Contributions to Model Risk

Der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
- Doctor rerum politicarum -

genehmigte Dissertation

von

Dipl.-Ök. Corinna Evers geb. Luedtke

geboren am 11. Juli 1981 in Langenhagen

2014

Referent: Prof. Dr. Philipp Sibbertsen

Koreferent: Prof. Dr. Daniel Rösch

Tag der Promotion: 20. März 2013

Die finale Version der vorliegenden Arbeit möchte ich zum Anlass nehmen, den zahlreichen Personen zu danken, die mich über viele Jahre begleitet und unterstützt haben:

Meinem Doktorvater Professor Dr. Philipp Sibbertsen danke ich für die jahrelange fachliche Betreuung und Begleitung im Rahmen von Studium und Promotion sowie das stets offene Ohr für alle die Arbeit begleitenden Anliegen.

Herrn Professor Dr. Daniel Rösch bin ich zum Dank verpflichtet zum einen durch die übernommene Zweitgutachterschaft und für die Schlüsselrolle bei der Veröffentlichung der ersten zwei Hauptkapitel dieser Arbeit.

Die Zusammenarbeit mit meinen Koautoren Dr. Gerhard Stahl, Grigoriy Tymchenko und Johannes Rohde hat die Arbeit in jeder individuellen Art und Weise geprägt. Insbesondere meinem Kollegen Johannes Rohde danke ich für seine stets kritische und reflektierende Sicht auf das Werk.

Meinen Kollegen Meike Gebel, Juliane Willert, Florian Heinen, Philip Bertram, Hendrik Kaufmann, und Claudia Grote danke ich für die gemeinsame Zeit und die anregenden fachlichen Diskussionen.

Allen weiteren Wegbegleitern am Institut für Statistik möchte ich auf diesem Wege für die Unterstützung danken, die indirekt durch die Übernahme kleiner Aufgaben doch in der Summe sehr groß war. Ganz besonders erwähnen möchte ich Gudrun Westphal, die mir mit Rat und Tat bei den administrativen Schritten zur Seite stand und immer für heitere Momente gesorgt hat. Den studentischen Hilfskräften im Erasmus-Büro danke ich für die stets exzellente Arbeit und Unterstützung in meiner dortigen Funktion.

Meinem Ehemann Kai danke ich in jeglicher Hinsicht, insbesondere jedoch für das unerschütterliche Vertrauen, nimmermüde Motivation und Glaube an mich.

Meinen Eltern gebührt der Dank für die Grundlegung meiner Ausbildung.

Zusammenfassung

Diese Arbeit liefert eine genaue Definition des Begriffes Modellrisiko. Ein falsches Modell kann zu erheblicher Über- oder Unterschätzung des Risikos einer Finanzinstitution führen. Weil der zugrundeliegende datengenerierende Prozess in der Praxis unbekannt ist, ist die Bewertung des Modellrisikos eine große Herausforderung. Bislang zu findende Definitionen von Modellrisiko waren entweder anwendungsorientiert und beinhalteten das Risiko, welches vielmehr durch den Statistiker denn durch das statistische Modell selbst induziert wird oder zu wissenschaftlich und entsprechend zu abstrakt um in der Praxis umgesetzt zu werden. Wir führen einen datengetriebenen Modellrisikobegriff ein, der Merkmale des wissenschaftlichen Ansatzes um einen statistischen Modellierungsprozess erweitert. Ferner schlagen wir die Anwendung robuster Schätzer zur Reduzierung des Modellrisikos vor und empfehlen die Anwendung von Stresstests zur Portfoliobewertung.

Weiterhin untersuchen wir inwieweit die Fehlspezifikation eines zugrundeliegenden GARCH- und Copula-GARCH-Modells zu Modellrisiko bei der Untersuchung des Value at Risk führen kann. Es wird gezeigt, dass es wichtig ist, Phänomene wie Asymmetrie und langes Gedächtnis in den Daten korrekt zu modellieren wohingegen die Wahl einer falschen Randverteilung von geringerer Bedeutung ist. Diese Arbeit versucht die folgende Hypothese zu validieren: das Fehlspezifikationsrisiko hat eine geringere Wirkung als das Schätzrisiko auf Prognosefehler mit entsprechendem Einfluss auf die Value at Risk Prognose. Komplexere Modelle führen zu einem höheren Schätzrisiko und beinhalten für längere Prognosehorizonte ein höheres Modellrisiko. Es wird gezeigt, dass selbst Backtests darin scheitern, die Genauigkeit von Risikomaßen einzuschätzen, selbst in dem Fall in dem die asymptotische Varianz des Tests um das Fehlspezifikations- und Schätzrisiko korrigiert wird. Es werden multivariate Backtests zur Lösung dieses Problems vorgeschlagen.

Modellunsicherheiten entstehen bei der Anwendung von Modellen und der Modellanwender sich sollte daher Unsicherheiten und Nachteile der verwendeten Modelle im Klaren sein. Ein komplexes Modell ist nicht notwendigerweise eine einfacheren Modell überlegen, wenn es um die Prognose von Risikomaßen geht. Während man argumentieren kann, dass im Rahmen der Finanzmarktregulierung das Modellrisiko durch einen Multiplikationsfaktor ausreichend Rechnung getragen wird, haben Finanzinstitutionen selbst wie auch Interessengruppen wie Investoren und Ratingagenturen ein Interesse das Risiko durch die Modellanwendung zu bestimmen um ein realistisches Bild der Finanzstabilität der Institution zu erlangen.

Schlüsselwörter: Modellrisiko, Schätzrisiko, Fehlspezifikationsrisiko

Abstract

This thesis provides a concise definition of model risk. A wrong model can lead to serious over- or underestimation of a financial institution's risk. Because the underlying data generating process is unknown in practice evaluating model risk is a challenge. So far, definitions of model risk are either application-oriented including risk induced by the statistician rather than by the statistical model or research-oriented and too abstract to be used in practice. We introduce a data-driven notion of model risk which includes the features of the research-oriented approach by extending it by a statistical model building procedure. We furthermore suggest the application of robust estimates to reduce model risk and advocate the application of stress tests with respect portfolio evaluation.

It is further investigated in as how far the misspecification of an underlying GARCH-type and Copula-GARCH-type model might introduce model risk when evaluating the Value at Risk. We find that it is important to correctly specify phenomena such as asymmetry and long memory in the data whereas choosing the correct marginal distribution is of minor importance. This paper attempts to validate the following hypothesis: misspecification risk has a less serious impact than estimation risk on forecast errors with a corresponding impact on VaR forecasts. More complex models lead to a higher estimation risk and thus entail higher model risk for longer forecast horizons. Even when accounting for model risk by incorporating estimation and misspecification risk by adjusting the asymptotic variance of the test statistic, backtests fail to assess the accuracy of computed risk measures. We suggest to use multivariate backtests for getting more viable backtests.

Model uncertainties arise by the application of models and the user of models should be aware of the uncertainties and flaws of the models used. Not the most complex models are necessarily the best models in the context of forecasting risk measures. While in the context of regulation one can argue that the measurement of model risk is sufficiently made allowance for by the multiplication factor, financial institutions themselves as well as their stakeholders such as investors and rating agencies have an interest in determining the risk stemming from model application in order to get a realistic picture of the financial stability of the institution.

Keywords: model risk, estimation risk, misspecification risk

Contents

1 Introduction

The omnipresent term of globalisation is perceived in the first place as being the gradual integration of economies and financial markets. As economies and financial markets become integrated to a higher degree, spill-over effects of adverse developments to other countries have a more detrimental effect on previously more loosely linked economies. One of the by-products of globalisation therefore is a higher vulnerability of the financial system as a whole.

Guaranteeing a sound and stable financial system in the light of asymmetric incentives of governments and financial institutions has therefore been the task of regulatory bodies. However, regulatory competition between countries make financial institutions shift their activities to countries with lower capital requirements. In the 1990s, more efforts for the harmonisation of regulatory requirements have been taken and implemented by the associated countries. More refined methods for measuring the risk taken by financial institutions have been developed at that time as well. Capital requirements as a buffer against problems arising from changes in stock prices, interest rates, commodity prices or foreign exchange rates (market risk) and the risk that a counterparty cannot fulfil its financial obligations and defaults on its debts (credit risk) were the categories were the main attention of risk management efforts were turned upon.

During the last century other risk categories have been taken into account within the Basel II frameworks due to several striking events. One of these is operational risk, that is the "risk of loss resulting from inadequate or failed internal processes, people and systems or from external events". Liquidity risk is another risk category were more emphasis has been put upon after the credit crunch in the aftermath of the US subprime crisis. A new research issue concerns the development of risk aggregation methods which account for the possible correlation between different risk categories. Although the measurement of credit risk is still a challenging objective, one should assume that the vulnerability of the financial system should have decreased to some extent due to the action taken by the responsible institutions. However, the example of uncertainties in the measurement of credit risk already suggests that there are other sources of risk that can lead to a biased risk measures. Traditionally, these errors have been taken into account by the introduction of a multiplication factor applied to the risk measure depending on the accuracy of the model used for risk quantification.

Although the term of model uncertainty and the problems of estimation errors are a very common phenomenon in the context of risk management problems arising from the application of models as such have been more or less neglected until recently. Models are an approximation of the complex reality and thus more or less simplify the real pattern of the underlying data generating process. Thus, using models to explain and predict developments in social sciences have the flaw of the models being only partly correct. During the last three centuries the risk management environment has become model-prone and the quantification of risk factors is regarded as essential in supervision efforts. However, the mere application of models itself introduces model risk through estimation and misspecification risk. The following papers are dedicated to this more recently introduced risk category.

The first paper provides a concise definition of model risk and summarises methods for its quantification. Model risk as part of the operational risk is a serious problem for financial institutions. As the pricing of derivatives as well as the computation of the market or credit risk of an institution depends on statistical models the application of a wrong model can lead to a serious over- or underestimation of the institution's risk. Because the underlying data generating process is unknown in practice evaluating model risk is a challenge. So far, definitions of model risk are either application-oriented including risk induced by the statistician rather than by the statistical model or research-oriented and too abstract to be used in practice. Especially, they are not data-driven. We introduce a data-driven notion of model risk which includes the features of the research-oriented approach by extending it by a statistical model building procedure and therefore compromises between the two definitions at hand. We furthermore suggest the application of robust estimates to reduce model risk and advocate the application of stress tests with respect to the valuation of the portfolio.

Evaluating market risk by means of the Value at Risk means to evaluate the forecast distribution of a suitable model for the return distribution of the underlying financial asset. The most popular models for this purpose are GARCH-type models for the returns of financial assets. Model specification mainly aims at obtaining a good in-sample fit to the data. In terms of measuring the model risk involved within a model the forecast distribution and thus the out-of-sample fit is the most important criteria. We investigate in how far the misspecification of an underlying GARCH-type model might introduce a model risk when evaluating the Value at Risk. In the second paper, we find that it is important to correctly specify phenomena such as asymmetry and long memory in the data whereas choosing the correct marginal distribution is of minor importance. Neglecting asymmetry and long memory in the data can lead to a serious forecasting error and therefore to serious model risk.

The effect of model risk on Value at Risk (VaR) forecasts by using Copula-GARCH models is examined in the third part of the thesis. Copula-GARCH models allow for the specification of the dependence structure of return series. This paper attempts to validate the following hypothesis: misspecification risk has a less serious impact than estimation risk on forecast errors with a corresponding impact on VaR forecasts. We conduct a Monte Carlo study where different Copula-GARCH models with different marginal distribution assumptions are simulated and used for forecasting the true as well as the other wrong models. We find that misspecification of the dependence structure as well as of the variance specification has a negligible effect on forecast accuracy. The effect of the marginal distributional assumptions is found to be more pronounced. More complex models lead to a higher estimation risk and thus entail higher model risk for longer forecast horizons.

Even when accounting for model risk by incorporating estimation and misspecification risk by adjusting the asymptotic variance of the test statistic by the model risk incurred may fail to produce correct type I errors when regulatory approaches restrict required backtests for assessing the accuracy of computed risk measures. Together with my co-author Johannes Rohde I analyse these problems in the fifth chapter of the thesis. We suggest to use multivariate backtests as being better solutions for getting more viable backtests.

