Three Essays on Estimation, Forecasting and Evaluation of Financial Risk

Dissertation

zur Erlangung des akademischen Grades eines Doktors der
Wirtschaftswissenschaften (Dr.rer.pol.)

vorgelegt von
Timo Dimitriadis

Universität
Konstanz

Sektion Politik – Recht – Wirtschaft
Fachbereich Wirtschaftswissenschaften
Konstanz, 2018

Tag der mündlichen Prüfung: 11. Dezember 2018

1. Referent: Dr. Roxana Halbleib
2. Referent: Prof. Dr. Winfried Pohlmeier
3. Referent: Prof. Dr. Yarema Okhrin

Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mich bei der Entstehung dieser Arbeit mit Hilfe, Unterstützung, Rat aber auch durch Unterhaltung im Privatleben begleitet haben.

Mein tiefster Dank gilt zunächst meiner Doktormutter Dr. Roxana Halbleib, die mir die Möglichkeit gegeben hat diese Arbeit in den letzten vier Jahren anzufertigen, mich dabei stets gefördert hat und mir mit hilfreichen Ratschlägen - sowohl akademisch als auch privat - zur Seite stand. Dies gilt sowohl für unser gemeinsames Forschungsprojekt als auch für die Freiheiten, die sie mir gelassen hat und Denkanstöße, die sie mir gegeben hat, eigene akademische Ideen zu entwickeln und auszuarbeiten. Des weiteren möchte ich Prof. Dr. Winfried Pohlmeier als Zweitbetreuer der Dissertation und Leiter der Arbeitsgruppe danken, insbesondere für die vielen hilfreichen Ratschläge in akademischen Diskussionen und in Seminaren. Für die Bereitschaft zur Übernahme der Drittbetreuung danke ich Prof. Dr. Yarema Okhrin. Sowohl bei der Heidelberger Akademie der Wissenschaften als auch bei der Graduate School of Decision Sciences möchte ich mich für die finanzielle Unterstützung bedanken.

Bei meinen Koautoren, Dr. Roxana Halbleib, Dr. Sebastian Bayer, Dr. Tobias Fissler und Prof. Dr. Johanna F. Ziegel möchte ich mich ebenfalls herzlichst bedanken. Es war mir stets ein Vergnügen mit solch wissensbegierigen, kompetenten und sympathischen Menschen zusammen zu arbeiten, sei es an der Universität Konstanz, in London bzw. Bern oder auch in Telefonanten. Für fachliche Diskussionen, gemeinsame Seminare, Korrekturlesen von Arbeiten und eine wunderbare Zusammenarbeit möchte ich mich bei den Mitgliedern der Ökonometrie Lehrstühle an der Universität Konstanz bedanken, insbesondere jedoch auch bei Jun.-Prof. Dr. Lyudmila Grigoryeva, Frederick Menninger, Patrick Schmidt und Britta Steffens.

Ein weiteres großes Dankschön geht an meine Freunde an der Universität Konstanz mit denen ich viele unterhaltsame Mittagessens-, Kaffee-, und Tischtennispausen verbringen

durfte, die mir den Arbeitsalltag versüßt haben. Ebenso danke ich meinen Mitbewohnern in Konstanz für das angenehme Zusammenleben und für die vielen gemeinsamen und spaßigen Abende. Für unzählige amüsante Nachmittage und Abende möchte mich ich sowohl bei meinem Volleyballteam als auch der Beachvolleyballgemeinschaft am Hörnle bedanken. Sowohl das sportliche Messen auf dem Spielfeld also auch das ein oder andere isotonische Kaltgetränk nach dem Sport haben mir den notwenigen Arbeitsausgleich verschafft.

Bei meiner Familie, insbesondere bei meinen Eltern, möchte ich mich zutiefst bedanken. Sowohl durch die mathematische und naturwissenschaftliche Erziehung als auch durch die Unterstützung während des Studiums und der Promotion wurde mir diese universitäre Ausbildung erst ermöglicht.

Zu guter Letzt kommt wie immer die wichtigste Person. Eine Beziehung mit einem Mathematiker, Ökonometriker oder allgemein Wissenschaftler zu führen mag nicht immer einfach sein. Gerade deshalb geht mein tiefster Dank an meine Lebensgefährtin Sandra, die mir die Freiheiten gelassen hat diesen Weg einzuschlagen, mich gleichzeitig dabei stets unterstützt, aber auch in gesunden Maße davon abgelenkt hat.

Timo Dimitriadis Konstanz, September 2018

Table of Contents

Summary

Motivated by the huge losses many firms encountered during the previous financial crisis between 2007 and 2010, there is a vast demand for reliable methods for generating financial risk forecasts as well as for the evaluation of these forecasts. As such a global financial crisis does not only affect the banks and financial institutions, but also global industrial companies and even whole economies, the Basel Committee of Banking Supervision (BCBS) as the regulatory authority of the international financial markets has to impose and enforce rules for an adequate risk management of the market participants. This set of rules contains the obligation to report risk forecasts which determine the capital requirements the banks have to hold as reserves. Thus, these risk forecasts (which are issued by the banks) have to be evaluated and tested for correctness by the BCBS. The scientific community consequently aims at providing the financial institutions with better estimation and forecasting techniques for financial risks as well as the BCBS with suitable and powerful tools for the evaluation of these risk forecasts.

This dissertation addresses the problem of both, generating forecasts and developing evaluation tools for the latter for the currently most popular risk measures Value at Risk (VaR) and Expected Shortfall (ES). The VaR is defined as the α-quantile of the return distribution where α is usually chosen to be 1% or 5%. This risk measure is generally criticized for being not *coherent* (see e.g. Artzner et al. (1999)) and for not accurately capturing tail risks beyond itself. This second critique aims at the mathematical fact that quantiles do not capture the exact form of the distribution beyond themselves. As a consequence of these drawbacks, starting two decades ago the academic community proposed the ES as an alternative risk measure. The ES is defined as the mean of the returns, given they are smaller than the respective α-quantile. This risk measure overcomes the theoretical drawbacks of the VaR as it is coherent and by its definition as a truncated tail mean, it is sensitive to extreme events in the tails. Consequently, the BCBS recently decided to introduce the ES into the regulatory framework of banking regulation (Basel Committee, 2016).

This cumulative thesis consists of three individual research papers which focus on different problems in estimation, forecasting and evaluation of these risk measures. The first chapter results from joint work with my supervisor, Dr. Roxana Halbleib from the University of Konstanz. The second and third chapters are both joint projects with Dr. Sebastian Bayer, a former PhD student at the University of Konstanz.

The first chapter, *How Informative is High-Frequency Data for Tail Risk Estimation and Forecasting? An Intrinsic Time Perspective* proposes a novel and simple approach to compute daily VaR and ES directly from high-frequency data. This estimation approach is based on the assumption that financial logarithm prices are subordinated unifractal processes in intrinsic time, which is a stochastic transformation of the clock time in accordance with the market activity. This assumption is very general and allows for a simple computation of the daily VaR and ES by scaling up their intraday counterparts computed from data sampled in intrinsic time. We denote these estimators by *Scaling-up VaR (ScaVaR)* and *Scaling-up ES (ScaES)*. This project is a first step in obtaining estimates for daily quantiles (and ES) directly from intraday high-frequency data, such as equivalently, the *Realized Volatility* estimates daily volatility directly from high-frequency data. As modern financial risk management is mainly concerned with the risk measures VaR and ES instead of the formerly used volatility, obtaining quantile (and ES) equivalents to Realized Volatility is of high importance. In the empirical exercise, we discuss the statistical and dynamic properties of the resulting daily VaR and ES estimates and show that our method outperforms standard ones in accurately estimating and forecasting VaR and ES.

In the second chapter, *A Joint Quantile and Expected Shortfall Regression Framework*, we introduce a novel regression technique which simultaneously models the VaR and the ES of a response variable given a set of covariates. This is the first instance in the literature where joint modeling of different functionals is crucial for the following reason: as shown by Gneiting (2011), the ES is not *elicitable* which implies that there does not exist a (strictly consistent) loss function which allows for consistently estimating parameters for a semiparametric regression framework for the ES stand-alone. However, in their seminal work, Fissler and Ziegel (2016) show that if one considers the ES jointly with the quantile at the same probability level, there exists a whole class of (strictly consistent) loss functions which can be used for the estimation of a joint regression procedure. In this article, we utilize these loss functions for M-estimation of the joint regression parameters. Furthermore, we introduce Z- (Method of Moments-) estimation of the parameters by using the (weak) derivatives of the loss functions as moment conditions. We show consistency and asymptotic normality for both, the M-estimator and the Z-estimator under weak regularity conditions for the full classes of loss and identification functions. The employed loss (and identification) functions depend on two specification functions, whose choices affect the properties of the resulting estimators. Extensive simulations verify the asymptotic properties and analyze the small sample behavior of the estimators for different choices of these specification functions. The simulation results show that estimators based on positively homogeneous loss functions perform superiorly and we further propose to rely on M-estimation of the model parameters as we find that the M-estimator is numerically more robust than the Z-estimator. This joint regression framework allows for various applications in estimating, forecasting and backtesting ES including dynamic ES models in the sense of the CAViaR model for quantiles (Engle and Manganelli, 2004), dynamic ES forecasting models (Žikeš and Baruník, 2016), ES forecast combinations, ES encompassing tests (Giacomini and Komunjer, 2005) and ES backtests based on Mincer-Zarnowitz regressions. We provide one application of this regression framework by introducing a novel Mincer-Zarnowitz backtest for the ES in the third chapter of this thesis.

The third chapter, *Regression Based Expected Shortfall Backtesting* introduces two novel backtests for the risk measure ES which are based on the joint regression framework for the quantile and the ES introduced in the second chapter of this thesis. These backtests are based on a Mincer-Zarnowitz regression for the ES, where we introduce a one-sided as well as a two-sided variant of this test. Our tests are the first in the literature which solely backtest the ES in the sense that they only require ES forecasts as input parameters. This property is crucial for the applicability of the tests for the BCBS as the respective financial institutions are only obligated to report ES forecasts and thus, the BCBS only has the issued ES forecasts at hand. We compare the empirical performance of our backtests to existing approaches (which backtest the VaR and the ES jointly) in terms of their empirical size and power through extensive simulation setups containing different data generating processes. These simulations show that our ES backtests outperform the competitors throughout all considered simulation designs in terms of both, the empirical size and the (size-adjusted) power. In practice, these backtests can be used by the BCBS in order to evaluate ES forecasts which is especially important as the BCBS decided to impose ES as standard risk measure without currently having a suitable backtest for ES forecasts at hand.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). “Coherent Measures of Risk”. *Mathematical Finance* 9 (3), 203–228 (see pp. 6, 10, 54, 97).

Basel Committee (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d352.pdf`. Basel Committee on Banking Supervision (see pp. 6, 11, 15, 26, 54, 97, 105).

Engle, R. and S. Manganelli (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles". *Journal of Business and Economic Statistics* 22 (4), 367–381 (see pp. 8, 12, 54, 73).

Fissler, T. and J. F. Ziegel (2016). "Higher order elicitability and Osband's principle". *The Annals of Statistics* 44 (4), 1680–1707 (see pp. 7, 12, 28, 54, 56–58, 60, 72, 78, 97, 119, 121–123).

Giacomini, R. and I. Komunjer (2005). "Evaluation and Combination of Conditional Quantile Forecasts". *Journal of Business & Economic Statistics* 23, 416–431 (see pp. 8, 12).

Gneiting, T. (2011). "Making and Evaluating Point Forecasts". *Journal of the American Statistical Association* 106 (494), 746–762 (see pp. 7, 11, 28, 54–58, 63, 97, 98, 102, 122).

Žikeš, F. and J. Baruník (2016). "Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility". *Journal of Financial Econometrics* 14 (1), 185–226 (see pp. 8, 12, 54, 73).

Zusammenfassung

Motiviert durch die enormen Verluste, die viele Unternehmen in der vorangegangenen Finanzkrise zwischen 2007 und 2010 erlitten haben, besteht ein enormer Bedarf an verlässlichen Methoden zur Erstellung von finanziellen Risikoprognosen sowie zur Auswertung dieser Prognosen. Da sich solch eine globale Finanzkrise nicht nur auf die Banken und Finanzinstitute, sondern auch auf globale Industriekonzerne und sogar auf ganze Volkswirtschaften auswirkt, muss der Basler Ausschuss für Bankenaufsicht (Basel Committee of Banking Supervision, BCBS) als Bankenaufichtsbehörde der internationalen Finanzmärkte Regeln für ein angemessenes Risikomanagement für diese Finanzinstitute erlassen und durchsetzen. Dieses Regelwerk enthält die Meldepflicht für Risikoprognosen, die die Höhe des nötigen Eigenkapitals für Rücklagen der Banken bestimmt. Daher müssen diese (von den Banken herausgegebenen) Risikoprognosen vom BCBS ausgewertet und auf ihre Korrektheit hin überprüft werden. Die wissenschaftliche Gemeinschaft hat es sich daher zum Ziel gesetzt, den Finanzinstituten bessere Schätz- und Prognoseverfahren für finanzielle Risiken sowie dem BCBS geeignete und wirksame Instrumente zur Auswertung dieser Risikoprognosen an die Hand zu geben.

Die vorliegende Dissertation befasst sich sowohl mit der Erstellung von Prognosen als auch mit der Entwicklung von Evaluationsmethoden für die derzeit gängigsten Risikomaße Value at Risk (VaR) und Expected Shortfall (ES). Der VaR ist definiert als das α-Quantil der Renditeverteilung, wobei α üblicherweise als 1% oder 5% gewählt wird. Dieses Risikomaß wird allgemein kritisiert, weil es nicht *kohärent* ist (Artzner u. a., 1999) und weil es extreme Risiken, die größer als der VaR selbst sind, nicht akkurat erfassen kann. Diese zweite Kritik zielt auf die mathematische Tatsache, dass Quantile nicht die genaue Form der Verteilung jenseits von sich selbst erfassen können. Als Folge dieser Nachteile schlug die akademische Gemeinschaft vor etwa zwei Jahrzehnten den ES als alternatives Risikomaß vor. Der ES (zum Level α) ist definiert als der Mittelwert der Renditen, die kleiner sind als das jeweilige α-Quantil (der VaR). Dieses Risikomaß überwindet die theoretischen Nachteile des VaR, da es kohärent ist und durch seine Definition als abgeschnittener Mittelwert extreme Ereignisse in den Rändern der Verteilung besser erfassen kann. Infolgedessen hat das BCBS vor kurzem beschlossen, den ES in die Richtlinien der internationalen Bankenregulierung aufzunehmen (Basel Committee, 2016).

Diese kumulative Dissertation besteht aus drei einzelnen Forschungsartikeln, die sich auf unterschiedliche Probleme bei der Schätzung, Vorhersage und Evaluierung dieser Risikomaße konzentrieren. Das erste Kapitel ist eine gemeinsame Arbeit mit meiner Betreuerin, Dr. Roxana Halbleib von der Universität Konstanz. Das zweite und dritte Kapitel sind jeweils gemeinsame Projekte mit Dr. Sebastian Bayer, einem ehemaligen Doktoranden an der Universität Konstanz.

Das erste Kapitel, *How Informative is High-Frequency Data for Tail Risk Estimation and Forecasting? An Intrinsic Time Perspective* schlägt einen neuen und einfachen Ansatz zur Berechnung des täglichen VaR und ES vor, welcher direkt auf hochfrequenten Daten basiert ist. Dieser Schätzungsansatz basiert auf der Annahme, dass die finanzielle Logarithmuspreise einem unifraktalen Prozess in *intrinsischer Zeit* folgen. Diese intrinsische Zeitdimension transformiert die klassische Zeit stochastisch entsprechend der Marktaktivität. Diese Annahme ist sehr allgemein und ermöglicht eine einfache Berechnung des täglichen VaR und ES durch Aufskalierung ihrer Intratages-Pendants, welche aus hochfrequenten Intratagesdaten (in intrinsischer Zeit) geschätzt werden. Wir nennen diese Schätzer *Scaling-up VaR (ScaVaR)* and *Scaling-up ES (ScaES)*. Dieses Projekt ist ein erster Schritt zur Erlangung von Schätzungen für tägliche Quantile (und ES) direkt aus Intratages-Hochfrequenzdaten im Sinne des *Realized Volatility* Schätzers, welcher die tägliche Volatilität direkt aus Hochfrequenzdaten schätzt. Da sich das moderne finanzielle Risikomanagement anstelle der früher verwendeten Volatilität vor allem mit den Risikomaßen VaR und ES beschäftigt, sind Quantil- (und ES-) Pendants zur *Realized Volatility* aktuell von sehr großer Bedeutung. Im empirischen Abschnitt dieses Arbeitspapiers diskutieren wir die statistischen und dynamischen Eigenschaften der sich ergebenden täglichen VaR- und ES-Schätzungen und zeigen, dass unsere Methode die Standardmethoden bei der Schätzung und Vorhersage von VaR und ES übertrifft.

Im zweiten Kapitel, *A Joint Quantile and Expected Shortfall Regression Framework*, stellen wir eine neue Regressionstechnik vor, die gleichzeitig den VaR und den ES einer abhängigen Variable gegeben einem Vektor von erklärenden Variablen modelliert. Dies stellt die erste Situation in der Literatur dar, in der die gemeinsame Modellierung verschiedener statistischer Funktionale von entscheidender Bedeutung ist. Der Grund dafür ist, dass der ES nicht *elicitable* ist (Gneiting, 2011). Dies bedeutet, dass es keine (streng konsistente) Verlustfunktion gibt, die eine konsistente Schätzung von Parametern für eine semiparametrische Regression für den ES alleine ermöglicht. In ihrer bahnbrechenden

Arbeit zeigen Fissler und Ziegel (2016), dass wenn man den ES gemeinsam mit dem Quantil basierend auf dem gleichen Wahrscheinlichkeitsniveau betrachtet, gibt es eine ganze Klasse von (streng konsistenten) Verlustfunktionen, die für die Schätzung eines gemeinsamen Regressionsverfahrens verwendet werden können. In diesem Artikel verwenden wir diese Verlustfunktionen für die M-Schätzung der gemeinsamen Regressionsparameter für das Quantil und den ES. Darüber hinaus stellen wir die Z- (Method of Moments-) Schätzung der Parameter unter Verwendung der (schwachen) Ableitungen der Verlustfunktionen als Momentenbedingungen vor. Wir zeigen Konsistenz und asymptotische Normalität für sowohl den M-Schätzer als auch den Z-Schätzer unter schwachen Regularitätsbedingungen für die vollen Klassen von Verlust- und Identifikationsfunktionen. Die verwendeten Verlust- (und Identifikations-) Funktionen hängen von zwei Spezifikationsfunktionen ab, deren Wahl die Eigenschaften der resultierenden Schätzer beeinflussen. Umfangreiche Simulationen verifizieren die asymptotischen Eigenschaften und analysieren das Verhalten der Schätzer für kleine Stichproben für verschiedene Wahlen dieser Spezifikationsfunktionen. Wir finden, dass Schätzer die auf positiv homogenen Verlustfunktionen basieren, eine bessere Schätzgenauigkeit aufweisen. Darüber hinaus schlagen wir die M-Schätzung der Modellparameter vor, da der M-Schätzer numerisch robuster ist als der Z-Schätzer. Diese gemeinsame Regressionsmethode ermöglicht verschiedene Anwendungen in der Schätzung, Vorhersage und Evaluierung (Testen) des ES. Beispiele dieser Anwendungen sind dynamische ES Modelle im Sinne des CAViaR-Modells für Quantile (Engle und Manganelli, 2004), dynamische ES-Prognosemodelle (Žikeš und Baruník, 2016), Kombinationen verschiedener ES Vorhersagen, ES-Encompassing Tests (Giacomini und Komunjer, 2005) und ES-Backtests auf Basis von Mincer-Zarnowitz-Regressionen. Wir illustrieren eine Anwendung dieses Regressionsverfahrens durch die Einführung eines neuartigen Mincer-Zarnowitz-Backtests für den ES im dritten Kapitel dieser Dissertation.

Das dritte Kapitel, *Regression Based Expected Shortfall Backtesting*, führt zwei neuartige Backtests für das Risikomaß ES ein, die auf der im zweiten Kapitel dieser Arbeit vorgestellten gemeinsamen Regressionsmethode für das Quantil und das ES aufgebaut sind. Diese Backtests basieren auf einer Mincer-Zarnowitz-Regression für den ES, bei der wir sowohl eine einseitige als auch eine zweiseitige Variante dieses Tests einführen. Unsere Tests sind die Ersten in der Literatur, die ausschließlich den ES in dem Sinne backtesten, dass die Tests nur ES-Prognosen als Information benötigen. Diese Eigenschaft ist ausschlaggebend für die Anwendbarkeit der Tests für das BCBS, da die jeweiligen Finanzinstitute nur zur Abgabe von ES-Prognosen verpflichtet sind und das BCBS somit nur diese ES-Prognosen zur Verfügung hat. Die empirischen Eigenschaften unserer Backtests vergleichen wir mit bestehenden Ansätzen (die den VaR und die ES gemeinsam backtesten) in Bezug auf ihre empirische

Spezifität und Sensitivität durch umfangreiche Simulationen basierend auf unterschiedlichen typischen (simulierten) Finanzdaten. Diese Simulationen zeigen, dass unsere ES-Backtests bei allen betrachteten Simulationen sowohl in Bezug auf die empirische Spezifität als auch auf die (Spezifität-berichtigte) Sensitivität die konkurrierenden Backtests übertreffen. In der Praxis können diese Backtests vom BCBS zur Evaluierung von ES-Prognosen verwendet werden. Dies ist zur Zeit besonders relevant, da das BCBS sich entschieden hat den ES als Standardrisikomaß zu verwenden ohne derzeit über einen geeigneten Backtest für ES-Prognosen zu verfügen.

Literatur

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). "Coherent Measures of Risk". *Mathematical Finance* 9 (3), 203–228 (see pp. 6, 10, 54, 97).

Basel Committee (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d352.pdf`. Basel Committee on Banking Supervision (see pp. 6, 11, 15, 26, 54, 97, 105).

Engle, R. and S. Manganelli (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles". *Journal of Business and Economic Statistics* 22 (4), 367–381 (see pp. 8, 12, 54, 73).

Fissler, T. and J. F. Ziegel (2016). "Higher order elicitability and Osband's principle". *The Annals of Statistics* 44 (4), 1680–1707 (see pp. 7, 12, 28, 54, 56–58, 60, 72, 78, 97, 119, 121–123).

Giacomini, R. and I. Komunjer (2005). "Evaluation and Combination of Conditional Quantile Forecasts". *Journal of Business & Economic Statistics* 23, 416–431 (see pp. 8, 12).

Gneiting, T. (2011). "Making and Evaluating Point Forecasts". *Journal of the American Statistical Association* 106 (494), 746–762 (see pp. 7, 11, 28, 54–58, 63, 97, 98, 102, 122).

Žikeš, F. and J. Baruník (2016). "Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility". *Journal of Financial Econometrics* 14 (1), 185–226 (see pp. 8, 12, 54, 73).

Chapter 1

How Informative is High-Frequency Data for Tail Risk Estimation and Forecasting? An Intrinsic Time Perspective

---

For helpful comments we would like to thank Sebastian Bayer, Mark Hallam, Jens Jackwerth, Winfried Pohlmeier and the members of the Chair of Econometrics at the Department of Economics at the University of Konstanz, Germany. All remaining errors are ours. Timo Dimitriadis gratefully acknowledges financial support from the WIN-Kolleg of the Heidelberg Academy of Sciences and Humanities, Germany and from the Graduate School of Decision Sciences (GSDS) at University of Konstanz, Germany. Roxana Halbleib gratefully acknowledges financial support from the European Social Fund, Ministry of Science, Research and Arts of Baden–Württemberg, Germany, from the Deutsche Forschungsgemeinschaft through the project PO375/11-1 titled ”Robust Risk Measures in Real Time Settings” and from the Zukunftskolleg at the University of Konstanz.

1.1. Introduction

The huge losses during and following the previous financial crisis have shown that the existing risk models have failed when they were needed most. Since its implementation as a standard approach by the Basel Committee on Banking Supervision in 1996 (Basel Committee, 1996), Value at Risk (VaR) has become the most known risk measure in today's financial world. In 2016, the Basel Committee recommended a shift from VaR to Expected Shortfall (ES) as a "more prudent" measure of risk given that it accounts for the distribution of financial losses beyond the VaR (Basel Committee, 2016). Although very appealing, the literature on estimating and forecasting ES relies heavily on the one of the VaR. The standard way of computing e.g., (daily) VaR is by using a location-scale model that involves the estimation and forecasting of the mean, of the standard deviation (or volatility) and of the quantiles of standardized residuals of (daily) financial returns from data sampled in calendar time and, usually, at low (daily) frequencies.[1] The location-scale approach stems originally from the assumption that daily financial returns are normally distributed, which has proven to be an unrealistic assumption during the last decades. To overcome this problem, the common practice is to replace the quantile of standardized residuals with one of a fat-tailed distribution, such as Student's t. However, Halbleib and Pohlmeier (2012), among others, provide empirical evidence on the general pitfalls of the location-scale model for the VaR, especially during financial turbulent times, and the necessity for more flexible and precise approaches.

This paper proposes a novel and simple method of estimating and forecasting daily VaR and ES directly from high-frequency (HF) data. It consists in scaling-up quantiles and ES of intraday returns sampled not in calendar time, but in a time dimension that aims at capturing the real "heartbeat" of the market. This new time dimension, which we denote here the intrinsic time, allows to sample more often during active/turbulent financial periods and less often during calm ones. Thus, the resulting intraday returns become very informative about the market's activity and riskiness, differently from the ones sampled in the standard clock (or calendar) time, which are picked independently of any type of information regrading the market's activity. Our method, which we denote the Scaled-up VaR (ScaVaR) and the Scaled-up ES (ScaES), is the first in the literature that directly exploits the rich information content of HF data from the intrinsic time perspective for tail risk estimation and forecasting.

The theoretical background of our approach is built on the assumption that financial logarithm price processes are unifractal in intrinsic time. The unifractality assumption

---

[1] Intraday data sampled in calendar time has found application in estimating and forecasting daily VaR by using Realized Volatility estimates for the standard deviation in the location-scale model (see Giot and Laurent (2004) among others).

allows for describing the relation between the distributions of financial returns computed at different aggregation (frequency) levels through one single scaling law driven by one scaling index. Based on this assumption, we are able to compute daily estimates of VaR and ES by scaling up their intraday counterparts by the single scaling index estimated from the data. This makes our procedure very simple compared to the location-scale models, as it avoids heavy parametric specifications for the mean, volatility and the distribution of the standardized returns.

We show that the assumption of unifractality in intrinsic time implies that the logarithm price process is multifractal and characterized by different scaling laws in accordance with the aggregation (frequency) level. As shown by Mandelbrot (1999), Calvet and Fisher (2008), Lux and Segnon (2018), multifractality is able to accommodate various features such as volatility clustering and persistence as well as fat-tailedness of the distributions, which are typical for financial returns. This makes our assumption more general and more realistic than the existent ones, such as Brownian Motion or fractional Brownian motion, which impose that financial returns are Gaussian.

The application of the fractal theory in finance has mainly focused on estimating and forecasting daily conditional volatilities by Mandelbrot et al. (1997), Calvet et al. (1997) and Calvet and Fisher (2002) who assume that the logarithm price process is a (fractional) Brownian motion in intrinsic time and by Calvet and Fisher (2001) who use multi-scaling relationships typical to multifractal processes.[2] Hallam and Olmo (2014b) and Hallam and Olmo (2014a) apply scaling relationships typical to fractality to estimate and forecast the whole probability density function (pdf) of daily financial returns from their intraday counterparts computed from data sampled in calendar time. The outcome of these two approaches is mainly driven by the information from the central region of the distribution instead of the tails, which makes them inappropriate for risk estimation and forecasting.

The concept of intrinsic time is not new in the economic literature and it has circulated under the name of "time deformation" (Mandelbrot et al., 1997), "deformed time" (Gouriéroux and Jasiak, 2001) or "operational time" (Clark, 1973), among others. It transforms the clock time according to an intensity measure characterizing the activity on the market (such as trading pattern, volatility level or volume) in the following manner: it accelerates the clock time when the intensity is high and it slows it down when the intensity is low. It is first used by Burns and Mitchell (1946) and formally introduced by Mandelbrot and Taylor (1967). So far, it has only sporadically found application, however for other purposes than the ones of this paper: to explain the fat-tailedness of financial returns (Clark, 1973), to describe the

---

[2] For an overview of these conditional volatility models see Calvet and Fisher (2008) and Lux and Segnon (2018).

evolution of macroeconomic variables (Stock, 1987, 1988), to capture the daily and weekly (seasonality) patterns in the volatility of foreign exchanges (Dacorogna et al., 1993; Müller et al., 1995), to model stochastic volatilities (Ghysels et al., 1997), to reparametrise diffusion equations (Gouriéroux and Jasiak, 2001), to increase the accuracy of Realized Volatility estimators (Corsi et al., 2001; Dong and Tse, 2014; McAleer and Medeiros, 2008; Oomen, 2005, 2006) or to describe the price change trend on foreign exchange markets (Guillaume et al., 1997).

Our ScaVaR and ScaES procedure consists in several simple steps: (1) sample HF prices in the intrinsic time based on a choice of intensity measure and at a certain frequency[3]; (2) estimate quantiles and ES of the resulting intraday logarithm returns; (3) estimate the scaling index that describes the unique scaling law of the unifractal price process and (4) compute daily estimates of VaR and ES by scaling up the intraday counterparts from point (2) with the scaling index obtained at point (3).

In our empirical application, we apply ScaVaR and ScaES to the exchange rate EURUSD and two stocks traded at NYSE (IBM and BAC). For our purposes, we apply the intraday trading pattern as the intensity measure driving the intrinsic time sampling scheme: i.e., the sampled prices are "equidistant" in terms of the number of trades. Further alternative intensity measures such as price changes, volumes or intraday volatility patterns may also be considered. For the sake of comparison, we also implement the standard clock-time sampling scheme, which samples prices equidistantly in time.

We provide a comprehensive description of the empirical properties of scaled-up daily VaR and ES estimates computed directly from HF data and we find that the estimates of foreign exchanges are less extreme, less volatile and less persistent than the ones of the stocks. Among the stocks, these effects are mostly pronounced for the asset from the financial sector (BAC). All VaR and ES estimates exhibit long memory and some skewness and overkurtosis.

Furthermore, we evaluate and compare the accuracy of ScaVaR and ScaES to estimate and forecast daily VaR and ES against a standard location-scale model and the fractal-based approach of Hallam and Olmo (2014a). We demonstrate that ScaVaR and ScaES significantly outperform the alternative methods, especially the method of Hallam and Olmo (2014a), in accurately estimating and forecasting daily VaR and ES. Comparing the two sampling schemes, we show that sampling the HF data in accordance with the trading activity is much more valuable for risk estimation and forecasting compared to the one sampled in clock time.

---

[3] Please note that in this paper we do not deal with the market microstructure noise, which affects HF prices. As it is a very complex topic by itself (as depicted in the Realized Volatility literature: Hansen and Lunde (2004), Bandi and Russell (2008), Zhang et al. (2005), Barndorff-Nielsen et al. (2008), among others), we leave it for further research.

This paper is organized as follows: Section 1.2 presents the theory of our approach and its implementation. Section 1.3 provides empirical results from applying our method and the alternatives to real data. Section 1.4 concludes.

1.2. Theory

In Section 1.2.1, we first briefly introduce the key mathematical concepts of the fractal theory. Section 1.2.2 provides the formal description of our methodology and Section 1.2.3 presents the practical steps for implementing our approach.

1.2.1. Key Theoretical Concepts

The most basic concept at the basis of our approach is the one of *unifractality* (Mandelbrot, 1963, 1982), which is defined in the following manner:

Definition 1.2.1. A real-valued stochastic process $\{U(t), t \geq 0\}$ is said to be *unifractal* or *self-affine* if for some $H > 0$ and all $c \geq 0$,

$$U(ct) \overset{d}{=} c^H U(t), \tag{1.1}$$

where $\overset{d}{=}$ denotes equality in distribution.

Hereby, the parameter H is known as the *self-affinity* or *scaling index* and it gives the (unique) scaling law describing the unifracal process. c is the scaling or aggregation factor that defines the time scale change from t to ct. It holds that, if $E[|U(1)|] < \infty$, then $H \leq 1$ and if $H = 1$, then $U(t) = tU(1)$ almost surely (Embrechts and Maejima, 2002). For this reason, we focus in our paper on the cases where $H \in (0, 1)$.

Assuming that the increments are Gaussian and if $H = 0.5$, then $U(t)$ is a standard Brownian Motion (BM), while if $H \neq 0.5$, then $U(t)$ is a fractional Brownian Motion (fBM). While both BM and fBM have stationary (Gaussian) increments, they are independent in the case of BM and dependent in the case of fBM.[4] fBM with $H > 0.5$ exhibits long memory. As a consequence, the self-affinity index H is in the long memory literature also known as the Hurst coefficient (see Beran et al. (2013), among others).

---

[4] A self-affine process $\{U(t), t \geq 0\}$ with stationary and independent increments is an α-stable Lévi process if the distribution of $U(1)$ is the α-stable described by the tail index $0 < \alpha \leq 2$. Thus, $H = 1/\alpha$ and for $\alpha = 2$, the α-stable distribution becomes the Gaussian distribution (Embrechts and Maejima, 2002).

Proposition 1.2.2. Let $\{U(t), t \geq 0\}$ be a unifractal process as given in Equation (1.1) with $H \in (0, 1)$ and with stationary increments. Then, for all $c \geq 0$ and for all $\Delta > 0$, it holds that

$$U(t + c\Delta) - U(t) \overset{d}{=} c^{H}[U(t + \Delta) - U(t)]. \tag{1.2}$$

The proof of Proposition 1.2.2 is given in Appendix 1.A. The proposition states that, if a process is unifractal and has stationary increments, then the increments are also unifractal following the same scaling law as the process itself.

The scaling relationship describing a unifractal process (given in Equation (1.1)) remains unchanged w.r.t. c: i.e., H does not depend on the choice of c. This might be very restrictive in many applications, including in the financial ones, as the relationships among variables scaled (aggregated) at different time scales (frequencies) may not be simple (unifractal), but complex as they might change with the scale or aggregation factor. Therefore, Mandelbrot (1974) introduces a more general concept, namely the *multifractality* that, differently from the unifractality, allows for the scaling relationships to change with the scaling factor c. The multifractality can be defined in two equivalent ways as it follows (Calvet and Fisher, 2002; Mandelbrot et al., 1997):

Definition 1.2.3. A real-valued process $\{M(t), t \geq 0\}$ is said to be *multifractal*,

1. if there exists a real-valued random variable $H(c)$ such that for all $c \geq 0$,

$$M(ct) \overset{d}{=} c^{H(c)} M(t), \tag{1.3}$$

where $M(t)$ and $H(c)$ are independent random functions and $H(c)$ is time independent *or*

2. if $M(t)$ has stationary increments and satisfies

$$E[|M(t)|^{q}] = c(q) t^{\tau(q)+1}, \quad \forall q \in \mathcal{Q} \quad \text{and} \quad \forall t \geq 0, \tag{1.4}$$

where $c(q)$ and $\tau(q)$ are deterministic functions of q, $\tau(q)$ is denoted the *scaling function*, $\mathcal{Q}$ is an interval on the real line with positive lengths such that $[0, 1] \subseteq \mathcal{Q}$.

While the first one refers to the variation of the scaling relationships with respect to the choice of the scaling factor c, the second definition states that a multifractal process has stationary increments and its moments exhibit multi-scaling laws, denoted here as *moment-scaling relationships*. $H(c)$ is known as the generalized Hurst component and it allows for different scaling relationships among the moments of $M(t)$ sampled at different timescales (frequencies). From Definition 1.2.3, one can see that unifractality (with stationary increments) is a special case of multifractality: if $M(t)$ is unifractal, then $H(c) = H, \forall c \geq 0$ and $M(t) \overset{d}{=} t^H M(1)$. As a consequence, $E[|M(t)|^q] = t^{Hq} E[|M(1)|^q]$: i.e., $\tau(q) = Hq - 1$ and $c(q) = E[|M(1)|^q]$. Similar to Proposition 1.2.2, one can easily show that the increments of a multifractal process are also multifractal and have the same moment-scaling relationship as the process itself (Mandelbrot et al., 1997).

Multifractal theory has so far found a wide application in many disciplines as it is able to describe complex systems characterized by long dynamic dependencies, occurrence of extreme events and multi-scaling relationships (Kantelhardt, 2009). The multifractal theory is thus able to address all these features within a unified mathematical framework (Kobeissi, 2013; Lux and Segnon, 2018). The fact that financial returns exhibit such empirical properties (volatility clustering and fat-tails), makes the multifractal framework an attractive one in describing the price process of the financial assets as shown by Mandelbrot et al. (1997) and Calvet and Fisher (2004).

While the concept of multifractality may seem at a first glance simple to understand, it is difficult to implement in practice, especially in the form defined in Equation (1.3), due to the difficulty of computing $H(c)$ for all choices of c. Most of the applications of multifractality in the finance literature (e.g., Hallam and Olmo (2014a), Calvet and Fisher (2004)) make use of the second definition that involves concepts (stationarity and moment-scaling relationships), which are easier to deal with in practice.

1.2.2. ScaVaR and ScaES

Before presenting the theoretical assumptions and results supporting our methodology, we first introduce the concept of intrinsic time. Assume $t \in [0, T]$ where $T > 0$. Let λ denote a real-valued intensity measure on $[0, T]$, which measures the market's activity at a certain point of time. From this intensity measure, we generate the time transformation $\theta : [0, T] \longrightarrow [0, T]$, which transforms the calendar time t into the *intrinsic time* $\theta(t)$:

$$\theta(t) = \lambda\left([0, t]\right). \tag{1.5}$$

$\theta(t)$ is an increasing function of the clock time t in accordance with the intensity measure λ, i.e., given the market's intensity λ between two time points $s, t \in [0, T]$, $s < t$, which are close to each other, the transformed time $\theta(t)$ elapses such that $\theta(t) - \theta(s) = \lambda\left([0, t]\right) - \lambda\left([0, s]\right) = \lambda\left([s, t]\right)$. Thus, the intrinsic time aims at capturing the real "heartbeat" of the market in accordance with its market intensity: it accelerates the time when the intensity of the market's activity is high and it slows the time down when the intensity is low. Choices of intensity markers could be: the number of trades, the number of price changes, trading volumes and intraday volatility patterns (see Clark (1973), Gouriéroux and Jasiak (2001), Oomen (2005), Dong and Tse (2014), Boudt et al. (2011), Mandelbrot et al. (1997), among others).

In what follows, we assume that the financial logarithm price process follows a unifractal process in the transformed time $\theta(t)$. More formally:

Assumption 1.2.4. Let $P(t)$ be the financial price process and $p(t) = \ln P(t)$.[5] We assume that:

$$p(t) = U(\theta(t)), \tag{1.6}$$

where $U(t)$ is a unifractal process with the Hurst coefficient H and stationary increments, $\theta(t)$ is defined above and it is independent of $U(t)$.

Assumption 1.2.4 is a very general one for financial logarithm price processes and includes as special cases: (1) the BM assumption when $H = 0.5$, $\theta(t) = t$ and the increments are independent and Gaussian, which is the most widespread assumption in the finance theory (Black and Scholes (1973), Markowitz (1952), etc.); (2) the fBM assumption when $H \neq 0.5$, $\theta(t) = t$ and the increments are Gaussian, but dependent, which is promoted by Comte and Renault (1998), Comte and Renault (1996) and Mandelbrot (1997), among others and (3) the assumption that the logarithm price process is a BM or fBM in $\theta(t)$ as in the multifractal model of asset returns (MMAR) of Mandelbrot et al. (1997) and Calvet and Fisher (2002). A major concern regarding all these previous assumptions is about the Gaussianity of the increments (or returns), which is empirically unrealistic for the current financial returns as they exhibit fat-tailed distributions. In contrast, Assumption 1.2.4 only assumes that $p(t)$ follows a unifractal law not in the clock time t, but in its transformation $\theta(t)$: i.e., $U(c\theta(t)) \overset{d}{=} c^H U(\theta(t))$. This allows for a simple derivation of the distribution and, for our purposes, of the quantiles of aggregated (daily) returns from their intraday counterparts sampled in intrinsic time by simply scaling them up by c^H.

In what follows we show that, by assuming that $\lambda(t)$ is a multifractal measure, the logarithm price process assumed in Equation (1.6) is multifractal, which emphasizes once again its very general character compared to the existing ones (BM or fBM), as the multifractality accommodates many important empirical stylized facts of financial returns, such as volatility clustering and fat-tailedness (Lux and Segnon, 2018). For this reason, we first define the concept of a multifractal measure.

---

[5] Notice that we normalize the logarithm price process by subtracting $p(0)$ from $p(t)$ as Definition 1.2.1 implies that for a unifractal process $U(t)$, it must hold that $U(0) = 0$ almost surely.

Definition 1.2.5. Let μ be a random measure on $[0, T]$. μ is a *multifractal measure* if

$$\mathbb{E}\left[\mu\left([t, t+\Delta]\right)^q\right] = c_\mu(q)(\Delta)^{\tau_\mu(q)+1}, \tag{1.7}$$

for all $t, \Delta \in [0, T]$ and all $q \in \mathcal{Q}$, where $c_\mu(q)$ and $\tau_\mu(q)$ are deterministic functions of q and $\mathcal{Q}$ is specified in Definition 1.2.3.

Thus, a multifractal measure is characterized by moment-scaling relationships similar to the ones given in Equation (1.4). In what follows, we assume that the intensity measure λ is a multifractal measure:

Assumption 1.2.6. The intensity measure λ defined on $[0, T]$ is a multifractal measure i.e., $\mathbb{E}\left[\lambda\left([t, t+\Delta]\right)^q\right] = c_\lambda(q)(\Delta)^{\tau_\lambda(q)+1}$, where $c_\lambda(q)$ and $\tau_\lambda(q)$ are deterministic functions of q.

Based on Assumption 1.2.6, we show now that $\theta(t)$ is a multifractal process as defined in Section 1.2.1:

Theorem 1.2.7. Let $\lambda(t)$ be a multifractal measure on $[0, T]$. If there exists $\varepsilon \in (0, 1)$ such that $(-\varepsilon, 1] \subseteq \mathcal{Q}$, then, $\theta(t) = \lambda\left([0, t]\right)$ is a multifractal process with continuous and non-decreasing paths.

The proof of the theorem is in Appendix 1.A. Based on Theorem 1.2.7, one can now show that the price process we assume in Assumption 1.2.4 is a multifractal one:

Theorem 1.2.8. The price process $p(t)$ given in Equation (1.6) is multifractal with the scaling function $\tau_p(q) = \tau_\lambda(Hq)$ and the scaling factor $c_p(q) = c_\lambda(Hq)\mathbb{E}\left[|U(1)|^q\right]$.

See Appendix 1.A for the proof. From Proposition 1.2.2, we get that the increments of the logarithm price process assumed in Equation (1.6) are also unifractal in the intrinsic time $\theta(t)$, and consequently, based on Theorem 1.2.8, also multifractal. In particular, by replacing t with $\theta(t)$ in Proposition 1.2.2, one can show that their distribution follows the same unifractal scaling law as the logarithm price itself:

Corollary 1.2.9. For all $c \geq 0$ and for all $\Delta > 0$, it holds that

$$U(\theta(t) + c\Delta) - U(\theta(t)) \overset{d}{=} c^H[U(\theta(t) + \Delta) - U(\theta(t))]. \tag{1.8}$$

From Corollary 1.2.9, one gets that the quantiles (VaR) and ES at probability level p of the increments of the logarithm price sampled in intrinsic time have the following scaling relationships:

$$Q_p\big(U(\theta(t) + c\Delta) - U(\theta(t))\big) \quad = c^H Q_p\big(U(\theta(t) + \Delta) - U(\theta(t))\big), \tag{1.9}$$

$$ES_p\big(U(\theta(t)+c\Delta)-U(\theta(t))\big) \quad = c^H ES_p\big(U(\theta(t)+\Delta)-U(\theta(t))\big), \tag{1.10}$$

for all $p \in (0,1)$, all $c \geq 0$ and for all $\Delta > 0$. Thus, equations (1.9) and (1.10) allow us to compute quantiles (VaR) and ES of increments sampled at the daily frequency by scaling up their counterparts computed from increments sampled at a higher frequency (e.g., 5 minute) and in intrinsic time. We denote these approaches by **Scaling-up VaR (ScaVaR)** and **Scaling-up ES (ScaES)**.

1.2.3. Implementation of ScaVaR and ScaES

Based on the theoretical assumptions and results presented in Section 1.2.2, we show here how to obtain daily estimates of quantiles (VaR) and ES from HF data based on ScaVaR and ScaES by implementing the following steps:

Step 1. Based on choices of the intensity measure λ, we generate the time transformation function $\theta(t) = \lambda([0,t])$ for all $t \in [0,T]$.[6] In order to sample $c+1$ HF logarithm prices (logarithm ticks) according to the transformed time $\theta(t)$, we fix $c \in \mathbb{N}$.[7] We define the equally spaced calendar time grid $\boldsymbol{t} = \{t_0, t_1, t_2, \ldots, t_c\}$, where $t_0 = 0$, $t_c = T$ and $t_i - t_{i-1}$ is constant for all $i = 1,\ldots,c$. The intrinsic time grid that samples $c+1$ prices according to the intrinsic time $\theta(t)$ is computed as:

$$\begin{aligned}
\tilde{\boldsymbol{t}} \;&:=\; \{\theta^{-1}(t), t \in \boldsymbol{t}\} \\
&=\; \{\theta^{-1}(t_0), \theta_d^{-1}(t_1), \ldots, \theta^{-1}(t_c)\} \\
&:=\; \{\tilde{t}_0, \tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_c\},
\end{aligned} \tag{1.11}$$

where we fix the first and the last observation, i.e., $\theta^{-1}(t_0) = \tilde{t}_0 = 0$ and $\theta^{-1}(t_c) = \tilde{t}_c = T$.[8] From the ticks sampled on the time grid $\tilde{\boldsymbol{t}}$, we compute the c intraday log-returns for each day d, denoted here by $\boldsymbol{r}_{f,d} = \{r_{1,d}, \ldots, r_{c,d}\}$ with $d = 1,\ldots \mathcal{D}$.

---

[6] We provide concrete examples on how to choose λ in the empirical exercise in Section 1.3.

[7] Alternatively, one can fix the frequency f of sampling the HF data, which is related to c in the following manner: $f = T/c$, where T is the total amount (measured in units of time: e.g., minutes) of calendar time per day (e.g., on NYSE $T = 390$ minutes). For example, on NYSE, if $c = 78$, then $f = 5$ minute. Note that the term "frequency" here does not necessarily have the classical calendar interpretation. E.g., a frequency of 5 minutes does not necessarily imply that we sample every 5 minutse, but the sampling points vary with the choice of the intensity measure such that, for instance on NYSE, one samples a total of 78 observations per day.

[8] A special case of this procedure is the Calendar Time Sampling (CTS) scheme that samples price observations that are equidistant in calendar time. Thus, the intensity measure λ is fixed and equal to 1 and $\tilde{\boldsymbol{t}} = \boldsymbol{t}$.

Step 2. From the intraday log-return series $\boldsymbol{r}_{f,d}$ of day d and for a given probability $p \in (0, 1)$, we compute the p-th quantiles and corresponding ES, denoted here by $\widehat{Q}_p(\boldsymbol{r}_{f,d})$ and $\widehat{ES}_p(\boldsymbol{r}_{f,d})$, respectively.

Step 3. We estimate the Hurst coefficient H from the intraday log-return series (see Kantelhardt (2009) for a survey of estimation methods of H).

Step 4. Based on equations (1.9) and (1.10), we compute the p-th quantiles and p-th ES of the daily returns on day d with $p \in (0, 1)$, denoted by $\widehat{Q}_{p,d}$ and $\widehat{ES}_{p,d}$, by scaling up their counterparts computed in Step 2 with the scale index H estimated in Step 3 as it follows:

$$\widehat{Q}_{p,d} = c^{\hat{H}}\widehat{Q}_p(\boldsymbol{r}_{f,d}), \tag{1.12}$$

$$\widehat{ES}_{p,d} = c^{\hat{H}}\widehat{ES}_p(\boldsymbol{r}_{f,d}). \tag{1.13}$$

1.3. Empirical Application

1.3.1. Data Description

In this section, we apply the ScaVaR and ScaES approaches to both, stock and foreign exchange rate data. In particular, we use the tick data of International Business Machines Corporation (IBM) and Bank of America (BAC) collected from TAQ database of NYSE from January 2, 2001 to July 24, 2017. We have a total of $\mathcal{D} = 4124$ trading days for each stock. The raw data consists of the best bid and best ask quotes during the trading period, which is from Monday to Friday from 9:30:00am to 16:00:00pm. From the best bid and best ask we compute the midquote as their average. The choice of the stocks aims at providing some diversification in what regards the liquidity of trading as well as the sector and the profile of the companies considered: IBM is a common choice in the related literature due to its high-liquidity, while BAC is from the financial sector and has experienced some serious losses, especially during the previous financial crisis.

