LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
CENTER FOR QUANTITATIVE RISK ANALYSIS
CEQURA

Stefan Mittnik

Solvency II Calibrations: Where Curiosity Meets Spuriosity

Working Paper Number 04, 2011
Center for Quantitative Risk Analysis (CEQURA)
Department of Statistics
University of Munich

http://www.cequra.uni-muenchen.de

Solvency II Calibrations: Where Curiosity Meets Spuriosity*

Stefan Mittnik

Chair of Financial Econometrics, Department of Statistics, and

Center for Quantitative Risk Analysis, Ludwig–Maximilians–Universität Munich, Germany

September 19, 2011

Abstract

The European Union's Solvency II regulatory framework, which is currently under development, specifies procedures and parameters for determining solvency capital requirements (SCR) for insurance companies. The proposed standard SCR calculations involve two steps. First, the risks of all individual business units, expressed in terms of Value–at–Risk (VaR), are measured and then, in a second step, aggregated to the company's overall SCR, using a so–called *Standard Formula* provided by the regulator. The Standard Formula has two inputs: the individual VaRs of the risk components and their correlations. The appropriate calibration of these input parameters has been the purpose of various Quantitative Impact Studies that have been conducted during recent years.

In this paper, we demonstrate that the derivation of the calibration parameters for the equity–risk module—with about 25%, on average, the most significant risk component of insurance companies' total SCR—is seriously flawed and gives rise to spurious and highly erratic parameters. As a consequence, an implementation of the Standard Formula with the currently proposed calibration settings is likely to produce inaccurate, erratic and biased capital requirements for equity–risk and, thus, to defeat the purpose of the EU's Solvency II Directive.

*I would like to thank Elmar Helten, Sandra Paterlini and participants at the Financial Econometrics Seminar at Ludwig–Maximilians–Universität and the Quantitative Finance Seminar at Stony Brook University for constructive comments. They are, however, not responsible for any errors or omissions. *Address for correspondence:* Center for Quantitative Risk Analysis, University of Munich, Akademiestr. 1/I, 80799 Munich, Germany; Tel.: +49 (0)89 2180-3224; Email: finmetrics@stat.uni-muenchen.de

Contents

1 Introduction **1**

2 Rolling–window Annualization **3**

3 Annualization and Temporal Dependence **5**

3.1 Return Dynamics 5

3.2 Volatility Dynamics 7

3.3 Consequences for Stress Factors 9

4 Annualization and Asset Dependence **14**

4.1 Correlation Concepts 15

4.2 Annualization and Correlations 20

4.2.1 Correlations from Simulated Daily and Annualized Returns . . 20

4.2.2 Bias and Efficiency 24

4.3 Heavy Tails 30

4.4 Tail Dependence 32

5 Conclusions **35**

6 Appendices **40**

1 Introduction

In June 2007 the European Commission (European Commission, 2007a) proposed a revision of the insurance law in the European Union with the objective[1]

> *"... to ensure the financial soundness of insurance undertakings, and in particular that they can survive difficult periods. This is to protect policyholders (consumers, businesses) and the stability of the financial system as a whole."*

To achieve this, the Solvency II Directive (European Parliament, 2009) aims at linking regulatory and economic capital more closely and improving risk management practices, i.e., the identification, measurement and control of risks. In addition to pure insurance risks, Solvency II also includes Solvency Capital Requirements associated with market risk, credit risk and operational risk. Moreover, the EU Directive specifies in detail the kind of losses the capital requirements need to be capable of absorbing:[2]

> *"... the Solvency Capital Requirement should be determined as the economic capital to be held by insurance and reinsurance undertakings in order to ensure that ... those undertakings will still be in a position, with a probability of at least 99.5%, to meet their obligations to policy holders and beneficiaries over the following 12 months. That economic capital should be calculated on the basis of the true risk profile of those undertakings, taking account of the impact of possible risk-mitigation techniques, as well as diversification effects."*

In other words, the Solvency Capital Requirement (SCR) represents the amount of own funds that would potentially be consumed by unexpected loss events, whose probability of occurrence within a one–year period is 0.5% or less. This concept equates the SCR directly to the Value–at–Risk (VaR) risk measure for the 99.5% confidence level and a time horizon of one year. Moreover, the Directive requires that the determination of risk capital takes diversification effects into account

To determine its SCR, an insurer can use the *Standard Formula* and parameters provided by the regulator, use its own internal model, or a use combination of the two. The Standard Formula has a modular structure and is to be applied in a stepwise, bottom–up procedure. First, capital charges are derived for each risk (sub–)module which are then, step–by–step, aggregated to the overall SCR. To

[1]European Commission (2007b), §1.
[2]European Parliament (2009), §65.

allow for diversification effects among the risk components, their correlations enter the calculations. The following main risk modules are considered:[3]

1. market risk
2. counterparty risk
3. life underwriting risk
4. health underwriting risk
5. non–life underwriting risk

According to the report (EIOPA, 2011) on the Fifth Quantitative Impact Study (QIS5), initiated by the Committee of the European Insurance and Occupational Pension Supervisors (CEIOPS),[4] the market–risk module, having a weight of more than 60% of overall SCR, is the most important module. The market–risk module consists of several submodules, of which, the equity–risk is the largest submodule.[5] It makes up about 40% of market risk and, thus, contributes about 25% to the overall SCR.[6]

In the analysis below, we focus on equity risk; but it is to be expected that the findings may also apply to other submodules within the market–risk module.

The *Basic Solvency Capital Requirements* (BSCR) includes the five main modules listed above. They are aggregated, allowing for diversification effects, by use of the Standard Formula:

$$BSCR = \sqrt{\sum_{i=1}^{5}\sum_{j=1}^{5} \rho_{ij} \times SCR_i \times SCR_j}, \tag{1}$$

where SCR_i represents the ith risk module's capital charge, which is given by the 99.5% VaR of that module; and $\rho_{i,j}$ denotes the correlation between the risk modules i and j. If a main module is segregated into submodules, the latter are aggregated, analogous to the Standard Formula (1), to obtain the main module's SCR from the submodules' SCRs.

The Standard Formula will play a crucial role in future regulation and management of insurers' risk as, for reasons of simplicity and cost efficiency, it will be fully or partially adopted by most insurance companies. Only for large and/or "sophisticated" companies will it be efficient to develop an internal model. But even in this case, the Standard Formula will, in one way or another, represent a kind of anchor for any (partial) internal model. Therefore, a proper calibration of the input parameters entering the Standard Formula, i.e., risk–specific SCR factors and correlations, are of ultimate importance to ensure a sound regulatory framework.

[3] These five modules make up the *Basic SCR*. By adding the SCR of the sixth main module, the operational–risk module, to the Basic SCR (without allowing for any diversification effects), one obtains the company's overall SCR.

[4] See CEIOPS (2010).

[5] The other submodules are: interest rate risk, currency risk, property risk, spread risk, and concentration risk.

[6] See Graph 11 in EIOPA (2011) for the relative weights of the individual risk components.

In the following, focusing on the equity–risk submodule, we will demonstrate that the QIS5 calibration procedure for risk assessments leads to SCR risk factors and correlations that are "spurious" and far from reliable. In present context, the expression "spurious correlation" refers to the situation, where the observed correlation between two variables is not genuine, but the result of "... the special case in which a correlation is not present in the original observations but is produced by the way the data are handled;" see Voigt (2005).

It turns out that the annualization procedure, transforming daily return data into annual returns, causes the QIS calibration parameters to be distorted. The chosen annualization strategy has serious implications as it affects risk and dependence structures in the data used for calibration. For one, it induces spurious dependence patterns, which are not genuinely present in the observed data. Secondly, it may alter or destroy dependence patterns that factually determine the riskiness of individual asset classes. Specifically, two types of dependencies matter in risk assessment: (i) temporal or dynamic dependencies, describing an asset's return and risk behavior over time; and (ii) cross–sectional dependencies, i.e., the relationship between assets at a given point in time. The dependencies along both dimensions need to be understood and properly modeled in order to reliably assess the risk of equity portfolios. It turns out that the currently proposed Solvency II calibrations equity risk obstruct both the understanding and modeling of risk and, thus, obfuscate insurers' equity–risk assessment.

Our findings differ from the criticism against specific QIS–calibration choices that has been raised before[7] in that it is more fundamental, calling virtually all calibration parameters specified for the equity–risk module into question, as they are largely a product of chance.

The organization of this paper follows the two possible dimensions dependencies can take affect, namely temporal and cross–sectional dependencies. After reviewing the annualization procedure chosen for QIS calibrations, Section 3 investigates consequences for return and risk dynamics that arise from the chosen annualization procedure, and Section 4 those for the dependence among asset classes. Section 5 summarizes the implications of our findings.

2 Rolling–window Annualization

Solvency II calibrations for the equity–risk module are designed for assessing the risk of various asset classes assuming a one–year holding period. Therefore, all SCR or, for that matter, VaR–calibrations refer to that horizon and, accordingly, the inputs for the Standard Formula need to be VaRs and correlations associated with annual returns. As a consequence, a straightforward calibration strategy would preferably

[7] See, for example, EIOPA (2011).

rely on annual return data for deriving the inputs for the Standard Formula. However, most of the asset classes considered in the equity–risk module have a rather short history, so that the analysis would rest on only very few annual return observations. Specifically, having daily data histories ranging from about 8 to almost to 40 years,[8] it is not possible to assess risks associated with once–in–two–hundred–years events as the $\text{VaR}_{99.5}$ measure implies. Given 8 to 40 non–overlapping annual return observations, we cannot directly derive VaR–estimates at a 99.5% confidence level nor the type of correlation, namely tail correlations, employed in QIS calibrations.

