *Tail Probability Approximations*. International Statistic Review, Vol. 55, pp. 37-48.

[29] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G., 2001. *Using a bootstrap method to choose the sample fraction in tail index estimation*. Journal of Multivariate analysis, Vol. 76, No. 2, pp. 226-248.

[30] Danielsson, J. and de Vries, C.G., 1997b. *Tail index and quantile estimation with very high frequency data.* Journal of Empirical Finance, Vol. 4, pp. 241-257.

[31] Danielsson, J. and De Vries, C. G., 2000. *Value-at-risk and extreme returns.* Annales d'Economie et de Statistique, pp. 239-270.

[32] Degiannakis, S., Floros, C. and Dent, P., 2013. *Forecasting value-at-risk and expected shortfall using fractionally integrated models of conditional volatility: International evidence.* International Review of Financial Analysis, Vol. 27, pp. 21-33.

[33] Diebold, F.X., Gunther, T.A. and Tay, A.S., 1998. *Evaluating density forecasts with applications to financial risk management.* International Economic Review, Vol. 39, No. 4, pp. 863-883.

[34] Diebold, F.X., Schuermann, T. and Stroughair, J.D., 2000. *Pitfalls and opportunities in the use of extreme value theory in risk management.* The Journal of Risk Finance, Vol. 50, pp. 264-272.

[35] Ding, Z., Granger, C.W.J. and Engle, R.F., 1993. *A long memory property of stock market returns and a new model.* Journal of Empirical Finance, Vol.1, pp. 83-106.

[36] Du, Z. and Escanciano, J.C., 2015. *Backtesting Expected Shortfall: Accounting for Tail Risk.* Available at SSRN 2548544.

[37] Embrechts, P., Lambrigger, D. and Straumann, D., 2002. *Correlation and dependence in risk management: properties and pitfalls.* In M. Dempster and A. Howarth, editors, Risk Management: Value-at-Risk and beyond, Cambridge University Press, pp. 176-223.

[38] Embrechts, P, Lambrigger, D. and Wüthrich, M., 2009. *Multivariate extremes and the aggregation of dependence risks: examples and counter-examples.* Extremes, Vol. 12, No. 2, pp. 107-127.

[39] Emmer, S., Kratz, M. and Tasche, D., 2015. *What is the best risk measure in practice? A comparison of standard measures.* Journal of Risk, Vol. 18, No. 2.

[40] Engle, R., 2004. *Risk and volatility: Econometric models and financial practice.* The American Economic Review, Vol. 94, No.3, pp. 405-420.

[41] Ergen, I., 2015. *Two-step methods in VaR prediction and the importance of fat tails.* Quantitative Finance, Vol. 15, No. 6, pp. 1013-1030.

[42] Ergün, A., Jun, J., 2010. *Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall.* Quarterly Review of Economics and Finance, Vol. 50, pp. 264-272.

[43] Fernandez, C. and Steel, M., 1998. *On Bayesian Modelling of Fat Tails and Skewness.* Journal of the American Statistical Association, Vol. 93, No. 441, pp. 359-371.

[44] Ferreira, A., de Haan, L. and Peng, L., 2003. *On optimising the estimation of high quantiles of a probability distribution.* Statistics, Vol. 37, No. 5, pp. 401-434.

[45] Fissler, T., Ziegel, J.F. and Gneiting, T., 2015. *Expected Shortfall is jointly elicitable with value at risk-implications for backtesting.* arXiv preprint arXiv:1507.00244.

[46] Garcia-Jorcano, L. and Novales, A., 2017. *Volatility specifications versus probability distributions in VaR estimation.* Manuscript.

[47] Gençay, R. and Selçuk, F., 2004. *Extreme value theory and Value-at-Risk: Relative performance in emerging markets.* International Journal of Forecasting, Vol. 20, No.2, pp. 287-303.

[48] Giot, P. and Laurent, S., 2003a. *Value-at-Risk for Long and Short Trading Positions.* Journal of Applied Econometrics, Vol. 18, pp. 641-664.

[49] Gneiting T., 2011 *Making and evaluation point forecasts.* Journal of the American Statistical Association, Vol. 106, pp. 746-762.