Thinking about model risk there are several crucial points to bear in mind: model uncertainties arise by the application of models and the user of models should be aware of the uncertainties and flaws of the models used. Not the most complex models are necessarily the best models in the context of forecasting risk measures. When it comes to determining the accuracy of models by using methods of backtesting it should be kept in mind that even when accounting for model uncertainties regulatory prescriptions can restrict the accurate measurement of models. While in the context of regulation one can argue that the measurement of model risk is sufficiently made allowance for by the multiplication factor, financial institutions themselves as well as their stakeholders such as investors and rating agencies have an interest in determining the risk stemming from model application in order to get a realistic picture of the financial stability of the institution.

2 Measuring Model Risk

Joint with Philipp Sibbertsen and Gerhard Stahl

Published in The Journal of Risk Model Validation (2008) **2**, *pp. 65–81*

3 Model Risk in GARCH-Type Financial Time Series

Joint with Philipp Sibbertsen

Published in Rösch, D. and Scheule, H. (ed.) Model Risk - Identification, Measurement and Management (2010), *pp. 75–89*

4 Model Risk in Copula-GARCH Type Financial Time Series

Joint with Grigoriy Tymchenko

4.1 Introduction

The last decades have seen a steadily growing model universe for the sake of describing the evolution of stochastic processes. Particularly in the context of financial management, statistical models have been developed that account for empirically justified facts and characteristics of financial time series. These include fat tails in return distributions, volatility clusters, asymmetries and long memory in volatility as well as non-linear dependence structures, see e.g. Cont [2001] and Embrechts et al. [2001] for more detailed descriptions. Striving for including these facts by defining new models goes along with rising complexity of models and numerosity of included parameters. For an applier it has thus become an increasingly difficult task to select and fit models to a given time series and to use them for the purpose of forecasting densities as well as determining quantiles of distributions in a risk measurement context. Therefore, it is questionable whether more intricate models are necessarily superior to simpler ones in predicting the price or the risk of a financial asset. Concerns are primarily related to the uncertainty of the additional risk incurred by using more complex models. The development towards a more model-prone statistical world has thus given rise to a new category of risk called model risk. Sibbertsen et al. [2009] define model risk as the risk occurring at the central steps of the statistical modeling process, namely model choice, specification of the functional form as well as model estimation.

Model risk should not be confounded with conventional risk categories such as credit, market and operational risk as its source is the risk incurred by the modeling of risk measures like Value at Risk (VaR) as such. Nevertheless, it is regarded as a distinct part of operational risk but can be more clearly distinguished from these risk categories by defining it as an uncertainty, see Cont [2004]. It has latterly achieved broader attention in the research community (see Kerkhof et al. [2010] and Escanciano and Olmo [2010] among others). According to the statistical modeling procedure model risk can be decomposed into misspecification and estimation risk. Alternative approaches for the quantification of model risk have already been proposed by Cont [2004] who uses a Bayesian as well as a worst-case approach for model risk measurement. Kerkhof et al. [2010] define misspecification risk as the difference between estimated VaR and the upper bound of the confidence region of the VaR estimate.

Due to the recent financial market crisis and a series of prominent bank failures as well as uncertainties induced by the budget crisis in several countries in the European Union more effective mechanisms of regulation and for handling model risk in particular have been called for. So far, the Basel II regulations implicitly deal with model risk by a multiplication factor ranging between three and four times the amount set aside as a capital buffer for market risk depending on the number of VaR breaches ([BCBS, 1996]). Besides system relevance, model risk is a non-negligible issue for financial institutions as Basel II allows for the internal calculation of risk capital. However, selecting models that take into account inherent characteristics of financial time series involves a trade-off between misspecification and estimation risk. In the process of setting up internal models and determining risk measures financial institutions have to decide which stylized facts have to be modeled and which the appropriate model is. We will argue that some of the aforementioned characteristics need not to be modeled and one can use simpler models instead due to lower variance of parameter estimates. Estimation risk has a more pronounced impact on out-of-sample forecasting performance than misspecification risk. Therefore, lower variance of parameter estimates and thus estimation risk is more important in this regard. However, in some circumstances which will be defined, modeling certain financial time series characteristics cannot be neglected and need to be taken into account by appropriate model classes. In these cases less parsimonious specifications including parameters that account for non-negligible facts should be preferred, thus reducing model risk. The purpose of this paper is to find out which these important characteristics and data situations are.

Important financial market data characteristics can be modeled by the class of copula-GARCH models which have recently been introduced for the purpose of risk forecasting, see e.g. Lee and Long [2009], Patton [2006], Fantazzini [2009]. These models combine the merits of the class of GARCH models with the possibility of modeling non-linear dependence structures between assets by means of copula models. Within the class of these models several studies consider the effect of underfitted models that beat less parsimonious models in a forecasting contest. An extensive study by Hansen and Lunde [2005] yields that simple GARCH$(1,1)$ beat other intricate GARCH specifications in the context of VaR forecasting. Hamerle and Rösch [2005] find that Gaussian copulas do not perform worde than Student-t copulas for the purpose of credit risk measurement. Our study is closely linked to the one of Fantazzini [2009] who investigated the accuracy of copula-GARCH models.

We simulated paths of different copula-GARCH models each including five stylized facts and possible combinations of them yielding eleven specifications overall. We then forecasted these processes with the true model as well as with simpler specifications. We find that when forecasting VaR, asymmetry in volatility is a non-negligible fact no matter whether it is the only fact present in the data or whether it occurs in combination with any of the other stylized facts. When forecasting volatility, however, there are only very special combinations of characteristics to be found that are not to be misspecified. For lower degrees of asymmetry ($\gamma < 0.3$) for the Asymmetric Power GARCH the GARCH model which does not take asymmetry in volatility into account will perform as good. However, neglecting fat tails or tail dependence and using models that do no take these facts into account will not deteriorate forecast performance. Thus, when taking forecast errors as a model risk measure underfitting will not lead to worse forecasts in many data situations.

In the next section copula-GARCH models are introduced. The third section is dedicated to the description and quantification of model risk sources and the bias-variance trade-off in copula-GARCH models. In section 4 the results of the conducted Monte-Carlo study are illustrated. Section 5 wraps up the findings.

4.2 Copula-GARCH Models

GARCH models. While conventional time series models assume the variance of stochastic processes to be constant over time, empirical evidence finds them to be time-varying instead. Models of the GARCH class are able to implicitly model the conditional variance and volatility clusters in financial market data. A vast number of extensions allow for other stylized facts such as long memory to be incorporated, see Bollerslev [2007] for an exhaustive overview. Time-varying volatility is introduced through multiplicative heteroskedasticity of the innovation term of the stochastic process y_t:

$$\varepsilon_t = \sigma_t \eta_t$$
$$\eta_t | \Psi_{t-i} \overset{iid}{\sim} (0,1)$$

where $\Psi_{t-i} = y_{t-1}, y_{t-2}, \ldots$ is a σ-algebra. While η_t is commonly assumed to be normally distributed, Bollerslev [1987] suggests that the marginals be t-distributed ($\eta_t \sim t(\nu)$) thus taking into account fat-tailed margins. The conditional variance σ_t^2 of the GARCH(p,q) by Bollerslev [1986] model depends on the lagged returns and variance

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

with parameters restricted $\omega > 0$, $\alpha_i \geq 0 \ \forall i = 1, \ldots, p$ and $\beta_j \geq 0 \ \forall j = 1, \ldots, q$ thus ensuring that σ_t^2 remains positive.

Copula models. During the end of the 1990s, copula models emerged in the field of risk management due to awareness of the fact that common risk models neglected the complexity of the dependence structure among assets. The attractiveness of copulas is mainly traced back to a theorem formulated by Sklar [1959] which establishes the decomposition of a joint distribution $F(x_1, \ldots, x_d)$ with random variables $x_1, \ldots, x_d$ into its d marginal distributions $F_i \ \forall \ i = 1, \ldots, d$ and their dependence structure by combining them via a coupling function C, called copula,

$$F(x_1, \ldots, x_d) = C(F_1(u_1), \ldots, F_d(u_d)) = C(u_1, \ldots, u_d).$$

A d-dimensional copula is a multivariate joint distribution defined on the d-dimensional unit hypercube $[0,1]^d$ such that every marginal distribution is uniform on the interval $[0,1]$. It is unique if the marginal distributions are continuous. The copula can thus be seen as the joint distribution of the inverse transform of the marginal distributions of x_i, $F^{-1}(u_i)$:

$$C(u_1,...,u_d) = F(F^{-1}(u_1),...,F^{-1}(u_d)).$$

Copula-GARCH models. Copula and GARCH models can be easily combined to form a new model class, copula-GARCH models. A straightforward way is to transform the marginal distributions η_t of the residuals into uniformly distributed marginals, so that $\eta_i = x_i$ in the above definition of the copula. Let the joint distribution of $\eta_1,...,\eta_d$ be

$$F(\eta_1,...,\eta_d;\theta) = C(F_1(\eta_1),...,F_d(\eta_d),\xi)$$

where θ denotes the copula and ξ denotes the marginal parameters. Several methods have been suggested for the estimation of copula-GARCH models. Although simultaneous estimation methods of marginal and copula parameters are available due to Sklar's Theorem estimation is preferred to be conducted in sequential steps. Among them is the Inference Functions for Margins (IFM) method by Joe [1997] where the copula as well as marginal parameters are separately estimated by maximum likelihood estimation. Genest et al. [1995] and Kim et al. [2007] suggest a semi-parametric pseudo maximum likelihood estimation (PML) of the dependence structure. The marginal parameters ξ are estimated in the first step. The copula parameters, θ, are estimated from fitting them to the empirical distributions of the marginals $\hat{F}_i$:

$$F(\eta_1,...,\eta_d;\theta) = C(\hat{F}_1(\eta_1),...,\hat{F}_d(\eta_d)).$$

Another additional time-varying feature can be incorporated by letting the dependence parameter of the copula vary over time, see among others Jondeau and Rockinger [2006].

4.3 Model Risk in Copula-GARCH Models

Model risk is defined as the risk induced by the choice, specification and estimation of a statistical model for risk forecasting, thus occuring at each step of the statistical modeling cycle, Cuthbertson et al. [1992]. Forecasting risk measures by means of copula-GARCH models includes the selection of an approriate estimation method for copula parameter estimation. The paper by Fantazzini [2009] suggests that IFM estimation leads to copula misspecification caused by the misspecification of marginals. This is why using the IFM estimation we are not able to disentangle the marginal and dependence misspecification effects on VaR. Another motivation for using the PML method is a huge reduction in complexity compared with simultaneous estimation, see Kim et al. [2007]. The estimated parameter vectors θ and ξ separately affect quantile mapping and VaR estimation. For this reason we favor a semi-parametric approach in our study and do not consider model uncertainty in this respect. We rather focus on the occurrence of model risk in other modeling steps, namely marginal and copula parameter estimation and their impact on forecasting volatility and risk measures.

Choosing a model that fits a time series adequately so that the risk of misspecifying the true underlying process is relatively small induces high estimation risk as a higher number of parameters needs to be determined. This induces low bias and high variance of parameter estimates through overfitting. If more parsimonious models are chosen at the expense of adequate specification estimation risk should decrease giving rise to a bias-variace trade-off.

Within a forecasting framework overfitting decreases the in-sample error. For the out-of-sample period on the other hand high variance of an estimator through overfitting increases the forecast error. Thus, in a risk management forecasting context one should consequently expect that estimation risk is more severe than misspecification risk. The bias-variance trade-off suggests that the choice of simpler models by misspecifying the true model does not decrease the accuracy of risk measures. Our following Monte Carlo study will investigate whether this statement is universally true and otherwise describe situations where departures are advisable.

4.4 Simulation Study

4.4.1 Stylized Facts and Specifications

Characteristics which are recognized as important empirical facts are fat-tailed distributions, asymmetries in volatility and (lower) tail dependence, see Figure 1 for an illustration of these facts. These can be modeled by models of the copula-GARCH class. Fat tails are commonly accounted for by student-t distributed margins in contrast to normally distributed ones. Volatility clusters and fat tails in conditional variance are accounted for by fitting a GARCH model. The Asymmetric Power ARCH (APARCH) allows for including asymmetric responses in volatility. While the Gaussian copula allows for combining different marginals, the Student-t copula incorporates tail dependence in addition. Lower tail dependence meaning that in market downturns correlations tend to rise can be modeled by the Clayton copula.

