Essays on Market Risk Measures

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

ICMA CENTRE

HENLEY BUSINESS SCHOOL

THE UNIVERSITY OF READING

Xiaohan Xue

May 2022

"取法乎上，仅得其中；取法于中，故为其下。"

"*Shoot for the moon. Even if you miss, you'll land among the stars.*"

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Xiaohan Xue

Acknowledgements

I want to express my gratitude to my supervisors, Dr Emese Lazar and Dr Shixuan Wang, for their patient guidance and continuous support throughout the thesis. Emese has provided her invaluable expertise in formulating the research questions and guided me to write research papers with an academic style. Her feedback always sharpens my thinking and leads me to a higher level. Shixuan always can provide optimal solutions when I suffer from problems in my journey. His research outputs and hardworking always motivate me to continue my academic career. Apart from being my supervisors, they have been very inspiring coauthors with fruitful insights into this research and good friends in my life. They always provide constructive comments on my presentation skills and writing style. I also greatly appreciate their support and encouragement in my job seeking.

I sincerely thank the ICMA Centre for its generous financial support and the Academic Computing Cluster at the University of Reading for support with the supercomputing clusters. I am very grateful to my family, friends, and colleagues for being very kind and supportive.

Finally, I would like to thank the examiners, Professor Michael Clements and Dr Ruijun Bu, for their precious time reviewing this thesis and their constructive comments.

Abstract

The thesis investigates topics on how to improve the estimation and forecasting for market risk measures (focusing on Value at Risk and Expected Shortfall, denoted by VaR and ES, jointly) by building superior models with extra information and by detecting structural changes in risk models (in a retrospective manner and a real-time manner).

The first contribution is introducing a new framework by incorporating intra-day information into dynamic semiparametric models to forecast VaR and ES. We consider the intraday measures including the realized variance and overnight returns. In the practical application, we apply the proposed models to international stock market indices, then evaluate the forecasting performance via various backtests. Our results show that our models outperform the benchmarks consistently across all indices and various significance levels.

Secondly, this thesis develops a test that can efficiently captures change points in the (VaR, ES) estimated by (semi)parametric models. We derive the asymptotic distribution of the test statistic and adopt a stationary bootstrapping technique to obtain the p-values of the test statistic. Monte Carlo simulation results show that our proposed test has better size control and higher power than the alternative tests. An empirical study of risk measures based on the S&P 500 index illustrates that our proposed test can detect change points associated with well-known market events.

The third main contribution is proposing a sequential monitoring method to detect changes in semiparametric risk models for (VaR, ES). We derive the asymp-

totic theorem for the monitoring scheme under the null hypothesis. Our Monte Carlo simulations with finite sample sizes show that this test has reasonable size control under the null hypothesis and high power under alternative hypotheses. Empirical applications based on the S&P 500 index and the GBP/EUR exchange rate illustrate that the detected change points often precede the actual market crashes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation for the Thesis

In the context of the ongoing COVID-19 recession and recent financial crises, the measurement and forecasting of financial risk have attracted unprecedented attention from academia and the financial industry. According to Basel III, financial risk can be divided into three main categories: market risk, credit risk and operational risk. Market risk measures the potential losses of portfolios exposed to the fluctuation of market indices or prices at a given significance level. Credit risk refers to the possibility of a loss caused by a failure to repay loans or default of contractual obligations by a borrower. Operational risk is associated with a potential loss resulting from the ineffectiveness or failures in the internal activities, procedures and systems. In this thesis, we focus on the concept of market risk and on approaches that accurately estimate and forecast market risk measures.

From the perspective of financial risk managers, a risk measure can be considered a map from spaces of probability distributions to actual losses. Risk measures can provide banks and other financial institutions with specific values of potential losses so that risk managers can adjust their capital reserves against the downside risk. Value at Risk and Expected Shortfall are the prevailing financial risk

measures that dominate current financial regulation.

Value at Risk (abbreviated as VaR henceforth), as a simple financial risk measure, can be traced back as far as 1922, when capital requirements were imposed by the New York Stock Exchange (NYSE) on member firms (Holton, 2004). Since 1996, VaR has been adopted as a market risk measure and attracted broad interest from market participants and academic researchers in multiple disciplines. VaR provides banks and other financial institutions with a loss level that occurs in the worst situation at a given significance level over a certain period. VaR also facilitates capital requirements computation for practitioners and regulators to take efficient capital allocation and risk management actions. Using VaR to measure risk has the main advantages of being intuitive and easily understood.

However, VaR is sometimes criticized because it cannot capture the tail's structure beyond the quantile. Additionally, VaR has another inherent deficiency due to the absence of *sub-additivity* for a portfolio, meaning that the portfolio cannot benefit from asset diversification when we use VaR as the risk measure. Due to the lack of *sub-additivity*, VaR is not a *coherent* risk measure.[1] Thus, to overcome the inadequacy of this risk measure, Artzner et al. (1999) propose a supplementary measure to VaR, named Expected Shortfall (abbreviated as ES henceforth). ES, computed as the expected value of exceedances beyond VaR, can capture the size of losses above a certain threshold. Also, ES fulfils all the properties related to *coherence*. The Basel Committee on Banking Supervision (2019) has proposed a transition from 1% VaR to 2.5% ES to formulate the capital requirements for banks and other financial institutions.

Another property of risk measures discussed in this thesis refers to *elicitability*. Based on Gneiting (2011), a risk measure is *elicitable* if there exists a loss function (or a scoring function) for the risk measure that can be used to comparatively evaluate the performance of models. In several semiparametric risk models, the parameter estimates are obtained by minimizing a consistent loss function, e.g.,

the "Lin-Lin" (or "tick" loss function) for VaR (more details about the estimation for VaR can be found in Engle and Manganelli (2004)). Gneiting (2011) documents that ES is elicitable only jointly with VaR[2], and thus individual ES cannot be estimated by this approach minimizing a loss function, nor can models be compared in terms of ES forecasting performance by evaluating average loss values. Fissler and Ziegel (2016) propose a class of loss functions (the FZ loss functions) for VaR and ES considered jointly. Consequently, throughout the thesis, we focus on the (VaR, ES) tuple and employ the FZ loss functions for these market risk measures.

Models used to estimate and forecast VaR and ES in the existing literature can be classified into three main categories: parametric, semiparametric and nonparametric (Engle and Manganelli, 2004; Taylor, 2008). Parametric models jointly predict VaR and ES via a conditional volatility forecast, which commonly relies on the assumption of the distribution of asset returns. The selection of density function impacts the estimation and forecasting of risk measures, especially when using unstable data. Conversely, nonparametric methods make no assumptions about the conditional distribution of asset returns. These methods estimate VaR and ES as quantiles of the chosen sample of returns over a specific window at a given significance level. Nonparametric methods are model-free and easy to implement (Engle and Manganelli, 2004), but they are often criticized because of the sensitivity to window size selection. Semiparametric models impose a parametric structure on the dynamics of VaR and ES through a dynamic framework but require no assumptions on the conditional distributions of financial returns (Patton et al., 2019).

Regarding the extreme losses during the recent crises, e.g., the COVID-19 recession and the cryptocurrency crash in 2021, building improved risk models that can capture the significant losses in holding portfolios has become one of the central questions in risk management. Particularly, estimating market risk can be disrupted by the presence of breaks and dependence on the dataset. Ignoring

structural breaks can cause estimation bias and forecasting errors, which negatively affect the decisions of regulators and risk managers. To mitigate the effect of changes in model parameter values, extensive literature documents various solutions, including proposing time-varying parameter models and using statistical tests to detect changes in the model parameter values.

On the one hand, models with time-varying parameters are well designed to fit such time series. Creal et al. (2013) propose a set of observation-driven models, as the generalized autoregressive score (GAS) models based on characteristics modelled as a function of the scaled score of the likelihood function. Inspired by Creal et al. (2013), Patton et al. (2019) construct a class of GAS models to estimate VaR and ES in a dynamic score-driven framework, where the parameters are estimated by minimizing a specified loss function. Motivated by Patton et al. (2019), we improve the semiparametric GAS models for VaR and ES by incorporating intraday and overnight measures of return variation.

On the other hand, a strand of literature uses statistical tests to detect changes in the parameter values of models.[3] It is worth mentioning that ignoring the presence of change points may cause misleading statistical inferences. A groundbreaking change detection method has been proposed by Page (1954). This method is a sequential scheme that compares the local and proportional global mean within a historical sample. Following this seminal work, structural breaks have been extensively investigated in the variance and correlation dynamics, as well as in quantile regressions. In risk management, after detecting change points in a risk model within a historical sample, practitioners can re-estimate the parameters by considering the presence of change points. Thus, we propose a new detection method for change points in the tuple (VaR, ES) estimated by (semi)parametric models in Chapter 3.

In practical applications, risk managers are more concerned about the timing to update model parameter values in order to suit the newly arriving data. However, the tests discussed above are designed to detect change points within

a historical time series dataset only, rather than continually monitoring whether a structural change has occurred as new data is revealed. To address this issue, Chu et al. (1996) propose a novel test paradigm for changes in a time series based on a detector and a boundary function. Inspired by this work, in Chapter 4, we extend the sequential monitoring scheme to joint (VaR, ES) risk models.

In the next section, we present an overview of the thesis.[4]

1.2   Overview of the Thesis

In Chapter 2, we propose a new framework for the joint modelling and forecasting of dynamic VaR and ES by incorporating intraday information into the semiparametric models introduced by Patton et al. (2019). Creal et al. (2013) formulate the idea by including the scaled score of the conditional observation density with respect to the time-varying parameters. To examine the improvement of incorporating intraday and overnight measures into the GAS framework in forecasting VaR and ES, we consider four intraday measures: the realized volatility at 5-min and 10-min sampling frequencies and the overnight return incorporated into these two realized volatilities. Combining the overnight returns with the realized volatility can also capture the intraday and overnight information. Thus, Chapter 2 of this thesis sheds light on the extensions of the GAS model and the improvement of forecasting performance by considering models that include intraday and overnight return information.

In the empirical study, the newly proposed semiparametric models are applied to four international stock market indices and compared with a range of parametric, nonparametric and semiparametric models, including historical simulations, GARCH and the original GAS models. To evaluate the performance of the risk models, we first employ backtesting approaches for VaR or ES forecasts individually. The backtests we consider include the unconditional coverage test introduced by Kupiec (1995) and the dynamic quantile test proposed by Engle

and Manganelli (2004) for VaR individually and the dynamic ES regression test used by Patton et al. (2019) for ES individually. Moreover, regarding the joint (VaR, ES) backtests, we compare the average loss values generated by the FZ0 loss function proposed by Fissler and Ziegel (2016), then employ the Diebold-Mariano test (Diebold and Mariano, 2002) and the Model Confidence Set test (Hansen et al., 2011) for the loss values.

The in-sample estimation results indicate that the coefficients for the intraday and overnight measures are all statistically significantly positive at both 1% and 5% significance levels. Intuitively, larger realized or overnight volatility will lead to a lower quantile in the next trading day. Our out-of-sample results show that the GAS models, enhanced with the realized volatility measures, outperform the benchmark models consistently across all indices and various probability levels, $\alpha = 10\%$, 5%, 2.5% and 1%. We found that the two-factor GAS model combined with 10-min realized volatility and the overnight returns can provide more accurate risk measures for risk management purposes than other models.

The semiparametric GAS framework, which captures time variation in parameters of risk models, is designed to fit the time series with structural breaks. In order to mitigate the effect of changes through another channel, we investigate change detection in risk models for in-sample and real-time manners in the following chapters.

Chapter 3 develops a new test to detect change points in (semi)parametric models for VaR and ES within a historical sample, based on the Wilcoxon test for the FZ loss values. We explain the intuition behind the test as follows. If the risk model parameters are well estimated in a stationary process, the optimal values for VaR and ES are corresponding to the minimum point on the FZ loss function and generate a stable loss series. However, if there is a change point in the process, the parameter values estimated from the whole sample are not suitable for each subsample (before and after the change point), which will result in breaks in the loss series. Thus, we aim to identify the change points in the loss

series instead of the time series of the tuple (VaR, ES).

We propose a new test statistic and derive the asymptotic behaviour of this statistic under weak dependence. To improve the finite sample performance of the test, we adopt a stationary bootstrap method based on Politis and Romano (1994) and prove the validity of the bootstrapping method for this test.

Next, we apply the proposed change point detection method in various designed scenarios to evaluate its performance. The Monte Carlo (MC) simulation results indicate that this test has better size control under the null hypothesis and higher power under the alternative hypotheses with finite sample sizes than other benchmark tests. Additionally, our empirical study on the risk measures of the S&P 500 index returns shows that this test can detect change points within the selected sample, which are consistent with well-known financial and economic events.

However, most tests for structural breaks in the literature (including the test introduced in Chapter 3) are designed to detect change points only within a given historical dataset rather than for newly arriving data. If we use the historical observations to obtain the optimal parameter values for a risk model, are the parameters estimated yesterday able to explain today's data? When do we need to adjust the model parameter values to fit the changes in the model parameters? Can we identify which parameter dominates the change in a risk model? To answer the questions stated above, we propose a sequential monitoring test for the structural change in the M-estimators of semiparametric risk models.

Thus, Chapter 4 of this thesis introduces a sequential monitoring procedure to detect changes in the parameter values of semiparametric VaR and ES joint risk models. The monitoring scheme depends on a proposed detector and an adequately selected boundary function, i.e., a change is detected when the detector crosses the boundary, following the study of Chu et al. (1996). In our case, the detector is based on the cumulative sequence of gradients of the FZ loss function with respect to (w.r.t.) the model parameters. The boundary function is chosen

such that the probability of a false detection under the null hypothesis of stable parameters is fixed. Our test is uniquely proposed for sequentially monitoring changes in the parameter values of the semiparametric models for VaR and ES.

Following this, we derive the asymptotic behaviour of the stopping time under defined assumptions. We can numerically obtain the critical values with the corresponding significance levels by simulating the independent Wiener processes based on our asymptotic theorem. In a simulation analysis, we use MC simulations to show the advantages of the proposed sequential monitoring test in identifying change points in the parameters of risk models. We show that our proposed test has a reasonable size control under the null hypothesis in finite samples. Additionally, we consider various scenarios for the post-break process under the alternative hypotheses. Our simulation results reveal that this test can detect the changes in the parameters of risk models with high empirical power.

We evaluate the empirical density of the stopping time estimated by the sequential monitoring manner. The findings show that there is no long delay in the empirically detected stopping time compared with the location of the simulated real change point. Moreover, we propose a novel method to identify the dominant source of the change points in the parameters. The simulation results show that most of the dominant sources of change points are identified correctly.

In an empirical study, we explore the applications of the sequential monitoring test on real data. We consider the S&P 500 index returns and the GBP/EUR exchange returns. The empirical results illustrate that our proposed test can detect change points associated with well-known market events, and often the detection precedes the actual market crashes. In particular, the sequential monitoring test can detect the change points earlier than the beginning of the financial crisis and the Black Thursday (12 March 2020) and Black Monday II (16 March 2020) in the COVID-19 pandemic period. According to our findings, we conclude that practitioners can improve their risk management strategies by monitoring for change points in their risk models and then adjusting the parameters of the

models based on the identified change points. Thus, significant losses can be avoided by adjusting parameter values seasonably.

1.3   Original Contributions

This thesis, including the following three main chapters, contributes to improving the forecasts of risk measures and proposing in-sample and real-time structural break detection methods in risk models for VaR and ES.

(1) The first set of original contributions in terms of risk measures forecasting is:

• we propose a set of novel semiparametric models to forecast VaR and ES jointly via incorporating the intraday and overnight measures into a GAS framework;

• we construct four proxies for the intraday and overnight information;

• we provide solid empirical evidence that the extended semiparametric models outperform other benchmarks via various backtesting methods;

• we compare the performance of the intraday and overnight measures with regard to forecasting VaR and ES when added to the GAS models.

(2) The second set of original contributions, in terms of proposing an in-sample structural break detection test for (VaR, ES) in risk models, is:

• we propose a test to detect change points in both (semi)parametric VaR and ES risk measures simultaneously based on the FZ loss functions;

• we derive the asymptotic behavior of the test statistic under weak dependence;

• we verify that the stationary bootstrap method is valid for calculating critical values for this test;

• we show the advantages of the proposed test in detecting change points in risk measures in different scenarios;

• we apply the test on risk measures of the S&P500 index returns, and the detected change points can be associated with well-known market events.

(3) Our third set of original contributions, in terms of proposing a real-time

detection method for change points in risk models for VaR and ES is given below:

• we develop a new test to sequentially monitor change points in the M-estimators of semiparametric risk models for VaR and ES joint risk measures;

• we contribute to the current literature on sequential monitoring studies;

• we derive the asymptotic behavior of the test statistic;

• we apply the proposed sequential monitoring test to designed simulations;

• we consider changes in the second to fourth moments (variance, skewness and kurtosis) for the post-break process, respectively, in the simulation study;

• we show that this test has a good size control under the null hypothesis and high power under various alternative hypotheses in finite samples;

• we provide empirical applications to demonstrate the practical usage of our proposed test;

• we show that this test is able to detect change points associated with well-known market events and often the detection precedes the actual market crashes.

1.4   Outline of the Thesis

The rest of this thesis proceeds as follows: Chapter 2 studies the extended GAS models that incorporate intraday information for VaR and ES forecasting; Chapter 3 proposes a new test to detect change points in the (semi)parametric (VaR, ES) tuple; Chapter 4 develops a sequential monitoring test to detect structural changes in the M-estimators of risk model parameters. Chapter 5 summarizes the main findings and discusses further research that builds on the findings presented in this thesis.

For an improved reading experience, we make each chapter self-contained. We (re)introduce variables and abbreviations in each chapter. Whenever possible, we endeavour to follow consistent notations throughout this thesis.

Notes

[1]According to Artzner et al. (1999), a risk measure is *coherent* if it satisfies the following four properties: *translation invariance*, *sub-additivity*, *positive homogeneity* and *monotonicity*.

[2]There is no (strictly) consistent loss function for ES that does not also contain VaR (Fissler and Ziegel, 2016). ES is typically estimated or predicted jointly with VaR.

[3]The terminology "change point" has, in general, the same meaning as the "structural break" in econometrics.

[4]In this thesis, the nonparametric approaches for VaR and ES discussed above are not included in the central part of our study. However, we consider several of them as benchmarks when comparing model performance and leave them for robustness check for our proposed tests.

Chapter 2

Forecasting Risk Measures Using Intraday Data in a Generalized Autoregressive Score Framework[1]

2.1 Introduction

From the perspective of financial risk managers, a risk measure can be considered a map from the space of probability distributions to real numbers. Risk measures can provide banks and financial institutions with specific values of potential losses so that risk managers can adjust their capital reserves against the downside risk. Value-at-Risk (VaR) and Expected Shortfall (ES) are two prevailing measures of financial risk that dominate contemporary financial regulation. VaR provides banks and investment institutions with a loss level that occurs in the

worst situation at a given confidence level, and it can be defined as:

$$VaR_t^\alpha \equiv \inf\{r_t \in \mathbb{R} | F(r_t | \mathscr{F}_{t-1}) \geq \alpha\},$$

where $F(\cdot | \mathscr{F}_{t-1})$ is the cumulative distribution function of asset returns r_t over a horizon given the information set $\mathscr{F}_{t-1}$, and $\alpha \in (0, 1)$ is a given significance level. As a quantile, VaR can be expressed directly in terms of the inverse cumulative distribution function: $VaR_t^\alpha = F^{-1}(\alpha | \mathscr{F}_{t-1})$, and as a risk measure, it has the advantage of being intuitive and easily understood.

However, VaR has inherent deficiencies as it ignores the shape and structure of the tail and is not a coherent risk measure in the sense of Artzner et al. (1999). Thus, after the financial crisis of 2007-08, the Basel Committee on Banking Supervision has proposed a transition from VaR with a confidence level of 99% to ES with a confidence level of 97.5% (Basel Committee on Banking Supervision, 2013). ES is the expectation of returns, conditional on its realization lying below VaR, and it can be defined as:

$$ES_t^\alpha \equiv \mathbb{E}[r_t | r_t \leq VaR_t^\alpha, \mathscr{F}_{t-1}].$$

ES is a coherent risk measure (Roccioletti, 2015), and it has been suggested as an alternative to VaR in risk management applications due to its superior mathematical properties.

Normally, ES is estimated via a two-stage approach based on VaR estimation. Whilst ES is itself not elicitable, Fissler et al. (2016) have shown that the pair $(VaR_t^\alpha, ES_t^\alpha)$ is elicitable (see also Acerbi and Székely, 2014). This means that ES can be estimated jointly with VaR by minimizing a loss function (Ziegel, 2016; Fissler and Ziegel, 2016).

Following the classification of Engle and Manganelli (2004), models in the current literature on estimating and forecasting risk measures can be divided

into three main categories: parametric, nonparametric and semiparametric models. Previous studies using parametric models to predict VaR and ES assume that financial returns follow a certain distribution, such as the standard normal (Gaussian) distribution. In reality, however, it is hardly reasonable to make such strong assumptions. Nonparametric models do not make assumptions about the distribution of financial returns, and have the advantage of being model free. While it is not necessary for such models to make a distributional assumption, an inherent problem is the difficulty in finding the optimal size of the estimation window (Engle and Manganelli, 2004). Semiparametric models impose a parametric structure on the dynamics of VaR and ES through their relationship with lagged information, but require no assumptions on the conditional distribution of financial returns (Patton et al., 2019).

Quantile regression, as an approach for estimating risk measures, has only recently been considered: Engle and Manganelli (2004) extend the basic quantile regression model to conditional autoregressive value at risk (CAViaR) models; these models focus solely on the estimation of VaR, and it is not obvious how they can be used for ES estimation. In order to estimate ES jointly with VaR in a semiparametric framework, Taylor (2008) proposes conditional autoregressive expectile (CARE) models, based on a simple function of expectiles.[1] Following this, Taylor (2019) synthesizes the quantile regression with the maximum likelihood estimation based on an Asymmetric Laplace density proposed by Koenker and Machado (1999), and estimates VaR and ES jointly. A growing literature documents a significant improvement in VaR and ES estimation in a quantile regression framework (Halbleib and Pohlmeier, 2012; Žikeš and Baruník, 2014; Wang and Zhao, 2016; Bayer, 2018).

Following the results of Fissler and Ziegel (2016), Patton et al. (2019) present several novel dynamic models for the joint estimation of VaR and ES. Specifically, they propose four dynamic semiparametric models for VaR and ES, based on the generalized autoregressive score (GAS) framework introduced by Creal

et al. (2013). This model has been successfully applied in risk measures estimation (Patton et al., 2019); CDS spread modelling (Oh and Patton, 2018); systemic risk modelling (Cerrato et al., 2017; Eckernkemper, 2017; Bernardi and Catania, 2019); and high-frequency data modelling (Gorgi et al., 2018; Lucas and Opschoor, 2018).[2] However, no studies on risk measures incorporating realized volatilities into the GAS framework have been considered so far.[3] This prompted the research question of this chapter, namely whether adding intraday measures of volatility into the GAS framework improves the accuracy of joint VaR and ES forecasts.

The question whether intraday data can improve the predictive accuracy of risk measures has already been addressed by academics.[4] Several studies extend quantile regression methods and other semiparametric models by using information variables generated from high-frequency data.[5] Many realized volatility measures have been confirmed to perform efficiently. The realized volatility proposed by Andersen and Bollerslev (1998) and Alizadeh et al. (2002) is one of the most widely used intraday volatility measures. Inspired by Engle and Manganelli (2004), Fuertes and Olmo (2013) propose a conditional quantile forecast method combining an effective device to deal with the inter-daily/intra-daily information. Meng and Taylor (2018) extend the CAViaR model and the Quantile Regression HAR model with realized volatility, overnight return and intraday range. In terms of ES estimation, the CARE models of Taylor (2008) have been extended to allow intraday measures as explanatory variables (Gerlach and Chen, 2014; Gerlach and Wang, 2022; Gerlach and Chen, 2017; Gerlach and Wang, 2020).

While the improvement from adding intraday variables into a semiparametric framework has been widely documented, evidence on using the score-driven model as the framework to estimate risk measures still remains hard to come by. Therefore, in our study, the first contribution is that we extend the set of semiparametric GAS models of Patton et al. (2019): the two-factor GAS model, the one-factor GAS model, the GARCH-FZ model, and the hybrid GAS/GARCH

model, to investigate whether realized measures can improve the predictive accuracy of GAS models. This chapter is the first one to estimate and forecast VaR and ES jointly by using intraday data in a GAS framework. We shed light on the potential improvement in risk forecasting from adding intraday information in the GAS framework for four stock indices using a long forecasting period (that includes the financial crisis period). Then we perform a thorough analysis to compare our forecasts with those generated from prevailing benchmarks in the current literature. Our results show that incorporating intraday data into the GAS framework outperform other (VaR, ES) forecasts in most cases.

Thus, our second contribution to the literature is that we provide empirical evidence that semiparametric models enhanced with realized volatility measures outperform other benchmark models via various backtesting methods. Our proposed models, especially the GAS-2F model, extended with realized volatilities dominate other benchmarks consistently. Thirdly, we compare four different types of realized measures with regard to their forecasting ability for risk measures, when added to GAS models.

The chapter is structured as follows: Section 2.2 briefly introduces the new GAS models that incorporate intraday information; the data used in our empirical study and the in-sample estimation results are presented in Section 2.3; Section 2.4 presents the forecasting study and backtesting results; and finally, Section 2.5 concludes the chapter.

2.2 Models

2.2.1 GAS Models for VaR and ES

Several extensions of the GAS models introduced by Creal et al. (2013) are proposed in Patton et al. (2019), which can be estimated by minimizing the loss

function of Fissler and Ziegel (2016) called FZ0:

$$\ell^{FZ0}(r, v, e; \alpha) = -\frac{1}{\alpha e}\mathbf{1}\{r \le v\}(v - r) + \frac{v}{e} + \log(-e) - 1, \qquad (2.2.1)$$

where r denotes the daily return, v and e represent the values of VaR and ES, respectively, and $\mathbf{1}$ is an indicator function which returns 1 when $r \le v$ (i.e., the VaR is exceeded), otherwise it returns zero. Fissler and Ziegel (2016) show the joint elicitability for VaR and ES, despite the fact that ES was known to be not elicitable on its own. They introduce loss functions, which should be used for VaR and ES measures, because it allows for this pair to be jointly evaluated. Based on the article, Patton et al. (2019) introduce new semiparametric models and derive the consistency and asymptotic normality of the parameter estimators. Patton et al. (2019) propose four models: the two-factor GAS model, the one-factor GAS model, the GARCH-FZ model, and the hybrid GAS/GARCH model, to estimate VaR and ES jointly by minimizing the loss function FZ0. The key novelty in their framework is the use of the scaled score (that can be computed as the first order derivative of the objective function[6]) to drive the time variation in the target parameter. Patton et al. (2019) present a "news impact curve" to show the impact of past observations on current forecasts of VaR and ES through the score variable. When $r > v$, the realized returns do not affect the estimation. But when $r \le v$, forecasts of ES and VaR react to realized returns through the score variable. In the one-factor GAS, the GARCH-FZ and the hybrid GAS/GARCH models, we treat the ES as a constant multiple of the VaR. This condition is reasonable when the ES and VaR vary together. This is naturally implied by some of the models with the error term following a distribution in the location-scale family (with zero mean), for example, GARCH models with Normal or Student's t distributions. However, this condition can be relaxed by allowing a time-varying factor between VaR and ES, such as in Taylor (2019, 2022). The GAS-FZ models are specified as below:

(1.A) One-factor GAS model (GAS-1F):

$$v_t = a \exp\{\kappa_t\},$$

$$e_t = b \exp\{\kappa_t\}, \quad b < a < 0, \tag{2.2.2}$$

$$\kappa_t = \beta_0 + \beta_1 \kappa_{t-1} + \beta_2 H_{t-1}^{-1} s_{t-1},$$

where the score variable s_t is defined as:

$$s_t \equiv \frac{\partial \ell^{FZ0}(r_t, a \exp\{\kappa_t\}, b \exp\{\kappa_t\}; \alpha)}{\partial \kappa} = -\frac{1}{e_t}\left(\frac{1}{\alpha}\mathbf{1}\{r_t \le v_t\}r_t - e_t\right), \tag{2.2.3}$$

and the Hessian factor H_t is set to one for simplicity;

(1.B) Two-factor GAS model (GAS-2F):

$$\begin{bmatrix} v_t \\ e_t \end{bmatrix} = \mathbf{w} + \mathbf{B}\begin{bmatrix} v_{t-1} \\ e_{t-1} \end{bmatrix} + \mathbf{A}\begin{bmatrix} \lambda_{v,t-1} \\ \lambda_{e,t-1} \end{bmatrix}, \tag{2.2.4}$$

where $\mathbf{w}$ is a (2×1) vector, $\mathbf{A}$ is a (2×2) matrix, and $\mathbf{B}$ is defined as a diagonal matrix for parsimony, and

$$\lambda_{v,t} \equiv -v_t(\mathbf{1}\{r_t \le v_t\} - \alpha), \tag{2.2.5}$$

$$\lambda_{e,t} \equiv \frac{1}{\alpha}\mathbf{1}\{r_t \le v_t\}r_t - e_t; \tag{2.2.6}$$

(1.C) GARCH-FZ model (GARCH-FZ):

$$v_t = a \cdot \sigma_t,$$

$$e_t = b \cdot \sigma_t, \quad b < a < 0, \tag{2.2.7}$$

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2,$$

where σ_t^2 is the conditional variance and is assumed to follow a GARCH(1,1) process. The parameters of this model are estimated by minimizing the loss

function FZ0 in (2.2.1), instead of using (Q)MLE.

(1.D) A hybrid GAS/GARCH model (Hybrid):

$$v_t = a \exp\{\kappa_t\},$$

$$e_t = b \exp\{\kappa_t\}, \quad b < a < 0,$$

$$\kappa_t = \beta_0 + \beta_1 \kappa_{t-1} + \beta_2 \left(-\frac{1}{e_{t-1}} \left(\frac{1}{\alpha} \mathbf{1}\{r_{t-1} \le v_{t-1}\} r_{t-1} - e_{t-1} \right) \right) + \beta_3 \log |r_{t-1}|,$$

$$(2.2.8)$$

where the variable κ_t is the log-volatility, described by the one-day lagged log-volatility, score factor and the logarithm of absolute return.

2.2.2 Realized Measures

This section provides a brief introduction to various intraday realized measures (RM) used in this chapter. The most popular measure is the realized volatility (RV), defined as:

$$RV\Delta_t = \sqrt{\sum_{i=1}^{N} (P_{t,i\cdot\Delta} - P_{t,(i-1)\cdot\Delta})^2},$$

$$\Delta = \frac{S}{N},$$

$$(2.2.9)$$

where $RV\Delta_t$ denotes the realized volatility calculated from the sum of N intraday squared returns, at frequency Δ, within day t. Here, the intraday frequency Δ divides the whole span of market opening hours S into N equal intervals, and $P_{t,i\cdot\Delta}$ denotes the log price at time $i \cdot \Delta$ of day t. However, the realized volatility ignores the information from the market overnight return, which is defined as:

$$overnight_t = log(P_{t,0}) - log(P_{t-1,S}),$$

$$(2.2.10)$$

where $P_{t,0}$ and $P_{t-1,S}$ denote the opening price on day t and the closing price on the previous day, respectively. Several studies have proven that incorporating the overnight return can lead to a more accurate realized measure. In this chapter,

we consider the approach of incorporating the overnight return in the realized volatility of Blair et al. (2001), Hua and Manzan (2013) and Meng and Taylor (2018) as follows:

$$RN\Delta_t = \sqrt{RV\Delta_t^2 + (overnight_t)^2}. \tag{2.2.11}$$

In the following, we will use frequencies of $\Delta = 5min$ and $\Delta = 10min$. As such, in the next section, RM can signify any of the following four realized measures of volatility: $RV5_t$, $RV10_t$, $RN5_t$, and $RN10_t$, and we extend the models with these measures.

2.2.3 GAS Models for VaR and ES with Realized Measures

Salvatierra and Patton (2015) propose a GAS model enhanced with high frequency measures to obtain a GRAS model, which has the equation for the dependence parameter, similar to the last row of (2.2.2), replaced with:

$$\kappa_t = \beta_0 + \beta_1\kappa_{t-1} + \beta_2 H_{t-1}^{-1}s_{t-1} + c\log(RM_{t-1}). \tag{2.2.12}$$

They use the realized covariance as RM_t, computed from the intraday prices $P_{t,i\cdot\Delta}$ of a set of assets. The authors find that the inclusion of 5-minute realized covariance significantly improves the in-sample fit and out-of-sample forecasts of the copula models.

Motivated by the set of GAS models and the GRAS model, our new models are proposed as:

(2.A) One-factor GAS model with realized measures (GAS-1F-Re):

$$\begin{aligned} v_t &= a\exp\{\kappa_t\} \\ e_t &= b\exp\{\kappa_t\}, \quad b < a < 0, \end{aligned} \tag{2.2.13}$$

where κ_t is defined in (2.2.12), and the score variable s_t is defined in (2.2.3). Here, the Hessian factor H_t is set to one for simplicity; $\log(RM_t)$ is the logarithm of a realized measure which can be: the realized volatility at 5-min and 10-min sampling frequencies ($RV5$ and $RV10$), and these two realized volatilities with the overnight return incorporated into them ($RN5$ and $RN10$), as defined in Section 2.2. **(2.B) Two-factor GAS model with realized measures (GAS-2F-Re):**

$$\begin{bmatrix} v_t \\ e_t \end{bmatrix} = \mathbf{w} + \mathbf{B} \begin{bmatrix} v_{t-1} \\ e_{t-1} \end{bmatrix} + \mathbf{A} \begin{bmatrix} \lambda_{v,t-1} \\ \lambda_{e,t-1} \end{bmatrix} + \mathbf{C}\, RM_{t-1}, \qquad (2.2.14)$$

where $\mathbf{w}$ and $\mathbf{C}$ are (2×1) vectors, $\mathbf{A}$, and $\mathbf{B}$ are both (2×2) matrices, $\mathbf{B}$ is defined as a diagonal matrix to simplify computation. Following Patton et al. (2019), we also define the forcing variables $\lambda_{v,t}$ and $\lambda_{e,t}$ as the partial derivatives of the given loss function ℓ^{FZ0} with respect to v_t and e_t, as in (2.2.5) and (2.2.6).

Hansen et al. (2012) and Hansen et al. (2014) introduce a new framework, Realized (Beta) GARCH, where the variance follows a GARCH(1,1) process, with the squared returns replaced with a realized measure of volatility. Following this model, we propose a GARCH-FZ-Realized model:

(2.C) GARCH-FZ model with realized measures (GARCH-FZ-Re):

$$\begin{aligned} v_t &= a \cdot \sigma_t, \\ e_t &= b \cdot \sigma_t, \quad b < a < 0, \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + cRM_{t-1}^2, \end{aligned} \qquad (2.2.15)$$

where the daily return r_{t-1} in the GARCH(1,1) variance equation in (2.2.7) is replaced with the realized measure RM_{t-1}. This model is estimated by minimizing the FZ0 loss function.

(2.D) A hybrid GAS/GARCH model with realized measures (Hybrid-

Re):

$$v_t = a \exp\{\kappa_t\},$$

$$e_t = b \exp\{\kappa_t\}, \quad b < a < 0,$$

$$\kappa_t = \beta_0 + \beta_1 \kappa_{t-1} + \beta_2 \left(-\frac{1}{e_{t-1}} \left(\frac{1}{\alpha} \mathbf{1}\{r_t \leq v_t\} r_{t-1} - e_{t-1} \right) \right)$$

$$+ \beta_3 \log|r_{t-1}| + c \log(RM_{t-1}),$$

(2.2.16)

where the log-volatility κ_t follows the hybrid GARCH model with one-day lagged log-volatility, score factor, realized measures and absolute daily return.

2.3 Data and Empirical Study

2.3.1 Data Description

To evaluate the forecasting performance of the new models and to compare them with benchmark models, we collected daily opening and closing prices of four international stock market indices: the S&P 500 (US); Dow Jones Industrial Average (US); NIKKEI 225 (Japan) and FTSE 100 (UK), from January 2000 to June 2019, from DataStream. To ensure the applicability of every model, we remove market-specific non-trading days and exactly zero returns from each index series. Panel A in Table 2.3.1 presents the summary statistics on the four daily equity return series over the full sample period. From the top panel, average annualized returns range from 0.544% for the NIKKEI 225 to 4.377% for the DJIA, and the annualized standard deviation ranges from 18% for the DJIA to about 24% for the NIKKEI 225. All daily return series exhibit substantial kurtosis at around 10. The second and third panels of this table show the sample VaR and ES for four different α levels: 1%, 2.5%, 5% and 10%. The NIKKEI 225 index proves to be different from the rest since its quantile and ES are lower than the sample risk measures of the other three indices.

Panel B presents the estimated parameters of the ARMA(p,q) models where the lags (p,q) are optimally selected via the BIC method. The ARMA models for the indices only include a constant except for the S&P 500, which contains an MA term with one lag. Panel C shows the estimated parameters of the GARCH(1,1) model, where the residuals are assumed to follow the skewed t distribution. Panel D presents the parameters of the degree of freedom and skewness in the skewed t distribution.

The percentage log overnight returns are generated as in (2.2.10). For the realized volatility, the data is obtained at 5-min and 10-min sampling frequencies from the Oxford-Man Institute's realized library[7](see Heber et al., 2009). To generate the new realized measure incorporating the overnight return in realized volatility, we use (2.2.11).

The entire sample is divided into an in-sample for estimation and an out-of-sample to backtest the estimated results. We employ a rolling window approach, where each model is re-estimated every five trading days using a rolling window of 2000 observations. Then the rest of the period until June 2019 of approximately 2900 days, is the out-of-sample period to evaluate one-day ahead VaR and ES estimates.

2.3.2 Forecasting Models

VaR and ES are predicted via the score forecast for one trading day ahead in the out-of-sample period for each series, using the proposed GAS-Realized models and the FZ-GARCH-Realized model, as well as nonparametric models and parametric models as benchmarks. For nonparametric models, historical simulations are widely used because of their advantages of being model free and easy to implement. In our study, we select three commonly used rolling window sizes to forecast VaR and ES: 125, 250 and 500 days. Two popular GARCH models are employed in this chapter, including the Gaussian (GARCH-G) and skewed t

Table 2.3.1: Summary statistics and marginal distribution estimates

	S&P 500	DJIA	NIKKEI	FTSE
Panel A: Summary statistics				
Mean (Annualized)	3.685	4.377	0.544	0.606
Std dev (Annualized)	18.900	17.821	23.748	18.105
Skewness	-0.208	-0.125	-0.429	-0.170
Kurtosis	11.176	10.980	9.341	9.487
VaR-1%	-3.427	-3.294	-4.111	-3.264
VaR-2.5%	-2.525	-2.361	-3.051	-2.409
VaR-5%	-1.885	-1.777	-2.360	-1.788
VaR-10v	-1.284	-1.182	-1.682	-1.233
ES-1%	-4.849	-4.568	-6.021	-4.546
ES-2.5%	-3.678	-3.453	-4.492	-3.457
ES-5%	-2.922	-2.750	-3.576	-2.764
ES-10%	-2.236	-2.096	-2.788	-2.120
Panel B: Conditional mean				
Constant	-0.001	0.007	-0.021	-0.003
AR(1)	-	-	-	-
MA(1)	-0.039	-	-	-
Panel C: Conditional variance				
Constant	0.010	0.010	0.025	0.014
ARCH	0.065	0.069	0.082	0.116
GARCH	0.926	0.922	0.910	0.874
Panel D: skewed t density				
DoF	9.020	8.130	12.204	22.177
Skewness	-0.092	-0.089	-0.089	-0.162

Note: This table presents the summary statistics of the four daily equity return series studied, over the full sample period from January 2000 to June 2019, and marginal distribution estimates over the in-sample period. Panel A reports the annualized mean, standard deviation of these returns in percentages, skewness, kurtosis, the sample VaR and ES estimates for four choices of α; Panel B presents the parameter estimates for AR(m) models of the conditional means of these returns; Panel C shows parameter estimates for GARCH-skewed $t(1,1)$ models of the conditional variance; Panel D presents parameter estimates for the skewed t density for the standardized residuals.

(GARCH-Skt) models as parametric model benchmarks. We also consider other established models that use high-frequency data (i.e., 5-min realized volatility), considered to be well-suited to forecast VaR and ES: the HAR model of Corsi et al. (2008), and the HEAVY model of Shephard and Sheppard (2010). In each model, after the conditional volatility estimation, we estimate VaR and ES with Gaussian and skewed t distributions of the errors (HAR-Skt-RV5, HEAVY-G-RV5, HEAVY-Skt-RV5). We also take the semiparametric approach of Taylor (2019) based on the asymmetric Laplace distribution (AL-CAViaR-Sym), into our benchmark set.

To evaluate the performance of the GAS models enhanced with realized measures, we also implement the four models proposed by Patton et al. (2019) as benchmarks. Differently from Patton et al. (2019) who used certain parameters estimated from a fixed in-sample period, we use a rolling window approach, where each model is re-estimated every five trading days using a window of size 2000 trading days. In this chapter, we consider four sets of GAS models extended with different realized measures: RV5, RV10, RN5 and RN10 as in Section 2.2. In the following section, we will show estimation results in these proposed models.

2.3.3   In-sample Estimation

The parameters of the GAS models and the proposed four sets of GAS-Realized models are estimated by minimizing the loss function in (2.2.1). The existence of an indicator function in the FZ0 loss function necessitates the use of a numerical search algorithm, but this algorithm is sensitive to the starting values used in the search. We optimize the proposed models using the following procedure: for each model, we first generate 10^5 vectors of parameters from predetermined intervals randomly for the parameters of the GAS models. For example, for the parameters (a and b) used to generate VaR and ES in GAS-1F, GARCH-FZ, and Hybrid models, we set the intervals as [-2, -3] and [-3, -4], respectively, to

ensure that ES is always less than VaR.[8] We compute the average loss value for each vector, then select the 10 vectors that generate the lowest average loss as initial values for the optimization routine. The vectors are selected as the initial values of the search algorithm for all windows in order to shorten computational time. We compute the optimal parameters by using a quasi-Newton method and the function *fminunc* as optimization algorithms, which are similar routines to the one used by Engle and Manganelli (2004). Alternatively, one can use the maximum likelihood estimates of a closely matching parametric model as the set of starting values.

Table 2.3.2 presents the estimated parameters together with their standard errors of the GAS models for the S&P 500, estimated using an estimation period of 2000 days from the beginning of January 2000 for $\alpha = 5\%$. The parameters of the three two-factor GAS models (GAS-2F, GAS-2F-RV5, and GAS-2F-RN5) are presented in the first panel of Table 2.3.2; we separate the parameters of VaR and ES. The b parameters are statistically significantly different from zero at both 1% and 5% significance levels for both VaR and ES,[9] which can be explained by the volatility clustering effect. The four columns on the right side of this panel show the parameters of GAS-2F extended with the 5-minute realized measures. Due to adding 5-min realized measures, the degree of clustering decreases for VaR and ES. Also, the parameters of score a_v and a_e experience a significant decrease after adding the realized measures. The parameters of the one-day lagged realized measures RM_{t-1}, c, are statistically significantly negative at the 5% significance level for both VaR and ES, indicating that larger values of these realized variables will result in a lower estimated quantile or ES, which is intuitive. The average loss generated by the GAS-2F model is 0.756, which is larger than the loss of the GAS-2F models extended with realized measures (0.735 and 0.734).

The second panel in Table 2.3.2 shows the estimated parameters of the other GAS models extended with the 5-minute realized measures using an estimation

period of 2000 days from the beginning of January 2000 for the S&P 500, for $\alpha = 5\%$. Similarly to the b parameters of the GAS-2F models, the β_1 parameters of the other models are also statistically significantly different from zero at both 1% and 5% significance levels, which means that the current estimated risk measures rely heavily on the previous estimation. Also, we find that the parameters of realized measures (c for the GAS-1F model, the GARCH-FZ model, and the Hybrid model) are all statistically significantly positive at both 1% and 5% significance levels. Intuitively, a large realized volatility will lead to a low quantile through the score variable in these models. We obtain that the inclusion of realized measures in the updating models results in smaller coefficients of the GAS shocks (β_2), which is intuitive. Later, we will see the role that the score variable plays in forecasting VaR and ES. In the following sections we compare the forecasting performance of these four sets of extended models, which gives a total of 16 models, with the 13 benchmark models enlisted above.

2.4 Out-of-sample Forecasting and Backtesting

We evaluate one day-ahead VaR and ES forecasts for the four international stock indices, and for the following four probability levels: 1%, 2.5%, 5% and 10%. One-day ahead VaR and ES forecasts are made with parameter values estimated every 5 days, for each model and probability level, using rolling windows of size 2000 (except for historical simulations). The forecasting sample period for each index is approximately 2900 days. In this section, we backtest the VaR and ES forecasts of the proposed models and compare their performance with that of benchmark models. First, we backtest VaR and ES individually via the Dynamic Quantile (DQ) regression and the Dynamic Expected Shortfall (DES). Following these tests, we employ a method based on the FZ0 loss function to backtest VaR and ES jointly.

Table 2.3.2: The estimated parameters of the GAS models for the S&P 500 for $\alpha = 5\%$

	GAS-2F		GAS-2F-RV5		GAS-2F-RN5	
	VaR	ES	VaR	ES	VaR	ES
w	-0.009	-0.012	-0.009	-0.016	-0.011	-0.023
(s.e.)	(0.002)	(0.003)	(0.030)	(0.053)	(0.033)	(0.045)
b	0.995	0.995	0.833	0.810	0.814	0.849
(s.e.)	(0.105)	(0.108)	(0.084)	(0.092)	(0.098)	(0.072)
a_v	-0.129	-0.140	-0.125	-0.066	-0.114	-0.118
(s.e.)	(0.070)	(0.103)	(0.304)	(0.629)	(0.416)	(0.466)
a_e	0.002	0.003	0.002	0.001	0.001	0.001
(s.e.)	(0.003)	(0.004)	(0.011)	(0.024)	(0.015)	(0.017)
c	-	-	-0.323	-0.477	-0.353	-0.360
(s.e.)	-	-	(0.148)	(0.208)	(0.190)	(0.158)
Avg loss	0.756		0.735		0.733	

	GAS-1F	GCH-FZ	Hybrid	GAS-1F	GCH-FZ	Hybrid	GAS-1F	GCH-FZ	Hybrid
					5min RV			5min RN	
β_1	0.993	0.922	0.993	0.857	0.857	0.875	0.851	0.761	0.872
(s.e.)	(0.002)	(0.088)	(0.002)	(0.116)	(0.081)	(0.072)	(0.143)	(0.077)	(0.096)
β_2	0.008	0.032	0.008	0.004	-	0.004	0.004	-	0.004
(s.e.)	(0.001)	(0.007)	(0.001)	(0.009)	-	(0.007)	(0.013)	-	(0.011)
β_3	-	-	4.393e-08	-	-	0.010	-	-	0.009
(s.e.)	-	-	(1.552e-09)	-	-	(0.016)	-	-	(0.018)
c	-	-	-	0.127	0.095	0.141	0.133	0.084	0.142
(s.e.)	-	-	-	(0.013)	(0.012)	(0.056)	(0.016)	(0.009)	(0.051)
a	-1.774	-2.269	-1.752	-1.973	-2.818	-2.150	-1.962	-2.987	-2.053
(s.e.)	(4.451)	(0.393)	(5.726)	(2.529)	(0.410)	(2.160)	(3.422)	(0.430)	(2.294)
b	-2.401	-3.043	-2.355	-2.599	-3.610	-2.779	-2.601	-3.822	-2.709
(s.e.)	(5.987)	(0.765)	(7.709)	(3.310)	(0.670)	(2.819)	(4.467)	(0.672)	(3.029)
Avg loss	0.761	0.780	0.761	0.737	0.727	0.753	0.734	0.722	0.749

Note: This table presents the parameter estimates and standard errors of the four GAS models proposed in Patton et al. (2019) and eight GAS models enhanced with 5-min realized volatility (and overnight returns), for VaR and ES, for the S&P 500 index using the first rolling window of 2000 days starting with January 2000. The top panel presents the estimated parameters of the two-factor GAS models. The bottom panel presents the parameters of the one-factor GAS model, the GARCH model, and the hybrid-factor GAS model, estimated using the FZ0 loss minimization. The bottom row of each panel presents the average (in-sample) losses from these models.

2.4.1 Backtesting VaR

The most popular procedures evaluating the performance of VaR forecasts are mainly based on VaR exceptions, i.e.,

$$I_t = \mathbf{1}\{r_t \leq v_t^\alpha\}.$$

The commonly used VaR backtesting method, known as the unconditional coverage (UC) test, is proposed by Kupiec (1995), and uses the proportion of exceptions as its main tool. In this test, the hit percentage is defined as the proportion of the returns below the estimated VaR, then the difference between the hit percentage and its theoretical value of α is examined. Thus, the VaR model is rejected or not according to the null hypothesis of the UC test below, based on which the Likelihood Ratio (LR) test is performed:

$$H_{UC}^{VaR} : \mathbb{E}_{t-1}[I_t] = \alpha.$$

Table 2.4.1 presents the number of model rejections of the above null hypothesis for four daily equity return series, over the out-of-sample period, for the 29 different forecasting models, at 1% and 5% significance levels, respectively, and for different probability levels. To obtain these columns, we perform the unconditional backtest above for all indices, and count the number of rejections for each model.

The third and fourth columns of Table 2.4.1 show that the proposed new GAS models extended with realized measures generally tend to have a lower number of UC test rejections as compared to the number of rejections of the GAS-FZ models of Patton et al. (2019), for $\alpha = 1\%$. The GARCH model and HEAVY model with a skewed t distribution also tend to have a lower number of rejections at 1% significance level. At 5% significance level, several GAS-FZ models with overnight returns incorporated in the realized volatility have zero rejections of the

UC test. In general, adding realized measures into GAS models for predicting VaR achieves a lower number of test rejections, based on our results on the hit percentage test.

However, the UC test is statistically weak for small sample size, and is criticized by several studies (see Nieto and Ruiz, 2016) that it ignores the clustering of VaR exceptions. To address these drawbacks, the conditional coverage (CC) test is considered, in which the null hypothesis is:

$$H_{CC}^{VaR} : \mathbb{E}_{t-1}[I_t | I_{t-1}] = \alpha.$$

We employ the dynamic quantile (DQ) test proposed by Engle and Manganelli (2004) to implement the CC test. The DQ test has power against the misspecification of ignoring conditionally correlated probabilities and can be extended to examine other explanatory variables. The DQ test examines whether the hit variable defined as $Hit_{v,t} = \mathbf{1}\{r_t \le v_t^\alpha\} - \alpha$, follows an i.i.d. Bernoulli distribution with probability level α and whether it is independent of the VaR estimator; the expected value of $Hit_{v,t}$ is 0. Furthermore, from the definition of the quantile function, the conditional expectation of v_t^α given any information known at $t-1$ must also be 0, which means that the hit function cannot be correlated with other lagged variables. Also, the $Hit_{v,t}$ must not be autocorrelated. If $Hit_{v,t}$ satisfies the conditions stated above, then there will be no autocorrelation in the hits, and no measurement error. We include one lag of $Hit_{v,t}$ in the regression of the test. Consider the following DQ regression:

$$Hit_{v,t} = a_0 + a_1 Hit_{v,t-1} + a_2 v_{t-1} + u_{v,t}, \tag{2.4.1}$$

where $\mathbf{a} = [a_0, a_1, a_2]$ is the set of parameters of the regression. Based on the null hypothesis, we test whether all parameters in the set $\mathbf{a}$ are zero. Performing this DQ test gives a test statistic, which is distributed $\mathcal{X}^2(3)$ asymptotically.

Table 2.4.1: The number of model rejections based on hit percentages of VaR forecasts (UC test) for the four indices for different α levels

Number	Model	1% VaR		2.5% VaR		5% VaR		10% VaR	
		1%	5%	1%	5%	1%	5%	1%	5%
1	RW-125	3	3	**0**	**0**	**0**	**0**	**0**	**0**
2	RW-250	1	2	**0**	1	**0**	**0**	**0**	**0**
3	RW-500	**0**	2	1	1	**0**	1	**0**	**0**
4	GARCH-G	4	4	3	3	1	1	**0**	1
5	GARCH-Skt	**0**	1	**0**	3	**0**	**0**	**0**	**0**
6	HAR-Skt-RV5	4	4	4	4	4	4	4	4
7	HEAVY-G-RV5	4	4	4	4	**0**	3	**0**	**0**
8	HEAVY-Skt-RV5	**0**	1	**0**	**0**	**0**	**0**	**0**	**0**
9	AL-CAViaR-Sym	2	3	1	3	**0**	**0**	**0**	**0**
10	GAS-2F	3	3	2	2	**0**	**0**	1	2
11	GAS-1F	**0**	3	**0**	**0**	**0**	**0**	1	1
12	GARCH-FZ	1	2	1	3	**0**	**0**	**0**	1
13	Hybrid	2	2	**0**	1	**0**	**0**	1	1
14	GAS-2F-RV5	**0**	1	1	1	1	1	1	1
15	GAS-1F-RV5	**0**	1	**0**	1	**0**	1	**0**	**0**
16	GARCH-FZ-RV5	**0**	1	**0**	1	**0**	**0**	**0**	**0**
17	Hybrid-RV5	2	3	**0**	1	**0**	**0**	**0**	**0**
18	GAS-2F-RV10	1	1	1	1	1	1	1	1
19	GAS-1F-RV10	**0**	2	1	1	**0**	1	**0**	**0**
20	GARCH-FZ-RV10	1	1	1	1	**0**	**0**	**0**	**0**
21	Hybrid-RV10	2	3	1	1	**0**	**0**	**0**	1
22	GAS-2F-RN5	2	3	**0**	1	**0**	**0**	**0**	**0**
23	GAS-1F-RN5	**0**	1	**0**	**0**	**0**	**0**	**0**	1
24	GARCH-FZ-RN5	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
25	Hybrid-RN5	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1
26	GAS-2F-RN10	**0**	1	**0**	**0**	**0**	**0**	**0**	**0**
27	GAS-1F-RN10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1
28	GARCH-FZ-RN10	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
29	Hybrid-RN10	**0**	1	**0**	**0**	**0**	**0**	1	1

Note: This table presents the number of model rejections based on hit percentages of VaR forecasts (UC test) for the four daily equity return series, over the out-of-sample period, for 29 different forecasting models. The first three rows (Models 1-3) correspond to rolling window historical forecasts, the next two rows (Models 4 and 5) correspond to GARCH forecasts based on different distributions for the standardized residuals, the next four rows (Models 6-9) correspond to forecasts using high-frequency data and the CAViaR model based on the asymmetric Laplace distribution. The next four rows (Models 10-13) correspond to GAS models proposed by Patton et al. (2019). The last 16 rows (Models 14-29) correspond to the GAS models extended with the 5-min and 10-min realized measures, respectively.