To still make use of historical market data, QIS calibrations employ rolling one–year data–windows to obtain annual returns at a daily frequency. Letting P_t denote the price of an asset at day t and w the window length (measured in trading days)[9] for which the multi–period return, denoted by R_t^w, is to be computed, we have

$$R_t^w = \frac{P_t - P_{t-w}}{P_{t-w}}, \quad w \geq 1, \quad t = w+1, w+2, \ldots \; . \tag{2}$$

Given, say 10 years of daily return data, the rolling–window approach gives rise to 9 years of annual return observations at a daily frequency. However, the annual returns that are generated in this manner overlap to a large extent. Annual returns computed for two consecutive days have more than 99% of daily return information in common and differ only by two daily return data that are not in common. Clearly, the use of *non–overlapping* annual return data is essential, because only they represent independent pieces of information about the underlying data generating process. CEIOPS analysts were well aware of this problem and write:[10]

> *"There is a balance to be struck between an analysis based on the richest possible set of relevant data and the possibility of distortion resulting from autocorrelation. In this case, we have chosen to take a rolling one-year window in order to make use of the greatest possible quantity of relevant data ."*

As will be demonstrated below, the "distortions" induced by the rolling one–year window approach are not as inconsequential as the above quotation may suggest. The most damaging implication is that the approach tends to induce spurious dependence patterns, both over time and across assets, which, in turn, produce artifactual risk structures.

[8]To compute correlations among the asset–class indices used in the sub–module equity risk there are about 40 years of daily observations for the asset pair Global Equity/Commodities, about 15 for the pairs Global Equity/Private Equity and Global Equity/Emerging–Markets Equity, and about 8 for the pair Global Equity/Hedge Fonds.

[9]In the simulations discussed below, we choose window lengths ranging from $w = 1$, to indicate no temporal aggregation, and $w = 259$, representing an aggregation over one calendar year. The latter corresponds to the average number of trading days recorded for the MSCI World index, which plays a prominent role in the Solvency II equity–risk calibrations.

[10]See §3.56 in CEIOPS (2010).

Before entering this discussion, a remark on the notation adopted below is in order. Returns calculated via (2) are referred to as *discrete* returns. For reasons of analytical tractability, empirical and theoretical analyses in finance typically employ approximations in form of *continuous* returns, defined by $r_t^w = \log P_t - \log P_{t-w}$. However, for annual returns, this approximation may be poor.[11] Throughout the paper, we denote discrete returns by upper case R_t and continuous returns by lower case r_t. Whenever we examine theoretical issues analytically, we will resort to the continuous approximation, r_t. All simulations, however, are conducted with exact, discrete returns, R_t.

3 Annualization and Temporal Dependence

Our analyses of the impact on temporal dependence when conducting equity–risk calibrations with annualized rolling–window returns are threefold. We, first, investigate the consequences of the chosen annualization for the dynamic properties of the *returns* and, then, for the *volatility* of asset returns. Finally, we investigate the implications on the calibration of the SCR stress factors which enter the Standard Formula.

3.1 Return Dynamics

The determination of VaR values from historical rolling–window return data may, at first sight, seem reasonable, as this amounts to searching for worst–case outcomes over all possible one–year holding periods in the sample at hand. However, construction of a daily series of annual returns via overlapping rolling–windows causes the resulting return series to be highly autocorrelated. The autocorrelation between consecutive (continuous) multi–period returns, denoted by r_t^w and r_{t-1}^w, becomes stronger as the length of the rolling window, w, increases, so that

$$\mathrm{Corr}(r_t^w, r_{t-1}^w) \overset{w\to\infty}{\longrightarrow} +1. \tag{3}$$

As w increases, the times series r_t^w, $t = 1, 2, \ldots$, approaches a random–walk–like process and, thus, approaches nonstationarity. This, in turn, implies that the joint distribution of a set of consecutive observations on r_t^w tends to vary over time. A random–walk, say x_t, in its purest form is generated by the stochastic first–order difference equation

$$x_t = a x_{t-1} + u_t, \tag{4}$$

with $a = 1$, and u_t being a white–noise series, i.e., an independent and identically distributed (iid) time series with $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t^2) = \sigma^2 < \infty$ and $\mathrm{E}(u_s u_t) = 0$, for $s \neq t$. Process (4) with $a = 1$ is also referred to as a *unit-root process.*[12]

[11] See Appendix A for a discussion on this issue.

[12] The term “unit root” is used, because the autoregressive polynomial has a root of size one.

Expressing the rolling–window returns, r_t^w, $t = 1, 2, \ldots, T$, by the first–order process

$$r_t^w = a r_{t-1}^w + v_t, \tag{5}$$

the ordinary least–squares (OLS) estimator of autoregressive coefficient, $\hat{a}_T$, approaches $(w-1)/w$ as the sample size, T, grows, i.e.,[13]

$$\hat{a}_T \overset{T\to\infty}{\longrightarrow} \frac{w-1}{w}. \tag{6}$$

It is well known that temporal and cross–sectional correlation analysis with unit–root processes will produce spurious and highly erratic results due to the peculiar dependence patterns that may arise.[14] To investigate the extent to which rolling–window annualization induces autocorrelation in finite samples, we conduct a Monte Carlo simulation and generate 10,000 daily (continuous) return series, r_t, $t = 1, 2, \ldots, T$, of length $T = 2,590$ and $T = 5,180$, with returns being independent and identically normally distributed, i.e., $r_t \overset{iid}{\sim} \mathrm{N}(0, 1)$. The chosen sample sizes, T, corresponds to about 10 and 20 years of daily observations, respectively. From each of the series we compute (discrete) rolling–window returns, R_t^w, with the window length, w, assuming values $w \in \{5, 22, 65, 130, 259\}$.[15] These values correspond more or less to aggregating daily returns to weekly, monthly, quarterly, semi–annual, and annual returns. By letting the window length grow, we can assess how the severity of the problems increases as the aggregation level grows. We, then, estimate the first–order autoregressive coefficient and, using the ADF–test (Dickey and Fuller, 1979), formally test for the presence of a unit–root.

The test results are summarized in Table 1, where the first column states the length of the aggregation window; Column 2 indicates the asymptotic value of the autoregressive (AR) coefficient, $\hat{a}$ in (6), associated with that window length; Columns 3 and 4 show the mean values of the 10,000 AR–coefficient estimates for the two sample sizes, respectively. The last two columns report the means of the ADF–statistics. The critical values of the ADF–statistic for the 99%, 95% and 90% levels are -3.4583, -2.8710, and -2.5937, respectively. If the value of the ADF–statistic lies above the critical value, we cannot reject the null hypothesis of a unit root.

The results in Table 1 indicate that—in line with the asymptotic counterpart—the estimated first–order AR–coefficient quickly increases as the window lengths, w, exceeds unity. Weekly aggregation produces a value of about 0.80, and monthly aggregation already 0.95. With a mean AR–coefficient of 0.996, annual aggregation produces a nearly perfect random walk. According to the ADF test, for the one–year rolling–window aggregation (i.e., $w = 259$) and the 10–year sample, we cannot reject the null hypothesis of a unit root at any conventional significance level. For the larger, 20–year sample, we can reject at the 90% and 95% levels, but not at the

[13]See Appendix B for details.

[14]See Granger and Newbold (1974). We will return to this issue in Section 4 below.

[15]See Appendix A for a description of the simulation of discrete multi–period returns.

Table 1: Asymptotic and simulated near–unit–root behavior of rolling–window returns

	AR Coefficient			ADF Statistic	
Window Length	Asymptotic	10 years	20 years	10 years	20 years
1	0	0.0002	0.0002	-20.7662	-29.3696
5	0.8	0.7997	0.7997	-13.6088	-19.2407
22	0.9545	0.9539	0.9542	-8.2773	-11.6922
65	0.9846	0.9841	0.9843	-4.6196	-6.5061
130	0.9923	0.9917	0.9920	-3.2865	-4.5858
259	0.9961	0.9955	0.9958	-2.3976	-3.2892

99% level. These findings suggest that for large samples, i.e., 20 years or more, a formal test is likely to reject the presence of a unit root. The outcome of the test is, however, merely a question of sample size. The nature of the rolling–window return series will be determined by the value of w or, for that matter, the implied AR coefficient. A value of $w = 259$ turns out ot induce strong temporal dependence and to distort calibration exercises.

To illustrate this, we simulate 40 years of daily return data with a normally distributed white–noise structure and perform rolling–window annualization. The top graph in Figure 1 shows a typical sample autocorrelation function (SACF) for the two series, i.e., $\mathrm{Corr}(R_t, R_{t-k})$ and $\mathrm{Corr}(R_t^{259}, R_{t-k}^{259})$, $k = 1, 2, \ldots, 259$. The SACF for daily returns looks like what we expect from white noise: it is close to zero for all lags and remains pretty much within the approximate 95% confidence band. The SACF for the annualized returns resembles that of a unit–root series. It starts near one, decays in a very slow and linear fashion, and is significantly different from zero. The behavior of the SACFs is compatible with the scatter plots of the two series (Figure 1, bottom).

These simulations demonstrate that rolling–window annualization alters the temporal dependence structure of the returns in a substantial way. We will see in Section 3.3 that this is not just a theoretical problem, but it has practical consequences.