The Asymmetric Power ARCH model (APARCH) of order (p,q) proposed by Ding et al. [1993] accounts for the stylized fact of asymmetric responses of volatility to shocks,

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i [|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}]^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta$$

with $\omega > 0$ and α_i as well as β_j being non-negative. The power parameter $\delta \geq 0$ is a Box-Cox transformation thereby linearizing the non-linear model and $-1 < \gamma < 1 \; \forall i = 1, ..., p$ is the parameter that incorporates the leverage effect so that negative shocks have a higher impact on the conditional variance than positive ones. Note that when setting $\delta = 2$ this model yields the GJR-GARCH model by Glosten et al. [1993] and further restricting $\gamma = 0$ results in the above specification of a GARCH(p,q) process.

Among the most popular copulas in risk management are elliptical copulas such as the Gaussian copula where $C_\Phi(u_1, ..., u_d; \rho_\Phi) = (\Phi(u_1), ..., \Phi(u_d))$ where Φ is the cdf of the Gaussian distribution and the Student-t copula $C_{t_\nu}(u_1, ..., u_d) = (t(u_1), ..., t(u_d); \rho_{t_\nu})$ with t_ν being the cdf of the Student-t distribution and ρ is correlation coefficient of the copula. In contrast to the Gaussian copula, the Student-t copula results in a star-shaped scatterplot for low degrees of freedom ν with its highest density on the main diagonal and allows for modeling higher dependence in the tails of the multivariate distribution (tail dependence). It tends towards a Gaussian copula for increasing values of ν. While advantageously one can easily specify different correlation patterns between the margins of elliptical copulas, their main obstacle is their radial symmetry which does not allow elliptical copulas for modeling asymmetric dependency structures, i.e. increasing dependencies among assets in periods of market downturns which are broadly observable among financial market data. The Clayton copula (Clayton [1978]) has been suggested to account for lower tail dependence in the sense of increasing concordance of random variables in the lower tails of the distribution. It belongs to the Archimedean copula class which is constructed by means of a convex copula generator $\psi(\cdot)$,

$$C(u_1, ..., u_d) = \psi^{-1}[\Sigma_{i=1}^n \psi(u_i)].$$

For the Clayton copula this generator is defined as

$$\psi(u_i) = \frac{1}{\kappa}(u_i^{-\kappa} - 1)$$

which by insertion in the Archimedean copula function leads to the Clayton copula with

$$C_{Cl}(u_1, ..., u_d, \kappa) = [\sum_{i=1}^{n} u_i^{-\kappa} - n - 1]^{-1/\kappa},$$

defined for $\kappa \in [-1, 0] \cup (0, \infty]$. A copula has lower tail dependence if the tail index is $\lambda \in (0, 1]$ and for the Clayton copula the tail index

$$\lambda = 2^{-1/\kappa}$$

results. The higher the copula parameter κ, the more pronounced is the dependence of the random variables in the lower tails. Therefore, the Clayton copula seems to be a promising model as it should be able to reflect the dependence structure in financial risk measurement much better due to increasing correlation of risk factors in adverse market situations. The following figures exemplarily show plots of the dependence structure produced by the respective copulas. Figure 2a) and b) show simulated draws from the Gaussian and the Student-t copula with same correlation coefficient which result in a different dependence structures. Figure 2c) displays the Clayton copula for $\kappa = 3$ where the asymmetric nature of this copula type becomes evident.

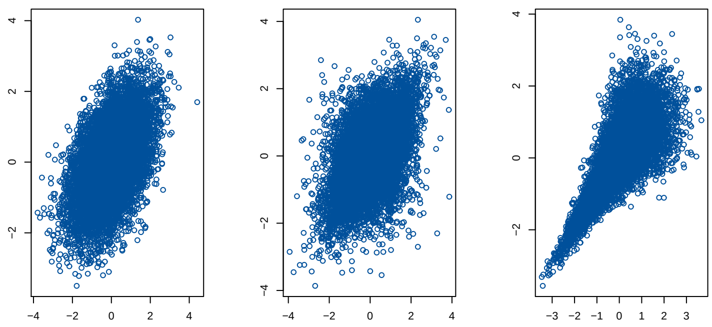


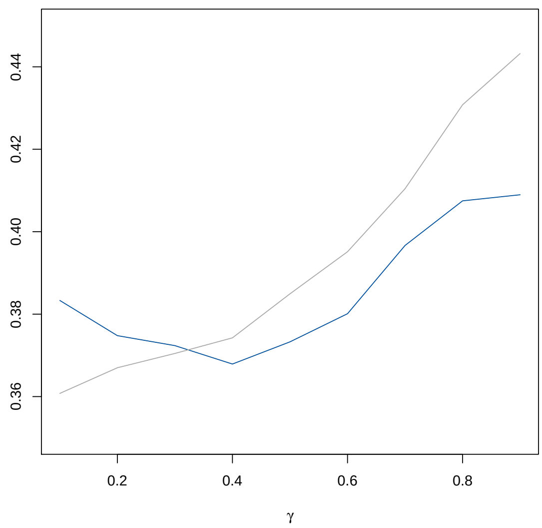
Figure 1: Normally distributed marginals, 10,000 random draws from a) Gaussian copula ($\rho = 0.5$), b) Student-t copula ($\rho = 0.5,\ \nu = 5$), c) Clayton copula ($\kappa = 3$)

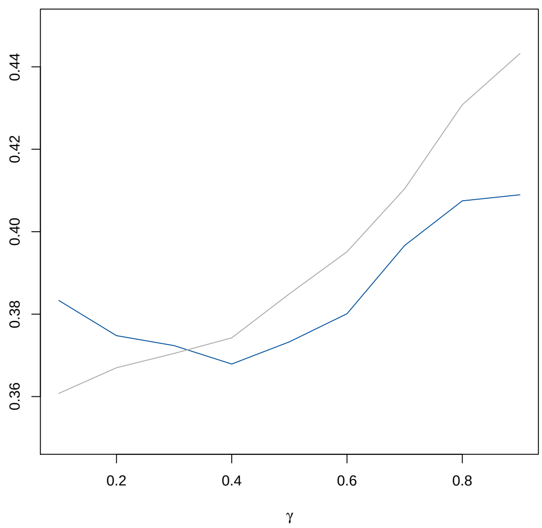
The following study will investigate misspecifications of underlying processes by underfitting and / or underparametrization of the true model. Our most basic specification is the GARCH model with normally distributed marginals without tail dependence (Gaussian copula). In a first step we investigate whether neglecting one characteristic leads to better or equally good forecasts. If this is true for instance if marginals are fat-tailed forecasting with normally distributed marginals will perform not worse than forecasting with t-distributed ones, forecasting with GARCH should lead to as good or even better forecasts than forecasts with APARCH although asymmetric volatility is present in the data.

However, even if one of the characteristic can be neglected when it is present in the data conditioned on the existence of other characteristics that are present it has to be taken into account that there are all kinds of fact combinations thinkable in which these characteristics cannot be neglected and underfitting will lead to higher forecast errors. As an example this means that when fat tails and asymmetric volatility are present in the data, it is to be determined whether GARCH models with normally distributed residuals, APARCH models with normally distributed residuals and GARCH models with student-t distributed residuals produce lower forecast errors. The most complex spefication when fat tails, asymmetric volatility and lower tail dependence are present in the data may lead to the situation where none of the facts can be neglected.

4.4.2 Simulation Design and Forecast Methodology

In a pre-analysis we determine a reasonable choice of the asymmetry parameter of the APARCH model for simulation. An APARCH(1,1) model with $\omega = 0.01,\ \alpha = 0.05,\ \beta = 0.85,\ \delta = 2$ and $\gamma \in (0.1, 0.2, ..., 0.9, 1.0)$ has been simulated and a GARCH as well as APARCH has been fitted to the simulated series and used for prediction of volatility and VaR. The following figure displays the forecast error for varying degrees of the asymmetry parameter γ of the underlying DGP. For $\gamma \to 0$ one should expect that the forecast errors resulting from fitting a GARCH model are as high or less than those from fitting an APARCH model as the asymmetry effect vanishes for smaller γ. For increasing γ one would expect that APARCH forecast errors are gradually becoming less than those resulting from fitting and predicting with the GARCH model. A Monte Carlo study has been conducted to evaluate the point where both models produce forecast of equal quality as far as forecast errors are concerned. Each step is replicated 100 times. The following figures show the forecast errors resulting from predicting volatility and VaR of the APARCH series with the true as well as the GARCH model.





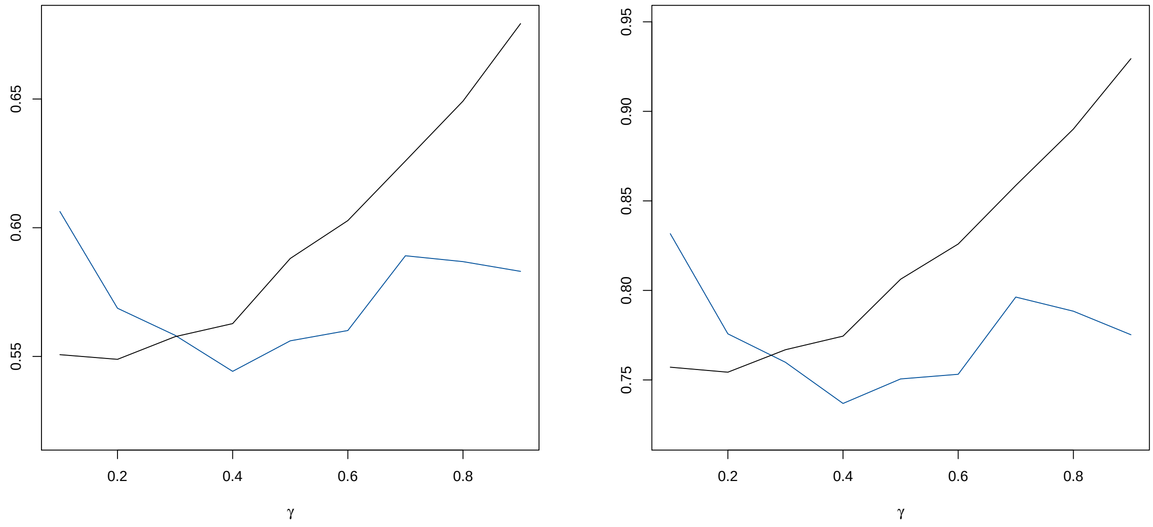


Figure 2: Mean Squared Forecast Error, upper left: return, upper right: volatility, bottom left: VaR95%, bottom right: VaR 99% , blue: APARCH, grey: GARCH

As expected, the GARCH model performs better for lower degrees of asymmetry and the APARCH is superior to GARCH for higher asymmetry in volatility. Mean squared forecast errors (MSFEs) are equal for both models when the degree of asymmetry of the underlying process is $\gamma = 0.4$ approximately. When forecasting VaR the difference between GARCH and APARCH becomes more significant for higher degrees of γ in comparison to those for volatility forecasts. We therefore set $\gamma = 0.5$ in our Monte Carlo study when simulating an APARCH model so that the asymmetry in volatility characteristic is pronounced in a reasonable way.

For our Monte-Carlo study we simulated eleven different bivariate data generating processes (DGP) with length $t = 980$ where each of the following specifications were combined:

- GARCH(p,q) or APARCH(p,q)
- Standard normally distributed or Student-t distributed marginals;
- Gaussian, Student-t or Clayton copula.

The only specification that was not simulated is the most basic specification from which we cannot depart to any simpler specification for the purpose of forecasting in our framework. The following table contains parameter choices for simulation. For the mean equation an AR(1) process was chosen and the order of the GARCH and APARCH process was set to $p = 1$ and $q = 1$. As we set $\delta = 2$ for the APARCH model we do not consider the power property of the APARCH but rather refer to the GJR-GARCH model and solely focus on the asymmetry of the process volatility.

	GARCH(1,1)	APARCH(1,1)	η
μ	0.15	0.15	-
ϕ	0.50	0.50	-
ω	0.01	0.01	-
α	0.05	0.05	-
β	0.85	0.85	-
γ	0.00	0.50	-
δ	2.00	2.00	-
μ_Φ	-	-	0
σ^2_Φ	-	-	1
ν	-	-	5
ρ_{Φ/t_ν}	-	-	0.5
ν	-	-	5
κ	-	-	3

Table 1: DGP Specification

The time series were split into an in-sample ($t_1 = 700$) and an out-of-sample ($t_2 = 280$) period and the ratio of in-sample to out-of-sample horizon is $\pi = 0.4$.

