Value-at-Risk estimates

TRAN MANH HA

Doctor of Philosophy

ASTON UNIVERSITY

September 2017

Value-at-Risk estimates

Tran Manh Ha

A thesis presented for the degree of

Doctor of Philosophy

2017

This thesis consists of three empirical essays on the Value-at-Risk (VaR) estimates. The first empirical study (Chapter 2) evaluates the performance of bank VaRs. The second empirical study (Chapter 3) investigates the predictive power of various VaR models using bank data. The third empirical study (Chapter 4) explores VaR estimates with high-frequency data.

The first study examines the performance of VaR estimates at seven international banks from 2001 to 2012. Using statistical tests, we find that bank VaRs were conservatively estimated in pre-crisis and post-crisis periods. During financial crisis, while some banks continued to overstate their VaRs, the others significantly underestimated their risk. The potential causes of the poor performance of bank VaRs are also discussed.

The second study investigates the predictive power of various VaR models using bank data. We find that the GARCH-based models are superior in estimating bank VaRs in both normal and crisis periods. We conclude that good VaR estimates at banks can be obtained using simple, accessible models rather than the complicated approach or banks' internal model. Thus, we argue that VaR should not be blamed for misleading risk estimates during financial crisis.

The third study evaluates VaR estimates using 5-minute sampling data of WTI Futures. First, we acknowledge the value of high-frequency data on the measure of volatility to characterize the quantile forecast of asset returns. Second, we find that quantile combination can improve the forecast accuracy. With the VaR implication, we show that VaR combination provides more accurate and robust results than individual VaR estimates.

Key words: Value at Risk, Backtesting, Commercial banks, Quantile forecasts.

Acknowledgement

Completing a PhD program has been the greatest challenge to my academic and personal life to date. My PhD journey would not have gone this far without the great support of people around me.

First and foremost, I'm greatly indebted to my supervisor Dudley Gilder. Thank you for having confidence in me and giving me a chance to pursue the PhD at Aston University. Your commitment, your academic suggestions and encouragement have contributed to my academic as well as personal development. I also cannot express enough gratitude toward my supervisor Nathan Joseph. Your patience, your deep academic insights and your advice helped me much in improving the quality of the PhD thesis. I would like to thank Margaret Woods for her guidance, support and encouragement in the first year of my PhD. I would like to thank the Research Degree Program coordinators, especially Jeanette Ikuomola and Ranjit Judge for their great support throughout my PhD. Finally, I would like to thank Vietnam International Education Development for their financial support of my PhD.

I also indebted to Dr Kevin Evans and Dr Rakesh Bissoondeeal, the examiners of my viva voce. Their advices and comments help me much in improving not only the thesis but also my academic career. I also grateful for the helpful comments from the participants of the 2015 Financial Engineering and Banking Society. I would like to thank Marcelo Fernandez for his discussions and suggestions at the 2016 Vietnam International Conference in Finance. I also thank the participant and discussants of the 2017 Vietnam International Conference in Finance.

Last but not least, I would like to thank my family, without whom it would be impossible for me to reach this stage. I am thankful to my father Dung, my mother Phuong, my sister Ngan for their understanding, support and encouragement. I am grateful to my wife, Mai Tran, who always stay beside me with unconditional love and understanding. Finally, I am grateful to my little son, Bobby, to whom I dedicate this thesis. You may not be aware, but your smiles, your love have motivated me to go through the end of this journey.

Table of content

List of Tables

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

CAViaR .... Conditional Autoregressive Value at Risk
CC .... Conditional Coverage
CQOM .... Conditional quantile optimization method
DQ .... Dynamic Quantile
EWMA .... Exponentially Weighted Moving Average
EVT .... Extreme Value Theory
GARCH .... Generalized Autoregressive Conditional Heteroskedasticity
GPD .... Generalized Pareto Distribution
HAR .... Heterogenous Autoregressive
HEAVY .... High Frequency-based Volatility Model
HS .... Historical Simulation
IND .... Independence
LF .... Loss Function
LQR .... Linear Quantile Regression
LR .... Likelihood Ratio
LS .... Logarithmic Score
MLE .... Maximum Likelihood Estimation
MRA .... Market Risk Amendment
MUC .... Multivariate Unconditional Coverage
P/L .... Profit and Loss
POT .... Peak over Threshold
QS .... Quantile Score
RV .... Realized Volatility
UC .... Unconditional Coverage
VaR .... Value at Risk
VCV .... Variance - Covariance
WQS .... Weighted Quantile Score
WTI .... West Texas Intermediate

Chapter 1: Introduction

During the 1970s and 1980s, a number of financial institutions built internal models to measure, aggregate and manage exposures across their business lines. However, as the activities of financial institutions became more and more complex, aggregating exposures across business lines become increasingly difficult because of the high correlations amongst risk factors. Financial institutions also lacked the means to managing risk across progressively diverse positions. The absence of integrated risk management demanded tools that measure the probability of loss at institutions-wide level. This led to the development of Value-at-Risk (VaR). VaR is a comprehensive solution to the problem of how to measure the risk taken by an increasingly complicated global bank.

The first VaR model was developed in the early 1990s in the back-office of JP Morgan. Following an order from the CEO to develop a system that measures risk across different trading positions, a single risk measure was developed. Early in 1994, JP Morgan introduced its Riskmetrics service — a simplified version of their internal VaR model. Later that year, JP Morgan published Riskmetrics system, and gave free access to it on the internet. The promotion of Riskmetrics provided a major boost to the ideas surrounding VaR system. Indeed, the adoption of VaR was rapid, amongst security firms, investment and commercial banks and other financial institutions. The VaR concept became increasingly popular, and by the mid-1990s, it was regarded as the dominant measure of market risk (Down, 2005).

VaR aims to capture the market risk of a trading portfolio. VaR is a numerical measure that determines the maximum potential loss on a portfolio within a given time horizon, and using a given level of confidence (Jorion, 2006). The concept is attractive and transparent and is widely regarded as a benchmark for measuring market risk. Beside profitability, a manager only needs to worry about the regulatory boundary measured by VAR, at the tails of their profit and loss (P/L) distribution. At institutional level, VaR has been used for risk management, and in public disclosure and financial reporting, as well as in the computation of regulatory capital requirements. Accurate VaR estimates are crucial for financial institutions as misleading VaR estimates can lead to sub-optimal capital allocation.

VaR is also used as a regulatory tool for ensuring the soundness of the financial systems. In 1996, the Basel Committee on Banking Supervision (BCBS) issued the Market Risk Amendment (MRA) to the first Basel Capital Accord, placing a milestone on the use of VaR. Indeed, the MRA allows financial institutions to use their internal VaR model to measure and disclose market risk. Following the BCBS support for VaR, regulators demanded that all financial institutions estimate and disclose their VaR measures in their financial reports.[1] The first time VaR was recognised in financial regulation was in 1997, when the U.S. Securities and Exchange Commission ruled that public corporations must disclose quantitative information about their derivatives trading and derivatives position. Major banks and dealers implemented the ruling by including VaR information in the notes to their financial statements. The Basel II Accord, which came into effect in January

[1] In the first pillar of Basel II, for market risk measurement the preferred approach is VaR.

2007[2], also strongly promotes VaR estimates as the preferred market risk management approach.[3]

VaR has some key drawbacks. Most VaR estimates rely on the normality assumption even if the observations from the empirical distribution is not normal. Risk is a matter of the behaviour in the extreme tails of a distribution. As Greenspan (1997) notes, that the biggest problem with risk management is the measurement of the fat-tailedness of a distribution. In fact, the occurrence of the recent financial crisis has partly been attributed to the failure to acknowledge the role of the fat-tails in VaR estimates (Triana, 2009). The use of VaR has also been blamed for providing little warning of the potential loss for banks during crisis periods (Nocera, 2009). Although details regarding the poor performance of bank VaRs is not new, it is surprising that VaR estimates of banks have received very little attention in empirical work. Therefore, this thesis aims to provide an empirical evaluation on the performance of VaR estimates provided by banks and the factors that affect the reliability of their VaR estimates.

This thesis consists of three empirical essays on the VaR estimates at banks. Chapter 2 contributes to the VaR literature by investigating the performance of bank VaRs for a set of international banks. Our dataset includes the daily trading P/L and VaR of seven commercial banks from January 2001 to December 2012, covering the

---

[2] Basel II came into effect in the European Union on 1 January 2007 under the Capital Requirements Directive (CRD) and all lenders covered by the CRD have had to implement it from the beginning of 2008. The US delayed this date to January 2009.

[3] The Basel II Capital Accord is a set of recommendations on banking regulation that is applicable to all banks in order to stimulate the improvement of risk management. Clients' commitment to Basel II compliance can be demonstrated to regulators through their evidence of systematic VaR backtesting. In the first pillar of Basel II, for market risk the preferred approach is specified as Value at Risk.

pre-crisis, financial crisis and post-crisis periods. Using the coverage tests, we find that bank VaRs were conservatively estimated in pre-crisis and post-crisis periods. During financial crisis, while some banks continued to overstate their VaR, the others remarkably understated their risk. We quantify the VaR distortions for seven sample banks and find that the VaR overstatement/understatement levels at large banks are more serious than small banks. We also find evidence of extreme losses during financial crisis which probably exceed the market risk capital requirements of banks. We attribute the poor performance of bank VaRs to three main causes: the use of contaminated data, the choice of VaR model and the benefit of VaR overstatement. The distortions of bank VaRs, which are popular across banks, make VaR a poor risk management tool.

The second empirical study (Chapter 3) investigates the forecasting power of VaR models using dataset of trading P/L of seven banks from January 2001 to December 2012. We compare the performance of internal VaR model at banks to alternative VaR approaches, including the Historical simulation (HS), the Variance-Covariance (VCV) and the Extreme Value Theory (EVT) approaches. To compare model performance, we develop a two-stage backtesting that examines the absolute and comparative performance of VaR models. The empirical analysis shows two main points. First, we find that the alternative VaR models can easily outperform banks' internal model in both normal and crisis periods. Second, we document the superiority of the GARCH-type models in providing good VaR estimates at banks. While the HS models perform inconsistently, none of the banks' internal model accurately capture the bank risk. The EVT approach, which was shown to be superior in VaR estimation with market data, performs very poorly with bank data.

Thus, we argue that good bank VaRs can be obtained using simple and accessible models rather than other sophisticated models or internal VaR model at banks.

The third empirical study (Chapter 4) explores VaR estimates with high-frequency data. This chapter provides the first assessment of quantile combinations, using high-frequency data. In this study, we use the 5-minute sampling data of WTI Crude Oil Futures as oil market is important for desk level trading at banks. The findings of this study are twofold. First, we acknowledge the value of the high-frequency data on the measure of volatility to characterize the quantile forecast of asset returns. Second, we find that the use of quantile combination can improve the accuracy of quantile forecasts. To find whether the use of high-frequency data can improve quantile forecast accuracy, we compare the performance of the GARCH(1,1) models to the Realized Volatility (RV) - based models, including the Heterogenous Autoregressive model (HAR-RV) of Corsi (2009), the High-frEquency-bAsed VolatilitY (HEAVY) model of Shephard and Sheppard (2010) and the RV-based Linear quantile regression (LQR-RV) of Zikes and Barunik (2016). Evaluating their absolute and comparative performance, we find that the HEAVY and HAR-RV model outperform the GARCH(1,1) models across quantile levels and forecast horizons. To to examine the power of quantile combination in improving forecast accuracy, this chapter uses the Conditional quantile optimization method (CQOM) of Halbleib and Pohlmeier (2012) to combine individual quantile forecasts. We find that the combined forecasts are superior stand-alone forecasts in providing accurate and robust results, not only at 1%-quantile (VaR), but also across all quantile thresholds and forecast horizons.

The thesis proceeds with three empirical studies in Chapter 2, Chapter 3 and Chapter 4, while Chapter 5 summarizes the findings of the thesis.

Chapter 2: Value-at-Risk models and commercial banks

2.1 Introduction

In the financial industry, Value-at-Risk (VaR) has become a standard risk measurement technique in finance. VaR specifically aims to capture the market risk of portfolios (Jorion, 2006), which is regarded as the maximum potential loss for a given period, normally one-day-ahead, using a certain level of confidence, typically the 95% or 99%. The VaR idea provided a comprehensive solution to identify an acceptable level of risk for an increasingly complicated global bank. In 1992, JP Morgan introduced its Riskmetrics service, in which it published the methodology and gave free access to the estimates of the necessary underlying parameter. Since then, the use of VaR has been promoted widely. However, only when the Basel Committee put forward the Market Risk Amendment (MRA) to the Capital Accord in 1996 made VaR become a benchmark for measuring market risk. VaR became a part of regulatory banking in 1997, when the U.S. Securities and Exchange Commission ruled that public corporations must disclose quantitative information about their derivatives trading activity. Major banks and dealers chose to implement the rule by disclosing VaR information on their financial statements.

The requirement of VaR disclosure has some main objectives. Firstly, it presents an aggregated estimate of the market risk value under taken by a bank. VaR also presents asymmetric information about the bank to market participants and investors. Secondly, result of VaR estimates can be converted into a capital charge to provide an adequate cushion for cumulative losses caused by adverse market conditions. More importantly, the disclosure of VaR estimates allows financial regulators to examine the validity of bank's internal VaR models. An assessment of these internal VaR models is provided in this thesis, using a procedure called "backtesting".

Backtesting is a technique which aims to investigate the forecasting power of a VaR model by periodically comparing the VaR estimates generated by the model with the actual P/L (or "trading outcome").[4] Our test is performed using daily data. Backtesting therefore identifies situations where VaR is underestimated (or a VaR exception), meaning that the portfolio has experienced a loss greater than the estimated VaR. The comparison between the risk measures with trading outcome, simply means that the financial institution counts the number of VaR exceptions. That is, the number of occasions that losses exceed the estimated VaR. The frequency of VaR exceptions is then compared with the intended level of coverage to assess the performance of the risk model.

An estimate of the number of VaR exception is asymmetric in nature. Recall that a VaR exception occurs only when the risk is underestimated. Therefore, simply over-estimating VaR can reduce the number of VaR exceptions. This creates an incentive to overstate VaR. Indeed, the current Basel backtesting framework relies on the number of VaR exception to evaluate the performance of internal VaR models.[5] Specifically, banks are penalized if their VaR model produces too many

---

[4] BCBS, 1996b

[6] A trading book consists of positions in financial instruments and commodities held either with trading intent or in order to hedge other elements of the trading book. To be eligible for trading book capital treatment, financial instruments must either be free of any restrictive covenants on their tradability or able to be hedged completely. In addition, positions should be frequently and accurately valued, and the portfolio should be actively managed (Basel Committee on Banking Supervision, 2004)

exceptions. However, there is no capital charge for banks that have no VaR exceptions. Therefore, even using the Basel backtesing framework incentivizes banks to overstate their VaR estimates.

Investigating the accuracy of bank VaRs is an important area of academic research. It is obvious that VaR models are only useful when they accurately forecast risk. If not, an inaccurate VaR estimate can cause financial institutions to underestimate (or overestimate) their risk, thereby providing misleading indicators of risk. The performance of VaR models also has direct impacts on the calculation of the market capital risk. According to the MRA, the accuracy of VaR models results in the use of a multiplier to convert VaR estimates into the minimum capital requirement for market risk. Banking regulators only allow a bank's internal VaR model to be used for regulatory capital computation, if the model provides satisfactory backtesting results. Specifically, a VaR model that fails a backtest will be reviewed and will either be disallowed in computing regulatory capital, or be subject to high capital multiplier.

In the literature, bank VaRs perform variously. As Lucas (2001) notes, banks have incentives to under-report VaR estimates to lower their cost of capital, although this can lead to an increase in the probability of VaR exceptions. On the other hand, there are evidences showing that banks excessively overstate their VaR estimates (see Berkowitz and O'Brien, 2002; Perignon et al., 2008; Perignon and Smith, 2010; O'Brien and Szerszen, 2014) as they want to minimise the likelihood of having many VaR exceptions to avoid reputational costs. Bank VaRs are also controversial since they cannot outperform VaR forecasts produced by simple econometric models, such as GARCH(1,1) (Berkowitz and O'Brien, 2002; Perignon et al., 2008; O'Brien and Szerszen, 2014).

The recent global crisis raised a number of questions regarding to the reliability of VaR estimates at financial institutions. Indeed, VaR was blamed for the financial crisis, as it dangerously produces too low risk figures by borrowing the past data and using the improper probabilistic assumptions (Triana, 2009). Although the poor performance of bank VaRs is not new, there has been little empirical study on this topic, due to the proprietary nature of the P/L and VaR data. Indeed, the performance of bank VaRs in crisis period has only been examined by O'Brien and Szerszen (2014) with the empirical evidence of US banks. To the best of our knowledge, there has been no empirical study on the accuracy of bank VaRs on international level. Besides, the performance of bank VaRs in post-crisis period still has not been covered in the literature.

Chapter 2 contributes to the literature as the first investigation on the performance of bank VaRs covering the pre-crisis, financial crisis and post-crisis periods. Our dataset includes the daily P/L and VaR of seven commercial banks from 2001 to 2012. Instead of focusing on a specific country, this chapter investigates bank VaRs on international level. Our dataset includes Scotia Bank, Royal Bank of Canada (Canada), Banca Intesa (Italy), Banco Santander (Spain), Societe Generale (France), Deutsche Bank (Germany) and Bank of America (USA). Compared to prior studies, we use longer and more diversified dataset. The rich dataset not only increases the power of the statistical tests, but also allows us to have a comprehensive evaluation of bank VaRs in international perspective. Besides, this study employs the innovative Multivariate Unconditional Coverage (MUC) test of Colletaz et al. (2013) to examine the presence of extreme losses that exceed VaR at very low coverage rates e.g. VaR(0.07%), VaR (0.03%) or even VaR(0.01%). To the best of our knowledge, the investigation of extreme losses during financial crisis has not been studied in the literature.

Our empirical analysis shows the poor performance of bank VaRs. We find that bank VaRs tend to be conservatively estimated in normal periods. During financial crisis, while some banks continue to overstate their VaR, the others significantly underestimate their risk. In case of VaR understatement, the number of VaR exceptions are excessively high and tend to cluster together. We find evidence of extreme losses during financial crisis that can exceed VaR at very low coverage rates e.g. VaR(0.07%), VaR(0.03%) or even VaR(0.01%). We attribute the poor performance of bank VaRs to the use of contaminated data, the choice of VaR model and the benefit of VaR overstatement. Due to the fact that the performance of bank VaRs does not improve overtime, we argue that banks accept their inferior VaR models and conservatively estimate their risk to have the economic merit of VaR overstatement. The manipulations of bank VaRs, which are popular across banks, make VaR a poor risk management tool.

Chapter 2 is presented as follows: Section 2.2 introduces background of VaR and their backtesting. Section 2.3 reports the preliminary analysis of sample banks and their daily trading P/L. Section 2.4 presents the empirical investigation of bank VaRs. Section 2.5 discusses the potential causes of the poor performance of bank VaRs, while Section 2.6 summarizes the main results of the chapter.

2.2 Value-at-risk and Backtesting Value-at-risk

2.2.1 Value-at-risk background

From an economic point of view, risk is uncertainty about future outcomes. Risk is a key element in the financial world, since firms usually make investment decision under uncertainty as they do not know whether predicted cash flows will materialize. From the 1970s, risk management has become one of the most important tasks of financial institutions since market volatility has increased remarkably after the collapse of Bretton Woods system. Besides, the unexpected catastrophes in 1990s, such as the demise of Barings Bank, Orange County, Long Term Capital Management after significant changes in market conditions, highlighted the demand of effective market risk management. The severity of the recent global financial crisis has also raised the importance of risk management.

The idea of VaR was originated in the late 1970s, when a number of financial institutions developed their internal models to measure and aggregate risk across business lines as a whole. As institutions were becoming more global and complex, with the development of financial products, the aggregation of risks became more demanding when taking into account of how they correlate with each other. During the late 1980s, JP Morgan set up their firm-wide risk management system based on portfolio theory that could model several hundred risk factors and aggregate them into a single financial risk measure. Their measure named Value-at-Risk, the maximum value that can be lost over the next trading day, placed a milestone on the development of market risk measurement. In 1992, at a time of global concerns about leverage and derivatives, JP Morgan introduced their Riskmetrics service to the public, comprising a detailed technical document and a covariance matrix for several hundred key factors, which was updated daily. With the launch of Riskmetrics, VaR was publicized and has been promoted widely.

VaR was originally developed to measure the maximum expected loss caused by movements in the volatility of asset prices for a given portfolio of financial assets. As Linsmeier and Pearson (1996) note, VaR presents as the loss which is expected to be surpassed with the given percent of probability during the next **t**-day periods. According to Alexander (2009), VaR is the potential loss that will not be exceeded if the given portfolio is held over some periods of time. With a given significant level α and set $p = 1- \alpha$ as level of confidence, and denote q_α as the α -quantile of the P/L of a portfolio over a holding period h_t, then the $VaR_{\alpha,ht}$ of the portfolio is defined as:

$$VaR_{\alpha,ht} = q_\alpha \tag{2.1}$$

The magnitude of VaR depends on two fundamental parameters: the significant level α (or 1- α level of confidence) and the holding horizon h_t. For instance, suppose a bank discloses their VaR on trading portfolio is $50 million with 99% level of confidence and the forecast horizon is one-day-ahead. It means that over the next working day, there is a 99% probability that if the bank suffers a loss, its magnitude will not exceed $50 million. In other words, there is only 1% probability that over the next working day, the loss will be greater than $50 million.

The significance level α is usually set by an external body (banking regulator, credit agency). Under the Basel Accord, banks are required to use 1% significant level (or 99% level of confidence) to determine their market risk capital requirement. Besides, credit rating agencies may be stricter in setting their significance level, which means a higher confidence level (e.g. 99.7% or 99.9%). In the absence of a regulated financial environment and external agencies, the confidence level for the VaR model can depend on the attitude to risk of managers. The more conservative the risk manager, the lower the value of α, i.e. the higher the level of confidence applied (Alexander, 2009). Dowd (1998) indicates that if investors want to validate a VaR model, the high level of confidence (such as 99% or 99.5%) should be avoided to be able to observe enough VaR exceptions. On the other hand, the risk appetite of senior management plays a significant role in selecting the level of confidence. Besides, the choice of risk horizon h_t also depends on the nature of each asset's volatility and degree of liquidity. For the investors who actively make money via trading in equity market, typically the 1-day risk horizon is appropriate, whereas institutional users and non-financial institutions may prefer longer risk horizon (Linsmeier and Pearson, 1996).

One of the main attractions of VaR is that it is presented in a simplest way with most understood unit of measure. By simplifying the assumptions used in its computation, VaR aggregates the diversification effects, leverage effects and probabilities of adverse price movements to single dollar value that is suitable for application and communication with interested parties. Besides, VaR has number of attractions over traditional risk measures. Firstly, VaR can be universally applied to any types of portfolio. It allows us to compare the risks between two different portfolios. Secondly, VaR enables us to aggregate the risk of desk-level portfolios into an overall measure of portfolio risk while taking into account the way different risk factors interact with each other. This characteristic is especially important, as financial institutions are exposed to a variety of market risks. Besides, VaR is probabilistic, which provides risk managers helpful information on the probability associated with specified loss. Thus, VaR is regarded as the most important quantitative risk measure used by financial institutions (Woods et al., 2008).

VaR also has limitations. The main drawback is that as a measure of market risk, it tells you nothing about the potential loss that would break the VaR threshold when extreme event occurs. A VaR figure at 99% level of confidence has no information about how much the unexpected loss might be in the 1% tail of the P/L distribution. The omission of VaR in order to take into account the magnitude of exceeded loss makes it unable to differentiate between two positions being given the same VaR figure but have very different risk exposures. Besides, the dependency on the normality assumption is other serious limitation of VaR. In normal market condition, VaR works reasonably well. But VaR is not reliable when financial markets are excessively volatile and during financial crises (Dowd, 2005).

VaR information can be used in various ways. First, senior risk managers can use VaR to set their overall risk target, and from that determine risk figures and position limits for each business line. If they want to increase the risk of financial institution, they would increase the overall VaR threshold or, vice versa. Secondly, VaR can be useful to calculate capital requirements, both at the institutional scale and business-unit level with the driving principle: the riskier the trading activity, the higher the VaR figure and the greater the capital requirement. Besides, VaR figures can be used for the purpose of reporting and disclosing financial risk, and financial institutions increasingly make a point of disclosing VaR information in their annual reports (Dowd, 2000; Jorion, 2002; Woods et al*.,* 2008). VaR is also informative not only for making decisions of investment, trading, hedging and portfolio management, but also setting limits on the rule that apply when rewarding managers in order to discourage excessive risk taking (Dowd, 1999; 2005).

2.2.2 Basel Capital Accord and market risk management

The Basel Committee on Banking Supervision (BSBS) was originated during the period of financial market turmoil after the breakdown of the Bretton Woods system in 1973. Aiming to improve financial stability by enhancing supervisory knowhow and the quality of global banking supervision, the Committee sets minimum standard of supervision and regulations of banks, where capital adequacy becomes the main focus. Backed by G10 Governors, the first Basel Capital Accord (Basel I) was published in 1988 to set up a minimum ratio of bank's capital to risk-weighted assets of 8% and has always been improved overtime to make it adapt to current financial situations. It is worth to emphasize that this framework only covers the credit risk in evaluating capital adequacy. Under this accord, all assets in bank's balance sheet are given a judgemental risk classification with a fixed risk weight from 0 to 100 percent.

Various amendments were made over years to extend the definition and effects of credit risk (in 1991 and 1995 Amendments), but the most important one was the MRA to the Capital Accord issued in 1996. These evolutions took place after the number of failures in risk management (e.g. Orange County, Procter and Gamble) that raised the concern about financial derivatives and called for effective regulation of financial market. During that period, industry also issued a number of best practice reports of how to manage the risk of financial derivatives, one of those was appeared in G30 report in 1993. This influential report then became the handbook of financial risk management discipline, and its best practice guidelines were globally accepted by both industry and regulators, which were also presented in 1996 Amendment. In particular, the 1996 MRA specified banks' exposure to interest rate related instruments and equities in the trading book[6] and commodities risk and foreign exchange risk throughout the bank. In this Amendment, banks have option to choose either the standardised method or internal model based approach for meeting market risk capital requirements. The first option specifies the measurement of risk for each category from foreign exchange, equities, commodities and derivatives.

In MRA, banks are allowed to use their internal VaR models to measure their market risk, which is subject to strict qualitative and quantitative standards. In internal model approach, banks are flexible in developing their own models, but required to follow minimum standards for the purpose of their capital charge. For some main points, VaR must be calculated on a daily basis with the minimum holding period of 10 trading days[7] at 99th percentile and one-tailed confidence interval. A minimum length of one year for historical data is needed to compute VaR, which is also subject to the update and revision requirement no less frequently than once every three months. Banks that use internal model approach are also subject

[6] A trading book consists of positions in financial instruments and commodities held either with trading intent or in order to hedge other elements of the trading book. To be eligible for trading book capital treatment, financial instruments must either be free of any restrictive covenants on their tradability or able to be hedged completely. In addition, positions should be frequently and accurately valued, and the portfolio should be actively managed (Basel Committee on Banking Supervision, 2004)

[7] Banks can use VaR numbers computed regarding to shorter holding periods scaled up to ten days by the square root of time (Basel Committee on Banking Supervision, 1996a)

to use comprehensive stress testing program to identify extreme events that could significantly impact banks. The issue of MRA to the Capital Accord made the popularity of VaR reach to its peak, where VaR model was globally accepted as a benchmark of market risk management and only few people could recognize its weaknesses. With the adoption of MRA, it was the signal that bank regulators accepted and stood on the principle of risk-based regulation.

By second half of 1990s, banking activities were becoming more sophisticated and risk modelling was quickly evolving that called for a major advancement of Basel Accord. In 1999, the new capital adequacy framework was proposed to better capture bank's risk taking and reflect the financial innovation. The Revised Capital Framework, generally known as Basel II, was officially released in 2004. Basel 2 consists of three pillars: (i) The first pillar: minimum capital requirements, which sought to develop and expand the standardised rules set out in the 1988 Capital Accord; (ii) The second pillar: supervisory review of capital adequacy and internal assessment process; and (iii) The third pillar: effective use of disclosure to enhance market discipline and encourage sound banking practices.

Basel II made a major enhancement in credit risk modelling and for the first time, covered the operational risk within the first pillar of the accord, but for market risk measurement there was no noticeable changes from Basel I to Basel II. In Basel II, the market risk disclosure requirement was clearly presented in the third pillar in both qualitative and quantitative aspects. For banks using standardised approach, the disclosure of capital requirements for all sources of risk (e.g. interest rate risk, equity position risk, foreign exchange risk and commodity risk) is needed.

Banks using internal model approach are required to disclose their qualitative information about the features of models employed, the description of portfolio's stress testing and the characteristics of backtesting process. In term of quantitative disclosure, they are required to present VaR values (e.g. high, mean and low VaR) over the reporting period and period-end, coming with the comparison of VaR estimates with actual P/L and the analysis of remarkable VaR exceptions.

According to the Basel capital accord, VaR estimates are based on one of two theoretical assumptions about a trading portfolio. First, the trading portfolio is assumed to be rebalanced over the risk horizon to keep the risk factor sensitivities or asset weights constant. Second, the trading portfolio is assumed to be held static such that no trading occurs during the period. However, both assumptions are unrealistic. In practice, trading portfolios are actively managed, and the actual P/L on the trading portfolio is not equal to the hypothetical P/L, on which VaR estimates are based. Indeed, the hypothetical P/L is the mark-to-market P/L, while the actual P/L includes all the P/L amounts from intraday trading, fees and commissions (using actual prices). Therefore, it is important to regularly revise VaR models and their performance using backtesting.

2.2.3 Value-at-Risk at commercial banks

VaR has become the standard market risk measurement at commercial banks since the publication of the MRA. Allowed to use internal models to estimate VaR, banks can be flexible in improving their existing models although they are required to follow minimum standard.

The MRA has no legal force. Instead, it formulates supervisory guidelines and standards. Corresponding to the international guidelines, domestic financial regulators require commercial banks to compute and publicly disclose their VaRs in annual reports. For example, the US commercial banks are required to publicly disclose their market risk under Financial Reporting Release Number 48, issued by SEC in 1997. In Canada, the Office of the Superintendent of Financial Institutions has required financial institutions to compute and publicly disclose their VaR since the late 1990s. The disclosure of VaR has some main targets. First, VaR provides a universal measure of amount of market risks suffered by a bank, aiming to reduce the asymmetric information between the bank and the public. Second, VaR can be translated into capital charge, which is supposed to be a cushion for cumulative losses. In addition, VaR disclosure allows financial regulator to evaluate the performance of bank's internal VaR model through the process of "backtesting".

2.2.4 Backtesting Value-at-Risk

It is obvious that VaR model is only useful when it forecasts risk reasonably well. If not, it can lead financial institutions to underestimate (or overestimate) their risk and hence provide misleading managerial guidelines. Therefore, after constructing VaR model, it is important to regularly evaluate the adequacy of the VaR model used.

Backtesting is a technique which aims to investigate the forecasting power of the VaR model by comparing the risk estimates generated by the model against actual daily changes in portfolio value over longer periods of time. Backtesting therefore identifies situations where VaR has been underestimated, meaning a portfolio has experienced a loss greater than the original estimated VaR. The comparison between the risk measures with trading outcome simply means that the financial institution counts the number of VaR exceptions, a situation when a loss is greater than the estimated VaR. The frequency of VaR exceptions is then compared with the intended level of coverage to assess the performance of the risk model. The results of the backtesting can be informative to refine the model used for the VaR predictions, making them more accurate and reducing the risk of unexpected losses.

The backtesting framework recommended by Basel committee includes the use of one-day-ahead VaR estimates and one-day trading outcomes, excluding any intra-day revenues, fees and commissions. It is straightforward to implement the Basel backtest, simply by computing the frequency of VaR exceptions – the cases that the trading P/L are not covered by the VaR estimates. With approximately 250 trading days in a fiscal year and 99% level of confidence, banks are expected to have about three VaR exceptions per year. Depending on the actual number of VaR exceptions, the backtest result can be classified into one of three zones. That is Green zone (0 to 4 exceptions), Yellow (5 to 9 exceptions) or Red (10 exceptions or above). Different zones are subject to different level of capital charge. Thus, the more VaR exceptions, the higher the charge.

The three-zone backtesting approach of Basel committee has some drawbacks. First, it is specified for evaluating VaR models for one year, with approximately 250 observations. Second, it primarily assumes that each day's test outcome is independent from each other'. This assumption is not realistic, due to the fact that VaR exceptions tend to cluster together, especially during crisis period.

Besides, the Basel framework does not penalize banks that overstate their VaRs to maintain low number of VaR exceptions. From risk management perspective, the estimated return at the lower tail defines the amount of capital that banks allocate to cover the possible loss, which is called the economic capital for market risk. The low number of VaR exceptions means that the tail is overestimated and therefore signals an excessive amount of allocated capital.

To overcome these shortcomings but still preserve the frequency-based approach of Basel framework, this chapter uses several coverage tests to evaluate the performance of bank VaRs, including: (i) the Unconditional Coverage (UC) test of Kupiec (1995); (ii) the Conditional Coverage (CC) test of Christoffersen (1998); and (iii) the Multivariate Unconditional Coverage (MUC) test of Colletaz et al. (2013). Each test has different power. The widely-used UC test aims to quantify and investigate the frequency of VaR exceptions, while the CC test additionally examines the assumption of independence of VaR exceptions. Besides, the MUC test allows us to jointly examine the coverage and comparative magnitude of extreme losses.

2.2.4.1 The Unconditional Coverage test

As the name suggests, the UC test examines the coverage of VaR exceptions based on the actual number of VaR exceptions. It is worth noting that the frequency of VaR exceptions plays a crucial role in validating the adequacy of VaR estimates. The Group of Thirty (G30) recommends that financial institutions carry out reality checks to evaluate the performance of their VaR models.[8] Specifically, it is recommended that an institution's VaR estimates are compared against actual

[9] BCBS, 1996b

portfolio outcomes. In line with the G30's idea, the Basel Committee proposes a "traffic light" backtesting framework, which relies on the number of VaR exceptions, to verify the accuracy of internal VaR models.[9]

Let r_t denote P/L of a given portfolio at time t and $VaR_{t|t-1}(\alpha)$ the one-day ahead VaR forecast for a given α coverage rate conditional on an information set Ω_{t-1}. If the VaR estimates are adequate, the UC property states that:

$$\Pr\left[r_t < VaR_{t|t-1}(\alpha)\right] = \alpha \tag{2.2}$$

Let $I_t(\alpha)$ be a hit indicator variable associated with the VaR exception at time t, in which:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

The null hypothesis of UC property is that the observed frequency of VaR exceptions is consistent with the nominal frequency α, thus:

$$H_0\text{: } E\left[I_t(\alpha)\right] = \alpha \quad \text{and} \quad H_1\text{: } E\left[I_t(\alpha)\right] \neq \alpha \tag{2.4}$$

To satisfy the UC property, the likelihood of which actual P/L on day t exceeds its corresponding $VaR_{t|t-1}(\alpha)$ must be equal to α x 100%, or $\Pr[I_t(\alpha) = 1] = \alpha$. The null hypothesis is rejected if the difference between observed frequency of VaR exceptions and the expected rate α, is statistically significant. Therefore, there are two occasions when UC hypothesis can be rejected. The first is the case of overestimation, when the realized frequency of VaR exceptions is smaller than the nominal rate α, or $\Pr[I_t(\alpha) = 1] < \alpha$. The second is when the VaR underestimation

[9] BCBS, 1996b

occurs and the realized frequency of VaR exceptions is higher than the nominal rate α, or Pr $[I_t(\alpha) = 1] > \alpha$.

Let T_0 and T_1 be the number of zeros and ones corresponding to the hit indicator variable (2.3). Let T be the total number of observations, thus $T = T_0 + T_1$. The Log-likelihood ratio statistics of the UC hypothesis, denoted as LR_{UC}, is given by:

$$LR_{UC}(\alpha) = -2\ln[(1-\alpha)^{T_0}\alpha^{T_1}] + 2\ln\left[(1-\frac{T_1}{T})^{T_0}(\frac{T_1}{T})^{T_1}\right] \rightarrow \chi^2(1) \quad (2.5)$$

The LR_{UC} converges to an asymptotically Chi-square distribution with one degree of freedom. Recall that UC text examines whether the empirical frequency of VaR exceptions N/T is sufficiently close to the predicted α. The null hypothesis H_0 is not rejected if the empirical frequency of VaR exceptions is close enough to the coverage rate α. There are some examples of non-rejection regions for the UC test. If the sample size T = 250 (one year), for a 5% nominal size the UC assumption is not rejected if the total number of VaR(1%) violations is strictly from 1 to 6. In case T = 500, the total number of exceptions must range from 1 to 11 in order to satisfy the UC hypothesis.

2.2.4.2 The Conditional Coverage test

The CC test extends the UC test by jointly examining the IND and UC properties of VaR exceptions. As Christoffersen (1998) notes, the IND hypothesis holds when VaR exceptions occurring at two different times for the same coverage rate are independently distributed. In other words, the hit indicator $\{I_t(\alpha) = 1\}$ corresponding to VaR exception at time t, at coverage rate α, is independent from hit process $\{I_{t-k}(\alpha) = 1\}$ associated with the VaR exception at time t-k, $\forall k \neq 0$. This assumes that past exceptions do not hold information on current and future VaR exceptions. Thus, the rejection of the IND hypothesis may imply a clustering of VaR exceptions.

