Saddle-point approach: backtesting VaR models in the presence of extreme losses

Ricardo João da Silva Gouveia

Dissertation submitted as partial requirement for the conferral of

Master in Finance

Supervisor:

Prof. António Manuel Barbosa, Prof. Auxiliar, ISCTE Business School, Departamento de Finanças

September 2018

Abstract

The Basel Committee for Banking Supervision requires every financial institution to carry out efficient Risk Management practices, so that these are able to face adverse days in the market and, thus, avoid another potential meltdown of the financial system, such as the 'Black Monday' in 1987 or the 'Subprime' crisis in 2007. To do so, traditional backtesting techniques assess the quality of commercial banks' risk forecasts based on the number of the exceedances. However, these backtests are not sensitive to the size of the exceedances, which could lead to inaccurate risk models to be accepted.

This way, this dissertation presents the Saddle-point backtest, a size-based procedure developed by Wong (2008) that evaluates risk models through the Tail-Risk-of-VaR.

This approach is believed to constitute a reliable size counterpart to the Basel II Agreements, hence deserving an important role in backtesting. However, the Saddle-point backtest shows some drawbacks regarding its application to non-parametric risk models, which is explored throughout this dissertation's empirical analysis.

Key Words: Backtesting, Value-at-Risk, Time Series, Risk Management

JEL Classification: C58, G32

Resumo

O Comité de Basileia para a Supervisão Bancária requer a todas as instituições financeiras que levem a cabo práticas de Gestão de Risco eficientes, de modo a que estas sejam capazes de enfrentar dias adversos no mercado e, desta forma, evitar outro eventual colapso do sistema financeiro, tal como a 'Segunda-feira Negra' em 1987 ou a crise do 'Subprime' em 2007. Para tal, as técnicas tradicionais de avaliação de modelos de risco aferem a qualidade das previsões dos bancos com base no número de excedências. No entanto, estes métodos não são sensíveis ao tamanho das excedências, o que pode levar a que modelos de risco pouco fiáveis sejam aceites.

Assim sendo, esta dissertação apresenta o teste de Saddle-point, um procedimento baseado no tamanho das excedências desenvolvido por Wong (2008), que avalia modelos de risco através do Risco-da-Cauda do Valor em Risco.

Crê-se que esta abordagem baseada no tamanho das excedências constitui uma fiável contraparte dos Acordos de Basileia II, merecendo, portanto, desempenhar um papel importante na avaliação de modelos de risco. No entanto, o teste de Saddle-point apresenta algumas falhas no que toca à sua aplicação a modelos de risco não paramétricos, algo que é explorado no decorrer da análise empírica desta dissertação.

Palavras-chave: Backtesting, Valor em Risco, Séries Temporais, Gestão de Risco

Classificação JEL: C58, G32

Acknowledgements

I would like to thank my coordinator Prof. António Manuel Barbosa, for his guidance and support throughout this entire project.

In addition, I would like to express my gratitude to both my parents, Claudia Gouveia and João Gouveia, for they have sponsored my entire academic career so far.

Table of Contents

Index of Tables

Index of Figures

List of Abbreviations

Basel Accords II – Second Amendment of the Basel Accords, in 1996

BCBS – Basel Committee for Banking Supervision

CDF – Cumulative Distribution Function

CC – Conditional Coverage

CR – Capital Requirements

EWMA – Exponentially Weighted Moving Average

EGARCH – Exponential General Autoregressive Conditional Heteroskedasticity

GARCH – General Autoregressive Conditional Heteroskedasticity

HS – Historical Simulation

IGARCH – Integrated General Autoregressive Conditional Heteroskedasticity

i. e. – id est (Latin word whose meaning in English is ‘that is’)

i.i.d. – independent and identically distributed

LR – Likelihood Ratio

L.T.C.M. – Long Term Capital Management hedge fund

P&L – Profit and Loss

TR-statistic – Tail-Risk-of-VaR-statistic

UC – Unconditional Coverage

VaR – Value-at-Risk

Vol. Adj. – Volatility Adjusted

1. Introduction

This dissertation aims to analyse the performance of a recently developed method of backtesting Value-at-Risk models, the Saddle-point backtest, in comparison to the main existing methodologies in academic literature.

This new framework appears as an alternative to the traditional set of Value-at-Risk backtests established by the Basel Accords for Banking Supervision, whose core consists in evaluating the performance of risk models according to the number and independence of exceedances, neglecting its size. The Saddle-point, created by Wong (2007), considers the number, as well as the size of the exceedances, for a more powerful way of assessing VaR models.

The lack of power observed in traditional methods for rejecting models showing extreme losses is the main motivation for this approach. Recent crises, such as the Black Monday in 1987 and the Subprime in 2007, constitute evidence that models fail to measure risk correctly when it comes to extreme events, and current backtests have seldom enough power to reject them.

Therefore, this dissertation seeks an answer to the following research question: "Does the Saddle-point test outperform traditional backtesting techniques for very low quantiles?".

Furthermore, coming up with a test that overcomes the previous feature does not suffice. It is also important that it performs correctly under "normal" market environment, i.e. during periods of low volatility in which extreme losses do not occur. Hence, a second research question arises: "Is this method a reliable backtest procedure over restful market climate?".

At last, Wong has applied the Saddle-point backtest to many parametric risk models, but never to non-parametric ones. Thus, another interesting research question to explore is: "Does the Saddle-point backtest work for non- and semi-parametric risk models as well?".

In this sense, section 2 states the literature review, which provides an overview on the Risk Management subject, as well as a theoretical contextualization of the Saddle-point backtest.

Additionally, sections 3, 4 and 5 offers a theoretical description of the backtesting techniques and risk models employed throughout this dissertation.

Moreover, section 6 introduces the financial data chosen to implement the risk models and draws some guidelines as how to apply all these theories empirically, whereas section 7 summarizes the main results taken from the practical application.

Finally, the last section focuses on some concluding remarks regarding risk model performance and, most importantly, the prominence of Saddle-point backtest in Risk Management.

2. Literature Review

Risk is defined as the probability of some adverse event to occur at some point in time.

In financial markets, risk takes the form of eventual losses. There are many components that contribute to the uncertainty underlying financial markets, each one representing a different source of risk, such as market risk, liquidity risk, credit risk and operational risk. Thus, Risk Management is the approach each financial institution takes, aiming to supervise and control the existing sources of risk which their portfolio of investments is exposed to.

A number of different approaches to measure the risk inherent to a portfolio of assets have been developed throughout the years. Alongside, several ways of testing these measures have gained a great deal of importance given that the Basel Accords rely on them to compute the Risk Capital Charge's multiplier.

Since the 1970's that many financial institutions started to develop quantitative models trying to measure market risk. In the beginning of the 1990's, JP Morgan released a technical document explaining how to put in practice an innovative way of quantifying risk, denominated Value-at-Risk. According to MacAleer (2009), Value-at-Risk may be seen as the "worst-case scenario on a typical day", as it provides maximum expected loss forecasts, subject to past financial information.

Krause (2003) points out that Value-at-Risk models – hereinafter denominated VaR, for simplicity - have become widespread amongst financial institutions due to its straightforwardness. Furthermore, VaR is currently a universal way of quantifying market risk, having been set as the standard methodology for risk assessment world-wide by the Basel Committee on Banking Supervision (BCBS). Additionally, VaR is used to determine the daily Capital Requirements for financial institutions, which is the amount of capital a bank must hold in order to prevent bankruptcy on a given day.

The institutionalization of VaR as the main risk measure amongst financial institutions was followed by designated ways to test it. Not every risk model works properly for any portfolio composition or period of time, which is why the Basel Committee on Banking Supervision leave this choice for the financial institution to make. However, these models are subject to regulatory backtests in order to be approved. To assess the suitability of risk models, Basel Committee (1996b) amendment implemented the 'traffic light' backtest, a global standard coverage test that determines the penalty attributed to the bank's Risk Capital Charge, based on the number of exceedances verified during the previous 250 working days (more or less a working year of financial data). In addition, other coverage tests have been recommended and widely put into practice as well, such as the Unconditional Coverage backtest developed by Kupiec (1995), that computes the number of exceedances observed over a given period and assesses whether the verified exceedance rate corresponds to the expected exceedance rate; and Christoffersen's Conditional Coverage test, which added an independence testing feature to Kupiec's UC test, hence constituting an improved procedure that takes into account both the number and independence of exceedances.

Additionally, there have been developed plenty of different backtesting methodologies that rely on the number or/and independence of exceedances. For instance, the Dynamic Quantile test of Engle & Manganelli (2004); the Berkowitz, Christoffersen and Pelletier (2011) test; the Regression Quantile backtest developed by Gaglianone et al (2011); and many others. For a summarized description of backtesting methodologies carried out in academic literature, see Nieto & Ruiz (2016).

Even though the traditional backtesting framework has gained popularity due to its plainness and easiness of implementation, it has also been subject of plenty criticism. For instance, Pritsker (2006) documented that Kupiec's test lacks power to reject unfeasible VaR models, and Escaciano & Olmo (2010) point out that Christoffersen's test procedure, joining UC and independence, may not consider the uncertainty inherent to the parameter estimation carried out by VaR models with some degree of parameterization, thus producing misleading results in a complex environment.

Nevertheless, according to Dias (2013) and Halbheib & Pohlmeier (2012), these backtesting procedures are still the most popular ones amongst both practitioners and scientific literature.

In addition to the already mentioned shortcomings, some authors such as Wong (2007, 2008) and Colletaz et al (2013), further criticize traditional backtests for overlooking the size of exceedances. Colletaz et al (2013) stress the fact that a risk model showing a pair of exceedances during the test period, is most likely to be accepted by the traditional backtesting procedures, even though these exceedances may be large enough to make the financial institution to go bankrupt.

The proliferation of complex financial products and the global rise in leverage have contributed to the increasing occurrence of extreme events, having led some authors to measure risk through the size of tail losses. For instance, Artzner et al (1999) suggest Expected Tail Loss (ETL) as an alternative to Value-at-Risk, which has been shown by Basak & Shapiro (1998) to produce lower losses than interval-based forecasts. Furthermore, Yamai & Yoshiba (2005), Taasche (2002) and Lucas & Siegmann (2008) study the relevance of Expected Shortfall as a reliable risk measure and further discuss its advantages over Value-at-Risk.

Likewise, backtesting through tail losses have been subject of thorough research. Berkowitz (2001) and Kerkhof & Melenberg (2004) have designed innovative backtests suitable for ETL and Expected Shortfall, respectively. However, as referred in Wong (2007, 2008), these tests require a large data set in order to be applied properly, which may not be available.

More recently, Colletaz et al (2013) developed the Risk Map, a three-dimensional representation of the number and magnitude of exceedances, as well as an overall graphical summary of the risk model's performance. The intuition behind this approach is not only to account for the regular – usually 1% – VaR exceedances, but also for the super-exceedances, which assume much lower VaR quantiles than 1%. This way, Colletaz et al (2013) propose to compute both sequences of exceedances and super-exceedances, in order to jointly test whether these satisfy the conditions stated in Kupiec's Unconditional Coverage backtest.

Although there are quite a few backtesting methodologies based on tail losses, the scope of this dissertation lies on Wong's Saddle-point backtest. In his first article, Wong (2007) alerts to the need of backtesting VaR models through the size of tail losses as an alternative to the frequency-based tests. Hence, the stated author proposes an innovative way of backtesting Expected Shortfall – in Wong (2007) – and Tail-Risk-of-VaR – in Wong (2008) –, accounting for the magnitude of exceedances rather than its frequency. Wong (2008) resorts to the Saddle-point technique, derived through the Lugannani & Rice (1980) formula, in order to asymptotically approximate tail loss distribution to the normal gaussian one. The major assumption over which this approach is built is that financial returns are independent and identically distributed, following a standard normal distribution. Furthermore, when the risk model assumes non-normal distributions for the returns, the Saddle-point backtest applies the same procedure as in Berkowitz & O'Brien (2002) and Kerkhof & Melenberg (2004), thus transforming the non-normal density into a normal one.

However, in his pioneer study, Mandelbrot (1963) found evidence of volatility clustering and autocorrelation in financial markets, which jeopardizes the reliability of the previous assumption. Nonetheless, the Saddle-point backtest was shown in Wong (2007, 2008) to increase test power and accuracy, when compared either to the traditional frequency-based framework or to the magnitude tests of Berkowitz (2001) and Kerkhof & Melenberg (2004).

Additionally, the Saddle-point approach is also suitable for small samples, which constitutes an advantage over the two previously stated tail losses backtesting procedures.

Overall, there is a vast variety of VaR models to choose from; see Kuester et al (2006) for an extensive comparative analysis of existing models. Attending to Gerlach et al (2011), these VaR models may be further divided into three different parametrization categories: non-parametric, semi-parametric and parametric. Amongst these, the parametric category has a much wider number of models to choose from, including the widespread RiskMetrics, Autoregressive Moving Average models with Bollerslev's Generalized Autoregressive Conditional Heteroskedasticity process – hereinafter denominated GARCH –, as well as various distribution functions for innovations, return processes and volatility dynamics. Conversely, the non- and semi-parametric categories have lower representation in what comes to variety of models. However, according to a study put forward by Pérignon & Smith (2010), thereabout three quarters of commercial banks worldwide prefer to use non-parametric Historical Simulation rather than any other method.

Following his first two papers, Wong has published two more articles with regard to the Saddle-point backtest – Wong (2009) and Wong et al (2012) – where he empirically applied this methodology to the exchange and stock markets, respectively. However, the analyses were always carried out based solely on parametric models. In fact, to my knowledge, neither Wong nor any other author has applied the Saddle-point backtest to non-parametric risk models so far, which constitutes a novelty put forward by this dissertation.

3. Risk Management according to the Basel Committee

In 1974, shortly after the collapse of the Bretton Woods international monetary system, central bankers of the G10 countries gathered down and created the Basel Committee on Banking Supervision. Aiming to prevent major crashes from happening and to promote responsible Risk Management practices, this organization designed a list of guidelines every financial institution should follow. Amongst these, one of the most relevant guidelines is the Market Risk Charge, which is the obligation of setting aside a certain amount of capital in order to face eventual short-term losses.

Although the BCBS gives autonomy to commercial banks regarding the choice of the risk model, it requires the following compliance criteria to be met: computation of daily VaR figures; VaR significance levels of 1%; minimum 250 working days of in-sample financial data; data sample updates at least every 3 months and a minimum holding period of 10 days, although it is accepted the square root scaling rule up until 10 days-ahead-VaR. See Basel Committee (1996) for details.

Following the previous steps, financial institutions must hold a minimum capital requirement established by the formula below:

$$MRC_t = Max\left[VaR_{t+1}^{\alpha};\ k \times \frac{1}{60} \times \sum_{i=1}^{60} VaR_{t-i}^{\alpha}\right] + c \qquad (1)$$

Where α represents VaR's significance level, which according to the BCBS should be 1%; k is a penalty multiplier that varies between 3 and 4, subject to the backtesting performance of the risk model and, at last, c accounts for an additional idiosyncratic capital charge.

3.1. Backtesting

Backtesting constitutes one of the most relevant practices risk managers have to take on. On the one hand, backtesting allows Risk Management departments at commercial banks to assess whether their risk models are correctly capturing the risk inherent to its portfolio of assets, thus assisting in the decision-making process. On the other hand, as mentioned in the previous section, the BCBS relies on backtesting procedures for the calculation of banks' capital requirement, which also makes backtesting a tool for rewarding and penalizing good and poor Risk Management practices, respectively.

This way, given the opportunity cost imposed by the BCBS to financial institutions whose risk models show insufficient risk coverage, commercial banks have an incentive to keep working on better ways to quantify the risk and improve their Risk Management practices.

3.1.1. Traffic Light Backtest

Alongside with Capital Requirements calculation, the Basel Committee (1996b) also specified a global standard backtesting methodology to assess banks' proprietary VaR models, named Traffic Light Coverage test. The implementation of this backtest requires the financial institution to respect the previously stated criteria regarding the computation of its VaR figures.

In order to perform this test, it is first necessary to sum up the number of exceedances produced by the VaR model during the test period. An exceedance takes place any time an observed loss in the market surpasses the Value-at-Risk predicted by the model.

After comparing the daily VaR forecasts with the actual returns observed in the market throughout the test period, the number of exceedances is recorded so that the VaR model can be scored inside one of three different coloured categories described in figure 1.

Figure 1: Traffic Light backtest – penalties

Zone	Number of Exceedances	k
Green	0	3.00
	1	3.00
	2	3.00
	3	3.00
	4	3.00
Yellow	5	3.40
	6	3.50
	7	3.65
	8	3.75
	9	3.85
Red	10 or more	4.00 or rejection

On the one hand, the green zone includes VaR forecasts that registered a maximum of 4 exceedances during the 250 days test period, constituting no concern to the regulation authority in what comes to wrongly accept inaccurate models and, therefore, are not subject to any penalty. On the other hand, the yellow zone indicates that the VaR model has come up short in predicting market risk between 5 and 9 times throughout the test sample, which does not allow any definitive conclusion about its accuracy. Thus, as shown in figure 1, the Basel Committee (1996b) recommends attributing penalties to the financial institutions whose VaR models stand within this intermedium category. At last, any model producing 10 or more exceedances end up falling in the red zone, resulting either in a maximum penalty of 4 being applied to the bank's Market Risk Charge, or in the rejection of the risk model by the regulatory authority. Actually, in these cases, the second alternative is the most commonly seen in practice and requires the bank to change its VaR model.

Given its simplicity, the BCBS adopted the Traffic Light Coverage test as the global backtesting procedure amongst financial institutions. However, as a result of its plainness, it lacks power to reject inaccurate models.

Moreover, this test is based upon the assumption that returns are independent from each other - which has been disproved by Mandelbrot (1963) - and, in addition, it also neglects the sizes of tail losses, which makes it suitable only to a preliminary analysis on the performance of risk models. There are other backtesting methodologies known to provide higher accuracy and test power than the Traffic Light backtest.

3.1.2. Kupiec's Unconditional Coverage test

The Unconditional Coverage backtest, created by Kupiec (1995) and further discussed in Christoffersen (1998), assesses whether the exceedance rate verified throughout the test period (π_{obs}) corresponds to the expected exceedance rate (π_{exp}), defined as the VaR significance level α. Hence, the null hypothesis of the test takes the form:

$$H_0: \pi_{obs} = \pi_{exp} \equiv \alpha \tag{2}$$

Specifically, the observed rate π_{obs} is defined as $\frac{x}{n}$, where x is the number of exceedances observed in the sample test and n represents the total number of observations in the sample. Therefore, alternatively one may write down the null hypothesis as follows:

$$H_0: \pi_{obs} = \frac{x}{n} = \alpha \tag{3}$$

Furthermore, Kupiec (1995) resorts to the likelihood-ratio test presented below, as to assess whether the exceedances sample estimate is statistically consistent with the significance level of the underlying VaR model.

$$-2\ln\left(\left(\frac{\alpha}{\pi_{obs}}\right)^{x} \times \left(\frac{1-\alpha}{1-\pi_{obs}}\right)^{n-x}\right) \sim \boldsymbol{\chi}_1^2 \tag{4}$$

Whenever the test statistic exceeds the critical value established according to its significance level, the Unconditional Coverage backtest rejects the null hypothesis, thus rejecting the VaR model. Moreover, if no exceedance is recorded throughout the entire sample test period, then this test cannot be applied.

Overall, even though Kupiec's coverage test constitutes a useful benchmark in backtesting VaR models, the fact that it is not applicable to a scenario where there are no exceedances during the test period, makes it a poor methodology for high confidence levels and small sample analysis, which is especially concerning given that the BCBS requires solely a year of past financial data for backtesting purposes. In addition, Kupiec also does not assess the trustworthiness of the independence assumption of the exceedances. Nevertheless, in spite of showing low power to reject inaccurate models - especially when the number of exceedances is zero -, this methodology has Basel Committee's stamp of approval, being still one of the most widespread backtests amongst financial institutions for Capital Requirements calculation purposes.

3.1.3. Christoffersen's Independence test

As previously stated, Kupiec's Unconditional Coverage test considers the number of exceedances observed in the test sample, but disregards whether they are clustered in time. Aiming to overcome this issue, Christoffersen (1998) implemented the Independence backtest, based on the null hypothesis that Kupiec's observed exceedance rate π_{obs} follows an independent and identically distributed (i.i.d.) Bernoulli process.

$$\text{H}_0\text{: } \pi_{obs} \overset{i.i.d.}{\sim} Bern \tag{5}$$

In order to assess the independent assumption, the author created the following likelihood-ratio statistic:

$$LR_{ind} = \frac{\pi_{obs}^{n_1} \times (1 - \pi_{obs})^{n_0}}{\pi_{01}^{n_{01}} \times (1 - \pi_{01})^{n_{00}} \times \pi_{11}^{n_{11}} \times (1 - \pi_{11})^{n_{10}}} \tag{6}$$

On the one hand, n_0 represents the number of days in which the model made a correct assessment of the risk. On the other hand, n_1 stands for the number of days an exceedance took place. Moreover, following the same reasoning, n_{11} accounts for the frequency with which two consecutive exceedances occurred, n_{00} represents how often two consecutive days without recording exceedances took place, n_{10} stands for the number of times an exceedance was followed by a non-exceedance and, conversely, n_{01} sums up the situations in which an exceedance followed a non-exceedance.

In turn,

- $\pi_{01} = \frac{n_{01}}{n_{01}+n_{00}}$ represents the proportion of exceedances given that a non-exceedance took place the previous day;
- $\pi_{00} = \frac{n_{11}}{n_{01}+n_{00}}$ represents the proportion of exceedances given that an exceedance took place the previous day.

What's more, resembling the Unconditional Coverage test, Christoffersen's independence test statistic and p-value are calculated through a Chi-square distribution function with one degree of freedom, as stated below:

$$-2\ln(LR_{ind}) \sim \chi_1^2 \tag{7}$$

In conclusion, this procedure constitutes a reliable assessment tool for the independence assumption. However, performed alone, this test is not relevant for VaR backtesting purposes since it only takes into account the occurrence of consecutive exceedances and not its size.

3.1.4. Christoffersen's Conditional Coverage test

After overcoming the independence pitfall showed in Kupiec's test, Christoffersen (1998) carried on and built another test that could assess both assumptions at the same time: correct risk coverage and independence of the exceedances. This way, the author created the Conditional Coverage backtest, underlying the following null hypothesis:

$$H_0: \pi_{obs} = \pi_{exp} \equiv \alpha \ \wedge \ \pi_{obs} \overset{i.i.d.}{\sim} Bern \tag{8}$$

The true purpose for which Christoffersen (1998) came up with the Independence test in the first place, was so that its likelihood-ratio statistic could be added to the one in Kupiec (1995), thus creating the Conditional Coverage test statistic, which follows a Chi-square distribution with 2 degrees of freedom, as shown in the appendix of Christoffersen (1998).

$$LR_{CC} = LR_{UC} + LR_{ind} \tag{9}$$

$$-2\ln(LR_{CC}) \sim \boldsymbol{\chi}_2^2 \tag{10}$$

In a similar way as in section 3.2, according to this methodology, the null hypothesis is rejected anytime the test statistic overshoots the critical value, leading to the rejection of the VaR model. Otherwise, the risk model is believed to simultaneously respect the independence property and to show correct risk coverage.