The time series are estimated and forecasted with the true model as well as the other eleven models. The bivariate time-series models are estimated with Maximum Likelihood with normally and t-distributed errors. The computed residuals are used for the estimation of the copula parameters by means of a pseudo-ML approach by converting the empirical distribution of margins into uniformly distributed ones which includes the computation of $u_i = \hat{F}_i(\eta_i)$. From these computed u_i we estimate the copula parameters ρ, ν and κ and the parameters of the marginal distributions, μ, σ^2 and ν. Only the estimation of the marginals requires the distributional assumptions. This is why we expected no violations in estimated copula parameters with respect to the marginal assumptions. Thus, the possible differences in predictions of VaR, return and volatility cannot be any more explained by means of violations of the copula parameters, caused by the marginal assumptions, see Fantazzini [2009]. However, these assumptions can be crucial for estimation of the GARCH model. We then computed one-step ahead forecasts of volatility and VaR at confidence levels $\alpha = (0.95,\ 0.99)$ by using a rolling window forecasting scheme of length 700. These steps are replicated 1,000 times.

4.4.3 Results

Although not necessarily that important for our argumentation we calculated deviances of the parameter estimates by using wrong models from those that have been used for simulation. Overall, estimation with misspecified models having less parameters and not taking into account the complexity of the underlying process results in partly heavily biased parameter estimates.[1]

Volatility Forecast

Results for volatility out-of-sample mean squared forecast errors are provided in Table 2. If one characteristic feature is included in the data, that is the GARCH-*t* for fat tails in margins with Gaussian copula the APARCH-*N* with Gaussian copula thus (asymmetry in volatility) and the GARCH-*N* with Student-*t* and Clayton copula (with (lower) tail dependence) has been simulated, then forecasting with a model that neglects this characteristic leads to MSFEs of lower or comparable size. Thus, forecasting with the basic model when marginals are fat-tailed does not lead to an increase in model risk. The same is true if two or even three features are included in the DGP. However, in certain combinations it is crucial not to miss the effect of the occurence of two characteristics at once which lead to a huge increase in MSFEs: when the DGP is an APARCH model with *t*-distributed margins and (lower) tail dependence, then forecasting with a GARCH model with fat-tailed margins no matter whether the dependence structure is transformed with a Gaussian, Student-*t* or Clayton copula will make the MSFEs rise considerably. Although not in every case, it seems to be important to pay attention to an asymmetric volatility structure when present in the data especially when assets have stronger dependence in the (lower) tails.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
GARCH-N (G)	0.1214	0.1214	0.8008	1.0281	1.0228	0.5993	0.5983	0.7306	1.8891	1.8510	2.3128
GARCH-N (T)	**0.1213**	0.1213	0.8006	1.0304	1.0269	0.5989	0.5983	0.7337	1.8890	1.8499	2.3223
GARCH-N (C)	0.1153	**0.1153**	0.7547	0.9917	1.0201	0.5627	0.5703	0.7262	1.7791	1.7692	2.2993
GARCH-t (G)	0.1157	0.1157	**1.0882**	1.3573	1.3685	0.5748	0.5750	0.7019	2.6102	2.5777	3.1805
GARCH-t (T)	0.1156	0.1156	1.0900	**1.3680**	1.3857	0.5749	0.5753	0.7051	2.6163	2.6112	3.2290
GARCH-t (C)	0.1101	0.1101	1.0323	1.3112	**1.3527**	0.5407	0.5487	0.6973	2.4771	2.4854	3.1320
APARCH-N (G)	0.1209	0.1209	0.7631	0.9221	0.9676	**0.6108**	0.6080	0.7454	1.9007	1.8458	2.3076
APARCH-N (T)	0.1209	0.1209	0.7630	0.9223	0.9715	0.6108	**0.6079**	0.7484	1.9006	1.8447	2.3171
APARCH-N (C)	0.1149	0.1149	0.7187	0.8870	0.9618	0.5738	0.5799	**0.7425**	1.7913	1.7655	2.3003
APARCH-t (G)	0.1854	0.1854	1.0822	1.2820	1.3427	0.6113	0.6093	0.7471	**2.3636**	2.3232	2.8752
APARCH-t (T)	0.1855	0.1855	1.0846	1.2969	1.3599	0.6113	0.6095	0.7504	2.3682	**2.3475**	2.9126
APARCH-t (C)	0.1756	0.1756	1.0258	1.2388	1.3221	0.5744	0.5813	0.7437	2.2432	2.2393	**2.8448**

Table 2: MSFE, Volatility Forecast

[1] More detailed results as well as result tables are available upon request.

Forecast Error Accuracy Test

Table XY provides results of Diebold Mariano test for error accuracy of volatility forecasts for non-nested models and *ENC-NEW* test by Clark and McCracken [2001] rolling scheme for nested models for $\pi = 0.4$ and $k = 1$ if only fat tails were neglected, $k = 2$ if asymmetric volatility was not modelled or $k = 3$ if both. Test statistics in bold indicate higher forecast errors:

True model	Reduced model	ENC-NEW	DM
GARCH-T + Clayton	GARCH-T + Gaussian		0.6856
	GARCH-T + Student-t		**0.0000**
	GARCH-N + Clayton	**76.5080**	
	GARCH-N + Gaussian		**0.0000**
	GARCH-N + Student-t		**0.0000**
APARCH-T + Gaussian	APARCH-N + Gaussian	**89.7712**	
	GARCH-T + Gaussian	**60.3272**	
	GARCH-N + Gaussian	**88.6559**	
APARCH-N+ Student-t	APARCH-N + Gaussian		0.2820
	GARCH-N + Student-t	**8.2977**	
	GARCH-N + Gaussian	**8.2033**	
APARCH-N + Clayton	APARCH-N + Gaussian		0.6706
	APARCH-N + Student-t		0.3796
	GARCH-N + Clayton	**3.9323**	
	GARCH-N + Gaussian		**0.0000**
	GARCH-N + Student-t		**0.0079**

Table 3: Forecast error accuracy tests for selected reduced models

The results suggest no significant increase in models risk due to higher forecast errors if (a)symmetric tail dependence was neglected, but sufficient increase if asymmetric volatility was misspecified. This effect, however, reinforces if two of the characteristics were not accounted for as this is the fact in case of APARCH-N + Clayton modelled with APARCH-N Gaussian and GARCH-N Gaussian.

Value at Risk forecast

Results for Value at Risk out-of-sample forecast errors are provided in Table 4 and 5. If the true model contains tail dependence (GARCH-N-Student-t copula), lower tail dependence (GARCH-N-Clayton copula) or fat-tailed margins (GARCH-t-Gaussian copula), forecasting Value at Risk with more parsimonious models neglecting each one of these facts will lead to lower forecast errors suggesting that misspecification leads to better forecasts when these characteristics are present in the data. However, if the feature of asymmetry in volatility (APARCH-N-Gaussian copula) is characteristic for the data, using a GARCH process to forecast Value at Risk will lead to an increase of forecast error in comparison to forecasting with the true model.

No matter whether the data-generating process has fat tails and tail dependence (GARCH-t-Student-t copula), fat tails and asymmetric tail dependence (GARCH-t-Clayton copula), fat tails and asymmetric volatility (APARCH-t Gaussian copula), asymmetric volatility and tail dependence (APARCH-N-Student-t copula) or asymmetric volatility and lower tail dependence (APARCH-N-Clayton copula) forecasting Value at Risk with GARCH models, when asymmetry in volatility is present, forecast errors will increase. The only exception is the latter combination where neglecting asymmetry in volatility is tolerable when forecasting is done with a model that does not take asymmetry in volatility into account.

If the data contains asymmetric volatility, fat tails and (lower) tail dependence choosing simpler specifications for forecasting will not have a deterioriating effect on model risk. The only fact that increases forecast errors is asymmetry in volatility. Another case is 95% VaR, neglecting the tail dependence in the data by using Gaussian copula to forecast leads to higher forecast errors.

A crucial characteristic that is not to be missed when forecasting Value at Risk is the asymmetry in volatility. If this feature is present in the data – no matter whether it occurs in combination with other more complex specification or alone – using simpler models will lead to higher forecast errors. Neglecting fat-tailedness of the marginal distributions as well as tail dependence or even lower tail dependence will in general induce no problems regarding the reliability of forecasts. Forecast errors might even decrease due to lower estimation risk. Interpretations do hardly differ between 95% and 99% VaR.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
GARCH-N (G)	0.7274	0.8574	3.3170	3.4345	4.2251	2.6079	2.5808	3.1417	6.7794	6.9241	7.7902
GARCH-N (T)	**0.7263**	0.8607	3.3149	3.4306	4.2575	2.6062	2.5773	3.1522	6.7750	6.9149	7.8183
GARCH-N (C)	0.7597	**0.9108**	3.3990	3.5757	4.4795	2.6798	2.6926	3.3280	6.9651	7.2306	8.2898
GARCH-t (G)	0.6984	0.8209	**4.1134**	4.2713	5.2167	2.4805	2.4520	2.9880	8.4746	8.6472	9.6878
GARCH-t (T)	0.6968	0.8236	4.1062	**4.2197**	5.2269	2.4786	2.4460	2.9965	8.4553	8.5246	9.6654
GARCH-t (C)	0.7282	0.8706	4.1631	4.3925	**5.5222**	2.5446	2.5536	3.1613	8.5755	8.8882	10.2522
APARCH-N (G)	0.7272	0.8570	3.1989	3.3127	4.0144	**2.5309**	2.5027	3.0521	6.4195	6.5498	7.4395
APARCH-N (T)	0.7262	0.8603	3.1970	3.3106	4.0276	2.5294	**2.4991**	3.0624	6.4155	6.5407	7.4730
APARCH-N (C)	0.7593	0.9103	3.2790	3.4506	4.2367	2.6003	2.6116	**3.2352**	6.5960	6.8372	7.9206
APARCH-t (G)	0.9608	0.8656	4.0831	4.1988	5.0071	2.5305	2.4957	3.0303	**7.5077**	7.6913	8.6576
APARCH-t (T)	0.9583	0.8686	4.0757	4.1446	4.9962	2.5286	2.4898	3.0396	7.4925	**7.5908**	8.6325
APARCH-t (C)	1.0052	0.9194	4.1294	4.3074	5.2982	2.5971	2.6016	3.2143	7.5990	7.9059	**9.1589**

Table 4: MSFE, 95% VaR Forecast

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
GARCH-N (G)	1.1964	1.4156	5.6075	5.7709	7.2030	4.3072	4.2668	5.1713	11.4449	11.6516	13.2443
GARCH-N (T)	**1.2539**	1.4533	5.6368	6.0286	7.4051	4.3357	4.4730	5.2960	11.5203	12.1994	13.5862
GARCH-N (C)	1.3494	**1.5386**	6.1680	6.4646	7.8003	4.7778	4.8114	5.6127	12.6956	13.1459	14.4414
GARCH-t (G)	1.1544	1.3645	**9.1975**	9.4914	12.1075	4.1637	4.1185	5.0042	19.2818	19.5902	22.9106
GARCH-t (T)	1.2089	1.4004	9.2381	**9.8342**	12.3396	4.1909	4.3143	5.1286	19.3872	20.3052	23.4748
GARCH-t (C)	1.3023	1.4839	10.1791	10.7427	**13.3126**	4.6246	4.6524	5.4426	21.5454	22.3604	25.4730
APARCH-N (G)	1.1959	1.4148	5.3673	5.5232	6.7321	**4.1829**	4.1417	5.0287	10.8019	11.0036	12.6581
APARCH-N (T)	1.2531	1.4522	5.3972	5.7661	6.8905	4.2086	**4.3398**	5.1563	10.8880	11.5188	12.9902
APARCH-N (C)	1.3487	1.5376	5.9158	6.1885	7.2968	4.6404	4.6721	**5.4579**	12.0150	12.4202	13.8075
APARCH-t (G)	1.6879	1.4555	9.1496	9.3071	11.6037	4.2486	4.1944	5.0785	**16.4738**	16.8080	19.7634
APARCH-t (T)	1.7682	1.4944	9.1928	9.6322	11.9157	4.2744	4.3933	5.2071	16.5566	**17.4408**	20.2769
APARCH-t (C)	1.9121	1.5849	10.1385	10.5424	12.8339	4.7201	4.7401	5.5263	18.3921	19.1805	**21.9716**