The LR test statistic for IND hypothesis is given by:

$$LR_{IND} = -2\ln[(1 - \frac{T_1}{T})^{T_0} (\frac{T_1}{T})^{T_1}] + 2\ln[(1 - \hat{\pi}_{01})^{T_{00}} \hat{\pi}_{01}^{T_{01}} \rightarrow \chi^2\ (1) \qquad (2.6)$$

where $T_{i,j}$, i, j =0, 1 is the number of observations with a j following a i in the hit indicator variable I_t, and $\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}}$. The LR_{IND} follows Chi-square distribution with one degree of freedom.

The CC test simultaneously examines the UC and CC hypotheses. The test statistic for CC hypothesis is presented as:

$$LR_{CC} = LR_{UC} + LR_{IND} \rightarrow \chi^2\ (2) \qquad (2.7)$$

The LR_{CC} test statistic converges to the Chi-square distribution with two degrees of freedom. The null hypothesis of CC holds when VaR model satisfies both UC and IND hypotheses. Therefore, the rejection of the CC hypothesis may come from the rejection of the UC and/or IND tests. The CC test allows us to test each hypothesis separately to see whether the model provides incorrect coverage or causes violation clustering.

Although being widely applied by both academia and practitioners, the CC has the key drawback, as it does not account for the magnitude of excessive losses. Specifically, the CC test cannot identify and evaluate the presence of extreme loss, which is far beyond normal VaR. This drawback can be overcome with the use of MUC test.

2.2.4.3 The Multivariate Unconditional Coverage test

The MUC test was proposed by Colletaz et al. (2013). The innovation of the MUC is that it gives another view to the performance of a risk model by jointly examining the number and magnitude of extreme losses, in case that an actual loss might not only violate the normal VaR defined with a coverage rate α (i.e. 1%), but also exceed VaR at lower probability α' (i.e. 0.2%).

In this approach, a VaR exception is the state when $r_t < VaR_t(\alpha)$, while a VaR super exception take places when $r_t < VaR_t(\alpha')$, with α' smaller than α. In other words, a super exception for VaR defines a situation of excessive losses which not only exceed normal VaR(α), but also break VaR at rare coverage rate, VaR(α'). Therefore, if the frequency of super exceptions is remarkably high, this means the magnitude of the losses associated with VaR(α) is too large.

Estimation of VaR(α')

Firstly, we demean the P/L series by subtracting the unconditional mean μ from each observation to get the new return series u_t, with $E(u_t) = 0$. To estimate VaR(α') of the P/L series, we firstly estimate VaR(α') of the demeaned series u_t. Then, the VaR(α') of the P/L series is obtained by adding the unconditional mean μ to the VaR(α') of the demeaned series u_t. We present the method to estimate VaR(α') of the zero-mean return series u_t as following.

As $E(u_t) = 0$, the conditional VaR of the demeaned series u_t can be presented as a function of the conditional variance of the return series, denoted as h_t:

$$VaR_{t|t-1}(\alpha,\beta) = -\sqrt{h_t}\, F^{-1}(\alpha;\beta) \tag{2.8}$$

where $F^{-1}(\alpha;\ \beta)$ is the α-quantile of a standardized conditional P/L distribution, which is assumed to be parametric associated with a set of parameters β.

Colletaz et al. (2013) propose a method to extract VaR(α') from VaR(α). Given the disclosed bank VaRs, VaR(α), an implied P/L conditional variance is defined as:

$$\sqrt{\hat{h}_t} = -\frac{VaR_{t|t-1}(\alpha)}{F^{-1}(\alpha; \hat{\beta})} \tag{2.9}$$

With the idea of option pricing, VaR(α') is now obtained by:

$$VaR_{t|t-1}(\alpha', \hat{\beta}) = -\sqrt{\hat{h}_t}\, F^{-1}(\alpha; \hat{\beta}) = VaR_{t|t-1}(\alpha)\, \frac{F^{-1}(\alpha'; \hat{\beta})}{F^{-1}(\alpha;\ \hat{\beta})} \tag{2.10}$$

To implement this method, we need two ingredients: (i) the auxiliary model for the conditional volatility h_t and (ii) the P/L distribution which is conditional on a set of parameters β. We follow Colletaz et al. (2013) to employ GARCH as auxiliary model and Student t distribution for $F^{-1}(\alpha; \hat{\beta})$. Therefore, the set of parameters β, which corresponds to the degree of freedom of the Student t distribution, can be obtained by Quasi-Maximum Likelihood estimation. This calibration procedure was shown to provide robust VaR(α') estimates (Colletaz et al., 2013).

The Log-likelihood ratio test of MUC hypothesis

Colletaz et al. (2013) propose a LR test to examine the MUC hypothesis. Let I_t (α) be a hit indicator variable corresponding to the $VaR_t(\alpha)$:

$$I_t(\alpha) = \begin{cases} 1 \ \text{ if } r_t < -VaR_{t|t-1}\,(\alpha) \\ 0 \qquad\qquad \text{otherwise} \end{cases} \tag{2.11}$$

and I_t (α') denotes a hit indicator variable associated with $VaR_t(\alpha')$:

$$I_t(\alpha') = \begin{cases} 1 \ \text{ if } r_t < -VaR_{t|t-1}\,(\alpha') \\ 0 \qquad\qquad \text{otherwise} \end{cases} \quad \text{with } \alpha' < \alpha \tag{2.12}$$

It is possible to jointly test the MUC null hypothesis of VaR exceptions and VaR super exceptions:

$$H_0: E[I_t(\alpha)] = \alpha \quad \text{and} \quad E[I_t(\alpha')] = \alpha' \tag{2.13}$$

Denote $J_{1,t}$ and $J_{2,t}$ hit indicator variables, and $J_{0,t} = 1 - J_{1,t} - J_{2,t} = 1 - I_t(\alpha)$ where:

$$J_{1,t} = I_t(\alpha) - I_t(\alpha') = \begin{cases} 1 & \text{if } -VaR_{t|t-1}(\alpha') < r_t < -VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

$$J_{2,t} = I_t(\alpha') = \begin{cases} 1 & \text{if } r_t < -VaR_{t|t-1}(\alpha') \\ 0 & \text{otherwise} \end{cases} \tag{2.15}$$

Now $\{J_{i,t}\}_{i=0}^{2}$ are the Bernoulli random variables that take value of 1 with given probabilities 1-α, α-α' and α' respectively. Let $N_i = \sum_{t=1}^{T} J_{i,t}$ be the count variable corresponding to each $J_{i,t}$ variable. The Log-likelihood ratio test statistics for MUC null hypothesis (LR_{MUC}) is given by:

$$LR_{MUC}(\alpha, \alpha') = -2\ln[(1-\alpha)^{N_0}(\alpha-\alpha')^{N_1}(\alpha')^{N_2}]$$

$$+ 2\ln\left[\left(\frac{N_0}{T}\right)^{N_0}\left(\frac{N_1}{T}\right)^{N_1}\left(\frac{N_2}{T}\right)^{N_2}\right] \xrightarrow{d} \chi^2(2) \tag{2.16}$$

where T is the total number of VaR estimates. The LR_{MUC} statistic converges to Chi-square distribution with two degrees of freedom.

2.3 Literature on the performance of bank VaRs

Berkowitz and O'Brien (2002) provide the first evidence on the performance of VaR models at banks. Using private daily data of P/L and VaR of six large banking institutions from January 1998 to March 2000, they show that bank VaRs at the 99th percentile are conservative and in some cases, are extremely inaccurate. In particular, the average exception rate is less than a half of one percent across banks, and in some cases VaR estimates were considerably removed from the lower bound of the P/L. Berkowitz and O'Brien (2002) attribute the cause of VaR overstatement to the method of aggregating VaRs across the entire trading portfolio. Indeed, of the banks whose VaRs are more conservative, the global VaR is simply the summation of sub-VaRs across sources of risk. Besides, the regulatory need of higher capital requirement when banks fail the backtest may motivate banks to overstate their VaR. When comparing with the performance of the ARMA and GARCH as the benchmark models, the banks' VaR forecasts are not better because they could not adequately capture the changes in the volatility of P/L. In brief, Berkowitz and O'Brien (2002) argue that the reported VaR figures are not useful as measures of the actual risk of bank's portfolio.

Perignon et al. (2008) confirm the continuous conservativeness in VaR estimates of commercial banks. With the use of non-anonymous daily data of the six largest Canadian banks, they provide evidence of systematic VaR overstatement. Some banks experienced one exceedance, while others even had no VaR exceptions during the 6-year period, from 1999 to 2005. This result contrasts to the common wisdom that banks intend to understate their VaR to lower the market risk capital requirement (Cuoco and Liu, 2006). The VaR conservativeness at Canadian banks is consistent with US evidence reported by Berkowitz and O'Brien (2002). Besides, when compared with the simple HS and GARCH(1,1) models, bank VaR models are not superior in producing accurate risk forecasts. Thus, the internal models used by commercial banks do not provide VaR estimates that are reliable to determine capital charges.

Instead of using time series of daily P/L and bank VaRs, Perignon and Smith (2010) collect the number of VaR exceptions disclosed on banks' annual reports to evaluate the performance of bank VaRs. Using the dataset from 66 commercial banks, they confirm the systematic VaR overstatement at banks. Specifically, during the period of 1996-2005, US largest banks experienced only 10 actual exceptions in comparison to an expected exception of 68. In 2005, Canadian banks even overstated their VaR figures more seriously: only 1 exception over the expectation of 13, while in case of international banks, it is 3 over 53 respectively.

O'Brien and Szerszen (2014) investigate the specifications and performance of the VaR estimates of five anonymous US banks before and during financial crisis. The dataset used in their study include daily P/L and bank VaRs reported to the Federal Reserve Board. Their bank data are standardized to keep the anonymity, hence some valuable information about the absolute magnitude of daily P/L and VaR are absent. Consistent with the prior studies, O'Brien and Szerszen (2014) show that bank VaRs were excessively conservative before financial crisis (with very few VaR exceptions). For the 2007-2008 period, bank VaRs were substantially underestimated and exhibited exception clustering. Besides, it is evident that standard GARCH and HS models can produce more accurate VaR estimates compared to those provided by the banks.

It can be summarized that bank VaRs were conservative before financial crisis, but substantially underestimated during crisis period. Besides, prior studies mostly focus on the evaluation of bank VaRs in tranquil period, the condition that VaR normally works well. However, the recent financial crisis witnessed the failure of

VaR on unimaginable scale. Given that VaR has been widely criticised, it is surprising that the literature on the performance of bank VaRs during the financial crisis has only been investigated O'Brien and Szerszen (2014). This study provides empirical evidence for US banks. No study has examined the performance of bank VaR models during the post-crisis period. To fill the gap in the literature, this chapter aims to investigate the performance of bank VaRs for a set of banks across six countries. The analysis covers both crisis and post-crisis periods. Given that bank VaRs were intentionally and systematically overstated in the pre-crisis period, we do not expect that banks change their behaviour during and post-crisis period. The following sections investigate the performance of bank VaRs and present the results.

2.4 Empirical analysis

2.4.1 Data collection method and methodology

It is important to note that the daily VaR and P/L data we are concerned with is from the trading book of the bank. Positions of commercial banks can be categorized into two books, which all exposed to market risk: the banking book, including on-balance sheet and off-balance sheet activities, and the trading book, covering positions of instruments in traded market. However, trading book is exposed to market risk in a more transparent way, as its instruments will hold some direct exposure to market risk (Woods et al., 2008). As the VaR figures disclosed by commercial banks are the risk measure of their trading activities, thus we particularly focus on the market risk exposures of the trading book of a bank.

To perform the evaluation of bank VaRs, we require time series data of daily VaR and P/L of banks. However, the P/L and VaR data are one of the most confidential data of banks, which only report to top managers and financial regulators. We find that the sources of assessible VaR information are bank's annual reports and filings. Therefore, to obtain time series data from these sources, we decide to use the data extraction method, which allows us to convert the graph in banks' annual reports into time series data. Our data collection strategy is inspired by Perignon et al. (2008).

Initially we seek for these largest banks in North America and Europe countries. For each bank, we then look up whether they present a graph of daily VaR and trading revenues in their annual reports. We end up with a sample of banks that disclose VaR backtesting graph in their public documents. This includes seven commercial banks of six major countries: Royal Bank of Canada, Scotia Bank (Canada) Societe Generale (France), Bank of America (USA), Deutsche Bank (Germany), Intesa Sanpaolo (Italy) and Banco Santander (Spain). Their backtesting graphs are then put into our Matlab-based data extraction program.

The technique we used to collect daily VaR and P/L from banks' public data sources is simply the "Click and Collect" strategy. Firstly, the backtesting graphs are converted using Matlab to define the images as objects. The images are then re-sharpened and re-scaled to make them visibly clear. Adding the vertical lines to the image, we then zoom in, look up and precisely click on each data point. After one "click", we collect the two-dimensional Matlab coordinates of a data point, which are then converted into graph coordinates. Repeating this procedure, we obtain time-series of daily P/L and bank VaRs. In order to check the robustness of the data collection method, we plot the extracted data in Microsoft Excel and superimpose this graph with the original graph obtained from bank's annual report. If two graphs match perfectly, this means that our extracted data are accurate. Any mismatches between the two graphs can be fixed by manually adjusting the extracted data until a complete match is achieved.

We present the graphic example of our data collection method in Figure 2.1. The first is the original backtesting graph we collected from 2012 annual report of Scotia bank, including the series of daily P/L (red line) and daily VaR (yellow line), correspondingly. Using the Matlab-based data extraction program, the time-series of daily P/L and bank VaRs are generated from the upper graph. We then put the extracted data into Excel and plot the graph to compare with the former.

The data extraction program helps us collect the dataset of seven commercial banks, starting from 2001 to 2012. However, we cannot get the full dataset of several banks due to the lack of the backtesting graphs in banks' annual reports in some periods. This includes Intesa Sanpaolo (2001-2004), Banco Santander (2001-2004), Scotia bank (2001) and Societe Generale (2001, 2010-2012). The sample is then divided into three sub-periods: pre-crisis, crisis and post-crisis periods. The pre-crisis period is from the start of the period to May 2007, while the global financial crisis period is from June 2007 to June 2009. The end of the financial crisis period is determined using the National Bureau of Economic Research indicator as the start of the economic recovery[10]. The post-crisis period is after July 2009.

[10] Available at: http://www.nber.org/cycles.html

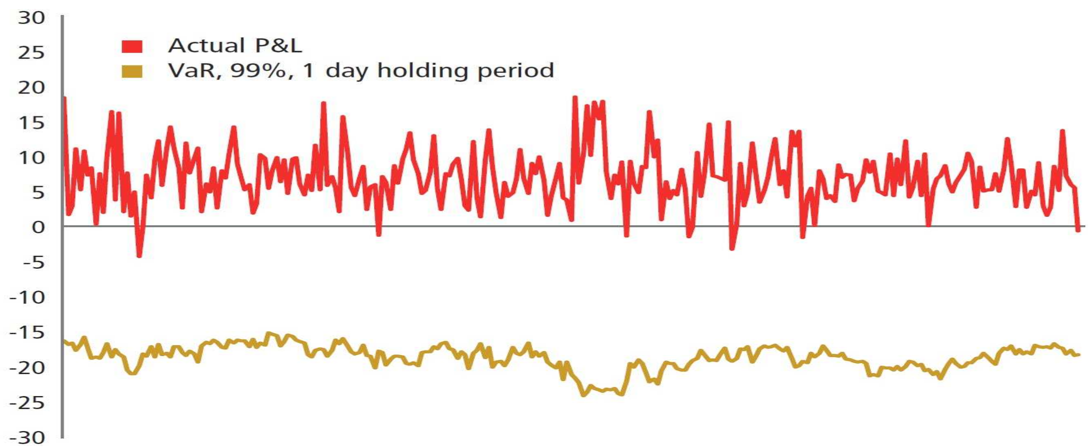
Prior studies focus on VaR performance of banks in a single country. Our study uses the largest data set to date and covers six developed countries. However, there might be potential sources of bias in our data sample. First, market movements in major countries tend to be highly correlated. Second, the choice of banks may not be representatives of the behaviour of all the banks in a specific country. To resolve this bias, future studies should use a larger set of banks across developed and developing countries.

It is also noteworthy that the disclosed bank VaR estimates are available in aggregate form from their financial reports. Constituents of the VaR estimates include interest rate, equity, foreign exchange, commodities and credit spread VaR. Although the nature of trading portfolio varies across banks, we find that interest rate changes is a major source of risk in VaR estimates, followed by equity, credit spread, commodities and foreign exchange rate changes. Commodities and foreign exchange VaR estimates become increasingly important to the aggregated bank VaR amount, especially during and after the financial crisis. Future studies should examine more closely the constituents of bank VaR estimates and simulate the trading portfolio, using market data.

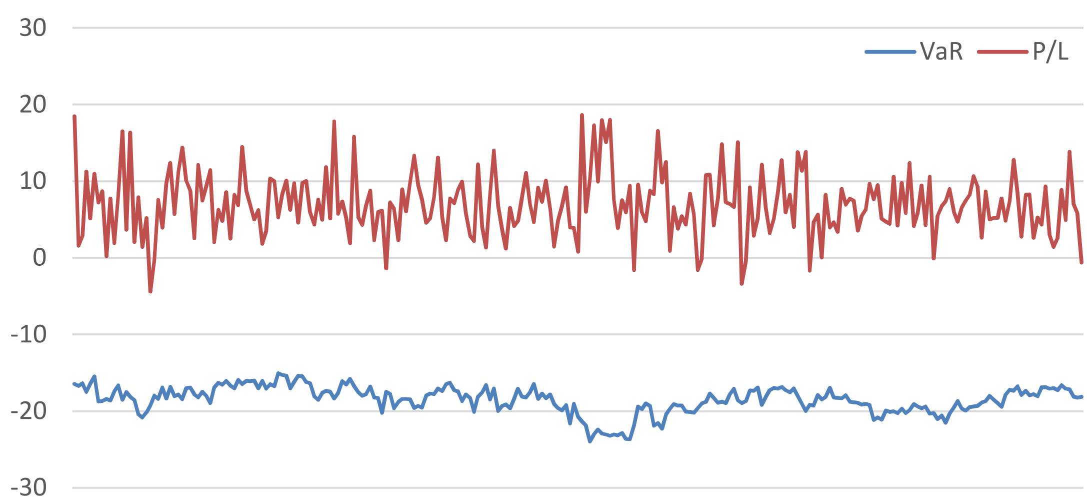
Figure 2.1: Example of data collection method

Notes: Figure 2.1 presents an example of our data collection method. The upper graph presents the daily VaR and P/L of Scotia bank in 2012, which was collected from bank's annual report. The time-series data of P/L and VaRs are generated from the original graph via our graphical data extraction program. To check the robustness of the extracted data, we superimpose the graphic presentation of time-series data, displayed in lower graph, to the original graph. If two graphs are perfectly matched, the extracted time-series data are robust.

Graphic presentation of trading P/L and VaRs in 2012 annual report of Scotia bank



Graphic presentation of extracted trading P/L and daily VaR of Scotia bank



2.4.2 Preliminary analysis of bank trading P/L

2.4.2.1 Descriptive statistics of sample banks

We present the descriptive statistics of seven commercial banks in Table 2.1. Most of figures are collected from banks' annual reports, except the data of banks' market capitalization are collected from Datastream. In Table 2.1, Panel A reports bank's market capitalization, total assets, total revenue, trading assets and P/L on trading portfolio. Panel B presents Tier 1, Tier 2 and Total regulatory capital (equal to Tier 1 plus Tier 1 minus adjustment), total risk-weight assets and the capital adequacy ratio (total regulatory capital divided by total risk-weighted assets). Panel C presents the internal VaR model at banks, the type of P/L data used in backtesting, the size of historical window and confidence level used to estimate VaR. The start date, end date and number of observations are also presented in Panel C.

Panel A reports the size of banks in our sample in some criteria: Market capitalization, total assets, total revenue, trading assets and trading P/L. These figures are presented in millions of local currencies, including US Dollar (Bank of America), Canadian Dollar (Royal Bank of Canada, Scotia Bank) and Euro (Deutsche Bank, Societe Generale, Intesa Sanpaolo, Banco Santander) and as of the end of fiscal year 2012. Based on the size of total assets, the largest banks in our sample are Deutsche Bank and Bank of America, while the smallest is Scotia Bank, after currency conversion. Besides, Deutsche Bank was also ranked fourth in the 2013 Global Finance's top 50 biggest banks in the world[11], while the smallest bank in our sample was ranked 42th in the list. The trading assets and trading P/L varies across banks, which reflects the size of trading activities of banks. Panel B presents the regulatory capital of banks, including the Tier 1 and Tier 2 capital, their risk-weighted assets and the total capital ratio. Complied with the minimum capital

[11] The report was based on the total assets as of the end of fiscal year 2012

requirement of Basel II[12], all banks maintain a generous amount of regulatory capital ratio, ranging from Societe Generale (12.7%) to Deutsche Bank (17.1%).

Panel C reports some brief descriptions of banks' internal VaR models. Among seven banks in our sample, six banks use Historical Simulation (HS) as their internal model, except Deutsche Bank using Monte Carlo Simulation. This fact is consistent with the finding of Perignon and Smith (2010), confirming that 73% of commercial banks favour Historical Simulation to estimate VaR. In consistent with Basel guidelines on VaR backtesting, all sample banks estimate and disclose their one-day-ahead VaR with 99% level of confidence on a daily basis. The length of moving window of historical data varies across banks. There are some banks using short windows of less than 300 daily historical observations to estimate VaR[13]. While Banco Santander and Royal Bank of Canada estimate their VaR based on 2 years of historical data, Bank of America use longest historical window as of 3 years. In HS approach, the length of historical window plays an important role to the accuracy of VaR estimates. Indeed, the small sample size could result in very few observations in the lower tail of the distribution and thus VaR. at 99% level of confidence, would be imprecise[14].

The time series of daily P/L and VaRs are mostly collected from January 2001 to December 2012. It is worth noting that the fiscal year of Canadian banks begins on 1st October and ends on 30th September next year. Thus, the time series data of

[12] According to the Pillar 1 of Basel 2, minimum capital requirements are set at 8% of the sum of risk-weighted assets.
[13] Intesa Sanpaolo and Deutsche Bank use one year of historical data (or 250 trading days equivalently). For Societe Generale and Scotia Bank, the window size is 260 and 300 respectively.
[14] As Alexander (2009) notes, even four years of daily historical data are not enough for stable VaR estimates.

Canadian banks in a calendar year are partly merged between two continuous fiscal years. For example, the time series trading P/L of Royal Bank of Canada in calendar year 2012 are the combination of the daily data from January 2012 to September 2012 of the fiscal year 2012, and the daily data from October 2012 to December 2012 of the fiscal year 2013.

Depending on each bank, the total number of trading days we collected range from 1,777 to 3,027, with the average of 2,525 daily observations per bank. Compared to prior studies in the literature (Berkowitz and O'Brien, 2002; Perignon et al., 2008; O'Brien and Szerszen, 2014), this study uses the longest dataset with the broadest sample of commercial banks investigated. This rich dataset gives us that advantage of rigorous backtesting VaR to demonstrate the power and accuracy of the statistical tests.

Table 2.1: Descriptive statistics of sample banks

Notes: Table 2.1 presents some descriptive statistics for seven commercial banks collected on 31st December 2012. Panel A reports bank's market capitalization, total assets, total revenue, trading assets and P/L on trading portfolio. Panel B presents Tier 1, Tier 2 and Total regulatory capital (equal to Tier 1 plus Tier 1 minus adjustment), total risk-weight assets and the capital adequacy ratio (total regulatory capital divided by total risk-weighted assets). Panel C presents the internal VaR model at banks, the type of P/L data used in backtesting, the size of historical window and confidence level to estimate VaR. The start date, end date and number of observations of banks are also presented in Panel C.

	Intesa Sanpaolo	Scotia Bank	Banco Santander	Bank of America	Royal Bank of Canada	Deutsche Bank	Societe Generale
Currency	mil €	mil C$	mil €	mil $	mil C$	mil €	mil €
Panel A: Sample banks							
Market capitalization	20,151.66	68,169.63	62,959.16	125,135.60	86,547.56	30,684.63	22,112.88
Total assets	673,472	668,225	1,282,880	2,209,974	825,100	2,012,329	1,250,696
Total revenue	17,881	19,646	44,553	83,334	29,772	21,490	23,110
Trading assets	63,546	87,596	177,917	465,836	120,783	245,538	484,206
Trading P/L	2,182	1,299	1,460	5,870	1,298	5,199	7,025
Panel B: Regulatory capital							
Tier 1 capital	36,133	34,450	57,558	155,461	36,807	50,483	40,499
Tier 2 capital	3,255	7,853	15,378	41,219	5,565	6,532	7,738
Total regulatory capital	39,388	42,303	72,936	196,680	42,372	57,015	41,308
Risk-weighted assets	298,619	253,309	557,030	1,206,051	280,609	333,605	378,495
Total capital ratio	13.60%	16.70%	13.09%	16.31%	15.10%	17.10%	12.70%
Panel C: Value-at-Risk							
Internal VaR model	Historical Simul.	Historical Simul.	Historical Simul.	Historical Simul.	Historical Simul.	Monte Carlo Simul.	Historical Simul.
Type of P/L data	Uncontaminated	Contaminated	Uncontaminated	Contaminated	Contaminated	Contaminated	Contaminated
Moving window	250 days	300 days	2 years	3 years	2 years	1 year	260 days
Confidence level	99%	99%	99%	99%	99%	99%	99%
Start date	Jan 2005	Jan 2002	Jan 2005	Jan 2001	Jan 2001	Jan 2001	Jan 2002
End date	Dec 2012	Dec 2012	Dec 2011	Dec 2012	Dec 2012	Dec 2012	Dec 2009
Number of observations	2017	2785	1777	3027	3020	3012	2035

2.4.2.2 Analysis of trading P/L of commercial banks

Recall that the trading P/L measures the daily gain or loss on the trading portfolio of a bank. As bank VaRs are relied on the assumption that the portfolio has no change during the holding period, the P/L presents the hypothetical changes in the trading portfolio assumed that end-of-day positions remain unchanged. Among seven commercial banks in our sample, six banks compute hypothetical P/L, except Scotia bank which report using actual P/L. As the use of different types of P/L can have impact on the backtesting result, therefore it is necessary to examine the properties of trading P/L disclosed by banks.

We follow Fresard et al. (2011) to define two types of P/L. The first type is Clean (or uncontaminated) P/L, which presents the change in the value of trading portfolio arising from previous-day positions. The second type is Contaminated P/L, which measures the change in trading portfolio based on previous day positions plus intraday revenues and/or any fees and commissions. To classify whether banks use contaminated or uncontaminated data, we rely on banks' annual report and carefully seek for any information related to trading P/L. Based on this information, we define bank as using uncontaminated P/L if they clearly state that fees, commissions and intraday revenues are excluded when generating trading P/L. Otherwise, the they are considered to be using contaminated data. Applying to seven banks in our sample, we find that there are only two banks reported using uncontaminated P/L. These include Intesa Sanpaolo and Banco Santander. The remaining five banks either reporting adding fees, incomes and intraday revenues or not mentioning these statements are considered as of using contaminated data.

The descriptive statistics of daily P/L of commercial banks are shown in Table 2.2. In each case, we show the 1%-quantile of the P/L distribution and the ratio of number of negative P/L days over total observations. The summary statistics indicate that during pre-crisis period, all banks experienced positive average trading P/L. For the financial crisis and post-crisis period, six over seven banks continued to have positive trend, except Banco Santander reports negative average trading P/L. The P/L on trading portfolio varies across banks. For example, Intesa Sanpaolo and Banco Santander exhibit many hypothetical trading losses, with the ratio of losses approximate 50%, while other banks keep low loss rates over the whole period. Besides, the magnitude of losses differs across banks. It can be seen from Figure 2.3 that visually Royal Bank of Canada suffers a small number, but excessively huge losses, while other Canadian banks experience a high number of losses with moderate magnitude.

Table 2.2: Analysis of daily P/L of sample banks

Notes: Table 2.2 presents the descriptive statistics of seven commercial banks from January 2001 to December 2012, divided into three sub-periods. The Jarque-Berra test statistics for normality are shown in the last column.

	Date	Mean	Min	Max	Std Dev	0.01 quantile	P/L < 0	Skewness	Kurtosis	JB test
Pre-crisis										
Intesa Sanpaolo	Jan 2005 - May 2007	0.1048	-23.370	22.796	5.5406	-18.3865	0.4802	-0.1056	6.6124	364.63**
Scotia Bank	Jan 2002 - May 2007	3.6668	-13.568	16.959	3.0208	-3.7391	0.0869	0.0390	4.4596	121.86**
Banco Santander	Jan 2005 - May 2007	1.9705	-66.886	38.366	8.0447	-19.9544	0.3762	-1.2819	15.8284	4321.2**
Bank of America	Jan 2001 - May 2007	15.4258	-57.388	96.759	14.4822	-21.4695	0.1059	0.1899	5.2973	364.63**
Royal Bank of Canada	Jan 2001 - May 2007	6.4211	-18.383	56.961	4.6188	-1.7968	0.0355	2.7184	26.5186	38942.6**
Deutsche Bank	Jan 2001 - May 2007	47.1337	-63.290	318.28	30.5518	-16.9191	0.0370	1.0950	8.5577	2371.4**
Societe Generale	Jan 2002 - May 2007	13.2450	-35.005	80.110	12.5048	-12.5724	0.1259	0.6977	5.4367	454.01**
Crisis										
Intesa Sanpaolo	June 2007 - June 2009	0.0509	-5.080	7.0252	1.7078	-3.7468	0.4838	0.3254	4.0149	31.975**
Scotia Bank	June 2007 - June 2009	5.2835	-36.936	36.5763	7.1991	-14.6100	0.1905	-0.0584	6.6055	284.66**
Banco Santander	June 2007 - June 2009	-1.1390	-89.445	61.0342	13.2276	-35.5109	0.5295	-0.5197	9.0367	820.79**
Bank of America	June 2007 - June 2009	22.239	-171.64	320.76	63.1064	-141.9280	0.3238	0.5986	5.3894	156.24**
Royal Bank of Canada	June 2007 - June 2009	5.3605	-730.00	296.00	55.2473	-281.6133	0.2514	-6.7171	76.4666	122014.8**
Deutsche Bank	June 2007 - June 2009	26.433	-360.69	571.746	86.9326	-272.9535	0.2781	-0.4254	7.8923	539.39**
Societe Generale	June 2007 - June 2009	3.1587	-275.22	128.44	43.5052	-160.5753	0.4343	-1.1347	9.3338	990.22**
Post-crisis										
Intesa Sanpaolo	July 2009 - Dec 2012	0.0663	-6.3547	10.114	1.8116	-4.5536	0.5169	0.5782	5.4293	267.21**
Scotia Bank	July 2009 - Dec 2012	6.3029	-14.574	27.211	5.0684	-6.2871	0.0920	0.1901	4.5185	90.974**
Banco Santander	July 2009 - Dec 2011	-0.0746	-40.096	63.035	9.3675	-29.1801	0.4582	-0.1495	7.1068	456.38**
Bank of America	July 2009 - Dec 2012	67.953	-119.75	317.29	58.0780	-62.5096	0.0923	0.5442	4.3031	106.66**
Royal Bank of Canada	July 2009 - Dec 2012	12.480	-91.314	179.74	15.8587	-31.7899	0.1178	1.2119	24.7167	17726.8**
Deutsche Bank	July 2009 - Dec 2012	7.3383	-128.85	226.66	33.9413	-79.5686	0.4215	0.6298	6.6070	542.52**
Societe Generale	July 2009 - Dec 2009	11.375	-57.398	70.434	23.2439	-38.8821	0.3281	-0.1174	2.8261	0.4554

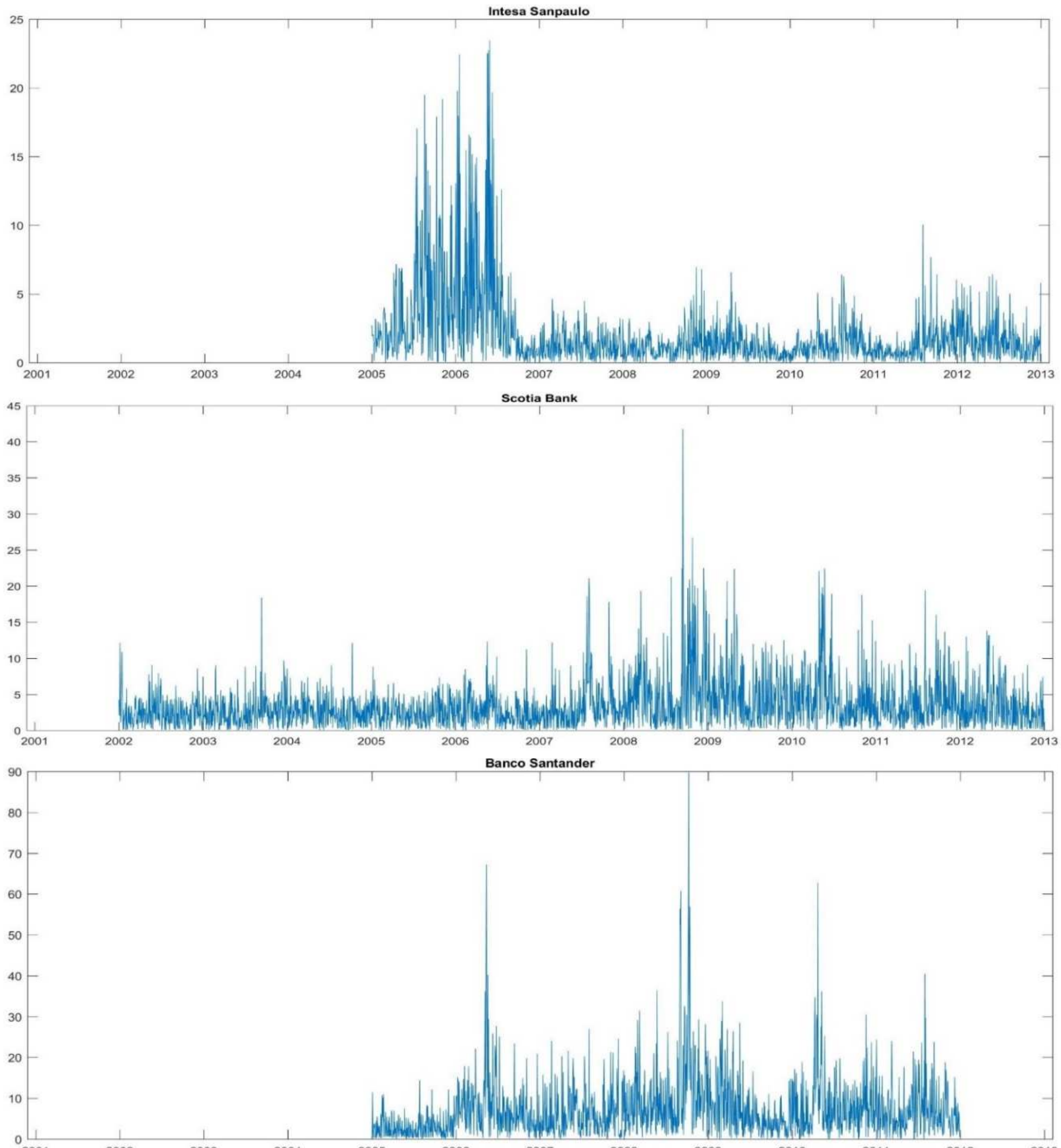
**: rejected at 95% level of confidence.

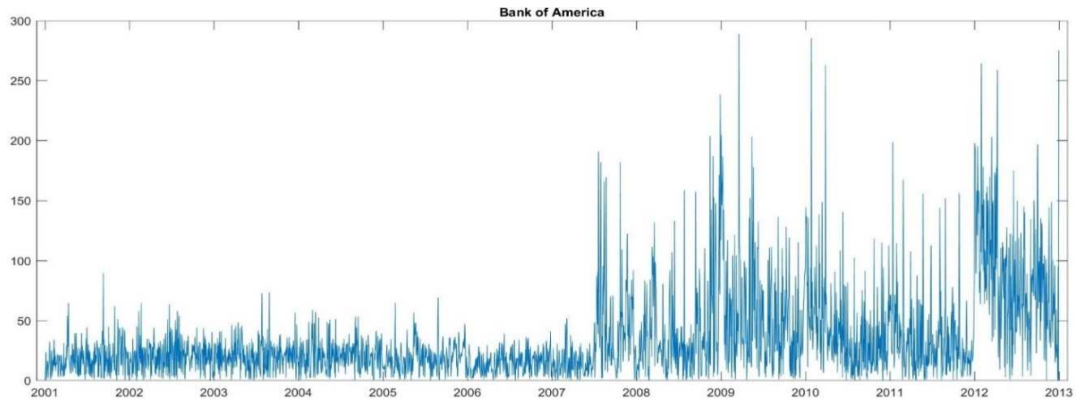
As presented in Table 2.2, the distribution of P/L is negatively skewed for financial crisis, while it is positively skewed in pre-crisis and post-crisis period. The high kurtosis of seven sample banks indicates that banks exhibit fat tails in their trading P/L distribution, which implies the presence of extreme losses. Moreover, the banks had higher mean and correspondingly low standard deviation in tranquil periods compared to those for the financial crisis. To visually present the volatility of daily P/L, we plot the absolute demeaned P/L series of seven commercial banks in Figure 2.2.