4. Wong's Saddle-point backtest

Having performed an outlook to BCBS's Risk Management requirements, as well as to the traditional backtesting techniques in section 3, it will now be put forward an alternative methodology to assess risk models based on the sizes of tail losses proposed by Wong (2008).

4.1. Backtesting through sizes of exceedances

Wong et al (2012) point out that recent financial crisis arose some pitfalls regarding traditional frequency-based backtests described in the previous section, claiming that these showed not to perform adequately when big losses occur, contrarily to the widespread belief that they correctly evaluate VaR models. In order to improve VaR backtesting, the stated author introduces the Saddle-point backtest, a size-based procedure that allows to assess the performance of a risk model, attending to Tail-Risk-of-Var and Expected Shortfall measures as alternative Basel II counterparts.

4.1.1. Tail Risk of VaR

Given a random portfolio return series R and denoting G as its absolutely continuous cumulative distribution function (CDF), the number of exceedances over a sample period of T days required by the BCBS is:

$$n = \sum_{t=1}^{T} I(r_t < q) \tag{11}$$

Where *I(.)* is an indicator function and $-q = -G^{-1}(\alpha)$ stands for the VaR at the α significance level, which, according to Basel II, should be set at 1% for market risk. However, as seen in Wong (2007 & 2008), traditional frequency-based backtesting framework lacks power compared to size-based backtests, leading Wong (2008) to use a Tail-Risk-of-VaR (TR) test statistic:

$$\widehat{TR} = -\frac{1}{T}\sum_{t=1}^{T}(r_t - q) \cdot I(r_t < q) \tag{12}$$

Tail Risk may be seen as a size counterpart of Basel II rules instead of the previously stated frequency-based one, since it sums up the sizes of all losses exceeding VaR. By doing so, $\widehat{TR}$ provides information to risk managers about the potential loss the institution may incur in beyond the VaR boundary.

4.1.2. Expected Shortfall

Moreover, Expected Shortfall has gained some popularity since the subprime crisis took place, appearing as an alternative to VaR. As already referred, some authors like Artzner et al (1999) present its advantages over VaR, pointing it as a more reliable risk measurement tool. In fact, in his first article regarding the Saddle-point backtest, Wong (2007) used Expected Shortfall as the size-based risk measure to be tested. Assuming again the series of portfolio returns *R* and VaR figures given by $-q$, ES may be regarded as the loss the financial institution expects to incur in when VaR is breached. Thus, setting a series of T random returns $\{r_1, \dots, r_T\}$ and assuming there are *n* exceedances $\{r_{(1)}, \dots, r_{(n)}\}$ inside the sample, then Expected Shortfall can be estimated just as in Kerkhoff & Melenberg (2004) and Wong (2007):

$$\widehat{ES} = -E(R|R < q) = -\frac{1}{n}\sum_{i=1}^{n} r_{(i)} \tag{13}$$

4.1.3. Relationship between TR and ES

Both approaches presented above take into account the size of the exceedances. Although, while ES is regarded as the average loss beyond VaR, TR is proportional to the aggregate sum of all tail losses. Furthermore, as Wong (2008) and Wong et al (2012) point out, these are related in the following way:

$$\widehat{TR} = \frac{n}{T} \cdot \widehat{ES} - \frac{nq}{T} \Leftrightarrow \alpha^{-1}\widehat{TR} = \widehat{ES} - VaR \tag{14}$$

Where the α for market risk is set at 1%, according to Basel II amendment. In his latest article, Wong (2012) explains that Kerkhoff & Melenberg (2004) compute ES using the 100α% worse returns in the sample, which makes $\frac{n}{T}$ constantly equal to α for any sample, resulting in the previous formula to hold regardless of the parameters. Such an approach would lead to the conclusion that there is no difference between backtesting Tail Risk or Expected Shortfall.

However, VaR is the risk metric sanctioned by the BCBS, and the procedures for risk model evaluation stipulated by this committee, require that VaR figures are compared with its homologous returns on a daily basis. All in all, it means that TR and ES should be calculated using ex-ante VaR figures, causing n to vary from sample to sample and, thus, $\frac{n}{T}$ not to be necessarily equal to α for any sample. Consequently, under this assumption, TR and ES are regarded as complementary in backtesting.

Tail-Risk-of-VaR represents the risk at the part of the tail that VaR does not account for, and even though it is closely related to Expected Shortfall, there are statistical differences between them in backtesting. On the one hand, by summing up the sizes of all losses in excess of VaR, TR can work as a direct size counterpart for the BCBS's traffic light test, whereas ES, which accounts only for the average of losses in days where the loss has exceeded VaR, cannot. On the other hand, due to the fact that VaR has become the worldwide standard risk metric, having been target of thorough research in academic literature for the last 20 years, Wong (2009) suggests it might be more useful to focus on improving VaR backtesting through Tail Risk, rather than delving into backtesting an alternative measure such as Expected Shortfall. This way, it is important to note that Tail-Risk-of-VaR Saddle-point backtest works as a complement to the Basel II rules, since the latter counts the number of all exceedances while the former sums up its sizes.

Nevertheless, ES plays an important role in backtesting since it is able to detect some inaccuracies that TR cannot. For instance, Wong (2012) refers that if a risk model records a very small number of exceedances during the test period, but each one revealing a huge tail loss, then TR might not capture this pitfall as effectively as ES, which has more power to reject such model. However, if the risk model registers too many small exceedances, then the average of losses would probably be too small to be rejected by ES, in which case TR would be preferred, given that it sums up the sizes of all exceedances.

Overall, the Saddle-point backtest may be applied to both Tail-Risk-of-VaR and Expected Shortfall. However, even though Wong (2012) demonstrated that these can be complementary to each other for backtesting, this dissertation focuses solely on VaR backtesting, for it is the most prominent approach among financial institutions and academic literature. Therefore, hereinafter the TR Saddle-point backtest will be described in detail and empirically analysed, along with the traditional backtesting framework.

4.2. Normality assumption

The whole Saddle-point backtest is based upon the assumption that the series of portfolio returns R are independent and identically distributed (i.i.d.), following a normal distribution. In fact, theoretically, according to the central limit theorem, the sampling distribution of *TR* can be approximated by a normal gaussian distribution when the sample size is large. However, Tail Risk is measured by summing up the sizes of exceedances which, following BCBS's rules for market risk, should occur only 1% of the times, making it virtually impossible to gather a data sample large enough to allow the practical application of the central limit theorem. Therefore, this sub-section introduces the Saddle-point technique used by Wong (2007), which constitutes a small sample asymptotic approach to accurately approximate the sampling distribution of *TR* to the normal distribution.

4.2.1. TR-statistic

As already mentioned, the test statistic of the Saddle-point backtest applied in the next chapters of this dissertation, is the Tail-Risk-of-VaR, hereinafter denominated TR-statistic.

Assume again a series of T normally-distributed returns as stated above $R \in \{r_1, \dots, r_T\}$, whose CDF and PDF are the ones of a standard normal distribution, being represented by Φ and ϕ, respectively. Note that, at the moment, VaR is computed as $q = \Phi^{-1}(\alpha)$.

Moreover, define the mixed discrete-continuous random variable X with realization x described below:

$$x = \begin{cases} r - q & if\ r < q \\ 0 & if\ r \geq q \end{cases} \tag{15}$$

Given that $TR = -E(X)$ and $x \in]-\infty, 0]$ is the summand in equation (12), TR-statistic can be defined as:

$$\widehat{TR} = -\bar{x} = -T^{-1} \cdot \sum_{t=1}^{T} x_t. \tag{16}$$

Additionally, the CDF and PDF of X are given, respectively, by the functions below:

$$F(x) = \begin{cases} \Phi(x+q) & if\ x < 0 \\ 1 & if\ x = 0 \end{cases} \tag{17}$$

$$f(x) = \begin{cases} \phi(x+q) & if\ x < 0 \\ 1-\alpha & if\ x = 0 \end{cases} \tag{18}$$

Furthermore, Wong (2008) shows that, assuming TR-statistic – X – is a mixed discrete-continuous random variable whose CDF is given by equation (17), then its moment-generating function is calculated in the following manner:

$$M(t) = \exp\left(-qt + \frac{t^2}{2}\right) \cdot \Phi(q-t) + 1 - \alpha \tag{19}$$

Moreover, the derivatives with respect to *t* are represented by the expressions below:

$$M'(t) = \exp\left(-qt + \frac{t^2}{2}\right) \cdot [(t-q) \cdot \Phi(q-t) - \phi(q-t)] \tag{20}$$

$$M''(t) = (t-q) \cdot M'(t) + M(t) - (1-\alpha) \tag{21}$$

$$M^{(n)}(t) = (t-q) \cdot M^{(n-1)}(t) + (n-1) \cdot M^{(n-2)}(t), \qquad where\ n \geq 3 \tag{22}$$

Finally, the first two moments of TR-statistic are given by:

$$\mu_X = E(X) = -q\alpha - \frac{e^{-\frac{q^2}{2}}}{\sqrt{2\pi}} \tag{23}$$

$$\sigma_X^2 = var(X) = \alpha + q^2\alpha + \frac{q \cdot e^{-\frac{q^2}{2}}}{\sqrt{2\pi}} - \mu_X^2 \tag{24}$$

For mathematical proof regarding TR-statistic's moments, one can consult Appendix A.1 and Appendix A.2 in Wong (2008).

In the next sub-section, it will be presented the formula which Wong (2007 & 2008) has based his work upon, that allows the asymptotical approximation of distribution X to a standard normal distribution and, thus, the computation of the Saddle-point p-value for statistical inference.

4.2.2. Saddle-point p-value

Lugannani and Rice (1980), as explained by Wong (2007), came up with an elegant formula that accurately determines the tail probability of the sample mean of an independent and identically distributed (i.i.d.) random sample $\{X_1, \dots, X_n\}$ from a continuous distribution F with density f.

In order to understand the reasoning behind the saddle-point technique, Wong (2007, 2008) draws our attention to the inversion formula for the PDF of TR-statistic's sample mean:

$$f_{\bar{x}}(\bar{x}) = \frac{T}{2\pi} \int_{-\infty}^{\infty} e^{T \cdot [K(it) - it\bar{x}]} \, dt, \;\; where\ \bar{x} < 0 \tag{25}$$

Above, i represents an imaginary number and $K(t) = \ln M(t)$ is the cumulant generating function of X. Additionally, by setting the derivative (with respect to it) of $K(it) - it\bar{x}$ to zero, one gets the saddle-point ϖ that must satisfy the condition:

$$K'(\varpi) = \bar{x} \tag{26}$$

Once the saddle-point ϖ is settled, the p-value of the test can be obtained through the Lugannani and Rice's formula following the procedure described below.[1]

Defining $\eta = \varpi \cdot \sqrt{T \cdot K''(\varpi)}$ and $\varsigma = sgn(\varpi) \cdot \sqrt{2T \cdot (\varpi\bar{x} - K(\varpi))}$, where $sgn(\varpi)$ is equal to zero when $\varpi = 0$ or takes the same sign of ϖ when $\varpi \neq 0$, then the p-value of the test can be constructed denoting the probability of $\bar{X}$ being lower than the realized sample $\{\bar{x}\}$.

[1] This formula has already been established by Lugannani & Rice (1980) and proved by Daniels (1987).

Moreover, since TR-statistic is nonnegative, the Saddle-point's p-value may also be regarded as the likelihood with which $\bar{X}$ will be greater than the test statistic:

$$P(\bar{X} < \bar{x}) = P(\bar{X} > \widehat{TR}) \approx \Phi(\varsigma) - \phi(\varsigma) \cdot \left[\frac{1}{\eta} - \frac{1}{\varsigma}\right] \tag{27}$$

Finally, in the cases where the test statistic is zero – $\bar{x} = \widehat{TR} = 0$ – the associated p-value for the saddle-point backtest is then given by:

$$P(\bar{X} = 0) = (1 - \alpha)^T \tag{28}$$

4.3. Hypothesis testing

The Basel Committee solicits banks to carry out one-tail frequency-based backtests such as the Traffic Light test, which means risk models cannot be rejected for being over-conservative, i. e. risk models yielding very few or zero exceedance. According to this regulatory entity, only in case the number of exceedances is too large should the VaR model fail the test. In turn, being two-tailed backtests, Kupiec and Christoffersen's coverage tests do not fulfil this requirement, for they take the risk of rejecting a model for having too few exceedances, whereas Wong's saddle-point backtest can either perform as a one-tailed or a two-tailed backtest, subject to the choice of the alternative hypothesis.

In order to comply with BCBS's criteria, this dissertation applies the Saddle-point backtest considering the following null and alternative hypotheses for statistical inference described in Wong (2008):

$$\begin{cases} H_0: TR = TR_0 \\ H_1: TR > TR_0 \end{cases} \tag{29}$$

Where TR is the empirical Tail-Risk-of-VaR collected from the sample test period, which represents the Saddle-point test-statistic; and TR_0 defines the value TR should denote in order not to reject the test's null hypothesis. Furthermore, admitting a standard normal distribution for the null hypothesis and sticking to Basel Committee's approach for market risk, $\alpha = 0.01$, then $\alpha^{-1} \cdot TR_0 = \alpha^{-1} \cdot \mu_X = 0.3389$.

If the observed tail risk (TR) is not significantly larger than TR_0, the null hypothesis is not rejected, implying that the risk model is consistent with the empirically observed Tail-Risk-of-VaR. Conversely, in case TR is in fact considerably larger than TR_0, then the risk model is believed to have inconsistent tail risk coverage, leading to the rejection of the test's null hypothesis and, consequently, to the rejection of the risk model.

In addition, if one wants to perform two-tail backtesting analysis using the saddle-point technique, then the statistical hypotheses may take the form:

$$\begin{cases} H_0: TR = TR_0 \\ H_1: TR \neq TR_0 \end{cases} \quad (30)$$

However, this dissertation will not perform such statistical inference, for it is not the standard procedure recommended by the regulatory entity.

4.4. Non-normal null hypothesis

The whole Saddle-point backtest is based upon the assumption that the series of portfolio returns is independent and normally distributed. Although, if the null hypothesis assumes a non-normal distribution for the returns, then it is possible to transform the non-normal series of returns into a normal one by following Berkowitz & O'Brien (2002) and Kerkhof & Melenberg (2004). Given, for instance, some return series R following a student-t distribution with 5 degrees of freedom whose standard cumulative distribution function (CDF) is denoted by $f_5(R)$, can easily be transformed into a normalized variable as described below:

$$Z = \Phi^{-1}(f_5(R)) \quad (31)$$

Note that, once again, Φ^{-1} represents the standard normal CDF.

Moreover, Wong (2008) adds that when the return distribution of the risk model assumes fatter tails, then the transformation above results in thinner tails than the one of the standard normal distribution, which in turn leads to TR-statistic significantly lower than if it had been assumed a standard normal distribution from the beginning. Oppositely, in case the risk model predicts a thinner-tailed distribution for the returns, the transformed variable will denote fatter tails than the normal distribution, ergo resulting in considerably larger TR-statistic.

Now that the Saddle-point backtest has already been contextualized and described in detail, along with the BCBS Risk Management rules and the traditional backtesting framework, it is time to apply it empirically to the financial markets. Hence, the next section provides a comprehensive overview on the Value-at-Risk methodology, which is the risk modelling followed by this dissertation in order to put forward the empirical analysis of the Saddle-point backtest. In addition, section 5 also explains how to apply the referred statistical test to each risk model, including non-parametric VaR.

5. Value-at-Risk

Value-at-Risk is defined as some loss threshold that, for a given time horizon h, will not be exceeded $100(1\text{-}\alpha)\%$ of the times, assuming that the portfolio's composition remains unchanged during this period. This way, it is important to reinforce that VaR denotes a lower boundary for the portfolio's P&L that is expected to be surpassed only $100\alpha\%$ of the times, rather than the maximum possible portfolio loss.

Thus, VaR consists in the α-quantile of the distribution of returns, conditional on past information, such that there is a $100\alpha\%$ probability to occur a portfolio loss R_t that will exceed it:

$$\text{Prob}(\text{-}R_t > \text{VaR}_t^{\alpha} |\ t\text{-}1) = \alpha \tag{32}$$

In order to construct a risk model following the VaR methodology described above, it is important to make certain assumptions and establish some parameters a priori, specifically:

- Holding period (h)
- Significance level (α)
- Type of model

Note that BCBS's rules specify $\alpha = 1\%$ for market risk, which is the value to be used in this dissertation. Furthermore, this regulatory authority requires a holding period of at least 10 days, allowing the scaling computation from 1-day-ahead-VaR to 10-days-ahead VaR by multiplying the former by the square root of the holding period as follows:

$$\text{VaR}_{10\text{ days}}^{\alpha} = \text{VaR}_{1\text{ day}}^{\alpha} \cdot \sqrt{h} \tag{33}$$

However, BCBS does not regard this scaling operation as being applicable for time horizons greater than 10 days.

Krause (2003) points out that VaR models have become widespread amongst financial institutions due to its simplicity, being nowadays the universal way of quantifying market risk for Regulatory Capital purposes.

Although there are many VaR models to choose from, this dissertation uses five models often seen in literature and used in practise amongst commercial banks:

- JP Morgan's RiskMetrics
- Student-t VaR
- Skewed-Student VaR
- Historical Simulation
- Vol. Adj. Historical Simulation

Moreover, according to Gerlach (2011), these models may be inserted into different parametrization categories. The first three models are regarded as parametric since these assign a certain distribution function to the returns and its volatility dynamics, assuming the market behaves accordingly. The fourth VaR model is inside the non-parametric category, as it simply takes the empirical α-quantile out of the sample of returns and defines it as the VaR figure for the next day, without making any assumption about the returns' distribution or its volatility. Finally, although the last model makes certain assumptions concerning the volatility of the errors, it does not parametrize the returns distribution, thus belonging to the semi-parametric category.

5.1. JP Morgan's RiskMetrics

In 1994, JP Morgan in partnership with Reuters, developed a technical document introducing the RiskMetrics, a fully parametrized way of measuring market risk, whose goal was to become the global benchmark approach to compute VaR. This document was updated several times throughout the years, but this dissertation bases on Morgan (1996).

The main assumption present in RiskMetrics methodology is that financial returns are normally-generated, following the autoregressive process described below:

$$r_{t+1} = \mu_t + \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim N(0,1) \tag{34}$$

Where μ_t is the mean of the returns, σ_t represent the volatility of returns and ε_t is the innovation process which follows a standard normal distribution. In addition, RiskMetrics methodology structures volatility considering the Exponential Weighted Moving Average (EWMA) technique, whose goal consists in attributing greater preponderance to recent market conditions in comparison to distant observations:

$$\hat{\sigma}^2_{t+1|t} = \lambda\hat{\sigma}^2_t + (1-\lambda)r^2_t, \qquad \lambda = 0{,}94 \tag{35}$$

Here, $\hat{\sigma}^2_t$ stands for the latest variance estimator, r^2_t represents the square of the most recent realized return and λ is a scaling parameter that assigns different weights to the previous variables for the forecast of the next day's volatility.

This method resembles the Integrated GARCH (IGARCH) model to forecast volatility. However, while IGARCH writes down a statistical model and computes its parameters through Maximum Likelihood estimation, RiskMetrics keeps $\lambda = 0{,}94$ for the entire analysis, for its goal is to derive a simple way of modelling returns that is consistent empirically; for extensive analysis regarding the IGARCH model, see Lumsdaine (1995).

All in all, RiskMetrics VaR model can be formulated in a straightforward way:

$$\mathrm{VaR}^{\alpha}_{t+1} = -\mu_t - \sigma_t \cdot \Phi^{-1}(\alpha) \tag{36}$$

Being Φ^{-1} the inverse standard normal CDF, which for market risk is set at $\Phi^{-1}(0.01) \approx -2.33$; and the volatility being estimated through EWMA as described above.

Shortly after its release, RiskMetrics rapidly took off and many commercial banks around the world adopted this market risk measure, mainly due to its success during the first years of implementation. Although, after the fall of some important hedge funds and investment companies in the late 1990's, such as L.T.C.M., RiskMetrics was no longer seen as foolproof. However, that was not reason enough for the world of Finance to disregard RiskMetrics completely, having remained as one of the most widespread risk measures so far. Moreover, Pafka & Kondor (2008) suggest that RiskMetrics performs satisfactorily much of the times, but only for confidence levels no higher than 95%. For lower quantiles it underestimates market risk more frequently than would be expected, which constitutes a concern since it was designed to provide accurate forecasts at the 99% confidence level.

Nevertheless, as long as the market does not exhibit extreme returns, RiskMetrics is believed to perform rather well.

At last, the application of the saddle-point backtest in evaluating RiskMetrics VaR is quite simple since it already follows a normal distribution from the start. Hence, one simply needs to standardize the returns as described below, and then assume the α-quantile of the normal distribution $\Phi^{-1}(\alpha)$ as the VaR figure.

$$z_t = \frac{r_t - \mu_t}{\sigma_t} \tag{37}$$

5.2. Student-t VaR

In its pioneer findings, Mandelbrot (1963) highlighted that the behaviour of financial time series deviated from standard gaussian assumptions, specially the one that regards returns as being independent and identically distributed. In fact, this author gathered empirical evidence confirming that extreme financial returns tend to take place more frequently than predicted by a gaussian distribution, which brings up the need to account for fatter tails in the returns distribution. With that goal in mind, there have been used a number of different distributions in order to properly model returns, one of which being the student-t distribution. Within this scope, Lin & Shen (2006) show that Student-t Value-at-Risk helps improving financial markets forecasts in comparison to the normal distribution, mostly due to the fact that it accounts for fatter tails.

Table 1: Student-t distribution

	Quantile				
v	**0.1%**	**0.3%**	**0.5%**	**1%**	**5%**
2	-22.33	-12.85	-9.92	-6.96	-2.92
3	-10.21	-6.99	-5.84	-4.54	-2.35
10	-4.14	-3.47	-3.17	-2.76	-1.81
50	-3.26	-2.87	-2.68	-2.40	-1.68
100	-3.17	-2.81	-2.63	-2.36	-1.66
500	-3.11	-2.76	-2.59	-2.33	-1.65
∞ (Normal)	-3.11	-2.76	-2.58	-2.33	-1.65

Table 1 displays values computed from the student-t distribution's CDF for different quantiles and degrees of freedom (ν), which allows us to conclude that, on the one hand, as the number of degrees of freedom approach infinity, the student-t becomes asymptotically normal. On the other hand, the lower the quantile of the distribution, the higher is the discrepancy between them, especially for low degrees of freedom. Therefore, it is clear that student-t expects extreme occurrences much more often than the normal distribution does.