3.2 Volatility Dynamics

Rolling–window annualization not only affects the dynamics of the return series in terms of autocorrelations. Also the volatility dynamics, i.e., risk dynamics, will be altered. Volatility reflects the extent to which the return process can deviate from its expected value; and variations in the return volatility reflect variations in the riskiness of an asset. If volatility dynamics exhibit particular patterns over time, prudent risk assessment needs to take these into account. If such patters are, however, spurious and only the consequence certain data transformations rather

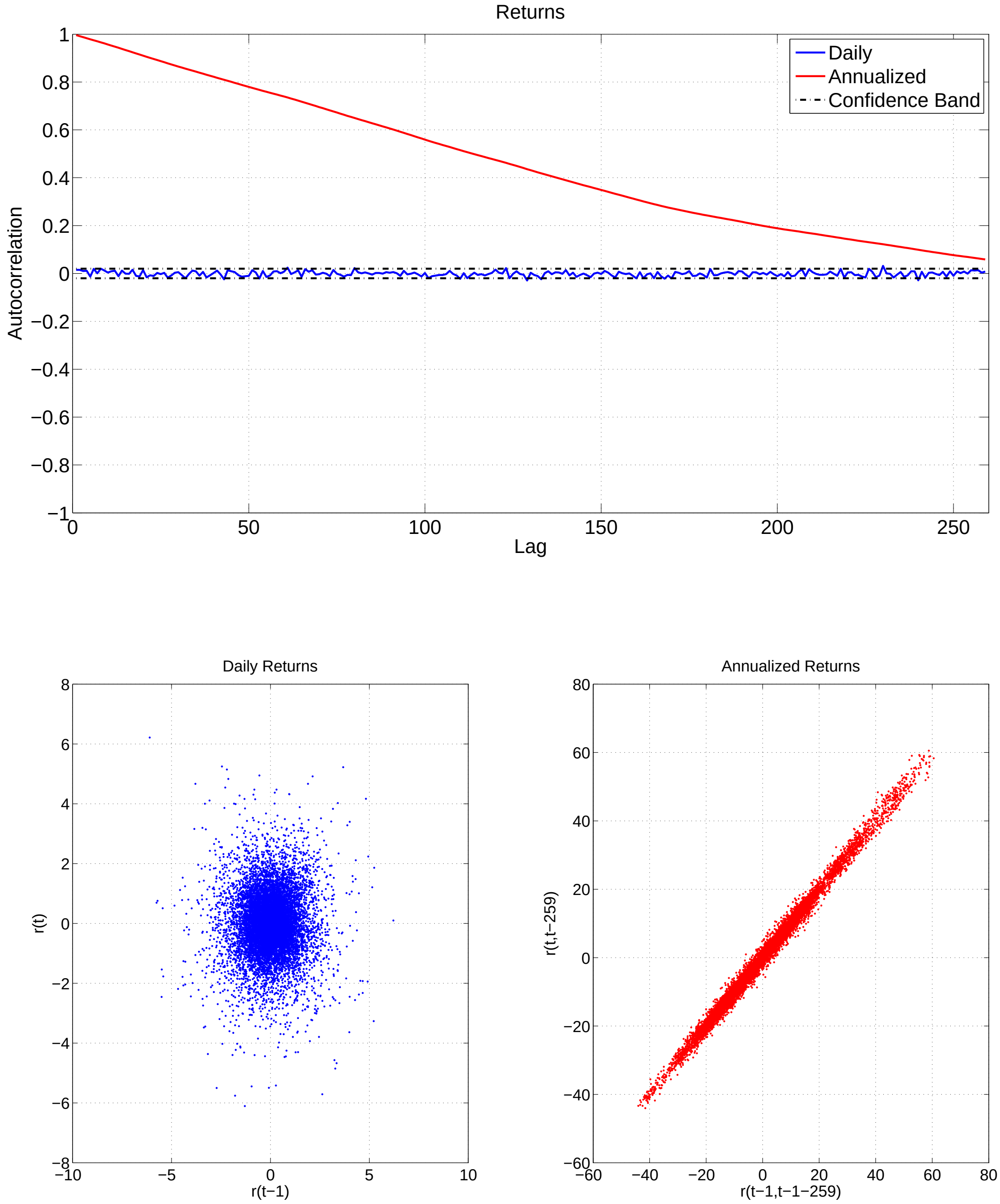


Figure 1: Sample autocorrelations (top) and scatter plots of simulated daily (bottom left) and annualized (bottom right) returns

than a genuine property of the underlying return process, efforts toward systematic risk management will be seriously undermined.

The class of Generalized Autoregressive Conditional Heteroskedasticty (GARCH) models, introduced by Engle (1982) and Bollerslev (1986), is the most common model for approximating volatility dynamics of financial assets. To investigate the impact of rolling–window annualization on volatility dynamics, we simulate a standard GARCH(1,1) model of the form

$$r_t = \mu + \sigma_t u_t, \tag{7}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 (r_{t-1} - \mu)^2 + \beta_1 \sigma_{t-1}^2, \tag{8}$$

where u_t is a normally distributed white–noise process with $\mathrm{E}(u_t) = 0$ and $\mathrm{Var}(u_t) = 1$, for all t. For the simulation, we use the GARCH parameters we obtain when fitting model (7)–(8) to the daily returns on the MSCI World Index,[16] the index employed in QIS5 to calibrate the asset class "global equity" within the equity–risk module.

Figure 2 plots the SACFs of the absolute daily and annualized returns, i.e., $\mathrm{Corr}(|R_t|, |R_{t-k}|)$ and $\mathrm{Corr}(|R_t^{259}|, |R_{t-k}^{259}|)$, derived from 40 years of simulated data.[17] The resulting SACF of the absolute daily returns is typical of what we observe for daily stock–index returns. There is a significant positive autocorrelation, starting at about 0.2, which gradually declines, becoming more or less insignificant after a lag of about 80 days. Thus, a (negative or positive) return shock carries over next period's volatility with a correlation of 0.2. The impact gradually vanishes for higher lags. For absolute annualized returns, autocorrelations are much stronger. They start at one and—though gradually decaying—stay much higher than those from absolute daily returns, to become insignificant after about 170 days.

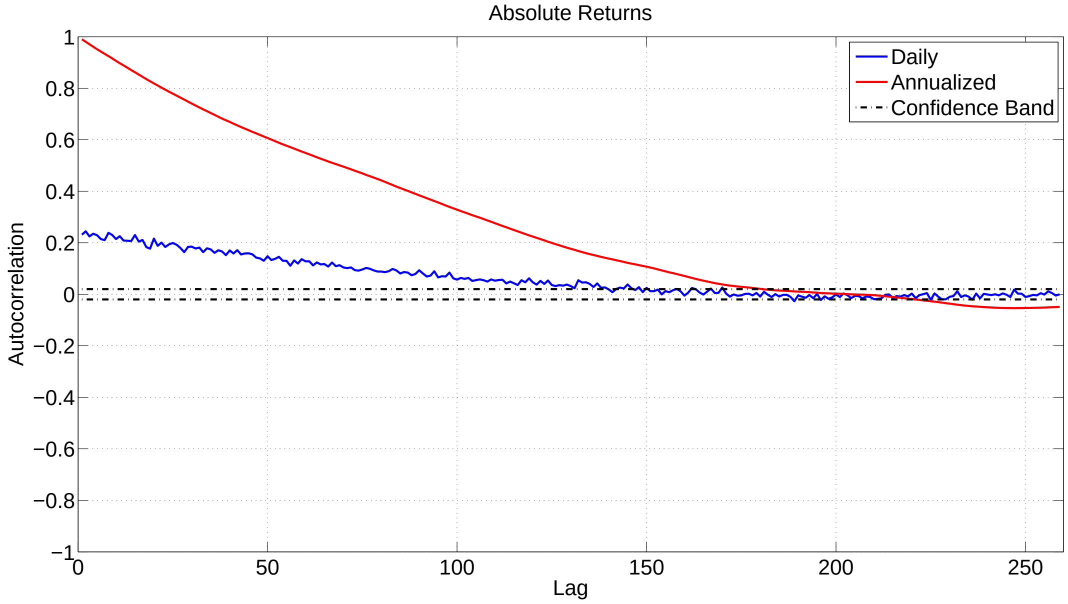
This shows that rolling–window annualization not only affects the temporal correlation of a return series, but it also alters the risk dynamics by inducing much stronger and more persistent temporal risk structures. Consequently, as will be shown next, the calibration of stress factors for individual equity classes can produce extremely misleading results.

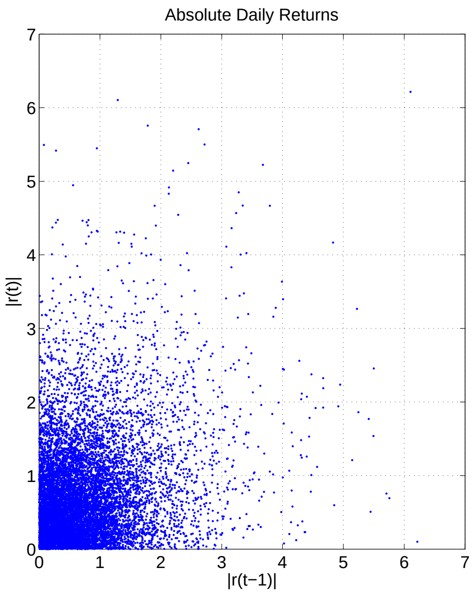
3.3 Consequences for Stress Factors

The presence of unit roots or near–unit roots has implications for both ingredients of the Standard Formula. If a return series is nonstationary, past behavior will be a poor indicator for its future behavior. As a consequence, even if the nature of the return process remains unchanged, past VaR–statistics, for example, do not provide an indication for those encountered in the future. To illustrate this, we conduct a

[16] Specifically, we use the daily MSCI World Price Index in U.S. dollar with the sample ranging from January 4, 1972 to January 31, 2011.

[17] We use *absolute* rather than *squared* returns or unobserved conditional variance, because the absolute returns tend to exhibit superior forecastability; see Granger and Sin (2000).





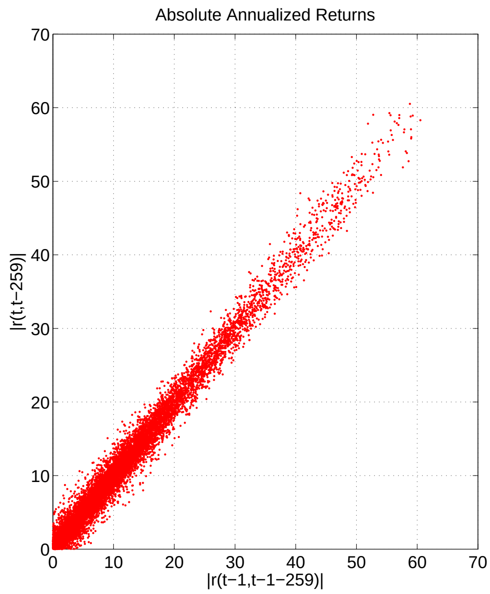


Figure 2: Sample autocorrelations (top) and scatter plots of GARCH–simulated absolute daily (bottom left) and absolute annualized (bottom right) returns

Monte Carlo experiment, generating independent and identically distributed white–noise data, $r_t \overset{iid}{\sim} \mathrm{N}(0, 1)$. Specifically, we simulate two independent risk–factor series, each of length $100 \times 259 = 25,900$ observations, which corresponds to about 100 years of daily return data. We "annualize" these by computing discrete, one–year rolling–window returns, leaving us with 99×259 overlapping annual return data at a daily frequency. Figure 3 plots the daily and annualized returns of the two simulated risk factors.[18]

As was to be expected, being generated by the same process, the two daily return series look pretty much the same. Both annualized versions fluctuate between –50% and +60%, but the locations of their peaks and troughs differ considerably.