Table 5: MSFE, 99% VaR Forecast

Backtests

As the Basel II framework stipulates the application of backtests in the sense of testing whether the fraction of times 99% VaR exceeds the return in a period of 250 days equals the VaR level that exceeds returns only 1% of the time. The test proposed by Kupiec tests whether the exceedance series unconditionally keeps the level. A test suggested by Christoffersen demands that also unconditionally the level holds and takes into account whether models have the ability to adjust or build up exceedance clusters. Table XY provides results of number of exceedances within out-of-sample periods, as well as p-values of both tests for 99% VaR:

True model	Reduced model	EXC	KT	CT
GARCH-T + Clayton	GARCH-T + Clayton	2	0.3048	**0.0241**
	GARCH-T + Gaussian	3	0.3312	**0.0412**
	GARCH-T + Student-t	3	0.3286	**0.0328**
	GARCH-N + Clayton	7	0.3105	0.2083
	GARCH-N + Gaussian	8	0.2758	0.2545
	GARCH-N + Student-t	8	0.2856	0.2483
APARCH-T + Gaussian	APARCH-T + Gaussian	2	0.2981	**0.0391**
	APARCH-N + Gaussian	7	0.2757	0.2228
	GARCH-T + Gaussian	2	0.3030	**0.0279**
	GARCH-N + Gaussian	7	0.2568	0.2581
APARCH-N+ Student-t	APARCH-N+ Student-t	6	0.2670	0.1944
	APARCH-N + Gaussian	7	0.2460	0.2275
	GARCH-N + Student-t	7	0.2492	0.2467
	GARCH-N + Gaussian	8	0.2310	0.2728
APARCH-N + Clayton	APARCH-N + Clayton	6	0.2520	0.2023
	APARCH-N + Gaussian	7	0.2146	0.2579
	APARCH-N + Student-t	7	0.2254	0.2367
	GARCH-N + Clayton	7	0.2285	0.2660
	GARCH-N + Gaussian	8	0.1966	0.3170
	GARCH-N + Student-t	8	0.2073	0.3029

Table 6: VaR Forecast accuracy: h=1, VaR=99%

Results suggest using more parsimonious models leads to better forecasts performance. Especially concerning the fat tails which can be adequately reflected with normal distribution of residuals of GARCH or APARCH model. Choosing the Student-t distribution instead increases the estimation risk, the predicts become more conservative. These models are rejected by Christoffersen test, i.e. their ability to build up exceedance clusters is very poor.

4.5 Conclusion

This paper investigates the trade-off between estimation and misspecification risk in a forecasting framework with attention focused on the forecasting of extreme quantiles of distributions of a portfolio of bivariate time series. It is argued that by utilizing the bias-variance trade-off through underfitting better forecasts and less model risk in forecasting through decreasing estimation risk results. On the other hand, if certain empirical data features that are process-relevant are not being modeled misspecification errors increase so that higher forecast errors result from estimation with more parsimonious models. This study looked at the characteristics and combinations of characteristics that need to be modeled when present in the data.

It is left open for further research whether other facts such as long memory in volatility or time dependence of copula parameters which have been found in financial time series need explicitly be taken into account in a copula-GARCH framework. Luedtke and Sibbertsen [2010] find that long memory in GARCH alone does play a role. Furthermore, it would be interesting whether the results hold in a multivariate modeling framework where the time-variation of the covariance and / or correlation between assets is modeled by respective models such as the Constant or Dynamic Conditional Correlation models (DCC, CCC).

5 Model Risk in Backtesting Risk Measures

Joint with Johannes Rohde

5.1 Introduction

Backtesting is a mean to analyse whether a model used for calculating risk measures is accurate. It is at the core of supervisory activity regarding the resilience of financial institutions in alleviating the impact of financial crisis as the accuracy of risk measures has implications for the solvency capital that financial institutions have to calculate.

BCBS [1996] regulations state that the calculation of a financial institutions' market capital requirement for preventing losses resulting from adverse market conditions be the maximum of either the 0.01% Value at Risk (VaR) or the average VaR reported during the previous 60 days multiplied by a factor depending on the sum of VaR violations during the reporting period (traffic-light approach). Thus, the accuracy of the VaR model is closely linked to the regulatory framework. An accurate VaR model satisfies two properties as defined by Kupiec [1995] and Christoffersen [1998].

The unconditional coverage property, formally

$$Pr(I(\alpha) = 1) = \alpha,$$

where $\{I_t\}$ is the hit sequence indicating if a violation occurred or not, claims that the probability of violations during the reporting period equals the α level set for VaR calculation. The VaR model is deemed inaccurate in the sense of failing to be able to account for the incurred risk if the number of violations exceeds the number of expected losses. The risk model is too conservative when the VaR model yields less violations than to be expected.

A second claim is the independence of elements of the hit sequence. If the violations occur in a cluster, the financial institution might not be able to tackle the losses in contrast to a situation where the violations are spread out evenly over the reporting horizon. An accurate VaR model is therefore characterized by satisfying the property of unconditional coverage as well as the independence property,

$$I_t(\alpha) \overset{iid}{\sim} Ber(\alpha),$$

ie that the hit sequence is identically and independently distributed with probability α. Backtests are statistical tests designed for determining the accuracy of VaR models. While several tests have been proposed for each of the two properties, joint tests determine whether the VaR model is accurate as a whole. However, joint tests are not to be gauged as being universally preferable to mono-property tests as the ability to detect the violation of one of the two properties is decreasing (Campbell [2005]).

A type I error arises when an accurate model with a coverage of 99% is erroneously rejected. When the VaR model is inaccurate with lower coverage, eg 2% type II error is the probability that the inaccurate model is not rejected. If the power of the backtest is low, then the probability of classifying an inaccurate model as accurate (not rejecting the null) is comparatively high. Backtests should have high power and not be over- or undersized. In a Monte Carlo study we analyse the problems of common backtest procedures. The main result of this paper will be that even when accounting for model risk, regulation sets restrictions to backtesting.

The paper is organized as follows: the next section describes relevant backtesting categories. It serves a starting point for further derivations of multivariate backtests which will be suggested as a mean to overcome problems resulting from supervisory restrictions. In the third chapter we conduct a Monte Carlo study and analyse the problems that arise when conducting univariate backtests in the course of regulation aspects.

5.2 Overview of backtests

Backtests can be distinguished into frequency-based as well as size-based tests. While the former tests examine the sequence obtained from the exceedance of VaR above the realized profit and losses series, the latter tests are constructed from the size of the exceedance conditioned on the violations. As the regulatory framework is based upon the violations and not on their size, size-based tests are relatively few in the literature due to regulatory constraints (Lopez [1999]).

The most basic backtests for testing the unconditional coverage property, the time until first failure (TUFF) test and its generalization, the proportion of failures (POF) test, were suggested by Kupiec [1995]. As shown in Kupiec [1995] the simplicity of the TUFF test ignores the total number of failures since the start of the monitoring, the POF test should always be run to verify potential loss estimates in place or in addition. In contrast to the TUFF framework, where only the elapsed time until the first failure is considered, the POF uses the entire information. To this (and all further analyses) consider a hit sequence $\{I_t\}_{t=1}^{n}$ of size n, where $\forall t : I_t \in \{0,1\}$, n_1 denotes the number of hits (ie $I_t = 1$) and $n_0 = n - n_1$ (ie $n_0 = \sharp(I_t = 0)$). The probability of observing n_1 hits in a sample of size n is given by the the probability function of the binomial distribution,

$$Pr(\sharp(I_t = 1) = n_1) = \binom{n}{n_1} (1-\alpha)^{n_0} \alpha^{n_1}.$$

For the null hypothesis of the POF test, $H_0 : \alpha = \hat{\Pi}$ with $\hat{\Pi} = \frac{n_1}{n}$, the associated test is a Likelihood Ratio (LR) test and its test statistics is given by

$$K = -2 \log \left(L(\alpha)/L(\hat{\Pi}) \right)$$

where α denotes the failure probability under the null, $L(\cdot)$ the corresponding Likelihood function and $\hat{\Pi} = \frac{n_1}{n}$.

However, when the sample size is relatively small both tests appear to have poor ability to distinguish between the underlying failure probability in the null hypothesis and failure probabilities that are slightly higher (see Kupiec [1995]). Thus, these frameworks might not be adequate for the analysis of the accuracy of VaR estimates covering only one trading year. Furthermore, a frequently arising problem is the non-existence of violations during the reporting period. This issue becomes most important when VaR models with a small failure probability are evaluated. In these cases the Kupiec tests are not computable.

When testing the *iid* hypothesis of the hit sequence the autocorrelation of the sequence itself or the equidistance of the time span between consecutive violations is examined. These tests require the complete specification of the alternative hypotheses in the sense that the way how violation clusters occur has to be specified exactly. Autocorrelation-based tests can be constructed by testing on the autocorrelation structure in the hit sequence itself, $\{I_t\}$, or in the demeaned sequence, $\{I_t - \alpha\}$, that forms a sequence of martingale difference summands (Berkowitz et al. [2009]).

The test by Christoffersen [1998] was the first test of this kind. The basic idea behind this LR-type test consists in the following comparison: If there is no dependence between two consecutive observations, then the probability of monitoring no violation on the day after a violation took place should be equal to the probability of monitoring no violation when on the day before no violation was observed, too.

As in Kupiec [1995] the LR framework is used and built on Markov chains. The independence of the observations of the hit sequence is tested under the null against the alternative of a first-order Markov chain where the stochastic matrix

$$\Pi_1 = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}$$

represents the transition matrix and $\pi_{i,j} = P(I_t = j | I_{t-1} = i), i, j \in \{0,1\}$ the transition probabilities. Let n_{ij} be the number of observations with value i and previous value j. Then the likelihood function for the hit sequence $\{I_t\}$ yields

$$L(\Pi_1) := L(\Pi_1; \{I_t\}) = \pi_{00}^{n_{00}} \pi_{01}^{n_{01}} \pi_{10}^{n_{10}} \pi_{11}^{n_{11}}$$

This is the likelihood under validness of the alternative model while the likelihood for the null model can be computed by considering the stochastic matrix

$$\Pi_2 = \begin{pmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2 \end{pmatrix}.$$

Employing this model under the null it is easy to see that the independence of the hit sequence is tested by this means since the rows have all the same entries. Under the null previous observations do not influence the probability of monitoring a violation or not. Matrix entries π_2 represent the probability of a violation and according to this the number of observations are aggregated over index j as the past value j has no influence on the present value i, $\pi_2 = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}$. Thus,

$$L(\Pi_2) := L(\Pi_2; \{I_t\}) = (1-\pi_2)^{(n_{00}+n_{10})} \pi_2^{n_{01}+n_{11}}$$

is the likelihood function under the null model.

Using $L(\Pi_1)$ and $L(\Pi_2)$ the LR test statistic for the Christoffersen test of independence is

$$LR.IND = -2 \log \left(\frac{L(\Pi_1)}{L(\Pi_2)} \right)$$

which is χ^2 distributed with one degree of freedom. Note that the Christoffersen [1998] test provides no possibility for testing conditional coverage as LR.IND does not depend on the true coverage probability α. A joint test for both testing the independence and the conditional coverage property as well is provided below.

A problem that arises with using this backtest is that the Christoffersen test of independence only examines for dependence between two consecutive observations. Campbell [2005] notes that it is also possible that the probability of monitoring a violation today is not influenced by yesterday's observation but indeed could be influenced by prior observations.

Next to the test for proving independence of observations of the hit sequence Christoffersen [1998] introduced a test of unconditional coverage, testing $E[I_t] = \alpha$ against its alternative $E[I_t] \neq \alpha$. The joint test of conditional coverage and independence by Christoffersen [1998] combines those tests to examine whether both properties of a VaR measure are jointly fulfiled.