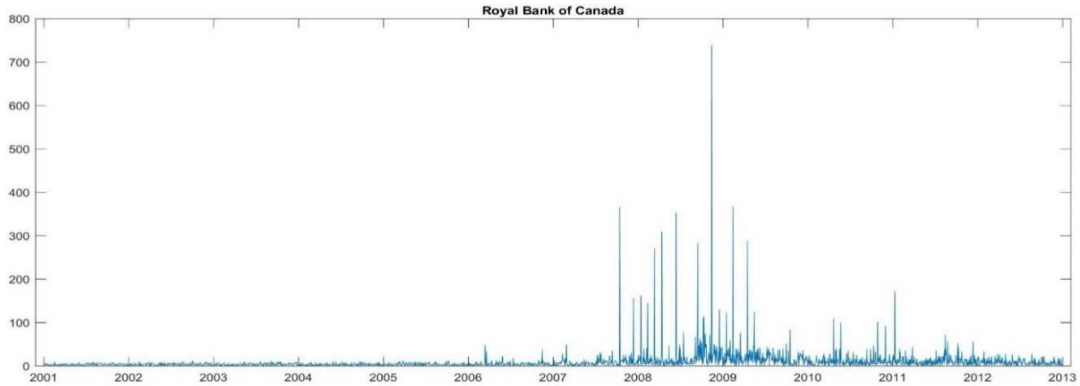
Figure 2.2 shows that from 2000s to June 2007, the trading P/L of t banks keeps stable. However, after June 2007 with the start of global financial crisis, the volatility of P/L increases significantly. We witness the visual evidence of extreme events which mostly occurred before the end of the financial crisis. Indeed, the extreme values are largest for Royal Bank of Canada, Deutsche Bank and Societe Generale. The crisis period also witnessed higher loss rates in trading portfolio compared with normal periods. Indeed, the average loss rate of seven sample banks in pre-crisis and post-crisis period are 0.1782 and 0.2895 respectively, while this ratio in crisis period roars to 0.3559. As a result, the 1%-quantile of trading P/L in crisis period is significantly higher than the tranquil periods in term of absolute value. Besides, the Jacque-Berra test statistics for normality of daily P/L are rejected at all cases.

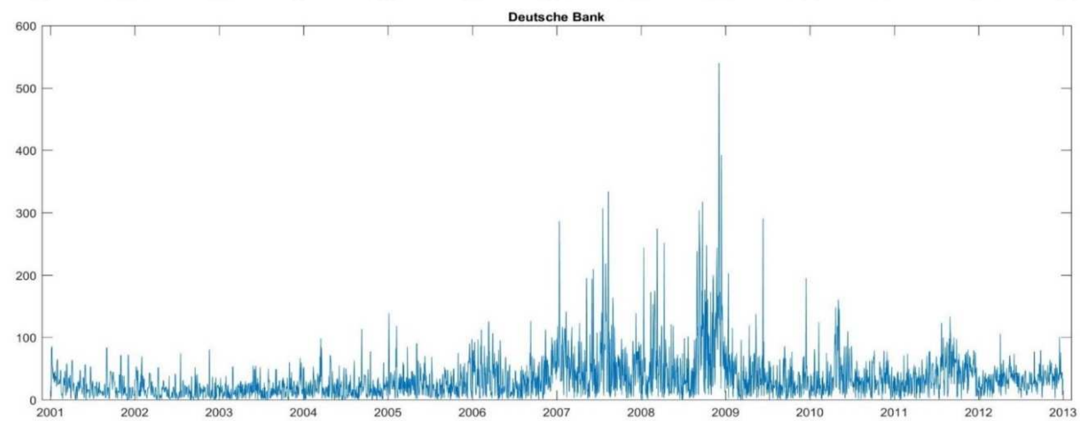
Figure 2.2: Absolute demeaned P/L of sample banks

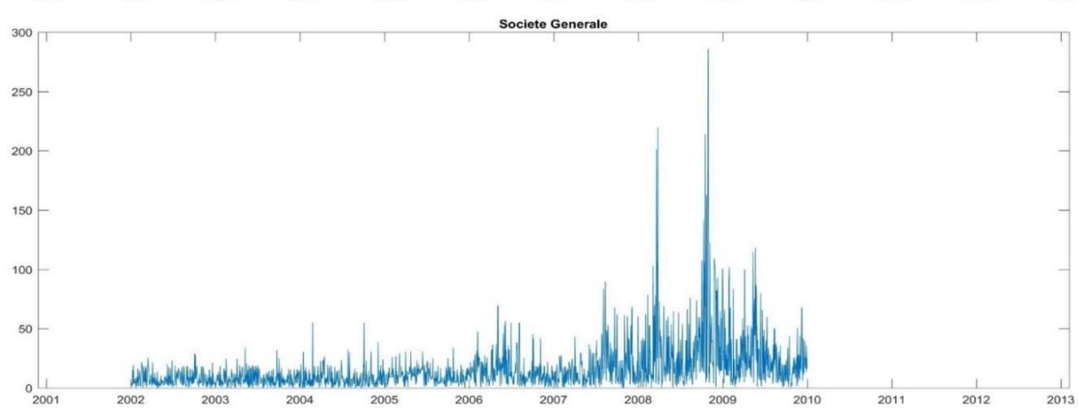
Notes: Figure 2.2 presents the absolute demeaned P/L of seven commercial banks from January 2001 to December 2012, which can be used as proxy of volatility of P/L.




Bank of America
300
250
200
150
100
50
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013


Royal Bank of Canada
800
700
600
500
400
300
200
100
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013


Deutsche Bank
600
500
400
300
200
100
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

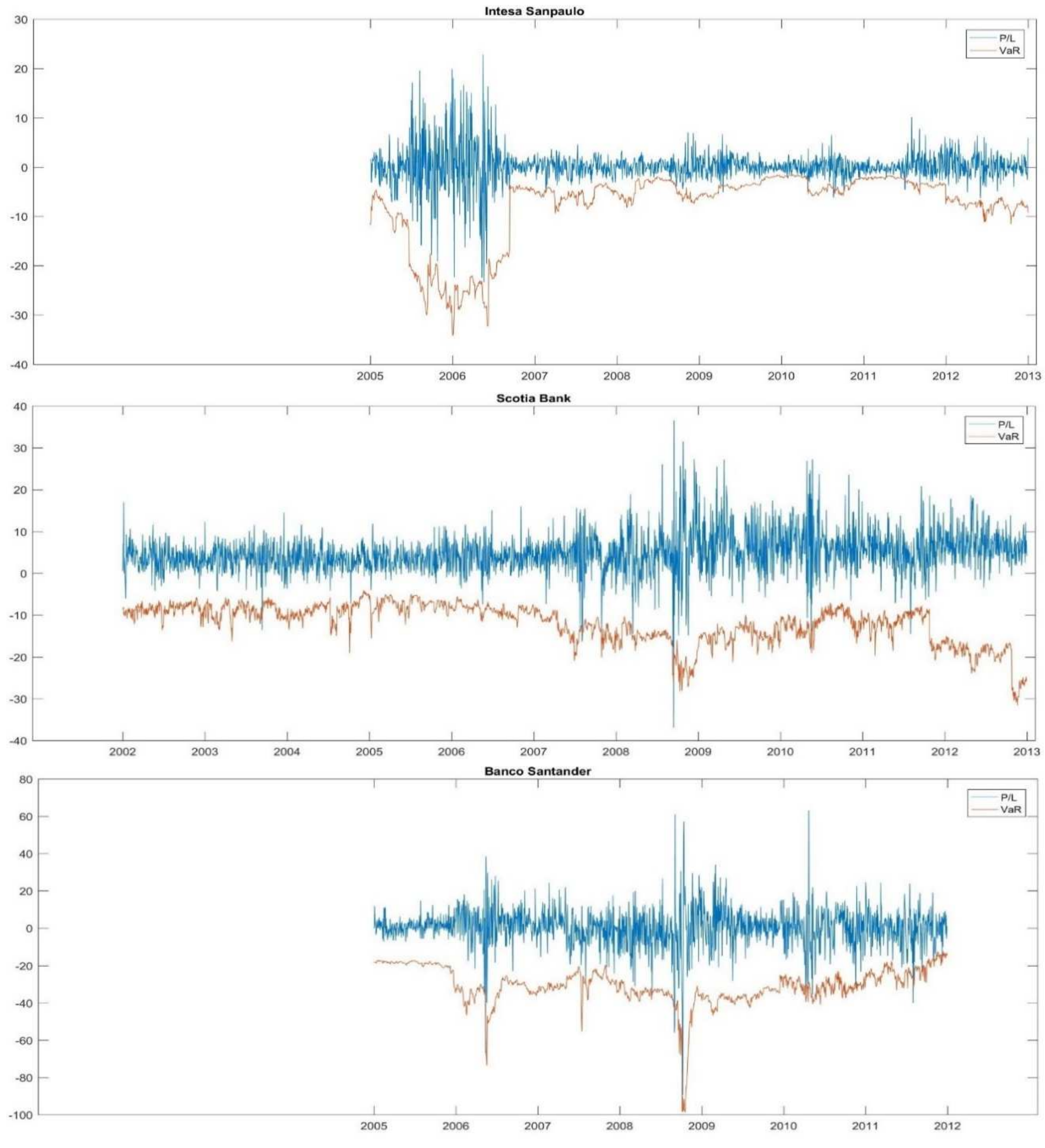

Societe Generale
300
250
200
150
100
50
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

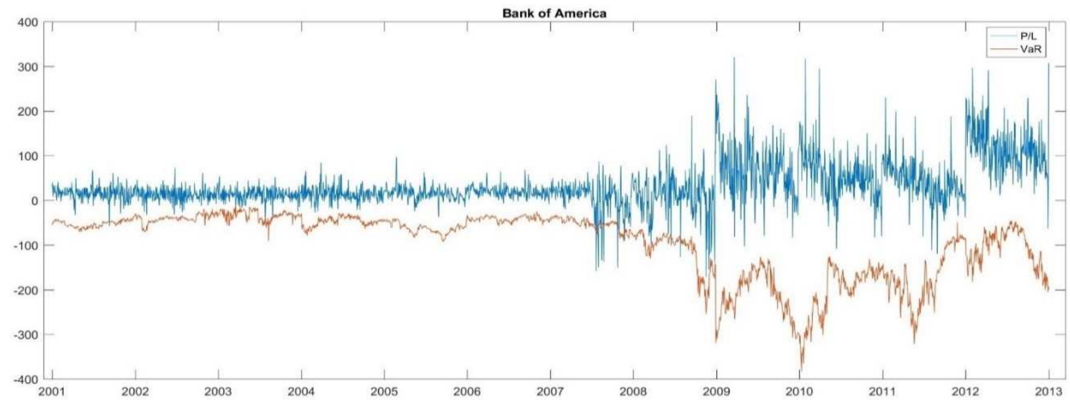
2.5 Evaluation of bank VaRs

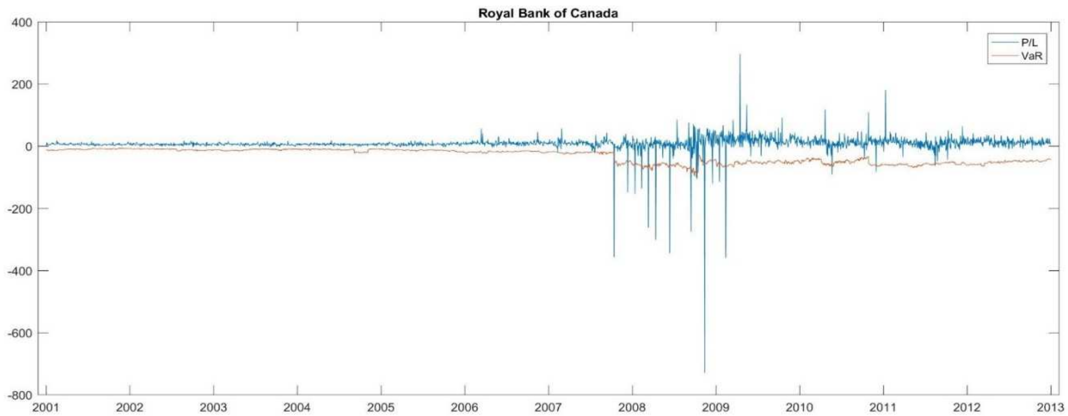
2.5.1 Preliminary analysis of bank VaRs

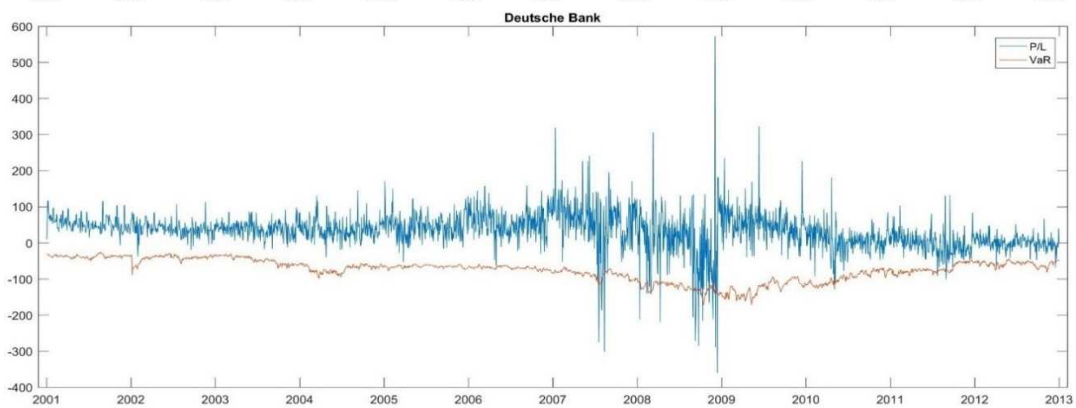
Figure 2.3: Daily trading P/L and VaRs of sample banks

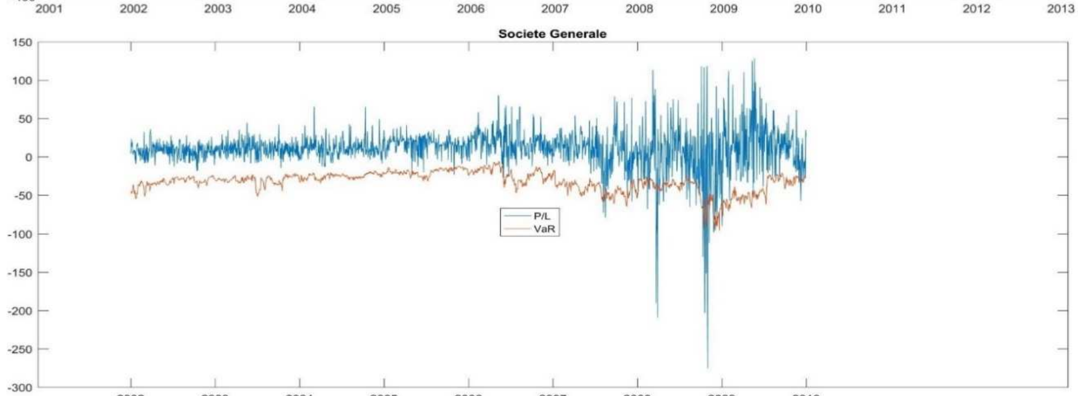
Notes: Figure 2.3 presents the daily trading P/L and VaRs of seven commercial banks from 2001 to 2012. The blue line depicts the P/L and the orange line presents the VaR.




Bank of America
P/L
VaR


Royal Bank of Canada
P/L
VaR


Deutsche Bank
P/L
VaR


Societe Generale
P/L
VaR

In Figure 2.3, we plot the daily one-day-ahead VaRs and daily P/L for seven commercial banks in our sample. In each graph, the blue line presents the daily P/L and the orange line depicts the VaR. We keep the same timeline, starting from 1st January 2001 and ending on 31st December 2012, for all banks for the ease of making comparison. The descriptive statistics of bank VaRs, the number of VaR exceptions and the exception rates of all sample banks are presented in Table 2.3.

Table 2.3: Preliminary analysis of bank VaRs

Notes: Table 2.3 presents the descriptive statistics of seven sample banks in three sub-periods: pre-crisis, crisis and post-crisis periods. Besides, we report the 1%-quantile of the P/L, the number of VaR exceptions and the total observations in the last three columns.

	Min	Max	Mean	1%-quantile	Number of exceptions	Number of observations
Pre-crisis period						
Intesa Sanpaolo	-34.175	-3.5654	-15.4865	-18.3865	0	606
Scotia Bank	-19.101	-4.0573	-8.7388	-3.7391	1	1369
Banco Santander	-73.550	-17.241	-27.702	-19.9544	3	606
Bank of America	-91.996	-11.384	-46.076	-21.4695	4	1614
Royal Bank of Canada	-26.748	-6.5418	-13.328	-1.7968	0	1604
Deutsche Bank	-99.090	-26.176	-57.318	-16.9191	0	1595
Societe Generale	-54.434	-5.4932	-27.091	-12.5724	3	1382
Crisis period						
Intesa Sanpaolo	-8.6191	-1.9340	-4.9081	-3.7468	1	525
Scotia Bank	-28.332	-10.350	-16.001	-14.6100	2	525
Banco Santander	-98.667	-19.767	-36.489	-35.5109	3	525
Bank of America	-319.21	-38.246	-120.318	-141.928	15	525
Royal Bank of Canada	-97.208	-16.241	-51.383	-281.613	17	525
Deutsche Bank	-172.98	-70.988	-118.134	-272.953	31	525
Societe Generale	-95.901	-25.225	-46.659	-160.575	35	525
Post-crisis period						
Intesa Sanpaolo	-11.569	-1.4946	-4.441	-4.5536	10	886
Scotia Bank	-31.616	-7.0771	-14.680	-6.2871	2	891
Banco Santander	-42.623	-12.317	-28.658	-29.1801	4	646
Bank of America	-382.92	-45.937	-168.514	-62.5096	0	888
Royal Bank of Canada	-70.269	-33.372	-53.287	-31.7899	3	891
Deutsche Bank	-136.20	-44.279	-81.463	-79.5686	7	892
Societe Generale	-62.042	-20.547	-30.617	-38.8821	4	128

Figure 2.3 shows that the number of VaR exceptions is extremely low before the financial crisis. With 99% level of confidence and 8,776 daily observations of bank VaRs and P/L, we expect to have 88 VaR exceptions in pre-crisis period. However, the actual number we record in Table 2.3 is only 11 exceptions. There are several banks even experiencing no VaR exceptions during this period, including Intesa Sanpaolo, Royal Bank of Canada and Deutsche Bank, while other banks suffered very few VaR exceptions. We find that at six over seven banks, the absolute value of mean VaR is much higher than the 1%-quantile of the P/L distribution, which indicates the VaR overstatement in the pre-crisis period.

Prior studies and empirical evidences suggest that banks are likely to have high number of VaR exceptions during financial crisis (see Nocera, 2009; O'Brien and Szerszen, 2014). However, Table 2.3 shows that not all banks in our sample underestimate their VaR in crisis period. With 525 trading days, the expected number of VaR exceptions is 5 for every bank. According to the actual number of VaR exceptions, we find that three over seven still overstated their VaRs during financial crisis. These are Intesa Sanpaolo, Scotia Bank and Banco Santander. Besides, there are banks that produced an excessive exception ratio, including Bank of America, Royal Bank of Canada, Deutsche Bank and Societe Generale. Their number of exceptions ranges from 15 (Bank of America) to 35 (Societe Generale). It is also worth to note that at these banks, the absolute value of mean VaR is significantly smaller than the actual 1%-quantile of the P/L distribution.

Banks come back to VaR conservativeness in post-crisis period. Indeed, we record a low exception rate compared to the prior period. With total 5,222 observations, the expected number of VaR exceptions is 52, while the actual number is 30. It is worth to note that these four banks which understated VaR and experience excessive number of VaR exceptions during financial crisis now suffer a very rare (or even zero) VaR exceptions in post-crisis.

Briefly, the preliminary analysis shows that bank VaRs were overstated in pre-crisis and post-crisis periods. During financial crisis, it is evident that not all banks underestimated their VaR. Indeed, we find that even the VaR overstatement still occurs at three over seven banks in crisis period. Thus, we argue that the common knowledge of VaR underestimation during financial crisis might not be universally correct.

2.5.2 The Coverage tests

First, we use the CC test to examine the UC and IND hypothesis of VaR exceptions. Recall that the UC hypothesis states that the observed frequency of VaR exceptions is equal to the expected rate. The null hypothesis of UC property is rejected if the difference between the realized and expected rates of VaR exceptions is statistically significant. Therefore, there are two cases under which the UC hypothesis can be rejected. If the realized frequency of VaR exceptions is smaller than the nominal rate of 1%, and the LR_{UC} statistic is significant. This implies that the VaR amount is overstated. Similarly, if the realized frequency of VaR exceptions is higher than the nominal rate of 1% and the LR_{UC} statistic is significant, the VaR estimate is now underestimated.

The null hypothesis of IND property of VaR exceptions states that VaR exceptions should be independently distributed. Therefore, a significant value of LR_{IND}, implies a rejection of the null hypothesis of IND and indicates the clustering of VaR exceptions. The CC hypothesis, which jointly examines the UC and IND properties, is rejected either when LR_{UC} or LR_{IND} is statistically significant.

In Table 2.4, we present the actual exception rate, the LR_{UC}, LR_{IND}, LR_{CC} test statistics and the summary of the backtesting results. There are some blanks in the test statistics due to the problem of no VaR exception. Table 2.4 shows that Banco Santander is the only bank which successfully passes the UC during the pre-crisis period but fails the IND test. This implies that their number of VaR exceptions is adequate but not independently distributed, or in other words, tends to cluster together. Besides, we notice that the UC hypothesis is statistically rejected at six remaining banks, which also leads to the rejection of the CC hypothesis. The systematic rejection of the coverage tests at these banks is not unexpected, as we have noted a very low number of VaR exceptions during this period. This result is consistent with the findings of prior studies in pre-crisis period, which confirm the VaR overstatement of US banks (Berkowitz and O'Brien, 2002; O'Brien and Szerszen, 2014), Canadian banks (Perignon et al., 2008) or top 50 largest commercial banks globally (Perignon and Smith, 2010).

We document the mixed performance of bank VaRs during financial crisis. It is interesting to note that some banks still overstate their VaR. This is a new finding since banks tended to underestimate their VaR during the financial crisis period. Indeed, our tests reject the UC hypothesis for Intesa Sanpaolo, as they continue to overstate their VaRs. Scotia Bank and Banco Santander are two banks that successfully pass the UC test, although the IND test is rejected in the case of Banco Santander due to clustering of VAR exceptions. Noticeably, there are four banks that also fail the statistical tests, due to the problem of excess exception rates and exception clustering. These include Bank of America, Royal Bank of Canada, Societe Geneale and Deutsche Bank. As seen in Table 2.4, their exceedance rates range from 0.0285 to 0.0667, which are much higher than the expected rate of 0.01. Besides, the LR_{IND} statistics show exceptions due to clustering at three (Bank of America, Deutsche Bank and Societe Generale) out of four banks, thereby overstating the VaR estimates.

Bank VaRs tend to be conservatively biased in post-crisis period. For banks experiencing VaR overstatement in the crisis period, they continue to inflate their VaRs in post-crisis period, from a moderate level (Banco Santander) to an excessively high level (Scotia Bank). While VaR understatement was very popular during financial crisis, it no longer appears in post-crisis period. Indeed, Bank of America and Royal Bank of Canada, which experienced an excessive number of VaR exceptions during financial crisis, surprisingly have very few number of VaR exceptions in post-crisis period. Banco Santander and Deutsche Bank performed reasonably well using our test statistics. The validity of the VaR estimates for Societe Generale and Intesa Sanpaolo is rejected using our CC test due to excessive clustering.

Table 2.4: Backtesting results of bank VaRs

Notes: The table shows the backtesting results of bank VaRs, including the test statistics of the Unconditional Coverage test (LR_{UC}), the Independence test (LR_{IND}) and the Conditional Coverage test (LR_{CC}). The brief evaluation of bank VaRs are presented in the next column, followed by the VaR distortion coefficient ρ. ** denote statistical significance at a 5% level.

	Exception ratio	LR_{UC}	LR_{IND}	LR_{CC}
Pre-crisis period				
Intesa Sanpaolo	0			
Scotia Bank	0.0007	20.246**	0.0015	20.247**
Banco Santander	0.0049	1.9068	7.1839**	9.0908**
Bank of America	0.0025	13.196**	0.0199	13.216**
Royal Bank of Canada	0			
Deutsche Bank	0			
Societe Generale	0.0022	12.544**	0.0131	12.557**
Crisis period				
Intesa Sanpaolo	0.0019	5.2019**	0.0038	5.2057**
Scotia Bank	0.0038	2.6475	0.0153	2.6628
Banco Santander	0.0057	1.1434	6.8983**	8.0417**
Bank of America	0.0285	12.216**	18.674**	30.981**
Royal Bank of Canada	0.0323	16.763**	0.3183	17.081**
Deutsche Bank	0.0590	59.996**	10.426**	70.422**
Societe Generale	0.0667	75.152**	4.8693**	80.022**
Post-crisis period				
Intesa Sanpaolo	0.0112	0.1449	8.4954**	8.6403**
Scotia Bank	0.0022	7.8823**	0.0090	7.8913**
Banco Santander	0.0062	1.0871	0.0499	1.1370
Bank of America	0			
Royal Bank of Canada	0.0033	5.3148**	0.0203	5.3351
Deutsche Bank	0.0078	0.4464	4.1818	4.6282
Societe Generale	0.0312	3.7779	9.5472**	13.325**

2.5.3 Measure of VaR distortions

We follow Perignon et al. (2008) to compute the VaR distortion parameter of banks which failed the UC test in previous section. The idea of the VaR distortion parameter is to quantify the level of VaR overstatement or VaR understatement at banks. Through the VaR distortion parameter, we can simply answer the question:

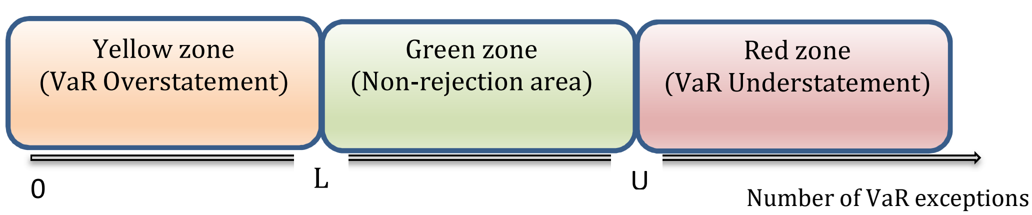
How much bank VaRs were overstated/understated? The following presents the method to estimate the VaR distortion parameter ρ.

The three-zone approach

We develop the three-zone approach to estimate VaR distortion parameter. This approach is inspired by the Basel backtesting framework. Based on the number of VaR exceptions, we classify the results of the UC test into three zones: The Green zone, the Yellow zone and the Red zone. The Green zone is the interval of VaR exceptions that satisfy the UC test. The Yellow zone and Red zone cover the number of VaR exceptions that failed the UC test. Specifically, the Yellow zone is the interval of VaR exceptions corresponding to the VaR overstatement e.g. bank VaRs are rejected due to rare number of VaR exceptions. The Red zone covers the interval of VaR exceptions corresponding with VaR understatement. We graphically present our three-zone approach in Figure 2.4.

Figure 2.4: The three-zone approach

Notes: Figure 2.4 presents the three-zone approach based on the number of VaR exceptions. The Green zone covers the number of exceptions that satisfy the UC test. The Yellow zone is the interval of exceptions corresponding to the VaR overstatement, while the Red zone is for VaR understatement.



Recall that the UC test relies on the frequency of VaR exceptions to validate the performance of a VaR model. In order to satisfy the UC test, the number of VaR exceptions generated by a VaR model must be within the Green interval [L, U]. As shown in Figure 2,4, L is the lower bound of the Green zone, while U is the upper bound. Thus, the Yellow interval corresponding to VaR overstatement is [0, L-1] while the Red zone is [U+1, ∞). For example, consider 2-year bank VaRs including 504 VaR estimates. To be not rejected by the UC test, the number of VaR exceptions must be within the interval [2,10]. The Yellow zone is now [0,1], while the Red zone is [11, ∞).

To estimate L and U, we base on the number of bank VaRs, denoted as N, the coverage rate α, and the simulated number of VaR exceptions, denoted as n. We start with n=0 and perform the UC test. We continue performing UC test with an increment of 1 in the value of n and save the results of the UC test. The select L as the first simulated value of n which satisfies the UC test, while U is the final simulated value of n that is not rejected by the UC test.

Estimation of VaR distortion parameters

Denote $DVaR_t$ as the disclosed bank VaRs with 1- α level of confidence and r_t is the time series of daily trading P/L. To be in the Green zone, $DVaR_t$ must satisfy the UC hypothesis:

$$\Pr\,(r_t < DVaR_t) = \alpha \qquad (2.17)$$

If $DVaR_t$ generates too many VaR exceptions (higher than U), it will be in Red zone and (2.17) becomes:

$$\Pr\,(r_t < DVaR_t) > \alpha \qquad (2.18)$$

The risk distortion parameter ρ in this case is the multiplier that helps $DVaR_t$ satisfy the UC test. In other words, it places $DVaR_t$ in the Green zone:

$$\Pr\,(r_t < DVaR_t \text{ x } (1+\rho)) = \alpha \qquad (2.19)$$

To satisfy (2.19), the number of VaR exceptions generated by $DVaR_t$ x $(1+ \rho)$ must be from L to U. Thus, the VaR distortion parameter $\hat{\rho}$ in this case is the one that makes $DVaR_t$ x $(1+ \hat{\rho})$ produce U number of exceptions. To estimate $\hat{\rho}$, we simulate a number of trials, each with 0.001 increment in the value of ρ and save the one that firstly deliver U number of exceptions.

In case of VaR overstatement, we repeat the procedure above. However, the expected number of exceptions is now L (the lower bound of the Green zone) instead of U. It is important to note that the VaR distortion parameter in case of VaR overstatement receives negative value, while it is positive if DVaRt were understated.

We report the estimation results of VaR distortion parameter in the last column of Table 2.5. Recall that we do not compute VaR distortion parameters for bank VaRs not rejected by the UC test. Table 2.5 shows that before financial crisis, bank VaRs were systematically overstated, from the moderate levels of 0.046 (Intesa Sanpaolo) to extremely high level of 0.791 (Royal Bank of Canada). This result is consistent with the finding of Perignon et al. (2008) showing the significant VaR overstatement of Canadian banks in the pre-crisis period. In the post-crisis period, we document that three over seven banks exhibit significant VaR overstatement. The lowest VaR overstatement coefficient we record in this period is 0.054 for Royal Bank of Canada, while the highest is 0.451 for Bank of America.

Table 2.5: Measurement of VaR distortion coefficients

Notes: Table 2.5 presents the measurement of VaR distortion coefficients of seven sample banks in three sub-periods. For each bank, we report the actual number of exceptions, the green-zone interval and the classification of bank VaRs according to the number of exceptions. If the number of VaR exceptions is below the green zone, bank VaR will be placed in Yellow zone (VaR overstatement). The Red zone is for bank VaRs that have higher number of exception than the Green zone. The estimated VaR distortion coefficients are shown in the last column.

	Number of exceptions	Green zone	Classification	VaR distortion coefficient ρ
Pre-crisis period				
Intesa Sanpaolo	0	[2 11]	Yellow	-0.046
Scotia Bank	1	[8 21]	Yellow	-0.514
Banco Santander	3	[2 11]	Green	-
Bank of America	4	[9 24]	Yellow	-0.335
Royal Bank of Canada	0	[9 24]	Yellow	-0.791
Deutsche Bank	0	[9 24]	Yellow	-0.563
Societe Generale	3	[8 21]	Yellow	-0.345
Crisis period				
Intesa Sanpaolo	1	[2 10]	Yellow	-0.068
Scotia Bank	2	[2 10]	Green	-
Banco Santander	3	[2 10]	Green	-
Bank of America	15	[2 10]	Red	0.283
Royal Bank of Canada	17	[2 10]	Red	1.263
Deutsche Bank	31	[2 10]	Red	0.754
Societe Generale	35	[2 10]	Red	0.937
Post-crisis period				
Intesa Sanpaolo	10	[4 15]	Green	-
Scotia Bank	2	[4 15]	Yellow	-0.377
Banco Santander	4	[3 12]	Green	-
Bank of America	0	[4 15]	Yellow	-0.451
Royal Bank of Canada	3	[4 15]	Yellow	-0.054
Deutsche Bank	7	[4 15]	Green	-
Societe Generale	4	[1 4]	Green	-

During financial crisis, it is evident that Banca Intesa still slightly overstated their VaR. Besides, the sign and the magnitude of the VaR under/overstatement show that four banks significantly underestimated their VaR. The level of VaR understatement ranges from moderate level of 0.283 (Bank of America) to the excessive level of 1.263 (Royal Bank of Canada). It is noticeable that Royal Bank of

Canada has the highest level of VaR understatement in crisis period, while they were the most conservative bank in VaR estimation in pre-crisis period. We find that the level of VaR understatement does not depend on the number of VaR exceptions, but on the magnitude of exceeded losses. Indeed, the magnitude of VaR distortion parameter at Royal Bank of Canada is much higher than Deutsche Bank and Societe Generale, although they experience nearly a half of the number of VaR exceptions comparatively. This serious VaR understatement of Royal Bank of Canada is not unexpected, as we document an abnormally high kurtosis in the descriptive statistics of their daily trading P/L.

2.5.4 The Multivariate Unconditional Coverage test

We now apply the MUC test to all banks during financial crisis. In this sub-section, we aim to identify and comparatively measure the magnitude of the extreme losses and its potential impact on the economic capital of banks.

Recall that the MUC test is designed to jointly examine the likelihood of a loss which exceeds not only regulatory VaR(1%), but also VaR at lower coverage rate VaR(α'), with $\alpha' < 1\%$. The consideration of coverage rate α' plays a crucial role in the evaluation, as it defines the extreme level that a loss can exceed. Our choice of the rare coverage rate α' is inspired by the 2011 McKinsey Market Risk Survey and Benchmarking[15]. Specifically, their report finds that 85% of global banks do not use discrete economic model to estimate their economic capital for market risk[16]. Instead, they base on VaR estimates at very high levels of confidence, typically from

[15] Managing market risk. Today and tomorrow - McKinsey Working Papers on Risk, Number 32.
[16] Banks' economic capital for market risk is designed to capture the potential for extreme events.

99.93% to 99.97%, to set up the level of economic capital. Therefore, we select the rare coverage rates α' at 0.07%, 0.03% and 0.01% to capture the presence of extreme losses during financial crisis. Therefore, if an extreme loss exceeds VaR at these significant levels, this implies that bank might not put enough economic capital to cover this loss.

Table 2.6: The MUC test of bank VaRs during financial crisis

Notes: Table 2.6 presents the number of VaR exceptions and its corresponding LR test statistics (in parenthesis) of seven commercials banks in crisis period. The second column shows the number of exceptions and the LR_{UC} test statistics at VaR(1%). The third, fourth and fifth column present the number of exceptions and the LR_{MUC} test statistics at VaR(0.07%), VaR(0.03%) and VaR(0.01%) respectively. ** denote statistical significance at a 5% level.

	VaR(1%)	VaR(0.07%)	VaR(0.03%)	VaR(0.01%)
Crisis period				
Intesa Sanpaolo	1 (5.201**)	0 (3.1784)	0 (3.1784)	0 (3.1784)
Scotia Bank	2 (2.647)	0 (1.4556)	0 (1.4556)	0 (1.4556)
Banco Santander	3 (1.1434)	0 (0.4808)	0 (0.4808)	0 (0.4808)
Bank of America	15 (12.216**)	5 (22.287**)	1 (14.282**)	0 (13.819**)
Royal Bank of Canada	17 (16.763**)	6 (29.454**)	2 (21.034**)	2 (24.786**)
Deutsche Bank	31 (59.996**)	6 (67.521**)	2 (63.561**)	1 (63.597**)
Societe Generale	39 (75.132**)	8 (86.549**)	5 (85.988**)	3 (85.666**)

Table 2.6 presents the MUC test bank VaRs during crisis period. We find the connection between the distortion parameter ρ and the results of the MUC test. First, we find that banks having negative distortion parameter ρ suffered no VaR super exceptions during financial crisis. These include Banco Santander, Intesa Sanpaolo and Scotia Bank. It can be explained that these banks with negative ρ intentionally overstated their VaR. Thus, it is not likely that they will experience excessive losses exceeding their inflated VaR.

Second, banks having positive distortion coefficient experienced VaR super exceptions and therefore are rejected by the LR_{MUC} test statistics. These banks are Bank of America, Royal Bank of Canada, Deutsche Bank and Societe Generale. Table 2.6 shows that at VaR(0.07%), the number of exceptions ranges from 5 for Bank of America to 8 for Societe Generale. At VaR(0.03%), Bank of America suffers one exception, while the corresponding numbers at Royal Bank of Canada, Deutsche Bank and Societe Generale are 2, 2 and 5 respectively. It is interesting that even at VaR(0.01%), three over four banks still experience VaR exceptions. These are Royal Bank of Canada, Deutsche Bank and Societe Generale which suffer two, one and three exceptions at VaR(0.01%) respectively. Recall that among four banks understating VaR in crisis period, Royal Bank of Canada and Societe Generale have the most serious VaR distortion parameters, followed by Deutsche Bank and Bank of America. This evidence shows a relation between VaR distortion parameter and the number of VaR super exceptions. We find that the higher the magnitude of VaR distortion parameter, the more number of VaR super exceptions at extreme coverage rates.

The MUC test implies the adequacy of the economic capital for market risk of banks. The economic capital models for market risk at banks are designed to deal with potential for extreme events. It is typical that the economic capital models aim to identify the maximum potential loss within a one-year horizon and level of confidence lies between 99.95% to 99.99% (Methta et al., 2012). In their research on the McKinsey working papers on risk, Methta et al. (2012) show that 85% of banks in their survey do not use a discrete economic capital model. Instead, banks use their VaR to compute economic capital for market risk, with the confidence interval ranges from 99.91% to 99.97% for global banks. Recall that we still find extreme losses in crisis period that exceed VaR(0.07%), VaR(0.03%) and even VaR(0.01%). Therefore, we argue that during financial crisis, these banks might not put enough economic capital for market risk.

2.6 Discussions

The purpose of this section is to discuss about the systemic VaR overstatement of commercial banks in normal periods and the poor performance of bank VaRs in financial crisis. The systemic VaR overstatement can be attributed to: the use of contaminated data, the choice of VaR model and the benefit of VaR overstatement.