The foreign exchange rate data we use is for the Euro against the US Dollar (EURUSD) from July 5, 2008 until July 20, 2016 ($\mathcal{D} = 2125$). The raw tick data is obtained from TICKDATA[9] that collects foreign exchange rate transactions from multiple data sources covering a variety of market places around the world. The foreign exchange rates are traded

---

[9] https://www.tickdata.com/

around the clock from Sunday 6pm ET until Friday 5:59:59.999pm ET[10], adding up to a total of five days a week with 24 hours of trading activity.

According to Step 1 presented in Section 1.2.3, the tick data is sampled at the frequency $f = 5$ minute leading to $c = 288$ for the foreign exchange rates and $c = 78$ for the stocks. The choice of this frequency attempts to accommodate the fact that intraday returns are contaminated with market microstructure noise (MMN) (see Andersen, Bollerslev, Diebold, and Ebens (2001) for a similar procedure when estimating daily volatilities by means of Realized Volatility measures).[11] Below we shortly describe another attempt to account for MMN when estimating the quantiles of intraday returns. However, a theoretical and empirically deeper analysis of the impact of the MMN noise on our estimates is left for further research.

For our purposes, we implement two sampling schemes, namely (1) the Calendar Time Sampling (CTS) scheme that samples price observations equidistantly in clock time and (2) the Time Transformation Sampling (TTS) scheme that is an intrinsic time scheme that samples price observations that are equidistant in terms of the number of ticks averaged over all past trading days. In this case, the intensity measure $\lambda(t)$ is given by the number of ticks at time t averaged over all days prior to and including the present day. This sampling scheme accounts for the intraday periodicity in the trading activity as a common pattern across all trading days (Wu, 2012).[12]

Descriptive statistics of HF returns sampled by the two sampling schemes are given in Table 1.3 in Appendix 1.B.[13] As one may observe from the table, the magnitude of the returns is very small, but their kurtosis is very large and much above the value of 3. However, the value of the kurtosis reduces from the calendar time to the intrinsic one.[14] This indicates that the intrinsic time sampling scheme leads to a reduction in the sampled extreme values compared to CTS as it picks more price observations during highly active periods (with high fluctuations) and less during calm one (with low fluctuations). In contrast, CTS samples the data independently of the market's activity.

---

[10] Before August 2, 2014, the trading data provided by TICKDATA covers trades from Sunday 4pm ET until Friday 3:59:59.999pm ET every week.

[11] Results from implementing other frequencies, such as 1 minute, 3 minute, 10 minute and 30 minute can be obtained from the authors upon request.

[12] Implementing further intrinsic sampling schemes based on other types of intensity measures, such as non-zero changes in the prices, volumes or volatility patterns (Clark (1973), Engle and Russell (1998), Griffin and Oomen (2008), Oomen (2005), Oomen (2006), Boudt et al. (2011), Dong and Tse (2014)) provides, in general, similar results to TTS when compared to CTS and the alternatives described below. They can be obtained from the authors upon request.

[13] Similar results for the other sampling schemes can be obtained from the authors upon request.

[14] Similar empirical findings are documented by Clark (1973) and Mandelbrot and Taylor (1967).

1.3.2. Descriptive Statistics of ScaVaR and ScaES Estimates

In what follows, we compute the VaR and ES at $p = 1\%, 2.5\%$ and 5%: $p = 1\%$ is the probability specified by the Basel Committee to compute VaR (Basel Committee, 1996), $p = 2.5\%$ is the probability to compute ES according to Basel Committee (2016) and $p = 5\%$ is the probability that is most popular among researches and practitioners (e.g., Kuester et al. (2006), among others).

In order to implement Step 2 presented in Section 1.2.3, we estimate the quantiles of the intraday returns by means of empirical quantiles by employing the MATLAB function $'quantile(x)'$. The ES estimator is computed as the average of the observations smaller than the empirical quantile estimator. The empirical quantile and ES estimators are little precise in small samples (see Harrell and Davis (1982), among others), which is also our case, as the number of intraday observations is limited. One way of improving the quality of these estimators is to increase the number of intraday returns, namely c. This leads to increasing the frequency of sampling the HF information, which may induce a bias in the estimates due to the MMN present in the intraday prices (due to bid-ask spreads, discrete price changes and asymmetric market information, among others). This issue is treated at length in the Realized Volatility literature that finds that sampling returns at frequencies between 5 and 30 minute provide an optimal trade-off between accuracy and efficiency of the Realized Volatility estimators (for an overview of the literature see Hansen and Lunde (2006)).

A very popular way of dealing with the MMN effects in the Realized Volatility literature is to use the subsampling estimator of Zhang et al. (2005), which averages over the Realized Volatilities computed from observations sampled in such a way that the sampling starting point is shifted by some seconds. As we are not aware of any similar technique to compute quantiles of intraday data, we follow the procedure of Zhang et al. (2005) and compute them by averaging (or taking the median) over the empirical intraday quantiles computed from returns sampled with shifted starting points. Thus, the new quantile estimator accounts for more intraday information and deals with MMN. In our empirical exercise, this type of subsampling quantile estimator improves in general the in-sample and out-of-sample results of ScaVaR and ScaES compared to the standard empirical quantiles.[15] This is a promising outcome that one should explore in more detail in the future research.

To implement Step 3 in Section 1.2.3, we estimate the Hurst coefficient by means of the detrended moving average (DMA) estimator of Alessio et al. (2002). We also implement further estimators, such as the Detrended Fluctuation Analysis (DFA) of Peng et al. (1994) and the estimators of Sánchez-Granero et al. (2012). However, given that the in-sample and

---

[15] They can be obtained from the authors upon request.

out-of-sample results do not change significantly by changing the estimation method of H, we stick here to the results computed from implementing DMA.

Based on the estimate of H and of the intraday empirical quantiles and ES, we are now able to implement Step 4 presented in Section 1.2.3 in order to compute the daily scaled-up estimates of VaR and ES as given in equations (1.12) and (1.13). Tables 1.4 and 1.5 in Appendix 1.B provide descriptive statistics of the logarithm transformations of their negative values namely of $\hat{q}_{p,d} := \ln(-\widehat{Q}_{p,d})$ and $\widehat{es}_{p,d} := \ln(-\widehat{ES}_{p,d})$. From both tables, we observe that the $\widehat{q}_{p,d}$ and $\widehat{es}_{p,d}$ exhibit overkurtosis and skewness and one rejects the normal distribution assumption in all cases. In general, there is a tendency in reducing the kurtosis and JB-test values from CTS to TTS, which may indicate a reduction in the extreme values in the estimated VaR and ES series. However, the most evident result is that exchange rates exhibit lower average kurtosis and JB-test values than the stocks, with BAC exhibiting the largest ones. This can be explained by the fact that the stock markets are more risky and that the financial crises, which is included in the sample, affects the stock market more heavily, especially the stocks from the financial sector.

A graphical inspection of the series in Figures 1.1 and 1.2 confirms these findings. The estimates of the stocks are much more volatile, with higher volatility clustering and higher peaks than the ones of the exchange rates, with the mostly pronounced ones for BAC. However, the clustering effects decrease by increasing p. The ACF of the series, depicted in Figures 1.3 and 1.4, indicate that the transformed VaR and ES estimates exhibit slowly decaying ACF, which is typical for long-memory, however with different decaying patterns: the fastest for the exchange rate, followed by the ones of the stocks, with BAC exhibiting the slowest decaying ACF pattern.

The histograms and the QQ-plots given in Figures 1.5-1.16 show that the estimates for the exchange rates are closer to the normal distribution than the ones of the stocks as they exhibit less extreme values, which is in line with the results presented in Tables 1.4 and 1.5. Among the stocks, BAC seems to have the most pronounced tails in VaR and ES estimates, which indicates that this asset is particularly risky compared to the other two.

1.3.3. In-Sample Results

For the assessment of the empirical performance of our new methods to accurately estimate daily VaR and ES, we compare them against (1) a standard location-scale approach with constant conditional mean, the conditional variance stemming from a GARCH(1,1) specification and the Student's t with estimated degrees of freedom for the distribution of the standardized residuals of daily returns (we denote this specification the GARCH-t) and (2) the multiscaling approach of Hallam and Olmo (2014a) (denoted here HOMF) that estimates the whole pdf of daily returns from matching the scaled up moments of intraday returns sampled in CTS to the moments derived from a parametric distributional assumption for daily returns. From these estimated pdf's, we compute the quantiles and ES that we need for our comparison. While the GARCH-t approach is a standard choice within the location-scale framework to estimate and forecast VaR (see Halbleib and Pohlmeier (2012), among others), HOMF may be regarded as being the closest to our methods, as it develops from the assumption that the logarithm price process is multifractal. However, different from our approach, it uses the moment-multiscaling relationships typical to multifractality applied on data sampled in the clock time, ignoring, thus, any type of market activity information.[16]

To evaluate the performance of the models in accurately estimating VaR and ES, we use strictly consistent scoring (loss) functions as described by Gneiting (2011). These scoring functions allow for a reliable ranking of the forecasts, assuring that asymptotically, the forecast with the smallest loss value is the best.

Thus, we employ the strictly consistent *asymmetric piece-wise linear scoring function* (sometimes also called *check loss* function or the *tick loss* function), which is the most popular scoring function for quantiles according to Gneiting (2011) and which is given by

$$S_p^{VaR}\big(\widehat{Q}_{p,d}, r_d\big) = (r_d - \widehat{Q}_{p,d})(p - \mathbb{1}_{\{\widehat{Q}_{p,d} \geq r_d\}}). \tag{1.14}$$

As there exists no strictly consistent scoring function for ES alone, we choose to use a joint scoring function for the pair VaR and ES as proposed by Fissler and Ziegel (2016), namely:

$$S_p^{VaRES}\big(\widehat{Q}_{p,d}, \widehat{ES}_{p,d}, r_d\big) = \frac{1}{\widehat{ES}_{p,d}}\left(\widehat{ES}_{p,d} - \widehat{Q}_{p,d} + \frac{(\widehat{Q}_{p,d} - r_d)\mathbb{1}_{\{r_d \leq \widehat{Q}_{p,d}\}}}{p}\right) + \ln(-\widehat{ES}_{p,d}). \tag{1.15}$$

Thus, over the whole evaluation window composed of $\mathcal{D}$ days, we compute the final scoring functions as the average over all days:

$$\bar{S}_p^{VaR} = \frac{1}{\mathcal{D}} \sum_{d=1}^{\mathcal{D}} S_p^{VaR}\big(\widehat{Q}_{p,d}, r_d\big), \tag{1.16}$$

---

[16] For comparison reasons, we also implement a location-scale model where the conditional variance stems from RV specifications instead of GARCH, as well as the unifractal-based method of Hallam and Olmo (2014b) that estimates pdf's of daily returns by scaling up the pdf's of intraday returns sampled in CTS. However, given that the results for these methods are not significantly different from the ones of the comparative models mentioned above, we choose to disregard them for the present analysis.

Table 1.1: In-sample results: VaR and VaRES scores. The entries in bold correspond to the smallest score values. The entries in parentheses give the p-values of the DM test with the ScaVaR/ScaES computed from data sampled in TTS as the benchmark. The VaR score values are scaled by 10^4.

Model	1%		2.5%		5%	
	VaR	VaRES	VaR	VaRES	VaR	VaRES
IBM						
ScaVaR/ScaES-TTS	**3.4056**	-3.5346	**6.4961**	**-3.7759**	**10.4120**	**-3.9944**
	(-)	(-)	(-)	(-)	(-)	(-)
ScaVaR/ScaES-CTS	3.4059	**-3.5356**	6.5118	-3.7754	10.4510	-3.9912
	(0.9841)	(0.8291)	(0.5618)	(0.8886)	(0.4599)	(0.3753)
HOMF	3.7214	-3.2112	7.3741	-3.5301	12.2920	-3.7559
	(0.0624)	(0.0004)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
GARCH-t	3.7186	-3.3284	7.4421	-3.5654	12.3910	-3.7676
	(0.0990)	(0.0024)	(0.0006)	(0.0000)	(0.0000)	(0.0000)
BAC						
ScaVaR/ScaES-TTS	**5.1814**	**-3.2071**	**9.7204**	**-3.4406**	15.417	-3.6576
	(-)	(-)	(-)	(-)	(-)	(-)
ScaVaR/ScaES-CTS	5.2077	-3.2021	9.7367	-3.4438	**15.451**	**-3.66**
	(0.2714)	(0.1493)	(0.8397)	(0.5134)	(0.7922)	(0.6010)
HOMF	6.8742	-2.6057	13.1520	-3.0221	21.2900	-3.2861
	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GARCH-t	6.4913	-2.8521	13.183	-3.1053	21.592	-3.3196
	(0.0025)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
EURUSD						
ScaVaR/ScaES-TTS	**1.8739**	**-4.0261**	**3.5462**	**-4.2857**	5.8535	**-4.4889**
	(-)	(-)	(-)	(-)	(-)	(-)
ScaVaR/ScaES-CTS	1.9234	-3.9945	3.5713	-4.2712	**5.8474**	-4.4865
	(0.0000)	(0.0000)	(0.3474)	(0.0003)	(0.8938)	(0.5620)
HOMF	2.0127	-3.8267	4.0978	-4.0852	6.9601	-4.2643
	(0.2812)	(0.1089)	(0.0011)	(0.0035)	(0.0000)	(0.0000)
GARCH-t	1.9607	-3.9002	4.0799	-4.1044	6.991	-4.2682
	(0.4922)	(0.2077)	(0.0029)	(0.0037)	(0.0000)	(0.0000)

$$\bar{S}_p^{VaRES} = \frac{1}{\mathcal{D}} \sum_{d=1}^{\mathcal{D}} S_p^{VaRES}\left(\widehat{Q}_{p,d}, \widehat{ES}_{p,d}, r_d\right), \tag{1.17}$$

whereas the model with the smallest average score exhibits the best performance. In order to test for differences in the score functions among the models, we implement the Diebold-Mariano (DM) test originally developed by Diebold and Mariano (1995).

Table 1.1 provides the VaR and VaRES scores and the results from the DM test with the ScaVaR/ScaES computed from data sampled in TTS as the benchmark. As one may see from the table, the ScaVaR/ScaES provides the smallest score values for all p's and assets. The TTS provides in general the most improvements compared to CTS.

For the stocks, the DM-test cannot reject that the quality of the estimates between the two sampling schemes is the same. However, for the exchange rate compared to the stocks, it seems that TTS provides more improvements in terms of score values compared to CTS, especially when estimating extreme risks ($p = 1\%$). This may be due to the fact that the intraday trading pattern may not be as informative for stocks as it is for the exchange rates. To improve the results for stocks, one should consider other types of intrinsic time sampling schemes, built on intensity measures, which may be more informative for stocks, such as, for instance, the intraday volatility pattern.[17]

Among the alternatives, the HOMF performs worst in terms of the score values, which indicates that estimating the whole pdf by scaling up intraday moments and using data sampled in CTS seems not to be appropriate for tail risk estimation.

1.3.4. Out-of-Sample Results

We compute one-step ahead forecasts of daily VaR and ES by using the rolling-window technique. The in-sample and out-of-sample windows are chosen such that they have the same lengths: i.e., for the foreign exchange rates, the out-of-sample window contains the last $\mathcal{D}^* = 1062$ days (for the period from May 7, 2008 until July 20, 2016) and for the stock data, the out-of-sample window contains the last $\mathcal{D}^* = 2062$ days (for the period from January 2, 2001 to July 24, 2017).

In order to make forecasts based on the daily estimates of VaR and ES stemming from ScaVaR and ScaES, we apply the HAR model of Corsi (2009) on the log-transformed series $\hat{q}_{p,d} := \log(-\widehat{Q}_{p,d})$ and $\widehat{es}_{p,d} := \log(-\widehat{ES}_{p,d})$ with $d = 1, \ldots, \mathcal{D} - \mathcal{D}^* + 1$ in order to account for their long persistence depicted in Figures 1.3 and 1.4.

Thus, the one-step ahead forecasts $\widehat{q}_{p,d+1|d}$ and $\widehat{es}_{p,d+1|d}$ are obtained in the following manner[18]:

$$\hat{q}_{p,d+1|d} = \hat{\beta}^{(0)} + \beta^{(1)}\hat{q}_{p,d} + \hat{\beta}^{(5)}\hat{q}^{(5)}_{p,d} + \hat{\beta}^{(22)}\hat{q}^{(22)}_{p,d}, \tag{1.18}$$

$$\widehat{es}_{p,d+1|d} = \hat{\gamma}^{(0)} + \hat{\gamma}^{(1)}\widehat{es}_{p,d} + \hat{\gamma}^{(5)}\widehat{es}^{(5)}_{p,d} + \hat{\gamma}^{(22)}\widehat{es}^{(22)}_{p,d}, \tag{1.19}$$

---

[17] It is a well-known fact that, during the trading period, stocks exhibit a pronounced "U"-shape in the intraday volatility pattern (see Harris (1986) and Andersen and Bollerslev (1997), among others). For this reason, in a preliminary attempt, besides further trading-based sampling schemes following Engle and Russell (1998), Griffin and Oomen (2008), Oomen (2005) Oomen (2006), we also implement some sampling schemes based on the intraday volatility pattern as described by Boudt et al. (2011) and Dong and Tse (2014). Our preliminary results show the volatility-based sampling scheme provides better score values for estimating and forecasting the VaR and ES of stocks compared to the exchange rate. A deeper empirical analysis of which type of sampling scheme is more valuable for which type of asset is left, however, for further research.

[18] Alternatively, one could consider other specifications such as Autoregressive Fractional Integrated Moving Average (ARFIMA) model as applied by Andersen, Bollerslev, Diebold, and Labys (2001) or MIDAS approach of Ghysels et al. (2006) to capture the long persistence dynamics similarly to Realized Volatilities.

where $\hat{q}^{(m)}_{p,d} = \frac{1}{m}\sum_{j=1}^{m}\hat{q}_{p,d-j+1}$ and $\widehat{es}^{(m)}_{p,d} = \frac{1}{m}\sum_{j=1}^{m}\widehat{es}_{p,d-j+1}$ are the averages over the past $m = 1, 5, 22$ days. The one-step ahead forecasts of $\widehat{Q}_{p,d+1|d}$ and $\widehat{ES}_{p,d+1|d}$ are obtained by applying the transformation[19]:

$$\widehat{Q}_{p,d+1|d} = -\exp\left(\hat{q}_{p,d+1|d}\right), \tag{1.20}$$

$$\widehat{ES}_{p,d+1|d} = -\exp\left(\widehat{es}_{p,d+1|d}\right). \tag{1.21}$$

The forecast evaluation is done similarly to the ones from the in-sample analysis, i.e., by implementing the two score functions defined in equations (1.16) and (1.17) along with the DM test with the ScaVaR and ScaES with TTS as a benchmark. Table 1.2 provides the out-of-sample results. As one may see from the table, our models perform very good compared to the two alternatives: in almost all cases, they provide the smallest score values. Between the two sampling schemes, it seems that the intrinsic one (TTS) provides the best results, especially for IBM. Whenever CTS delivers the smallest score values, the DM test cannot reject that they are equal to the ones stemming from TTS. The very good performance of our approaches is regardless of the choice of p. Among the alternatives, GARCH-t provides the smallest scores for all p's. In one case, GARCH-t has a score value smaller than our ScaVaR/ScaES model based on TTS, however the DM test does not reject the null hypothesis that the models have equal forecasting power.

1.4. Conclusions

We provide a novel approach of estimating and forecasting daily VaR and ES directly from HF data. Our method assumes that financial logarithm price processes are unifractal (which is driven by a single scaling law) in an alternative time dimension that we denote the intrinsic time. The intrinsic time transforms the clock time according to an intensity measure that captures the real "heartbeat" of the market's activity: it allows to sample tick prices more often when the market's activity is intense and less often when it is calm. We show theoretically that this assumption translates in assuming that logarithm price processes are multifractal, which is a very general and realistic assumption compared to the well-spread existent ones, such as the Brownian Motion or the fractional Brownian Motion: i.e., it accommodates important empirical features of financial returns, such as volatility clustering and fat-tailedness.

---

[19] Note that, due to the transformation, the forecasts are biased. A bias correction as the one proposed by Bianchi and Calzolari (1980) and Oomen (2001) adds, however, further estimation noise to the forecasts (see e.g., Halbleib and Voev (2011)) and, therefore, for the moment, we decide to ignore it in our empirical exercise.

Table 1.2: Forecasting results: VaR and VaRES scores. The entries in bold correspond to the smallest score values. The entries in parentheses give the p-values of the DM test with the ScaVaR/ScaES computed from data sampled in TTS as the benchmark. The VaR score values are scaled by 10^4.

Model	1%		2.5%		5%	
	VaR	VaRES	VaR	VaRES	VaR	VaRES
IBM						
ScaVaR/ScaES-TTS	**2.8612**	**-3.5967**	**5.6670**	**-3.8291**	**9.4745**	**-4.0152**
	(-)	(-)	(-)	(-)	(-)	(-)
ScaVaR/ScaES-CTS	2.8819	-3.5890	5.6899	-3.8235	9.4769	-4.0128
	(0.0216)	(0.0001)	(0.1461)	(0.0212)	(0.8821)	(0.1882)
HOMF	3.1472	-3.2739	6.0910	-3.6506	9.9882	-3.9038
	(0.1825)	(0.0280)	(0.0677)	(0.0162)	(0.0270)	(0.0112)
GARCH-t	2.9129	-3.5381	5.9644	-3.7435	9.8456	-3.9478
	(0.7721)	(0.4937)	(0.2068)	(0.1276)	(0.1576)	(0.0871)
BAC						
ScaVaR/ScaES-TTS	5.8081	-2.9161	**11.8621**	**-3.1344**	19.4769	**-3.3272**
	(-)	(-)	(-)	(-)	(-)	(-)
ScaVaR/ScaES-CTS	**5.7742**	**-2.9201**	11.9429	-3.1298	**19.4757**	-3.3252
	(0.2834)	(0.2633)	(0.9154)	(0.9082)	(0.9014)	(0.4988)
HOMF	6.9563	-2.4055	12.9338	-2.8758	20.7714	-3.1553
	(0.0251)	(0.0026)	(0.0196)	(0.0010)	(0.0009)	(0.0001)
GARCH-t	6.3756	-2.7287	12.5178	-3.0156	20.5219	-3.2265
	(0.1701)	(0.0888)	(0.0973)	(0.0352)	(0.0046)	(0.0045)
EURUSD						
ScaVaR/ScaES-TTS	1.7749	**-4.0216**	**3.3912**	**-4.3099**	5.6807	-4.5006
	(-)	(-)	(-)	(-)	(-)	(-)
ScaVaR/ScaES-CTS	1.8141	-3.9928	3.4053	-4.2956	**5.6523**	**-4.5015**
	(0.0134)	(0.0746)	(0.1146)	(0.0326)	(0.5552)	(0.1819)
HOMF	1.8317	-3.8836	3.5365	-4.2225	5.9195	-4.4249
	(0.5706)	(0.3541)	(0.1911)	(0.2282)	(0.0193)	(0.0525)
GARCH-t	**1.7657**	-3.9625	3.4905	-4.2533	5.8475	-4.4479
	(0.9655)	(0.6083)	(0.3270)	(0.3666)	(0.0737)	(0.1167)

In practice, our method is very simple to implement and consists in computing daily estimates of VaR and ES by scaling up their intraday empirical counterparts computed from returns sampled in intrinsic time. In our empirical exercise, we describe the statistical and dynamic properties of these estimates and show that our method outperforms the standard ones in accurately estimating and forecasting daily VaR and ES. In particular, the empirical results show that data sampled in intrinsic time in accordance with the intraday trading pattern is very valuable compared to the calendar one when estimating and forecasting extreme risks.

These results are very promising and of great importance in the light of the current challenges on forecasting financial risks, but also in the light of the huge amount of high frequency data (recorded in nanoseconds) that has become available lately. They open a new agenda of research on how to exploit at best the richness of the information content of high-frequency data from another time perspective for risk estimation and forecasting, but also for other financial purposes.

The multifractal theory of B. Mandelbrot has in general received so far very little attention from the economics and finance community with only a few people researching in this direction. We provide here new opportunities, but also new challenges to exploit it for a better understanding and handling of financial markets. In general, it would be interesting to research which are the implications of our multifractal assumption for the logarithm price process on its stochastic properties, on the (infill) asymptotics of related measures or on the statistical inference. In particular, from the current analysis, various immediate questions arise, such as: How one should deal with the market microstructure noise in our framework? How well the multifractal assumption can accommodate the empirical features of the real prices? Which intensity measure (such as based on volatility patterns, volumes, price changes, transaction patterns) and resulting intrinsic time-based sampling scheme are mostly useful to forecast financial risks? Which are the dynamic and statistical properties of the data sampled in intrinsic time and of the measures (volatilities, quantiles, etc) computed from it? All these questions await further research and aim at providing a new understanding from another (time) perspective on how the financial markets work.

Appendix 1.A Proofs

Proof of Proposition 1.2.2. Let $t' = t/c$. Then, for all $t \geq 0$ and for all $\Delta > 0$, it holds that

$$U(t + c\Delta) - U(t) = U\big(c(\tilde{t} + \Delta)\big) - U(c\tilde{t}). \tag{1.22}$$

From the unifractal assumption of U, we get that

$$U\big(c(\tilde{t} + \Delta)\big) - U(c\tilde{t}) \overset{d}{=} c^H \left(U\big(\tilde{t} + \Delta\big) - U(\tilde{t})\right). \tag{1.23}$$

Given that we assume that the increments of U are stationary, we get that

$$c^H \left(U\big(\tilde{t} + \Delta\big) - U(\tilde{t})\right) \overset{d}{=} c^H \left(U\big(t + \Delta\big) - U(t)\right), \tag{1.24}$$

which completes the proof. ☐

Proof of Theorem 1.2.7. Based on the definition of θ in Equation (1.5), we get that, for all $t \in [0, T]$ and for all $h \geq 0$,

$$\theta(t + h) - \theta(t) = \lambda\left([0, t + h]\right) - \lambda\left([0, t]\right) = \lambda\left([t, t + h]\right). \tag{1.25}$$

As $\lambda\left([t, t + h]\right) \geq 0$, then $\theta(t + h) \geq \theta(t)\ \forall t, \forall h$ almost surely, i.e., $\theta(t)$ has almost surely non-decreasing paths.

In order to show that $\theta(t)$ is a multifractal process, we make use of the second definition of the multifractality in Section 1.2.1 and show that (a) it has moment-scaling relationships as the ones given in Equation (1.4) and (b) it has stationary increments.

a) From Assumption 1.2.6 we get that for all $q \in \mathcal{Q}$ and for all $t \geq 0$,

$$\mathbb{E}\left[\theta(t)^q\right] = \mathbb{E}\left[\lambda\left([0, t]\right)^q\right] = c_\lambda(q) t^{\tau_\lambda(q)+1}, \tag{1.26}$$

where $\mathcal{Q}$ is some interval on the real line and there exists $\varepsilon > 0$ such that $(-\varepsilon, 1] \in \mathcal{Q}$ and $c_\lambda(q)$ and $\tau_\lambda(q)$ are inherited from the measure λ.

b) From Equation (1.25), we get that

$$\theta(t + h) - \theta(t) = \lambda\left([t, t + h]\right) \quad \text{and} \tag{1.27}$$

$$\theta(s + h) - \theta(s) = \lambda\left([s, s + h]\right), \tag{1.28}$$

for all $s, t \in [0, T]$.

Define $\eta[s,t] := \log(\lambda[s,t])$ and let $M_{\eta[t,t+h]}(q)$ be its moment generating function. Then, for all $q \in (-\varepsilon, \varepsilon)$, it holds that

$$M_{\eta[t,t+h]}(q) = \mathbb{E}\left[\exp(q \cdot \eta[t, t+h])\right] = \mathbb{E}\left[\exp(q \cdot \log(\lambda[t, t+h]))\right] \tag{1.29}$$

$$= \mathbb{E}\left[\lambda[t, t+h]^q\right] = c_\lambda(q) h^{\tau_\lambda(q)+1}, \tag{1.30}$$

and analogously

$$M_{\eta[s,s+h]}(q) = \mathbb{E}\left[\lambda[s, s+h]^q\right] = c_\lambda(q) h^{\tau_\lambda(q)+1}, \tag{1.31}$$

i.e., $M_{\eta[t,t+h]}(q) = M_{\eta[s,s+h]}(q)$ for all $q \in (-\varepsilon, \varepsilon)$. As the moment generating function on any interval $(-\varepsilon, \varepsilon)$ uniquely charactrizes the distribution of the underlying variable, we get that the distributions of $\eta[s, s+h]$ and $\eta[t, t+h]$ are identical for all $s, t \in [0, T]$ and for all $h > 0$. As the logarithm is a strictly increasing function, we also get that the distributions of $\lambda[s, s+h]$ and $\lambda[t, t+h]$ are identical for all $s, t \in [0, T]$ and for all $h > 0$, which implies that the process $\theta(t)$ defined in Equation (1.5) has stationary increments.

□

Proof of Theorem 1.2.8. For all $t \geq 0$, it holds that

$$\mathbb{E}\left[\,|p(t)|^q \,|\, \theta(t) = u\right] = \mathbb{E}\left[\,|U(u)|^q \,|\, \theta(t) = u\right]. \tag{1.32}$$

For the unifractal process U, it holds that $U(u) \overset{d}{=} u^H U(1)$ and therefore

$$\mathbb{E}\left[\,|U(u)|^q \,|\, \theta(t) = u\right] = \mathbb{E}\left[\,|u^H U(1)|^q \,\middle|\, \theta(t) = u\right] = \theta(t)^{Hq}\, \mathbb{E}[|U(1)|^q], \tag{1.33}$$

as $U(t)$ and $\theta(t)$ are independent. By the law of iterated expectations, we get that

$$\mathbb{E}\left[|p(t)|^q\right] = \mathbb{E}\left[\mathbb{E}\left[\,|p(t)|^q \,|\, \theta(t) = u\right]\right] \tag{1.34}$$

$$= \mathbb{E}\left[\theta(t)^{Hq} \mathbb{E}\left[|U(1)|^q\right]\right] \tag{1.35}$$

$$= \mathbb{E}\left[\theta(t)^{Hq}\right] \mathbb{E}\left[|U(1)|^q\right]. \tag{1.36}$$

From the Theorem 1.2.7, we know that $\theta(t)$ is a multifractal process, i.e. $\mathbb{E}\left[\theta(t)^{Hq}\right] = c_\lambda(Hq)\, t^{\tau_\lambda(Hq)+1}$ and thus

$$\mathbb{E}\left[\theta(t)^{Hq}\right] \mathbb{E}\left[|U(1)|^q\right] = c_\theta(Hq) \mathbb{E}\left[|U(1)|^q\right] \cdot t^{\tau_\lambda(Hq)+1}, \tag{1.37}$$

which implies that the process $p(t)$ satisfies the multi-scaling law typical to multifractality with $\tau_p(q) = \tau_\lambda(Hq)$ and $c_p(q) = c_\lambda(Hq)\mathbb{E}\left[|U(1)|^q\right]$.

In the following, we show that the process $p(t) = U\big(\theta(t)\big)$ has stationary increments: i.e., for all $s, t \geq 0$, $t > s$ and for all $h > 0$, it holds that

$$U\big(\theta(t+h)\big) - U\big(\theta(s+h)\big) \overset{d}{=} U\big(\theta(t)\big) - U\big(\theta(s)\big). \tag{1.38}$$

In order to prove Equation (1.38), we show that

$$\begin{align}
U\big(\theta(t+h)\big) - U\big(\theta(s+h)\big) &\overset{d}{=} U\big(\theta(t+h) - \theta(s+h)\big) - U(0) \tag{1.39}\\
&\overset{d}{=} U\big(\theta(t) - \theta(s)\big) - Y(0) \tag{1.40}\\
&\overset{d}{=} u\big(\theta(t)\big) - U\big(\theta(s)\big). \tag{1.41}
\end{align}$$

Formally, the equality in distribution in (1.39) can be shown by conditioning on $\theta(s+h)$. For that, we assume that the real-valued stochastic process $p(t)$ is equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ and is defined on some complete probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$. Let the measure of $\theta(s+h)$ be denoted by $\nu_{\theta(s+h)}$. Then, for all $s, t \geq 0$, $t > s$, for all $h > 0$ and for all $A \in \mathcal{B}(\mathbb{R})$, it holds that

$$\begin{align}
&\mathbb{P}\left(U\big(\theta(t+h)\big) - U\big(\theta(s+h)\big) \in A\right) \nonumber\\
=& \int_{\mathbb{R}_+} \mathbb{P}\left(U\big(\theta(t+h)\big) - U\big(\delta\big) \in A \,\middle|\, \theta(s+h) = \delta\right) \mathrm{d}\nu_{\theta(s+h)}(\delta) \tag{1.42}\\
=& \int_{\mathbb{R}_+} \mathbb{P}\left(U\big(\theta(t+h) - \delta\big) - U\big(0\big) \in A \,\middle|\, \theta(s+h) = \delta\right) \mathrm{d}\nu_{\theta(s+h)}(\delta) \tag{1.43}\\
=& \, \mathbb{P}\left(U\big(\theta(t+h) - \theta(s+h)\big) - U\big(0\big) \in A\right), \tag{1.44}
\end{align}$$

where we use in (1.43) the fact that the process $U(t)$ has stationary increments. The third equality in (1.41) can be shown equivalently by conditioning on $\theta(s)$.

For the equality in (1.40), it holds that for all $s, t \in \mathbb{R}_+$, $t > s$, for all $h > 0$ and for all $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{align}
&\mathbb{P}\left(U\big(\theta(t+h) - \theta(s+h)\big) - U\big(0\big) \in A\right) \nonumber\\
=& \int_{\mathbb{R}_+} \mathbb{P}\left(U\big(\eta\big) - U\big(0\big) \in A \,\middle|\, \theta(t+h) - \theta(s+h) = \eta\right) \mathrm{d}\nu_{\theta(t+h)-\theta(s+h)}(\eta) \tag{1.45}\\
=& \int_{\mathbb{R}_+} \mathbb{P}\left(U\big(\eta\big) - U\big(0\big) \in A\right) \mathrm{d}\nu_{\theta(t)-\theta(s)}(\eta) \tag{1.46}\\
=& \int_{\mathbb{R}_+} \mathbb{P}\left(U\big(\eta\big) - U\big(0\big) \in A \,\middle|\, \theta(t) - \theta(s) = \eta\right) \mathrm{d}\nu_{\theta(t)-\theta(s)}(\eta) \tag{1.47}
\end{align}$$

$$= \mathbb{P}\left(U\big(\theta(t) - \theta(s)\big) - U\big(0\big) \in A\right), \tag{1.48}$$

where we use in (1.46) the fact that the process $\theta(t)$ has stationary increments and it is independent of $U(t)$ and in (1.47) the fact that the processes $U(t)$ and $\theta(t)$ are independent. □

Appendix 1.B Tables

Table 1.3: Descriptive statistics of the intraday returns for the foreign exchange rates over the window May 7, 2008 - July 20, 2016 and for the stocks over the window January 2, 2001-July 24, 2017 computed at 5 minute frequency and sampled by two sampling schemes.

	EURUSD		IBM		BAC	
	CTS	TTS	CTS	TTS	CTS	TTS
Mean	-5.49E-07	-5.49E-07	9.54E-06	9.54E-06	-1.10E-05	-1.10E-05
Median	0.0000	0.000	0.0000	0.0000	0.0000	0.0000
Maximum	0.0150	0.0137	0.0759	0.0759	0.0684	0.0765
Minimim	-0.0117	-0.0121	-0.0380	-0.0380	-0.0929	-0.0783
Std. dev	0.0004	0.0004	0.0014	0.0014	0.0025	0.0025
Skewness	0.1013	-0.0015	0.5781	0.5781	-0.7884	-0.1087
Kurtosis	29.6087	20.1280	48.6243	48.6243	76.1225	56.9787

Table 1.4: Descriptive statistics of daily estimates of the log transformation of the negative VaR, $\widehat{q}_{p,d}$, for EURUSD from May 7, 2008 until July 20, 2016 and for the stocks from January 2, 2001 until July 24, 2017.

Sampling scheme	CTS			TTS		
Probability *p*	1%	2.5%	5%	1%	2.5%	5%
	EURUSD					
mean	-6.9670	-7.2699	-7.5406	-7.0042	-7.2884	-7.5406
variance	0.1778	0.1713	0.1743	0.1840	0.1732	0.1696
skewness	-0.0710	-0.0461	-0.0593	-0.0072	-0.0196	-0.0340
kurtosis	3.6626	3.5845	3.5237	3.6346	3.5856	3.4557
JB test-statistic	40.6535	31.0038	25.5263	35.6722	30.4994	18.7972
	IBM					
mean	-5.8222	-6.0944	-6.3459	-5.8254	-6.0991	-6.3494
variance	0.2497	0.2297	0.2642	0.2479	0.2296	0.2275
skewness	0.5722	0.6603	0.6946	0.5468	0.6244	0.6576
kurtosis	3.6888	3.7661	3.6402	3.6367	3.7629	3.6778
JB test-statistic	306.556	400.517	402.02	275.191	367.986	376.185
	BAC					
mean	-5.5273	-5.8033	-6.0581	-5.5334	-5.8069	-6.0575
variance	0.4382	0.3920	0.4583	0.4369	0.3965	0.3834
skewness	0.7855	0.8503	0.8179	0.7952	0.8300	0.8301
kurtosis	4.1658	4.3489	4.1832	4.2087	4.3102	4.3156
JB test-statistic	657.632	809.626	700.317	685.622	768.469	771.082

Table 1.5: Descriptive statistics of daily estimates of the log transformation of the negative ES, $\widehat{es}_{p,d}$, for EURUSD from May 7, 2008 until July 20, 2016 and for the stocks from January 2, 2001 until July 24, 2017.

Sampling scheme	CTS			TTS		
Probability *p*	1%	2.5%	5%	1%	2.5%	5%
	EURUSD					
mean	-6.7249	-6.9328	-7.1361	-6.7673	-6.9683	-7.1625
variance	0.2018	0.1779	0.1687	0.2091	0.1847	0.1728
skewness	0.0256	-0.0529	-0.0802	0.0762	0.0057	-0.0238
kurtosis	3.6180	3.6077	3.6256	3.4211	3.5194	3.5613
JB test-statistic	34.0504	33.6855	36.9249	17.7567	23.8947	28.0986
	IBM					
mean	-5.7586	-5.8774	-6.0389	-5.7617	-5.8807	-6.0428
variance	0.2642	0.2406	0.2253	0.2630	0.2384	0.2230
skewness	0.5408	0.6029	0.6707	0.5162	0.5785	0.6384
kurtosis	3.6064	3.7502	3.8106	3.5437	3.7139	3.7617
JB test-statistic	264.228	346.577	422.147	233.925	317.584	379.783
	BAC					
mean	-5.4636	-5.5838	-5.7493	-5.4703	-5.5891	-5.7522
variance	0.4583	0.4248	0.4024	0.4573	0.4234	0.3975
skewness	0.7563	0.8094	0.8516	0.7740	0.8084	0.8523
kurtosis	4.0682	4.2408	4.3109	4.1412	4.2564	4.3626
JB test-statistic	589.184	714.833	793.821	635.514	720.376	818.307

Appendix 1.C Figures

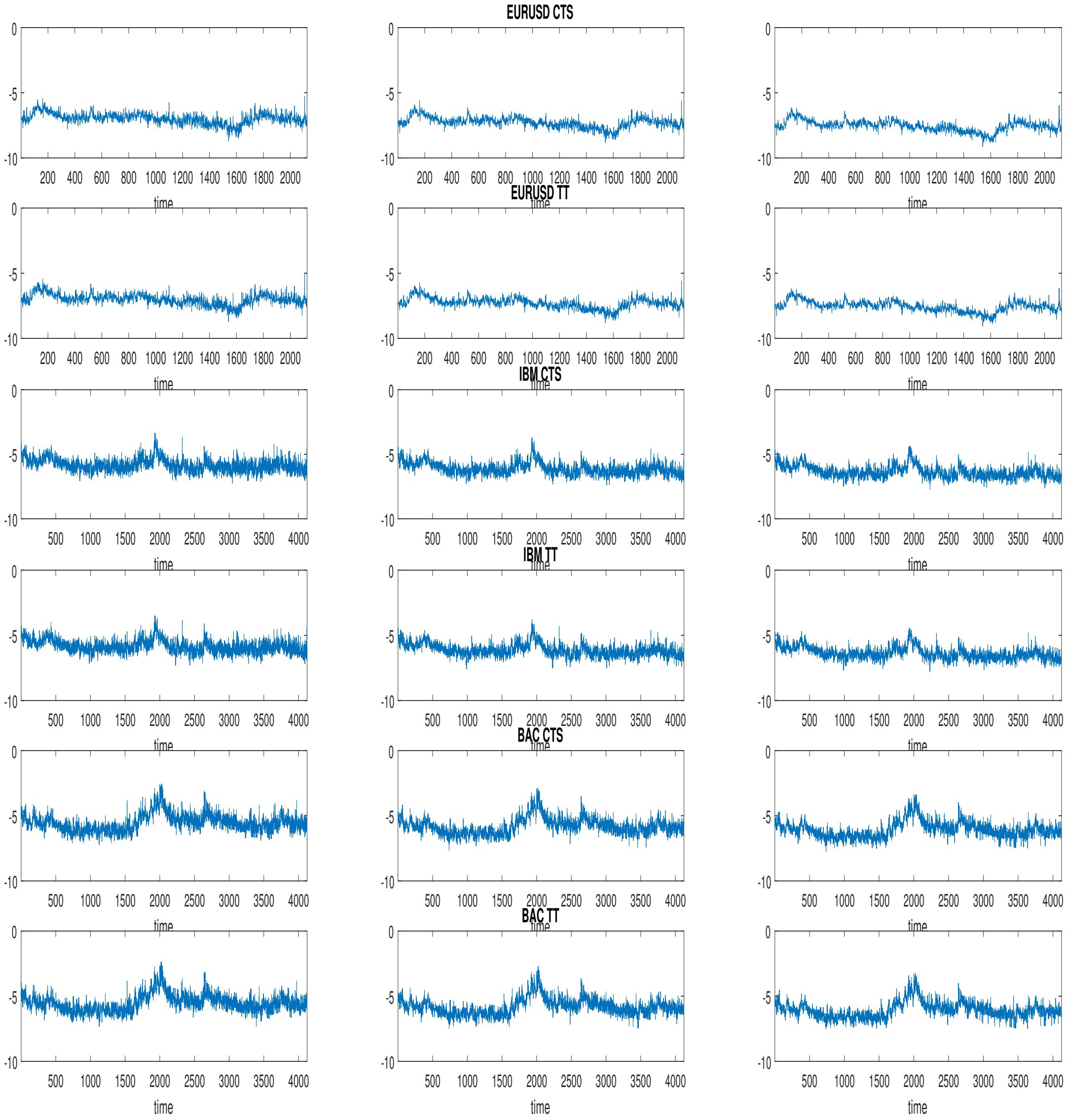


Figure 1.1: Line graph of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$, computed from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The first column corresponds to $p = 1\%$, the second column corresponds to $p = 2.5\%$ and the third column corresponds to $p = 5\%$.

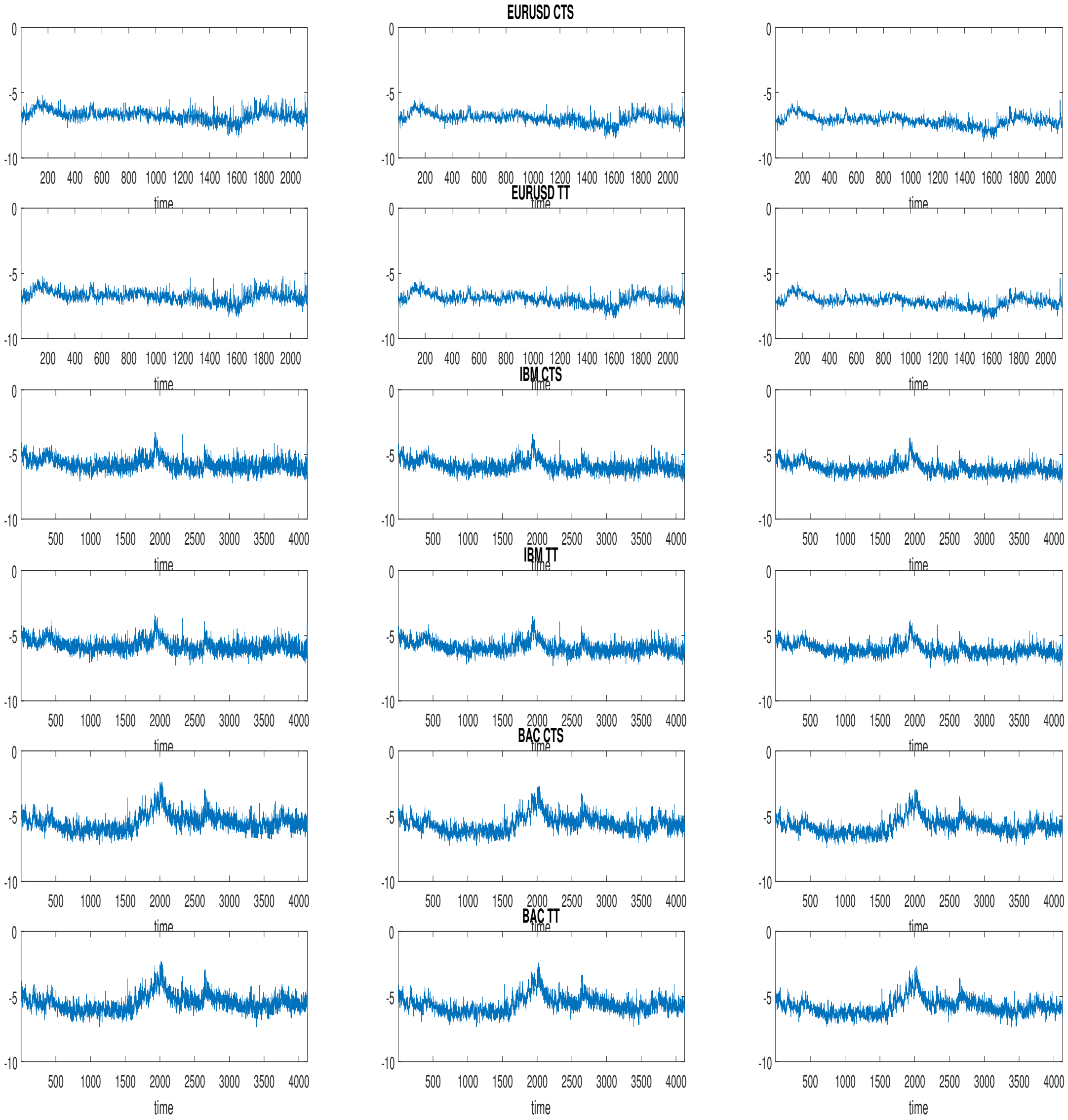


Figure 1.2: Line graph of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$, computed from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The first column corresponds to $p = 1\%$, the second column corresponds to $p = 2.5\%$ and the third column corresponds to $p = 5\%$.