Moreover, constructing VaR models using the student-t distribution is performed in a similar matter as described in the previous section. However, instead of assuming that the returns are normally distributed, this time a standard student-t distribution is implemented instead:

$$\mathrm{VaR}_{t+1}^{\alpha} = -\mu_t - \sigma_t \cdot f_{\nu}{}^{-1}(\alpha) \cdot \sqrt{\frac{\nu - 2}{\nu}} \tag{38}$$

Where $\nu = [2, +\infty[$ stands for the number of degrees of freedom of the student-t distribution and the volatility is, once again, calculated through the EWMA methodology. Furthermore, it is important to highlight that in equation (37) the inverse of the student-t's cumulative density function is multiplied by a square root in order to standardize the distribution, i.e. to transform it into a homogenous distribution with zero mean and standard deviation equal to one.

What's more, note that the methodology is similar to the one of RiskMetrics, except a standard student-t distribution with ν degrees of freedom replaces the standard gaussian approach, which permits to account for fatter tails than a normal distribution does.

Additionally, this dissertation resorts to Maximum Likelihood estimation in order to compute the optimal degrees of freedom for the returns' distribution, at each point in the sample. To do so, firstly, it is necessary to standardize the returns in the sample in the same way as in equation (37):

Then, defining $\Gamma(\cdot)$ as the Gamma function, the optimal number of degrees of freedom is calculated by maximizing the following log-likelihood function, which is the log-function of the standardized (mean zero and unit variance) student-t probability density function:

$$\max_{\nu}\left\{\sum_{i=1}^{n} \ln\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \left(\left(\frac{\nu+1}{2}\right)\cdot \ln\left[\Gamma\left(1+\frac{z_{t-i}^{2}}{\nu-2}\right)\right]\right) - \frac{\ln(\nu-2)-\ln(\pi)}{2}\right\} \qquad (39)$$

Since the VaR forecast is performed using a rolling window that changes every day, this maximum likelihood estimation is performed the same way.

Finally, in order to apply the saddle-point backtest to such a distribution, it is necessary to normalize it using the same technique described in section 4.4. In this case, financial returns are assumed by the risk model to follow a student-t distribution with ν degrees of freedom. Therefore, through the standard student-t cumulative density function stated below, it is possible to transform afterwards student-t returns into normal ones by applying equation (31).

$$f_{\nu}(z_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\cdot\left(1+\frac{{z_t}^{2}}{\nu-2}\right)^{-\left(\frac{\nu+1}{2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)\cdot\sqrt{(\nu-2)\cdot\pi}} \qquad (40)$$

Overall, although it is a theoretical improvement comparing to the normal distribution, this method still requires plenty assumptions regarding the properties of returns, not to mention that the parameters of the distribution are again entirely on the hands of the analysts, who get the fall in case the model performs poorly.

5.3. Skewed-Student VaR

In real financial markets, it is rare to find a symmetrically distributed sample of returns. Usually, it has some sort of bias, either to the right (positively skewed) or to the left (negatively skewed). In other words, it means that, typically, financial returns either have frequent small gains and few large losses or few large gains and several small losses. Therefore, since normal and student-t are symmetrical (in relation to zero) distributions, these may not be appropriate to model the risk inherent to a P&L portfolio.

Thus, the third and last parametric VaR model used in this dissertation assumes a skewed-student distribution, which adds a skewness parameter to the student-t distribution.

First of all, in order to apply the Skewed-Student VaR model, one needs to define the quantile function required to perform the Value-at-Risk computation. Hence, the skewed-student quantile is defined below as stated by Lambert & Laurent (2001).

$$skst^{-1}{}_{\xi,\nu}(\alpha)\begin{cases}\dfrac{\frac{1}{\xi}\cdot f_{\nu}{}^{-1}\left[\frac{\alpha}{2}\cdot(1+\xi^2)\right]-m}{s}, & if \quad \alpha<\dfrac{1}{1+\xi^2}\\ \dfrac{-\xi\cdot f_{\nu}{}^{-1}\left[\frac{1-\alpha}{2}\cdot(1+\xi^{-2})\right]-m}{s}, & if \quad \alpha\geq\dfrac{1}{1+\xi^2}\end{cases} \tag{41}$$

Where α represents the quantile, ξ is the skewness parameter and $f_{\nu}{}^{-1}(\cdot)$ is the standardized quantile student-t function as stated in equation (40), and m and s^2 represent the non-standardized mean and variance of the skewed-student distribution:

$$m=\frac{\Gamma\left(\frac{\nu-1}{2}\right)\cdot\sqrt{(\nu-2)}}{\Gamma\left(\frac{\nu}{2}\right)\cdot\sqrt{\pi}}\cdot\left(\xi-\frac{1}{\xi}\right) \tag{42}$$

$$s^2=\left(\xi^2-\frac{1}{\xi^2}-1\right)-m^2 \tag{43}$$

Moreover, this dissertation computes the Skewed-Student Value-at-Risk in a similar manner as RiskMetrics, i.e. using EWMA volatility estimates described in equation (35) and modelling the α-quantile VaR as follows:

$$\mathrm{VaR}_{t+1}^{\alpha}=-\mu_t-\sigma_t\cdot skst^{-1}{}_{\xi,\nu}(\alpha) \tag{44}$$

In addition, Fernandez & Steel (1998) introduced a location scale skewed-student CDF, which constituted the pilar of the conditional mean and conditional variance skewed-student CDF brought up by Lambert & Laurent (2001). Furthermore, Trottia & Ardia (2016) have re-parametrized this last CDF, in order to create the standardized skewed-student CDF used in this dissertation and stated below.

$$skst_{\xi,\nu}(z_t)\begin{cases}\dfrac{2}{\xi+\frac{1}{\xi}}\cdot\left[\xi\cdot f_\nu\left(\dfrac{1}{\xi}\cdot(s\cdot z_t+m)\right)+\dfrac{1}{\xi}\right]-1, & if \quad z\geq-\dfrac{m}{s}\\ \dfrac{2}{1+\xi^2}\cdot f_\nu[\xi\cdot(s\cdot z_t+m)], & if \quad z<-\dfrac{m}{s}\end{cases} \tag{45}$$

Where $f_\nu(\cdot)$ is the standardized student-t CDF described in equation (40) and z_t represents the standardized returns as defined in equation (37).

Finally, in order to test this risk model using the Saddle-point approach, one needs to normalize the returns in the same way as in the previous section. In this sense, taking into account the CDF stated by equation (45), the entire sample of returns can be normalized resorting again to equation (31), thus allowing for the Saddle-point analysis to be carried out.

Nevertheless, regardless of the distribution chosen in order to better reflect financial returns, the parameters choice represents a huge responsibility for risk managers, who take the fall if the model ends up failing to predict reliable forecasts.

5.4. Historical Simulation

Historical Simulation is perhaps one of the oldest ways of measuring VaR and, according to many authors such as Pritsker (2006) and Pérignon & Smith (2010), it is also the method commercial banks resort to more often. This is mostly due to the fact it does not require any assumption in order to be put forward, which takes the weight of responsibility off risk managers’ shoulders, regarding the choice of distribution’s parameters.

The Historical Simulation method is quite simple to understand and implement, since one only needs to extract the α-quantile of an empirical time series of portfolio returns. Hence, it is a pure non-parametric model, which means that neither assumptions regarding the distribution of returns, nor of the volatility dynamics, need to be made.

Nonetheless, despite its simplicity, Historical Simulation suffers from a few pitfalls that must be considered when being applied. For instance, HS assumes the past will always repeat itself in the future, thus neglecting the future possibility of occurring higher losses than in the sample test period. Moreover, Barone-Adesi et al (1999) bring our attention to the fact that this approach is inappropriate unless the returns are actually independent and identically distributed, condition that is not always verified in practice in financial markets. And finally, this methodology attributes the same weight to every observation in the sample, which means that when some extreme loss entries the sample, it increases the following risk forecasts for a long time, until one day it no longer counts for its calculation, which is especially worrying for large sample sizes.

Furthermore, similarly to the VaR calculation procedure using parametric models, this dissertation considers a rolling window that is moved forward every day until the end of the test period. In addition, when using Historical Simulation, the size of the window constitutes an important choice that may ultimately determine whether the model fails or succeeds.

For large data sets, the model is not very precise in capturing current market conditions, whereas for small samples it is unlikely to accurately predict the empirical distribution of returns. Thus, the analysts face a trade-off between, on the one hand, big samples with good forecast accuracy, but poor representation of current market conditions; and on the other hand, small samples that, although being more likely to capture actual market movements, may fall short in modelling extreme quantiles correctly.

Overall, in comparison to the number of assumptions and parameters choice inherent to fully-parametric methods, this approach entails much smaller responsibilities for risk managers.

5.5. Volatility Adjusted Historical Simulation

Although Historical Simulation has several up sides compared to parametric Value-at-Risk, its shortcomings incentivised new approaches trying to overcome them, amongst which the Volatility Adjusted Historical Simulation VaR model.

Seeking to transform clustered and autocorrelated financial time series into i.i.d. returns, Hull & White (1998) modelled volatility changes over time through EWMA and GARCH processes, which they then used to smooth the series of returns, and extract α-quantiles from that returns series. This variant of Historical Simulation is denominated in this dissertation as Volatility Adjusted Historical Simulation.

Furthermore, an empirical study conducted by Pagan & Schwert (1990) suggest that the exponential GARCH– or EGARCH (1, 1) – tends to capture volatility dynamics more efficiently than the GARCH model, for it regards the relation between volatility and past financial returns as being asymmetric. Therefore, instead of EWMA or GARCH processes, this dissertation resorts to the EGARCH (1, 1) methodology with normal gaussian innovation process. This volatility model was firstly introduced by Nelson (1991), and is stated below as described by Pagan & Schwert (1990)

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{j=1}^{q} \beta_j \cdot \ln(\sigma_{t-j}^2) + \sum_{k=1}^{p} \alpha_k \cdot \left[\theta \cdot \psi_{t-k} + \gamma \cdot \left(|\psi_{t-k}| - \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \right) \right] \tag{46}$$

Where $\psi_t = \frac{u_t}{\sigma_t}$ and $u \sim N(0, \sigma_t^2)$, while the EGARCH parameters β, α, and γ are calculated through maximum likelihood.

Afterwards, instead of attributing different weights to the returns, the volatility estimates of the return series are adjusted so that the whole sample reflects current market conditions.

$$r_{t,j}^* = \frac{\sigma_{T,j}}{\sigma_{t,j}} \cdot r_{t,j} \tag{47}$$

Where $r_{t,j}$ is the realized return in the market at day t, $\sigma_{T,j}$ represents the volatility estimate for day T – the most recent day in the sample – and $\sigma_{t,j}$ stands for the volatility estimate at time t, which is prior to or coincidental with day T.

Once the steps above are fulfilled, Value-at-Risk can be computed by extracting the α-quantile of the smoothed portfolio returns $R^* \in \{r_{t,j}^*, r_{t-1,j}^*, \dots, r_{t-T,j}^*\}$, just like in the simple Historical Simulation.

On the one hand, empirically, Vol. Adj. Historical Simulation is shown to improve VaR forecasts according to Pritsker (2006), who concludes that this method overcomes the non-responsiveness of traditional Historical Simulation to changes in volatility. On the other hand, the same author highlights two flaws in this methodology: the first one regards to non-proper modelling of the time-varying correlations between returns; and the second one concerns to the same trade-off inherent to the choice of the data window's length described in the previous sub-section.

In conclusion, although Vol. Adj. Historical Simulation makes assumptions when estimating the volatility series used to smooth the sample returns, it does not parametrize the returns' distribution, thus belonging to the semi-parametric models' category.

5.6. Saddle-Point Application to Non-Parametric Distributions

Since Wong has developed the Saddle-point backtest, he has extensively tested it empirically. For instance, the author applies it to the EUR/USD exchange market in Wong (2009), and to several market indices in Wong et al (2012). However, at every point in the analyses, the risk models considered were fully parametric. Furthermore, to the best knowledge of the author of this dissertation, to this date, no author has published any article applying Wong's Saddle-point backtest to non-parametric risk models. This way, given how widespread non-parametric models are among financial institutions, this dissertation fills an important gap in the Saddle-point analysis.

According to Wong (2008), a non-normal distribution assumed by the risk model may be transformed into a normal distribution by applying equation (31), where the empirical returns are homogenised into percentiles through the risk model's CDF, and then normalized through the inverse normal CDF. Following this reasoning, the same transformation may be performed for non-parametric risk models if, instead of some parametric distribution's CDF, one takes into account the empirical cumulative distribution of the sample and then, similarly to equation (31), transform the empirical non-parametric quantile into a standard normally-distributed variable.

Further ahead, sections 6.2.4 and 6.2.5 provide a detailed description of how to apply the previous procedure to the non- and semi-parametric risk models employed in this dissertation.

6. Data and Methodology

Now that both the risk models and backtests have been specified and theoretically discussed in the previous sections, it is time to apply them empirically to the financial markets, in order to assess how well these behave in different circumstances.

6.1. Data description

With that goal in mind, this dissertation applied all five risk models to ten distinct market indices, each one representing either some economic activity sector or a determined country or region of the globe:

- NIKKEI225, the Japanese Market Index that includes the 225 main companies in the Tokyo Stock Exchange;
- S&P500, the USA Market Index that includes the 500 main companies in the New York Stock Exchange;
- FTSE100, the British Market Index that includes the 100 main companies in the London Stock Exchange;
- MSCI World Index, a broad global equity index that represents large and mid-cap equity performance across 23 developed markets countries, covering approximately 85% of the free float-adjusted market capitalization in each country;
- DAX30, the German Market Index that includes the 30 main companies in the Frankfurt Stock Exchange;
- S&P/TSX60, the Canadian Market Index that includes the 60 main companies in the Toronto Stock Exchange;
- Dow Jones, the USA Industrial Index that includes the 30 main companies operating in the Industry sector in the New York Stock Exchange;
- NASDAQ, the USA Technological Index that includes the 100 main companies operating in the Technology sector in the United States of America;
- AEX, the Dutch Market Index that includes the 25 main companies in the Amsterdam Stock Exchange;

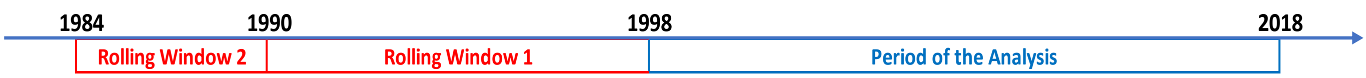
- HSI, the Hang Seng Market Index that includes the 50 main companies in the Hong Kong Stock Exchange.

Moreover, the empirical analysis was performed throughout 20 years of financial data from January 1st, 1998 until December 31st, 2017. However, it was necessary to obtain data prior to 1998 in order to apply a proper rolling window for the computation of the risk models. Therefore, the total length of data that has been collected from the Bloomberg Platform for each of the market indices specified above, was 8870 days, from January 2nd, 1984 to December 31st, 2017.

The length of data chosen to carry out the analysis is believed to allow an efficient assessment of the actual performance of the Saddle-point backtest and the traditional framework, since it covers high volatility periods and extreme events, as well as calm moments with no abrupt market movements.

Figure 2: Time Frame

This figure illustrates the length of the financial data used throughout the empirical analysis, as well as how has the data been divided in order to compute the VaR figures according to the different risk models. Rolling Windows 1 and 2 have 2000 and 1653 observations, respectively.



As shown in figure 2, although the analysis is put forward for the 20-year time period between 1998 and 2018, it has also been necessary to use two rolling windows, whose applicability will be further detailed in the next sub-section.

6.2. Methodology

This dissertation had its entire practical application carried out using Microsoft Excel and RStudio software, which includes the computation of both the risk models and backtesting methodologies.

First of all, given that the scope of this dissertation is to evaluate the Saddle-point backtest in opposition to the traditional backtesting framework, it only makes sense to compare it directly to Kupiec's Unconditional Coverage test, and not to Christoffersen's Independence and

Conditional Coverage tests. As pointed out in section 3, similarly to Wong's Saddle-point backtest, Kupiec's test centres its analysis on the risk coverage showed by the VaR model, whereas Christoffersen's backtests consider the independence of the exceedances, hence belonging to a different class of backtests.

Nevertheless, although the Saddle-point backtest assesses the performance of a VaR model based on its risk coverage – just as Kupiec's test does – these backtests differ significantly since the former accounts for the exceedance size, while the latter considers solely its frequency. Consequently, the goal of the current analysis is to put both tests against each other in several markets and under different economical environments.

In this sense, the analysis put forward by this dissertation follows the approach detailed in the remainder of this section.

First of all, the financial returns have been computed equally across all market indices and VaR methodologies, i.e. arithmetically calculating the returns from the adjusted (to dividends) closing prices – P_i –, as shown by the equation below.

$$r_n = \frac{P_n - P_{n-1}}{P_{n-1}} \tag{48}$$

Furthermore, the Saddle-point backtest requires the standardization of the market returns and the respective VaR figures, implying the computation of the mean and standard deviation parameters for every risk model. Hence, this dissertation has computed both parameters using a rolling window of 2000 days prior to January 1st, 1998, staring on May 3rd, 1990. This rolling window is specified in Figure 2 as "Rolling Window1". What's more, the standardization is carried out both for the arithmetic returns and VaR figures, respectively, as follows:

$$z_t = \frac{r_t - \mu_t}{\sigma_t}; \quad VaR_t^* = \frac{VaR_t - \mu_t}{\sigma_t} \tag{49}$$

Moreover, in order to apply the Saddle-point backtest, these measures go through the normalization process described in equation (31). However, the procedure for the computation of the TR-statistic is different between parametric and non-parametric risk models.

On the one hand, when applying Wong's test to parametric risk models, the Tail-Risk-of-VaR is computed by, firstly, subtracting each standardized risk measure by the homologous standardized return every time an exceedance occurs – $Min(VaR_t^* - z_t; 0)$ –, and then calculate its average as equation (16) indicates. On the other hand, if the Saddle-point backtest is to be applied to non-parametric risk models, there is no need to standardize the VaR figures. As explained in sections 5.4. and 5.5., Historical Simulation and Vol. Adj. Historical Simulation VaR forecasts are computed through the 1% empirical quantile. Therefore, given that the Saddle-point backtest assumes that the market follows a standard normal distribution, one may simply use the 1% standard normal quantile for every day in the test sample for the computation of the TR-statistic – $Min(\Phi^{-1}(0.01) - z_t; 0)$.

The corresponding p-value of the test is obtained through Microsoft Excel's tool 'Solver', as explained in further detail in sub-section 6.2.6.

Additionally, the traditional backtests have all been carried out in Microsoft Excel. In fact, these were performed almost exactly as they have been described throughout section 3, except for the cases where the number of exceedances registered during the year were zero. In these situations, the Unconditional Coverage backtest cannot be applied exactly as theoretically stated by Kupiec (1995), which thwarts the empirical analysis. Therefore, instead of zero, whenever such occurrence comes across, it is considered a marginal value very close to zero for the exceedances – 0.0000001 – in order to overcome this obstacle. This way, for a 260-day sample (more or less a working year), Kupiec's test ends up rejecting, at the 95% confidence level, models showing zero exceedances during the test period.

Hereinafter, one by one, each sub-section is going to briefly describe the practical application of all five risk models chosen to incorporate the empirical analysis at this dissertation, whose theoretical details can be found in section 5.

Finally, the last sub-section focuses on the Saddle-point's p-value calculation.

6.2.1. RiskMetrics application

Rewinding to its theoretical explanation, the RiskMetrics model required the computation of two parameters: mean and standard deviation. Moreover, the standard deviation follows an Exponentially Weighted Moving Average process, with lambda parameter fixed at 0.94 – for more details, see section 5.1. This way, all the computations have been performed in Microsoft Excel, using Rolling Window 1 to estimate the (normal gaussian) distribution's parameters.

In addition, VaR figures have been estimated according to the RiskMetrics methodology through Microsoft Excel, using the process and formulas described in section 5.1.

6.2.2. Student-t VaR application

According to its theoretical explanation, the student-t model underlines the computation of three parameters: mean, standard deviation and degrees of freedom. Besides that, the standard deviation is set as an Exponentially Weighted Moving Average process, with lambda parameter fixed at 0.94 – for more details, see section 5.1. Therefore, although most of the computations have been performed in Microsoft Excel, the number of degrees of freedom had to be computed for every day in the sample through Maximum Likelihood using RStudio software [2]. In addition, every parameter of the distribution was calculated considering Rolling Window 1 described in Figure 2.

Finally, every step of VaR calculation through the student-t quantile was performed in Microsoft Excel, relying on the process and formulas described in section 5.2.

6.2.3. Skewed-Student VaR application

In line with its theoretical explanation, the skewed-student model takes the computation of four parameters: mean, standard deviation, degrees of freedom and skewness. On top of that, the standard deviation is set as an Exponentially Weighted Moving Average process, with lambda parameter fixed at 0.94 – for more details, see section 5.1. Thus, the mean and standard deviation have been calculated in Microsoft Excel using Rolling Window 1, while the degrees of freedom and skewness parameters had to be computed using RStudio software[3], since its estimation through Maximum Likelihood is too computationally intensive to be carried out by Microsoft Excel. What's more, these last two parameters have also been estimated using Rolling Window 1.

In turn, VaR calculation through the skewed quantile was carried out solely in Microsoft Excel using the process and formulas described in section 5.3.

[2] The R codes used for the computation of the degrees of freedom are made available by this author on request.
[3] The R codes used for the computation of the degrees of freedom and skewness parameters are made available by this author on request.

6.2.4. Historical Simulation application

In accordance with its theoretical explanation, the Historical Simulation model does not require any parameter estimation. However, the standardization needed to perform Wong's test implies the calculation of the mean and standard deviation, which have been carried out using Rolling Window 1. Furthermore, both the empirical quantile function and empirical CDF have been computed in Microsoft Excel as follows:

- The empirical quantile function is available at Microsoft Excel under the name "quantile";

- The empirical CDF of a series of financial returns can be obtained through a simple two-step operation: first, count the number of days in the sample whose return was higher than or equal to the most recent return; and second, divide this number by the total days in the sample – in this case it has been considered a sample of 2000 observations.

However, by computing the cumulative distribution function as described above, there may occur some situations in which the CDF returns a value of 0% or 100%, thus preventing the transformation to the normal distribution, since the 0%- and 100%-quantile of the normal distribution returns a value of $-\infty$ and ∞, respectively. Therefore, in order to overcome this obstacle, when such cases come across it is summed up or subtracted a marginal value from the CDF – considering 0.0000001% instead of 0% and substituting 100% for 99.99999% – so that the transformation could be performed.

All in all, even though the previous procedure may change the true results of the extreme edges in the tails of the distribution, it is the only way to apply Wong's test to non-parametric models.

Moreover, Historical Simulation VaR has been computed in Microsoft Excel following the process described in section 5.4.

6.2.5. Volatility Adjusted Historical Simulation application

As pointed out during its theoretical overview, Volatility Adjusted Historical Simulation adjusts the realized returns to current volatility trends using, for that matter, a volatility generating process – in this dissertation it has been chosen an EGARCH (1,1) with normal innovation process. Hence, in order to compute daily volatility estimates through this method, this dissertation resorted to RStudio software[4]. Additionally, this VaR methodology requires volatility estimates for the entire sample used to apply the risk models, which in this case starts 2000 days before the analysis period, on May 3rd, 1990. This way, there has been employed Rolling Window 2 – which is illustrated in Figure 2 –, a 1653-day length rolling window, starting on January 2nd, 1984, so that the EGARCH volatility estimates could be accurately computed for every day in the sample.