The basic idea is as simple as for the independence test: First, if the unconditional coverage property is fulfiled then $\frac{n_{00}+n_{10}}{n_{00}+n_{01}+n_{10}+n_{11}} = \alpha$ must hold implying that the proportion of observed violation matches with the hit probability α. Furthermore, as stated previously, the probability of a non-violation following a previous hit equals the probability of a non-violation following a previous non-violation, i.e. $\frac{n_{00}}{n_{00}+n_{01}} = \frac{n_{10}}{n_{10}+n_{11}}$, when the independence property is on the hand. Combining this, if the VaR measure fulfils the independence property these probabilities should match the total proportion of non-violations, which, provided the unconditional property is valid, leads to

$$\frac{n_{00}}{n_{00}+n_{10}} = \frac{n_{10}}{n_{10}+n_{11}} = \frac{n_{00}+n_{01}}{n_{00}+n_{01}+n_{10}+n_{11}} = \alpha$$

that is tested under the null. In terms of the LR framework the likelihood of the null of the unconditional coverage test is tested here against the alternative of the independence test, forming a test of conditional coverage in effect. Thus, the test statistics results in

$$LR.CC = -2\log\left(\frac{L(\alpha)}{L(\Pi_1)}\right).$$

Christoffersen [1998] shows that the limiting distribution of the joint test is $\chi^2(2)$. However, even if running a joint test might seem always preferable over running the unconditional coverage test and the independence test separately, one has to note that joint tests dismiss VaR measures that violate only one property. As a result the joint test may detect the violation of either unconditional coverage or independence in less cases than a test that focuses on only one of these properties does. According to Campbell [2005] the employment of a test that covers only a sole property might be preferable when prior information over the VaR measure is available.

Escanciano and Olmo [2010] provide a test of unconditional coverage as well as a test of conditional coverage. Their analysis bases on a Monte Carlo study, where the unconditional and the conditional coverage tests are compared to a corrected version of these tests. These corrected releases account for the impact of estimation risk arising when forecasts are carried out. All tests are based on the demeaned hit sequence $\{I_t - \alpha\}$.

The test of unconditional coverage are derived from the validity of $E[I_t] = \alpha$ under the null model. Its test statistics is presented by

$$S_P = \frac{1}{\sqrt{n}} \sum_{t+R=1}^{P} (I_t - \alpha)$$

and is predicated on the unconditional coverage tests by Kupiec [1995] and Christoffersen [1998]. It can easily be checked that $\frac{1}{\sigma}S_P$ is converging against a standard normal distribution, where $\sigma = \sqrt{\alpha(1-\alpha)}$ is nothing else than the standard deviation of the binomial distribution for I_t. This holds as S_P is the standardized version of $\{I_t\}$ with

$$\frac{1}{\sigma P^{-\frac{1}{2}}} S_P = \frac{\frac{1}{P}\sum_{t+R=1}^{P}(I_t - \alpha)}{\sigma P^{-\frac{1}{2}}} = \frac{1}{\sqrt{P}\sigma} \sum_{t+R=1}^{P} (I_t - \alpha) \longrightarrow N(0;1)$$

When adjusting σ for estimation risk it can be shown that the term of the estimated standard deviation gets the form

$$\sigma_{corr} = \left(\alpha(1-\alpha) + \pi \hat{A}\hat{V}\hat{A}'\right)^{-\frac{1}{2}}$$

when the applied forecast scheme is set fixed and the underlying DGP is a GARCH process of order (1,1). Note that Escanciano and Olmo [2010] also provide adjusted tests for rolling and recursive forecast schemes. For $\pi\hat{A}\hat{V}\hat{A}' = 0$ the impact of estimation risk is asymptotically irrelevant.

The parameter $\pi = \lim_{n\to\infty} \frac{P}{R}$ denotes the relation between the length P of the out-of-sample series and the first R observations which are used to estimate the process parameters. It is quiet intuitive that for a large value of R in relation to P and thus a relatively long in-sample series the influence of estimation risk becomes negligibly small. The matrix V is of dimension (3×3) and contains the variances and co-variances of the data generating process, while A denotes a (3×1)-vector containing the first derivations of the DGP wrt the GARCH parameters respectively. $\hat{A}$ and $\hat{V}$ are the consistent estimators of A and V respectively. For a detailed derivation of A and V see Appendix.

The resulting test statistics

$$\tilde{S}_P = \frac{1}{\sqrt{n}\,\sigma_{corr}} \sum_{t=1}^{n} (I_t - \alpha)$$

is $N(0;1)$ distributed for $n \to \infty$.

The leadoff duration-based backtesting approach was proposed by Christoffersen and Pelletier [2004] with the motivation to overcome the pitfall of small power of backtests existing by then in small sample sizes and to uncover not only first order Markov dependencies such as the independence test by Christoffersen [1998]. This approach is justified by the authors by no-hit periods which are either relatively short by reason of high market volatility or relatively long when the market is calmed down. For this, we define $d_i = t_i - t_{i-1}, i = 1,\ldots,I$ as the duration between the hit number $i-1$ and i occurring at dates t_{i-1} and t_i $(t \in \{1,\ldots,n\})$, respectively.

To construct the test that emanates from the independence of the durations and thus, from a correct specified VaR model, a memoryless probability distribution is needed to model the durations. The only continuous distribution that accounts for a constant failure probability α is the exponential distribution with the density

$$f^{Exp}(d) = \alpha \exp(-\alpha\, d).$$

Note that the corresponding hazard function for the exponential distribution is $\lambda^{Exp}(d) = \alpha$ which can be interpreted than the probability that a violation occurs at date d past the last hit after having already waited for $d-1$ days is constantly α and independent from d, ie memoryless. Thus, the null of independence is that the durations d_i come from an exponential distribution with likelihood function

$$\ln L(\alpha) = n\ln(\alpha) - \alpha\overline{d}.$$

For the alternative model a duration distribution with a non-constant hazard rate is required. The simplest case should be the Weibull distribution with density

$$f^W(d) = \alpha^b\, b\, d^{b-1} \exp(-(\alpha\, d)^b)$$

where $b \in \mathbb{R}_{>0}$ is a shape parameter. Note that the exponential distribution is nested by the Weibull distribution for $b = 1$. The hazard rate can easily be obtained by

$$\lambda^W(d) = \alpha^b\, b\, d^{b-1}.$$

For $b < 1$ the Weibull hazard rate is decreasing. Transferred to the financial market a decreasing λ^W indicates that the market tends to more extreme durations, i.e. periods of relatively short or relatively long duration. The log-Likelihood function under the alternative is then given by

$$\ln L(\alpha; k) = \ln \lambda + \ln k + (k-1) \sum_i \ln d_i - \lambda \sum_i d_i^k.$$

Thereby, the pair of hypotheses can be reformulated in terms of the shape parameter b by $H_0 : b = 1$ versus $H_1 : b \neq 1$.

The null of independence can be tested by a Likelihood ratio test by evaluation of

$$LR_{Dur} = -2 \frac{\ln L(\alpha)}{\ln L(\alpha; b)}$$

which follows a χ^2 distribution with two degrees of freedom.

To conduct the test it is necessary to transform the hit sequence $\{I_t\}$ into a duration sequence $\{d_i\}_{i=1}^{I}$. While doing the transformation it has to be kept into account that the first and last duration is possibly censored, ie the duration of the first no-hit period is longer than d_1 as there is no data available before. Of course, the only exception consists in the case that the first observation is already a hit. Likewise the last duration could be longer than d_I when the last observation of $\{I_t\}$ is not a hit.

In the above spanned framework it is possible to model dependencies of higher order than the Markov test. However, this test contains no information about the exact order of dependence, but could only be captured by the EACD framework by Engle and Russell [1998].

Another test of independence that does not exploit the hit sequence directly, but the properties of the durations between consecutive hits was recently proposed by Candelon et al. [2011]. The major motivation behind the construction of this test is to overcome the drawback of low power in realistic sample sizes when conducting backtests.

The idea behind this test is as follows: To each distribution belonging to the Pearson family an orthonormal polynomial can be associated. Orthonormal polynomials build a sequence of polynomials at which each two polynomials are pairwise orthonormal under the L^2-inner product. Considering the duration sequence $\{d_i\}$ as being discrete, the orthonormal polynomial associated with the geometric distribution can be employed.

Define the number of employed polynomials h, the orthonormal polynomial associated to the memoryless geometric distribution follows the recursion

$$M_h = M_{j+1}(d;\alpha) = \frac{(1-\beta)(2j+1)+\beta(j-d+1)}{(j+1)\sqrt{(1-\beta)}}\, M_j(d;\alpha) - \frac{j}{j+1} M_{j-1}(d;\beta)$$

for any $j \in \mathbb{N}_0$, $\forall d \in \mathbb{N}_0$, $d := d_i \forall i \in \{1,\dots,I\}$ and initial values $M_{-1}(d;\alpha) = 0$, $M_0(d;\beta) = 1$. Using the method of moments to estimate the parameters of this polynomial regression efficient and consistent estimates can be obtained. Thus, under the null of conditional coverage the moment condition

$$H_0 : E[M_j(d;\alpha)] = 0$$

is tested. Thus, under the null model the duration sequence follows a geometric distribution with hit probability α, meaning that there is no correlation between two consecutive hits as the geometric distribution provides the only memoryless discrete probability distribution.

In contrast to the duration-based test by Christoffersen and Pelletier [2004], this framework allows to test separately for unconditional coverage and the independence hypothesis. The reasoning is straightforward: As the expectation of a geometric distributed random variable with parameter α is equal to $\frac{1}{\alpha}$, it is easily shown that this is equivalent to the condition for the orthonormal polynomial of order $h = 1$ that is tested under H_0 of unconditional coverage:

$$\begin{aligned} E[M_1(d;\alpha)] &= E\left[\frac{1-\alpha d}{\sqrt{1-\alpha}}\right] \\ &= E\left[\frac{1-\alpha\frac{1}{\alpha}}{\sqrt{1-\alpha}}\right] = 0 \end{aligned}$$

The usage of orthonormal polynomials enables to run the test within the GMM framework with known asymptotic covariance matrices. The test statistics employing the polynomial order h is

$$C_{CC}^{G}(h) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_j(d_i;\alpha)\right)' \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_j(d_i;\alpha)\right)$$

following a χ^2 limiting distribution with h degrees of freedom and $j = 1,\dots,h$. Note that for the special case of unconditional coverage and $h = 1$ the test statistics becomes

$$C_{CC}^{G}(1) = C_{UC}^{G} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} M_1(d_i;\alpha)\right)^2.$$

When presuming that $\{d_t\}$ is continuous the tests are run with the same conditions adjusted for the exponential distribution and its corresponding orthonormal polynomials following the recursion

$$L_h = L_{j+1}(d;\alpha) = \frac{1}{n+1}\left[(2n+1-\alpha d)L_j(d;\alpha) - nL_{n-1}(d;\alpha)\right]$$

with initial values $L_{-1} = 1$ and $L_1 = 1 - \alpha d$ and L being polynomials of the Laguerre family. The test statistics for the continuous case and the orthonormal polynomials associated with the exponential distribution is then

$$C_{CC}^{Exp}(h) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_j(d_i;\alpha)\right)' \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_j(d_i;\alpha)\right)$$

again following a $\chi^2(h)$ distribution under the null.

5.3 Simulation Study

The following simulation studies aim at detecting the problems arising from conducting backtests with univariate time series. For this purpose we simulated GARCH(1,1) processes

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \theta_0 + \theta_1 Y_{t-1}^2 + \theta_2 \sigma_{t-1}^2. \end{aligned}$$

with parameter vector $\theta' = (\theta_0, \theta_1, \theta_2) = (0.1, 0.1, 0.85)$ and different lengths of in-sample R and out-of-sample horizon P. The in-sample period with $R = (250; 500; 750; 1,000; 1,500)$ is used for the estimation of the respective parameters and the out-of-sample period $P = (250; 500; 750; 1,000; 1,500)$ is used for the evaluation of the backtest. The VaR for the respective series with confidence level of $\alpha = 0.01$ is calculated in the next step. Following this, the hit sequence $\{I_t\}$ is computed. For testing the accuracy of the VaR computation the test statistics of the aforementioned backtests are calculated. The procedure is replicated 5,000 times. Table 18 shows the results of the Monte Carlo study. For each combination of in-sample and out-of-sample length, the respective empirical size is calculated from the computed test statistics and the nominal coverage is chosen as amounting to $\alpha = 0.05$. The first three columns summarise the results for the Kupiec test and the tests suggested by Christoffersen (independence and conditional coverage test), while the remaining columns show size results for duration-based backtests for which the sequence $\{d_t\}$ of the time span between the respective hits of sequence $\{I_t\}$ has been taken into account. While tests (4) to (6) are based on the null of a geometric distribution with $h = 1, 3, 5$, tests (7) to (9) report the results for the tests where the distribution under the null is supposed to be continuous with the same number of orthogonal polynomials as under the discrete assumption.