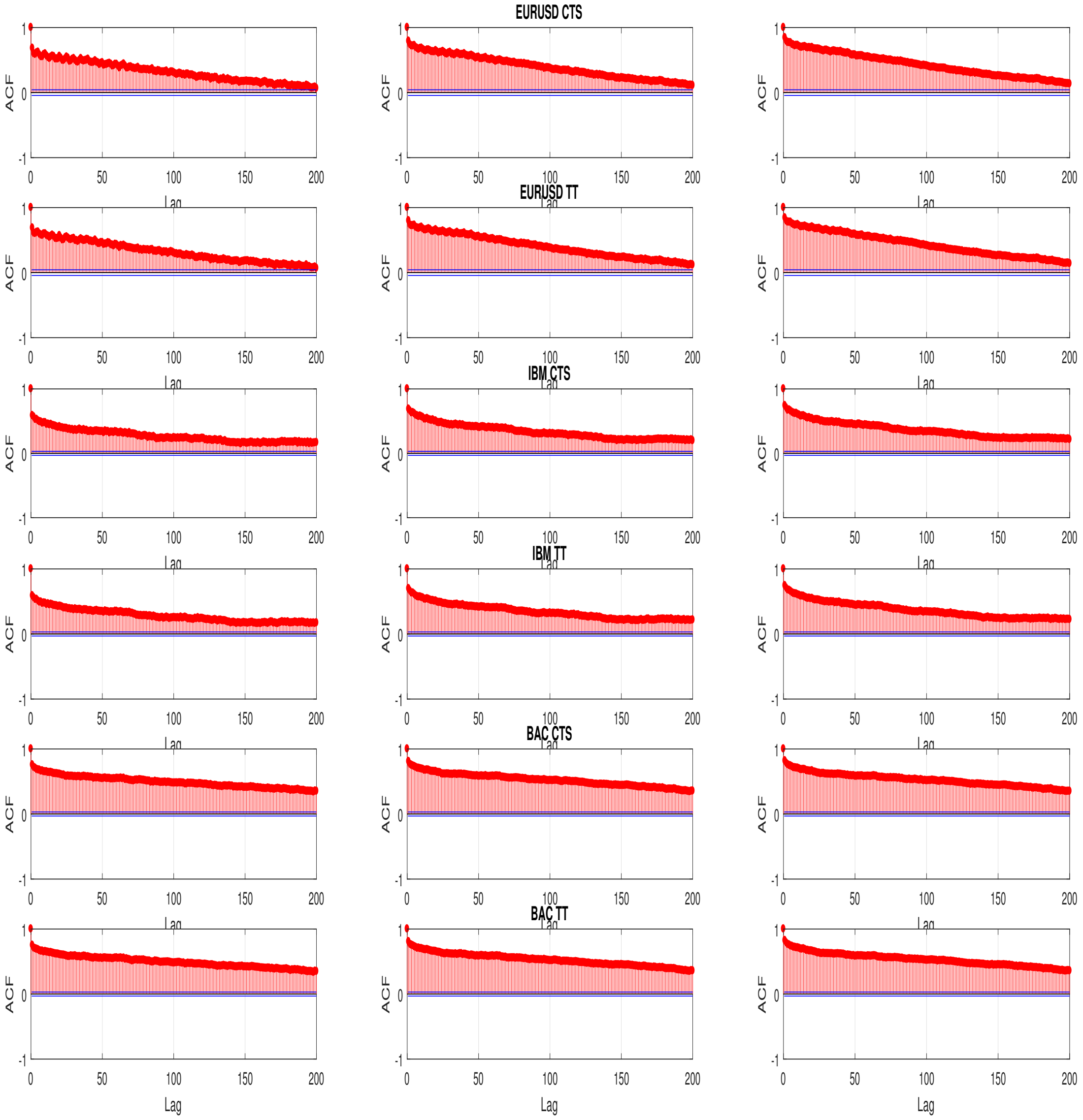


Figure 1.3: ACF of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$, computed from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The first column corresponds to $p = 1\%$, the second column corresponds to $p = 2.5\%$ and the third column corresponds to $p = 5\%$. The blue line represents the 95% confidence interval.

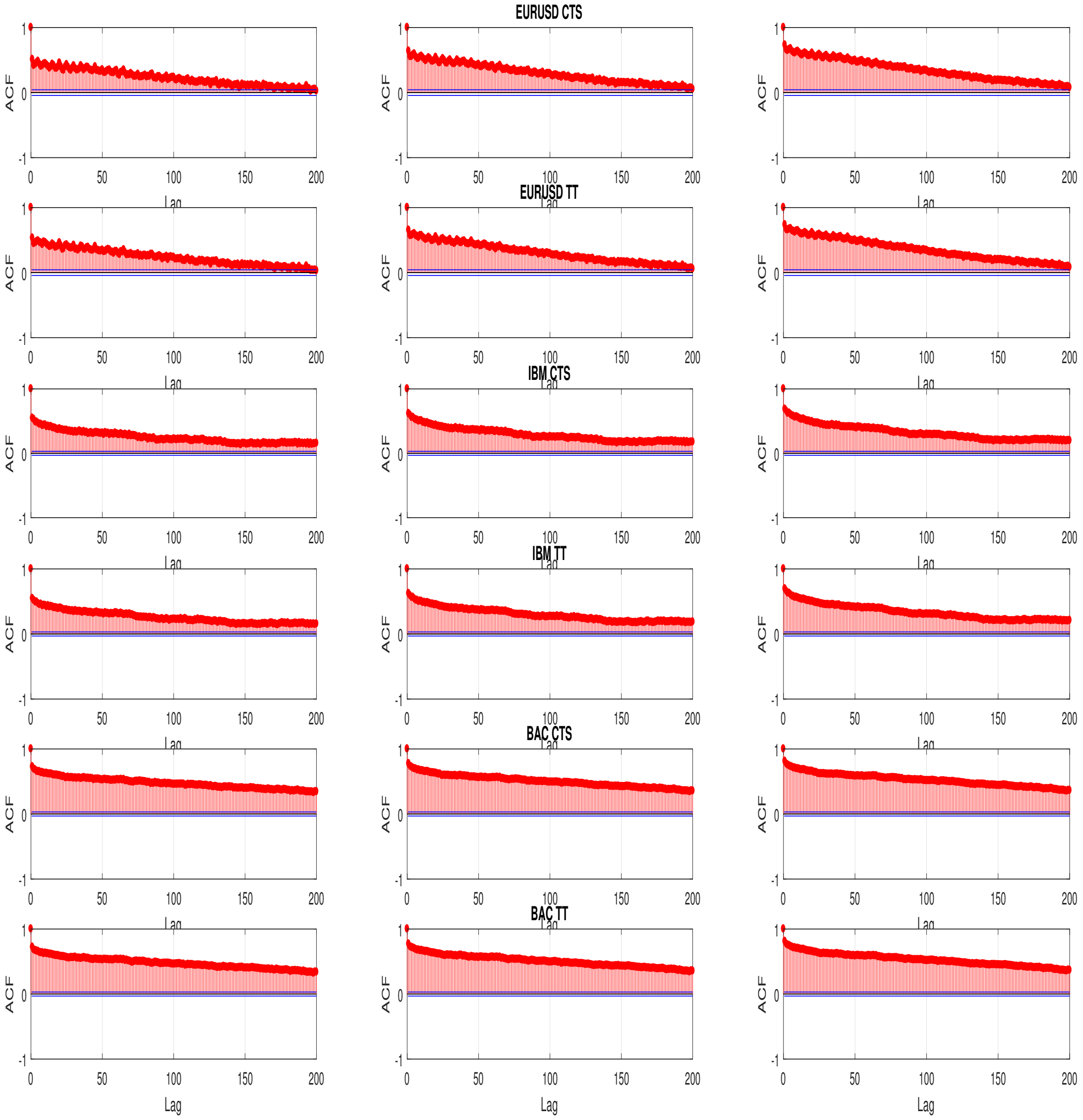


Figure 1.4: ACF of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$, computed from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The first column corresponds to $p = 1\%$, the second column corresponds to $p = 2.5\%$ and the third column corresponds to $p = 5\%$. The blue line represents the 95% confidence interval.

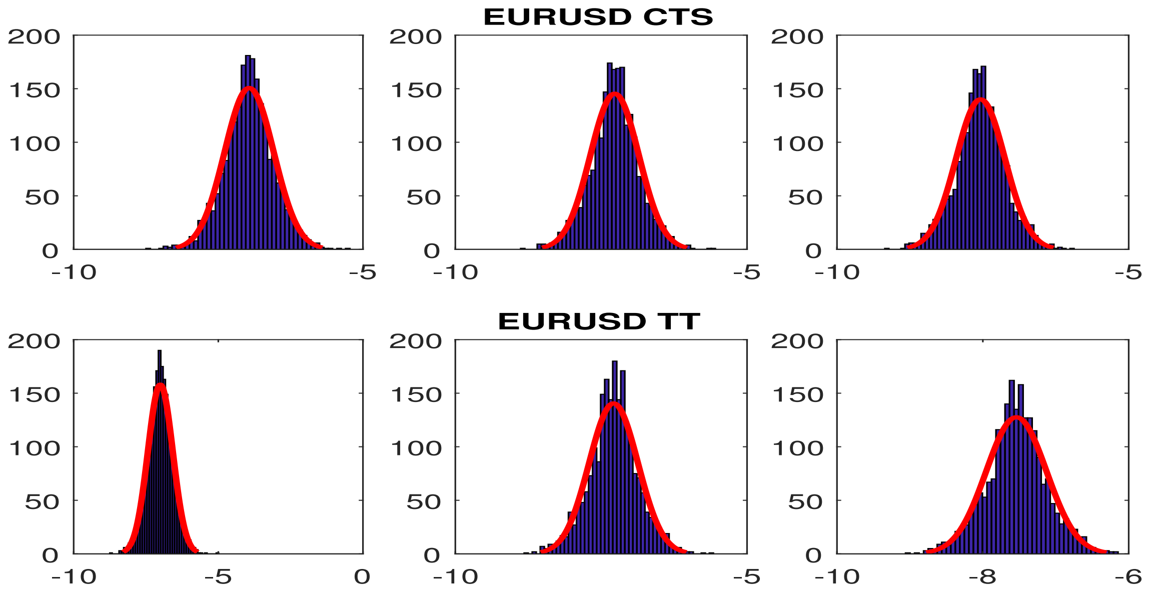


Figure 1.5: Histogram of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$ of EURUSD, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line represents the pdf of a normal distribution.

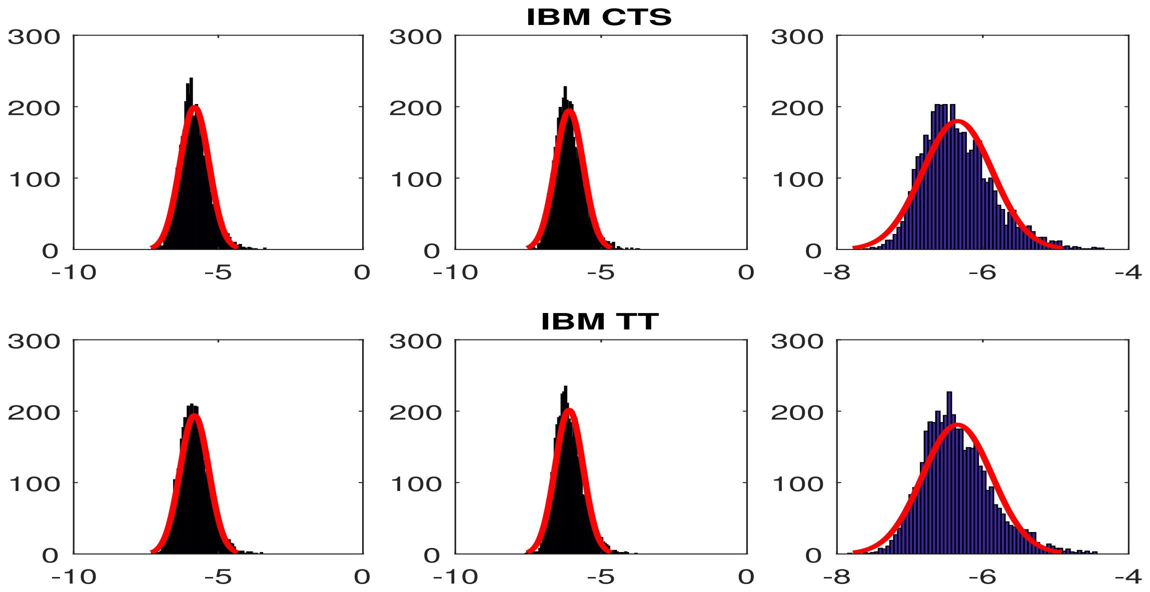


Figure 1.6: Histogram of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$ of IBM, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line represents the pdf of a normal distribution.

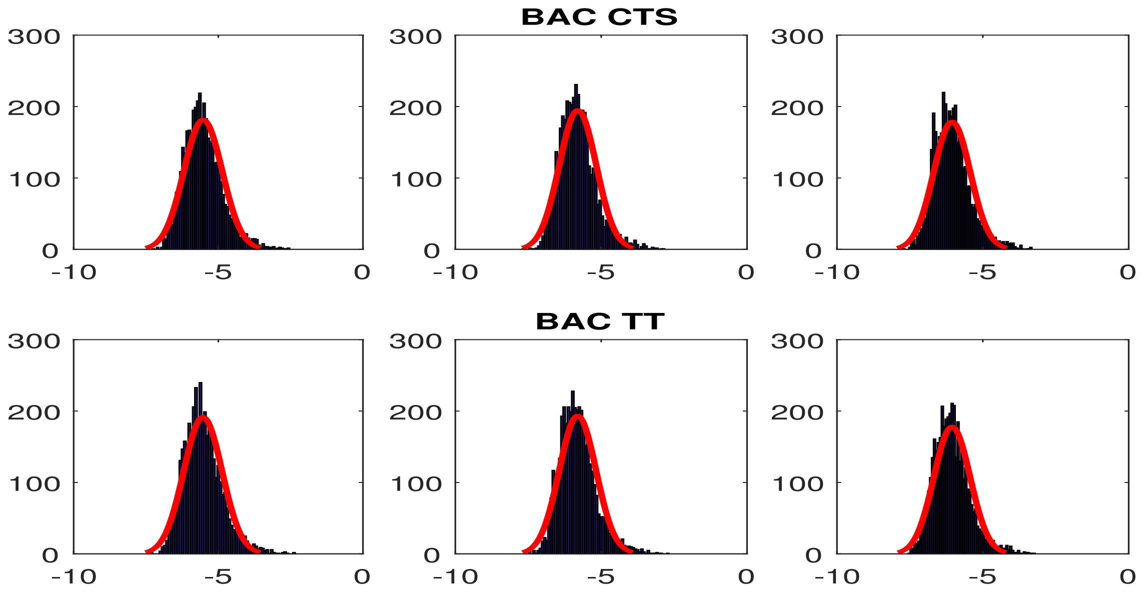


Figure 1.7: Histogram of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$ of BAC, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line represents the pdf of a normal distribution.

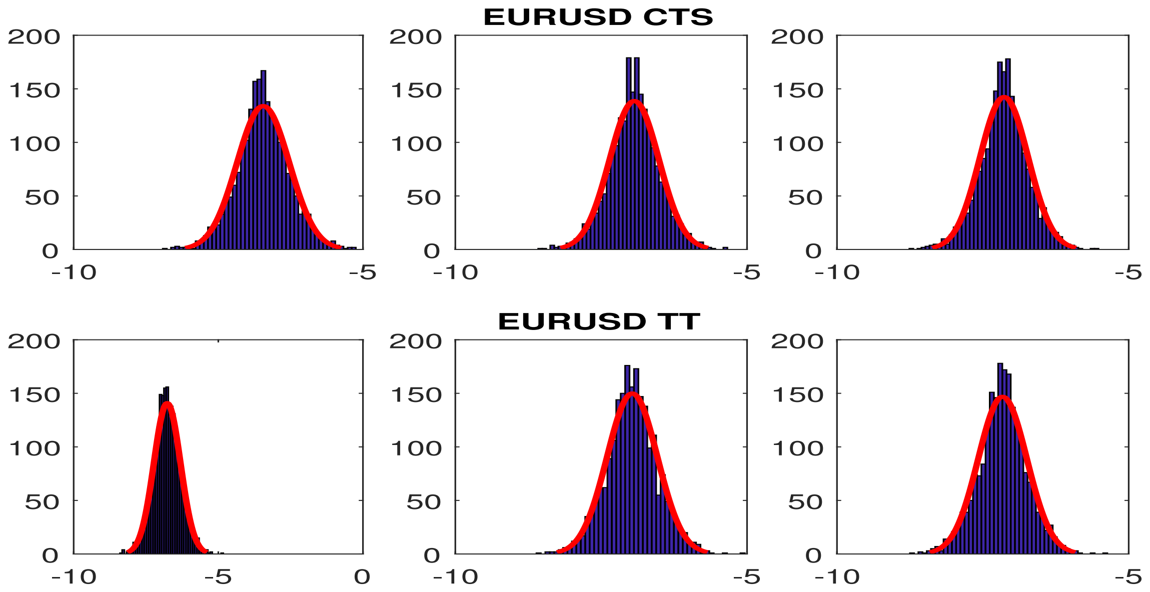


Figure 1.8: Histogram of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$ of EURUSD, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line represents the pdf of a normal distribution.

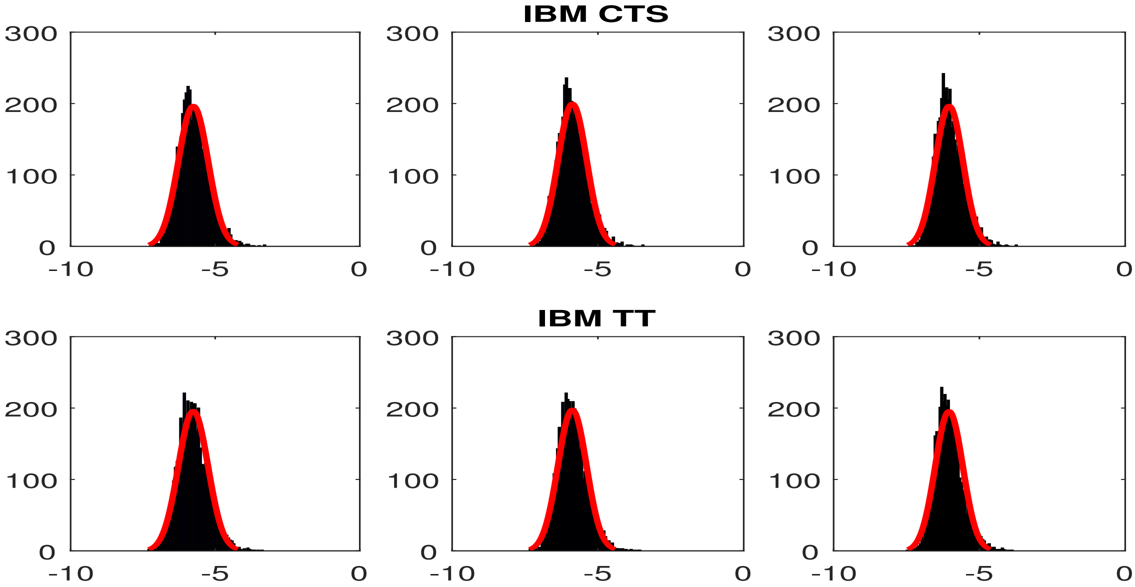


Figure 1.9: Histogram of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$ of IBM, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line represents the pdf of a normal distribution.

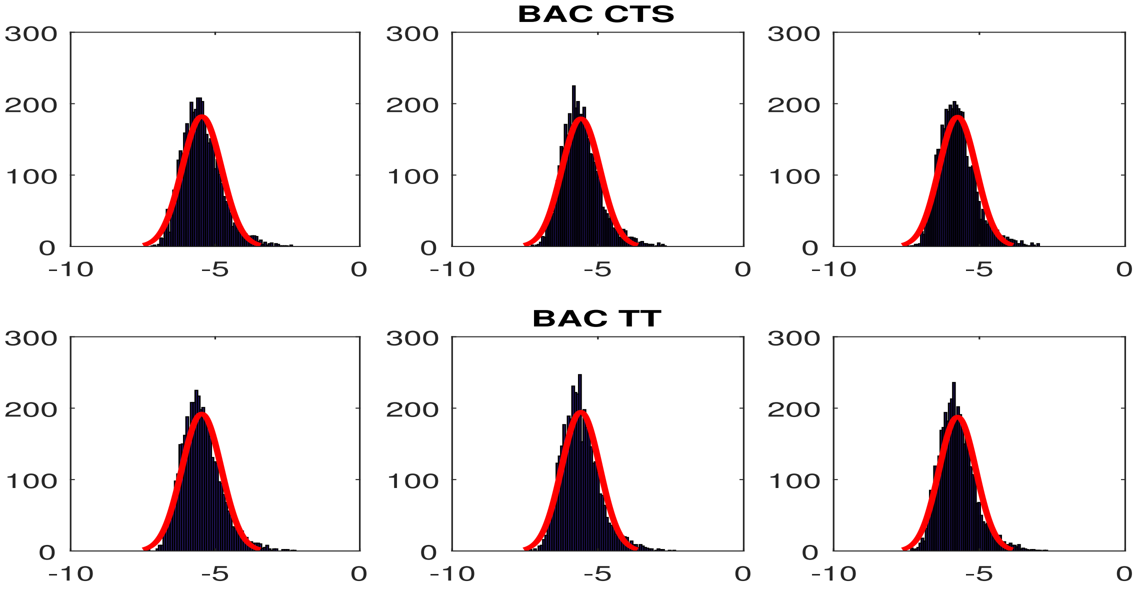


Figure 1.10: Histogram of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$ of BAC, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line represents the pdf of a normal distribution.

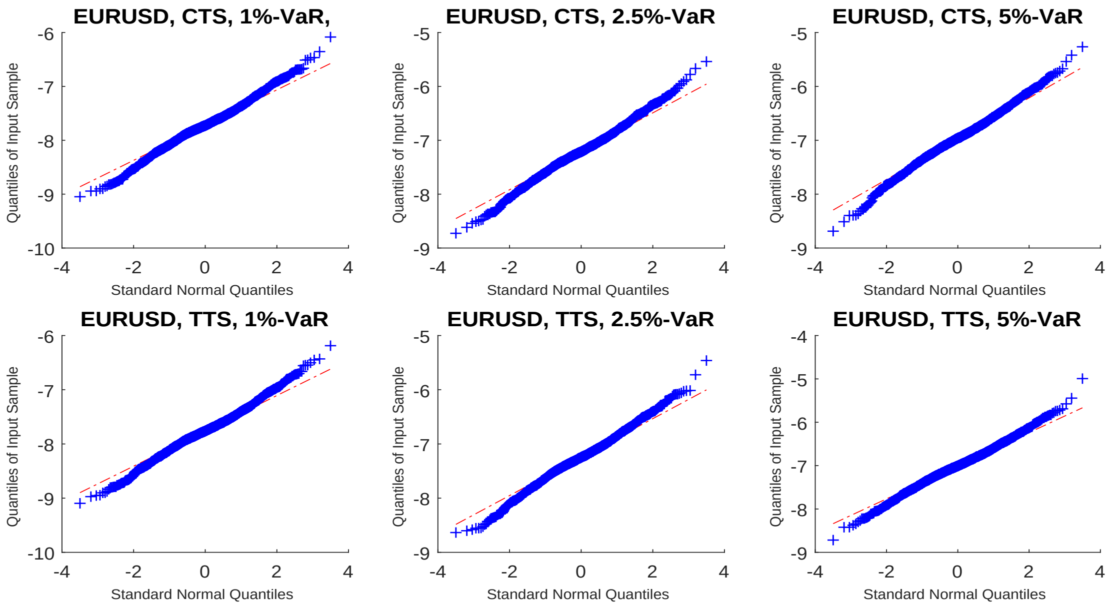


Figure 1.11: QQ-plots of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$ of EURUSD, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line corresponds to the standard normal distribution.

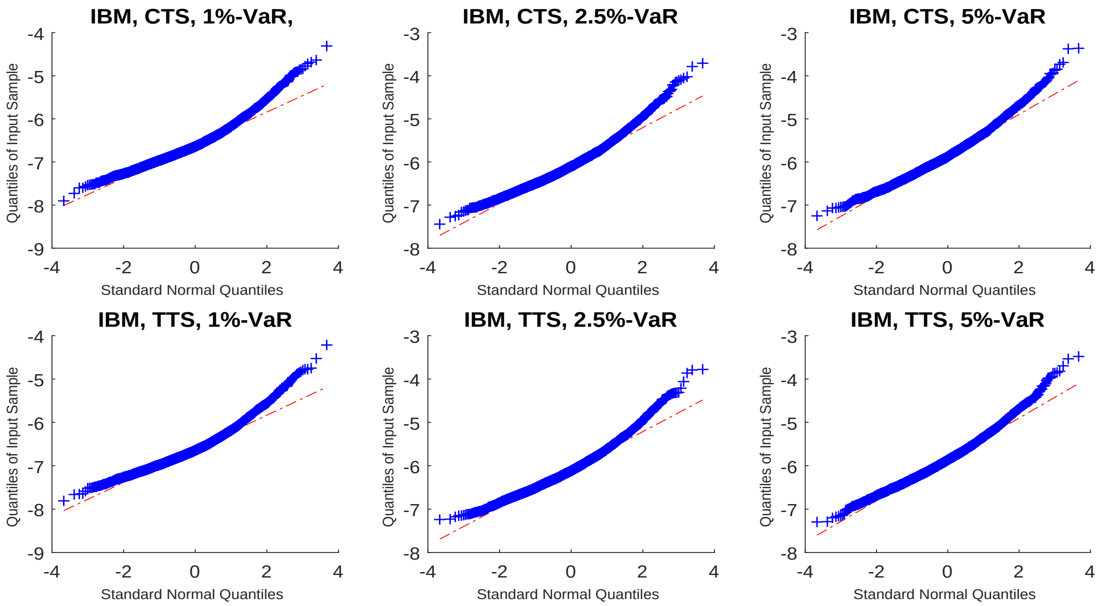


Figure 1.12: QQ-plots of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$ of IBM, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line corresponds to the standard normal distribution.

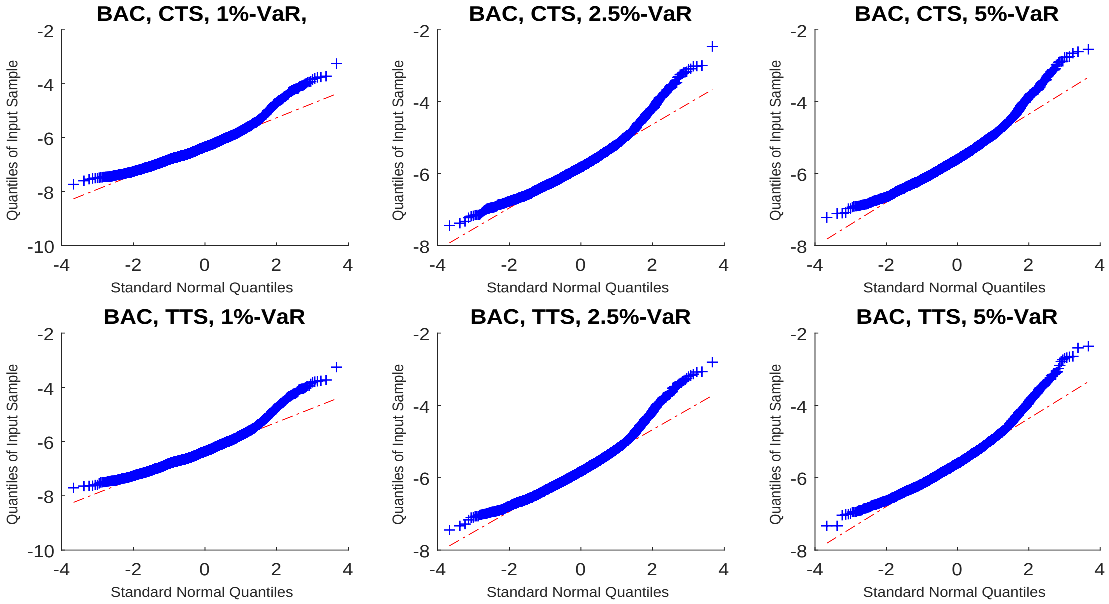


Figure 1.13: QQ-plots of the log-transformation of the negative estimates of daily quantiles, $\widehat{q}_{p,d}$ of BAC, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line corresponds to the standard normal distribution.

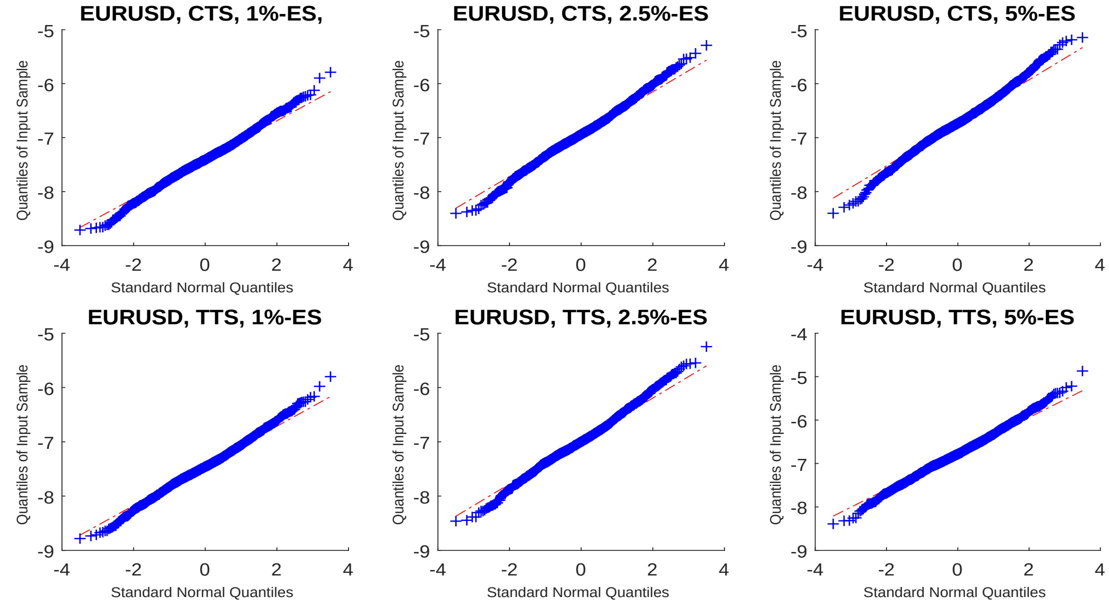


Figure 1.14: QQ-plots of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$ of EURUSD, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line corresponds to the standard normal distribution.

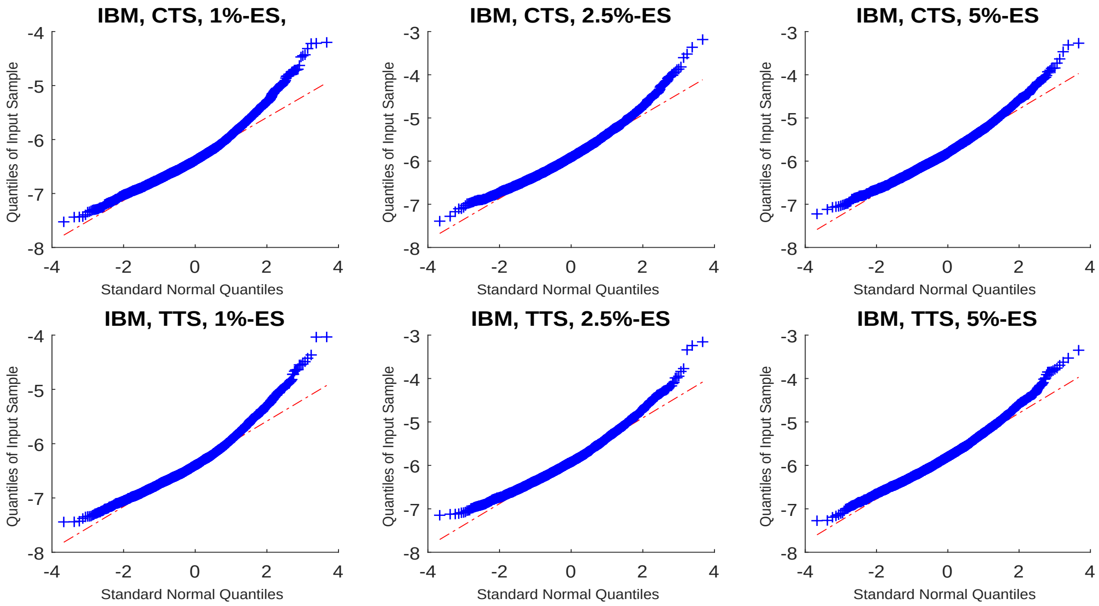


Figure 1.15: QQ-plots of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$ of IBM, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line corresponds to the standard normal distribution.

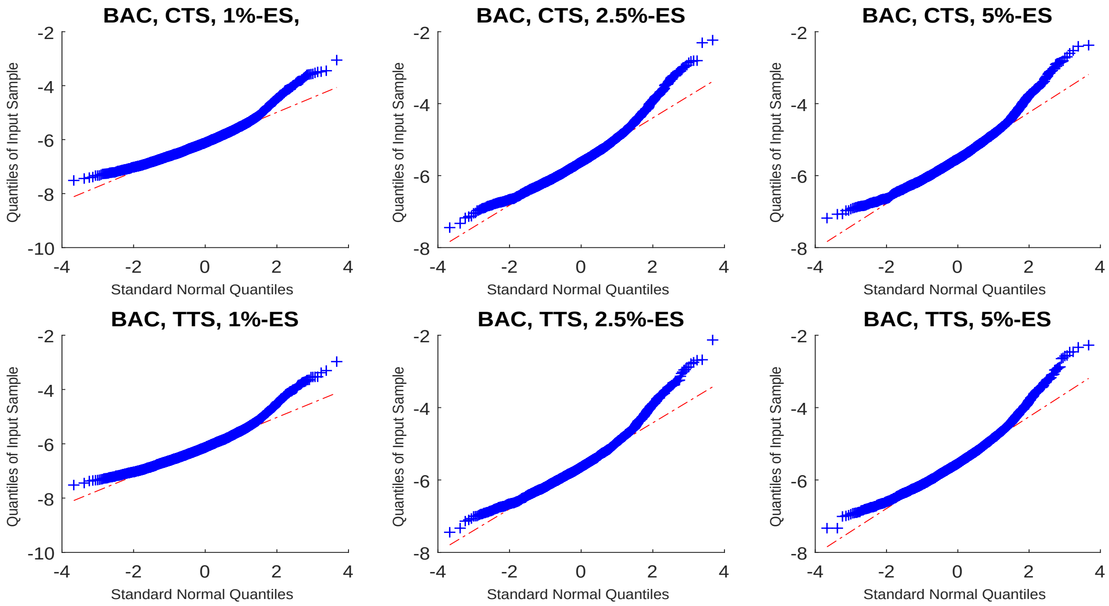


Figure 1.16: QQ-plots of the log-transformation of the negative estimates of daily ES, $\widehat{es}_{p,d}$ of BAC, computed based on the daily estimates of the quantiles as described in the main text from scaling up estimates of intraday quantiles estimated from data sampled in TTS and CTS (use the DMA estimator of Hurst coefficient and the frequency of 5 minute to sample the data). The left, middle an right column corresponds to $p = 1\%$, $p = 2.5\%$ and $p = 5\%$ respectively. The red line corresponds to the standard normal distribution.

References

Alessio, E., A. Carbone, G. Castelli, and V. Frappietro (2002). “Second-order moving average and scaling of stochastic time series”. *The European Physical Journal B - Condensed Matter and Complex Systems* 27 (2), 197–200 (see p. 26).

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). “The Distribution of Realized Exchange Rate Volatility”. *Journal of the American Statistical Association* 96, 42–55 (see p. 30).

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). “The Distribution of Realized Stock Return Volatility”. *Journal of Financial Economics* 61, 43–76 (see p. 25).

Andersen, T. G. and T. Bollerslev (1997). “Intraday Periodicity and Volatility Persistence in Financial Markets”. *Journal of Empirical Finance* 4, 115–158 (see p. 30).

Bandi, F. M. and J. R. Russell (2008). “Microstructure Noise, Realized Volatility and Optimal Sampling”. *The Review of Economic Studies* 75 (2), 339–369 (see p. 17).

Barndorff-Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard (2008). “Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise”. *Econometrica* 76 (6), 1481–1536 (see p. 17).

Basel Committee (1996). *Overview of the Amendment to the Capital Accord to Incorporate Market Risks*. Tech. rep. Available at `http://www.bis.org/publ/bcbs23.pdf`. Basel Committee on Banking Supervision (see pp. 15, 26, 98, 104).

— (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d352.pdf`. Basel Committee on Banking Supervision (see pp. 6, 11, 15, 26, 54, 97, 105).

Beran, J., Y. Feng, S. Ghosh, and R. Kulik (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer (see p. 18).

Bianchi, C. and G. Calzolari (1980). “The One-Period Forecast Errors in Nonlinear Econometric Models”. *International Economic Review* 21, 201–208 (see p. 31).

Black, F. and M. S. Scholes (1973). “The Pricing of Options and Corporate Liabilities”. *Journal of Political Economy* 81 (3), 637–654 (see p. 21).

Boudt, K., C. Croux, and S. Laurent (2011). “Robust Estimation of Intraweek Periodicity in Volatility and Jump Detection”. *Journal of Empirical Finance* 18 (2), 353–367 (see pp. 21, 25, 30).

Burns, A. F. and W. C. Mitchell (1946). “Measuring Business Cycles”. New York, New York: National Bureau of Economic Research (see p. 16).

Calvet, L. E. and A. J. Fisher (2001). “Forecasting Multifractal Volatility”. *Journal of Econometrics* 105 (1), 27–58 (see p. 16).

— (2002). “Multifractality in Asset Returns: Theory and Evidence”. *Review of Economics and Statistics* 84 (3), 381–406 (see pp. 16, 19, 21).

— (2004). “How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes”. *Journal of Financial Econometrics* 2 (1), 49–83 (see p. 20).

— (2008). *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press (see p. 16).

Calvet, L. E., A. J. Fisher, and B. Mandelbrot (1997). “Large Deviation and the Distribution of Price Changes”. Cowles Foundation Discussion Paper (see p. 16).

Clark, P. K. (1973). “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices”. *Econometrica* 41 (1), 135–155 (see pp. 16, 21, 25).

Comte, F. and E. Renault (1996). “Long Memory Continuous Time Models”. *Journal of Econometrics* 73 (1), 101–149 (see p. 21).

— (1998). “Long Memory in Continuous-Time Stochastic Volatility Models”. *Mathematical Finance* 8 (4), 291–323 (see p. 21).

Corsi, F. (2009). “A Simple Approximate Long-Memory Model of Realized Volatility”. *Journal of Financial Econometrics* 7 (2), 174–196 (see p. 30).

Corsi, F., G. Zumbach, U. A. Müller, and M. A. Dacorogna (2001). “Consistent High-Precision Volatility from High-Frequency Data”. *Economic Notes by Banca Monte dei Paschi die Siena SpA* 30 (2), 183–204 (see p. 17).

Dacorogna, M. M., U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet (1993). “A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market”. *Journal of International Money and Finance* 12 (4), 413–438 (see p. 17).

Diebold, F. and R. Mariano (1995). “Comparing Predictive Accuracy”. *Journal of Business and Economic Statistics* 13 (3), 253–63 (see p. 29).

Dong, Y. and Y.-K. Tse (2014). “Business Time Sampling Scheme and Its Application”. Working Paper (see pp. 17, 21, 25, 30).

Embrechts, P. and M. Maejima (2002). *Selfsimilar Processes*. Princeton University Press (see p. 18).

Engle, R. F. and J. R. Russell (1998). “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data”. *Econometrica* 66 (5), 1127–1162 (see pp. 25, 30).

Fissler, T. and J. F. Ziegel (2016). “Higher order elicitability and Osband’s principle”. *The Annals of Statistics* 44 (4), 1680–1707 (see pp. 7, 12, 28, 54, 56–58, 60, 72, 78, 97, 119, 121–123).

Ghysels, E., C. Gouriéroux, and J. Jasiak (1997). “Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets”. In: *Nonlinear Modelling in Foreign Exchange Markets*. Ed. by Dunis, C. and Zhou, B. New York: Wiley (see p. 17).

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). “Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies”. *Journal of Econometrics* 131 (1-2), 59–95 (see p. 30).

Giot, P. and S. Laurent (2004). “Modelling daily Value-at-Risk using realized volatility and ARCH type models”. *Journal of Empirical Finance* 11 (3), 379–398 (see p. 15).

Gneiting, T. (2011). “Making and Evaluating Point Forecasts”. *Journal of the American Statistical Association* 106 (494), 746–762 (see pp. 7, 11, 28, 54–58, 63, 97, 98, 102, 122).

Gouriéroux, C. and J. Jasiak (2001). *Financial Econometrics*. Princeton: Princeton University Press (see pp. 16, 17, 21).

Griffin, J. and R. Oomen (2008). “Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?” *Econometric Reviews* 27 (1-3), 230–253 (see pp. 25, 30).

Guillaume, D. M., M. M. Dacorogna, R. R. Dave, U. A. Müller, R. B. Olsen, and O. V. Pictet (1997). “From the Bird’s Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets". *Finance and Stochastics* 1 (2), 95–129 (see p. 17).

Halbleib, R. and W. Pohlmeier (2012). "Improving the Value at Risk Forecasts: Theory and Evidence from the Financial Crisis". *Journal of Economic Dynamics and Control* 36 (8), 1212–1228 (see pp. 15, 28, 54, 73).

Halbleib, R. and V. Voev (2011). "Forecasting Multivariate Volatility Using the VARFIMA Model on Realized Covariance Cholesky Factors". *Journal of Economics and Statistics* 231 (1), 134–152 (see p. 31).

Hallam, M. and J. Olmo (2014a). "Forecasting Daily Return Densities from Intraday Data: A Multifractal Approach". *International Journal of Forecasting* 30 (4), 863–881 (see pp. 16, 17, 20, 27).

— (2014b). "Semiparametric Density Forecasts of Daily Financial Returns from Intraday Data". *Journal of Financial Econometrics* 12 (2), 408–432 (see pp. 16, 28).

Hansen, P. R. and A. Lunde (2004). "Realized Variance and IID Market Microstructure Noise". Brown University Working Paper (see p. 17).

— (2006). "Realized Variance and Market Microstructure Noise". *Journal of Business and Economic Statistics* 24, 127–218 (see p. 26).

Harrell, F. and E. Davis (1982). "A New Distribution-Free Quantile Estimator". *Biometrika* 69 (3), 635–640 (see p. 26).

Harris, L. (1986). "A transaction data study of weekly and intradaily patterns in stock returns". *Journal of Financial Economics* 16 (1), 99–117 (see p. 30).

Kantelhardt, J. W. (2009). "Fractal and Multifractal Time Series". In: *Mathematics of Complexity and Dynamical Systems*. Ed. by Meyers, R. A. New York: Springer New York, 463–487 (see pp. 20, 24).

Kobeissi, Y. H. (2013). *Multifractal Financial Markets*. Springer (see p. 20).

Kuester, K., S. Mitinik, and M. S. Paolella (2006). "Value-at-Risk Prediction: A Comparison of Alternative Strategies". *Journal of Financial Econometrics* 4 (1), 53–89 (see p. 26).

Lux, T. and M. Segnon (2018). "Multifractal models in finance: their origin, properties, and applications". In: *The Oxford Handbook of Computational Economics and Finance*. Ed. by Shu-Heng Chen, M. K. and Du, Y.-R. Oxford University Press. Chap. 6 (see pp. 16, 20, 21).

Mandelbrot, B. B. (1963). "The Variation of Certain Speculative Prices". *Journal of Business* 36, 394–419 (see p. 18).

— (1974). "Intermittent Turbulence in Self Similar Cascades: Divergence of High Moments and Dimension of the Carrier". *Journal of Fluid Mechanics* 62 (2), 331–358 (see p. 19).

— (1982). *The Fractal Geometry of Nature*. W. H. Freeman and Company, New York (see p. 18).

— (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, New York, 394–419 (see p. 21).

— (1999). *Multifractals and 1/f Noise*. Springer Verlag, New York (see p. 16).

Mandelbrot, B. B., A. Fisher, and L. Calvet (1997). "A Multifractal Model of Asset Returns". Cowles Foundation Discussion Paper (see pp. 16, 19–21).

Mandelbrot, B. B. and H. M. Taylor (1967). "On the Distribution of Stock Price Differences". *Operations Research* 15 (6), 1057–1062 (see pp. 16, 25).

Markowitz, H. (1952). "Portfolio Selection". *Journal of Finance* 7 (1), 77–91 (see p. 21).

McAleer, M. and M. C. Medeiros (2008). “Realized Volatility: A Review”. *Econometric Reviews* 27 (1-3), 10–45 (see p. 17).

Müller, U. A., M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen, and J. R. Ward (1995). “Fractals and Intrinsic Time - A Challenge to Econometricians”. Working Paper Series, Olsen&Associates (see p. 17).

Oomen, R. C. A. (2001). “Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realized Return Variance”. European University, Economics Discussion Paper No. 2001/6 (see p. 31).

— (2005). “Properties of Bias-Corrected Realized Variance under Alternative Sampling Schemes”. *Journal of Financial Econometrics* 3 (4), 555–577 (see pp. 17, 21, 25, 30).

— (2006). “Properties of Realized Variance Under Alternative Sampling Schemes”. *Journal of Business and Economic Statistics* 24 (2), 219–237 (see pp. 17, 25, 30).

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994). “Mosaic organization of DNA nucleotides”. *Phys. Rev. E* 49 (2), 1685–1689 (see p. 26).

Sánchez-Granero, M., M. Fernández-Martínez, and J. Trinidad-Segovia (2012). “Introducing fractal dimension algorithms to calculate the Hurst exponent of financial time series”. *The European Physical Journal B* 85 (3) (see p. 26).

Stock, J. (1987). “Measuring Business Cycle Time”. *Journal of Political Economy* 95 (6), 1240–1261 (see p. 17).

— (1988). “Estimating Continuous Time Processes Subject to Time Deformation”. *Journal of the American Statistical Association* 83 (401), 77–85 (see p. 17).

Wu, Z. (2012). “On the Intraday Periodicity Duration Adjustment of High-Frequency Data”. *Journal of Empirical Finance* 19 (4), 282–291 (see p. 25).

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data”. *Journal of the American Statistical Association* 100 (472), 1394–1411 (see pp. 17, 26).

Chapter 2

A Joint Quantile and Expected Shortfall Regression Framework

We thank Tobias Fissler, Lyudmila Grigoryeva, Roxana Halbleib, Phillip Heiler, Frederic Menninger, Winfried Pohlmeier, Patrick Schmidt, Johanna Ziegel and the participants of the Stochastics Colloquium on 11/30/2016 at the University of Konstanz for fruitful discussions and suggestions which inspired some of the results of this paper. Financial support by the Heidelberg Academy of Sciences and Humanities (HAW) within the project "Analyzing, Measuring and Forecasting Financial Risks by means of High-Frequency Data", by the German Research Foundation (DFG) within the research group "Robust Risk Measures in Real Time Settings" and general support by the Graduate School of Decision Sciences (University of Konstanz) is gratefully acknowledged. The computation in this work was performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC.

2.1. Introduction

We introduce a novel semiparametric joint regression framework for the Expected Shortfall (ES) by jointly modeling both, regression equations for the conditional quantile and the conditional ES. The functional ES is defined as the expected value of a random variable, given that its realizations are smaller than some quantile of the underlying distribution. We propose both, an M- and a Z-estimator for the joint regression parameters and show that these estimators are consistent and asymptotically normal under weak regularity conditions. Modeling simultaneous regression equations for the quantile and the ES is necessary as M- and Z-estimation of regression parameters of a stand-alone regression framework for the ES is infeasible. The underlying reason is that there does not exist an appropriate loss function that the ES minimizes in expectation and which could be used as the objective function for M-estimation of the regression parameters (Gneiting, 2011). However, Fissler and Ziegel (2016) show that there exists such a loss function if one considers the pair consisting of the quantile and the ES at the same probability level. This result gives rise to the idea of jointly modeling semiparametric models for both, the quantile and the ES and for jointly estimating the regression parameters through M-estimation. This paper is the first to propose such a joint regression framework in the sense that we model multiple (different) functionals at the same time. The situation for the Z-estimator and the availability of underlying identification functions (moment conditions) is equivalent to the loss functions used for the M-estimator and consequently only allows for joint Z-estimation of both regression equations.

Such a regression framework for the ES is essential for a variety of academic disciplines which consider measuring, forecasting and the evaluation of extreme risks. The most prominent example for this is financial risk management, where the Basel Accords recently proposed to use ES as the standard risk measure (Basel Committee, 2016). The previously used risk measure is the Value-at-Risk (VaR), which is defined as the α-quantile of the return distribution and which has several drawbacks as it is not coherent and fails to capture tail risks beyond itself (Artzner et al., 1999). These deficiencies are overcome by the ES as it has the desired ability to capture information from the whole left tail of the return distribution, which is particularly important for measuring extreme financial risks. Modeling a regression equation for the ES opens up the possibility to extend the existing applications of quantile regression on VaR in the financial literature to ES, such as e.g. in Chernozhukov and Umantsev (2001), Engle and Manganelli (2004), Koenker and Xiao (2006), Gaglianone et al. (2011), Halbleib and Pohlmeier (2012), Komunjer (2013), Xiao et al. (2015) and Žikeš and Baruník (2016). Such estimation, forecasting, and backtesting methods for the ES are particularly sought-after in light of the recent shift from VaR to ES in the Basel Accords.

A further possible field of application for this regression framework arises in microeconometrics where researchers are interested in non-central features of the conditional distribution such as e.g. in income economics and the analysis of social inequalities. In these fields, a traditional method for the comparison of different regions of the conditional distribution is quantile regression (see e.g. Koenker (2005), Section 1.5.). However, the interpretation of the ES as the mean of the worst α percent is more intuitive as the rather technical interpretation of quantiles, which motivates the use of an ES regression technique in these fields.

Nadarajah et al. (2014) provide an overview of existing estimation methods for the ES. However, the reviewed approaches are only applicable for univariate data and not suitable for estimating the conditional ES through a regression technique. Nevertheless, there are some approaches for the ES which incorporate explanatory variables through indirect estimation procedures. Taylor (2008b) proposes an implicit approach for forecasting ES using exponentially weighted quantile regression and Taylor (2008a) introduces a procedure based on expectile regression and a relationship between the ES and expectiles. Taylor (2017) suggests a joint modeling technique for the quantile and the ES based on maximum likelihood estimation of the asymmetric Laplace distribution. However, asymptotic statistical theory for these estimation approaches of the conditional ES is not available. Barendse (2017) proposes a two-step estimation approach for a regression framework for the interquantile expectation and Patton et al. (2017) use an ES regression framework in order to estimate dynamic models for the ES.