What's more, every calculation besides the EGARCH volatility estimates have been carried out in Microsoft Excel and, as to the CDF calculation, the procedure is exactly the same as the one for the Historical Simulation model.

Finally, VaR estimates have been produced by following every step detailed throughout section 5.5.

6.2.6. Capital Risk Charge

Additionally, in order to improve the comparative analysis of the risk models and backtests carried out by this dissertation, it has been calculated the daily Capital Risk Charge figures as detailed by Basel II Agreements, for every risk model and across all market indices. To do so, BCBS's Traffic Light backtest has been performed exactly as stated in sub-section 3.1.1., using a 250-day sample. This way, these Capital Risk Charges have been computed for every year of the sample analysis except 1998.

6.2.7. Saddle-point p-value application

In contrast to the traditional backtests that follow a Chi-square distribution for its p-value calculation, the Saddle-point's p-value requires more complicated steps in order to be determined, all of which are explained in detail in sub-section 4.2.2. In that sub-section, it has been suggested that the condition stated by equation (26) must be satisfied in order to apply the Lugannani & Rice (1980) formula. However, there is no exact solution for the equation, only close approximations.

Therefore, this dissertation used the 'Solver' tool in Microsoft Excel to minimize the condition stated below, subject to upper and lower boundaries of the equation.

[4] The R codes used for the computation of the volatility estimates through EGARCH (1, 1) model with normal innovations, are made available by this author on request.

$$|K'(\varpi) - \bar{x}| \tag{50}$$

Where $|\cdot|$ represents the absolute value and, after a "trial and error" process, ϖ was found to belong to the interval $[-3.9; 35.15]$.

Overall, by minimizing equation (50), subject to ϖ's upper and lower boundaries, one gets the optimal ϖ parameter that ultimately allows the computation of the Saddle-point p-value through equation (27).

7. Empirical Analysis

After applying all five risk models described in section 5 to ten market indices, the corresponding backtesting results have been summarized in Tables 3-7 in the Appendix.

For instance, Table 3 summarizes the UC and Saddle-point tests' results for the RiskMetrics model, in the form of p-values. Moreover, it is considered three different lengths for the test sample: twenty 1-year tests, four 5-year tests and one 20-year test. Likewise, Tables 4-7 display the same information for the remaining risk models.

Moreover, Table 2 not only allows to assess the rejection behaviour of both backtesting procedures carried out in this dissertation, but also provides an outlook on the yearly performance of each risk model according to both backtests, just as described below.

Table 2: Backtests' frequency of rejection

This table includes (by descending order) the following items: the frequency with which a risk model has not been rejected by any of the tests (row 1); how often has one test rejected a given risk model when the other has not (rows 2 and 3); the number of occasions where both backtests have rejected a risk model, but one of them has rejected it at a greater confidence level (rows 4 and 5); and, finally, how many times have the p-values of both tests been the same (last row).[5]

p-value		Risk Models					
UC	Saddle-point	RiskMetrics	Student-t VaR	Skewed-Student VaR	Historical Simulation	Vol. Adj. Historical Simulation	TOTAL
>10%	>10%	69	115	133	61	123	501
<10%	>10%	15	25	13	7	7	67
>10%	<10%	32	24	27	3	29	115
< Saddle-point	<10%	32	18	17	107	33	207
<10%	< UC	51	17	10	5	8	91
UC =	Saddle-point	1	1	0	17	0	19

In this sense, this section will be split into two parts: the first one includes the analysis of Wong's Saddle-point backtest in contrast to Kupiec's Unconditional Coverage test; and the second part provides an overview on the general performance of the risk models.

[5] These items represent 1-year sample test analysis.

7.1. Backtesting Analysis

As the scope of this dissertation is to analyse the performance of Wong's Saddle-point backtest in contrast to Kupiec's Unconditional Coverage test, this sub-section explores the results obtained from the empirical parsing carried out as described in section 6.

First of all, it is a note-worthy mention that, as Table 2 indicates, from all 1-year sample test results obtained throughout the empirical analysis, both backtests have unanimously accepted a risk model about half of the times. In fact, from the entire 1-year analysis – which totals 1000 test results – only 182 times have these backtests differed from each other when it came to a rejection decision. Amongst these occasions, there were 67 times where Kupiec's test decided to reject a risk model against Wong's backtest verdict, while the inverse situation was registered 115 times.

Figures 3 and 4 in the Appendix illustrate the exceedances that took place over the year of 2014 in the British and Canadian markets, respectively. While the former considers the Skewed-Student VaR model, the latter refers to the Student-t VaR model. These figures depict two similar situations, taking place in two different markets and considering two different risk models. As observed, both situations have registered plenty exceedances during one single year, having both risk models failed to predict market risk more than 3% of the times during that length of time, which is more than 3 times the expected coverage rate. Therefore, Kupiec's test has rejected the risk model at the 99% confidence level in both situations, given that it bases its backtesting results solely on the coverage rate. However, Wong's Saddle-point considers the sizes of the exceedances as the main object for risk model assessment, which are shown in Figures 3 and 4 to be quite small. Hence, even though there are several exceedances in the sample, Wong's test ended up accepting the VaR model in both cases.

On the other hand, Figures 5 and 6 in the Appendix display inverse situations to the previous ones. Figure 5 depicts the exceedances that occurred throughout 2011 in the Japanese market, considering Student-t VaR model, whereas Figure 6 summarizes the exceedances taking place during the year of 2000 in the Hong Kong market, considering RiskMetrics model. As pointed out by these Figures, in both cases the VaR has been breached four times in a year, which corresponds to a 1.54% exceedance rate. Given the features of Unconditional Coverage test, it comes as no surprise that it accepts the risk model in either case. In contrast, Wong's backtest is shown to reject the VaR model in both situations, for the sizes of the exceedances are pretty large. Although there are only a few of them, the exceedances in these cases have enough dimension to lead to the rejection of the risk model according to the Saddle-point methodology.

Additionally, based on Tables 3-7 in the Appendix, 5-year sample testing results show that backtesting results present similar findings for parametric risk models to the ones stated in Table 2. Albeit there can be cited some cases where backtests' resolutions differed, the bottom line is that UC and Saddle-point seldom present opposite results across market indices for every parametric VaR model. The exception, however, lies on non- and semi-parametric risk models, whose test results are summarized in the bottom rows of Tables 6 and 7, respectively.

According to the empirical results for the simple Historical Simulation model displayed in Table 6, during overall calmer periods such as 2003-2007 and 2013-2017, Kupiec and Wong's backtesting results tend to diverge considerably, whereas for periods of financial distress, e. g. 1998-2002 and 2008-2012, these two backtests present similar conclusions with regard to the feasibility of this non-parametric model.

The previous finding is supported by theoretical proof, for Kupiec's test rejects risk models for having too few exceedances in the sample, contrarily to Wong's Saddle-point backtest, in which the fewer and smaller the exceedances – provided that there are exceedances in the sample – the likelier it is for the test to accept the risk model. Thus, because Historical Simulation is very responsive to large shifts in the market – as detailed in section 5.4. –, its VaR figures tend to suffer sharp and long-lasting rises when an extreme loss occurs. Consequently, these Historical Simulation VaR figures are bound to remain quite large for long periods of time after the shock, especially if the rolling window used to calculate VaR is large – which is definitely the case of this empirical analysis. Therefore, when the markets go through turmoil periods, the number and size of the exceedances are expected to rise as a consequence of the financial distress, leading to a likely rejection of the risk model by the Unconditional Coverage and Saddle-point backtests. But when the markets cool down, VaR forecasts tend to become too conservative in comparison to the realized returns, resulting almost certainly in very low exceedance rate – i. e. the relative number of exceedances in the sample – and, ergo, in the rejection of the risk model by the UC test for being too conservative. In turn, the Saddle-point backtest returns the opposite feedback, due to its inability to reject a risk model for having too few exceedances – provided that there are exceedances in the sample, and that these are not too large.

The situation discussed above illustrates one of the main differences between the traditional backtesting framework and the Saddle-point backtest. On the one hand, at the eyes of the regulatory entity, whose only concern is to ensure that financial institutions are properly protected against adverse scenarios, Wong's backtest is seen as a preferred methodology to be employed. On the other hand, commercial banks and other financial institutions see in Capital Requirements an opportunity cost imposed by the BCBS, making its minimization one of their main goals. For this reason, financial institutions may regard traditional backtests as a preferential choice due to its rejection properties.

Nevertheless, even though commercial banks continuously seek the minimization of the Capital Risk Charges established by Basel II Agreements, these should also consider their exposure to extreme events. There is a generalized belief that Capital Requirements are so conservative that banks are not usually worried that it could be surpassed at any time, for it sets a "security bar" extremely high – see equation (1). However, as Figures 7-10 in the Appendix point out, there could be times when Capital Requirements are not large enough to prevent the financial institution from going bankrupt.

Figures 7-10 in display each of these situations through a graph, containing all year's realized returns, VaR figures and Capital Risk Charge numbers. In addition, below each graph there is a small table with the p-values of Kupiec and Wong's tests, for all three test samples length – 1, 5 and 20 years.

On March 15th, 2011 the Japanese market index NIKKEI225 registered a loss of 10.55%, the third largest loss in its history, due to the earth-quake and tsunami that stroke the country just two days before. As previously stated, from all three parametric risk models used in this empirical analysis, none has been able to produce sufficiently large Capital Risk Charge figures to face the post-tsunami crash loss on this day. Moreover, Figures 7, 9 and 10 display this occurrence – specified on the circled area in the graphs – for each of the risk models, that were not able to produce risk forecasts that would prepare the financial institution to such market loss.

This way, it is very important that the backtesting procedure carried out by commercial banks, whatever it may be, has the power to reject these models, at least for the year of 2011. However, as one may observe in Figures 7, 9 and 10, Kupiec's test does not reject any of the models for NIKKEI225 in 2011, for the exceedance rate is considered acceptable by this frequency-based backtest. On the contrary, Wong's Saddle-point backtest rejects all three models for the Japanese market index in 2011, with a confidence level of 99%.

Besides the Japanese post-tsunami crash event, there has only been registered one other super-exceedance throughout the entire empirical analysis, taking place in the German market index on January 21st, 2008. In this occasion, the loss observed in the market rounded 7.16%, which could not be covered by Student-t VaR-based Capital Risk Charge. As shown in Figure 8, this time the Unconditional Coverage backtest has actually been able to reject this risk model in 2008, although not at the 1% significance level, but rather at the 5% instead. As to the Saddle-point test, it has again rejected the risk model at the 1% significance level.

Additionally, considering a 5 and 20-year test sample, the Saddle-point backtest has also managed to reject all three parametric risk models referred above, at a minimum 95% confidence level for the NIKKEI225 market index in 2011. In contrast, considering the super-exceedance of Student-t VaR model in the German market, Figure 8 suggests that Wong's test failed to reject it for test samples larger than 1 year of financial data.

Overall, Figures 7-10 in the Appendix documented empirical evidence supporting the idea that the Saddle-point backtest has greater power to reject inaccurate models than Kupiec's UC test, given its ability to reject risk models containing super-exceedances. Nonetheless, it is interesting to assess whether the Saddle-point backtest would also have rejected the same risk models, if the test sample had contained the super-exceedance alone, i. e. if every other exceedance in the sample simply did not exist. In this sense, Figures 11-13 in the Appendix, illustrate these exact scenarios for the 2011 super-exceedances in the Japanese market for each of the parametric risk models, showing that only in the case of RiskMetrics – displayed in Figure 13 – has the Saddle-point backtest rejected the model when the super-exceedance has been left alone in the test sample. The same does not happen for the Student-t and Skewed-Student VaR models. Therefore, albeit Wong's test has rejected models containing super-exceedances in its sample throughout this empirical analysis, it may not do so every time this situation comes across. In addition, the larger the test sample used to carry out backtesting analyses, the likelier it is for super-exceedances to go unnoticed.

Moreover, Table 7 indicates that the conclusions drawn by the backtests for Vol. Adj. Historical Simulation model tend to differ for 5-year sample tests and, to some extent, even for 1-year sample as well. In fact, according to 1-year sample backtesting analysis summarized in Table

2, Wong's test rejects this semi-parametric model 29 times without the consensus of Kupiec's UC test, which is more than four times the frequency of opposite situations. At first sight, one might reason that, in such cases, the few exceedances recorded throughout the test sample are considerably large, which assumes a significant Tail-Risk-of-VaR that ultimately leads to the rejection of the model according to Wong's Saddle-point test, just like the cases depicted in Figures 5 and 6. However, after taking a closer look to the results in Table 7, it is possible to identify a different motive to explain this phenomenon. Figure 14 in the Appendix displays the Vol. Adj. Historical Simulation exceedances recorded in 2014 for the AEX Market Index, along with the backtesting results – which dictated the acceptance by Kupiec's test and a clear rejection by Wong's test. In contrast to the cases illustrated by Figures 5 and 6, this figure shows that none of the exceedances in this situation seem to be that large. What's more, by displaying the Tail-Risk-of-VaR figures computed for each of the previously mentioned exceedances, Figure 15 in the Appendix highlights important inconsistencies in the application of Wong's test to non-parametric models. Looking at the sizes of the exceedances that took place on March, 3rd and on October, 2nd in Figure 14, although the latter seems to be only slightly larger than the former, it is expected that the Tail-Risk-of-VaR measure does not differ too much amongst them. Nevertheless, as verified in Figure 15, the discrepancy between the Tail-Risk-of-VaR in both exceedances is extremely large.

The inconsistency underlined by Figures 14 and 15 can be explained by the process used for the calculation of the non- and-semi parametric CDFs. As explained in sections 6.2.4. and 6.2.5., the calculation of VaR figures through these non- and semi-parametric models, consists of extracting the 1%-quantile out of a returns sample at each day, and the computation of the empirical CDF consists of similar procedure. This way, there are some days where the realized return in the market is lower or higher than any other day in the rolling window's sample, thus generating empirical CDF values of 0% and 100%. In case the former situation comes across, the normalized return computed through equation (31) assumes a CDF value very close to 0% – 0.000000001% – in order to perform the Saddle-point analysis. Nonetheless, the normalized return in such cases is still quite small, resulting in large Tail-Risk-of-VaR.

On the one hand, considering the simple Historical Simulation model, if the rolling window sample does not incorporate any big loss – which might happen during calm periods –, then the appearance of a slightly larger loss in the market leads to the situation described above, thus resulting in huge Tail-Risk-of-VaR measures without the existence of real significant empirical tail size. However, as observed in Tables 2 and 6, this kind of situation does not seem to take place very often, given that only trice in the entire 1-year sample analysis has the Saddle-point backtest rejected the Historical Simulation model while Kupiec's test accepted it. In this sense, there may be two motives why such situations are not usually identified by the backtesting analysis: first, as previously suggested, the Historical Simulation model performs so poorly in frequency-based backtesting analysis that it is rejected the majority of the times, allowing for these situations to pass unremarked in a comparative analysis between these two backtests; and second, even though extreme losses may occur in the market, these are not common at all, and if the appearance of a market loss higher than any other loss in the rolling window sample is already a rare event, then the appearance of loss that is only slightly larger than the worst loss in the sample is even rarer.

On the other hand, as described in sections 5.5. and 6.2.5., the Vol. Adj. Historical Simulation model computes its VaR figures and CDF in a distinct way, given that instead of extracting the 1%-quantile directly from the sample of returns, these are previously adjusted to current volatility estimates generated through an EGARCH process with normal innovations. Therefore, during calm market periods, the returns that take part of the sample are smoothed down by low current volatility estimates, but when the market suffers wide shifts the opposite effect occurs. Figure 14 in the appendix suggest that, in the beginning of 2014, the Dutch market index was going through a calm period, with no relevant shifts occurring for quite some time, for there were neither big losses taking part in the rolling window, nor high EGARCH volatility estimates for the market at that point. On January 24th, the market has registered a loss of 2.49% that, although it cannot be regarded as a huge loss, it was higher than any other adjusted (to current volatility) loss in the rolling window up to that point. In addition, before the year ended, a similar situation occurred on October 2nd, in which a 2.54% loss surpassed every adjusted return in the rolling window sample. What's more, only a couple of weeks later, on October 15th, has a 3.46% loss taken place that, although it was higher than the previous exceedance-generating losses, it was not ranked below the worst adjusted return in the sample, due to higher volatility estimate for current volatility on that day. Hence, when calculating the Vol. Adj. Historical Simulation CDF for the most recent return in the market, it is not necessary for this return to be lower than every other return in the sample in order to have a 0% CDF value. As long as the volatility estimates are low enough, the adjusted returns in the sample decrease to a point where the realized return is lower than all the adjusted returns in the sample, which confirms that the volatility estimates play a decisive role in this process.

All in all, the lack of accuracy in measuring Tail-Risk-of-VaR for non-parametric models exposed above, jeopardizes the application of Wong's Saddle-point backtest. In particular, this dissertation shows that this drawback gains even greater impact for the Vol. Adj. Historical Simulation model. In turn, although the simple Historical Simulation model shares the same problem, it has much smaller impact empirically.

7.2. Risk Models Performance

The assessment to the performance of the parametric risk models is made through the analysis of both backtests, whereas given the information provided in the previous sub-section, only Kupiec's test results were contemplated for the non- and semi-parametric risk models.

First of all, after a quick glance over Tables 2-7 it is quite straightforward to suggest that, according to Kupiec's Unconditional Coverage and Wong's Saddle-point backtests, the Skewed-Student VaR model – whose test results are represented in Table 5 – outperforms the other risk models. In fact, such remark was, to some extent, already expected a priori to the empirical application of the risk models, for several authors, such as Wong (2009) and Wong et al (2012), have already documented greater accuracy for the Skewed-Student VaR model. What's more, as depicted in Table 2, even though the Vol. Adj. Historical Simulation showed to predict market risk rather well, its forecasts seem to have come up short in comparison to the Skewed-Student VaR ones.

On the other hand, conversely to the good performance of the Skewed-Student VaR, Table 3 shows that RiskMetrics has failed both risk coverage tests in several occasions, and for almost every market. In fact, RiskMetrics has been rejected in every market index for the overall 20-year sample test, from 1998 to 2018, and for almost every 5-year sample test as well. Additionally, Table 2 suggests that the 1-year sample test results of RiskMetrics are also mediocre, which confirms the conclusions drawn by Pafka & Kondor (2008), who have studied the performance of the RiskMetrics methodology for the 95% and 99% confidence levels, having found that it shows significant lack of accuracy when it comes to the latter.

In turn, carrying out the same parsing to Table 4, Student-t VaR seems to render improved test results in comparison to RiskMetrics, especially for small length test samples. However, for the total aggregate period analysis accounting from 1998 to 2018, this risk model shows very similar test results to the ones of RiskMetrics, having been rejected throughout every market index, except by the Saddle-point backtest in the German market index DAX30. Nevertheless, as one may observe in Table 2, the fact that Student-t VaR accounts for fatter tails in its distribution, appears to constitute a more reliable alternative to the previous parametric model, especially when contemplating 1-year period analysis – which is the standard test period stipulated by Basel II Agreement.

Furthermore, it is relevant to highlight that Historical Simulation has not proven to constitute a reliable alternative to the parametric models, especially in comparison to the Student-t and Skewed-Student VaR models. In particular, the exceedance rates observed throughout the sample are quite bipolar, i. e. some years it is very low – in many occasions it is actually zero – whereas in other years it is quite large. From a 1-year and a 5-year sample analysis viewpoint, Tables 2 and 5 show that, more often than not, this VaR model does not seem to provide proper risk coverage. Nonetheless, considering the 20- and 5-year sample analyses illustrated in Tables 2-5, Historical Simulation does seem to constitute an improvement to the normally-distributed RiskMetrics, and even to the Student-t VaR model. However, except for the 20-year test sample, Skewed-Student VaR performs better than the non-parametric model.

In addition, Tables 2, 6 and 7 show that Vol. Adj. Historical Simulation performs better than the simple Historical Simulation model for every sample length, which goes in accordance to other authors findings, such as Pritsker (2006), who have gathered empirical evidence regarding the improvement in VaR forecasting brought up by this semi-parametric model, in comparison to the non-parametric Historical Simulation.

At last, throughout the entire risk analysis put forward by this dissertation, and considering the time period between January 1st, 1999 and December 31st, 2017, Figures 7-10 reveal that all three parametric VaR models have registered at least one super-exceedance in its sample, i.e. a day in which BCBS's Capital Risk Charge has been exceeded by a return in the market, leading to a bankruptcy event, while parametric models have not. Such remark may raise a doubt on whether Skewed-Student VaR really does outperform both non- and semi-parametric risk models, given the lack of ability to prevent bankruptcy shown by the former.

8. Conclusions

This dissertation's goal was to assess the quality of Wong's Saddle-point test in comparison to the traditional backtesting framework. Particularly, given that the Saddle-point is a risk coverage test, it has been put against Kupiec's Unconditional Coverage, for the latter is the only risk coverage backtest amongst the traditional backtesting methodologies.

To begin with, section 2 provided a theoretical contextualization of Wong's Saddle-point backtest on the broad subject of Risk Management, while sections 3 and 5 gave, respectively, a description of the main backtesting procedures and risk models carried out by financial institutions worldwide.

In turn, section 4 introduced Wong's Saddle-point theory as an alternative to the traditional frequency-based backtests, given that it accounts mainly for the size of the exceedances observed during the sample test period.

At last, once section 6 specified the data used for the implementation of the models along with the methodology employed for the empirical application of the concepts presented on the previous sections, the empirical analysis, whose main results are summarized in section 7, could finally be performed.

From the entire empirical analysis put forward by this dissertation, there are some interesting findings worth highlighting.

First of all, the most important remark is that Wong's Saddle-point backtest cannot be properly applied to non-parametric risk models, especially not to the Vol. Adj. Historical Simulation model. The fact that Wong's test requires the returns to be normally distributed, causes major drawbacks in calculating Tail-Risk-of-VaR through empirical CDFs, which does not allow the application of the Saddle-point backtest to both non- and- semi-parametric risk models mentioned in this dissertation.

On the one hand, for simple Historical Simulation VaR model, when the realized return at some day is even slightly lower (or larger) than every return in the rolling window, then Tail-Risk-of-VaR overestimates the size of the empirical distribution's left tail. Thus, the lower the rolling window's length used for the calculation of Historical Simulation VaR figures, the more frequent the previous drawback becomes.

On the other hand, Vol. Adj. Historical Simulation tends to overestimate the size of the empirical distribution's left tail whenever current volatility estimates are too small, leading to the shortening of the adjusted returns in the rolling window's sample, which exalts the number of situations described in the previous paragraph. Therefore, the calmer the market environment, the higher the risk of misrepresenting the tail's size.

As to the performance of both backtesting techniques for parametric risk models, when considering long test samples, such as 20 or even 5 years, the results do not vary much between these two methodologies, whether during quiet market moments or through turmoil periods.