We derive SCR estimates for the two simulated risk factors. We do this for the daily and annualized discrete returns by computing day by day the historical $\mathrm{VaR}_{99.5}$–values—in other words, the 0.5%–quantiles of the series—using 10–year rolling samples. Figure 4 shows that the VaRs for the daily returns are rather stable; they hover around the expected value (solid vertical line) and, with a range from 2.3% to 2.8%, stay about 95% of the time within the 95% confidence bounds.[19] Compared to this, VaR estimates from annualized returns vary dramatically. They assume values from 16% to 46% during the 89 years sampled and deviate considerably from the expected value.[20] They deviate by more the than ten standard deviations in either direction and stay for long periods far away from the expected value, so that it is the exception rather than the rule that the estimates fall inside the confidence band.

Given that the data were generated by independent iid white noise processes, i.e., a processes without any temporal dependence structure, the VaR series for annualized returns appears to exhibit distinctive patterns, which may be mistaken for structurally inherent properties. Such SCR patterns may easily trigger specific regulatory actions. Relying on historical VaR estimates from annualized returns, a regulator could be tempted to set the stress factor for Asset 1 much too low during years 29 through 81, just to ratchet it up to an excessively high level after year 88, while, at the same time, inappropriately lowering the stress factor for Asset 2. Similarly disturbing is the fact that, although annual–return VaRs exhibit strong persistence, they can change very abruptly. An insurance company's reliance on annual–return VaRs is bound to induce sudden and erratic portfolio adjustments, without any change in the underlying market processes.

[18] By generating *two* independent series with identical properties we get an impression of the variability of the dynamic properties due to rolling–window annualization. Moreover, below we will use the two series to demonstrate the consequences of annualization on the dependence structure across assets.

[19] Note that Figure 4 plots, against common convention, negative VaR–values to bring them in line with CEIOPS' usage. The sign switch is also compatible with the QIS–documentations' convention of sign reversion.

[20] We obtain a sample of 89 years because we lose the initial 11 years of the sample, namely, one year due to the annualization and ten years to calculate historical VaRs.

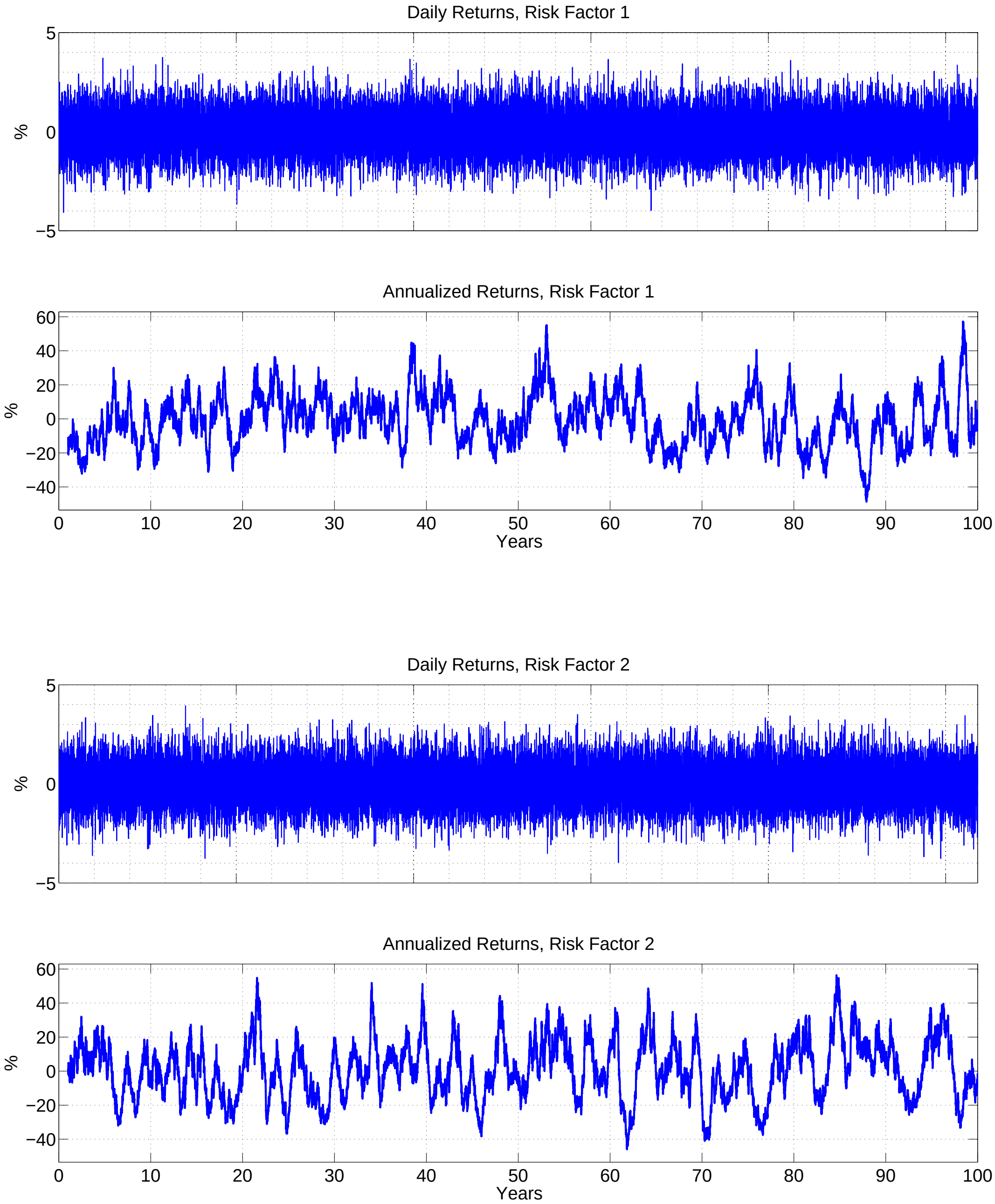


Figure 3: Time Series Plots of two simulated independent daily return series and corresponding annual rolling–window returns

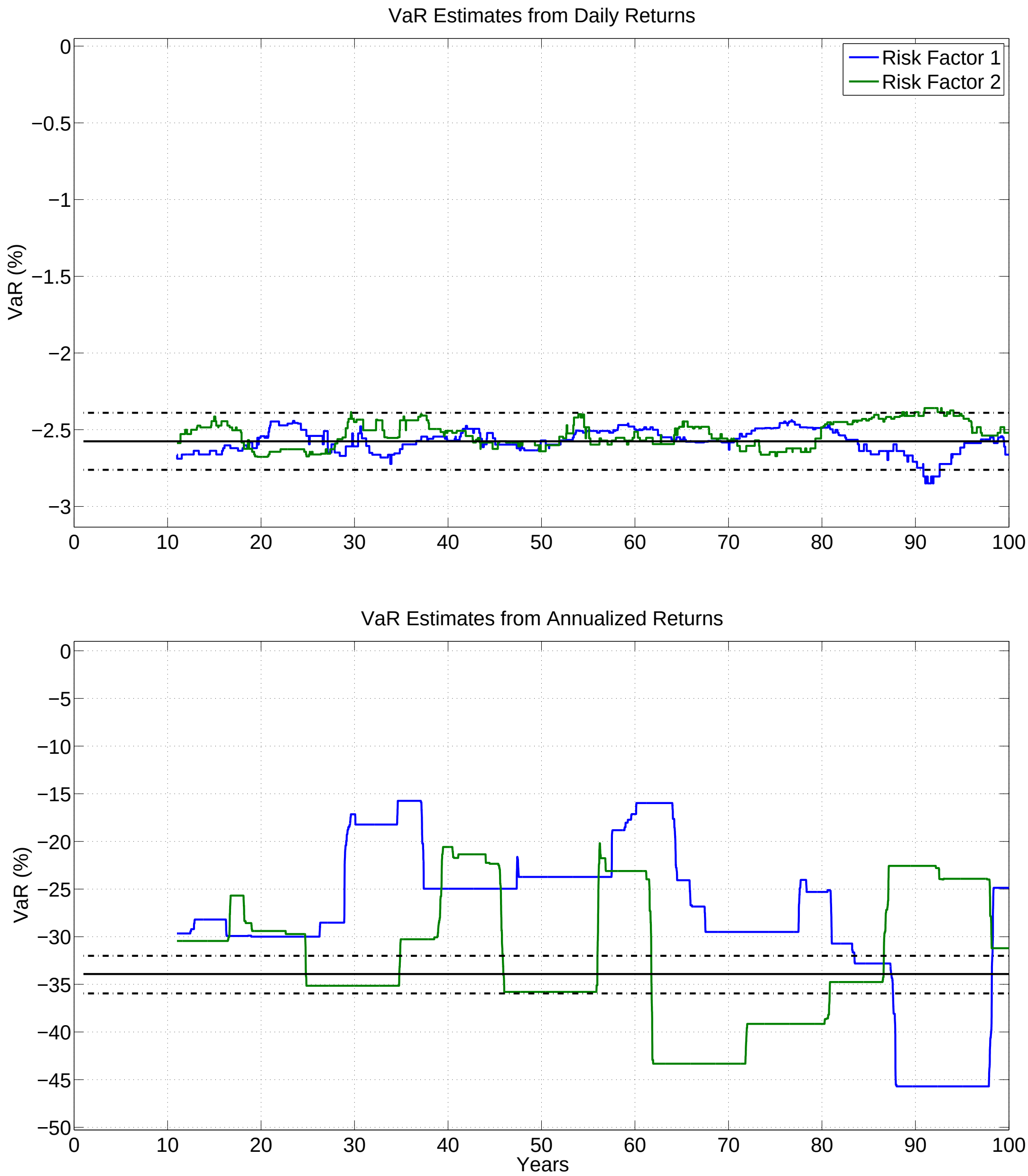


Figure 4: Historical VaRs for daily (top) and annualized (bottom) returns with theoretical VaRs (solid vertical line) and 95% confidence intervals

From all this, it follows that a reliance on VaR estimates derived from one–year rolling–window returns in regulatory or firms' investment processes will produce arbitrary outcomes.