M-estimation (Z-estimation) of regression frameworks for different functionals can usually be applied based on different choices of loss (identification) functions. E.g. mean regression parameters can be estimated by employing any loss function from the Bregman class of loss functions and quantile regression parameters can be estimated by employing a member of the class of generalized piecewise linear loss functions (Gneiting, 2011). Equivalently, the possible loss and identification functions we employ for the M- and Z-estimator in this paper are not unique as they depend on two specification functions which can be chosen freely subject to some conditions. Even though consistency and asymptotic normality hold for all applicable choices of these specification functions, they affect the asymptotic covariance of the estimators, the necessary moment conditions, the numerical stability of the optimization algorithm and the required computation (optimization) times. We discuss the choice of these functions in a theoretical context with respect to asymptotic efficiency and necessary regularity conditions, and with respect to the numerical properties of the optimization algorithm.

The estimation of the asymptotic covariance matrix of the regression parameters imposes some difficulties. The first occurs in the estimation of the density quantile function, analogous to quantile regression (cf. Koenker, 2005) and thus, we utilize estimation procedures stemming from this literature. The second issue is the estimation of the variance of the negative quantile residuals conditional on the covariates, a nuisance quantity which is new to the literature. We introduce several estimators for this quantity which are able to cope with limited sample sizes and which can model the dependency of the negative quantile residuals on the covariates. Furthermore, we estimate the covariance matrix using the bootstrap. For ease of application, we provide an R package (Bayer and Dimitriadis, 2017b) which contains the implementation of the M- and Z-estimator and where the user can choose the specification functions, the numerical optimization procedure and the estimation method for the covariance matrix of the parameter estimates.

We conduct a Monte-Carlo simulation study where we consider three data generating processes with different properties. We numerically verify consistency and asymptotic normality of the M-estimator for a range of different choices of the specification functions. Furthermore, we find that the Z-estimator is numerically unstable due to the redescending nature of the utilized identification functions and consequently, we rely on M-estimation of the regression parameters. Moreover, we find that the performance of the M-estimator strongly depends on the specification functions, where choices resulting in positively homogeneous loss functions (Efron, 1991; Nolde and Ziegel, 2017) lead to a superior performance in terms of asymptotic efficiency, computation times, and mean squared error of the estimator.

The rest of the paper is organized as follows. In Section 2.2, we introduce the joint regression framework, the underlying regularity conditions together with the asymptotic properties of our estimators and discuss the choice of the specification functions. Section 2.3 provides details on the numerical implementation of the estimators and on the estimation of the asymptotic covariance matrix. Section 2.4 presents an extensive simulation study and Section 2.5 provides concluding remarks. The proofs are deferred to Appendix 2.B.

2.2. Methodology

2.2.1. The Joint Regression Framework

Following Lambert et al. (2008), Gneiting (2011) and Fissler and Ziegel (2016), we introduce the concept of (multivariate) p-elicitability. We consider a random variable $Z : \Omega \to \mathbb{R}^d$, defined on some complete probability space $(\Omega, \mathcal{F}, P)$, a class of distributions $\mathcal{P}$ on $\mathbb{R}^d$, equipped with the Borel σ-field and a functional $T : \mathcal{P} \to D$ with its domain of action $D \subseteq \mathbb{R}^p, p \in \mathbb{N}$. We call an integrable loss function $\rho : \mathbb{R}^d \times D \to \mathbb{R}$ *strictly consistent* for the functional T relative to the class of distributions $\mathcal{P}$, if T is the unique minimizer of $\mathbb{E}\big[\rho(Z, \cdot)\big]$ for all distributions $F \in \mathcal{P}$, where F is the distribution of Z. Furthermore, we call a p-dimensional functional T *p-elicitable* relative to the class $\mathcal{P}$, if there exists a loss function ρ which is strictly consistent for T relative to $\mathcal{P}$. If the dimension p is clear from the context, we simply call the functional elicitable instead of p-elicitable.

Given the generalized α-quantile $Q_\alpha(Z) = F^{-1}(\alpha) = \inf\big\{z \in \mathbb{R} : F(z) \geq \alpha\big\}$ for some $\alpha \in (0, 1)$, the ES of the random variable Z at level α is defined as $\mathrm{ES}_\alpha(Z) = \frac{1}{\alpha}\int_0^\alpha Q_u(Z)\,\mathrm{d}u$. If the distribution function of Z is continuous at its α-quantile, this definition can be simplified to the conditional tail expectation $\mathrm{ES}_\alpha(Z) = \mathbb{E}\big[Z \,\big|\, Z \leq Q_\alpha(Z)\big]$. Gneiting (2011) shows that the ES is not 1-elicitable with respect to any class $\mathcal{P}$ of probability distributions on intervals $I \subseteq \mathbb{R}$, which contains measures with finite support or finite mixtures of absolutely continuous distributions with compact support (see also Weber (2006)). This result has several consequences for the risk measure ES. First, consistent and meaningful ranking of competing forecasts for the functional ES is infeasible. Second, and more consequential for this work, estimating the parameters of a stand-alone regression model for the functional ES in the sense that $\mathrm{ES}_\alpha(Y|X) = X'\theta_0^e$ by means of M-estimation, i.e. by minimizing some strictly consistent loss function, is infeasible. Even though the ES is not 1-elicitable, Fissler and Ziegel (2016) show that the pair consisting of the ES and the quantile at common probability level α is 2-elicitable relative to the class of distributions with finite first moments and unique α-quantiles and they characterize the full class of strictly consistent loss functions for this pair subject to some regularity conditions. Since the definition of the ES already depends on the respective quantile, the fact that the ES is only elicitable jointly with the quantile is not surprising.

We utilize this joint elicitability result for the introduction of a new joint regression framework for the quantile and the ES where the aforementioned class of strictly consistent loss functions serves as the basis for the M-estimation of the joint regression parameters. For this, let $Y : \Omega \to \mathbb{R}$ and $X : \Omega \to \mathbb{R}^k$ be random variables defined on the some complete probability space $\big(\Omega, \mathcal{F}, P\big)$ as above. Henceforth, the transpose of X will be denoted by X', the cumulative distribution function of Y given X by $F_{Y|X}$ and the conditional density function by $f_{Y|X}$. For an l-times differentiable real-valued function $G : \mathbb{R} \to \mathbb{R}$, we denote the l-th derivative by $G^{(l)}(\cdot)$.

Assumption 2.2.1 (The joint regression model). The regression framework which jointly models the conditional quantile and ES of Y given X for some fixed level $\alpha \in (0, 1)$ is given by

$$Y = X'\theta_0^q + u^q \qquad \text{and} \qquad Y = X'\theta_0^e + u^e, \tag{2.1}$$

where $Q_\alpha(u^q|X) = 0$ and $\mathrm{ES}_\alpha(u^e|X) = 0$. The model is parametrized by $\theta_0 = (\theta_0^{q\prime}, \theta_0^{e\prime})' \in \Theta \subset \mathbb{R}^{2k}$, where the parameter space Θ is compact, convex and has nonempty interior, $\mathrm{int}(\Theta) \neq \emptyset$.

This assumption implies that the model is correctly specified in the sense that there exists a true parameter vector $\theta_0 \in \Theta$ for which our joint model is equal to the true quantile and ES of the conditional distribution of Y given X. This model is semiparametric in the sense that we assume that there exists a parametric model for the quantile and ES of the conditional distribution $F_{Y|X}$ without fully specifying this conditional distribution through parametric restrictions. We propose both, an M-estimation and a Z-estimation procedure for the compound regression parameter vector θ_0. For the M-estimation, we adapt the class of strictly consistent joint loss functions[1] for the quantile and ES as given in Fissler and Ziegel (2016) such that it can be used in a regression framework,

$$\begin{aligned}\rho(Y, X, \theta) = &\left(\mathbb{1}_{\{Y \le X'\theta^q\}} - \alpha\right) G_1(X'\theta^q) - \mathbb{1}_{\{Y \le X'\theta^q\}} G_1(Y) \\ &+ G_2(X'\theta^e)\left(X'\theta^e - X'\theta^q + \frac{(X'\theta^q - Y)\mathbb{1}_{\{Y \le X'\theta^q\}}}{\alpha}\right) - \mathcal{G}_2(X'\theta^e) + a(Y),\end{aligned} \tag{2.2}$$

where the function G_1 is twice continuously differentiable, $\mathcal{G}_2$ is three times continuously differentiable, $\mathcal{G}_2^{(1)} = G_2$, G_2 and $G_2^{(1)}$ are strictly positive, G_1 is increasing and a and G_1 are integrable functions. Fissler and Ziegel (2016) also show that given some unimportant regularity conditions, there exist no strictly consistent loss functions outside the class of functions given in (2.2) which implies that this is the most general class of objective functions for the M-estimator of this regression framework. We discuss the choice of the *specification functions* G_1 and $\mathcal{G}_2$ in a theoretical context in Section 2.2.3 and by their numerical performance in Section 2.4.2. The function a only depends on Y and thus, it does not influence the estimated regression parameters and is usually set to zero. The corresponding (ρ-type) M-estimator is formally defined as a sequence $\hat{\theta}_{\rho,n}$ such that

$$\hat{\theta}_{\rho,n} = \mathrm{argmin}_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \rho(Y_i, X_i, \theta). \tag{2.3}$$

Instead of minimizing some objective function $\rho(Y, X, \theta)$ such as in (2.2) and (2.3), we can also define the corresponding Z-estimator (or ψ-type M-estimator), which sets a vector

---

[1] One can interpret the structure of this loss function as follows (Fissler et al., 2016): The first summand in (2.2) is a strictly consistent loss function for the quantile (Gneiting, 2011) and hence only depends on the quantile, whereas the second summand cannot be split into a part depending only on the quantile and one depending only on the ES. This illustrates the fact that the ES itself is not 1-elicitable, but 2-elicitable together with the respective quantile.

of identification functions (moment conditions) to zero. In the case of our joint quantile and ES regression, these identification functions are given by

$$\psi(Y,X,\theta) = \begin{pmatrix} \psi_1(Y,X,\theta) \\ \psi_2(Y,X,\theta) \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha}\left(\mathbb{1}_{\{Y \le X'\theta^q\}} - \alpha\right)\left(\alpha X G_1^{(1)}(X'\theta^q) + X G_2(X'\theta^e)\right) \\ X G_2^{(1)}(X'\theta^e)\left(X'\theta^e - X'\theta^q + \frac{1}{\alpha}(X'\theta^q - Y)\mathbb{1}_{\{Y \le X'\theta^q\}}\right) \end{pmatrix}, \tag{2.4}$$

where the functions G_1 and G_2 are given as above. More generally, it suffices that these identification functions converge to zero almost surely and thus, we formally define the Z-estimator as a sequence $\hat{\theta}_{\psi,n}$, such that

$$\frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, X_i, \hat{\theta}_{\psi,n}) \to 0 \quad \text{a.s.}. \tag{2.5}$$

Identification functions for a regression framework are usually obtained as the derivative of some corresponding loss function. However, for this joint quantile and ES regression, the loss functions $\rho(Y,X,\theta)$ given in (2.2) are only differentiable for the points where $Y \neq X'\theta^q$. As these points of non-differentiability form a nullset with respect to the absolutely continuous distribution of Y given X, the identification functions $\psi(Y,X,\theta)$ are almost surely the derivative of $\rho(Y,X,\theta)$. When the loss function $\rho(Y,X,\theta)$ is continuously differentiable in θ, it is obvious that the M- and Z-estimation approaches are equivalent. However, in this case the loss function $\rho(Y,X,\theta)$ is not differentiable everywhere and $\psi(Y,X,\theta)$ is discontinuous at the points where $Y = X'\theta^q$. Thus, we treat these two estimation approaches as different estimators and show their asymptotic behavior separately.

2.2.2. Asymptotic Properties

In this section, we present the asymptotic properties of the M- and Z-estimator of the regression parameters. Consistency and asymptotic normality hold under the following set of weak regularity conditions, which are natural for this regression framework.

Assumption 2.2.2 (Regularity Conditions).

($\mathcal{A}$-1) The data (Y_i, X_i) for $i = 1, \dots, n$ is an iid series of random variables, distributed such as (Y, X) given above. Furthermore, the conditional distribution $F_{Y|X}$ has finite second moments and is absolutely continuous with probability density function $f_{Y|X}$, which is strictly positive, continuous and bounded in a neighbourhood of the true conditional quantile, $X'\theta_0^q$.

($\mathcal{A}$-2) The matrix $\mathbb{E}\left[XX'\right]$ is positive definite.

($\mathcal{A}$-3) The functions $\rho(Y, X, \theta)$ and $\psi(Y, X, \theta)$ are given as in (2.2) and (2.4), where the function G_1 is twice continuously differentiable, $\mathcal{G}_2$ is three times continuously differentiable, $\mathcal{G}_2^{(1)} = G_2$, G_2 and $G_2^{(1)}$ are strictly positive, G_1 is increasing and a and G_1 are integrable.

Remark 2.2.3 (Finite Moment Conditions). We further have to assume that certain moments of X are finite. For the sake of space, we specify the Finite Moment Conditions ($\mathcal{M}$-1) - ($\mathcal{M}$-4) in Appendix 2.A. Note that these general moment conditions simplify substantially for sensible choices of the specification functions G_1 and $\mathcal{G}_2$ as further outlined in Section 2.2.3.

Assumption ($\mathcal{A}$-1) is a combination of typical regularity conditions of mean and quantile regression. Absolute continuity of $F_{Y|X}$ with a strictly positive, bounded and continuous density function in a neighborhood of the true conditional quantile is also imposed for the asymptotic theory of quantile regression. Existence of the conditional moments of Y given X is subject to the conditions of mean regression and is included in our regularity conditions since ES is a truncated mean. The positive definiteness (full rank condition) in ($\mathcal{A}$-2) is common for any regression design with stochastic regressors in order to exclude perfect multicollinearity of the regressors. The conditions for the specification functions G_1 and $\mathcal{G}_2$ in ($\mathcal{A}$-3) mainly originate from the conditions for the joint elicitability of the quantile and ES in Fissler and Ziegel (2016). Differentiability of these functions is required in this setup for obtaining the identification functions and for the differentiations in the computation of the asymptotic covariance in Theorem 2.2.6 and Theorem 2.2.7. The existence of certain moments of the explanatory variables as in conditions ($\mathcal{M}$-1) - ($\mathcal{M}$-4) in Appendix 2.A is also standard in any regression design relying on stochastic regressors. Even though compactness of the parameter space Θ in Assumption 2.2.1 generally simplifies the proofs, in this setup it is crucial for consistency of the Z-estimator as the identification functions ψ_2 are redescending to zero for many reasonable choices of the G_2 function such as e.g. the choices resulting in positively homogeneous loss functions. For details on this, we refer to Section 2.3.1.

Theorem 2.2.4. Assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-1) in Appendix 2.A hold true. Then, for every sequence $\hat{\theta}_{\psi,n} \in \Theta$ satisfying $\frac{1}{n} \sum_{i=1}^{n} \psi(Y_i, X_i, \hat{\theta}_{\psi,n}) \xrightarrow{\mathbb{P}} 0$, it holds that $\hat{\theta}_{\psi,n} \xrightarrow{\mathbb{P}} \theta_0$.

The proof of the Theorem is given in Appendix 2.B.

Theorem 2.2.5. Assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-2) in Appendix 2.A hold true. Then, for the M-estimator defined in (2.3), it holds that $\hat{\theta}_{\rho,n} \xrightarrow{\mathbb{P}} \theta_0$.

The proof of the Theorem is given in Appendix 2.B. For the validity of this Theorem, it remains to assume that $\hat{\theta}_{\rho,n}$ *nearly* minimizes the loss function ρ in the sense that $\frac{1}{n}\sum_{i=1}^{n}\rho(Y_i, X_i, \hat{\theta}_{\rho,n}) \leq \frac{1}{n}\sum_{i=1}^{n}\rho(Y_i, X_i, \theta_0) + o_P(1)$ rather than the strict definition of the M-estimator in (2.3).

Theorem 2.2.6. Assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-3) in Appendix 2.A hold true. Then, for every sequence $\hat{\theta}_{\psi,n} \in \Theta$ satisfying $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(Y_i, X_i, \hat{\theta}_{\psi,n}) \xrightarrow{\mathbb{P}} 0$, it holds that

$$\sqrt{n}\big(\hat{\theta}_{\psi,n} - \theta_0\big) \xrightarrow{d} \mathcal{N}\left(0, \Lambda^{-1}C\Lambda^{-1}\right), \tag{2.6}$$

with

$$\Lambda = \begin{pmatrix}\Lambda_{11} & 0 \\ 0 & \Lambda_{22}\end{pmatrix} \qquad \text{and} \qquad C = \begin{pmatrix}C_{11} & C_{12} \\ C_{21} & C_{22}\end{pmatrix}, \tag{2.7}$$

where

$$\Lambda_{11} = \frac{1}{\alpha}\mathbb{E}\left[(XX')f_{Y|X}(X'\theta_0^q)\big(\alpha G_1^{(1)}(X'\theta_0^q) + G_2(X'\theta_0^e)\big)\right], \tag{2.8}$$

$$\Lambda_{22} = \mathbb{E}\big[(XX')G_2^{(1)}(X'\theta_0^e)\big], \tag{2.9}$$

$$C_{11} = \frac{1-\alpha}{\alpha}\mathbb{E}\left[(XX')\big(\alpha G_1^{(1)}(X'\theta_0^q) + G_2(X'\theta_0^e)\big)^2\right], \tag{2.10}$$

$$C_{12} = C_{21} = \frac{1-\alpha}{\alpha}\mathbb{E}\left[(XX')\big(X'\theta_0^q - X'\theta_0^e\big)\big(\alpha G_1^{(1)}(X'\theta_0^q) + G_2(X'\theta_0^e)\big)G_2^{(1)}(X'\theta_0^e)\right], \tag{2.11}$$

$$C_{22} = \mathbb{E}\left[(XX')\big(G_2^{(1)}(X'\theta_0^e)\big)^2\left(\frac{1}{\alpha}\operatorname{Var}\big(Y - X'\theta_0^q \big| Y \leq X'\theta_0^q, X\big) + \frac{1-\alpha}{\alpha}\big(X'\theta_0^q - X'\theta_0^e\big)^2\right)\right]. \tag{2.12}$$

The proof of the Theorem is given in Appendix 2.B. For asymptotic normality of the Z-estimator, we have to strengthen the condition (2.5) to $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(Y_i, X_i, \hat{\theta}_{\psi,n}) \xrightarrow{\mathbb{P}} 0$.

Theorem 2.2.7. Assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-4) in Appendix 2.A hold true. Then, for the M-estimator defined in (2.3), it holds that

$$\sqrt{n}\big(\hat{\theta}_{\rho,n} - \theta_0\big) \xrightarrow{d} \mathcal{N}\big(0, \Lambda^{-1}C\Lambda^{-1}\big), \tag{2.13}$$

where the matrices Λ and C are given as in Theorem 2.2.6.

The proof of the Theorem is given in Appendix 2.B. Similar to the consistency statement of Theorem 2.2.5, it is possible to relax the minimization condition (2.3) to some *near* minimization. However, for this theorem it is required that the sequence $\hat{\theta}_{\rho,n}$ is such that $\frac{1}{n}\sum_{i=1}^{n}\rho(Y_i, X_i, \hat{\theta}_{\rho,n}) \leq \inf_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\rho(Y_i, X_i, \theta) + o_P(n^{-1})$, which is a stronger condition as required for Theorem 2.2.5.

Remark 2.2.8 (Quantile Regression). Notice that the asymptotic covariance matrix of the quantile-specific parameter estimates $\hat{\theta}^q$ is given by $\alpha(1-\alpha)D_1^{-1}D_0D_1^{-1}$, where

$$D_1 = \mathbb{E}\left[(XX')f_{Y|X}(X'\theta_0^q)\big(\alpha G_1^{(1)}(X'\theta_0^q) + G_2(X'\theta_0^e)\big)\right] \qquad \text{and} \tag{2.14}$$

$$D_0 = \mathbb{E}\left[(XX')\big(\alpha G_1^{(1)}(X'\theta_0^q) + G_2(X'\theta_0^e)\big)^2\right]. \tag{2.15}$$

This simplifies to the covariance matrix of quantile regression parameter estimates by setting $G_1(z) = z$ and $G_2(z) = 0$, which means ignoring the ES-specific part of our loss function and identification functions. This demonstrates that the quantile regression method is nested in our regression procedure, also in terms of its asymptotic distribution.

Remark 2.2.9 (Asymptotic Covariance of the ES and the Oracle Estimator). The ES-specific part of the asymptotic covariance is mainly governed by the term C_{22}, which depends on the quantity

$$\frac{1}{\alpha}\operatorname{Var}\left(Y - X'\theta_0^q \middle| Y \leq X'\theta_0^q, X\right) + \frac{1-\alpha}{\alpha}\left(X'\theta_0^q - X'\theta_0^e\right)^2 = \frac{1}{\alpha^2}\operatorname{Var}\left((Y - X'\theta_0^q)\mathbb{1}_{\{Y\leq X'\theta_0^q\}}\middle| X\right). \tag{2.16}$$

It is reasonable that the asymptotic covariance of ES regression parameters depends on the truncated variance of Y given X as the asymptomatic covariance of mean regression parameters is driven by the conditional (non-truncated) variance of Y given X. The second term $\left(X'\theta_0^q - X'\theta_0^e\right)^2$ in (2.16) is included since the ES represents a truncated mean where the truncation point itself is a statistical functional (the quantile). In comparison, we consider an oracle M-estimator for the ES-specific regression parameters θ^e, given by the loss function

$$\rho_{\text{Oracle}}(Y, X, \theta^e) = (Y - X'\theta^e)^2\mathbb{1}_{\{Y\leq X'\theta_0^q\}}, \tag{2.17}$$

where we assume that the true quantile regression parameters θ_0^q are known. The resulting asymptotic covariance is given by

$$\operatorname{AVar}\left(\widehat{\theta}^e_{\text{Oracle}}\right) = \frac{1}{\alpha}\mathbb{E}\left[XX'\right]^{-1}\cdot\mathbb{E}\left[(XX')\operatorname{Var}\left(Y - X'\theta_0^e\middle| Y \leq X'\theta_0^q, X\right)\right]\cdot\mathbb{E}\left[XX'\right]^{-1}, \tag{2.18}$$

which shows that the additional term $\big(X'\theta_0^q - X'\theta_0^e\big)^2$ is not included for this estimator with fixed truncation point $X'\theta_0^q$.

Remark 2.2.10 (Joint Estimation of the Sample Quantile and ES). We can use this regression framework to jointly estimate the quantile and ES of an identically distributed sample $Y_1, \dots, Y_n$ by regressing on a constant only. The asymptotic covariance matrix given in Theorem 2.2.6 and Theorem 2.2.7 then simplifies to Σ with components

$$\Sigma_{11} = \frac{\alpha(1-\alpha)}{f_Y^2(\theta_0^q)}, \tag{2.19}$$

$$\Sigma_{12} = \Sigma_{21} = (1-\alpha)\frac{\theta_0^q - \theta_0^e}{f_Y(\theta_0^q)}, \tag{2.20}$$

$$\Sigma_{22} = \frac{1}{\alpha}\operatorname{Var}(Y - \theta_0^q | Y \le \theta_0^q) + \frac{1-\alpha}{\alpha}(\theta_0^q - \theta_0^e)^2, \tag{2.21}$$

where θ_0^q and θ_0^e are the true quantile and ES of Y. The same result is obtained by Zwingmann and Holzmann (2016), who further allow for a distribution function for Y which is not differentiable at the quantile with strictly positive derivative. Notice that in this simplified case without covariates, the asymptotic covariance matrix is independent of the specification functions G_1 and $\mathcal{G}_2$ used in the loss and identification functions. Furthermore, (2.19) implies that quantile estimates stemming from our joint estimation procedure have the same asymptotic efficiency as quantile estimates stemming from minimizing the generalized piecewise linear loss (Gneiting, 2011) and as sample quantiles (cf. Koenker (2005)). The same holds true for the efficiency of the sample ES estimators (based on the sample quantile) of Brazauskas et al. (2008) and Chen (2008).

Remark 2.2.11 (Pseudo-R^2 and the choice of $a(Y)$). By choosing $a(Y) = \alpha G_1(Y) + \mathcal{G}_2(Y)$ in (2.2), we can guarantee non-negative losses $\rho(Y, X, \theta) \ge 0$. This choice enables us to define a pseudo-R^2 for our joint regression framework in the sense of Koenker and Machado (1999),

$$R^{QE} = 1 - \frac{\rho(Y, X, \hat{\theta})}{\rho(Y, X, \tilde{\theta})}, \tag{2.22}$$

where $\hat{\theta}$ denotes the parameter estimates of the full regression model and $\tilde{\theta}$ denotes the parameter estimates of a regression model restricted to an intercept term only. However, this choice of $a(Y)$ comes at the cost of more restrictive moment conditions, since we need to impose that $\mathbb{E}\big[G_1(Y) + \mathcal{G}_2(Y)\big] < \infty$.

2.2.3. Choice of the Specification Functions

The loss and identification functions given in (2.2) and (2.4) depend on two specification functions, G_1 and $\mathcal{G}_2$ (with derivative G_2), which have to fulfill the regularity conditions ($\mathcal{A}$-3) in Assumption 2.2.2. Fissler et al. (2016) already mention the feasible choices $G_1(z) = 0$, $G_1(z) = z$, $G_2(z) = \exp(z)$ and $G_2(z) = \exp(z)/\big(1 + \exp(z)\big)$ in order to show that this class is non-empty. In contrast to the loss functions of mean, quantile and expectile regression, there is no natural choice for these specification functions for the quantile and ES yet (Nolde and Ziegel, 2017). However, as the choice of these functions strongly influences the performance of our regression procedure in terms of its asymptotic efficiency, the necessary moment conditions of the regressors and the numerical performance of the optimization algorithm, we discuss sensible selection criteria in the following.

Efron (1991) and Nolde and Ziegel (2017) argue that for M-estimation of regression parameters it is crucial that the utilized loss function is positively homogeneous of some order $b \in \mathbb{R}$ in the sense that

$$\rho(cY, X, c\theta) = c^b \rho(Y, X, \theta) \tag{2.23}$$

for all $c > 0$. This is an important property for loss functions since the ordering of the losses should be independent of the unit of measurement, e.g. the currency we measure the prices and risk forecasts with. Loss functions following this property guarantee that we can change the scaling and still obtain the same optima and consequently the same parameter estimates. For the pair consisting of the quantile and the ES, Nolde and Ziegel (2017) characterize the full class of positively homogeneous[2] loss functions of order b for the case where we restrict the domain of $\mathcal{G}_2$, i.e. the conditional ES to the negative real line[3],

$$b < 0: \quad G_1(z) = -c_0, \qquad \mathcal{G}_2(z) = c_1(-z)^b + c_0, \tag{2.24}$$

$$b = 0: \quad G_1(z) = d_0 \mathbb{1}_{\{z \le 0\}} + d_0' \mathbb{1}_{\{z > 0\}}, \qquad \mathcal{G}_2(z) = -c_1 \log(-z) + c_0, \tag{2.25}$$

$$b \in (0,1): \quad G_1(z) = \big(d_1 \mathbb{1}_{\{z \le 0\}} + d_1' \mathbb{1}_{\{z > 0\}}\big)|z|^b - c_0, \qquad \mathcal{G}_2(z) = -c_1(-z)^b + c_0, \tag{2.26}$$

for some constants $c_0, d_0, d_0' \in \mathbb{R}$ with $d_0 \le d_0'$, $d_1, d_1' \ge 0$ and $c_1 > 0$. There are no positively homogeneous loss functions for the cases $b \ge 1$. Our numerical simulations show that there is no gain in efficiency or numerical accuracy by deviating from the choice $G_1(z) = 0$ (see

---

[2] For $b = 0$, only the loss differences are positively homogeneous. However, the ordering of the losses is still unaffected under this slightly weaker property.

[3] Since the conditional ES of financial assets for small probability levels is always negative, this is no critical restriction. However, for the numerical parameter estimation, we have to restrict the parameter space Θ such that $X_i'\theta^e < 0$ for all $\theta \in \Theta$ and for all X_i in the underlying sample. For details on this, we refer to Section 2.3.1.

also Fissler et al. (2016), Nolde and Ziegel (2017), and Ziegel et al. (2017)), which is also consistent with the homogeneity result. Consequently, we use $G_1(z) = 0$ in the following.

A different natural guiding principle for selecting the specification functions is induced by choosing $\mathcal{G}_2$ (and G_1) such that the moment conditions ($\mathcal{M}$-1) - ($\mathcal{M}$-4) in Appendix 2.A are as least restrictive and as parsimonious as possible. For instance, choosing $\mathcal{G}_2$ such that G_2 and its first and second derivatives are bounded functions (and $G_1(z) = 0$) results in the moment condition $\mathbb{E}\left[||X||^5 + ||X||^4\mathbb{E}\big[|Y|\big|X\big] + ||X||^3\mathbb{E}\big[Y^2\big|X\big] + |a(Y)|\right] < \infty$. This motivates the usage of bounded functions[4] for G_2 such as e.g. the second example of Fissler et al. (2016), $G_2(z) = \exp(z)/\big(1 + \exp(z)\big)$, which is the distribution function of the standard logistic distribution. Further examples of bounded G_2 functions include the distribution functions of absolutely continuous distributions on the real line. In the simulation study in Section 2.4.2, we compare the performance of different specification functions in terms of the mean squared error and the asymptotic efficiency of the estimator.

2.3. Numerical Estimation of the Model

In this section, we discuss the difficulties one encounters and the solutions we propose for estimating the joint regression model. Section 2.3.1 illustrates the numerical optimization procedure we employ for estimating the regression parameters and Section 2.3.2 discusses different estimation methods for the covariance matrix of the estimator.

2.3.1. Optimization

Theorem 2.2.6 and Theorem 2.2.7 imply that both, M-estimation and Z-estimation of the regression parameters θ have the same asymptotic efficiency and consequently, we discuss these estimation approaches in terms of their numerical performance in the following. The numerical implementation of the Z-estimator relies on root-finding of the identification functions given in (2.4), which we implement as in GMM-estimation by minimizing the inner product $\sum_i \psi(Y_i, X_i, \theta)' \cdot \sum_i \psi(Y_i, X_i, \theta)$. However, the identification functions are redescending to zero for many attractive choices of $\mathcal{G}_2$ in the sense that $\psi_2(Y, X, \theta) \longrightarrow 0$ for $X'\theta^e \longrightarrow -\infty$. Consequently, for θ such that $\theta^q = \theta_0^q$ and $X'\theta^e \rightarrow -\infty$, we get the same minimal value of the Z-estimation objective function $\sum_i \psi(Y_i, X_i, \theta)' \cdot \sum_i \psi(Y_i, X_i, \theta)$ as for the true regression parameters θ_0. Thus, the Z-estimator is numerically unstable and diverges in many setups.

Consequently, we rely on M-estimation of the regression parameters in the following. As the loss functions given in (2.2) are not differentiable and non-convex for all applicable

---

[4] Note that the positively homogeneous loss functions exhibit unbounded $\mathcal{G}_2$ functions. However, as the function $\mathcal{G}_2(z)$ does not grow faster than linear as z tends to infinity, the resulting finite moment conditions are not too restrictive.

choices of the specification functions (Fissler, 2017), we apply a derivative-free global optimization technique. More specifically, we use the Iterated Local Search (ILS) meta-heuristic of Lourenço et al. (2003), which successively refines the parameter estimates by repeated optimizations with iteratively perturbed starting values. Our exact implementation consists of the following steps. First, we obtain starting values for θ^q and θ^e from two quantile regressions of Y on X for the probability levels α and $\tilde{\alpha}$, where we choose $\tilde{\alpha}$ such that the $\tilde{\alpha}$-quantile and the α-ES coincide under normality. Second, using these starting values we minimize the loss function with the derivative-free and robust Nelder-Mead Simplex algorithm (Nelder and Mead, 1965). Third, we perturb the resulting parameter estimates by adding normally distributed noise with zero mean and standard deviation equal to the estimated asymptotic standard errors of the initial quantile regression estimates. Fourth, we re-optimize the model with the perturbed parameter estimates as new starting values. If the loss is further decreased by this re-optimization, we update the estimates and otherwise, we retain the previous ones. Fifth, we iterate over the previous two steps until the loss does not decrease in $m = 10$ consecutive iterations. Our numerical experiments indicate that this repeated optimization procedure yields estimates very close to the ones stemming from other global optimization techniques such as e.g. simulated annealing, whereas the major advantage of ILS is the considerably lower computation time.

For the choices of the specification functions which result in positively homogeneous loss functions, we have to restrict the domain of $\mathcal{G}_2$ to the negative real line as already discussed in Section 2.2.3. Thus, we have to restrict Θ such that $X_i'\theta^e < 0$ for all $\theta \in \Theta$ and for all $i = 1, \ldots, n$ during the optimization process. Even though in financial risk management the response variable Y is usually given by financial returns where the true (conditional) ES is strictly negative, there might still be some outliers X_i such that $X_i'\theta_0^e \geq 0$. In such a case, imposing the restriction $X_i'\theta^e < 0$ for all $i = 1, \ldots, n$ during the optimization process generates substantially biased estimates for θ^e. In order to avoid this, we estimate the regression model for the transformed dependent variables $Y - \max(Y)$ for the positively homogeneous loss functions and add $\max(Y)$ to the estimated intercept parameters to undo the transformation[5].

We provide an R package for the estimation of the regression parameters (see Bayer and Dimitriadis, 2017b). This package contains an implementation of both, the M- and the Z-estimator, where different optimization algorithms can be chosen (ILS, simulated annealing).

---

[5] Note that this data transformation changes the average loss function as the applied loss functions are in general not translation invariant. Thus, optimizing the translated loss function can lead to different parameter estimates. However, we do not face the risk of obtaining substantially biased estimates in cases where $X_i'\theta_0^e \geq 0$ for some $i \in \{1, \ldots n\}$. Our numerical experiments indicate that the difference between estimating the model for Y and for $Y - \max(Y)$ is small when $X_i'\theta_0^e < 0$ for all $i \in \{1, \ldots n\}$, but can be quite substantial if there is an outlier for X_i such that $X_i'\theta_0^e \geq 0$.

The package allows for choosing the specification functions G_1 and $\mathcal{G}_2$ and it includes an option to estimate the model either with or without the translation of the dependent variable. Furthermore, the covariance matrix of the parameter estimates can be estimated either by using the asymptotic theory and the resulting techniques we discuss in the next section, or by using the nonparametric iid bootstrap (Efron, 1979). We recommend applying the M-estimator with the ILS algorithm as this procedure exhibits the best performance in our numerical experiments with respect to accuracy, stability and computation times.

2.3.2. Asymptotic Covariance Estimation

While most parts of the asymptotic covariance matrix given in Theorem 2.2.6 and Theorem 2.2.7 are straightforward to estimate, two nuisance quantities impose some difficulties. The first is the density quantile function $f_{Y|X}(X'\theta_0^q)$, which is already well investigated in the quantile regression literature. In particular, we consider the estimators proposed by Koenker (1994), henceforth denoted by *iid* and by Hendricks and Koenker (1992), henceforth denoted by *nid*. The main difference between these is that the first is based on the assumption that the quantile residuals are independent of the covariates, whereas the second allows for a linear dependence structure. Both approaches depend on a bandwidth parameter which we choose according to Hall and Sheather (1988).

The second nuisance quantity is the variance of the quantile residuals, conditional on the covariates and given that these residuals are negative,

$$\operatorname{Var}\left(Y - X'\theta_0^q \middle| Y \leq X'\theta_0^q, X\right) = \operatorname{Var}\left(u^q \middle| u^q \leq 0, X\right). \tag{2.27}$$

Estimation of this quantity is demanding for two reasons. First, for very small probability levels which are typical in financial risk management such as e.g. $\alpha = 2.5\%$, the truncation $u^q \leq 0$ cuts off all but very few (about $\alpha \cdot n$) observations. Second, modeling this truncated variance conditional on the covariates X is challenging, especially considering the very small sample sizes. Under the assumption of homoscedasticity, i.e. that the distribution of u^q is independent of the covariates X, we can simply estimate (2.27) by the sample variance of the negative quantile residuals and we refer to this estimator as *ind* in the following.

We propose two further estimators which allow for a dependence of the quantile residuals on the covariates. For this purpose, we assume a location-scale process with linear[6]

---

[6] This approach can further be generalized by considering more general specifications for the conditional mean and standard deviation. However, our numerical experiments indicate that the estimation accuracy for the asymptotic covariance matrix does not increase by deviating from these linear specifications.

specifications of the conditional mean and standard deviation in order to explicitly model the conditional relationship of u^q on X,

$$u^q = X'\zeta + X'\phi \cdot \varepsilon, \tag{2.28}$$

for some parameter vectors $\zeta, \phi \in \mathbb{R}^k$ and where $\varepsilon \sim G(0,1)$ follows a zero mean, unit variance distribution, such that $u^q|X \sim G\big(X'\zeta, (X'\phi)^2\big)$ with distribution function F_G and density f_G. As we need to estimate the truncated variance of u^q given $u^q \leq 0$, i.e. a truncated variant of $(X'\phi)^2$, one possibility is to estimate (2.28) only for those observations where $u^q \leq 0$. However, this approach particularly suffers from the very few negative quantile residuals as we need to estimate additional parameters compared to the *ind* approach.

We present a feasible alternative by estimating the parameters ζ and ϕ using all available observations of u^q and X by quasi generalized pseudo maximum likelihood (Gouriéroux and Monfort, 1995, Section 8.4.4) and we obtain the truncated conditional variance by the scaling formula $\mathrm{Var}\left(u^q|u^q \leq 0, X\right) = \int_{-\infty}^{0} z^2 h(z)\, dz - \left(\int_{-\infty}^{0} z h(z)\, dz\right)^2$, where $h(z) = f_G(z)/F_G(0)$ is the truncated conditional density of u^q given X and $u^q \leq 0$. We propose one parametric estimator, henceforth denoted by *scl-N*, where we assume that the distribution G is the normal distribution and apply a closed-form solution to the scaling formula. We further propose a semiparametric estimator, henceforth denoted by *scl-sp*, where we estimate the distribution G nonparametrically and then apply the scaling formula for this estimated density by numerical integration.

2.4. Simulation Study

In this section, we investigate the finite sample behavior of the M-estimator and verify the asymptotic properties derived in Section 2.2.2 through simulations. Furthermore, we compare the performance of different choices for the specification functions and evaluate the precision of the different covariance matrix estimators described in Section 2.3.2.

2.4.1. Data Generating Process

In order to assess the numerical properties of estimating the joint regression model, we simulate data from a linear location-scale data generating process (DGP),

$$Y = X'\gamma + (X'\eta) \cdot v, \tag{2.29}$$

where $v \sim F(0,1)$ has zero mean and unit variance, $X = \left(1, X_2, \dots, X_k\right)'$ and $\gamma, \eta \in \mathbb{R}^k$. For this process, the true conditional quantile and ES are linear functions in X, given by

$$Q_\alpha\left(Y|X\right) = X'(\gamma + z_\alpha \eta) \qquad \text{and} \qquad \mathrm{ES}_\alpha\left(Y|X\right) = X'(\gamma + \xi_\alpha \eta), \tag{2.30}$$

where z_α and ξ_α are the quantile and ES of the distribution $F(0,1)$, which implies that $\theta_0^q = \gamma + z_\alpha \eta$ and $\theta_0^e = \gamma + \xi_\alpha \eta$. Furthermore, the conditional distributions of the quantile- and ES-residuals are given by

$$u^q|X \sim F\left(-z_\alpha(X'\eta),\, (X'\eta)^2\right) \qquad \text{and} \qquad u^e|X \sim F\left(-\xi_\alpha(X'\eta),\, (X'\eta)^2\right). \tag{2.31}$$

For the simulation study, we want to assess the performance of our regression procedure in various setups. Thus, we specify γ, η and F in the following such that we get data which is homoscedastic (DGP-(1)) and heteroskedastic (DGP-(2)). Furthermore, we include a regression setup with multiple, correlated regressors and a leptocurtic conditional distribution (DGP-(3)),

$$\begin{aligned}
&\text{DGP-(1)}: \quad X = (1,\, X_2), \qquad X_2 \sim \chi_1^2 \quad \text{and} \quad Y|X \sim \mathcal{N}\big(-X_2,\, 1\big)\\
&\text{DGP-(2)}: \quad X = (1,\, X_2), \qquad X_2 \sim \chi_1^2 \quad \text{and} \quad Y|X \sim \mathcal{N}\big(-X_2,\, (1+0.5X_2)^2\big)\\
&\text{DGP-(3)}: \quad X = (1,\, X_2,\, X_3) \quad X_2,\, X_3 \sim U[0,1] \quad \text{with} \quad \mathrm{corr}(X_2, X_3) = 0.5 \quad \text{and}\\
&\qquad\qquad\quad Y|X \sim t_5\left(X_2 - X_3,\, (1 + X_2 + X_3)^2\right).
\end{aligned}$$

We simulate all three processes 25,000 times with varying sample sizes of $n = 250$, 500, 1000, 2000 and 5000 observations. For each replication and for each of the sample sizes we regress the simulated Y's on the covariates X using our joint regression method for the probability level $\alpha = 2.5\%$.

2.4.2. Comparing the Specification Functions

We start the discussion of the simulation results by investigating the numerical performance of the M-estimator based on different choices of the specification function[7] $\mathcal{G}_2$ used in the loss function in (2.2). We use three natural examples resulting in positively homogeneous

---

[7] Following the reasoning of Section 2.2.3 and Nolde and Ziegel (2017) and Ziegel et al. (2017), we fix $G_1(z) = 0$ throughout the simulation study.

loss functions of order $b = -1$, $b = 0$ and $b = 0.5$ respectively[8], a bounded G_2 function and the (unbounded) exponential function:

$$\begin{aligned} &\mathcal{G}_2(z) = -1/z, \qquad \mathcal{G}_2(z) = -\log(-z), \qquad \mathcal{G}_2(z) = -\sqrt{-z}, \\ &\mathcal{G}_2(z) = \log\big(1 + \exp(z)\big), \qquad \text{and} \qquad \mathcal{G}_2(z) = \exp(z). \end{aligned} \tag{2.32}$$

Figure 2.1 presents the sum (over the $2k$ regression parameters) of the mean squared errors (MSE) of the regression parameters for the three DGPs described above, different sample sizes and for the five choices of the specification functions given in (2.32). As implied by the asymptotic theory, we obtain consistent parameter estimates for all five choices of the specification functions as the MSEs converge to zero for all three DGPs. However, they differ substantially with respect to their small sample properties. The three positively homogeneous specifications result in the most accurate estimates, whereas the choices $\mathcal{G}_2(z) = -\sqrt{-z}$ and $\mathcal{G}_2(z) = -\log(-z)$ tend to perform slightly better than the choice $\mathcal{G}_2(z) = -1/z$. Furthermore, the bounded choice $\mathcal{G}_2(z) = \log\big(1 + \exp(z)\big)$ still performs better than the unbounded exponential function.

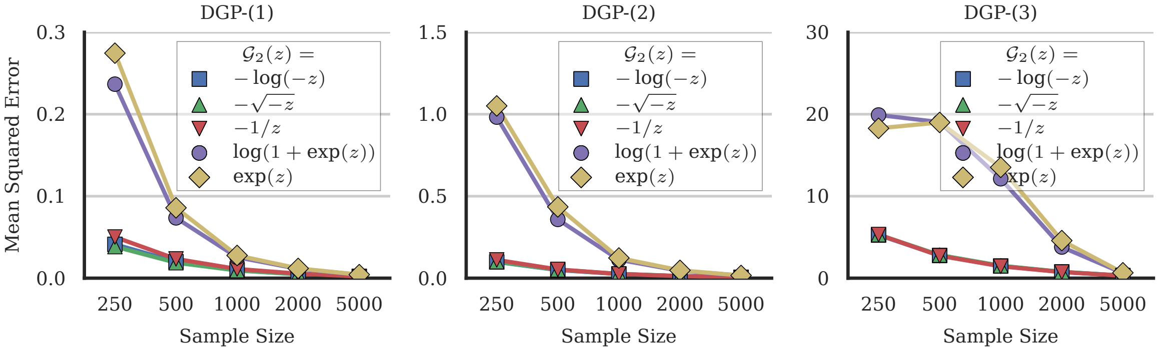


Figure 2.1: Sum of the mean squared errors of the parameter estimates for all three DGPs. The results are shown for the five choices of the specification functions given in (2.32) and a range of sample sizes.

Table 2.1 reports the Frobenius norms of the lower triangular parts of the true asymptotic covariance matrices and of the respective (lower triangular) quantile-specific and the ES-specific sub-matrices for the three DGPs and for the five choices of the specification functions given in (2.32). For comparison, we also report the Frobenius norm of the lower triangular part of the asymptotic covariance of the quantile regression estimator. We approximate the true asymptotic covariance matrix through Monte-Carlo integration with a sample size of 10^9 using the formulas in Theorem 2.2.6 and by using the true density and conditional truncated variance. On average, the specification functions $\mathcal{G}_2(z) = -\log(-z)$ and $\mathcal{G}_2(z) = -\sqrt{-z}$ exhibit the smallest asymptotic covariances, closely followed by the third choice for a

---

[8] Our numerical simulations show that the numerical results are unaffected by different choices of the associated constants in (2.24) - (2.26).

positively homogeneous loss function, $\mathcal{G}_2(z) = -1/z$. The non-homogeneous choices lead to considerably larger asymptotic variances for all considered DGPs and sub-matrices. Furthermore, by comparing the quantile-specific parameters of the joint estimation approach (from the positively homogeneous loss functions) to quantile regression estimates, we roughly obtain the same asymptotic efficiency.

Table 2.1: This table reports the Frobenius norms of the lower triangular parts of the asymptotic covariance matrices and the respective quantile-specific and the ES-specific sub-matrices for the three DGPs and for the five choices of the specification functions given in (2.32). For comparison, we report the same quantity for the asymptotic covariance of the quantile regression estimator.

	DGP-(1)			DGP-(2)			DGP-(3)		
	Q	ES	Full	Q	ES	Full	Q	ES	Full
$\mathcal{G}_2(z) = -\log(-z)$	7.5	13.1	9.2	17.9	26.9	20.0	581.1	1739.1	1053.0
$\mathcal{G}_2(z) = -\sqrt{-z}$	7.0	11.8	8.4	18.0	25.4	19.3	584.5	1740.1	1054.4
$\mathcal{G}_2(z) = -1/z$	9.1	16.9	11.8	24.1	39.4	28.5	613.7	1851.9	1119.8
$\mathcal{G}_2(z) = \log(1 + \exp(z))$	15.4	21.5	16.6	72.4	80.1	67.1	987.9	2393.0	1496.4
$\mathcal{G}_2(z) = \exp(z)$	15.8	22.6	17.2	74.6	84.5	70.0	1001.9	2440.4	1524.6
Quantile Regression	6.8	–	–	21.4	–	–	600.5	–	–

2.4.3. Comparing the Variance-Covariance Estimators

In this section, we compare the empirical performance of the asymptotic covariance estimators discussed in Section 2.3.2. For the comparison of their precision, Figure 2.2 reports the average of the Frobenius norm of the lower triangular part of the differences between the estimated covariances and the empirical covariance of the estimated parameters. We report results for the three homogeneous loss functions and the three DGPs, where each of the plots presents the average norm differences for the four covariance estimators (*iid/nid*, *nid/scl-N*, *nid/scl-sp* and the iid bootstrap) depending on the sample size.

We find that the *iid/nid* estimator performs well for the first, homoscedastic DGP whereas for the other two DGPs, it fails to capture the underlying more complicated dynamics of the data. The *nid/scl-N* estimator outperforms the other estimation approaches in the first two DGPs, where the underlying conditional distribution follows a normal distribution whereas its performance drops for the third DGP, which follows a Student-*t* distribution. The performance of the flexible *nid/scl-sp* estimator is the most stable throughout all three DGPs. Eventually, the bootstrap estimator accurately estimates the covariance for all three DGPs, whereas in comparison to the other estimators, it is particularly good in small samples. The provided R package contains all four covariance estimators.