However, for small sample backtesting the results tend to diverge in several occasions. Given the characteristics of both tests, their individual criteria for risk model evaluation sometimes leads to opposite results. On the one hand, during years when the VaR forecasts have been surpassed several times by realized market returns, Kupiec's test rejects the model, while the Saddle-point backtest may not reject it if the exceedances are small enough. On the other hand, when the exceedance rate is quite close to what is expected – in the case of this dissertation it is expected 1% exceedance rate – the Unconditional Coverage test confirms that the risk model provides proper risk coverage, whereas Wong's test could reject it in case the exceedances are too large.

Additionally, considering the results shown by Figures 7-10, one can conclude that Wong's Saddle-point backtest has greater power to reject non-proper risk models, especially for small test samples. This way, besides practically matching, by and large, Kupiec's test results, the Saddle-point seems to appear as a better alternative for backtesting parametric VaR models during crashes.

From a regulatory authority's viewpoint, such remarks suggest that Unconditional Coverage test should be replaced by Wong's Saddle-point when backtesting parametric models. Nonetheless, the role played by the traditional backtesting framework should not be minimized, given that the Saddle-point does not account for the number of exceedances as effectively as Kupiec's test. In fact, as pointed out by Wong (2008), the Saddle-point backtest has been designed to be complementary to the traditional backtesting framework, working as a size counterpart of the Basel II Agreement.

Furthermore, given the pitfalls of Wong's Saddle-point test, the conclusions drawn regarding non- and semi-parametric risk model's performance in this dissertation, consider solely Kupiec's test. In this sense, there are some relevant observations to point out on this matter.

Firstly, RiskMetrics struggles to provide accurate 1%-VaR forecasts, leading to the conclusion that the normal distribution does not capture financial markets' dynamics, especially during turmoil periods.

Secondly, although the other two parametric models – Student-t and Skewed-Student VaR – seem to get improved backtesting results, none of these has been able to prevent bankruptcy on the post-tsunami Japanese crash of 2011.

What's more, although the simple Historical Simulation managed to prevent super-exceedances, it shows some inconsistencies regarding its risk forecasts, for these tend either to underestimate market risk, or to be overconservative, thus coming short in producing reliable risk assessment.

Finally, the semi-parametric Vol. Adj. Historical Simulation seems to outperform every risk model in this dissertation, except for the Skewed-Student VaR – which shows slightly better test results. However, in contrast to the parametric model, the Vol. Adj. Historical Simulation does not register any super-exceedance anywhere throughout the entire empirical analysis. Overall, it is reasonable to conclude that the semi-parametric model showed a superior performance.

To conclude, this dissertation confirms Wong's framework as an important contribution to Risk Management, and to work quite satisfactorily as a complement to the frequency-based traditional backtests, but only when assessing parametric risk models. The Saddle-point methodology is shown in this dissertation not to be applicable to non- and semi-parametric VaR models.

This way, even though Wong's test can work as a reliable size counterpart to the Basel II rules for parametric risk models, the fact that it is not applicable to non-parametric VaR models make it an incomplete test, especially bearing in mind that, according to Pérignon & Smith (2010), Historical Simulation is the most used VaR model amongst financial institutions.

References

Atzner, P., Delbaen, F., Eber, J.-M. & Heath, D., 1999. *"Coherent Measures of Risk"*. **Mathematical Finance**, 9, 203-228.

Basel Committee on Banking Supervision, 1996. *"Amendment to the Capital Accord to Incorporate Market Risk"*.

Basel Committee on Banking Supervision, 1996b. "*Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*".

Barone-Adesi, G., Giannopoulos, K. & Vosper, L., 1999. *"VaR without correlations for portfolios of derivative securities"*. **Journal of Future Markets**. 19(5), 583-602.

Basak, S. & Shapiro, A. (1998). *"Value-at-Risk based Risk Management: optimal policies and asset prices"*. **Review of Financial Studies**, 14, 371-405.

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

Berkowitz, J., Christoffersen, P. & Pelletier, D., 2011. *"Evaluating Value-at-Risk models with desk level-data"*. **Management Science**, 57(2), 2213-2227.

Berkowitz, J. & O'Brien, J., 2002. *"How accurate are value-at-risk models at commercial banks?"*. **The Journal of Finance**, 2002 – Wiley Online Library.

Bollerslev, T., 1986. *"Generalized Autoregressive Conditional Heteroskedasticity"*. **Journal of Econometrics**. 31(3), 307-327.

Christoffersen, P., 1998. *"Evaluating internal forecasts"*. **International Economic Review**, 39, 841-862.

Colletaz, G., Hurlin, C. & Pérignon, C., 2013. *"The risk map: a new tool for validating risk models"*. **Journal of Banking and Finance**, 37(10), 3843-3856.

Daniels, H.E., 1987. *"Tail Probability Approximations"*. **International Statistical Review**, 55(1), 37-48.

Dias, A., 2013. *"Market capitalization and value-at-risk"*. **Journal of Banking and Finance**, 37(12), 5248-5260.

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

Escaciano, J. C. & Olmo, J., 2010. *"Backtesting parametric value-at-risk with estimation risk"*. **Journal of Business and Economic Statistics**, 28(1), 36-51.

Gaglianone, W.P., Lima, L.R., Linton, O. & Smith, D.R., 2011. *"Evaluating Value-atRisk models via quantile regression"*. **Journal of Business and Economic Statistics**, 29(1), 150-160.

Gerlach, R. H., Chen, C. W. S. & Chan, N. Y. C., (2011). *"Bayesian time-varying quantile forecasting for Value-at-Risk in financial markets"*. **Journal of Business & Economic Statistics**, 29(4), 481-492.

Giot, P. & Laurent, S., 2003. *"Value-at-Risk for Long and Short Trading Positions"*. **Journal of Applied Econometrics**, 18, pp. 641-664.

Halbleib, R. & Pohlmeier, W., 2012. *"Improving the value-at-risk forecasts: Theory and evidence from financial crisis"*. **Journal of Economic Dynamic and Control**, 36(8), 1212-1228.

Hull, J. & White, A. 1998. "*Incorporating Volatility Updating into the Historical Simulation method for Value-at-Risk*", **Journal of Risk 1** (Fall): 5–19.

Kerkhof, J. & Melenberg, M., 2004. *"Backtesting for risk-based regualatory capital"*. **Journal of Banking and Finance**, 28(8), 1845-1865.

Krause, A., 2003. *"Exploring the limitations of Value at Risk: How good is it in practise?"*. **Journal of Risk and Finance**. 19-28. Winter.

Kuester, K., Mittnik, S. & Paolella, M. S., 2006. *"Value at Risk prediction: A comparison of alternative strategies"*. **Journal of Financial Econometrics**. 4, 53-89.

Kupiec, P., 1995. *"Techniques for verifying the accuracy of Risk Management Models"*. **Journal of Derivatives**, 3, 73-84.

Lambert P, Laurent S. 2001. *"Modelling financial time series using GARCH-type models and a skewed Student density"*. Stat Discussion Paper; 0125 (2001) 21 pages. Mimeo, Universite de Liège

Lin, C. S. & Shen, S.S., 2006. *"Can the student-t distribution provide accurate value at risk?"*. **The Journal of Risk Finance**. Vol. 7, Issue:3, pp. 292-300.

Lucas, A. & Siegmann, A., 2008. *"The effect of shortfall as a risk measure for portfolios with hedge funds"*. **Journal of Business Finance and Accounting**, 35, 200-226.

Lugannani, R. & Rice, S.O., 1980. *"Saddlepoint approximation for the distribution of the sum of independent random variables"*. **Advanced Applied Probability**. 12, 475-490.

Lumsdaine, R.L., 1995. *"Finite-sample properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: a Monte Carlo investigation"*. **Journal of Business and Economic Statistics**, 1, pp.1-9.

Mandelbrot, B., 1963. *"The variation of certain speculative prices"*. **Journal of Business**, 36, 394-419.

McAleer, M., 2009. *"The ten commandments for optimizing value-at-risk and daily capital charges"*. **Journal of Economic Surveys**, 23(5), 831-849.

Morgan Guaranty Trust Company & Reuters Ltd, 1996. "*RiskMetrics—Technical Document"*, 4th ed., New York: Morgan Guaranty Trust Company.

Nelson, D. B. 1991, "*Conditional Heteroskedasticity in Asset Pricing: A New Approach*". **Econometrica**, 59, 347–370.

Nieto, M.R. & Ruiz, E. 2016. *"Frontiers in VaR forecasting and backtesting"*. **International Journal of Forecasting**. 32. 475-501.

Pafka, S. & Kondor, I., 2008. *"Evaluating the RiskMetrics methodology in measuring volatility and value-at-risk in financial markets"*. **Physica A: Statistical Mechanics and its applications**. Elsevier. Hungary.

Pagan, A. R., and Schwert, G. W. 1990, "*Alternative Models for Conditional Stock Volatility*," **Journal of Econometrics**, 45, 267–290.

Pérignon, C., Deng, Z.Y. & Wang, Z.Y., 2008. *"Do banks overstate their value-at-risk?"*. **Journal of Banking and Finance**, 34, 55-66.

Pérignon, C. & Smith, D.R., 2010. *"The level and quality of Value-at-Risk disclosure by commercial banks"*. **Journal of Banking and Finance**, 34, 362-377.

Pritsker, M., 2006. *"The hidden dangers of Historical Simulation"*. **Journal of Banking and Finance**, 30(2), 561-582.

Taasche, D., 2002. *"Expected Shortfall and beyond"*. **Journal of Banking and Finance**, 26(7), 1519-1533.

Venkataram, S., 1997. *"Value at Risk for a mixture of normal distributions: the use of Quasi-Bayesian estimation techniques"*. **Economic Perspectives - Federal Reserve**. Vol: 21(2). (1997). ISSN: 1048-115X.

Wong, W.K., 2007. *"Backtesting trading risk of commercial banks using expected shortfall"*. **Journal of Banking and Finance** 32 (2008), 1404-1415.

Wong, W.K., 2008. *"Backtesting value-at-risk based on tail losses"*. **Journal of Empirical Finance**. 17, 526-538.

Wong, W.K., 2009. *"Backtesting the tail risk of VaR in holding US dollar"*. **Applied Financial Economics**, 19, 327-337.

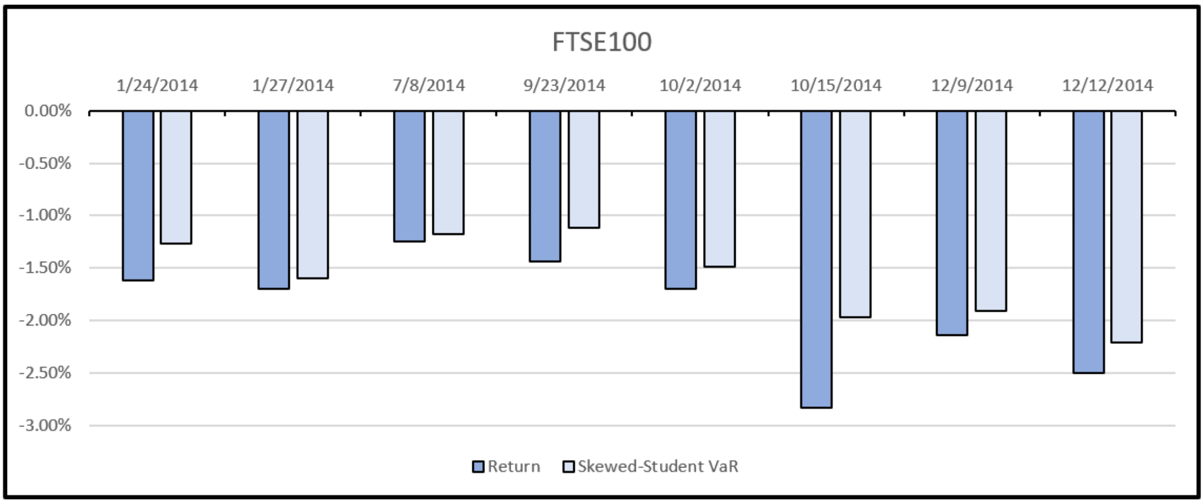
Wong, K. W., Fan, G. & Zeng, Y., 2012. "Capturing Tail Risks beyond VaR". **Review of Pacific Basin Financial Markets and Policies** Vol. 15, No. 3 (2012) 1250015 (25 pages).

Yamai, Y. & Yoshiba, T., 2005. *"Value at Risk versus Expected Shortfall: a practical perspective"*. **Journal of Banking and Finance**, 29(4), 997-1015.

Appendix

Figure 3: FTSE100 – 2014

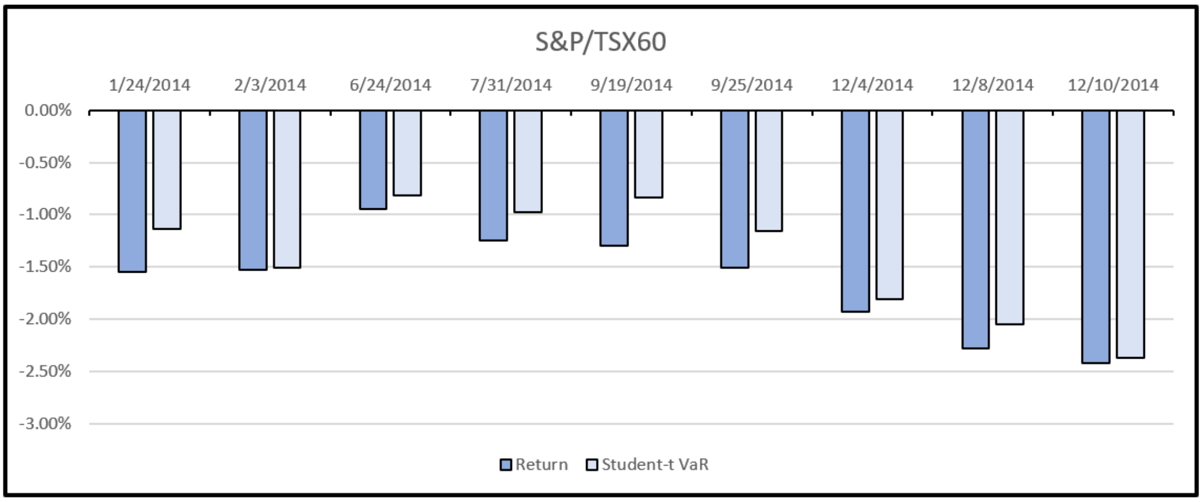
This figure illustrates Skewed-Student VaR exceedances that occurred during 2014 in the British Market Index, as well as Kupiec's UC and Wong's Saddle-point test results for the year.



Exceedance Rate	Unconditional Coverage	Saddle-point
3.07%	0.71%***	16.09%

Figure 4: S&P/TSX60 – 2014

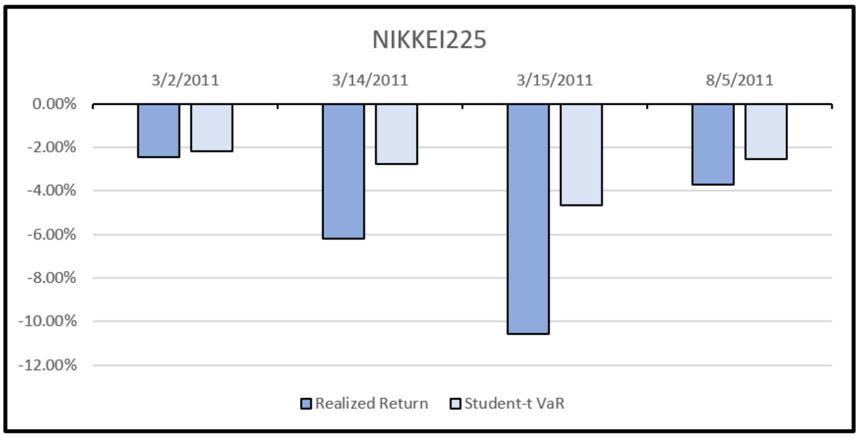
This figure illustrates Student-t VaR exceedances that occurred during 2014 in the Canadian Market Index, as well as Kupiec's UC and Wong's Saddle-point test results for the year.



Exceedance Rate	Unconditional Coverage	Saddle-point
3.45%	0.19%***	16.46%

Figure 5: NIKKEI225 – 2011

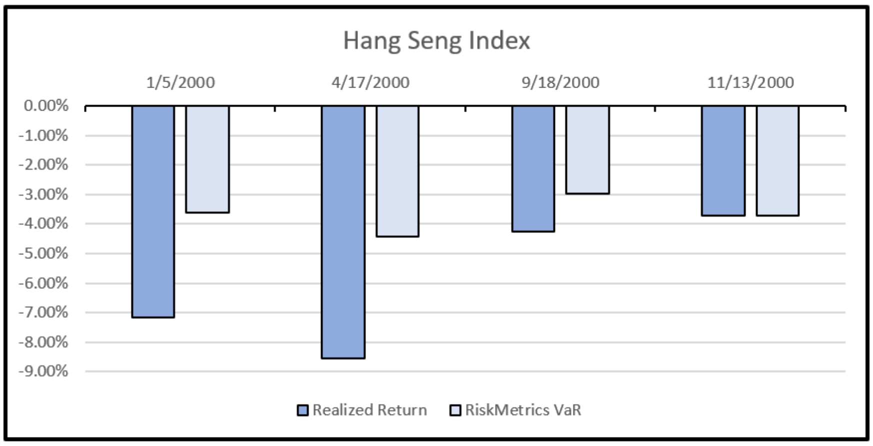
This figure illustrates Student-t VaR exceedances that occurred during 2011 in the Japanese Market Index, as well as Kupiec's UC and Wong's Saddle-point test results for the year.



Exceedance Rate	Unconditional Coverage	Saddle-point
1.54%	41.87%	0.51%***

Figure 6: Hang Seng – 2000

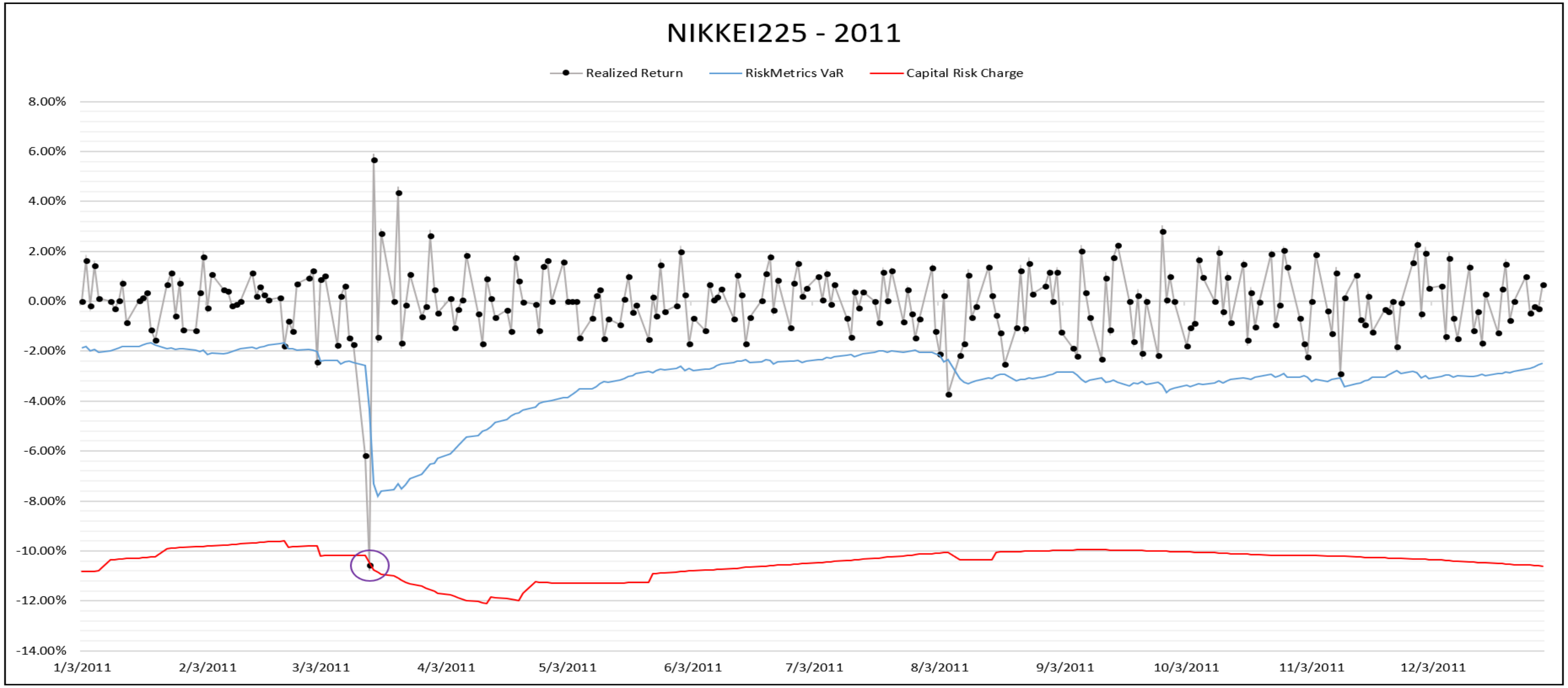
This figure illustrates Student-t VaR exceedances that occurred during 2000 in the Hong Kong Market Index, as well as Kupiec's UC and Wong's Saddle-point test results for the year.



Exceedance Rate	Unconditional Coverage	Saddle-point
1.54%	41.87%	0.03%***

Figure 7: Super-exceedance NIKKEI225 - 2011

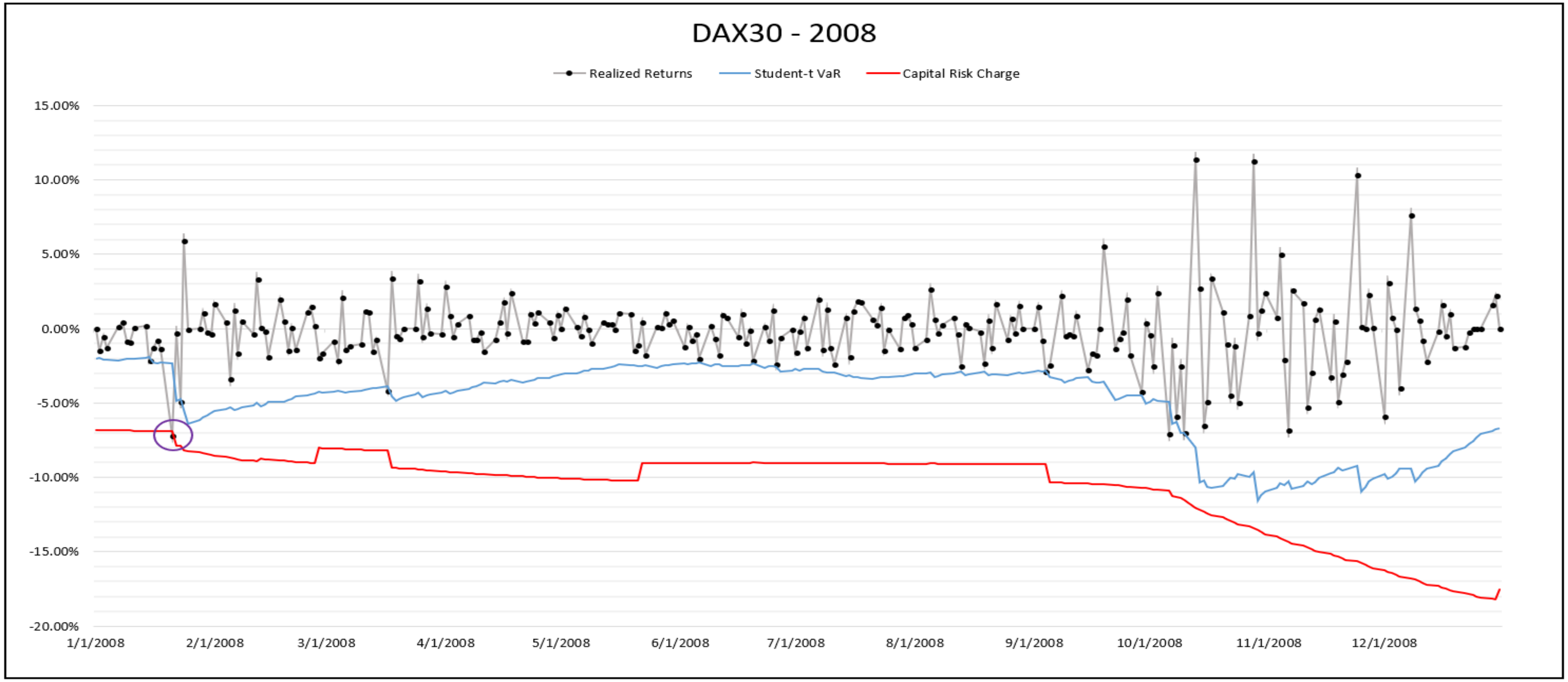
This figure consists of two complementary components, both concerning the analysis of NIKKEI225 in 2011: firstly, a graph displaying the Realized Returns, RiskMetrics VaR figures and Capital Risk Charge determined in accordance to Basel II Agreements; and secondly, a small table depicting UC and Saddle-point p-values for 2011, as well as for the 5-year period from 2008 to 2012 and the overall 20-year period between 1998 and 2017.