The use of contaminated data

Recall that the statistical tests were designed to help risk managers and banking regulators to evaluated bank VaRs and penalize banks with poorly performing VaR models. However, Fresard et al. (2013) show that the use of contaminated data can distort the number of VaR exceptions and thus the backtesting result of bank VaR. The contaminated terms, such as fees and commissions, are mostly risk-free and do not reflect the exposure to market risk of trading portfolio. We find that among seven sample banks, five banks are considered using contaminated data in their backtesting (see Panel C of Table 2.1). These banks are Scotia Bank, Bank of America, Royal Bank of Canada, Deutsche Bank and Societe Generale. The inclusion of the positive, risk-free incomes may inflate the trading P/L and as a result, lead to the problem of having low number of VaR exceptions.

The choice of internal VaR model

The choice of VaR internal model might be the cause of the poor performance of bank VaRs in both normal and crisis period. Table 2.1 shows that among seven sample banks, six banks use Historical simulation (HS) as their internal VaR model, except Deutsche Bank uses Monte Carlo simulation. To examine whether the choice of VaR model leads to the poor performance of bank VaRs, we use the simplest Naïve HS as the benchmark model to compare the performance of bank VaRs.

In the Naïve HS model, the VaR(1%) is simply the 0.01-quantile of the empirical return distribution of an portfolio. Denote r_t as the P/L on trading portfolio on day t. For time horizon T, we have the sequence of daily P/L $\{r_t\}_{i=1}^{T}$. The 1-day VaR(1%) is defined as the α-quantile of the sequence of the historical returns:

$$\mathrm{VaR_t(1\%) = Quantile\ \{\{r_{t-i}\}_{i=1}^{T}, \alpha\}} \tag{2.20}$$

We use rolling window technique to estimate one-day-ahead VaR. For the size of rolling window, we select T = 504 days (or two years of historical data, equivalently). The performance of bank VaRs and our benchmark model will be evaluated based on the out-of-sample forecasts. As we use the first 504 observations to start our forecast, hence the length of the forecast horizon for pre-crisis period will be shorten by 2 years, while the other sub-periods keep remained. We continue using CC test to evaluate the performance of bank VaRs and the benchmark model. The backtesting results are reported in Table 2.7.

From the test results, we find two important points. The first is the failure of bank VaRs and the benchmark model in crisis period. Indeed, the Naïve HS cannot capture the change in market volatility during this period and hence generating too many exceptions. As a result, its validity is rejected by the test statistics in six over seven banks. The second point is the superiority of our benchmark model to bank VaRs in providing accurate risk figures in normal periods. While bank VaRs tend to be overstated and thus produce very rare (even zero) number of VaR exceptions, the benchmark model delivers appropriate number of VaR exceptions according to the coverage tests. Although both using HS, the difference in the performance of banks' internal model and our benchmark model can be explained that bank HS updates the historical window in daily frequency with current position exposures, while in case of Naïve HS, only the newest and oldest observations are updated.

This result shows evidence that even the simplest VaR model can outperform bank VaRs in providing good VaR estimates. Besides, it is the fact that the performance of bank VaRs does not improve over time. Therefore, we argue that banks are adopting their inferior model to overstate their VaRs. The intentional VaR overstatement can bring banks some economic merits, which will be discussed in the next section.

Table 2.7: Performance evaluation of bank VaRs and the benchmark model

Notes: Table 2.7 presents the number of VaR exceptions and the LR test statistics of bank VaRs and the benchmark model in three sub-periods. The benchmark model is the Naïve Historical Simulation, in which VaR is estimated using moving window of two years of historical data.

	Total observations	Bank VaRs				Benchmark model			
		No of exceptions	LR_{UC}	LR_{IND}	LR_{CC}	No of exceptions	LR_{UC}	LR_{IND}	LR_{CC}
Pre-crisis period									
Intesa Sanpaolo	606	0	-	-	-	0	-	-	-
Scotia Bank	1369	0	-	-	-	8	0.0490	0.1490	0.1980
Banco Santander	606	0	-	-	-	0	-	-	-
Bank of America	1614	3	13.196**	0.0199	13.216**	6	2.8320	0.0650	2.8970
Royal Bank of Canada	1604	0	-	-	-	18	3.7970	1.0890	4.8770
Deutsche Bank	1595	0	-	-	-	12	0.1080	0.2670	0.3750
Societe Generale	1382	3	12.544**	0.0131	12.557**	11	0.5300	0.2790	0.8090
Crisis period									
Intesa Sanpaolo	525	1	5.2019**	0.0038	5.2057**	6	0.1060	3.8040	3.9100
Scotia Bank	525	2	2.6475	0.0153	2.6628	11	4.858**	5.719**	10.578**
Banco Santander	525	3	1.1434	6.8983**	8.0417**	14	10.145**	20.292**	30.437**
Bank of America	525	15	12.216**	18.674**	30.981**	17	16.763**	21.896**	38.659**
Royal Bank of Canada	525	17	16.763**	0.3183	17.081**	21	27.268**	3.7980	31.067**
Deutsche Bank	525	31	59.996**	10.426**	70.422**	18	19.222**	14.600**	33.822**
Societe Generale	525	35	75.152**	4.8693**	80.022**	25	39.370**	16.327**	55.697**
Post-crisis period									
Intesa Sanpaolo	886	10	0.1449	8.4954	8.6403	9	0.0020	9.4070	9.4100
Scotia Bank	891	2	7.8823**	0.0090	7.8913**	3	5.314**	0.0203	5.335**
Banco Santander	646	4	1.0871	0.0499	1.1370	3	2.3250	0.0280	2.3530
Bank of America	888	0	0.0000	0.0000	0.0000	5	2.0240	0.0560	2.0810
Royal Bank of Canada	891	3	5.3148**	0.0203	5.3351	5	2.0510	0.0560	2.1070
Deutsche Bank	892	7	0.4464	4.1818	4.6282	3	5.328**	0.0200	5.3480**
Societe Generale	128	4	3.7779	9.5472**	13.325**	0	-	-	-

**: rejected at 95% confidence level

The benefits of VaR manipulation

Banks may have benefits from both VaR overstatement and VaR understatement. Over-reporting VaR helps bank to have rare VaR exceptions, which looks attractive to investors and banking regulators. Besides, the Basel backtesting framework does not penalize bank that overstate their VaR, therefore it may signal that VaR overstatement is preferable to VaR understatement. However, there is a cost of overstating VaR. Under Basel framework[17], banks are subjected to market risk capital requirements, which directly depends on the magnitude of its VaRs. Specifically, the capital requirements are computed by taking the higher value between previous day VaR and the average VaR of the preceding 60 business days, multiplied by the scaling factor of 3. Therefore, the more VaR overstatement, the higher the market risk charges to the bank.

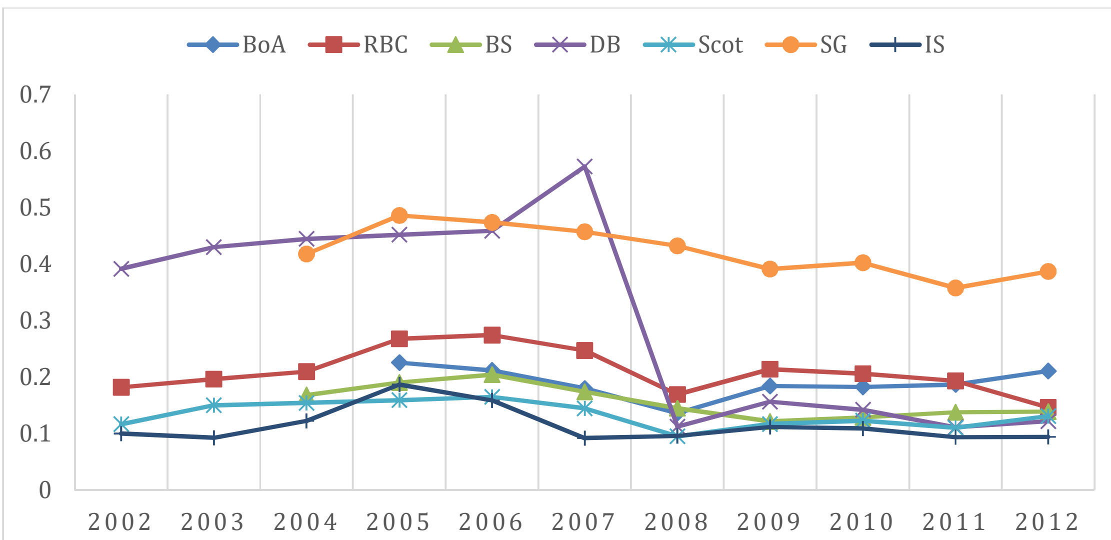
Understating VaR may cost banks more. According to the Basel backtesting framework, if a bank has an excessive number of VaR exceptions, their internal VaR model will be investigated to improve future backtesting outcomes. There will also be a capital fine for bank that does not meet the desired number of VaR exceptions, which is the multiplier adding to the scaling factor. More important, bank that suffers a number of VaR exceptions might put their reputation at risk. Having too many VaR exceptions may be considered as the signal of poor risk management. As a result, it is likely that the market will penalize bank which is unable to accurately manage its exposure to market risk. Besides, we find that the trading activities do

[17] Basel Committee on Banking Supervision, 1996b.

not account much in banking business. Indeed, Figure 2.5 shows that on average, most banks keep their trading portfolio account for less than 20% of their total assets. With the small contribution of trading activities, banks have more incentive to overstate their VaRs to avoid the reputational risk in the trade off with higher capital requirement associated with over-reporting VaRs. The manipulation in VaR backtesting results makes VaR itself less informative as a useful risk management tool.

Figure 2.5: Contribution of trading portfolio to banks' total assets

Notes: Figure 2.5 presents the ratio of the value of bank's trading portfolio over their total assets from 2002 to 2012 of seven commercial banks. There are some missing points in the graph due to lack of data.



2.7 Concluding remarks

Chapter 2 contributes to literature on the performance of bank VaRs before, during and after global financial crisis. Using non-anonymous daily data of seven commercial banks from early 2001 to 2012, we examine the accuracy of bank VaRs associating with their daily trading P/L. Our backtesting result shows evidence of systemic VaR overstatement in normal periods. During financial crisis, while some banks continued to overstate their VaR, the others significantly underestimated their risk. The VaR exceptions at these banks were excessively high and tend to cluster together. We also find evidence of extreme losses during financial crisis, which were likely to exceed VaR at very low coverage rates. We attribute the poor performance of bank VaRs to several factors: (i) the use of contaminated data, as we find that banks do not exclude fees, commissions and intraday revenues from their trading P/L, (ii) the choice of VaR model, as the simplest Naïve HS can easily outperform internal VaR model at banks and (iii) the incentives of VaR manipulation, as banks tend to overstate their VaR to avoid reputational risk.

This chapter finds that the systemic VaR overstatement at banks is intentional. We find evidence that banks continue using the inferior models to inflate their VaR estimates to get the economic merit of VaR overstatement. The distortions of bank VaRs, which are popular across banks, make VaR a poor risk management tool. Therefore, this chapter suggests several recommendations to banks, financial regulators and investors. For commercial banks, we suggest that banks' risk managers should stop manipulating their VaR estimates, as these are the informative risk figures being used by both financial regulators and public investors. For investors, we recommend that the they should not rely on the disclosed VaR and backtesting results to make investment decision, as these figures are distorted and may result in misleading financial decisions. This chapter also has recommendation for financial regulators. To deal with the problem of VaR overstatement, we suggest that the financial authorities should treat VaR overstatement equal to VaR understatement. Therefore, bank having rare or no VaR exceptions should be penalized the same as having excessive VaR exceptions.

Chapter 3:

The predictive power of VaR models at commercial banks

3.1 Introduction

In the financial industry, VaR has been widely used and regarded as a benchmark for measuring market risk (Jorion, 2006). VaR measures the maximum potential loss on a given period of time, commonly one-day or ten-day ahead, at a certain level of confidence (e.g. 95% or 99% typically). While the concept of VaR is attractive and simple, there is no unique methods that VaR computation adopts. Indeed, there are a number of alternative approaches that VaR can be implemented. VaR estimates are also sensitive to the choice of the risk model and the nature of the pre-sample data (Boucher et al., 2014). For that reason, evaluating the predictive power of VaR models has been increasingly of interest.

The choice of the VaR model is important in estimating VaR. A good model should capture the true data generating process and therefore provide good VaR estimates. According to Cairns (2000), model risk is also an issue since the uncertainty in estimating risk arising from the choice of an inappropriate model (i.e. incorrect assumptions about the distributional form of the statistical model) and parameter uncertainty (i.e. estimation error in the parameters of the chosen model). This double uncertainty causes plausible risk forecasts (Beder, 1995) and more generally, the inability to forecast risk with acceptable accuracy. These issues should be considered in future work.

The literature on the performance of VaR models has traditionally devoted on the use of market data e.g. stock indices, simulated portfolios, commodity prices to estimate VaR. However, it is the fact that commercial banks, the primary users of VaR measurement and the main objects of regulatory framework of market risk management, do not use market data to estimate their VaRs. Indeed, they use their own P/L data on trading portfolios to obtain VaR estimates[18]. To the best of our knowledge, there has been very little empirical research on the performance of VaR models with the use of bank data. Using daily data of six US banks from January 1998 through March 2000, Berkowitz and O'Brien (2002) show that the complicated structure models cannot outperform the ARMA-GARCH model with normal distribution in forecasting portfolio VaR. Besides, they find that the simple GARCH produces less conservative VaR estimates than banks' internal model. O'Brien and Szerszen (2014) extend the dataset with five US banks that covers the pre-crisis and crisis period. They compare the performance of bank VaRs to the Naïve and Filtered HS, the EVT approach and the GARCH models with alternative distributional assumptions and confirm the superiority of GARCH models.

It is important to note that the literature on the performance of VaR models using bank data is limited to these studies above. This is because of the data availability, as most researchers cannot get access to P/L and VaR data of commercial banks. As a result, to date there has been no study on bank VaRs that

[18] The information is available on bank's annual reports under market risk management section. For example, Scotia bank disclose their information under the section "Risk measurement summary" on page 63 of their 2012 annual report.

covers the post-crisis period. Thus, the literature demands a comprehensive study on the predictive power of VaR models with the use of bank data that covers the pre-crisis, financial crisis and post-crisis period. This is the gap in the literature we want to fill.

This chapter investigates the performance of various VaR models using the non-anonymous daily P/L and VaR data of seven commercial banks. The sample period starts from January 2001 and ends in December 2012, which is divided into three subperiods: pre-crisis, crisis and post-crisis. Compared to prior studies, this research uses wider and longer dataset. The rich dataset not only increases the power of the statistical test, but also gives us a more comprehensive evaluation on the performance of alternative VaR models in both normal and crisis periods. This chapter employs various VaR models, including the Naïve and Filtered HS, the Riskmetrics, the GARCH(1,1) and GJR-GARCH(1,1) models with alternative distributional assumptions and the Extreme Value Theory (EVT) approach. To compare model performance, we develop a two-stage backtesting framework. The first stage investigates the absolute performance of VaR estimates using the coverage tests. The second stage quantifies the comparative performance of VaR models using the magnitude loss function (LF).

Our empirical analysis shows two main points. First, we find that the alternative VaR models can easily outperform banks' internal model in both normal and crisis periods. Second, we acknowledge the superiority of the GARCH-type models in forecasting bank VaRs. Specifically, the unconditional models with Gaussian distribution outperform other models in normal periods, while incorporating Student t is by far the best in estimating VaR during financial crisis.

While the HS models perform inconsistently, none of the banks' internal model accurately capture the bank risk. The EVT approach, which was shown to be superior in VaR estimation with market data, performs very poorly with bank data. Thus, we argue that good bank VaRs can be obtained using simple and accessible models rather than other sophisticated models.

The outline of the chapter is as follows: Section 3.2 briefly introduces various VaR models used in this chapter. Section 3.3 gives overview of our two-stage backtesting procedure, followed by the empirical analysis in the Section 3.4. Finally, section 3.5 summarizes the findings of this chapter.

3.2 Models to estimate VaR

Chapter 3 employs three approaches with six VaR models to forecast VaR: The Naïve and Filtered HS models, the Riskmetrics model, the GARCH(1,1) specifications with normal distribution (GARCHn) and with Student t distribution (GARCHt), the GJR-GARCH(1,1) specifications with normal distribution assumption (GJR-GARCHn) and Student t innovation (GJR-GARCHt) and the EVT approach. The CAViaR model of Engle and Manganelli (2004), although having been widely used in recent studies, will not be used in this chapter due to its incompatibility when working with the trading P/L data of banks.

3.2.1 The Historical simulation models

The HS approach is one of the simplest method to estimate VaR. It is also the most popular VaR approach used by commercial banks, as 73% of commercial banks that disclose using this method to estimate their VaRs (Perignon and Smith, 2010b). Within this HS approach, the Naïve HS and the Filtered HS are the most widely-used models to compute VaR. These model specifications are presented in the following.

The Naïve historical simulation model

The Naïve HS was initially introduced by Boudoukh et al. (1998) and Barone-Adesi et al. (1998, 1999). In value terms, a 1-day historical VaR at coverage rate α is defined as the α-quantile of the empirical 1-day P&L distribution. Specifically, suppose we have the historical data from day 1 to day T, and denote r_t the return of portfolio on day t, then we get a series of returns $\{r_t\}_{t=1}^{T}$. The 1-day VaR at coverage rate α is computed as the α-quantile of the sequence of the past portfolio returns:

$$VaR_t = \text{Quantile}\,\{\{r_{t-i}\}_{i=1}^{T}, \alpha\} \qquad (3.1)$$

One of the main attractions of Naïve HS is that it does not require any parametric assumptions about the distribution of the risk factor. In historical simulation approach, the dependencies and dynamic evolutions of the risk factors are directly inferred from historical data. More specific, this method implicitly assumes that future outcomes have been mirrored in the past, and that the historically simulated returns distribution is identical to the empirical distribution over the forecast horizon. As it does not depend on parametric assumptions about the behaviour of market variables, historical VaR models can accommodate fat tails, skewness and any other non-normal features that cannot be properly fitted with parametric approaches. Thus, this model is ideal to assess the risk of complicated path-dependent financial products but still keep the dynamic behaviour or risk factors in realistic manner (Alexander, 2009). The popularity of historical VaR also comes from the ease-of-use, intuition and simplicity of the model itself. Historical simulation approach is fairly easy to implement and can accommodate any type of portfolio position, in which historical data is readily accessible either in public or internal sources.

The main drawback of the Naïve HS is that its results are significantly dependent on the input data set. It arises from the assumption of historical model that future outcomes will behave like the past. As a result, historical scenarios might not properly reflect the next cycle of the market if there is a significant change in market conditions. The second concern is data limitations, as we demand a considerable amount of historical observations to get risk forecasts at acceptable precision, particularly when applying this model to estimate VaR with high level of confidence (e.g. 99% or higher). On the other hand, there are also problems with long data set, as the longer the historical data, the more the problems with aged data. The long data set might include past events, which are unlikely to occur again, but can distort the risk estimates. Besides, innovations in current market observations are likely to be diluted by older events, making the forecasting results less responsive to current market conditions.

The Filtered Historical Simulation

The Filtered HS model was proposed by Hull and White (1998) and Barone-Adesi et al. (1999) to remedy some of the shortcomings of the Naïve HS model. Indeed, it is the combination of the Naïve HS and the GARCH specifications. This semi-parametric approach holds the advantage of the Naïve HS method which does not make any distribution assumption of asset returns, while its variances are conditionally obtained via volatility models.

In this approach, firstly we fit the normal GARCH process to the return series. The estimated model will be used to infer conditional variance $\hat{\sigma}^2{}_t$ for each day in the sample. The second step is to obtain standardized return $\hat{\varphi}_t$ by dividing the historical returns r_t by the inferred conditional volatility $\hat{\sigma}_t$. The standardized returns $\{\hat{\varphi}_t\}$ should be independent and identically distributed, thus they are suitable to be historically simulated. The next step is to use bootstrap technique for the dataset of standardized returns to get a large number of simulated standardized returns $\{\widehat{\Psi}_t\}$. Each of simulated standardized returns is then multiplied by the today's forecast of tomorrow's volatility $\hat{\sigma}_{t+1|t}$ obtained by previous GARCH model to get a large sample of simulated returns. Finally, a one-day-ahead VaR with α significant level will be estimated as a α^{th} quantile of simulated return series:

$$VaR_t = \text{Quantile}\,\{\{\Psi_{t-i}\}_{i=1}^{T}, \alpha\} \tag{3.2}$$

3.2.2 The conditional volatility models

We present the data generating process of the VaR estimates using conditional volatility models:

$$r_t = \mu + \varepsilon_t \tag{3.3}$$

$$\varepsilon_t = z_t \sigma_t, \;\; z_t \sim D(0,1) \tag{3.4}$$

$$VaR_t = \hat{\mu}_t + \phi^{-1}_{(\alpha)} * \hat{\sigma}_t \tag{3.5}$$

where r_t is the time series of daily P/L with constant mean μ, and $\phi^{-1}_{(\alpha)}$ is the α-quantile of the assumed distribution function. Equation (3.5) shows that there are two essential components to obtain VaR: (i) the assumption of probability distribution and (ii) the estimated parameters of the conditional mean and conditional volatility. The distributional assumption is used not to fit the whole sample to, but to adjust the specific quantile for VaR[19], following the equation (3.5). In order to obtain VaR, estimating conditional mean μ_t and conditional volatility σ_t is essential. There are several models that capture the dynamic of the conditional volatility, in which we employ the Exponential Weighted Moving Average (EWMA), the GARCH and GJR-GARCH models.

The Riskmetrics model

The Riskmetrics model was firstly proposed by JP Morgan in 1989, then became publicly accessible in 1992. It is one of the simplest ways to estimate VaR. In the Riskmetrics, the dynamic of σ_t follows the EWMA:

$$\sigma^2{}_{t} = \lambda\sigma^2{}_{t-1} + (1-\lambda)\ \varepsilon^2{}_{t-1} \tag{3.6}$$

Regarding to the choice of decay factor λ, we select the value of 0.94 as recommended by Riskmetrics. The conditional volatility σ_t is now obtained as following:

$$\sigma^2{}_{t} = 0.94\sigma^2{}_{t-1} + 0.06\ \varepsilon^2{}_{t-1} \tag{3.7}$$

The GARCH(1,1) models

The second model to conditionally estimate σ_t is the GARCH(p,q) model, which is designed to capture the change in market volatility including its clustering characteristics of financial time series (Engle, 1982; Bollerslev, 1986). In GARCH(p,q) model, the conditional variance $\hat{\sigma}^2{}_t$ is obtained as:

$$\sigma^2{}_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon^2{}_{t-1} + \sum_{j=1}^{q} \beta_i \sigma^2{}_{t-1} \tag{3.8}$$

We select the basic GARCH(1,1) model to estimate VaR, as Handen and Lunde (2005) note, there is no evidence that GARCH(1,1) is outperformed by other more

[19] This is different to the approach used in Chapter 4, in which we assume the distributional assumption and adjust the whole sample to obtain the quantile forecasts.

sophisticated models in forecasting volatility of financial time series. The conditional volatility σ_t is now obtained as following:

$$\sigma^2{}_t = \omega + \alpha\varepsilon^2{}_{t-1} + \beta\sigma^2{}_{t-1} \tag{3.9}$$

To complete GARCH specification, the first choice of distributional assumption is the Gaussian distribution, as studies have shown that the GARCHn is superior to banks' internal model in providing accurate risk figures (Berkowitz and O'Brien, 2002; Perignon et al., 2008). We also investigate the forecasting power of GARCH(1,1) model with Student t innovation due to its ability to cope with fat-tailed distribution. Besides, studies have shown that the GARCHt performs well in VaR estimates for both equity markets and currency returns (Angelidis et al., 2004; So and Yu, 2006).

The GJR-GARCH(1,1) models

The GJR-GARCH model was proposed by Glosten, Jagannathan and Runkle (1993) to take into account the leverage effect, in which past negative returns have more impact on today's volatility than past positive returns. In the GJR-GARCH(1,1) specifications, the conditional volatility σ_t is now obtained as:

$$\sigma^2{}_t = \omega + \alpha\varepsilon^2{}_{t-1} + \beta\sigma^2{}_{t-1} + \xi I[\,\varepsilon_{t-1} < 0]\,\varepsilon^2{}_{t-1} \tag{3.10}$$

Similar to the GARCH(1,1) model, we incorporate the GJR-GARCH(1,1) specification with two distributional assumptions to estimate the conditional volatility σ_t, including the Gaussian distribution (GJR-GARCHn) and the Student t distribution (GJR-GARCHt).

3.2.3 The Extreme Value Theory approach

Due to their ease of use and intuitive appeal, the conditional volatility models have been popularly used to compute VaR. However, the recent financial crises witnessed the failure of these conventional VaR models in capturing the extreme losses, especially during the 1997 Asian financial crisis and the 2007-2009 global crisis. This demands for an alternative approach that directly investigates the tail behaviour of the return distribution rather than the whole distribution. It leads to the popular application of EVT to estimate VaR in recent studies on risk modelling.

EVT is a type of statistics which focuses on the tail behaviour of the asset returns and is designed to cope with events that are extreme in nature. In general, there are two approaches to identify extremes in return series. The first approach, which is called Block Maxima method, divides dataset into blocks and only takes into account the maxima of each block. The second approach, which is called Peak Over Threshold (POT), focuses only on the extreme observations that exceeds a given threshold. Between two approaches, the POT method uses data more efficiently and therefore has been widely used in the computation of extreme risk measures (Gilli and Kellezi, 2006). For that reason, we use POT method to estimate bank VaRs.

The main idea of POT method is to extract extremes from a time series data, X_t, t=1, 2…n with a distribution function $F(x) = Pr\{X_t \leq x\}$ by taking all exceedances over a given threshold u. An exceedance occurs when $X_t > u$ for any t in t=1, 2…n.

In POT method, the critical step in estimating parameters of Generalized Pareto Distribution (GPD) is to determine threshold value u. Theories suggest that u should be high enough in order to make the conditional excess distribution function converges to GPD (Pickands, 1975; Balkema and de Haan, 1974). However, the high threshold u will leave less observations for the estimation of the tail distribution function's parameters. To date, there has been no automatic algorithm for the selection of the right threshold u, although this issue was discussed by Danielsson and de Vries (1997) and Dupuis (1998), Danielsson et al. (2001). Following the ideas of McNeil and Frey (2000) and Fernandez (2003), we select the threshold at the 10%- quantile of return distribution (or 10 percent of observations in the left tail) as the choice of u. We also test some different thresholds around the vicinity of 10% and found out that the estimation results did not change noticeably.

Given the 10%-quantile threshold, the next step is to fit the exceedances above this threshold into GPD. In order to estimate the parameters of GPD, we employ the Maximum Likelihood Estimation (MLE).

For a sample $X = \{x_1, x_2 \ldots x_n\}$, the log-likelihood function $L(\xi,\sigma|x)$ for the GPD is the logarithm of the joint density of the n observations:

$$L(\xi,\sigma|x) = = \begin{cases} -n\log\sigma - \left(\frac{1}{\xi} + 1\right)\sum_{i=1}^{n} \log\left(1 + \frac{\xi}{\sigma} x_i\right) & \text{if } \xi \neq 0 \\ -n\log\sigma - \frac{1}{\sigma}\sum_{i=1}^{n} x_i & \text{if } \xi = 0 \end{cases} \tag{3.11}$$

We obtain the values of $\hat{\xi}$ and $\hat{\sigma}$ which maximize the log-likelihood function for the sample of exceedances over a given threshold u. Given the estimated coefficients, VaR using EVT approach is now computed as:

$$VaR_{EVT,t}(\alpha) = u + \frac{\hat{\xi}}{\hat{\sigma}}\left[\left(\frac{n}{n_u}\alpha\right)^{-\hat{\xi}} - 1\right] \tag{3.12}$$

Since being applied to financial risk forecasting, EVT has gained a popular use. Indeed, there are number of studies which confirm the superiority of EVT approach in modelling extreme risk. With the dataset of emerging markets covering both normal and crisis periods, Gencay and Selcuk (2004) find out that EVT approach outperforms a variety VaR models in estimating VaR at high quantile. Chan and Gray (2006) report the success of EVT-based model for the purpose of forecasting tail quantiles and estimating VaR for electricity spot prices. Using reality check to compare predictive power of VaR models, Bao et al. (2006) find evidence that EVT models do better than alternative VaR models in crisis period. The superiority of EVT-based models to conventional VaR models is consistently reported by Bekiros and Georgoutsos (2005); Aloui et al. (2011), Schaumburg (2012) and Adrian and Shin (2014).

3.3 Methods of evaluating VaR estimates

In literature, there are two main approaches to examine the performance of a VaR model. The first is to conduct statistical hypothesis testing in order to find whether a VaR model outperforms others in fitting with theoretical statistical properties. The most popular tests of this approach are the UC, the IND and the CC test[20]. The second approach is to perform loss function (LF), which reflects the loss from a failure of VaR model in correspondent with their realized loss (Lopez, 1998,1999; Sarma et al., 2003). As each type of test captures one type of potential misspecification of the VaR models, hence we design a two-stage evaluation to comprehensively examine the adequacy of VaR models when fitting with banks' data. The first stage evaluates the absolute performance of VaR models by using the statistical tests. The second stage, which is designed for VaR estimates that pass the first stage backtesting, aims to assess the comparative performance of VaR models.

[20] See section 2.4 in chapter 2.

3.3.1 The evaluation of absolute performance

In the first stage of the backtesting procedure, we aim to evaluate the absolute performance of VaR models by examining the behaviour of VaR exceptions. In line with the first study, we continue using the coverage tests to examine the UC, IND and CC hypothesis of the hit sequence of a VaR estimates.

Recall that the UC test examines the coverage of VaR exceptions based on the actual number of VaR exceptions. The null hypothesis of UC test states that the observed frequency of VaR exceptions is consistent with the expected coverage rate.[21] The CC test extends the UC test by jointly examining the IND and UC properties of VaR exceptions. The CC hypothesis holds when both UC and IND hypotheses are simultaneously satisfied. Specifically, a VaR model is not rejected by the CC test if its VaR exceptions are independently distributed and the actual frequency of VaR exceptions is not significantly different to the expected coverage rate.[22]

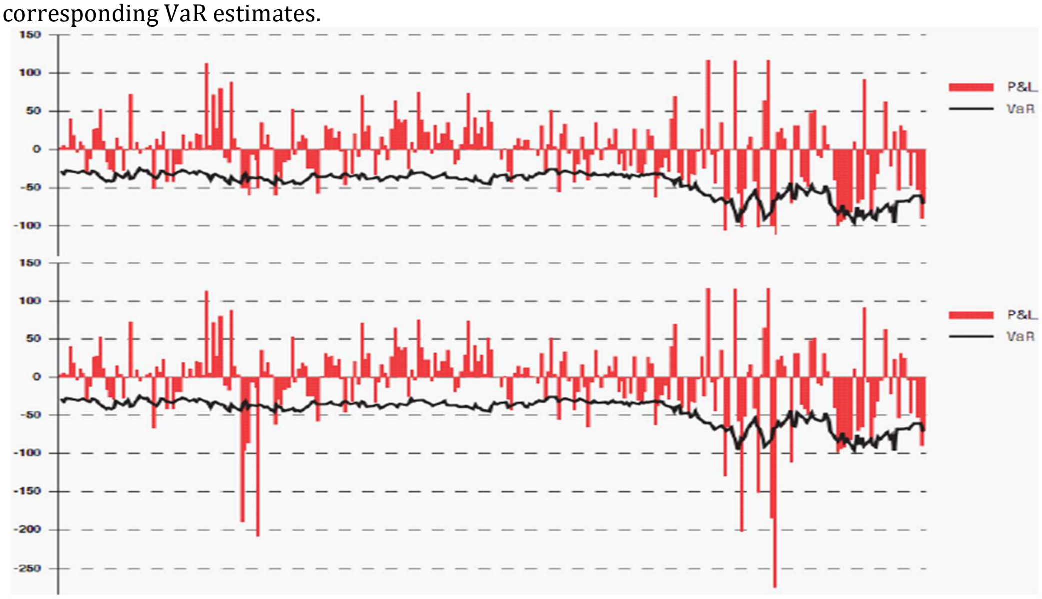
The main shortcoming of the coverage tests is that their results are not comparable. Indeed, the CC test only shows us whether a VaR model satisfies the CC hypothesis. It is clear that coverage test does take into account the magnitude of losses beyond VaR. Therefore, it does not allow us to compare the performance of two VaR models that both pass the statistical test. We show an example in Figure 3.1, which shows the P/L and VaR of bank 1 and bank 2 presumably. According to the CC test, two banks perform equally since they have the same number of VaR exceptions with identical hit sequence. However, it can be seen that the excessive

[21] See Section 2.2.4.1
[22] See Section 2.2.4.2

losses of bank 2 are more serious than bank 1, implying that the performance of two VaR models are not comparatively equal. To overcome this drawback, we employ the LF, which is the magnitude-based test, to evaluate the comparative performance of VaR models.

Figure 3.1: Example of coverage evaluation

Notes: Figure 3.1 shows example of two banks that have the same number of VaR exceptions but different excessive losses. The red columns present the trading P/L while the black line shows the corresponding VaR estimates.



3.3.2 The evaluation of comparative performance

According to the risk management perspective, both of the number of VaR exceptions and their magnitude are matters of concern. To deal with these specific concerns of the risk managers, Lopez (1998, 1999) proposes the LF approach to quantify the magnitude of excessive losses beyond VaR. The general form of the LF for VaR model i at time t is defined as:

$$F_{i,\,t+1} = \begin{cases} f(r_{t+1}, VaR_{i,t}) & \text{if } r_{t+1} < VaR_{i,t} \\ g(r_{t+1}, VaR_{i,t}) & \text{if } r_{t+1} \geq VaR_{i,t} \end{cases} \tag{3.13}$$

where r_{t+1} is the trading P/L on day t+1.

To take into account the magnitude of VaR exceptions, we use the magnitude LF (Lopez, 1998), which is the quadratic form of the loss function:

$$LF_{i,\,t+1} = \begin{cases} (r_{t+1} - VaR_{i,t})^2 & \text{if } r_{t+1} < VaR_{i,t} \\ 0 & \text{otherwise} \end{cases} \qquad (3.14)$$

Suppose we have N observations. The single score to compare the performance between VaR models is computed as:

$$LFS = \frac{1}{N}\sum_{i=1}^{N} LF(VaR_t, r_{t+1}) \qquad (3.15)$$

It is important to note that the quadratic loss function is not a statistical test of model adequacy but a procedure to rank models by their scores. The idea of the quadratic loss function is expressed in a negative orientation: the more serious the failure, the lower the score. Thus, the performance of VaR models are examined by comparing the values of the their LFS in which the preferred model is the one that minimizes the loss.

The LF also has the shortcoming. Recall that the LF measures the magnitude of losses beyond the VaR and prefers model that minimizes the losses. It is obvious that the best VaR model is the one generating no VaR exceptions. It is the one-sided nature of the LF, which only accounts for the magnitude of excessive losses but ignores the frequency of VaR exceptions. Therefore, the LF prefers the VaR overstatement to the VaR accuracy and VaR understatement. It may give banks incentive to overstate their VaRs to have some economic merits, which were discussed in Section 2.6 in previous chapter. To overcome the drawback of the LF, we combine the LF to these statistical tests presented above. Specifically, we only implement the LF for VaR models which successfully pass the coverage test in the first stage. This will help us discard the case of highly-ranked LF score due to VaR overstatement.

3.4 Empirical analysis

3.4.1 Data description

We present the descriptive statistics of daily P/L of seven sample banks in Table 3.1, divided into three sub-periods. The pre-crisis is from the start date of the series to May 2007, while the financial crisis is from June 2007 to June 2009, and the post-crisis is from July 2009 onward. At each bank, we present the mean, minimum value (min), maximum value (max), the standard deviation (Std dev), the ratio of negative trading days (P/L <0), the skewness, kurtosis and Jarque-Berra test statistics for normality of the P/L series.

Table 3.1 shows that in pre-crisis period, for average all banks report to have good performance on their trading activities. In crisis and post-crisis period, most banks witness positive trend in their daily P/L, except Banco Santander which suffers a decrease in the trading revenues. Besides, the distribution of daily trading P/L of banks is highly skewed. In pre-crisis and post-crisis period, the P/L distribution tends to be positively skewed, but it is negatively skewed during financial crisis. The high kurtosis shows that banks have fat tails in their P/L distribution, in which the highest is recorded for Royal Bank of Canada. The fat-tailed distribution indicates the sign of extreme losses on bank's trading portfolio. The Jarque-Berra test for normality is rejected in most cases, except Societe Generale in post-crisis period due to the problem of short dataset.

Table 3.1: Descriptive statistics of daily P/L of sample banks

Notes: Table 3.1 presents the descriptive statistics of seven commercial banks from January 2001 to December 2012, divided into three sub-periods. The Jarque-Berra test statistics for normality are shown in the last column. The sign ** indicates that the test statistics is rejected at 95% confidence level.