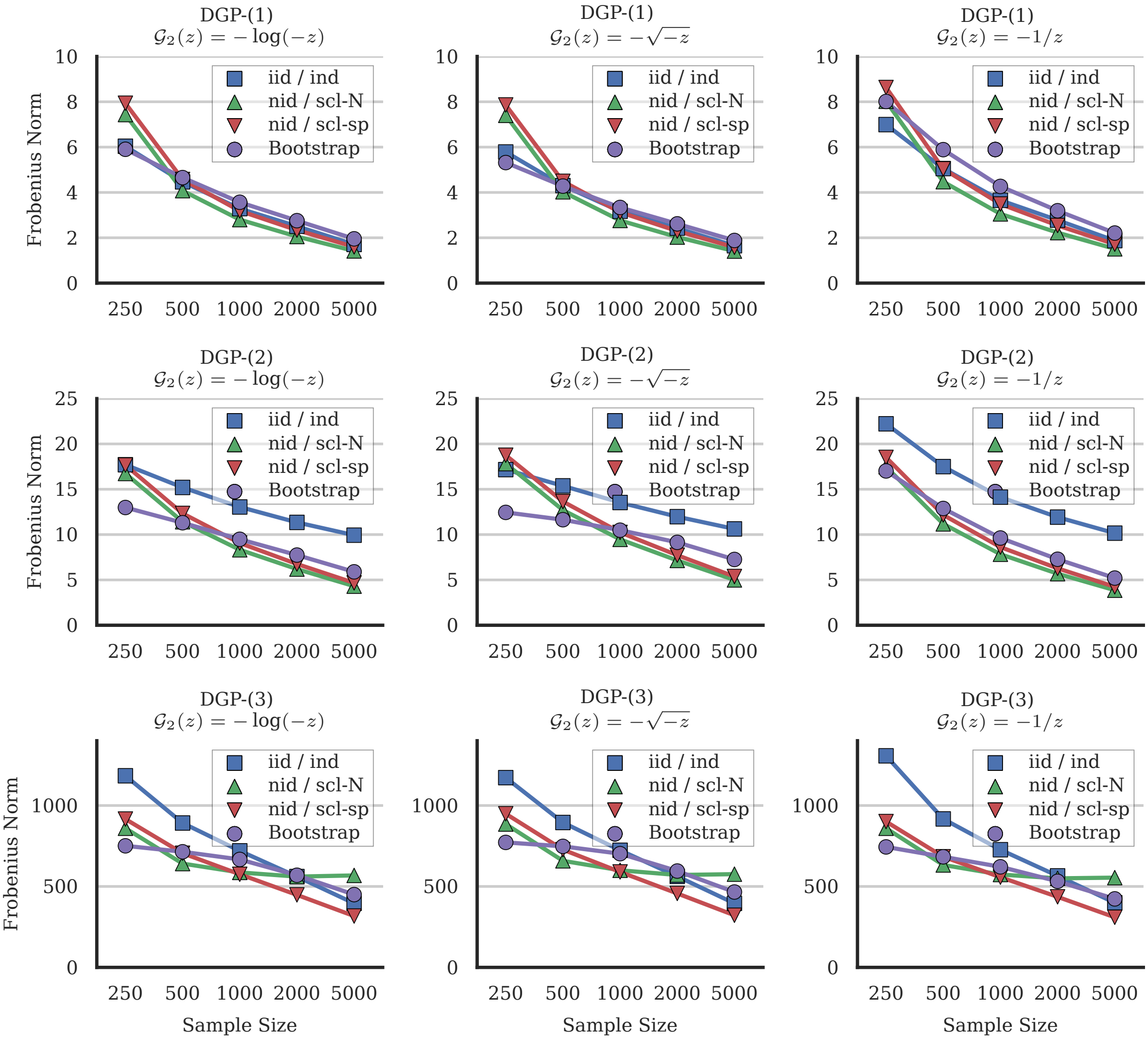


Figure 2.2: This figure compares four covariance estimation approaches described in Section 2.3.2 for the three data generating processes, a range of sample sizes and the three positively homogeneous choices of the $\mathcal{G}_2$-functions. We report the average of the Frobenius norm of the lower triangular part of the differences between the estimated asymptotic covariances and the empirical covariance of the M-estimator.

2.5. Conclusion

In this paper, we introduce a joint regression technique for the quantile (the VaR) and the ES. This regression approach relies on the class of strictly consistent joint loss functions introduced by Fissler and Ziegel (2016), which permits the joint elicitation of the quantile and the ES. We introduce an M- and a Z-estimator for the parameters of the joint regression model. Given a set of standard regularity conditions, we show consistency and asymptotic normality for both estimators, which we also verify numerically through extensive simulations. The underlying loss and identification functions and the asymptotic covariance matrices of the estimators depend on the choice of two specification functions, which we investigate in terms of the resulting moment conditions, asymptotic efficiency, numerical performance and computation times. In our numerical simulations, we find that choices resulting in positively homogeneous loss functions dominate other choices with respect to the aforementioned criteria. Furthermore, we propose several estimation methods for the asymptotic covariance matrix, which are able to cope with different properties of the underlying data. We provide an R package (see Bayer and Dimitriadis, 2017b) that implements the M- and Z-estimation procedures and where one can choose the underlying specification functions, the numerical optimization approach and the estimation method for the asymptotic covariance matrix.

Our new joint regression technique allows for a wide range of applications for the risk measures VaR and ES. This regression approach can be used to model the ES (jointly with the VaR) by generalizing existing applications of quantile regression on VaR, such as e.g. in Koenker and Xiao (2006), Engle and Manganelli (2004), Chernozhukov and Umantsev (2001), Žikeš and Baruník (2016), Halbleib and Pohlmeier (2012), Komunjer (2013) and Xiao et al. (2015). Patton et al. (2017) use this regression for the estimation of autoregressive models for the ES and Bayer and Dimitriadis (2017c) use this regression to develop an ES backtest which is particularly relevant in light of the recent introduction of ES into the Basel regulatory framework and the present lack of accurate backtesting methods for the ES.

Appendix 2.A Finite Moment Conditions

For convenience of the supremum notation, for all $\theta \in \text{int}(\Theta)$ and for $d > 0$, we define the open neighborhood $U_d(\theta) = \{\tau \in \Theta : ||\tau - \theta|| < d\}$ and its closure $\bar{U}_d(\theta) = \{\tau \in \Theta : ||\tau - \theta|| \leq d\}$.

($\mathcal{M}$-1) For Theorem 2.2.4, we assume that the following moments are finite for some $d_0 > 0$:

- $\mathbb{E}[||X||^2 \sup_{\theta \in U_{d_0}(\theta_0)} |G_1^{(1)}(X'\theta^q)|]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in U_{d_0}(\theta_0)} |G_1^{(2)}(X'\theta^q)|]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in U_{d_0}(\theta_0)} |G_2(X'\theta^e)|]$
- $\mathbb{E}[||X||^3 \sup_{\theta \in U_{d_0}(\theta_0)} |G_2^{(1)}(X'\theta^e)|]$
- $\mathbb{E}[||X||^3 \sup_{\theta \in U_{d_0}(\theta_0)} |G_2^{(2)}(X'\theta^e)|]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in U_{d_0}(\theta_0)} |G_2^{(1)}(X'\theta^e)| \; \mathbb{E}[|Y||X]]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in U_{d_0}(\theta_0)} |G_2^{(2)}(X'\theta^e)| \; \mathbb{E}[|Y||X]]$

($\mathcal{M}$-2) For Theorem 2.2.5, we assume that the following moments are finite:

- $\mathbb{E}[||X||^2]$
- $\mathbb{E}[\sup_{\theta \in \Theta} |G_1(X'\theta^q)|]$
- $\mathbb{E}[|G_1(Y)|]$
- $\mathbb{E}[|a(Y)|]$
- $\mathbb{E}[||X|| \sup_{\theta \in \Theta} |G_2(X'\theta^e)|]$
- $\mathbb{E}[\sup_{\theta \in \Theta} |G_2(X'\theta^e)| \; \mathbb{E}[|Y||X]]$
- $\mathbb{E}[\sup_{\theta \in \Theta} |\mathcal{G}_2(X'\theta^e)|]$

($\mathcal{M}$-3) For Theorem 2.2.6, we assume that the following moments are finite for some constant $d_0 > 0$ and for all $\theta \in \bar{U}_{d_0}(\theta_0)$:

- $\mathbb{E}[||X||^3 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_1^{(1)}(X'\tau^q))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_1^{(2)}(X'\tilde{\tau}^q))]$
- $\mathbb{E}[||X||^3 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_1^{(1)}(X'\tau^q))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tilde{\tau}^e))]$
- $\mathbb{E}[||X||^3 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2(X'\tau^e))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_1^{(2)}(X'\tilde{\tau}^q))]$
- $\mathbb{E}[||X||^3 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2(X'\tau^e))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tilde{\tau}^e))]$
- $\mathbb{E}[||X||^3 \sup_{\tau \in \bar{U}_{d_0}(\theta_0)} (G_1^{(1)}(X'\tau^q))^2]$
- $\mathbb{E}[||X||^3 \sup_{\tau \in \bar{U}_{d_0}(\theta_0)} (G_2(X'\tau^e))^2]$
- $\mathbb{E}[||X||^3 \sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_1^{(1)}(X'\tau^q) G_2(X'\tau^e)]$
- $\mathbb{E}[||X||^5 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tau^e))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_2^{(2)}(X'\tilde{\tau}^e))]$
- $\mathbb{E}[||X||^5 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tau^e))^2]$
- $\mathbb{E}[||X||^4 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tau^e))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_2^{(2)}(X'\tilde{\tau}^e)) \mathbb{E}[|Y||X]]$
- $\mathbb{E}[||X||^3 G_2^{(1)}(X'\theta^e)(\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tau^e)) \mathbb{E}[|Y||X]]$
- $\mathbb{E}[||X||^3 G_2^{(1)}(X'\theta^e)(\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2^{(2)}(X'\tau^e)) \mathbb{E}[Y^2|X]]$

- $\mathbb{E}[||X||^3 (\sup_{\tau \in \bar{U}_{d_0}(\theta_0)} G_2^{(1)}(X'\tau^e))(\sup_{\tilde{\tau} \in \bar{U}_{d_0}(\theta_0)} G_2^{(2)}(X'\tilde{\tau}^e))\mathbb{E}[Y^2|X]]$

($\mathcal{M}$-4) For Theorem 2.2.7, we assume that the following moments are finite for some constant $d_0 > 0$:

- $\mathbb{E}[|G_1(Y)|]$
- $\mathbb{E}[|a(Y)|]$
- $\mathbb{E}[||X|| \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} |G_1^{(1)}(X'\theta^q)|]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} (G_1^{(1)}(X'\theta^q))^2]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} |G_1^{(1)}(X'\theta^q) G_2(X'\theta^e)|]$
- $\mathbb{E}[||X|| \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} |G_2(X'\theta^e)|]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} |G_2^{(1)}(X'\theta^e)|]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} (G_2(X'\theta^e))^2]$
- $\mathbb{E}[||X||^4 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} (G_2^{(1)}(X'\theta^e))^2]$
- $\mathbb{E}[||X|| \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} |G_2^{(1)}(X'\theta^e)| \ \mathbb{E}[|Y||X]]$
- $\mathbb{E}[||X||^3 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} (G_2^{(1)}(X'\theta^e))^2 \ \mathbb{E}[|Y||X]]$
- $\mathbb{E}[||X||^2 \sup_{\theta \in \bar{U}_{d_0}(\theta_0)} (G_2^{(1)}(X'\theta^e))^2 \ \mathbb{E}[Y^2|X]]$

Appendix 2.B Proofs

Henceforth, $||v||$ denotes the maximum norm for a vector $v \in \mathbb{R}^k$ and for a matrix A, $||A||$ denotes the row-sum matrix norm which is induced by the maximum norm for vectors. For convenience of the supremum notation, for all $\theta \in \text{int}(\Theta)$ and for some $d > 0$, we define the open neighborhood $U_d(\theta) = \{\tau \in \Theta : ||\tau - \theta|| < d\}$ and its closure $\bar{U}_d(\theta) = \{\tau \in \Theta : ||\tau - \theta|| \leq d\}$.

Proof of Theorem 2.2.4. We apply Theorem 2 from Huber (1967) and show that the function $\psi(Y, X, \theta)$ as given in (2.4) satisfies the respective assumptions of this theorem. Note that the parameter space Θ is assumed to be compact and thus, we do not have to show condition (B-4) in the notation of Huber (1967). As the product of continuous functions and the indicator function $\mathbb{1}_{\{Y \leq X'\theta^q\}}$, the function ψ is measurable and regarded as a stochastic process in θ, ψ is separable in the sense of Doob as it is almost surely continuous in θ (Gikhman and Skorokhod (2004), p.164). This condition assures measurability of the suprema[9] given below and in Lemma 2.B.1.

In oder to show that ψ has a unique root at θ_0, let us first define the sets

$$U_\theta = \left\{\omega \in \Omega \middle| X(\omega)'\theta^q \neq X(\omega)'\theta_0^q\right\}, \qquad \text{and} \qquad W_\theta = \left\{\omega \in \Omega \middle| X(\omega)'\theta^q = X(\omega)'\theta_0^q\right\}, \tag{2.33}$$

---

[9] Many other authors such as e.g. Andrews (1994), Newey and McFadden (1994), and van der Vaart (1998) rely on outer probability in order to avoid these measurability issues.

for all $\theta \in \Theta$ such that $\Omega = W_\theta \cup U_\theta$ and $W_\theta \cap U_\theta = \emptyset$. We first show that $\mathbb{P}(U_\theta) > 0$ for all $\theta \neq \theta_0$. In order to see this, we assume the converse, i.e. let us assume that for a fixed $\theta \neq \theta_0$, it holds that $\mathbb{P}(W_\theta) = \mathbb{P}\big(X'\theta^q = X'\theta_0^q\big) = 1$, which implies that

$$(\theta^q - \theta_0^q)' \, \mathbb{E}[XX'] \, (\theta^q - \theta_0^q) = \mathbb{E}\big[\big(X'\theta^q - X'\theta_0^q\big)^2\big] = 0. \tag{2.34}$$

However, since $\theta^q \neq \theta_0^q$, this contradicts the assumption that the matrix $\mathbb{E}[XX']$ is positive definite and we can conclude that $\mathbb{P}(U_\theta) > 0$.

The quantity

$$\lambda_1(\theta) = \mathbb{E}\big[\psi_1(Y, X, \theta)\big] = 1/\alpha \, \mathbb{E}\left[X\big(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e)\big)\big(F_{Y|X}(X'\theta^q) - F_{Y|X}(X'\theta_0^q)\big)\right]$$

exists under the moment conditions ($\mathcal{M}$-1) in Appendix 2.A and if $\theta^q = \theta_0^q$, it holds that $\lambda_1(\theta) = 0$. Now, we assume that $\theta \in \Theta$ such that $\theta^q \neq \theta_0^q$. By splitting the expectation, we get that

$$\begin{aligned}
&\lambda_1(\theta)'(\theta^q - \theta_0^q) \\
&= 1/\alpha \, \mathbb{E}\left[\big(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e)\big)\big(X'\theta^q - X'\theta_0^q\big)\big(F_{Y|X}(X'\theta^q) - F_{Y|X}(X'\theta_0^q)\big)\mathbb{1}_{\{\omega \in W_\theta\}}\right] \\
&+ 1/\alpha \, \mathbb{E}\left[\big(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e)\big)\big(X'\theta^q - X'\theta_0^q\big)\big(F_{Y|X}(X'\theta^q) - F_{Y|X}(X'\theta_0^q)\big)\mathbb{1}_{\{\omega \in U_\theta\}}\right].
\end{aligned}$$

The first summand is obviously zero since for all $\omega \in W_\theta$, $F_{Y|X}(X'\theta^q) - F_{Y|X}(X'\theta_0^q) = 0$. Since the distribution of Y given X has strictly positive density in a neighbourhood of $X'\theta_0^q$, we get that $F_{Y|X}$ is strictly increasing in a neighbourhood of $X'\theta_0^q$ and thus

$$\big(X'\theta^q - X'\theta_0^q\big)\big(F_{Y|X}(X'\theta^q) - F_{Y|X}(X'\theta_0^q)\big) > 0 \tag{2.35}$$

for all $\omega \in U_\theta$. Furthermore, since $\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e) > 0$ for all $\theta \in \Theta$ and $\mathbb{P}(U_\theta) > 0$, we get that

$$\begin{aligned}
&\lambda_1(\theta)'(\theta^q - \theta_0^q) \\
&= 1/\alpha \, \mathbb{E}\left[\big(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e)\big)\big(X'\theta^q - X'\theta_0^q\big)\big(F_{Y|X}(X'\theta^q) - F_{Y|X}(X'\theta_0^q)\big)\mathbb{1}_{\{\omega \in U_\theta\}}\right] > 0,
\end{aligned}$$

and consequently $\lambda_1(\theta) \neq 0$. This implies that $\lambda_1(\theta) = 0$ if and only if $\theta^q = \theta_0^q$. Furthermore,

$$\lambda_2(\theta) = \mathbb{E}\left[X G_2^{(1)}(X'\theta^e)\,\big(X'\theta^q\,\big(F_{Y|X}(X'\theta^q) - \alpha\big)/\alpha + X'\theta^e - 1/\alpha \, \mathbb{E}\big[Y\mathbb{1}_{\{Y \le X'\theta^q\}}\big|X\big]\big)\right]. \tag{2.36}$$

Assuming that $\theta^q = \theta_0^q$, which results from $\lambda_1(\theta) = 0$, we get that $F_{Y|X}(X'\theta^q) = F_{Y|X}(X'\theta_0^q) = \alpha$ and $1/\alpha\, \mathbb{E}\big[Y \mathbb{1}_{\{Y \le X'\theta_0^q\}} \big| X\big] = X'\theta_0^e$. Thus, (2.36) simplifies to $\mathbb{E}\big[(XX')G_2^{(1)}(X'\theta^e)\big]\big(\theta^e - \theta_0^e\big)$ and by applying Lemma 2.B.2, we get that the matrix $\mathbb{E}\big[(XX')G_2^{(1)}(X'\theta^e)\big]$ is positive definite for all $\theta \in \Theta$. Consequently, $\lambda_2(\theta) = 0$ if and only if $\theta^e = \theta_0^e$ and together with the arguments for λ_1, we get that $\lambda(\theta) = 0$ if and only if $\theta = \theta_0$. Eventually, assumption (B-2)' from Theorem 2 of Huber (1967) follows directly from Lemma 2.B.1, which concludes this proof. □

Proof of Theorem 2.2.5. For this proof, we apply Theorem 5.7 from van der Vaart (1998) and show that the respective assumptions of this theorem hold. As in the proof of Theorem 2.2.6, we can conclude measurability of the suprema since the process ρ is continuous and consequently separable in the sense of Doob. Thus, we do not have to rely on outer probability measures such as in van der Vaart (1998). We start by showing uniform convergence in probability of the empirical mean of the objective function by the help of Lemma 2.4 of Newey and McFadden (1994). Since we have iid data, a compact parameter space Θ and $\rho(Y, X, \theta)$ is continuous for all $\theta \in \Theta$, it remains to show that there exists a dominating function $d(Y, X) \ge |\rho(Y, X, \theta)\big|$ for all $\theta \in \Theta$ with $\mathbb{E}\big[d(Y, X)\big] < \infty$. We define

$$\begin{aligned} d(Y, X) = \sup_{\theta \in \Theta} |G_1(X'\theta^q) + 1/\alpha\, G_2(X'\theta^e)(X'\theta^q - Y)| + \big|G_1(Y)\big| \\ + \sup_{\theta \in \Theta} \big|G_2(X'\theta^e)\big(X'\theta^e - X'\theta^q\big)\big| + \sup_{\theta \in \Theta} |\mathcal{G}_2(X'\theta^e)| + \big|\alpha G_1(Y) + a(Y)\big| \end{aligned} \tag{2.37}$$

and it holds that $d(Y, X) \ge \big|\rho(Y, X, \theta)\big|$ for all $\theta \in \Theta$ and consequently, we can conclude uniform convergence in probability.

We now show that $\mathbb{E}\big[\rho(Y, X, \theta)\big]$ has a unique and global minimum at $\theta = \theta_0$. For this, we assume that $\theta \in \Theta$ such that $\theta \ne \theta_0$ and we define the sets

$$U_\theta = \big\{\omega \in \Omega \big| X(\omega)'\theta^q \ne X(\omega)'\theta_0^q \quad \text{or} \quad X(\omega)'\theta^e \ne X(\omega)'\theta_0^e\big\} \quad \text{and} \tag{2.38}$$

$$W_\theta = \big\{\omega \in \Omega \big| X(\omega)'\theta^q = X(\omega)'\theta_0^q \quad \text{and} \quad X(\omega)'\theta^e = X(\omega)'\theta_0^e\big\}, \tag{2.39}$$

such that $\Omega = U_\theta \cup W_\theta$ and $U_\theta \cap W_\theta = \emptyset$. We first show that $\mathbb{P}(U_\theta) > 0$ for all $\theta \ne \theta_0$. In order to see this, we assume the converse, i.e. we assume that $\mathbb{P}(W_\theta) = 1$, which implies that $(\theta^q - \theta_0^q)'\, \mathbb{E}[XX']\, (\theta^q - \theta_0^q) = \mathbb{E}\left[\big(X'\theta^q - X'\theta_0^q\big)^2\right] = 0$, since $\mathbb{P}\big(X'\theta^q = X\theta_0^q\big) = 1$ and equivalently $(\theta^e - \theta_0^e)'\mathbb{E}[XX'](\theta^e - \theta_0^e) = 0$. However, since $\theta \ne \theta_0$ and consequently either $\theta^q \ne \theta_0^q$ or $\theta^e \ne \theta_0^e$, this contradicts the assumption that the matrix $\mathbb{E}[XX']$ is positive definite and it follows that $\mathbb{P}(U_\theta) > 0$.

From the joint elicitability property of the quantile and ES of Fissler and Ziegel (2016), Corollary 5.5 we get that for all $x \in \mathbb{R}^k$ such that $x'\theta^q \neq x'\theta_0^q$ or $x'\theta^e \neq x'\theta_0^e$, it holds that

$$\mathbb{E}\big[\rho(Y, X, \theta_0)\big|X = x\big] < \mathbb{E}\big[\rho(Y, X, \theta)\big|X = x\big], \tag{2.40}$$

since the distribution of Y given X has a finite first moment and a unique α-quantile. Thus, for all $\omega \in U_\theta$,

$$\mathbb{E}\big[\rho(Y, X, \theta_0)\big|X\big](\omega) < \mathbb{E}\big[\rho(Y, X, \theta)\big|X\big](\omega). \tag{2.41}$$

We now define the random variable

$$h(X, \theta, \theta_0)(\omega) = \mathbb{E}\big[\rho(Y, X, \theta_0)\big|X\big](\omega) - \mathbb{E}\big[\rho(Y, X, \theta)\big|X\big](\omega), \tag{2.42}$$

and (2.41) implies that $h\big(X, \theta, \theta_0\big)(\omega) < 0$ for all $\omega \in U_\theta$. Since $\mathbb{P}(U_\theta) > 0$, this implies that $\mathbb{E}\left[h(X, \theta, \theta_0)\mathbb{1}_{\{\omega \in U_\theta\}}\right] < 0$. Furthermore, for all $\omega \in W_\theta$, it obviously holds that $h(X, \theta, \theta_0)(\omega) = 0$ and consequently $\mathbb{E}\left[h(X, \theta, \theta_0)\mathbb{1}_{\{\omega \in W_\theta\}}\right] = 0$. Thus, we get that

$$\mathbb{E}\big[h(X, \theta, \theta_0)\big] = \mathbb{E}\left[h(X, \theta, \theta_0)\mathbb{1}_{\{\omega \in U_\theta\}}\right] + \mathbb{E}\left[h(X, \theta, \theta_0)\mathbb{1}_{\{\omega \in W_\theta\}}\right] < 0 \tag{2.43}$$

for all $\theta \in \Theta$ such that $\theta \neq \theta_0$, which shows that $\mathbb{E}\big[\rho(Y, X, \theta)\big]$ has a unique minimum at $\theta = \theta_0$. As we define $\hat{\theta}_{\rho,n} = \operatorname{argmin}_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \rho(Y_i, X_i, \theta)$ in (2.3), it obviously holds that $\frac{1}{n} \sum_{i=1}^n \rho(Y_i, X_i, \hat{\theta}_{\rho,n}) \leq \frac{1}{n} \sum_{i=1}^n \rho(Y_i, X_i, \theta_0) + o_P(1)$ which concludes this proof. □

Proof of Theorem 2.2.6. We apply Theorem 3 of Huber (1967) for the ψ-function as given in (2.4) and show the respective assumptions of this theorem. Consistency of the Z-estimator is shown in Theorem 2.2.4. For the measureability and separability of the ψ function, we refer to the proof of Theorem 2.2.4. It is already shown in the proof of Theorem 2.2.4 that there exists a $\theta_0 \in \Theta$ such that $\lambda(\theta_0) = 0$. For the technical conditions (N-3), we apply Lemma 2.B.3, Lemma 2.B.1 and Lemma 2.B.4. It remains to show that $\mathbb{E}\big[||\psi(Y, X, \theta_0)||^2\big] < \infty$, which follows from the subsequent computation of C and the Moment Conditions ($\mathcal{M}$-3) in Appendix 2.A. The asymptotic covariance matrix is given by $\Lambda^{-1}C\Lambda^{-1}$, where $C = \mathbb{E}\big[\psi(Y, X, \theta_0)\,\psi(Y, X, \theta_0)'\big]$ and

$$\Lambda = \frac{\partial \lambda(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial \lambda_1(\theta)}{\partial \theta^q}\Big|_{\theta_0} & \frac{\partial \lambda_1(\theta)}{\partial \theta^e}\Big|_{\theta_0} \\ \frac{\partial \lambda_2(\theta)}{\partial \theta^q}\Big|_{\theta_0} & \frac{\partial \lambda_2(\theta)}{\partial \theta^e}\Big|_{\theta_0} \end{pmatrix}. \tag{2.44}$$

Straightforward calculations yield the matrix C as given in (2.10) - (2.12). For the computation of Λ, we first notice that the function

$$\mathbb{E}\big[\psi(Y,X,\theta)\big|X\big] = \begin{pmatrix} \frac{1}{\alpha}\big(F_{Y|X}(X'\theta^q)-\alpha\big)\big(\alpha G_1^{(1)}(X'\theta^q)+G_2(X'\theta^e)\big) \\ X G_2^{(1)}(X'\theta^e)\Big(X'\theta^e - X'\theta^q + \frac{1}{\alpha}\mathbb{E}\big[(X'\theta^q - Y)\mathbb{1}_{\{Y\le X'\theta^q\}}\big|X\big]\Big) \end{pmatrix} \tag{2.45}$$

is continuously differentiable for all θ in some neighborhood $U_d(\theta_0)$ around θ_0, since the distribution $F_{Y|X}$ has a density which is strictly positive, continuous and bounded in this area. Let us choose a value $\tilde{\theta} \in U_d(\theta_0)$ such that $X'\tilde{\theta} \le X'\theta$. Then,

$$\begin{aligned} \frac{\partial}{\partial\theta^q}\mathbb{E}\big[Y\mathbb{1}_{\{Y\le X'\theta^q\}}\big|X\big] &= \frac{\partial}{\partial\theta^q}\mathbb{E}\big[Y\mathbb{1}_{\{Y\le X'\tilde{\theta}^q\}}\big|X\big] + \frac{\partial}{\partial\theta^q}\mathbb{E}\big[Y\mathbb{1}_{\{X'\tilde{\theta}^q<Y\le X'\theta^q\}}\big|X\big] \\ &= \frac{\partial}{\partial\theta^q}\int_{X'\tilde{\theta}^q}^{X'\theta^q} y f_{Y|X}(y)\mathrm{d}y = X(X'\theta^q)f_{Y|X}(X'\theta^q). \end{aligned} \tag{2.46}$$

We consequently get that for all $\theta \in U_d(\theta_0)$,

$$\begin{aligned} \frac{\partial}{\partial\theta^q}\mathbb{E}\big[\psi_1(Y,X,\theta)\big|X\big] &= 1/\alpha\,(XX')\Big[\big(\alpha G_1^{(1)}(X'\theta^q)+G_2(X'\theta^e)\big)f_{Y|X}(X'\theta^q) \\ &\qquad\qquad +G_1^{(2)}(X'\theta^q)\big(F_{Y|X}(X'\theta^q)-\alpha\big)\Big], \\ \frac{\partial}{\partial\theta^e}\mathbb{E}\big[\psi_1(Y,X,\theta)\big|X\big] &= \frac{\partial}{\partial\theta^q}\mathbb{E}\big[\psi_2(Y,X,\theta)\big|X\big] = 1/\alpha\,(XX')G_2^{(1)}(X'\theta^e)\big(F_{Y|X}(X'\theta^q)-\alpha\big), \\ \frac{\partial}{\partial\theta^e}\mathbb{E}\big[\psi_2(Y,X,\theta)\big|X\big] &= 1/\alpha\,(XX')G_2^{(2)}(X'\theta^e)\,\big(X'\theta^q\big(F_{Y|X}(X'\theta^q)-\alpha\big) \\ &\qquad + \alpha(X'\theta^e) - \mathbb{E}\big[Y\mathbb{1}_{\{Y\le X'\theta^q\}}\big|X\big]\big) + (XX')G_2^{(1)}(X'\theta^e). \end{aligned}$$

In order to conclude that $\frac{\partial}{\partial\theta}\mathbb{E}\big[\mathbb{E}\big[\psi(Y,X,\theta)\big|X\big]\big] = \mathbb{E}\big[\frac{\partial}{\partial\theta}\mathbb{E}\big[\psi(Y,X,\theta)\big|X\big]\big]$, we apply a measure-theoretical version of the Leibniz integration rule, which requires that the derivative of the integrand exists and is absolutely bounded by some integrable function $d(Y,X)$, independent of θ. For the first term, this can easily be obtained by defining

$$\begin{aligned} d(Y,X) = \sup_{\theta\in U_d(\theta_0)} \Big\| 1/\alpha\,(XX')\Big[\big(\alpha G_1^{(1)}(X'\theta^q)+G_2(X'\theta^e)\big)f_{Y|X}(X'\theta^q) \\ +\; G_1^{(2)}(X'\theta^q)\big(F_{Y|X}(X'\theta^q)-\alpha\big)\Big]\Big\|, \end{aligned}$$

which has finite expectation by the Moment Conditions ($\mathcal{M}$-3). The other two terms follow the same reasoning. Inserting $\theta=\theta_0$ eventually shows (2.8) and (2.9). □

Proof of Theorem 2.2.7. For this proof, we apply Theorem 5.23 from van der Vaart (1998) and show that the respective assumptions of this theorem hold. Theorem 2.2.5 shows consistency of the M-estimator. The map $(Y, X) \mapsto \rho(Y, X, \theta)$ is obviously measurable as the sum of measurable functions. Furthermore, the map $\theta \mapsto \rho(Y, X, \theta)$ is almost surely differentiable since the only point of non-differentiability occurs where $Y = X'\theta^q$, which is a nullset with respect to the joint distribution of Y and X and for all $\theta \in \Theta$ such that $Y \neq X'\theta^q$, its derivative is given by $\psi(Y, X, \theta)$. Local Lipschitz continuity with square-integrable Lipschitz-constant follows from Lemma 2.B.5. We have already seen in the proof of Theorem 2.2.5 that the function $\mathbb{E}\big[\rho(Y, X, \theta)\big]$ is uniquely minimized at the point θ_0 and is twice continuously differentiable and consequently admits a second-order Taylor expansion at θ_0. The condition $\frac{1}{n}\sum_{i=1}^n \rho(Y_i, X_i, \hat{\theta}_{\rho,n}) \leq \inf_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^n \rho(Y_i, X_i, \theta) + o_P(n^{-1})$ is obviously fulfilled as the definition of the M-estimator in (2.3) implies that $\frac{1}{n}\sum_{i=1}^n \rho(Y_i, X_i, \hat{\theta}_{\rho,n}) = \inf_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^n \rho(Y_i, X_i, \theta)$ as Θ is compact. Thus, we have shown the necessary assumptions of Theorem 5.23 from van der Vaart (1998).

For the computation of the covariance matrix, we notice that the distribution of Y given X has a density $f_{Y|X}$ in a neighborhood of $X'\theta_0$, which is strictly positive, continuous and bounded. Therefore, by the same arguments as in (2.46), we get that $\frac{\partial}{\partial\theta^q}\mathbb{E}\big[G_1(Y)\mathbb{1}_{\{Y \leq X'\theta^q\}}\big|X\big] = XG_1(X'\theta^q)f_{Y|X}(X'\theta^q)$. Thus, straight-forward calculations yield that for all $\theta \in U_d(\theta_0)$, it holds that $\frac{\partial}{\partial\theta}\mathbb{E}\big[\rho(Y, X, \theta)\big|X\big] = \mathbb{E}\big[\psi(Y, X, \theta)\big|X\big]$ and by applying the Leibniz integration rule such as in the proof of Theorem 2.2.6, we finally get that

$$\frac{\partial}{\partial\theta}\mathbb{E}\big[\rho(Y, X, \theta)\big] = \mathbb{E}\big[\psi(Y, X, \theta)\big]. \tag{2.47}$$

Consequently, the asymptotic covariance matrix equals the one given in Theorem 2.2.6. □

Lemma 2.B.1. Let

$$u(Y, X, \theta, d) = \sup_{\tau\in\bar{U}_d(\theta)} \big\|\psi(Y, X, \tau) - \psi(Y, X, \theta)\big\| \tag{2.48}$$

and assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-1) in Appendix 2.A hold. Then, there are strictly positive real numbers b and d_0, such that

$$\mathbb{E}\big[u(Y, X, \theta, d)\big] \leq b \cdot d \qquad \text{for} \quad ||\theta - \theta_0|| + d \leq d_0, \tag{2.49}$$

and for all $d \geq 0$.

Proof of Lemma 2.B.1. For measurability of the suprema, we refer to the proof of Theorem 2.2.4. Let in the following $d > 0$ and $\theta \in \Theta$ such that $||\theta - \theta_0|| + d \le d_0$. We first notice that for some fixed $X \in \mathbb{R}^k$ and for all $\tau \in \bar{U}_d(\theta)$, it holds that

$$\left|\mathbb{1}_{\{Y \le X'\theta^q\}} - \mathbb{1}_{\{Y \le X'\tau^q\}}\right| \le \mathbb{1}_{\{X'\theta^q_- \le Y \le X'\theta^q_+\}} \tag{2.50}$$

for all $Y \in \mathbb{R}$ and for some $\theta^q_-, \theta^q_+ \in \bar{U}_d(\theta)$. Since $\bar{U}_d(\theta)$ is compact, we get that

$$\sup_{\tau \in \bar{U}_d(\theta)} \left|\mathbb{1}_{\{Y \le X'\theta^q\}} - \mathbb{1}_{\{Y \le X'\tau^q\}}\right| \le \mathbb{1}_{\{X'\theta^q_- \le Y \le X'\theta^q_+\}} \tag{2.51}$$

for all $Y \in \mathbb{R}$ and for some values $\theta^q_-, \theta^q_+ \in \bar{U}_d(\theta)$. Note that the values θ^q_- and θ^q_+ depend on X and θ, however they are independent of Y. Consequently, it holds that

$$\begin{aligned} &\mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left|\mathbb{1}_{\{Y \le X'\theta^q\}} - \mathbb{1}_{\{Y \le X'\tau^q\}}\right| \,\middle|\, X\right] \le \mathbb{E}\left[\mathbb{1}_{\{X'\theta^q_- \le Y \le X'\theta^q_+\}} \,\middle|\, X\right] \\ &= F_{Y|X}\big(X'\theta^q_+\big) - F_{Y|X}\big(X'\theta^q_-\big) = f_{Y|X}(X'\tilde{\theta}^q)\big(X'\theta^q_+ - X'\theta^q_-\big) \\ &\le 2||X|| \cdot \sup_{\tau \in \bar{U}_d(\theta)} f_{Y|X}(X'\tau^q) \cdot d, \end{aligned} \tag{2.52}$$

where we apply the mean value theorem for some $\tilde{\theta}^q$ on the line between θ^q_- and θ^q_+, i.e. $\tilde{\theta}^q \in \bar{U}_d(\theta)$.

For the first component of ψ, we get that

$$\begin{aligned} &\mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\|\psi_1(Y, X, \theta) - \psi_1(Y, X, \tau)\right\|\right] \\ &\le \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\|X\left(G_1^{(1)}(X'\theta^q) - G_1^{(1)}(X'\tau^q) + \frac{G_2(X'\theta^e) - G_2(X'\tau^e)}{\alpha}\right)\right\|\right] \\ &\qquad + \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\|X\left(G_1^{(1)}(X'\tau^q) + \frac{G_2(X'\tau^e)}{\alpha}\right)\right\| \cdot \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left|\mathbb{1}_{\{Y \le X'\theta^q\}} - \mathbb{1}_{\{Y \le X'\tau^q\}}\right| \,\middle|\, X\right]\right]. \end{aligned} \tag{2.53}$$

The first term in (2.53) is $\mathcal{O}(d)$ since $G_1^{(1)}(X'\theta^q)$ and $G_2(X'\theta^e)$ are continuously differentiable functions w.r.t θ and thus, by the mean value theorem we get that

$$\begin{aligned} \sup_{\tau \in \bar{U}_d(\theta)} \left|G_1^{(1)}(X'\theta^q) - G_1^{(1)}(X'\tau^q)\right| &\le \sup_{\tilde{\tau} \in \bar{U}_d(\theta)} \left\|X G_1^{(2)}(X'\tilde{\tau}^q)\right\| \cdot \sup_{\tau \in \bar{U}_d(\theta)} \left\|\theta^q - \tau^q\right\| \\ &\le \sup_{\tilde{\tau} \in \bar{U}_d(\theta)} \left\|X G_1^{(2)}(X'\tilde{\tau}^q)\right\| \cdot d, \end{aligned} \tag{2.54}$$

and the respective moments are finite by assumption. The same arguments hold for the function G_2. For the second term in (2.53), we apply (2.52) and thus get that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left\|X\left(G_1^{(1)}(X'\tau^q)+\frac{G_2(X'\tau^e)}{\alpha}\right)\right\|\cdot\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left|\mathbb{1}_{\{Y\le X'\theta^q\}}-\mathbb{1}_{\{Y\le X'\tau^q\}}\right|\,\middle|\,X\right]\right]\\
\le\;&\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left\|X\left(G_1^{(1)}(X'\tau^q)+\frac{G_2(X'\tau^e)}{\alpha}\right)\right\|\,||X||\cdot\sup_{\tau\in\bar{U}_d(\theta)}f_{Y|X}(X'\tau^q)\right]\cdot d.
\end{aligned}
\tag{2.55}
$$

Since the density $f_{Y|X}$ is bounded in a neighborhood of $X'\theta_0^q$ and the respective moments are finite by assumption, we get that this term is also $\mathcal{O}(d)$.

For the second component of ψ, we get that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left\|\psi_2(Y,X,\theta)-\psi_2(Y,X,\tau)\right\|\right]\\
\le\;&\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left\|X(X'\theta^e-X'\theta^q)G_2^{(1)}(X'\theta^e)-X(X'\tau^e-X'\tau^q)G_2^{(1)}(X'\tau^e)\right\|\right]\\
&+\mathbb{E}\left[\left\|\frac{XG_2^{(1)}(X'\theta^e)X'\theta^q}{\alpha}\right\|\cdot\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left|\left(\mathbb{1}_{\{Y\le X'\theta^q\}}-\mathbb{1}_{\{Y\le X'\tau^q\}}\right)\right|\,\middle|\,X\right]\right]\\
&+\mathbb{E}\left[\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left\|\mathbb{1}_{\{Y\le X'\tau^q\}}\left(\frac{XG_2^{(1)}(X'\theta^e)X'\theta^q}{\alpha}-\frac{XG_2^{(1)}(X'\tau^e)X'\tau^q}{\alpha}\right)\right\|\,\middle|\,X\right]\right]\\
&+\mathbb{E}\left[\left\|\frac{XG_2^{(1)}(X'\theta^e)}{\alpha}\right\|\cdot\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left|Y\left(\mathbb{1}_{\{Y\le X'\theta^q\}}-\mathbb{1}_{\{Y\le X'\tau^q\}}\right)\right|\,\middle|\,X\right]\right]\\
&+\mathbb{E}\left[\mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)}\left\|\frac{Y\mathbb{1}_{\{Y\le X'\tau^q\}}}{\alpha}\left(XG_2^{(1)}(X'\theta^e)-XG_2^{(1)}(X'\tau^e)\right)\right\|\,\middle|\,X\right]\right]\\
=\;&\text{(i)}+\text{(ii)}+\text{(iii)}+\text{(iv)}+\text{(v)}.
\end{aligned}
$$

The first, third and fifth term are linearly bounded by (2.54) since the functions $(X'\theta^e-X'\theta^q)G_2^{(1)}(X'\theta^e)$ and $(X'\theta^q)G_2^{(1)}(X'\theta^e)$ and $G_2^{(1)}(X'\theta^e)$ are continuously differentiable. For the second term, we use the arguments from (2.52). For the fourth term, we use similar arguments as in (2.52), and get that there exist some $\theta^q_-, \theta^q_+ \in \bar{U}_d(\theta)$ and a value $\tilde{\theta}^q$ on the line between θ^q_- and θ^q_+, such that

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{XG_2^{(1)}(X'\theta^e)}{\alpha}\right\| \mathbb{E}\left[\sup_{\tau\in\bar{U}_d(\theta)} \left|Y\left(\mathbb{1}_{\{Y\le X'\theta^q\}} - \mathbb{1}_{\{Y\le X'\tau^q\}}\right)\right| \,\middle|\, X\right]\right] \\
\le\, &\mathbb{E}\left[\left\|\frac{XG_2^{(1)}(X'\theta^e)}{\alpha}\right\| \mathbb{E}\left[|Y|\, \mathbb{1}_{\{X'\theta^q_- \le Y \le X'\theta^q_+\}} \,\middle|\, X\right]\right] \\
=\, &\mathbb{E}\left[\left\|\frac{XG_2^{(1)}(X'\theta^e)}{\alpha}\right\| \int_{X'\theta^q_-}^{X'\theta^q_+} |y| f_{Y|X}(y)\mathrm{d}y\right] \\
\le\, &\mathbb{E}\left[\left\|\frac{XG_2^{(1)}(X'\theta^e)}{\alpha}\right\| |X'\tilde{\theta}^q| f_{Y|X}(X'\tilde{\theta}^q)\big(X'\theta^q_+ - X'\theta^q_-\big)\right] \\
\le\, &\frac{2}{\alpha}\mathbb{E}\left[G_2^{(1)}(X'\theta^e)\big\|X\big\|^2 \sup_{\tau\in\bar{U}_d(\theta)} |X'\tau^q| f_{Y|X}(X'\tau^q)\right] \cdot d = \mathcal{O}(d)
\end{aligned} \tag{2.56}$$

since $f_{Y|X}$ is bounded in a neighborhood of $X'\theta_0$ and the respective moments exist by assumption. This concludes the proof of the lemma. □

Lemma 2.B.2. Let the random variable $X \in \mathbb{R}^k$ with distribution $\mathbb{P}$ be such that its second moments exist and the matrix $\mathbb{E}[XX']$ is positive definite. Furthermore, let $\tilde{\Theta} \subset \mathbb{R}^k$ be a compact subspace with nonempty interior and let $g : \mathbb{R}^k \times \tilde{\Theta} \to \mathbb{R}$ be a strictly positive function. Then, the matrix

$$\mathbb{E}\big[(XX')g(X,\theta)\big] \tag{2.57}$$

is also positive definite.

Proof of Lemma 2.B.2. Since $\mathbb{E}[XX']$ is positive definite, we know that for all $z \in \mathbb{R}^k$ with $z \neq 0$, it holds that $0 < z'\mathbb{E}[XX']z = \mathbb{E}[z'(XX')z] = \mathbb{E}[(X'z)^2]$ and consequently $\mathbb{P}\big(X'z \neq 0\big) > 0$. Since $\sqrt{g(X,\theta)}$ is a strictly positive scalar for all $\theta \in \tilde{\Theta}$, it also holds that $\mathbb{P}\big((X'z)\sqrt{g(X,\theta)} \neq 0\big) > 0$ and thus, for all $z \neq 0$,

$$z'\mathbb{E}\big[(XX')g(X,\theta)\big]z = \mathbb{E}\left[\left(X'z\sqrt{g(X,\theta)}\right)^2\right] > 0. \tag{2.58}$$

This positivity statement holds since $\big(X'z\sqrt{g(X,\theta)}\big)^2$ is a non-negative random variable and $\mathbb{P}\big((X'z)\sqrt{g(X,\theta)} \neq 0\big) > 0$. This shows that the matrix $\mathbb{E}\big[(XX')g(X,\theta)\big]$ is positive definite. □

Lemma 2.B.3. Assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-3) in Appendix 2.A hold. Then, for

$$\lambda(\theta) = \mathbb{E}\big[\psi(Y, X, \theta)\big], \tag{2.59}$$

there are strictly positive numbers a, d_0, such that

$$||\lambda(\theta)|| \geq a \cdot ||\theta - \theta_0|| \qquad \text{for} \quad ||\theta - \theta_0|| \leq d_0. \tag{2.60}$$

Proof of Lemma 2.B.3. Let $d_0 > 0$ and let $||\theta - \theta_0|| \leq d_0$. Then, applying the mean value theorem, we get that

$$\lambda_1(\theta) = \frac{1}{\alpha}\mathbb{E}\left[(XX')\big(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e)\big) f_{Y|X}(X'\tilde{\theta}^q)\right](\theta^q - \theta_0^q) \tag{2.61}$$

for some $\tilde{\theta}^q$ on the line between θ^q and θ_0^q. Similarly, for the second component we get that

$$\begin{aligned}\lambda_2(\theta) = \mathbb{E}&\left[X\frac{G_2^{(1)}(X'\theta^e) f_{Y|X}(X'\tilde{\theta}^q)}{\alpha}\big[X'(\theta^q - \theta_0^q)\big]\big[X'(\tilde{\theta}^q - \theta^q)\big]\right] \\ &+ \mathbb{E}\big[(XX')G_2^{(1)}(X'\theta^e)\big](\theta^e - \theta_0^e),\end{aligned} \tag{2.62}$$

where $\tilde{\theta}^q$ lies on the line between θ^q and θ_0^q.

We first assume that $||\theta - \theta_0|| = ||\theta^q - \theta_0^q||$, i.e. $||\theta^q - \theta_0^q|| \geq ||\theta^e - \theta_0^e||$. Since the matrix

$$A(\theta) := \mathbb{E}\left[(XX')\frac{\big(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e)\big)}{\alpha} f_{Y|X}(X'\tilde{\theta}^q)\right] \tag{2.63}$$

exists and has full rank for all $\theta \in \Theta$ by Lemma 2.B.2 and is obviously symmetric, A has strictly positive real Eigenvalues $\gamma_1(\theta), \dots, \gamma_k(\theta)$ with minimum $\gamma_{(1)}(\theta)$ and we thus get that[10]

$$\begin{aligned}||\lambda(\theta)|| \geq ||\lambda_1(\theta)|| = ||A(\theta)(\theta^q - \theta_0^q)|| &\geq \gamma_{(1)}(\theta) \cdot ||\theta^q - \theta_0^q|| &(2.64)\\ &\geq \left(\inf_{||\theta - \theta_0|| \leq d_0} \gamma_{(1)}(\theta)\right) \cdot ||\theta^q - \theta_0^q|| = c_1\, ||\theta - \theta_0||. &(2.65)\end{aligned}$$

---

[10] For a symmetric matrix A with full rank, we can find an orthogonal basis of Eigenvectors $\{v_1, \dots, v_k\}$ with corresponding nonzero Eigenvalues $\{\gamma_1(\theta), \dots, \gamma_k(\theta)\}$ such that $x = \sum b_j v_j$ with $b_j \in \mathbb{R}$. Then, $||Ax|| = ||A \sum b_j v_j|| = ||\sum b_j A v_j|| = ||\sum b_j \gamma_j v_j|| \geq \min|\gamma_j| \cdot ||\sum b_j v_j|| = \min|\gamma_j| \cdot ||x||$.

Since $||\theta - \theta_0|| \leq d_0$ is a compact set and the function $\theta \mapsto \inf_{||\theta-\theta_0||\leq d_0} \gamma_{(1)}(\theta)$, where $\gamma_{(1)}(\theta)$ is the smallest Eigenvalue of the matrix $A(\theta)$, is continuous[11], we get that the infimum coincides with the minimum and thus, the constant $c_1 := \inf_{||\theta-\theta_0||\leq d_0} \gamma_{(1)}(\theta)$ is strictly positive and does not depend on θ.