	Exceedances	Exceedance Rate	Unconditional Coverage	Saddle-point
2011	5	1.92%	18.44%	0%***
2008-2012	25	1.92%	0.32%***	0%***
1998-2017	103	1.97%	0%***	0%***

Figure 8: Super-exceedance DAX30 - 2008

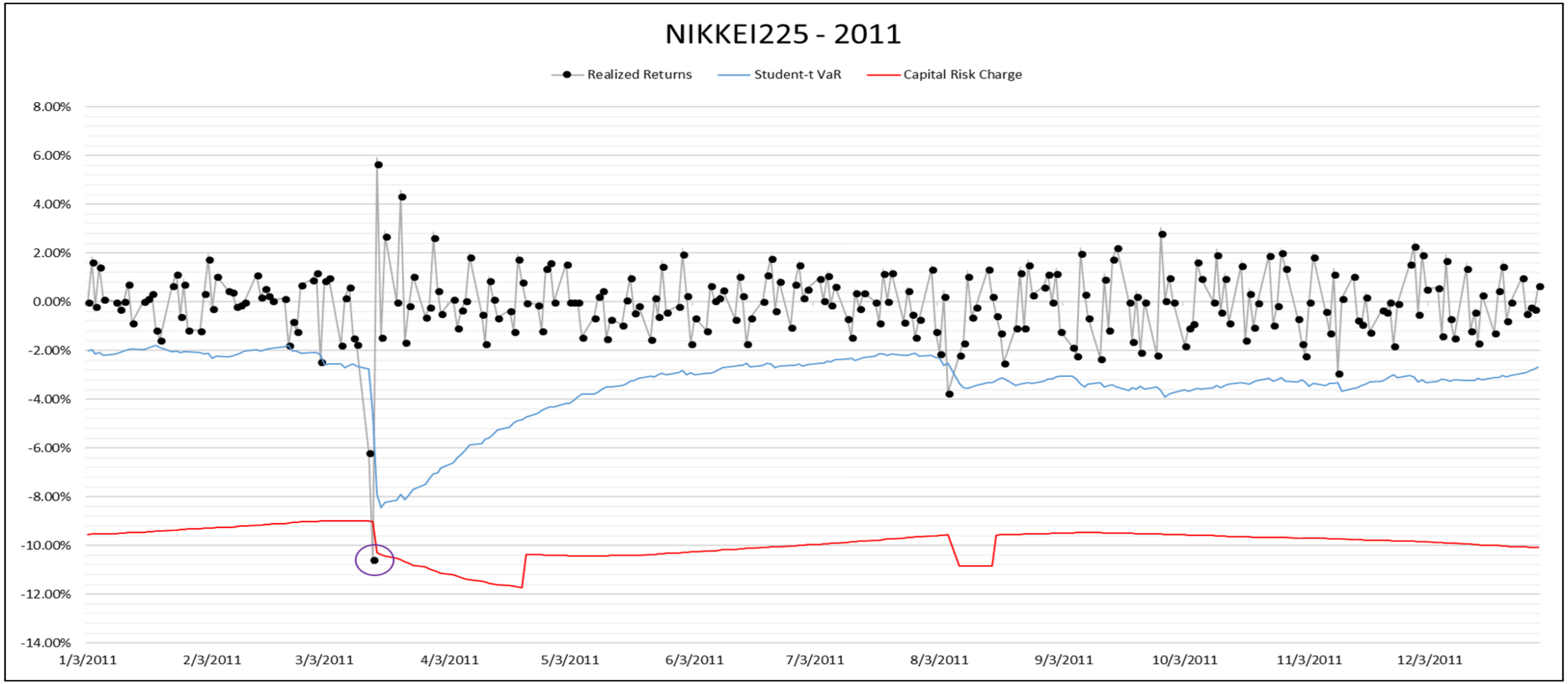
This figure consists of two complementary components, both concerning the analysis of DAX30 in 2008: firstly, a graph displaying the Realized Returns, Student-t VaR figures and Capital Risk Charge determined in accordance to Basel II Agreements; and secondly, a small table depicting UC and Saddle-point p-values for 2008, as well as for the 5-year period from 2008 to 2012 and the overall 20-year period between 1998 and 2017.



	Exceedances	Exceedance Rate	Unconditional Coverage	Saddle-point
2008	7	2.67%	2.43%**	0.77%***
2008-2012	20	1.53%	7.3%*	17.65%
1998-2017	67	1.28%	4.81%**	12.78%

Figure 9: Super-exceedance NIKKEI225 – 2011

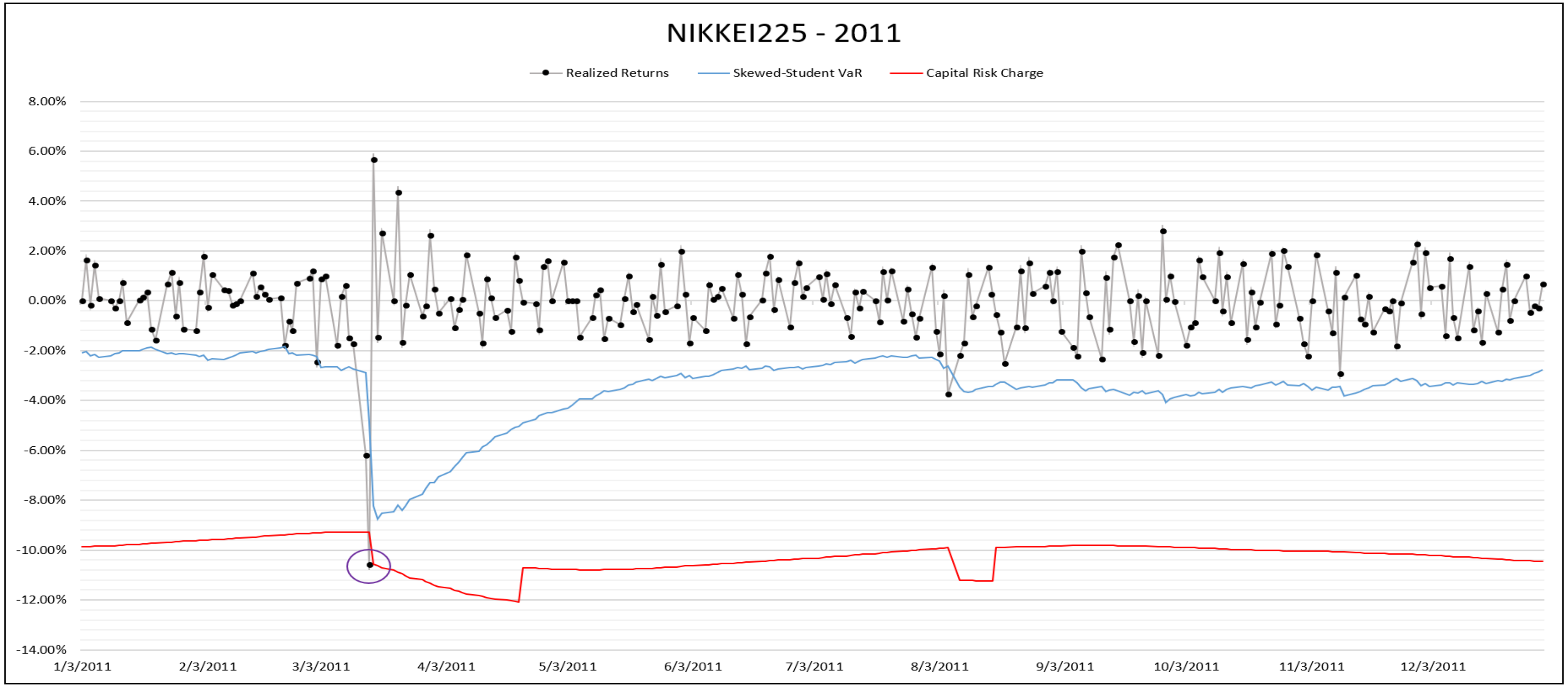
This figure consists of two complementary components, both concerning the analysis of NIKKEI225 in 2011: firstly, a graph displaying the Realized Returns, Student-t VaR figures and Capital Risk Charge determined in accordance to Basel II Agreements; and secondly, a small table depicting UC and Saddle-point p-values for 2011, as well as for the 5-year period from 2008 to 2012 and the overall 20-year period between 1998 and 2017.



	Exceedances	Exceedance Rate	Unconditional Coverage	Saddle-point
2011	4	1.54%	41.87%	0.51%***
2008-2012	18	1.38%	19.28%	2.53%**
1998-2017	79	1.51%	0.05%***	0.29%***

Figure 10: Super-exceedance NIKKEI225 - 2011

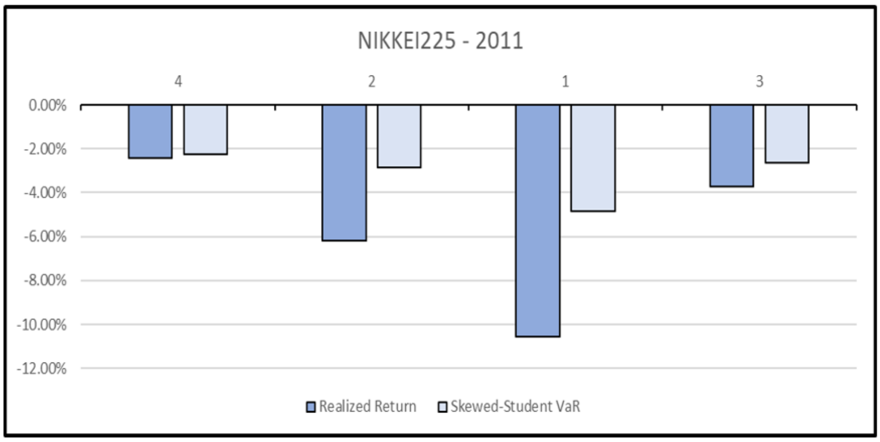
This figure consists of two complementary components, both concerning the analysis of NIKKEI225 in 2011: firstly, a graph displaying the Realized Returns, Skewed-Student VaR figures and Capital Risk Charge determined according to Basel II Agreements; and secondly, a small table depicting UC and Saddle-point p-values for 2011, as well as for the 5-year period from 2008 to 2012 and the overall 20-year period between 1998 and 2017.



	Exceedances	Exceedance Rate	Unconditional Coverage	Saddle-point
2011	4	1.54%	41.87%	0.72%***
2008-2012	15	1.15%	59.62%	4.86%**
1998-2017	71	1.56%	1.3%**	1.04%**

Figure 11: Saddle-point sensitivity analysis 1

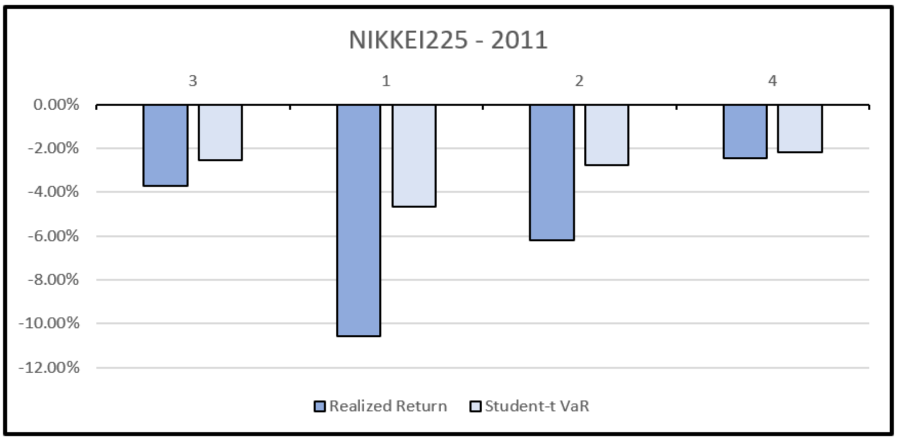
This figure illustrates Skewed-Student VaR exceedances that occurred during 2011 in the Japanese Market Index. Then, one by one, from the smallest to the largest, the exceedances are taken out from the sample, until there is only the biggest exceedance left – the one that has surpassed Basel II Capital Risk Charge.



Exceedances	Saddle-point
{1, 2, 3, 4}	0.77%***
{1, 2, 3}	2.00%**
{1, 2}	2.45%**
1	21.15%

Figure 12: Saddle-point sensitivity analysis 2

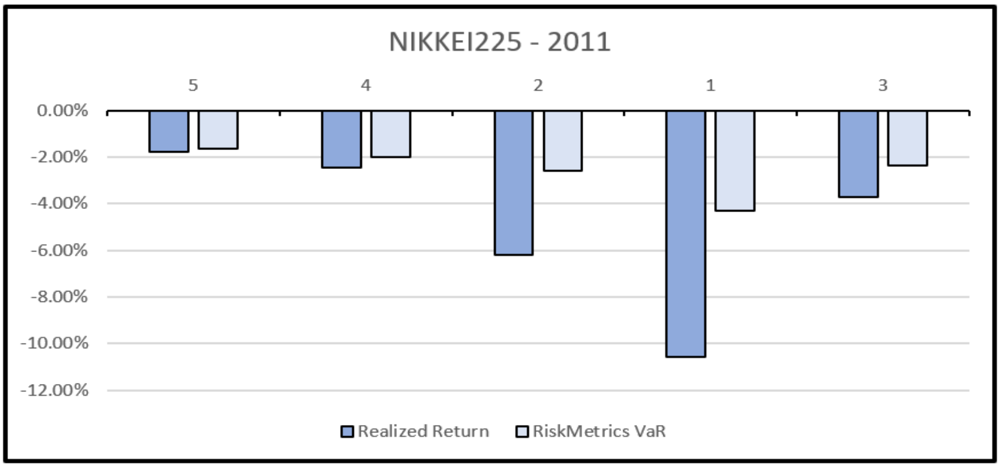
This figure illustrates Student-t VaR exceedances that occurred during 2011 in the Japanese Market Index. Then, one by one, from the smallest to the largest, the exceedances are taken out from the sample, until there is only the biggest exceedance left – the one that has surpassed Basel II Capital Risk Charge.



Exceedances	Saddle-point
{1, 2, 3, 4}	0.51%***
{1, 2, 3}	0.69%***
{1, 2}	2.14%**
1	20.62%

Figure 13: Saddle-point sensitivity analysis 3

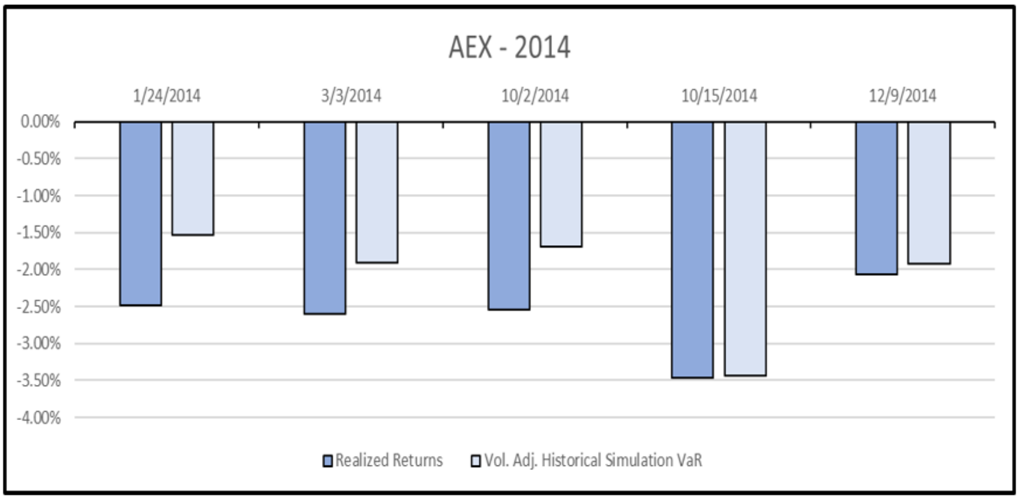
This figure illustrates RiskMetrics VaR exceedances that occurred during 2011 in the Japanese Market Index. Then, one by one, from the smallest to the largest, the exceedances are taken out from the sample, until there is only the biggest exceedance left – the one that has surpassed Basel II Capital Risk Charge.



Exceedances	Saddle-point
{1, 2, 3, 4, 5}	0.00%***
{1, 2, 3, 4}	0.00%***
{1, 2, 3}	0.00%***
{1, 2}	0.00%***
1	0.19%***

Figure 14: Amsterdam Market Index – 2014

This figure illustrates Vol. Adj. Historical Simulation VaR exceedances that occurred during 2014 in the Hong Kong Market Index, as well as Kupiec's UC and Wong's Saddle-point test results for the year.



Exceedance Rate	Unconditional Coverage	Saddle-point
1.54%	41.87%	0.03%***

Figure 15: Tail-Risk-of-VaR in Amsterdam Index – 2014

This figure illustrates the Tail-Risk-of-VaR measures for the Hong Kong Market Index during 2014, when the risk model at use is the Vol. Adj. Historical Simulation.

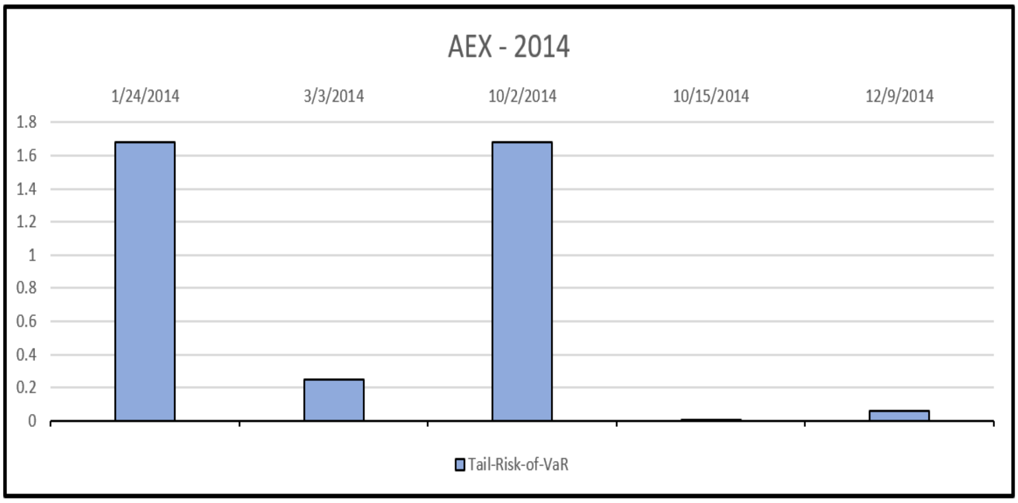


Table 3: Backtesting Results for RiskMetrics VaR

This table summarizes the performance of the RiskMetrics model in constructing one-day-ahead Value-at-Risk forecasts for ten different market indices, according to two distinct backtests: Kupiec's Unconditional Coverage and Wong's Saddle-point Backtest. The results displayed below regard to the exceedance rate (light blue coloured) and p-values of the tests, and the symbols [*; **; ***] represent the rejection of the VaR model at the 10%, 5% and 1% significance level, respectively.