4 Annualization and Asset Dependence

We now turn to the second ingredient of the Standard Formula (1), the correlation parameters that need to be specified in order to aggregate the modules' SCRs to the next higher level.

The most common approach to measure and model dependencies between random variables is to compute the Pearson correlation coefficient. Not only is it easily computed, Pearson correlation is also the cornerstones of modern portfolio theory, which underlies widely adopted risk–diversification concepts, including the Standard Formula. However, Pearson correlation is a measure of *linear* dependence and, thus not appropriate for nonlinear or non-Gaussian risk structures. This limitation has been recognized when developing the Solvency II guidelines. To particularly capture the joint behavior of risk factors in situations of extreme stress, Solvency II calibrations are based on "tail correlations" rather than conventional Pearson–correlation estimates.

Since Granger and Newbold (1974) it is well known that regression analysis involving unit–root processes will produce spurious and highly erratic results.[21] They showed that estimated correlations between two independent random walks, say $x_t = x_{t-1} + u_{xt}$ and $y_t = y_{t-1} + u_{yt}$, with u_{xt} and u_{yt} being two independent white noise series, can assume values far away from zero, even though the two series are totally independent. Clearly, if this is the case, any correlation estimate between two nonstationary times series is unreliable, and its use becomes highly questionable.

Figure 5 indicates the potential problem with assessing the dependence structure between risk factors when the analysis relies on returns derived from rolling–window annualization. The graph in the top half overlays the two independently simulated series of annualized returns plotted in Figure 3. We observe periods where both series seem to run pretty much in sync as well as periods where they are very dissimilar. The scatter plots of the two risk factors in the bottom half of Figure 5 illustrate the difference in the dependence pattern for the daily and the annualized return data. The former (bottom left) is very homogeneous and looks like what we expect from uncorrelated data. In comparison, the scatter plot of the annualized returns (bottom right) looks rather inhomogeneous and is somewhat splattered. This spottiness arises from the fact that the apparent common behavior varies over

[21] Note that the findings for regression analysis between random–walk–type processes immediately carry over to correlation analysis. For a theoretical analysis of regressions with random–walk–like processes see Phillips (1987).

time.

An illustrative selection of subsamples of the bivariate annualized return series is presented in Figure 6. The top panel shows the time series of the subsamples; the bottom panel the corresponding scatter plots. We observe that, over fairly long periods, the two series may exhibit rather strong positive (left and right panels) but also strong negative dependency (center panel). The (sub–)sample correlations for the three cases are 0.42 (left subsample), -0.65 (center) and 0.75 (right). Such variations are typical for pairs of independent (near–)unit root process.

In the following, we investigate the implications of rolling–window annualization on calibrating asset dependence. We begin with an introduction of the alternative correlation concepts considered in Solvency II calibrations. Then, we investigate three specific issues in more depth using, again, Monte Carlo simulations. First, we take a closer look at the consequences of annualization on bias and efficiency of the various correlation estimates. These analyses are simulation–based and limited to normally distributed risk factors. In a second step, we examine to what extent heavy–tailedness may affect the calibration of correlations. We do so by drawing from bivariate t–distributions, such that we still remain in the elliptical world, justifying the use of the Standard Formula. Finally, we investigate how annualization affects the tail–dependence properties between two equity classes.

4.1 Correlation Concepts

QIS calibrations for equity risk are based "tail correlation." One approach to obtain such estimates is to compute the conventional Pearson correlation from joint tail observations. The joint tail observations associated with a given VaR_α–level consist of those return pairs for which both assets fall simultaneously below their respective $(1-\alpha)$–quantile. This approach, illustrated in Figure 7, is referred to as the "data–cutting method" in CEIOPS (2010) and amounts to computing the conditional correlation

$$\rho_\alpha^{DCQ} = \mathrm{Corr}(r_i, r_j \,|\, r_i < -\mathrm{VaR}_\alpha(r_i),\, r_j < -\mathrm{VaR}_\alpha(r_j)). \tag{9}$$

The problem with the data-cutting approach is that, even for large data samples, the number of data point entering the estimation may be extremely small. For example, given 40 years of daily return data, i.e., about 10,000 observations, and adopting the Solvency II convention of using the 99.5%–VaR, only observations falling below the 0.5%–quantile matter. This leaves us with 50 tail–observations for each of the assets. The intersection of these two subsets, data pairs where *both* components fall below the 0.5%–quantile, defines the set of *joint* tail observations. Depending on the degree of dependence, this will leave us with much fewer than 50 observations.

Figure 8 illustrates for a bivariate normal distribution how the portion of common

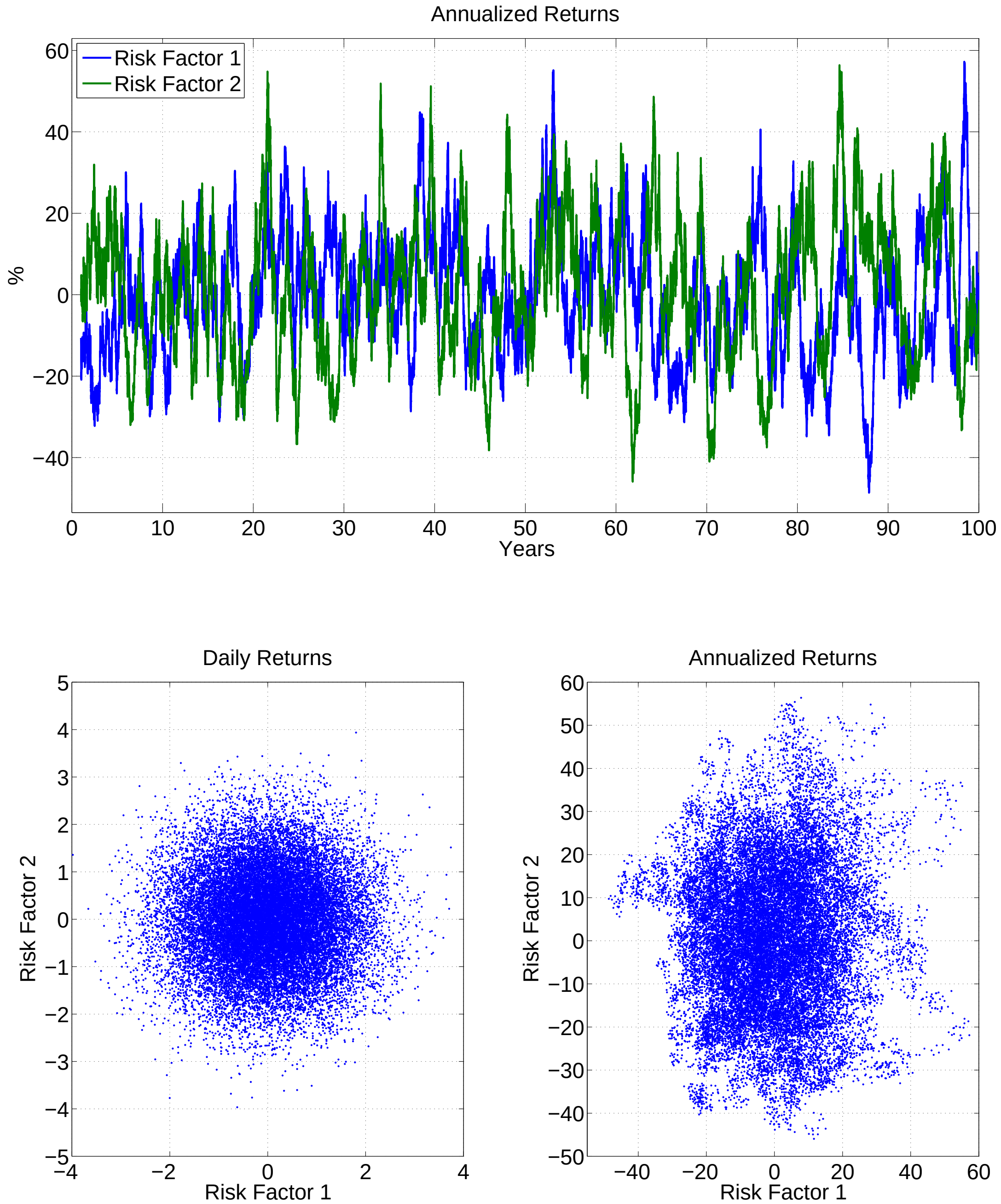


Figure 5: Time series plots of simulated, annualized returns (top) and scatter plots of daily (bottom left) and annualized (bottom right) return series

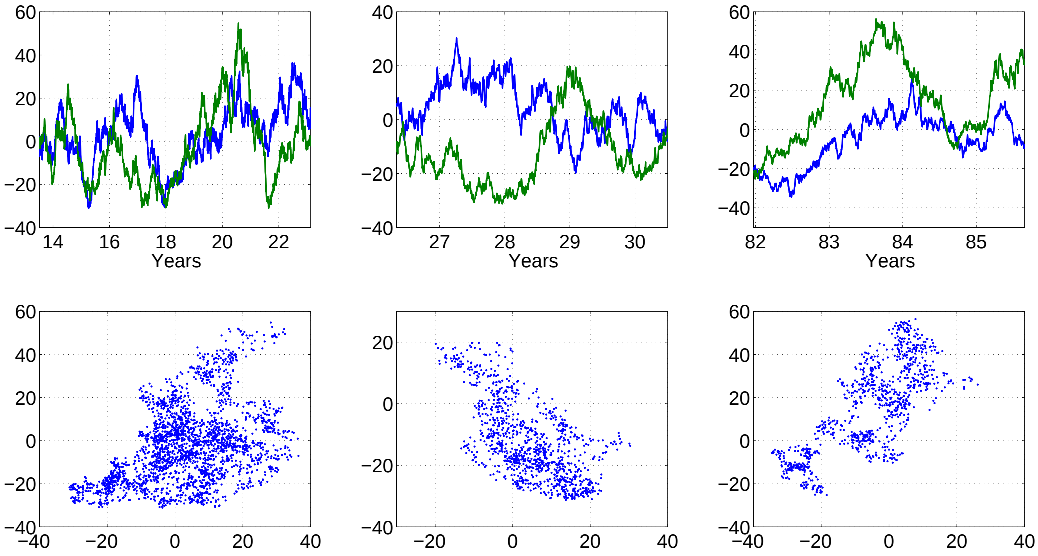


Figure 6: Selected subsets of the annualized returns with time series plots (top) and scatter plots (bottom) with sample correlations $\hat{\rho} = 0.42$, $\hat{\rho} = -0.65$ and $\hat{\rho} = 0.75$ (from left to right)

tail observations quickly drops as we move away from perfect positive correlation. For example, given a correlation of $\rho = 0.75$ and having observations on 10,000 return pairs, we can expect to have 14 joint tail observations. Despite the rare luxury of having such a large data set, the tail–correlation estimate obtained by the data–cutting approach will be based on an extremely small number of data points and, thus, lead to highly unstable estimates.