Now, we assume that $||\theta - \theta_0|| = ||\theta^e - \theta^e_0|| \leq d_0$, i.e. $||\theta^e - \theta^e_0|| \geq ||\theta^q - \theta^q_0||$. For the first term of $\lambda_2(\theta)$, given in (2.62), we define the vector

$$b(\theta) := \mathbb{E}\left[X \frac{G_2^{(1)}(X'\theta^e) f_{Y|X}(X'\tilde{\theta}^q)}{\alpha} \left[X'(\theta^q - \theta^q_0)\right]\left[X'\tilde{\theta}^q - X'\theta^q\right]\right], \tag{2.66}$$

and for its l-th component, we get that

$$\begin{aligned} |b_l(\theta)| &= \left|\sum_{i,j} (\theta^q_i - \theta^q_{0i})(\tilde{\theta}^q_j - \theta^q_j)\mathbb{E}\left[X_i X_j X_l \frac{G_2^{(1)}(X'\theta^e) f_{Y|X}(X'\tilde{\theta}^q)}{\alpha}\right]\right| \\ &\leq \sum_{i,j} \mathbb{E}\left[\left|X_i X_j X_l \frac{G_2^{(1)}(X'\theta^e) f_{Y|X}(X'\tilde{\theta}^q)}{\alpha}\right|\right] \cdot |\theta^q_i - \theta^q_{0i}| \cdot |\tilde{\theta}^q_j - \theta^q_j| \\ &\leq c_2 \sum_{i,j} |\theta^q_i - \theta^q_{0i}| \cdot |\tilde{\theta}^q_j - \theta^q_j| \\ &\leq c_2 k^2 ||\theta - \theta_0||^2, \end{aligned} \tag{2.67}$$

for all $l = 1, \dots, k$, which implies that

$$||b(\theta)|| \leq c_3 ||\theta - \theta_0||^2, \tag{2.68}$$

for some $c_3 > 0$. For $D(\theta) := \mathbb{E}\big[(XX')G_2^{(1)}(X'\theta^e)\big]$, it holds that $||D(\theta)(\theta^e - \theta^e_0)|| \geq c_4||\theta^e - \theta^e_0|| = c_4||\theta - \theta_0||$ for $c_4 > 0$ by the same arguments as in (2.64). From (2.67), we can choose d_0 small enough such that

$$2||b(\theta)|| \leq 2c_3||\theta - \theta_0||^2 \leq c_4||\theta - \theta_0|| \leq ||D(\theta)(\theta^e - \theta^e_0)||. \tag{2.69}$$

---

[11] This follows since the entries of the matrix $A(\theta)$ are continuous in θ as the expectation of a continuous function which is dominated by an integrable function is again continuous by the dominated convergence theorem. Furthermore, the Eigenvalues of a matrix are the solution of the characteristic polynomial, which has continuous coefficients since our matrix entries are continuous in θ. Eventually, since the roots of any polynomial with continuous coefficients are again continuous, we can conclude that the Eigenvalues of $A(\theta)$ are continuous in θ.

Furthermore, by the submultiplicativity of the matrix norm, we also get that $||D(\theta)(\theta^e - \theta_0^e)|| \leq ||D(\theta)|| \cdot ||\theta^e - \theta_0^e|| = c_5 ||\theta^e - \theta_0^e||$ and by the inverse triangle inequality, we get that

$$||\lambda(\theta)|| \geq ||\lambda_2(\theta)|| = \left\|D(\theta)(\theta^e - \theta_0^e) + b(\theta)\right\| \geq \Big| ||D(\theta)(\theta^e - \theta_0^e)|| - ||b(\theta)|| \Big|. \tag{2.70}$$

From (2.69), we can choose d_0 small enough such that $||D(\theta^e - \theta_0^e)|| > 2||b||$ and thus

$$\Big| ||D(\theta^e - \theta_0^e)|| - ||b|| \Big| = ||D(\theta^e - \theta_0^e)|| - ||b|| \geq \frac{1}{2} ||D(\theta^e - \theta_0^e)|| \tag{2.71}$$

$$\geq \frac{c_4}{2} ||\theta^e - \theta_0^e|| = \frac{c_4}{2} ||\theta - \theta_0||. \tag{2.72}$$

□

Lemma 2.B.4. Let

$$u(Y, X, \theta, d) = \sup_{\tau \in \bar{U}_d(\theta)} \left\| \psi(Y, X, \tau) - \psi(Y, X, \theta) \right\|. \tag{2.73}$$

and assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-3) in Appendix 2.A hold. Then, there are strictly positive numbers c and d_0, such that

$$\mathbb{E}\left[u(Y, X, \theta, d)^2\right] \leq c \cdot d \qquad \text{for} \quad ||\theta - \theta_0|| + d \leq d_0, \tag{2.74}$$

and for all $d \geq 0$.

Proof of Lemma 2.B.4. Let in the following $d > 0$ and $\theta \in \Theta$ such that $||\theta - \theta_0|| + d \leq d_0$. It holds that

$$\left(\sup_{\tau \in \bar{U}_d(\theta)} \left\| \psi(Y, X, \tau) - \psi(Y, X, \theta) \right\| \right)^2 = \sup_{\tau \in \bar{U}_d(\theta)} \left\| \psi(Y, X, \tau) - \psi(Y, X, \theta) \right\|^2 \tag{2.75}$$

and consequently, we show that

$$\mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\| \psi_j(Y, X, \tau) - \psi_j(Y, X, \theta) \right\|^2 \right] = \mathcal{O}(d) \tag{2.76}$$

for both components $j = 1, 2$ and for some $d > 0$ small enough.

For the first squared component, we get that

$$\mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\| \psi_1(Y, X, \tau) - \psi_1(Y, X, \theta) \right\|^2 \right]$$

$$
\begin{aligned}
\leq\ & \max\left(\left| \frac{1-\alpha}{\alpha} \right|^2, 1 \right) \cdot \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\| X\left(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e) - \alpha G_1^{(1)}(X'\tau^q) - G_2(X'\tau^e)\right) \right\|^2 \right] \\
& + \frac{2}{\alpha^2} \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\| X\left(\alpha G_1^{(1)}(X'\tau^q) + G_2(X'\tau^e)\right) \right\|^2 ||X|| \sup_{\tau \in \bar{U}_d(\theta)} f_{Y|X}(X'\tau^q) \right] \cdot d \\
& + \frac{2}{\alpha^2} \max\left(1-\alpha, \alpha\right) \mathbb{E}\Bigg[\sup_{\tau \in \bar{U}_d(\theta)} \left\| X\left(\alpha G_1^{(1)}(X'\theta^q) + G_2(X'\theta^e) - \alpha G_1^{(1)}(X'\tau^q) - G_2(X'\tau^e)\right) \right\| \\
& \qquad \cdot \left\| X\left(\alpha G_1^{(1)}(X'\tau^q) + G_2(X'\tau^e)\right) \right\| \Bigg],
\end{aligned}
$$

where we apply (2.52) for the second summand. The remaining two summands can be bounded linearly by the arguments given in (2.54) since $G_1^{(1)}$ and G_2 are continuously differentiable functions and the respective moments are finite.

For the second component of ψ, we get that

$$
\begin{aligned}
& \left\| \psi_2(Y, X, \tau) - \psi_2(Y, X, \theta) \right\| \\
& \leq \left\| X(X'\theta^e - X'\theta^q) G_2^{(1)}(X'\theta^e) - X(X'\tau^e - X'\tau^q) G_2^{(1)}(X'\tau^e) \right\| \\
& \quad + \left\| \frac{X G_2^{(1)}(X'\theta^e) X'\theta^q}{\alpha} \left(\mathbb{1}_{\{Y \leq X'\theta^q\}} - \mathbb{1}_{\{Y \leq X'\tau^q\}} \right) \right\| \\
& \quad + \left\| \mathbb{1}_{\{Y \leq X'\tau^q\}} \left(\frac{X G_2^{(1)}(X'\theta^e) X'\theta^q}{\alpha} - \frac{X G_2^{(1)}(X'\tau^e) X'\tau^q}{\alpha} \right) \right\| \\
& \quad + \left\| \frac{X G_2^{(1)}(X'\theta^e)}{\alpha} Y \left(\mathbb{1}_{\{Y \leq X'\theta^q\}} - \mathbb{1}_{\{Y \leq X'\tau^q\}} \right) \right\| \\
& \quad + \left\| \frac{Y \mathbb{1}_{\{Y \leq X'\tau^q\}}}{\alpha} \left(X G_2^{(1)}(X'\theta^e) - X G_2^{(1)}(X'\tau^e) \right) \right\| \\
& = \text{(i)} + \text{(ii)} + \text{(iii)} + \text{(iv)} + \text{(v)}.
\end{aligned}
\tag{2.77}
$$

Thus, in order to evaluate $\mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\| \psi_2(Y, X, \tau) - \psi_2(Y, X, \theta) \right\|^2 \right]$, we have to consider all the cross products out of the five summands in (2.77). Since the techniques applied are very similar, we only show details for two of the cross products.

$$
\begin{aligned}
& \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \text{(ii)} \cdot \text{(v)} \right] \\
= & \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left\| \frac{X G_2^{(1)}(X'\theta^e) X'\theta^q}{\alpha} \left(\mathbb{1}_{\{Y \leq X'\theta^q\}} - \mathbb{1}_{\{Y \leq X'\tau^q\}} \right) \right\| \right.
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \left\| \frac{Y \mathbb{1}_{\{Y \leq X'\tau^q\}}}{\alpha} \left(X G_2^{(1)}(X'\theta^e) - X G_2^{(1)}(X'\tau^e) \right) \right\| \Bigg] \\
&\leq \frac{1}{\alpha^2} \mathbb{E}\left[\left\| X G_2^{(1)}(X'\theta^e) X'\theta^q \right\| \cdot \mathbb{E}\big[|Y| \big| X \big] \cdot ||X|| \cdot \sup_{\tau \in \bar{U}_d(\theta)} \left\| G_2^{(1)}(X'\theta^e) - G_2^{(1)}(X'\tau^e) \right\| \right] \\
&\leq \frac{1}{\alpha^2} \mathbb{E}\left[\left\| X G_2^{(1)}(X'\theta^e) X'\theta^q \right\| \cdot \mathbb{E}\big[|Y| \big| X \big] \cdot ||X|| \cdot \sup_{\tau \in \bar{U}_d(\theta)} \left\| X G_2^{(2)}(X'\tau^e) \right\| \right] \cdot d \\
&= \mathcal{O}(d),
\end{aligned}
$$

by (2.54) since $G_2^{(1)}$ is continuously differentiable.

The following crossproducts can be bounded analogously by bounding the indicator functions and by applying the mean value theorem as in (2.54): (i)^2, (iii)^2, (v)^2, (i) $\cdot$ (iii), (i) $\cdot$ (iv), (i) $\cdot$ (v), (ii) $\cdot$ (iv), (ii) $\cdot$ (v), (iii) $\cdot$ (iv), (iii) $\cdot$ (v) and (iv) $\cdot$ (v).

A second type of technique, similar to the arguments in (2.56) arises in the cases (ii)^2, (iv)^2 and (ii) $\cdot$ (iv). We get that there exists $\theta_-^q, \theta_+^q \in \bar{U}_d(\theta)$ and a value $\tilde{\theta}^q$ on the line between θ_-^q and θ_+^q, such that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} (\text{iv})^2 \right] &\leq \mathbb{E}\left[\left\| \frac{X G_2^{(1)}(X'\theta^e)}{\alpha} \right\|^2 \mathbb{E}\left[\sup_{\tau \in \bar{U}_d(\theta)} \left| Y \left(\mathbb{1}_{\{Y \leq X'\theta^q\}} - \mathbb{1}_{\{Y \leq X'\tau^q\}} \right) \right|^2 \middle| X \right] \right] \\
&\leq \mathbb{E}\left[\left\| \frac{X G_2^{(1)}(X'\theta^e)}{\alpha} \right\|^2 \mathbb{E}\left[Y^2 \, \mathbb{1}_{\{X'\theta_-^q \leq Y \leq X'\theta_+^q\}} \middle| X \right] \right] \\
&= \mathbb{E}\left[\left\| \frac{X G_2^{(1)}(X'\theta^e)}{\alpha} \right\|^2 \int_{X'\theta_-^q}^{X'\theta_+^q} y^2 f_{Y|X}(y) \mathrm{d}y \right] \\
&\leq \mathbb{E}\left[\left\| \frac{X G_2^{(1)}(X'\theta^e)}{\alpha} \right\|^2 (X'\tilde{\theta}^q)^2 f_{Y|X}(X'\tilde{\theta}^q) \left(X'\theta_+^q - X'\theta_-^q \right) \right] \\
&\leq \frac{2}{\alpha} \mathbb{E}\left[\left\| X \right\|^3 \left(G_2^{(1)}(X'\theta^e) \right)^2 \cdot \sup_{\tau \in \bar{U}_d(\theta)} (X'\tau^q)^2 f_{Y|X}(X'\tau^q) \right] \cdot d \\
&= \mathcal{O}(d),
\end{aligned}
$$

where we apply a multivariate version of the mean value theorem and notice that $f_{Y|X}$ is bounded. □

Lemma 2.B.5. Assume that Assumption 2.2.1, Assumption 2.2.2 and the Moment Conditions ($\mathcal{M}$-4) in Appendix 2.A hold. Then, the function $\rho(Y, X, \theta)$, given in (2.2) is locally

Lipschitz continuous in θ in the sense that for all $\theta_1, \theta_2 \in U_d(\theta_0)$ in some neighborhood of θ_0, it holds that

$$\left|\rho(Y, X, \theta_1) - \rho(Y, X, \theta_2)\right| \leq K(Y, X) \cdot \left\|\theta_1 - \theta_2\right\|, \tag{2.78}$$

where $\mathbb{E}\left[K(Y, X)^2\right] < \infty$.

Proof. We start the proof by splitting the ρ function into two parts,

$$\rho(Y, X, \theta) = \rho_1(Y, X, \theta) + \rho_2(Y, X, \theta), \tag{2.79}$$

where

$$\rho_1(Y, X, \theta) = \mathbb{1}_{\{Y \leq X'\theta^q\}} \left(G_1(X'\theta^q) - G_1(Y) + \frac{1}{\alpha} G_2(X'\theta^e)(X'\theta^q - Y) \right), \tag{2.80}$$

$$\rho_2(Y, X, \theta) = G_2(X'\theta^e)\big(X'\theta^e - X'\theta^q\big) - \mathcal{G}_2(X'\theta^e) - \alpha G_1(X'\theta^q) + a(Y). \tag{2.81}$$

Local Lipschitz continuity of ρ_2 follows since it is a continuously differentiable function and thus locally Lipschitz. We consequently get that for some $d > 0$ and for all $\theta_1, \theta_2 \in U_d(\theta_0)$, it holds that

$$\left|\rho_2(Y, X, \theta_1) - \rho_2(Y, X, \theta_2)\right| \leq \left\|\theta_1 - \theta_2\right\| \cdot \sup_{\theta \in U_d(\theta_0)} \left\| \begin{pmatrix} -XG_2(X'\theta^e) - \alpha X G_1^{(1)}(X'\theta^q) \\ XG_2^{(1)}(X'\theta^e)\big(X'\theta^e - X'\theta^q\big) \end{pmatrix} \right\|, \tag{2.82}$$

with Lipschitz-constant

$$K(Y, X) = \sup_{\theta \in U_d(\theta_0)} \left\| \begin{pmatrix} -XG_2(X'\theta^e) - \alpha X G_1^{(1)}(X'\theta^q) \\ XG_2^{(1)}(X'\theta^e)\big(X'\theta^e - X'\theta^q\big) \end{pmatrix} \right\|, \tag{2.83}$$

which is square-integrable by the moment conditions ($\mathcal{M}$-4).

For the function ρ_1, we consider three cases. First, let $\theta_1, \theta_2 \in \Theta$ such that $X'\theta_1^q \leq X'\theta_2^q < Y$. Then it holds that,

$$\rho_1(Y, X, \theta_1) = \rho_1(Y, X, \theta_2) = 0, \tag{2.84}$$

since $\mathbb{1}_{\{Y \leq X'\theta_1^q\}} = \mathbb{1}_{\{Y \leq X'\theta_2^q\}} = 0$, which is obviously a Lipschitz continuous function.

Second, let $\theta_1, \theta_2 \in \Theta$ such that $Y \leq X'\theta_1^q \leq X'\theta_2^q$. Then, for $\theta = \theta_1, \theta_2$,

$$\rho_1(Y, X, \theta) = G_1(X'\theta^q) - G_1(Y) + \frac{1}{\alpha} G_2(X'\theta^e)(X'\theta^q - Y), \tag{2.85}$$

which is a continuously differentiable function and thus

$$\left|\rho_1(Y,X,\theta_1)-\rho_1(Y,X,\theta_2)\right| \le \left\|\theta_1-\theta_2\right\| \cdot \sup_{\theta\in U_d(\theta_0)} \left\| \begin{pmatrix} XG_1^{(1)}(X'\theta^q)+\frac{1}{\alpha}XG_2(X'\theta^e) \\ \frac{1}{\alpha}XG_2^{(1)}(X'\theta^e)(X'\theta^q-Y) \end{pmatrix} \right\|. \tag{2.86}$$

Finally, let $\theta_1,\theta_2\in\Theta$ such that $X'\theta_1^q < Y \le X'\theta_2^q$. Then, since G_1 is increasing, we get that

$$\begin{aligned} \left|\rho_1(Y,X,\theta_1)-\rho_1(Y,X,\theta_2)\right| &= \left|G_1(X'\theta_2^q)-G_1(Y)+\frac{1}{\alpha}G_2(X'\theta_2^e)(X'\theta_2^q-Y)\right| \\ &\le \left|G_1(X'\theta_2^q)-G_1(X'\theta_1^q)\right| + \left|\frac{1}{\alpha}G_2(X'\theta_2^e)(X'\theta_2^q-X'\theta_1^q)\right| \\ &\le \left\|\theta_1^q-\theta_2^q\right\| \cdot \sup_{\theta\in U_d(\theta_0)} \left(\left\|XG_1^{(1)}(X'\theta^q)\right\| + \frac{1}{\alpha}\left\|XG_2(X'\theta^e)\right\|\right). \end{aligned}$$

Thus, the function $\rho(Y,X,\theta)$ is locally Lipschitz continuous in θ with square-integrable Lipschitz constants, $\mathbb{E}\left[K(Y,X)^2\right] < \infty$ by the Moment Conditions ($\mathcal{M}$-4) in Appendix 2.A. □

Proposition 2.B.6. Let Y be a real-valued random variable with distribution function F, finite first and second moments and a unique α-quantile $q_\alpha = F^{-1}(\alpha)$. Then,

$$\frac{1}{\alpha^2}\int_{-\infty}^{q_\alpha}\int_{-\infty}^{q_\alpha} F(x\wedge y)-F(x)F(y)\mathrm{d}x\mathrm{d}y = \frac{1}{\alpha}\operatorname{Var}(Y|Y\le q_\alpha)+\frac{1-\alpha}{\alpha}\left(q_\alpha-\xi_\alpha\right)^2, \tag{2.87}$$

where $\xi_\alpha = \mathbb{E}\left[Y\middle|Y\le q_\alpha\right]$ denotes the α-ES of Y.

Proof. We first notice that for a distribution F with finite second moment und unique α-quantile, it holds that

$$\mathbb{E}\left[Y\middle|Y\le q_\alpha\right] = -\frac{1}{\alpha}\int_{-\infty}^{q_\alpha} F(x)\mathrm{d}x + q_\alpha \qquad \text{and} \tag{2.88}$$

$$\mathbb{E}\left[Y^2\middle|Y\le q_\alpha\right] = -\frac{2}{\alpha}\int_{-\infty}^{q_\alpha} xF(x)\mathrm{d}x + q_\alpha^2, \tag{2.89}$$

which can be obtained by using the identity

$$Y\mathbb{1}_{\{Y\le q_\alpha\}} = \mathbb{1}_{\{Y\le q_\alpha\}}\left(\int_0^\infty \mathbb{1}_{\{Y>t\}}\,\mathrm{d}t - \int_{-\infty}^0 \mathbb{1}_{\{Y\le t\}}\,\mathrm{d}t\right) \tag{2.90}$$

and by taking expectations on both sides. By applying (2.88), we get that

$$\int_{-\infty}^{q_\alpha}\int_{-\infty}^{q_\alpha} F(x)F(y)\mathrm{d}x\mathrm{d}y = \left(\int_{-\infty}^{q_\alpha} F(x)\mathrm{d}x\right)^2 = \left(\alpha q_\alpha - \alpha\mathbb{E}\left[Y \middle| Y \le q_\alpha\right]\right)^2 = \alpha^2\left(q_\alpha - \xi_\alpha\right)^2. \tag{2.91}$$

Furthermore, notice that

$$\int_{-\infty}^{q_\alpha}\int_{-\infty}^{q_\alpha} F(x \wedge y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{q_\alpha}\int_{-\infty}^{y} F(x)\mathrm{d}x\mathrm{d}y + \int_{-\infty}^{q_\alpha}\int_{y}^{q_\alpha} F(y)\mathrm{d}x\mathrm{d}y, \tag{2.92}$$

and by rearranging the order of integration for the first term in (2.92), we get that

$$\begin{aligned}\int_{-\infty}^{q_\alpha}\int_{-\infty}^{y} F(x)\,\mathrm{d}x\mathrm{d}y &= \iint\limits_{\{(x,y):\, y\le q_\alpha,\, x\le y\}} F(x)\,\mathrm{d}x\mathrm{d}y = \iint\limits_{\{(x,y):\, x\le q_\alpha,\, y\ge x\}} F(x)\,\mathrm{d}y\mathrm{d}x \\ &= \int_{-\infty}^{q_\alpha}\int_{x}^{q_\alpha} F(x)\,\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{q_\alpha} F(x)(q_\alpha - x)\,\mathrm{d}x.\end{aligned} \tag{2.93}$$

Thus, by first using (2.92) and (2.93) and by plugging in (2.88) and (2.91), we obtain

$$\begin{aligned}\int_{-\infty}^{q_\alpha}\int_{-\infty}^{q_\alpha} F(x \wedge y)\mathrm{d}x\mathrm{d}y &= 2\int_{-\infty}^{q_\alpha}\int_{y}^{q_\alpha} F(y)\,\mathrm{d}x\mathrm{d}y \\ &= 2\int_{-\infty}^{q_\alpha} F(y)(q_\alpha - y)\,\mathrm{d}y \\ &= 2q_\alpha\int_{-\infty}^{q_\alpha} F(y)\,\mathrm{d}y - 2\int_{-\infty}^{q_\alpha} yF(y)\,\mathrm{d}y \\ &= 2q_\alpha\left(\alpha q_\alpha - \alpha\xi_\alpha\right) + \alpha\mathbb{E}\left[Y^2 \middle| Y \le q_\alpha\right] - \alpha q_\alpha^2 \\ &= \alpha\mathbb{E}\left[Y^2 \middle| Y \le q_\alpha\right] + \alpha q_\alpha^2 - 2\alpha q_\alpha\xi_\alpha.\end{aligned} \tag{2.94}$$

Eventually, using (2.91) and (2.94), straight-forward calculations yield that

$$\frac{1}{\alpha^2}\int_{-\infty}^{q_\alpha}\int_{-\infty}^{q_\alpha} F(x \wedge y) - F(x)F(y)\mathrm{d}x\mathrm{d}y = \frac{1}{\alpha}\operatorname{Var}(Y|Y \le q_\alpha) + \frac{1-\alpha}{\alpha}\left(q_\alpha - \xi_\alpha\right)^2, \tag{2.95}$$

which concludes the proof. □

Appendix 2.C Separability of Almost Surely Continuous Functions

Definition 2.C.1 (Separability of a Stochastic Process). A stochastic process $\psi(x,\theta)$: $\Omega \times \Theta \to \mathcal{Y}$ is called separable in the sense of Doob, if there exists in Ω an everywhere dense countable set I, and in Ω a nullset N such that for any arbitrary open set $G \subset \Theta$ and every closed set $F \subset \mathcal{Y}$, the two sets

$$\{x|\psi(x,\theta) \in F,\ \forall \theta \in G\} \qquad \text{and} \tag{2.96}$$

$$\{x|\psi(x,\theta) \in F,\ \forall \theta \in G \cap I\} \tag{2.97}$$

differ from each other at most by a subset of N.

Proposition 2.C.2 (Gikhman and Skorokhod (2004)). Let Θ and $\mathcal{Y}$ be metric spaces, Θ be a separable space. The sets (2.96) and (2.97) coincide for all $x \in \Omega$ for which the stochastic process $\psi(x,\theta)$ is continuous in θ.

Proof. It is clear that $\{x|\psi(x,\theta) \in F,\ \forall \theta \in G\} \subseteq \{x|\psi(x,\theta) \in F,\ \forall \theta \in G \cap I\}$. We thus only show the reverse.

Let $G \subset \Theta$ be an arbitrary open set and $F \subset \mathcal{Y}$ an arbitrary closed set. Let furthermore $x \in \Omega$ such that $\psi(x,\theta) \in F$ for all $\theta \in G \cap I$. We have to show that $\psi(x,\tilde{\theta}) \in F$ for all $\tilde{\theta} \in G$ but $\tilde{\theta} \notin I$.

Thus, let $\tilde{\theta} \in G \setminus I$. Since I is a dense set in Θ, there exists a sequence $(\theta_n)_{n \in \mathbb{N}} \in \Theta \cap I$, such that $\theta_n \longrightarrow \tilde{\theta}$ and since G is an open set in Θ and $\tilde{\theta} \in G$, we can conclude that for $m \in \mathbb{N}$ large enough, $\theta_n \in G$ for all $n \geq m$. Furthermore, by continuity at θ, it holds that $\psi(x,\theta_n) \longrightarrow \psi(x,\tilde{\theta})$ and since $\theta_n \in G \cap I$ for all n large enough, $\psi(x,\theta_n) \in F$ by assumption. Eventually, since F is a closed set, $\psi(x,\tilde{\theta}) \in F$ which proves the proposition. □

Corollary 2.C.3 (Separability of continuous functions). Let Θ and $\mathcal{Y}$ be metric spaces, Θ be a separable space, and let the stochastic process $\psi(x,\theta)$ be almost surely continuous. Then, ψ is separable.

Proof. Since $\psi(x,\theta)$ is continuous for all $x \in \Omega \setminus N$ for some $N \subset \Omega$ with $\mathbb{P}(N) = 0$. We get from Proposition 2.C.2 that the sets (2.96) and (2.97) coincide for all $x \in \Omega \setminus N$, i.e. they differ only by a subset of N. □

References

Andrews, D. (1994). “Empirical Process Methods in Econometrics”. In: *Handbook of Econometrics*. Ed. by Engle, R. and McFadden, D. Vol. 4. Elsevier. Chap. 37, 2247–2294 (see p. 75).

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). “Coherent Measures of Risk”. *Mathematical Finance* 9 (3), 203–228 (see pp. 6, 10, 54, 97).

Barendse, S. (2017). “Interquantile Expectation Regression”. Available at `https://ssrn.com/abstract=2937665` (see pp. 55, 98, 102, 120–123).

Basel Committee (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d352.pdf`. Basel Committee on Banking Supervision (see pp. 6, 11, 15, 26, 54, 97, 105).

Bayer, S. and T. Dimitriadis (2017b). *esreg: Joint (VaR, ES) Regression*. R package version 0.2.0, available at `https://github.com/BayerSe/esreg` (see pp. 56, 66, 73).

— (2017c). “Regression-based Expected Shortfall Backtesting”. Working Paper (see p. 73).

Brazauskas, V., B. L. Jones, M. L. Puri, and R. Zitikis (2008). “Estimating conditional tail expectation with actuarial applications in view”. *Journal of Statistical Planning and Inference* 138 (11), 3590–3604 (see p. 63).

Chen, S. X. (2008). “Nonparametric Estimation of Expected Shortfall”. *Journal of Financial Econometrics* 6 (1), 87–107 (see p. 63).

Chernozhukov, V. and L. Umantsev (2001). “Conditional value-at-risk: Aspects of modeling and estimation”. *Empirical Economics* 26 (1), 271–292 (see pp. 54, 73).

Efron, B. (1979). “Bootstrap Methods: Another Look at the Jackknife”. *The Annals of Statistics* 7 (1), 1–26 (see pp. 67, 103).

— (1991). “Regression percentiles using asymmetric squared error loss”. *Statistica Sinica* 1 (1), 93–125 (see pp. 56, 64, 98, 103, 122).

Engle, R. and S. Manganelli (2004). “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles”. *Journal of Business and Economic Statistics* 22 (4), 367–381 (see pp. 8, 12, 54, 73).

Fissler, T. (2017). “On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space”. PhD thesis. Universität Bern (see p. 66).

Fissler, T. and J. F. Ziegel (2016). “Higher order elicitability and Osband’s principle”. *The Annals of Statistics* 44 (4), 1680–1707 (see pp. 7, 12, 28, 54, 56–58, 60, 72, 78, 97, 119, 121–123).

Fissler, T., J. F. Ziegel, and T. Gneiting (2016). “Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting”. *Risk Magazine* January, 58–61 (see pp. 58, 64, 65, 97, 121).

Gaglianone, W. P., L. R. Lima, O. Linton, and D. R. Smith (2011). “Evaluating Value-at-Risk Models via Quantile Regression”. *Journal of Business & Economic Statistics* 29 (1), 150–160 (see pp. 54, 98, 103, 112, 124).

Gikhman, I. and A. Skorokhod (2004). *The Theory of Stochastic Processes I*. Vol. 210. Classics in Mathematics. Springer Berlin Heidelberg (see pp. 75, 92).

Gneiting, T. (2011). “Making and Evaluating Point Forecasts”. *Journal of the American Statistical Association* 106 (494), 746–762 (see pp. 7, 11, 28, 54–58, 63, 97, 98, 102, 122).

Gouriéroux, C. and A. Monfort (1995). *Statistics and Econometric Models: Volume 1, General Concepts, Estimation, Prediction and Algorithms*. Cambridge University Press (see p. 68).

Halbleib, R. and W. Pohlmeier (2012). “Improving the Value at Risk Forecasts: Theory and Evidence from the Financial Crisis”. *Journal of Economic Dynamics and Control* 36 (8), 1212–1228 (see pp. 15, 28, 54, 73).

Hall, P. and S. J. Sheather (1988). “On the Distribution of a Studentized Quantile”. *Journal of the Royal Statistical Society. Series B (Methodological)* 50 (3), 381–391 (see p. 67).

Hendricks, W. and R. Koenker (1992). “Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity”. *Journal of the American Statistical Association* 87 (417), 58–68 (see p. 67).

Huber, P. (1967). “The behavior of maximum likelihood estimates under nonstandard conditions”. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 221–233 (see pp. 75, 77, 78).

Koenker, R. (1994). “Confidence Intervals for Regression Quantiles”. In: *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium, held from September 4–9, 1993*. Ed. by Mandl, P. and Hušková, M. Physica-Verlag Heidelberg, 349–359 (see p. 67).

— (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press (see pp. 55, 56, 63, 122).

Koenker, R. and J. A. F. Machado (1999). “Goodness of Fit and Related Inference Processes for Quantile Regression”. *Journal of the American Statistical Association* 94 (448), 1296–1310 (see p. 63).

Koenker, R. and Z. Xiao (2006). “Quantile Autoregression”. *Journal of the American Statistical Association* 101 (475), 980–990 (see pp. 54, 73).

Komunjer, I. (2013). “Quantile Prediction”. In: *Handbook of Economic Forecasting*. Vol. 2. Elsevier. Chap. 17, 961–994 (see pp. 54, 73, 100).

Lambert, N. S., D. M. Pennock, and Y. Shoham (2008). “Eliciting Properties of Probability Distributions”. In: *Proceedings of the 9th ACM Conference on Electronic Commerce*. ACM, 129–138 (see p. 56).

Lourenço, H. R., O. C. Martin, and T. Stützle (2003). “Iterated Local Search”. In: *Handbook of Metaheuristics*. Ed. by Glover, F. and Kochenberger, G. A. Boston, MA: Springer US, 320–353 (see p. 66).

Nadarajah, S., B. Zhang, and S. Chan (2014). “Estimation methods for expected shortfall”. *Quantitative Finance* 14 (2), 271–291 (see pp. 55, 100).

Nelder, J. A. and R. Mead (1965). “A Simplex Method for Function Minimization”. *The Computer Journal* 7 (4), 308–313 (see p. 66).

Newey, W. and D. McFadden (1994). “Large sample estimation and hypothesis testing”. In: *Handbook of Econometrics*. Ed. by Engle, R. and McFadden, D. Vol. 4. Elsevier. Chap. 36, 2111–2245 (see pp. 75, 77).

Nolde, N. and J. F. Ziegel (2017). “Elicitability and backtesting: Perspectives for banking regulation”. *Annals of Applied Statistics* 11 (4), 1833–1874 (see pp. 56, 64, 65, 69, 97–99, 105, 107, 108, 110–112, 115, 117–119, 121, 123–125).

Patton, A. J., J. F. Ziegel, and R. Chen (2017). “Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)”. arXiv:1707.05108 [q-fin.EC] (see pp. 55, 73, 98, 102, 103, 120–123).

Taylor, J. W. (2008a). “Estimating Value at Risk and Expected Shortfall Using Expectiles”. *Journal of Financial Econometrics* 6 (2), 231–252 (see p. 55).

— (2008b). “Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall”. *Journal of Financial Econometrics* 6 (3), 382–406 (see p. 55).

— (2017). “Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution”. *Journal of Business & Economic Statistics* 0 (0), 1–13 (see pp. 55, 123).

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press (see pp. 75, 77, 79, 80).

Weber, S. (2006). “Distribution Invariant Risk Measures, Information, and Dynamic Consistency”. *Mathematical Finance* 16 (2), 419–441 (see pp. 57, 97).

Xiao, Z., H. Guo, and M. S. Lam (2015). “Quantile Regression and Value at Risk”. In: *Handbook of Financial Econometrics and Statistics*. Ed. by Lee, C.-F. and Lee, J. C. Springer, 1143–1167 (see pp. 54, 73).

Ziegel, J. F., F. Krüger, A. Jordan, and F. Fasciati (2017). “Murphy Diagrams: Forecast Evaluation of Expected Shortfall”. arXiv:1705.04537 [q-fin.RM] (see pp. 65, 69).

Žikeš, F. and J. Baruník (2016). “Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility”. *Journal of Financial Econometrics* 14 (1), 185–226 (see pp. 8, 12, 54, 73).

Zwingmann, T. and H. Holzmann (2016). “Asymptotics for the expected shortfall”. arXiv:1611.07222 [math.ST] (see p. 63).

Chapter 3

Regression Based Expected Shortfall Backtesting

We thank Tobias Fissler, Lyudmila Grigoryeva, Roxana Halbleib, Phillip Heiler, Ekaterina Kazak, Winfried Pohlmeier, James Taylor, and Johanna Ziegel for suggestions which inspired some results of this paper. Financial support by the Heidelberg Academy of Sciences and Humanities (HAW) within the project “Analyzing, Measuring and Forecasting Financial Risks by means of High-Frequency Data”, by the German Research Foundation (DFG) within the research group “Robust Risk Measures in Real Time Settings” and general support by the Graduate School of Decision Sciences (University of Konstanz) is gratefully acknowledged. The computation in this work was performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC.

3.1. Introduction

Through the transition from Value at Risk (VaR) to Expected Shortfall (ES) as the primary market risk measure in the Basel Accord (Basel Committee, 2016), there is a great demand for reliable methods for estimating, forecasting and backtesting the ES. Formally, the ES at level $\tau \in (0, 1)$ is defined as the mean of the returns smaller than the respective τ-quantile (the VaR), where τ is usually chosen to be 2.5% as stipulated by the Basel Accord. The ES is introduced into the banking regulation because it overcomes several shortcomings of the VaR, such as not being coherent and its inability to capture tail risks beyond the τ-quantile (Artzner et al., 1999; Basel Committee, 2013; Danielsson et al., 2001). In contrast to estimation and forecasting of ES where most of the existing models for the VaR can easily be adapted and generalized to the ES, such a generalization is unfortunately not as straight-forward for backtesting of ES forecasts (Emmer et al., 2015). In general, backtesting of a risk measure is the process of testing whether given forecasts for this risk measure are correctly specified, which is carried out by comparing the history of the issued risk forecasts with the corresponding realized returns. The primary reason for the difficulty to directly backtest ES is its non-elicitability and non-identifiability (Fissler and Ziegel, 2016; Fissler et al., 2016; Gneiting, 2011; Weber, 2006) as consequently, there is no analog to the hit sequence which is the natural identification function of quantiles and which lies at the heart of most VaR backtests.[1]

As a consequence, most of the proposed procedures in the growing literature on backtesting ES use indirect approaches, e.g. based on forecasts for the entire tail distribution or by linear approximations of the ES with VaR forecasts at several probability levels. However, these approaches are in fact either joint backtests for a vector of risk measures such as the triple containing the VaR, the ES, and the volatility or even for the whole tail distribution (Nolde and Ziegel, 2017). As the proposed backtests require further input parameters such as forecasts for the volatility, the tail distribution beyond some quantile, or even the entire distribution, they are not applicable for the regulatory authorities because this additional information is not reported by the financial institutions (Aramonte et al., 2011; Basel Committee, 2016, 2017).

In this paper, we propose a novel backtest for ES forecasts which is based on a regression framework which models the conditional ES as a linear function, where we use financial returns as the response variable and ES forecasts as the explanatory variable including an intercept term. For correct ES forecasts, the intercept and slope parameters should be equal

---

[1] See Acerbi and Székely (2014), Carver (2013), Emmer et al. (2015), Fissler et al. (2016), Kerkhof and Melenberg (2004), Nolde and Ziegel (2017), Yamai and Yoshiba (2002), and Ziegel (2016) for the ongoing discussion on backtestability of the ES.

to 0 and 1, respectively. We use a Wald statistic to test for these parameter values, where we apply both, an asymptotic test using the covariance estimator introduced in Dimitriadis and Bayer (2017) and a bootstrap hypothesis test. We call this novel test the *bivariate ESR backtest*. This procedure is the first that backtests the risk measure ES stand-alone, i.e. the first that only uses ES forecasts as input parameters.[2] Through this feature, our new test is the first backtest for the ES which is practicably applicable for the regulatory authorities who only have ES forecasts at hand.

Such regression-based forecast evaluation approaches are already used for testing mean forecasts (Mincer and Zarnowitz, 1969), quantile forecasts (Gaglianone et al., 2011; Guler et al., 2017), and expectile forecasts (Guler et al., 2017). In contrast to these functionals where regression techniques are easily available (see e.g. Koenker and Bassett (1978), Efron (1991)), estimating regression parameters for an ES specific regression equation is more difficult as the ES is not elicitable (Gneiting, 2011). We overcome this difficulty by estimating the parameters of a joint regression procedure for the quantile and the ES, recently proposed by Dimitriadis and Bayer (2017), Patton et al. (2017) and Barendse (2017).

We also introduce a second regression-based ES backtest by fixing the slope parameter in the regression to one, and by only estimating and testing the intercept term, where we call this test the *intercept ESR backtest*. This second backtest allows for both, one-sided and two-sided hypotheses which contrasts with the first backtest that only allows for a two-sided hypothesis as it is generally unclear how underestimated and overestimated ES forecasts respectively influence the slope and intercept parameters. In the following, we use the convention that *overestimation* of risk measures is meant in the mathematical sense, i.e. as reporting too large real numbers which implies that the associated market risk is *underestimated*. Because the capital requirements that the financial institutions must keep as a reserve depend on the reported risk forecasts, the market participants have an incentive to overestimate the risk forecasts to minimize these expensive capital requirements. In contrast, underestimation of the forecasts results in too conservative risk forecasts and larger capital reserves, which does not have to be punished by the regulatory authorities. Thus, the regulators only have to prevent and penalize the overestimation of risk forecasts, which demonstrates the necessity of one-sided testing procedures. For example, the currently applied traffic light system (Basel Committee, 1996) is in fact a one-sided VaR backtest. As the bivariate ESR backtest, this intercept ESR test also has the desired characteristic to only require ES forecasts as input parameters and consequently is the first procedure in the literature that solely backtests the

---

[2] The backtests which come closest to our procedure in this regard are the exceedance residual backtests of McNeil and Frey (2000) and the conditional coverage backtests of Nolde and Ziegel (2017) which are in fact joint backtests for the VaR and the ES.

ES against a one-sided alternative. We provide implementations of the two ESR backtests proposed in this paper in our R package `esback` (Bayer and Dimitriadis, 2017a).

We introduce several simulation setups to evaluate the empirical properties of our novel ES backtests and compare them to the existing joint VaR and ES backtests of McNeil and Frey (2000) and Nolde and Ziegel (2017). In the first setup, we implement the classical size and power analysis for backtesting risk measures, where we simulate data stemming from a realistic data generating process and evaluate the empirical rejection frequencies of the backtests for forecasts stemming from the true and from some misspecified forecasting models. In the second setup, we introduce a novel technique for evaluating the power of backtests for financial risk measures, where we continuously misspecify certain model parameters of the data generating process to obtain a continuum of alternative models with a gradually increasing degree of misspecification. Misspecifying the different model parameters separately allows us to misspecify certain model characteristics (such as the reaction to shocks) in isolation, which permits a closer examination of the proposed backtesting procedures. To the best of our knowledge, this evaluation technique is new to the literature.

From these simulations, we find that the bivariate and the intercept ESR backtests that we propose in this paper are reasonable sized, especially when the tests are applied using the bootstrap. Moreover, they are more powerful than the existing backtests of McNeil and Frey (2000) and Nolde and Ziegel (2017) in almost all of the considered simulation designs for both, testing against one-sided and two-sided alternatives. Notably, throughout all simulation designs, the two ESR backtests are able to detect the various different misspecifications of the forecasts. In contrast, the existing backtests sometimes completely fail to detect certain misspecifications, for instance when the forecaster reports risk forecasts for a misspecified probability level.

The rest of this paper is organized as follows. Section 3.2 introduces the theory of our new backtests, and Section 3.3 reviews the existing ES backtesting techniques. Section 3.4 contains several simulation studies, and Section 3.5 applies the backtests to the risk forecasts of the S&P500 index. Section 3.6 concludes.

3.2. Theory

3.2.1. Setup and Notation

Let us consider a stochastic process

$$Z = \left\{ Z_t : \Omega \to \mathbb{R}^{k+1}, k \in \mathbb{N}, t = 1, \ldots, T \right\}, \tag{3.1}$$

defined on some complete probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$, with the filtration $\mathcal{F} = \big\{\mathcal{F}_t,\ t = 1, \dots T\big\}$ and $\mathcal{F}_t = \sigma\{Z_s,\ s \le t\}$ for all $t = 1, \dots, T$. We partition the stochastic process $Z_t = (Y_t, X_t)$, where Y_t is an absolutely continuous random variable of interest and X_t is a k-dimensional vector of explanatory variables. We denote the conditional cumulative distribution function of Y_t given the past information $\mathcal{F}_{t-1}$ by $F_t(y) = \mathbb{P}(Y_t \le y \mid \mathcal{F}_{t-1})$ and the corresponding probability density function by f_t. Whenever they exist, the mean and the variance of F_t are denoted by $\mathbb{E}_t[\cdot]$ and $\mathrm{Var}_t(\cdot)$.

For financial applications, the variable Y_t denotes the daily log returns of a financial asset (for instance, a stock or a portfolio), i.e. $Y_t = \log P_t - \log P_{t-1}$, where P_t denotes the price of the asset at day $t = 1, \dots, T$. This means that throughout this paper, we use the sign convention that positive returns denote profits, and negative returns denote losses. The vector X_t contains further variables that are used to produce forecasts for certain functionals (usually risk measures) of the random variable Y_t.

We are interested in testing whether forecasts for a certain d-dimensional, $d \in \mathbb{N}$ functional (risk measure) $\rho = \rho(F_t)$ of the conditional distribution F_t are correctly specified. For that, we define the most frequently used functionals for financial risk management in the following. The conditional quantile of Y_t given the information set $\mathcal{F}_{t-1}$ at level $\tau \in (0, 1)$ is defined as

$$Q_\tau\big(Y_t \mid \mathcal{F}_{t-1}\big) = F_t^{-1}(\tau) = \inf\big\{y \in \mathbb{R} : F_t(y) \ge \tau\big\}, \tag{3.2}$$

which is called the VaR at level τ in financial applications. Furthermore, we define the functional ES at level τ of Y_t given $\mathcal{F}_{t-1}$ as

$$\mathrm{ES}_\tau\big(Y_t \mid \mathcal{F}_{t-1}\big) = \frac{1}{\tau} \int_0^\tau F_t^{-1}(s)\, \mathrm{d}s. \tag{3.3}$$

If the distribution function F_t is continuous at its τ-quantile, this definition can be simplified to the truncated tail mean of Y_t,

$$\mathrm{ES}_\tau\big(Y_t \mid \mathcal{F}_{t-1}\big) = \mathbb{E}_t\left[Y_t \mid Y_t \le Q_\tau\big(Y_t \mid \mathcal{F}_{t-1}\big)\right]. \tag{3.4}$$

We denote an $\mathcal{F}_{t-1}$-measurable one-step-ahead forecast for day t for the risk measure ρ of the distribution F_t, stemming from some external forecaster or from some given forecasting model[3] by $\hat{\rho}_t = \hat{\rho}_t(\mathcal{F}_{t-1})$. Following this notation, we denote forecasts for the τ-VaR by $\hat{v}_t$

---

[3] For recent overviews on VaR and ES forecasting approaches, see Komunjer (2013) and Nadarajah et al. (2014).

and for the τ-ES by $\hat{e}_t$ for some fixed level $\tau \in (0, 1)$. For simplicity of the notation, we drop the dependence on τ as it is a fixed quantity.

As both, the incentive of the forecaster and the underlying method used to generate the forecasts are in general unknown, these forecasts are not necessarily correctly specified. The focus of this paper is to develop statistical tests for correctness of a given series of forecasts $\left\{\hat{\rho}_t,\ t = 1, \dots, T\right\}$ for the risk measure ρ relative to the realized return series $\left\{Y_t,\ t = 1, \dots, T\right\}$. This is in the literature usually referred to as *backtesting* of the risk measure ρ without strictly defining this terminology. We provide such a definition in the following. The core message of this definition is that besides the realized return series, a backtest for some risk measure is only allowed to require forecasts for this risk measure as input parameters.

Definition 3.2.1. A *backtest* for the series of forecasts $\left\{\hat{\rho}_t,\ t = 1, \dots, T\right\}$ for the d-dimensional risk measure (functional) ρ relative to the realized return series $\left\{Y_t,\ t = 1, \dots, T\right\}$ is a function

$$f : \mathbb{R}^T \times \mathbb{R}^{T \times d} \to (0, 1), \tag{3.5}$$

which maps the return and forecast series onto the respective p-value of the test.

This strict differentiation becomes relevant in the context of backtesting ES as, in contrast to the existing VaR backtests, the recently proposed ES backtests require further input parameters such as forecasts for the VaR, the volatility, or the entire tail distribution. The demand for these further quantities induces the following practical problems. First, the regulatory authorities who rely on such backtesting methods do not necessarily receive forecasts from the financial institutions for the additional information required by these tests, which makes such backtests inapplicable for the regulatory authorities. Second, a rejection of the tests does not necessarily imply that the ES is misspecified, but that the forecasts for any of the input components are misspecified. Consequently, these tests are in fact not backtests for the ES, but rather backtests for vectors of risk measures, or the entire tail distribution.

The two novel regression based ES backtests we propose in the next section are the first backtests for the ES which follow Definition 3.2.1 as they only require forecasts for the ES. This makes these tests the first ES backtests in this sense.

3.2.2. The Bivariate ESR Backtest

We propose a new backtest for the risk measure ES that tests whether a series of ES forecasts $\{\hat{e}_t, t = 1, \dots T\}$, stemming from some external forecaster or forecasting model is correctly specified relative to a series of in due course realized returns $\{Y_t, t = 1, \dots T\}$. For that, we regress the returns Y_t on the forecasts $\hat{e}_t$ including an intercept term by using a regression equation designed specifically for the functional ES,

$$Y_t = \alpha + \beta \hat{e}_t + u_t^e, \tag{3.6}$$

where $\text{ES}_\tau(u_t^e \mid \mathcal{F}_{t-1}) = 0$. Given the structure in (3.6) and since $\hat{e}_t$ is generated by using the information set $\mathcal{F}_{t-1}$, this condition on the error term is equivalent to

$$\text{ES}_\tau \left(Y_t \mid \mathcal{F}_{t-1}\right) = \alpha + \beta \hat{e}_t. \tag{3.7}$$

We then test the hypothesis

$$\mathbb{H}_0 : (\alpha, \beta) = (0, 1) \qquad \text{against} \qquad \mathbb{H}_1 : (\alpha, \beta) \neq (0, 1). \tag{3.8}$$

Under $\mathbb{H}_0$, the ES forecasts are correctly specified as it holds that $\hat{e}_t = \text{ES}_\tau \left(Y_t \mid \mathcal{F}_{t-1}\right)$.[4] Since this ES backtest is based on a regression procedure and simultaneously tests the parameters α and β, we call this test the *bivariate ESR backtest*.