[50] Gneiting, T. and Katzfuss, M., 2014. *Probabilistic forecasting.* Annual Review of Statistics and its Applications, Vol. 1, pp. 125-151.

[51] Gomes, M.I., de Haan, L. and Rodrigues, L.H., 2008. *Tail index estimation for heavy-tailed models: Accomodation of bias in weighted log-excesses.* Journal of the Royal Statistical Society, Series B, Vol. 70, No. 1, pp. 31-52.

[52] Gomes, M.I., Figueiredo, F., Rodrigues, L.H. and Miranda, M.C., 2012. *A computational sturdy of a quasi-PORT methodology for VaR based on second-order reduced-bias estimation.* Journal of Statistical Computation and Simulation, Vol. 82, No. 4, pp. 587-602.

[53] Gomes, M.I., Matins, M.J. and Neves, M.M., 2007. *Improving second order reduced-bias tail index estimator.* Review of Statistics, Vol. 5, No. 2, pp. 177-207.

[54] Gomes, M.I. and Pestana. D., 2007. *A sturdy reduced bias extreme quantile (VaR) estimator.* Journal of the american Statistical Association, Vol. 102, pp. 477, pp. 280-292.

[55] Gonzalo, J. and Olmo, J., 2004. *Which extreme values are really extreme?.* Journal of Financial Econometrics, Vol. 2, No. 3, pp. 349-369.

[56] Gourieroux, C. and Jasiak, J., 2010a. *Value at Risk.* In Y. Aït-Sahalia, & L.P. Hansen (Eds.), Handbook of Financial Econometrics. Amsterdam; North-Holland.

[57] Graham, A. and Pál, J., 2014. *Backtesting value-at-risk tail losses on a dynamic portfolio.* Journal of Risk Model Validation, Vol. 8, No. 2, pp. 59-96.

[58] Halbleib, R. and Pohlmeier, W., 2012. *Improving the value at risk forecasts: Theory and evidence from the financial crisis.* Journal of Economic Dynamics and Control, Vol. 36, No. 8, pp. 1212-1228.

[59] Hansen, B., 1994. *Autorregressive conditional density estimation.* International Economic Review, Vol. 35, pp. 705-730.

[60] Hill, B.M., 1975. *A simple general approach to inference about the tail of a distribution.* The Annals of Statistic, Vol. 3, No. 5, pp. 1163-1174.

[61] Huisman, R., Koedijk, K. G., Kool, C. J. M. and Palm, F., 2001. *Tail-index estimates in small samples.* Journal of Business and Economic Statistics, Vol. 19, No.2, pp. 208-216.

[62] Jalal, A. and Rockinger, M., 2008. *Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data.* Journal of Empirical Finance, Vol. 15, pp. 868-877.

[63] Johnson, N.L., 1949. *Systems of frequency curves generated by methods of translations.* Biometrika, Vol. 36, pp. 149-176.

[64] Jondeau, E., Poon, S.H. and Rockinger, M., 2000. *Financial Modeling Under Non-Gaussian Distributions.* Springer.

[65] Kerkhof, J. and Melenberg, B., 1995. *Backtesting for risk-based regulatory capital.* Journal of Banking and Finance, Vol. 28, pp. 1845-1865.

[66] Kourouma, L., Dupre, D., Sanfilippo, G. and Taramasco, O., 2010. *Extreme value at risk and expected shortfall during financial crisis.* Available at SSRN 1744091.

[67] Kuester, K., Mittnik, S. and Paolella, M.S., 2006. *Value-at-Risk prediction: A comparison of alternative strategies.* Journal of Financial Econometrics, Vol. 4, pp. 53-89.

[68] Lambert, N.S., Pennock, D.M. and Shoham, Y. , 2008. *Eliciting properties of probability distributions.* In proceedings of the 9th ACM Conference on Electronic Commerc, ACM, pp. 129-138.

[69] Li, D., Peng, L. and Yang, J., 2010. *Bias reduction for high quantiles.* Journal of Statistical Planning and Inference, Vol. 140, No. 9, pp. 2433-2441.

[70] Longin, F., 2005. *The choice of the distribution of asset returns: How extreme value theory can help?* Journal of Banking and Finance, Vol. 29, pp. 1017-1035.