The middle panel of Table 2.4.2 shows the p-values of the DQ test of VaR forecasts for $\alpha = 1\%$, for the four stock indices. The p-values that are greater than 5% indicate no evidence against the optimality at 5% significance level (in bold), and values between 1% and 5% are in italics. For the S&P 500, all of our newly proposed models pass the DQ test at 1% significance level. When we consider the NIKKEI 225 and FTSE 100 index, we see significant improvements after adding realized measures in the GAS models. For the DJIA index, using realized measures we obtain that fewer models fail the DQ test, while the historical simulations pass the test and the GARCH model with the skewed t distribution performs well. But for this index, all of the GAS-1F models extended with realized measures are able to pass the DQ test for all four indices. Overall, adding realized measures enables GAS-FZ models to reduce the number of rejections of the DQ test for $\alpha = 1\%$.

For $\alpha = 2.5\%$ (see Table 2.4.3), we obtain similar results, namely that adding realized measures generally reduces the number of rejections of the DQ test. For the DJIA index, the two-factor GAS model can pass the test after adding realized measures RN5 and RN10. For $\alpha = 5\%$, in Table 2.4.4, we can see that all original GAS-FZ models can pass the DQ test across the four indices except the Hybrid model for the S&P 500 index. After adding realized measures in the GAS models, it can be seen that the p-values increase and the DQ test is generally passed. Table 2.4.5 presents the number of model rejections at 1% and 5% significance levels for quantile regression VaR backtests across the four markets, for different probability levels. It can be concluded that the set of GAS models extended with realized measures tend to have a lower number of rejections than the original GAS models and several other benchmarks. It should be noted that the four GAS-1F model extended with different realized measures have the least number of rejections of the DQ test, especially for low values of α.

Table 2.4.2: Out-of-sample average losses and dynamic regression tests (α=1%) for the (VaR,ES) forecasts

	Average Loss				DQ test (VaR) p-values				DES test (ES) p-values			
	S&P	DJIA	NIK	FTSE	S&P	DJIA	NIK	FTSE	S&P	DJIA	NIK	FTSE
RW-125	1.479	1.400	1.864	1.298	**0.063**	**0.109**	*0.017*	**0.087**	*0.032*	**0.056**	0.008	**0.082**
RW-250	1.522	1.473	1.928	1.377	0.350	0.302	*0.042*	*0.043*	0.255	0.204	*0.024*	**0.075**
RW-500	1.633	1.550	1.998	1.464	0.128	0.159	*0.017*	*0.028*	0.170	0.162	*0.049*	0.058
GARCH-G	1.380	1.246	1.636	1.190	0.001	0.004	*0.031*	0.000	0.000	0.001	*0.012*	0.000
GARCH-Skt	1.246	1.128	1.565	1.105	*0.043*	0.114	0.550	0.265	*0.036*	*0.049*	0.433	0.268
HAR-Skt-RV5	1.306	1.118	2.735	1.132	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
HEAVY-G-RV5	1.233	1.164	1.609	1.137	0.000	0.000	0.003	0.000	0.000	0.000	0.001	0.000
HEAVY-Skt-RV5	1.117	1.047	**1.507**	1.065	0.063	*0.021*	0.414	0.145	0.053	*0.028*	0.310	0.166
AL-CAViaR-Sym	1.306	1.158	1.529	1.102	0.004	0.095	0.255	0.296	0.007	0.131	0.182	0.314
GAS-2F	1.244	1.260	1.670	1.217	*0.035*	**0.075**	0.493	0.001	*0.059*	*0.044*	*0.409*	*0.011*
GAS-1F	1.222	1.184	1.650	1.184	0.209	0.219	*0.019*	0.629	0.371	0.274	*0.025*	0.832
GARCH-FZ	1.241	1.147	1.521	1.088	*0.044*	0.230	0.260	0.000	0.055	0.331	0.288	0.000
Hybrid	1.242	1.140	1.533	1.180	0.457	0.146	0.235	0.087	0.367	0.360	0.206	0.098
GAS-2F-RV5	*1.106*	1.016	1.562	1.080	0.687	0.000	0.121	0.000	0.675	0.000	0.198	0.000
GAS-1F-RV5	1.109	1.008	1.521	1.060	0.304	0.252	0.403	0.706	0.242	0.283	0.328	0.610
GARCH-FZ-RV5	1.118	1.031	1.518	1.113	0.349	0.525	0.296	0.112	0.253	0.268	0.251	0.089
Hybrid-RV5	1.141	1.087	1.575	1.069	0.253	0.000	*0.033*	0.324	0.205	0.000	*0.036*	0.364
GAS-2F-RV10	1.121	1.015	1.610	1.066	0.685	0.000	0.000	0.671	0.812	0.000	0.005	0.720
GAS-1F-RV10	1.117	1.024	1.557	1.071	0.239	0.450	0.327	0.538	0.212	0.391	0.248	0.510
GARCH-FZ-RV10	1.116	1.052	1.534	1.104	0.496	0.830	0.140	0.078	0.391	0.692	0.137	0.064
Hybrid-RV10	1.131	1.097	1.617	**1.054**	0.126	0.000	0.000	0.868	0.129	0.000	0.001	0.836
GAS-2F-RN5	1.165	*1.001*	1.553	1.076	*0.028*	0.000	0.006	0.499	*0.019*	0.000	*0.023*	0.703
GAS-1F-RN5	1.109	**0.995**	*1.518*	1.063	0.295	0.429	0.164	0.243	0.262	0.445	0.190	0.223
GARCH-FZ-RN5	1.123	1.012	1.598	1.109	0.250	0.000	0.659	0.058	0.157	0.000	0.435	*0.050*
Hybrid-RN5	1.118	1.026	1.582	1.071	0.319	0.000	0.286	0.286	0.237	0.000	0.262	0.269
GAS-2F-RN10	1.133	1.005	1.565	1.065	0.193	0.502	0.258	0.377	0.225	0.555	0.308	0.486
GAS-1F-RN10	**1.102**	1.014	1.586	*1.060*	0.790	0.696	0.548	0.340	0.717	0.702	0.371	0.334
GARCH-FZ-RN10	1.123	1.021	1.620	1.113	0.697	0.000	0.192	0.115	0.570	0.000	0.093	0.080
Hybrid-RN10	1.118	1.031	1.549	1.062	0.261	0.000	0.382	0.818	0.274	0.000	0.354	0.784

Note: The left panel of this table presents the average losses, using the FZ0 loss function, for the four daily equity return series, over the out-of-sample period for α=1%. The lowest average loss in each column is highlighted in bold, the second-lowest is highlighted in italics. The middle and right panels of this table present p-values from dynamic regression tests for the VaR and ES forecasts respectively. Values that are greater than 0.05 (indicating no evidence against optimality at the 0.05 level) are in bold, and values between 0.01 and 0.05 are in italics.

Table 2.4.3: Out-of-sample average losses and dynamic regression tests (α=2.5%) for the (VaR,ES) forecasts

	Average Loss				DQ test (VaR) p-values				DES test (ES) p-values			
	S&P	**DJIA**	**NIK**	**FTSE**	**S&P**	**DJIA**	**NIK**	**FTSE**	**S&P**	**DJIA**	**NIK**	**FTSE**
RW-125	1.198	1.120	1.522	1.063	**0.147**	*0.014*	**0.067**	*0.013*	**0.113**	*0.025*	**0.069**	*0.036*
RW-250	1.238	1.167	1.550	1.128	*0.025*	**0.059**	*0.024*	*0.030*	**0.145**	**0.212**	**0.054**	**0.110**
RW-500	1.347	1.281	1.623	1.235	0.001	0.005	0.006	0.000	*0.018*	*0.018*	*0.025*	*0.020*
GARCH-G	1.080	0.982	1.341	0.989	*0.026*	*0.028*	**0.305**	0.000	0.003	0.005	**0.086**	0.000
GARCH-Skt	1.034	0.942	1.320	0.950	**0.179**	**0.215**	**0.794**	**0.095**	**0.128**	**0.234**	**0.551**	**0.088**
HAR-Skt-RV5	1.044	0.925	2.053	0.959	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.002
HEAVY-G-RV5	0.964	0.898	1.327	0.946	*0.018*	*0.016*	*0.016*	*0.021*	0.002	0.003	0.003	0.006
HEAVY-Skt-RV5	0.926	0.863	**1.291**	0.918	**0.253**	**0.274**	**0.235**	**0.095**	**0.127**	**0.158**	**0.164**	**0.091**
AL-CAViaR-Sym	1.064	0.957	1.311	0.945	**0.075**	**0.172**	**0.530**	**0.067**	*0.035*	**0.235**	**0.434**	**0.072**
GAS-2F	1.057	1.001	1.414	0.979	**0.383**	0.004	**0.103**	0.004	**0.178**	*0.032*	*0.017*	*0.014*
GAS-1F	1.041	0.971	1.356	0.970	**0.765**	**0.316**	**0.073**	**0.873**	**0.654**	**0.284**	**0.083**	**0.899**
GARCH-FZ	1.033	0.929	1.308	0.956	**0.091**	**0.201**	**0.468**	**0.084**	**0.067**	**0.264**	**0.387**	**0.070**
Hybrid	1.020	0.943	1.300	0.954	**0.651**	**0.305**	**0.348**	**0.166**	**0.478**	**0.533**	**0.358**	**0.138**
GAS-2F-RV5	0.947	0.851	1.331	0.909	**0.378**	**0.356**	0.001	**0.886**	**0.740**	**0.360**	0.000	**0.977**
GAS-1F-RV5	0.936	0.844	1.319	0.913	**0.517**	**0.519**	**0.207**	**0.152**	**0.434**	**0.543**	**0.139**	**0.194**
GARCH-FZ-RV5	**0.924**	0.855	*1.297*	0.919	**0.785**	**0.642**	**0.185**	**0.193**	**0.605**	**0.407**	**0.123**	**0.185**
Hybrid-RV5	0.950	0.871	1.315	0.912	**0.419**	**0.338**	**0.185**	**0.862**	**0.367**	**0.343**	**0.147**	**0.896**
GAS-2F-RV10	0.934	0.846	1.338	*0.908*	**0.528**	**0.772**	0.000	**0.792**	**0.800**	**0.722**	0.000	**0.789**
GAS-1F-RV10	0.934	0.869	1.305	0.914	**0.174**	**0.491**	**0.154**	**0.733**	**0.237**	**0.489**	**0.101**	**0.621**
GARCH-FZ-RV10	0.931	0.856	1.311	0.915	**0.795**	**0.695**	**0.103**	**0.130**	**0.631**	**0.465**	**0.075**	**0.120**
Hybrid-RV10	0.946	0.883	1.306	0.913	**0.339**	**0.714**	**0.137**	**0.775**	**0.416**	**0.704**	**0.112**	**0.756**
GAS-2F-RN5	0.942	0.845	1.311	0.910	**0.085**	**0.362**	**0.411**	**0.751**	**0.206**	**0.335**	**0.171**	**0.787**
GAS-1F-RN5	0.939	0.843	1.320	0.914	**0.419**	**0.536**	**0.588**	**0.717**	**0.449**	**0.559**	**0.395**	**0.650**
GARCH-FZ-RN5	*0.925*	0.844	1.330	0.917	**0.816**	**0.876**	**0.738**	**0.224**	**0.696**	**0.659**	**0.516**	**0.220**
Hybrid-RN5	0.942	0.870	1.305	0.912	**0.229**	**0.679**	**0.571**	**0.814**	**0.306**	**0.654**	**0.406**	**0.804**
GAS-2F-RN10	0.937	**0.831**	1.305	**0.907**	*0.029*	**0.804**	**0.429**	**0.855**	**0.116**	**0.839**	**0.210**	**0.824**
GAS-1F-RN10	0.929	0.845	1.318	0.911	**0.391**	**0.493**	**0.730**	**0.233**	**0.402**	**0.508**	**0.506**	**0.245**
GARCH-FZ-RN10	0.930	*0.840*	1.330	0.913	**0.810**	**0.860**	**0.793**	**0.120**	**0.737**	**0.721**	**0.542**	**0.118**
Hybrid-RN10	0.938	0.881	1.305	0.914	**0.286**	**0.401**	**0.644**	**0.545**	**0.381**	**0.457**	**0.452**	**0.438**

Note: The left panel of this table presents the average losses, using the FZ0 loss function, for the four daily equity return series, over the out-of-sample period for α=2.5%. The lowest average loss in each column is highlighted in bold, the second-lowest is highlighted in italics. The middle and right panels of this table present *p*-values from dynamic regression tests for the VaR and ES forecasts respectively. Values that are greater than 0.05 (indicating no evidence against optimality at the 0.05 level) are in bold, and values between 0.01 and 0.05 are in italics.

Table 2.4.4: Out-of-sample average losses and dynamic regression tests (α=5%) for the (VaR,ES) forecasts

	Average Loss				DQ test (VaR) p-values				DES test (ES) p-values			
	S&P	DJIA	NIK	FTSE	S&P	DJIA	NIK	FTSE	S&P	DJIA	NIK	FTSE
RW-125	0.977	0.894	1.282	0.876	0.008	**0.075**	0.000	0.002	**0.074**	**0.200**	*0.045*	*0.014*
RW-250	1.011	0.950	1.288	0.931	0.008	**0.065**	**0.072**	0.001	**0.093**	**0.213**	**0.113**	0.007
RW-500	1.104	1.058	1.348	0.993	0.003	0.001	0.000	0.000	0.006	0.003	0.004	0.001
GARCH-G	0.849	0.775	1.142	0.808	**0.715**	**0.840**	**0.949**	*0.024*	**0.243**	**0.273**	**0.448**	0.004
GARCH-Skt	0.836	0.764	1.135	0.794	**0.857**	**0.968**	**0.979**	**0.255**	**0.722**	**0.738**	**0.830**	**0.201**
HAR-Skt-RV5	0.826	0.733	1.613	0.791	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.002
HEAVY-G-RV5	0.755	0.698	1.123	0.779	**0.477**	**0.265**	**0.299**	**0.144**	**0.079**	*0.046*	*0.048*	*0.032*
HEAVY-Skt-RV5	**0.743**	0.686	1.110	0.768	**0.624**	**0.548**	**0.582**	**0.432**	**0.497**	**0.490**	**0.344**	**0.364**
AL-CAViaR-Sym	0.854	0.770	1.133	0.794	**0.573**	**0.452**	**0.959**	**0.125**	**0.221**	**0.367**	**0.692**	**0.090**
GAS-2F	0.861	0.801	1.151	0.796	**0.624**	**0.455**	**0.703**	**0.264**	**0.206**	**0.200**	**0.562**	**0.456**
GAS-1F	0.848	0.782	1.140	0.786	**0.059**	**0.280**	**0.131**	**0.560**	*0.028*	**0.159**	**0.121**	**0.659**
GARCH-FZ	0.839	0.762	1.134	0.793	**0.441**	**0.456**	**0.973**	**0.255**	**0.388**	**0.417**	**0.732**	**0.218**
Hybrid	0.853	0.770	1.113	0.794	*0.046*	**0.108**	**0.986**	**0.507**	*0.016*	**0.155**	**0.897**	**0.319**
GAS-2F-RV5	*0.744*	0.679	1.120	*0.764*	**0.237**	**0.503**	0.002	**0.782**	**0.136**	**0.580**	0.000	**0.934**
GAS-1F-RV5	*0.744*	0.684	1.113	0.769	**0.779**	**0.945**	**0.286**	**0.177**	**0.822**	**0.908**	**0.174**	**0.275**
GARCH-FZ-RV5	0.746	0.689	**1.109**	0.771	**0.931**	**0.914**	**0.617**	**0.538**	**0.957**	**0.718**	**0.322**	**0.417**
Hybrid-RV5	0.765	0.693	1.118	0.767	**0.677**	**0.801**	**0.424**	**0.474**	**0.937**	**0.784**	**0.258**	**0.577**
GAS-2F-RV10	0.747	*0.674*	1.119	0.766	**0.078**	**0.520**	0.004	**0.410**	**0.063**	**0.593**	0.001	**0.666**
GAS-1F-RV10	0.751	0.683	*1.109*	0.774	**0.589**	**0.909**	**0.230**	**0.374**	**0.786**	**0.863**	**0.156**	**0.314**
GARCH-FZ-RV10	0.750	0.689	1.123	0.773	**0.808**	**0.821**	**0.655**	**0.386**	**0.857**	**0.743**	**0.294**	**0.265**
Hybrid-RV10	0.754	0.696	1.108	0.767	**0.561**	**0.578**	**0.435**	**0.349**	**0.750**	**0.887**	**0.274**	**0.563**
GAS-2F-RN5	0.749	**0.671**	1.116	0.767	**0.317**	**0.755**	**0.921**	**0.648**	**0.221**	**0.770**	**0.660**	**0.573**
GAS-1F-RN5	0.747	0.681	1.123	0.766	**0.641**	**0.838**	**0.983**	**0.424**	**0.904**	**0.824**	**0.866**	**0.454**
GARCH-FZ-RN5	0.745	0.679	1.137	0.770	**0.788**	**0.779**	**0.972**	**0.492**	**0.887**	**0.822**	**0.780**	**0.394**
Hybrid-RN5	0.753	0.702	1.117	0.771	**0.668**	**0.820**	**0.998**	**0.445**	**0.923**	**0.813**	**0.815**	**0.599**
GAS-2F-RN10	0.748	*0.674*	1.118	**0.763**	**0.082**	**0.261**	**0.904**	**0.611**	**0.087**	**0.370**	**0.644**	**0.862**
GAS-1F-RN10	0.751	0.682	1.115	0.769	**0.596**	**0.775**	**0.856**	**0.384**	**0.923**	**0.830**	**0.966**	**0.379**
GARCH-FZ-RN10	0.746	0.676	1.134	0.771	**0.729**	**0.621**	**0.990**	**0.378**	**0.923**	**0.826**	**0.900**	**0.277**
Hybrid-RN10	0.765	0.709	1.112	0.768	**0.272**	**0.563**	**0.546**	**0.428**	**0.777**	**0.837**	**0.975**	**0.626**

Note: The left panel of this table presents the average losses, using the FZ0 loss function, for the four daily equity return series, over the out-of-sample period for α=5%. The lowest average loss in each column is highlighted in bold, the second-lowest is highlighted in italics. The middle and right panels of this table present *p*-values from dynamic regression tests for the VaR and ES forecasts respectively. Values that are greater than 0.05 (indicating no evidence against optimality at the 0.05 level) are in bold, and values between 0.01 and 0.05 are in italics.

Table 2.4.5: Rejections at 1% and 5% significance levels for dynamic quantile and ES regression backtests across the four markets

	α = 1%				α = 2.5%				α = 5%				α = 10%			
	VaR		ES		VaR		ES		VaR		ES		VaR		ES	
	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
RW-125	0	1	1	2	0	2	0	2	3	3	0	2	3	4	1	2
RW-250	0	2	0	1	0	3	0	0	2	2	1	1	1	3	1	2
RW-500	0	2	0	1	4	4	0	4	4	4	4	4	3	4	4	4
GARCH-G	3	4	3	4	1	3	3	3	0	1	1	1	0	2	0	0
GARCH-Skt	0	1	0	2	0	0	0	0	0	0	0	0	0	1	0	0
HAR-Skt-RV5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
HEAVY-G-RV5	4	4	4	4	0	4	4	4	0	0	0	3	0	0	0	0
HEAVY-Skt-RV5	0	1	0	1	0	0	0	0	0	0	0	0	0	2	0	1
AL-CAViaR-Sym	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0
GAS-2F	1	2	0	2	2	2	0	3	0	0	0	0	0	1	1	1
GAS-1F	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1	1
GARCH-FZ	1	2	1	1	0	0	0	0	0	1	0	0	1	2	0	0
Hybrid	0	0	0	0	0	0	0	0	0	1	0	1	2	2	0	1
GAS-2F-RV5	2	2	2	2	1	1	1	1	1	1	1	1	2	3	1	3
GAS-1F-RV5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
GARCH-FZ-RV5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Hybrid-RV5	1	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0
GAS-2F-RV10	2	2	2	2	1	1	1	1	1	1	1	1	1	3	3	3
GAS-1F-RV10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
GARCH-FZ-RV10	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Hybrid-RV10	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
GAS-2F-RN5	2	3	1	3	0	0	0	0	0	0	0	0	2	2	0	2
GAS-1F-RN5	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
GARCH-FZ-RN5	1	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1
Hybrid-RN5	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
GAS-2F-RN10	0	0	0	0	0	1	0	0	0	0	0	0	2	2	1	2
GAS-1F-RN10	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1
GARCH-FZ-RN10	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1
Hybrid-RN10	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0

Note: This table presents the number of data series (out of 4) with model rejections at 1% and 5% significance levels for the DQ (VaR) and DES (ES) backtests across the four markets, for four probability levels. Smaller numbers of model rejections are preferable (values of 0 in bold).

2.4.2 Backtesting ES

All models that we consider produce both VaR and ES forecasts. From an economic point of view, for example, when we compare the 2.5% ES forecasts of the GAS-1F-RV5 and the 2.5% ES forecasts of the GAS-1F, the first one has, on average, an ES forecast lower with 13.29% (S&P 500), 17.49% (DJIA), 8.40% (NIKKEI), and 5.31% (FTSE 100). The results indicates that ignoring realized measures overestimates risk on average. Looking at the significance of these values, we follow the backtesting method of Patton et al. (2019) to evaluate the ES estimates individually, using a dynamic ES (DES) regression test:

$$\lambda_{e,t}^s = b_0 + b_1 \lambda_{e,t-1}^s + b_2 e_{t-1} + u_{e,t}, \tag{2.4.2}$$

where $\lambda_{e,t}^s$ is the standardized version of $\lambda_{e,t}$ defined in (2.2.6) ($\lambda_{e,t}^s = \frac{\lambda_{e,t}}{e_t} = \frac{1}{\alpha}\mathbf{1}\{r_t \leq v_t\}\frac{r_t}{e_t} - 1$), and $\mathbf{b} = [b_0, b_1, b_2]$ is the set of parameters of the regression. Based on the null hypothesis, we test whether all parameters in set $\mathbf{b}$ are zero.

The right panel of Table 2.4.2 shows the p-values from the DES test of the ES forecasts for $\alpha = 1\%$, for the four stock indices. Similarly to the result of the DQ test, incorporating the realized measure RN10 in GAS models seems to reduce the number of backtest rejections for the NIKKEI 225 and the FTSE 100 indices. GAS-1F models with realized measures can pass the DES test at 5% significance level for all indices, which is consistent with the result of the DQ test. The two-factor GAS model, after adding the risk measure RN10, passes the DES test for all indices. Almost all of our new models pass the DES test across the four indices for $\alpha = 2.5\%$, except the GAS-2F for the NIKKEI 225, as can be seen in the right panel of Table 2.4.3. Table 2.4.4 presents similar results across four indices using an α of 5%, whilst some benchmarks also have p-values higher than 5%, for example, the HEAVY model with a skewed t distribution. Table 2.4.5 summarizes the total number of model rejections at 1% and 5% significance levels

for the Dynamic ES regression backtests, across the four markets, for different probability levels. The GAS-1F models enhanced with realized measures have the smallest number of backtest rejections.

2.4.3 Joint Backtesting of the (VaR, ES) Risk Measures

In order to compare jointly the VaR and ES forecasts generated by different models, in this section, a loss function proposed in Fissler and Ziegel (2016) is employed. The authors discuss how VaR and ES are jointly elicitable and present a group of loss functions for risk measure estimation and backtesting. We follow the choice of Patton et al. (2019) for the loss function FZ0, as defined in (2.2.1). To compare the performance of each model using the FZ0 loss function, we calculate the average loss value

$$L_M^{FZ0} = \frac{1}{M} \sum_{1 \leq t \leq M} \ell_t^{FZ0}$$

for different α values across the four indices.

The left panel of Table 2.4.2 presents the average losses for the four equity return series, over the out-of-sample period, for 13 different benchmark forecasting models and 16 newly proposed models that use the 5-min and 10-min realized measures. The lowest average loss in each column is highlighted in bold, whilst the second lowest is highlighted in italics. For $\alpha = 1\%$, the GAS-FZ models enhanced with the realized volatility using overnight returns and the HEAVY-Skt model perform well, overall.

For $\alpha = 2.5\%$ (see Table 2.4.3), the GAS-2F model employing the 10-min realized volatility and overnight returns (GAS-2F-RN10) outperforms the other models, with lower loss than most other models for most series and being consistently ranked well, being the best model for the DJIA and FTSE 100 index. In Table 2.4.4 ($\alpha = 5\%$), the GAS-2F-RN5 and GAS-2F-RN10 models outperform the other models with the lowest loss for the DJIA and the FTSE 100 index, respectively. The HEAVY-Skt model has the lowest loss value for the S&P 500.

Table 2.4.6 presents the rankings (with the best performing model ranked 1 and the worst ranked 29) based on average losses using the FZ0 loss function, for the four index return series, over the out-of-sample period, for the 29 different forecasting models. The last two columns in each panel represent the average rank across the four series and the rank of the average, respectively. For $\alpha = 1\%$, the best-performing model is the GAS-1F model with the 5-min realized volatility and overnight returns, followed by the GAS-1F models extended with the other two realized measures. Considering $\alpha = 2.5\%$, the GAS-2F-RN10, GARCH-FZ-RV5, and GAS-1F-RN10 are the three models having the lowest average loss values. For $\alpha = 5\%$ and $\alpha = 10\%$, our proposed models have a relatively higher rank than the benchmarks, except the HEAVY model with a skewed t distribution, which is ranked second for $\alpha = 5\%$.

Another observation here is that the losses generated from the GAS-FZ models with realized measures are generally lower than the loss generated from most benchmark approaches. However, the HEAVY-Skt is always one of best 5 models considered in the overall ranking for all four probability levels. This suggests that the variables extracted from intraday data provide useful information for risk measure forecasting.

In order to analyse the relative performance of each model, we employ the Diebold-Mariano (DM) test to compare any two models using differences in average losses. In this chapter, t-statistics from the DM test compare the average losses, using the FZ0 loss function, for four indices, and for different probability levels, over the out-of-sample period. A negative t-statistic indicates that the row model outperforms the column model with a significant loss difference. The absolute values greater than 1.96 (2.575 or 1.64) indicate that the average loss difference is significantly different from zero at 95% (99% or 90%) significance level. In Figure 2.4.1, we present the results for the S&P 500 with the null hypothesis that the row model and the column model have equal values for the loss function. The numbering of the models used in the figure is given in the first column of

Table 2.4.6: Out-of-sample performance rankings for various levels of α

	$\alpha = 1\%$						$\alpha = 2.5\%$					
	S&P	DJIA	NIK	FTSE	Avg	Rank	S&P	DJIA	NIK	FTSE	Avg	Rank
RW-125	27	27	26	27	26.8	27	27	27	26	27	26.8	27
RW-250	28	28	27	28	27.8	28	28	28	27	28	27.8	28
RW-500	29	29	28	29	28.8	29	29	29	28	29	28.8	29
GARCH-G	26	25	23	25	24.8	25	26	25	23	26	25.0	25
GARCH-Skt	23	19	14	17	18.3	21	21	21	17	20	19.8	22
HAR-Skt-RV5	24	18	29	21	23.0	24	23	19	29	23	23.5	23
HEAVY-G-RV5	19	23	19	22	20.8	22	18	18	18	19	18.3	20
HEAVY-Skt-RV5	7	14	1	7	7.3	4	**3**	12	**1**	16	8.0	4
AL-CAViaR-Sym	25	22	6	15	17.0	19	25	23	11	18	19.3	21
GAS-2F	22	26	25	26	24.8	26	24	26	25	25	25.0	26
GAS-1F	18	24	24	24	22.5	23	22	24	24	24	23.5	24
GARCH-FZ	20	21	4	14	14.8	17	20	20	9	22	17.8	19
Hybrid	21	20	7	23	17.8	20	19	22	**3**	21	16.3	18
GAS-2F-RV5	**2**	8	12	13	8.8	5	16	9	21	**3**	12.3	15
GAS-1F-RV5	**3**	4	5	**3**	3.8	**2**	9	4	15	9	9.3	8
GARCH-FZ-RV5	8	12	**3**	19	10.5	10	**1**	10	**2**	17	7.5	**2**
Hybrid-RV5	16	16	15	9	14.0	16	17	15	13	7	13.0	17
GAS-2F-RV10	11	7	20	8	11.5	12	8	8	22	**2**	10.0	10
GAS-1F-RV10	6	10	11	11	9.5	8	7	13	4	12	9.0	7
GARCH-FZ-RV10	5	15	8	16	11.0	11	6	11	12	14	10.8	12
Hybrid-RV10	14	17	21	**1**	13.3	14	15	17	8	10	12.5	16
GAS-2F-RN5	17	**2**	10	12	10.3	9	14	6	10	4	8.5	5
GAS-1F-RN5	4	**1**	**2**	5	3.0	**1**	12	**3**	16	13	11.0	13
GARCH-FZ-RN5	12	5	18	18	13.3	15	**2**	5	19	15	10.3	11
Hybrid-RN5	10	11	16	10	11.8	13	13	14	6	6	9.8	9
GAS-2F-RN10	15	**3**	13	6	9.3	7	10	**1**	5	**1**	4.3	**1**
GAS-1F-RN10	**1**	6	17	**2**	6.5	**3**	4	7	14	5	7.5	**3**
GARCH-FZ-RN10	13	9	22	20	16.0	18	5	**2**	20	8	8.8	6
Hybrid-RN10	9	13	9	4	8.8	6	11	16	7	11	11.3	14

(Continued on the next page)

(Continued) Out-of-sample performance rankings for various levels of α

	\multicolumn{6}{c}{$\alpha = 5\%$}						\multicolumn{6}{c}{$\alpha = 10\%$}				
	S&P	DJIA	NIK	FTSE	Avg	Rank	S&P	DJIA	NIK	FTSE	Avg	Rank
RW-125	27	27	26	27	26.8	27	27	27	27	27	27.0	27
RW-250	28	28	27	28	27.8	28	28	28	26	28	27.5	28
RW-500	29	29	28	29	28.8	29	29	29	28	29	28.8	29
GARCH-G	23	24	24	26	24.3	25	23	26	25	25	24.8	26
GARCH-Skt	20	21	21	24	21.5	21	20	23	23	22	22.0	23
HAR-Skt-RV5	19	19	29	20	21.8	22	15	15	29	16	18.8	18
HEAVY-G-RV5	16	16	16	18	16.5	18	5	14	14	10	10.8	12
HEAVY-Skt-RV5	1	11	4	8	6.0	**2**	1	10	9	7	6.8	4
AL-CAViaR-Sym	25	23	18	22	22.0	23	24	21	22	24	22.8	25
GAS-2F	26	26	25	25	25.5	26	16	18	16	26	19.0	20
GAS-1F	22	25	23	19	22.3	24	26	20	20	18	21.0	22
GARCH-FZ	21	20	19	21	20.3	20	22	22	24	21	22.3	24
Hybrid	24	22	7	23	19.0	19	21	16	10	17	16.0	16
GAS-2F-RV5	9	5	14	**2**	7.5	6	6	7	5	4	5.5	**2**
GAS-1F-RV5	**2**	10	6	11	7.3	5	7	**1**	**2**	5	3.8	**1**
GARCH-RV5	4	12	**2**	15	8.3	7	**3**	11	6	12	8.0	8
Hybrid-RV5	18	14	12	6	12.5	15	18	24	**3**	23	17.0	17
GAS-2F-RV10	6	4	13	**3**	6.5	**3**	14	9	11	**3**	9.3	10
GAS-1F-RV10	12	9	**3**	17	10.3	11	11	8	**1**	8	7.0	5
GARCH-FZ-RV10	11	13	17	16	14.3	17	9	12	17	14	13.0	14
Hybrid-RV10	15	15	**1**	5	9.0	9	19	19	4	19	15.3	15
GAS-2F-RN5	10	**1**	9	7	6.8	4	8	**3**	12	**2**	6.3	**3**
GAS-1F-RN5	7	7	15	4	8.3	8	13	13	8	9	10.8	13
GARCH-FZ-RN5	**3**	6	22	12	10.8	13	**2**	**2**	19	11	8.5	9
Hybrid-RN5	14	17	10	13	13.5	16	17	25	18	15	18.8	19
GAS-2F-RN10	8	**2**	11	**1**	5.5	**1**	10	6	13	**1**	7.5	6
GAS-1F-RN10	13	8	8	10	9.8	10	12	5	7	6	7.5	7
GARCH-FZ-RN10	5	**3**	20	14	10.5	12	4	4	21	13	10.5	11
Hybrid-RN10	17	18	5	9	12.3	14	25	17	15	20	19.3	21

Note: This table presents the rankings (with the best performing model ranked 1 and the worst ranked 29) based on the average losses using the FZ0 loss function, for four daily equity return series, over the out-of-sample period, for 29 different forecasting models. The last column in each panel represents the average rank across the four equity return series.

Table 2.4.1. Positive test statistics corresponding to darker colors mean that the row model has larger losses than the column model. The white blocks mean that the row model dominates the column model in loss comparison at 95% significance level; the light green (below white in the color bar) blocks mean that the row model has lower average loss than the column model, but not significantly so; and the dark red blocks mean that the row model has higher loss than the column model at 95% significance level. In Figure 2.4.1, at 1% level, the rows for Model 8 (HEAVY-Skt-RV5), Model 23 (GAS-1F-RN5), and Model 27 (GAS-1F-RN10) have lighter blocks compared to the other rows, therefore, these are the three best performing models for the S&P 500 index at 1% level. For 2.5% level, Model 8, Model 24 (GARCH-FZ-RN5), and Model 27 outperform the others. At 5% and 10% levels, Model 3, Model 24, and Model 28 (GARCH-FZ-RN10) are the three best performing models for the S&P 500 index.

Following Gerlach and Wang (2020) and Taylor (2019), we use the model confidence set (MCS) test introduced by Hansen et al. (2011) to compare the forecasting models via the FZ0 loss function. This approach builds model confidence sets using one-sided elimination based on the Diebold-Mariano test. In this chapter, we consider the 75% confidence level[10] and employ two methods: the R method using sums of absolute values for calculating the test statistic for MCS; and the SQ method uses the summed squares.[11] Table 2.4.7 presents the number of models within the MCS test using the block bootstrap with the block length of 12 and 10,000 replications, based on the losses generated from the FZ0 loss function. The GAS-2F-RN10 is the best performing model, overall, and the GAS models extended with realized measures perform better than most of the benchmark models. The main finding generated from the MCS test echo the results from the other backtesting methods. The result that some GAS models enhanced with realized measures end up more often in the MCS than HAR and HEAVY models highlights the usefulness of the score function that the GAS models build on, and we also show evidence that the use of realized measures enhances the risk

Figure 2.4.1: Color map based on the Diebold-Mariano (DM) test

(a) 1% S&P 500

(b) 2.5% S&P 500

(c) 5% S&P 500

(d) 10% S&P 500
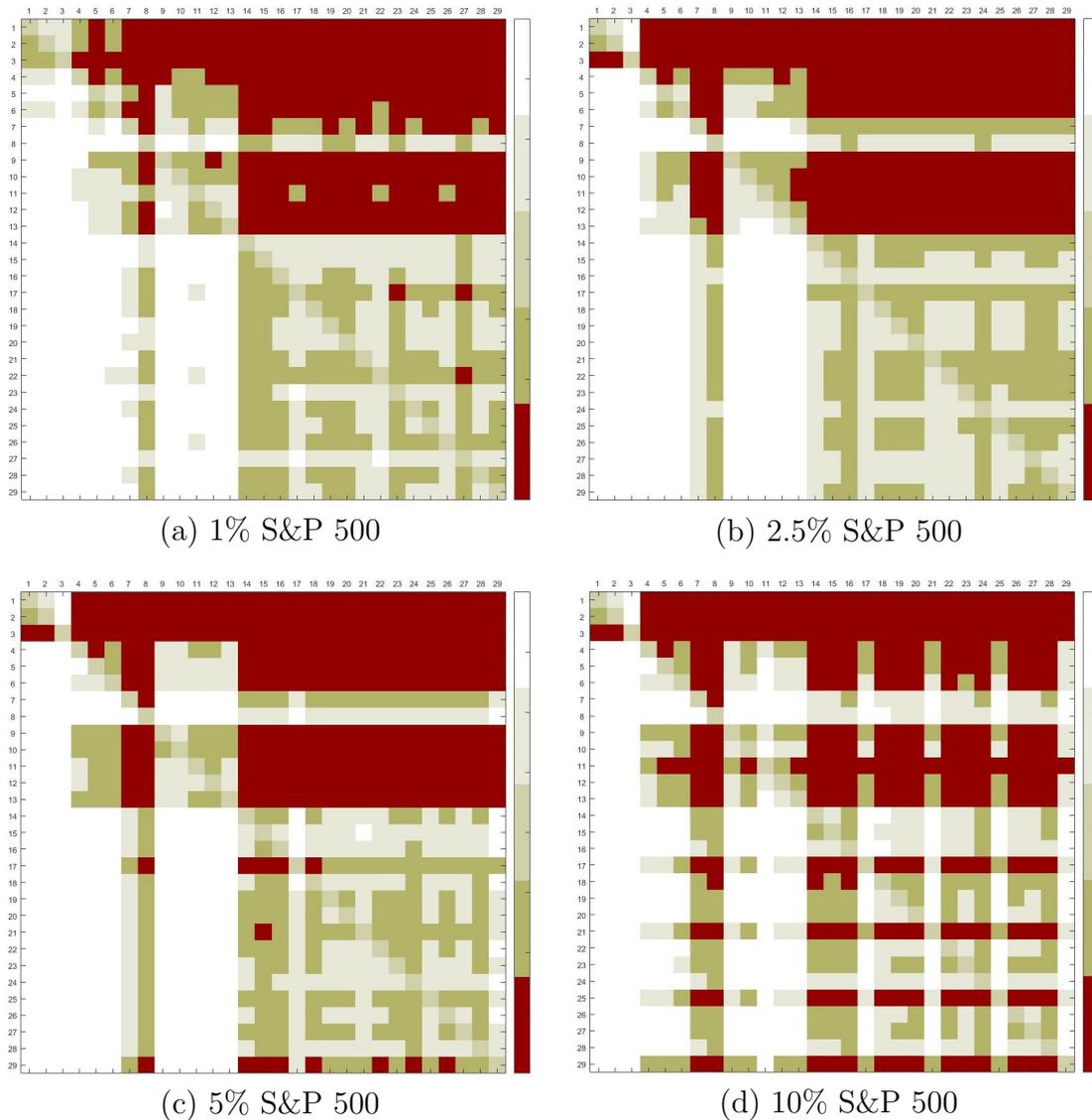
Note: This figure presents the color maps based on the Diebold-Mariano (DM) test comparing the average losses using the FZ0 loss function over the out-of-sample period for 29 different models, for the S&P 500. White blocks mean that the row model has lower average loss than the column model at 5% significance level; light green (below white in the color bar) blocks mean that the row model has lower average loss than the column model, but not significantly different from it, and so on. Darker color blocks mean that the row model has higher average loss than the column model.

forecasts of GAS models.

2.5   Conclusions

Patton et al. (2019) proposed a set of semiparametric models (GAS-FZ) in a generalized autoregressive score (GAS) framework to estimate risk measures. This chapter provides an extension of this, using exogenous information from high frequency data, in order to improve on the prediction of VaR and ES. This provides a new semiparametric framework named GAS-FZ-Realized, proposed for estimating and forecasting VaR and ES jointly. Through incorporating four realized measures (5-min and 10-min realized volatility with or without the overnight return) into the GAS-FZ models, we observe an improvement in forecasting risk measures over both in-sample and out-of-sample periods.

We employ the newly proposed models to estimate the VaR and ES of four international stock indices empirically, over the period 2000 to 2019. The parameters of the models are estimated by minimizing the FZ loss function of Fissler and Ziegel (2016). Then VaR and ES forecasts are built and individually backtested using the unconditional coverage test and the dynamic quantile (and ES) regression tests, as well as the joint loss function is computed. The main finding is that forecasts generated from the GAS-FZ-Realized models outperform forecasts based on GARCH models or historical simulations, even those based on the original GAS-FZ models. The only exception is the HEAVY-Skt-RV5 which is difficult to beat.

To conclude, the GAS-FZ-Realized models, especially the GAS-2F combined with the 10-min realized volatility and the overnight return, can provide more accurate risk measures for risk management across different stock indices and probability levels when compared to other models. This work could be potentially extended by improving the ES component, as the dynamics of VaR may not change simultaneously with ES, for example, by modelling an AR relationship

Table 2.4.7: The 75% model confidence set for the R and SQ methods across the four stock indices

	Summed absolute values (R method)					Summed squares (SQ method)				
	1%	2.5%	5%	10%	TOTAL	1%	2.5%	5%	10%	TOTAL
RW-125	0	0	0	0	0	0	0	0	0	0
RW-250	0	0	0	0	0	0	0	0	0	0
RW-500	0	0	0	0	0	0	0	0	0	0
GARCH-N	0	0	0	0	0	0	0	0	0	0
GARCH-Skt	0	0	0	0	0	0	1	0	0	1
HAR-Skt-RV5	0	0	0	0	0	0	0	0	0	0
HEAVY-G-RV5	0	0	0	0	0	0	1	2	1	4
HEAVY-Skt-RV5	3	3	3	2	11	3	3	3	**2**	11
AL-CAViaR-Sym	2	1	0	0	3	2	2	0	0	4
GAS-2F	0	0	0	0	0	0	0	0	0	0
GAS-1F	0	0	0	0	0	0	0	0	0	0
GARCH-FZ	2	1	0	0	3	2	1	0	0	3
Hybrid	1	1	1	0	3	1	1	1	0	3
GAS-2F-RV5	**4**	1	3	**3**	11	**4**	2	3	**2**	11
GAS-1F-RV5	**4**	2	3	2	11	**4**	3	3	**2**	**12**
GARCH-FZ-RV5	2	3	3	**3**	11	3	3	3	**2**	11
Hybrid-RV5	1	1	2	1	5	2	2	2	1	7
GAS-2F-RV10	3	2	**4**	2	11	3	2	**4**	**2**	11
GAS-1F-RV10	2	3	2	2	9	2	3	2	**2**	9
GARCH-FZ-RV10	2	2	1	1	6	3	3	3	1	10
Hybrid-RV10	2	2	2	1	7	2	2	3	1	8
GAS-2F-RN5	2	2	**4**	2	10	2	3	**4**	**2**	11
GAS-1F-RN5	**4**	2	2	0	8	**4**	3	3	1	11
GARCH-FZ-RN5	2	2	2	2	8	2	2	2	**2**	8
Hybrid-RN5	2	3	3	0	8	2	3	3	0	8
GAS-2F-RN10	2	**4**	**4**	2	**12**	3	**4**	**4**	**2**	**13**
GAS-1F-RN10	2	2	3	1	8	3	3	3	1	10
GARCH-FZ-RN10	1	2	3	1	7	2	2	3	**2**	9
Hybrid-RN10	3	3	2	0	8	3	3	2	0	8

Note: This table presents the number of indices for which each method is within the model confidence set (MCS) at the 75% confidence level based on the FZ0 loss function. The highest value (in bold) means that the model is the most favored one across four stock indices and for different probability levels.

between VaR and ES (Taylor, 2019) or by assuming a dynamic Omega ratio to describe the relationship between the two measures (Taylor, 2022). Moreover, this chapter can be extended by using realized volatility at different frequencies or via other proposed realized measures, for example those found in Meng and Taylor (2018).

Notes

[1]The connection between quantiles, expectiles and ES is originally found in Aigner et al. (1976), and considered further by Newey and Powell (1987).

[2]More studies related to the GAS model can be found on: http://www.gasmodel.com/.

[3]Salvatierra and Patton (2015) use measures of realized covariance to build forecasts for copula models.

[4]Both parametric - see Giot and Laurent (2004), Hansen et al. (2012), and Louzis et al. (2014) - and semiparametric models - see Clements et al. (2008), Fuertes and Olmo (2013), Žikeš and Baruník (2014), and Gerlach and Wang (2016).

[5]See Clements et al. (2008); Fuertes and Olmo (2013); Žikeš and Baruník (2014); Gerlach and Chen (2014); Gerlach and Wang (2022), and Gerlach and Chen (2017).

[6]Normally, the objective function is a probability density function, but here the loss function FZ0 acts as the objective.

[7]This realized library can be accessed by: https://realized.oxford-man.ox.ac.uk/

[8]For parameters in the GAS-2F models, the predetermined intervals for w, b, a_v, a_e, and c are [-0.1, 0.1], [0.8, 1], [-0.1, 0.1], [-0.1, 0.1], and [-1, 0], respectively. For parameters in the GAS-1F, GARCH-FZ, and Hybrid models, the predetermined intervals for β_1, β_2, β_3, c, a, and b are [0.8, 1], [0, 0.1], [0, 0.1], [0, 0.5], [-2, -3], and [-3, -4], respectively.

[9]We use the Student's t-test for significance testing.

[10]The 95% confidence level was considered as well with similar results (results available on request).

[11]Details can be found on page 465 of Hansen et al. (2011); and the Matlab code for MCS testing can be downloaded from www.kevinsheppard.com/MFE Toolbox.

Chapter 3

Loss Function-based Change Point Detection in Risk Measures

3.1 Introduction

Measuring market risk plays a central role not only in the area of risk management but also in the broader context of financial markets. Value-at-Risk (VaR) and Expected Shortfall (ES) are two prevalent risk measures dominating in contemporary financial regulation (Leung et al., 2021). VaR provides financial institutions with a loss level that occurs in the worst situations at a given significance level; ES, as an alternative to VaR, is the expectation of losses, conditional on their realization lying below VaR. As for the estimation of these two measures, Engle and Manganelli (2004) distinguish three main categories: nonparametric, parametric, and semiparametric approaches. In a univariate framework, some of the models for financial risk measures include GARCH family models (Bali and Theodossiou, 2007), score-driven models (Patton et al., 2019), and CAViaR-ES models (Taylor, 2019)[1].

It is worth mentioning that the presence of change points in time series may cause misleading statistical inference under the assumption of stationarity (Stock and Watson, 1996; Clements and Hendry, 1996; Diebold and Inoue, 2001; Mikosch

and Stărică, 2004; Loschi et al., 2007). Related empirical evidence has been extensively documented, especially in stock returns (Pástor and Stambaugh, 2001), volatility (Inclan and Tiao, 1994), correlation dynamics (Barassi et al., 2020), and macroeconomic time series (Pesaran and Timmermann, 2007). There is a vast literature of tests for change points in time series; some of these detect changes in a historical dataset (Csörgő and Horváth, 1997; Aue et al., 2009), whereas others monitor changes in a sequential manner (Berkes et al., 2004; Dette and Gösmann, 2020; Horváth et al., 2020a). Also, these tests can differ in terms of their objective function given by, e.g., the likelihood for volatility models (Chen and Hong, 2016) and copula models (Ye et al., 2012) or the loss function for quantile regressions (Qu, 2008). We refer the readers to Aue and Horváth (2013) for a detailed literature review.

In applications of risk management, the existence of change points can cause estimation errors for VaR and ES, as argued in Hoga (2017) and Fan et al. (2018). These papers use an innovative self-normalized estimator *à la* Zhang and Lavitas (2018) to detect change points when the risk measures are estimated in a non-parametric way. Specifically, Hoga (2017) investigates change points in the VaR process, and Fan et al. (2018) consider changes in the ES process. Since regulatory capital requirements in Basel Committee on Banking Supervision (2019) are linked to ES estimates, it would be prudent to detect change points in this process. Also, if change points are detected in the ES series alone, then the effect of VaR on ES is ignored. Since ES is elicitable only jointly with VaR[2], it is meaningful to detect change points in the (VaR, ES) tuple.

To fill this gap, our study extends the current literature by proposing a test to detect change points in the VaR and ES series simultaneously, which are estimated by (semi)parametric models. We construct this test using the FZ loss functions proposed by Fissler and Ziegel (2016). Since the FZ loss functions are minimized for the true values of VaR and ES, changes in the parameter values of the model cause breaks in the process of the VaR and ES estimates, which will result in

breaks in the loss series. Our framework of detecting change points in the VaR and ES series based on loss values is general and can accommodate for any type of (semi)parametric estimation models.

Our first contribution is to propose a test to detect change points in both VaR and ES risk measures simultaneously, based on the FZ loss functions. The general framework is closely related to the likelihood ratio test to detect changes in volatility, and the test for structural changes in quantile regressions proposed by Qu (2008). Due to the dominance of the indicator term in the FZ loss functions, the presence of extreme values (spikes), when returns exceed VaR, is one of the main characteristics of the loss series. However, the commonly used CUSUM test can be affected by the presence of outliers (Fearnhead and Rigaill, 2019). To address this problem, we adopt a more suitable alternative, namely the Wilcoxon test (Dehling et al., 2013b) to detect change points in the loss process.[3] We call this procedure the loss-based Wilcoxon test, and we shed light on its advantages in detecting joint change points in time series of VaR and ES simultaneously.

Secondly, this chapter contributes to the current literature by deriving the asymptotic behavior of our test statistic under weak dependence. Also, to improve the finite sample performance of the proposed test, we adopt a stationary bootstrap method based on Politis and Romano (1994), which follows the strand of literature in the area (Hušková and Kirch, 2008; Quaedvlieg, 2021). Furthermore, we prove that the stationary bootstrap is valid for the loss-based Wilcoxon test.

Thirdly, using Monte Carlo (MC) simulations, we show the advantages of the loss-based Wilcoxon test in detecting change points in risk measures. These advantages include good size control and higher power in finite samples compared with the alternatives. Additionally, our study on risk measures of the S&P 500 index returns provides an empirical application to demonstrate the practical usage of our proposed test. We present evidence that the loss-based Wilcoxon test can detect change points that are consistent with well-known market events.

The chapter is structured as follows: Section 3.2 briefly introduces the FZ family of loss functions and the Wilcoxon test statistic, and presents some theoretical results related to the asymptotic distribution of this statistic under weak dependence; Section 3.3 discusses the stationary bootstrap method and provides the validity of bootstrapping; Section 3.4 shows the simulation results; Section 3.5 contains an empirical application based on the S&P 500 index; Section 3.6 concludes the chapter.

3.2 Test Statistic for Change Point Detection

3.2.1 Loss Functions

Let $\{r_t\}_{t=1}^M$ be a series of observed returns measured over an arbitrary frequency, such as daily. (Semi)parametric models can be used to estimate the corresponding conditional risk measures, VaR and ES, denoted by $\{v_t(\boldsymbol{\theta})\}_{t=1}^M$ and $\{e_t(\boldsymbol{\theta})\}_{t=1}^M$, at a specified significance level α, where $\boldsymbol{\theta}$ denotes the unknown parameter vector of the model. Fissler and Ziegel (2016) introduce the FZ family of loss functions stated below, used to evaluate the (VaR, ES) tuple of risk measures:

$$\ell^{FZ}(r_t, v_t, e_t, \boldsymbol{\theta}; \alpha) = (\mathbf{1}\{r_t \leq v_t(\boldsymbol{\theta})\} - \alpha) \left(G_1(v_t(\boldsymbol{\theta})) - G_1(r_t) + \frac{1}{\alpha} v_t(\boldsymbol{\theta}) G_2(e_t(\boldsymbol{\theta})) \right)$$
$$- G_2(e_t(\boldsymbol{\theta})) \left(\frac{1}{\alpha} \mathbf{1}\{r_t \leq v_t(\boldsymbol{\theta})\} r_t - e_t(\boldsymbol{\theta}) \right) - \mathcal{G}_2(e_t(\boldsymbol{\theta})),$$
$$(3.2.1)$$

where G_1 is weakly increasing, G_2 is strictly increasing and strictly positive, and $\mathcal{G}'_2 = G_2$ (for more details, see Patton et al., 2019).

For the specification function G_1 in (3.2.1), we use $G_1(z) = 0$, which follows the reasoning of Nolde and Ziegel (2017). We consider the second specification function $\mathcal{G}_2$ with different degrees of positive homogeneity[4] $\tilde{b} = -1$, $\tilde{b} = 0$, and $\tilde{b} = 0.5$, which follow the choices of Dimitriadis and Bayer (2019), specified as: $\mathcal{G}_2(z) = -\frac{1}{z}$, $\mathcal{G}_2(z) = -\log(-z)$, and $\mathcal{G}_2(z) = -\sqrt{-z}$, respectively, where z

must be negative. Positive homogeneity is a crucial property of loss functions, which means that linear rescaling of the input variables does not alter the ranking of the losses (Dimitriadis and Bayer, 2019). Nolde and Ziegel (2017) state the importance of the choice of loss functions for M-estimations, and indicate that positively homogeneous loss functions outperform others in terms of asymptotic efficiency.

In our study, we use the three loss functions corresponding to the above specifications, detailed in Table 3.2.1, to compute the time series of loss values.