Apart from the lack–of–data problem, focusing solely on tail, especially, far–tail correlations may provide a misleading picture about possible dependencies between assets. If, for example, two assets follow a nondegenerate joint normal distribution, no matter how strong the correlation is, tail correlations approach zero the further into the tails we go (see Rosenbaum, 1961),[22] suggesting the absence of dependence.

Alternatively, a different data–cutting strategy coule be adopted. Rather than computing correlations from joint tail observations, we could condition on only one risk factor and compute

$$\rho_\alpha^{DCH} = \mathrm{Corr}(r_i, r_j \,|\, r_i < -\mathrm{VaR}_\alpha(r_i)). \tag{10}$$

With this, the two–dimensional return plane is not segmented into quadrants but rather into half–planes. This ensures that we do not end up with an insufficient number of tail observations, as the remaining sample size corresponds to the chosen VaR–quantile. The use of (10) is particularly appropriate when the asset on which

[22]Note that the data–cutting approach is equivalent to the concept of "excess correlation" used in Longin and Solnik (2001) who condition on percentage rather than quantile levels.

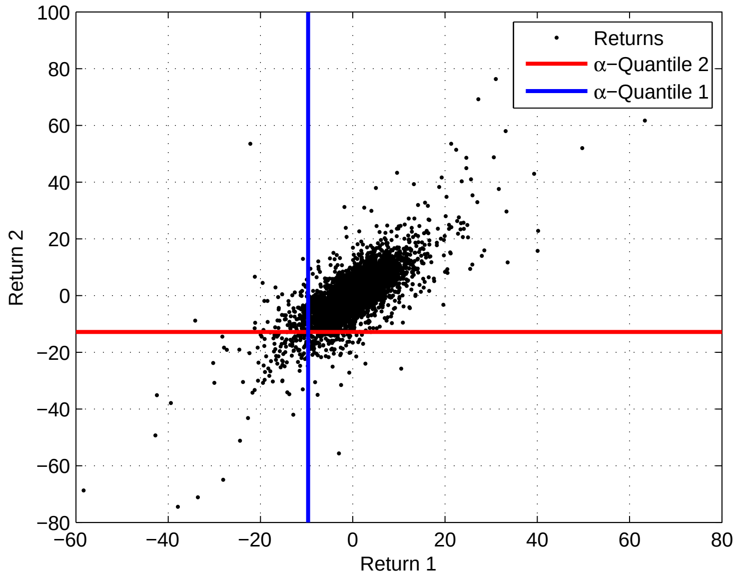


Figure 7: Tail Correlation via Data–Cutting Approach

we condition is regarded as the underlying risk driver.

Because of the small number of data points available to compute tail correlations—even in the presence of large data sets—, QIS calibrations do not, or not exclusively rely on the data–cutting method (9). They (also) seem to adopt what we, below, refer to as VaR–implied correlations,[23] which simply results from an inversion of the Standard Formula.

For two risk components, the Standard Formula reduces to[24]

$$\mathrm{VaR}_\alpha(r_i + r_j) = \sqrt{\mathrm{VaR}_\alpha(r_i)^2 + \mathrm{VaR}_\alpha(r_j)^2 + 2\rho \mathrm{VaR}_\alpha(r_i)\mathrm{VaR}_\alpha(r_j)}. \tag{11}$$

CEIOPS (2010)[25] suggests to use that value for ρ which minimizes the "aggregation error"

$$\left|\mathrm{VaR}_\alpha(r_i + r_j)^2 - \mathrm{VaR}_\alpha(r_i)^2 - \mathrm{VaR}_\alpha(r_j)^2 - 2\rho \mathrm{VaR}_\alpha(r_i)\mathrm{VaR}_\alpha(r_j)\right|.$$

Having empirical VaR–estimates, denoted by $\widehat{\mathrm{VaR}}_\alpha(\cdot)$, for assets i and j as well as

[23]Ultimately, it is not clear what particular method has been used for deriving the correlations entering the Solvency II Standard Formula.

[24]Equation (11) assumes that both return series have mean zero. In practice, this assumption is typically violated. Ignoring this fact, use of (11) will, generally, lead to biased VaR–implied correlations. CEIOPS (2010) justifies the simplifying zero–mean assumption by arguing that their "... calibration intends to quantify unexpected losses" (Footnote 113, p. 338). However, it is left open where the expected means should come from.

[25]See § 3.1251 in CEIOPS (2010).

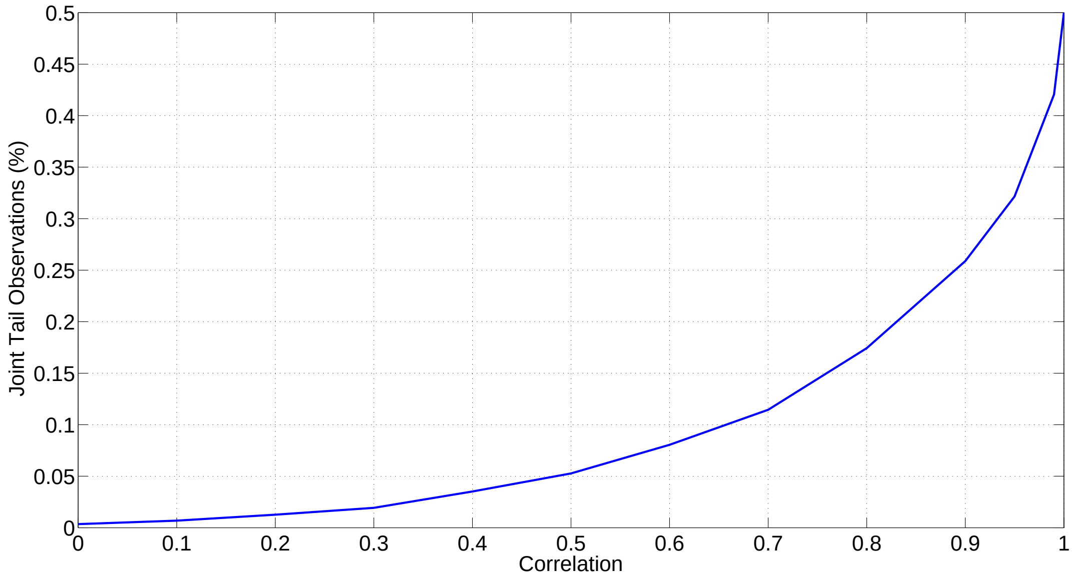


Figure 8: Percentage of joint tail observations in data–cutting approach

for the sum of the two, the minimization amounts to

$$\hat{\rho}_\alpha^{VaR} = \begin{cases} +1, & \text{if } \widehat{\mathrm{VaR}}_\alpha(r_i + r_j) > \widehat{\mathrm{VaR}}_\alpha(r_i) + \widehat{\mathrm{VaR}}_\alpha(r_j) \\ -1, & \text{if } \widehat{\mathrm{VaR}}_\alpha(r_i + r_j) < \left| \widehat{\mathrm{VaR}}_\alpha(r_i) - \widehat{\mathrm{VaR}}_\alpha(r_j) \right| \\ \frac{\widehat{\mathrm{VaR}}^2_\alpha(r_i + r_j) - \widehat{\mathrm{VaR}}^2_\alpha(r_i) - \widehat{\mathrm{VaR}}^2_\alpha(r_j)}{2\widehat{\mathrm{VaR}}_\alpha(r_i)\widehat{\mathrm{VaR}}_\alpha(r_j)}, & \text{otherwise.} \end{cases} \tag{12}$$

The first condition in (12) arises in the presence of superadditivity, i.e., when subadditivity[26] fails. The second condition could be referred to as "superdiversification," i.e., the seemingly unusual situation where the risks of two individual positions are more than offset by the risk (or, better, "chance") of the combined positions. Only if neither of the two cases apply, will the VaR–implied–correlation estimate be strictly between ± 1. Although superadditivity and superdiversification may be rather unrealistic, the coarseness of extreme–quantile estimates may, in empirical analysis, lead to such pathological situations.[27]

[26] See Artzner et al. (1999) on the VaR–measure's lack of subadditivity.

[27] See Mittnik et al. (2011) on the potential of superadditivity in the context of aggregating operational risk components.