	NIKKEI225			S&P500			FTSE100			MSCI World Index			DAX30		
	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point
1998	1.92%	18.68%	37.09%	3.07%	0.71%***	0%***	2.68%	2.38%**	0.46%***	3.07%	0.71%***	0%***	2.68%	2.38%**	0%***
1999	1.92%	18.68%	39.58%	1.15%	81.27%	52.79%	1.15%	81.27%	13.07%	1.92%	18.68%	26.65%	1.53%	42.26%	16.25%
2000	2.69%	2.34%**	0.01%***	2.31%	7.01%*	0%***	1.92%	18.44%	1.12%**	1.92%	18.44%	0%***	1.54%	41.87%	60.84%
2001	1.15%	81.27%	3.78%**	1.92%	18.68%	0.06%***	1.92%	18.68%	0.01%***	1.92%	18.68%	0.21%***	1.53%	42.26%	0.27%***
2002	0.77%	69.23%	58.30%	1.15%	81.27%	17.13%	2.30%	7.13%*	5.2%*	1.92%	18.68%	9.7%*	0.38%	25.22%	60.23%
2003	1.53%	42.26%	3.97%**	1.15%	81.27%	88.35%	0.77%	69.23%	16.06%	0.38%	25.22%	85.21%	0.77%	69.23%	63.99%
2004	1.15%	81.76%	13.65%	1.91%	18.91%	61.07%	2.29%	7.24%*	9.51%*	1.53%	42.65%	30.98%	2.67%	2.43%**	20.86%
2005	1.54%	41.87%	8.45%*	1.15%	80.77%	65.87%	2.69%	2.34%**	32.32%	1.54%	41.87%	68.27%	1.54%	41.87%	18.76%
2006	1.92%	18.44%	18.88%	2.31%	7.01%*	21.56%	1.54%	41.87%	14.42%	1.15%	80.77%	46.66%	1.54%	41.87%	17.03%
2007	3.45%	0.19%***	0.01%***	4.60%	0%***	0.01%***	3.45%	0.19%***	0.1%***	2.68%	2.38%**	0.95%***	1.53%	42.26%	12.10%
2008	3.05%	0.72%***	0%***	3.44%	0.19%***	0%***	3.05%	0.72%***	0%***	3.44%	0.19%***	0%***	3.05%	0.72%***	0%***
2009	0.77%	69.23%	40.59%	0.77%	69.23%	73.41%	1.15%	81.27%	14.72%	1.15%	81.27%	29.64%	0.77%	69.23%	84.26%
2010	2.68%	2.38%**	33.36%	3.45%	0.19%***	2.81%**	2.30%	7.13%*	10.65%	3.07%	0.71%***	1.01%**	2.30%	7.13%*	28.63%
2011	1.92%	18.44%	0%***	2.69%	2.34%**	0%***	2.31%	7.01%*	8.83%*	3.85%	0.04%***	0.04%***	3.46%	0.18%***	0.31%***
2012	1.15%	81.27%	40.00%	1.92%	18.68%	20.60%	1.53%	42.26%	56.50%	1.53%	42.26%	36.74%	1.53%	42.26%	20.86%
2013	1.92%	18.68%	1.1%**	1.92%	18.68%	17.59%	2.68%	2.38%**	10.51%	3.07%	0.71%***	5.35%*	3.07%	0.71%***	3.21%**
2014	2.68%	2.38%**	3.17%**	3.83%	0.05%***	1.37%**	4.21%	0.01%***	1.2%**	3.07%	0.71%***	6.16%*	3.07%	0.71%***	5.38%*
2015	3.07%	0.71%***	0.19%***	2.68%	2.38%**	1.09%**	2.68%	2.38%**	0.34%***	3.07%	0.71%***	0.31%***	2.68%	2.38%**	8.64%*
2016	3.07%	0.71%***	0%***	1.53%	42.26%	0.74%***	1.53%	42.26%	25.17%	1.92%	18.68%	0.1%***	1.53%	42.26%	0.76%***
2017	1.15%	80.77%	37.59%	1.54%	41.87%	2.11%**	1.54%	41.87%	4.74%**	1.15%	80.77%	39.32%	1.15%	80.77%	55.66%
1998-2002	1.69%	2.32%**	0.12%***	1.92%	0.31%***	0%***	1.99%	0.15%***	0%***	2.15%	0.03%***	0%***	1.53%	7.25%*	0%***
2003-2007	1.92%	0.31%***	0%***	2.22%	0.01%***	3.28%**	2.15%	0.03%***	0.12%***	1.46%	12.05%	20.65%	1.61%	4.18%**	8.98%*
2008-2012	1.92%	0.32%***	0%***	2.45%	0%***	0%***	2.07%	0.07%***	0%***	2.61%	0%***	0%***	2.22%	0.01%***	0.01%***
2013-2017	2.38%	0%***	0%***	2.30%	0.01%***	0%***	2.53%	0%***	0.01%***	2.45%	0%***	0%***	2.30%	0.01%***	0.09%***
1998-2017	1.97%	0%***	0%***	2.22%	0%***	0%***	2.19%	0%***	0%***	2.17%	0%***	0%***	1.92%	0%***	0%***

	S&P/TSX60			Dow Jones			NASDAQ			AEX			HSI		
	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point
1998	4.98%	0%***	0%***	3.83%	0.05%***	0%***	2.68%	2.38%**	0%***	2.68%	2.38%**	0.00%***	3.07%	0.71%***	0%***
1999	1.53%	42.26%	0.43%***	1.15%	81.27%	72.70%	1.15%	81.27%	4.56%**	0.77%	69.23%	28.21%	0.38%	25.22%	84.32%
2000	2.31%	7.01%*	0%***	2.69%	2.34%**	0%***	1.15%	80.77%	0.03%***	2.69%	2.34%**	0.07%***	1.54%	41.87%	0.03%***
2001	1.53%	42.26%	0%***	1.15%	81.27%	0%***	0.38%	25.22%	12.60%	2.68%	2.38%**	0.00%***	2.30%	7.13%*	0.2%***
2002	1.92%	18.68%	30.61%	1.15%	81.27%	7.76%*	0.00%	2.2%**	7.26%*	0.77%	69.23%	35.61%	1.15%	81.27%	68.82%
2003	1.15%	81.27%	38.11%	0.00%	2.2%**	7.26%*	1.15%	81.27%	79.33%	1.53%	42.26%	57.70%	1.15%	81.27%	32.88%
2004	1.15%	81.76%	9.28%*	1.15%	81.76%	51.16%	0.76%	68.79%	82.72%	1.91%	18.91%	12.54%	1.15%	81.76%	25.42%
2005	2.69%	2.34%**	21.81%	1.15%	80.77%	36.98%	1.15%	80.77%	68.45%	2.31%	7.01%*	27.29%	2.31%	7.01%*	20.94%
2006	2.69%	2.34%**	2.33%**	1.92%	18.44%	11.29%	1.92%	18.44%	29.07%	2.31%	7.01%*	9.13%*	1.92%	18.44%	6.32%*
2007	3.45%	0.19%***	0.1%***	3.45%	0.19%***	0.02%***	2.68%	2.38%**	9.18%*	2.68%	2.38%**	4.7%**	2.30%	7.13%*	4.72%**
2008	4.20%	0.01%***	0%***	2.67%	2.43%**	0.03%***	2.67%	2.43%**	0.05%***	2.67%	2.43%**	0.00%***	3.05%	0.72%***	0%***
2009	1.15%	81.27%	35.86%	0.77%	69.23%	71.77%	1.53%	42.26%	45.37%	0.38%	25.22%	74.72%	0.38%	25.22%	32.73%
2010	2.30%	7.13%*	20.26%	3.07%	0.71%***	1.91%**	3.45%	0.19%***	5.4%*	2.68%	2.38%**	6.95%*	1.15%	81.27%	58.14%
2011	2.69%	2.34%**	0.48%***	3.46%	0.18%***	0.01%***	3.08%	0.69%***	0%***	1.92%	18.44%	8.20%*	3.08%	0.69%***	0.17%***
2012	1.92%	18.68%	20.68%	2.30%	7.13%*	7.86%*	2.30%	7.13%*	29.94%	1.92%	18.68%	24.81%	1.53%	42.26%	22.00%
2013	2.30%	7.13%*	2.29%**	2.30%	7.13%*	18.17%	3.07%	0.71%***	9.18%*	2.68%	2.38%**	9.52%*	2.68%	2.38%**	7.6%*
2014	3.83%	0.05%***	4.53%**	3.83%	0.05%***	3.21%**	3.83%	0.05%***	3.56%**	2.68%	2.38%**	0.47%***	3.07%	0.71%***	7.3%*
2015	3.07%	0.71%***	2.54%**	2.30%	7.13%*	2.24%**	3.07%	0.71%***	0.91%***	2.68%	2.38%**	0.35%***	1.53%	42.26%	0.63%***
2016	3.07%	0.71%***	9.12%*	1.15%	81.27%	1.26%**	1.53%	42.26%	0.25%***	1.15%	81.27%	5.46%*	3.07%	0.71%***	0.69%***
2017	2.31%	7.01%*	12.14%	1.92%	18.44%	4.42%**	2.69%	2.34%**	0.14%***	1.54%	41.87%	54.74%	2.31%	7.01%*	20.24%
1998-2002	2.45%	0%***	0%***	1.99%	0.15%***	0%***	1.07%	79.18%	0%***	1.92%	0.31%***	0.00%***	1.69%	2.32%**	0%***
2003-2007	2.22%	0.01%***	0.04%***	1.53%	7.25%*	1.44%**	1.53%	7.25%*	56.81%	2.15%	0.03%***	2.96%**	1.76%	1.23%**	1.91%**
2008-2012	2.45%	0%***	0%***	2.45%	0%***	0%***	2.61%	0%***	0%***	1.92%	0.32%***	0.02%***	1.84%	0.64%***	0%***
2013-2017	2.91%	0%***	0.03%***	2.30%	0.01%***	0.01%***	2.84%	0%***	0%***	2.15%	0.03%***	0.01%***	2.53%	0%***	0.01%***
1998-2017	2.51%	0%***	0%***	2.07%	0%***	0%***	2.01%	0%***	0%***	2.03%	0.00%***	0.00%***	1.96%	0%***	0%***

Table 4: Backtesting Results for Student-t VaR

This table summarizes the performance of the Student-t distribution model in constructing one-day-ahead Value-at-Risk forecasts for ten different market indices, according to two distinct backtests: Kupiec's Unconditional Coverage and Wong's Saddle-point Backtest. The results displayed below regard to the exceedance rate (light blue coloured) and p-values of the tests, and the symbols [*; **; ***] represent the rejection of the VaR model at the 10%, 5% and 1% significance level, respectively.

	NIKKEI225			S&P500			FTSE100			MSCI World Index			DAX30			S&P/TSX60			Dow Jones			NASDAQ			AEX			HSI		
	Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value	
		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point
1998	1.15%	81.27%	80.22%	2.68%	2.38%**	1.2%**	2.30%	7.13%*	10.99%	2.68%	2.38%**	3.2%**	1.92%	18.68%	4%**	4.60%	0%***	0%***	2.68%	2.38%**	0.69%***	2.30%	7.13%*	0.65%***	1.92%	18.68%	4.62%**	1.92%	18.68%	0.7%***
1999	1.15%	81.27%	85.70%	0.38%	25.22%	92.31%	0.77%	69.23%	44.86%	0.77%	69.23%	73.07%	1.15%	81.27%	60.02%	1.15%	81.27%	30.08%	0.00%	2.2%**	7.26%*	0.77%	69.23%	58.55%	0.77%	69.23%	73.41%	0.00%	2.2%**	7.26%*
2000	2.31%	7.01%*	6.57%*	1.54%	41.87%	11.69%	1.54%	41.87%	8.66%*	1.54%	41.87%	4.33%**	0.77%	69.67%	92.81%	1.54%	41.87%	15.74%	1.54%	41.87%	4.85%**	1.15%	80.77%	25.79%	1.54%	41.87%	3.21%**	1.15%	80.77%	9.5%*
2001	1.15%	81.27%	47.82%	1.15%	81.27%	15.36%	1.92%	18.68%	1.5%**	1.53%	42.26%	16.48%	0.77%	69.23%	23.63%	1.15%	81.27%	4.83%**	1.15%	81.27%	3.97%**	0.38%	25.22%	36.78%	1.53%	42.26%	0.61%***	1.53%	42.26%	18.61%
2002	0.77%	69.23%	89.13%	0.77%	69.23%	48.46%	1.92%	18.68%	25.18%	1.53%	42.26%	56.31%	0.38%	25.22%	75.55%	0.77%	69.23%	77.23%	1.15%	81.27%	46.40%	0.00%	2.2%**	7.26%*	0.77%	69.23%	58.17%	0.77%	69.23%	88.43%
2003	1.15%	81.27%	32.07%	0.00%	2.2%**	7.26%*	0.77%	69.23%	39.18%	0.00%	2.2%**	7.26%*	0.77%	69.23%	76.28%	0.77%	69.23%	66.50%	0.00%	2.2%**	7.26%*	0.77%	69.23%	89.26%	0.77%	69.23%	86.16%	0.38%	25.22%	56.16%
2004	0.38%	24.99%	37.01%	1.15%	81.76%	78.89%	2.29%	7.24%*	18.12%	1.53%	42.65%	56.69%	1.53%	42.65%	34.29%	0.76%	68.79%	31.99%	1.15%	81.76%	67.11%	0.38%	24.99%	88.05%	1.91%	18.91%	22.24%	1.15%	81.76%	50.02%
2005	1.54%	41.87%	22.62%	1.15%	80.77%	80.42%	2.69%	2.34%**	47.29%	0.38%	25.44%	81.89%	1.54%	41.87%	23.03%	2.31%	7.01%*	55.03%	1.15%	80.77%	51.50%	0.77%	69.67%	74.96%	1.92%	18.44%	36.82%	1.92%	18.44%	47.16%
2006	1.54%	41.87%	53.51%	1.54%	41.87%	34.79%	1.15%	80.77%	23.86%	1.15%	80.77%	59.41%	1.54%	41.87%	24.76%	2.31%	7.01%*	15.60%	1.92%	18.44%	22.97%	1.54%	41.87%	33.28%	1.92%	18.44%	16.39%	1.54%	41.87%	28.22%
2007	2.30%	7.13%*	1.39%**	3.83%	0.05%***	0.94%***	3.07%	0.71%***	1.77%**	1.92%	18.68%	5.05%*	1.15%	81.27%	18.22%	2.68%	2.38%**	2.2%**	2.68%	2.38%**	1.29%**	1.92%	18.68%	16.17%	2.68%	2.38%**	15.45%	1.15%	81.27%	39.92%
2008	2.67%	2.43%**	1.4%**	2.67%	2.43%**	10.08%	3.05%	0.72%**	1.34%**	2.67%	2.43%**	1.65%**	2.67%	2.43%**	0.77%***	2.29%	7.24%*	3.23%**	1.91%	18.91%	16.89%	2.67%	2.43%**	4.13%**	2.29%	7.24%*	1.64%**	2.67%	2.43%**	17.82%
2009	0.77%	69.23%	57.90%	0.38%	25.22%	85.17%	0.77%	69.23%	45.97%	1.15%	81.27%	61.07%	0.00%	2.2%**	7.26%*	0.77%	69.23%	74.67%	0.38%	25.22%	86.38%	0.77%	69.23%	58.00%	0.38%	25.22%	83.48%	0.38%	25.22%	59.07%
2010	0.77%	69.23%	70.72%	2.30%	7.13%*	24.29%	1.53%	42.26%	27.71%	2.68%	2.38%**	6.76%*	1.53%	42.26%	64.52%	1.53%	42.26%	47.95%	1.92%	18.68%	17.35%	2.30%	7.13%*	22.45%	2.30%	7.13%*	26.03%	0.38%	25.22%	85.41%
2011	1.54%	41.87%	0.51%***	2.31%	7.01%*	1.2%**	1.92%	18.44%	45.03%	2.31%	7.01%*	4.21%**	2.69%	2.34%**	19.97%	2.31%	7.01%*	6.7%*	2.69%	2.34%**	2.71%**	3.08%	0.69%***	1.01%**	1.54%	41.87%	35.72%	2.31%	7.01%*	7.27%*
2012	1.15%	81.27%	62.63%	1.53%	42.26%	50.39%	1.15%	81.27%	77.42%	1.15%	81.27%	56.55%	0.77%	69.23%	53.43%	1.53%	42.26%	41.73%	2.30%	7.13%*	29.59%	1.15%	81.27%	55.23%	1.53%	42.26%	50.48%	1.53%	42.26%	43.54%
2013	1.53%	42.26%	22.60%	1.92%	18.68%	46.83%	1.92%	18.68%	28.26%	1.15%	81.27%	19.88%	1.92%	18.68%	12.75%	1.53%	42.26%	6.81%*	1.92%	18.68%	44.26%	2.30%	7.13%*	28.89%	2.30%	7.13%*	24.63%	1.92%	18.68%	20.50%
2014	2.30%	7.13%*	15.66%	3.07%	0.71%***	10.73%	3.83%	0.05%***	8.1%*	2.30%	7.13%*	18.28%	1.53%	42.26%	22.47%	3.45%	0.19%***	16.46%	2.30%	7.13%*	14.71%	2.68%	2.38%**	23.84%	2.30%	7.13%*	2.62%**	2.30%	7.13%*	21.48%
2015	3.07%	0.71%***	3.31%**	1.53%	42.26%	11.36%	1.92%	18.68%	7.18%*	2.30%	7.13%*	6.34%*	1.53%	42.26%	60.00%	2.68%	2.38%**	15.99%	1.53%	42.26%	19.65%	1.92%	18.68%	9.15%*	2.30%	7.13%*	10.77%	1.53%	42.26%	12.17%
2016	1.92%	18.68%	1.27%**	0.77%	69.23%	5.13%*	1.15%	81.27%	66.82%	1.53%	42.26%	5.86%*	0.77%	69.23%	21.12%	1.92%	18.68%	28.74%	0.77%	69.23%	6.95%*	1.53%	42.26%	4.59%**	0.77%	69.23%	39.47%	3.07%	0.71%***	16.05%
2017	1.15%	80.77%	58.43%	1.54%	41.87%	7.78%*	1.54%	41.87%	14.95%	0.77%	69.67%	54.10%	0.77%	69.67%	76.78%	1.92%	18.44%	35.03%	1.92%	18.44%	16.00%	1.92%	18.44%	2.56%**	0.77%	69.67%	72.78%	1.54%	41.87%	57.88%
1998-2002	1.30%	29.24%	71.41%	1.30%	29.24%	5.3%*	1.69%	2.32%**	0.95%***	1.61%	4.18%**	4.35%**	1.00%	99.11%	40.82%	1.84%	0.63%***	0.06%***	1.30%	29.24%	0.92%***	0.92%	76.92%	13.64%	1.30%	29.24%	0.61%***	1.07%	79.18%	7.85%*
2003-2007	1.38%	19.18%	6.2%*	1.53%	7.25%*	39.02%	1.99%	0.15%***	4.59%**	1.00%	99.11%	63.88%	1.30%	29.24%	24.04%	1.76%	1.23%**	10.23%	1.38%	19.18%	23.18%	1.07%	79.18%	76.28%	1.84%	0.63%***	20.80%	1.23%	42.63%	49.82%
2008-2012	1.38%	19.28%	2.53%**	1.84%	0.64%***	6.04%*	1.69%	2.34%**	14.63%	1.99%	0.15%***	1.31%**	1.53%	7.3%*	17.65%	1.69%	2.34%**	9.38%*	1.84%	0.64%***	6.73%*	1.99%	0.15%***	2.12%**	1.61%	4.21%**	15.20%	1.46%	12.12%	29.51%
2013-2017	1.99%	0.15%***	0.7%***	1.76%	1.23%**	1.06%**	2.07%	0.07%***	4.75%**	1.61%	4.18%**	2.58%**	1.30%	29.24%	26.29%	2.30%	0.01%***	4.25%**	1.69%	2.32%**	3.69%**	2.07%	0.07%***	0.47%***	1.69%	2.32%**	6.32%*	2.07%	0.07%***	8.52%*
1998-2017	1.51%	0.05%***	0.29%***	1.61%	0%***	0.23%***	1.86%	0%***	0.05%***	1.55%	0.02%***	0.26%***	1.28%	4.81%**	12.78%	1.90%	0%***	0.01%***	1.55%	0.02%***	0.09%***	1.51%	0.05%***	0.45%***	1.61%	0%***	0.2%***	1.46%	0.19%***	5.52%*

Table 5: Backtesting Results for Skewed-Student VaR

This table summarizes the performance of the Skewed-Student distribution model in constructing one-day-ahead Value-at-Risk forecasts for ten different market indices, according to two distinct backtests: Kupiec's Unconditional Coverage and Wong's Saddle-point Backtest. The results displayed below regard to the exceedance rate (light blue coloured) and p-values of the tests, and the symbols [*; **; ***] represent the rejection of the VaR model at the 10%, 5% and 1% significance level, respectively.

	NIKKEI225			S&P500			FTSE100			MSCI World Index			DAX30		
	Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value	
		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point
1998	1.15%	81.27%	82.29%	2.68%	2.38%**	0.97%***	2.30%	7.13%*	9.62%*	2.30%	7.13%*	6.5%*	1.92%	18.68%	5.82%*
1999	1.15%	81.27%	87.67%	0.38%	25.22%	92.08%	0.77%	69.23%	45.66%	0.77%	69.23%	84.78%	0.77%	69.23%	69.87%
2000	1.92%	18.44%	7.69%*	1.54%	41.87%	12.47%	1.54%	41.87%	11.84%	0.77%	69.67%	8.07%*	0.00%	2.22%**	7.33%*
2001	1.15%	81.27%	49.76%	0.77%	69.23%	17.01%	1.53%	42.26%	2.8%**	1.53%	42.26%	31.18%	0.77%	69.23%	31.21%
2002	0.77%	69.23%	87.66%	0.77%	69.23%	51.60%	1.53%	42.26%	38.02%	0.77%	69.23%	75.45%	0.38%	25.22%	81.66%
2003	1.53%	42.26%	31.46%	0.00%	2.2%**	7.26%*	0.77%	69.23%	47.41%	0.00%	2.2%**	7.26%*	0.38%	25.22%	86.97%
2004	0.38%	24.99%	37.15%	1.15%	81.76%	82.14%	1.91%	18.91%	25.23%	1.15%	81.76%	70.44%	1.15%	81.76%	48.19%
2005	1.54%	41.87%	21.65%	0.77%	69.67%	84.49%	2.69%	2.34%**	61.99%	0.38%	25.44%	84.62%	1.54%	41.87%	31.29%
2006	1.54%	41.87%	49.93%	1.54%	41.87%	38.66%	1.15%	80.77%	30.65%	1.15%	80.77%	64.63%	1.54%	41.87%	33.18%
2007	2.30%	7.13%*	1.5%**	3.83%	0.05%***	1.8%**	2.30%	7.13%*	5.45%*	1.53%	42.26%	7.19%*	1.15%	81.27%	24.10%
2008	1.91%	18.91%	1.83%**	1.53%	42.65%	16.47%	1.91%	18.91%	4.02%**	2.67%	2.43%**	4.22%**	1.53%	42.65%	1.78%**
2009	0.77%	69.23%	63.83%	0.38%	25.22%	87.76%	0.77%	69.23%	53.71%	0.77%	69.23%	71.99%	0.00%	2.2%**	7.26%*
2010	0.77%	69.23%	77.04%	1.53%	42.26%	37.31%	1.15%	81.27%	40.06%	2.68%	2.38%**	17.89%	0.77%	69.23%	78.18%
2011	1.54%	41.87%	0.72%***	2.31%	7.01%*	2.75%**	0.77%	69.67%	60.14%	1.92%	18.44%	9.13%*	1.15%	80.77%	36.52%
2012	0.77%	69.23%	69.93%	1.53%	42.26%	63.84%	0.77%	69.23%	85.00%	0.77%	69.23%	68.93%	0.77%	69.23%	61.22%
2013	0.38%	25.22%	25.48%	1.15%	81.27%	58.68%	1.53%	42.26%	38.56%	0.77%	69.23%	27.50%	1.53%	42.26%	18.87%
2014	1.92%	18.68%	24.62%	2.68%	2.38%**	21.37%	3.07%	0.71%***	16.09%	1.92%	18.68%	34.06%	1.53%	42.26%	30.53%
2015	2.68%	2.38%**	5.92%*	1.53%	42.26%	18.16%	1.92%	18.68%	12.15%	1.92%	18.68%	15.77%	1.15%	81.27%	69.70%
2016	1.92%	18.68%	1.99%**	0.77%	69.23%	6.93%**	0.77%	69.23%	77.92%	0.77%	69.23%	8.9%*	0.77%	69.23%	24.73%
2017	1.15%	80.77%	64.16%	1.54%	41.87%	11.13%	1.54%	41.87%	22.06%	0.77%	69.67%	61.63%	0.77%	69.67%	81.06%
1998-2002	1.23%	42.63%	75.26%	1.23%	42.63%	5.58%*	1.53%	7.25%*	2.27%**	1.23%	42.63%	19.64%	0.77%	37.76%	57.55%
2003-2007	1.46%	12.05%	5.84%*	1.46%	12.05%	51.77%	1.76%	1.23%	16.69%	0.84%	55.96%	76.92%	1.15%	59.42%	46.79%
2008-2012	1.15%	59.62%	4.86%**	1.46%	12.12%	19.17%	1.07%	79.39%	38.85%	1.76%	1.25%**	9.54%*	0.84%	55.78%	40.06%
2013-2017	1.61%	4.18%**	2.34%**	1.53%	7.25%*	4.6%**	1.76%	1.23%	16.94%	1.23%	42.63%	12.13%	1.15%	59.42%	43.36%
1998-2017	1.36%	1.3%**	1.04%**	1.42%	0.43%***	2.26%**	1.53%	0.03%	2.07%**	1.27%	6.45%*	10.04%	0.98%	87.02%	48.88%