As outlined in Dimitriadis and Bayer (2017), estimating the parameters (α, β) in (3.6) by M- or Z-/GMM-estimation stand-alone using a semiparametric method without specifying the full conditional distribution of the error term u_t^e is not possible since the functional ES is not elicitable (Gneiting, 2011). However, these parameters can be estimated through a regression technique which jointly models a regression equation for the quantile and the ES proposed by Dimitriadis and Bayer (2017), Patton et al. (2017) and Barendse (2017), which we briefly review in Appendix 3.A. We use this joint regression framework for the semiparametric estimation of (3.6) by estimating the joint system,

$$Y_t = \gamma + \delta \hat{e}_t + u_t^q, \tag{3.9}$$

$$Y_t = \alpha + \beta \hat{e}_t + u_t^e, \tag{3.10}$$

where $Q_\tau(u_t^q \mid \mathcal{F}_{t-1}) = 0$. and $\text{ES}_\tau(u_t^e \mid \mathcal{F}_{t-1}) = 0$. This means we choose Y_t as the response variable and $\left(1, \hat{e}_t\right)$ as explanatory variables for this regression procedure. Because our null

---

[4] Given that the ES forecasts are correctly specified, i.e. $\hat{e}_t = \text{ES}_\tau \left(Y_t \mid \mathcal{F}_{t-1}\right)$, the correct specification condition (3.7) is equivalent to $\alpha = (1 - \beta)\hat{e}_t$. This results in the remark of Holden and Peel (1990), who claim that the null hypothesis, given in (3.8) is only a sufficient, but not a necessary condition for correctly specified forecasts as $\alpha = (1 - \beta)\hat{e}_t$ is the required necessary condition. However, this more general condition implies that the forecasts $\hat{e}_t$ are constant for all $t = 1, \dots, T$, which is highly unrealistic given the dynamic nature of financial time series. Consequently, we employ the hypotheses given in (3.8) for our backtesting procedure.

hypothesis is based on only testing the parameters (α, β) in the ES regression equation given in (3.10), we use a Wald statistic which only incorporates these parameters,

$$T_{\text{ESR}} = \left((\hat{\alpha}, \hat{\beta})' - (0, 1)' \right)' \widehat{\Sigma}_{\text{ES}}^{-1} \left((\hat{\alpha}, \hat{\beta})' - (0, 1)' \right)', \tag{3.11}$$

where $\widehat{\Sigma}_{\text{ES}}$ is an estimator for the (asymptotic) covariance matrix of the M-estimator of the parameters (α, β). Patton et al. (2017) show consistency and asymptotic normality for the M-estimator of the regression parameters for α-mixing time series. Using this, and given that $\hat{\Sigma}_{\text{ES}} \xrightarrow{\mathbb{P}} \Sigma_{\text{ES}}$, the test statistic asymptotically follows a χ^2 distribution with two degrees of freedom,

$$T_{\text{ESR}} \xrightarrow{d} \chi_2^2. \tag{3.12}$$

We implement both, backtests based on estimates for the asymptotic covariance matrix and based on the bootstrap (Efron, 1979). For the asymptotic version, we employ the *scl-sp* covariance estimation method discussed in Dimitriadis and Bayer (2017). We further implement the bootstrap hypothesis testing procedure[5] where in each bootstrap sample, we estimate the model parameters and the asymptotic covariance matrix to compute a total of $B = 1000$ bootstrap Wald statistics as in (3.11), where the bootstrap estimates are centered around the estimate for the original sample. Finally, the bootstrap p-value is the share of the B bootstrap test statistics that are larger than or equal to the test statistic for the original sample. As neither the underlying loss function of the M-estimator, given in (3.28), nor the asymptotic covariance, given in (3.31) - (3.35), depend on the temporal ordering of the pairs $(Y_t, \hat{e}_t)$, we apply the iid bootstrap resampling technique of Efron (1979).

Similar tests are already implemented for backtesting of forecasts for the mean (Mincer and Zarnowitz, 1969), for quantiles (Gaglianone et al., 2011) and for expectiles (Guler et al., 2017). As these functionals are elicitable, M-estimation of regression parameters for mean, quantile (Koenker and Bassett, 1978) and expectile (Efron, 1991) regressions is straight-forward. This section shows that introducing the same concept for backtesting ES forecasts is possible, but technically more demanding as we have to estimate the regression parameters through a joint system as given in (3.9) and (3.10).

---

[5] This approach provides an asymptotic refinement, i.e. the error in the rejection probability decreases faster compared to both, the asymptotic distribution and the bootstrapped covariance matrices for the test, see e.g. MacKinnon (2009). In the construction of confidence intervals, this is also known as the percentile-t method.

3.2.3. The One-Sided Intercept ESR Backtest

The bivariate ESR backtest introduced in the previous section only allows for testing two-sided hypotheses as specified in (3.8) because it is generally unclear how too small or too large risk forecasts influence the parameters α and β. Because the capital requirements the financial institutions have to keep as a reserve depend on the reported risk forecasts, the market participants have an incentive to overestimate[6] the risk forecasts in order to keep as little capital requirements as possible. In contrast, underestimation of the risk measures results in too conservative risk forecasts and consequently higher capital requirements, which does not have to be punished by the regulatory authorities.[7] Thus, the regulatory authorities only have to prevent and consequently penalize the overestimation of risk measures, which can be done by using one-sided backtesting procedures. For example, the traffic light system (Basel Committee, 1996), currently implemented in the Basel Accords, is in fact a one-sided backtest for the hit ratios of VaR forecasts.

Consequently, we also introduce a regression-based backtesting procedure for the ES that allows for both, specifying one-sided and two-sided hypotheses. This backtest is based on regressing the forecast errors, $Y_t - \hat{e}_t$, on an intercept term only,

$$Y_t - \hat{e}_t = \alpha + u_t^e, \tag{3.13}$$

where $\mathrm{ES}_\tau(u_t^e \mid \mathcal{F}_{t-1}) = 0$ and testing whether the parameter α is zero. Estimation of the parameter α in (3.13) is carried out by computing the empirical ES of the forecast errors $Y_t - \hat{e}_t$. By using this restricted regression equation, we can define a one-sided and a two-sided hypothesis,

$$\begin{aligned} &\mathbb{H}_0^{2s}: \alpha = 0 \quad &&\text{against} \quad &&\mathbb{H}_1^{2s}: \alpha \neq 0, \quad \text{and} \\ &\mathbb{H}_0^{1s}: \alpha \geq 0 \quad &&\text{against} \quad &&\mathbb{H}_1^{1s}: \alpha < 0, \end{aligned} \tag{3.14}$$

which we test by using a t-test based on the asymptotic covariance and based on the bootstrap procedure described above. Note that this is equivalent to setting the slope parameter of the bivariate ESR test given in (3.6) to one and only estimating and testing the intercept term. Consequently, we call this backtest the *intercept ESR backtest*. Both, the bivariate and the intercept ESR backtests proposed in this paper are implemented in our R package `esback` (Bayer and Dimitriadis, 2017a).

---

[6] Overestimation of a risk measure is meant in the mathematical sense, which means reporting a too large real number. As the ES forecasts are strictly negative, this implies in fact underestimation of the associated market risk.

[7] One could interpret the higher capital requirements as a punishment for too conservative risk forecasts.

3.3. Existing Backtests

Over the past two decades and especially driven by the recent transition from VaR to ES in the Basel regulatory framework (Basel Committee, 2016, 2017), a large literature on backtesting the ES has emerged. These backtests are usually introduced with financial regulators in mind who need to verify the risk forecasts they receive from the financial institutions. To be applicable for the regulatory authorities, a backtest for the risk measure ES thus follows Definition 3.2.1 and only requires the observed return series and the ES forecasts as input variables. However, many of the proposed backtests for the ES fail to have this property. In particular, several tests require the whole return distribution (Acerbi and Székely, 2014; Berkowitz, 2001; Graham and Pál, 2014; Kerkhof and Melenberg, 2004; Wong, 2008), the cumulative violation process $\int_0^\tau \mathbb{1}_{\{Y_t \le \hat{v}_t(p)\}} \, \mathrm{d}p$ (Costanzino and Curran, 2015; Du and Escanciano, 2017; Emmer et al., 2015; Kratz et al., 2018), the volatility (McNeil and Frey, 2000; Nolde and Ziegel, 2017; Righi and Ceretta, 2013, 2015), or the VaR (McNeil and Frey, 2000; Nolde and Ziegel, 2017) in addition to the ES forecasts. However, this information (except the VaR) is not reported by the financial institutions and therefore, most of these tests can not be used by the regulators (Aramonte et al., 2011; Basel Committee, 2017).

Furthermore, when more information than solely the ES forecasts is used for backtesting, a rejection of the null hypothesis does not necessarily imply that the ES forecasts are wrong. More precisely, a rejection of the null implies that *some* component of the input parameters is incorrect (cf. Nolde and Ziegel, 2017). A related concern is raised by Aramonte et al. (2011), who note that financial institutions could be tempted to submit forecasts of this additional information chosen such that the tests have particularly low power, so that correctness of their internal model (or their issued ES forecasts) is not doubted.

Strictly following Definition 3.2.1, we would have to distinguish between backtests for the ES and joint backtests for the pair VaR and ES. However, as the ES is strongly intertwined with the VaR (through its definition and through the joint elicitability), sensible forecasts for the ES are based on correctly specified VaR forecasts. Consequently, it is reasonable to backtest both quantities jointly and thus, we compare the performance of our ES backtests to existing joint VaR and ES backtests in the literature. In the following, we describe the exceedance residual test of McNeil and Frey (2000) and the conditional calibration tests of Nolde and Ziegel (2017) in more detail, since both have versions that only require VaR forecasts in addition to the ES.

3.3.1. Testing the Exceedance Residuals

One of the first and still most frequently used tests for the ES is the exceedance residual (ER) backtest of McNeil and Frey (2000). This approach is based on the ES residuals that exceed the VaR, $er_t = \big(Y_t - \hat{e}_t\big)\mathbb{1}_{\{Y_t \le \hat{v}_t\}}$, which form a martingale difference sequence given that $\hat{v}_t$ and $\hat{e}_t$ are the true $\mathcal{F}_{t-1}$-measurable quantile and ES respectively. McNeil and Frey (2000) furthermore consider a second version that uses exceedance residuals standardized by the volatility, i.e. $er_t/\hat{\sigma}_t$.

This backtest tests whether the expected value of the (raw or standardized) ER, $\mu = \mathbb{E}[er_t]$, is zero using the estimate $\hat{\mu} = 1/(\sum_{t=1}^{T} \mathbb{1}_{\{Y_t \le \hat{v}_t\}}) \sum_{t=1}^{T} er_t$ in conjunction with a bootstrap hypothesis test (see Efron and Tibshirani, 1994, p. 224). In the original paper, McNeil and Frey (2000) propose to test μ against the one-sided alternative that μ is negative, i.e. that the ES is overestimated. However, in this paper we discuss both, tests based on one-sided and two-sided hypotheses, so that in addition to the original proposal, we also include a two-sided test,

$$\begin{aligned} &\mathbb{H}_0^{2s}: \mu = 0 \qquad \text{against} \qquad \mathbb{H}_1^{2s}: \mu \neq 0, \quad \text{and} \\ &\mathbb{H}_0^{1s}: \mu \geq 0 \qquad \text{against} \qquad \mathbb{H}_1^{1s}: \mu < 0. \end{aligned} \tag{3.15}$$

By Definition 3.2.1, the test using the standardized ER is in fact a joint backtest for the triple VaR, ES and volatility, whereas the test using the raw ER is a joint backtest for the pair VaR and ES. In light of the discussion above, the test using the raw ER is therefore preferred. Nevertheless, in the simulation studies and the empirical application we apply both approaches and find that they perform alike.

Even though the intercept ESR test introduced in Section 3.2.3 and the ER backtest appear to be similar, there is a subtle difference between the two test statistics. For the intercept ESR test, we compute the empirical ES of $Y_t - \hat{e}_t$, i.e. the average of $Y_t - \hat{e}_t$ given that $Y_t - \hat{e}_t$ is smaller than its empirical τ-quantile. In contrast, the ER backtest computes the average of $Y_t - \hat{e}_t$, given that Y_t is smaller than the respective forecast for its τ-quantile $\hat{v}_t$. This difference seems marginal, but it has severe consequences for the theoretical and empirical properties of the tests.

As we can write $\hat{\mu} = 1/\tilde{T} \sum_{t=1}^{T} Y_t \mathbb{1}_{\{Y_t \le \hat{v}_t\}} - 1/\tilde{T} \sum_{t=1}^{T} \hat{e}_t \mathbb{1}_{\{Y_t \le \hat{v}_t\}}$, where $\tilde{T} = \sum_{t=1}^{T} \mathbb{1}_{\{Y_t \le \hat{v}_t\}}$, the ER backtest in fact compares the empirical average of Y_t truncated at $\hat{v}_t$ to the average ES forecast $\hat{e}_t$, whenever there is a VaR violation. Thus, this backtest rejects whenever the distance/relation between the VaR and ES-forecasts is incorrect. However, simultaneous misspecifications of both forecasts, such as e.g. generated by misspecification of the volatility process in location scale models cannot be detected. In the same spirit, the ER backtest cannot distinguish between correct forecasts for the VaR and ES at level τ and (correct) forecasts for a misspecified probability level $\tilde{\tau} \neq \tau$, as the given level τ does not influence the ER test statistic at all. In contrast, by computing the empirical τ-quantile of $Y_t - \hat{e}_t$ (instead of using the forecast $\hat{v}_t$), the intercept ESR test does not suffer from these shortcomings as can be observed in the simulation results in Section 3.4.2.

3.3.2. Conditional Calibration Backtests

Nolde and Ziegel (2017) introduce the concept of conditional calibration (CC) based on strict identification functions (also known as moment conditions or estimating equations) of the respective functional and show that many classical backtests for risk measures can be unified using this concept. For the pair VaR and ES at level $\tau \in (0, 1)$, they choose the strict identification function

$$V(Y, v, e) = \begin{pmatrix} \tau - \mathbb{1}_{\{Y \le v\}} \\ e - v + \mathbb{1}_{\{Y \le v\}}(v - Y)/\tau \end{pmatrix}, \tag{3.16}$$

whose expectation is zero if and only if v and e equal the true VaR and ES of Y respectively. The CC backtest for forecasts for the VaR, $\hat{v}_t$ and for the ES, $\hat{e}_t$ is based on the hypotheses

$$\begin{aligned} &\mathbb{H}_0^{2s}: \mathbb{E}\big[V(Y_t, \hat{v}_t, \hat{e}_t) \mid \mathcal{F}_{t-1}\big] = 0 \quad \text{against} \quad \mathbb{E}\big[V(Y_t, \hat{v}_t, \hat{e}_t) \mid \mathcal{F}_{t-1}\big] \neq 0, \quad \text{and} \\ &\mathbb{H}_0^{1s}: \mathbb{E}\big[V(Y_t, \hat{v}_t, \hat{e}_t) \mid \mathcal{F}_{t-1}\big] \ge 0 \quad \text{against} \quad \mathbb{E}\big[V(Y_t, \hat{v}_t, \hat{e}_t) \mid \mathcal{F}_{t-1}\big] < 0, \end{aligned} \tag{3.17}$$

component-wise and almost surely for all $t = 1, \dots, T$. This is equivalent to testing $\mathbb{E}\big[h_t' V(Y_t, \hat{v}_t, \hat{e}_t)\big] = 0$ for all $\mathcal{F}_{t-1}$ measurable $\mathbb{R}^2$-valued functions h_t. As this is infeasible, Nolde and Ziegel (2017) propose to use an $\mathcal{F}_{t-1}$-measurable sequence of $q \times 2$-matrices of test functions $\boldsymbol{h}_t$ for some $q \in \mathbb{N}$ and to use the Wald-type test statistic

$$T_{\mathrm{CC}} = T \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h}_t V(Y_t, \hat{v}_t, \hat{e}_t) \right)' \widehat{\Omega}^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h}_t V(Y_t, \hat{v}_t, \hat{e}_t) \right), \tag{3.18}$$

where $\widehat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{h}_t V(Y_t, \hat{v}_t, \hat{e}_t)) (\boldsymbol{h}_t V(Y_t, \hat{v}_t, \hat{e}_t))'$ is a consistent estimator of the covariance of the q-dimensional vector $\boldsymbol{h}_t V(Y_t, \hat{v}_t, \hat{e}_t)$. Under $\mathbb{H}_0$, the test statistic asymptotically follows a χ_q^2 distribution with q degrees of freedom.

Nolde and Ziegel (2017) propose two versions of this test, where the first uses no information besides the risk forecasts (termed *simple CC test*), and where the second additionally requires volatility forecasts (termed *general CC* test). For the simple CC test, the test function is the identity matrix, $\boldsymbol{h}_t = I_2$, for both, the one- and two-sided hypotheses. For the general CC test, they propose to choose

$$\boldsymbol{h}_t = \hat{\sigma}_t \big((\hat{e}_t - \hat{v}_t) / \tau, \, 1 \big) \quad \text{and} \quad \boldsymbol{h}_t = \begin{pmatrix} 1 & |\hat{v}_t| & 0 & 0 \\ 0 & 0 & 1 & \hat{\sigma}_t^{-1} \end{pmatrix}', \tag{3.19}$$

for the two-sided and for the one-sided test, respectively, where $\hat{\sigma}_t$ is a forecast for the volatility. As with the standardized ER test, the general CC test is strictly speaking a backtest for the triple VaR, ES, and volatility, but we nevertheless include both versions in our empirical comparisons. We provide implementations of the two ESR backtests proposed in this paper, both ER backtests of McNeil and Frey (2000) and both CC backtests of Nolde and Ziegel (2017) in our R package `esback` (Bayer and Dimitriadis, 2017a).

3.4. Monte-Carlo Simulations

In this section, we evaluate the empirical performance of our proposed ES backtests and compare them to the tests of McNeil and Frey (2000) and Nolde and Ziegel (2017). For that, we first assess the empirical size of the tests, defined as the rejection frequency of the test under the null hypothesis and which should equal the nominal significance level. Then, we analyze the empirical power of the tests which is defined as the rejection frequency of forecasts stemming from some misspecified model and which is optimally as close to one as possible.

This comparison is conducted using two different approaches. The first, presented in Section 3.4.1, follows the typical strategy in the related literature of first assessing the size of the backtests with some realistic location-scale data generating process (DGP), followed by an evaluation of the power by backtesting forecasts stemming from overly simplified models, in our case the Historical Simulation model. In the second setup, presented in Section 3.4.2, we continuously misspecify certain parameters of the true model and thereby obtain alternative models with a continuously increasing degree of misspecification. This approach of evaluating backtests has two main advantages. First, we obtain power curves which can be used to draw conclusions how an increasing model misspecification influences the test decisions. Second, misspecifying the different model parameters separately allows us to misspecify certain model characteristics while leaving the remaining model unchanged. Thus, we can evaluate the capability of the backtests to identify certain model misspecifications, which allows for a closer examination of the backtesting procedures.

3.4.1. Traditional Size and Power Comparisons

For the first simulation study, we simulate returns from an EGARCH(1,1) model (Nelson, 1991) with t-distributed innovations, where the parameter values are calibrated using daily returns of the S&P 500 index. This model is given by

$$\begin{aligned} Y_t &= \sigma_t z_t, \\ \log(\sigma_t^2) &= -0.160 - 0.125 z_{t-1} + 0.130\,(|z_{t-1}| - E[|z_{t-1}|]) + 0.983\log(\sigma_{t-1}^2), \\ z_t &\sim t_{7.24}, \end{aligned} \tag{3.20}$$

where z_t are innovations stemming from the standardized Student-t distribution with 7.24 degrees of freedom. As the EGARCH model is highly flexible and due to its calibrated parameter values, this DGP accurately replicates the distributional properties of daily financial returns. Conditional VaR and ES forecasts at level τ for the DGP in (3.20) are given by

$$\hat{v}_t = \hat{\sigma}_t q_z(\tau) \quad \text{and} \quad \hat{e}_t = \hat{\sigma}_t \xi_z(\tau), \tag{3.21}$$

where $\hat{\sigma}_t$ is a volatility forecast generated through the model given in (3.20) and $q_z(\tau)$ and $\xi_z(\tau)$ are the τ-quantile, respectively the τ-ES of the innovations z_t. For the following size and power analysis of the backtests, we simulate the process (3.20) 10,000 times with sample sizes of 250, 500, 1000, 2500, and 5000 observations and 250 additional pre-sample values required for the power analysis. As stipulated by the Basel Accords, we forecast the two risk measures for the probability level $\tau = 2.5\%$. In this part of the study, we focus on two-sided hypotheses and defer the one-sided case to Section 3.4.3.

Table 3.1 presents the empirical sizes of the considered backtests for the different sample sizes and for nominal test sizes of 1%, 5%, and 10%. We find that in large samples, all backtests display rejection rates close to the respective nominal sizes. However, in small samples all backtests are oversized and they differ with respect to their speed of convergence. Looking at the individual tests in greater detail, we find that especially the tests relying on asymptotic quantities (i.e. the ESR and CC tests) are substantially oversized in small samples and converge to the nominal sizes comparably slow. However, by using the bootstrap for the intercept and bivariate ESR tests (indicated by (b) in the table), the empirical sizes are much closer to the nominal sizes in small samples than for the asymptotic versions. Comparing the intercept and the bivariate ESR test, we find that the former has better size properties in small samples, presumably because less parameters need to be estimated and the covariance is simpler. Furthermore, also the two ER tests (which also rely on bootstrapping) exhibit good

Table 3.1: Empirical sizes of the backtests

Nominal Size	Sample Size	bivariate ESR (b)	bivariate ESR	intercept ESR (b)	intercept ESR	General CC	Simple CC	Std. ER	ER
1%	250	0.03	0.18	0.02	0.10	0.01	0.21	0.04	0.05
	500	0.02	0.09	0.02	0.06	0.03	0.12	0.00	0.01
	1000	0.02	0.05	0.01	0.03	0.03	0.07	0.01	0.01
	2500	0.01	0.03	0.01	0.02	0.02	0.04	0.01	0.01
	5000	0.01	0.02	0.01	0.01	0.02	0.03	0.01	0.01
5%	250	0.09	0.27	0.07	0.16	0.08	0.28	0.07	0.08
	500	0.07	0.17	0.06	0.11	0.10	0.19	0.04	0.06
	1000	0.06	0.11	0.05	0.07	0.08	0.13	0.04	0.06
	2500	0.05	0.07	0.04	0.06	0.07	0.09	0.05	0.06
	5000	0.06	0.06	0.05	0.07	0.06	0.07	0.05	0.06
10%	250	0.15	0.33	0.12	0.20	0.17	0.32	0.12	0.14
	500	0.13	0.22	0.10	0.16	0.16	0.24	0.09	0.11
	1000	0.12	0.16	0.09	0.12	0.13	0.19	0.10	0.12
	2500	0.11	0.12	0.09	0.12	0.12	0.14	0.10	0.11
	5000	0.11	0.11	0.11	0.13	0.11	0.12	0.10	0.11

Notes: The table reports the empirical sizes of the backtests for the EGARCH(1,1)-t process given in (3.20). The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000).

empirical sizes and there are hardly any differences between the raw and the standardized version.

For a comparison of the power of the backtests, we evaluate their ability to reject the null hypothesis for risk models producing incorrect ES forecasts. We utilize the Historical Simulation approach which forecasts the VaR and ES by using their empirical counterparts from previous trading days,

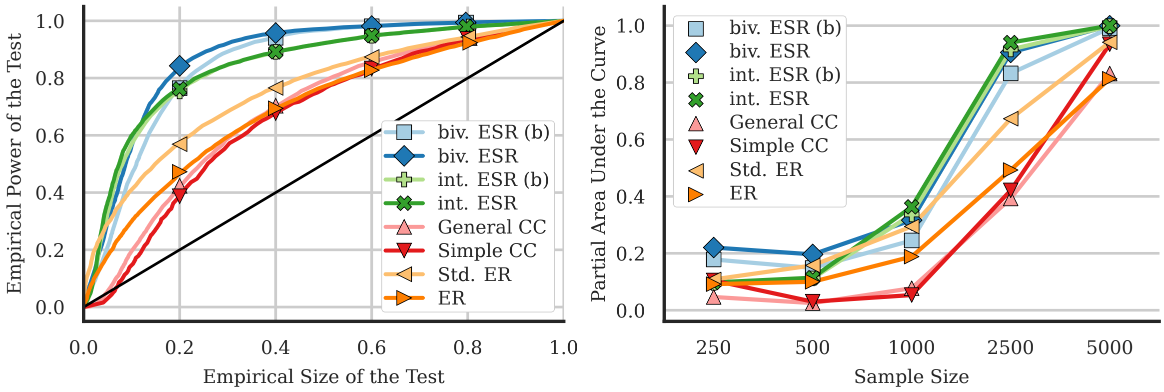
$$\hat{v}_t = \widehat{Q}_\tau \left(Y_{t-1}, Y_{t-2}, \cdots, Y_{t-w}\right) \quad \text{and} \quad \hat{e}_t = \frac{1}{\sum_{i=1}^{w} \mathbb{1}_{\{Y_{t-i} \le \hat{v}_{t-i}\}}} \sum_{i=1}^{w} Y_{t-i} \cdot \mathbb{1}_{\{Y_{t-i} \le \hat{v}_{t-i}\}}, \tag{3.22}$$

where $\widehat{Q}_\tau$ is the empirical τ-quantile and w is the length of a rolling window, that we set to 250, i.e. one year of data. Since the standardized ER and the general CC backtest both require forecasts of the volatility, we estimate this quantity with the sample standard deviation of the returns over the same rolling window.

For a meaningful and fair comparison of the power of the backtests to reject the null hypothesis, we compare the *size-adjusted power*[8] of the backtests (Lloyd, 2005). For this,

---

[8] A comparison of the *raw power*, i.e. the raw rejection rate of the null hypotheses, could be misleading due to the differences in the empirical sizes of the backtests. In particular, an oversized test would exhibit unrealistically large rejection rates.



(a) Size-adjusted Power

(b) Partial Area Under the Curve

Figure 3.1: Size-adjusted power and Partial Area Under the Curve plots against Historical Simulation for a sample size of 1000 days. The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000).

the original critical values of the tests are either increased or decreased such that the rejection frequencies of the true model equal the nominal test sizes. The size-adjusted power is then given by the rejection frequencies of the alternative models using these modified critical values.

Figure 3.1a contains the size-adjusted power of the backtests for all empirical sizes in the unit interval for a sample size of 1000.[9] The black line depicts the case of equal empirical size and power, which can be seen as a lower bound for any reasonable test: whenever the power is below this line, randomly guessing the test decision is more accurate than performing the test. We observe that both, the bivariate and intercept version of the ESR backtest clearly dominate the others at almost all empirical sizes, including the most relevant region of test sizes between 1% and 10%. Furthermore, the ESR tests using asymptotic quantities are slightly more powerful than their bootstrap versions (indicated by (b)), but the loss in power is negligible compared to the improvements in the sizes we find in Table 3.1.

In order to present results for all considered sample sizes in condensed form for the relevant area of empirical sizes between 1% and 10%, we summarize the size-adjusted power by the partial area under the curve (PAUC), as proposed by Lloyd (2005). For that, we numerically compute the area under each power curve for the empirical sizes between 1% and 10% which is thus the power to reject a false model averaged over the considered test sizes. In Figure 3.1b, we present the PAUC for all backtests and sample sizes. As expected,

---

[9] These plots are known as the receiver operating characteristic (ROC) curves and origin from the psychometrics literature (Lloyd, 2005). They are an effective presentation method for general binary classification tasks such as hypothesis testing as they show the size-adjusted power simultaneously for all significance levels.

Table 3.2: Empirical sizes for the second simulation study.

Null Hypothesis	bivariate ESR (b)	bivariate ESR	intercept ESR (b)	intercept ESR	General CC	Simple CC	Std. ER	ER
Two-Sided	0.06	0.07	0.05	0.06	0.08	0.09	0.05	0.06
One-Sided	–	–	0.07	0.03	0.02	0.02	0.06	0.06

Notes: This table shows the empirical sizes of the backtests for the GARCH(1,1)-t model given in (3.23), for a nominal test size of 5% and for both, one-sided and two-sided hypotheses. The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000). Note that the bivariate ESR test does not permit a one-sided hypothesis and therefore, we only present sizes for the two-sided hypothesis.

the average power increases with the sample size, so that using more information leads to more reliable decisions about the quality of a forecast. We find that for all considered sample sizes, the ESR backtests dominate the other testing approaches. As a robustness check for these findings, we repeat this simulation experiment with the DGP used by Gaglianone et al. (2011) and find that the results, presented in Appendix 3.B, are similar to the findings of this section.

3.4.2. Continuous Model Misspecification

In the second simulation study, we use a GARCH(1,1) model with standardized Student-t distributed innovations,

$$\begin{aligned} Y_t &= \sigma_t z_t, \\ \sigma_t^2 &= \gamma_0 + \gamma_1 Y_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \\ z_t &\sim t_\nu, \end{aligned} \tag{3.23}$$

with the parameter values $\gamma_0 = 0.01$, $\gamma_1 = 0.1$, $\gamma_2 = 0.85$, and $\nu = 5$ for the true model. For the analysis of the backtests, we simulate 10,000 times from this model with a fixed sample size of 2500 observations and consider the probability level $\tau = 2.5\%$ for the VaR and the ES.

Table 3.2 presents the empirical sizes of the backtests for a nominal size of 5% for both, the two- and one-sided hypotheses. As in the first simulation study, we find that most of the backtests are reasonably sized with rejection frequencies close to the nominal value. However, the two CC tests reject the true model slightly too often in the two-sided, respectively too rarely in the one-sided case.

For a detailed analysis of the power of the backtests, we continuously misspecify the true model according to the following five designs:

(a) We misspecify how the conditional variance reacts to the squared returns by varying the ARCH parameter γ_1. We choose $\tilde{\gamma}_1$ between 0.03 and 0.2 and let $\tilde{\gamma}_2 = 0.95 - \tilde{\gamma}_1$, such that the persistence of the GARCH process remains constant. When $\tilde{\gamma}_1 < \gamma_1$, there is too little variation in the ES forecasts due to the reduced response to shocks and the GARCH process approaches a constant volatility model.

(b) We alter the unconditional variance of the GARCH process $\mathbb{E}[\sigma_t^2] = \gamma_0/(1 - \gamma_1 - \gamma_2)$ between 0.5 and 0.01 by varying the parameter γ_0 while holding γ_1 and γ_2 constant. Since the conditional variance is a weighted combination of the unconditional variance, the past squared returns and the past conditional variance, this change implies that the ES is always underestimated when the unconditional variance is larger as its true value, and vice versa.

(c) We vary the persistence of shocks between 0.9 and 0.999 by setting $\tilde{\gamma}_1 = c \cdot \gamma_1$ and $\tilde{\gamma}_2 = c \cdot \gamma_2$ for a constant c that we vary, and $\tilde{\gamma}_0 = \mathbb{E}[\sigma_t^2](1 - \tilde{\gamma}_1 - \tilde{\gamma}_2)$ to keep the unconditional variance constant. A higher persistence causes a stronger and longer reaction to shocks.

(d) We vary the degrees of freedom of the underlying Student-t distribution between 3 and ∞. Since the conditional variance is unaffected, this modification implies a relative horizontal shift of the ES forecasts.

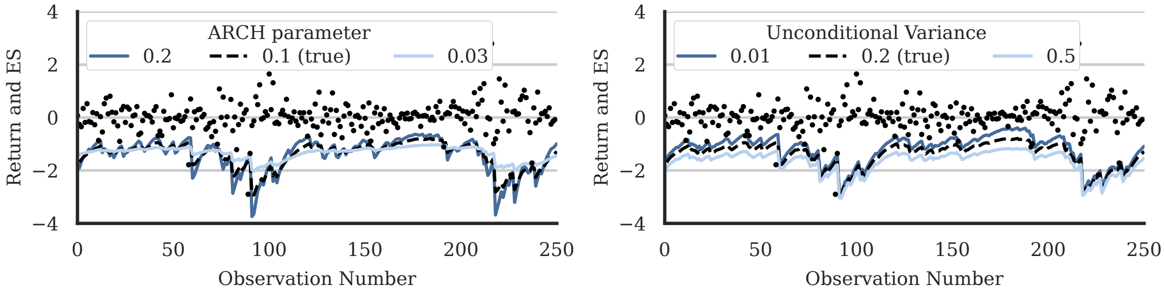
(e) We misspecify the probability level $\tilde{\tau}$ of the ES forecasts between 0.5% and 5%. This represents the scenario that a forecaster submits (accidentally or on purpose) predictions for some level $\tilde{\tau} \neq \tau$. Similar to changing the degrees of freedom, this modification implies a relative horizontal shift of the ES forecasts.

As an illustrative example of these misspecifications, Figures 3.2a to 3.2e show 250 realizations of the returns of the true DGP (3.23), together with the corresponding ES forecasts of the true model (black dashed line) and two models following the parameter misspecifications described in the points (a) to (e) above.

We present the size-adjusted rejection rates plotted against the respective misspecified parameters for these five designs in Figures 3.3a to 3.3e. The true model is indicated by the gray vertical line and, induced by the results of Figure 3.2, the x-axis is oriented such that too small (too risky) ES forecasts are on the left side of the true model.[10] Even though there is no backtest that dominates the others throughout all considered designs, several conclusions can be drawn from this figure.

---

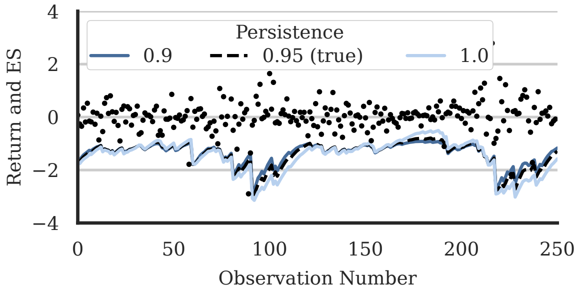
[10] Notice that this inequality of the forecast magnitude only holds on average in the cases of Figures 3.3a and 3.3c whereas it holds strictly for Figures 3.3b, 3.3d and 3.3e.



(a) Changing the reaction to the squared returns

(b) Changing the unconditional variance

(c) Changing the persistence



(d) Changing the degrees of freedom

(e) Changing the probability level

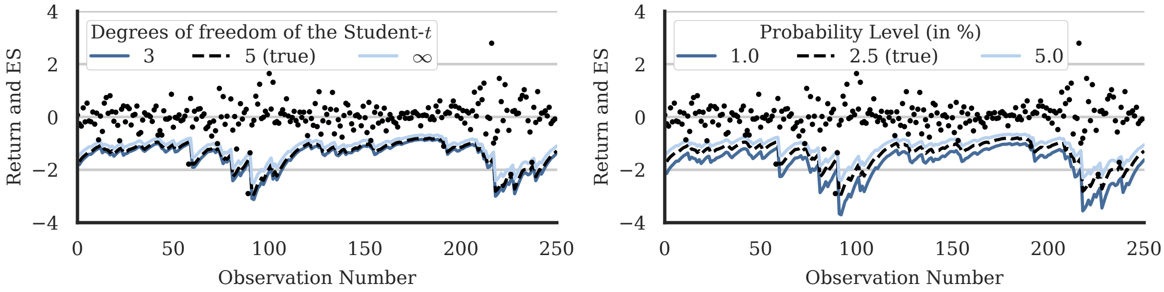
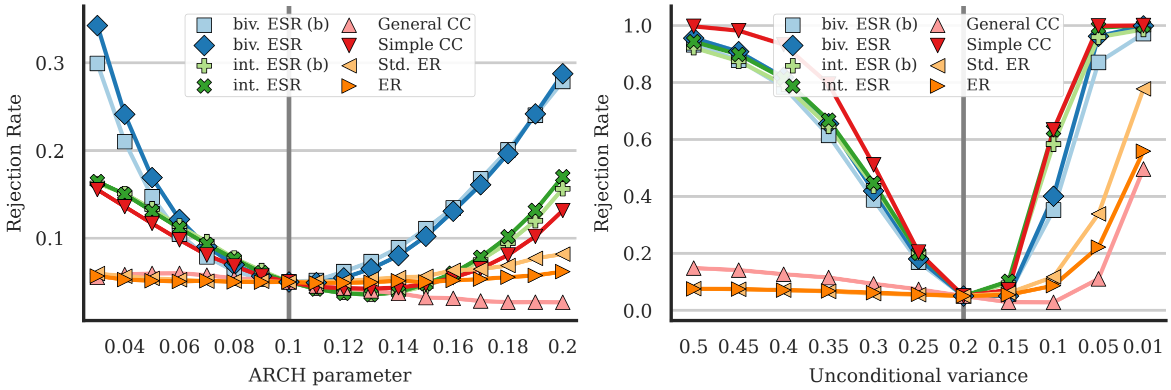
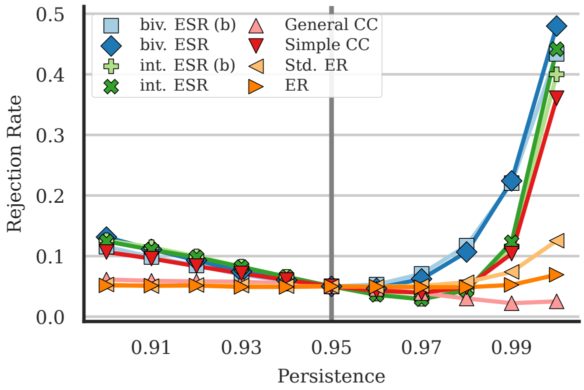


Figure 3.2: These plots show exemplary simulated return series with 250 observations for the DGP given in (3.23) and for the five parameter misspecifications illustrated in the points (a) - (e) in Section 3.4.2. In each of the subfigures, the black dashed line corresponds to the true model parameters.

(a) Changing the ARCH parameter (b) Changing the unconditional variance



(c) Changing the persistence



(d) Changing the degrees of freedom (e) Changing the probability level

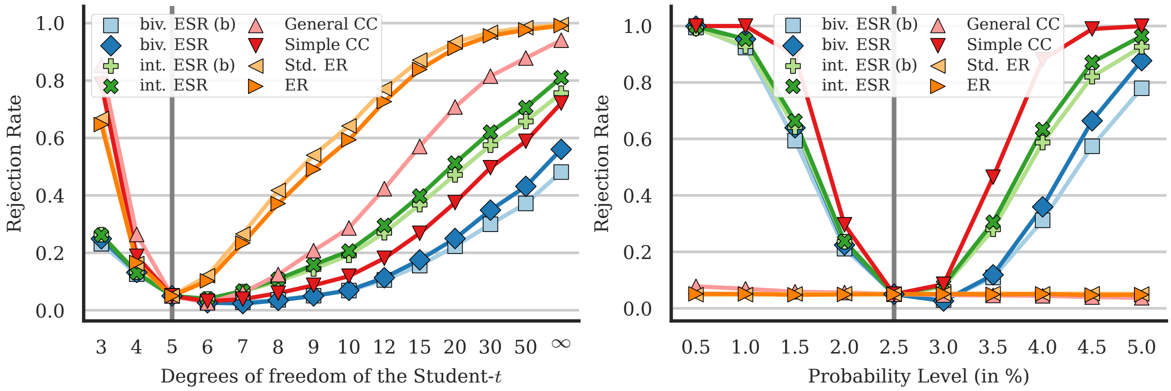


Figure 3.3: Size-adjusted rejection rates for various types of misspecification. The gray vertical line depicts the true model. The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000).

(1) Overall, the bivariate and intercept ESR tests perform similar and in four out of the five considered designs, their performance is superior compared to the general CC and both ER backtesting approaches. (Figures 3.3a to 3.3c and 3.3e). The ESR backtests outperform the competitors especially when we misspecify the volatility dynamics of the underlying GARCH process (Figures 3.3a to 3.3c). This shows that, in contrast to the existing approaches, our new ESR backtests can be used to detect misspecifications in the dynamics used to construct the ES forecasts which go beyond level shifts.
(2) The application of the bootstrap for our ESR tests mainly affects the empirical sizes whereas the empirical size-adjusted power of the asymptotic and the bootstrap ESR tests is similar throughout all designs.
(3) The two ER tests (and the general CC test that is constructed to be similar to the ER backtest) cannot discriminate between forecasts for the VaR and ES issued through misspecified volatility processes (Figures 3.3a to 3.3c) and through misspecified probability levels $\tilde{\tau} \neq \tau$ (Figure 3.3e). This confirms the theoretical results discussed in Section 3.3.1 that these backtests only reject misspecifications which affect the relation (distance) between the VaR and ES forecasts. In contrast, these backtests perform well in the case of misspecified tails of the residual distribution, which affects the relative distance between the VaR and ES forecasts (Figure 3.3d). If these backtests would be used by the regulatory authorities, banks could submit joint VaR and ES forecasts for some level $\tilde{\tau} > \tau$ or some (too small) volatility process in order to minimize their capital requirements without facing the risk of being detected by these backtests. In comparison, our intercept ESR backtest which is similar to the ER backtests by construction is clearly able to identify these misspecified probability levels.
(4) Throughout all five misspecifications, the simple CC backtest also exhibits good power properties, similar to our proposed backtests. However, our two ESR backtests exhibit much better size properties (see Tables 3.1 and 3.2) and in contrast to the simple CC test, they do not fail to reject the Historical Simulation forecasts in the first simulation study (see Figure 3.1).

Together with the results from the first simulation study, these findings show that our proposed ESR backtests are a powerful choice for backtesting ES forecasts. They are reasonably sized and exhibit good power properties against a variety of misspecifications. Notably, in contrast to the existing backtests, there is no single type of misspecification where our ESR tests are unable to discriminate between forecasts of the true and the misspecified models.

3.4.3. Testing One-sided Hypotheses

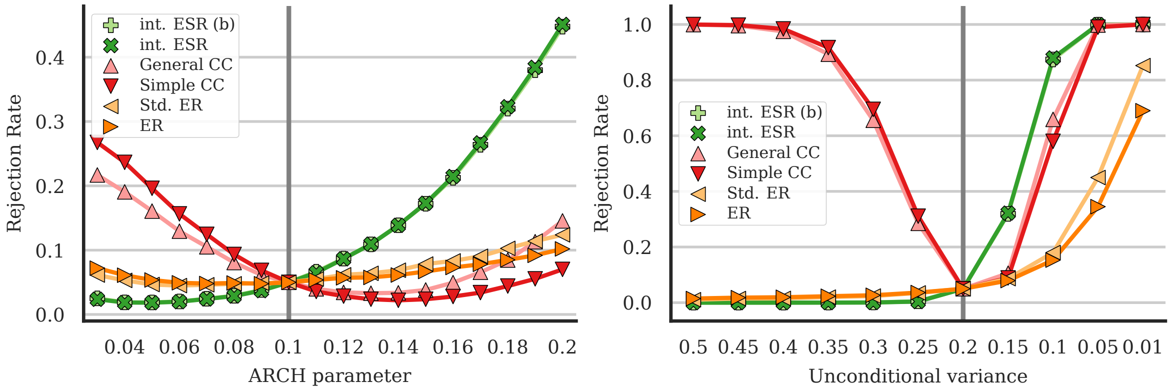
For the regulatory authorities, testing against a one-sided alternative might be more meaningful than the two-sided version we consider in the previous section. Holding more money than stipulated in the Basel accords is no concern for regulators as it is only important that banks keep enough monetary reserves to cover the risk from their market activities. As all backtests (with exception of the bivariate ESR test) allow for testing against one-sided alternatives, we assess their ability to reject the null hypothesis that the issued ES forecasts are smaller or equal to the true ES, i.e. that the associated market risk is not underestimated.

In Figures 3.4a to 3.4e, we present the size-adjusted rejection rates for the one-sided versions of the considered backtests and for the five continuous parameter misspecifications described in the points (a) - (e) from the previous section. The structure of these figures is analog to the two-sided case where the x-axis is oriented such that too small (too risky) ES forecasts are on the left side of the true model (vertical gray line). As it can be seen in Figures 3.2a to 3.2e, the five modifications of the true model exhibit clear patterns when they are over-, respectively underestimating the true ES, where the overestimation holds strictly for the cases (b), (d) and (e) and on average for the cases (a) and (c). Thus, the one-sided backtests should only reject the null hypothesis for ES forecasts that overestimate the true ES, i.e. which are on the right side of the true model in Figures 3.4a to 3.4e.

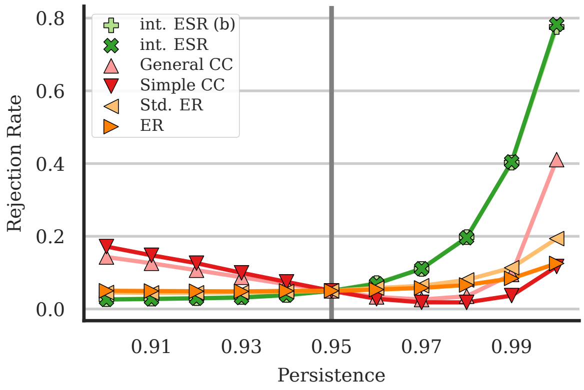
We find that our intercept ESR backtest (in the asymptotic and the bootstrap version) is reasonably sized (compare Table 3.2) and clearly dominates the ER and the CC tests in terms of their power in four out of five misspecification designs. Only when changing the degrees of freedom, the ER test is slightly more powerful than the intercept ESR test. Surprisingly, we see that in four out of the five cases, the one-sided CC tests (both, the simple and the general version) also reject too small (too risky) ES forecasts, even though these should not be rejected by the specifications of the one-sided tests.[11] Furthermore, as for the two-sided tests, both ER backtests fail to detect misspecifications of the underlying volatility process and of the underlying probability level. Summarizing these results, the proposed intercept ESR backtest is a powerful backtest with good size properties for testing one-sided hypotheses which clearly dominates the existing one-sided backtesting techniques in the literature.

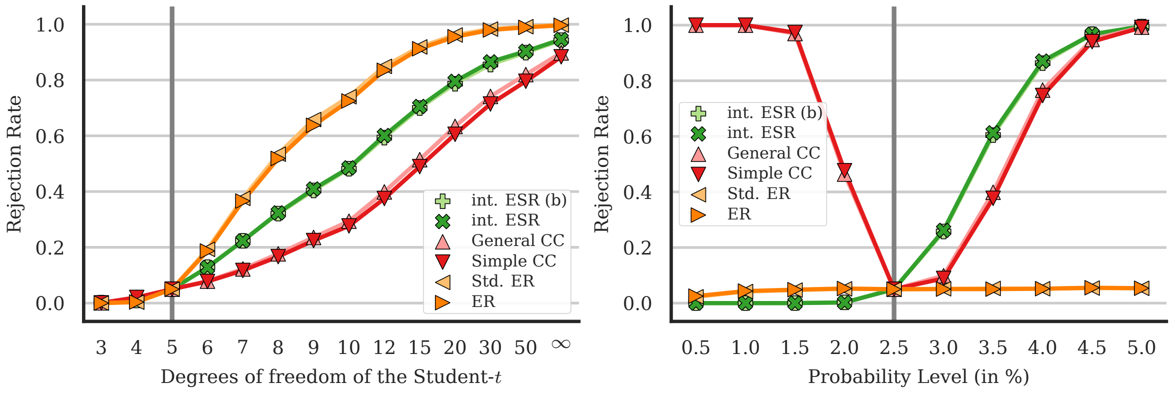
---

[11] We verified our implementation of the CC tests with the codes provided by Nolde and Ziegel (2017) at `https://github.com/nnolde/Elicitability-and-Backtesting/`.


(a) Changing the reaction to the squared returns

(b) Changing the unconditional variance


(c) Changing the persistence


(d) Changing the degrees of freedom

(e) Changing the probability level

Figure 3.4: Size-adjusted rejection rates for various types of misspecification with a one-sided hypothesis. The gray vertical line depicts the true model. The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000).

Table 3.3: Results of the empirical application.

Model	bivariate ESR (b)	bivariate ESR	intercept ESR (b)	intercept ESR	General CC	Simple CC	Std. ER	ER	Mean Loss	MCS p-value
GJR-GARCH-skew-t	0.07	0.12	0.78	0.77	0.85	0.08	0.87	0.14	0.951	1.00
GJR-GARCH-FHS	0.13	0.20	0.30	0.30	0.39	0.69	0.34	0.55	0.953	0.70
GJR-GARCH-t	0.06	0.10	0.06	**0.04**	0.28	0.11	0.23	0.90	0.963	0.28
GJR-GARCH-N	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.982	**0.02**
GARCH-skew-t	0.12	0.19	0.92	0.92	0.33	**0.05**	0.38	0.10	0.986	**0.05**
GARCH-FHS	0.09	0.14	0.19	0.17	0.67	0.31	0.68	0.69	0.993	**0.03**
GARCH-t	**0.03**	**0.05**	**0.05**	**0.03**	0.57	**0.02**	0.58	0.60	1.000	**0.03**
GARCH-N	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.030	**0.01**
RiskMetrics	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.075	**0.00**
Historical Simulation	**0.01**	**0.00**	**0.01**	**0.00**	0.11	**0.01**	0.06	0.06	1.132	**0.01**

Notes: In this table, p-values smaller than 5% are printed bold-faced and the models are sorted by the average loss. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000). We compute the MCS p-values using the R-statistic of Hansen et al. (2011).

3.5. Empirical Application

In the empirical application, we predict the market risk for the daily close-to-close log-returns of the S&P500 index for the time period from January 3, 2000 to October 18, 2017, totaling up to 4478 days. We predict the ES (and the VaR for application of the existing tests) for this return series using 10 different risk models. The first two are the Historical Simulation estimated with a rolling window of 250 days and RiskMetrics. The other 8 models follow the volatility specifications of the GARCH(1,1) model and the asymmetric GJR-GARCH(1,1) model of Glosten et al. (1993) and use four different assumptions on the conditional distribution of the innovations. These are the standard normal distribution (abbreviated by N), the standardized Student-t (t), the standardized skewed Student-t (skew-t) and the semi-parametric filtered historical simulation approach (FHS) of Barone-Adesi et al. (1999), where the quantile, respectively the ES of the innovations is estimated from the standardized returns. We estimate these 8 models on a rolling window of 1000 days.