Table 3.2.1: Loss functions in the FZ family with different degrees of positive homogeneity $\tilde{b}$

$\tilde{b}$	FZ loss function
0	$\ell^{FZ0}(r, v, e, \boldsymbol{\theta}; \alpha) = -\frac{1}{\alpha e(\boldsymbol{\theta})} \mathbf{1}\{r \le v(\boldsymbol{\theta})\}(v(\boldsymbol{\theta}) - r) + \frac{v(\boldsymbol{\theta})}{e(\boldsymbol{\theta})} + \log(-e(\boldsymbol{\theta})) - 1$
-1	$\ell^{FZ1}(r, v, e, \boldsymbol{\theta}; \alpha) = \frac{1}{e(\boldsymbol{\theta})^2} \left(\frac{1}{\alpha} \mathbf{1}\{r \le v(\boldsymbol{\theta})\}(v(\boldsymbol{\theta}) - r) - (v(\boldsymbol{\theta}) - e(\boldsymbol{\theta})) \right) + \frac{1}{e(\boldsymbol{\theta})}$
0.5	$\ell^{FZ2}(r, v, e, \boldsymbol{\theta}; \alpha) = \frac{1}{2\sqrt{-e(\boldsymbol{\theta})}} \left(\frac{1}{\alpha} \mathbf{1}\{r \le v(\boldsymbol{\theta})\}(v(\boldsymbol{\theta}) - r) - (v(\boldsymbol{\theta}) - e(\boldsymbol{\theta})) \right) + \sqrt{-e(\boldsymbol{\theta})}$

To provide some intuition, ℓ^{FZ0} can be reformulated as:

$$\ell^{FZ0}(r, v, e, \boldsymbol{\theta}; \alpha) = \begin{cases} -\frac{1}{\alpha e(\boldsymbol{\theta})}(v(\boldsymbol{\theta}) - r) + \frac{v(\boldsymbol{\theta})}{e(\boldsymbol{\theta})} + \log(-e(\boldsymbol{\theta})) - 1, & \text{if } r \le v(\boldsymbol{\theta}), \\ \frac{v(\boldsymbol{\theta})}{e(\boldsymbol{\theta})} + \log(-e(\boldsymbol{\theta})) - 1, & \text{if } r > v(\boldsymbol{\theta}). \end{cases}$$

The probability of the first outcome is α, and the probability of the second one is $1 - \alpha$. Thus, the distribution of the loss value can be generally considered as a mixing distribution with mixing parameter α.

To get a better understanding of the time series properties of the risk measures and loss series, we test, using simulations based on a GARCH(1,1)-skewed t data generating process (DGP), for the presence of (1) autocorrelation, (2) conditional heteroskedasticity, (3) unit root, and (4) outliers against the normal distribution in these series.[5] The results show that the loss series possibly has weak autocor-

relation, but we found no evidence of conditional heteroskedasticity. Also, in our setup we found that the loss series is stationary. Furthermore, the VaR exceptions cause spikes in the loss series, which lead to rejections of the normality test.

3.2.2 Risk Measure Estimation

In this section, we consider two types of estimation approaches for VaR and ES: parametric estimation (Francq and Zakoïan, 2015) and semiparametric estimation (Patton et al., 2019). In the first case, parameters are estimated using conditionally heteroskedastic models by maximizing a specific likelihood function. Following Francq and Zakoïan (2015), we have:

Assumption 3.2.1. *(A) $\{r_t\}_{t=1}^M$ is strictly stationary, and satisfies $\mathbb{E}[|r_t|^{\tilde{s}}] < \infty$, for some $\tilde{s} > 4$; $\{r_t\}_{t=1}^M$ is also an ergodic solution of the model $r_t = \sigma_t u_t$, $\sigma_t = \sigma(r_{t-1}, r_{t-2}, \ldots; \boldsymbol{\theta}^*)$, where σ_t is a volatility process, $\boldsymbol{\theta}^* \in \mathbb{R}^d$ is a set of volatility parameters belonging to a parameter space Θ, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0, \infty)$;*

(B) There exists a function $\tilde{H}$ such that for any volatility parameters $\boldsymbol{\theta}^* \in \Theta$, for any $\tilde{K} > 0$, and any sequence $\{x_t\}_{t=1}^M$

$$\tilde{K}\sigma(x_1, x_2, \ldots; \boldsymbol{\theta}^*) = \sigma(x_1, x_2, \ldots; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = \tilde{H}(\boldsymbol{\theta}^*, \tilde{K})$, which is the risk parameters; the model in (A) can be reparameterized as $r_t = \sigma_t^* u_t^*$, $\sigma_t^* = \sigma(r_{t-1}, r_{t-2}, \ldots; \boldsymbol{\theta})$;

(C) For any real sequence $\{x_t\}_{t=1}^M$, the function $\sigma(x_1, x_2, \ldots; \boldsymbol{u})$ is continuous. Almost surely, $\sigma_t(\boldsymbol{u}) \in (\underline{\omega}, \infty)$ for any $\boldsymbol{u} \in \Theta \subset \mathbb{R}^d$ and for some $\underline{\omega} > 0$. Additionally, $\sigma_t(\boldsymbol{\theta})/\sigma_t(\boldsymbol{u}) = 1$ a.s. iff $\boldsymbol{\theta} = \boldsymbol{u}$;

(D) $\mathbb{E}[\tilde{g}(u_0^*, \sigma)] < \mathbb{E}[\tilde{g}(u_0^*, 1)]$, $\forall \sigma > 0$, $\sigma \neq 1$, where $\tilde{g}$ is a specific likelihood function;

(E) There exist a random variable C_1 measurable with respect to $\{r_u\}_{u \leq 0}$ and a constant $\tilde{\rho} \in (0, 1)$ such that $\sup_{\boldsymbol{\theta}^* \in \Theta} |\sigma_t(\boldsymbol{\theta}^*) - \tilde{\sigma}_t(\boldsymbol{\theta}^*)| \leq C_1 \tilde{\rho}^t$, where $\tilde{\sigma}_t(\boldsymbol{\theta}^*)$

denotes the dynamic of the volatility with arbitrary initial values $\{\tilde{y}_u\}_{u \leq 0}$.

Proposition 3.2.1. *(Francq and Zakoïan, 2015) Under Assumption 3.2.1, if the $\hat{\boldsymbol{\theta}}_M$ parameters in a GARCH parametric model are estimated via (Q)MLE, the estimators converge to the true values of the parameters $\boldsymbol{\theta}$ when M goes to infinity, such that $\hat{\boldsymbol{\theta}}_M \xrightarrow{P} \boldsymbol{\theta}$ as $M \to \infty$;*

On the other hand, in the semiparametric approach, parameters in a parametric structure are estimated by minimizing the FZ0 loss function. Patton et al. (2019) prove the consistency of parameters based on the following assumptions:

Assumption 3.2.2. *(A) The loss series $\{\ell_t\}_{t=1}^M$ obeys the uniform law of large numbers;*

(B) The process $\{r_t\}_{t=1}^M$ is strictly stationary and satisfies $\mathbb{E}[r_t] = 0$, $\mathbb{E}[|r_t|^{\tilde{s}}] < \infty$, for some $\tilde{s} > 4$; Conditional on all past information $\mathscr{F}_{t-1}$, the distribution of r_t is $F(\cdot|\mathscr{F}_{t-1})$ which belongs to a class of distribution functions on $\mathbb{R}$ with unique α-quantiles;

(C) The vector of unknown parameters to be estimated is $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, where Θ is a compact subset of $\mathbb{R}^d$ for $d \in \mathbb{N}$;

(D) For any t, both $v_t(\boldsymbol{u})$ and $e_t(\boldsymbol{u})$ are $\mathscr{F}_{t-1}$-measurable and a.s. continuous in $\boldsymbol{u}$;

(E) If $P\left((v_t(\boldsymbol{u}) = v_t(\boldsymbol{\theta})) \cap (e_t(\boldsymbol{u}) = e_t(\boldsymbol{\theta}))\right) = 1 \ \forall t$, then $\boldsymbol{u} = \boldsymbol{\theta}$.

Proposition 3.2.2. *(Patton et al., 2019) Under Assumption 3.2.2, if the $\hat{\boldsymbol{\theta}}_M$ parameters in a model are estimated by minimizing the FZ0 loss function, the estimators converge to the true values of the parameters $\boldsymbol{\theta}$ when M goes to infinity, such that $\hat{\boldsymbol{\theta}}_M \xrightarrow{P} \boldsymbol{\theta}$ as $M \to \infty$.*

Propositions 3.2.1 and 3.2.2 indicate that $v_t(\hat{\boldsymbol{\theta}}_M)$ and $e_t(\hat{\boldsymbol{\theta}}_M)$ converge to the true values of the risk measures (VaR and ES, respectively).[6] Thus, in the following sections, we assume that the parameters in a selected (semi)parametric model are well estimated.

3.2.3 Hypotheses and Test Statistic

The distribution of $\{r_t\}_{t=1}^M$ and the values of $\{v_t(\boldsymbol{\theta})\}_{t=1}^M$ and $\{e_t(\boldsymbol{\theta})\}_{t=1}^M$ all depend on the model parameter vector which can be time varying, hence it will be denoted by $\boldsymbol{\theta}_t$. Thus, in this case, a procedure for detecting a change point can be conducted by testing the null hypothesis: $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_M$, against the alternative hypothesis that there is one unknown break point k^*, that is: $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_{k^*} \neq \boldsymbol{\theta}_{k^*+1} = \cdots = \boldsymbol{\theta}_M$. The true values of VaR and ES will lead to the minimal loss values for the given returns. If there is a change point, the parameter values estimated using the time period between 1 and k^* will be different from the parameter values estimated from the whole sample, so the VaR and ES estimates based on the parameters obtained from the whole sample will deviate from the true values, leading to an increase in their loss values.

The setup above can be translated into a testing framework using the loss series. Thus, we consider the framework with the loss values generated by a stochastic process $\ell_t = \mu_t + \varepsilon_t$, where $\{\mu_t\}_{t=1}^M$ is the unknown mean and the noise $\{\varepsilon_t\}_{t=1}^M$ has zero mean and finite variance. The null hypothesis of no change point in the loss series can be written as:

$$H_0 : \mu_1 = \cdots = \mu_M,$$

versus the alternative hypothesis[7] of one change point in the loss series:

$$H_1 : \quad \mu_1 = \cdots = \mu_{k^*} \neq \mu_{k^*+1} = \cdots = \mu_M, \quad 1 < k^* < M.$$

The CUSUM test is commonly used to detect change points of a process. However, this test has the limitation that it can be disturbed by the presence of outliers or extremely heavy-tailed noise (Fearnhead and Rigaill, 2019; Gerstenberger, 2018). As shown in Section 3.2.1, spikes (against normality) commonly exist in the loss series, due to the VaR exception, and thus making the CUSUM

test less suitable to be applied directly on the "raw" loss series. As highlighted by Gerstenberger (2018), the Wilcoxon test statistic is a rank-type statistic and has the inherent advantage that it is not affected by spikes. Therefore, we employ the Wilcoxon test to detect change points in the rank of the loss series. The general form of the Wilcoxon test statistic is defined as:

$$W_M := \max_{1 \leq k \leq M} |W_{k,M}|, \text{ where } W_{k,M} := \sum_{i=1}^{k} R_i - \frac{k}{M} \sum_{i=1}^{M} R_i, \qquad (3.2.2)$$

where $R_i = \text{rank}(\ell_i) = \sum_{j=1}^{M} \mathbf{1}\{\ell_j \leq \ell_i\}$, for $i = 1, \ldots, M$. Inspired by Betken (2016), our test statistic based on ranks is given below:

$$W_M = \max_{1 \leq k \leq M} \left| \sum_{i=1}^{k} R_i - \frac{k}{M} \sum_{i=1}^{M} R_i \right| = \max_{1 \leq k \leq M} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{M} \left(\mathbf{1}\{\ell_i \leq \ell_j\} - \frac{1}{2} \right) \right|.$$
$$(3.2.3)$$

The location estimator of the Wilcoxon test $\hat{k}_W$ is defined as the value that maximizes the loss-based Wilcoxon test statistic,

$$\hat{k}_W = \hat{k}_W(M) := \min \{k : |W_{k,M}| = W_M\}. \qquad (3.2.4)$$

3.2.4 Asymptotic Distribution of the Test Statistic

In this section, we investigate the asymptotic distribution of our proposed Wilcoxon-type statistic in (3.2.3). This can be treated as a U-statistic (Csörgő and Horváth, 1988; Dehling et al., 2017) with the kernel:

$$h_W(X, Y) = \mathbf{1}\{X \leq Y\} - \frac{1}{2}. \qquad (3.2.5)$$

We can define the U-process as below:

$$U_M(u) = \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor+1}^{M} h_W(\ell_i, \ell_j) = \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor+1}^{M} \left(\mathbf{1}\{\ell_i \leq \ell_j\} - \frac{1}{2} \right), \quad (3.2.6)$$

where $0 \leq u \leq 1$, and $\lfloor \cdot \rfloor$ denotes the integer part of a number. Thus the Wilcoxon change point test statistic in (3.2.3) can be written as:

$$W_M = \max_{0 \leq u \leq 1} \left| U_M(u) \right|. \qquad (3.2.7)$$

The kernel $h_W(X, Y)$ is antisymmetric, so it satisfies:

$$h_W(X, Y) = -h_W(Y, X). \qquad (3.2.8)$$

In this case, $\mathbb{E}[h_W(\ell_i, \ell_j)] = 0$ and similarly to the symmetric case we let $\tilde{h}_W(X) = \mathbb{E}[h_W(X, \ell_i)]$. Following Csörgő and Horváth (1988), it is reasonable to assume that:

$$0 < \mathbb{E}[h_W^2(\ell_i, \ell_j)] < \infty \quad \text{and} \quad 0 < \sigma_W^2 = \mathbb{E}[\tilde{h}_W^2(\ell_i)] < \infty. \qquad (3.2.9)$$

To derive the asymptotic distribution of the process $U_M(u)$, we consider the following assumptions.

Assumption 3.2.3. *(A) The loss series $\{\ell_t\}_{t=1}^M$ is strictly stationary and satisfies $\mathbb{E}[|\ell_t|^\xi] < \infty$, for some $\xi > 0$;*

(B) For any integer $1 \leq t \leq M$, the cumulative distribution function F of ℓ_t is continuous on the real line with a density f that is bounded;

(C) $h_W(\ell_1, \ell_2)$ given in (3.2.5) is an antisymmetric kernel, such that for a $\delta > 0$, $K > 0$:

$$\int \int |h_W(\ell_1, \ell_2)|^{2+\delta} \, dF(\ell_1) dF(\ell_2) \leq K,$$

$$\forall k \in \mathbb{N}_0 : \int |h_W(\ell_1, \ell_{1+k})|^{2+\delta} \, dP(\ell_1, \ell_{1+k}) \leq K;$$

(D) $\{r_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta})\}_{t=1}^M$ is strong mixing (α-mixing) with $\alpha^*(M) = O\left(M^{-(\tilde{z}-2)/\tilde{z}}\right)$ for some $\tilde{z} > 2$; $\{\ell_t(r_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta}))\}_{t=1}^M$ is strong mixing with the coefficient $\alpha^*(M) = O\left(M^{-\rho}\right)$ for a $\rho > \frac{3\xi\delta + \delta + 5\xi + 2}{2\xi\delta}$.

Assumption 3.2.3 (A) is a standard moment and stationarity condition for the loss series. Assumption 3.2.3 (B) is the condition on the continuous and bounded density of the loss series, which supports the proof of $\mathscr{P}$-Lipschitz-continuity for the kernel. Assumption 3.2.3 (C) requires the moment bound for the given kernel $h_W(\ell_1, \ell_2)$, which is consistent with Borovkova et al. (2001) and Dehling and Wendler (2010). Patton et al. (2019) provide the same dependence condition as Assumption 3.2.3 (D) for $\{r_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta})\}_{t=1}^M$ to support the central limit theorem for the loss series; if the first half of Assumption 3.2.3 (D) holds, the sequence of loss $\ell_t(r_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta}))$ is α-mixing with a decay rate at least as fast as that of $\{r_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta})\}_{t=1}^M$ (Patton et al., 2019). Thus, it is reasonable to assume the mixing condition for the loss series with the coefficient provided by Dehling and Wendler (2010). In the following, we discuss the $\mathscr{P}$-Lipschitz-continuity property for the kernel $h_W(X, Y)$.

Definition 3.2.1. *($\mathscr{P}$-Lipschitz-continuity) Let $\{X_t\}_{t\in\mathbb{N}}$ be a stationary process. A kernel h is called $\mathscr{P}$-Lipschitz-continuous if there is a constant $\tilde{a} > 0$ with*

$$\mathbb{E}\left[|h(X,Y) - h(X',Y)|\, \boldsymbol{1}\{|X - X'| \leq \epsilon\}\right] \leq \tilde{a}\epsilon,$$

for every $\epsilon > 0$, every pair X and Y with the common distribution $\mathscr{P}_{X_1, X_m}$ for $m \in \mathbb{N}$ with $m > 1$ or $\mathscr{P}_{X_1} \times \mathscr{P}_{X_1}$ and X' and Y also with one of these common distributions.

Proposition 3.2.3. *If Assumption 3.2.3 (B) holds, then the antisymmetric kernel $h_W(X,Y) = \boldsymbol{1}\{X \leq Y\} - \frac{1}{2}$ for the test statistic is $\mathscr{P}$-Lipschitz-continuous.*

The proof of this proposition can be found in Appendix 3.A. We then provide the asymptotic behavior of the $U_M(u)$ process.

Theorem 3.2.1. *Under the null hypothesis, assume that (3.2.8), (3.2.9), and Assumption 3.2.3 hold. Additionally, assume that:*
i) under a parametric setting, Assumption 3.2.1 holds,

or

ii) under a semiparametric setting, Assumption 3.2.2 holds.

Then as $M \to \infty$, we have:

$$\sup_{0 \leq u \leq 1} \left| \frac{1}{M^{3/2}} U_M(u) - \sigma_W B_M(u) \right| = o_P(1),$$

where $B_M(u), 0 \leq u \leq 1$ is a Brownian bridge, and

$$\sigma_W^2 = Var\left(F(\ell_1)\right) + 2 \sum_{j=2}^{\infty} Cov\left(F(\ell_1), F(\ell_j)\right).$$

The proof of Theorem 3.2.1 is provided in Appendix 3.A. One way to implement such a test is by estimating the long-run variance and using the asymptotic limit to obtain the p-values. However, as often found in the literature, the empirical size obtained when relying on the asymptotic limit in finite samples may differ significantly from the prespecified significance level. Table 3.B.2 of Appendix 3.B shows that the loss-based Wilcoxon test based on the asymptotic distribution with two long-run variance estimators is generally oversized, especially for small samples. As such, instead of estimating the long-run variance σ_W^2 above, we are going to use bootstrapping to obtain the p-values. The following section will elaborate the bootstrapping algorithm.

3.3 Stationary Bootstrap for p-values

3.3.1 Bootstrap Method

It is well known that bootstrap techniques have been widely used to avoid the finite sample size distortions (see Chen and Hong, 2016; Chen and Fang, 2019; Barendse and Patton, 2021, for more examples). Thus, we propose to obtain the p-values of the test statistic W_M by using stationary bootstrapping in the

following way. For a given return series $\{r_t\}_{t=1}^M$, we calculate the test statistic W_T using (3.2.7). Then, we adopt the stationary bootstrap method of Politis and Romano (1994) to generate N_B bootstrapped return series $\{r_t^*\}_{t=1}^M$ using the expected block length b_{opt}.[8] For each bootstrapped series, we estimate the bootstrapped VaR and ES denoted by $\{v_t^*(\hat{\boldsymbol{\theta}}_M^*)\}_{t=1}^M$ and $\{e_t^*(\hat{\boldsymbol{\theta}}_M^*)\}_{t=1}^M$, where $\hat{\boldsymbol{\theta}}_M^*$ is the parameter vector estimated from the bootstrapped returns $\{r_t^*\}_{t=1}^M$. Then we compute the loss series denoted by $\{\ell_t^*\}_{t=1}^M$. Applying (3.2.6) and (3.2.7) for each bootstrapped series j, we compute the bootstrapped U-process, $U_M^{*(j)}$ and the bootstrapped statistic $W_M^{*(j)}$. Then, we define the set of the bootstrapped statistics $\mathcal{W}_M^* = \{W_M^{*(1)}, \ldots, W_M^{*(N_B)}\}$. After that, we calculate the frequency that the statistic W_M is below $W_M^{*(j)}$, and this is the bootstrapped p-value. The detailed procedure can be found in Algorithm 1.

Algorithm 1 Bootstrap procedure to obtain p-value, $\texttt{Bootstrap}(\{r_t\}_{t=1}^M, W_M, N_B)$

Input: $\{r_t\}_{t=1}^M, W_M, N_B$
Output: p-value (p)
Initialization: $j = 0$
repeat //Bootstrap j//
  $\quad j = j + 1$
  $\quad$ Generate the bootstrapped returns $\{r_t^*\}_{t=1}^M$ using the stationary bootstrap
  $\quad$ Estimate the bootstrapped risk measure series $\{v_t^*(\hat{\boldsymbol{\theta}}_M^*)\}_{t=1}^M$ and $\{e_t^*(\hat{\boldsymbol{\theta}}_M^*)\}_{t=1}^M$
  $\quad$ Compute the bootstrapped loss series $\{\ell_t^*\}_{t=1}^M$
  $\quad$ Compute the bootstrapped statistic $W_M^{*(j)}$
until $j = N_B$;
Using $\{W_M^{*(1)}, \ldots, W_M^{*(N_B)}\}$ compute $p = \frac{1}{N_B} \sum_{j=1}^{N_B} \mathbf{1}\{W_M^{*(j)} > W_M\}$
return p.

3.3.2 Validity of the Bootstrap Method

To verify the validity of the bootstrap method, we obtain the asymptotic distribution of the bootstrapped statistic W_M^*, which is computed based on (3.2.7) using the bootstrapped data. Then we show that it asymptotically converges to the limit distribution of the statistic W_M under the null hypothesis. To conduct the verification, we consider the following proposition, which is needed for the

proof of our results.

Assumption 3.3.1. *$\{r_t^*\}_{t=1}^M$ is generated by the stationary bootstrap with geometric block lengths with success probability $p_M = \phi M^{-m}$ and $\phi, m \in (0, 1)$.*

Proposition 3.3.1. *(Politis and Romano, 1994) If Assumption 3.2.1 (A) or Assumption 3.2.2 (B) holds, and additionally Assumption 3.3.1 holds, then the pseudo time series $\{r_t^*\}_{t=1}^M$ is stationary.*

This proposition implies that the stationary bootstrapping ensures the stationarity of the process. In this chapter, we resample the return series $\{r_t\}_{t=1}^M$ instead of resampling the loss series $\{\ell_t\}_{t=1}^M$ directly.[9] The following theorem states the asymptotic behavior of the statistics of the bootstrapped loss series.

Theorem 3.3.1. *Under the null hypothesis, assume that Assumptions 3.2.3 and 3.3.1 hold. Additionally, assume that:*
i) under a parametric setting, Assumption 3.2.1 holds,
or
ii) under a semiparametric setting, Assumption 3.2.2 holds.
Let b_{opt} be the expected block length with $b_{opt} \to \infty$ and also $M/b_{opt} \to \infty$ as $M \to \infty$. Then we have the following convergence result for the bootstrapped process U_M^* obtained with expected block length b_{opt}:

$$|Var^*(M^{-3/2}U_M^*(u)) - Var(M^{-3/2}U_M(u))| \xrightarrow{P} 0, \qquad (3.3.1)$$

$$\sup_{x \in \mathbb{R}} |P^*(M^{-3/2}U_M^*(u) \leq x) - P(M^{-3/2}U_M(u) \leq x)| \xrightarrow{P} 0, \qquad (3.3.2)$$

where Var^* and P^* denote the variance and probability with respect to the probability measure induced by the stationary bootstrap.

The proof of this theorem can be found in Appendix 3.A.

Recall that $W_M^{*(j)}$, $1 \leq j \leq N_B$, denotes the bootstrapped statistic calculated similarly to W_M defined in (3.2.7). Next, we show that the asymptotic distribution

of the bootstrapped statistic W_M^* coincides with the asymptotic distribution of W_M under the null hypothesis. The empirical distribution function of $W_M^{*(j)}$ is calculated as:

$$Q_{M,N_B}(w) = \frac{1}{N_B} \sum_{1 \leq j \leq N_B} \mathbf{1}\{W_M^{*(j)} \leq w\}, \quad w \in \mathbb{R}. \tag{3.3.3}$$

Based on equations (3.2.7) and (3.3.3), as well as Theorems 3.2.1 and 3.3.1, we obtain the following result:

Corollary 3.3.1. *If the assumptions of Theorem 3.3.1 hold, then under H_0 we have:*

$$\sup_{w \in \mathbb{R}} |P(W_M \leq w) - Q_{M,N_B}(w)| \xrightarrow{P} 0, \quad \text{where } N_B \to \infty \text{ and } M \to \infty. \tag{3.3.4}$$

This corollary demonstrates that the proposed bootstrap methodology is appropriate to be used to obtain the p-value of the loss-based Wilcoxon test statistic. In the next section, we implement a simulation study to show that the bootstrap methodology has the correct size under the null hypothesis and has high power under the alternative hypothesis.

3.4 Simulation Analysis

3.4.1 Simulation Design

We perform a simulation study to investigate the size and power of the proposed test in finite samples. Under the null hypothesis, the DGP of the return series is a univariate GARCH process as given below:

$$\begin{aligned} r_t &= \sigma_t u_t, \quad u_t \sim i.i.d. \text{ skewed } t \ (\nu, \lambda), \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2, \quad t = 1, ..., M, \end{aligned} \tag{3.4.1}$$

where r_t is the simulated return process generated by the product of u_t, which follows the standardized skewed t distribution of Hansen (1994), with degree of freedom (DoF) ν and skewness λ, and conditional volatility σ_t given by a GARCH(1,1) specification. For the simulations, we choose the sample sizes of $M \in \{1000, 3000\}$ to study the finite sample properties and convergence of the test.[10]

Under the alternative hypothesis, the DGP of returns is the process $r_t = \sigma_t u_t$ with:

$$\begin{cases} \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2, & u_t \sim i.i.d. \text{ skewed } t\ (\nu, \lambda), \quad \text{if } 1 < t \le \lfloor \pi M \rfloor, \\ \sigma_t^2 = \beta_0^* + \beta_1^* \sigma_{t-1}^2 + \beta_2^* r_{t-1}^2, & u_t \sim i.i.d. \text{ skewed } t\ (\nu^*, \lambda^*), \quad \text{if } \lfloor \pi M \rfloor < t \le M, \end{cases}$$
$$(3.4.2)$$

where one of the parameters changes its value after $\lfloor \pi M \rfloor$ which is the location of the change point in the process. In this chapter, we consider $\pi \in \{0.5, 0.75\}$. This change in the return series will eventually cause a change point in the VaR and ES as well, and our main purpose is to investigate the detection of change points in the VaR and ES processes at $\alpha = 1\%$.[11]

Regarding parameter values, we set $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 16.5, -0.5)$. Under the null hypothesis, $(\beta_0^*, \beta_1^*, \beta_2^*, \nu^*, \lambda^*) = (\beta_0, \beta_1, \beta_2, \nu, \lambda)$ in (3.4.2), meaning no change points in the process. For the alternative hypothesis, we consider six different scenarios of change points to evaluate the empirical power of the proposed test. Each break consists of a change in the value of one parameter as follows:

- H_1^{A1}: an increase of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.94$;

- H_1^{A2}: a decrease of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.86$;

- H_1^{B1}: an increase of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.09$;

- H_1^{B2}: a decrease of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.01$;

- H_1^{C1}: a decrease of 13.5 in the DoF parameter, i.e. $\nu^* = 3$;

- H_1^{C2}: a decrease of 14 in the DoF parameter[12], i.e. $\nu^* = 2.5$.

In addition to the above alternatives, we follow Andreou and Ghysels (2002) to examine whether the presence of outliers affects our test results under the null hypothesis. We conjecture that the existence of outliers should not lead to the rejection of the test, i.e. an effective test would not mistakenly consider outliers as change points:

- H_0^D: $(\beta_0^*, \beta_1^*, \beta_2^*, \nu^*, \lambda^*) = (\beta_0, \beta_1, \beta_2, \nu, \lambda)$, when 12 randomly selected returns in the simulated process are multiplied by 5.

In the simulation, we consider the eight DGPs detailed above. For the estimation of VaR and ES, we use the following three (semi)parametric models: GARCH(1,1)-skewed t (G-Skt), GARCH(1,1)-Gaussian (G-G) and GAS-Hybrid (Hybrid).[13] In terms of the loss function, we choose loss functions with three different degrees of positive homogeneity: ℓ^{FZ0}, ℓ^{FZ1}, and ℓ^{FZ2}, given in Table 3.2.1.

For each combination of (DGP, estimation model, loss function) we follow Algorithm 2 to compute the rejection rates of the proposed test, as explained in detail below. For each simulation i, we simulate return series of length M, denoted by $\{r_t\}_{t=1}^M$. We then estimate the VaR and ES series using the given model, and we denote the estimated risk series as $\{v_t(\hat{\boldsymbol{\theta}}_M)\}_{t=1}^M$ and $\{e_t(\hat{\boldsymbol{\theta}}_M)\}_{t=1}^M$. Following this, we calculate the loss series $\{\ell_t\}_{t=1}^M$ for the given loss function. Then, based on (3.2.7) we compute the loss-based Wilcoxon statistic W_T for the loss series. By calling the bootstrap procedure in Algorithm 1 with $N_B = 1000$, we obtain the p-value of simulation i, denoted by $p(i)$. If $p(i)$ is below the test significance level q, then the null hypothesis is rejected for simulation i.[14] By

repeating this simulation $N_S = 1000$ times, we obtain the rejection rate ζ as the frequency of $p(i)$ being lower than q in the total number of simulations.

Algorithm 2 Monte Carlo simulation procedure for Loss-based Wilcoxon test

---

Input: N_S, N_B, M, q
Output: rejection rate (ζ)
Initialization: $i = 0$
repeat //Simulation i//
    $i = i + 1$
    Simulate $\{r_t\}_{t=1}^M$ using the specified DGP with sample size M
    Estimate the risk measure series $\{v_t(\hat{\boldsymbol{\theta}}_M)\}_{t=1}^M$ and $\{e_t(\hat{\boldsymbol{\theta}}_M)\}_{t=1}^M$
    Calculate the loss values $\{\ell_t\}_{t=1}^M$
    Calculate:
$$W_T = \max_k \left| \sum_{m=1}^{k} \sum_{n=k+1}^{M} \left(\mathbf{1}\{\ell_m \leq \ell_n\} - \frac{1}{2} \right) \right|.$$
    Obtain p-value by calling Algorithm 1: $p(i) = \texttt{Bootstrap}(\{r_t\}_{t=1}^M, W_M, N_B)$.
until $i = N_S$;
Using the p-values: $\{p(1),\ldots,p(N_S)\}$ compute the rejection rate $\zeta = \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{1}\{p(i) < q\}$
return ζ.

In terms of the simulation results, we expect that the empirical size converges to q, the test significance level under the null hypothesis, as the number of observations increases. Under the alternative hypothesis, the expectation is that the empirical power is high and converges to 1 with the sample size. When adding outliers to the process without change points, the empirical rejection rate should be close to q if the change point test is not sensitive to outliers. Our setup allows us to explore the sensitivity of the test to the choice of risk estimation model, loss function, type and location of break and sample size.[15]

3.4.2   Simulation Results

The simulation results commence with the evaluation of the proposed loss-based Wilcoxon test in identifying change points in risk measures when the underlying process is generated from the DGP in (3.4.1) and (3.4.2) with the parameter values given in Section 3.4.1. Table 3.4.1 shows the size and power of the test based on the bootstrapping procedure at 5% test significance level. The top panel

of the table shows the empirical sizes under the null hypothesis. As expected, all of the empirical sizes for the Wilcoxon test are close to the test significance level. As the sample size increases, the empirical size gets closer to 5% in general.

For the alternative hypotheses, we consider the change points detailed in Section 3.4.1. The results in Table 3.4.1 reveal that our test has a strong power in detecting change points in the volatility parameters ($H_1^{A_1}$, $H_1^{A_2}$, $H_1^{B_1}$, $H_1^{B_2}$) and reasonable power in detecting change points in the DoF ($H_1^{C_1}$, $H_1^{C_2}$). The power of the test improves when M increases from 1000 to 3000 for all DGPs and loss functions. The table also shows that the power of the test is sensitive to the location of change point. The rejection rate modestly falls when the location of change point moves to $\lfloor 0.75M \rfloor$. However, as the sample size increases, the test can successfully detect the change point that occurs even at $\lfloor 0.75M \rfloor$. Also, the results show that our test is not sensitive to the presence of outliers (H_0^D).

In the following, we compare our proposed loss-based Wilcoxon test with five alternative tests in terms of size and power, under the same simulation settings and hypotheses as detailed before. For the first two alternative tests, we consider (i) the self-normalized CUSUM (SN-CUSUM) test for VaR and (ii) the SN-CUSUM for ES, which detect change points in the VaR and ES processes individually. Following Shao and Zhang (2010), the two test statistics are defined as:

$$V_M^v = \sup_k \frac{\left(M^{-\frac{1}{2}} \sum_{t=1}^k v_t(\hat{\boldsymbol{\theta}}_M) - \frac{k}{M} \sum_{t=1}^M v_t(\hat{\boldsymbol{\theta}}_M)\right)^2}{M^{-2}\left[\sum_{t=1}^k S_{v,t}^2(1,k) + \sum_{t=k+1}^M S_{v,t}^2(k+1,M)\right]^{\frac{1}{2}}}, \qquad (3.4.3)$$

$$V_M^e = \sup_k \frac{\left(M^{-\frac{1}{2}} \sum_{t=1}^k e_t(\hat{\boldsymbol{\theta}}_M) - \frac{k}{M} \sum_{t=1}^M e_t(\hat{\boldsymbol{\theta}}_M)\right)^2}{M^{-2}\left[\sum_{t=1}^k S_{e,t}^2(1,k) + \sum_{t=k+1}^M S_{e,t}^2(k+1,M)\right]^{\frac{1}{2}}}, \qquad (3.4.4)$$

where $v_t(\hat{\boldsymbol{\theta}}_M)$ and $e_t(\hat{\boldsymbol{\theta}}_M)$ are VaR and ES, and $S_{v,t}(j,k) = \sum_{i=j}^t (v_i(\hat{\boldsymbol{\theta}}_M) - \bar{v}_{j,k})$, $\bar{v}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^k v_t(\hat{\boldsymbol{\theta}}_M)$, as well as $S_{e,t}(j,k) = \sum_{i=j}^t (e_i(\hat{\boldsymbol{\theta}}_M) - \bar{e}_{j,k})$, $\bar{e}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^k e_t(\hat{\boldsymbol{\theta}}_M)$. Table 3.4.2 presents the empirical size and power simulation

Table 3.4.1: Empirical size and power of the loss-based Wilcoxon test for a change point

	$\pi = 0.5$						$\pi = 0.75$					
	M=1000			M=3000			M=1000			M=3000		
	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid
H_0: Univariate GARCH(1,1)-skewed t, with $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 16.5, -0.5)$												
ℓ^{FZ0}	0.045	0.044	0.030	0.047	0.063	0.039	0.045	0.044	0.030	0.047	0.063	0.039
ℓ^{FZ1}	0.045	0.044	0.030	0.047	0.064	0.039	0.045	0.044	0.030	0.047	0.064	0.039
ℓ^{FZ2}	0.045	0.044	0.030	0.047	0.064	0.038	0.045	0.044	0.030	0.047	0.064	0.038
H_1^{A1}: An increase of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.94$												
ℓ^{FZ0}	0.992	0.992	0.923	1.000	1.000	1.000	0.788	0.769	0.326	0.996	0.975	0.919
ℓ^{FZ1}	0.992	0.992	0.923	1.000	1.000	1.000	0.788	0.770	0.326	0.996	0.975	0.918
ℓ^{FZ2}	0.992	0.992	0.923	1.000	1.000	1.000	0.788	0.773	0.328	0.996	0.975	0.918
H_1^{A2}: A decrease of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.86$												
ℓ^{FZ0}	0.627	0.623	0.373	0.988	0.963	0.770	0.232	0.220	0.118	0.712	0.695	0.355
ℓ^{FZ1}	0.627	0.623	0.373	0.988	0.963	0.770	0.232	0.220	0.117	0.712	0.695	0.354
ℓ^{FZ2}	0.627	0.623	0.373	0.988	0.963	0.770	0.232	0.221	0.117	0.712	0.695	0.355
H_1^{B1}: An increase of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.09$												
ℓ^{FZ0}	0.912	0.911	0.715	1.000	1.000	0.985	0.608	0.605	0.225	0.939	0.898	0.747
ℓ^{FZ1}	0.912	0.911	0.715	1.000	1.000	0.985	0.608	0.604	0.225	0.939	0.898	0.747
ℓ^{FZ2}	0.912	0.911	0.715	1.000	1.000	0.985	0.608	0.604	0.225	0.939	0.898	0.747
H_1^{B2}: A decrease of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.01$												
ℓ^{FZ0}	0.529	0.524	0.350	0.987	0.951	0.731	0.196	0.194	0.114	0.694	0.644	0.363
ℓ^{FZ1}	0.529	0.524	0.349	0.987	0.951	0.731	0.196	0.194	0.114	0.694	0.643	0.363
ℓ^{FZ2}	0.528	0.524	0.349	0.987	0.951	0.730	0.196	0.194	0.113	0.694	0.643	0.363
H_1^{C1}: A decrease of 13.5 in the DoF parameter, i.e. $\nu^* = 3$												
ℓ^{FZ0}	0.293	0.290	0.176	0.777	0.758	0.283	0.164	0.159	0.092	0.393	0.354	0.169
ℓ^{FZ1}	0.293	0.290	0.176	0.777	0.758	0.281	0.165	0.159	0.093	0.393	0.354	0.169
ℓ^{FZ2}	0.293	0.290	0.176	0.776	0.757	0.282	0.166	0.159	0.093	0.393	0.354	0.169
H_1^{C2}: A decrease of 14 in the DoF parameter, i.e. $\nu^* = 2.5$												
ℓ^{FZ0}	0.636	0.627	0.358	0.996	0.988	0.449	0.330	0.326	0.164	0.593	0.552	0.389
ℓ^{FZ1}	0.636	0.627	0.357	0.996	0.987	0.448	0.331	0.324	0.165	0.593	0.552	0.389
ℓ^{FZ2}	0.636	0.627	0.358	0.996	0.987	0.448	0.331	0.324	0.164	0.593	0.552	0.389
H_0^D: 12 randomly selected returns in the simulated process multiplied by 5												
ℓ^{FZ0}	0.041	0.042	0.041	0.035	0.052	0.039	0.041	0.042	0.041	0.035	0.052	0.039
ℓ^{FZ1}	0.041	0.042	0.040	0.035	0.052	0.039	0.041	0.042	0.040	0.035	0.052	0.039
ℓ^{FZ2}	0.041	0.042	0.041	0.035	0.052	0.039	0.041	0.042	0.041	0.035	0.052	0.039

Note: Empirical size and power, for $q = 5\%$, of the loss-based Wilcoxon test under various hypotheses via 1000 simulations, for three types of risk measures (GARCH(1,1)-skewed t, GARCH(1,1)-Gaussian and GAS-Hybrid) and three FZ loss functions with different degrees of positive homogeneity. We consider two sample sizes: 1000 and 3000, and different locations of the change point at $\lfloor \pi M \rfloor$ with $\pi = 0.5$ and 0.75.

results of the SN-CUSUM tests for VaR and ES. The sizes of the SN-CUSUM tests are close to the test significance level, but their powers are generally less than the power of our test for all loss functions considered.

One disadvantage of the standard CUSUM test is the low power in detecting change points occurring in relatively early or late segments of the sample period. As an alternative, Horváth et al. (2020b) propose a Rényi-type statistic for break detection to mitigate this problem. However, when the change point happens around the middle of the sample period, the detecting power of the Rényi-type test is relatively low. The Rényi-type test works under the assumption that there is no change point occurring within the two trimmed domains, at the beginning and at the end of the sample defined by the trimming parameter u_0. Thus, we consider the alternative test (iii) a Rényi-type test based on the rank of loss values. Specifically, the test statistic is a Rényi-type formulation of the loss-based Wilcoxon test statistic:

$$D_M := \max_{\lfloor u_0 M \rfloor \leq k \leq M - \lfloor u_0 M \rfloor} \left| \frac{1}{k} \sum_{i=1}^{k} R_i - \frac{1}{M-k} \sum_{i=k+1}^{M} R_i \right| \qquad (3.4.5)$$

with trimming parameter u_0.[16]

In addition to these, we consider the following two recently developed tests: (iv) the break point test for VaR of Hoga (2017), and (v) the break point test for ES of Fan et al. (2018).[17] These two tests are based on the self-normalized variance estimator of Shao and Zhang (2010).

Table 3.4.3 presents the simulations results for alternative tests (iii) to (v) (in columns Rényi, Hoga, and FGP, respectively). The results highlight that our test outperforms tests (iv) and (v) in all cases. We outperform the Rényi-type test (iii) when the change point occurs at $\lfloor 0.5M \rfloor$, but when the change point occurs at $\lfloor 0.75M \rfloor$, test (iii) has better power properties than our test. This meets our expectation that the Rényi-type test has high power in detecting change points occurring relatively early or late in the sample, but has lower power

Table 3.4.2: Empirical size and power of the SN-CUSUM test for a change point

	$\pi = 0.5$						$\pi = 0.75$					
	M=1000			M=3000			M=1000			M=3000		
	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid
H_0: Univariate GARCH(1,1)-skewed t, with $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 16.5, -0.5)$												
VaR	0.034	0.034	0.037	0.066	0.066	0.044	0.034	0.034	0.037	0.066	0.066	0.044
ES	0.034	0.034	0.043	0.066	0.066	0.043	0.034	0.034	0.043	0.066	0.066	0.043
H_1^{A1}: An increase of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.94$												
VaR	0.629	0.629	0.548	0.772	0.772	0.765	0.724	0.724	0.595	0.924	0.924	0.884
ES	0.629	0.629	0.546	0.772	0.772	0.765	0.724	0.724	0.600	0.924	0.924	0.884
H_1^{A2}: A decrease of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.86$												
VaR	0.361	0.361	0.307	0.783	0.783	0.676	0.100	0.100	0.104	0.353	0.353	0.317
ES	0.361	0.361	0.305	0.783	0.783	0.675	0.100	0.100	0.100	0.353	0.353	0.313
H_1^{B1}: An increase of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.09$												
VaR	0.372	0.372	0.362	0.558	0.558	0.847	0.518	0.518	0.421	0.776	0.776	0.744
ES	0.372	0.372	0.361	0.558	0.558	0.847	0.518	0.518	0.421	0.776	0.776	0.743
H_1^{B2}: A decrease of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.01$												
VaR	0.295	0.295	0.303	0.769	0.769	0.683	0.091	0.091	0.092	0.338	0.338	0.296
ES	0.295	0.295	0.296	0.769	0.769	0.687	0.091	0.091	0.093	0.338	0.338	0.297
H_1^{C1}: A decrease of 13.5 in the DoF parameter, i.e. $\nu^* = 3$												
VaR	0.161	0.161	0.208	0.268	0.268	0.225	0.076	0.076	0.155	0.171	0.171	0.226
ES	0.161	0.161	0.210	0.268	0.268	0.223	0.076	0.076	0.157	0.171	0.171	0.225
H_1^{C2}: A decrease of 14 in the DoF parameter, i.e. $\nu^* = 2.5$												
VaR	0.392	0.392	0.352	0.606	0.606	0.322	0.180	0.180	0.225	0.337	0.340	0.303
ES	0.392	0.392	0.341	0.606	0.606	0.319	0.180	0.180	0.223	0.337	0.340	0.302
H_0^D: 12 randomly selected returns in the simulated process multiplied by 5												
VaR	0.033	0.033	0.047	0.046	0.046	0.048	0.033	0.033	0.047	0.046	0.046	0.048
ES	0.033	0.033	0.047	0.046	0.046	0.047	0.033	0.033	0.047	0.046	0.046	0.047

Note: Empirical size and power, for $q = 5\%$, of the SN-CUSUM test for VaR and ES, considered individually, under various hypotheses via 1000 simulations, for three types of risk measures (GARCH(1,1)-skewed t, GARCH(1,1)-Gaussian and GAS-Hybrid). We consider two sample sizes: 1000 and 3000, and different locations of the change point at $\lfloor \pi M \rfloor$ with $\pi = 0.5$ and 0.75.

in the middle. For our simulation setup, we find the Hoga and FGP tests to be oversized under the null hypothesis and to have less power than the Rényi-type loss-based Wilcoxon test.[18]

To offer a visual demonstration, Figure 3.4.1 compares the loss-based Wilcoxon test using the FZ0 loss function with tests (i) to (v), from the point of view of size and power. The five alternatives are denoted by VaR, ES, Rényi-type, Hoga, and FGP, respectively. For the loss-based Wilcoxon test and alternative tests (i)-(iii), the VaR and ES are obtained using the GARCH(1,1)-skewed t model. The tests are performed at 5% test significance level, and we assume that the change point occurs at $\lfloor 0.5M \rfloor$ under the alternative hypotheses. Based on the empirical sizes of the Hoga and FGP tests under H_0 and H_0^D, it can be concluded that these tests are oversized for the DGP considered. The loss-based Wilcoxon test has higher power than the alternatives for all scenarios of change points corresponding to the different alternative hypotheses. The SN-CUSUM tests work relatively well when volatility changes, but have lower power when the DoF parameter decreases. Overall, our proposed test can identify change points in the risk measures of time series with the correct size and stronger power than all five alternatives considered.[19]

3.5 Empirical Application

In this section, we apply our proposed Wilcoxon change point test to S&P 500 index daily log returns. The index data is collected from Datastream and spans the period from January 2, 1990 to December 31, 2019, in total 7559 observations. We apply the proposed loss-based Wilcoxon test to detect change points in the VaR and ES risk measures estimated by the GARCH(1,1)-skewed t model. Based on our simulations above that consider the Wilcoxon tests based on loss functions with different degrees of positive homogeneity, it can be concluded that our test is not sensitive to the choice of loss function. As such, in the empirical section

Table 3.4.3: Empirical size and power of alternative tests for a change point

| | $\pi = 0.5$ | | | | | | $\pi = 0.75$ | | | | | |
| | M=1000 | | | M=3000 | | | M=1000 | | | M=3000 | | |
Rényi	Hoga	FGP	Rényi	Hoga	FGP	Rényi	Hoga	FGP	Rényi	Hoga	FGP
H_0: Univariate GARCH(1,1)-skewed t, with $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 16.5, -0.5)$											
0.045	0.148	0.104	0.044	0.107	0.105	0.045	0.148	0.104	0.044	0.107	0.105
H_1^{A1}: An increase of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.94$											
0.966	0.602	0.596	1.000	0.736	0.749	0.967	0.758	0.728	1.000	0.934	0.924
H_1^{A2}: A decrease of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.86$											
0.396	0.318	0.291	0.854	0.507	0.465	0.391	0.186	0.137	0.934	0.208	0.162
H_1^{B1}: An increase of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.09$											
0.794	0.488	0.486	0.968	0.617	0.644	0.799	0.631	0.646	0.988	0.863	0.863
H_1^{B2}: A decrease of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.01$											
0.338	0.344	0.290	0.851	0.559	0.529	0.337	0.164	0.134	0.914	0.213	0.172
H_1^{C1}: A decrease of 13.5 in the DoF parameter, i.e. $\nu^* = 3$											
0.193	0.209	0.211	0.461	0.194	0.264	0.268	0.200	0.244	0.648	0.196	0.348
H_1^{C2}: A decrease of 14 in the DoF parameter, i.e. $\nu^* = 2.5$											
0.445	0.263	0.210	0.833	0.210	0.220	0.576	0.196	0.243	0.907	0.173	0.279
H_0^D: 12 randomly selected returns in the simulated process multiplied by 5											
0.044	0.144	0.115	0.039	0.104	0.099	0.044	0.144	0.115	0.039	0.104	0.099

Note: Empirical size and power, for $q = 5\%$, of three alternative tests (iii), (iv) and (v) under various hypotheses via 1000 simulations. We consider two sample sizes: 1000 and 3000, and different locations of the change point at $\lfloor \pi M \rfloor$ with $\pi = 0.5$ and 0.75. For the Rényi-type test, we choose the loss values computed by the FZ0 loss function with VaR and ES estimated by the GARCH(1,1)-skewed t model.

Figure 3.4.1: Size and power of the loss-based Wilcoxon test and alternatives
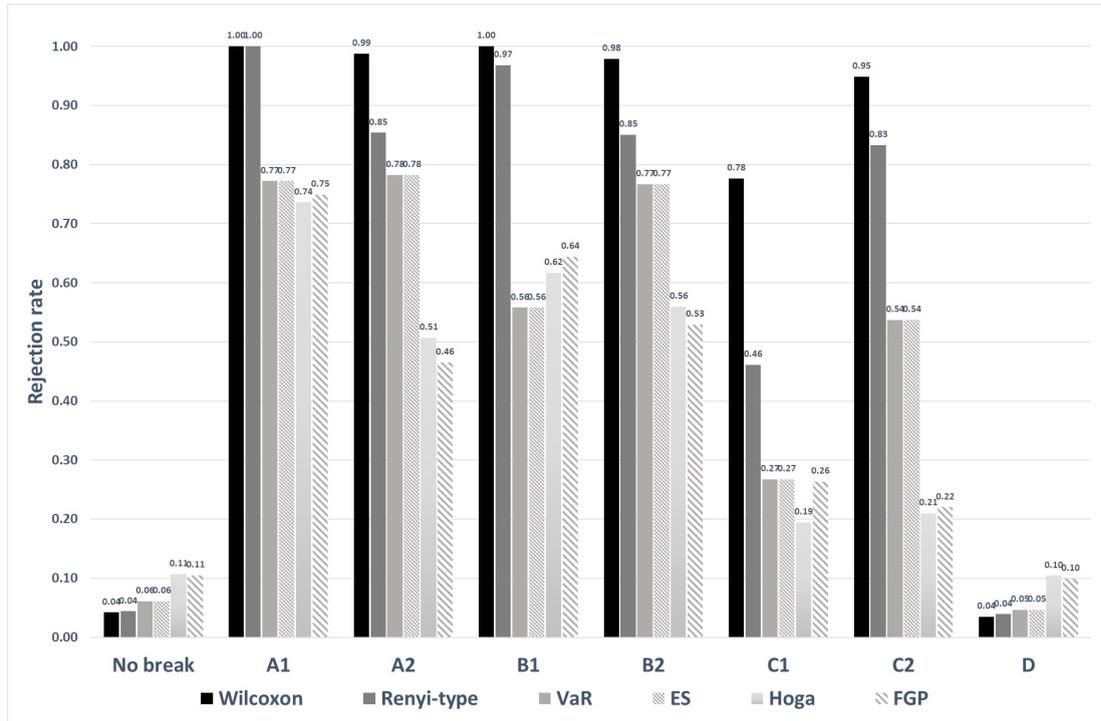
Note: Empirical size and power of the loss-based Wilcoxon test with five alternative tests (i)-(v) under various hypotheses described in Section 3.4.1 at 5% test significance level when the change point occurs at $\lfloor 0.5M \rfloor$. For the Wilcoxon test and Rényi-type test, we use the FZ0 loss function to compute the loss values. For all tests except Hoga and FGP, VaR and ES are estimated by the GARCH(1,1)-skewed t model.

we only use the FZ0 loss function to compute loss values, following Dimitriadis and Schnaitmann (2021).

In order to find change points in the risk measures, we first compute the loss-based Wilcoxon test statistic W_M. Then, we bootstrap the return process 1000 times via the stationary bootstrap method with the optimal block length (Politis and White, 2004; Patton et al., 2009), obtain the empirical distribution of the Wilcoxon statistic and get the 95% critical values. If the test statistic W_M is larger than the critical value, we reject the null hypothesis of no change. In such cases, a change point is detected, and we follow the binary segmentation method discussed by Inclan and Tiao (1994) and Ye et al. (2012) to find further change points. Specifically, the data can be split into sub-periods according to

the locations of the detected change points until no further change point can be found. The detailed algorithm and procedure of detecting multiple change points can be found in Figure 3.B.5 of Appendix 3.B.

Based on our test, the earliest change point we detect in the estimated risk measures occurred in June 1992 (following the early 1990s recession in the United States). The second change point occurred in December 1996 (the start of the dot-com bubble). Then another change point is identified in June 2003 (after the burst of the dot-com bubble), and the following change points are in July 2007 (the beginning of the subprime mortgage crisis), September 2008 (the bankruptcy of Lehman Brothers), July 2009, and January 2012 (the start and end of the European debt crisis). We also successfully detect change points associated with the 2015–16 stock market selloff[20] and the 2018 cryptocurrency crash[21]. Figure 3.5.1 presents the returns as well as the risk estimates, highlighting the detected change points. Additionally, we apply this test for other estimation approaches (GAS-Hybrid and historical simulations) and compare the empirical results with alternative tests applied for the same sample (more details can be found in Tables 3.B.6 and 3.B.7 of Appendix 3.B).