	S&P/TSX60			Dow Jones			NASDAQ			AEX			HSI		
	Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value	
		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point
1998	4.60%	0%***	0.01%***	2.30%	7.13%*	1.55%**	2.30%	7.13%*	2.21%**	1.92%	18.68%	7.13%*	2.30%	7.13%*	0.45%***
1999	1.15%	81.27%	40.59%	0.00%	2.2%**	7.26%*	0.38%	25.22%	67.47%	0.77%	69.23%	81.16%	0.00%	2.2%**	7.26%*
2000	1.54%	41.87%	25.59%	1.54%	41.87%	8.72%*	1.15%	80.77%	42.91%	1.15%	80.77%	5.48%*	1.15%	80.77%	7.76%*
2001	0.77%	69.23%	7.11%*	0.77%	69.23%	9.09%*	0.38%	25.22%	42.86%	1.53%	42.26%	1.37%**	1.53%	42.26%	15.59%
2002	0.38%	25.22%	85.41%	1.15%	81.27%	65.19%	0.00%	2.2%**	7.26%*	0.38%	25.22%	69.18%	0.77%	69.23%	86.05%
2003	0.38%	25.22%	72.19%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	93.81%	0.38%	25.22%	54.94%
2004	0.76%	68.79%	38.25%	0.76%	68.79%	78.99%	0.38%	24.99%	90.31%	1.91%	18.91%	33.84%	1.15%	81.76%	47.36%
2005	1.15%	80.77%	68.21%	1.15%	80.77%	61.81%	0.77%	69.67%	79.06%	1.92%	18.44%	47.60%	1.92%	18.44%	41.24%
2006	2.31%	7.01%*	25.98%	1.54%	41.87%	28.50%	1.54%	41.87%	36.40%	1.54%	41.87%	25.36%	1.54%	41.87%	24.49%
2007	1.92%	18.68%	4.75%**	2.30%	7.13%*	2.18%**	1.92%	18.68%	18.80%	1.92%	18.68%	29.44%	1.15%	81.27%	39.25%
2008	2.29%	7.24%*	9.05%*	1.53%	42.65%	20.23%	2.67%	2.43%**	7.36%*	1.91%	18.91%	4.27%**	2.67%	2.43%**	19.00%
2009	0.77%	69.23%	89.90%	0.38%	25.22%	86.93%	0.77%	69.23%	63.25%	0.38%	25.22%	87.30%	0.38%	25.22%	59.34%
2010	1.15%	81.27%	72.40%	1.92%	18.68%	17.14%	1.92%	18.68%	36.73%	1.53%	42.26%	38.27%	0.38%	25.22%	85.87%
2011	1.92%	18.44%	18.64%	2.69%	2.34%**	1.76%*	1.92%	18.44%	2.66%**	1.15%	80.77%	49.23%	2.31%	7.01%*	9.72%*
2012	1.15%	81.27%	59.92%	2.30%	7.13%*	28.62%	1.15%	81.27%	66.72%	1.53%	42.26%	64.56%	1.53%	42.26%	50.10%
2013	1.53%	42.26%	14.39%	1.92%	18.68%	44.28%	2.30%	7.13%*	44.32%	1.92%	18.68%	37.41%	1.92%	18.68%	25.10%
2014	2.30%	7.13%*	41.59%	2.30%	7.13%*	16.90%	2.30%	7.13%*	41.01%	1.92%	18.68%	5.01%*	2.30%	7.13%*	26.72%
2015	1.92%	18.68%	43.42%	1.53%	42.26%	22.53%	1.92%	18.68%	17.13%	1.92%	18.68%	18.89%	1.15%	81.27%	14.23%
2016	1.15%	81.27%	48.09%	0.77%	69.23%	8.3%*	1.53%	42.26%	8.23%*	0.77%	69.23%	44.73%	3.07%	0.71%***	19.47%
2017	1.54%	41.87%	54.22%	1.92%	18.44%	16.66%	1.92%	18.44%	5.57%*	0.77%	69.67%	78.31%	1.54%	41.87%	63.89%
1998-2002	1.69%	2.32%**	0.41%***	1.15%	59.42%	4.97%**	0.84%	55.96%	34.62%	1.15%	59.42%	2.6%**	1.15%	59.42%	4.88%**
2003-2007	1.30%	29.24%	26.55%	1.15%	59.42%	38.80%	0.92%	76.92%	83.47%	1.53%	7.25%*	49.51%	1.23%	42.63%	42.65%
2008-2012	1.46%	12.12%	45.69%	1.76%	1.25%**	5.93%*	1.69%	2.34%**	9.23%*	1.30%	29.37%	39.47%	1.46%	12.12%	36.76%
2013-2017	1.69%	2.32%**	36.57%	1.69%	2.32%**	5.07%*	1.99%	0.15%***	4.1%**	1.46%	12.05%	18.27%	1.99%	0.15%***	14.59%
1998-2017	1.53%	0.03%***	2.94%**	1.44%	0.29%***	0.63%***	1.36%	1.3%**	9.82%*	1.36%	1.3%**	6.34%*	1.46%	0.19%***	6.08%*

Table 6: Backtesting Results for Historical Simulation VaR

This table summarizes the performance of the Historical Simulation model in constructing one-day-ahead Value-at-Risk forecasts for ten different market indices, according to two distinct backtests: Kupiec's Unconditional Coverage and Wong's Saddle-point Backtest. The results displayed below regard to the exceedance rate (light blue coloured) and p-values of the tests, and the symbols [*; **; ***] represent the rejection of the VaR model at the 10%, 5% and 1% significance level, respectively.

	NIKKEI225			S&P500			FTSE100			MSCI World Index			DAX30		
	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point
1998	0.77%	69.23%	20.12%	5.36%	0%***	0.02%***	5.75%	0%***	0%***	5.36%	0%***	0%***	4.98%	0%***	0.04%***
1999	0.00%	2.2%**	7.26%*	1.92%	18.68%	53.27%	2.30%	7.13%*	43.48%	1.15%	81.27%	85.57%	0.77%	69.23%	67.40%
2000	1.54%	41.87%	0.04%***	3.08%	0.69%***	2.3%**	2.31%	7.01%*	13.88%	1.92%	18.44%	7.06%*	1.54%	41.87%	86.97%
2001	1.92%	18.68%	10.82%	2.30%	7.13%*	7.22%*	2.68%	2.38%**	0%***	3.07%	0.71%***	2.27%**	2.30%	7.13%*	0%***
2002	0.77%	69.23%	86.92%	3.45%	0.19%***	11.10%	5.36%	0%***	0%***	3.07%	0.71%***	1.77%**	6.51%	0%***	1.26%**
2003	1.53%	42.26%	56.87%	0.38%	25.22%	79.23%	0.77%	69.23%	51.90%	0.38%	25.22%	87.85%	0.38%	25.22%	60.64%
2004	0.38%	24.99%	79.37%	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*
2005	0.38%	25.44%	90.42%	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*
2006	0.38%	25.44%	74.37%	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*
2007	0.38%	25.22%	48.48%	1.15%	81.27%	64.28%	1.15%	81.27%	65.22%	0.38%	25.22%	86.39%	0.00%	2.2%**	7.26%*
2008	8.40%	0%***	0%***	9.16%	0%***	0%***	5.34%	0%***	0%***	9.16%	0%***	0%***	4.20%	0.01%***	0.06%***
2009	0.77%	69.23%	90.60%	2.30%	7.13%*	50.46%	0.77%	69.23%	63.06%	2.30%	7.13%*	43.34%	0.77%	69.23%	92.74%
2010	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2011	0.77%	69.67%	26.30%	1.54%	41.87%	53.12%	1.15%	80.77%	73.83%	1.54%	41.87%	47.71%	2.31%	7.01%*	17.54%
2012	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2013	1.15%	81.27%	47.61%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2014	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2015	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	84.82%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2016	1.53%	42.26%	24.13%	0.38%	25.22%	94.08%	0.77%	69.23%	92.74%	0.38%	25.22%	64.95%	0.38%	25.22%	38.32%
2017	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*
1998-2002	1.00%	99.11%	2.17%**	3.22%	0%***	0.01%***	3.68%	0%***	0%***	2.91%	0%***	0%***	3.22%	0%***	0%***
2003-2007	0.61%	13.09%	94.53%	0.31%	0.32%***	99.83%	0.38%	1.05%**	99.19%	0.15%	0.01%***	99.99%	0.08%	0%***	99.93%
2008-2012	1.99%	0.15%***	0%***	2.61%	0%***	0%***	1.46%	12.12%	0%***	2.61%	0%***	0%***	1.46%	12.12%	15.43%
2013-2017	0.54%	6.52%	93.72%	0.08%	0%***	100.00%	0.23%	0.08%***	100.00%	0.08%	0%***	99.96%	0.08%	0%***	99.60%
1998-2017	1.04%	80.01%	0%***	1.55%	0.02%***	0%***	1.44%	0.29%***	0%***	1.44%	0.29%***	0%***	1.21%	14.45%	10.66%

	S&P/TSX60			Dow Jones			NASDAQ			AEX			HSI		
	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point	Exc. Rate	p-value UC	p-value Saddle-point
1998	6.51%	0%***	0%***	3.83%	0.05%***	0.05%***	6.13%	0%***	0%***	5.36%	0%***	0%***	3.07%	0.71%***	6.54%*
1999	1.92%	18.68%	56.70%	0.77%	69.23%	75.07%	3.07%	0.71%***	6.21%*	1.15%	81.27%	17.21%	0.00%	2.2%**	7.26%*
2000	4.62%	0%***	0%***	3.46%	0.18%***	2.07%**	10.00%	0%***	0%***	0.38%	25.44%	64.77%	0.77%	69.67%	31.20%
2001	1.53%	42.26%	9.6%*	2.30%	7.13%*	3.74%**	3.07%	0.71%***	3.97%**	2.30%	7.13%*	0%***	0.38%	25.22%	48.48%
2002	0.38%	25.22%	89.22%	2.30%	7.13%*	15.96%	0.00%	2.2%**	7.26%*	4.60%	0%***	0.32%***	0.00%	2.2%**	7.26%*
2003	0.00%	2.2%**	7.26%*	0.38%	25.22%	83.11%	0.00%	2.2%**	7.26%*	1.53%	42.26%	26.66%	0.00%	2.2%**	7.26%*
2004	0.38%	24.99%	74.68%	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*
2005	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*
2006	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*
2007	0.77%	69.23%	85.67%	0.77%	69.23%	79.22%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	2.68%	2.38%**	5.2%*
2008	10.31%	0%***	0%***	8.40%	0%***	0%***	6.49%	0%***	0%***	4.20%	0.01%***	0%***	10.31%	0%***	0%***
2009	1.92%	18.68%	60.44%	1.92%	18.68%	73.62%	0.77%	69.23%	66.16%	0.77%	69.23%	90.60%	0.77%	69.23%	90.60%
2010	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	94.08%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2011	0.38%	25.44%	87.76%	1.15%	80.77%	48.23%	1.15%	80.77%	27.65%	1.15%	80.77%	84.53%	1.15%	80.77%	56.02%
2012	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2013	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2014	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2015	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	79.23%	0.77%	69.23%	62.09%
2016	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	84.82%	0.38%	25.22%	48.48%	0.38%	25.22%	94.08%
2017	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*	0.00%	2.22%**	7.33%*
1998-2002	2.99%	0%***	0%***	2.53%	0%***	0.02%***	4.45%	0%***	0%***	2.76%	0%***	0%***	0.84%	55.96%	55.55%
2003-2007	0.23%	0.08%***	99.96%	0.23%	0.08%***	99.97%	0.00%	0%***	0%***	0.31%	0.32%***	99.02%	0.54%	6.52%*	91.46%
2008-2012	2.53%	0%***	0%***	2.30%	0.01%***	0%***	1.76%	1.25%	0.02%***	1.23%	42.80%	0.01%***	2.45%	0%***	0%***
2013-2017	0.00%	0%***	0%***	0.00%	0%***	0%***	0.08%	0%***	100.00%	0.15%	0.01%***	99.52%	0.23%	0.08%***	99.94%
1998-2017	1.44%	0.29%***	0%***	1.27%	6.45%*	0.04%***	1.57%	0.01%***	0%***	1.11%	42.55%	0%***	1.02%	90.83%	8%*

Table 7: Backtesting Results for Volatility Adjusted Historical Simulation VaR

This table summarizes the performance of the Vol. Adj. Historical Simulation model in constructing one-day-ahead Value-at-Risk forecasts for ten different market indices, according to two distinct backtests: Kupiec's Unconditional Coverage and Wong's Saddle-point Backtest. The results displayed below regard to the exceedance rate (light blue coloured) and p-values of the tests, and the symbols [*; **; ***] represent the rejection of the VaR model at the 10%, 5% and 1% significance level, respectively.

	NIKKEI225			S&P500			FTSE100			MSCI World Index			DAX30		
	Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value	
		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point
1998	2.68%	2.38%**	2.78%**	2.30%	7.13%*	8.75%*	2.30%	7.13%*	21.72%	2.30%	7.13%*	8.41%*	3.07%	0.71%***	14.66%
1999	3.07%	0.71%***	6.32%*	0.38%	25.22%	86.40%	1.15%	81.27%	47.32%	1.15%	81.27%	46.92%	0.77%	69.23%	75.83%
2000	3.08%	0.69%***	0%***	0.77%	69.67%	13.06%	1.15%	80.77%	0.05%***	1.15%	80.77%	0.01%***	0.38%	25.44%	74.38%
2001	1.15%	81.27%	11.83%	0.77%	69.23%	54.78%	1.15%	81.27%	15.53%	0.38%	25.22%	48.49%	0.38%	25.22%	71.76%
2002	3.45%	0.19%***	0.11%***	0.00%	2.2%**	7.26%*	0.38%	25.22%	86.40%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*
2003	0.77%	69.23%	18.14%	0.38%	25.22%	79.24%	0.38%	25.22%	84.83%	0.38%	25.22%	89.23%	0.38%	25.22%	91.70%
2004	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*	0.76%	68.79%	40.28%	0.00%	2.17%**	7.18%*	0.38%	24.99%	74.69%
2005	0.00%	2.22%**	7.33%*	0.38%	25.44%	92.69%	0.77%	69.67%	51.31%	0.00%	2.22%**	7.33%*	0.77%	69.67%	19.99%
2006	0.38%	25.44%	83.00%	2.69%	2.34%**	22.64%	1.15%	80.77%	42.81%	1.15%	80.77%	41.87%	1.54%	41.87%	0.01%***
2007	2.68%	2.38%**	0.06%***	4.21%	0.01%***	0%***	3.07%	0.71%***	0.99%***	1.92%	18.68%	3.6%**	1.53%	42.26%	0.01%***
2008	1.53%	42.65%	0.01%***	1.15%	81.76%	10.68%	1.53%	42.65%	0.01%***	2.67%	2.43%**	2.6%**	1.91%	18.91%	0.02%***
2009	1.15%	81.27%	40.30%	1.15%	81.27%	54.42%	1.53%	42.26%	24.13%	2.68%	2.38%**	15.48%	1.92%	18.68%	22.85%
2010	0.77%	69.23%	84.29%	1.92%	18.68%	10.05%	0.77%	69.23%	49.40%	1.92%	18.68%	4.59%**	1.15%	81.27%	75.45%
2011	2.31%	7.01%*	16.39%	1.92%	18.44%	9.78%*	0.38%	25.44%	84.71%	1.54%	41.87%	33.13%	0.77%	69.67%	50.49%
2012	0.77%	69.23%	63.08%	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	55.36%	0.77%	69.23%	32.73%
2013	0.00%	2.2%**	7.26%*	0.38%	25.22%	74.54%	0.77%	69.23%	20.13%	0.77%	69.23%	29.05%	1.53%	42.26%	10.56%
2014	0.00%	2.2%**	7.26%*	0.77%	69.23%	65.21%	1.53%	42.26%	38.30%	0.77%	69.23%	51.92%	0.77%	69.23%	33.87%
2015	1.53%	42.26%	0.02%***	1.15%	81.27%	50.52%	1.15%	81.27%	12.19%	1.53%	42.26%	14.58%	1.53%	42.26%	44.34%
2016	0.38%	25.22%	89.23%	1.15%	81.27%	0.01%***	1.53%	42.26%	33.25%	1.53%	42.26%	0%***	1.15%	81.27%	0.03%***
2017	0.00%	2.22%**	7.33%*	0.77%	69.67%	51.99%	0.77%	69.67%	9.5%*	0.77%	69.67%	51.99%	0.38%	25.44%	86.30%
1998-2002	2.68%	0%***	0%***	0.84%	55.96%	44.72%	1.23%	42.63%	1.52%**	1.00%	99.11%	0.83%***	0.92%	76.92%	86.09%
2003-2007	0.77%	37.76%	13.78%	1.53%	7.25%*	2.49%**	1.23%	42.63%	18.64%	0.69%	23.37%	69.44%	0.92%	76.92%	0%***
2008-2012	1.30%	29.37%	1.2%**	1.23%	42.80%	12.83%	0.84%	55.78%	3.54%**	1.84%	0.64%***	1.74%**	1.30%	29.37%	1.53%**
2013-2017	0.38%	1.05%**	32.53%	0.84%	55.96%	3.31%	1.15%	59.42%	5.95%*	1.07%	79.18%	0.17%***	1.07%	79.18%	1.14%**
1998-2017	1.28%	4.81%**	0%***	1.11%	42.55%	0.64%***	1.11%	42.55%	0.09%***	1.15%	28.73%	0.02%***	1.05%	69.63%	0%***

	S&P/TSX60			Dow Jones			NASDAQ			AEX			HSI		
	Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value		Exc. Rate	p-value	
		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point		UC	Saddle-point
1998	3.45%	0.19%***	2.64%**	2.30%	7.13%*	12.90%	2.30%	7.13%*	3.58%**	1.15%	81.27%	22.36%	1.53%	42.26%	43.84%
1999	0.77%	69.23%	46.21%	0.00%	2.2%**	7.26%*	1.15%	81.27%	37.01%	0.77%	69.23%	65.87%	0.38%	25.22%	91.70%
2000	1.54%	41.87%	43.36%	1.54%	41.87%	14.12%	0.77%	69.67%	0.07%***	1.15%	80.77%	0.02%***	0.77%	69.67%	15.76%
2001	0.77%	69.23%	26.66%	0.77%	69.23%	15.88%	0.38%	25.22%	60.65%	1.15%	81.27%	17.10%	0.77%	69.23%	47.45%
2002	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.00%	2.2%**	7.26%*	0.38%	25.22%	84.83%	0.00%	2.2%**	7.26%*
2003	0.38%	25.22%	92.76%	0.38%	25.22%	84.83%	0.00%	2.2%**	7.26%*	0.38%	25.22%	74.54%	0.77%	69.23%	36.51%
2004	0.38%	24.99%	77.16%	0.00%	2.17%**	7.18%*	0.00%	2.17%**	7.18%*	0.76%	68.79%	70.60%	1.15%	81.76%	77.55%
2005	0.77%	69.67%	78.25%	0.77%	69.67%	85.44%	0.38%	25.44%	84.71%	1.15%	80.77%	48.15%	1.92%	18.44%	20.70%
2006	1.54%	41.87%	34.37%	1.92%	18.44%	8.51%*	1.15%	80.77%	18.31%	1.54%	41.87%	13.22%	1.54%	41.87%	19.28%
2007	1.92%	18.68%	6.84%*	4.21%	0.01%***	0%***	3.07%	0.71%***	0%***	1.15%	81.27%	15.50%	2.68%	2.38%**	15.14%
2008	3.82%	0.05%***	3.78%**	1.15%	81.76%	33.01%	2.67%	2.43%**	4.64%**	1.15%	81.76%	15.33%	1.91%	18.91%	17.24%
2009	0.38%	25.22%	92.76%	1.15%	81.27%	77.41%	1.53%	42.26%	18.71%	1.53%	42.26%	33.15%	0.38%	25.22%	60.65%
2010	0.38%	25.22%	87.87%	1.92%	18.68%	8.4%*	2.68%	2.38%**	6.17%*	1.15%	81.27%	82.91%	0.00%	2.2%**	7.26%*
2011	0.77%	69.67%	73.99%	1.92%	18.44%	15.39%	1.92%	18.44%	5.71%*	0.38%	25.44%	91.62%	1.15%	80.77%	15.55%
2012	0.38%	25.22%	74.54%	0.38%	25.22%	84.83%	0.00%	2.2%**	7.26%*	1.15%	81.27%	52.54%	0.77%	69.23%	70.57%
2013	0.77%	69.23%	46.98%	0.38%	25.22%	71.76%	0.38%	25.22%	83.12%	1.53%	42.26%	12.54%	1.15%	81.27%	23.43%
2014	1.53%	42.26%	17.74%	1.15%	81.27%	56.26%	0.77%	69.23%	72.69%	1.92%	18.68%	0%***	1.53%	42.26%	59.70%
2015	0.77%	69.23%	76.63%	1.15%	81.27%	60.14%	1.53%	42.26%	42.71%	1.92%	18.68%	0%***	1.15%	81.27%	0.02%***
2016	1.15%	81.27%	0.03%***	0.77%	69.23%	0.02%***	1.15%	81.27%	0.01%***	1.53%	42.26%	0.03%***	2.30%	7.13%*	0.03%***
2017	0.38%	25.44%	60.46%	0.38%	25.44%	64.78%	1.54%	41.87%	0.02%***	0.38%	25.44%	86.30%	0.77%	69.67%	62.39%
1998-2002	1.30%	29.24%	22.80%	0.92%	76.92%	33.04%	0.92%	76.92%	1.52%**	0.92%	76.92%	1.53%**	0.69%	23.37%	73.86%
2003-2007	1.00%	99.11%	62.85%	1.46%	12.05%	2.31%**	0.92%	76.92%	2.5%**	1.00%	99.11%	37.23%	1.61%	4.18%**	19.32%
2008-2012	1.15%	59.62%	74.27%	1.30%	29.37%	30.54%	1.76%	1.25%**	1.96%**	1.07%	79.39%	70.34%	0.84%	55.78%	54.02%
2013-2017	0.92%	76.92%	2.68%**	0.77%	37.76%	6.75%*	1.07%	79.18%	0.02%***	1.46%	12.05%	0%***	1.38%	19.18%	0.01%***
1998-2017	1.09%	50.79%	18.76%	1.11%	42.55%	1.53%**	1.17%	23.15%	0%***	1.11%	42.55%	0%***	1.13%	35.19%	1.31%**