[71] Louzis, D. P., Xanthopoulos-Sisinis, S. and Refenes, A. P., 2013. *The Role of High-Frequency Intra-daily Data, Daily Range and Implied Volatility in Multi-period Value-at-Risk Forecasting.* Journal of Forecasting, Vol. 32, No. 6, pp. 561-576.

[72] Lugannani, R. and Rice, S.O., 1980. *Saddlepoint approximation for the distribution of the sum of independent random variables.* Advanced Applied Probability, Vol. 12, pp. 475-490.

[73] Mabrouk, S. and Saadi, S., 2012. *Parametric value-at-risk analysis: Evidence from stock indices.* The Quartely Review of Economics and Fiannce, Vol. 52, pp. 305-321.

[74] Marinelli, C., D'Addona, S. and Rachev, S., 2007. *A comparison of some univariate models for value-at-risk and expected shortfall.* International Journal of Theoretical and Applied Finance, Vol. 10, pp. 1043-1075.

[75] McNeil, A. and Frey, R., 2000. *Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach*. Journal of Empirical Finance, Vol. 7, pp. 271-300.

[76] McNeil, A., Frey, R. and Embrechts, P., 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

[77] Osband, K.H., 1985. *Providing incentives for better cost forecasting*. PhD Thesis. University of California, Berkeley.

[78] Pascual, L., Romo, J. and Ruiz, E., 2006. *Bootstrap prediction for returns and volatilities in garch models*. Computational Statistics and Data Analysis, Vol. 50, No. 9, pp. 2293-2312.

[79] Pérignon, C. and Smith, D.R., 2010a. *Diversification and Value-at-Risk*. Journal of Banking and Finance, Vol. 34, pp. 55-66.

[80] Pickands, J., 1975. *Statistical inference using extreme order statistics*. The Annals of Statistics, Vol. 3, pp. 119-131.

[81] Pritsker, M., 2006. *The hidden dangers of historical simulation*. Journal of Banking and Finance, Vol. 30, No. 2, pp. 561-582.

[82] Righi, M.B. and Ceretta, P.S., 2015. *A comparison of Expected Shortfall estimation models*. Journal of Economics and Business, Vol. 78, pp. 14-47.

[83] Righi, M.B. and Ceretta, P.S., 2013. *Individual and Flexible Expected Shortfall Backtesting*. Journal of Risk Model Validation, Vol. 7, No. 3, pp. 3-20.

[84] Rocco, M., 2014. *Extreme value theory in finance: A survey*. Journal of Economic Surveys, Vol. 28, No. 1, pp. 82-108.

[85] Rosenblatt, M., 1952. *Remarks on a multivariate transformation*. Annals of Mathematical Statistics, Vol. 23, pp. 470-472.

[86] Ruiz, E. and Pascual, L., 2002. *Bootstrapping financial time series*. Journal of Economic Surveys, Vol. 16, No. 3, pp. 271-300.

[87] Sajjad, R., Coakley, J. and Nankervis, J. C., 2008. *Markov-switching garch modelling of value-at-risk*. Studies in Nonlinear Dynamics and Econometrics, Vol. 12, No.3, pp. 7.

[88] Smith, R., 1987. *Estimating tails of probability distributions*. The Annals of Statistics, Vol. 15, pp. 1174-1207.

[89] Stahl, G., Zheng, R., Kiesel, R. and Rühlicke, 2012. *Conceptualizing Robustness in Risk Management*. Available at SSRN 2065723.

[90] Theodossiou, P., 1998. *Financial data and skewed generalized t distribution.*. Management Science, Vol. 44, pp. 1650-1661.

[91] Tolikas, K., 2014. *Unexpected tails in risk measurement: Some international evidence.* Journal of Banking and Finance, Vol. 40, pp. 476-493.

[92] Venter, J.H. and Jongh, P.J. De., 2004. *Selecting and innovation distribution for GARCH models to improve efficiency of risk an volatility estimation.* Journal of Risks, Vol. 6, pp. 27-52.

[93] Wong, W.K., Fan, G. and Zeng, Y., 2012. *Capturing tail risks beyond VaR.* Review of Pacific Basin Financial Markets and Polices, Vol. 15, 1250015-1-1250015-25.