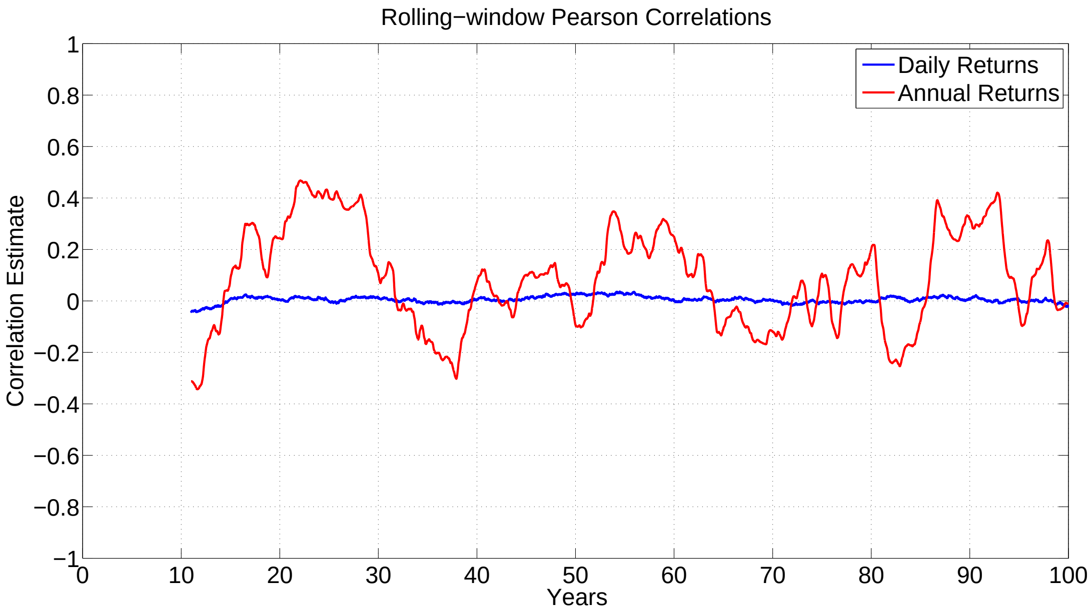


Figure 9: Historical Pearson Correlations VaRs for Daily and Annual Returns

4.2 Annualization and Correlations

4.2.1 Correlations from Simulated Daily and Annualized Returns

In the following, we assess the consequences of rolling–window annualization on correlation estimates. First, we compute the Pearson correlation for the two uncorrelated return series shown in Figure 3. We do this for both daily and annualized return series using, analogous to the VaR calculations in Figure 4, 10–year rolling windows to derive correlation estimates at each day in the 100–year period, starting in year 11.

The results are shown in Figures 9. The Pearson–correlation estimates based on daily data behave as expected. They hover tightly around zero, with a range of ± 0.05. The correlation estimates derived from the one–year rolling–window returns behave very differently. They vary considerably and assume values between –0.4 and +0.5. Given that the two annualized return series are independent, the correlation estimates are remarkably large.

Because Solvency II calibrations of equity–risk components are based on tail correlations rather than just Pearson correlations, we also compute both data–cutting and VaR–implied tail correlations, ρ_α^{DCQ} and ρ_α^{VaR} from the simulated returns. When applying the data–cutting approach and adopting the 99.5% confidence level specified in Solvency II, we run into the problem that—for both daily and annualized returns—there are practically no joint tail observations. In other words, ten years or 2,590 observations are far from sufficient for the southwest quadrant, depicted in

Figure 7, to contain any data, so that no tail correlations, can be computed.

If the data–cutting approach is to be adopted, one can no longer stick to the 99.5% confidence level, as demanded by the EU Directive (European Parliament, 2009). Therefore, in the simulations discussed below, we report results for lower levels. The CEIOPS analysts also experimented with alternative confidence levels[28] Analyzing the dependence between equity and fixed income, confidence levels from 99% down to 80% are considered. It turns out that the 99% confidence level ist still too ambitious to obtain sufficient joint tail observations. Therefore, we compute data–cutting tail correlations for the 95% and 80% confidence levels.[29]

The number of available joint tail observations (top) and the tail–correlation estimates (bottom) for both daily and annualized returns and the 95%–level are shown in Figure 10. For daily returns the number of joint tail observations lies between 3 and 13—sample sizes much too low to obtain reliable estimates. As a consequence, the tail correlation estimates (bottom of Figure 10) range from –1 to +1. The picture looks even bleaker for annualized returns. Although the number joint tail observations can move up to almost 60, it is zero for most of the available 89–year period. As a result, the tail–correlation plot (again, bottom of Figure 10) has large gaps. In the few occasions where we can compute tail correlations, the estimates also range from –1 to +1.

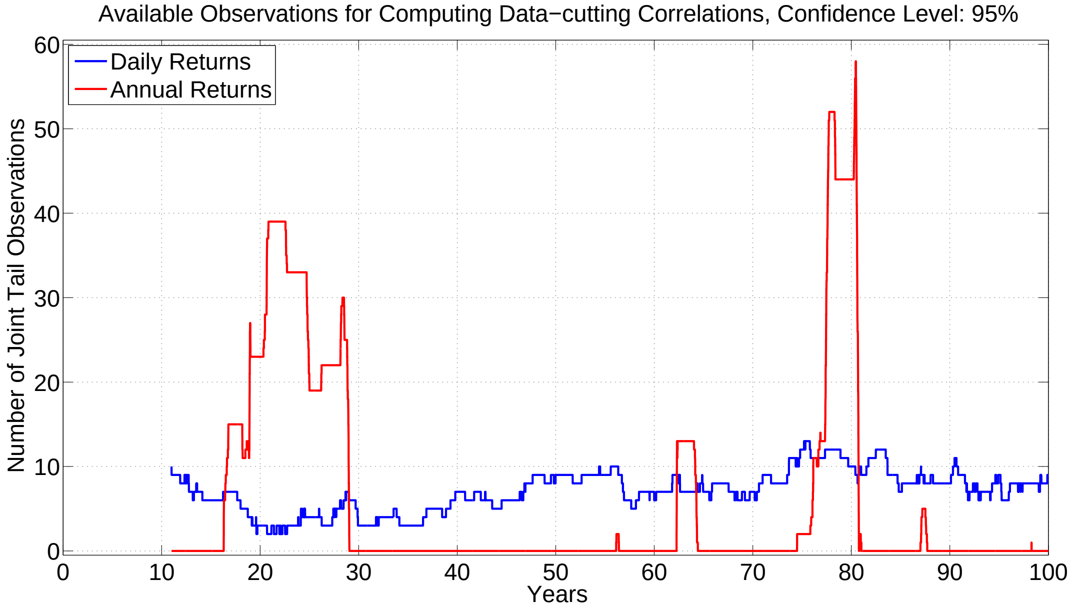
Given these findings, we cannot expect the data–cutting approach to produce prudent correlation parameters that can be used for risk aggregation via the Standard Formula. Relying on annualized returns, the problem will not vanish even when working with much longer than 10–year sample sizes. One option would be to substantially lower the confidence level—although this contradicts the EU Directive, which explicitly prescribes the 99.5% level. But even for the 80%–level, the number of observations can be insufficient to obtain reliable estimates. As Figure 11 indicates, though most of the time there is a reasonable number of joint tail observations, there is no guarantee for this to hold throughout a sample.

More of a concern is the fact that the tail–correlation estimates jump erratically, assuming values between –1 and +0.7. Clearly, with this performance, the DCQ–

[28]See Paragraphs 3.1376–3.1385 in CEIOPS (2010).

[29]The difficulty of deriving tail–correlation estimates using the data–cutting approach is acknowledged in Paragraph 3.1384 in CEIOPS (2010) which states: "... the choice of percentile is important in determining the correct correlation coefficient." In an attempt to define the meaning of "correct," Paragraph 3.1385 continues:

> *It is key to strike a balance between being adequately in the tail, and having enough data points for a reliable analysis. ... [T]he overall correlation matrix should produce a level of stress equivalent to a 99.5% VaR event, so each individual pair can be equivalent to significantly less than a 99.5th percentile stress, but still should be firmly in the tail. The analysis must be subject to sensitivities for different percentiles, and should be taken as providing an indication of the correct correlation.*



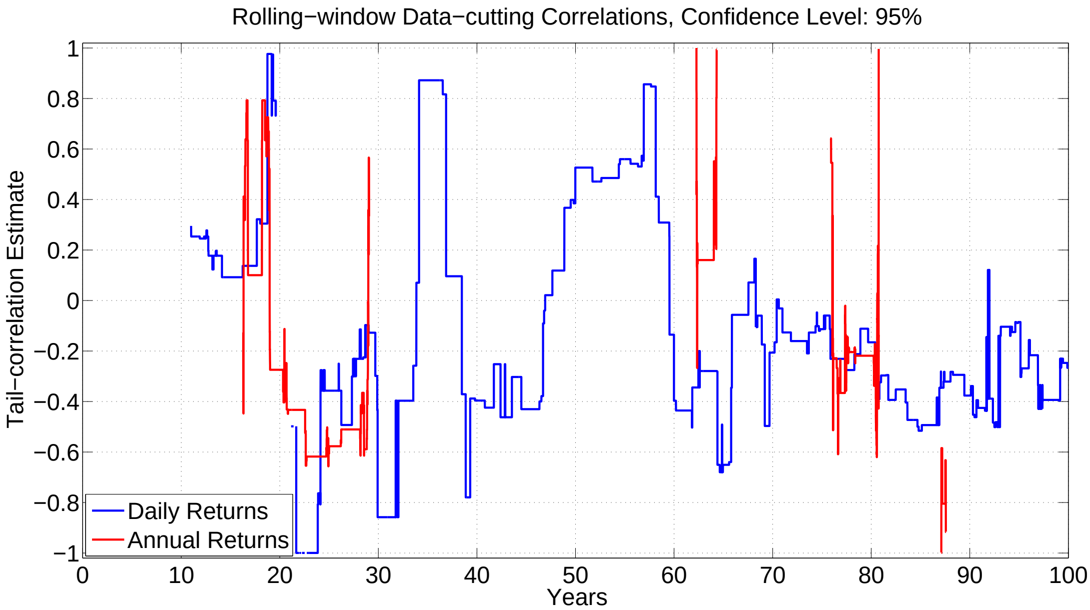


Figure 10: Available number of observations (top) and data–cutting tail–correlation estimates, ρ_{95}^{DCQ}, applied to daily and annualized returns; 95% confidence level and 10–year estimation window