	P	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
R=250	250	0.0930	0.0322	0.0808	0.0486	0.0512	0.0334	0.0138	0.0134	0.0118
	500	0.2240	0.0428	0.1208	0.1758	0.1020	0.0730	0.0344	0.0390	0.0366
	750	0.2262	0.0578	0.1832	0.1840	0.1696	0.1392	0.0718	0.0746	0.0660
	1,000	0.2786	0.0684	0.2286	0.2396	0.2016	0.1660	0.0962	0.0952	0.0816
	1,500	0.3452	0.0756	0.3148	0.3454	0.2828	0.2426	0.1472	0.1458	0.1224
R=500	250	0.0664	0.0328	0.0622	0.0350	0.0388	0.0246	0.0066	0.0080	0.0072
	500	0.1682	0.0412	0.0802	0.1250	0.0682	0.0468	0.0224	0.0270	0.0250
	750	0.1612	0.0640	0.1300	0.1198	0.1128	0.0936	0.0470	0.0574	0.0524
	1,000	0.2138	0.0652	0.1712	0.1746	0.1454	0.1192	0.0666	0.0698	0.0600
	1,500	0.2472	0.0694	0.2296	0.2478	0.1834	0.1500	0.0872	0.0854	0.0744
R=750	250	0.0628	0.0368	0.0582	0.0314	0.0348	0.0236	0.0056	0.0064	0.0074
	500	0.1576	0.0414	0.0680	0.1102	0.0610	0.0456	0.0168	0.0234	0.0252
	750	0.1460	0.0605	0.1216	0.1065	0.0998	0.0849	0.0399	0.0514	0.0448
	1,000	0.1973	0.0621	0.1502	0.1581	0.1247	0.1000	0.0523	0.0589	0.0507
	1,500	0.2058	0.0748	0.2104	0.2064	0.1550	0.1260	0.0652	0.0764	0.0628
R=1,000	250	0.2058	0.0748	0.2104	0.2064	0.1550	0.1260	0.0652	0.0764	0.0628
	500	0.1430	0.0424	0.0634	0.1036	0.0556	0.0412	0.0166	0.0222	0.0230
	750	0.1300	0.0556	0.1076	0.0956	0.0918	0.0734	0.0378	0.0466	0.0394
	1,000	0.1678	0.0690	0.1440	0.1366	0.1096	0.0968	0.0568	0.0574	0.0508
	1,500	0.1877	0.0757	0.1941	0.1877	0.1522	0.1208	0.0673	0.0743	0.0625
R=1,500	250	0.1678	0.0690	0.1440	0.1366	0.1096	0.0968	0.0568	0.0574	0.0508
	500	0.1404	0.0378	0.0624	0.1000	0.0534	0.0384	0.0160	0.0224	0.0236
	750	0.1206	0.0620	0.1058	0.0890	0.0844	0.0674	0.0316	0.0402	0.0358
	1,000	0.1486	0.0604	0.1188	0.1152	0.0952	0.0822	0.0444	0.0494	0.0434
	1,500	0.1652	0.0752	0.1856	0.1656	0.1318	0.1062	0.0622	0.0678	0.0558

Table 7: Results - Size, $\alpha = 0.01$

The first observation to be made is that the majority of the backtests are oversized and hence reject the null too often. Thus, even if the null is true the backtests classify the VaR to be inaccurate. However, some of the duration-based backtests tend to be undersized especially if P and R are both small. Secondly, the smaller the ratio $\pi = P/R$ of out-of-sample length to in-sample length, the lower is the distortion, that is the difference between the empirical and nominal size. For example, for $R = 250$ the Kupiec test is distorted by 29.52% for $P = 1,500$ and the lower the in-sample period the smaller is the distortion. When the out-of-sample length is reduced to $P = 250$ the size is distorted by 4.3%. This is due to the reason that the smaller the amount of data available for estimation of parameters in comparison to P the higher is the estimation risk involved which leads to less accurate projections of VaR. Duration-based backtests tend to have lower size distortions in general.

Acknowledging model risk, Escanciano and Olmo [2010] provided tests corrected for estimation risk. When correcting the variance of the backtest by Kupiec and taking into account the demeaned hit sequence $\{I_t\}$ the test should not be rejected as often as is the case with the uncorrected test. Therefore, it should be expected that the size distortions decrease by applying the estimation risk corrected backtest by Escanciano and Olmo [2010]. We again conducted a Monte Carlo experiment as outlined above with 500 replications and $R, P = (250; 500; 750; 1,000)$ and computed S_P and $\tilde{S}_P$. Size results are reported in Table 19.

	R = 250				R = 500			
P	250	500	750	1,000	250	500	750	1,000
S_P	0.138	0.182	0.250	0.268	0.108	0.154	0.228	0.194
$\tilde{S}_P$	0.088	0.096	0.082	0.118	0.074	0.078	0.092	0.074

	R = 750				R = 1,000			
P	250	500	750	1,000	250	500	750	1,000
S_P	0.128	0.142	0.228	0.184	0.100	0.090	0.180	0.156
$\tilde{S}_P$	0.090	0.098	0.084	0.064	0.084	0.062	0.078	0.084

Table 8: Results

For each combination of R and P the effect of the variance correction results in a much lower empirical coverage for $\tilde{S}_P$ and for low π empirical and nominal coverage do hardly deviate from each other.

In Figure 10, the density of the true asymptotic distribution of S_P and $\tilde{S}_P$, ie the normal distribution, as well as the kernel density estimation of the test statistic S_P as well as $\tilde{S}_P$ of the corrected test for $R = 250$ and $P = 500$ and $\alpha = 0.05$ are plotted. Whereas the density of S_P deviates considerably from its asymptotic distribution, the kernel density of the corrected backtest comes much closer to it.

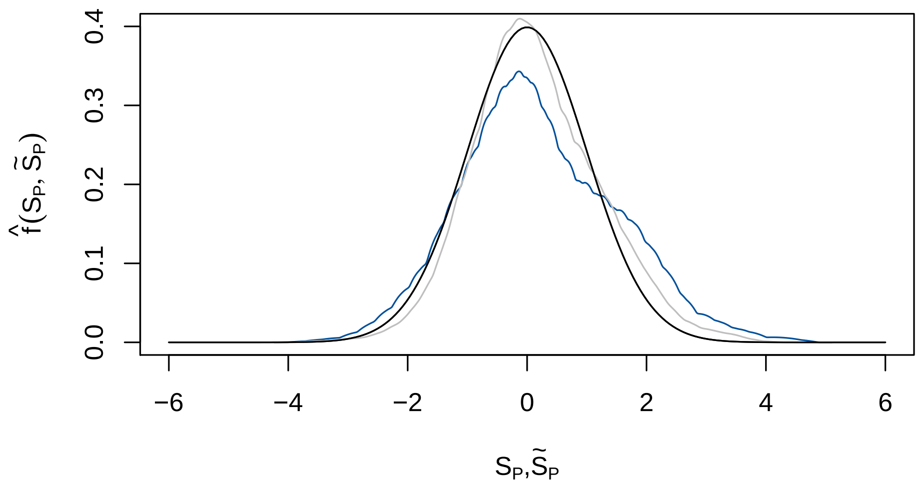


Figure 3: Density of normal distribution ($\mu = 0$, $\sigma = 1$) (black), Kernel density estimate of S_P (blue), Kernel density estimate of $\tilde{S}_P$ (gray) for $R = 250$, $P = 500$ and $\alpha = 0.05$

However, for the Basel II relevant period length of $R = 250$ and the VaR level of $\alpha = 0.01$ size distortions remain at a considerable level of about 3%. The problem therefore remains that the test rejects too often. Looking at the size distortions of the tests proposed by Escanciano and Olmo [2010] we can see that even when accounting for estimation risk the problem prevails. In their follow-up paper for including misspecification risk in their backtest, Escanciano and Olmo [2011] acknowledge that their modified test still suffers from problems of high size distortions also in case of very small in-sample lengths. To put it in a nutshell, all classes of univariate backtests proposed (although duration-based backtests to a lesser extent) have problems when it comes to short in-sample horizons.

Although the corrected backtests result in a reduction of the size distortion, the tests tend to reject too often. Even though the correction for estimation risk has been conducted the problem especially prevails in the Basel II scenario for $R = 250$ and VaR confidence level of $\alpha = 0.01$. In this set-up duration-based backtests with orthonormal approximation of the distribution under the null seem to be the most promising alternative.

5.4 Conclusion

In our paper we analysed the problems of backtests that have been suggested so far. Backtests based on hit and duration sequences in an univariate framework show heavy size distortions. A solution for this is to account for model risk and correct the asymptotic variance of the backtest and thereby reduce the distortion. The problems of univariate backtesting resulting in considerable size distortions for the relevant Basel II set-up however cannot be alleviated by modifying backtests in a way that account for estimation risk or misspecification risk. When financial institutions conduct backtesting, they face restrictions from the regulation side where the in-sample length is set to $R = 250$. A reduction of the out-of-sample length does not suffice to reduce the empirical size. Using inaccurate backtests has severe implications and higher risk-based capital results as the factor for its calculation of directly linked to the number of hits.

A solution suggested by Danciulescu [2010] as well as Berkowitz et al. [2009] is to conduct multivariate backtesting as a mean to overcome these problems. They argue that the sample size is thereby increased and information is more efficiently used for this purpose. In our Monte Carlo study backtests based on orthonormal polynomials performed best. Extending these backtest in a multivariate surrounding would therefore be an alternative to the common approaches. Backtesting with multivariate orthonormal polynomials includes the assumption that under the null the duration sequences follow a respective discrete or continuous multivariate distribution and that this distribution is approximated by Laguerre polynomials in the continuous case. The idea of multivariate backtesting with Laguerre polynomials is a topic to be pursued in further research.

5.5 Appendix

Quasi-Maximum-Likelihood estimation of GARCH(1,1)

As in Francq and Zakoïan [2004] and Escanciano and Olmo [2007].

Model is a pure GARCH(1,1) $Y_t = \mu + \sigma_t \varepsilon_t$ with $\sigma_t^2 = \theta_0 + \theta_1 Y_{t-1}^2 + \theta_2 \sigma_{t-1}^2$ with $\mu = 0$, innovation $\varepsilon_t = Y_t / \sigma_t \overset{iid}{\sim} t(\nu)$ and parameter vector $\theta = (\theta_0, \theta_1, \theta_2)$.

Asymptotic normality of QMLE:

$$\sqrt{T}(\hat{\theta} - \theta)' \xrightarrow{d} N(0, V)$$
$$V = J^{-1} I J^{-1}$$

Conditional Gaussian quasi-log-likelihood:

$$L = \sum \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left(-\frac{Y_t^2 - \mu}{2\sigma_t^2}\right)$$
$$\tilde{l}_t = -\frac{1}{2} log(2\pi) - \frac{1}{2} log(\sigma_t^2) - \frac{1}{2}\frac{Y_t^2}{\sigma_t^2} = -\frac{1}{2}\left\{ log(2\pi) + log(\sigma_t^2) + \frac{Y_t^2}{\sigma_t^2} \right\}$$

Score:

$$\frac{\partial \tilde{l}_t}{\partial \theta} = -\frac{1}{2}\left\{ \frac{\partial(log(\sigma_t^2))}{\partial \theta} + \frac{\partial(\frac{Y_t^2}{\sigma_t^2})}{\partial \theta} \right\} = -\frac{1}{2}\left\{ \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta} - \frac{Y_t^2}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta} \right\}$$
$$= -\frac{1}{2}\left\{ 1 - \frac{Y_t^2}{\sigma_t^2} \right\}\left\{ \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta} \right\} = -\frac{1}{2}\{1 - \varepsilon_t^2\}\left\{ \frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta} \right\}$$