	Date	Mean	Min	Max	Std Dev	P/L < 0	Skewness	Kurtosis	JB test
Pre-crisis period									
Intesa Sanpaolo	Jan 2005 - May 2007	0.1048	-23.370	22.796	5.5406	0.4802	-0.1056	6.6124	364.63**
Scotia Bank	Jan 2002 - May 2007	3.6668	-13.568	16.959	3.0208	0.0869	0.0390	4.4596	121.86**
Banco Santander	Jan 2005 - May 2007	1.9705	-66.886	38.366	8.0447	0.3762	-1.2819	15.8284	4321.2**
Bank of America	Jan 2001 - May 2007	15.4258	-57.388	96.759	14.4822	0.1059	0.1899	5.2973	364.63**
Royal Bank of Canada	Jan 2001 - May 2007	6.4211	-18.383	56.961	4.6188	0.0355	2.7184	26.5186	38942.6**
Deutsche Bank	Jan 2001 - May 2007	47.1337	-63.290	318.28	30.5518	0.0370	1.0950	8.5577	2371.4**
Societe Generale	Jan 2002 - May 2007	13.2450	-35.005	80.110	12.5048	0.1259	0.6977	5.4367	454.01**
Crisis period									
Intesa Sanpaolo	June 2007 - June 2009	0.0509	-5.080	7.0252	1.7078	0.4838	0.3254	4.0149	31.975**
Scotia Bank	June 2007 - June 2009	5.2835	-36.936	36.5763	7.1991	0.1905	-0.0584	6.6055	284.66**
Banco Santander	June 2007 - June 2009	-1.1390	-89.445	61.0342	13.2276	0.5295	-0.5197	9.0367	820.79**
Bank of America	June 2007 - June 2009	22.239	-171.64	320.76	63.1064	0.3238	0.5986	5.3894	156.24**
Royal Bank of Canada	June 2007 - June 2009	5.3605	-730.00	296.00	55.2473	0.2514	-6.7171	76.4666	122014.8**
Deutsche Bank	June 2007 - June 2009	26.433	-360.69	571.746	86.9326	0.2781	-0.4254	7.8923	539.39**
Societe Generale	June 2007 - June 2009	3.1587	-275.22	128.44	43.5052	0.4343	-1.1347	9.3338	990.22**
Post-crisis period									
Intesa Sanpaolo	July 2009 - Dec 2012	0.0663	-6.3547	10.114	1.8116	0.5169	0.5782	5.4293	267.21**
Scotia Bank	July 2009 - Dec 2012	6.3029	-14.574	27.211	5.0684	0.0920	0.1901	4.5185	90.974**
Banco Santander	July 2009 - Dec 2011	-0.0746	-40.096	63.035	9.3675	0.4582	-0.1495	7.1068	456.38**
Bank of America	July 2009 - Dec 2012	67.953	-119.75	317.29	58.0780	0.0923	0.5442	4.3031	106.66**
Royal Bank of Canada	July 2009 - Dec 2012	12.480	-91.314	179.74	15.8587	0.1178	1.2119	24.7167	17726.8**
Deutsche Bank	July 2009 - Dec 2012	7.3383	-128.85	226.66	33.9413	0.4215	0.6298	6.6070	542.52**
Societe Generale	July 2009 - Dec 2009	11.375	-57.398	70.434	23.2439	0.3281	-0.1174	2.8261	0.4554

It is important to note that the data used in this chapter is the daily P/L on trading portfolio of banks, which is in absolute value with currency unit. Recall that the literature on the performance of VaR models is entirely dependent on the use of daily returns of market indices, typically in percentage term. The banks' P/L data has some specific properties over the market data. The first is the non-zero expected daily return of the trading portfolio, which is presented on Table 3.1. The second is the highly skewed P/L distribution. This is due to the fact that the likelihood of having positive daily P/L is much higher than negative P/L. Indeed, we count the number of days that banks record positive trading P/L from 2001 to 2012 and find that for average, 76% of trading days are reported to have positive P/L. This is remarkably higher than the likelihood of having positive daily returns on popular market indices. In comparison, from 2001 to 2012 the S&P 500 index is recorded to have a ratio of 53% positive trading days, while at FTSE 100 it is 51%.

3.4.2 Forecasting methodology

We evaluate the predictive performance of alternative VaR models based on their out-of-sample forecasts. As commercial banks are required to disclose their daily VaR at 99% level of confidence, we estimate VaR(1%) with one-day-ahead forecast horizon. This chapter adopts the moving window technique to estimate VaR. We select two sizes of moving window, which are 252 trading days (or one year of historical data) and 504 trading days (two year of historical data), as they are most popular at banks. The moving window technique is simply described as following. With a window size of 504, the first window is placed between the day 1 and day 504. The estimation of VaR on day 505 is based on the first window, including 504 historical observations. Next, the window is moved one-step-ahead, starting from day 2 through day 505, to form a new in-sample data. With the new dataset, we re-estimate VaR model to obtain the VaR on day 506. This procedure is repeated until getting the full out-of-sample forecasts.

3.4.3 Preliminary analysis

3.4.3.1 Unit root test

First, we conduct the Augmented Dicky-Fuller test (ADF) to examine the existence of the unit root process in the time series of P/L. Recall that the ADF tests the existence of unit root in the time series, in which the null model:

$$r_t = r_{t-1} + \beta_1 \Delta r_{t-1} + \beta_2 \Delta r_{t-2} + \ldots + \beta_n \Delta r_{t-n} + \varepsilon_t \qquad (3.16)$$

against the alternative model:

$$r_t = \phi r_{t-1} + \beta_1 \Delta r_{t-1} + \beta_2 \Delta r_{t-2} + \ldots + \beta_n \Delta r_{t-n} + \varepsilon_t \quad \text{with } \phi < 1 \qquad (3.17)$$

We apply the ADF test firstly to the whole period (from 2001 to 2012) and then to all three sub-periods (pre-crisis, financial crisis and post-crisis period). The test results are presented in Table 3.2.

Table 3.2 shows that there is sufficient evidence to reject the null hypothesis of unit root at all cases. In other words, the data exhibits no unit root process. This implies that all time series of trading P/L are stationary and can be modelled directly without making any further transforms.

Table 3.2: Unit root test of daily P/L of sample banks

Notes: Table 3.2 presents the Augmented Dickey- Fuller test statistic and its corresponding p-value of seven sample banks. The second column shows the test result of the entire period, while the following columns report the test results of three sub-periods: pre-crisis, crisis and post-crisis. The p-value less than 0.001 is presented as 0.001.

	Entire sample	Pre-crisis	Crisis period	Post-crisis
Intesa Sanpaolo	-39.1477	-21.3698	-18.6206	-27.3719
	0.001**	0.001**	0.001**	0.001**
Scotia Bank	-27.4614	-17.5409	-13.8356	-14.4249
	0.001**	0.001**	0.001**	0.001**
Banco Santander	-33.1435	-20.5139	-17.1239	-20.7141
	0.001**	0.001**	0.001**	0.001**
Bank of America	-20.9289	-19.4858	-11.9500	-9.3743
	0.001**	0.001**	0.001**	0.001**
Royal Bank of Canada	-46.8952	-15.1763	-21.6421	-18.4330
	0.001**	0.001**	0.001**	0.001**
Deutsche Bank	-28.0853	-11.9358	-15.7510	-18.9095
	0.001**	0.001**	0.001**	0.001**
Societe Generale	-30.134	-17.0826	-17.5837	-7.4434
	0.001**	0.001**	0.001**	0.001**

**: significant at 95% level of confidence

3.4.3.2 Parameter estimation of GARCH-type models

We report the parameter estimation of GARCH-type models in Table 3.3 to examine its specifications. Due to limitation in space, we only present the parameter estimation of the first 2-year moving window in three sub-periods. Table 3.3 shows that the estimated value of constant mean $\hat{\mu}$ are statistically significant for banks that experienced positive mean P/L. Besides, the estimated values of ARCH and GARCH parameters are statistically significant in most cases. The GARCH parameter $\hat{\beta}$ ranges highly from 0.7 to 0.99, implying that the today's volatility has significant memory of past day volatility. In GJR-GARCH(1,1) specifications, the leverage effect is positively significant at several banks, which shows that past negative return has more impact on today volatility than past positive return.

Table 3.3: Parameter estimation of GARCH-type models

Notes: Table 3.3 presents the estimation results of GARCH-type models in three sub-periods. We report the estimated value of parameter and its associated t-statistics in parenthesis. The sign ** means that the estimated value of parameter is statistically significant at 95% level of confidence

Pre-crisis period	GARCHn				GJR-GARCHt					
	μ	ω	α	β	μ	ω	α	β	ξ	dof
Intesa Sanpaolo	0.142	0.0305	0.1525	0.848	0.109	0.047	0.132	0.821	0.093	12.358
	(1.413)	(0.696)	(6.169)**	(44.961)**	(0.998)	(0.7942)	(3.668)**	(30.376)**	(1.6735)	(1.991)**
Scotia Bank	3.598	0.448	0.051	0.899	3.561	0.426	0.042	0.901	0.019	12.382
	(44.153)**	(2.243)**	(3.225)**	(26.457)**	(44.314)**	(1.754)	(2.044)**	(22.723)**	(0.775)	(4.368)**
Banco Santander	1.3897	1.0891	0.206	0.783	1.351	0.865	0.194	0.804	-0.008	10.233
	(6.175)**	(2.972)**	(6.653)**	(27.246)**	(6.124)**	(2.102)**	(3.600)**	(22.750)**	(-0.131)	(1.921)
Bank of America	15.027	41.506	0.131	0.676	15.009	26.440	0.037	0.804	0.070	15.009
	(39.931)**	(4.656)**	(7.272)**	(13.770)**	(44.191)**	(2.553)**	(2.218)**	(12.776)**	(2.024)**	(44.192)**
Royal Bank of Canada	5.599	0.066	0.049	0.951	5.620	0.177	0.039	0.952	-0.009	6.081
	(66.072)**	(1.975)	(9.105)**	(150.02)**	(68.029)**	(2.071)**	(3.861)**	(79.124)**	(-0.709)	(7.778)**
Deutsche Bank	43.156	10.676	0.054	0.932	42.364	4.842	0.056	0.937	0.008	6.710
	(67.168)**	(3.485)**	(6.065)**	(81.343)**	(70.855)**	1.4645	(4.462)**	(68.646)**	(0.703)	(5.949)**
Societe Generale	11.956	4.127	0.050	0.922	11.808	4.549	0.060	0.917	-0.023	9.210
	(36.229)**	(3.687)**	(6.049)**	(65.950)**	(37.851)**	(2.415)**	(3.679)**	(41.008)**	(-1.298)	(4.750)**

	GJR-GARCHn					GARCHt				
	μ	ω	α	β	ξ	μ	ω	α	β	dof
Intesa Sanpaolo	0.121	0.029	0.119	0.836	0.090	0.121	0.048	0.169	0.830	11.219
	(1.126)	(0.635)	(4.703)**	(41.728)**	(1.981)	(1.119)	(0.822)	(4.735)**	(30.415)**	(2.134)**
Scotia Bank	3.586	0.479	0.039	0.892	0.034	3.565	0.402	0.049	0.906	12.236
	(43.470)**	(2.304)**	(2.311)**	(26.25)**	(1.911)	(44.895)**	(1.726)**	(2.641)**	(23.305)**	(4.615)**
Banco Santander	1.377	1.054	0.194	0.785	0.021	1.347	0.854	0.189	0.804	10.321
	(6.004)**	(2.895)**	(4.318)**	(26.558)**	(0.408)	(6.183)**	(2.120)**	(4.901)**	(23.149)**	(1.946)
Bank of America	15.031	38.666	0.113	0.692	0.033	14.926	50.457	0.111	0.650	6.166
	(39.870)**	(4.823)**	(6.767)**	(15.248)**	(1.252)	(43.955)**	(2.775)**	(3.652)**	(6.474)**	(6.001)**
Royal Bank of Canada	5.635	0.1605	0.0541	0.9512	-0.034	5.620	0.147	0.038	0.953	6.031
	(63.685)**	(3.263)**	(8.618)**	(144.92)**	(-4.728)**	(68.1)**	(2.113)**	(3.957)**	(81.358)**	(8.042)**
Deutsche Bank	43.161	10.581	0.054	0.932	0.001	42.337	5.248	0.059	0.938	6.728
	(65.279)**	(3.487)**	(5.976)**	(81.240)**	(0.143)	(71.406)**	(1.569)	(4.607)**	(68.493)**	(5.920)**
Societe Generale	11.933	4.236	0.049	0.934	-0.031	11.820	3.893	0.057	0.916	8.892
	(35.908)**	(4.211)**	(5.451)**	(79.685)**	(-3.444)**	(38.088)**	(2.231)**	(3.753)**	(39.973)**	(5.316)**

Table 3.3: continued

Crisis period	GARCHn				GJR-GARCHt					
	μ	ω	α	β	μ	ω	α	β	ξ	dof
Intesa Sanpaolo	0.1421	0.030	0.152	0.847	0.108	0.046	0.132	0.820	0.093	12.358
	(1.413)	(0.696)	(6.169)**	(44.961)**	(0.998)	(0.794)	(3.668)**	(30.376)**	(1.673)	(1.991)**
Scotia Bank	4.865	2.658	0.204	0.744	4.902	1.463	0.177	0.816	-0.033	5.806
	(23.068)**	(5.214)**	(8.018)**	(25.707)**	(26.938)**	(2.382)**	(3.605)**	(20.390)**	(-0.657)	(4.665)**
Banco Santander	0.141	4.937	0.1717	0.789	0.229	2.503	0.103	0.840	0.0779	8.3388
	(0.425)	(3.135)**	(6.497)**	(23.028)**	(0.741)	(2.040)**	(2.946)**	(25.970)**	(2.089)**	(2.831)**
Bank of America	22.815	48.596	0.228	0.772	21.167	39.303	0.243	0.746	0.021	6.403
	(19.371)**	(4.705)**	(9.363)**	(44.493)**	(19.488)**	(2.652)**	(4.349)**	(21.201)**	(0.428)	(6.828)**
Royal Bank of Canada	4.586	5.781	0.003	0.997	9.118	24.437	0.006	0.994	-0.006	2.043
	(0.8992	(0.588)	(2.536)**	(244.97)**	(19.732)**	(1.130)	(1.490)	(122.17)**	(-0.465)	(19.73)**
Deutsche Bank	44.658	166.728	0.153	0.829	44.658	167.657	0.082	0.851	0.074	4.953
	(21.191)**	(4.114)**	(6.557)**	(36.411)**	(22.899)**	(2.594)**	(2.532)**	(26.271)**	(2.099)**	(6.029)**
Societe Generale	10.896	17.158	0.133	0.863	10.542	27.692	0.123	0.841	0.0467	6.1596
	(10.399)**	(2.316)**	(7.172)**	(47.104)**	(10.745)**	(2.150)**	(3.485)**	(27.609)**	(1.099)	(4.165)**

	GJR-GARCHn					GARCHt				
	μ	ω	α	β	ξ	μ	ω	α	β	dof
Intesa Sanpaolo	0.121	0.029	0.118	0.836	0.090	0.121	0.048	0.169	0.830	11.219
	(1.126)	(0.635)	(4.703)**	(41.728)**	(1.981)**	(1.119)	(0.822)	(4.735)**	(30.415)**	(2.134)**
Scotia Bank	4.870	2.737	0.226	0.738	-0.038	4.900	1.386	0.1597	0.821	5.756
	(22.955)**	(5.086)**	(5.536)**	(23.182)**	(-0.925)	(27.020)**	(2.363)**	(4.189)**	(21.195)**	(4.656)**
Banco Santander	0.061	4.961	0.106	0.800	0.096	0.285	2.470	0.158	0.832	7.896
	(0.180)	(3.405)**	(3.496)**	(23.364)**	(2.742)**	(0.932)	(1.860)	(4.879)**	(25.479)**	(3.014)**
Bank of America	22.704	43.776	0.197	0.779	0.048	21.179	40.293	0.255	0.745	6.346
	(17.792)**	(3.555)**	(5.792)**	(36.624)**	1.4767	(19.796)**	(2.739)**	(5.709)**	(22.238)**	(6.888)**
Royal Bank of Canada	-0.410	82.956	0.1052	0.947	-0.105	9.129	24.858	0.028	0.972	2.075
	(-0.125)	(2.700)**	(2.605)**	(50.071)**	(-2.539)**	(19.711)**	(1.534)	(1.973)**	(90.203)**	(60.316)**
Deutsche Bank	44.658	243.276	0.090	0.823	0.084	44.658	142.95	0.147	0.839	4.770
	(18.516)**	(4.569)**	(4.112)**	(31.270)**	(3.310)**	(23.221)**	(2.293)**	(3.906)**	(24.283)**	(5.972)**
Societe Generale	10.967	17.722	0.116	0.863	0.032	10.823	25.626	0.149	0.843	6.076
	(9.792)**	(2.469)**	(6.387)**	(46.455)**	(1.406)	(10.76)**	(2.057)**	(4.340)**	(28.216)**	(4.124)**

Table 3.3: continued

Post-crisis period	GARCHn				GJR-GARCHt					
	μ	ω	α	β	μ	ω	α	β	ξ	dof
Intesa Sanpaolo	0.101	0.023	0.109	0.884	0.082	0.012	0.067	0.901	0.061	200
	(1.707)	(1.022)	(3.926)**	(31.568)**	(1.347)	(0.588)	(2.669)**	(36.213)**	(1.743)	(0.098)
Scotia Bank	6.461	1.727	0.131	0.797	6.340	1.152	0.109	0.841	0.007	9.051
	(40.263)**	(3.216)**	(5.472)**	(21.162)**	(40.138)**	(2.138)**	(3.085)**	(19.752)**	(0.176)	(2.855)**
Banco Santander	0.249	6.460	0.136	0.799	0.261	15.672	0.025	0.687	0.217	30.867
	(0.607)	(2.271)**	(4.011)**	(15.973)**	(0.638)	(2.753)**	(0.468)	(8.270)**	(2.705)**	(0.654)
Bank of America	62.739	93.668	0.063	0.911	60.243	63.071	0.088	0.893	0.017	7.061
	(29.932)**	(2.564)**	(4.616)**	(46.476)**	(31.310)**	(1.337)	(3.095)**	(28.37)**	(0.788)	(4.577)**
Royal Bank of Canada	11.764	2.048	0.056	0.943	11.257	10.175	0.005	0.925	0.0543	3.067
	(27.562)**	(4.877)**	(7.617)**	(190.69)**	(14.855)**	(2.009)**	(0.087)	(29.831)**	(2.285)**	(9.508)**
Deutsche Bank	1.401	20.560	0.085	0.889	0.946	15.978	0.086	0.891	0.015	5.257
	(1.618)	(3.343)**	(5.808)**	(45.762)**	(1.162)	(1.952)	(3.147)**	(32.712)**	(0.561)	(4.867)**
Societe Generale	5.574	100.266	0.138	0.808	5.025	128.172	0.157	0.779	0.015	4.644
	(3.513)**	(2.753)**	(4.735)**	(19.469)**	(3.369)**	(1.853)	(2.448)**	(11.085)**	(0.230)	(3.749)**

	GJR-GARCHn					GARCHt				
	μ	ω	α	β	ξ	μ	ω	α	β	dof
Intesa Sanpaolo	0.083	0.012	0.066	0.902	0.061	0.096	0.026	0.111	0.880	55.693
	(1.368)	(0.605)	(2.712)**	(37.063)**	(1.793)	(1.616)	(1.011)	(3.726)**	(28.970)**	(0.373)
Scotia Bank	6.439	1.704	0.112	0.798	0.041	6.342	1.140	0.112	0.842	8.993
	(39.723)**	(3.106)**	(3.583)**	(20.441)**	(1.091)	(40.319)**	(2.133)**	(3.794)**	(19.92)**	(3.033)**
Banco Santander	0.227	15.115	0.019	0.694	0.291	0.293	7.262	0.126	0.799	14.783
	(0.551)	(3.021)**	(0.402)	(9.420)**	(3.199)**	(0.723)	(2.008)**	(3.190)**	(13.271)**	(1.494)
Bank of America	62.866	87.938	0.061	0.912	0.007	59.963	78.911	0.097	0.887	7.099
	(28.491)**	(2.132)**	(3.627)**	(40.575)**	(0.499)	(31.461)**	(1.659)	(3.515)**	(28.272)**	(4.552)**
Royal Bank of Canada	11.714	3.076	0.032	0.943	0.039	11.422	9.416	0.021	0.9351	2.975
	(20.416)**	(5.653)**	(4.403)**	(173.40)**	(4.483)**	(31.013)**	(1.613)	(1.861)	(28.453)**	(9.622)**
Deutsche Bank	1.345	21.845	0.079	0.884	0.019	0.919	15.333	0.090	0.895	5.231
	(1.463)	(3.275)**	(4.437)**	(42.580)**	(1.172)	(1.139)	(1.951)	(3.585)**	(34.041)**	(4.884)**
Societe Generale	5.590	96.3072	0.148	0.812	-0.021	5.007	124.907	0.164	0.781	4.645
	(3.303)**	(2.768)**	(4.192)**	(20.138)**	(-0.659)	(3.364)**	(1.887)	(2.786)**	(11.440)**	(3.759)**

3.4.4 Evaluating the predictive power of VaR models

3.4.4.1 The evaluation of absolute performance

In the first stage of our backtesting, we evaluate the absolute performance of VaR models. Specifically, we use the statistical tests to examine the UC, IND and CC hypotheses of the hit sequence corresponding to VaR estimates respectively. Recall that the UC test examines the coverage of VaR exceptions based on the actual number of VaR exceptions. The null hypothesis of UC test states that the observed frequency of VaR exceptions is consistent with the expected coverage rate[23]. The CC test extends the UC test by jointly examining the IND and UC properties of VaR exceptions. The CC hypothesis holds when both UC and IND hypotheses are simultaneously satisfied. A VaR model is not rejected by the CC test if the VaR exceptions are independently distributed and the actual frequency of VaR exceptions is not significantly different to the expected coverage rate[24].

The test results of the pre-crisis period are presented in Table 3.4, while Table 3.5 and Table 3.6 report the evaluation of the crisis and post-crisis period. Although the performance of VaR models substantially varies across modelling approaches and distributional assumption, some clear points emerge. We first discuss the performance of banks' internal modes, then the HS-based models, the GARCH-based models and finally the EVT-based model.

[23] See Section 2.2.4.1

[24] See Section 2.2.4.2

Table 3.4: The absolute performance of VaR estimates in pre-crisis period

Notes: Table 3.4 presents the CC test results of VaR estimates in pre-crisis period. VaRs are estimated using 2-year moving window (504 trading days) and one-year moving window (252 trading days equivalently). At each sample bank, we present the actual exception rate (Ratio), the test statistics of the UC hypothesis (LR_{UC}), the IND hypothesis (LR_{IND}) and the CC hypothesis (LR_{CC}). We do not compute the test statistics for banks that experience no VaR exceptions.

2-year moving window		Bank VaRs	Naïve HS	Filtered HS	Riskmetrics	GARCHn	GARCHt	GJR-GARCHn	GJR-GARCHt	EVT
Intesa Sanpaolo	Ratio	0	0	0	0.0194	0.0097	0.0097	0.0097	0.0097	0
	LR_{UC}	-	-	-	0.7429	0.0004	0.0004	0.0004	0.0004	-
	LR_{IND}	-	-	-	0.0800	0.0198	0.0198	0.0198	0.0198	-
	LR_{CC}	-	-	-	0.8229	0.0202	0.0202	0.0202	0.0202	-
Scotia bank	Ratio	0	0.0092	0.0103	0.0080	0.0056	0.0023	0.046	0.023	-
	LR_{UC}	-	0.0506	0.0141	0.3401	1.8343	7.4939**	3.1551	7.4939**	-
	LR_{IND}	-	0.1494	0.1893	0.1142	0.0581	0.0093	0.0372	0.0093	-
	LR_{CC}	-	0.2000	0.2034	0.4543	1.8925	7.5031**	3.1923	7.5031**	-
Banco Santander	Ratio	0	0	0.029	0.0194	0.0194	0.0097	0.019	0.0097	0
	LR_{UC}	-	-	2.5519	0.7429	0.7429	0.0004	0.7429	0.0004	-
	LR_{IND}	-	-	3.6828	0.0800	0.0800	0.0198	0.0800	0.0198	-
	LR_{CC}	-	-	6.2347	0.8229	0.8229	0.0202	0.8229	0.0202	-
Bank of America	Ratio	0.0027	0.0054	0.0090	0.0126	0.0081	0.0018	0.0108	0.0027	0
	LR_{UC}	8.4096**	2.8414	0.1139	0.7068	0.4290	11.420**	0.0718	8.4096**	-
	LR_{IND}	0.0163	0.0652	0.1818	1.9219	3.5816	0.0072	2.4756	0.0163	-
	LR_{CC}	8.4258**	2.9066	0.2957	2.6287	4.0107	11.427**	2.5474	8.4258**	-
Royal Bank of Canada	Ratio	0	0.0163	0.0172	0.0127	0.0036	0.0018	0.0054	0.0018	0
	LR_{UC}	-	3.7742	4.8276**	0.7608	5.9521	11.255**	2.7493	11.255**	-
	LR_{IND}	-	1.0912	0.9341	0.3610	0.0292	0.0073	0.0658	0.0073	-
	LR_{CC}	-	4.8654	5.7617	1.1218	5.9813	11.262**	2.8151	11.262**	-
Deutsche Bank	Ratio	0	0.0109	0.0128	0.0128	0.0045	0.0018	0.0045	0.0018	0
	LR_{UC}	-	0.1065	0.8114	0.8114	4.0499**	11.107**	4.0499**	11.107**	-
	LR_{IND}	-	0.2669	0.3640	0.3640	0.0460	0.0073	0.0460	0.0073	-
	LR_{CC}	-	0.3735	1.1754	1.1754	4.0959**	11.114**	4.0959**	11.114**	-
Societe Generale	Ratio	0.0034	0.0136	0.0204	0.0079	0.0045	0	0.0045	0.0011	0
	LR_{UC}	5.1552**	1.0703	1.7848	0.3917	3.2968	-	3.2968	11.284**	-
	LR_{IND}	0.0206	0.3326	0.3908	0.0964	0.0366	-	0.0366	0.0023	-
	LR_{CC}	5.1757	1.4028	2.1756	0.4881	3.3334	-	3.3334	11.286**	-

**: significant at 95% confidence level

Table 3.4: continued

1-year moving window		Bank VaRs	Naïve HS	Filtered HS	Riskmetrics	GARCHn	GARCHt	GJR-GARCHn	GJR-GARCHt	EVT
Intesa Sanpaolo	Ratio	0	0.0112	0.0508	0.0141	0.0084	0.0028	0.0112	0.0056	0.0141
	LR_{UC}	-	0.0606	30.314**	0.5476	0.0847	2.5557	0.0606	0.7941	0.5476
	LR_{IND}	-	4.7843	3.4849	3.8140	0.0514	0.0057	0.0917	0.0228	3.8140
	LR_{CC}	-	4.8449	33.799**	4.3616	0.1361	2.5614	0.1523	0.8169	4.3616
Scotia bank	Ratio	0	0.0125	0.0134	0.0134	0.0062	0.0044	0.0107	0.0071	0.0008
	LR_{UC}	-	0.6755	1.2048	1.2048	1.8057	4.3253	0.0623	1.0027	15.588**
	LR_{IND}	-	1.9310	6.0404	1.6944	0.0884	0.0450	0.2609	0.1155	0.0018
	LR_{CC}	-	2.6065	7.2452	2.8992	1.8941	4.3703	0.3232	1.1182	15.590**
Banco Santander	Ratio	0	0.0031	0.0649	0.0225	0.0254	0.0169	0.0225	0.0169	0.0169
	LR_{UC}	-	10.226**	48.378**	4.2077**	5.9927	1.4430	4.2077	1.4430	1.4430
	LR_{IND}	-	9.1740**	13.765**	1.9575	1.5443	3.0607	1.9575	3.0607	3.0607
	LR_{CC}	-	19.399**	62.144**	6.1653**	7.5370	4.5038	6.1653	4.5038	4.5038
Bank of America	Ratio	0.0027	0.0088	0.0095	0.0146	0.0110	0.0029	0.0168	0.0022	0
	LR_{UC}	8.4096**	0.2004	0.0280	2.6475	0.1388	9.4923**	5.4213**	12.230**	-
	LR_{IND}	0.0163	2.8425	7.9431**	1.0884	2.0267	0.0236	3.5672	0.0133	-
	LR_{CC}	8.4258**	3.0428	7.9711**	3.7359	2.1654	9.5159**	8.9885**	12.243**	-
Royal Bank of Canada	Ratio	0	0.0155	0.0155	0.0110	0.0051	0.0014	0.0081	0.0029	0
	LR_{UC}	-	3.5879	3.5879	0.1603	3.8464**	15.477**	0.5029	9.3504**	-
	LR_{IND}	-	8.6907**	8.6907**	0.3368	0.0729	0.0059	0.1806	0.0238	-
	LR_{CC}	-	12.278**	12.278**	0.4971	3.9193	15.483**	0.6835	9.3741**	-
Deutsche Bank	Ratio	0	0.0013	0.0013	0.0016	0.0044	0.0007	0.0044	0.0007	0
	LR_{UC}	-	1.4263	1.4263	4.6446**	5.2215**	19.762**	5.2215**	19.7623**	-
	LR_{IND}	-	1.3914	1.3914	3.8302	0.0539	0.0015	0.0539	0.0015	-
	LR_{CC}	-	2.8177	2.8177	8.4748**	5.2754	19.763**	5.2754	19.7637**	-
Societe Generale	Ratio	0.0034	0.0106	0.0115	0.0123	0.0035	0.0009	0.0061	0.0026	0
	LR_{UC}	5.1552**	0.0442	0.2494	0.6105	6.3265**	15.826**	1.9043	8.6895**	-
	LR_{IND}	0.0206	0.2578	0.3029	0.3263	0.0284	0.0018	0.0873	0.0160	-
	LR_{CC}	5.1757	0.3020	0.5523	0.9368	6.3549**	15.828**	1.9917	8.7055**	-

**: significant at 95% confidence level

Table 3.5: The absolute performance of VaR estimates in crisis period

Notes: Table 3.5 presents the CC test results of VaR estimates in crisis period. VaRs are estimated using 2-year moving window (504 trading days) and one-year moving window (252 trading days equivalently). At each sample bank, we present the actual exception rate (Ratio), the test statistics of the UC hypothesis (LR_{UC}), the IND hypothesis (LR_{IND}) and the CC hypothesis (LR_{CC}). We do not compute the test statistics for banks that experience no VaR exceptions.

2-year moving window		Bank VaRs	Naïve HS	Filtered HS	Riskmetrics	GARCHn	GARCHt	GJR-GARCHn	GJR-GARCHt	EVT
Intesa Sanpaolo	Ratio	0.0019	0.0114	0.0171	0.0095	0.0095	0.0057	0.0095	0.0057	0.0057
	LR_{UC}	5.2019**	0.1064	2.2436	0.0113	0.0113	1.1434	0.0113	1.1434	1.1434
	LR_{IND}	0.0038	3.8041	14.1854	4.5747	0.0963	0.0345	0.0963	0.0345	6.8983
	LR_{CC}	5.2057**	3.9105	16.4289	4.5860	0.1076	1.1780	0.1076	1.1780	8.0417
Scotia bank	Ratio	0.0038	0.0209	0.0304	0.0171	0.0152	0.0057	7/525	0.0095	0.0076
	LR_{UC}	2.6475	4.8588**	14.425**	2.2436	1.2646	1.1434	0.5402	0.0113	0.3227
	LR_{IND}	0.0153	5.7194**	6.7405	0.3146	0.2481	0.0345	0.1896	0.0963	5.5583
	LR_{CC}	2.6628	10.5782**	21.165**	2.5582	1.5127	1.1780	0.7298	0.1076	5.8811
Banco Santander	Ratio	0.0057	0.0266	0.0704	0.0438	0.0438	0.0190	0.0361	0.0152	0.0171
	LR_{UC}	1.1434	10.145**	79.099**	33.137**	33.137**	3.4491	21.796**	1.2646	2.2436
	LR_{IND}	6.8983**	20.292**	24.368**	9.5964**	2.9338	1.8329	1.8368	0.2481	22.337**
	LR_{CC}	8.0417**	30.437**	103.46**	42.733**	36.070**	5.2820	23.633**	1.5127	24.581**
Bank of America	Ratio	0.0285	0.0323	0.0971	0.0228	0.0419	0.0209	0.0495	0.0266	0.04
	LR_{UC}	12.216**	16.763**	144.74**	6.4545**	30.155**	4.8588**	42.614**	10.145**	27.268**
	LR_{IND}	18.674**	21.896**	32.983**	0.5626	3.3476	1.5070	1.8921	0.7740	11.385**
	LR_{CC}	30.981**	38.659**	177.72**	7.0171**	33.502**	6.3658**	44.506**	10.919**	38.653**
Royal Bank of Canada	Ratio	0.0323	0.04	0.0552	0.0304	0.0304	0.0247	0.0304	0.0190	0.0342
	LR_{UC}	16.763**	27.268**	52.821**	14.425**	14.425**	8.2210**	14.425**	3.4491	19.222**
	LR_{IND}	0.3183	3.7988	5.4765**	2.9242	0.4444	0.6615	0.4444	0.3891	2.1619
	LR_{CC}	17.081**	31.067**	58.298**	17.349**	14.869**	8.8825**	14.869**	3.8382	21.3841
Deutsche Bank	Ratio	0.0590	0.0342	0.0514	0.0266	0.0419	0.0133	0.04	0.0152	0.0095
	LR_{UC}	59.996**	19.222**	45.939**	10.145**	30.155**	0.5402	27.268**	1.2646	0.0113
	LR_{IND}	10.426**	14.600**	18.612**	3.8640**	6.5378**	0.1896	0.0306	2.6544	0.0963
	LR_{CC}	70.422**	33.822**	64.551**	14.009**	36.692**	0.7298	27.298**	3.9190	0.1076
Societe Generale	Ratio	0.0667	0.0476	0.0933	0.0342	0.0323	0.0095	0.0323	0.0114	0.0247
	LR_{UC}	75.152**	39.370**	135.35**	19.222**	16.763**	0.0113	16.763**	0.1064	8.2210**
	LR_{IND}	4.8693**	16.327**	27.794**	2.1619	0.3183	0.0963	0.3183	0.1390	9.2609**
	LR_{CC}	80.022**	55.697**	163.15**	21.384**	17.081**	0.1076	17.081**	0.2454	17.481**

**: significant at 95% confidence level

Table 3.5: continued

1-year moving window		Bank VaRs	Naïve HS	Filtered HS	Riskmetrics	GARCHn	GARCHt	GJR-GARCHn	GJR-GARCHt	EVT
Intesa Sanpaolo	Ratio	0.0019	0.0095	0.0247	0.0095	0.0057	0.0057	0.0076	0.0076	0.0133
	LR_{UC}	5.2019**	0.0113	8.2210**	0.0113	1.1434	1.1434	0.3227	0.3227	0.5402
	LR_{IND}	0.0038	12.840**	0.9829	4.5747**	0.0345	0.0345	0.0615	0.0615	27.884**
	LR_{CC}	5.2057**	12.851**	9.2039**	4.5860	1.1780	1.1780	0.3843	0.3843	28.424**
Scotia bank	Ratio	0.0038	0.0190	0.0247	0.0190	0.0209	0.0076	0.0266	0.0095	0.0076
	LR_{UC}	2.6475	3.4491	8.2210**	3.4491	4.8588**	0.3227	10.145**	0.0113	0.3227
	LR_{IND}	0.0153	1.8329	0.9829	0.3891	0.4718	0.0615	0.7687	0.0963	5.5583**
	LR_{CC}	2.6628	5.2820	9.2039**	3.8382	5.3306	0.3843	10.914**	0.1076	5.8811
Banco Santander	Ratio	0.0057	0.0285	0.0380	0.0419	0.0380	0.0209	0.0323	0.0171	0.0342
	LR_{UC}	1.1434	12.216**	24.480**	30.155**	24.480**	4.8588**	16.763**	2.2436	19.222**
	LR_{IND}	6.8983**	7.5018**	23.199**	10.459**	1.5451	5.7194**	0.3183	0.3146	9.6069**
	LR_{CC}	8.0417**	19.718**	47.679**	40.614**	26.025**	10.578**	17.081**	2.5582	28.829**
Bank of America	Ratio	0.0285	0.0323	0.0647	0.0209	0.0380	0.0152	0.0380	0.0190	0.0342
	LR_{UC}	12.216**	16.763**	71.2674**	4.8588**	24.480**	1.2646	24.480**	3.4491	19.222**
	LR_{IND}	18.674**	15.846**	18.9496**	0.4718	1.5451	2.6544	0.0727	1.8329	14.600**
	LR_{CC}	30.981**	32.609**	90.2170**	5.3306	26.025**	3.9190	24.553**	5.2820	33.822**
Royal Bank of Canada	Ratio	0.0323	0.0209	0.0247	0.0304	0.0419	0.0133	0.0266	0.0133	0.0380
	LR_{UC}	16.763**	4.8588**	8.2210**	14.425**	30.155**	0.5402	10.145**	0.5402	24.480**
	LR_{IND}	0.3183	1.5070	0.9829	2.9242	0.0067	0.1896	0.7687	0.1896	1.5451
	LR_{CC}	17.081**	6.3658**	9.2039**	17.349**	30.161**	0.7298	10.914**	0.7298	26.025**
Deutsche Bank	Ratio	0.0590	0.0304	0.0380	0.0285	0.0266	0.0133	0.0304	0.0152	0.0114
	LR_{UC}	59.996**	14.425**	24.480**	12.216**	10.145**	0.5402	14.425**	1.2646	0.1064
	LR_{IND}	10.426**	11.548**	29.520**	3.3694	0.7740	0.1896	0.4444	0.2481	3.8041
	LR_{CC}	70.422**	25.973**	54.001**	15.586**	10.919**	0.7298	14.869**	1.5127	3.9105
Societe Generale	Ratio	0.0667	0.0380	0.0533	0.0323	0.0304	0.0114	0.0361	0.0152	0.0209
	LR_{UC}	75.152**	24.480**	49.343**	16.763**	14.425**	0.1064	21.796**	1.2646	4.8588
	LR_{IND}	4.8693**	7.9548**	22.127**	2.5231	0.4444	0.1390	0.1339	2.6544	5.7194**
	LR_{CC}	80.022**	32.435**	71.470**	19.286**	14.869**	0.2454	21.930**	3.9190	10.578**

**: significant at 95% confidence level

Table 3.6: The absolute performance of VaR estimates in post-crisis period

Notes: Table 3.6 presents the CC test results of VaR estimates in post-crisis period. VaRs are estimated using 2-year moving window (504 trading days) and one-year moving window (252 trading days equivalently). At each sample bank, we present the actual exception rate (Ratio), the test statistics of the UC hypothesis (LR_{UC}), the IND hypothesis (LR_{IND}) and the CC hypothesis (LR_{CC}). We do not compute the test statistics for banks that experience no VaR exceptions.