[94] Wong, W.K., 2010. *Backtesting value-at-risk based on tail losses.* Journal of Empirical Finance, Vol. 17, No. 3, pp. 526-538.

[95] Wong, W.K., 2008. *Backtesting trading risk of commercial banks using expected shortfall.* Journal of Banking and Finance, Vol. 3, pp. 1404-1415.

[96] Zhou, J., 2012.*Extreme risk measures for REITs: A comparison among alternative methods.* Applied Financial Economics, Vol. 22, pp. 113-126.

[97] Ziegel, J.F., 2016. *Coherence and elicitability.* Mathematical Finance, Vol. 26, No. 4, pp. 901-918.

Appendices

A Elicitability of VaR

We say that ψ is elicitable if it is the minimised value of some scoring function $s(x, y)$ according to

$$\psi = arg \min_x \mathbb{E}[s(x, Y)] \tag{21}$$

where Y is the distribution representing verified observations. The distribution can be empirical, parametric or simulated. Furthermore, as we saw, for elicitability to hold, the scoring function has to be strictly consitent.

We can show that $VaR_\alpha(Y)$ is elicitable through the scoring function

$$s(x, y) = (\mathbb{1}_{(x \geq y)} - \alpha)(x - y) \tag{22}$$

According to (21) this is true if we can show that

$$VaR_\alpha(Y) = arg \min_x \mathbb{E}[(\mathbb{1}_{(x \geq y)} - \alpha)(x - Y)]$$

Hence, if we minimise $\mathbb{E}[(\mathbb{1}_{(x \geq y)} - \alpha)(x - Y)]$ and show that get $VaR_\alpha(Y)$ as the minimizer, this proves that VaR is elicitable through its scoring function (22). We use $\mathbb{1}_{(x \geq y)} = \theta(x - y)$ where $\theta(x)$ is the Heaviside step function equal to one when $x \geq 0$ and zero otherwise. We can write (22) as

$$s(x, y) = (\theta(x - y) - \alpha)(x - y)$$

From this we get,

$$\mathbb{E}[s(x, Y)] = \mathbb{E}[(\theta(x - Y) - \alpha)(x - Y)]$$

We can write this as

$$\mathbb{E}[(\theta(x - Y) - \alpha)(x - Y)] = \int (\theta(x - y) - \alpha)(x - y) f_Y(y) dy$$

$$= (1 - \alpha) \int_{-\infty}^{x} (x - y) f_Y(y) dy - \alpha \int_{x}^{\infty} (x - y) f_Y(y) dy$$

We now want to take the first derivative of $\mathbb{E}[s(x, Y)]$, set it equal to 0 and solve for x. We want to calculate

$$\frac{d}{dx} \left((1 - \alpha) \int_{-\infty}^{x} (x - y) f_Y(y) dy - \alpha \int_{x}^{\infty} (x - y) f_Y(y) dy \right) \tag{23}$$

We take the derivative of the two terms in (23) independently. From the first term, by using Leibniz's rule, we get that

$$\frac{d}{dx} \left((1 - \alpha) \int_{-\infty}^{x} (x - y) f_Y(y) dy \right) = (1 - \alpha) \left(\int_{-\infty}^{x} f_Y(y) dy + (x - x) f_Y(x) - 0 f_Y(-\infty)(x + \infty) \right) =$$

$$= (1 - \alpha) \int_{\infty}^{x} f_Y(y) dy$$

Similarly for the second term, we get

$$\frac{d}{dx}\left(-\alpha\int_x^{\infty(x-y)} f_Y(y)dy\right) = -\alpha\int_x^{\infty} f_Y(y)dy$$

We can now add the two terms together and get

$$\frac{d}{dx}\mathbb{E}[s(x,y)] = (1-\alpha)\int_{-\infty}^{x} f_Y(y)dy - \alpha\int_x^{\infty} f_Y(y)dy =$$

$$= \int_{-\infty}^{x} f_Y(y)dy - \alpha$$

We set this equal to zero and find

$$\alpha = \int_{-\infty}^{x} f_Y(y)dy$$

$$x = F_Y^{-1}(\alpha)$$

which defines $VaR_\alpha(Y)$. Thus, we have prove that $VaR_\alpha(Y)$ is elicitable through its scoring function.