Table 3.5.1 reports the GARCH(1,1)-skewed t parameter estimates and standard errors obtained by the QMLE method for each sub-period, the average VaR and ES estimates, and the average loss values. Firstly, it can be seen that the volatility parameters and the DoF estimates experience large changes across the sub-periods, which leads to change points in the VaR and ES processes as well. For instance, after the burst of the dot-com bubble, we can observe a decline in the level of the volatility. Moreover, we can see a large reduction in the value of the DoF parameter from 11.1 to 6.5 during the European debt crisis period. Secondly, during a crisis or a crash period, VaR and ES are high in absolute values, as can be seen in the 2007-2008 financial crisis and the European debt crisis. The average loss values are also found to be generally higher during crisis periods than during stable periods. Finally, when change points are taken into

Figure 3.5.1: Daily S&P 500 index returns and risk estimates at 1% level
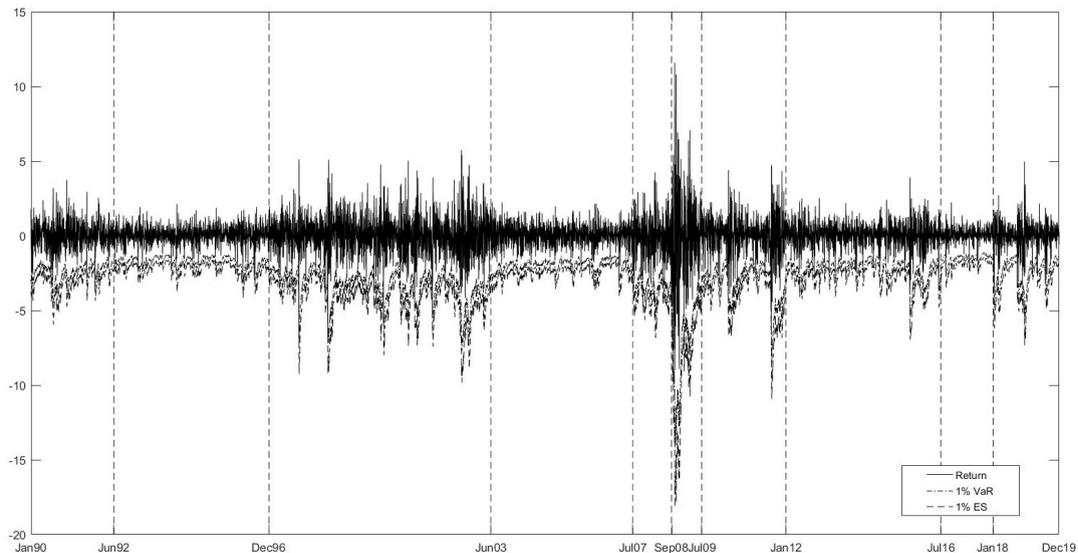
Note: Daily S&P 500 index returns and 1% VaR and ES estimates obtained by the GARCH skewed t model. The vertical dash lines are at the estimated change points.

consideration, the loss values are typically lower than the ones computed when the change points are ignored (this can be seen comparing Loss and Loss_NC in Table 3.5.1). According to our findings, it can be concluded that risk management practitioners can improve on the risk estimates by first identifying change points in the loss series of risk measures and then computing model parameter values based on the identified change points.

3.6   Conclusions

We propose a new test, named the loss-based Wilcoxon test, to detect change points in the series of VaR and ES risk measures considered jointly. Our test is based on the Wilcoxon test (Dehling et al., 2013b) applied to the FZ loss functions proposed by Fissler and Ziegel (2016). The framework of our test is general and can accommodate for any type of (semi)parametric estimation methods for VaR and ES. We perform extensive simulations based on various types of change point

Table 3.5.1: Subsample estimation results

	1990/01 - 1992/05	1992/05 - 1996/12	1996/12 - 2003/06	2003/06 - 2007/07	2007/07 - 2008/09
β_0	0.031 (0.012)	0.013 (0.006)	0.084 (0.022)	0.021 (0.007)	0.153 (0.075)
β_1	0.924 (0.023)	0.925 (0.024)	0.862 (0.024)	0.924 (0.023)	0.881 (0.032)
β_2	0.037 (0.013)	0.039 (0.013)	0.092 (0.017)	0.031 (0.014)	0.032 (0.036)
ν	7.343 (1.992)	5.567 (0.817)	9.943 (2.264)	11.137 (4.338)	14.078 (11.435)
λ	0.001 (0.002)	-0.008 (0.039)	-0.042 (0.037)	-0.076 (0.030)	-0.081 (0.066)
VaR	-2.237	-1.491	-3.270	-1.719	-3.424
ES	-2.805	-1.952	-4.000	-2.097	-4.095
Loss	1.038	0.747	1.431	0.702	1.244
Loss_NC	1.135	0.736	1.475	0.815	1.400

	2008/09 - 2009/07	2009/07 - 2012/01	2012/01 - 2016/07	2016/07 - 2018/01	2018/01 - 2019/12
β_0	0.012 (0.154)	0.036 (0.013)	0.075 (0.016)	0.059 (0.052)	0.042 (0.012)
β_1	0.930 (0.072)	0.860 (0.024)	0.723 (0.041)	0.692 (0.276)	0.773 (0.038)
β_2	0.063 (0.059)	0.119 (0.028)	0.170 (0.034)	0.047 (0.089)	0.181 (0.042)
ν	11.378 (9.630)	6.736 (1.683)	8.019 (1.819)	3.814 (0.684)	6.189 (1.830)
λ	-0.047 (0.067)	-0.151 (0.042)	-0.089 (0.035)	0.113 (0.059)	-0.211 (0.062)
VaR	-7.044	-3.223	-2.074	-1.071	-2.532
ES	-8.517	-4.138	-2.601	-1.517	-3.299
Loss	1.973	1.270	0.916	0.748	1.181
Loss_NC	2.109	1.362	1.011	0.827	1.435

Note: Estimated parameter values and standard errors for β_0, β_1, β_1, ν, and λ in the GARCH(1,1)-skewed t model: $\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2$, $u_t \sim i.i.d.$ *skewed* t (ν, λ) for the S&P 500 index in 10 sub-periods. We also report the average VaR and ES at 1% level and the associated average loss values using the parameters estimated within the sub-periods (Loss) and the average loss using parameters estimated over the whole sample period without consideration of change points (Loss_NC).

scenarios, including different locations for the change points and different changes in the volatility and DoF parameters. Our results show that the proposed test has better size under the null hypothesis and higher power properties under the considered alternative hypotheses, compared with five different alternative tests. We present an application of the loss-based Wilcoxon test on the S&P 500 index returns. The empirical results show that the test can detect the change points associated with well-known financial events.

Appendices

3.A Proofs

Proof of Theorem 3.2.1

Proof. In general, the Hoeffding decomposition can be applied to a U-statistic with a kernel $h(x, y)$, so that we have:

$$h(x, y) = \tilde{\theta} + h_1(x) + h_2(y) + \tilde{g}(x, y),$$

where $\tilde{\theta} = \mathbb{E}[h(X, Y)]$, $h_1(x) = \mathbb{E}[h(x, Y) - \tilde{\theta}]$, $h_2(y) = \mathbb{E}[h(X, y) - \tilde{\theta}]$ and $\tilde{g}(x, y) = h(x, y) - h_1(x) - h_2(y) - \tilde{\theta}$.

We have the properties for these three terms:

$$\mathbb{E}[h_1(X)] = \mathbb{E}[h_2(X)] = 0, \tag{3.A.1}$$

and

$$\mathbb{E}[\tilde{g}(x, Y)] = \mathbb{E}[\tilde{g}(X, y)] = 0. \tag{3.A.2}$$

The proof of Theorem 3.2.1 is based on a lemma introduced below.

Lemma 3.A.1. *(Dehling and Wendler (2010)) Let h be a $\mathscr{P}$-Lipschitz-continuous kernel with $2 + \delta$ moments for some $\delta > 0$, $\{X_t\}_{t \in \mathbb{N}}$ be a stationary strong mixing process with $\mathbb{E}[|X_1|^\xi] < \infty$ for some $\xi > 0$ and $\alpha^*(M) = O(M^{-\rho})$ with*

$\rho > \frac{3\xi\delta + 2\delta + 5\xi + 2}{2\xi\delta}$, *then for* $V_M(\tilde{g}) = \frac{2}{\sqrt{M}(M-1)} \sum_{1 \le i < j \le M} \tilde{g}(X_i, X_j)$, *we have*

$$\mathbb{E}\left[MV_M^2(\tilde{g})\right] \le \frac{4}{M(M-1)^2} \sum_{1 \le i_1 < i_2 \le M} \sum_{1 \le i_3 < i_4 \le M} |\mathbb{E}[\tilde{g}(X_{i_1}, X_{i_2})\tilde{g}(X_{i_3}, X_{i_4})]|$$

$$\le \frac{4}{M^3} \sum_{i_1, i_2, i_3, i_4 = 1}^{M} |\mathbb{E}[\tilde{g}(X_{i_1}, X_{i_2})\tilde{g}(X_{i_3}, X_{i_4})]| = O(M^{-\eta})$$

where $\eta = \min\left\{\rho \frac{2\xi\delta}{3\xi\delta + \delta + 5\xi + 2} - 1, 1\right\} > 0.$

The proof of this lemma can be found in Dehling and Wendler (2010) as the proof for Lemma 3.6.

Recall $h_W(\ell_i, \ell_j)$ is antisymmetric with $\tilde{\theta} = 0$. In order to prove the asymptotic normality of this U-process, we use the Hoeffding decomposition for the kernel $h_W(\ell_i, \ell_j)$:

$$h_W(\ell_i, \ell_j) = h_1(\ell_i) + h_2(\ell_j) + \tilde{g}(\ell_i, \ell_j).$$

Thus, based on (3.2.6), we have the decomposed U-process:

$$\frac{1}{M^{3/2}} U_M(u) = \frac{1}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor + 1}^{M} (h_1(\ell_i) + h_2(\ell_j) + \tilde{g}(\ell_i, \ell_j))$$

$$= \frac{1}{M^{3/2}} \left((M - \lfloor uM \rfloor) \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i) + \lfloor uM \rfloor \sum_{j=\lfloor uM \rfloor + 1}^{M} h_2(\ell_j) + \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor + 1}^{M} \tilde{g}(\ell_i, \ell_j)\right)$$

By Lemma 3.A.1, we have that for a given $u \in [0, 1]$, the upper boundary of the variance of $\frac{1}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor + 1}^{M} \tilde{g}(\ell_i, \ell_j)$:

$$\frac{1}{M^3} \sum_{\substack{i_1 = 1 : \lfloor uM \rfloor \\ i_2 = \lfloor uM \rfloor + 1 : M}} \sum_{\substack{i_3 = 1 : \lfloor uM \rfloor \\ i_4 = \lfloor uM \rfloor + 1 : M}} |\mathbb{E}[\tilde{g}(\ell_{i_1}, \ell_{i_2})\tilde{g}(\ell_{i_3}, \ell_{i_4})]|$$

$$\le \frac{1}{M^3} \sum_{1 \le i_1 < i_2 \le M} \sum_{1 \le i_3 < i_4 \le M} |\mathbb{E}[\tilde{g}(\ell_{i_1}, \ell_{i_2})\tilde{g}(\ell_{i_3}, \ell_{i_4})]| \qquad (3.A.3)$$

$$< \frac{1}{M^3} \sum_{i_1, i_2, i_3, i_4 = 1}^{M} |\mathbb{E}[\tilde{g}(\ell_{i_1}, \ell_{i_2})\tilde{g}(\ell_{i_3}, \ell_{i_4})]| = O(M^{-\eta}).$$

Hence, the variance of $\frac{1}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor+1}^{M} \tilde{g}(\ell_i, \ell_j)$ vanishes as M increases.

By (3.A.2) and (3.A.3), we have

$$\frac{1}{M^{3/2}} \sup_{0 \leq u \leq 1} \left| \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor+1}^{M} \tilde{g}(\ell_i, \ell_j) \right| \to 0$$

in probability.

Thus, by the Lemma of Slutsky, it is enough to show that the sum of the first two terms

$$\left(\frac{M - \lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i) + \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{j=\lfloor uM \rfloor+1}^{M} h_2(\ell_j) \right)_{0 \leq u \leq 1}$$

converges in distribution to the limit process of Theorem 3.2.1. Because the kernel $h_W(\ell_i, \ell_j)$ is antisymmetric, we have that $h_2(\ell_j) = -h_1(\ell_j)$. Thus, we can rewrite the representation as

$$\frac{M - \lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i) - \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{i=\lfloor uM \rfloor+1}^{M} h_1(\ell_i) = \frac{1}{M^{1/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i) - \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{M} h_1(\ell_i).$$

To obtain the limit of the process, we state the theorem below, which is a direct consequence of Theorem 4 in Borovkova et al. (2001) and Theorem 3.1 in Davidson and De Jong (2000).

Theorem 3.A.1. *Let $\{Y_k\}_{k \in \mathbb{Z}}$ be a L_2 near-epoch dependent (NED) with respect to a strong mixing process. Also, suppose that $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}[|Y_i|^{4+\delta}] \leq \infty$ for some $\delta > 0$. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{Y}} \sum_{i=1}^{Y} Y_i \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = Var(Y_1) + 2 \sum_{k=2}^{\infty} Cov(Y_1, Y_k)$.

The proof of the theorem follows immediately from Borovkova et al. (2001) and Davidson and De Jong (2000).

Applying Theorem 3.A.1 on the partial sum process, using similar arguments as in Chapter 4 of Csörgő and Horváth (1997) and Donsker's theorem, it can be shown that $\frac{1}{M^{1/2}}\sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i) - \frac{\lfloor uM \rfloor}{M^{3/2}}\sum_{i=1}^{M} h_1(\ell_i)$ converges to a limit process $\{\sigma_W(W(u) - uW(1))\}_{0 \leq u \leq 1}$, where $\{W(u)\}_{0 \leq u \leq 1}$ is a Wiener process, and

$$\sigma_W^2 = Var(h_1(\ell_1)) + 2\sum_{k=2}^{\infty} Cov(h_1(\ell_1), h_1(\ell_k)).$$

Additionally, we have that $h_1(x) = \frac{1}{2} - F(x)$. Thus,

$$\sigma_W^2 = Var(F(\ell_1)) + 2\sum_{k=2}^{\infty} Cov(F(\ell_1), F(\ell_k)).$$

By the Lemma of Slutsky, we obtain that as $M \to \infty$, $\frac{1}{M^{3/2}} U_M(u)$ converges in distribution to $\{\sigma_W B(u)\}_{0 \leq u \leq 1}$, where $B(u) = W(u) - uW(1)$ is a Brownian bridge and $\sigma_W^2 = Var(F(\ell_1)) + 2\sum_{k=2}^{\infty} Cov(F(\ell_1), F(\ell_k))$. $\qquad\square$

Proof of Proposition 3.2

Proof. The kernel $h_W(X, Y)$ is $\mathscr{P}$-Lipschitz-continuous, if there is a constant $\tilde{a} > 0$, so that for all $\epsilon > 0$ and every common distribution of X, X' and Y,

$$\mathbb{E}\left[|h_W(X, Y) - h_W(X', Y)|\mathbf{1}\{|X - X'| \leq \epsilon\}\right] < \tilde{a}\epsilon.$$

For random variables X, X' and Y, we have:

$$\mathbb{E}\left[|h_W(X, Y) - h_W(X', Y)|\mathbf{1}\{|X - X'| \leq \epsilon\}\right]$$
$$= \mathbb{E}\left[|\mathbf{1}\{X \leq Y\} - \mathbf{1}\{X' \leq Y\}|\mathbf{1}\{|X - X'| \leq \epsilon\}\right].$$

We have:

$$\mathbb{E}\left[|\mathbf{1}\left\{X \leq Y\right\} - \mathbf{1}\left\{X' \leq Y\right\}|\,\mathbf{1}\left\{|X - X'| \leq \epsilon\right\}\right] \leq P(-\epsilon \leq X - X' \leq \epsilon)$$

Based on Assumption 3.2.3 (B) on the continuous distribution function, there exists a constant $\tilde{a} = 2\sup(f)$ that satisfies the following:

$$P(X' - \epsilon \leq X \leq X' + \epsilon) = F(X' + \epsilon) - F(X' - \epsilon) = \int_{X'-\epsilon}^{X'+\epsilon} f(t)dt \leq \frac{\tilde{a}}{2} \cdot 2\epsilon = \tilde{a}\epsilon.$$

Thus, based on Definition 3.2.1, the antisymmetric kernel of the Wilcoxon test statistic $h_W(X, Y)$ is $\mathscr{P}$-Lipschitz-continuous. $\qquad\square$

Proof of Theorem 3.1

Proof. In order to obtain the asymptotic behavior of the bootstrapped U-process, we use the Hoeffding decomposition for the bootstrapped kernel $h_W(\ell_i^*, \ell_j^*)$:

$$h_W(\ell_i^*, \ell_j^*) = h_1(\ell_i^*) + h_2(\ell_j^*) + \tilde{g}(\ell_i^*, \ell_j^*).$$

Thus, we have the decomposed bootstrapped U-process:

$$\frac{1}{M^{3/2}}U_M^*(u) = \frac{1}{M^{3/2}}\sum_{i=1}^{\lfloor uM \rfloor}\sum_{j=\lfloor uM \rfloor+1}^{M}\left(h_1(\ell_i^*) + h_2(\ell_j^*) + \tilde{g}(\ell_i^*, \ell_j^*)\right)$$

$$= \frac{1}{M^{3/2}}\left((M - \lfloor uT \rfloor)\sum_{i=1}^{\lfloor uM \rfloor}h_1(\ell_i^*) + \lfloor uM \rfloor\sum_{j=\lfloor uM \rfloor+1}^{M}h_2(\ell_j^*) + \sum_{i=1}^{\lfloor uM \rfloor}\sum_{j=\lfloor uM \rfloor+1}^{M}\tilde{g}(\ell_i^*, \ell_j^*)\right).$$
$$(3.A.4)$$

In the following, we are going to use the result below:

Lemma 3.A.2. *(Hwang and Shin (2015)) Let h be a $\mathscr{P}$-Lipschitz-continuous kernel with $2 + \delta$ moments for some $\delta > 0$, $\{X_n^*\}_{n \in \mathbb{N}}$ be a stationary bootstrapped strong mixing process with $\mathbb{E}[|X_1^*|^\xi] < \infty$ for some $\xi > 0$ and $\alpha^*(M) = O(M^{-\rho})$*

with $\rho > \frac{3\xi\delta + 2\delta + 5\xi + 2}{2\xi\delta}$, then for $V_M^*(\tilde{g}) = \frac{2}{\sqrt{M}(M-1)} \sum_{1 \leq i < j \leq M} \tilde{g}(X_i^*, X_j^*)$:

$$\mathbb{E}\left[MV_M^{*2}(\tilde{g})\right] = O(M^{-\eta}),$$

where $\eta = \min\left\{\rho\frac{2\xi\delta}{3\xi\delta + \delta + 5\xi + 2} - 1, 1\right\} > 0$.

The proof of this lemma can be found in Hwang and Shin (2015) as the proof for Lemma 2.

As shown in Lemma 3.A.2, the variance of the last term in (3.A.4) vanishes as M increases:

$$Var^*\left(\frac{1}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} \sum_{j=\lfloor uM \rfloor + 1}^{M} \tilde{g}(\ell_i^*, \ell_j^*)\right) \xrightarrow{P} 0.$$

Thus, by Lemma of Slutsky and the property of kernel shown in (3.A.2), it is enough to show that

$$\left(\frac{M - \lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i^*) + \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{j=\lfloor uM \rfloor + 1}^{M} h_2(\ell_j^*)\right)_{0 \leq u \leq 1}$$

converges in distribution to the limit process of

$$\left(\frac{M - \lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i) + \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{j=\lfloor uM \rfloor + 1}^{M} h_2(\ell_j)\right)_{0 \leq u \leq 1}.$$

Because the kernel $h_W(\ell_i^*, \ell_j^*)$ is antisymmetric, we have that $h_2(\ell_j^*) = -h_1(\ell_j^*)$. Thus, we can rewrite the representation as:

$$\frac{M - \lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i^*) - \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{i=\lfloor uM \rfloor + 1}^{M} h_1(\ell_i^*) = \frac{1}{M^{1/2}} \sum_{i=1}^{\lfloor uM \rfloor} h_1(\ell_i^*) - \frac{\lfloor uM \rfloor}{M^{3/2}} \sum_{i=1}^{M} h_1(\ell_i^*).$$

To obtain the limit of the process, we state the theorems of Calhoun (2018).

Theorem 3.A.2. *Let $\{Y_k\}_{k\in\mathbb{Z}}$ be a L_2 near-epoch dependent (NED) with respect to a strong mixing process. Additionally, suppose that $\mu_{nt} - \bar{\mu}_n$ is uniformly bounded, where $\mu_{nt} = \mathbb{E}[Y_{nt}]$ and $\bar{\mu}_n = n^{-1}\sum_{t=1}^n \mu_{nt}$. Then we have:*

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(\bar{Y}_n^* - \mathbb{E}^*[\bar{Y}_n^*]) \le x) - P(\sqrt{n}(\bar{Y}_n - \mathbb{E}[\bar{Y}_n]) \le x) \right| \xrightarrow{P} 0,$$

where $\bar{Y}_n = \frac{1}{n}\sum_{t=1}^n Y_{nt}$, and $\bar{Y}_n^* = \frac{1}{n}\sum_{t=1}^n Y_{nt}^*$.

Theorem 3.A.3. *Suppose that the conditions of Theorem 3.A.2 hold and let d be any distance function that metricizes weak convergence. Then we have:*

$$P^*(d(Z_n^*, \sigma W) > \delta^*) \xrightarrow{P} 0, \qquad\qquad (3.A.5)$$

for all positive δ^*, where $Z_n^*(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un \rfloor}(Y_{nt}^* - \mathbb{E}^*[\bar{Y}_n^*])$, and σW denotes a Brownian motion scaled by the positive constant σ. If, in addition, $\sup_{t=1,\dots,n} |\mu_{nt} - \bar{\mu}_n| = o(n^{-1/2})$ and

$$n^{-1} \sum_{s,t=1}^{\lfloor \xi^* n \rfloor} Cov(Y_{ns}, Y_{nt}) \to \sigma^2 \xi^*$$

for all $\xi^* \in [0,1]$, then

$$P^*(d(Z_n, \sigma W) > \delta^*) \xrightarrow{P} 0, \qquad\qquad (3.A.6)$$

for any positive δ^*, where $Z_n(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un \rfloor}(Y_{nt} - \bar{\mu}_n)$.

If both (3.A.5) and (3.A.6) hold, then the distribution of bootstrapped values Z_n^* can be used to approximate the distribution of Z_n, because they have the same distribution asymptotically.

The assumptions listed in Theorem 1 of Calhoun (2018) are satisfied under Assumptions 3.2.3 and 3.3.1 in our study. Applying Theorem 3.A.2 and Theorem 3.A.3 for $h_1(\ell_t)$, we have:

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{\lfloor uM\rfloor}{M^{3/2}}\sum_{i=1}^{M}h_1(\ell_i^*)\leq x\right)-P\left(\frac{\lfloor uM\rfloor}{M^{3/2}}\sum_{i=1}^{M}h_1(\ell_i)\leq x\right)\right|\xrightarrow{P}0,$$

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{1}{M^{1/2}}\sum_{i=1}^{\lfloor uM\rfloor}h_1(\ell_i^*)\leq x\right)-P\left(\frac{1}{M^{1/2}}\sum_{i=1}^{\lfloor uM\rfloor}h_1(\ell_i)\leq x\right)\right|\xrightarrow{P}0,$$

Thus, based on the Lemma of Slutsky, we have:

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{1}{M^{1/2}}\sum_{i=1}^{\lfloor uM\rfloor}h_1(\ell_i^*)-\frac{\lfloor uM\rfloor}{M^{3/2}}\sum_{i=1}^{M}h_1(\ell_i^*)\leq x\right)\right.$$
$$\left.-P\left(\frac{1}{M^{1/2}}\sum_{i=1}^{\lfloor uM\rfloor}h_1(\ell_i)-\frac{\lfloor uM\rfloor}{M^{3/2}}\sum_{i=1}^{M}h_1(\ell_i)\leq x\right)\right|\xrightarrow{P}0,$$

Thus, we obtain the convergence in probability in (3.3.2):

$$\sup_{x\in\mathbb{R}}\left|P^*(M^{-3/2}U_M^*(u)\leq x)-P(M^{-3/2}U_M(u)\leq x)\right|\xrightarrow{P}0.$$

$\square$

3.B   Tables and Figures

Table 3.B.1: Rejection rates of tests on the properties of simulated loss series

	Ljung-Box	ARCH	ADF	Grubb's
ℓ^{FZ0}	0.167	0.063	1.000	1.000
ℓ^{FZ1}	0.194	0.085	1.000	1.000
ℓ^{FZ2}	0.113	0.042	1.000	1.000

Note: We simulate returns using the DGP of GARCH(1,1)-skewed t model ($\beta_0 = 0.05$, $\beta_1 = 0.9$, $\beta_2 = 0.05$, $\nu = 16.5$, $\lambda = -0.5$), then use the given model to estimate 1% VaR and ES, and substitute into the FZ loss functions to calculate the loss series. We run the simulation 1000 times with sample size of 3000. We calculate the rejection frequency as the number of rejections divided by the number of simulations. We show the rejection frequencies obtained from the Ljung-Box test for autocorrelation, Engle's ARCH test for conditional heteroskedasticity, the Augmented Dickey-Fuller (ADF) test for unit root and Grubbs's test for the existence of outliers respectively, with 10 lags. All tests are performed at 5% test significance level.

Figure 3.B.1: Autocorrelation function (ACF) plots for the risk measures (VaR, ES), loss values, and the square of loss values
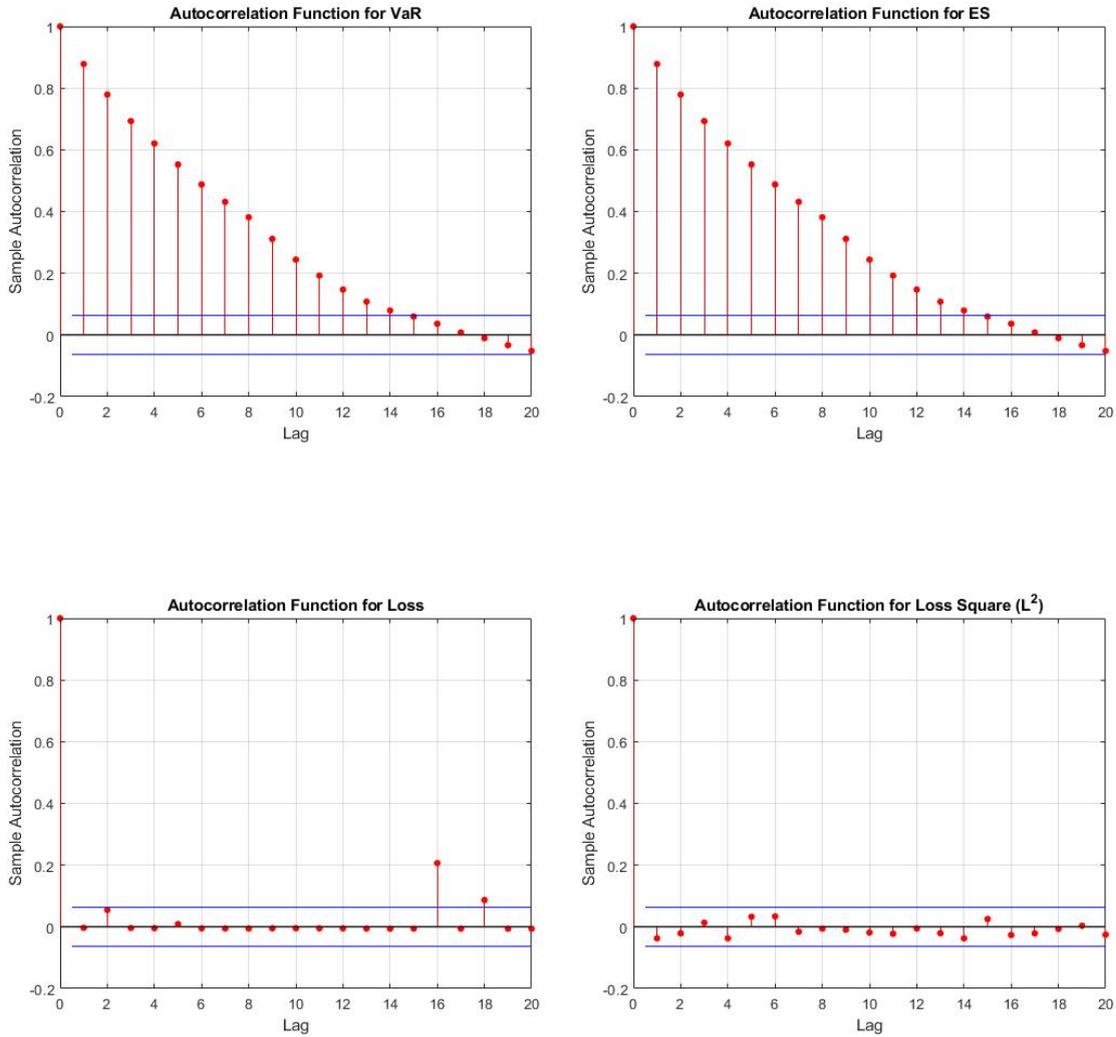
Note: This figure displays the autocorrelation function of the risk measures, the loss series, and the square of loss series under the null hypothesis. We simulate returns using the DGP of GARCH(1,1)-skewed t model ($\beta_0 = 0.05$, $\beta_1 = 0.9$, $\beta_2 = 0.05$, $\nu = 16.5$, $\lambda = -0.5$), then use the given model to estimate 1% VaR and ES, and substitute into the FZ loss functions to calculate the loss series. We run the simulation 1000 times with sample size of 3000.

Table 3.B.2: Empirical size for $q = 5\%$ of the loss-based Wilcoxon test using the asymptotic distribution with long-run variance estimators

	M=1000			M=3000		
	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid
Bartlett estimator with bandwidth $b_{BT} = \lfloor 4(M/100)^{2/9} \rfloor$						
ℓ^{FZ0}	0.747	0.739	0.733	0.736	0.737	0.736
ℓ^{FZ1}	0.747	0.740	0.733	0.736	0.737	0.736
ℓ^{FZ2}	0.747	0.739	0.733	0.736	0.737	0.736
Bartlett estimator with bandwidth $b_{BT} = 0.02M$						
ℓ^{FZ0}	0.315	0.313	0.373	0.184	0.173	0.241
ℓ^{FZ1}	0.315	0.313	0.371	0.184	0.173	0.241
ℓ^{FZ2}	0.315	0.313	0.368	0.184	0.173	0.240
Bartlett estimator with bandwidth $b_{BT} = 0.05M$						
ℓ^{FZ0}	0.107	0.112	0.126	0.079	0.080	0.106
ℓ^{FZ1}	0.107	0.112	0.124	0.079	0.080	0.106
ℓ^{FZ2}	0.107	0.112	0.123	0.079	0.080	0.105
DFSVW estimator with bandwidth $b_{DFSVW} = \lfloor 4(M/100)^{2/9} \rfloor$						
ℓ^{FZ0}	0.764	0.759	0.770	0.707	0.707	0.731
ℓ^{FZ1}	0.764	0.759	0.768	0.707	0.707	0.731
ℓ^{FZ2}	0.764	0.759	0.765	0.707	0.707	0.731
DFSVW estimator with bandwidth $b_{DFSVW} = 0.02M$						
ℓ^{FZ0}	0.229	0.230	0.301	0.096	0.096	0.102
ℓ^{FZ1}	0.229	0.230	0.300	0.096	0.096	0.102
ℓ^{FZ2}	0.229	0.230	0.301	0.096	0.096	0.102
DFSVW estimator with bandwidth $b_{DFSVW} = 0.05M$						
ℓ^{FZ0}	0.075	0.075	0.102	0.039	0.039	0.049
ℓ^{FZ1}	0.075	0.075	0.100	0.039	0.039	0.049
ℓ^{FZ2}	0.075	0.075	0.100	0.039	0.039	0.049

Note: We simulate returns using the DGP of GARCH(1,1)-skewed t model ($\beta_0 = 0.05$, $\beta_1 = 0.9$, $\beta_2 = 0.05$, $\nu = 16.5$, $\lambda = -0.5$). Empirical size for $q = 5\%$ of the loss-based Wilcoxon test under the null hypothesis is generated via 1000 simulations, for three types of risk measures (GARCH(1,1)-skewed t, GARCH(1,1)-Gaussian and GAS-Hybrid) and three FZ loss functions with different degrees of positive homogeneity. We consider two sample sizes: 1000 and 3000, and VaR and ES are computed at $\alpha = 1\%$ level. In this test, we implement the loss-based Wilcoxon test for a change point by using the asymptotic distribution with two long-run variance estimators: the Bartlett estimator and the DFSVW estimator of Dehling et al. (2013a). The bandwidth $b_{BT} = \lfloor 4(M/100)^{2/9} \rfloor$ for the Bartlett estimator according to Newey and West (1994); and the bandwidths $0.02T$ and $0.05T$ for the DFSVW estimator are recommended by Dehling et al. (2013a).

Table 3.B.3: Empirical size and power of the loss-based Wilcoxon test for a change point by bootstrapping loss series directly

	$\pi = 0.5$						$\pi = 0.75$					
	M=1000			M=3000			M=1000			M=3000		
	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid	G-Skt	G-G	Hybrid
H_0: Univariate GARCH(1,1)-skewed t, with $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 16.5, -0.5)$												
ℓ^{FZ0}	0.165	0.170	0.176	0.095	0.091	0.096	0.165	0.170	0.176	0.095	0.091	0.096
ℓ^{FZ1}	0.173	0.175	0.177	0.093	0.095	0.097	0.173	0.175	0.177	0.093	0.095	0.097
ℓ^{FZ2}	0.177	0.171	0.181	0.107	0.091	0.091	0.177	0.171	0.181	0.107	0.091	0.091
H_1^{A1}: An increase of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.94$												
ℓ^{FZ0}	0.998	0.999	0.979	1.000	1.000	1.000	0.914	0.920	0.780	0.998	0.999	0.973
ℓ^{FZ1}	0.997	0.999	0.978	1.000	1.000	1.000	0.919	0.920	0.785	0.999	0.999	0.977
ℓ^{FZ2}	0.997	0.999	0.980	1.000	1.000	1.000	0.922	0.922	0.774	0.999	0.999	0.978
H_1^{A2}: A decrease of 0.04 in the volatility persistence parameter, i.e. $\beta_1^* = 0.86$												
ℓ^{FZ0}	0.799	0.796	0.672	0.996	0.995	0.961	0.423	0.434	0.364	0.791	0.793	0.669
ℓ^{FZ1}	0.798	0.803	0.674	0.994	0.994	0.960	0.422	0.429	0.373	0.792	0.793	0.659
ℓ^{FZ2}	0.793	0.801	0.670	0.995	0.994	0.958	0.426	0.422	0.374	0.793	0.782	0.672
H_1^{B1}: An increase of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.09$												
ℓ^{FZ0}	0.940	0.942	0.886	0.999	0.999	0.989	0.740	0.748	0.592	0.952	0.962	0.889
ℓ^{FZ1}	0.945	0.944	0.885	0.999	0.999	0.993	0.741	0.749	0.593	0.959	0.960	0.878
ℓ^{FZ2}	0.939	0.945	0.888	0.999	0.999	0.990	0.744	0.753	0.589	0.953	0.961	0.883
H_1^{B2}: A decrease of 0.04 in the volatility reaction parameter, i.e. $\beta_2^* = 0.01$												
ℓ^{FZ0}	0.819	0.831	0.688	0.998	0.996	0.944	0.434	0.429	0.363	0.820	0.816	0.688
ℓ^{FZ1}	0.826	0.831	0.691	0.997	0.997	0.947	0.437	0.429	0.367	0.812	0.811	0.691
ℓ^{FZ2}	0.824	0.831	0.691	0.996	0.997	0.948	0.427	0.434	0.372	0.808	0.817	0.689
H_1^{C1}: A decrease of 13.5 in the DoF parameter, i.e. $\nu^* = 3$												
ℓ^{FZ0}	0.614	0.622	0.512	0.851	0.854	0.465	0.358	0.359	0.365	0.515	0.515	0.389
ℓ^{FZ1}	0.617	0.625	0.517	0.853	0.853	0.462	0.358	0.363	0.360	0.520	0.511	0.384
ℓ^{FZ2}	0.620	0.631	0.508	0.854	0.854	0.463	0.359	0.359	0.365	0.509	0.514	0.393
H_1^{C2}: A decrease of 14 in the DoF parameter, i.e. $\nu^* = 2.5$												
ℓ^{FZ0}	0.897	0.897	0.648	0.987	0.988	0.623	0.598	0.588	0.541	0.866	0.866	0.587
ℓ^{FZ1}	0.899	0.900	0.649	0.986	0.988	0.622	0.595	0.599	0.537	0.863	0.877	0.593
ℓ^{FZ2}	0.899	0.901	0.645	0.988	0.989	0.622	0.604	0.592	0.545	0.870	0.864	0.592
H_0^{D}: 12 randomly selected returns in the simulated process multiplied by 5												
ℓ^{FZ0}	0.182	0.183	0.173	0.112	0.112	0.116	0.182	0.183	0.173	0.112	0.112	0.116
ℓ^{FZ1}	0.175	0.183	0.185	0.116	0.113	0.122	0.175	0.183	0.185	0.116	0.113	0.122
ℓ^{FZ2}	0.183	0.183	0.179	0.114	0.117	0.119	0.183	0.183	0.179	0.114	0.117	0.119

Note: Empirical size and power, for $q = 5\%$, of the loss-based Wilcoxon test under various hypotheses via 1000 simulations, for three types of 1% risk measures (GARCH(1,1)-skewed t, GARCH(1,1)-Gaussian and GAS-Hybrid) and three FZ loss functions with different degrees of positive homogeneity. We consider two sample sizes: 1000 and 3000, and different locations of the change point: $\lfloor 0.5M \rfloor$ and $\lfloor 0.75M \rfloor$. In this test, we bootstrap the loss series directly, instead of resampling returns.

Table 3.B.4: Estimation Methods

| Models	Latent variables	Output
Parametric methods:		
GARCH(1,1)-Gaussian (Bali and Theodossiou, 2007)	$\hat{\sigma}_t^2 = \beta_0 + \beta_1\hat{\sigma}_t^2 + \beta_2 r_{t-1}^2$	$v_{\alpha,t} = \mu_t + F_G^{-1}(\alpha) \cdot \hat{\sigma}_t$ $e_{\alpha,t} = \mu_t + \frac{1}{\alpha}f_G(F_G^{-1}(\alpha)) \cdot \hat{\sigma}_t$
GARCH(1,1)-skewed t (Patton et al., 2019)	$\hat{\sigma}_t^2 = \beta_0 + \beta_1\hat{\sigma}_t^2 + \beta_2 r_{t-1}^2$	$v_{\alpha,t} = \mu_t + F_{skt}^{-1}(\alpha) \cdot \hat{\sigma}_t$ $e_{\alpha,t} = \mu_t + \frac{1}{\alpha}f_{skt}(F_{skt}^{-1}(\alpha)) \cdot \hat{\sigma}_t$
Semi-parametric methods:		
Hybrid (Patton et al., 2019)	$\kappa_t = \beta_0 + \beta_1\kappa_{t-1} + \beta_2(-\frac{1}{e_{t-1}}(\frac{1}{\alpha}\mathbf{1}\{r_t \le v_t\}r_{t-1} - e_{t-1})) + \beta_3 \log|r_{t-1}|$	$v_{\alpha,t} = a \cdot \exp\{\kappa_t\}$ $e_{\alpha,t} = b \cdot \exp\{\kappa_t\}$

Table 3.B.5: Empirical size and power of the loss-based Wilcoxon test for a change point in risk measures estimated by historical simulations

|	0.5M		0.75M	
	M=1000	M=3000	M=1000	M=3000
H_0: $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 16.5, -0.5)$				
ℓ^{FZ0}	0.064	0.042	0.064	0.042
ℓ^{FZ1}	0.092	0.075	0.092	0.075
ℓ^{FZ2}	0.076	0.065	0.076	0.065
H_1^{A1}: $\beta_1^* = 0.94$				
ℓ^{FZ0}	0.878	0.999	0.418	0.800
ℓ^{FZ1}	0.862	0.999	0.375	0.805
ℓ^{FZ2}	0.875	0.999	0.455	0.804
H_1^{A2}: $\beta_1^* = 0.86$				
ℓ^{FZ0}	0.479	0.914	0.270	0.429
ℓ^{FZ1}	0.415	0.913	0.213	0.417
ℓ^{FZ2}	0.424	0.917	0.255	0.431
H_1^{B1}: $\beta_2^* = 0.09$				
ℓ^{FZ0}	0.715	0.987	0.365	0.728
ℓ^{FZ1}	0.652	0.985	0.318	0.731
ℓ^{FZ2}	0.680	0.988	0.263	0.721
H_1^{B2}: $\beta_2^* = 0.01$				
ℓ^{FZ0}	0.490	0.951	0.252	0.439
ℓ^{FZ1}	0.479	0.945	0.253	0.434
ℓ^{FZ2}	0.521	0.952	0.252	0.448
H_1^{C1}: $\nu^* = 3$				
ℓ^{FZ0}	0.191	0.419	0.180	0.337
ℓ^{FZ1}	0.165	0.448	0.157	0.341
ℓ^{FZ2}	0.138	0.400	0.131	0.328
H_1^{C2}: $\nu^* = 2.5$				
ℓ^{FZ0}	0.229	0.336	0.173	0.312
ℓ^{FZ1}	0.233	0.339	0.177	0.310
ℓ^{FZ2}	0.228	0.337	0.178	0.313
H_0^D: 12 randomly selected returns multiplied by 5				
ℓ^{FZ0}	0.073	0.050	0.073	0.050
ℓ^{FZ1}	0.090	0.075	0.090	0.075
ℓ^{FZ2}	0.071	0.069	0.071	0.069

Note: Empirical size and power, for $q = 5\%$, of the loss-based Wilcoxon test under various hypotheses via 1000 simulations, for the 1% risk measures estimated by historical simulations and three FZ loss functions with different degrees of positive homogeneity. We consider two sample sizes: 1000 and 3000, and different locations of the change point: $\lfloor 0.5M \rfloor$ and $\lfloor 0.75M \rfloor$.

Figure 3.B.2: Power curves of the loss-based Wilcoxon test and alternative tests
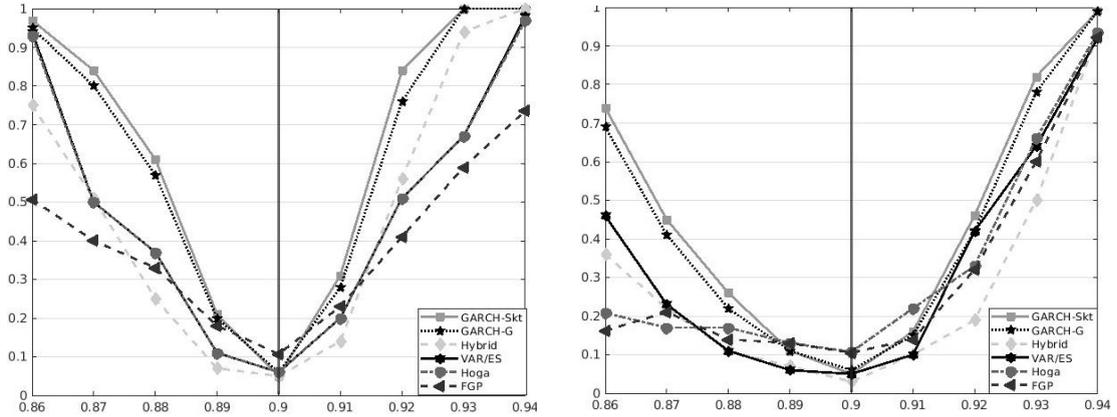
(a) The change in β_1, $\pi = 0.5$

(b) The change in β_1, $\pi = 0.75$
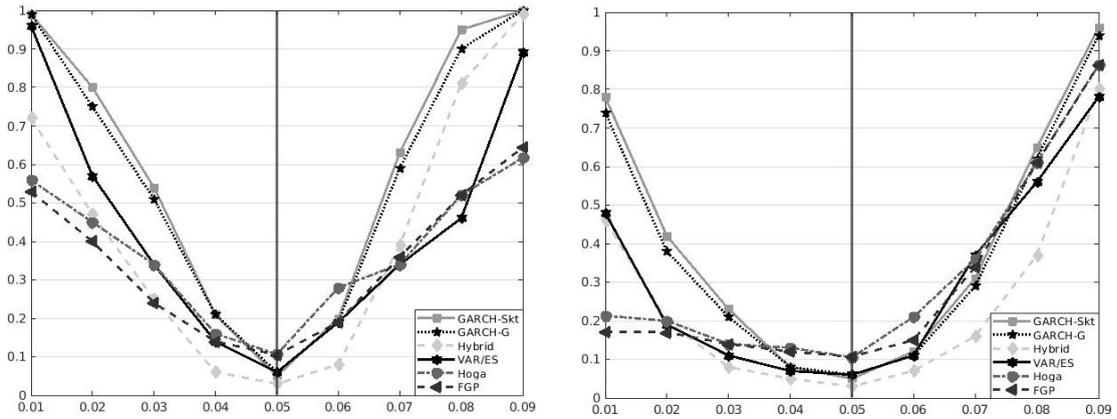
(c) The change in β_2, $\pi = 0.5$

(d) The change in β_2, $\pi = 0.75$

Note: Power curves (empirical size and power) of the loss-based Wilcoxon test for three (semi-)parametric estimation methods with alternative tests under various hypothesis at 5% test significance level in the given GARCH(1,1)-skewed t process via 1000 simulations. For the loss-based Wilcoxon test, we choose the loss values computed by the FZ0 loss function with 1% VaR and ES estimated by the GARCH(1,1)-skewed t, GARCH(1,1)-Gaussian, and GAS-Hybrid model (denoted by GARCH-Skt, GARCH-G and Hybrid in legend). Regarding the self-normalized CUSUM test for 1% VaR and ES, we use the 1% VaR and ES estimated by GARCH(1,1)-skewed t (denoted by VaR/ES in legend). Because the estimation results for VaR and ES are the same with each other, we select one of them. Panel (a) and (b) consider the change in parameter β_1 from 0.90 to 0.86 (and 0.94) when the change point occurs at $\lfloor 0.5M \rfloor$ and $\lfloor 0.75M \rfloor$, respectively; Panel (c) and (d) consider the change in parameter β_2 from 0.05 to 0.01 (and 0.09) when the change point occurs at $\lfloor 0.5M \rfloor$ and $\lfloor 0.75M \rfloor$, respectively.

Figure 3.B.3: Change point tests for AR(1) process and ARCH(1) process
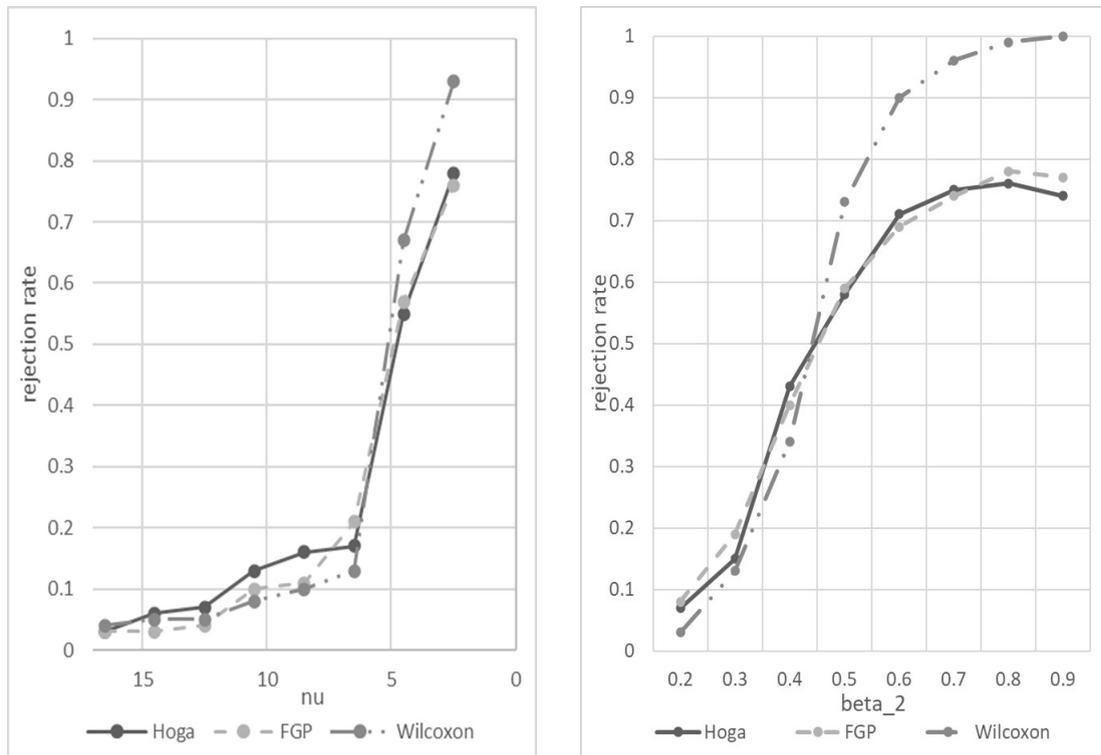
Note: Left: Change point tests for AR(1) process. Relationship between empirical detection probability and degrees of freedom ν of t-distributed innovations after the change point. Before the change point, $\nu = 16.5$. Right: Change point tests for ARCH(1) process. Relationship between empirical detection probability and ARCH(1) parameter β_2 after the change point. Before the change point, $\beta_2 = 0.2$. In both cases, change point testing is conducted with 5% test significance level using risk measures at the 1% level. For both the AR(1) and ARCH(1) processes, the abrupt change occurs at $\lfloor 0.5M \rfloor$. Each plotted point is an average over 1000 replications, and the sample size is 3000. We simulate AR(1) process: $r_t = 0.5r_{t-1} + t(\nu)$ and ARCH(1) process: $r_t = \sqrt{1 + \beta_2 r_{t-1}^2}\, u_{t-1}$, where u_{t-1} follows the Gaussian distribution. Under the null hypothesis, we set $\nu = 16.5$ and $\beta_1 = 0.2$. Under the alternative hypothesis, we adjust the parameter values ν and β_2 in the two processes from 16.5 to 2.5 and 0.2 to 0.9, respectively. For the loss-based Wilcoxon test, we choose the loss values computed by the FZ0 loss function with VaR and ES estimated by the GARCH(1,1)-skewed t model.

Figure 3.B.4: Size and power of the loss-based Wilcoxon test and alternatives with the change point at $\lfloor 0.75M \rfloor$
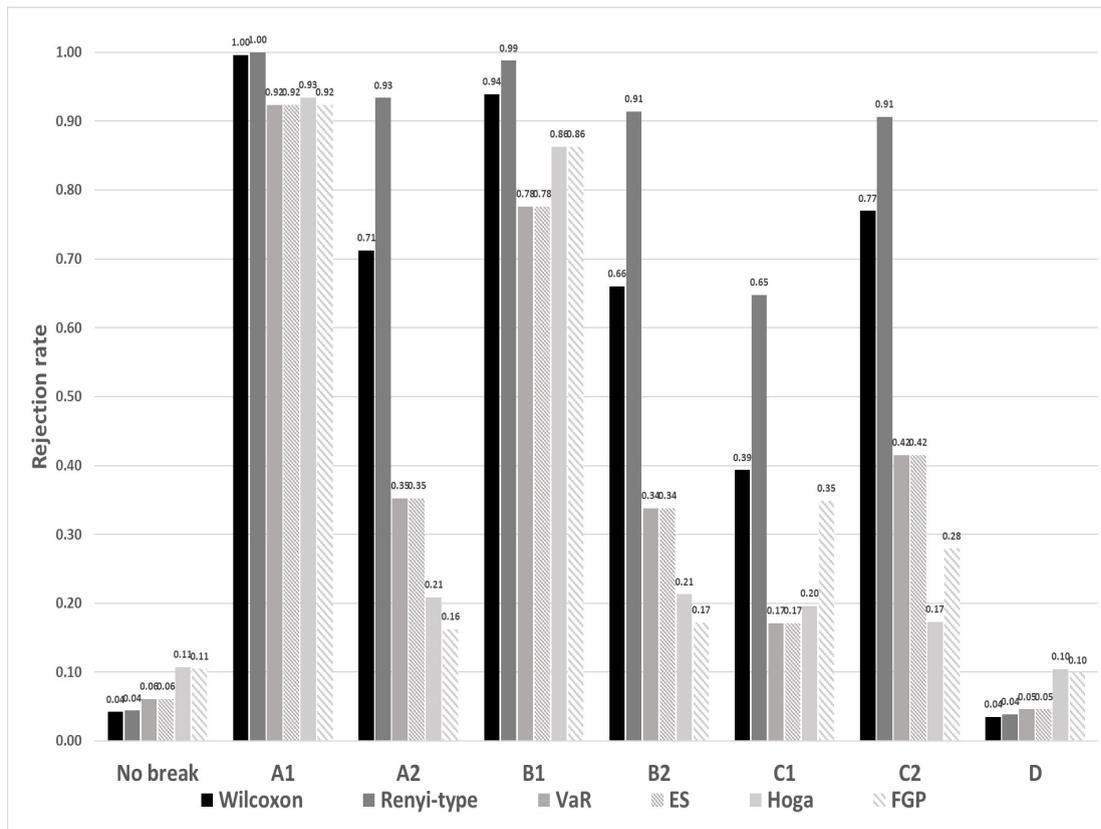
Note: Empirical size and power of the loss-based Wilcoxon test with five alternative tests under various hypotheses at 5% test significance level when the change point occurs at $\lfloor 0.75M \rfloor$ via 1000 simulations. For the Wilcoxon test and Rényi-type test, we use the FZ0 loss function to compute the loss values. For all tests except Hoga and FGP, 1% VaR and ES are estimated by the GARCH(1,1)-skewed t model.