2-year moving window		Bank VaRs	Naïve HS	Filtered HS	Riskmetrics	GARCHn	GARCHt	GJR-GARCHn	GJR-GARCHt	EVT
Intesa Sanpaolo	Ratio	0.0112	0.0101	0.0349	0.0124	0.0112	0.0056	0.0112	0.0067	0.0045
	LR_{UC}	0.1449	0.0026	33.986**	0.4898	0.1449	2.0071	0.1449	1.0454	3.3738
	LR_{IND}	8.4954**	9.4077	16.813**	2.3949	0.2286	0.0568	0.2286	0.0819	17.284**
	LR_{CC}	8.6403**	9.4102	50.799**	2.8847	0.3734	2.0639	0.3734	1.1273	20.657**
Scotia bank	Ratio	0.0022	0.0033	0.0044	0.0101	0.0044	0.0011	0.0044	0.0011	0
	LR_{UC}	7.8823**	5.3148**	3.4291	0.0011	3.4291	11.498**	3.4291	11.498**	-
	LR_{IND}	0.0090	0.0203	0.0361	3.1687	0.0361	0.0022	0.0361	0.0022	-
	LR_{CC}	7.8913**	5.3351	3.4653	3.1698	3.4653	11.500**	3.4653	11.500**	-
Banco Santander	Ratio	0.0062	0.0046	0.0263	0.0154	0.0185	0.0123	0.0154	0.0123	0.0015
	LR_{UC}	1.0871	2.3258	12.025**	1.6899	3.8482**	0.3495	1.6899	0.3495	7.2182**
	LR_{IND}	0.0499	0.0280	7.1462**	2.1888	0.4550	0.2009	0.3150	0.2009	0.0031
	LR_{CC}	1.1370	2.3538	19.172**	3.8786	4.3032	0.5505	2.0048	0.5505	7.2213**
Bank of America	Ratio	0	0.0056	0.0045	0.0101	0.0056	0.0056	0.0067	0.0056	0
	LR_{UC}	-	2.0247	3.3959	0.0019	2.0247	2.0247	1.0584	2.0247	-
	LR_{IND}	-	0.0567	0.0362	0.1845	0.0567	0.0567	0.0817	0.0567	-
	LR_{CC}	-	2.0814	3.4322	0.1864	2.0814	2.0814	1.1401	2.0814	-
Royal Bank of Canada	Ratio	0.0033	0.0056	0.0011	0.0101	0.0067	0	0.0101	0	0
	LR_{UC}	5.3148**	2.0511	11.498**	0.0011	1.0780	-	0.0011	-	-
	LR_{IND}	0.0203	0.0565	0.0022	0.1839	0.0814	-	0.1839	-	-
	LR_{CC}	5.3351	2.1076	11.500**	0.1850	1.1595	-	0.1850	-	-
Deutsche Bank	Ratio	0.0078	0.0033	0.0011	0.0201	0.0123	0.0033	0.0123	0.0033	0
	LR_{UC}	0.4464	5.3281**	11.516**	7.2291**	0.4608	5.3281**	0.4608	5.3281**	-
	LR_{IND}	4.1818	0.0203	0.0022	0.7423	0.2750	0.0203	0.2750	0.0203	-
	LR_{CC}	4.6282	5.3484	11.518**	7.9714**	0.7358	5.3484	0.7358	5.3484	-
Societe Generale	Ratio	0.0312	0	0	0.0078	0.0078	0	0.0078	0	0
	LR_{UC}	3.7779	-	-	0.0625	0.0625	-	0.0625	-	-
	LR_{IND}	9.5472**	-	-	0.0159	0.0159	-	0.0159	-	-
	LR_{CC}	13.325**	-	-	0.0784	0.0784	-	0.0784	-	-

**: significant at 95% confidence level

Table 3.6: continued

1-year moving window		Bank VaRs	Naïve HS	Filtered HS	Riskmetrics	GARCHn	GARCHt	GJR-GARCHn	GJR-GARCHt	EVT
Intesa Sanpaolo	Ratio	0.0112	0.0101	0.0383	0.0124	0.0112	0.0079	10/886	0.0112	0.0079
	LR_{UC}	0.1449	0.0026	41.953**	0.4898	0.1449	0.4208	0.1449	0.0026	0.4208
	LR_{IND}	8.4954**	3.1582	6.9201**	2.3949	0.2286	0.1116	2.7519	3.1582	21.026**
	LR_{CC}	8.6403**	3.1608	48.873**	2.8847	0.3734	0.5324	2.8967	3.1608	21.447**
Scotia bank	Ratio	0.0022	0.0078	0.0067	0.0112	0.0067	0.0022	0.0067	0.0022	0
	LR_{UC}	7.8823**	0.4421	1.0780	0.1321	1.0780	7.8823**	1.0780	7.8823**	-
	LR_{IND}	0.0090	4.1796	0.0814	2.7621	4.8244**	0.0090	4.8244**	0.0090	-
	LR_{CC}	7.8913**	4.6217	1.1595	2.8941	5.9024	7.8913**	5.9024	7.8913**	-
Banco Santander	Ratio	0.0062	0.0123	0.0232	0.0216	0.0201	0.0139	0.0185	0.0170	0.0139
	LR_{UC}	1.0871	0.3495	8.3341**	6.6890**	5.1900**	0.9068	3.8482**	2.6764	0.9068
	LR_{IND}	0.0499	3.0342	0.7144	1.0628	0.5348	0.2547	1.5511	0.3817	15.408**
	LR_{CC}	1.1370	3.3837	9.0485**	7.7518**	5.7249	1.1615	5.3994	3.0582	16.315**
Bank of America	Ratio	0	0.0090	0.0078	0.0146	0.0135	0.0056	0.0157	0.0101	0
	LR_{UC}	-	0.0891	0.4293	1.6986	1.0047	2.0247	2.5487	0.0019	-
	LR_{IND}	-	0.1456	0.1114	0.3867	2.0830	0.0567	1.5519	0.1845	-
	LR_{CC}	-	0.2348	0.5406	2.0853	3.0877	2.0814	4.1006	0.1864	-
Royal Bank of Canada	Ratio	0.0033	0.0101	0.0078	0.0134	0.0089	0.0022	0.0157	0.0011	0.0022
	LR_{UC}	5.3148**	0.0011	0.4421	0.9834	0.0952	7.8823**	2.5137	11.498**	7.8823**
	LR_{IND}	0.0203	3.1687	0.1110	0.3280	0.1451	0.0090	0.4475	0.0022	0.0090
	LR_{CC}	5.3351	3.1698	0.5531	1.3115	0.2403	7.8913**	2.9613	11.500**	7.8913**
Deutsche Bank	Ratio	0.0078	0.0179	0.0179	0.0190	0.0201	0.0033	0.0179	0.0044	0.0033
	LR_{UC}	0.4464	4.6104**	4.6104**	5.8598**	7.2291**	5.3281**	4.6104**	3.4402	5.3281**
	LR_{IND}	4.1818	15.570**	15.570**	0.6614	0.7974	0.0203	4.7270**	0.0361	0.0203
	LR_{CC}	4.6282	20.180**	20.180**	6.5211**	8.0265**	5.3484	9.3374**	3.4763	5.3484
Societe Generale	Ratio	0.0312	0	0	0.0156	0.0078	0.0078	0.0078	0.0078	0
	LR_{UC}	3.7779	-	-	0.3608	0.0625	0.0625	0.0625	0.0625	-
	LR_{IND}	9.5472**	-	-	0.0640	0.0159	0.0159	0.0159	0.0159	-
	LR_{CC}	13.325**	-	-	0.4248	0.0784	0.0784	0.0784	0.0784	-

**: significant at 95% confidence level

We first investigate the performance of bank's internal VaR models. Among seven banks in the sample, while only Deutsche Bank employ Monte Carlo simulation, all the remaining banks use HS as the internal VaR model[25]. In pre-crisis period, Table 3.4 shows that bank VaRs were conservatively estimated and therefore produce very rare, even no VaR exceptions. As a result, the statistical tests strongly reject both the UC and CC hypothesis of all seven banks. In crisis period, the performance of bank VaRs varies. While the VaR overstatements still exhibit at some banks (Scotia Bank, Intesa Sanpaolo and Banco Santander), the others remarkably underestimate their VaRs and suffer high exception rates. Besides, the LR_{IND} shows evidence of VaR violation clustering during crisis period. This is due to the fact that most banks use HS as their internal VaR model, which does not account for volatility clustering. In post-crisis period, bank VaRs still poorly perform. The statistical tests reject the validity of VaR models at four over seven banks due as they overstate VaR to keep small exception ratios.

The HS models, including the Naïve HS and the Filtered HS models, perform reasonably well in normal periods. With 2-year moving window, both the Naïve HS and the Filtered HS provide good VaR estimates for most banks in pre-crisis period. However, when we shorten the moving window to 1-year, the predictive power of Filtered HS is much worse. Indeed, the semi-parametric model provides poor coverage with a number of non i.i.d VaR exceptions. In post-crisis period, while the Naïve HS performs substantially well, the Filtered HS continues to provide poor coverage rate. It is also noticed that with shorter window size, the performance of

[25] This information can be obtained on banks' annual reports.

Naïve HS slightly improves, as it eliminates the impact of aged data in financial crisis to the VaR estimates in post-crisis period.

The shortcoming of the unconditional models is clearly shown in crisis period. Indeed, both HS models produce very high exception ratios and fail the statistical tests at most banks. Their performance does not improve when we change the size of moving windows. The poor performance of the HS models can be explained that they are not able to incorporate with the changes in market volatility during financial crisis. To catch up with the market turbulence, there are two potential approaches. The first approach, used in this chapter, is to use the time-varying conditional volatility models e.g. the GARCH-type models to capture daily volatility. The second approach is to use the Markov-switching models to characterize the market behaviours in different regimes, and therefore better capture the changes in market volatility in crisis period. Within this approach, the requirement for real-time indicator of when to switch regime is essential. We do not carry out the regime-switching models in this study and suggest them for future work.

The conditional volatility models with Gaussian assumption, including the Riskmetrics, GARCHn and GJR-GARCHn, seem to be the best forecasting models in normal market conditions. With 2-year moving window, the Gaussian-based models provide accurate VaR estimates and therefore pass the statistical tests in most cases. However, these models become much less accurate with shorter moving window. We find the similar results in post-crisis period, when they perform better with longer window size. During financial crisis, the Gaussian-based models produce very poor coverage rates. Table 3.5 documents their actual exception ratios, which are significantly larger than the nominal rate of 1%. The IND test indicates the existence of the clustering of VaR exceptions in crisis period. As a result, the statistical tests reject the accuracy of the conditional models with Gaussian innovation in most cases. The poor performance of these models in crisis period can be attributed to the assumption of normal distribution, which is not able to capture the significant changes in market conditions during crisis period.

We keep the same GARCH specifications but change the distributional assumption from Gaussian to Student t. Comparing the performance of these models, we find that incorporating the Student t significantly improves the predictive power of the GARCH-type models in crisis period. As having advantage of the fat-tailed distribution, the Student t-based conditional volatility models show their superiority in providing accurate and robust VaR estimates across banks and moving windows in crisis period, while the Gaussian-based models tend to underestimate the risk. Table 3.5 shows that the frequency of VaR exceptions produced by Student t-based models are close to the nominal rate of 1%, while it is remarkably higher than 1% in case of Gaussian-based models. Besides, the GARCH dynamics show their ability of dealing with volatility clustering. As a result, the performance of the Student t-based GARCH models are hardly rejected by any statistical tests.

While performing well in financial crisis, both GARCHt and GJR-GARCHt produce poor VaR estimates in normal periods. Indeed, the statistical tests show that these models noticeably inflate VaR estimates and consequently produce very low exception ratios in both pre-crisis and post-crisis periods. This can be explained that compared to the Gaussian distribution, the Student t has heavier tails. Therefore, VaR estimates using Student t assumption tend to be more conservative than using Gaussian assumption, which is appropriate in estimating VaR in turbulent period rather than in normal period.

The coverage tests show the poor performance of EVT model. In both pre-crisis and post-crisis period, the EVT seriously overestimates VaR and therefore generates very rare or no VaR exceptions. This result is not surprising as EVT was designed to deal with extreme events, which normally happen during financial crisis. Thus, the normal market conditions are not the ideal environment for this approach. However, the EVT model still poorly perform in crisis period. Table 3.5 shows that VaR estimates using EVT seem to be understated for most banks. By generating a high number of non i.i.d exceptions, VaR estimates using EVT are rejected across banks. Besides, the number of rejections increases when we shorten the size of moving window from two years to one year. The poor performance of the EVT in crisis period is inconsistent with the findings of prior studies, which show the superiority of EVT approach in estimating VaR using market returns (see Gencay and Selcuk, 2004; Chan and Gray, 2006, Aloui et al., 2011; Schaumburg, 2012; Adrian and Shin, 2013). However, it is important to note that prior studies estimate VaR using market returns instead of bank's data. Applying to the trading P/L data of commercial banks, we argue that good VaR estimates can be obtained by using more simple and accessible models rather than EVT approach or banks' internal models.

3.4.4.2 The comparative performance and the selection of VaR models

Using the magnitude LF, this section aims to compare the performance of VaR models that passes the first stage evaluation. Therefore, we do not compute the LF for models that were rejected by the coverage tests. Recall that the magnitude LF scores models based on their quadratic magnitude of excessive losses. As the LF is negatively oriented, we rank model performance based on their LF: the lower the LF, the higher the rank. A model is ranked as number 1 if it has the lowest LF, indicating the best model in line. We present the LF ranking of VaR estimates using 2-year moving window in Table 3.7, while the similar results using 1-year moving window are presented in Table 3.8.

Table 3.7 and Table 3.8 show that the comparative performance of VaR models varies across banks and the sizes of moving windows. In order to comprehensively compare and select the best VaR model in each period, we propose the model selection framework. Specifically, a VaR model is assessed following two criteria: (i) the number of accurate VaR estimates according to the coverage tests in the first stage evaluation and (ii) the ranking of the magnitude LF in the second stage evaluation. We argue that a good VaR model firstly has to provide VaR estimates that satisfy the coverage tests. Thus, we prefer VaR model that has higher number of accurate VaR estimates according to the statistical tests. If two VaR models have the same number of successful VaR estimates, we rely on their LF to select the more accurate model. As the LF is negatively oriendted, we compute the average LF ranking of a VaR model in each sub-period, in which the lower value is preferable[26]. We present the selection of VaR models in Table 3.9.

Table 3.9 documents the superiority of conditional volatility models in estimating VaR of commercial banks. In pre-crisis period, we find that the most

[26] As model ranking is negatively oriented.

accurate VaR models are the conditional volatility models with Gaussian assumption. These models, including the Riskmetrics, GARCHn and GJR-GARCHn, dominate the top performance models in both moving window sizes. While the Naïve HS performs consistently, the Filtered HS shows their poor performance in pre-crisis period. The worst performances in pre-crisis period are bank VaRs and EVT, as they significantly inflate VaR estimates. In the post-crisis period, the performance of alternative models does not change significantly. While Gaussian-based conditional models still perform best at 2-year moving window, we document the success of GARCHt, Riskmetrics and Naïve HS in VaR estimation using 1-year moving window.

Table 3.7: The ranking of magnitude LF of VaR estimates using 2-year moving window

Notes: Table 3.7 presents the ranking of various VaR estimates using 2-year moving window based on their LF. We do not compute LF for the cases that VaR model fails the statistical tests in the first stage. The ranking of VaR models are based on the magnitude of LF, which the lower LF receives the higher rank. In each row, the number 1 indicates the best model which has the lowest LF.

	Bank VaRs	Naïve HS	Filtered HS	RM	GARCH	GARCHt	GJR-GARCH	GJR-GARCHt	EVT
Pre-crisis period									
Intesa Sanpaolo	-	-	-	5	3	1	4	2	-
Scotia Bank	-	3	4	5	1	-	2	-	-
Banco Santander	-	-	6	3	4	2	5	1	-
Bank of America	-	3	1	5	2	-	4	-	-
Royal Bank of Canada	-	4	-	1	2	-	3	-	-
Deutsche Bank	-	3	2	1	-	-	-	-	-
Societe Generale	-	4	5	2	1	-	3	-	-
Crisis period									
Intesa Sanpaolo	-	7	8	2	6	4	5	3	1
Scotia Bank	1	-	-	7	5	2	6	3	4
Banco Santander	-	-	-	-	-	2	-	1	-
Bank of America	-	-	-	-	-	-	-	-	-
Royal Bank of Canada	-	-	-	-	-	-	-	1	-
Deutsche Bank	-	-	-	-	-	2	-	1	
Societe Generale		-	-	-	-	2	-	1	-
Post-crisis period									
Intesa Sanpaolo	-	5	-	6	3	2	4	1	-
Scotia Bank	-	-	3	4	1	-	2	-	-
Banco Santander	2	1	-	6	-	3	5	4	-
Bank of America	-	4	1	7	5	2	6	3	-
Royal Bank of Canada	-	1	-	4	2	-	3	-	-
Deutsche Bank	1	-	-	-	2	-	3	-	-
Societe Generale	-	-	-	3	2	-	1	-	-

Table 3.8: The ranking of magnitude LF of VaR estimates using 1-year moving window

Notes: Table 3.8 presents the ranking of various VaR estimates using 1-year moving window based on their LF. We do not compute LF for the cases that VaR model fails the statistical tests in the first stage. The ranking of VaR models are based on the magnitude of LF, which the lower LF receives higher rank. In each row, the number 1 indicates the best model which has the lowest LF

	Bank VaRs	Naïve HS	Filtered HS	RM	GARCH	GARCHt	GJR-GARCH	GJR-GARCHt	EVT
Pre-crisis period									
Intesa Sanpaolo	-	6	-	5	4	1	3	2	-
Scotia Bank	-	5	6	8	3	2	7	4	1
Banco Santander	-	-	-	-	3	1	4	2	5
Bank of America	-	2	-	3	1	-	-	-	-
Royal Bank of Canada	-	-	-	1	-	-	2	-	-
Deutsche Bank	-	2	1	-	-	-	-	-	-
Societe Generale	-	3	4	2	-	-	1	-	-
Crisis period									
Intesa Sanpaolo	-	-	-	-	2	1	4	3	-
Scotia Bank	1	6	-	5	4	2	-	3	-
Banco Santander	-	-	-	-	-	-	-	1	-
Bank of America	-	-	-	-	-	2	-	1	-
Royal Bank of Canada	-	-	-	-	-	1	-	2	-
Deutsche Bank	-	-	-	-	-	1	-	2	
Societe Generale	-	-	-	-	-	1	-	2	-
Post-crisis period									
Intesa Sanpaolo	-	6	-	5	2	1	4	3	-
Scotia Bank	-	3	1	2	-	-	-	-	-
Banco Santander	1	4	-	-	-	2	-	3	-
Bank of America	-	4	3	6	5	1	7	2	-
Royal Bank of Canada	-	2	1	3	4	-	5	-	-
Deutsche Bank	2	-	-	-	-	-	-	1	-
Societe Generale	-	-	-	1	5	3	4	2	-

Table 3.9: Selection of VaR models

Notes: Table 3.9 presents the selection of VaR models in three sub-periods using two moving window sizes. For each VaR model, we report the number of banks (N) that the VaR model can provide accurate VaR estimates that satisfy the statistical tests, the average value of LF ranking (ALF) and the overall ranking (Rank). We prefer VaR model that has the highest N. If two VaR models have the same N, we prefer the model that has lower ALF.

	2-year moving window			1-year moving window		
	N	ALF	Rank	N	ALF	Rank
Pre-crisis period						
Bank VaRs	0	-	8	0	-	9
Naïve HS	5	3.4	4	5	3.6	2
Filtered HS	5	3.6	5	3	3.67	7
Riskmetrics	7	3.14	1	5	3.8	3
GARCHn	6	2.17	2	4	2.75	4
GARCHt	2	1.5	6	3	1.33	5
GJR-GARCHn	6	3.5	3	5	3.4	1
GJR-GARCHt	2	1.5	6	3	2.33	6
EVT	0	-	8	2	3	8
Crisis period						
Bank VaRs	1	1	7	1	1	4
Naïve HS	1	7	8	1	6	7
Filtered HS	1	8	9	0	-	8
Riskmetrics	2	4.5	4	1	5	6
GARCHn	2	5.5	5	2	3	3
GARCHt	5	2.4	2	6	1.33	2
GJR-GARCHn	2	5.5	5	1	4	5
GJR-GARCHt	6	1.67	1	7	2	1
EVT	2	2.5	3	0	1	8
Post-crisis period						
Bank VaRs	2	1.5	7	1	2	8
Naïve HS	4	2.75	4	5	3.8	3
Filtered HS	2	2	8	3	1.67	7
Riskmetrics	6	5	3	5	3.4	2
GARCHn	6	2.5	2	4	4	35
GARCHt	3	2.33	5	4	1.75	4
GJR-GARCHn	7	3.43	1	4	5	6
GJR-GARCHt	3	2.33	5	5	2.2	1
EVT	0	-	9	0	-	9

We acknowledge the predictive power of GARCH-type models with Student t assumption in crisis period. While the GJR-GARCHt is the best VaR model in both window sizes, the GARCHt is consistently the runner-up. Besides, it is important to note that the difference in the performance of conditional volatility models with Student t assumption and the other models is noticeable. Table 3.9 reports that using 2-year moving window, the GARCHt can provide accurate VaR estimates for five over seven banks. The EVT model, ranked as number three, only succeeds in forecasting VaR of two banks. In case of 1-year moving window, the gap in performance is even more significant.

In brief, our empirical evaluation shows the superiority of conditional volatility models in estimating bank VaRs. Regarding to the choice of distributional assumption, we find that the conditional volatility models with Gaussian distribution perform well in normal periods, while incorporating Student t distribution significantly improves models' performance in crisis period. The EVT approach, which was shown to be powerful in literature, poorly perform in estimating bank VaRs. Coming with the poor performance of the bank VaRs, we argue that good VaR estimates at commercial banks can be obtained by using more simple and accessible models rather than the complicated approach or banks' internal models.

3.5 Concluding remarks

This chapter investigates the forecasting power of various VaR models using daily trading P/L of seven commercial banks. Our dataset is from 2001 to 2012, covering the pre-crisis, crisis and post-crisis period. The competing VaR models used in this chapter are banks' internal model, the Naïve and Filtered HS, the conditional volatility models[27] with different distributional assumptions and the EVT approach.

[27] These include the Riskmetrics, GARCHn, GJR-GARCHn, GARCHt and GJR-GARCHt

To comprehensively evaluate VaR models, we develop a two-stage backtesting framework. In the first stage, we assess the absolute performance of VaR models using coverage tests. The second stage quantifies the magnitude LF in order to compare model performances.

Our empirical evaluation shows the superiority of conditional volatility models in estimating bank VaRs. Regarding to the choice of distributional assumption, we find that the Gaussian distribution uniformly improves VaR predictive power in normal periods, while the Student t is by far the best in estimating VaR during financial crisis. While the HS models perform inconsistently, none of the banks' internal model accurately capture bank risk. The EVT approach, which was documented to be superior in estimating in VaR in prior studies, performs very poorly with bank data. Thus, we argue that good VaR estimates at commercial banks can be obtained using simple and accessible models rather than other complicated models.

The findings of this chapter leave a puzzle: what are banks doing and why? Banks rationally know the poor performance of their internal VaR models, and they are smart enough to find the alternative models to produce much better results. But as we witness, banks are still using their poorly performed VaR models, which generate poor VaR estimates. We assume that banks have incentives to use their inaccurate VaR models to get some economic merits of VaR overstatement, which were discuss in Section 2.6 of the previous chapter. Therefore, this study has several suggestions to the financial regulators and public investors. First, financial regulators should be concerned with the problem of VaR conservativeness by penalizing banks which inflate their VaRs. We suggest that the VaR overstatement should be treated equally as the VaR understatement. Second, public investors should not look at the VaR figures and other market risk disclosure to infer the risk profile of a bank. As shown in this study, banks tend to overstate their VaR to have very rare, or even no VaR exceptions. By doing this, risk model at banks seems to perform well and therefore, banks seem to look safe to public investors. However, we find that the disclosed risk figures are intentionally manipulated and do not present the real risk profile of a bank.

Chapter 4:

Improving quantile forecast accuracy

4.1 Introduction

Recent literature in financial econometrics has witnessed the wide range of parametric models that focus on density forecasting. This approach requires financial decisions to incorporate the estimation and simulation of the entire distribution of future changes in returns and volatility of financial assets. One of the prime applications in decision making is measuring and managing tail risk (Value-at-Risk). This can be defined as the specific quantile of forecasted density. However, the recent global crisis raised number of questions about the misleading results of quantile forecasts in general and VaR forecasts in particular. These criticisms have led to the growing interest in improving the accuracy of quantile forecast and especially, of Value-at-Risk estimates.

In the past decades, the predominant approach in estimating conditional return distribution was represented by the GARCH model proposed by Engle (1982) and Bollerslev (1986) and followed by a number of sophisticated specifications to the underlying model. The GARCH model proposes time variation in the return distribution mainly dependent on the conditional variance, and has received a great success in clarifying several empirical specifications of asset returns. However, the main drawback of the GARCH model is that its validity depends on the assumption of the dynamics of the underlying conditional distribution of returns. This problem of parametric model specification has now been helped with the availability of high-frequency data and the so-called realized volatility (RV) literature, an ex post measure of volatility. Andersen and Bollerslev (1998), Andersen et al. (2001a,b) use realized volatility extracted from high-frequency data as a proxy for the volatility of the low-frequency returns. The latent variance process in this approach is now observable and measurable, representing a model-free estimator of daily quadric variation of return dynamic. Correspondingly, a number of models have been proposed to incorporate with high-frequency data, including GARCHX-type models of Engle (2002), the Mixed data sampling regression model (MIDAS) proposed by Ghysels et al. (2004), the Multiplicative error model (MEM) of Engle and Gallo (2006), the Heterogeneous autoregressive model of realized volatility (HAR-RV) of Corsi (2009), the High-frEquency-bAsed VolatilitY (HEAVY) models of Shephard and Sheppard (2010), the RV-based bivariate models of Maheu and McCurdy (2011), the RV-based linear quantile regression (LQR-RV) model of Zikes and Barunik (2016).

The literature on economic forecasting has traditionally devoted on the estimation and evaluation of point forecasts for the conditional mean of economic variables. For a given information set at specific time, forecasts can be obtained by using linear formulation or other specific functional forms, with time varying or time-invariant coefficients, be sophisticated in dynamic specifications or using the simple ones. There are also a number of information sets available to be used in forecasting. Thus, there might be a variety of forecasts which can be obtained alternatively. Studies have suggested that a combination of alternative forecasts can improve the forecasts accuracy. A forecast combination can be considered as the way of pooling information from individual forecasts. The literature on forecast combination is not novel. It is common that forecast combinations produce superior forecasts than methods based upon individual forecasting models (Granger and Ramanathan, 1984; Clemen, 1989; Makridakis and Hibon, 2000; Stock and Watson, 2004).

Recent years have witnessed the growing interest of forecasting other characteristics of the forecast distribution rather than its conditional mean, such as specific conditional quantiles. A prime example of the increasing attention to quantile forecast is in risk management perspective with the popular application of VaR. Recall from Chapter 2 that VaR places a threshold of losses in sense that the amount of losses will exceed this threshold with a small target probability α. It is obvious that VaR(α%) is the α %-quantile forecast of the return distribution. Forecasting VaR is indeed forecasting the specific conditional quantile of the return distribution. Therefore, any methods to improve the accuracy of quantile forecasts can be applicable to increase the predictive power of VaR models. This chapter aims to examine potential methods to enhance the quantile forecast accuracy to extend the findings of the previous chapters, which focus on the evaluation of VaR estimates.

There are number of approaches to estimate conditional quantiles in general and VaR in particular, varying from parametric (e.g. Danielsson and de Vries, 1997; Barone-Adesi, 1998; Diebold et al., 1998; Embrechts et al., 1999; McNeil and Frey, 2000) to semi-parametric (Koenker and Zhao, 1996; Taylor, 1999; Christoffersen et al., 2000; Engle and Manganelli, 2004; Komunjer, 2005) to non-parametric (Bhattacharya and Gangopadhyay, 1990; White, 1992). Although there are number of studies on quantile forecasts, it is surprising that little empirical research has been conducted in the combination of conditional quantile forecasts. Timmermann (2005) studies the optimal combination of conditional mean forecasts at lengths and only suggests the optimization of combining conditional quantile forecasts as a further implication of the principle of optimal combination. Giacomini and Komunjer (2005) use Generalized Method-of-Moment estimation to obtain the optimal combination weights of two forecasts and find that a linear combination (with intercept) of two forecasts outperform both individual components. Halbleib and Pohlmeier (2012) propose two VaR optimal combination methods based on the framework of Conditional coverage and Quantile regression approaches. Applying to stock data in crisis period, the combined forecasts are superior to the stand-alone forecasts. Although both combination methods perform well, Halbleib and Pohlmeier (2012) show that the quantile regression-based optimization is more stable and provides better results than the conditional coverage-based optimization.

The contributions of this chapter to the literature are twofold. First, our empirical analysis acknowledges the value of the high-frequency data on the measure of volatility to characterize the quantile forecast of asset returns. Second, we find that the use of quantile combination can significantly improve the accuracy of quantile forecasts. To the best of our knowledge, this chapter is the first study investigating the quantile combination with the use of high-frequency data.

This chapter uses the 5-minute sampling frequency of WTI Crude Oil Futures from 3rd Jan 1995 through on 30th June 2016. The reason we select the data of oil market to carry out the quantile forecasts is that oil market becomes increasingly important for desk-level trading at banks. In order to determine whether the use of high-frequency data improves the accuracy of quantile forecast, we compare the performance of RV-based forecasting models to the GARCH(1,1). These models include the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) of

Corsi (2009), the High-frequency-based volatility (HEAVY) model of Shephard and Sheppard (2010) and the RV-based Linear quantile regression (LQR-RV) model of Zikes and Barunik (2016). The performance of competing models is evaluated at the realized returns. In the first stage, we examine the absolute performance of each model by focusing on the dynamic of the hit process at each quantile level. The second stage quantifies the comparative performance of models using the Quantile Scores (QS). We find evidence that the HAR-RV and HEAVY models remarkably outperform the GARCH(1,1) at most quantile levels and at different forecast horizons, while the LQR-RV performs poorly. Thus, we acknowledge the value of high-frequency data in estimating conditional volatility and quantile forecasting.

This study also examines whether the quantile combination can improve forecast accuracy. Using the Conditional quantile optimization method (CQOM) of Halbleib and Pohlmeier (2012), we find that the combined quantile forecasts remarkably outperform stand-alone forecasts across quantile levels and forecast horizons. With an VaR implication, we show that the VaR combinations are superior to individual VaR estimates in providing accurate and robust results.

This chapter is organized as followed. Section 4.2 sets out the theoretical framework and models for the construction of quantile forecast. Section 4.3 discusses about the method of evaluation, while section 4.4 briefly describes the data. The empirical application is carried out in Section 4.5. Section 4.6 presents some VaR implications, and Section 4.7 gives summary of the chapter.

4.2 Methodology

This section presents the approaches to quantile forecasting and quantile combination. The literature provided two approaches to quantile forecasting: the indirect forecasting and the direct forecasting. The first approach is commonly used in prior studies (Clements et al., 2007; Gallo and Brownless, 2010; Hua and Mazan, 2013). Here, conditional volatility is used to estimate quantile forecasts. In the second approach, quantile forecasts are obtained directly through the quantile regression framework of Koenker and Basset (1978). In this chapter, we employ both approaches in order to forecast the quantiles of the return distribution.

4.2.1 The indirect quantile forecasting approach

In the indirect approach, quantile forecast is the output of two factors: the econometric model used to estimate conditional volatility, and the method used extract the conditional quantiles from the volatility forecasts. That is, the forecasted quantiles are inferred from the corresponding predictive densities, which are obtained either by assuming or simulating the return process based on the estimation of conditional volatility. The econometric models we adopt are the conditional volatility models, in which the return process is conditionally heteroskedastic, generated in the following way:

$$r_t = \mu + z_t \sigma_t, \quad z_t \xrightarrow{i.i.d} D(0,1) \tag{4.1}$$

where $D(0,1)$ is a probability distribution with zero mean and unit variance, and μ and σ_t are the estimated constant mean and conditional standard deviation of the return process at day t respectively. In empirical work, the estimated value of constant mean μ is very close to zero for daily data, thus the assumption of zero mean is commonly used. Therefore, the return process is now only dependent on the estimated value of conditional volatility $\widehat{\sigma_t}$. This raises the important role of appropriate and accurate volatility forecasting, which is vital to financial asset return modelling. Consequently, the predictive α-quantile of r_{t+h} conditional on the information set at time t, denoted as $q_\alpha(r_{t+h}|\Omega_t)$, relies on the estimated conditional volatility $\widehat{\sigma_t}$ and the method of obtaining quantile forecasts from estimated conditional volatility $\widehat{\sigma_t}$. The literature on volatility measures, volatility forecasting and quantile extracting method will be presented in the following sections.

4.2.1.1 Volatility measures

One of the most popular measures of return variation is the variance, a mathematical expectation of the average deviation from the mean. For daily data, variance can be defined as squared return $r^2{}_t$. This computation, although of interest, has been widely criticized as a noisy measure of variance (Andersen and Bollerslev, 1998).

The use of high-frequency data in modelling and forecasting volatility has been widely promoted (Andersen et al., 2009). The availability of the high-frequency financial data allows us to extract useful information for inferences (Engle and Gallo, 2006; Brownlees and Gallo, 2010; Shephard and Sheppard, 2010; Maheu and McCurdy, 2011). Prior studies suggest that the inclusion of high-frequency data is beneficial as it enlarges the data set available for forecasting. Indeed, the tick-by-tick data contains detailed characterizations of the asset, including liquidity supply and demand, the trading dynamics, the dependence structure and contagion effects.

As the rich source of information, the high-frequency data allows researchers to construct nonparametric estimators, called realized measures, to estimate the variation of the price path of a financial asset during the time when the asset is traded frequently. The literature of realized measures include realized volatility (Andersen et al., 2001a,b), realized absolute variation (Ghysels, Santa-Clara and Valkanov, 2006), realized power variation (Forsberg and Ghysels, 2004), bi-power realized volatility (Barndoff-Nielsen and Shephard, 2004), two-scaled realized volatility (Zhang et al., 2005) and realized kernel (Barndorff-Nielsen et al., 2008). Among these measures, the realized volatility (Andersen et al., 2001) has been widely-used and is often the benchmark high-frequency volatility estimates. The realized volatility can simply be obtained by taking the sum of intraday squared returns. Denote realized volatility estimator on day t by RV_t. The realized volatility on day t, RV_t, can be computed as:

$$RV_t = \sum_{i=1}^{n} r^2{}_{i,t} \tag{4.2}$$

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i-1,t}) \tag{4.3}$$

in which $r_{i,t}$ is the intraday return between time i and i-1 within day t, and n is the number of intraday returns in day t. In the absence of jumps and microstructure effects, as sampling frequency n increases to infinity, the realized volatility RV_t converges to underlying integrated volatility, a natural volatility measure (Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002). For this reason, the choice of the sampling frequency n plays an important role in the construction of realized volatility. In our empirical analysis, we select a 5-minute sampling frequency, as it has been shown to provide a reasonably balance between the demand for finely sampled observations and the robustness to the contaminations of market microstructure effects[28] (Hua and Mazan, 2013). Besides, empirical studies have shown that 5-minute realized volatility is hardly outperformed by any other frequency measures (Liu et al., 2015).

4.2.1.2 The GARCH(1,1) models

In order to see whether the use of high-frequency data can improve the accuracy of quantile forecasts, we select the simplest GARCH(1,1) specification as a benchmark for the conditional volatility model using low-frequency data. The reason we choose the simplest GARCH(1,1) specification of Bollerslev (1986) among hundreds of GARCH-type models is due to its popularity, simplicity, ease-of-use and efficiency. In particular, Hansen and Lunde (2005) extensively compare the out-of-sample performance of 330 GARCH-type specifications in forecasting volatility of exchange rate and show that the simplest GARCH(1,1) outperforms other complicated specifications.

Denote r_t the close-to-close return on day t. Simply, r_t can be obtained by:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \tag{4.4}$$

in which P_t is the closing price of given asset on day t.

We assume the return process follow a time-varying location-scale model with zero mean assumption:

$$r_t = z_t \sigma_t \tag{4.5}$$

[28] Market microstructure effects includes order size, order arrival date, bid-ask spread, price impact, price resilience and market efficiency which vary across assets and across time (Kyle and Obizhaeva, 2016).

where z_t is an i.i.d error term assumed to follow a distribution function with zero mean and unit variance, and σ_t denotes the conditional standard deviation on day t. The dynamic of the conditional variance, according to GARCH(1,1) specification, is as followed:

$$\sigma^2{}_{t+1} = \omega + \alpha\varepsilon^2{}_t + \beta\sigma^2{}_t \tag{4.6}$$

where ω, α, β are parameters. The estimated parameters $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ can be obtained by Maximum Likelihood Estimation to generate the forecast value of σ_{t+1}.