Hessian:

$$\frac{\partial^2 \tilde{l}_t}{\partial \theta \partial \theta'} = -\frac{1}{2}\left\{ -Y_t^2 \frac{\partial \sigma_t^{-2}}{\partial \theta}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta} + \left(1 - \frac{Y_t^2}{\sigma_t^2}\right)\left(\frac{\partial \sigma_t^{-2}}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta} + \frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} \right) \right\}$$
$$= -\frac{1}{2}\left\{ -Y_t^2 \frac{\partial \sigma_t^{-2}}{\partial \theta}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta} + \left(1 - \frac{Y_t^2}{\sigma_t^2}\right)\left(\frac{\partial \sigma_t^{-2}}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta} \right) + \left(1 - \frac{Y_t^2}{\sigma_t^2}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} \right) \right\}$$
$$= -\frac{1}{2}\left\{ \frac{\partial \sigma_t^{-2}}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta}\left[-\frac{Y_t^2}{\sigma_t^2} + \left(1 - \frac{Y_t^2}{\sigma_t^2}\right) \right] + \left(1 - \frac{Y_t^2}{\sigma_t^2}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} \right) \right\}$$
$$= -\frac{1}{2}\left\{ -\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta}\left(1 - 2\frac{Y_t^2}{\sigma_t^2}\right) + \left(1 - \frac{Y_t^2}{\sigma_t^2}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} \right) \right\}$$
$$= -\frac{1}{2}\left\{ \left(1 - \frac{Y_t^2}{\sigma_t^2}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} \right) + \left(2\frac{Y_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta} \right\}$$
$$= -\frac{1}{2}\left\{ (1 - \varepsilon_t^2)\left(\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} \right) + (2\varepsilon_t^2 - 1)\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}\frac{\partial \sigma_t^2}{\partial \theta} \right\}$$

Expected value of Hessian, J:

$$J = E\left[-\frac{1}{2}\left\{(1-\varepsilon_t^2)\left(\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right)+(2\varepsilon_t^2-1)\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\right\}\right] = \frac{1}{2}\left\{E\left[\frac{1}{2}(2\varepsilon_t^2-1)\right]E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\right]\right\}$$
$$= \frac{1}{2}(2E(\varepsilon_t^2)-1)E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta}\right] = \frac{1}{2}E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right]$$

Expected value of squared score, I:

$$I = E\left[-\frac{1}{2}(1-\varepsilon_t^2)\frac{\partial\sigma_t^2}{\partial\theta}\frac{1}{\sigma_t^2}\left(-\frac{1}{2}(1-\varepsilon_t^2)\frac{\partial\sigma_t^2}{\partial\theta}\frac{1}{\sigma_t^2}\right)'\right] = E\left[\frac{1}{4}(1-\varepsilon_t^2)^2\right]E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right]$$
$$= \frac{1}{4}(E(\varepsilon_t^4)+1-E(\varepsilon_t^2))E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right] = \frac{1}{4}(E(\varepsilon_t^4)-1)E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right]$$
$$= \frac{1}{2}(E(\varepsilon_t^4)-1)\frac{1}{2}E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right] = \frac{1}{2}(E(\varepsilon_t^4)-1)J$$

Hence, asymptotic covariance matrix of QMLE, V:

$$V = J^{-1}\frac{1}{2}(E(\varepsilon_t^4)-1)JJ^{-1} = J^{-1}\frac{1}{2}(E(\varepsilon_t^4)-1)$$
$$= \frac{1}{2}(E(\varepsilon_t^4)-1)2\left[E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right]\right]^{-1} = (E(\varepsilon_t^4)-1)E\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right]^{-1}$$

Consistent estimate of V:

$$\hat{V} = (\kappa-1)\left[P^{-1}\sum_{t=R+1}^{n}\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right]^{-1}$$

where

$$\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'} = \begin{pmatrix} \psi^2 & \psi\sum_{j=1}^{\infty}\theta_2^{j-1}Y_{t-j}^2 & \psi\sum_{j=1}^{\infty}\theta_2^{j-1}\sigma_{t-j}^2 \\ \psi\sum_{j=1}^{\infty}\theta_2^{j-1}Y_{t-j}^2 & \left(\sum_{j=1}^{\infty}\theta_2^{j-1}Y_{t-j}^2\right)^2 & \sum_{j=1}^{\infty}\theta_2^{j-1}Y_{t-j}^2\sum_{j=1}^{\infty}\theta_2^{j-1}\sigma_{t-j}^2 \\ \psi\sum_{j=1}^{\infty}\theta_2^{j-1}\sigma_{t-j}^2 & \sum_{j=1}^{\infty}\theta_2^{j-1}Y_{t-j}^2\sum_{j=1}^{\infty}\theta_2^{j-1}\sigma_{t-j}^2 & \left(\sum_{j=1}^{\infty}\theta_2^{j-1}\sigma_{t-j}^2\right)^2 \end{pmatrix}$$

with $\psi \equiv (1-\theta_2)^{-1}$ and where κ is the unstandardized kurtosis.

Consistent estimate of A:

$$\hat{A} = f(F_\varepsilon^{-1})F_\varepsilon^{-1}\frac{1}{P}\sum(\frac{1}{\sigma_t}\frac{\partial\sigma_t}{\partial\theta} = f(F_\varepsilon^{-1})F_\varepsilon^{-1}\frac{1}{P}\sum\begin{bmatrix} \frac{1}{2\sigma_t^2(1-\theta} \\ \frac{1}{\sigma_t^2}\sum_{j=1}^{\infty}\theta^{j-1}y_{t-j}^2 \\ \frac{1}{\sigma_t^2}\sum_{j=1}^{\infty}\theta j-1\sigma_{t-j}^2 \end{bmatrix}$$

References

C. Alexander. The present and future of financial risk managment. *Journal of Financial Econometrics*, 3:3–25, 2003.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

R.T. Baillie, T. Bollerslev, and H.O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.

G. Bakshi, C. Cao, and Z. Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, 52:2003–2049, 1997.

G. Balkema and P. Embrechts. High risk scenarios and extremes. *European Mathematical Society*, 2007.

G. Baumgartner, H. B"uhlmann, and M. Koller. Multidimensional valuation of life insurance policies and fair value. Technical report, Schweizerische Aktuarsvereinigung, 2004.

BCBS. Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. Technical report, Basel Committee on Banking Supervision, 1996.

J. Berkowitz, P.F. Christoffersen, and D. Pelletier. Evaluating value-at-risk models with desk-level data. *Management Science, Articles in Advance*, 31:1–15, 2009.

T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69:542–547, 1987.

T. Bollerslev. Glossary to arch (garch). *Working paper*, 2007.

T. Bollerslev and H.O. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73:151–184, 1996.

G. Bonti, M. Kalkbrener, C. Lotz, and G. Stahl. Credit risk concentrations under stress. *The Journal of Credit Risk*, 2:115–136, 2007.

S.D. Campbell. A review of backtesting and backtesting procedures. *Finance and Economics Discussion Series, Federal Reserve Board*, 21, 2005.

B. Candelon, G. Colletaz, C. Hurlin, and S. Tokpavi. Backtesting value-at-risk: A gmm duration-based test. *Journal of Financial Econometrics*, 9:314–343, 2011.

P. F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39:841–862, 1998.

P. F. Christoffersen and D. Pelletier. Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2:84–108, 2004.

T.E. Clark and M.W. McCracken. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105:85–110, 2001.

D.G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151, 1978.

M.P. Clements and D.F. Hendry. *Forecasting in macro-economics.* Chapman and Hall, 1996.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

R. Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16:519–547, 2004.

R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. Technical report, Columbia Center for Financial Engineering, 2007.

V. Corradi and N.R. Swanson. Predictive density evaluation. *Handbook of Economic Forecasting*, 1:197–284, 2006.

M. Crouhy, D. Galai, and R. Mark. Model risk. *Journal of Financial Engineering*, 7:267–288, 1998.

K. Cuthbertson, S.G. Hall, and M.P. Taylor. *Applied econometric techniques.* Harvester Wheatsheaf, 1992.

C. Danciulescu. Backtesting value-at-risk models: A multivariate approach. Technical report, Indiana University, Center for Applied Economics & Policy Research Working Paper No. 004-2010, 2010.

J. Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statistics*, 22:16–29, 2004.

E. Derman. Model risk. *Quantitative Strategies Research Notes, Goldman Sachs*, 1996.

F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 20 (1):134–144, 1996.

Z. Ding, C.W.J. Granger, and R.F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.

P. Embrechts, A. McNeil, and D. Straumann. *Risk Management: Value at Risk and Beyond*, chapter Correlation and dependence in risk management: properties and pitfalls, pages 176–223. Cambridge University Press, 2001.

R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007, 1982.

R.F. Engle and J. Russell. Econometric analysis of discrete-valued irregularly-spaced financial transactions data using a new autoregressive conditional multinomial model. Technical report, University of California San Diego Dept. of Economics, Discussion paper no. 98-10, 1998.

J.C. Escanciano and J. Olmo. Estimation risk effects on backtesting for parametric value-at-risk models. Technical report, Indiana University, Center for Applied Economics and Policy Research, 2007.

J.C. Escanciano and J. Olmo. Backtesting parametric value-at-risk with estimation risk. *Journal of Business and Economis Statistics*, 28:36–51, 2010.

J.C. Escanciano and J. Olmo. Robust backtesting tests for value-at-risk models. *Journal of Financial Econometrics*, 9:132–161, 2011.

D. Fantazzini. The effects of misspecified marginals and copulas on computing the value at risk: A monte carlo study. *Computational Statistics and Data Analysis*, 53:2168–2188, 2009.

C. Francq and J.-M. Zakoïan. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10:605–637, 2004.

C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82 (3):543–552, 1995.

E. Ghysels, A. harvey, E. Renault, and G.S. Maddala. Stochastic volatility. *Handbook of Statistics: Statistical Methods in Finance*, 14:119–192, 1996.

L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801, 1993.

T.C. Green and S. Figlewski. Market risk and model risk for a financial institution. *Journal of Finance*, 54:1465–1499, 1999.

A. Hamerle and D. Rösch. Misspecified copulas in credit risk models: How good is gaussian? *Journal of Risk*, 8:41–58, 2005.

F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics: The approach based on influence functions*. Wiley, New York, 1986.

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

S. Jaschke, R. Stehle, and G. Stahl. Value-at-risk forecasts under scrutiny - the german experience. *Quantitative Finance*, 7:621–636, 2007.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman Hall, 1997.

J. Jondeau and M. Rockinger. The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25:827–853, 2006.

J. Kerkhof, B. Melenberg, and H. Schumacher. Model risk and regulatory capital. *Available at SSRN: http://ssrn.com/abstract=301531*, 2002.

J. Kerkhof, B. Melenberg, and H. Schumacher. Model risk and capital reserves. *Journal of Banking & Finance*, 34:267–279, 2010.

G. Kim, M.J. Silvapulle, and P. Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, 51:2836–2850, 2007.

P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3:73–84, 1995.

T.-H. Lee and X. Long. Copula-based multivariate garch model with uncorrelated dependent errors. *Journal of Econometrics*, 150:207–218, 2009.

F. Longin and B. Solnik. Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance*, 14:3–26, 1995.

J.A. Lopez. Methods for evaluating value-at-risk estimates. Technical report, Federal Reserve Bank of San Francisco, Economic Review, 1999.

C. Luedtke and P. Sibbertsen. *Model Risk, Identification, Measurement and Management*, chapter Model Risk in GARCH-Type Financial Time Series, pages 78–89. Riskbooks, 2010.

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

C. Marshall. *Measuring and Managing Operational Risks in Financial Institutions*. Wiley, 2001.

S. Morgenthaler and J.W. Tukey. *Configural Polysampling*. Wiles, New York, 1991.

D.B. Nelsen. Stationarity and persistence in the garch(1,1) model. *Ecomometric Theory*, 6: 318–334, 1990.

T. Padoa-Schioppa. *Regulating Finance*. Oxford Univ. Press, 2004.

A.J. Patton. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47:527–556, 2006.

R. Rebonato. Managing model risk. *Mastering Risk*, 2:82–115, 2001.

O. Rousseeuw and V.J. Yohai. Robust regression by means of s-estimators. *Springer Lecture Notes in Statistics*, 26:256–272, 1984.

P. Sibbertsen. S-estimation in the linear regression model with long-memory error terms under trend. *Journal of Time Series Analysis*, 22:353–363, 2001.

P. Sibbertsen, G. Stahl, and C. Luedtke. Measuring model risk. *Journal of Risk Model Validation*, 2:65–81, 2009.

A. Sklar. Fonctions de repartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Universite de Paris*, 8:229–231, 1959.

G. Stahl. Three cheers. *Risk Magazine*, 10:67–69, 1997.

G. Stahl, E. Nill, B. Siehl, and J. Willsberg. Risk modeling and model risk - the irba case. Technical report, Bundesanstalt f"ur Finanzdienstleistungsaufsicht, 2008.

UCG. Insurance: risk-adjusted insurance-company valuation. Technical report, UniCredit Group, 2008.