Figure 3.B.5: Sequence of the loss-based Wilcoxon algorithm based on the GARCH(1,1)-skewed t model, applied to the S&P 500 index
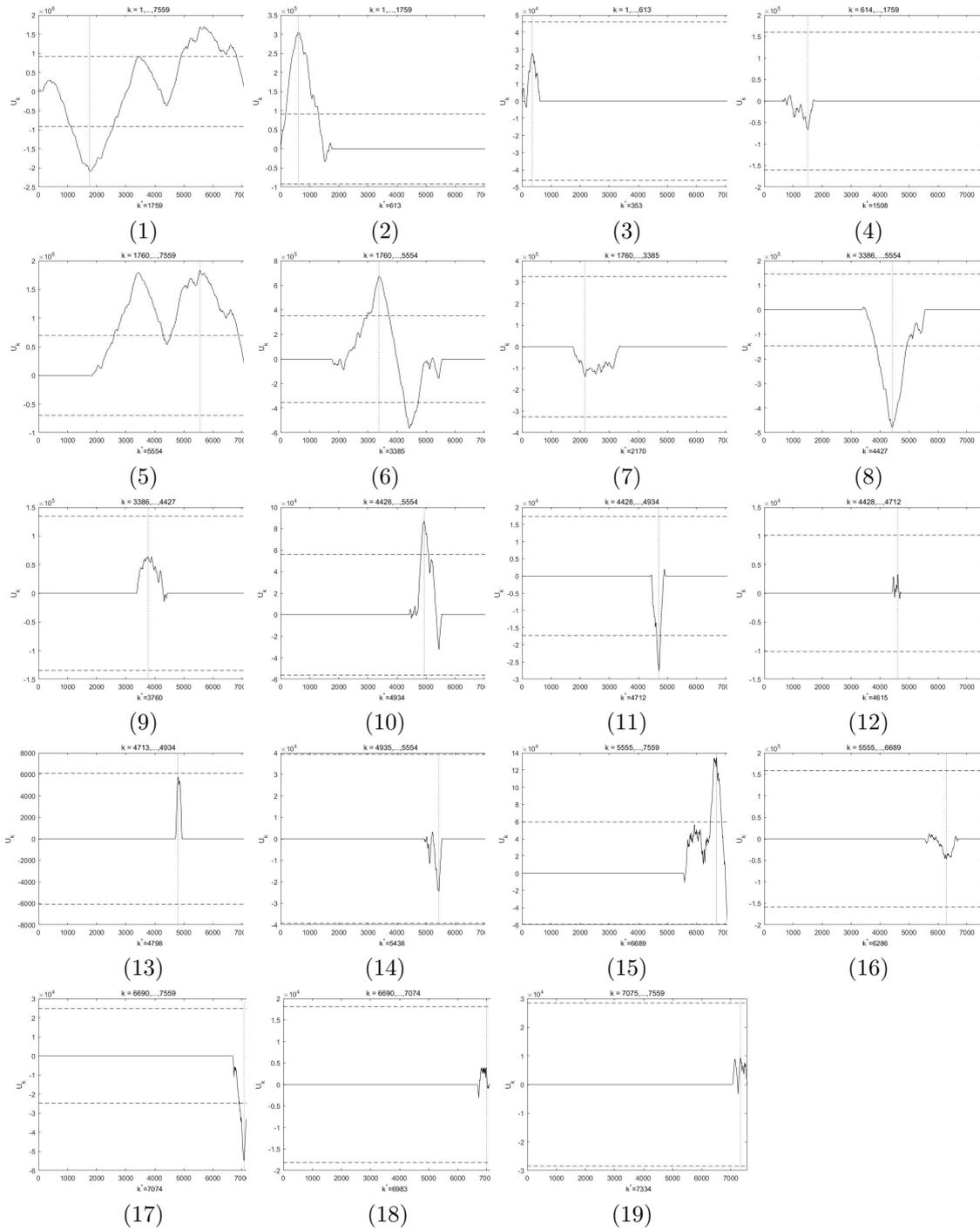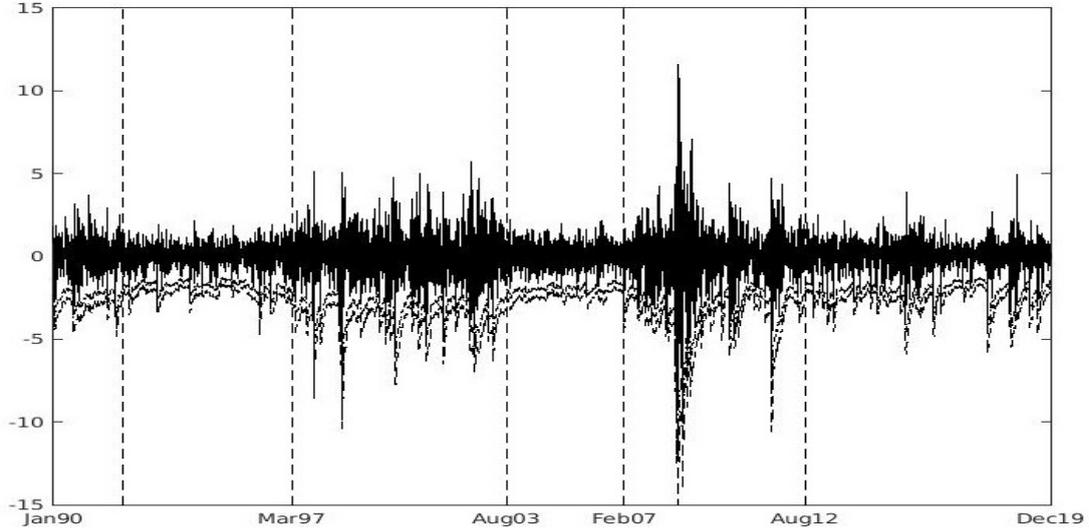
Figure 3.B.6: Daily returns and 1% (VaR, ES) estimated by the GAS-Hybrid
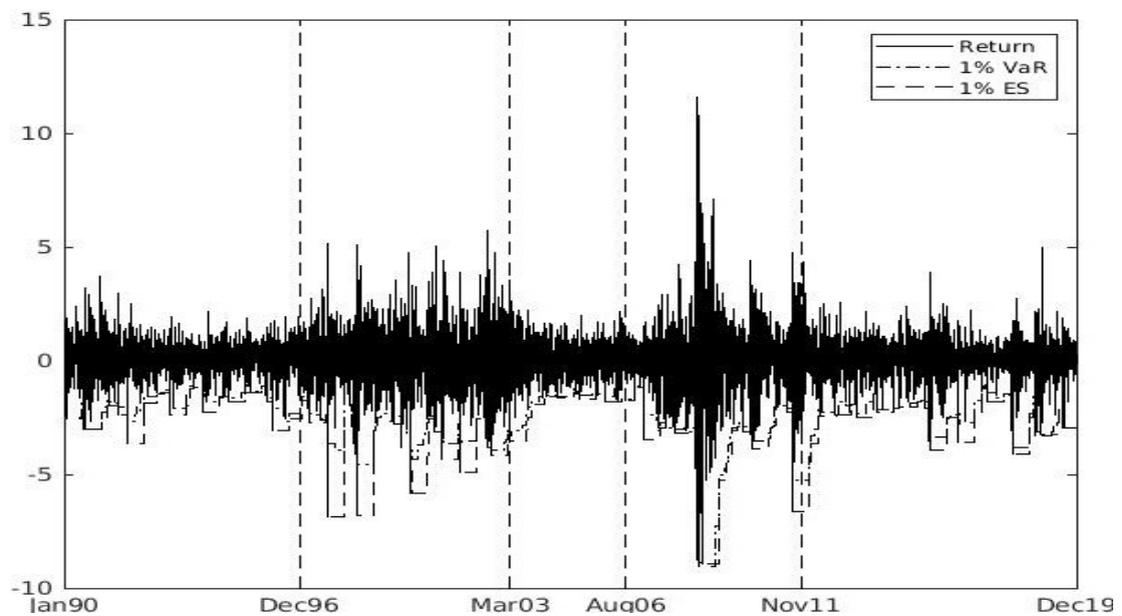

Note: The vertical dash lines are the detected change points by using the binary segmentation method.

Table 3.B.6: Subsample estimation results of GAS-Hybrid model

	1990/01-1992/01	1992/01-1997/02	1997/02-2003/08	2003/08-2007/02	2007/02-2012/08	2012/08-2019/12
β_1	0.820(0.097)	0.976(0.007)	0.912(0.083)	0.835(0.382)	0.942(0.068)	0.798(0.039)
β_2	0.012(0.005)	0.000(0.001)	0.005(0.004)	0.001(0.001)	0.006(0.004)	0.008(0.001)
β_3	1.09E-08(1.66E-08)	0.017(0.001)	0.006(0.006)	0.006(0.005)	0.028(0.049)	0.027(0.008)
a	-2.017(5.905)	-3.883(13.324)	-3.236(14.339)	-1.584(9.298)	-4.164(30.9862)	-2.493(0.624)
b	-2.825(8.271)	-4.919(16.960)	-4.314(19.165)	-1.785(10.441)	-4.843(37.294)	-3.074(0.879)
VaR	-2.465	-1.815	-3.216	-1.928	-3.312	-2.162
ES	-3.136	-2.309	-4.091	-2.452	-4.214	-2.751
Loss	1.036	0.684	1.411	0.515	1.299	0.973
Loss_NC	1.149	0.716	1.493	0.688	1.394	1.031

Note: Estimated parameter values for β_1, β_2, and β_3 in the GAS-Hybrid model: $\kappa_t = \beta_0 + \beta_1\kappa_{t-1} + \beta_2(-\frac{1}{e_{t-1}}(\frac{1}{\alpha}\mathbf{1}\{r_t \leq v_t\}r_{t-1} - e_{t-1})) + \beta_3 \log|r_{t-1}|$, $v_t = a \cdot \exp\{\kappa_t\}$, $e_t = b \cdot \exp\{\kappa_t\}$ for the S&P 500 index in 6 sub-periods. We also report the average VaR and ES at 1% level and the associated average loss values using the parameters estimated within the sub-periods (Loss) and the average loss using parameters estimated over the whole sample period without consideration of change points (Loss_NC).

Figure 3.B.7: Daily returns and 1% (VaR, ES) estimated by historical simulations

Note: The vertical dash lines are the estimated change points by the loss-based Wilcoxon test based on historical simulations with a rolling window of size 125.

Table 3.B.7: Change points detected using various tests in the S&P 500

Type of test	Loss value based tests				VaR based tests				ES based tests			
Event	G-Skt	GAS-H	HS	Renyi-G-Skt	G-Skt	GAS-H	HS	Hoga	G-Skt	GAS-H	HS	FGP
Early 1990s recession	3-Jun-92	28-Jan-92	-	16-Jul-91	9-Mar-92	28-Jan-92	-	-	9-Mar-92	28-Jan-92	-	-
Dot-com bubble (boom)	12-Dec-96	28-Feb-97	23-Dec-96	11-Dec-96	27-Mar-97	11-Apr-97	27-Oct-97	-	27-Mar-97	11-Apr-97	27-Oct-97	-
Dot-com bubble (burst)	3-Jun-03	15-Aug-03	24-Mar-03	2-Jun-03	2-Jun-03	18-Jun-03	19-Sep-03	-	2-Jun-03	18-Jun-03	19-Sep-03	-
Subprime mortgage crisis	24-Jul-07	26-Feb-07	24-Aug-06	7-Jun-07	7-Nov-07	27-Feb-07	27-Feb-07	-	7-Nov-07	27-Feb-07	27-Feb-07	-
Bankruptcy of Lehman Brothers	9-Sep-08	-	-	9-Sep-08	17-Sep-08	15-Sep-08	-	-	17-Sep-08	15-Sep-08	-	-
European debt crisis (start)	28-Jul-09	-	-	28-Jul-09	8-Jun-09	5-Jun-09	-	-	8-Jun-09	5-Jun-09	-	-
European debt crisis (end)	11-Jan-12	13-Aug-12	1-Nov-11	9-Jan-12	6-Jan-12	4-Jan-12	2-May-12	-	6-Jan-12	4-Jan-12	10-May-12	-
Stock market selloff	18-Jul-16	-	-	25-Jul-16	-	-	-	-	-	-	-	-
Cryptocurrency crash	26-Jan-18	-	-	17-Jan-18	-	-	-	-	-	-	-	-
Additional breaks				14-Jul-11	2-Aug-11	17-Jan-08	8-Mar-96	-	2-Aug-11	17-Jan-08	8-Mar-96	-
							12-Jan-11				27-Dec-10	
							8-Aug-11				4-Aug-11	
							21-Aug-15				21-Aug-15	

Note: "-" denotes no break point for this event. We use the loss-based Wilcoxon tests for GARCH(1,1)-skewed t, GAS-Hybrid and historical simulations, the Renyi-type Wilcoxon test for GARCH(1,1)-skewed t, self-normalized CUSUM tests for VaR and ES, and the tests proposed Hoga and FGP. *Additional breaks* in the first column means the detected breaks cannot be matched with economic events.

Notes

[1] This class of semiparametric models has been extended to incorporate the intraday or high-frequency information (Meng and Taylor, 2020; Lazar and Xue, 2020; Gerlach and Wang, 2020) and combine with networks (Bonaccolto et al., 2022).

[2] There is no (strictly) consistent loss function for ES that does not also contain VaR (Fissler and Ziegel, 2016).

[3] In order to construct the Wilcoxon test statistic, we initially obtain the ranks of loss values and then feed the ranks into the CUSUM procedure. More details can be found in Section 2.2.

[4] A loss function ℓ is called positive homogeneous of degree $\tilde{b}$ if for all r, v and e, $L(\tilde{c}r, \tilde{c}v, \tilde{c}e) = \tilde{c}^{\tilde{b}}L(r, v, e)$, for all $\tilde{c} > 0$.

[5] The simulation setup and results are reported in Table 3.B.1 of Appendix 3.B. We use the Ljung-Box test, Engle's ARCH test, the Augmented Dickey-Fuller (ADF) test and Grubb's test.

[6] Boucher et al. (2014) and Lazar and Zhang (2019) discuss that the model risk of risk measures can be decomposed into estimation bias and model misspecification. When using the true model (the one used to generate the data process) with the true parameters to estimate risk measures, model risk is not present.

[7] For simplicity, in this chapter we consider the alternative hypothesis that there exists only one change point k^* occurring in the series.

[8] In this chapter, we follow Hoga (2017) to set the expected block length as $0.08T$, which can consistently produce satisfactory results in various settings. It is possible to select the optimal block length for stationary bootstrapping, please see Politis and White (2004) and Patton et al. (2009) for more details.

[9] We found that resampling the loss series $\{\ell_t(r_t^*)\}_{t=1}^M$ directly would lead to a higher empirical size, especially for small sample sizes (the related simulation results can be found in Table 3.B.3 of Appendix 3.B).

[10] These sample sizes are in line with the sample sizes used in the literature on risk measurement (see Patton et al., 2019).

[11] Results for $\alpha = 5\%$ are consistent with the results reported here, and are available upon request.

[12] We are aware that these values of ν^* mean that the fourth moment of the simulated returns does not exist. Nevertheless, these values of ν^* are useful for illustrative purposes. The literature

considers DGPs with less than four finite moments, such as in Berkes et al. (2003).

[13]More details about the models can be found in Table 3.B.4 of Appendix 3.B.

[14]Here, we only consider the case of $q = 5\%$; the results for other values of q are available on request.

[15]If model misspecification risk is present, then the ordering of models is affected by the choice of loss function (Patton, 2020); in this case, the size and power properties of our proposed test might be affected by the choice of loss function.

[16]We use the FZ0 loss function to compute the loss values for GARCH(1,1)-skewed t risk estimates. We implement the stationary bootstrapping based MC simulation for the Rényi-type loss-based Wilcoxon test, instead of applying the asymptotic distribution that assumes normally distributed loss values.

[17]We choose the historical quantile to estimate VaR and ES, in line with Fan et al. (2018), instead of applying the Weissman estimator for VaR used by Hoga (2017). Based on our simulations, the critical value at 5% test significance level is 80.21 for $u_0 = 0.2$, which is very close to the one given by Hoga (2017).

[18]In Table 3.B.5 of Appendix 3.B, we show that the loss-based Wilcoxon test has strong power in detecting the change point in the series of VaR and ES estimated by historical simulations. In Figure 3.B.2 we present the power curves of this test for three (semi)parametric models with alternative tests to show the detection power in terms of the marginal change in parameters. Figure 3.B.3 compares our test with alternative tests in terms of size and power for AR(1) and ARCH(1) processes, which are the DGPs used by Fan et al. (2018). Our results are consistent with the results in Table 3.4.1 and 3.4.3.

[19]In Figure 3.B.4 of Appendix 3.B we present the results when the break occurs at $\lfloor 0.75M \rfloor$.

[20]Between August 2015 and early 2016, the S&P 500 and DJIA dropped more than 10% twice.

[21]The S&P 500 index dropped almost 20% between September and December 2018.

Chapter 4

Sequential Monitoring for Changes in M-estimators of Risk Models

4.1 Introduction

In the ongoing recession following the COVID-19 pandemic and past financial crises, there is an increasing demand for more effective risk measures. Risk measures have become essential in supporting asset management decisions for banks and other financial institutions, especially under market turmoil. Value-at-Risk (VaR) and Expected Shortfall (ES) are two prevailing measures of financial risk that dominate current financial regulation. VaR measures how much a certain portfolio can lose at a given significance level within a given period. As a supplementary measure to VaR, ES captures the expected value of exceedances beyond the quantile. In the current literature, the estimation and forecasting approaches for joint VaR and ES can be classified into three main categories: nonparametric, semiparametric and parametric. Regarding the (semi)parametric models, the estimation method is based on the theory for M-estimators (see White (1996) and Newey and McFadden (1994) for example).

Ignoring change points in model parameters will likely lead to biased statistical inference and inaccurate forecasts. This situation has often been encountered in time series analysis. A strand of literature considers tests for parameter stability in a wide class of both linear and nonlinear parametric models (Andrews, 1993; Stock and Watson, 1996; Bai and Perron, 1998). Empirically, the misleading results caused by change points in time series has been documented in stock returns (Smith and Timmermann, 2021), conditional and unconditional variance (Andreou and Ghysels, 2002; Inclan and Tiao, 1994), correlation dynamics (Barassi et al., 2020), quantile regression (Qu, 2008), VaR (Hoga, 2017) and (semi)parametric models for risk measures (Lazar et al., 2021). We refer the readers to Hansen (2001) and Aue and Horváth (2013) for a detailed literature review of the change point detection methods for historical observations.

However, most literature mentioned above is designed to detect change points within a given historical dataset. In addition to this paradigm of historical detection of change points, another practical research question is whether newly arriving data is consistent with a well-trained relationship between series based on historical data because change points in the model can trigger such inconsistencies. As a milestone in the literature of real-time detection, Chu et al. (1996) propose an innovative test for changes in a time series based on a sequential detector and a boundary function. This monitoring scheme detects a change point when the proposed detector exceeds the boundary curve. Following this seminal work, several studies document the use of sequential monitoring in other models, e.g., the GARCH (p,q) models (Berkes et al., 2004), the functional linear models with dependent errors (Aue et al., 2014) and the dynamic linear models for real estate prices (Horváth et al., 2021b). Additionally, Horváth et al. (2020a) develop sequential monitoring procedures for changes from stationarity to mild non-stationarity of a time series. Horváth et al. (2021a) propose a sequentially monitoring scheme for a change point in a sequence of distributions. Regarding sequential monitoring for changes in the tail behaviour of time series, Hoga and

Wied (2017) propose a real-time detection procedure for changes in the nonparametric tail index of weakly dependent random variables.

In academia and the financial industry, (semi)parametric models are widely applied to forecast VaR and ES jointly instead of nonparametric methods. Unlike parametric models, which require a conditional distribution of returns for estimation and prediction, the construction of semiparametric models eliminate the need to specify and estimate a conditional density but rely on minimizing a specified loss function (Engle and Manganelli, 2004; Patton et al., 2019). The timing of parameter adjustments for risk models has become an essential part of risk management. However, sequentially detecting change points in risk models for VaR and ES has remained unexplored in the current literature.

To fill this gap, our contribution is that we propose a sequential monitoring scheme to detect changes in the parameter values of (semi)parametric models. This procedure is based on evaluating the gradients of the FZ loss function (Fissler and Ziegel, 2016) with respect to (w.r.t.) the model parameters, instead of relying solely on the time series of risk measures. Our detection procedure mainly follows Chu et al. (1996) where a change is detected when a proposed detector exceeds a defined boundary function. In our case, the detector is based on the cumulative sum process of gradients of the FZ loss function. The boundary curve is chosen such that the probability of a false detection under the null hypothesis of stable parameters is fixed. Our proposed detector is similar to two tests for change points proposed by Qu (2008) and Berkes et al. (2004). A unifying view of the fluctuation-type statistic based on the gradient estimated from historical samples is presented in Qu (2008). However, this test is proposed for detecting change points *ex-post* instead of real-time detection. This gradient-based test for quantile regression can be extended to sequential monitoring for change points in nonlinear risk models for VaR and ES. This chapter is also relevant to Berkes et al. (2004), who derive a sequential test for the changes in the parameters of a GARCH sequence. However, we propose an extension of their test by considering risk

models for VaR and ES jointly.

Additionally, using Monte Carlo (MC) simulations, we show the advantages of the proposed sequential monitoring test in identifying change points in risk models. Under the null hypothesis, our test is shown to have suitable size control in finite samples. Regarding the alternative hypotheses, we consider changes in VaR and ES caused by the varying second to fourth moments (variance, skewness and kurtosis) individually, and our test has high power under those various scenarios.

Lastly, we empirically demonstrate the practical usage of our proposed test on risk measures of the S&P 500 index returns and the GBP/EUR exchange rate returns. We present evidence that the test can detect change points in risk models earlier than the financial crisis when we apply the test to the S&P 500 index returns. Also, the change points detected by the test are consistent with well-known market events, such as the Black Monday in the US stock market caused by the COVID-19 pandemic and the sterling depreciation after the Brexit referendum.

The chapter is structured as follows: Section 4.2 formulates the detection problem in the framework of sequential change point hypothesis testing and presents some theoretical results related to the asymptotic distribution of the proposed test statistic; Section 4.3 discusses the MC simulation setup under various scenarios, and presents the simulation results; Section 4.4 contains the empirical applications based on the S&P 500 index and the GBP/EUR exchange rate return; and Section 4.5 concludes the chapter.

4.2 Sequential Monitoring for Changes in Semi-parametric Risk Models

4.2.1 Semiparametric Models Formulation

We start by briefly reviewing the general risk model for VaR and ES estimation and forecasting. Suppose that we observe a series of asset returns in a training (historical) sample $\{r_t\}_{t=1}^{M}$. Let v_t^{α} and e_t^{α} denote the VaR and ES at a specified significance level α. Then for $\alpha \in (0,1)$, the conditional VaR and ES of y are given by

$$v_t^{\alpha} \equiv F^{-1}(\alpha|\mathscr{F}_{t-1}) = \inf\{r_t \in \mathbb{R}|F(r_t|\mathscr{F}_{t-1}) \geq \alpha\}, \qquad (4.2.1)$$

and

$$e_t^{\alpha} \equiv \mathbb{E}[r_t|r_t \leq v_t^{\alpha}, \mathscr{F}_{t-1}], \qquad (4.2.2)$$

respectively, where $F(\cdot|\mathscr{F}_{t-1})$ is the cumulative distribution function of observations r_t over a horizon given the information set $\mathscr{F}_{t-1}$.

Let $\boldsymbol{x}_t$ be a $(K \times 1)$ vector of time t exogenous variables given the information set $\mathscr{F}_t$, and let $\boldsymbol{\theta}^{\alpha}$ be a $(d \times 1)$ vector of unknown parameters. Then, we let $v_t^{\alpha}(\boldsymbol{\theta}^{\alpha}) \equiv v(\boldsymbol{\theta}^{\alpha}, \boldsymbol{x}_{t-1})$ and $e_t^{\alpha}(\boldsymbol{\theta}^{\alpha}) \equiv e(\boldsymbol{\theta}^{\alpha}, \boldsymbol{x}_{t-1})$ denote the estimated VaR and ES at the significance level α at time t, given past information up to time $t-1$. In the following, we suppress the superscript α from $\boldsymbol{\theta}^{\alpha}$, v_t^{α} and e_t^{α} for notational convenience.

Patton et al. (2019) define a generic dynamic semiparametric model for conditional VaR and ES at significance level α in the following framework:

$$\begin{bmatrix} v_t(\boldsymbol{\theta}) \\ e_t(\boldsymbol{\theta}) \end{bmatrix} \equiv \begin{bmatrix} v(r_{t-1}, h^*(\boldsymbol{x}_{t-1}), \dots, r_1, h^*(\boldsymbol{x}_1); \boldsymbol{\theta}) \\ e(r_{t-1}, h^*(\boldsymbol{x}_{t-1}), \dots, r_1, h^*(\boldsymbol{x}_1); \boldsymbol{\theta}) \end{bmatrix}, \quad t = 1, \dots, M, \quad (4.2.3)$$

where $h^*(\cdot)$ is a function that links the quantiles $v_t(\boldsymbol{\theta})$ and $e_t(\boldsymbol{\theta})$ to exogenous variables that belong to the information set.

Unlike using (Quasi-) maximum likelihood estimation ((Q)MLE) to estimate the parameters of a parametric model, the estimation of the unknown parameters $\boldsymbol{\theta}$ for a semiparametric risk model is based on the FZ0 loss function introduced by Fissler and Ziegel (2016):

$$\hat{\ell}_t\left(\boldsymbol{u}\right) = -\frac{1}{\alpha \hat{e}_t\left(\boldsymbol{u}\right)}\mathbf{1}\left\{r_t \leq \hat{v}_t(\boldsymbol{u})\right\}(\hat{v}_t(\boldsymbol{u}) - r_t) + \frac{\hat{v}_t(\boldsymbol{u})}{\hat{e}_t(\boldsymbol{u})} + \log\left(-\hat{e}_t(\boldsymbol{u})\right) - 1, \quad (4.2.4)$$

where $\boldsymbol{u}$ is a possible set of parameter values in the generic parameter space Θ, $\boldsymbol{u} \in \Theta \subset \mathbb{R}^d$. We also have that $\boldsymbol{\theta} \in \Theta$. For a given set $\boldsymbol{u}$ of parameter values, $\hat{v}_t(\boldsymbol{u})$ and $\hat{e}_t(\boldsymbol{u})$ are the estimated VaR and ES at time t.

The estimator for the unknown parameter $\boldsymbol{\theta}$ is obtained by

$$\hat{\boldsymbol{\theta}}_M = \arg\min\{\hat{L}_M\left(\boldsymbol{u}\right) : \boldsymbol{u} \in \Theta\}, \quad (4.2.5)$$

where

$$\hat{L}_M\left(\boldsymbol{u}\right) = \frac{1}{M}\sum_{1 \leq t \leq M}\hat{\ell}_t\left(\boldsymbol{u}\right). \quad (4.2.6)$$

When we construct our detector function based on the estimated parameters, we are exposed to parameter estimation uncertainty. This uncertainty could cause misleading detection results, for example, spurious break detection or biased detected locations. Thus, in this chapter, we would like to note that neglecting parameter estimation uncertainty in the calculation of the detector in finite samples means that we are effectively assuming that we have the true value. Patton et al. (2019) show that the M-estimation of these parameters via FZ loss minimization leads to a consistent and asymptotically normal estimator. The conditions for the proof are provided in the following section.

4.2.2 Sequential Monitoring for Change Points

Consider a financial return time series from which we have observed a "stable" historical sample of length M, $r_1, ..., r_M$, which is referred to the training sample. Based on this training sample, we can estimate a semiparametric model for (VaR, ES). As additional observations of the series, $r_{M+1}, r_{M+2}, ...$ are revealed in an online manner, we are interested in detecting whether the parameters of the semiparametric model for (VaR, ES) have changes in the incoming observations.

To sequentially monitor for changes in the parameters of a risk model, the procedure has the null hypothesis that:

$$H_0 : \boldsymbol{\theta}_{M+k} = \boldsymbol{\theta}, \quad 1 \leq k < \infty,$$

against the alternative hypothesis that a change in the parameters occurred at k^* and the parameters after k^* are $\boldsymbol{\theta}^*$:

$$H_A : \exists k^* \in \mathbb{N} : \boldsymbol{\theta}_{M+k} = \begin{cases} \boldsymbol{\theta}, & 1 \leq k \leq k^*, \\ \boldsymbol{\theta}^*, & k^* + 1 \leq k < \infty. \end{cases}$$

To test the hypotheses, we sequentially evaluate the change in the mean of gradient of the FZ loss function w.r.t. parameters $\boldsymbol{\theta}$. We construct a test based on the intuition that a parameter change must have occurred if the gradient of the loss function persistently deviate from zero. If such a change occurs in the monitoring horizon, then the parameter estimates based on the training sample cannot characterize the incoming observations, leading to a non-zero expectation of gradient. Consequently, it is no longer valid to use the parameters estimated based on the training sample to make risk forecasts after the change.

Here, we define $\hat{\boldsymbol{\ell}}'_t(\boldsymbol{u})$ as the vector of the first-order derivatives of the FZ0

loss function w.r.t. the parameters of the model at t:

$$
\begin{aligned}
\hat{\ell}'_t(\boldsymbol{u}) = \frac{\partial \hat{\ell}_t(\boldsymbol{u})}{\partial \boldsymbol{u}} = \hat{v}'_t(\boldsymbol{u}) \frac{1}{-\hat{e}_t(\boldsymbol{u})} &\left(\frac{1}{\alpha} \mathbf{1}\{r_t \le \hat{v}_t(\boldsymbol{u})\} - 1 \right) \\
&+ \hat{e}'_t(\boldsymbol{u}) \frac{1}{\hat{e}_t(\boldsymbol{u})^2} \left(\frac{1}{\alpha} \mathbf{1}\{r_t \le \hat{v}_t(\boldsymbol{u})\}(\hat{v}_t(\boldsymbol{u}) - r_t) - \hat{v}_t(\boldsymbol{u}) + \hat{e}_t(\boldsymbol{u}) \right),
\end{aligned}
$$

$$(4.2.7)$$

where $\hat{v}'_t(\boldsymbol{u})$ and $\hat{e}'_t(\boldsymbol{u})$ denote the first-order derivative of $\hat{v}_t(\boldsymbol{u})$ and $\hat{e}_t(\boldsymbol{u})$ w.r.t. the set of parameters $\boldsymbol{u}$, respectively.

To derive the asymptotic results for this test, we consider the following assumption for the time series of the gradient.

Assumption 4.2.1. $\{\ell'_t(\boldsymbol{u})\}$ *is a stationary ergodic martingale difference sequence.*

A key observation is that under the null hypothesis of no structural change, in the sequence $\{\ell'_t(\boldsymbol{u})\}$, $\mathbf{1}\{r_t \le \hat{v}_t(\boldsymbol{u})\}$ is a pivotal statistic, i.e., a sequence of independent binary random variables with mean α and variance $\alpha(1-\alpha)$.

Then, we consider the covariance matrix estimator:

$$
\hat{\boldsymbol{D}}_M = \frac{1}{M} \sum_{1 \le i \le M} \hat{\ell}'_i\left(\hat{\boldsymbol{\theta}}_M\right)^\top \hat{\ell}'_i\left(\hat{\boldsymbol{\theta}}_M\right),
\tag{4.2.8}
$$

where $^\top$ denotes the transpose of a vector.

The monitoring scheme follows from the nonnegative definite property of the matrix $\hat{\boldsymbol{D}}_M$ given in (4.2.8). In the following, we show that $\hat{\boldsymbol{D}}_M$ is nonsingular with probability tending to one as $M \to \infty$. Hence $\hat{\boldsymbol{D}}_M^{-1/2}$ exists with probability tending to one as $M \to \infty$.

We also define:

$$
\boldsymbol{D}(\boldsymbol{u}) = \mathbb{E}[\ell'_0(\boldsymbol{u})^\top \ell'_0(\boldsymbol{u})],
$$

and

$$
\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{\theta}).
\tag{4.2.9}
$$

To derive the main theoretical results, we consider the assumptions below.

Assumption 4.2.2. *(A) The loss sequence $\{\hat{\boldsymbol{\ell}}_t(\boldsymbol{u})\}$ obeys the uniform law of large numbers;*

(B) Θ is a compact subset of $\mathbb{R}^d$ for $d < \infty$;

(C) $\{r_t\}_{t=1}^{\infty}$ is a strictly stationary process. Conditional on all the past information $\mathscr{F}_{t-1}$, the distribution of r_t is $F(\cdot|\mathscr{F}_{t-1})$ which, for all t, belongs to a class of distribution functions on $\mathbb{R}$ with finite first moments and the unique α-quantiles.

(D) For any t, both $v_t(\boldsymbol{u})$ and $e_t(\boldsymbol{u})$ are $\mathscr{F}_{t-1}$-measurable and a.s. continuous in $\boldsymbol{u}$;

(E) If $P\left((v_t(\boldsymbol{u}) = v_t(\boldsymbol{\theta})) \cap (e_t(\boldsymbol{u}) = e_t(\boldsymbol{\theta}))\right) = 1 \; \forall t$, then $\boldsymbol{u} = \boldsymbol{\theta}$.

Assumption 4.2.3. *(A) For all t, we have (i) $v_t(\boldsymbol{u})$ and $e_t(\boldsymbol{u})$ are a.s. twice continuously differentiable in $\boldsymbol{u}$, (ii) $e_t(\boldsymbol{\theta}) < v_t(\boldsymbol{\theta}) \leq 0$;*

(B) For all t, we have (i) conditional on all the past information $\mathscr{F}_{t-1}$, r_t has a continuous density $f_t(\cdot|\mathscr{F}_{t-1})$ that satisfies $f_t(y|\mathscr{F}_{t-1}) \leq K < \infty$ and $|f_t(y'|\mathscr{F}_{t-1}) - f_t(y''|\mathscr{F}_{t-1})| \leq K|y' - y''|$, (ii) $\mathbb{E}\left[|r_t|^{4+\delta}\right] \leq K < \infty$, for some $0 < \delta < 1$;

(C) There exists a neighborhood of $\boldsymbol{\theta}$, $\mathscr{N}(\boldsymbol{\theta})$, such that for all t we have (i) $|1/e_t(\boldsymbol{u})| \leq K < \infty$, $\forall \boldsymbol{u} \in \mathscr{N}(\boldsymbol{\theta})$, (ii) there exist some (possibly stochastic) $\mathscr{F}_{t-1}$-measurable functions $V(\mathscr{F}_{t-1})$, $V_1(\mathscr{F}_{t-1})$, $H_1(\mathscr{F}_{t-1})$, $V_2(\mathscr{F}_{t-1})$, $H_2(\mathscr{F}_{t-1})$ that satisfy $\boldsymbol{u} \in \mathscr{N}(\boldsymbol{\theta}): |v_t(\boldsymbol{u})| \leq V(\mathscr{F}_{t-1}), ||v_t'(\boldsymbol{u})|| \leq V_1(\mathscr{F}_{t-1}), ||e_t'(\boldsymbol{u})|| \leq H_1(\mathscr{F}_{t-1}), ||v_t''(\boldsymbol{u})|| \leq V_2(\mathscr{F}_{t-1})$, and $||e_t''(\boldsymbol{u})|| \leq H_2(\mathscr{F}_{t-1})$;

(D) For $0 < \delta < 1$ and for all t we have (i) $\mathbb{E}\left[V_1(\mathscr{F}_{t-1})^{3+\delta}\right]$, $\mathbb{E}\left[H_1(\mathscr{F}_{t-1})^{3+\delta}\right]$, $\mathbb{E}\left[V_2(\mathscr{F}_{t-1})^{\frac{3+\delta}{2}}\right]$, $\mathbb{E}\left[H_2(\mathscr{F}_{t-1})^{\frac{3+\delta}{2}}\right] \leq t$, (ii) $\mathbb{E}\left[V(\mathscr{F}_{t-1})^{2+\delta}V_1(\mathscr{F}_{t-1})H_1(\mathscr{F}_{t-1})^{2+\delta}\right] \leq t$, (iii) $\mathbb{E}\left[H_1(\mathscr{F}_{t-1})^{1+\delta}H_2(\mathscr{F}_{t-1})|r_t|^{2+\delta}\right]$, $\mathbb{E}\left[H_1(\mathscr{F}_{t-1})^{3+\delta}|r_t|^{2+\delta}\right] \leq t$;

(E)
$$\boldsymbol{G} = \mathbb{E}\left[\frac{f_t(v_t(\boldsymbol{\theta})|\mathscr{F}_{t-1}))}{-e_t(\boldsymbol{\theta})\alpha}v_t'(\boldsymbol{\theta})^{\top}v_t'(\boldsymbol{\theta}) + \frac{1}{e_t(\boldsymbol{\theta})^2}e_t'(\boldsymbol{\theta})^{\top}e_t'(\boldsymbol{\theta})\right]$$

is a (strictly) positive definite, nonsingular matrix;

(F) $\{r_t, v_t(\boldsymbol{\theta}), e_t(\boldsymbol{\theta}), v'_t(\boldsymbol{\theta}), e'_t(\boldsymbol{\theta})\}$ is α-mixing with $\sum_{m=1}^{\infty} \alpha(m)^{(q-2)q} < \infty$ for some $q > 2$;

(G) $\sup_{\boldsymbol{u}} \sum_{t=1}^{M} \mathbf{1}\{r_t = v_t(\boldsymbol{u})\} \leq t$ a.s.

Assumption 4.2.4. *(A) The deterministic positive sequence c_T satisfies $c_T = o(1)$ and $c_T^{-1} = o(T^{1/2})$.*

(B) (i) $\frac{1}{M} \sum_{t=1}^{M} \boldsymbol{\ell}'_t(\boldsymbol{\theta})^\top \boldsymbol{\ell}'_t(\boldsymbol{\theta}) - \boldsymbol{D} \xrightarrow{p} \boldsymbol{0}$, where $\boldsymbol{D}$ is defined in (4.2.9).

(ii) $\frac{1}{M} \sum_{t=1}^{M} \frac{1}{e_t(\boldsymbol{\theta})^2} e'_t(\boldsymbol{\theta})^\top e'_t(\boldsymbol{\theta}) - \mathbf{E}\left[\frac{1}{e_t(\boldsymbol{\theta})^2} e'_t(\boldsymbol{\theta})^\top e'_t(\boldsymbol{\theta})\right] \xrightarrow{p} \boldsymbol{0}$.

(iii) $\frac{1}{M} \sum_{t=1}^{M} \frac{f_t(v_t(\boldsymbol{\theta})|\mathscr{F}_{t-1})}{-\alpha e_t(\boldsymbol{\theta})} v'_t(\boldsymbol{\theta})^\top v'_t(\boldsymbol{\theta}) - \mathbf{E}\left[\frac{f_t(v_t(\boldsymbol{\theta})|\mathscr{F}_{t-1})}{-\alpha e_t(\boldsymbol{\theta})} v'_t(\boldsymbol{\theta})^\top v'_t(\boldsymbol{\theta})\right] \xrightarrow{p} \boldsymbol{0}$.

Patton et al. (2019) derive the asymptotic normality of $\sqrt{M}\left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\right)$, if Assumptions 4.2.4 - 4.2.3 hold. Also, based on Theorem 3 in Patton et al. (2019), we have the following proposition:

Proposition 4.2.1. *If Assumptions 4.2.2 - 4.2.4 hold, then*

$$\hat{\boldsymbol{D}}_M - \boldsymbol{D} \xrightarrow{P} \boldsymbol{0}.$$

To simplify the monitoring scheme, we define a detector:

$$\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right) = \left\|\sum_{M < t \leq M+k} \hat{\boldsymbol{\ell}}'_t\left(\hat{\boldsymbol{\theta}}_M\right) \hat{\boldsymbol{D}}_M^{-1/2}\right\|_{\infty}, \quad (4.2.10)$$

and a boundary function:

$$g(M, k) = cM^{1/2}\left(1 + \frac{k}{M}\right) b\left(\frac{k}{M}\right), \quad (4.2.11)$$

where $\|\cdot\|_{\infty}$ denotes the maximum norm of a vector, i.e., for a generic vector $z = (z_1, \ldots, z_d)$, $\|z\|_{\infty} = \max\{|z_1|, \ldots, |z_d|\}$, where d measures the size of vector, c is the critical value and $b(\cdot)$ is a selected function, defined as:

$$b\left(\frac{k}{M}\right) = \left(\frac{k}{M+k}\right)^\gamma,$$

where $0 \leq \gamma < \frac{1}{2}$ (as in Chu et al., 1996; Horváth et al., 2020a, 2021a).

Here, we define the stopping time $\hat{k}^*$ as:

$$\hat{k}^* = \min \left\{ k : \Gamma \left(M, k; \hat{\boldsymbol{\theta}}_M \right) > cM^{1/2} \left(1 + \frac{k}{M} \right) b \left(\frac{k}{M} \right) \right\}, \qquad (4.2.12)$$

If $\hat{k}^* < \infty$, we say that a change occurs, so that under the null hypothesis,

$$\lim_{M \to \infty} P_{H_0} \left(\hat{k}^* < \infty \right) = \lim_{M \to \infty} P_{H_0} \left(\frac{\Gamma \left(M, k; \hat{\boldsymbol{\theta}}_M \right)}{g(M, k)} > 1 \text{ for some } k \geq 1 \right) = q, \tag{4.2.13}$$

where $0 < q < 1$ is a prescribed significance level, and under the alternative hypothesis,

$$\lim_{M \to \infty} P_{H_A} \left(\hat{k}^* < \infty \right) = \lim_{M \to \infty} P_{H_A} \left(\frac{\Gamma \left(M, k; \hat{\boldsymbol{\theta}}_M \right)}{g(M, k)} > 1 \text{ for some } k \geq 1 \right) = 1. \tag{4.2.14}$$

We impose the following conditions on the function $b(t)$:

Assumption 4.2.5.

$$b(t) \text{ is continuous on } (0, \infty), \tag{4.2.15}$$

and

$$\inf_{0 < t < \infty} b(t) > 0. \tag{4.2.16}$$

Theorem 4.2.1. *If Assumptions 4.2.1 - 4.2.5 hold, then*

$$\lim_{M \to \infty} P_{H_0}\{k^* < \infty\} = 1 - P \left(\sup_{0 \leq u \leq 1} \frac{|W(u)|}{b(u/(1-u))} \leq c \right)^d, \tag{4.2.17}$$

where $\{W(u), 0 \leq u \leq 1\}$ denotes a Wiener process.

An outline of the proof of this theorem is given in Appendix 4.B.

4.2.3 Critical Values

In this section, we will demonstrate how to obtain the critical value c in (4.2.17) and how to tune the detector and boundary to improve the false positive rate in finite samples.

Based on Theorem 4.2.1, the values of $c(\gamma, q)$ are defined below:

$$1 - P\left(\sup_{0 \leq u \leq 1} \frac{|W(u)|}{u^\gamma} \leq c(\gamma, q)\right)^d = q, \tag{4.2.18}$$

which depends on the selection of boundary functions with parameter γ, the prescribed significance level q and the dimension of the parameter vector d.

Table 4.2.1 provides the critical values $c(\gamma, q)$ defined in (4.2.18) based on 1,000,000 replications of $\sup_{0 \leq u \leq 1} |W(u)|/u^\gamma$. The Wiener process is approximated on a grid of 10,000 equally spaced points in $[0, 1]$.[1]

However, it is not realistic to monitor the change in an infinite horizon. In practice, the monitoring is done for $1 \leq k \leq T$. Thus, we consider the close-end procedure in which the length of the training sample M and the termination time T are asymptotically proportional.

Assumption 4.2.6. *The time to termination of the sequential procedure $T = T(M)$, and:*

$$\lim_{M \to \infty} \frac{T}{T + M} = \theta.$$

Remark 4.2.1. *Given Assumption 4.2.6, for $0 \leq \gamma < 1/2$, the scale transformation of the Wiener process gives:*

$$\sup_{0 \leq u \leq \theta} \frac{|W(u)|}{u^\gamma} = \sup_{0 \leq u^* \leq 1} \frac{|W(u^*\theta)|}{(u^*\theta)^\gamma} \overset{\mathcal{D}}{=} \theta^{1/2-\gamma} \sup_{0 \leq u^* \leq 1} \frac{|W(u^*)|}{u^{*\gamma}}.$$

We provide the distribution of $\theta^{1/2-\gamma}|W(u^*)|/u^{*\gamma}$ for $\gamma = 0, .15, .25, .35, .45$ and $.49$, $d = 4$ and $T = \tau M$, with $\tau \in \{0.2, 0.4, 0.6, 0.8, 1\}$ in Table 4.C.1 of Appendix 4.C.

Table 4.2.1: Critical values $c(\gamma, q)$ obtained via 1,000,000 simulations

	$\gamma = 0$			$\gamma = 0.15$		
q	0.10	0.05	0.01	0.10	0.05	0.01
d=3	2.378	2.630	3.142	2.444	2.691	3.198
d=4	2.480	2.725	3.226	2.544	2.785	3.283
d=5	2.558	2.796	3.289	2.622	2.857	3.341
	$\gamma = 0.25$			$\gamma = 0.35$		
q	0.10	0.05	0.01	0.10	0.05	0.01
d=3	2.513	2.755	3.254	2.633	2.866	3.349
d=4	2.613	2.850	3.337	2.728	2.960	3.433
d=5	2.688	2.919	3.400	2.801	3.028	3.497
	$\gamma = 0.45$			$\gamma = 0.49$		
q	0.10	0.05	0.01	0.10	0.05	0.01
d=3	2.925	3.143	3.594	3.280	3.491	3.928
d=4	3.013	3.226	3.672	3.365	3.572	4.001
d=5	3.081	3.288	3.730	3.430	3.633	4.055

Our preliminary results show that the monitoring scheme with the decision functions defined in (4.2.10) and (4.2.11) over rejects when H_0 holds. The oversized rejection rates are presented in the left panel of Table 4.3.2. To improve the false positive rates, we follow Horváth et al. (2006) to modify the detector function by normalizing the gradient by a sequentially updated covariance estimator $\tilde{\boldsymbol{D}}_{M,k}$, and we use this instead of the covariance estimator obtained based on the fixed initial M observations $\hat{\boldsymbol{D}}_M$. The sequentially updated covariance estimator is defined as:

$$\tilde{\boldsymbol{D}}_{M,k} = \frac{1}{M+k} \sum_{1 \leq t \leq M+k} \hat{\boldsymbol{\ell}}_t' \left(\hat{\boldsymbol{\theta}}_M \right)^\top \hat{\boldsymbol{\ell}}_t' \left(\hat{\boldsymbol{\theta}}_M \right). \tag{4.2.19}$$

Our modified detector function is:

$$\tilde{\Gamma} \left(M, k; \hat{\boldsymbol{\theta}}_M \right) = \left\| \sum_{M < t \leq M+k} \hat{\boldsymbol{\ell}}_t' \left(\hat{\boldsymbol{\theta}}_M \right) \tilde{\boldsymbol{D}}_{M,k}^{-1/2} \right\|_\infty. \tag{4.2.20}$$

Then, the corresponding version of the stopping time k^{**} can be defined as:

$$k^{**} = \min \left\{ k : \left\| \sum_{M < t \leq M+k} \hat{\boldsymbol{\ell}}_t' \left(\hat{\boldsymbol{\theta}}_M \right) \tilde{\boldsymbol{D}}_{M,k}^{-1/2} \right\|_\infty > M^{1/2} \left(1 + \frac{k}{M} \right) b \left(\frac{k}{M} \right) \right\}. \tag{4.2.21}$$

Remark 4.2.2. *If Assumptions 4.2.1 - 4.2.5 hold, then:*

$$\lim_{M \to \infty} P_{H_0}(k^{**} < \infty) = 1 - P \left(\sup_{0 \leq u \leq 1} \frac{|W(u)|}{b(u/(1-u))} \leq c \right)^d, \tag{4.2.22}$$

where $\{W(u), 0 \leq u \leq 1\}$ denotes a Wiener process.

Additionally, we follow Horváth et al. (2020a) and Horváth et al. (2021b) to tune the boundary function from (4.2.11) to:

$$\tilde{g}(M, k) = c \left(1 + \frac{\gamma}{M^{1/2}} \right) M^{1/2} \left(1 + \frac{k}{M} \right) \left(\frac{k}{k + M} \right)^\gamma. \tag{4.2.23}$$

Since the tuning term in (4.2.23) converges to 1:

$$\left(1 + \frac{\gamma}{M^{1/2}} \right) \xrightarrow{P} 1,$$

under the null hypothesis with the conditions of Theorem 4.2.1, we also have that:

$$\lim_{M \to \infty} P_{H_0} \left(\frac{\tilde{\Gamma} \left(M, k; \hat{\boldsymbol{\theta}}_M \right)}{\tilde{g}(M, k)} > 1 \text{ for some } k \geq 1 \right) = q, \tag{4.2.24}$$

and under the alternative hypothesis, we obtained that:

$$\lim_{M \to \infty} P_{H_A} \left(\frac{\tilde{\Gamma} \left(M, k; \hat{\boldsymbol{\theta}}_M \right)}{\tilde{g}(M, k)} > 1 \text{ for some } k \geq 1 \right) = 1. \tag{4.2.25}$$

4.2.4 Long-run Covariance Estimators

It is also possible to improve the finite sample performance by considering the weak dependence in $\{\boldsymbol{\ell}'_t(\boldsymbol{u})\}_{t=1}^{M}$ and using a long-run covariance given by:

$$\boldsymbol{D} = \sum_{j=-\infty}^{+\infty} \delta_j, \quad \delta_j = \mathbb{E}[\boldsymbol{\ell}'_t(\boldsymbol{u})^{\top} \boldsymbol{\ell}'_{t+j}(\boldsymbol{u})]. \tag{4.2.26}$$

The long-run covariance can be estimated by:

$$\hat{\boldsymbol{D}}_M = \frac{M}{M-1} \sum_{j=-M+1}^{M-1} k\left(\frac{j}{W}\right) \hat{\delta}_j(M), \tag{4.2.27}$$

and by the matrix below:

$$\tilde{\boldsymbol{D}}_{M+k} = \frac{M+k}{M+k-1} \sum_{j=-M-k+1}^{M+k-1} k\left(\frac{j}{W}\right) \hat{\delta}_j(M+k), \quad k \geq 0, \tag{4.2.28}$$

where $k(\cdot)$ is a real-valued kernel function, W is the bandwidth parameter, and $\hat{\delta}_j(M)$ is the sample autocovariance of $\{\boldsymbol{\ell}'_t(\boldsymbol{u})\}_{t=1}^{M}$. These estimators are extensions of (4.2.8) and (4.2.19).

In this chapter, we consider the following three kernel functions for the long-run covariance estimation:

(1) Bartlett (abbreviated by "BT"):

$$k_{BT}(z) = \begin{cases} 1 - |z|, & \text{for}|z| \leq 1, \\ 0, & \text{otherwise}, \end{cases} \tag{4.2.29}$$

(2) Truncated (abbreviated by "TR"):

$$k_{TR}(z) = \begin{cases} 1, & \text{for } |z| \leq 1, \\ 0, & \text{otherwise}, \end{cases} \tag{4.2.30}$$

(3) Quadratic spectral (abbreviated by "QS"):

$$k_{QS}(z) = \frac{25}{12\pi^2 z^2} \left(\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5) \right). \qquad (4.2.31)$$

We follow Newey and West (1994) to select the optimal bandwidth W for these kernels, shown as:

$$W_{BT} = \lfloor 4(M/100)^{2/9} \rfloor, \quad W_{TR} = \lfloor 4(M/100)^{1/5} \rfloor, \quad W_{QS} = \lfloor 4(M/100)^{2/25} \rfloor.$$

4.3 Monte Carlo Simulations

In this section, we investigate the finite sample performance of the proposed monitoring test in well-defined simulation setups, by illustrating how critical values are obtained, designing simulations and presenting the results on empirical size and power.

4.3.1 Simulation Design

Under the null hypothesis, we consider the following data generating process (DGP):

$$\begin{aligned} r_t &= \sigma_t u_t, \quad u_t \sim i.i.d. \ \mathcal{N}(0,1), \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2, \quad t = 1, ..., M, M+1, \ldots, T, \end{aligned} \qquad (4.3.1)$$

where r_t is the simulated return process generated as the product of innovation u_t, which follows the Gaussian distribution, and conditional volatility σ_t given by a GARCH(1,1) specification.

In our simulations, we use the boundary function $\tilde{g}(M, k)$ given in (4.2.23) with $\gamma = \{0, 0.15, 0.25, 0.35, 0.45, 0.49\}$, and the training sample sizes of $M = 1000$ corresponding to 4 years of daily returns. The large sample size enables us to consider

risk model estimations for quantiles as low as 1%, which are often used in risk management (Patton et al., 2019). In simulations, we consider the testing sample sizes of $T = \tau M$, with $\tau \in \{0.2, 0.4, 0.6, 0.8, 1\}$. The parameter values are set as $(\beta_0, \beta_1, \beta_2) = (0.05, 0.90, 0.05)$. Regarding the long-run covariance estimation, we consider the estimators with the following kernels: Bartlett, Truncated, and Quadratic Spectral, which are abbreviated as "BT", "TR" and "QS" in Table 4.3.2. The results are based on 5000 repetitions.

Under the alternative hypothesis, we consider the following DGP before the break $[1, M + k^*)$:

$$
\begin{aligned}
r_t &= \sigma_t u_t, \quad u_t \sim i.i.d. \ \mathcal{N}(0, 1), \\
\sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2, \quad 1 \le t < M + k^*,
\end{aligned} \tag{4.3.2}
$$

Regarding the post-break DGP, we consider the following setups:

- $H_{A,1} : r_t = \sigma_t u_t, \quad u_t \sim \mathcal{N}(0, 1), \quad \sigma_t^2 = \beta_0 + \beta_1^* \sigma_{t-1}^2 + \beta_2 y_{t-1}^2, \quad M + k^* \le t \le M + T$;

- $H_{A,2} : r_t = \sigma_t u_t, \quad u_t \sim t(\nu^*), \quad \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2, \quad M + k^* \le t \le M + T$;

- $H_{A,3} : r_t = \sigma_t u_t, \quad u_t \sim SN(\lambda^*), \quad \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2, \quad M + k^* \le t \le M + T$.

Under the hypothesis $H_{A,2}$, $t(\nu^*)$ denotes the Student's t distribution with the degree of freedom (DoF) parameter ν^*. Under the hypothesis $H_{A,3}$, $SN(\lambda^*)$ denotes the skewed Normal distribution with the skewness parameter λ^*. Regarding the location of change point in the monitoring horizon, we consider $k^* = 1$ and $0.5T$. The former choice indicates that the change happens at the beginning of the monitoring horizon, and the latter one means that the change occurs in the middle of the monitoring horizon.

We first elaborate on the behavior of the monitoring scheme under the various alternative scenarios described above. Under the alternative hypothesis $H_{A,1}$, the persistence parameter changes from $\beta_1 = 0.90$ in the GARCH(1,1) model to $\beta_1^* = 0.94$ after the training sample in increments of size 0.01, meaning that the unconditional variance level gradually increases from 1 to 5.