To completely specify the GARCH(1,1) models, we need to make assumption about the distribution of error term z_t. The first GARCH(1,1) model comes with the Gaussian distribution (denoted as GARCHn), as it has been the most widely-used statistical assumption and has become the benchmark in financial econometrics. We also investigate the forecasting power of GARCH(1,1) model with Student t innovation assumption (denoted as GARCHt) due to its ability to cope with fat-tailed distribution. Besides, the GARCHt model had confirmed its superiority in the previous chapters.

4.2.1.3 The realized volatility models

The high-frequency-based econometric models have been widely constructed in the studies of Andersen et al. (2003), Andersen et al. (2004), Koopman et al. (2005), Ghysels and Sinko (2006), Andersen et al. (2007), Clements et al. (2008), Corsi (2009), Shephard and Sheppard (2010), Brownlees and Gallo (2010), Maheu and McCurdy (2011), Zikes and Barunik (2016). One of the advantages of the RV is that it provides model-free estimates of volatility of returns of an asset, which do not rely on the assumption of parametric models e.g. GARCH-type models. Empirically, it is evident that the forecasting models using realized measures significantly outperform the popular GARCH-type models and stochastic volatility models in terms of out-of-sample forecasting (Andersen et al., 2003).

The literature on forecasting realized measures is not novel. One of the first works is from Andersen et al. (2001, 2003; 2007) who focus on applying least squares estimators of autoregressive specification of realized measures and their logarithm forms. Their papers provide support for long memory in the series of realized measures by calling the HAR-RV model. Engle (2002) develops a GARCHX-type model for return process by using realized variances based on 5-minute high-frequency data. His result shows that squared daily return is helpful in forecasting realized variance, although there might be some uncertainty over the statistical significance of this effect. This research was extended by Engle and Gallo (2006) who study multiple volatility measures in order to gather information across volatility indicators rather than concentrate on high-frequency based statistics solely. Their work develops the MEM which is shown to perform well in short to medium range forecast horizons. With the same model structure, Shephard and Sheppard (2010) extend the GARCH-X specification and propose HEAVY models incorporating the dynamic of both returns and realized measures. In literature, the HEAVY models are witnessed to perform systematically better than the GARCH model when apply to the comprehensive dataset of intraday data of stock indices worldwide. Because of its performance and is the comprehensive representative for the GARCH-type realized volatility models, it is selected to be one of the three high frequency-based models in this study.

Another popular approach to high-frequency data is the MIDAS model proposed by Ghysels et al. (2004). This model involves time series with different frequencies, which is shown to be more efficient than the traditional approach of aggregating all series with different frequency to the least frequent sampling. The favour of MIDAS regressions is also confirmed by Ghysels et al. (2006) when it outperforms other linear forecast (ARCH-type) models encompassing daily realized volatility in terms of both in-sample and out-of-sample forecasts of high-frequency exchange rate data. However, it is important to note that the MIDAS model is generally outperformed by the HAR-RV model in forecasting quantiles of the daily exchange rate returns of five pair of currencies (Clements et al., 2008). The superiority of HAR-RV model over other realized measure-based GARCH-type models is also confirmed in the recent studies (see Maheu and McCurdy, 2011; Hua and Mazan, 2013; Celik and Ergin, 2014; Vortelinos, 2015; Huang et al., 2016). Due to the its success in prior studies, we select the long-memory HAR-RV model as one of the high frequency-based conditional volatility models to use in this study, along with the HEAVY model.

The HAR-RV model

The HAR-RV model provides a method for volatility forecasting with different interval sizes. Inspired by the Heterogeneous Market Hypothesis (Muller et al., 1993), the HAR-RV model focuses on the heterogeneity generating from the difference in the time horizons, as financial market is combined as different participants with different trading frequency and volume. In particular, there are participants which have a huge spectrum of trading frequency, while institutional traders tend to trade less frequently but with higher volume. Participants with different time horizons react to, and generate different sorts of volatility elements: the short-term participants with daily of intraday trading frequency; the medium-term traders who normally restructure their positions weekly and the long-term institutional investors with a typical time of one month or longer. Thus, Corsi (2009) proposes a model with three volatility components matching the time horizons of different market participants: one-day, one-week and one-month. The HAR-RV model in integrated form is expressed in the following way:

$$r_t = z_t \sigma_t \tag{4.7}$$

$$\sigma^d{}_{t+1} = c + \beta_d \, RV^d{}_t + \beta_w \, RV^w{}_t + \beta_m \, RV^m{}_t + \nu^d{}_{t+1} \tag{4.8}$$

in which $RV^d{}_t$, $RV^w{}_t$, $RV^m{}_t$ are the daily, weekly and monthly observed realized volatilities respectively. Define $RV^d{}_t = RV_t$, where RV_t is the realized volatility on day t. The past weekly and monthly observed realized volatilities are defined as:

$$RV^w{}_t = \frac{1}{5}\left(RV^d{}_t + RV^d{}_{t-1} + \ldots + RV^d{}_{t-4}\right) \tag{4.9}$$

$$RV^m{}_t = \frac{1}{22}\left(RV^d{}_t + RV^d{}_{t-1} + \ldots + RV^d{}_{t-21}\right) \tag{4.10}$$

Besides, $\sigma^d{}_{t+1}$ can be written as:

$$\sigma^d{}_{t+1} = RV^d{}_{t+1} + \nu^d{}_{t+1} \tag{4.11}$$

Thus, we obtain a time series presentation of the HAR-RV model:

$$RV^d{}_{t+1} = c + \beta_d \, RV^d{}_t + \beta_w \, RV^w{}_t + \beta_m \, RV^m{}_t + \nu^d{}_{t+1} \tag{4.12}$$

At this point, the coefficients β(.) of HAR-RV model can be simply obtained by simple linear regression. This study uses the standard OLS regression estimators to obtain the estimated values of coefficients β(.) in the following section.

The HEAVY model

Base on the structure of the GARCHX-type models of Engle (2002), Shephard and Sheppard (2010) develop the HEAVY model to harness high frequency data to conduct multistep-ahead forecasts of the volatility of financial returns.

Suppose we have a data set of daily returns $r_1, r_2, \ldots r_T$. Let us denote the low-frequency data set available at day t by Ω_t^{LF}. In the original GARCH(1,1) model of Engle (1982) and Bollerslev (1986), the conditional variance on day t+1, denoted as $\mathrm{var}(r_{t+1}|\Omega_t^{LF})$, can be written as (4.6). To extend the GARCH(1,1) specification, Shephard and Sheppard (2010) add features of daily realized measure to form the class of High-Frequency-Based Volatility (HEAVY) models. The HEAVY specification is made up of two pillars: the HEAVY-r model to estimate close-to-close conditional variance $\mathrm{var}(r_{t+1}|\Omega_t^{HF})$ and the HEAVY-RM model to estimate the conditional expectation of open-to-close variation $E(RM_{t+1}|\Omega_t^{HF})$

HEAVY-r: $\mathrm{var}(r_{t+1}|\Omega_t^{HF}) = h_{t+1} = \omega + \alpha RV_t + \beta h_t$ (4.13)

HEAVY-RM: $E(RM_{t+1}|\Omega_t^{HF}) = \mu_{t+1} = \omega_R + \alpha_R RV_t + \beta_R \mu_t$ (4.14)

The HEAVY-r equation can be estimated using standard quasi-likelihood:

$$\log Q_1(\omega,\Psi) = \sum_{t=2}^{T} l^r{}_t \text{ in which } l^r{}_t = -\frac{1}{2}(\log h_t + r^2{}_t/h_t) \quad (4.15)$$

and $\Psi = (\alpha, \beta)'$ and the initial value $h_1 = T^{-1/2} \sum_{t=1}^{|T|^{-1/2}} r^2{}_t$

Similarly, the HEAVY-RM can be solved using a normal quasi-likelihood:

$$\log Q_2(\omega_R, \Psi_R) = \sum_{t=2}^{T} l^{RM}{}_t \text{ in which } l^{RM}{}_t = -\frac{1}{2}(\log \mu_t + RM_t/\mu_t) \quad (4.16)$$

and $\Psi_R = (\alpha_R, \beta_R)'$ where the initial value $\mu_1 = T^{-1/2} \sum_{t=1}^{|T|^{-1/2}} RM_t$

4.2.1.4 Methods for computing quantile forecasts

A quantile forecast is an output of two components: (i) the model used to estimate volatility and (ii) the method of generating quantile from the volatility forecast. In the first step, these models in previous section deliver the h-day forecasts of daily volatility. The next step is to compute the quantile forecasts from the estimated volatility $\hat{\sigma}_{t+h}$

The simplest method to generate conditional quantile $q_\alpha(r_{t+h}|\Omega_t)$ is to assume the distribution for the daily returns with the distribution function:

$$F_t(y) = \Pr(r_{t+h} \leq r|\ \Omega_t) \tag{4.17}$$

Assuming that the daily returns are unforecastable, the process of returns are described as: $r_{t+h} = \varepsilon_{t+h}$, where $\varepsilon_{t+h} = \hat{\sigma}_{t+h}\, z_{t+h}$ and z_{t+h} is i.i.d. Thus, the forecasted α-quantile is defined as:

$$q_\alpha(r_{t+h}|\Omega_t) = \hat{\sigma}_{t+h}\, F^{-1}{}_t(\alpha) \tag{4.18}$$

Another method to obtain the conditional quantile $q_\alpha(r_{t+h}|\Omega_t)$ from the estimated volatility $\hat{\sigma}_{t+h}$ is to simulate the density of returns on day t+h, with the dynamic of return process is described as below:

$$r_{t+h} = \varepsilon_{t+h} \text{ where } \varepsilon_{t+h} = \sigma_{t+h}\, z_{t+h} \text{ and } z_{t+h} \sim D(0,1) \tag{4.19}$$

where D(0,1) is the given choice of distributional assumption. The intuition here is that the simulated real-world density of r_{t+h} is the combination of the forecasted conditional volatility on day t+h, $\hat{\sigma}_{t+h}$, plus the simulated values of z_{t+h}. This is achieved by making a large number of independent draws from D(0,1). In line with previous chapters, we used the Gaussian and Student t distribution for D(0,1). In order to detect whether the forecast improvements emanate from the choice of distribution rather than the choice of model, we use cross comparison. Specifically, we compare the performance of the GARCH(1,1) model with alternative distributional assumptions to see which choice of distribution provide better forecast. We also compare performance of different forecasting models but with the same Gaussian assumption to detect the better model specification.

To obtain the density forecast of r_{t+h}, in the first step we make 100,000 random independent draws from D(0,1) and therefore simulate 100,000 random values of z_{t+1}. Given the set of parameters $\widehat{\Psi}$ and the conditional volatility $\hat{\sigma}_{t+1}$ estimated from the forecasting models, we obtain 100,000 simulated values of r_{t+1} by repeating the process: $r_{t+1} = z_{t+1}\, \hat{\sigma}_{t+1}$. In the second step, each simulated value r_{t+1} is then used to estimate $\hat{\sigma}_{t+2}$ using the set of parameters $\widehat{\Psi}$ and $\hat{\sigma}_{t+1}$. With a random drawn from D(0,1) and the estimated $\hat{\sigma}_{t+2}$, we obtain the simulated value of r_{t+2}. We continue this procedure to generate 100,000 simulated values of r_{t+n}. This defines the density forecast of r_{t+h}, denoted as $D(r_{t+h}|\ \Omega_t)$. Then, the quantile forecast $q_\alpha(r_{t+h}|\Omega_t)$ is simply acquired by taking the quantile of the density forecast $D(r_{t+h}|\ \Omega_t)$.

4.2.2 The direct quantile forecasting model

In the direct approach, the quantile forecasts are obtained directly through the quantile regression framework of Koenker and Basset (1978), in which the time series of specific quantiles q(α) is directly modelled using any information set claimed to be relevant. The basic idea of the quantile regression is that the α-quantiles $q_\alpha(r_{t+1}|\Omega_t)$ is modelled as a function of variables available at information set Ω_t:

$$q_\alpha(r_{t+1}|\Omega_t) = g_\alpha(X_t, \beta_\alpha) \tag{4.20}$$

where function $g_\alpha(X_t, \beta_\alpha)$ and parameter vector β_α are explicitly dependent on quantile level α (Koenker and Basset, 1978).

One of the most widely-used quantile regression-based models in finance is the Conditional Autoregressive Value at Risk (CAViaR) model proposed by Engle and Manganelli (2004). In particular, CAViaR specification is a dynamic non-linear quantile regression approach for quantiles of daily returns of an asset, in which the quantile forecast $q_\alpha(r_{t+1}|\Omega_t)$ is the function of past daily return and past quantile of daily returns.

$$q_\alpha(r_{t+1}|\Omega_t) = g_\alpha(X_t, \beta_\alpha) \text{ where } X_t = \{r_t, q_\alpha(r_t)\} \tag{4.21}$$

The CAViaR-based approach has gained popularity in the literature of financial econometrics and received great success in forecasting quantile of return distributions (Gourieroux and Jasiak, 2008; Ma and Pohlman, 2009; Huang et al., 2009; Taylor, 2016). However, the estimation of CAViaR models with rolling window technique require high level of processing resource that makes it computationally expensive. Besides, CAViaR models do not perform very well in comparison with the Realized volatility-based linear quantile regression models of Zikes and Barunik (2016). For these reasons, we do not include the CAViaR-based models in this chapter.

Zikes and Barunik (2016) suggest quantile regression models as a linear function of past-quadratic variation and exogenous variables. Taking advantage of high-frequency data, in their research the α-quantile $q_\alpha(r_{t+1}|\Omega_t)$ is expressed as a linear function of past realized measures:

$$\text{LQR1:} \quad q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \beta_{\mathbf{v}}(\alpha)' RV^{1/2}{}_t + \beta_z(\alpha)' z_t \tag{4.22}$$

LQR2: $q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \beta_1(\alpha)' \, IV^{1/2}{}_t + \beta_2(\alpha)' \, JV^{1/2}{}_t$

$$+ \beta_3(\alpha)' \, VIX^{1/2}{}_t + \beta_z(\alpha)' z_t \quad (4.23)$$

in which RV_t, IV_t, JV_t, VIX_t are Realized variance, Integrated variance, Jump variation and Implied volatility respectively. Applying to the 5-minute intraday data of S&P500 and WTI Crude Oil Futures, they show evidence that there is no remarkable difference between the performance of the LQR1 and LQR2 in out-of-sample forecast. In addition, it is also worth to note that the Linear quantile regression models of Zikes and Barunik (2016) outperform the CAViaR model of Engle and Manganelli (2004) both in left and right tails of the return distributions in one-step-ahead forecast, although at longer horizons, the differences are not significant. Due to its performance and simplicity, we select the LQR1 specification (hereafter, denoted as LQR-RV) to use in this research. Specifically, the LQR-RV specification is presented as:

$$q_\alpha(r_{t+1}|\Omega_t) = \beta_0(\alpha) + \beta_{\mathbf{v}}(\alpha)' \, RV^{1/2}{}_t + \beta_z(\alpha)' z_t \quad (4.24)$$

As Koenker and Bassett (1978) note, the parameters of linear quantile regression can be obtained by minimizing the objective function:

$$QR_T(\beta(\alpha)) = \frac{1}{T}\sum_{t=1}^{T} \rho_\alpha(r_{t+1} - \beta_0(\alpha) - \beta_{\mathbf{v}}(\alpha)' \, RV_{t,M} - \beta_z(\alpha)' \, z_t) \quad (4.25)$$

where $\rho_\alpha(\vartheta) = (\alpha - 1\{\vartheta < 0\})\vartheta$ and $\beta(\alpha) = (\beta_0, \beta_{\mathbf{v}}(\alpha)', \beta_z(\alpha)')'$

4.2.3 The quantile combination

The literature on the combination of conditional quantile forecasts is not as rich as of the conditional mean. In prior studies, the approaches to combining quantile forecasts are restricted as the linear combination function of two individual quantile forecasts. There are two main ways to estimate the weights of individual quantile forecasts in the linear combination. The first method, considered by Giacomini and Komunjer (2005) and Halbleib and Pohlmeier (2012), is based on the Conditional coverage hypotheses of Christoffersen (1998). Specifically, the weights of the individual forecasts are optimized to produce forecast combination which optimally satisfies both UC and IND properties e.g. the sequence of hit process must be independently distributed and the actual coverage of the hits must be equal to the theoretical one. The second method to combine quantile forecasts is the conditional quantile optimization method (CQOM). This method, proposed by Halbleib and Pohlmeier (2012), relies on the quantile regression framework of Koenker and Basset (1978) to select the optimal combination weights of individual VaR forecasts. Taking advantage of the quantile regression approach, the CQOM does not require any assumptions for the return distribution. Thus, it allows for a flexible approach of VaR forecast combinations to minimize the distance between the population quantiles and the combined VaR. Furthermore, this second approach is shown to provide better forecast combinations than the traditional conditional coverage-based optimization approach (Halbleib and Pohlmeier, 2012).

This chapter employs the CQOM of Halbleib and Pohlmeier (2012) to combine quantile forecasts. Suppose we have two individual quantile forecasts available at time t, denoted as $q_{\alpha,1}(r_{t+1}|\Omega_t)$ and $q_{\alpha,2}(r_{t+1}|\Omega_t)$. In general form, the CQOM defines the combined α-quantile, denoted as $q_{\alpha,12}(r_{t+1}|\Omega_t)$, as the linear function of a pair of stand-alone quantile forecasts:

$$q_{\alpha,12}(r_{t+1}|\Omega_t) = \lambda_0 + \lambda_1\, q_{\alpha,1}(r_{t+1}|\Omega_t) + \lambda_2\, q_{\alpha,2}(r_{t+1}|\Omega_t) \qquad (4.26)$$

Prior studies on optimal forecast combinations commonly set up the constrain to the sum of weights of individual forecasts to unity (Granger and Ramanathan, 1984; Halbleib and Pohlmeier, 2012). However, we do not place any boundary constraints on the loadings to retain the flexibility of the model. Besides, it is possible to extend the quantile combination of more than two individual forecasts:

$$q_{\alpha,N}(r_{t+1}|\Omega_t) = \lambda_0 + \lambda_1\, q_{\alpha,1}(r_{t+1}|\Omega_t) + \lambda_2\, q_{\alpha,2}(r_{t+1}|\Omega_t) + \ldots + \lambda_n\, q_{\alpha,n}(r_{t+1}|\Omega_t)$$
$$= q_{\alpha,N}(\lambda_N) \qquad (4.27)$$

where $q_{\alpha,N} = \{1, q_{\alpha,1}(r_{t+1}|\Omega_t), q_{\alpha,2}(r_{t+1}|\Omega_t), \ldots, q_{\alpha,n}(r_{t+1}|\Omega_t)\}$ is the vector of individual quantile forecasts and $\lambda_N = \{\lambda_0, \lambda_1, \ldots, \lambda_n\}'$ is the vector of optimal loadings of and. Based on the quantile regression framework, the vector of optimal loadings is given by solving this minimization:

$$\hat{\lambda}_t = \arg\min_{\lambda_N} \{ \textstyle\sum_{r_{t+1} \geq q_{\alpha,N}(r_{t+1}|\Omega_t)} \alpha |r_{t+1} - q_{\alpha,N}(\lambda_N)|$$
$$+ \textstyle\sum_{r_{t+1} < q_{\alpha,N}(r_{t+1}|\Omega_t)} (1-\alpha)|r_{t+1} - q_{\alpha,N}(\lambda_N)| \} \qquad (4.28)$$

where α is the quantile level, r_{t+1} is the realized returns on day t+1.

4.3 Methods of evaluation

We evaluate the accuracy of quantile forecasts in two steps. First, we assess the absolute performance of alternative forecasting models by taking the hit test and examining the behavior of the hit process at every quantile level. Second, we quantify their relative performance by using the Tick loss function to find which model performs best at each quantile level.

4.3.1 Evaluation of absolute performance

In the first stage, we evaluate the absolute performance of alternative models by calling the Dynamic Quantile (DQ) test of Engle and Manganelli (2004), a statistical test based on the properties of violations at a given quantile.

Let r_{t+1} be the realized cumulative return from day t to day t+1, and $q^i_\alpha(r_{t+1}|\Omega_t)$ is the α^{th} quantile forecast conditional on the information set available at time t. Statistically, the probability of a hit at given quantile is defined as:

$$\Pr[r_{t+h} < q^i_\alpha(r_{t+h}|\Omega_t)] = \alpha \tag{4.29}$$

Let us denote $I_t(\alpha)$ the hit process associated with the ex-post observations at α-quantile forecast at time t:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_{t+1} < q^i_\alpha(r_{t+1}|\Omega_t) \\ 0 & \text{otherwise} \end{cases} \tag{4.30}$$

Theoretically, the quantile forecasts are valid if and only if the hit process $I_t(\alpha)$ satisfies two properties: The UC and the IND hypothesis (Christoffersen, 1998). The UC property, firstly proposed by Kupiec (1995), states that the probability of a realized return hitting the α quantile forecast must be equal to α:

$$\Pr[I_t(\alpha)=1] = E[I_t(\alpha)] = \alpha. \tag{4.31}$$

Furthermore, to satisfy the IND property, the hit process $I_t(\alpha)$ must be independently distributed. When both UC and IND properties are simultaneously satisfied, quantile forecasts are said to be valid with CC property, which is the basis of most of the backtests in the literature (Christoffersen, 1998; Engle and Manganelli, 2004; Berkowitz et al., 2011; Colletaz et al., 2013).

To test the absolute performance of alternative forecast models, we investigate seven quantiles: 0.01; 0.05; 0.1; 0.5; 0.9; 0.95; 0.99 and perform the hit test at every quantile level. In particular, following the literature of Value at Risk (VaR), we consider each quantile as VaR(α), which α = {0.01; 0.05; 0.1; 0.5; 0.9; 0.95; 0.99}, then examine the behaviour of the hit process at each quantile.

Firstly, let $Hit_t(\alpha) = I_t(\alpha) - \alpha$ be the demeaned process of hit function. Thus $Hit_t(\alpha)$ will take the value 1- α if the realized return at day t is less than the forecasted value of quantile α:

$$Hit_{t+1}(\alpha) = \begin{cases} 1-\alpha & \text{if } r_{t+1} < q_\alpha(r_{t+1}|\Omega_t) \\ -\alpha & \text{otherwise} \end{cases} \tag{4.32}$$

If the quantile α is correctly specified, the conditional expectation of $Hit_t(\alpha)$ given the set of information at time t-1 must be equal to zero. In addition, under the CC assumption, variable $Hit_t(\alpha)$ must be uncorrelated with its own lagged values and lagged quantile $q_\alpha(r_{t-1}|\Omega_{t-2})$. The CC hypothesis can be examined in the following linear regression model:

$$Hit_{t+1}(\alpha) = \beta_0 + \beta_1\, Hit_t + \beta_2\, q_\alpha(r_t|\Omega_{t-1}) + u_{t+1} \tag{4.33}$$

Hence testing the null hypothesis of CC assumption is equivalent to testing the joint null hypothesis that the coefficients β_1, β_2 and the intercept β_o in the regression model are all equal to zero:

$$H_0: \beta_0 = \beta_1 = \beta_2 = 0 \tag{4.34}$$

Indeed, at time t the hit process is not correlated with past violations and past quantile if $\beta_1 = \beta_2 = 0$ (implied by the independence hypothesis), while the null constant β_0 will fulfil the UC hypothesis. If we denote by $\Psi = (\beta_0, \beta_1, \beta_2)'$ the parameter vector and by X the matrix of explanatory variables in the regression model, the test statistics DQ_{CC} to test the null is defined as:

$$\mathrm{DQ}_{CC} = \frac{\hat{\Psi}' Z' Z \hat{\Psi}}{\alpha(1-\alpha)} \xrightarrow{T \to \infty} \chi^2(3) \quad (4.35)$$

It is important to note that the DQ test can only be suitable for one-step-ahead forecasts. This is because the sequence of hit process $\{Hit_{t|t+h}\}$ is h-dependent, which breaks the assumptions underlying the likelihood ratio test of the DQ test, in multi-step-ahead forecasts. The literature does not seem to provide alternative, robust test for the multistep-ahead forecast of conditional quantiles.

4.3.2 Evaluation of comparative performance

Christoffersen (1998) and Diebold et al. (1998) early develop statistical tests to evaluate interval and density forecasts with null hypothesis that the forecasting model is specified correctly. Nevertheless, empirical models tend to be incorrectly specified, hence to a forecaster, their relative performance might be more interested rather than their absolute performance.

There are several methods to evaluate the performance of interval and density forecast, in which the main difference among those approaches is represented by the loss function or score function assumed in the forecast comparison. Among the score function approaches, the Logarithmic Score (LS) of Gneiting and Raftery (2007) is the widely used one in the literature of forecasting. In computation, LS of model i is defined as $LS^i_{t+h} = \ln f^i_{t+h}(r^i_{t+h} | \Omega_t)$ in which $f^i_{t+h}(.| \Omega_t)$ is the density forecast of the h-day return conditional on an information set available at day t and r^i_{t+h} is the realized cumulative return from day t to day t+h. The score function LS^i_{t+h} evaluates the density forecast at time t in comparison with the realized cumulative return at day t+h. To relatively compare the performance between two models, suppose model A and model B, we take the difference in their log-score function:

$$\Delta LS^{A,B}_{t+h} = LS^{A}_{t+h} - LS^{B}_{t+h} \tag{4.36}$$

ΔLS is positive orientation. Thus, positive value of $\Delta LS^{A,B}_{t+h}$ means that model A is superior to model B and vice versa. Example of the application of this approach included Sheppard and Shephard (2010) and Maheu and McCurdy (2011) amongst others. Nevertheless, the use of LS has been criticized since it only captures the overall examination of the performance of a model but does not provide the evaluation of the specific intervals of the return distribution. Indeed, risk managers and forecasters might be interested in assessing models based on specific quantiles of the distribution e.g. VaR(1%). For this reason, the methods of evaluation based on the properties of the violation process at given quantile, assumed that the model is correctly specified, are of interest. Christoffersen (1998) propose an assessment of interval forecast using the unconditional and independence properties of the violation series, which is then considered in the studies of Kuester et al. (2006) and Brownless and Gallo (2010). Another approach to evaluate and compare quantile forecasts is the tick loss function (see Giacomini and Konmunjer, 2005; Gneiting and Raftery, 2007; Clements et al., 2008). The tick loss function is defined as:

$$L_{\alpha,i} = E((\alpha\text{-}1\{e^{i}_{t+h} < 0\})e^{i}_{t+h} \tag{4.37}$$

where $e^{i}_{t+h} = r_{t+h} - q^{i}_{\alpha}(r_{t+h}|\Omega_t)$ and $q^{i}_{\alpha}(r_{t+h}|\Omega_t)$ is the α^{th} quantile forecast of model i. The tick loss function penalizes quantile violations more strictly with the increase in the magnitude of the violation. Therefore, the tick loss function is negative orientation, in which we prefer model that has lower loss function value.

This chapter evaluates the relative performance of forecast models using Quantile Score (QS). Specifically, the QS incorporates the idea of the asymmetric absolute loss function obtained in quantile regression estimation to the context of out-of-sample assessment (Giacomini and Komunjer, 2005; Hua and Mazan, 2013[29]). To illustrate, let $q^{i}_{t+h}(\alpha)$ denote conditional quantile forecasts of model *i* based on information set at time t and r^{i}_{t+h} denote the h-period cumulative return, the QS at α-quantile is defined as:

$$QS^{i}_{t+h}(\alpha) = [\, r^{i}_{t+h} - q^{i}_{t+h}(\alpha)]\, [I(\, r^{i}_{t+h} \leq q^{i}_{t+h}(\alpha)) - \alpha] \tag{4.38}$$

in which I(.) denotes the hit indicator which is equal to 1 if the argument is true and takes the value of 0 otherwise. Compared to the equation (4.37), it is clear that the QS function is the negative form of the tick loss function. Therefore, the QS is positively oriented, thus between two competing models, we prefer the one with higher QS.

To evaluate the performance of model of the certain area of the return distribution rather than single quantile, Gneiting and Ranjan (2011) propose the statistical evaluation which involves the integration of the QS across different quantile levels. The Weight Quantile Score (WQS) of model i, denoted as WQS^{i}, is defined as:

$$WQS^{i}_{t+h}(\alpha) = \int_0^1 QS^{i}_{t+h}(\alpha)\, w(\alpha) d\alpha \tag{4.39}$$

in which w(α) presents the weight function corresponding to each quantile. The choice of w(α) depends on the purpose of evaluation. To comprehensively evaluate the accuracy of quantile forecasts, we follow Hua and Manzan (2013) to select five

[29] In the paper of Hua and Mazan (2013), the QS function was presented incorrectly. Specifically, their QS equation is indeed the Tick loss function, which is shown to be negatively oriented.

types of the weight function. (i) $w(\alpha) = 1$ that equally weights all quantiles to give the entire evaluation of the forecasted distribution. (ii) $w(\alpha) = \alpha(1-\alpha)$ that provides higher weights to the central quantile and lower weights to the tails. (iii) $w(\alpha) = (2\alpha - 1)^2$ to focus on the tails of the return distribution. (iv) $w(\alpha) = (1 - \alpha)^2$ which specifically focuses on the left tail and (v) $w(\alpha) = \alpha^2$ that gives higher weights to the right tail of the return distribution.

4.4 Data specifications

Our dataset includes daily settlement prices of WTI Crude Oil Futures traded at the New York Mercantile Exchange (NYMEX). Recall that 5-minute sampling frequency balances the demand for finely sampled observations and the robustness to the contaminations of market microstructure. Besides, the 5-minute realized volatility is shown to be hardly beaten by any other measures (Liu et al., 2015). For these reasons, we select the 5-minute sampling frequency to get 999,855 observations of five-minute settlement prices of WTI Crude Oil Futures, starting from 3rd Jan 1995 and ending on 30th June 2016. The data series is obtained from Tick Data.

The close-to-close daily return of day t, denoted by r_t, is computed as equation (4.4), whereas the 5-minute intraday return at point i on day t is computed as equation (4.3). The realized volatility on day t can be simply obtained by taking the sum of squared intraday returns within this day following the equation (4.2). We report the summary statistics of the daily returns, daily squared returns and the realized volatility in Table 4.1 and their time-series are presented in Figure 4.1.

It can be seen that the WTI Crude Oil Futures experience a positive trend in daily returns from Jan 1995 to June 2016. The distribution of r_t is negatively skewed and leptokurtic, implying the fat-tailed characteristics of the return distribution. Comparing two measures of variation, we find that the daily squared returns are more volatile and noisy than the realized measure. This evidence is also consistent with the literature of variance measures (Andersen and Bollerslev, 1997). Besides, the Jacque-Bera test for normality is also rejected at all series.

Table 4.1: Summary statistics of daily returns and RV of WTI Crude Oil Futures

Notes: Table 4.1 reports the summary statistics of daily returns, squared returns and realized volatility and the squared root of realized volatility.

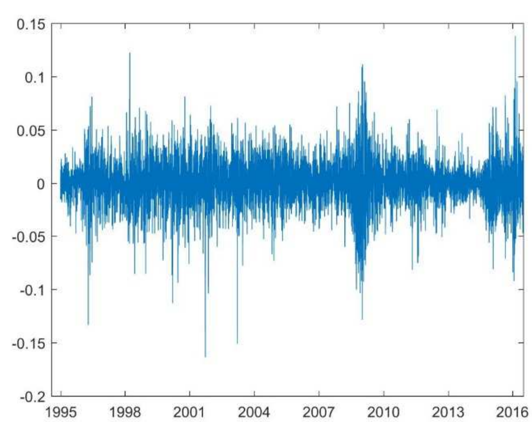
	Mean	Stdev	Skewness	Kurtosis	Min	Max	JB stats	p-value
r_t	1.649e-04	0.0213	-0.2295	6.8715	-0.1636	0.1382	3877	0.001**
${r^2}_t$	4.557e-04	0.0011	8.9314	137.001	0	0.0268	4662539	0.001**
RV_t	4.841e-04	9.082e-04	18.22	543.94	1.498e-06	0.0351	74992625	0.001**
${RV^{1/2}}_t$	0.0193	0.0107	3.1005	29.5525	0.0012	0.1874	189681	0.001**

**: significant at 95% level of confidence

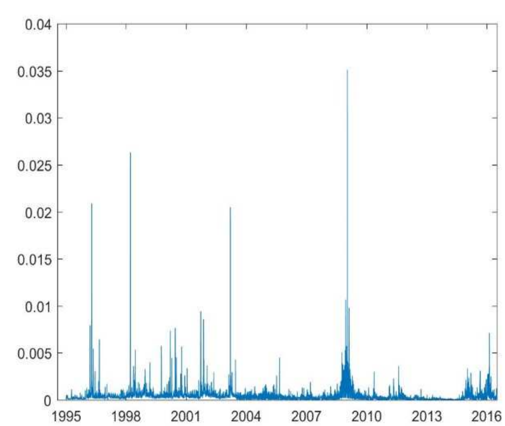
Figure 4.1: Time series of Daily returns and Realized Volatility of WTI Crude Oil Futures

Notes: Figure 4.1 plots the time series of daily returns and daily RV of WTI Futures from January 1st, 1995 to June 30th, 2016.

Daily returns



Daily realized volatilitiy



4.5 Evaluation of Quantile forecasts

We evaluate the predictive power of alternative models based on their out-of-sample performance. The in-sample period is from 1st January 1995 and lasts to 31st December 2004, while we start the out-of-sample forecasting experiment on 3rd January 2005 and end on 30th June 2016, including 3,513 days. We continue using the rolling window technique with the window size of 10 years of historical data (or 2,610 trading days, equivalently). In particular, the forecast on 3rd Jan 2005 will be obtained via an econometric model with parameters estimated from the historical data set I_{t-1} starting from 1st Jan 1995 and ending on 31st Dec 2004[30]. Then the window will move one-step ahead to produce the forecast on date 4th Jan 2005. This procedure is repeated continuously to get the out-of-sample forecasts. We consider the forecast horizons of 1-day, 5-day (or one-week-ahead forecast) and 22-day (or one-month-ahead forecast).

4.5.1 Model estimation

The first step of our empirical analysis is to estimate parameters of the conditional volatility models and coefficients of the LQR-RV model. Regarding to the choice of quantile levels, we focus on forecasting the 1%, 5%, 10%, 50%, 90%, 95% and 99% quantiles as these are the most interesting thresholds from economic point of view, especially in risk management. In addition, we employ the realized standard

---

[30] The last trading day before 3rd Jan 2005

deviation rather than variance, hence we take the square root of the realized variance RV_t discussed in the previous session. We estimate the parameters of GARCH(1,1) models by using Maximum Likelihood Estimation, while the HEAVY parameters are obtained by (4.15) and (4.16). Besides, we use OLS to estimate the coefficients β(.) of HAR-RV model, while the estimation of LQR-RV coefficients in (4.24) is executed by using the interior point method of Portnoy and Koenker (1997) to solve the optimization function (4.25). Due to limit in space, we only present the estimation results of the first rolling window, with the in-sample period starting from 1st Jan 1996 to 31st Dec 2005. The estimation results of the conditional volatility models are presented in Table 4.2, while Table 4.3 reports the estimated coefficients of the LQR-RV models.

Table 4.2: Parameter estimation of Conditional volatility models

Notes: Table 4.2 reports the estimated parameters of the GARCH-type models with the sample period from 1st Jan 1995 until 31st Dec 2005. The associated t-statistics are shown in parenthesis.

GARCHn		
ω	α	β
0.00001	0.0494	0.9477
(1.908)	(9.9697)	(171.92)

HEAVY					
HEAVY-r			HEAVY-RM		
ω	α	β	ω_R	α_R	β_R
0.0000	0.0566	0.9369	0.0000	0.2576	0.7187

GARCHt			
ω	α	β	dof
0.00004	0.0377	0.9534	5.902
(2.101)	(95.948)	(4.8486)	(8.877)

HAR-RV			
ω	β_d	β_w	β_m
0.0001	-0.2346	1.236	-0.0398
(0.7134)	(-12.163)	(30.219)	(-0.696)

Table 4.3: Coefficient estimation of the LQR-RV model

Notes: Table 4.3 presents the estimated coefficients of the LQR-RV specification as described in (4.24) at seven quantiles. The coefficients are estimated using the interior point method (Portnoy and Koenker, 1997) to minimize the objective function (4.25). The associated t-statistics are shown in parenthesis.

α	0.01	0.05	0.1	0.5	0.9	0.95	0.99
Constant	-0.0380	-0.0254	-0.0180	0.0015	0.0200	0.0253	0.0451
	(-4.705)	(-7.009)	(-6.528)	(1.117)	(8.978)	(8.498)	(7.009)

$RV^{1/2}_t$	-1.0538	-0.5150	-0.4034	-0.0436	0.3311	0.5013	0.3727
	(-1.988)	(-2.445)	(-2.820)	(-0.623)	(3.002)	(3.265)	(1.090)

Table 4.2 shows that the lagged conditional volatility parameters β of GARCH(1,1) models are close to one and statistically significant, indicating that the forecasted conditional volatility is significantly determined by its lagged value. In HAR-RV model, the estimated values of β_d and β_w are statistically significant, while the constant and β_m are not. This implies that the future volatility is mostly driven by the past-week and past-day volatilities, but has little memory of the past month volatility. As one of GARCH-type models, the estimation of HEAVY parameters report future volatility has a very considerable memory of the past, which is consistent to the literature. Indeed, Shephard and Sheppard (2010) find that the momentum parameter β in the HEAVY-r model typically ranges from 0.6 to 0.75 for stock indexes, but there are exceptions with higher momentum, e.g. exchange rate. In case of LQR-RV model, we find lagged realized volatility is statistically significant across different quantiles. The estimated coefficients also have the expected sign: the left-tail quantiles alter negatively with realized volatility, while the right-tail behaves positively.