Table 3.3 presents the p-values of the different ES backtests (for the two-sided hypothesis), the average losses of the strictly consistent 0-homogeneous loss function for the pair VaR and ES[12] (Fissler and Ziegel, 2016), and the p-value of the Model Confidence Set (MCS) of Hansen et al. (2011) applied to this loss function. With the MCS p-values, we can determine a set of models having equal predictive ability at a certain significance level with respect to the losses. The models are sorted according to the average loss.

From this table we can draw several conclusions. First, the MCS rejects 7 out of 10 models at the 5% significance level, i.e. only 3 models have equal predictive power with respect to the joint loss function. These three (GJR-GARCH-skew-t, -FHS, -t) share the

---

[12] This is in fact the loss function given in (3.28), applied to a scenario of forecast comparison.

same assumption on the volatility process and only differ with respect to the assumption on the innovations. Moreover, for these three models the null hypothesis of correct forecasts is not rejected by almost all backtests at the 5% significance level. Thus, the backtests and the MCS agree on which models predict the ES (and the VaR) well. Second, the CC and ER tests reject less forecasts at the 5% significance level than the two ESR backtests, which reflects the findings of the simulation studies where these backtests are often less powerful than our ESR tests. In particular, the null hypothesis is not rejected for the Historical Simulation model, although this approach yields large losses. Third, incorporating leverage into the volatility dynamics appears to be important, since mainly the models using the GJR-GARCH are not rejected by the backtests. Additionally, it is crucial to consider models with flexible tails, e.g. by using the skewed Student-t or the FHS approach, since the models based on conditionally normally distributed returns are collectively rejected by the backtests and the MCS.

3.6. Conclusion

In this paper, we introduce two novel backtests for ES forecasts which regress the realized returns on the issued ES forecasts using an appropriate regression method for the ES introduced in Barendse (2017), Dimitriadis and Bayer (2017), and Patton et al. (2017) and test the resulting parameter estimates. We introduce a bivariate version, denoted *bivariate ESR backtest*, where we test the intercept and the slope parameters for zero and one, and an intercept version, denoted *intercept ESR backtest*, that only incorporates an intercept term being estimated and tested for zero. The motivation for the latter test is the possibility to specify a one-sided hypothesis that is particularly relevant for the regulatory authorities. These backtests can be interpreted as ES-specific versions of the classical Mincer and Zarnowitz (1969) test for evaluating mean forecasts.

A unique feature of the backtests proposed in this paper is that they solely require forecasts for the ES and are consequently the first backtests for the ES stand-alone. In contrast to that, a common drawback of the existing backtests is that they need forecasts of further input parameters, such as the VaR, the volatility, the tail distribution or even the whole return distribution. Using more information than the ES forecasts is problematic for two reasons. First, these tests are not applicable for the regulatory authorities, who receive forecasts of the ES, but not of the additional information required by these tests. Second, rejecting the null hypothesis does not necessarily imply that the ES forecasts are incorrect as the rejection can be a result of a false prediction of any of the input parameters.

In several simulation studies, we assess the empirical size and power properties of our proposed backtests and compare them to the approaches of McNeil and Frey (2000) and Nolde and Ziegel (2017), which jointly backtest the VaR and the ES. We find that our regression-based tests are reasonably sized, especially when they are applied using the bootstrap. Moreover, in almost all simulation designs our two proposed backtests are more powerful than the existing tests. The backtests from the literature are often not able to distinguish between forecasts stemming from the true model and some misspecified model, for instance when we consider a misspecified volatility process or a wrong probability level of the ES. In contrast to that, our two ESR backtests detect the misspecifications in all considered simulation experiments. We provide an implementation of our backtests and of several approaches from the literature in the `esback` package for R (Bayer and Dimitriadis, 2017a).

This paper contributes to the ongoing discussion about which risk measure is the best in practice in the following way. As the VaR is criticized for not being subadditive and for not capturing tail risks beyond itself, the recent literature proposes both, the ES and expectiles as alternative risk measures. Expectiles are suggested as they are coherent, elicitable and are able to capture extreme risks beyond the VaR and thus, they simultaneously overcome the drawbacks of the VaR and the ES (Bellini et al., 2014; Ziegel, 2016). Unfortunately, as opposed to the VaR and ES, they lack a visual and intuitive interpretation (Emmer et al., 2015). In contrast, the ES is mainly criticized for its theoretical deficiencies of being not elicitable and of being backtestable only with difficulties. However, starting with the joint elicitability result of VaR and ES of Fissler and Ziegel (2016), there is a growing body of literature using this result for a regression procedure (Barendse, 2017; Dimitriadis and Bayer, 2017; Patton et al., 2017) and for relative forecast comparison (Fissler et al., 2016; Nolde and Ziegel, 2017), which is extended by this paper through introducing the ESR backtests, which are the first sensible backtests for the ES stand-alone. This shows that, even though technically more demanding, the ES can be modeled, evaluated and backtested in the same way as quantiles and expectiles. Combining this with its ability to capture extreme tail risks and its intuitive visual interpretation, the ES is an appropriate candidate for being the standard risk measure in practice.

Appendix 3.A The Joint Quantile and ES Regression Technique

Assume we have a response variable Y_t and a k-dimensional vector of covariates X_t following the definition of the stochastic process in (3.1) and assume that we are interested in a linear regression technique for the ES at level $\tau \in (0,1)$,

$$Y_t = X_t'\theta^e + u_t^e, \tag{3.24}$$

where $\mathrm{ES}_\tau(u_t^e \mid \mathcal{F}_{t-1}) = 0$ for all $t = 1, \dots, T$, $T \geq 1$. This means that we model the conditional ES through a linear function, $\mathrm{ES}_\tau(Y_t \mid \mathcal{F}_{t-1}) = X_t'\theta^e$. As outlined in Dimitriadis and Bayer (2017), Patton et al. (2017) and Barendse (2017), it is infeasible to estimate the parameter vector θ^e by M- or Z-/GMM- estimation using a semiparametric method without specifying the full conditional distribution of the error term u_t^e. The underlying reason for this is that the functional ES is not elicitable (Gneiting, 2011), i.e. there exists no strictly consistent loss function for the ES, which could be minimized in M-estimation of the regression parameters.[13]

However, as the ES and the quantile at common probability level τ are jointly elicitable (Fissler and Ziegel, 2016), the parameters θ^e in (3.24) can be estimated by jointly modeling a regression equation for the quantile and for the ES,

$$Y_t = X_t'\theta^q + u_t^q \qquad \text{and} \tag{3.25}$$

$$Y_t = X_t'\theta^e + u_t^e, \tag{3.26}$$

where $Q_\tau(u_t^q \mid \mathcal{F}_{t-1}) = 0$ and $\mathrm{ES}_\tau(u_t^e \mid \mathcal{F}_{t-1}) = 0$ for all $t = 1, \dots, T$, $T \geq 1$. Here, $\theta = \left(\theta^q, \theta^e\right)$ denotes the $2k$-dimensional vector of regression parameters of the joint model and the quantile and ES equations are modelled through the separate $k-$dimensional parameter vectors θ^q and θ^e. The M-estimator of the regression parameters θ is obtained by

$$\widehat{\theta}_T = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} \rho(Y_t, X_t, \theta), \tag{3.27}$$

---

[13] In contrast, classical mean regression is built on minimizing the squared loss function, which is a strictly consistent loss function for the mean. Furthermore, quantile regression (Koenker, 2005) is estimated by minimizing the asymmetric linear loss function, a strictly consistent loss function for quantiles and expectile regression (Efron, 1991) by the asymmetric squared error loss, a strictly consistent loss function for expectiles.

where the loss function[14] is given by

$$\rho(Y_t, X_t, \theta) = \frac{1}{-X_t'\theta^e}\left(X_t'\theta^e - X_t'\theta^q + \frac{(X_t'\theta^q - Y_t)\mathbb{1}_{\{Y_t \leq X_t'\theta^q\}}}{\tau} \right) + \log(-X_t'\theta^e). \tag{3.28}$$

Consistency and the asymptotic normality of the M-estimator of θ is shown by Patton et al. (2017) for an α-mixing stochastic process $Z_t = (Y_t, X_t)$. Under the further technical conditions in Assumption 1 and 2 in Patton et al. (2017), it holds that

$$\sqrt{T} C_T^{-1/2} \Lambda_T \big(\hat{\theta}_T - \theta_0\big) \xrightarrow{d} \mathcal{N}\left(0, I\right), \tag{3.29}$$

where θ_0 denotes the unknown true parameter value and where

$$\Lambda_T = \begin{pmatrix} \Lambda_{11,T} & 0 \\ 0 & \Lambda_{22,T} \end{pmatrix} \qquad \text{and} \qquad C_T = \begin{pmatrix} C_{11,T} & C_{12,T} \\ C_{21,T} & C_{22,T} \end{pmatrix}, \tag{3.30}$$

with

$$\Lambda_{11,T} = -\frac{1}{T} \sum_{t=1}^{T} \frac{1}{\alpha} \mathbb{E}\left[(X_t X_t') f_t(X_t'\theta_0^q) / (X_t'\theta_0^e) \right], \tag{3.31}$$

$$\Lambda_{22,T} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[(X_t X_t') / (X_t'\theta_0^e)^2 \right], \tag{3.32}$$

$$C_{11,T} = \frac{1}{T} \sum_{t=1}^{T} \frac{1-\alpha}{\alpha} \mathbb{E}\left[(X_t X_t') / (X_t'\theta_0^e)^2 \right], \tag{3.33}$$

$$C_{12,T} = C_{21,T} = -\frac{1}{T} \sum_{t=1}^{T} \frac{1-\alpha}{\alpha} \mathbb{E}\left[(X_t X_t') \big(X_t'\theta_0^q - X_t'\theta_0^e\big) / (X_t'\theta_0^e)^3 \right], \tag{3.34}$$

$$C_{22,T} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[(X_t X_t') / (X_t'\theta_0^e)^4 \left(\frac{1}{\alpha} \operatorname{Var}_t \big(Y_t - X_t'\theta_0^q \mid Y_t \leq X_t'\theta_0^q \big) + \frac{1-\alpha}{\alpha} \big(X_t'\theta_0^q - X_t'\theta_0^e \big)^2 \right) \right]. \tag{3.35}$$

---

[14] As shown by Dimitriadis and Bayer (2017), consistent and asymptotically normal M-estimation of these regression parameters can be obtained by employing loss functions from a whole class of functions, originally introduced by Fissler and Ziegel (2016) in the context of forecast evaluation. However, consensus seems to emerge on the 0-homogeneous loss function presented in (3.28), see e.g. Barendse (2017), Dimitriadis and Bayer (2017), Patton et al. (2017), and Taylor (2017) and Nolde and Ziegel (2017).

Table 3.B.4: Empirical sizes of the backtests

Nominal Size	Sample Size	bivariate ESR (b)	bivariate ESR	intercept ESR (b)	intercept ESR	General CC	Simple CC	Std. ER	ER
1%	250	0.02	0.10	0.01	0.07	0.01	0.17	0.04	0.04
	500	0.02	0.05	0.01	0.05	0.02	0.08	0.00	0.00
	1000	0.01	0.04	0.01	0.03	0.02	0.05	0.00	0.01
	2500	0.01	0.02	0.01	0.02	0.02	0.03	0.01	0.01
	5000	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01
5%	250	0.08	0.18	0.05	0.13	0.06	0.22	0.06	0.07
	500	0.06	0.12	0.05	0.10	0.07	0.14	0.04	0.04
	1000	0.06	0.09	0.05	0.07	0.07	0.10	0.04	0.04
	2500	0.06	0.07	0.05	0.06	0.06	0.07	0.05	0.05
	5000	0.05	0.06	0.05	0.06	0.05	0.06	0.05	0.05
10%	250	0.14	0.24	0.10	0.18	0.13	0.26	0.11	0.11
	500	0.12	0.18	0.10	0.14	0.13	0.19	0.08	0.08
	1000	0.11	0.14	0.10	0.12	0.12	0.15	0.09	0.09
	2500	0.10	0.12	0.10	0.11	0.11	0.12	0.10	0.10
	5000	0.10	0.11	0.10	0.11	0.11	0.11	0.10	0.10

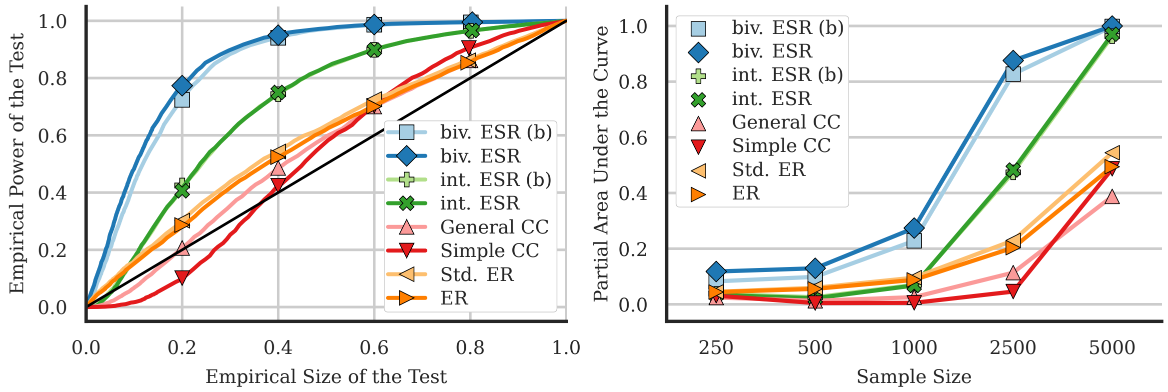
Notes: The table reports the empirical sizes of the backtests for a GARCH(1,1)-N process. The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000).

Appendix 3.B Robustness Check

The DGP used by Gaglianone et al. (2011) is a GARCH(1,1) model with standard normally distributed innovations,

$$\begin{aligned} Y_t &= \sigma_t z_t, \\ \sigma_t^2 &= 0.05 + 0.05 Y_{t-1}^2 + 0.90 \sigma_{t-1}^2, \\ z_t &\sim \mathcal{N}(0, 1). \end{aligned} \tag{3.36}$$

For this DGP, Table 3.B.4 and Figure 3.B.5 present the empirical sizes and the PAUC analog to the results provided in Section 3.4.1.



(a) Size-adjusted Power

(b) Partial Area Under the Curve

Figure 3.B.5: Size-adjusted power and Partial Area Under the Curve plots against Historical Simulation for a sample size of 1000 days. The number of Monte-Carlo repetitions is 10,000 and the probability level for the risk measures is $\tau = 2.5\%$. ESR refers to the backtests introduced in this paper with (b) indicating the bootstrap version, CC to the conditional calibration tests of Nolde and Ziegel (2017), and ER to the exceedance residuals tests of McNeil and Frey (2000).

References

Acerbi, C. and B. Székely (2014). “Backtesting Expected Shortfall”. *Risk* December (see pp. 97, 105).

Aramonte, S., P. Durand, S. Kobayashi, M. Kwast, J. A. Lopez, G. Mazzoni, P. Raupach, M. Summer, and J. Wu (2011). *Messages from the academic literature on risk measurement for the trading book*. Tech. rep. Working Paper No. 19, available at `http://www.bis.org/publ/bcbs_wp19.pdf`. Basel Committee on Banking Supervision (see pp. 97, 105).

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). “Coherent Measures of Risk”. *Mathematical Finance* 9 (3), 203–228 (see pp. 6, 10, 54, 97).

Barendse, S. (2017). “Interquantile Expectation Regression”. Available at `https://ssrn.com/abstract=2937665` (see pp. 55, 98, 102, 120–123).

Barone-Adesi, G., K. Giannopoulos, and L. Vosper (1999). “VaR without correlations for portfolios of derivative securities”. *Journal of Futures Markets* 19 (5), 583–602 (see p. 119).

Basel Committee (1996). *Overview of the Amendment to the Capital Accord to Incorporate Market Risks*. Tech. rep. Available at `http://www.bis.org/publ/bcbs23.pdf`. Basel Committee on Banking Supervision (see pp. 15, 26, 98, 104).

— (2013). *Fundamental review of the trading book: A revised market risk framework*. Tech. rep. Available at `http://www.bis.org/publ/bcbs265.pdf`. Basel Committee on Banking Supervision (see p. 97).

— (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d352.pdf`. Basel Committee on Banking Supervision (see pp. 6, 11, 15, 26, 54, 97, 105).

— (2017). *Pillar 3 disclosure requirements – consolidated and enhanced framework*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d400.pdf`. Basel Committee on Banking Supervision (see pp. 97, 105).

Bayer, S. and T. Dimitriadis (2017a). *esback: Expected Shortfall Backtesting*. R package version 0.1.1, available at https://github.com/BayerSe/esback (see pp. 99, 104, 108, 121).

Bellini, F., B. Klar, A. Müller, and E. R. Gianin (2014). "Generalized quantiles as risk measures". *Insurance: Mathematics and Economics* 54, 41–48 (see p. 121).

Berkowitz, J. (2001). "Testing Density Forecasts, With Applications to Risk Management". *Journal of Business & Economic Statistics* 19 (4), 465–474 (see p. 105).

Carver, L. (2013). "Mooted VAR substitute cannot be back-tested, says top quant". *Risk Magazine* March (see p. 97).

Costanzino, N. and M. Curran (2015). "Backtesting general spectral risk measures with application to expected shortfall". *Risk* March (see p. 105).

Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin (2001). "An Academic Response to Basel II". FMG Special Papers. Financial Markets Group Special Papers, available at `https://EconPapers.repec.org/RePEc:fmg:fmgsps:sp130` (see p. 97).

Dimitriadis, T. and S. Bayer (2017). "A Joint Quantile and Expected Shortfall Regression Framework". arXiv:1704.02213 [math.ST] (see pp. 98, 102, 103, 120–123).

Du, Z. and J. C. Escanciano (2017). "Backtesting Expected Shortfall: Accounting for Tail Risk". *Management Science* 63 (4), 940–958 (see p. 105).

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". *The Annals of Statistics* 7 (1), 1–26 (see pp. 67, 103).

— (1991). "Regression percentiles using asymmetric squared error loss". *Statistica Sinica* 1 (1), 93–125 (see pp. 56, 64, 98, 103, 122).

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*. New York: Chapman and Hall (see p. 106).

Emmer, S., M. Kratz, and D. Tasche (2015). "What Is the Best Risk Measure in Practice? A Comparison of Standard Measures". *Journal of Risk* 18 (2), 31–60 (see pp. 97, 105, 121).

Fissler, T. and J. F. Ziegel (2016). "Higher order elicitability and Osband's principle". *The Annals of Statistics* 44 (4), 1680–1707 (see pp. 7, 12, 28, 54, 56–58, 60, 72, 78, 97, 119, 121–123).

Fissler, T., J. F. Ziegel, and T. Gneiting (2016). "Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting". *Risk Magazine* January, 58–61 (see pp. 58, 64, 65, 97, 121).

Gaglianone, W. P., L. R. Lima, O. Linton, and D. R. Smith (2011). "Evaluating Value-at-Risk Models via Quantile Regression". *Journal of Business & Economic Statistics* 29 (1), 150–160 (see pp. 54, 98, 103, 112, 124).

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". *The Journal of Finance* 48 (5), 1779–1801 (see p. 119).

Gneiting, T. (2011). "Making and Evaluating Point Forecasts". *Journal of the American Statistical Association* 106 (494), 746–762 (see pp. 7, 11, 28, 54–58, 63, 97, 98, 102, 122).

Graham, A. and J. Pál (2014). "Backtesting value-at-risk tail losses on a dynamic portfolio". *The Journal of Risk Model Validation* 8 (2), 59–96 (see p. 105).

Guler, K., P. T. Ng, and Z. Xiao (2017). “Mincer–Zarnowitz quantile and expectile regressions for forecast evaluations under aysmmetric loss functions”. *Journal of Forecasting* 36 (6). for.2462, 651–679 (see pp. 98, 103).

Hansen, P. R., A. Lunde, and J. M. Nason (2011). “The Model Confidence Set”. *Econometrica* 79 (2), 453–497 (see p. 119).

Holden, K. and D. A. Peel (1990). “On Testing For Unbiasedness And Efficiency Of Forecasts”. *The Manchester School* 58 (2), 120–127 (see p. 102).

Kerkhof, J. and B. Melenberg (2004). “Backtesting for risk-based regulatory capital”. *Journal of Banking & Finance* 28 (8), 1845–1865 (see pp. 97, 105).

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press (see pp. 55, 56, 63, 122).

Koenker, R. and G. Bassett (1978). “Regression quantiles”. *Econometrica* 46 (1), 33–50 (see pp. 98, 103).

Komunjer, I. (2013). “Quantile Prediction”. In: *Handbook of Economic Forecasting*. Vol. 2. Elsevier. Chap. 17, 961–994 (see pp. 54, 73, 100).

Kratz, M., Y. H. Lok, and A. J. McNeil (2018). “Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall”. *Journal of Banking & Finance* 88 (C), 393–407 (see p. 105).

Lloyd, C. J. (2005). “Estimating test power adjusted for size”. *Journal of Statistical Computation and Simulation* 75 (11), 921–933 (see pp. 110, 111).

MacKinnon, J. G. (2009). “Bootstrap Hypothesis Testing”. In: *Handbook of Computational Econometrics*. Ed. by Belsley, D. A. and Kontoghiorghes, E. J. John Wiley & Sons, Ltd. Chap. 6, 183–213 (see p. 103).

McNeil, A. J. and R. Frey (2000). “Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach”. *Journal of Empirical Finance* 7 (3–4), 271–300 (see pp. 98, 99, 105, 106, 108, 110–112, 115, 118, 119, 121, 124, 125).

Mincer, J. and V. Zarnowitz (1969). “The Evaluation of Economic Forecasts”. In: *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. National Bureau of Economic Research, Inc, 3–46 (see pp. 98, 103, 120).

Nadarajah, S., B. Zhang, and S. Chan (2014). “Estimation methods for expected shortfall”. *Quantitative Finance* 14 (2), 271–291 (see pp. 55, 100).

Nelson, D. B. (1991). “Conditional Heteroskedasticity in Asset Returns: A New Approach”. English. *Econometrica* 59 (2), 347–370 (see p. 109).

Nolde, N. and J. F. Ziegel (2017). “Elicitability and backtesting: Perspectives for banking regulation”. *Annals of Applied Statistics* 11 (4), 1833–1874 (see pp. 56, 64, 65, 69, 97–99, 105, 107, 108, 110–112, 115, 117–119, 121, 123–125).

Patton, A. J., J. F. Ziegel, and R. Chen (2017). “Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)”. arXiv:1707.05108 [q-fin.EC] (see pp. 55, 73, 98, 102, 103, 120–123).

Righi, M. B. and P. S. Ceretta (2013). “Individual and flexible expected shortfall backtesting”. *Journal of Risk Model Validation* 7 (3), 3–20 (see p. 105).

— (2015). “A comparison of Expected Shortfall estimation models”. *Journal of Economics and Business* 78, 14–47 (see p. 105).

Taylor, J. W. (2017). “Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution”. *Journal of Business & Economic Statistics* 0 (0), 1–13 (see pp. 55, 123).

Weber, S. (2006). “Distribution Invariant Risk Measures, Information, and Dynamic Consistency”. *Mathematical Finance* 16 (2), 419–441 (see pp. 57, 97).

Wong, W. (2008). “Backtesting trading risk of commercial banks using expected shortfall”. *Journal of Banking & Finance* 32 (7), 1404–1415 (see p. 105).

Yamai, Y. and T. Yoshiba (2002). “On the validity of value-at-risk: comparative analyses with expected shortfall”. *Monetary and Economic Studies* 20 (1), 57–85 (see p. 97).

Ziegel, J. F. (2016). “Coherence and Elicitability”. *Mathematical Finance* 26 (4), 901–918 (see pp. 97, 121).

Complete References

Acerbi, C. and B. Székely (2014). “Backtesting Expected Shortfall”. *Risk* December (see pp. 97, 105).

Alessio, E., A. Carbone, G. Castelli, and V. Frappietro (2002). “Second-order moving average and scaling of stochastic time series”. *The European Physical Journal B - Condensed Matter and Complex Systems* 27 (2), 197–200 (see p. 26).

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). “The Distribution of Realized Exchange Rate Volatility”. *Journal of the American Statistical Association* 96, 42–55 (see p. 30).

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). “The Distribution of Realized Stock Return Volatility”. *Journal of Financial Economics* 61, 43–76 (see p. 25).

Andersen, T. G. and T. Bollerslev (1997). “Intraday Periodicity and Volatility Persistence in Financial Markets”. *Journal of Empirical Finance* 4, 115–158 (see p. 30).

Andrews, D. (1994). “Empirical Process Methods in Econometrics”. In: *Handbook of Econometrics*. Ed. by Engle, R. and McFadden, D. Vol. 4. Elsevier. Chap. 37, 2247–2294 (see p. 75).

Aramonte, S., P. Durand, S. Kobayashi, M. Kwast, J. A. Lopez, G. Mazzoni, P. Raupach, M. Summer, and J. Wu (2011). *Messages from the academic literature on risk measurement for the trading book*. Tech. rep. Working Paper No. 19, available at `http://www.bis.org/publ/bcbs_wp19.pdf`. Basel Committee on Banking Supervision (see pp. 97, 105).

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). “Coherent Measures of Risk”. *Mathematical Finance* 9 (3), 203–228 (see pp. 6, 10, 54, 97).

Bandi, F. M. and J. R. Russell (2008). “Microstructure Noise, Realized Volatility and Optimal Sampling”. *The Review of Economic Studies* 75 (2), 339–369 (see p. 17).

Barendse, S. (2017). “Interquantile Expectation Regression”. Available at `https://ssrn.com/abstract=2937665` (see pp. 55, 98, 102, 120–123).

Barndorff-Nielsen, O. E., P. Hansen, A. Lunde, and N. Shephard (2008). “Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise”. *Econometrica* 76 (6), 1481–1536 (see p. 17).

Barone-Adesi, G., K. Giannopoulos, and L. Vosper (1999). “VaR without correlations for portfolios of derivative securities”. *Journal of Futures Markets* 19 (5), 583–602 (see p. 119).

Basel Committee (1996). *Overview of the Amendment to the Capital Accord to Incorporate Market Risks*. Tech. rep. Available at `http://www.bis.org/publ/bcbs23.pdf`. Basel Committee on Banking Supervision (see pp. 15, 26, 98, 104).

Basel Committee (2013). *Fundamental review of the trading book: A revised market risk framework*. Tech. rep. Available at `http://www.bis.org/publ/bcbs265.pdf`. Basel Committee on Banking Supervision (see p. 97).

— (2016). *Minimum capital requirements for Market Risk*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d352.pdf`. Basel Committee on Banking Supervision (see pp. 6, 11, 15, 26, 54, 97, 105).

— (2017). *Pillar 3 disclosure requirements – consolidated and enhanced framework*. Tech. rep. Available at `http://www.bis.org/bcbs/publ/d400.pdf`. Basel Committee on Banking Supervision (see pp. 97, 105).

Bayer, S. and T. Dimitriadis (2017a). *esback: Expected Shortfall Backtesting*. R package version 0.1.1, available at https://github.com/BayerSe/esback (see pp. 99, 104, 108, 121).

— (2017b). *esreg: Joint (VaR, ES) Regression*. R package version 0.2.0, available at `https://github.com/BayerSe/esreg` (see pp. 56, 66, 73).

— (2017c). "Regression-based Expected Shortfall Backtesting". Working Paper (see p. 73).

Bellini, F., B. Klar, A. Müller, and E. R. Gianin (2014). "Generalized quantiles as risk measures". *Insurance: Mathematics and Economics* 54, 41–48 (see p. 121).

Beran, J., Y. Feng, S. Ghosh, and R. Kulik (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer (see p. 18).

Berkowitz, J. (2001). "Testing Density Forecasts, With Applications to Risk Management". *Journal of Business & Economic Statistics* 19 (4), 465–474 (see p. 105).

Bianchi, C. and G. Calzolari (1980). "The One-Period Forecast Errors in Nonlinear Econometric Models". *International Economic Review* 21, 201–208 (see p. 31).

Black, F. and M. S. Scholes (1973). "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy* 81 (3), 637–654 (see p. 21).

Boudt, K., C. Croux, and S. Laurent (2011). "Robust Estimation of Intraweek Periodicity in Volatility and Jump Detection". *Journal of Empirical Finance* 18 (2), 353–367 (see pp. 21, 25, 30).

Brazauskas, V., B. L. Jones, M. L. Puri, and R. Zitikis (2008). "Estimating conditional tail expectation with actuarial applications in view". *Journal of Statistical Planning and Inference* 138 (11), 3590–3604 (see p. 63).

Burns, A. F. and W. C. Mitchell (1946). "Measuring Business Cycles". New York, New York: National Bureau of Economic Research (see p. 16).

Calvet, L. E. and A. J. Fisher (2001). "Forecasting Multifractal Volatility". *Journal of Econometrics* 105 (1), 27–58 (see p. 16).

— (2002). "Multifractality in Asset Returns: Theory and Evidence". *Review of Economics and Statistics* 84 (3), 381–406 (see pp. 16, 19, 21).

— (2004). "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes". *Journal of Financial Econometrics* 2 (1), 49–83 (see p. 20).

— (2008). *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press (see p. 16).

Calvet, L. E., A. J. Fisher, and B. Mandelbrot (1997). "Large Deviation and the Distribution of Price Changes". Cowles Foundation Discussion Paper (see p. 16).

Carver, L. (2013). "Mooted VAR substitute cannot be back-tested, says top quant". *Risk Magazine* March (see p. 97).

Chen, S. X. (2008). "Nonparametric Estimation of Expected Shortfall". *Journal of Financial Econometrics* 6 (1), 87–107 (see p. 63).

Chernozhukov, V. and L. Umantsev (2001). “Conditional value-at-risk: Aspects of modeling and estimation”. *Empirical Economics* 26 (1), 271–292 (see pp. 54, 73).

Clark, P. K. (1973). “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices”. *Econometrica* 41 (1), 135–155 (see pp. 16, 21, 25).

Comte, F. and E. Renault (1996). “Long Memory Continuous Time Models”. *Journal of Econometrics* 73 (1), 101–149 (see p. 21).

— (1998). “Long Memory in Continuous-Time Stochastic Volatility Models”. *Mathematical Finance* 8 (4), 291–323 (see p. 21).

Corsi, F. (2009). “A Simple Approximate Long-Memory Model of Realized Volatility”. *Journal of Financial Econometrics* 7 (2), 174–196 (see p. 30).

Corsi, F., G. Zumbach, U. A. Müller, and M. A. Dacorogna (2001). “Consistent High-Precision Volatility from High-Frequency Data”. *Economic Notes by Banca Monte dei Paschi die Siena SpA* 30 (2), 183–204 (see p. 17).

Costanzino, N. and M. Curran (2015). “Backtesting general spectral risk measures with application to expected shortfall”. *Risk* March (see p. 105).

Dacorogna, M. M., U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet (1993). “A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market”. *Journal of International Money and Finance* 12 (4), 413–438 (see p. 17).

Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin (2001). “An Academic Response to Basel II”. FMG Special Papers. Financial Markets Group Special Papers, available at `https://EconPapers.repec.org/RePEc:fmg:fmgsps:sp130` (see p. 97).

Diebold, F. and R. Mariano (1995). “Comparing Predictive Accuracy”. *Journal of Business and Economic Statistics* 13 (3), 253–63 (see p. 29).

Dimitriadis, T. and S. Bayer (2017). “A Joint Quantile and Expected Shortfall Regression Framework”. arXiv:1704.02213 [math.ST] (see pp. 98, 102, 103, 120–123).

Dong, Y. and Y.-K. Tse (2014). “Business Time Sampling Scheme and Its Application”. Working Paper (see pp. 17, 21, 25, 30).

Du, Z. and J. C. Escanciano (2017). “Backtesting Expected Shortfall: Accounting for Tail Risk”. *Management Science* 63 (4), 940–958 (see p. 105).

Efron, B. (1979). “Bootstrap Methods: Another Look at the Jackknife”. *The Annals of Statistics* 7 (1), 1–26 (see pp. 67, 103).

— (1991). “Regression percentiles using asymmetric squared error loss”. *Statistica Sinica* 1 (1), 93–125 (see pp. 56, 64, 98, 103, 122).

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*. New York: Chapman and Hall (see p. 106).

Embrechts, P. and M. Maejima (2002). *Selfsimilar Processes*. Princeton University Press (see p. 18).

Emmer, S., M. Kratz, and D. Tasche (2015). “What Is the Best Risk Measure in Practice? A Comparison of Standard Measures”. *Journal of Risk* 18 (2), 31–60 (see pp. 97, 105, 121).

Engle, R. and S. Manganelli (2004). “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles”. *Journal of Business and Economic Statistics* 22 (4), 367–381 (see pp. 8, 12, 54, 73).

Engle, R. F. and J. R. Russell (1998). “Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data”. *Econometrica* 66 (5), 1127–1162 (see pp. 25, 30).

Fissler, T. (2017). “On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space”. PhD thesis. Universität Bern (see p. 66).

Fissler, T. and J. F. Ziegel (2016). “Higher order elicitability and Osband’s principle”. *The Annals of Statistics* 44 (4), 1680–1707 (see pp. 7, 12, 28, 54, 56–58, 60, 72, 78, 97, 119, 121–123).

Fissler, T., J. F. Ziegel, and T. Gneiting (2016). “Expected Shortfall is jointly elicitable with Value at Risk - Implications for backtesting”. *Risk Magazine* January, 58–61 (see pp. 58, 64, 65, 97, 121).

Gaglianone, W. P., L. R. Lima, O. Linton, and D. R. Smith (2011). “Evaluating Value-at-Risk Models via Quantile Regression”. *Journal of Business & Economic Statistics* 29 (1), 150–160 (see pp. 54, 98, 103, 112, 124).

Ghysels, E., C. Gouriéroux, and J. Jasiak (1997). “Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets”. In: *Nonlinear Modelling in Foreign Exchange Markets*. Ed. by Dunis, C. and Zhou, B. New York: Wiley (see p. 17).

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). “Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies”. *Journal of Econometrics* 131 (1-2), 59–95 (see p. 30).

Giacomini, R. and I. Komunjer (2005). “Evaluation and Combination of Conditional Quantile Forecasts”. *Journal of Business & Economic Statistics* 23, 416–431 (see pp. 8, 12).

Gikhman, I. and A. Skorokhod (2004). *The Theory of Stochastic Processes I*. Vol. 210. Classics in Mathematics. Springer Berlin Heidelberg (see pp. 75, 92).

Giot, P. and S. Laurent (2004). “Modelling daily Value-at-Risk using realized volatility and ARCH type models”. *Journal of Empirical Finance* 11 (3), 379–398 (see p. 15).

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”. *The Journal of Finance* 48 (5), 1779–1801 (see p. 119).

Gneiting, T. (2011). “Making and Evaluating Point Forecasts”. *Journal of the American Statistical Association* 106 (494), 746–762 (see pp. 7, 11, 28, 54–58, 63, 97, 98, 102, 122).

Gouriéroux, C. and J. Jasiak (2001). *Financial Econometrics*. Princeton: Princeton University Press (see pp. 16, 17, 21).

Gouriéroux, C. and A. Monfort (1995). *Statistics and Econometric Models: Volume 1, General Concepts, Estimation, Prediction and Algorithms*. Cambridge University Press (see p. 68).

Graham, A. and J. Pál (2014). “Backtesting value-at-risk tail losses on a dynamic portfolio”. *The Journal of Risk Model Validation* 8 (2), 59–96 (see p. 105).

Griffin, J. and R. Oomen (2008). “Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?” *Econometric Reviews* 27 (1-3), 230–253 (see pp. 25, 30).

Guillaume, D. M., M. M. Dacorogna, R. R. Dave, U. A. Müller, R. B. Olsen, and O. V. Pictet (1997). “From the Bird’s Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets”. *Finance and Stochastics* 1 (2), 95–129 (see p. 17).

Guler, K., P. T. Ng, and Z. Xiao (2017). “Mincer–Zarnowitz quantile and expectile regressions for forecast evaluations under aysmmetric loss functions”. *Journal of Forecasting* 36 (6). for.2462, 651–679 (see pp. 98, 103).

Halbleib, R. and W. Pohlmeier (2012). “Improving the Value at Risk Forecasts: Theory and Evidence from the Financial Crisis”. *Journal of Economic Dynamics and Control* 36 (8), 1212–1228 (see pp. 15, 28, 54, 73).

Halbleib, R. and V. Voev (2011). “Forecasting Multivariate Volatility Using the VARFIMA Model on Realized Covariance Cholesky Factors”. *Journal of Economics and Statistics* 231 (1), 134–152 (see p. 31).

Hall, P. and S. J. Sheather (1988). “On the Distribution of a Studentized Quantile”. *Journal of the Royal Statistical Society. Series B (Methodological)* 50 (3), 381–391 (see p. 67).

Hallam, M. and J. Olmo (2014a). “Forecasting Daily Return Densities from Intraday Data: A Multifractal Approach”. *International Journal of Forecasting* 30 (4), 863–881 (see pp. 16, 17, 20, 27).

— (2014b). “Semiparametric Density Forecasts of Daily Financial Returns from Intraday Data”. *Journal of Financial Econometrics* 12 (2), 408–432 (see pp. 16, 28).

Hansen, P. R., A. Lunde, and J. M. Nason (2011). “The Model Confidence Set”. *Econometrica* 79 (2), 453–497 (see p. 119).

Hansen, P. R. and A. Lunde (2004). “Realized Variance and IID Market Microstructure Noise”. Brown University Working Paper (see p. 17).

— (2006). “Realized Variance and Market Microstructure Noise”. *Journal of Business and Economic Statistics* 24, 127–218 (see p. 26).

Harrell, F. and E. Davis (1982). “A New Distribution-Free Quantile Estimator”. *Biometrika* 69 (3), 635–640 (see p. 26).

Harris, L. (1986). “A transaction data study of weekly and intradaily patterns in stock returns”. *Journal of Financial Economics* 16 (1), 99–117 (see p. 30).

Hendricks, W. and R. Koenker (1992). “Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity”. *Journal of the American Statistical Association* 87 (417), 58–68 (see p. 67).

Holden, K. and D. A. Peel (1990). “On Testing For Unbiasedness And Efficiency Of Forecasts”. *The Manchester School* 58 (2), 120–127 (see p. 102).

Huber, P. (1967). “The behavior of maximum likelihood estimates under nonstandard conditions”. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 221–233 (see pp. 75, 77, 78).

Kantelhardt, J. W. (2009). “Fractal and Multifractal Time Series”. In: *Mathematics of Complexity and Dynamical Systems*. Ed. by Meyers, R. A. New York: Springer New York, 463–487 (see pp. 20, 24).

Kerkhof, J. and B. Melenberg (2004). “Backtesting for risk-based regulatory capital”. *Journal of Banking & Finance* 28 (8), 1845–1865 (see pp. 97, 105).

Kobeissi, Y. H. (2013). *Multifractal Financial Markets*. Springer (see p. 20).

Koenker, R. (1994). “Confidence Intervals for Regression Quantiles”. In: *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium, held from September 4–9, 1993*. Ed. by Mandl, P. and Hušková, M. Physica-Verlag Heidelberg, 349–359 (see p. 67).

— (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press (see pp. 55, 56, 63, 122).

Koenker, R. and G. Bassett (1978). “Regression quantiles”. *Econometrica* 46 (1), 33–50 (see pp. 98, 103).

Koenker, R. and J. A. F. Machado (1999). “Goodness of Fit and Related Inference Processes for Quantile Regression”. *Journal of the American Statistical Association* 94 (448), 1296–1310 (see p. 63).

Koenker, R. and Z. Xiao (2006). “Quantile Autoregression”. *Journal of the American Statistical Association* 101 (475), 980–990 (see pp. 54, 73).

Komunjer, I. (2013). “Quantile Prediction”. In: *Handbook of Economic Forecasting*. Vol. 2. Elsevier. Chap. 17, 961–994 (see pp. 54, 73, 100).

Kratz, M., Y. H. Lok, and A. J. McNeil (2018). “Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall”. *Journal of Banking & Finance* 88 (C), 393–407 (see p. 105).

Kuester, K., S. Mitinik, and M. S. Paolella (2006). “Value-at-Risk Prediction: A Comparison of Alternative Strategies”. *Journal of Financial Econometrics* 4 (1), 53–89 (see p. 26).

Lambert, N. S., D. M. Pennock, and Y. Shoham (2008). “Eliciting Properties of Probability Distributions”. In: *Proceedings of the 9th ACM Conference on Electronic Commerce*. ACM, 129–138 (see p. 56).

Lloyd, C. J. (2005). “Estimating test power adjusted for size”. *Journal of Statistical Computation and Simulation* 75 (11), 921–933 (see pp. 110, 111).

Lourenço, H. R., O. C. Martin, and T. Stützle (2003). “Iterated Local Search”. In: *Handbook of Metaheuristics*. Ed. by Glover, F. and Kochenberger, G. A. Boston, MA: Springer US, 320–353 (see p. 66).

Lux, T. and M. Segnon (2018). “Multifractal models in finance: their origin, properties, and applications”. In: *The Oxford Handbook of Computational Economics and Finance*. Ed. by Shu-Heng Chen, M. K. and Du, Y.-R. Oxford University Press. Chap. 6 (see pp. 16, 20, 21).

MacKinnon, J. G. (2009). “Bootstrap Hypothesis Testing”. In: *Handbook of Computational Econometrics*. Ed. by Belsley, D. A. and Kontoghiorghes, E. J. John Wiley & Sons, Ltd. Chap. 6, 183–213 (see p. 103).

Mandelbrot, B. B. (1963). “The Variation of Certain Speculative Prices”. *Journal of Business* 36, 394–419 (see p. 18).

— (1974). “Intermittent Turbulence in Self Similar Cascades: Divergence of High Moments and Dimension of the Carrier”. *Journal of Fluid Mechanics* 62 (2), 331–358 (see p. 19).

— (1982). *The Fractal Geometry of Nature*. W. H. Freeman and Company, New York (see p. 18).

— (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, New York, 394–419 (see p. 21).

— (1999). *Multifractals and 1/f Noise*. Springer Verlag, New York (see p. 16).

Mandelbrot, B. B., A. Fisher, and L. Calvet (1997). “A Multifractal Model of Asset Returns”. Cowles Foundation Discussion Paper (see pp. 16, 19–21).

Mandelbrot, B. B. and H. M. Taylor (1967). “On the Distribution of Stock Price Differences”. *Operations Research* 15 (6), 1057–1062 (see pp. 16, 25).

Markowitz, H. (1952). “Portfolio Selection”. *Journal of Finance* 7 (1), 77–91 (see p. 21).

McAleer, M. and M. C. Medeiros (2008). “Realized Volatility: A Review”. *Econometric Reviews* 27 (1-3), 10–45 (see p. 17).

McNeil, A. J. and R. Frey (2000). “Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach”. *Journal of Empirical Finance* 7 (3–4), 271–300 (see pp. 98, 99, 105, 106, 108, 110–112, 115, 118, 119, 121, 124, 125).

Mincer, J. and V. Zarnowitz (1969). “The Evaluation of Economic Forecasts”. In: *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. National Bureau of Economic Research, Inc, 3–46 (see pp. 98, 103, 120).

Müller, U. A., M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen, and J. R. Ward (1995). “Fractals and Intrinsic Time - A Challenge to Econometricians”. Working Paper Series, Olsen&Associates (see p. 17).

Nadarajah, S., B. Zhang, and S. Chan (2014). “Estimation methods for expected shortfall”. *Quantitative Finance* 14 (2), 271–291 (see pp. 55, 100).

Nelder, J. A. and R. Mead (1965). “A Simplex Method for Function Minimization”. *The Computer Journal* 7 (4), 308–313 (see p. 66).

Nelson, D. B. (1991). “Conditional Heteroskedasticity in Asset Returns: A New Approach”. English. *Econometrica* 59 (2), 347–370 (see p. 109).

Newey, W. and D. McFadden (1994). “Large sample estimation and hypothesis testing”. In: *Handbook of Econometrics*. Ed. by Engle, R. and McFadden, D. Vol. 4. Elsevier. Chap. 36, 2111–2245 (see pp. 75, 77).

Nolde, N. and J. F. Ziegel (2017). “Elicitability and backtesting: Perspectives for banking regulation”. *Annals of Applied Statistics* 11 (4), 1833–1874 (see pp. 56, 64, 65, 69, 97–99, 105, 107, 108, 110–112, 115, 117–119, 121, 123–125).

Oomen, R. C. A. (2001). “Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realized Return Variance”. European University, Economics Discussion Paper No. 2001/6 (see p. 31).

— (2005). “Properties of Bias-Corrected Realized Variance under Alternative Sampling Schemes”. *Journal of Financial Econometrics* 3 (4), 555–577 (see pp. 17, 21, 25, 30).

— (2006). “Properties of Realized Variance Under Alternative Sampling Schemes”. *Journal of Business and Economic Statistics* 24 (2), 219–237 (see pp. 17, 25, 30).

Patton, A. J., J. F. Ziegel, and R. Chen (2017). “Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)”. arXiv:1707.05108 [q-fin.EC] (see pp. 55, 73, 98, 102, 103, 120–123).

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994). “Mosaic organization of DNA nucleotides”. *Phys. Rev. E* 49 (2), 1685–1689 (see p. 26).

Righi, M. B. and P. S. Ceretta (2013). “Individual and flexible expected shortfall backtesting”. *Journal of Risk Model Validation* 7 (3), 3–20 (see p. 105).

— (2015). “A comparison of Expected Shortfall estimation models”. *Journal of Economics and Business* 78, 14–47 (see p. 105).

Sánchez-Granero, M., M. Fernández-Martínez, and J. Trinidad-Segovia (2012). “Introducing fractal dimension algorithms to calculate the Hurst exponent of financial time series”. *The European Physical Journal B* 85 (3) (see p. 26).

Stock, J. (1987). “Measuring Business Cycle Time”. *Journal of Political Economy* 95 (6), 1240–1261 (see p. 17).

— (1988). “Estimating Continuous Time Processes Subject to Time Deformation”. *Journal of the American Statistical Association* 83 (401), 77–85 (see p. 17).

Taylor, J. W. (2008a). “Estimating Value at Risk and Expected Shortfall Using Expectiles”. *Journal of Financial Econometrics* 6 (2), 231–252 (see p. 55).

— (2008b). “Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall”. *Journal of Financial Econometrics* 6 (3), 382–406 (see p. 55).

— (2017). “Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution”. *Journal of Business & Economic Statistics* 0 (0), 1–13 (see pp. 55, 123).

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press (see pp. 75, 77, 79, 80).

Weber, S. (2006). “Distribution Invariant Risk Measures, Information, and Dynamic Consistency”. *Mathematical Finance* 16 (2), 419–441 (see pp. 57, 97).

Wong, W. (2008). “Backtesting trading risk of commercial banks using expected shortfall”. *Journal of Banking & Finance* 32 (7), 1404–1415 (see p. 105).

Wu, Z. (2012). “On the Intraday Periodicity Duration Adjustment of High-Frequency Data”. *Journal of Empirical Finance* 19 (4), 282–291 (see p. 25).

Xiao, Z., H. Guo, and M. S. Lam (2015). “Quantile Regression and Value at Risk”. In: *Handbook of Financial Econometrics and Statistics*. Ed. by Lee, C.-F. and Lee, J. C. Springer, 1143–1167 (see pp. 54, 73).

Yamai, Y. and T. Yoshiba (2002). “On the validity of value-at-risk: comparative analyses with expected shortfall”. *Monetary and Economic Studies* 20 (1), 57–85 (see p. 97).

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data”. *Journal of the American Statistical Association* 100 (472), 1394–1411 (see pp. 17, 26).

Ziegel, J. F. (2016). “Coherence and Elicitability”. *Mathematical Finance* 26 (4), 901–918 (see pp. 97, 121).

Ziegel, J. F., F. Krüger, A. Jordan, and F. Fasciati (2017). “Murphy Diagrams: Forecast Evaluation of Expected Shortfall”. arXiv:1705.04537 [q-fin.RM] (see pp. 65, 69).

Žikeš, F. and J. Baruník (2016). “Semi-parametric Conditional Quantile Models for Financial Returns and Realized Volatility”. *Journal of Financial Econometrics* 14 (1), 185–226 (see pp. 8, 12, 54, 73).

Zwingmann, T. and H. Holzmann (2016). “Asymptotics for the expected shortfall”. arXiv:1611.07222 [math.ST] (see p. 63).

Eigenabgrenzung

Das erste Kapitel, *How Informative is High-Frequency Data for Tail Risk Estimation and Forecasting? An Intrinsic Time Perspective*, ist in Zusammenarbeit mit meiner Betreuerin Dr. Roxana Halbleib von der Universität Konstanz entstanden. Meine individuelle Leistung bei der Erstellung dieses Kapitels liegt bei etwa 75%.

Das zweite und dritte Kapitel, *A Joint Quantile and Expected Shortfall Regression Framework* und *Regression Based Expected Shortfall Backtesting*, sind in Zusammenarbeit mit Dr. Sebastian Bayer entstanden, der ebenfalls Dokotorand an der Graduate School of Decision Sciences der Universität Konstanz war. Meine individuellen Leistungen bei der Erstellung der Kapitel betragen 60% für das zweite und 50% für das dritte Kapitel.