In the following setups, we consider changes in the conditional distribution of the returns. The alternative hypothesis $H_{A,2}$ considers a change in the heavy-tailedess of the underlying distribution of the DGP from the Gaussian distribution $\mathcal{N}(0,1)$ to the Student's t distribution $t(\nu^*)$ with the DoF parameters $\nu^* = \{9.5, \ldots, 4.5\}$. This means that the process becomes more heavy-tailed after the break. Additionally, under the alternative hypothesis $H_{A,3}$, we use the skewed Normal distribution with varying skewness parameters $\lambda^* = \{-0.1, \ldots, -0.5\}$ after the break to replace the original Gaussian distribution. In this scenario, the process becomes more negatively skewed after the training sample.

In the simulations, we employ the GARCH-FZ model proposed by Patton et al. (2019)[2] to forecast VaR and ES jointly at 5% significance level[3]:

$$
\begin{aligned}
v_t &= a \cdot \sigma_t, \\
e_t &= b \cdot \sigma_t, \quad b < a < 0, \\
\sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2,
\end{aligned}
\tag{4.3.3}
$$

where $\boldsymbol{\theta} = (\beta_1, \beta_2, a, b)$ is the parameter vector of this model.[4]

4.3.2 Simulation Results

This section presents the empirical size and power of the proposed sequential monitoring test under the null hypothesis and alternative hypotheses illustrated in Section 4.3.1.

Table 4.3.2 shows the empirical sizes of the proposed test for the DGP discussed above at test significance levels of 10%, 5% and 1% with a range of values

Table 4.3.2: Empirical size of $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ **for the GARCH-FZ model with** $\alpha = 5\%$

Kernel	q	$\gamma \setminus$ T	$\hat{\boldsymbol{D}}_M$					$\tilde{\boldsymbol{D}}_{M,k}$				
			0.2M	0.4M	0.6M	0.8M	M	0.2M	0.4M	0.6M	0.8M	M
Cov	10%	0	0.186	0.153	0.163	0.173	0.186	0.104	0.107	0.108	0.115	0.104
		0.15	0.155	0.144	0.135	0.148	0.155	0.098	0.104	0.098	0.097	0.098
		0.25	0.203	0.179	0.186	0.192	0.203	0.092	0.104	0.094	0.100	0.092
		0.35	0.202	0.181	0.191	0.197	0.202	0.160	0.164	0.158	0.156	0.160
		0.45	0.267	0.247	0.252	0.259	0.267	0.166	0.164	0.168	0.168	0.166
		0.49	0.251	0.225	0.236	0.243	0.251	0.152	0.150	0.151	0.152	0.152
	5%	0	0.132	0.109	0.124	0.132	0.132	0.062	0.072	0.075	0.062	0.062
		0.15	0.111	0.100	0.102	0.105	0.111	0.054	0.075	0.065	0.056	0.054
		0.25	0.166	0.138	0.140	0.158	0.166	0.054	0.065	0.063	0.061	0.054
		0.35	0.160	0.145	0.156	0.157	0.160	0.108	0.120	0.117	0.112	0.108
		0.45	0.215	0.201	0.208	0.209	0.215	0.127	0.128	0.127	0.130	0.127
		0.49	0.223	0.204	0.214	0.216	0.223	0.131	0.133	0.133	0.132	0.131
	1%	0	0.069	0.066	0.062	0.072	0.069	0.023	0.034	0.025	0.024	0.023
		0.15	0.066	0.059	0.059	0.062	0.066	0.018	0.026	0.026	0.023	0.018
		0.25	0.096	0.084	0.093	0.093	0.096	0.021	0.030	0.027	0.020	0.021
		0.35	0.105	0.095	0.093	0.100	0.105	0.049	0.066	0.059	0.051	0.049
		0.45	0.146	0.142	0.142	0.142	0.146	0.084	0.088	0.086	0.084	0.084
		0.49	0.178	0.163	0.173	0.176	0.178	0.089	0.091	0.091	0.091	0.089
BT	10%	0	0.152	0.153	0.173	0.183	0.188	0.092	0.090	0.083	0.085	0.085
		0.15	0.141	0.144	0.141	0.156	0.160	0.089	0.084	0.076	0.073	0.072
		0.25	0.167	0.192	0.193	0.204	0.214	0.105	0.083	0.081	0.074	0.071
		0.35	0.193	0.203	0.207	0.206	0.215	0.151	0.146	0.143	0.142	0.141
		0.45	0.245	0.252	0.260	0.266	0.270	0.158	0.152	0.152	0.154	0.156
		0.49	0.220	0.242	0.255	0.266	0.270	0.143	0.152	0.149	0.148	0.148
	5%	0	0.116	0.120	0.127	0.139	0.146	0.061	0.056	0.048	0.044	0.047
		0.15	0.103	0.111	0.112	0.120	0.127	0.058	0.048	0.049	0.040	0.040
		0.25	0.124	0.152	0.156	0.164	0.168	0.069	0.046	0.046	0.040	0.036
		0.35	0.159	0.153	0.162	0.168	0.172	0.115	0.101	0.101	0.098	0.093
		0.45	0.201	0.210	0.218	0.222	0.228	0.130	0.128	0.122	0.116	0.116
		0.49	0.192	0.215	0.228	0.236	0.242	0.112	0.111	0.111	0.110	0.110
	1%	0	0.066	0.067	0.075	0.083	0.076	0.029	0.022	0.015	0.016	0.016
		0.15	0.063	0.066	0.066	0.070	0.074	0.016	0.020	0.016	0.010	0.010
		0.25	0.074	0.087	0.101	0.102	0.104	0.029	0.020	0.016	0.015	0.014
		0.35	0.108	0.099	0.102	0.104	0.106	0.067	0.050	0.043	0.039	0.038
		0.45	0.130	0.148	0.150	0.151	0.154	0.092	0.084	0.081	0.081	0.080
		0.49	0.152	0.172	0.180	0.183	0.188	0.083	0.083	0.083	0.081	0.081

(Continued on the next page)

(Continued) Empirical size of $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha = 5\%$

Kernel	q	$\gamma \setminus$ T	$\hat{\boldsymbol{D}}_M$					$\tilde{\boldsymbol{D}}_{M,k}$				
			0.2M	0.4M	0.6M	0.8M	M	0.2M	0.4M	0.6M	0.8M	M
TR	10%	0	0.188	0.156	0.177	0.185	0.188	0.090	0.097	0.087	0.091	0.090
		0.15	0.159	0.147	0.141	0.159	0.159	0.077	0.091	0.078	0.082	0.077
		0.25	0.215	0.190	0.195	0.211	0.215	0.076	0.084	0.082	0.079	0.076
		0.35	0.212	0.204	0.210	0.212	0.212	0.145	0.152	0.151	0.148	0.145
		0.45	0.278	0.258	0.265	0.272	0.278	0.156	0.154	0.156	0.155	0.156
		0.49	0.274	0.247	0.259	0.269	0.274	0.149	0.150	0.150	0.149	0.149
	5%	0	0.146	0.121	0.133	0.143	0.146	0.050	0.057	0.055	0.050	0.050
		0.15	0.124	0.109	0.113	0.119	0.124	0.046	0.049	0.052	0.043	0.046
		0.25	0.170	0.154	0.156	0.165	0.170	0.039	0.052	0.049	0.046	0.039
		0.35	0.176	0.155	0.161	0.166	0.176	0.097	0.107	0.106	0.100	0.097
		0.45	0.230	0.212	0.222	0.228	0.230	0.123	0.129	0.126	0.124	0.123
		0.49	0.241	0.217	0.227	0.236	0.241	0.114	0.114	0.115	0.114	0.114
	1%	0	0.076	0.070	0.077	0.086	0.076	0.016	0.022	0.015	0.016	0.016
		0.15	0.073	0.066	0.069	0.071	0.073	0.012	0.022	0.021	0.013	0.012
		0.25	0.102	0.090	0.100	0.102	0.102	0.014	0.020	0.019	0.017	0.014
		0.35	0.106	0.099	0.101	0.103	0.106	0.040	0.051	0.047	0.042	0.040
		0.45	0.156	0.149	0.149	0.151	0.156	0.081	0.086	0.082	0.081	0.081
		0.49	0.190	0.173	0.182	0.184	0.190	0.083	0.084	0.083	0.083	0.083
QS	10%	0	0.192	0.154	0.176	0.182	0.192	0.093	0.099	0.090	0.101	0.093
		0.15	0.158	0.156	0.146	0.154	0.158	0.091	0.099	0.085	0.090	0.091
		0.25	0.217	0.190	0.198	0.212	0.217	0.081	0.092	0.084	0.089	0.081
		0.35	0.221	0.209	0.213	0.216	0.221	0.148	0.163	0.163	0.156	0.148
		0.45	0.290	0.278	0.282	0.283	0.290	0.155	0.156	0.163	0.157	0.155
		0.49	0.273	0.244	0.258	0.265	0.273	0.149	0.151	0.150	0.149	0.149
	5%	0	0.146	0.120	0.127	0.142	0.146	0.053	0.055	0.059	0.054	0.053
		0.15	0.125	0.115	0.111	0.117	0.125	0.050	0.057	0.055	0.045	0.050
		0.25	0.169	0.150	0.158	0.163	0.169	0.043	0.056	0.053	0.049	0.043
		0.35	0.174	0.161	0.168	0.169	0.174	0.098	0.111	0.109	0.104	0.098
		0.45	0.249	0.230	0.239	0.242	0.249	0.123	0.129	0.123	0.123	0.123
		0.49	0.247	0.218	0.232	0.240	0.247	0.124	0.126	0.124	0.124	0.124
	1%	0	0.077	0.069	0.075	0.082	0.077	0.018	0.025	0.018	0.019	0.018
		0.15	0.074	0.061	0.070	0.072	0.074	0.017	0.024	0.024	0.016	0.017
		0.25	0.103	0.090	0.099	0.103	0.103	0.016	0.022	0.020	0.019	0.016
		0.35	0.105	0.103	0.103	0.104	0.105	0.042	0.055	0.050	0.044	0.042
		0.45	0.178	0.165	0.171	0.174	0.178	0.087	0.095	0.089	0.088	0.087
		0.49	0.196	0.178	0.187	0.190	0.196	0.087	0.087	0.087	0.087	0.087

Note: Empirical size of the sequential monitoring scheme for the change points in the GARCH-FZ model via 5000 simulations generated by the GARCH(1,1)-Gaussian model. VaR and ES are jointly estimated at 5% level. We consider that $M = 1000$ and monitoring sample sizes of $\{0.2M, \ldots, M\}$. Regarding the long-run covariance estimator, we consider the following estimator kernels: Bartlett (BT), Truncated (TR) and Quadratic Spectral (QS).

Figure 4.3.1: Empirical size of $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha = 5\%$ and $\hat{D}_M$
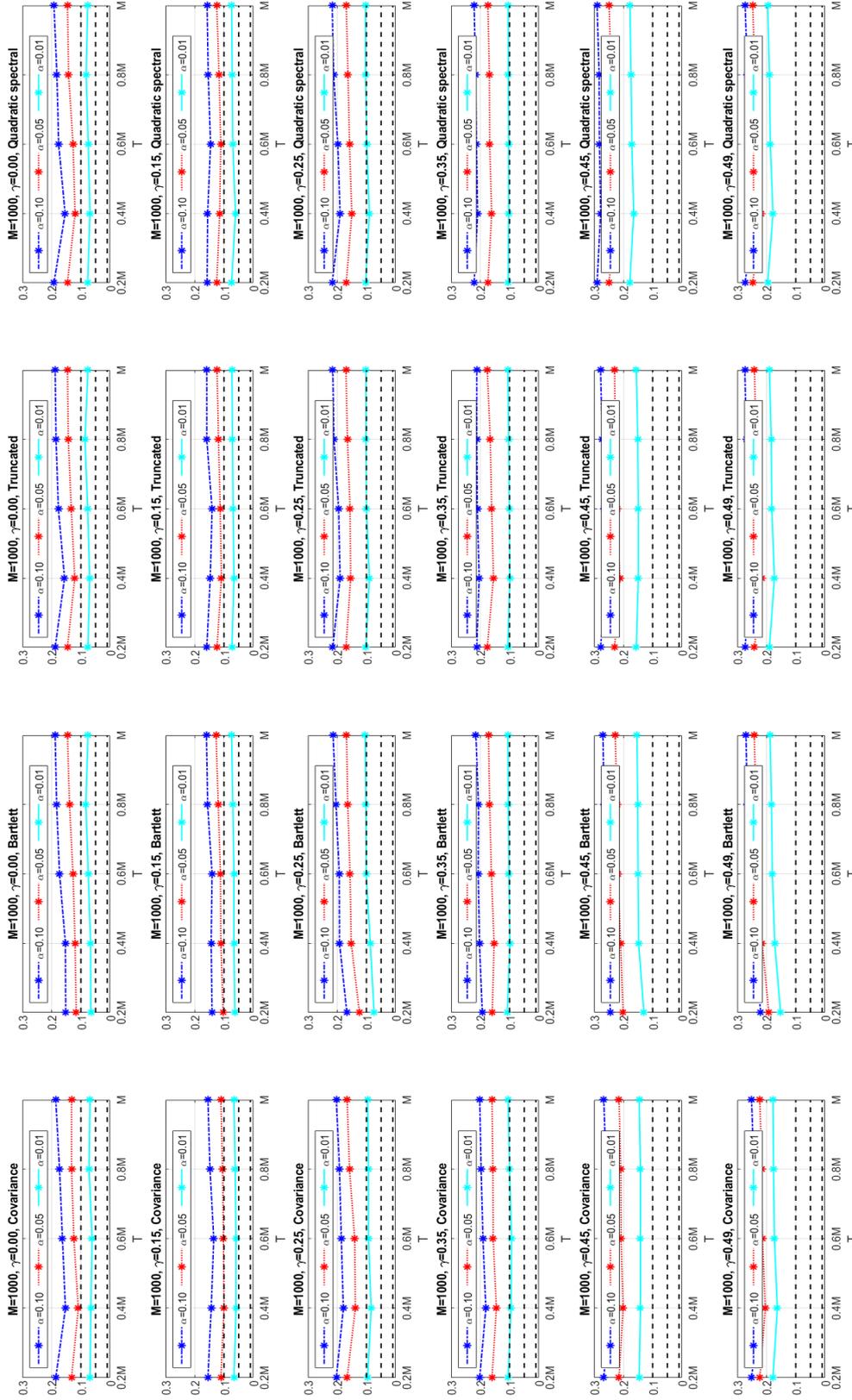
Note: The empirical size of $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha = 5\%$ and $\hat{D}_M$. We consider the training sample size $M = 1000$ and the monitoring horizon sizes $0.2M, \ldots, M$. The empirical sizes at 10%, 5% and 1% are shown in the blue dashed line, red dotted line and light blue solid line, respectively.

Figure 4.3.2: Empirical size of $\tilde{\Gamma}\left(M,k;\hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha=5\%$ and $\tilde{D}_{M,k}$
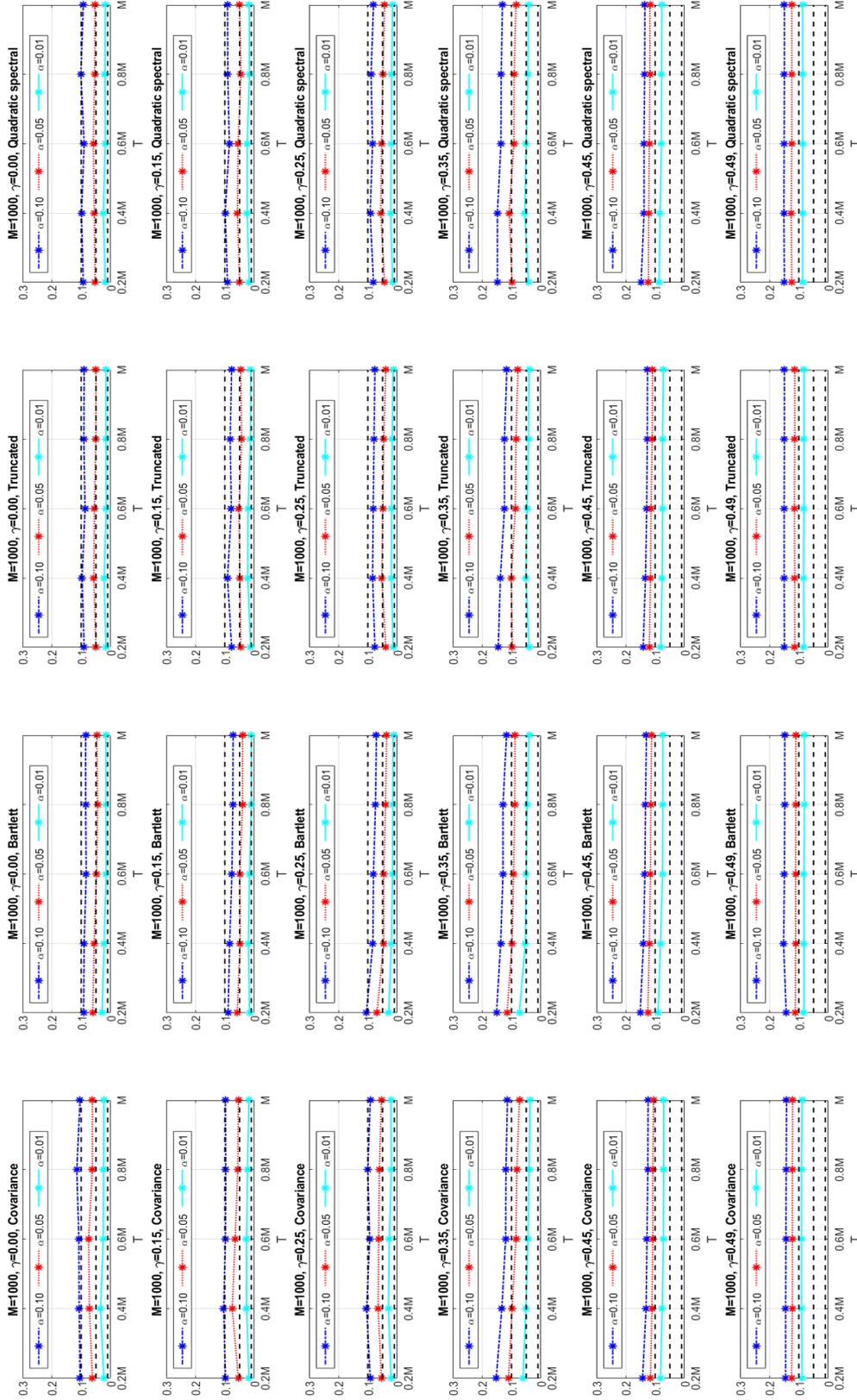
Note: The empirical size of $\tilde{\Gamma}\left(M,k;\hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha=5\%$ and $\tilde{D}_{M,k}$. We consider the training sample size $M=1000$ and the monitoring horizon sizes $0.2M,\ldots,M$. The empirical sizes at 10%, 5% and 1% are shown in the blue dashed line, red dotted line and light blue solid line, respectively.

Figure 4.3.3: Empirical size convergence of $\tilde{\Gamma}\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha = 5\%$ and $\tilde{D}_{M,k}$
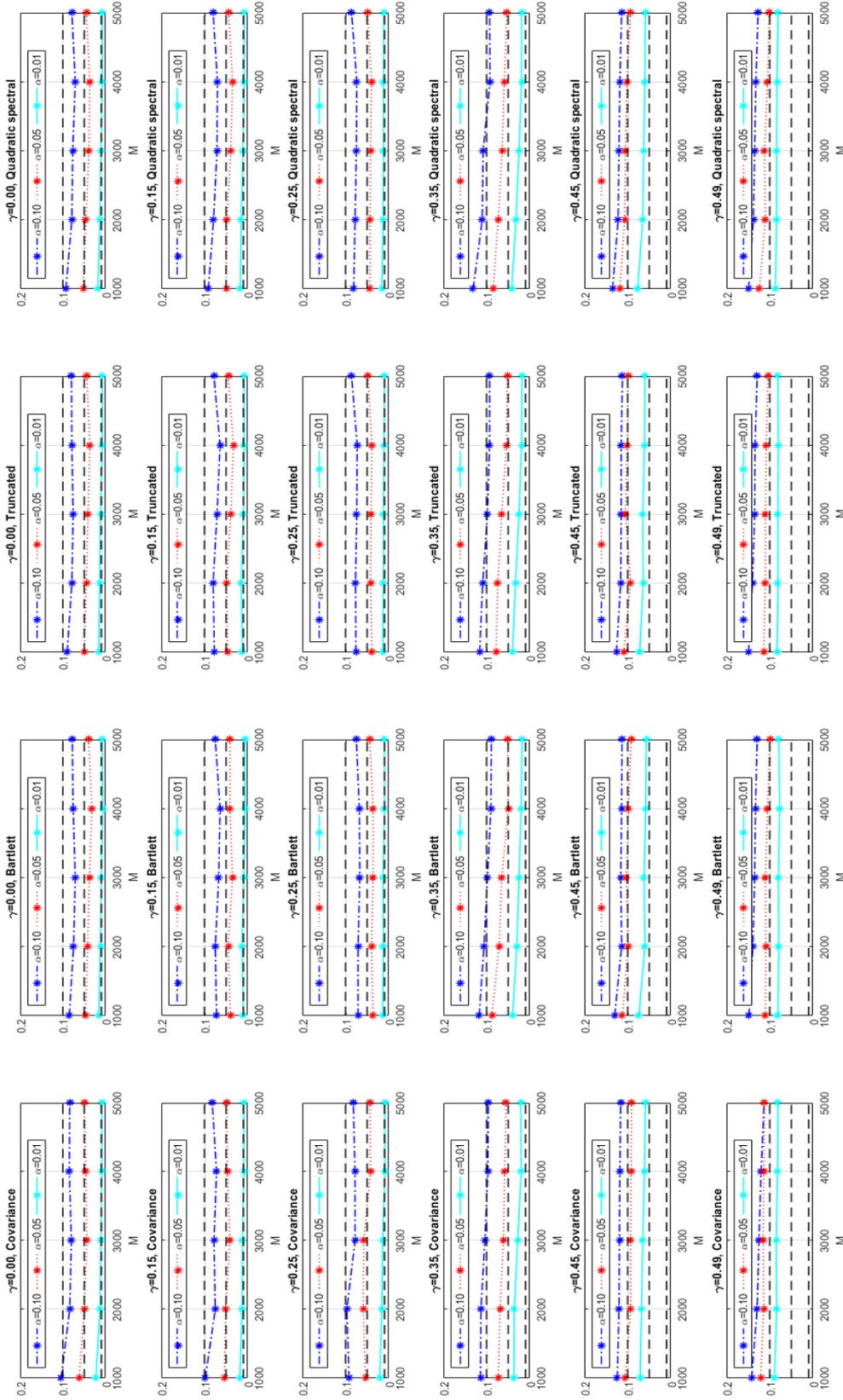
Note: The convergence of empirical sizes of $\tilde{\Gamma}\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for the GARCH-FZ model with $\alpha = 5\%$ and $\tilde{D}_{M,k}$, when M varies from 1000 to 5000. We consider the monitoring horizon sizes the same with the training sample sizes. The empirical sizes at 10%, 5% and 1% are shown in the blue dashed line, red dotted line and light blue solid line, respectively.

of γ and monitoring horizon sizes. Overall, there are four major findings. First, the monitoring manner with the fixed covariance estimator $\hat{\boldsymbol{D}}_M$ is oversized compared with the prescribed significance levels: 10%, 5% and 1%. However, when we replace the fixed estimator with the sequentially updated covariance estimator $\tilde{\boldsymbol{D}}_{M,k}$, we can see an improvement in the empirical size. Second, the selection of the boundary parameter matters for the empirical size results. We obtain an improved empirical size control when select lower values of the boundary parameter, e.g., $\gamma = 0$, 0.15 or 0.25. It is interesting to observe a discontinuity between the empirical size with $\gamma = 0.25$ and 0.35. This finding may be associated with model misspecification. In this case, using the GARCH-FZ model for empirical applications requires $\gamma = 0$, 0.15 or 0.25 to avoid false positives. The selection of γ based on other semiparametric risk models requires further investigations. Third, the results are consistent across different monitoring horizons. Thus, the proposed monitoring scheme can be applied to different monitoring horizons. Fourth, the choice of long-run covariance estimator for this test has noticeable impact on the empirical size. In particular, when the Bartlett kernel is used to construct the covariance estimator, the empirical sizes are reasonable, especially at 1% significance level. The empirical size results shown in the table are visualized in Figures 4.3.1 and 4.3.2.

Figure 4.3.3 shows the convergence of sample sizes of the detection method with the sequentially updated covariance estimator $\tilde{\boldsymbol{D}}_{M,k}$ for 5% (VaR, ES) to justify Theorem 4.2.1 and Remark 4.2.2. In this case, we consider that the training sample size M varies from 1000 to 5000, and the monitoring horizon has the same sample size of the training sample $(T = M)$. In general, it is noticeable that the empirical sizes are more likely to approach the prescribed significance levels q when M goes larger. Additionally, the results are consistent across different selections for the long-run covariance estimation kernel function and the boundary parameter. In particular, the convergence of the empirical size is more obvious when we select boundary parameter $\gamma = 0.35$. Whilst the rates of false detections

are generally oversized when we choose $\gamma = 0.45$ and 0.49, we still can notice decreasing trends of the empirical sizes when $M \to \infty$.

Figures 4.3.4 and 4.3.5 present the power curve for the sequential monitoring scheme for change points in the GARCH-FZ model for 5% (VaR, ES) under the alternative $H_{A,1}$ for $k^* = 1$ and $0.5T$ at the test significance level 5%. In this case, we consider the training sample size of $M = 1000$ and the same size for the monitoring horizon. The empirical probability of stopping under the alternative is high in all cases of γ we considered. It is clear that the empirical power is approaching 1 as β_1^* goes to 0.94, i.e., the unconditional variance increases to 5, which is consistent with our expectations.

Next, we evaluate the empirical distribution of the stopping time $\hat{k}^*$ estimated by the proposed sequential monitoring scheme. The empirical density functions of $\hat{k}^*$ for a change point in the GARCH-FZ for 5% (VaR, ES) with $q = 5\%$ with $M = T = 1000$ and $\beta_1^* = 0.94$ are exhibited in Figures 4.3.6 and 4.3.7.[5] In Figure 4.3.6, we observe that in the case of a change occurring immediately in the monitoring horizon, the higher value of γ, the faster the detection of the change point. However, the detection scheme with a higher γ is not recommended in empirical applications, due to possible spurious detections before the theoretical change point in the scenario of $k^* = 0.5T$.[6] This false positive problem is visualized as the "humps" at the start of the monitoring sample in Figure 4.3.7. Overall, for $k^* = 0.5T$, the monitoring scheme with $\gamma = 0.15$ provides the shortest delay in detection. Thus, in general, $\gamma = 0.15$ (displayed as the solid red line) is recommended because it gives a good balance between the length of delay and the proportion of false early detections.

In this chapter, we also propose a novel way to identify the dominant source of the change point by linking it to the parameter that has the largest absolute

value in the vector below:

$$\sum_{M < t \leq M + \hat{k}^*} \hat{\boldsymbol{\ell}}_t' \left(\hat{\boldsymbol{\theta}}_M \right) \hat{\boldsymbol{D}}_M^{-1/2},$$

after the stopping time $\hat{k}^*$ is first detected.[7] Figure 4.3.8 and 4.3.9 show the dominant source of the change point in the GARCH-FZ model for 5% (VaR, ES) at 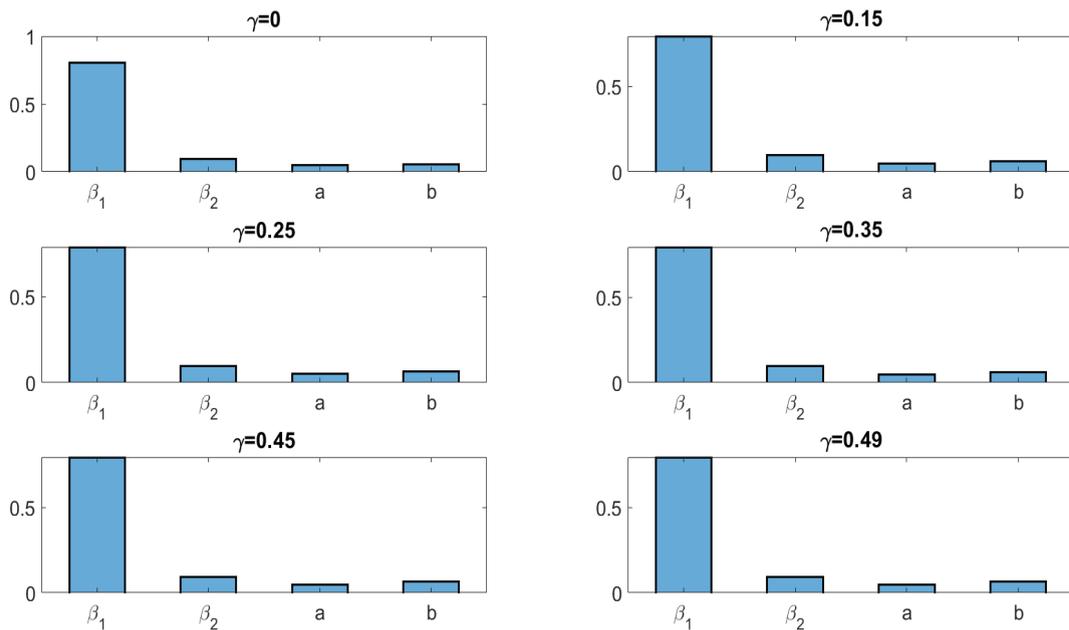5% test significance level with $M = T = 1000$ and $k^* = 1$. In this setting, we consider gradual increases in the value of β_1^* from 0.90 to 0.94 and keep the other parameter values unchanged. Also, we assume that the multipliers for VaR and ES stay constant. The simulation results displayed in Figure 4.3.8 indicate that the structural break is mainly caused by a change in β_1^* instead of the other parameters. Figure 4.3.9 shows that the dominant source is correctly identified as β_1^* for various locations of the change point. Overall, the results are consistent for different values of γ.

We next consider changes in the underlying distribution in the DGP, from the Gaussian distribution to the Student's t distribution with the DoF parameters $\nu^* = \{9.5, \ldots, 4.5\}$, meaning that the process becomes more heavy-tailed after the break. In this scenario, the change of tailedness is not likely to be reflected in the monitoring sample immediately after the break. To improve the simulation results, we use larger sample sizes for the historical sample and testing sample in this setting, i.e., $M = T = 2000$. Figure 4.3.10 shows the empirical power of the test with the change occurring at $k^* = 1$ in the GARCH-FZ model for 5% VaR and ES with different boundary curves. It is obvious from this figure that the sequential monitoring scheme with a lower value of γ generates a higher power than the scheme with $\gamma = 0.45$ or 0.49.

It is worthwhile to highlight the dominant source of the change point when the underlying distribution of the process switches from the Gaussian to the Student's t with $\nu^* = 4.5$. Based on the setup under the alternative hypothesis $H_{A,1}$, we

Figure 4.3.4: Empirical power of the monitoring scheme for a change point in the GARCH-FZ with $k^* = 1$ under $H_{A,1}$
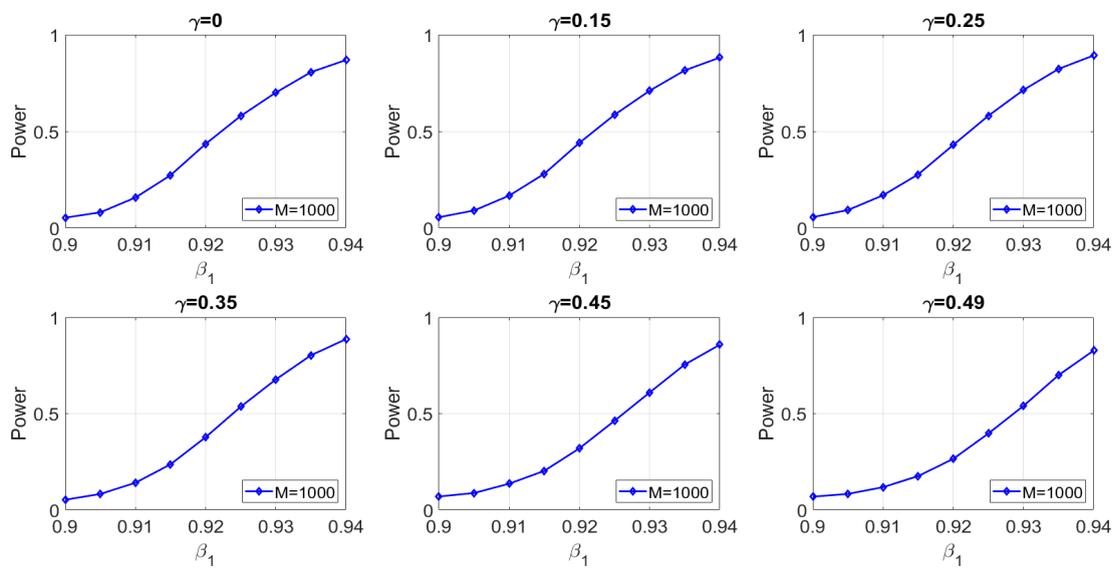
Note: This figure presents the empirical power of the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 1000$ and $k^* = 1$ under $H_{A,1}$

Figure 4.3.5: Empirical power of the monitoring scheme for a change point in the GARCH-FZ with $k^* = 0.5T$ under $H_{A,1}$
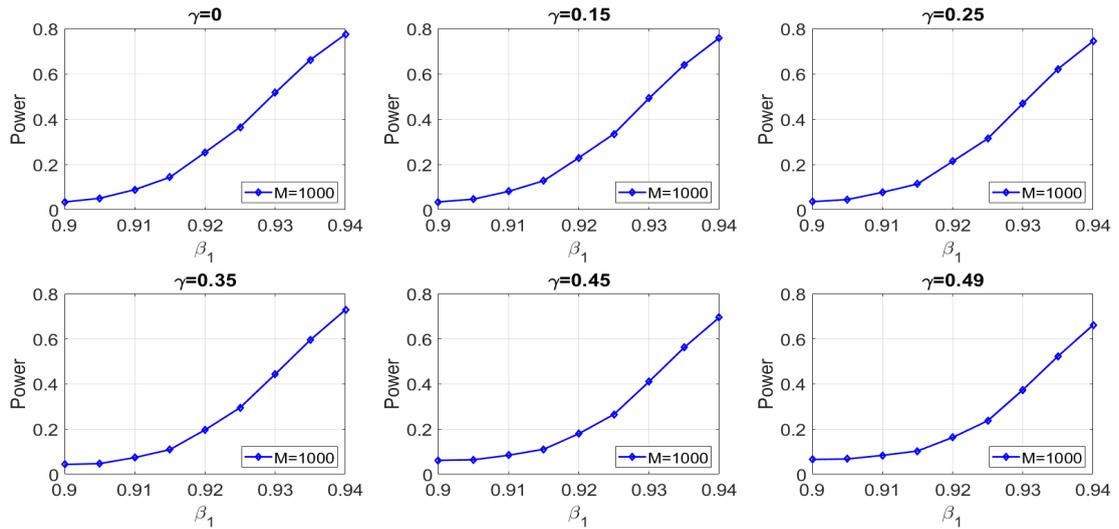
Note: This figure presents the empirical power of the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 1000$ and $k^* = 0.5T$ under $H_{A,1}$

Figure 4.3.6: Empirical density functions of the stopping time $\hat{k}^*$ for a change point in the GARCH-FZ with $k^* = 1$ and $\beta_1^* = 0.94$
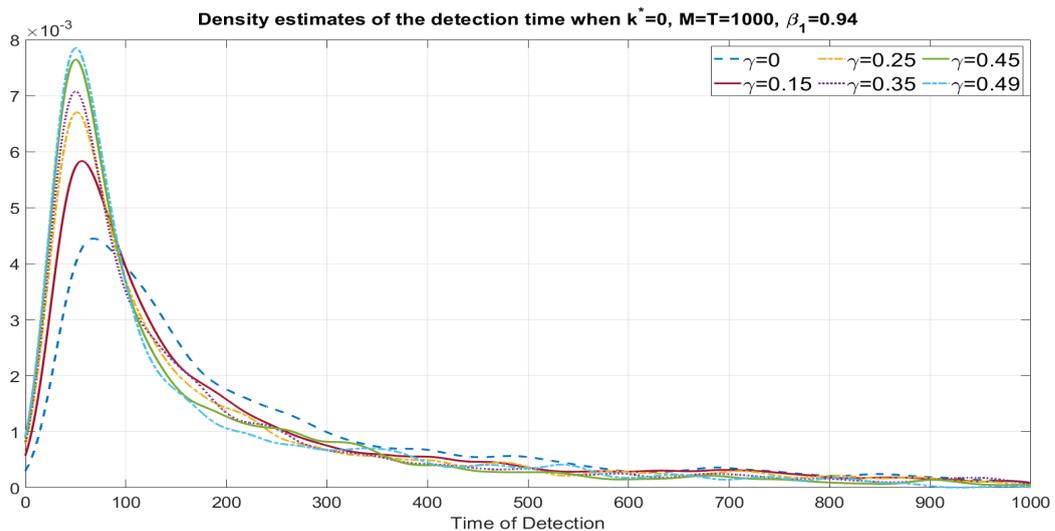
Note: This figure presents the empirical density functions of the stopping time $\hat{k}^*$ estimated by the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 1000$, $k^* = 1$ and $\beta_1^* = 0.94$

Figure 4.3.7: Empirical density functions of the stopping time $\hat{k}^*$ for a change point in the GARCH-FZ with $k^* = 0.5T$ and $\beta_1^* = 0.94$
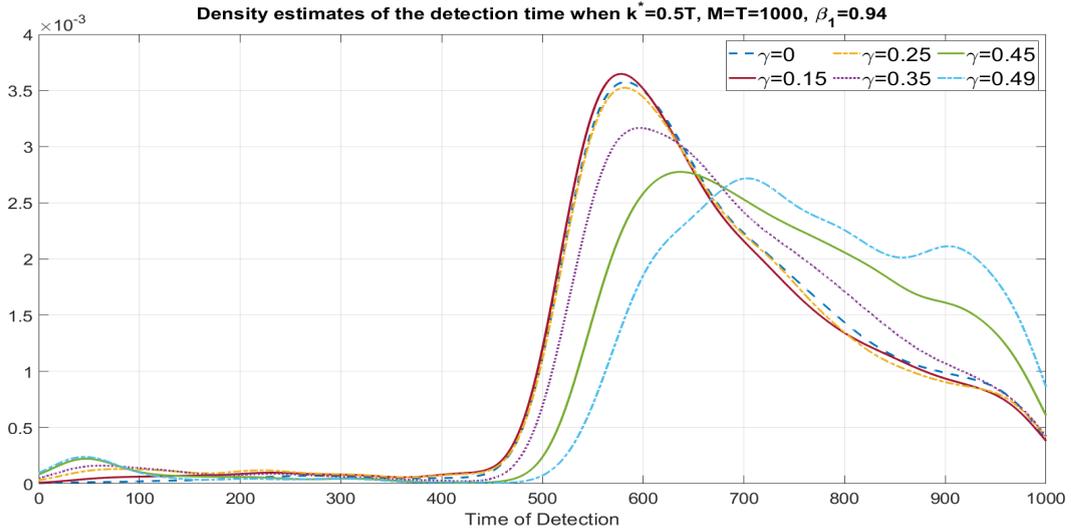
Note: This figure presents the empirical density functions of the stopping time $\hat{k}^*$ estimated by the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 1000$, $k^* = 0.5T$ and $\beta_1^* = 0.94$

Figure 4.3.8: The dominant source of the change in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 1000$, $k^* = 1$ and $\beta_1^* = 0.94$

Figure 4.3.9: The dominant source of the change in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 1000$, $k^* = 0.5T$ and $\beta_1^* = 0.94$
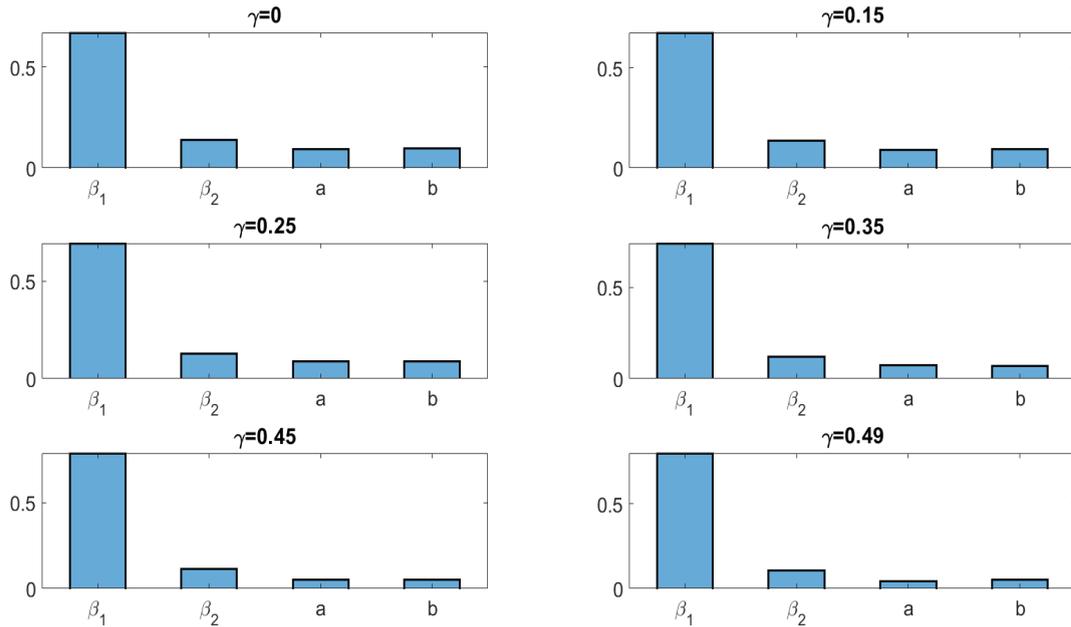

expect that the change point only causes a change in the multipliers for VaR and ES in the GARCH-FZ model, instead of the parameters in the GARCH process. Figure 4.3.11 reveals that the dominant source for the change point under the hypothesis $H_{A,2}$ is mainly the parameters a and b, which are the multipliers for VaR and ES in the model. Meanwhile, the identified dominant source of structural break is rarely β_1 or β_2. Additional figures illustrating the empirical power curve, the estimated densities of the stopping time and the identified dominant source of change points for another location of the change point, i.e., $k^* = 0.5T$, are provided in Appendix 4.D.

Under the alternative hypothesis $H_{A,3}$, we investigate the performance of the proposed test to identify changes in the skewness of the residuals. Figure 4.3.12 shows the empirical power curve for the alternative $H_{A,3}$ for $K^* = 1$ at 5% test significance level with the sample sizes $M = T = 2000$. It is clear that the empirical power of our test increases as the skewness gets more negative. In

general, the choice of γ has a minor effect on the empirical power, but when $\gamma =0$, 0.15 or 0.25, the test seems to have a higher power. This finding also motivates us to select γ with a lower value in practical applications. Regarding the dominant source of the change point, Figure 4.3.13 illustrates that the change point is mainly caused by a change in the multiplier b for ES in the GARCH-FZ model, and this finding is more salient with $\gamma =0.45$ or 0.49. While the test can identify the change of the other parameters, e.g., β_2 with a good success rate, in the simulations with a low value of γ, the multiplier b for ES is still identified most often as the dominant source of the change point. The simulation results with $k^* = 0.5T$ are reported in Appendix 4.D.

Figure 4.3.10: Empirical power of the monitoring scheme for a change point in the GARCH-FZ with $k^* = 1$ under $H_{A,2}$
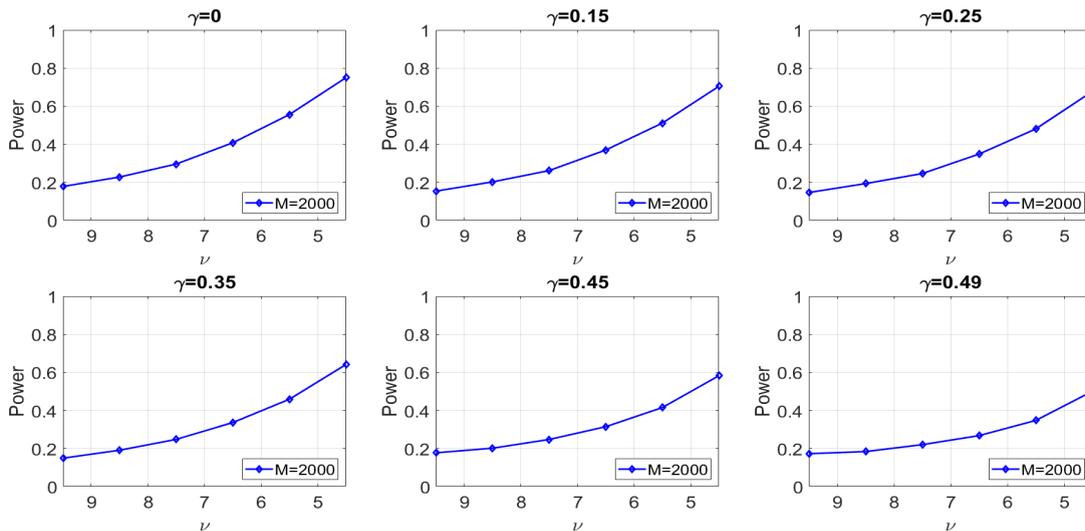
Note: This figure presents the empirical power of the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$ and $k^* = 1$ under $H_{A,2}$

Figure 4.3.11: The dominant source of the change in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 1$ and $\nu^* = 4.5$
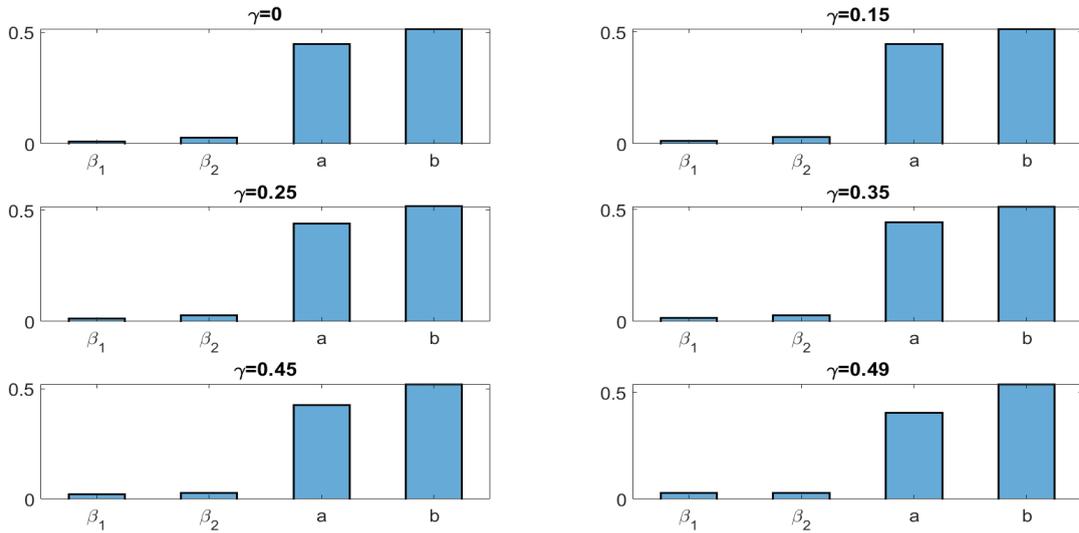
Figure 4.3.12: Empirical power of the monitoring scheme for a change point in the GARCH-FZ with $k^* = 1$ under $H_{A,3}$
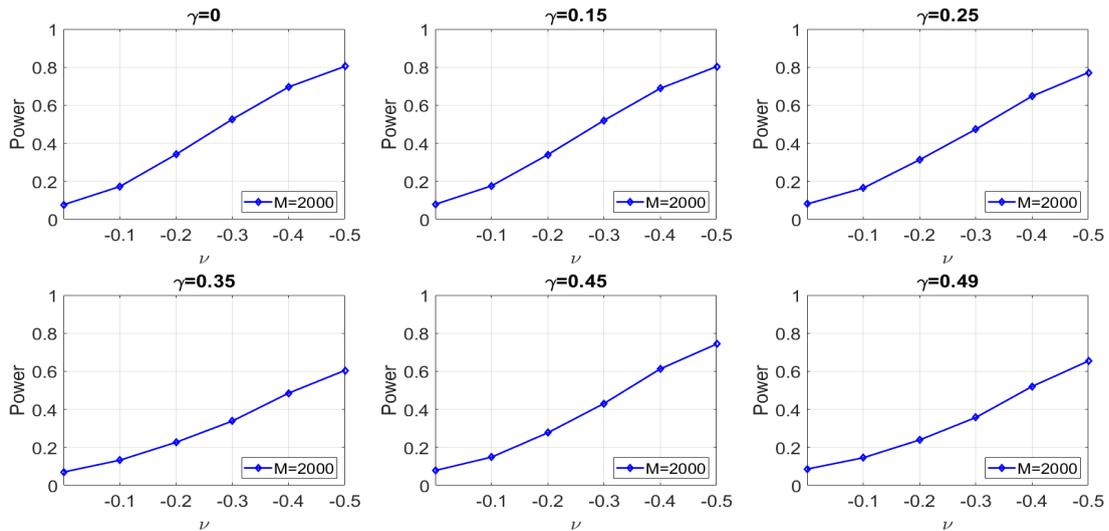
Note: This figure presents the empirical power of the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$ and $k^* = 1$ under $H_{A,3}$.

Figure 4.3.13: The dominant source of the change in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 1$ and $\lambda^* = -0.5$
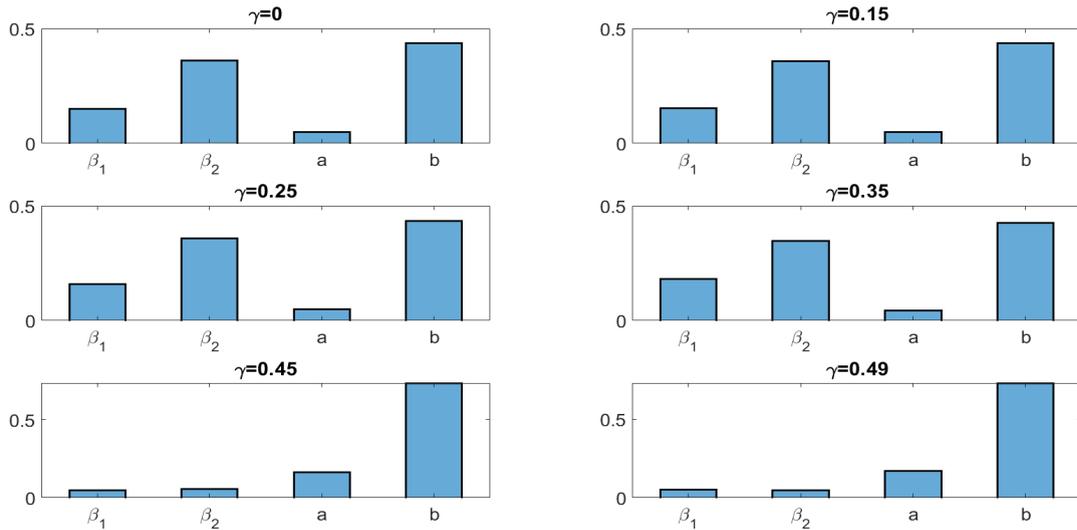
4.4 Empirical Applications

In this section, we apply the proposed monitoring scheme to identify change points in risk models applied for financial returns. Our main aim is to investigate whether this test is able to detect in real time well-known events that are known to have caused shocks in financial markets. Here, we consider the daily log returns of: the S&P 500 index and the GBP/EUR exchange rate, which are collected from Datastream and Bloomberg, respectively. The risk model for (VaR, ES) used in our empirical applications is the GARCH-FZ model proposed by Patton et al. (2019), which is the same as the one used in the simulation study. We consider the semiparametrically estimated (VaR, ES) at 5% test significance level.[8]

To verify the condition that there is no break in the selected training samples, we firstly apply the two-sample Kolmogorov-Smirnov (KS) test to check whether the first and second half of the training sample are identically distributed. Additionally, we test that there is no break in the time series of (VaR, ES) in training samples by employing the loss-based Wilcoxon test (Lazar et al., 2021). Table

4.4.1 displays the results for the selected periods, specifically the KS test statistics, loss-based Wilcoxon test statistics and the average values of 5% VaR and ES. These results suggest that there is no break in the series of training samples. In the subsections below, we apply our proposed test for change point detection.

Table 4.4.1: Description of the selected time series of training and testing samples

		Time span	KS Stat	Wilcoxon Stat	5% VaR	5% ES
Sample A: S&P 500 Index						
Pre financial crisis period	Training	23/May/2003-20/Dec/2005	0.058	0.782	-1.229	-1.498
	Testing	21/Dec/2005-23/Feb/2007	0.101	1.790	-1.037	-1.407
Event I: Financial crisis	Training	23/May/2003-20/Dec/2005	0.073	1.524	-1.121	-1.414
	Testing	21/Dec/2005-16/July/2010	0.146***	2.043***	-2.909	-4.476
Event II: COVID pandemic	Training	06/Feb/2018-01/Oct/2019	0.067	0.688	-1.877	-2.566
	Testing	02/Oct/2019-31/Dec/2020	0.137***	1.327**	-2.954	-5.092
Sample B: GBP/EUR exchange rate						
Event III: Brexit	Training	18/Aug/2010-12/Feb/2014	0.084	2.490	-0.795	-1.004
	Testing	13/Feb/2014-09/Apr/2018	0.097***	2.398**	-0.939	-1.357

Note: This table includes the time span, the Kolmogorov-Smirnov test statistics, the loss-based Wilcoxon test statistics and the 5% VaR and ES estimated by historical simulations. The two-sample Kolmogorov-Smirnov test is conducted for the first and second half of each training and testing samples. The loss-based Wilcoxon test is conducted for the training and testing samples based on Lazar et al. (2021), and the critical values can be found in Table 4.C.2 of Appendix 4.C. *** and ** indicate values significant at 1% and 5% significance level. All samples are daily log returns.