4.5.2 Evaluation of absolute performance

We evaluate the forecasting power of alternative models based on the out-of-sample performance at each quantile level: 1%, 5%, 10%, 50%, 90%, 95% and 99%. The reason we include the very low quantiles to the test (1%-quantile and 99%-quantile) is that we stand on the risk management perspective, which concern most about the left tail of the return distribution. We adopt the rolling window approach to forecast the one-day-ahead quantile and keep the window size fixed at 10-year historical data. The out-of-sample period is from 3rd Jan 2006 to 30th June 2016 including 3,513 daily observations. Recall that our evaluation is based on the properties of the hit process at specific quantile. In particular, we test whether the hit process satisfy both UC and IND hypotheses by calling the DQ test. It is also important to recall that the DQ approach is only valid for one-step-ahead quantile forecasts. The test results are reported in Table 4.4.

Table 4.4: Absolute performance of alternative forecasting models

Notes: Table 4.4 presents the absolute performance of alternative models. At each model, $\hat{\alpha}$ presents the actual coverage rate, which is computed as the number of hits over total number of observations. The DQ stats report the test statistics of the DQ test, while the third row presents its corresponding p-value.

	0.01	0.05	0.1	0.5	0.9	0.95	0.99
GARCHn							
$\hat{\alpha}$	0.0134	0.0421	0.0831	0.4944	0.9223	0.9556	0.9872
DQ stats	6.4681	6.8331	11.4218	3.9999	24.8381	4.4208	3.6431
p-value	0.0909	0.0774	0.0097	0.2615	0.0000	0.2195	0.3027
GARCHt							
$\hat{\alpha}$	0.0137	0.0444	0.0834	0.4942	0.9229	0.9553	0.9875
DQ stats	8.1104	3.2296	11.1543	5.7846	25.6538	4.0705	4.9679
p-value	0.0438	0.3576	0.0109	0.1226	0.0000	0.2539	0.1742
HEAVY							
$\hat{\alpha}$	0.0137	0.0461	0.0931	0.5050	0.9197	0.9562	0.9878
DQ stats	6.0855	2.3083	4.7381	3.6228	19.9759	5.5078	2.5525
p-value	0.1075	0.5109	0.1920	0.3052	0.0002	0.1382	0.4659
HAR-RV							
$\hat{\alpha}$	0.0148	0.0487	0.0939	0.4959	0.9061	0.9490	0.9858
DQ stats	13.3040	2.8102	2.3604	2.8926	3.3809	2.6777	6.5731
p-value	0.0040	0.4218	0.5010	0.4085	0.3365	0.4440	0.0868
LQR-RV							
$\hat{\alpha}$	0.0088	0.0430	0.0882	0.5107	0.9172	0.9593	0.9895
DQ stats	0.7302	3.6361	5.3365	4.5879	12.8378	8.9301	1.1304
p-value	0.6941	0.1623	0.0694	0.1009	0.0016	0.0115	0.5682

Across models and quantiles, we report the out-of-sample actual coverage rate $\hat{\alpha}$, the likelihood ratio test statistics and the p-value of the DQ test for the joint hypothesis of UC and IND properties. If a model passes the DQ test at a given quantile, it means that the sequence of hits at this quantile is not only quantitively adequate, but also independent from each other. Based on the DQ test, there are three main points emerged. First, we witness that the all models perform reasonably well across quantiles with the real coverage rate $\hat{\alpha}$ are very close to the nominal levels. There are some commonalities in the performance of the models. It can be seen that four over five models remarkably underestimate the 90%-quantile and fail the DQ test, while only HAR-RV is successful in capturing this quantile. Second, we find the asymmetries in the performance for the left tail versus the right tail across models. At 1-% quantile, four over five models seem to underestimate the left tail and produce high coverage rate, except LQR-RV which slightly overestimate left tail. The DQ test statistically rejects the validity of two over five forecasts at the 1%-quantile. At the right tail, we find that all models overestimate the 99%-quantile of the return distribution. However, the overestimation at the right tail is not significant, which has been statistically confirmed by the results of the DQ test.

Third, it can be seen that the RV-based models perform slightly better than the GARCH(1,1) models. Indeed, both GARCHn and GARCHt models perform fairly well at the middle and end-tails of the return distribution, but fails to accurately forecast the 10% and 90% quantiles. In case of HAR-RV, HEAVY and LQR-RV, there is only one misspecification of each model over seven quantiles. While the long-memory HAR-RV model seems to underestimate the 1%-quantile of the return distribution, the HEAVY and LQR-RV models fail to properly capture the 90%-

quantile. Except these cases, the high frequency-based models perform accurately in forecasting quantiles of the return distribution.

4.5.3 Evaluation of comparative performance

This section aims to relatively evaluate the performance of the alternative forecasting models. To compare their performance, we select the GARCHn as the benchmark model, as it is the most popular GARCH-type specification and is also hard to beat by more sophisticated models (Hansen and Lunde, 2005).

As discussed in the previous section, we employ the QS approach to compare the out-of-sample predictive power of five models at seven quantile levels. Recall that the QS is positively oriented, which means that between two models, the one with higher QS is preferable. To compare the performance of model i with the benchmark model at α-quantile, we take the difference in their QS:

$$\Delta QS^{\alpha}{}_{i} = QS^{\alpha}{}_{i} - QS^{\alpha}{}_{GARCHn} \tag{4.40}$$

The positive value of $\Delta QS^{\alpha}{}_{i}$ means that model i outperforms the benchmark GARCHn model and vice versa. Additionally, to provide a comprehensive evaluation the quantile forecasts, we employ the WQS which specifically focus on different characteristics of return distribution: the entire distribution, the middle, two tails, the left and the right tails of the return distribution. Similar to ΔQS, we compute the ΔWQS to compare model performance to the benchmark model

$$\Delta WQS_{i} = WQS_{i} - WQS_{GARCHn} \tag{4.41}$$

The positive ΔWQS_i indicates that model i is superior to the GARCHn in providing good quantile forecasts. The ΔWQS_i is also comparative, which means that model with higher ΔWQS_i has the better performance. The $\Delta QS^{\alpha}{}_i$ and ΔWQS_i results are reported in Panel A and Panel B of Table 4.5.

We document the diversification in the performance of alternative forecasting models across quantile levels and forecast horizons. There are some models perform well, but the others show the poor predictive power in comparison with the benchmark model. We find that the differences in model performance mostly come from the tails of the return distribution, whereas at the centre, they are not significant. We will discuss the performance of each forecasting model in the following.

Table 4.5 shows the weak predictive power of the GARCHt model in comparison to the GARCHn. The poor performance of GARCHt at one-day-ahead forecasts comes from the left tail and the center of return distribution, while at the right tail, it performs slightly better than the benchmark model. At five-day-ahead quantile forecasts, the GARCHt even performs worse, as it provides less accurate forecasts at all quantiles compared to the GARCHn. The predictive power of GARCHt only excels at 22-day forecast horizon, when it dominates the benchmark at seven quantiles of the return distribution. Therefore, we argue that the use of Gaussian distributional assumption in GARCH(1,1) specification provides better quantile forecasts than the use of Student t distribution.

The worst performance in the comparative evaluation belongs to the LQR-RV model. Indeed, the direct quantile forecasting model is inferior to the GARCHn across quantiles and forecast horizons. It is also noticeable that the ΔQS and ΔWQS between the LQR-RV and GARCHn are remarkably higher than any other models, especially at two tails of the return distribution. Thus, the LQR-RV model might not be the good choice of model in application to risk management.

Contrary to the direct quantile forecasting model, the indirect forecasting models with the use of RV exhibit the robust and accurate quantile forecasts. Indeed, the ΔQS and ΔWQS show that the HAR-RV and HEAVY models are hardly beaten by the GARCHn at any quantiles. We find that the superiority of the RV-based indirect quantile forecasting models is stable across the tails and center of return distribution, at different forecast horizons. While the HEAVY is the most accurate model in one-day-head quantile forecasts, the HAR-RV excels its predictive power in longer forecasts horizons e.g. 5-day-ahead and 22-day-ahead. The superiority of HAR-RV model in quantile forecasts with long forecast horizon can be attributed to the use of long memory to capture the long-lag effects, which was previously confirmed in prior studies (see Clements et al., 2008; Hua and Mazan, 2013; Huang et al., 2016).

Table 4.5 shows that the good performance of alternative models is primarily due to the ability of the alternative models to capture the observations in the tails of the forecasted distribution. Indeed, the magnitude of ΔWDS in the tails of the distribution is much larger than in the centre of the distribution. Furthermore, there are asymmetries in the left and right tail of the forecasted density. We find that the HEAVY and HAR-RV models perform comparatively better in the left tail than in the right tail. These models dominate others on forecasting accurate quantiles. Therefore, it implies that the HAR-RV and HEAVY models are more suitable for measuring tail risks, especially VaR.

Table 4.5: Relative performance of the alternative forecasting models

Notes: Table 4.5 presents the comparative performance of the GARCHt, HEAVY, HAR-RV and LQR-RV to the benchmark GARCHn at seven quantiles and three forecast horizons. Panel A presents the difference in QS of a model to the benchmark GARCHn as shown in (4.40), while Panel B shows the difference in WQS as shown in (4.41). Both $\Delta QS^{\alpha}{}_{i}$ and ΔWQS_i are comparative and positively oriented, which means we prefer model with higher $\Delta QS^{\alpha}{}_{i}$ and ΔWQS_i.

	Panel A: ΔQS							Panel B: ΔWQS				
	0.01	0.05	0.1	0.5	0.9	0.95	0.99	Uniform	Center	Tails	Left	Right
h=1												
GARCHt	-0.0031	-0.0269	-0.0198	-0.0013	0.0156	0.0038	-0.0046	-0.0363	-0.0019	-0.0288	-0.0473	0.0109
HEAVY	0.0479	0.1701	0.1552	-0.0170	0.0866	0.1407	0.0296	0.6132	0.0330	0.4810	0.3893	0.2239
HAR-RV	-0.0109	0.1252	0.1301	0.0014	-0.0680	0.0304	-0.0080	0.2001	0.0131	0.1476	0.2337	-0.0336
LQR-RV	-0.2832	-0.5534	-0.5027	0.0128	-0.3035	-0.2643	-0.1436	-2.0378	-0.1124	-1.5881	-1.4095	-0.6283
h=5												
GARCHt	-0.0184	-0.0277	-0.0034	0.0054	-0.0154	-0.0234	0.0248	-0.0580	-0.0027	-0.0473	-0.0500	-0.0080
HEAVY	0.1335	0.2521	0.2531	0.0007	0.1433	0.1696	0.0811	1.0334	0.0580	0.8014	0.6814	0.3520
HAR-RV	0.1350	0.2987	0.3112	-0.0003	0.1904	0.2024	0.0805	1.2179	0.0710	0.9338	0.7983	0.4195
LQR-RV	0.0402	-0.2424	-0.2648	0.0194	-0.3493	-0.3241	-0.1899	-1.3110	-0.0788	-0.9958	-0.5510	-0.7600
h=22												
GARCHt	0.1099	0.0885	0.0296	-0.0007	0.0044	0.0221	0.0045	0.2583	0.0093	0.2212	0.2300	0.0283
HEAVY	0.1718	0.2426	0.2179	-0.0005	0.0517	0.1044	0.1000	0.8878	0.0433	0.7147	0.6511	0.2368
HAR-RV	0.2005	0.3287	0.2963	0.0050	0.2303	0.2154	0.1272	1.4034	0.0777	1.0924	0.8927	0.5106
LQR-RV	0.0790	-0.0426	-0.0629	0.0196	-0.3022	-0.3115	-0.0390	-0.6596	-0.0444	-0.4821	-0.0997	-0.5600

4.6 Evaluation of forecast combinations

This section aims to investigate whether quantile combination can improve forecast accuracy. In the first sub-section, we combine quantile forecasts at all seven quantile levels to find whether the combined quantile forecasts outperform individual forecasts. In line with previous section, we continue using the QS framework to compare forecast accuracy. The second sub-section presents the VaR implication of this research. Specifically, we investigate whether the combination of stand-alone VaR(1%), or 1%-quantile, can improve the accuracy of VaR forecasts. To evaluate VaR combinations, we employ the two-stage backtesting framework which were previously presented in Chapter 3.

4.6.1 Evaluation of quantile combinations

As mentioned in section 2.4, we employ the CQOM to combine a pair of one-day-ahead stand-alone quantile forecasts with the combination function (4.27) and optimization function (4.28). Recall that the literature commonly sets unity sum constraint to the loadings of individual forecasts (see Granger and Ramanathan, 1984; Halbleib and Pohlmeier, 2012). However, we set no constraints to the loadings to relax model flexibility. We also set no standard on the combination of quantile forecasts. Specifically, we combine models with different distributional assumptions, with both low-frequency and high-frequency models and with direct and indirect quantile forecasts. With five stand-alone models, we have 10 forecast combinations. At every quantile, we compute the QS and WQS of the combined forecasts and compare its value to the benchmark GARCHn. The comparative performance of combined quantile forecasts is presented in Table 4.6.

Table 4.6: Evaluation of combined quantile forecasts

Notes: Table 4.6 presents the comparative performance of the combined quantile forecasts to the benchmark GARCHn at seven quantiles and three forecast horizons. Panel A presents the difference in QS of these four models to the benchmark GARCHn as shown in (4.40), while Panel B shows the difference in WQS as shown in (4.41). Both $\Delta QS^{\alpha}{}_{i}$ and ΔWQS_i are comparative and positively oriented, which means we prefer model that has higher $\Delta QS^{\alpha}{}_{i}$ and ΔWQS_i.

	Panel A: ΔQS							Panel B: ΔWQS				
	0.01	0.05	0.1	0.5	0.9	0.95	0.99	Uniform	Centre	Tails	Left	Right
h=1												
GARCHt + HAR-RV	0.0544	0.1639	0.1636	0.0016	0.0970	0.0900	0.0279	0.5985	0.0367	0.4515	0.3353	0.1896
GARCHt + LQR-RV	0.0253	-0.0088	-0.0015	0.0039	0.1386	0.0963	0.0266	0.2806	0.0180	0.2086	0.0183	0.2263
HAR-RV + LQR-RV	0.0632	0.1232	0.1504	0.0097	0.0435	0.0848	0.0353	0.5102	0.0307	0.3872	0.2981	0.1506
HEAVY + HAR-RV	0.0643	0.1849	0.1767	0.0016	0.1108	0.1317	0.0407	0.7108	0.0424	0.5413	0.3749	0.2512
HEAVY + LQR-RV	0.0705	0.1605	0.1512	0.0039	0.1729	0.1761	0.0534	0.7886	0.0474	0.5991	0.3396	0.3543
GARCHt + HEAVY	0.0587	0.1272	0.1586	0.0017	0.1606	0.1337	0.0353	0.6758	0.0425	0.5059	0.3032	0.2877
GARCHn + HARRV	0.0541	0.1600	0.1659	0.0016	0.0875	0.0945	0.0307	0.5944	0.0361	0.4498	0.3333	0.1887
GARCHn + HEAVY	0.0588	0.1600	0.1429	0.0016	0.1533	0.1531	0.0348	0.7045	0.0429	0.5330	0.3201	0.2987
GARCHn + LQR-RV	0.0261	0.0108	0.0149	0.0039	0.1330	0.1115	0.0293	0.3296	0.0207	0.2469	0.0500	0.2382
GARCHn + GARCHt	0.0199	0.0015	0.0359	0.0017	0.0893	0.0270	0.0138	0.1891	0.0134	0.1356	0.0513	0.1110

The Panel A of Table 4.6 presents the difference in the QS and WQS of combined forecasts with the benchmark model, while Panel B reports the difference in the WQS. It is clear that the quantile combinations outperform the benchmark model in most cases. The only exception is the combination of LQR-RV and GARCHt, when it poorly performs at 5% and 10% quantile. Comparing the value of ΔQS and ΔWQS between Table 4.6 and Table 4.5, we find that the quantile combinations outperform the stand-alone models in producing accurate quantile forecasts. Compared to stand-alone forecasts, the combined quantile forecasts provide very good results, which remain robust in respect of the choices of model and distributional assumptions.

We also find that the good combination does not necessarily come from two good stand-alone forecasts. Indeed, the optimal combination is between the HEAVY, the best stand-alone model, and LQR-RV, the worst one. It can be explained that different models have their own dynamics and use different sets of information. The forecast combination is to pool the information contained in individual components. Thus, the good combination comes from the diversification gains from the stand-alone forecasts.

4.6.2 Evaluation of Value-at-Risk combinations

Recall that VaR measures the maximum amount of will not be exceeded with a specific time interval and level of confidence. Conditional on the information set at time t, denoted as Ω_t, the next day VaR is defined as the negative α-quantile of the return distribution:

$$\mathrm{VaR}(\alpha)_{t+1} = -q_\alpha(r_{t+1}|\Omega_t) \tag{4.42}$$

It is obvious that VaR(1%) is the 1%-quantile forecast of the return distribution. Therefore, the combined VaR(1%) is indeed the 1%-quantile combinations in previous section. To investigate whether VaR combination can improve forecast accuracy, we compare its performance to the stand-alone 1%-quantile forecasts.

We continue using our two-stage backtesting framework in Chapter 3 to evaluate VaR combinations. Recall that our backtesting procedure uses the statistical tests to evaluate the absolute performance and the magnitude LF to compare the accuracy of VaR estimates. We do not use QS approach to evaluate VaR forecasts, as the QS quantifies the asymmetric loss $q^{i}_{t+h}(\alpha|\ \Omega_t)$ - r^{i}_{t+h} in both cases that $r^{i}_{t+h} \leq q^{i}_{t+h}(\alpha|\ \Omega_t)$ and $r^{i}_{t+h} > q^{i}_{t+h}(\alpha|\ \Omega_t)$ (see Equation 4.38). However, risk managers are only concerned with the losses that exceed the given quantile e.g. when $r^{i}_{t+1} < q^{i}_{t+1}(\alpha|\ \Omega_t)$, which is. Standing on the risk management perspective, we argue that magnitude LF is more appropriate to evaluate VaR estimates than the QS. We present the evaluation of VaR combinations in Table 4.7.

Table 4.7: Evaluation of VaR combinations

Notes: Table 4.7 presents the evaluation of VaR forecasts, including actual exception ratio ($\hat{\alpha}$), the test statistics of the UC hypothesis (LR_{UC}), the IND hypothesis (LR_{IND}), the CC hypothesis (LR_{CC}) and the magnitude LF. Panel A presents the test results of stand-alone VaR estimates, while Panel B reports the evaluation of VaR combinations.

	$\hat{\alpha}$	LR_{UC}	LR_{IND}	LR_{CC}	LF
GARCHn	0.0134	3.6701	0.1913	3.8613	0.0108
GARCHt	0.0137	4.2811**	0.1602	4.4413	0.0111
HEAVY	0.0137	4.2811**	0.1602	4.4413	0.0094
HAR-RV	0.0148	7.1393**	1.4183	8.5575**	0.0098
LQR-RV	0.0088	0.5083	0.5521	1.0604	0.0129
GARCHt + HAR-RV	0.0131	0.0782	0.1481	0.0745	0.0090
GARCHt + LQR-RV	0.0097	0.8486	0.4149	0.7043	0.0099
HAR-RV + LQR-RV	0.0134	0.0554	0.1607	0.0597	0.0088
HEAVY + HAR-RV	0.0131	0.0782	0.1481	0.0745	0.0085
HEAVY + LQR-RV	0.0097	0.8486	0.4149	0.7043	0.0083
GARCHt + HEAVY	0.0105	0.7519	0.4081	0.6756	0.0080
GARCHn + HARRV	0.0131	0.0782	0.1481	0.0745	0.0089
GARCHn + HEAVY	0.0100	0.9838	0.3626	0.6606	0.0077
GARCHn + LQR-RV	0.0094	0.7164	0.4288	0.6846	0.0091
GARCHn + GARCHt	0.0097	0.8486	0.4149	0.7043	0.0088

**: significant at 95% confidence level

Table 4.7 shows the poor performance of stand-alone VaR(1%), as three over five estimates are rejected by the UC test due to VaR understatement. Besides, we witness the clustering of VaR exceptions in case of HAR-RV, which leads to the rejection of the CC test. The GARCHt and HEAVY models, although being marginally rejected by the UC test, still be able to pass the CC test. The magnitude LF shows that HEAVY is the best stand-alone VaR estimates, while the LQR-RV is less accurate in capturing the excessive losses.

It is clear that VaR combinations are superior to the individual VaR forecasts. Indeed, the exception ratios of combined VaR forecasts are closer to the 1% nominal rate than the stand-alone estimates. From Table 4.6, the failure rate of stand-alone

VaR ranges from 0.0088 to 0.0148 with average value of 0.01264. After the combination, the actual coverage rate of VaR exceptions reduces to 0.01117, which is closer to the nominal rate of 0.01. Recall that the stand-alone VaR estimate from HAR-RV model fails the coverage tests at the 1%-quantile (see Table 4.4 and Table 4.7) as the model underestimates the left tail. However, when combining with other stand-alone forecasts, its performance remarkably makes a progress. Indeed, there is no evidence that the combinations including HAR-RV forecasts fail the tests of UC and IND hypotheses.

The LF results given in Table 4.7 confirm the value of VaR combination in improving forecast accuracy. Put aside the individual VaR forecasts, VaR combinations are noticeable superior in minimizing the magnitude of exceeded losses. Not only reduce the failure rate, we find that VaR combinations also generate lower LF than any of its components. Quantitatively, the average LF of fives individual forecasts is 0.0104, while the corresponding value of 10 VaR combinations reduces to 0.0087. Thus, we support the use of VaR combinations in improving the accuracy of VaR forecast.

4.7 Concluding remarks

This chapter examines the value of intraday return information on the accuracy of quantile forecasts by comparing the out-of-sample performance of high-frequency-based models to the GARCH(1,1) models. The realized volatility-based models excelling in this chapter include HAR-RV, the HEAVY and the LQR-RV model. The quantile forecasts are obtained using indirect approach (including the GARCH(1,1), HAR-RV and HEAVY) and direct approach (the LQR-RV model) using the 5-minute intraday data of WTI Crude Oil Futures. To investigate the accuracy of quantile forecasts, we employ the two-stage evaluation, including the DQ test and the QS method. In comparison with the GARCH(1,1), we find the HAR-RV and HEAVY provide better quantile forecasts across quantile levels and forecast horizons, while the LQR-RV performs poorly.

Chapter 4 also examines the power of quantile combination in improving forecast accuracy. Using the CQOM method, we combine one-day-ahead quantile forecasts in pair. We find that quantile combinations noticeably improve the accuracy of individual forecasts. In line with previous chapter, this study has an VaR implication. It is evident that VaR combination not only helps individual forecasts satisfy the UC and CC hypothesis, but also improve the forecast accuracy.

Chapter 5: Conclusions

This thesis consists of three empirical studies on the estimation of VaR at banks. In the first study, we aim to investigate the empirical performance of VaR estimates at commercial banks. While prior studies only focus on bank VaRs in specific market, this study contributes to the literature as the first investigating the performance of bank VaRs on international level. We also use much richer dataset, covering the pre-crisis, financial crisis and post-crisis periods. Our empirical analysis shows that banks were systematically conservative in estimating their VaR in the pre-crisis and post-crisis periods. During the global financial crisis, we witnessed different behaviours of bank VaRs. While some banks continued to overstate their VaRs, other banks remarkably understated their risk. Thus, the number of VaR exceptions for these banks are excessively high and tend to cluster together. We find evidence of extreme losses during financial crisis which likely exceeded the economic capital of banks. We attribute the causes of the poor performance of bank VaRs to the use of contaminated data, the choice of VaR model and the benefit of VaR manipulation.

The second empirical study contributes to the literature as one of the very few studies investigating the forecasting power of VaR models using bank data. Compared to prior studies, our rich and international dataset helps increasing the power of the statistical tests, which allows us to have more comprehensive evaluation. In this chapter, we compare the performance of internal VaR model at banks and alternative VaR approaches, including the HS, VCV and EVT. Our two-stage backtest shows the superiority of the GARCH-type models in estimating bank VaRs. Regarding to the choice of distributional assumption, we find that the

Gaussian distribution uniformly improves VaR predictive power in normal periods, while the Student t is by far the best in estimating VaR during financial crisis. While the HS models perform inconsistently, the worst performances in our empirical analysis belong to the banks' internal model and the EVT. It is surprising that the EVT approach, which was shown to be superior in estimating in VaR in prior studies, performs very poorly with bank data. We argue that good bank VaRs can be obtained by the simple GARCH-type models rather than the internal model at banks or other complicated models.

The third study focuses on the accuracy of VaR estimates with high-frequency data. Specifically, we answered two questions: (i) does the use of high-frequency data helps improving forecast accuracy and (ii) does the use of quantile combination enhance the performance of individual quantile forecasts. Our dataset includes the time series of 5-minute sampling frequency of WTI Crude Oil Futures from Jan 1995 to June 2016. Comparing the performance of the GARCH(1,1) models with the RV-based models, we find that the use of high-frequency data can improve the accuracy of quantile forecasts. Besides, we examine the power of quantile combination in improving forecast accuracy. Using the CQOM method, we combine one-day-ahead quantile forecasts in pair. We find that the quantile combinations noticeably improve the accuracy of individual forecasts.

We find that the daily P/L and bank VaRs are not integrated across the entire trading portfolio. Indeed, they are estimated for subgroup of positions, including foreign exchange, interest rate, equities, commodities and credit portfolio. The information about the decompositions of bank VaRs and the VaR diversification effect can be found on banks' annual report in the market risk management section. Therefore, it is possible to quantitatively estimate the average contribution of each subgroup of VaR estimates and the impact of portfolio diversification resulting from the aggregated bank VaR. Using market data as proxy for each subgroup and considering the diversification effect, one can simulate the trading portfolio to use as the input of VaR models.

We suggest that future research should use high-frequency data for quantile forecasting. One of the limitations of the thesis is that we only implement the real-world density forecasts. Therefore, it is possible to carry out a risk-neutral density forecast and compare its performance to the real-world density forecast. Furthermore, future studies can examine the combination of stand-alone density forecasts and compare their accuracy to the performance of quantile combinations. Finally, future research should focus on a larger dataset comprise of banks in both developed and developing countries.

References

Adrian, T. and Shin, H.S., 2014. Procyclical leverage and value-at-risk. *Review of Financial Studies, 27(2),* pp.373-403.

Alexander, C., 2009. *Market Risk Analysis, Value at Risk Models (Vol. 4):* John Wiley & Sons.

Aloui, R., Aïssa, M.S.B. and Nguyen, D.K., 2011. Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking & Finance, 35(1),* pp.130-141.

Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25,* 177–190.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885– 905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001a. The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001b. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica 71*, 529–626.

Andersen,T. G.,T. BollerslevandF. X. Diebold. 2009. Parametric and Nonparametric Volatility Measurement. In: L.P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics.* Amsterdam: North-Holland.

Angelidis, T., Benos, A. and Degiannakis, S., 2004. The use of GARCH models in VaR estimation. *Statistical methodology*, *1*(1-2), pp.105-128.

Balkema, A.A. and De Haan, L., 1974. Residual life time at great age. *The Annals of probability*, pp.792-804.

Bao, Y., Lee, T.H. and Saltoglu, B., 2006. Evaluating predictive performance of value-at-risk models in emerging markets: a reality check. *Journal of forecasting*, *25*(2), pp.101-128.

Barndorff-Nielsen, O.E. and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the*

Royal Statistical Society: Series B (Statistical Methodology), *64*(2), pp.253-280.

Barndorff-Nielsen, O.E. and Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, *2*(1), pp.1-37.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica, 76(6),* pp.1481-1536.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K., 1998. Don't look back.

Barone-Adesi, G., Giannopoulos, K. and Vosper, L., 1999. VaR without correlations for portfolio of derivative securities.

Basel Committee on Banking Supervision, 1995. ''An Internal Model-Based Approach to Market Risk Capital Requirements.''

Basel Committee on Banking Supervision, 1996a. Amendment to the Capital Accord to Incorporate Market Risks, Bank for International Settlements.

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, Bank for International Settlements.

Beder, T.S., 1995. VAR: Seductive but dangerous. *Financial Analysts Journal*, *51*(5), pp.12-24.

Bekiros, S.D. and Georgoutsos, D.A., 2005. Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance. *Journal of International Financial Markets, Institutions and Money*, *15*(3), pp.209-228.

Berkowitz, J. and O'brien, J., 2002. How accurate are value-at-risk models at commercial banks? *The journal of finance, 57(3),* pp.1093-1111.

Best, P., 2000. *Implementing value at risk*. John Wiley & Sons.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), pp.307-327.

Bollerslev, T., Kretschmer, U., Pigorsch, C., Tauchen, G., 2009. A discrete-time model for daily S&P500 returns and realized variations: jumps and leverage effects. *Journal of Econometrics 150 (2),* 151–166.

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

Boudoukh, J., Richardson, M. and Whitelaw, R., 1998. The best of both worlds. *Risk*, *11*(5), pp.64-67.

Brownlees, C.T. and Gallo, G.M., 2010. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics, 8(1),* pp.29-56.

Cairns, A.J., 2000. A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, *27*(3), pp.313-330.

Chan, K.F. and Gray, P., 2006. Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of forecasting*, *22*(2), pp.283-300.

Christoffersen, P.F., 1998. Evaluating interval forecasts. *International economic review*, pp.841-862.

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. *International journal of forecasting, 5(4),* pp.559-583.

Clements M. P., Galvao, A.B., Kim, J. H., 2008. Quantile forecasts of Daily exchange rate returns from Forecasts of realized volatility. *Journal of Empirical Finance 15,* 174-196.

Colletaz, G., Hurlin, C. and Pérignon, C., 2013. The Risk Map: A new tool for validating risk models. *Journal of Banking & Finance, 37(10),* pp.3843-3854.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*, 174–196.

Cuoco, D. and Liu, H., 2006. An analysis of VaR-based capital requirements*. Journal of Financial Intermediation, 15(3)*, pp.362-394.

da Silva, A.L. and de Melo Mendes, B.V., 2003. Value-at-risk and extreme returns in Asian stock markets. *International Journal of Business*, *8*(1), p.17.

Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G., 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate analysis*, *76*(2), pp.226-248.

Danielsson, J. and de Vries, C.G., 1997. Tail index and quantile estimation with very high frequency data. *Journal of empirical Finance*, *4*(2-3), pp.241-257.

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

Dowd, K., 1998. *Beyond value at risk: the new science of risk management (Vol. 3).* Wiley Chichester.

Dowd, K., 1999. A value at risk approach to risk-return analysis. *The Journal of Portfolio Management, 25(4)*, 60-67.

Dowd, K., 2000. Accounting for value at risk. *Journal of Risk Finance 2 (Fall):* 51–58.

Dowd, K., 2005. *Measuring market risk, Second edition.* John Wiley & Sons.

Dupuis, D.J., 1999. Exceedances over high thresholds: A guide to threshold selection. *Extremes*, *1*(3), pp.251-261.

Embrechts, P., Resnick, S.I. and Samorodnitsky, G., 1999. Extreme value theory as a risk management tool. *North American Actuarial Journal, 3(2),* pp.30-41.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pp.987-1007.

Engle, R., 2002. New frontiers for ARCH models. *Journal of Applied Econometrics*, *17*(5), pp.425-446.

Engle, R.F., Gallo, G.M., 2006. A multiple indicators model for volatility using intradaily data. *Journal of Econometrics 131*, 3–27.

Engle, R.F., Manganelli, S., 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics 22*, 367–381.

Fernandez, V., 2003. Extreme value theory and value at risk.

Frésard, L., Pérignon, C. and Wilhelmsson, A., 2011. The pernicious effects of contaminated data in risk management. *Journal of Banking & Finance, 35(10),* pp.2569-2583.

Forsberg, L. and E. Ghysels, 2004. Jumps or no jumps, smooth or rough: It does not matter for predicting volatility - Choosing the right regressor does," *Discussion paper UNC.*

Gaglianone, W.P., Lima, L.R., Linton, O., Smith, D.R., 2011. Evaluating value-at-risk models via quantile regression. *Journal of Business and Economic Statistics 29*, 150–160.

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

Ghysels, E., Santa-Clara, P. and Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics, 131(1)*, pp.59-95.

Ghysels, E., Sinko, A. and Valkanov, R., 2007. MIDAS regressions: Further results and new directions*. Econometric Reviews, 26(1),* pp.53-90.

Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics 23*, 416–431.

Gilli, M. and Kellezi, E., 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics, 27(2-3)* 207-228

Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance. Vol. 48, No. 5*, pp. 1779–1801.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*, 359-378.

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

Granger, C.W. and Ramanathan, R., 1984. Improved methods of combining forecasts. *Journal of forecasting*, 3(2), pp.197-204.

Greenspan, A., 1997. Discussion at Symposium: Maintaining Financial Stability in a Global Economy. *The Federal Reserve Bank of Kansas City*, *54*.

Hansen, P.R. and Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of applied econometrics*, *20*(7), pp.873-889.

Halbleib, R., Pohlmeier, W., 2012. Improving the value at risk forecasts: Theory and evidence from the financial crisis. *Journal of economic dynamics and control.* 1212-1228.

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

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics 24*, 127–218.

Hansen, P.R., Huang, Z., Shek, H.H., 2012. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*, 877–906.

Hua, J., Manzan, S., 2013. Forecasting the return distribution using high-frequency volatility measures. *Journal of Banking and Finance*. 4381-4403.

Huang, D., Yu, B., Fabozzi, F.J. and Fukushima, M., 2009. CAViaR-based forecast for oil price risk. *Energy Economics, 31(4),* pp.511-518.

Huang, Z., Liu, H. and Wang, T., 2016. Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model. *Economic Modelling, 52*, pp.812-821.

Hull, J. and White, A., 1998. Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of risk*, *1*(1), pp.5-19.

Jeon, J. and Taylor, J.W., 2013. Using CAViaR Models with Implied Volatility for Value-at-Risk Estimation. *Journal of Forecasting, 32(1),* pp.62-74.

Jorion, P., 2002. How informative are value-at-risk disclosures? *The Accounting Review, 77(4)*, pp.911-931.

Jorion, P., 2006. *Value at risk: the new benchmark for managing financial risk (Vol.2).* McGraw-Hill New York.

Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica 46,* 33–50.

Koenker, R. and Zhao, Q., 1996. Conditional quantile estimation and inference for ARCH models. *Econometric Theory, 12(05),* pp.793-813.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised, and implied volatility measurements. *Journal of Empirical Finance 12,* 445–475.

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.H., 1995. Techniques for verifying the accuracy of risk measurement models. *The journal of Derivatives, 3(2),* pp.73-84.

Kyle, A. S. and Obizhaeva, A. A., 2016. Market Microstructure Invariance: Empirical Hypotheses. *Econometrica, 84*: 1345–1404.

Lee, T.-H., Bao, Y., & Saltoglou, B. (2006). Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. *Journal of Forecasting, 25*, 101–128.

Linsmeier, T.J. and Pearson, N.D., 1996. Risk measurement: An introduction to value at risk.

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

Lopez, J. A. (1997). Regulatory evaluation of value-at-risk models. Federal Reserve Bank of New York.

Lopez, J. A. (1998). Methods for evaluating value-at-risk estimates: Federal Reserve Bank of New York.

Lucas, A., 2001. Evaluating the Basle guidelines for backtesting banks' internal risk management models. *Journal of Money, Credit and Banking*, pp.826-846.

Ma, L. and Pohlman, L., 2008. Return forecasts and optimal portfolio construction: a quantile regression approach. *The European Journal of Finance, 14(5),* pp.409-425.

Maheu, J.M., McCurdy, T.J., 2011. Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics 160*, 69–76.

Makridakis, S. and Hibon, M., 2000. The M3-Competition: results, conclusions and implications*. International journal of forecasting, 16(4),* pp.451-476.

Methta, A., Neukirchen, M., Pfetsch, S., Poppensieker, T., 2012. Managing market risk: Today and tomorrow*. McKinsey working papers on Risk, number 32*.

McNeil, A.J. and Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, *7*(3-4), pp.271-300.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Pictet, O.V., Olsen, R.B. and Ward, J.R., 1993. Fractals and intrinsic time: A challenge to econometricians. Unpublished manuscript, Olsen & Associates, Zürich.

Nocera, J., 2009. Risk management. New York Times.

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

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

Pickands III, J., 1975. Statistical inference using extreme order statistics. *The Annals of Statistics*, pp.119-131.

O'Brien, J.M. and Szerszen, P., 2014. An evaluation of bank var measures for market risk during and before the financial crisis. Finance and Economics Discussion Series 2014-21. Board of Governors of the Federal Reserve System (U.S.).

Sarma, M., Thomas, S. and Shah, A., 2003. Selection of VaR models. *Journal of Forecasting, Vol. 22 No. 4*, pp. 337-58.

Schaumburg, J., 2012. Predicting extreme value at risk: Nonparametric quantile regression with refinements from extreme value theory. *Computational Statistics & Data Analysis, 56(12),* pp.4081-4096.

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

So, M.K. and Philip, L.H., 2006. Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, *16*(2), pp.180-197.

Stock, J.H. and Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set*. Journal of Forecasting, 23(6),* pp.405-430.

Taylor, J., 2016. Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution. *Journal of Business and Economic Statistics.*

Timmermann, A., 2005. Forecast combinations. Working paper.

Vortelinos, D.I., 2015. Forecasting realized volatility: HAR against Principal Components Combining, neural networks and GARCH. *Research in international business and finance.*

Woods, M., Dowd, K., Humphrey, C., 2008. The value of risk reporting: a critical analysis of value-at-risk disclosures in the banking sector. *International Journal of Financial Services Management 3,* 45-64.

Zhang, L., Mykland, P. A. and Ait-Sahalia, Y., 2003. A tale of two time scales: Determining integrated volatility with noisy high-frequency data", Working Paper w10111, NBER.

Zikes, F., Barunik, J., 2016. Semiparametric Conditional Quantile Models for Financial Returns and Realized Volatility. *Journal of Financial Econometrics.* 185-226.