4.4.1 Application 1: the S&P 500 Index

In this application, we consider monitoring for change points in the GARCH-FZ model applied on the daily log returns of the S&P 500 index during three selected periods: (1) 23 May 2003 to 23 February 2007, (2) 23 May 2003 to 16 July 2010, and (3) 6 February 2018 to 31 December 2020. The testing samples of these selected periods cover the pre-financial crisis period, the great financial crisis and the COVID pandemic, respectively.[9]

Pre-financial Crisis Period

The training sample spans from 23 May 2003 to 20 December 2005, containing observations after the burst of the dot-com bubble. Then we monitor the detector from 21 December 2005 to 23 February 2007, which is before the financial crisis. The KS statistic indicates that the log returns in the training sample generally follow the same distribution. Also, there is no significant change in the time series of (VaR, ES) in both samples based on the loss-based Wilcoxon test statistics.

The daily log returns of the S&P 500 index are displayed in the upper panel of Figure 4.4.1, where the observations with the white background are in the training sample used for parameter estimation, and the observations with the gray background are in the testing sample used to monitor for change points. We present the trajectory of the detector and the boundary in the lower panel of Figure 4.4.1. The detector never crosses the boundary curve during the testing period, indicating that no changes in the risk model parameters can be detected during the monitoring horizon. The result can be interpreted to mean that the parameters in the GARCH-FZ model estimated from the training sample are still valid in the testing sample. For practitioners, it is not necessary to adjust the risk model parameters in this case.

Financial Crisis

Next, we extend the testing sample to include the Great Recession from the end of 2007 to 2009. In this application, the training sample spans from 23 May 2003 to 20 December 2005, which is the same as the one for the pre-financial crisis period. Here we consider the testing sample from the end of the training sample until 16 July 2010. We are interested in checking whether this monitoring scheme is able to detect the start of the financial crisis in December 2007. In Figure 4.4.2, the detector $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ exceeds the selected boundary curve $g(M, k)$ on 8 March 2007, which is earlier than the beginning of the financial crisis. This

Figure 4.4.1: Real-time detection for the S&P 500 index within the pre-financial crisis period
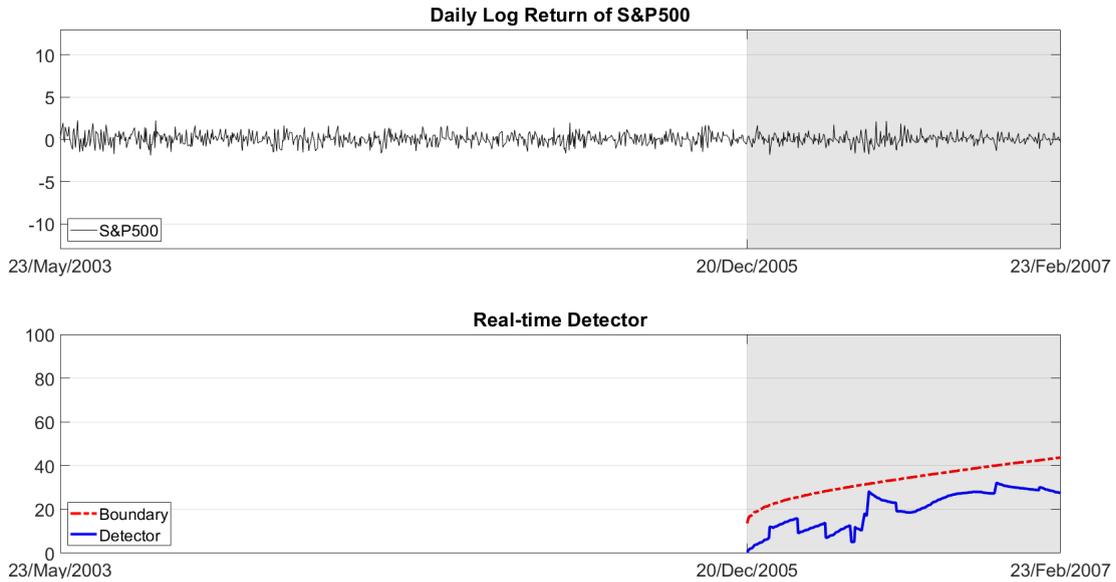
Note: Upper panel: The log return of the S&P 500 during 15/Oct/2003 to 23/Feb/2007; lower panel: Real-time detector based on the GARCH-FZ model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area).

finding indicates that the sequential monitoring scheme enables us to identify change points even before the actual crisis unfolds. This early detection provides practitioners with a timing to adjust the parameter values of their risk models in order to measure the risk more effectively during the financial crisis. Next, we estimate the parameters of the GARCH-FZ model and we calculate the average values of the time series of VaR, ES and loss values for both pre-break and post-break samples. Also, for the sample following the break detection, we compute the average loss for the (VaR, ES) estimated by the model with parameter values based on the pre-break sample, which is denoted by Loss_NC in the table. This is followed by their identification of the dominant source of the change points. All statistics and results discussed above are displayed in Table 4.4.2.

The left panel of Table 4.4.2 shows the parameter estimates. After the detected change point, we can observe an increase in the value of β_2 from 0.015 to 0.088, which leads to a higher level of volatility. This observation is consistent with the

identified dominant source of the change point, i.e., β_2 dominates the others in the parameters vector at the detected change point. There is no big difference in the multipliers a and b for VaR and ES before and after the identified change point, but still the higher volatility level post-break will result in a lower level of VaR and ES consequently. The average FZ loss value for (VaR, ES) based on the parameters estimated in the post-break sample is 1.261, which is almost half of the average loss value calculated by using the parameters based on the pre-break sample (2.481). The large difference in the average loss values indicates the importance of accurate change point detection.

Additionally, we provide the detection results based on the GARCH-Gaussian model for VaR and ES in Appendix 4.F. Figure 4.F.1 shows that the detected change point is 13 November 2007. The test based on the GARCH-Gaussian model can detect the change point later than the one based on the GARCH-FZ model, due to likely model misspecification issues.

Figure 4.4.2: Real-time detection for the S&P 500 index within the financial crisis period
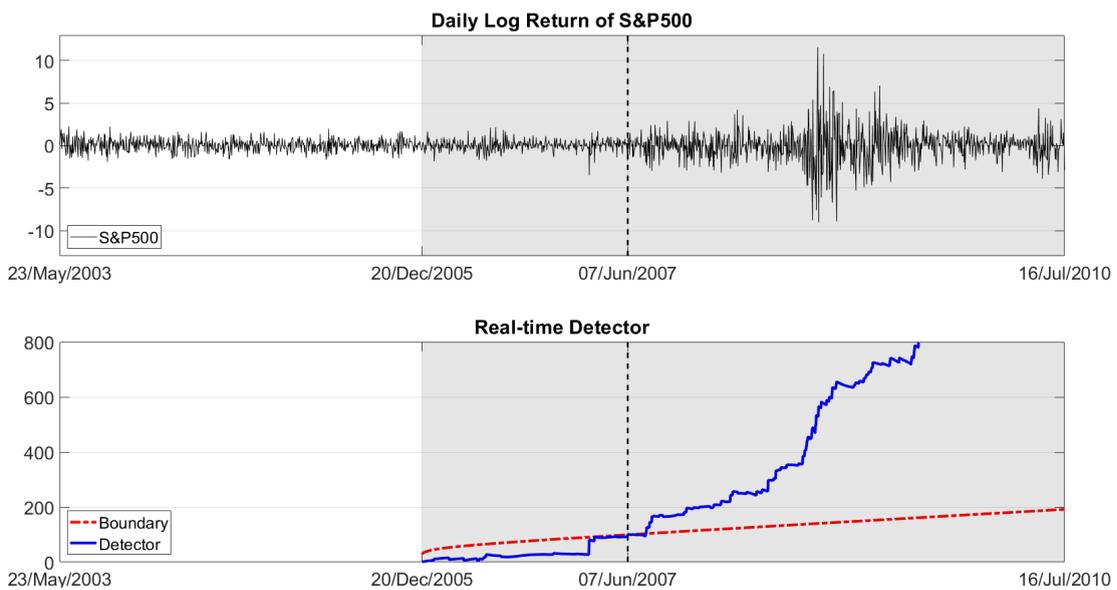
Note: Upper panel: The log return of the S&P 500 during 12/Jan/2004 to 09/Dec/2009; lower panel: Real-time detector based on the GARCH-FZ model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area). The vertical dash line denotes the estimated change point for 5% VaR and ES.

COVID Pandemic

In this section, we investigate the effect of the COVID pandemic on the US stock market by detecting change points in risk models. Regarding the training sample, we select a relatively more stable period before the pandemic from 6 February 2018 to 1 October 2019 to ensure that there is no structural break in the sample.[10] Then we consider the testing sample following the training sample up to 31 December 2020. Figure 4.4.3 shows that the detector $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for 1% VaR and ES is above the boundary curve after 9 March 2020, when the US stock market declined the most in a week since the financial crisis of 2007–2008. If risk managers adjusted their reserves against risk, they could have avoided the large losses that occurred on Black Thursday (12 March 2020) and Black Monday II (16 March 2020).

The middle panel of Table 4.4.2 indicates that after the detected change point, the multipliers for VaR and ES in this model experience an increase. Even though the unconditional volatility decreases after the detection, the change in the parameter values of a and b dominates and leads to a decline in the level of risk measures. It is worthwhile to mention that if we keep using the parameters estimated in the training sample, the average loss would be 1.892, which is much higher than the average loss computed taking the change point into consideration, 1.332.

Figure 4.F.2 shows that the detected change point based on the GARCH-Gaussian model is 16 March 2020, when the third trading curb occurred. The detected date is somewhat compared with the one based on the GARCH-FZ model. This finding indicates that the GARCH-FZ model is more efficient than the GARCH-Gaussian model in this case because it is able to detect change points much earlier.

Figure 4.4.3: Real-time detection for the S&P 500 index within the COVID-19 pandemic period
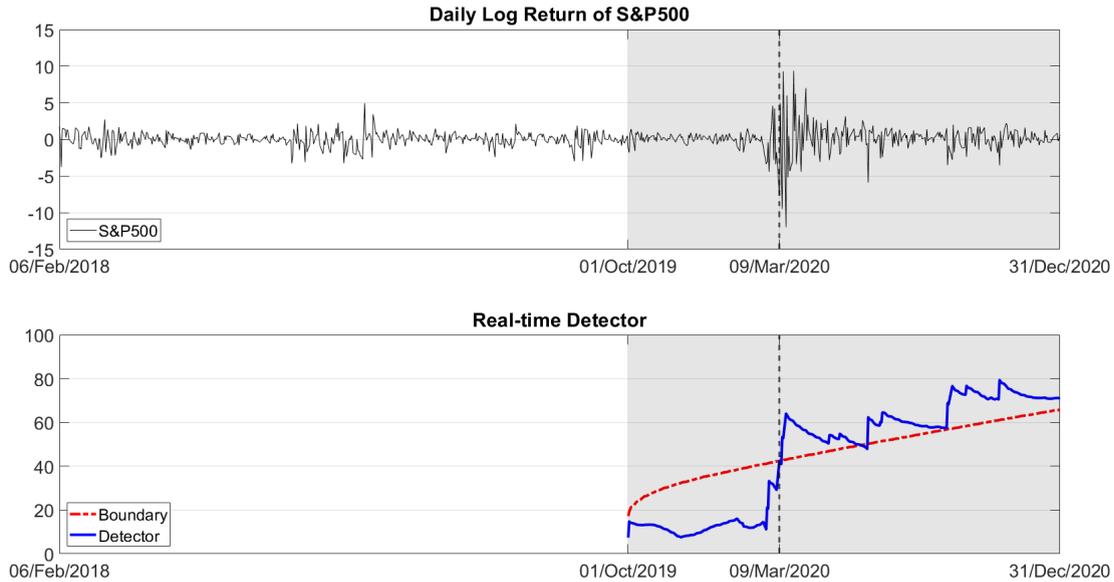
Note: Upper panel: The log return of the S&P 500 during 06/Feb/2018 to 31/Dec/2020; lower panel: Real-time detector based on the GARCH-FZ model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area). The vertical dash line denotes the estimated change point for 5% VaR and ES.

4.4.2 Application 2: the GBP/EUR Exchange Rate

At last, we focus on the impact of Brexit on the fluctuations of the GBP/EUR exchange rates. In this case, we select the training sample as spanning from 18 August 2010 to 12 February 2014 to ensure no change points in this period. Next, we apply the proposed monitoring test for the testing sample from 13 February 2014 to 9 April 2018. Figure 4.4.4 shows that the detectors $\Gamma\left(M, k; \hat{\boldsymbol{\theta}}_M\right)$ for 1%, 2.5% and 5% risk measures cross the boundary curve on 24 August 2015, 24 February 2016 and 24 June 2016, respectively.

When we consider $\alpha = 5\%$, we can identify the change point located on the day after the Brexit referendum, when the sterling was at a 31-year low, having fallen 11% in two trading days, and the FTSE 100 index had surrendered 85 billion pounds. However, if we consider risk estimates at lower significance levels, i.e., 1% and 2.5% risk measures, we can detect change points in the GARCH-FZ

model even earlier than the Brexit referendum. Also, the lower the significance level we consider, the earlier the change point can be detected. This observation implies that the GBP/EUR exchange rates already contained information about the uncertainty associated with the likely vote outcome before the Brexit referendum.

Figure 4.4.4: Real-time detection for the GBP/EUR exchange rate
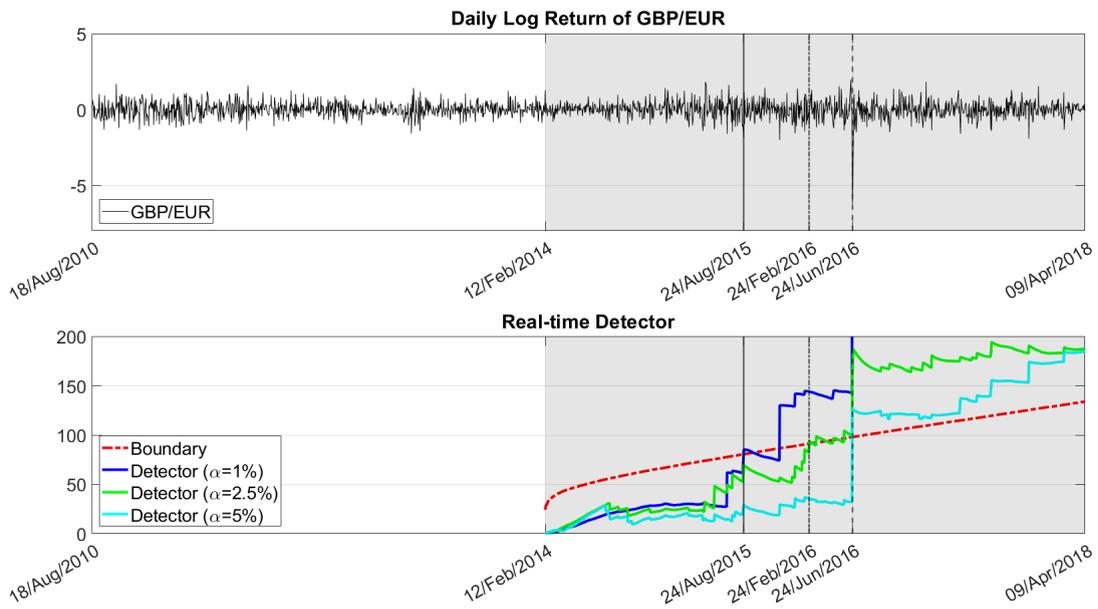
Note: Upper panel: The log return of the GBP/EUR exchange rate during 18/Aug/2010 to 09/Apr/2018; lower panel: Real-time detector based on the GARCH-FZ model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area). The vertical solid line denotes the estimated change point for the 1% risk measures; the vertical dash-dot line denotes the estimated change point for the 2.5% risk measures; the vertical dash line denotes the estimated change point for the 1% risk measures.

4.5   Conclusions

In this chapter, we propose a new test to sequentially monitor change points in the M-estimators of semiparametric risk models for VaR and ES risk measures jointly by evaluating the change in the gradient of the FZ loss function introduced by Fissler and Ziegel (2016). When the gradient-based detector exceeds a selected boundary function, a change point is detected. We perform MC simulations for

Table 4.4.2: Estimated coefficients of the GARCH-FZ model in both training and monitoring samples

Variables	Financial crisis		COVID Pandemic		Brexit ($\alpha = 5\%$)	
	Before detection	After detection	Before detection	After detection	Before detection	After detection
β_1	0.926	0.914	**0.707**	**0.494**	0.974	0.707
	(0.120)	(0.050)	(0.203)	(0.207)	(0.024)	(0.192)
β_2	**0.015**	**0.088**	0.210	0.053	0.015	0.013
	(0.021)	(0.050)	(0.138)	(0.030)	(0.011)	(0.031)
a	-2.108	-1.771	-1.957	-4.204	**-2.628**	**-4.117**
	(2.092)	(0.295)	(0.976)	(1.017)	(1.604)	(0.309)
b	-2.650	-2.177	-2.948	-5.888	-3.087	-4.905
	(2.609)	(0.218)	(1.255)	(1.384)	(1.861)	(0.530)
VaR	-1.100	-3.204	-1.641	-3.061	-1.174	-1.466
ES	-1.382	-3.938	-2.473	-4.287	-1.379	-1.817
Loss	0.322	1.261	0.820	1.332	0.308	0.595
Loss_NC	-	2.481	-	1.408	-	0.628
Dominant Source	$\boldsymbol{\beta_2}$		$\boldsymbol{\beta_1}$		$\boldsymbol{a}$	

Table 4.4.3: Estimated coefficients of the GARCH-FZ model in both training and monitoring samples of Case IV (COVID Pandemic) for risk measures with different α

Variables	$\alpha = 1\%$			$\alpha = 2.5\%$			$\alpha = 5\%$		
	Training sample	Before detection	After detection	Training sample	Before detection	After detection	Training sample	Before detection	After detection
β_1	0.930	0.963	0.753	0.930	0.969	0.776	0.930	0.974	0.707
	(0.162)	(0.036)	(0.049)	(0.162)	(0.011)	(0.048)	(0.162)	(0.024)	(0.192)
β_2	0.021	0.014	0.129	0.021	0.032	0.364	0.021	0.015	0.013
	(0.034)	(0.011)	(0.249)	(0.034)	(0.011)	(0.361)	(0.034)	(0.011)	(0.031)
a	-2.992	-2.995	-2.948	-2.992	-2.215	-1.986	-2.992	-2.628	-4.117
	(4.501)	(1.979)	(1.444)	(4.501)	(0.161)	(0.541)	(4.501)	(1.604)	(0.309)
b	-3.334	-3.324	-4.964	-3.334	-2.506	-3.191	-3.334	-3.087	-4.905
	(5.009)	(2.244)	(2.240)	(5.009)	(0.234)	(0.457)	(5.009)	(1.861)	(0.530)
VaR	-1.156	-1.149	-1.426	-1.156	-1.215	-1.458	-1.156	-1.174	-1.466
ES	-1.288	-1.275	-2.402	-1.288	-1.375	-2.342	-1.288	-1.379	-1.817
Loss	0.237	0.235	0.835	0.237	0.301	0.785	0.237	0.308	0.595
Dominant Source	$\boldsymbol{a}$			$\boldsymbol{\beta_1}$			$\boldsymbol{a}$		

various scenarios with finite sample sizes. The monitoring scheme exhibits a reasonable size control under the null hypothesis and high empirical power in all cases across different selections of boundary curves. We recommend a boundary function by comparing the empirical size, power, and stopping time distribution.

In an empirical study, we apply the monitoring scheme for the S&P 500 index and the GBP/EUR exchange rate to sequentially monitor the possible change points in the selected samples. For each case, we consider the VaR and ES estimated by a selected semiparametric model at 1%, 2.5% and 5% significance levels, respectively. The main finding indicates that the lower the significance level, the earlier the detected change point. In most cases, our test can identify structural changes even before a market crash occurs. Most detected change points can be associated with well-known financial or economic events, such as the day after the Brexit referendum and the Black Thursday in the COVID-19 recession period. According to our findings, we can conclude that practitioners can improve their risk management strategies by monitoring for change points in their risk models and then adjusting the parameters of the models based on the identified change points.

Our proposed sequential monitoring test for change points in the M-estimators of semiparametric risk models for VaR and ES contribute to the ongoing debate on the structural breaks in risk measures and the tail index. However, this chapter only considers models for VaR and ES jointly. It would be of interest to extend this test to models of other measures of uncertainty: volatility, individual VaR and expectile. It might also be interesting to formulate tests to sequentially monitor change points in the tail dependence, modelled by copula functions.

Appendices

4.A   First-order Derivatives of the FZ0 Loss Function

The FZ0 loss function is:

$$\ell(y, v, e; \alpha) = -\frac{1}{\alpha e}\mathbf{1}\{y \leq v\}(v - y) + \frac{v}{e} + \log(-e) - 1. \tag{4.A.1}$$

First, we take the first-order derivatives of this loss function w.r.t. risk measures v and e, respectively. These are:

$$v' = \frac{\partial \ell}{\partial v} = -\frac{1}{\alpha e}\mathbf{1}\{y \leq v\} + \frac{1}{e} = -\frac{1}{\alpha e}(\mathbf{1}\{y \leq v\} - \alpha) \tag{4.A.2}$$

$$e' = \frac{\partial \ell}{\partial e} = \frac{1}{\alpha e^2}\mathbf{1}\{y \leq v\}(v - y) - \frac{v}{e^2} + \frac{1}{e} \tag{4.A.3}$$

In the simulation study and empirical work, we select the GARCH-FZ model to forecast VaR and ES. The model can be expressed as:

$$v_t = a \cdot \sigma_t,$$
$$e_t = b \cdot \sigma_t, \quad b < a < 0, \tag{4.A.4}$$
$$\sigma_t^2 = \beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2 r_{t-1}^2,$$

where σ_t^2 is the conditional variance and is assumed to follow a GARCH(1,1)

process. The parameters of this model are estimated by minimizing the loss function FZ0, instead of using (Q)MLE.

In the second step, we take the first derivatives of the risk measures v_t and e_t w.r.t. each parameter in the GARCH-FZ model. Thus, we have the first-order derivatives of v_t:

$$\begin{aligned}
\frac{\partial v_t}{\partial \beta_1} &= a \cdot \frac{1}{2}(\beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \left(\sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1} \right) \\
&= \frac{1}{2} a \sigma_t^{-1} \cdot \left(\sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1} \right) \\
&= \frac{1}{2} \frac{v_t}{\sigma_t^2} \left(\sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1} \right);
\end{aligned}$$

(4.A.5)

$$\begin{aligned}
\frac{\partial v_t}{\partial \beta_2} &= a \cdot \frac{1}{2}(\beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \left(\beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_2} + r_{t-1}^2 \right) \\
&= \frac{1}{2} a \sigma_t^{-1} \cdot \left(\beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_2} + r_{t-1}^2 \right) \\
&= \frac{1}{2} \frac{v_t}{\sigma_t^2} \left(\beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_2} + r_{t-1}^2 \right);
\end{aligned}$$

(4.A.6)

$$\begin{aligned}
\frac{\partial v_t}{\partial a} &= \sigma_t + a \cdot \frac{1}{2}(\beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial a} \\
&= \sigma_t;
\end{aligned}$$

(4.A.7)

$$\begin{aligned}
\frac{\partial v_t}{\partial b} &= a \cdot \frac{1}{2}(\beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial b} \\
&= 0;
\end{aligned}$$

(4.A.8)

and the first-order derivatives of e_t:

$$\begin{aligned}
\frac{\partial e_t}{\partial \beta_1} &= b \cdot \frac{1}{2}(\beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \left(\sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1} \right) \\
&= \frac{1}{2} b \sigma_t^{-1} \cdot \left(\sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1} \right) \\
&= \frac{1}{2} \frac{e_t}{\sigma_t^2} \left(\sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1} \right);
\end{aligned}$$

(4.A.9)

$$\frac{\partial e_t}{\partial \beta_2} = b \cdot \frac{1}{2}(\beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \left(\beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_2} + r_{t-1}^2\right)$$

$$= \frac{1}{2}b\sigma_t^{-1} \cdot \left(\beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_2} + r_{t-1}^2\right) \tag{4.A.10}$$

$$= \frac{1}{2}\frac{e_t}{\sigma_t^2}\left(\beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_2} + r_{t-1}^2\right);$$

$$\frac{\partial e_t}{\partial a} = b \cdot \frac{1}{2}(\beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial a} \tag{4.A.11}$$

$$= 0;$$

$$\frac{\partial e_t}{\partial b} = \sigma_t + b \cdot \frac{1}{2}(\beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2 r_{t-1}^2)^{-1/2} \cdot \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial b} \tag{4.A.12}$$

$$= \sigma_t.$$

Finally, by using the Chain Rule for the derivations, we have the first-order derivatives of the FZ0 loss function w.r.t. each parameter in the GARCH-FZ model as follows:

$$\frac{\partial \ell}{\partial \beta_1} = \frac{\partial \ell}{\partial v_t}\frac{\partial v_t}{\partial \beta_1} + \frac{\partial \ell}{\partial e_t}\frac{\partial e_t}{\partial \beta_1}$$
$$\frac{\partial \ell}{\partial \beta_2} = \frac{\partial \ell}{\partial v_t}\frac{\partial v_t}{\partial \beta_2} + \frac{\partial \ell}{\partial e_t}\frac{\partial e_t}{\partial \beta_2}$$
$$\frac{\partial \ell}{\partial a} = \frac{\partial \ell}{\partial v_t}\frac{\partial v_t}{\partial a} + \frac{\partial \ell}{\partial e_t}\frac{\partial e_t}{\partial a} \tag{4.A.13}$$
$$\frac{\partial \ell}{\partial b} = \frac{\partial \ell}{\partial v_t}\frac{\partial v_t}{\partial b} + \frac{\partial \ell}{\partial e_t}\frac{\partial e_t}{\partial b}$$

4.B Outline of Proof of Theorem 4.2.1

The proof of Theorem 4.2.1 is based on the lemmas stated below.

Lemma 4.B.1. *If Assumptions 4.2.2 - 4.2.4 hold, then:*

$$\sup_{1\leq k<\infty} \frac{\left| \sum_{M<t\leq M+k} \boldsymbol{\ell}_t'(\hat{\boldsymbol{\theta}}_M) - \sum_{M<t\leq M+k} \boldsymbol{\ell}_t'(\boldsymbol{\theta}) \right|}{M^{1/2}\left(1+\frac{k}{M}\right)b\left(\frac{k}{M}\right)} \overset{a.s.}{=} O\left(M^{-1/2}\right),$$

as $M \to \infty$.

Proof. By the proof of Patton et al. (2019), we have:

$$\left| \sum_{M<t\leq M+k} \left(\boldsymbol{\ell}_t'(\hat{\boldsymbol{\theta}}_M) - \boldsymbol{\ell}_t'(\boldsymbol{\theta}) \right) \right| \overset{a.s.}{=} o\left(1\right),$$

implying Lemma 4.B.1. □

Let

$$\boldsymbol{\ell}_t''(\boldsymbol{u}) = \frac{1}{e_t(\boldsymbol{u})^2}e_t'(\boldsymbol{u})^\top e_t'(\boldsymbol{u}) + \frac{f_t(v_t(\boldsymbol{u})|\mathscr{F}_{t-1})}{-\alpha e_t(\boldsymbol{u})}v_t'(\boldsymbol{u})^\top v_t'(\boldsymbol{u}), \qquad (4.\text{B}.1)$$

and

$$\boldsymbol{\Lambda}(\boldsymbol{u}) = \mathbb{E}[\boldsymbol{\ell}_0''(\boldsymbol{u})]. \qquad (4.\text{B}.2)$$

Lemma 4.B.2. *If Assumptions 4.2.2 - 4.2.4 hold, then:*

$$\left| \frac{1}{M} \sum_{1\leq t\leq M} \boldsymbol{\ell}_t''(\boldsymbol{\theta}) - \boldsymbol{\Lambda}(\boldsymbol{\theta}) \right| \to 0 \ a.s.$$

Proof. Suppose that Assumptions 4.2.4 (B) (ii) and (iii) hold, we have:

$$
\left| \frac{1}{M} \sum_{t=1}^{M} \left(\frac{1}{e_t(\boldsymbol{\theta})^2} e_t'(\boldsymbol{\theta})^\top e_t'(\boldsymbol{\theta}) + \frac{f_t(v_t(\boldsymbol{\theta})|\mathscr{F}_{t-1})}{-\alpha e_t(\boldsymbol{\theta})} v_t'(\boldsymbol{\theta})^\top v_t'(\boldsymbol{\theta}) \right) \right.
$$
$$
\left. - \mathbb{E} \left[\frac{1}{e_t(\boldsymbol{\theta})^2} e_t'(\boldsymbol{\theta})^\top e_t'(\boldsymbol{\theta}) + \frac{f_t(v_t(\boldsymbol{\theta})|\mathscr{F}_{t-1})}{-\alpha e_t(\boldsymbol{\theta})} v_t'(\boldsymbol{\theta})^\top v_t'(\boldsymbol{\theta}) \right] \right| \xrightarrow{P} \mathbf{0}
$$

The proof follows Theorem 3 of Patton et al. (2019). □

Lemma 4.B.3. *If the assumptions of Theorem 4.2.1 are satisfied, then:*

$$
\sup_{1 \le k < \infty} \frac{\left| \sum\limits_{M < t \le M+k} \boldsymbol{\ell}_t'(\hat{\boldsymbol{\theta}}_M) - \left(\sum\limits_{M < t \le M+k} \boldsymbol{\ell}_t'(\boldsymbol{\theta}) + \left(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta} \right) k \boldsymbol{\Lambda}(\boldsymbol{\theta}) \right) \right|}{M^{1/2} \left(1 + \frac{k}{M} \right) b \left(\frac{k}{M} \right)} = o_P(1),
$$

as $M \to \infty$.

Proof. First we show that:

$$
\sup_{1 \le k < \infty} \frac{\left| \sum\limits_{M < t \le M+k} \boldsymbol{\ell}_t''(\boldsymbol{\theta}) - k \boldsymbol{\Lambda}(\boldsymbol{\theta}) \right|}{M \left(1 + \frac{k}{M} \right) b \left(\frac{k}{M} \right)} = o_P(1), \tag{4.B.3}
$$

as $M \to \infty$. Since $\boldsymbol{\ell}_t''(\boldsymbol{\theta})$ is a stationary sequence, by Assumption 4.2.5, (4.B.3) implies that:

$$
\sup_{1 \le k < \infty} \frac{\left| \sum\limits_{M < t \le M+k} \boldsymbol{\ell}_t''(\boldsymbol{\theta}) - k \boldsymbol{\Lambda}(\boldsymbol{\theta}) \right|}{M + k} = o_P(1), \tag{4.B.4}
$$

as $M \to \infty$. According to Lemma 4.B.2, we can derive (4.B.4).

Then Theorem 2 of Patton et al. (2019) implies that:

$$
\left| \hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta} \right| = O_P(M^{-1/2}). \tag{4.B.5}
$$

Using the mean value theorem for $\sum \boldsymbol{\ell}_t'(\hat{\boldsymbol{\theta}}_M)$ and (4.B.3) and (4.B.5) for $\sum \boldsymbol{\ell}_t''(\hat{\boldsymbol{\theta}}_M)$, we have Lemma 4.B.3. □

Lemma 4.B.4. *If Assumptions 4.2.2 - 4.2.4 hold, then:*

$$\sqrt{M}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}) = -\boldsymbol{\Lambda}^{-1}(\boldsymbol{\theta})\frac{1}{\sqrt{M}}\sum_{t=1}^{M}\boldsymbol{\ell}'_t(\boldsymbol{\theta}) + o_P(1).$$

Proof. Lemma 4.B.4 follows from Lemma 1 of Patton et al. (2019). □

Lemma 4.B.5. *If the assumptions of Theorem 4.2.1 hold, then:*

$$\sup_{1 \le k < \infty} \frac{\left|\sum_{M < t \le M+k} \boldsymbol{\ell}'_t(\hat{\boldsymbol{\theta}}_M) - \left(\sum_{M < t \le M+k} \boldsymbol{\ell}'_t(\boldsymbol{\theta}) - \frac{k}{M}\sum_{1 \le t \le M}\boldsymbol{\ell}'_t(\boldsymbol{\theta})\right)\right|}{M^{1/2}\left(1 + \frac{k}{M}\right)b\left(\frac{k}{M}\right)} = o_P(1),$$

as $M \to \infty$.

Proof. The Lemma 7 of Patton et al. (2019) states the asymptotic normality of $M^{-1/2}\sum_{1 \le t \le M}\boldsymbol{\ell}'_t(\boldsymbol{\theta})$. If Assumption 4.2.5 holds, we have:

$$\max_{1 < k \le M} \frac{\left|\frac{k}{M}\sum_{1 \le t \le M}\boldsymbol{\ell}'_t(\boldsymbol{\theta})\right|}{M^{1/2}\left(1 + \frac{k}{M}\right)b\left(\frac{k}{M}\right)}$$

$$\le \sup_{1 \le k < \infty} \frac{\frac{k}{M}}{1 + \frac{k}{M}} \sup_{1 \le j < \infty} \frac{1}{b\left(\frac{j}{k}\right)} M^{-1/2}\left|\sum_{1 \le t \le M}\boldsymbol{\ell}'_t(\boldsymbol{\theta})\right| = o_P(1).$$

Hence combining Lemmas 4.B.3 and 4.B.4, we have the results in Lemma 4.B.5.

□

Lemma 4.B.6. *If the assumptions of Theorem 4.2.1 hold, then:*

$$\sup_{1 \le k < \infty} \frac{\left|\sum_{M < t \le M+k} \boldsymbol{\ell}'_t(\boldsymbol{\theta}) - \frac{k}{M}\sum_{1 \le t \le M}\boldsymbol{\ell}'_t(\boldsymbol{\theta})\right|}{M^{1/2}\left(1 + \frac{k}{M}\right)b\left(\frac{k}{M}\right)} \xrightarrow{\mathcal{D}} \sup_{0 < u < \infty} \frac{|W_{\boldsymbol{D}}(1 + u) - (1 + u)W_{\boldsymbol{D}}(1)|}{(1 + u)b(u)},$$

as $M \to \infty$, *where* $W_{\boldsymbol{D}}(s)$ *is a Gaussian process with the mean* $\mathbb{E}[W_{\boldsymbol{D}}(s)] = 0$ *and* $\mathbb{E}[W_{\boldsymbol{D}}(s)^{\top}W_{\boldsymbol{D}}(s')] = \min(s, s')\boldsymbol{D}$.

Proof. As is shown in Assumption 4.2.1, $\boldsymbol{\ell}'_t(\boldsymbol{\theta})$ is a stationary ergodic martingale

difference sequence. We also have $Cov(\boldsymbol{\ell}'_t(\boldsymbol{\theta})) = \boldsymbol{D}$. Thus, based on the Cramér-Wold device, we have that, for any $\tilde{\tau} > 0$,

$$M^{-1/2} \sum_{1 \le t \le uM} \boldsymbol{\ell}'_t(\boldsymbol{\theta}) \xrightarrow{\mathcal{D}(0,\tilde{\tau})} W_{\boldsymbol{D}}(u) \quad \text{as } M \to \infty. \tag{4.B.6}$$

Hence,

$$M^{-1/2} \left(\sum_{M < t \le (1+u)M} \boldsymbol{\ell}'_t(\boldsymbol{\theta}) - u \sum_{1 \le t \le M} \boldsymbol{\ell}'_t(\boldsymbol{\theta}) \right) \xrightarrow{\mathcal{D}(0,\tilde{\tau})} W_{\boldsymbol{D}}(1+u) - (1+u)W_{\boldsymbol{D}}(1), \tag{4.B.7}$$

for any $\tilde{\tau} > 0$ as $M \to \infty$. By the Hájek-Rényi-Chow inequality, we have:

$$\lim_{\tilde{\tau} \to \infty} \limsup_{M \to \infty} P \left(\sup_{\tilde{\tau}M \le k < \infty} \left| \sum_{1 \le t \le M+k} \boldsymbol{\ell}'_t(\boldsymbol{\theta}) \right| \middle/ \left(M^{1/2} \left(1 + \frac{k}{M}\right) b\left(\frac{k}{M}\right) \right) \ge x \right) = 0, \tag{4.B.8}$$

for any $x > 0$. The coordinates of $W_{\boldsymbol{D}}(u)$ are Brownian motions, so by the law of the itegrated logarithm, we have:

$$\sup_{\tilde{\tau} \le u < \infty} \frac{W_{\boldsymbol{D}}(1+u)}{(1+u)b(u)} \to 0 \text{ a.s. } \tilde{\tau} \to \infty. \tag{4.B.9}$$

By using (4.B.7) and (4.B.9), we can prove the Lemma 4.B.6 $\qquad \square$

Proof. Proof of Theorem 4.2.1. Putting together Lemmas, we have:

$$\sup_{1 \le t < \infty} \frac{\left| \sum_{M < i \le M+k} \boldsymbol{\ell}'_t(\boldsymbol{\theta}_M) \hat{\boldsymbol{D}}_M^{-1/2} \right|}{M^{1/2} \left(1 + \frac{t}{M}\right) b\left(\frac{t}{M}\right)} \xrightarrow{\mathcal{D}} \sup_{0 < u < \infty} \frac{\left| (W_{\boldsymbol{D}}(1+u) - (1+u)W_{\boldsymbol{D}}(1)) \boldsymbol{D}^{-1/2} \right|}{(1+u)b(u)}.$$

Elementary arguments show that:

$$\mathbb{E}\left[\left((W_{\boldsymbol{D}}(1+u) - (1+u)W_{\boldsymbol{D}}(1)) \boldsymbol{D}^{-1/2}\right)^\top \left((W_{\boldsymbol{D}}(1+u) - (1+u)W_{\boldsymbol{D}}(1)) \boldsymbol{D}^{-1/2}\right) \right]$$
$$= u(1+s)\boldsymbol{I}_d, \quad u \le s,$$

where $\boldsymbol{I}_d$ is the identity matrix in $\mathbb{R}^d$. Computing the covariances one can verify that:

$$\left(\boldsymbol{W_D}(1+u) - (1+u)\boldsymbol{W_D}(1)\right)\boldsymbol{D}^{-1/2}$$
$$\stackrel{\mathcal{D}}{=} \left\{(1+u)W_1\left(\frac{u}{1+u}\right), \ldots, (1+u)W_d\left(\frac{u}{1+u}\right)\right\}, \quad u \geq 0,$$

where $W_1, W_2, \ldots, W_d$ are independent Wiener processes. Hence,

$$\sup_{0<u<\infty} \frac{\left|\left(\boldsymbol{W_D}(1+u) - (1+u)\boldsymbol{W_D}(1)\right)\boldsymbol{D}^{-1/2}\right|}{(1+u)b(u)} \stackrel{\mathcal{D}}{=} \max_{1\leq i\leq d}\sup_{0<s<1} \frac{|W_i(s)|}{b\left(\frac{s}{1-s}\right)},$$

completing the proof of Theorem 4.2.1. $\qquad\square$

4.C   Tables

Table 4.C.1: Critical values for $\theta^{1/2-\gamma}|W(u^*)|/u^{*\gamma}$ for $\gamma = 0, .15, .25, .35, .45$ and .49, with the parameter vector dimension of $d = 4$

γ	q \ T	0.2M	0.4M	0.6M	0.8M	M
0	10%	1.013	1.326	1.519	1.654	1.754
	5%	1.113	1.457	1.669	1.817	1.927
	1%	1.317	1.725	1.976	2.151	2.281
0.15	10%	1.359	1.641	1.805	1.916	1.996
	5%	1.487	1.796	1.975	2.097	2.185
	1%	1.753	2.117	2.329	2.471	2.575
0.25	10%	1.670	1.910	2.045	2.134	2.197
	5%	1.821	2.083	2.230	2.327	2.396
	1%	2.132	2.440	2.612	2.725	2.806
0.35	10%	2.085	2.261	2.355	2.416	2.459
	5%	2.262	2.453	2.555	2.621	2.668
	1%	2.624	2.845	2.963	3.040	3.094
0.45	10%	2.755	2.830	2.869	2.894	2.911
	5%	2.950	3.030	3.072	3.098	3.116
	1%	3.357	3.449	3.496	3.526	3.547
0.49	10%	3.305	3.323	3.332	3.338	3.342
	5%	3.508	3.527	3.537	3.543	3.547
	1%	3.930	3.951	3.962	3.968	3.973

Note: The critical values $c(\gamma, q)$ are based on 1,000,000 replications of $\sup_{0 \le u^* \le 1} |W(u^*)|/u^{*\gamma}$. The Wiener process is approximated on a grid of 10,000 equally spaced points in $[0, 1]$.

Table 4.C.2: Critical values of the loss-based Wilcoxon test for each selected training and testing samples

		Time span	10%	5%	1%
Sample A: S&P 500 Index					
Case I: no break	Training	23/May/2003-20/Dec/2005	2.442	2.469	2.525
	Testing	21/Dec/2005-23/Feb/2007	2.231	2.269	2.325
Case II: financial crisis	Training	23/May/2003-20/Dec/2005	3.104	3.142	3.197
	Testing	21/Dec/2005-16/July/2010	1.246	1.478	1.768
Case III: COVID Pandemic	Training	06/Feb/2018-01/Oct/2019	2.193	2.253	2.341
	Testing	02/Oct/2019-31/Dec/2020	0.991	1.125	1.649
Sample B: GBP/USD exchange rate					
Case IV: Brexit	Training	18/Aug/2010-12/Feb/2014	3.329	3.364	3.434
	Testing	13/Feb/2014-09/Apr/2018	1.451	1.675	3.597

Note: The critical values of the loss-based Wilcoxon test for each sample are computed via 1000 times stationary bootstrapping with the optimal block length proposed by Patton et al. (2009).

Table 4.C.3: Estimated coefficients of the GARCH-Gaussian model in both training and monitoring samples

Variables	Financial crisis			COVID Pandemic			Brexit		
	Training sample	Before detection	After detection	Training sample	Before detection	After detection	Training sample	Before detection	After detection
β_0	0.019	0.014	0.030	0.047	0.042	0.068	0.006	0.001	0.022
	(0.013)	(0.009)	(0.015)	(0.012)	(0.042)	(0.038)	(0.003)	(0.001)	(0.012)
β_1	0.913	0.923	0.885	0.771	0.720	0.777	0.932	0.961	0.893
	(0.044)	(0.034)	(0.022)	(0.046)	(0.039)	(0.058)	(0.022)	(0.007)	(0.052)
β_2	0.043	0.044	0.107	0.186	0.280	0.204	0.043	0.034	0.024
	(0.018)	(0.016)	(0.021)	(0.045)	(0.046)	(0.074)	(0.013)	(0.006)	(0.015)
VaR	-1.043	-1.027	-2.812	-1.472	-1.583	-2.437	-0.771	-0.816	-0.929
ES	-1.316	-1.296	-3.521	-1.854	-1.988	-3.101	-0.966	-1.024	-1.148
Loss	0.311	0.303	1.297	0.802	0.833	1.289	0.004	0.060	0.228
Dominant Source	$\boldsymbol{\beta_0}$			$\boldsymbol{\beta_2}$			$\boldsymbol{\beta_1}$		

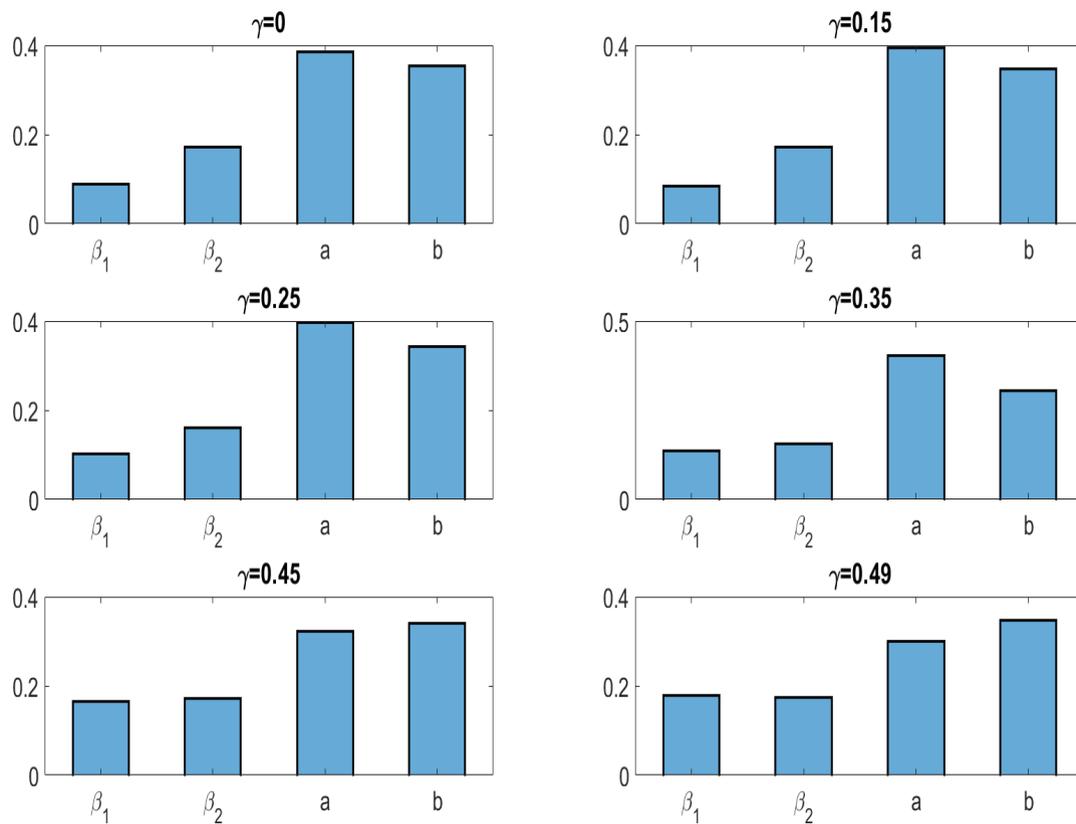## 4.D Figures for Simulations Results Based on the GARCH-FZ Model
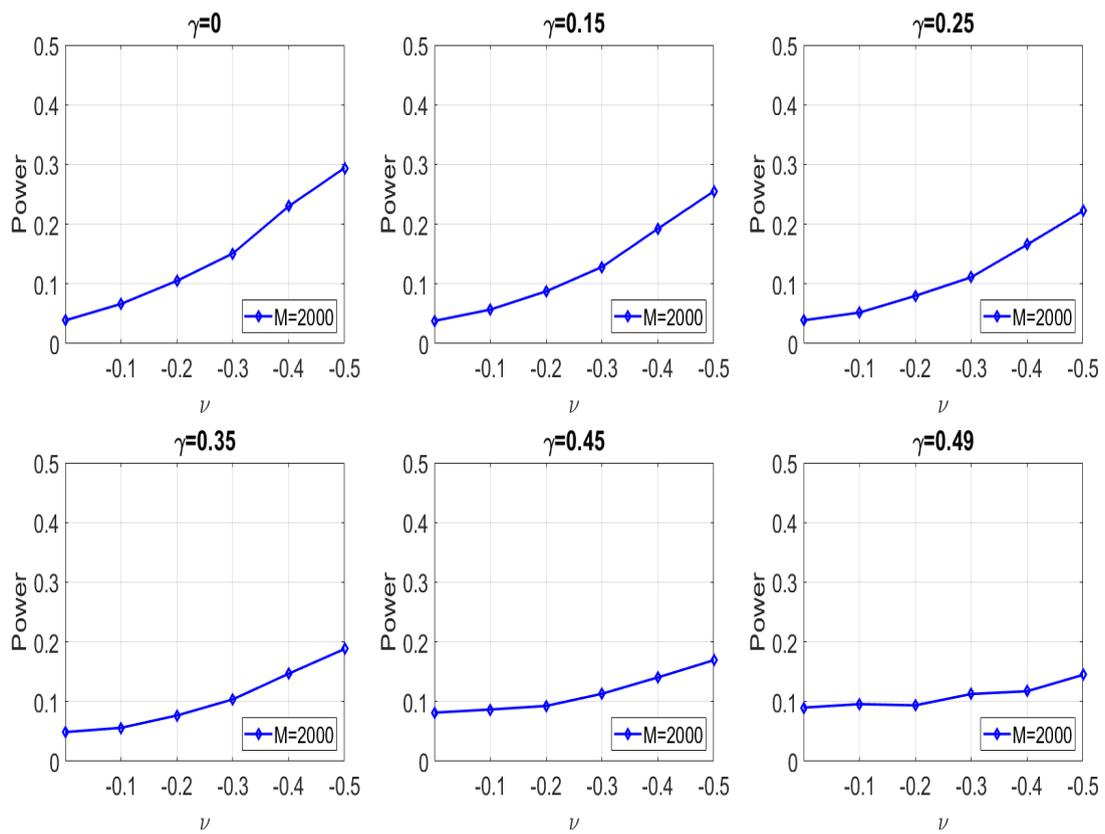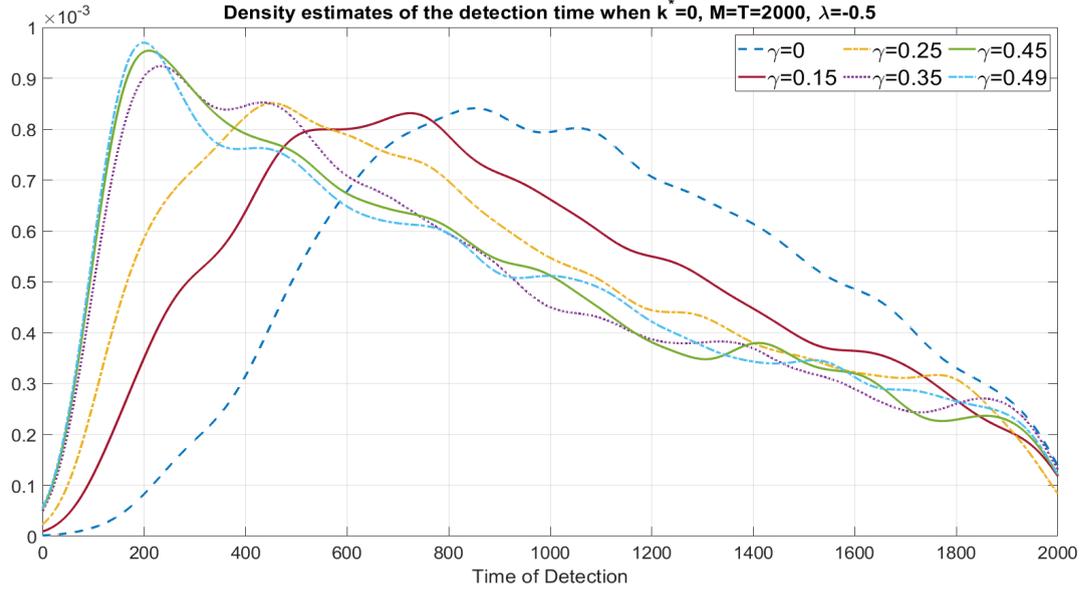
Figure 4.D.1: Empirical power of the monitoring scheme for a change point in the GARCH-FZ with $k^* = 0.5T$ under $H_{A,2}$
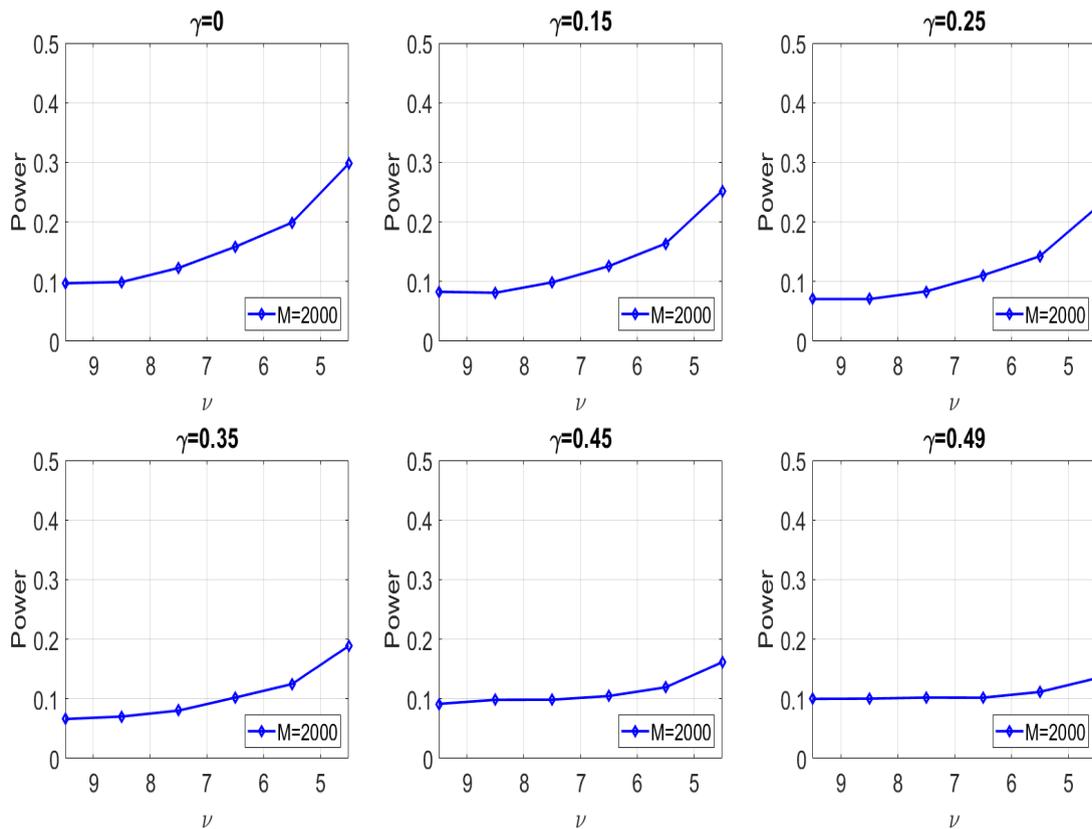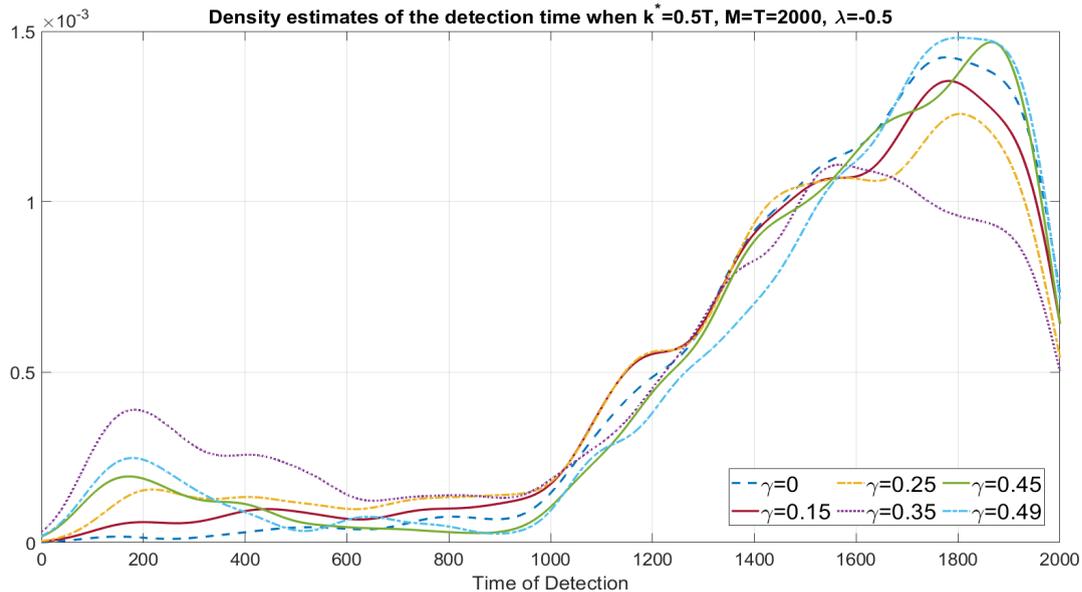

Note: This figure presents the empirical power of the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$ and $k^* = 0.5T$ under $H_{A,2}$.

Figure 4.D.2: Empirical density functions of the stopping time $\hat{k}^*$ for a change point in the GARCH-FZ with $k^* = 1$ and $\nu^* = 4.5$
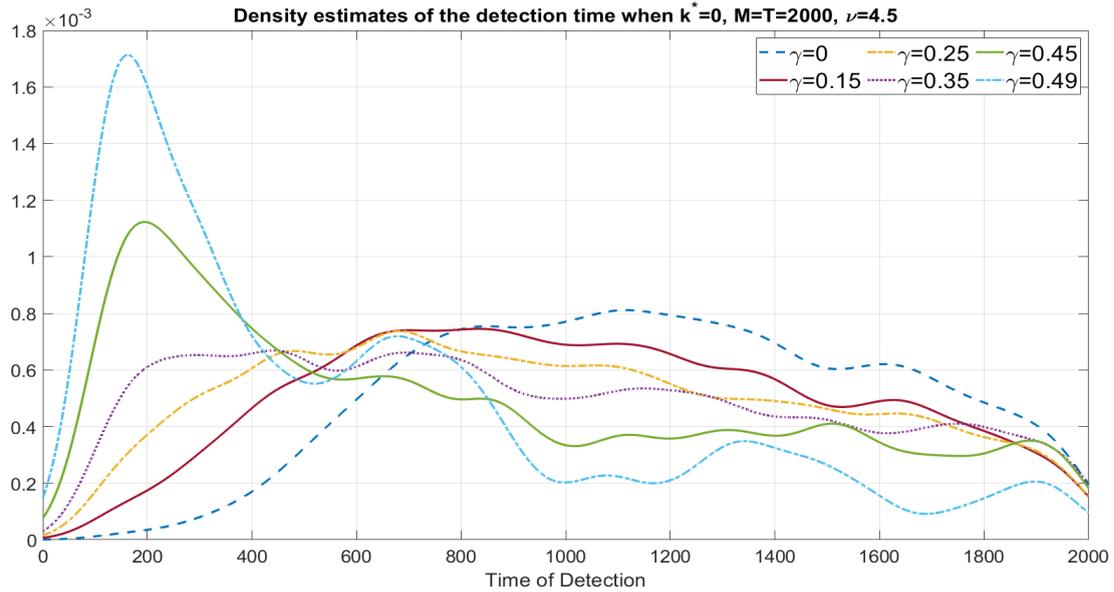

Note: This figure presents the empirical density functions of the stopping time $\hat{k}^*$ estimated by the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 1$ and $\nu^* = 4.5$.

Figure 4.D.3: Empirical density functions of the stopping time $\hat{k}^*$ for a change point in the GARCH-FZ with $k^* = 0.5T$ and $\nu^* = 4.5$
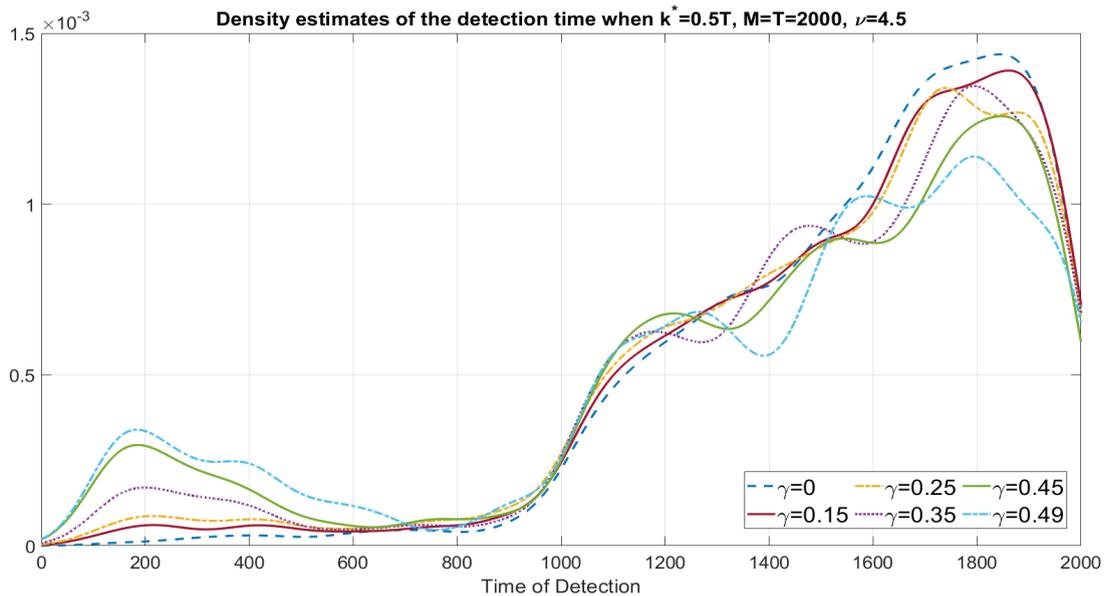
Note: This figure presents the empirical density functions of the stopping time $\hat{k}^*$ estimated by the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 0.5T$ and $\nu^* = 4.5$.

Figure 4.D.4: The dominant source of the change in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 0.5T$ and $\nu^* = 4.5$

Figure 4.D.5: Empirical power of the monitoring scheme for a change point in the GARCH-FZ with $k^* = 0.5T$ under $H_{A,3}$

Note: This figure presents the empirical power of the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$ and $k^* = 0.5T$ under $H_{A,3}$.

Figure 4.D.6: Empirical density functions of the stopping time $\hat{k}^*$ for a change point in the GARCH-FZ with $k^* = 1$ and $\lambda^* = -0.5$

Note: This figure presents the empirical density functions of the stopping time $\hat{k}^*$ estimated by the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 1$ and $\lambda^* = -0.5$.

Figure 4.D.7: Empirical density functions of the stopping time $\hat{k}^*$ for a change point in the GARCH-FZ with $k^* = 0.5T$ and $\lambda^* = -0.5$

Note: This figure presents the empirical density functions of the stopping time $\hat{k}^*$ estimated by the sequential monitoring scheme for a change point in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 0.5T$ and $\lambda^* = -0.5$.

Figure 4.D.8: The dominant source of the change in the GARCH-FZ for 5% (VaR, ES) at 5% test significance level with $M = T = 2000$, $k^* = 0.5T$ **and** $\lambda^* = -0.5$
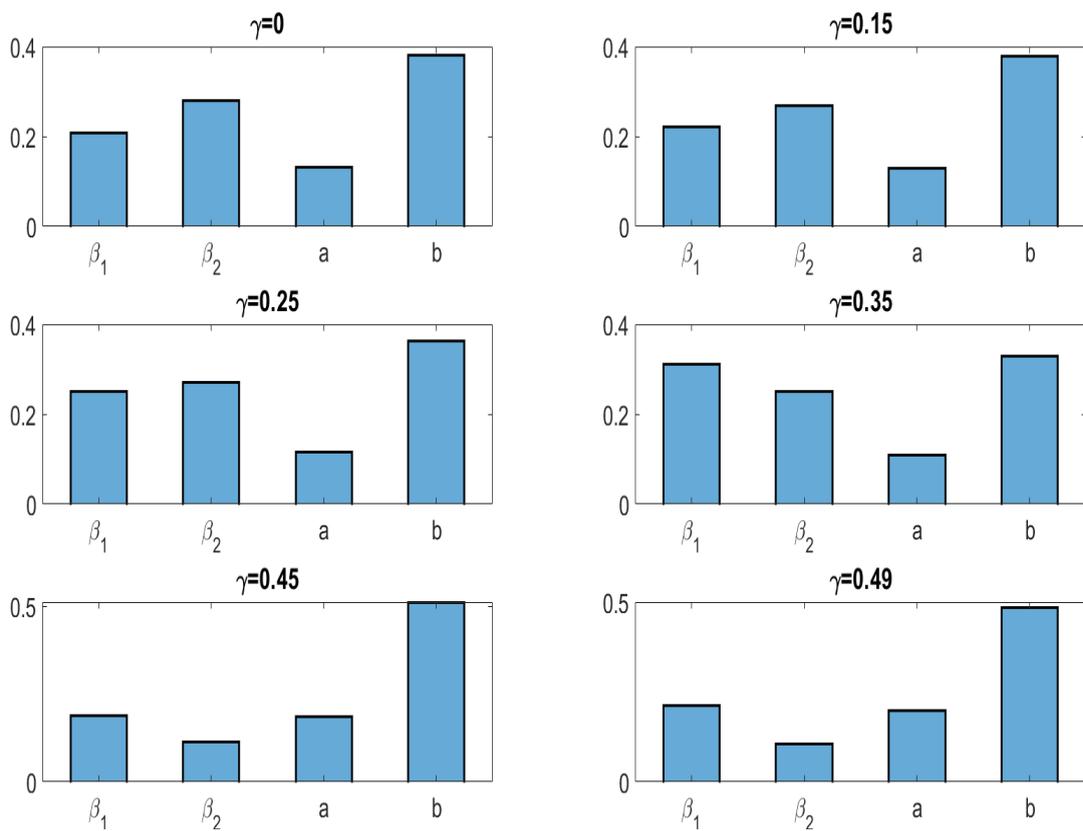
4.E   Figures for Simulations Results Based on the GARCH-skewed t Model

Figure 4.E.1: The dominant source of change point when β_1 increases to 0.94
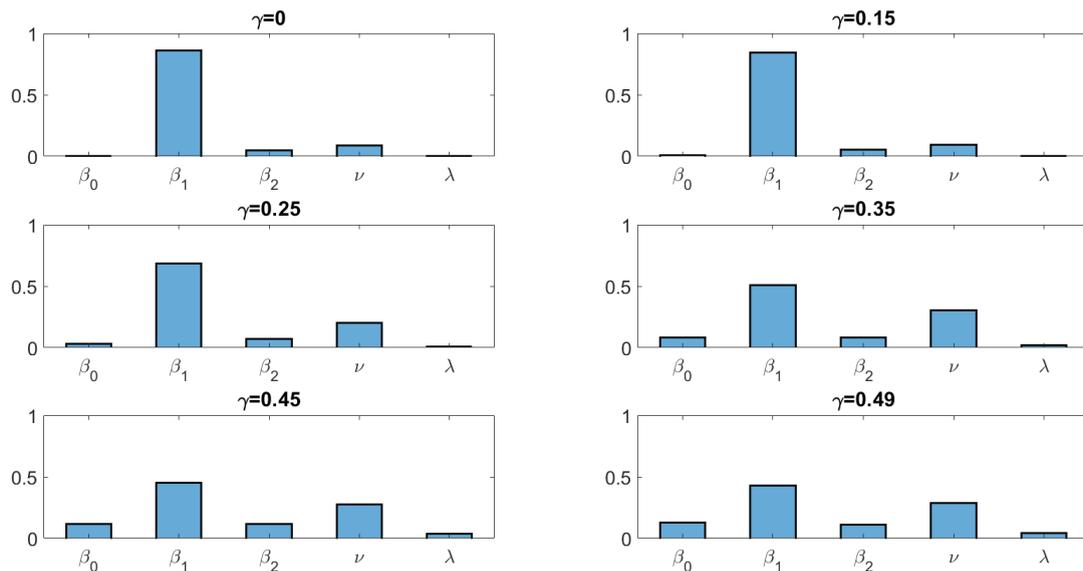
*Note:* We use the GARCH(1,1)-skewed t: $r_t = \sigma_t u_t, u_t \sim i.i.d.\ skewed\ t\ (\nu, \lambda), \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2$, where $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 20.5, 0)$ to simulate 3000 times for the training sample. After the change point $k^* = 1$, we increase the value of β_1 to 0.94. Here, the monitoring sample size T is equal to the training sample size M, which is 1000.

Figure 4.E.2: The dominant source of change point when β_2 increases to 0.09
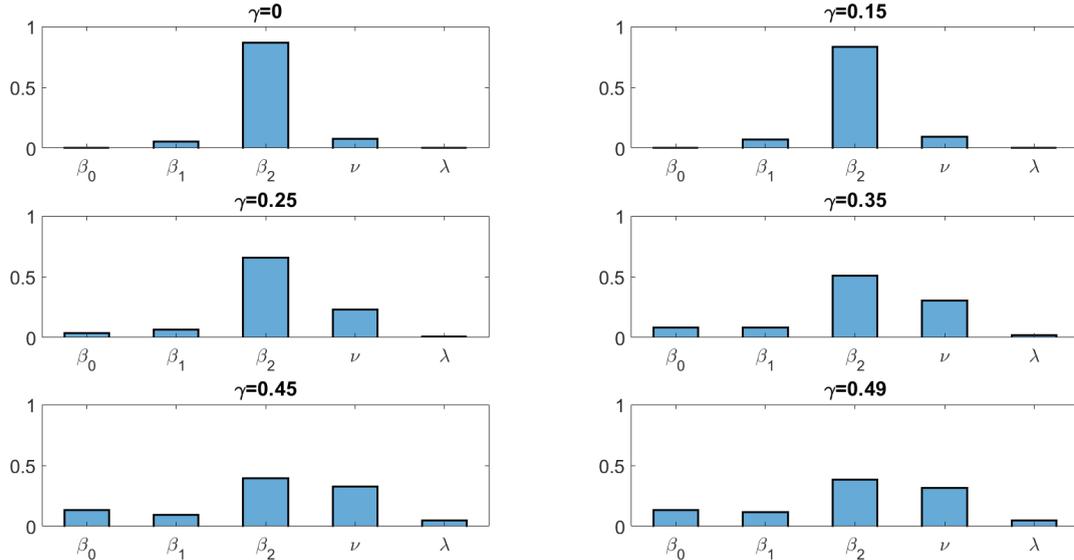
Note: We use the GARCH(1,1)-skewed t: $r_t = \sigma_t u_t, u_t \sim i.i.d.$ *skewed* $t\,(\nu, \lambda), \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2$, where $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 20.5, 0)$ to simulate 3000 times for the training sample. After the change point $k^* = 1$, we increase the value of β_2 to 0.09. Here, the monitoring sample size T is equal to the training sample size M, which is 1000.

Figure 4.E.3: The dominant source of change point when ν decreases to 4.5
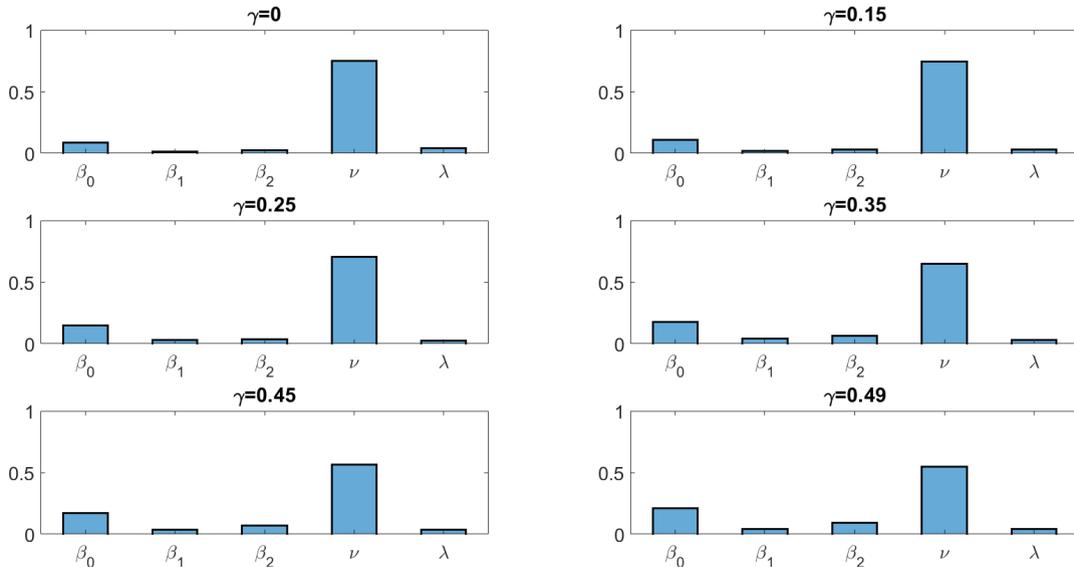
Note: We use the GARCH(1,1)-skewed t: $r_t = \sigma_t u_t, u_t \sim i.i.d.$ *skewed* $t\,(\nu, \lambda), \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2$, where $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 20.5, 0)$ to simulate 3000 times for the training sample. After the change point $k^* = 1$, we decrease the value of ν to 4.5. Here, the monitoring sample size T is equal to the training sample size M, which is 1000.

Figure 4.E.4: The dominant source of change point when λ decreases to -0.5
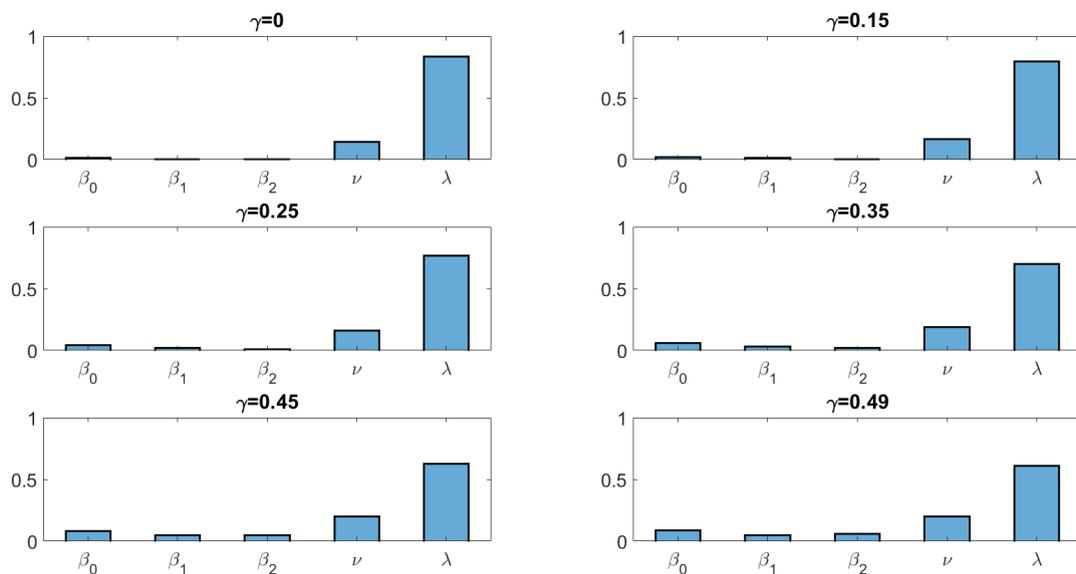
Note: We use the GARCH(1,1)-skewed t: $r_t = \sigma_t u_t, u_t \sim i.i.d.\ skewed\ t\ (\nu, \lambda), \sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 y_{t-1}^2$, where $(\beta_0, \beta_1, \beta_2, \nu, \lambda) = (0.05, 0.9, 0.05, 20.5, 0)$ to simulate 3000 times for the training sample. After the change point $k^* = 1$, we decrease the value of λ to -0.5. Here, the monitoring sample size T is equal to the training sample size M, which is 1000.

4.F Empirical Results Based on the GARCH-Gaussian Model

Figure 4.F.1: Real-time detection based on the GARCH-Gaussian model for the S&P 500 index within the financial crisis period
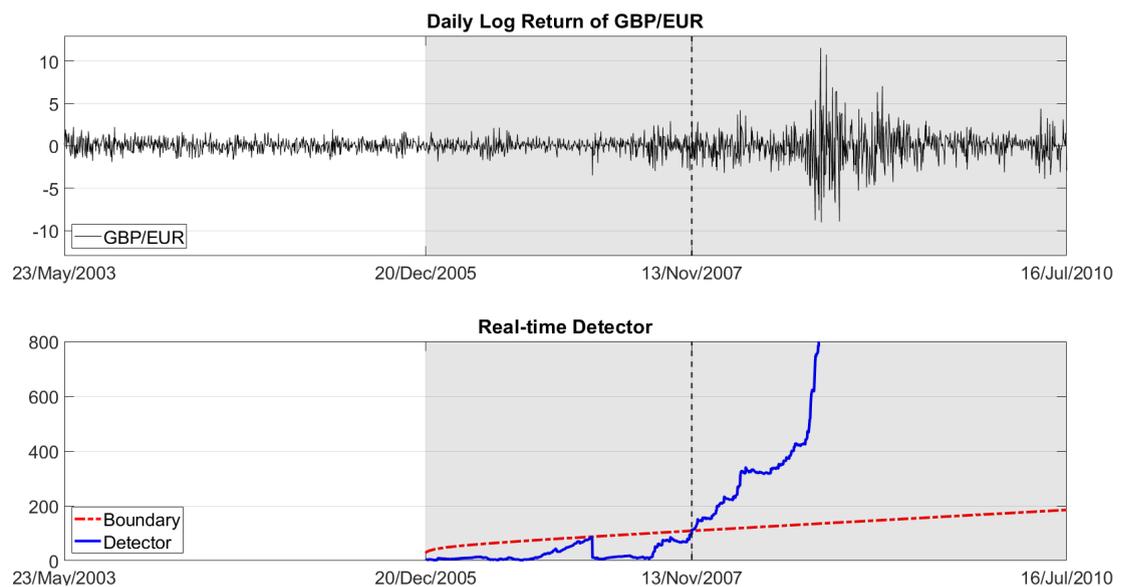
Note: Upper panel: The log return of the S&P 500 during 23 May 2003 to 16 July 2010; lower panel: Real-time detector based on the GARCH-Gaussian model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area). The vertical dash line denotes the estimated volatility change point.

Figure 4.F.2: Real-time detection based on the GARCH-Gaussian model for the S&P 500 index within the COVID-19 pandemic period
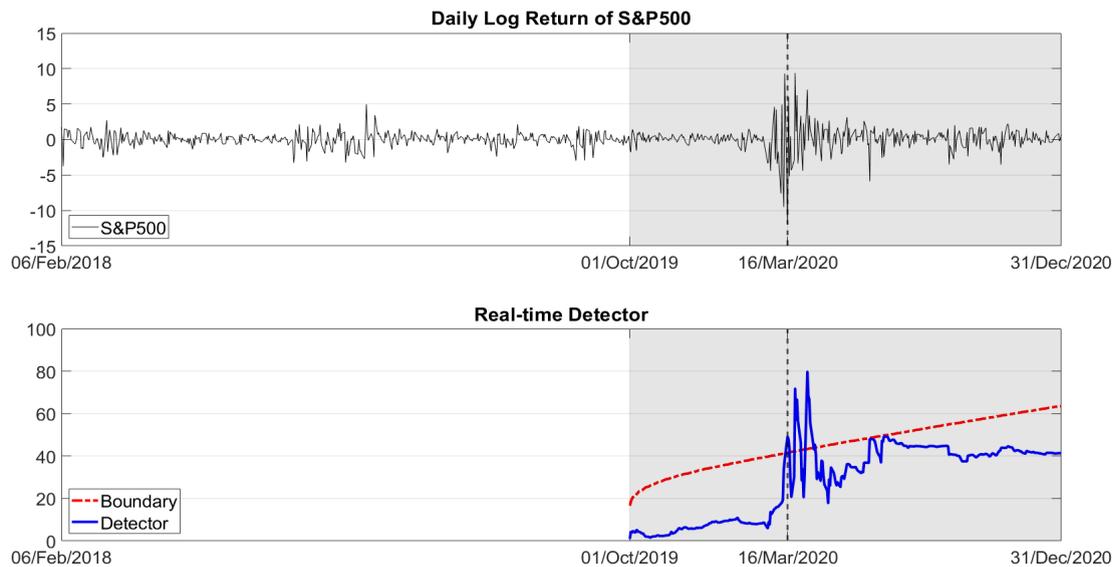
Note: Upper panel: The log return of the S&P 500 during 6 February 2018 to 31 December 2020; lower panel: Real-time detector based on the GARCH-Gaussian model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area). The vertical dash line denotes the estimated volatility change point.

Figure 4.F.3: Real-time detection based on the GARCH-Gaussian model for the GBP/EUR exchange rate
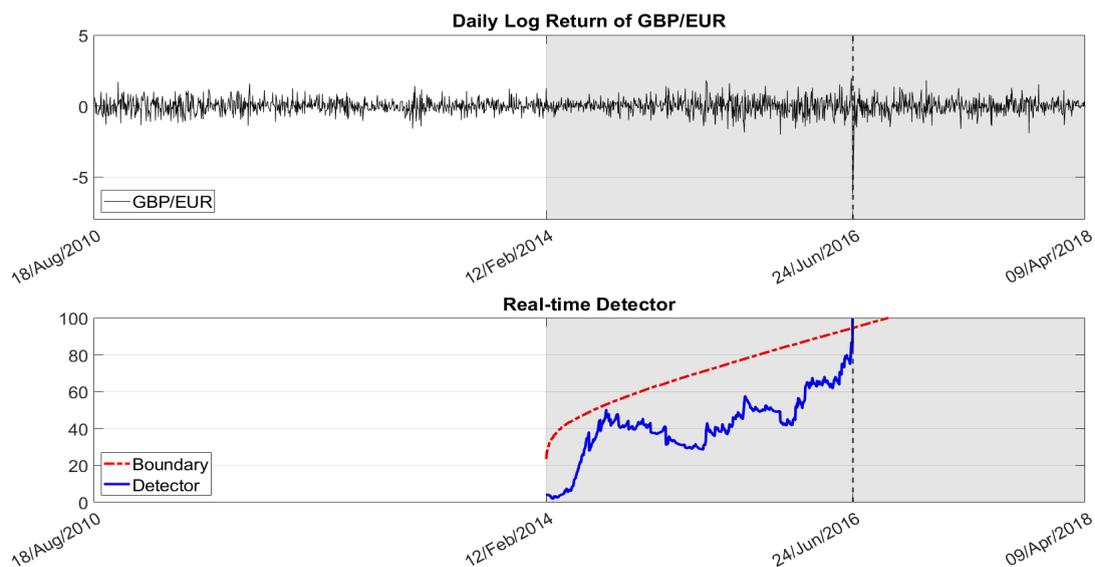
Note: Upper panel: The log return of the GBP/EUR exchange rate during 18 August 2010 to 9 April 2018; lower panel: Real-time detector based on the GARCH-Gaussian model versus the boundary function with $\gamma = 0.15$ in the testing period (gray shaded area). The vertical dash line denotes the estimated volatility change point.

Notes

[1]In the table, we only consider risk models with 3 to 5 parameters. More selected critical values are available upon request.

[2]The GARCH-FZ model is one of the semiparametric models for VaR and ES, which has a similar framework with the standard GARCH(1,1) model, but has an extension for VaR and ES modelling. Rather than estimating the parameters of this model using (Q)MLE, parameter estimates are obtained via FZ loss minimization.

[3]Results for other significance levels are available on request.

[4]In this model, we have $\beta_0 = 1 - \beta_1 - \beta_2$, so the parameter vector to be estimated is $\boldsymbol{\theta} = (\beta_1, \beta_2, a, b)$.

[5]The empirical densities of the stopping time are displayed by using kernel smoothing function estimates with a select bandwidth.

[6]The sequential monitoring scheme with a high value of γ is not applicable for the GARCH-FZ model to estimate 5% VaR and ES. The choice of the boundary parameter γ depends on model selection as well as the significance levels for VaR and ES.

[7]We also perform the identification of the dominant source of the change in the GARCH-skewed Normal model for 5% VaR and ES. More details can be found in Appendix 4.E.

[8]In the case of the effect of Brexit on the GBP/EUR exchange rate market, we compare change point detection in (VaR, ES) at 1%, 5% and 10% separately.

[9]According to the National Bureau of Economic Research (NBER), the Great Recession and the COVID-19 recession are during the period from December 2007 to June 2009, and the period from February 2020 to April 2020, respectively.

[10]The KS test statistic in Table 4.4.1 shows that the log returns in the training sample are probably from the same distribution. The loss-based Wilcoxon test indicates no change points detected in the time series of VaR and ES for the training sample.

Chapter 5

Conclusions and Further Research

5.1 Summary of the Findings

This thesis contributes to risk measurement and management by proposing improved estimation and forecasting methods for two widely used risk measures, VaR and ES. It also provides valuable suggestions regarding model construction and the timing of parameter adjustments for risk managers, regulators and other practitioners.

Chapter 2 introduces a set of risk models, which are extended from the dynamic semiparametric models proposed by Patton et al. (2019) to forecast the tuple (VaR, ES) jointly. These models incorporate the intraday and overnight information into the semiparametric GAS framework. We use the realized volatility at 5-min and 10-min frequencies and combine them with the overnight returns, respectively, to proxy the market fluctuation during the trading time as well as overnight. We observe an improvement in the estimation and forecasting performance of risk measures over both in-sample and out-of-sample horizons. In estimating the in-sample parameters, we show that, in general, the (VaR, ES) forecasts produced by the extended semiparametric models can generate rela-

tively low in-sample losses compared with the forecasts of the original models. Additionally, the coefficients of intraday and overnight variables in each comprehensive framework are statistically significant at 1% significance levels. This finding indicates that the in-sample estimation can benefit from adding intraday and overnight information into semiparametric risk models. We employ prevailing backtests in the current literature for the forecasts generated by our proposed models and benchmarks regarding the out-of-sample results. The out-of-sample results provide substantial evidence of the outperformance of our extended models for each backtest. Especially the semiparametric GAS two-factor model combined with the realized volatility at 10-min frequency and the overnight return can provide more accurate risk forecasts across different assets and probability levels. This chapter contributes to the literature on forecasting risk measures and the use of intraday information by providing solid empirical evidence.

Chapter 3 proposes an in-sample detection method for the change points in (semi)parametric models used for risk measure estimations. This method is based on the Wilcoxon test applied to the FZ loss functions for joint (VaR, ES), so we call it the loss-based Wilcoxon test. The general framework of the proposed test can accommodate any (semi)parametric models for VaR and ES if the consistency of the estimated parameters can hold. We derive the asymptotic behaviour of the proposed test statistic with specified conditions as the sample size converges to infinity. However, using the asymptotic limit to obtain p-values for a test statistic has been often criticized due to the oversized empirical results for small finite samples. To address the finite sample size distortions, we use the stationary bootstrap method to obtain the p-values. We also verify the validity of using stationary bootstrap for this test. The MC simulations are performed based on various setups: no change point in a stationary GARCH process, changes in the volatility and heavy-tailedness of the returns with different locations of change points. The simulation results reveal that the loss-based Wilcoxon test has better size control under the null hypothesis and higher power under alternative

hypotheses compared with alternative tests. Using the log returns of the S&P 500 index for the empirical application, we show the advantages of the proposed test with real data. Our test can identify the change points within the in-sample data, and each identified change point can be associated with a well-known market event. We contribute to the current literature on change point detection by proposing a method to identify changes in (semi)parametric (VaR, ES) jointly.

Chapter 4 investigates the sequential monitoring scheme for changes in M-estimators of semiparametric risk models for (VaR, ES). The proposed test is based on evaluating the change in the normalized gradient of the FZ loss function. This monitoring scheme detects a change point when the gradient-based detector crosses a selected boundary curve. In our simulations, we design a set of scenarios to evaluate the performance of our proposed test for finite sample sizes. The simulation results exhibit that this test has a reasonable size control under the null hypothesis and high empirical test power under various alternative hypotheses. In addition to examining the empirical size and power, we study the empirical density of the stopping time estimated by the proposed sequential monitoring scheme. The results point out that there is not much delay for the detected change point when we compare it with the location of the actual change point. Moreover, we explore the dominant source of the change point among the parameters of a semiparametric risk model in an innovative way. The dominant source of the change points can be identified when we modify the volatility, skewness or kurtosis of the simulated process. In the empirical analysis, we apply the proposed monitoring scheme for log returns of the S&P 500 index and the GBP/EUR exchange rate. Overall, the detected change point based on our proposed scheme can be associated with financial or economic events, and in some cases, the detection precedes the actual market crash, for example, the Black Thursday in the Covid-19 recession period. Our proposed sequential monitoring test can contribute to the ongoing debate on the structural breaks in risk measures and the tail index.

The main findings shown in this thesis can provide several practical suggestions for practitioners in financial markets. First, risk managers can incorporate the realized measures and overnight returns into the semiparametric observation-driven models to yield more robust VaR and ES forecasts. Also, banks and other financial institutions can adjust their minimum capital requirements according to the predicted 1% VaR and 2.5% ES. Based on the forecasts, asset managers can construct optimal portfolios by solving the mean-VaR or mean-ES optimization problem to achieve a reasonable trade-off between profit and risk. Furthermore, we conclude that risk management practitioners can improve the in-sample risk estimates by first identifying change points in the loss series for (VaR, ES) risk measures and then computing model parameter values based on the identified change points. Considering change points improves the plain (semi)parametric risk models in terms of risk measure estimation and forecasting. Finally, risk managers can benefit from using the sequential monitoring scheme to detect real-time change points in a risk model. By adjusting the parameter values of their risk models based on the timing identified by the monitoring scheme, practitioners can be aware of significant losses in their portfolios during market crashes and make corresponding preparations.

5.2   Suggestions for Future Research

While this thesis contributes to the ongoing debate on market risk measurement and the topics of risk estimation and forecasting from several perspectives, we still leave research gaps to be filled. We now discuss potential topics for further studies based on the main findings of this thesis.

Risk Forecasting Chapter 2 indicates that incorporating intraday and overnight information into a semiparametric risk model for joint (VaR, ES) can improve the forecasting accuracy of the risk models. However, our study only considers the

realized measures and overnight returns as explanatory variables. Along this line of research, one could rely on other intraday proxies, such as the realized volatility at other frequencies, realized bipower variation estimates (Barndorff-Nielsen and Shephard, 2004), good/bad realized volatility (Patton and Sheppard, 2015), or intraday ranges (Meng and Taylor, 2020). Additionally, one could use other exogenous information in the semiparametric risk model, e.g., the information from options markets (implied volatility or variance risk premium), the market sentiment extracted from news or firms announcements based on textual analysis, or low-frequency macroeconomic variables (one could solve the mixed frequency issue by employing the MIDAS framework by Engle et al. (2013)).

Another exciting research question is to develop the dynamic relationship between VaR and ES in a semiparametric model. One could follow the work of Taylor (2019) which models an AR relationship between VaR and ES and the study of Taylor (2022) which proposes a dynamic Omega ratio to describe the gap between VaR and ES. It would be helpful to construct new dynamics to model the ratio of VaR over ES.

Change Point Detection for Other Risk Measures Chapters 3 and 4 propose in-sample and real-time change points detection methods in risk measures. However, we only consider models for VaR and ES jointly in this thesis. Thus, it would be natural to propose change points detection methods for other market risk measures (volatility, expectile, and VaR individually) in both in-sample and real-time schemes. As discussed in the Introduction, elicitable risk measures have loss (scoring) functions. The forecasts of volatility are often backtested in a Q-LIKE framework that is based on the Gaussian or Student's t likelihood function. The generalized piecewise linear loss function introduced by Koenker and Bassett (1978) is the loss function to be used for VaR measures taken in isolation. Regarding the expectile, the asymmetric least squares loss function, proposed by Newey and Powell (1987), can be used. From a theoretical point of

view, the loss-based Wilcoxon test and the sequential monitoring test introduced in Chapters 3 and 4 can be extended to other risk measures based on the specified loss functions discussed above.

Change Point Detection for Tail Dependence In Chapters 3 and 4, we only consider the change points detection within a univariate process but ignore the dependence between the tails of bivariate processes. Ye et al. (2012) use an innovative change point testing method for structural changes in the dependence between time series of two equity index returns within an in-sample period. However, the sequential monitoring for real-time changes in tail dependence remains unexplored. To fill this gap, one may construct a novel method to identify the changes occurring in the tail dependence during the out-of-sample period, with the dependence modelled using copula functions.

Change Point Detection Method with Forward-looking Information To enhance the change point detection methods proposed in Chapters 3 and 4, it would be worthwhile to propose monitoring schemes that rely on forward-looking information (e.g., VIX and risk-neutral moments). In the existing literature, a strand of studies confirm the improvement of forecasting by using forward-looking information extracted from options markets (see Huggenberger et al., 2018; Molino and Sala, 2021, for example). Bauer and Huggenberger (2021) document that the risk estimates that incorporate information from options markets can quickly react to changing market conditions. The inclusion of forward-looking information is likely to improve the detection of change points, especially in fast moving markets.

Empirical Applications for Other Assets This thesis only considers stock indices and foreign exchange rates in the empirical studies. The in-depth theoretical results presented in each chapter could be applied to other types of assets, e.g.,

corporate bonds, options, commodities, commodity futures, and cryptocurrencies. As documented in several other studies (see Lazar and Zhang, 2019; Christoffersen et al., 2019; Liu et al., 2022, for more details), the log returns of commodities, commodity futures and cryptocurrencies are fat-tailed and negatively skewed. As such, it is worthwhile to apply the semiparametric models proposed in Chapter 2 for other asset classes. Also, another promising direction for further studies could be applying the change point detection methods in Chapters 3 and 4 to the market risk measures of the asset classes mentioned above.

References

Acerbi, C., Székely, B., 2014. Backtesting expected shortfall. Risk 27, 76–81.

Aigner, D. J., Amemiya, T., Poirier, D. J., 1976. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. International Economic Review 17, 377–396.

Alizadeh, S., Brandt, M. W., Diebold, F. X., 2002. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047–1091.

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

Andreou, E., Ghysels, E., 2002. Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics 17, 579–600.

Andrews, D. W., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.

Aue, A., Hörmann, S., Horváth, L., Hušková, M., 2014. Dependent functional linear models with applications to monitoring structural change. Statistica Sinica 24, 1043–1073.

Aue, A., Hörmann, S., Horváth, L., Reimherr, M., 2009. Break detection in the covariance structure of multivariate time series models. Annals of Statistics 37, 4046–4087.

Aue, A., Horváth, L., 2013. Structural breaks in time series. Journal of Time Series Analysis 34, 1–16.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Bali, T. G., Theodossiou, P., 2007. A conditional-SGT-VaR approach with alternative GARCH models. Annals of Operations Research 151, 241–267.

Barassi, M., Horváth, L., Zhao, Y., 2020. Change-point detection in the conditional correlation structure of multivariate volatility models. Journal of Business & Economic Statistics 38, 340–349.

Barendse, S., Patton, A., 2021. Comparing predictive accuracy in the presence of a loss function shape parameter. Journal of Business & Economic Statistics Forthcoming.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–37.

Basel Committee on Banking Supervision, 2013. Fundamental review of the trading book: A revised market risk framework. Available at `https://www.bis.org/publ/bcbs265.pdf`.

Basel Committee on Banking Supervision, 2019. Minimum capital requirements for market risk. Available at `https://www.bis.org/bcbs/publ/d457.pdf`.

Bauer, J., Huggenberger, M., 2021. Option-implied solvency capital requirements. working paper .

Bayer, S., 2018. Combining value-at-risk forecasts using penalized quantile regressions. Econometrics and Statistics 8, 56–77.

Berkes, I., Gombay, E., Horváth, L., Kokoszka, P., 2004. Sequential change-point detection in GARCH (p, q) models. Econometric Theory 20, 1140–1167.

Berkes, I., Horváth, L., Kokoszka, P., 2003. Estimation of the maximal moment exponent of a GARCH (1, 1) sequence. Econometric Theory 19, 565–586.

Bernardi, M., Catania, L., 2019. Switching generalized autoregressive score copula models with application to systemic risk. Journal of Applied Econometrics 34, 43–65.

Betken, A., 2016. Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. Journal of Time Series Analysis 37, 785–809.

Blair, B. J., Poon, S.-H., Taylor, S. J., 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. Journal of Econometrics 105, 5–26.

Bonaccolto, G., Caporin, M., Maillet, B. B., 2022. Dynamic large financial networks via conditional expected shortfalls. European Journal of Operational Research 298, 322–336.

Borovkova, S., Burton, R., Dehling, H., 2001. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. Transactions of the American Mathematical Society 353, 4261–4318.

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

Calhoun, G., 2018. Block bootstrap consistency under weak assumptions. Econometric Theory 34, 1383–1406.

Cerrato, M., Crosby, J., Kim, M., Zhao, Y., 2017. The joint credit risk of UK global-systemically important banks. Journal of Futures Markets 37, 964–988.

Chen, B., Hong, Y., 2016. Detecting for smooth structural changes in GARCH models. Econometric Theory 32, 740–791.

Chen, Q., Fang, Z., 2019. Improved inference on the rank of a matrix. Quantitative Economics 10, 1787–1824.

Christoffersen, P., Lunde, A., Olesen, K. V., 2019. Factor structure in commodity futures return and volatility. Journal of Financial and Quantitative Analysis 54, 1083–1115.

Chu, C.-S. J., Stinchcombe, M., White, H., 1996. Monitoring structural change. Econometrica 64, 1045–1065.

Clements, M. P., Galvão, A. B., Kim, J. H., 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. Journal of Empirical Finance 15, 729–750.

Clements, M. P., Hendry, D. F., 1996. Intercept corrections and structural change. Journal of Applied Econometrics 11, 475–494.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27, 46–78.

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

Csörgő, M., Horváth, L., 1988. Invariance principles for changepoint problems. Journal of Multivariate Analysis 27, 151–168.

Csörgő, M., Horváth, L., 1997. Limit theorems in change-point analysis. John Wiley & Sons Chichester.

Davidson, J., De Jong, R. M., 2000. The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally integrated processes. Econometric Theory 16, 643–666.

Dehling, H., Fried, R., Sharipov, O. S., Vogel, D., Wornowizki, M., 2013a. Estimation of the variance of partial sums of dependent processes. Statistics & Probability Letters 83, 141–147.

Dehling, H., Rooch, A., Taqqu, M. S., 2013b. Non-parametric change-point tests for long-range dependent data. Scandinavian Journal of Statistics 40, 153–173.

Dehling, H., Rooch, A., Taqqu, M. S., et al., 2017. Power of change-point tests for long-range dependent data. Electronic Journal of Statistics 11, 2168–2198.

Dehling, H., Wendler, M., 2010. Central limit theorem and the bootstrap for u-statistics of strongly mixing data. Journal of Multivariate Analysis 101, 126–137.

Dette, H., Gösmann, J., 2020. A likelihood ratio approach to sequential change point detection for a general class of parameters. Journal of the American Statistical Association 115, 1361–1377.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Diebold, F. X., Mariano, R. S., 2002. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Dimitriadis, T., Bayer, S., 2019. A joint quantile and expected shortfall regression framework. Electronic Journal of Statistics 13, 1823–1871.

Dimitriadis, T., Schnaitmann, J., 2021. Forecast encompassing tests for the expected shortfall. International Journal of Forecasting 37, 604–621.

Eckernkemper, T., 2017. Modeling systemic risk: Time-varying tail dependence when forecasting marginal expected shortfall. Journal of Financial Econometrics 16, 63–117.

Engle, R. F., Ghysels, E., Sohn, B., 2013. Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics 95, 776–797.

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

Fan, L., Glynn, P. W., Pelger, M., 2018. Change-point testing and estimation for risk measures in time series. arXiv preprint:1809.02303 .

Fearnhead, P., Rigaill, G., 2019. Changepoint detection in the presence of outliers. Journal of the American Statistical Association 114, 169–183.

Fissler, T., Ziegel, J., 2016. Higher order elicitability and Osband's principle. Annals of Statistics 44, 1680–1707.

Fissler, T., Ziegel, J. F., Gneiting, T., 2016. Expected shortfall is jointly elicitable with value at risk-implications for backtesting. Risk 29, 58–61.

Francq, C., Zakoïan, J.-M., 2015. Risk-parameter estimation in volatility models. Journal of Econometrics 184, 158–173.

Fuertes, A.-M., Olmo, J., 2013. Optimally harnessing inter-day and intra-day information for daily value-at-risk prediction. International Journal of Forecasting 29, 28–42.

Gerlach, R., Chen, C. W., 2014. Bayesian expected shortfall forecasting incorporating the intraday range. Journal of Financial Econometrics 14, 128–158.

Gerlach, R., Chen, C. W., 2017. Semi-parametric expected shortfall forecasting in financial markets. Journal of Statistical Computation and Simulation 87, 1084–1106.

Gerlach, R., Wang, C., 2016. Forecasting risk via realized GARCH, incorporating the realized range. Quantitative Finance 16, 501–511.

Gerlach, R., Wang, C., 2020. Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. International Journal of Forecasting 36, 489–506.

Gerlach, R., Wang, C., 2022. Bayesian semi-parametric realized conditional autoregressive expectile models for tail risk forecasting. Journal of Financial Econometrics 20, 105–138.

Gerstenberger, C., 2018. Robust Wilcoxon-type estimation of change-point location under short-range dependence. Journal of Time Series Analysis 39, 90–104.

Giot, P., Laurent, S., 2004. Modelling daily value-at-risk using realized volatility and arch type models. Journal of Empirical Finance 11, 379–398.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106, 746–762.

Gorgi, P., Hansen, P., Janus, P., Koopman, S., 2018. Realized Wishart-GARCH: a score-driven multi-asset volatility model. Journal of Financial Econometrics 17, 1–32.

Halbleib, R., Pohlmeier, W., 2012. Improving the value at risk forecasts: Theory and evidence from the financial crisis. Journal of Economic Dynamics and Control 36, 1212–1228.

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

Hansen, B. E., 2001. The new econometrics of structural change: dating breaks in US labour productivity. Journal of Economic Perspectives 15, 117–128.

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.

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

Hansen, P. R., Lunde, A., Voev, V., 2014. Realized beta GARCH: A multivariate GARCH model with realized measures of volatility. Journal of Applied Econometrics 29, 774–799.

Heber, G., Lunde, A., Shephard, N., 2009. Oxford-man institute's realized library version 0.1.

Hoga, Y., 2017. Testing for changes in (extreme) VaR. Econometrics Journal 20, 23–51.

Hoga, Y., Wied, D., 2017. Sequential monitoring of the tail behavior of dependent data. Journal of Statistical Planning and Inference 182, 29–49.

Holton, G. A., 2004. Value-at-Risk: Theory and Practice. Citeseer.

Horváth, L., Kokoszka, P., Wang, S., 2021a. Monitoring for a change point in a sequence of distributions. Annals of Statistics 49, 2271–2291.

Horváth, L., Kokoszka, P., Zhang, A., 2006. Monitoring constancy of variance in conditionally heteroskedastic time series. Econometric Theory 22, 373–402.

Horváth, L., Liu, Z., Lu, S., 2021b. Sequential monitoring of changes in dynamic linear models, applied to the US housing market. Econometric Theory Forthcoming.

Horváth, L., Liu, Z., Rice, G., Wang, S., 2020a. Sequential monitoring for changes from stationarity to mild non-stationarity. Journal of Econometrics 215, 209–238.

Horváth, L., Miller, C., Rice, G., 2020b. A new class of change point test statistics of Rényi type. Journal of Business & Economic Statistics 38, 570–579.

Hua, J., Manzan, S., 2013. Forecasting the return distribution using high-frequency volatility measures. Journal of Banking & Finance 37, 4381–4403.

Huggenberger, M., Zhang, C., Zhou, T., 2018. Forward-looking tail risk measures. Available at SSRN 2909808 .

Hušková, M., Kirch, C., 2008. Bootstrapping confidence intervals for the change-point of time series. Journal of Time Series Analysis 29, 947–972.

Hwang, E., Shin, D. W., 2015. Stationary bootstrap for U-statistics under strong mixing. Communications for Statistical Applications and Methods 22, 81–93.

Inclan, C., Tiao, G. C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association 89, 913–923.

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

Koenker, R., Machado, J. A., 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94, 1296–1310.

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

Lazar, E., Wang, S., Xue, X., 2021. Loss function-based change point detection in risk measures .

Lazar, E., Xue, X., 2020. Forecasting risk measures using intraday data in a generalized autoregressive score framework. International Journal of Forecasting 36, 1057–1072.

Lazar, E., Zhang, N., 2019. Model risk of expected shortfall. Journal of Banking & Finance 105, 74–93.

Leung, M., Li, Y., Pantelous, A. A., Vigne, S. A., 2021. Bayesian value-at-risk backtesting: The case of annuity pricing. European Journal of Operational Research 293, 786–801.

Liu, Y., Tsyvinski, A., Wu, X., 2022. Common risk factors in cryptocurrency. Journal of Finance Forthcoming.

Loschi, R. H., Iglesias, P. L., Arellano-Valle, R. B., Cruz, F. R., 2007. Full predictivistic modeling of stock market data: Application to change point problems. European Journal of Operational Research 180, 282–291.

Louzis, D. P., Xanthopoulos-Sisinis, S., Refenes, A. P., 2014. Realized volatility models and alternative value-at-risk prediction strategies. Economic Modelling 40, 101–116.

Lucas, A., Opschoor, A., 2018. Fractional integration and fat tails for realized covariance kernels. Journal of Financial Econometrics 17, 66–90.

Meng, X., Taylor, J. W., 2018. An approximate long-memory range-based approach for value at risk estimation. International Journal of Forecasting 34, 377–388.

Meng, X., Taylor, J. W., 2020. Estimating value-at-risk and expected shortfall using the intraday low and range data. European Journal of Operational Research 280, 191–202.

Mikosch, T., Stărică, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Review of Economics and Statistics 86, 378–390.

Molino, A., Sala, C., 2021. Forecasting value at risk and conditional value at risk using option market data. Journal of Forecasting 40, 1190–1213.

Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handbook of Econometrics 4, 2111–2245.

Newey, W. K., Powell, J. L., 1987. Asymmetric least squares estimation and testing. Econometrica 55, 819–847.

Newey, W. K., West, K. D., 1994. Automatic lag selection in covariance matrix estimation. The Review of Economic Studies 61, 631–653.

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

Nolde, N., Ziegel, J. F., 2017. Elicitability and backtesting: Perspectives for banking regulation. The Annals of Applied Statistics 11, 1833–1874.

Oh, D. H., Patton, A. J., 2018. Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. Journal of Business & Economic Statistics 36, 181–195.

Page, E. S., 1954. Continuous inspection schemes. Biometrika 41, 100–115.

Pástor, L., Stambaugh, R. F., 2001. The equity premium and structural breaks. Journal of Finance 56, 1207–1239.

Patton, A., Politis, D. N., White, H., 2009. Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. Econometric Reviews 28, 372–375.

Patton, A. J., 2020. Comparing possibly misspecified forecasts. Journal of Business & Economic Statistics 38, 796–809.

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Patton, A. J., Ziegel, J. F., Chen, R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). Journal of Econometrics 211, 388–413.

Pesaran, M. H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137, 134–161.

Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Politis, D. N., White, H., 2004. Automatic block-length selection for the dependent bootstrap. Econometric Reviews 23, 53–70.

Qu, Z., 2008. Testing for structural change in regression quantiles. Journal of Econometrics 146, 170–184.

Quaedvlieg, R., 2021. Multi-horizon forecast comparison. Journal of Business & Economic Statistics 39, 40–53.

Roccioletti, S., 2015. Backtesting Value at Risk and Expected Shortfall. Springer.

Salvatierra, I. D. L., Patton, A. J., 2015. Dynamic copula models and high frequency data. Journal of Empirical Finance 30, 120–135.

Shao, X., Zhang, X., 2010. Testing for change points in time series. Journal of the American Statistical Association 105, 1228–1240.

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

Smith, S. C., Timmermann, A., 2021. Break risk. Review of Financial Studies 34, 2045–2100.

Stock, J. H., Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business & Economic Statistics 14, 11–30.

Taylor, J. W., 2008. Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics 6, 231–252.

Taylor, J. W., 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. Journal of Business & Economic Statistics 37, 121–133.

Taylor, J. W., 2022. Forecasting Value at Risk and expected shortfall using a model with a dynamic Omega ratio. Journal of Banking & Finance Forthcoming.

Wang, C.-S., Zhao, Z., 2016. Conditional value-at-risk: Semiparametric estimation and inference. Journal of Econometrics 195, 86–103.

White, H., 1996. Estimation, inference and specification analysis. No. 22, Cambridge University Press.

Ye, W., Liu, X., Miao, B., 2012. Measuring the subprime crisis contagion: Evidence of change point analysis of copula functions. European Journal of Operational Research 222, 96–103.

Zhang, T., Lavitas, L., 2018. Unsupervised self-normalized change-point testing for time series. Journal of the American Statistical Association 113, 637–648.

Ziegel, J. F., 2016. Coherence and elicitability. Mathematical Finance 26, 901–918.

Žikeš, F., Baruník, J., 2014. Semi-parametric conditional quantile models for financial returns and realized volatility. Journal of Financial Econometrics 14, 185–226.