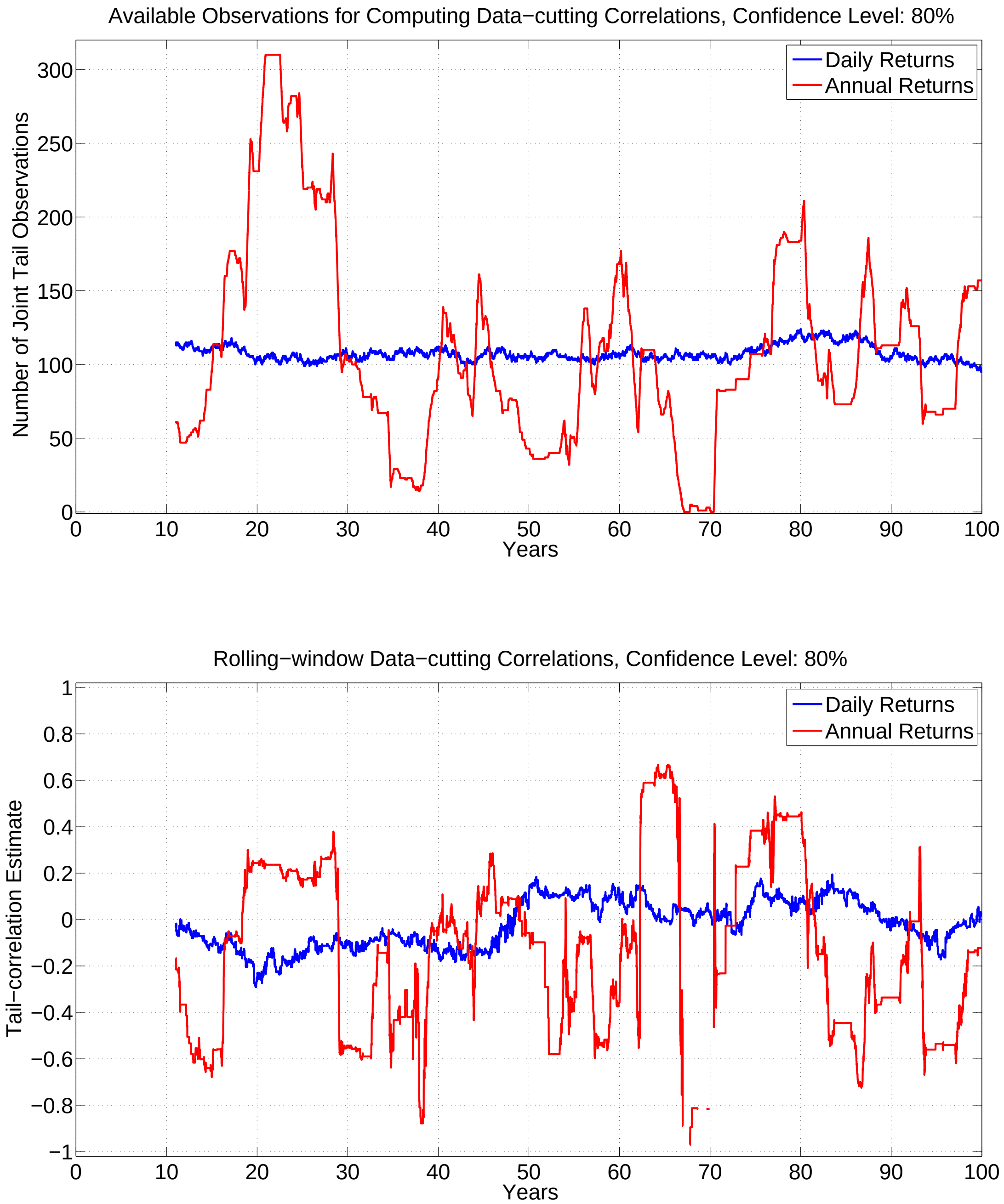


Figure 11: Available number of observations (top) and data–cutting tail–correlation estimates, ρ_{80}^{DCQ}, applied to daily and annualized returns; 80% confidence level and 10–year estimation window

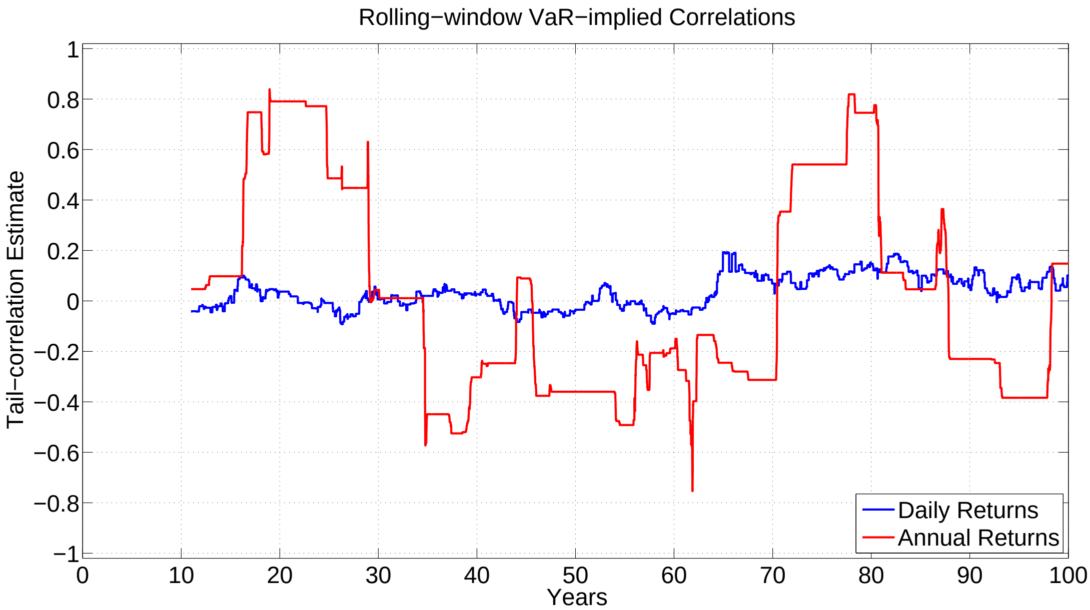


Figure 12: VaR–implied tail correlations for daily and annual returns

correlation approach does not qualify for regulatory purposes, unless much more observations from the center of the distribution are included. Then, however, we can no longer consider the estimates to be *tail* correlations.[30]

We also compute $\text{VaR}_{99.5}$–implied tail correlations from the simulated data (Figure 12). For daily returns, the estimates lie steadily between –0.1 and +0.2. For the annualized–returns, we obtain extremely erratic results. The tail–correlation estimates cover values from about –0.75 to +0.85 and exhibit sudden jumps and sign switches.

To summarize, the simulation results for data–cutting and VaR–implied correlations strongly indicate that overlapping annual rolling–window returns will seriously hamper any calibration effort towards designing prudent regulatory processes.

4.2.2 Bias and Efficiency

In empirical analysis it is commonly desired to employ unbiased and efficient estimators. That is, the estimator should, on average, produce accurate point estimates; and it should do so with little uncertainty, meaning that the interval estimates should be small. In the following we examine how the use of overlapping rolling–window returns affects the unbiasedness and efficiency of correlation estimates. We conduct Monte Carlo analyses to investigate both bias and efficiency of correlation

[30]It should be noted that the problem of insufficient joint tail observations for computing data–cutting correlations may be less dramatic when—as is usually the case—returns are heavy–tailed. We will turn to this issue in Section 4.3 below.

estimates as the window lengths increase, drawing the daily (continuous) returns from a bivariate normal distribution, i.e.,

$$r_t = \begin{pmatrix} r_{1t} \\ r_{2t} \end{pmatrix} \overset{iid}{\sim} \mathrm{N}(\mu, \Sigma), \quad \text{with } \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{13}$$

From (13) we generate 20,000 bivariate time series of length2 $259 \times n$, with $n = 10, 40$, derive rolling–window returns with windows of lengths $w \in \{1, 5, 22, 65, 130, 259\}$, and compute three types of correlations between R^w_{1t} and R^w_{2t}: the standard Pearson correlations based on all data, the half–plane data–cutting correlations, based on the 0,5%–portion of the largest losses, and VaR–implied correlations at the 99.5% level.

First, we generate independent series by setting, in (13), $\rho = 0$. Figure 13 plots the bias for the three correlation estimators as the window length varies. The Pearson correlation is estimated from the whole sample; for the data–cutting correlation and VaR–implied correlation we follow QIS5 specifications and obtain estimates for the 0.5%–quantile (i.e., $\mathrm{VaR}_{99.5}$). Whereas the conventional Pearson and the data–cutting correlations remain unbiased, the VaR–implied correlation estimate exhibits a systematic upward bias as the window length increases. For annual aggregation ($w = 259$) the bias reaches 0.09 for the 40–year sample. This means that even if the returns of two assets are uncorrelated and independent, the VaR–implied correlation estimates will on average produce a value of about 0.09, wrongly suggesting a positive dependence.

Turning to the efficiency of the correlation estimators, Figure 14 reveals that the confidence intervals around of the three estimators behave quite differently. The conventional Pearson correlation has the tightest intervals, but they grow considerably with the length of the aggregation window. Data–cutting correlations exhibit already for small window length extremely large interval spreads, ranging from –0.9 to +0.9. The confidence intervals for the VaR–implied correlations are not much better. They range from –0.5 to +1 for monthly aggregation, and cover the maximum possible range ± 1 for annual aggregation. The extreme range could be due to a couple of extreme outliers. But even the 90%–confidence interval ranges from –0.5 to about +0.8, suggesting that, apart from being biased, VaR–implied correlation estimates from rolling–window returns can be virtually all over the place and provide no information about the underlying data–generating process.

The seriousness of the spurious–correlation problem is evident from the plots in Figure 15. They show how the widths of the confidence intervals grows as the window length increases. Debating whether or nor two particular asset classes have a tail correlation of 0.3 or 0.8 is rather meaningless, given the blatant instability of data– cutting and VaR–implied correlation estimators, when based on overlapping rolling–window returns.

The histograms of the 20,000 VaR–implied correlation estimates are presented in Figure 16. They, too, demonstrate the quick increase of the estimates' dispersion as

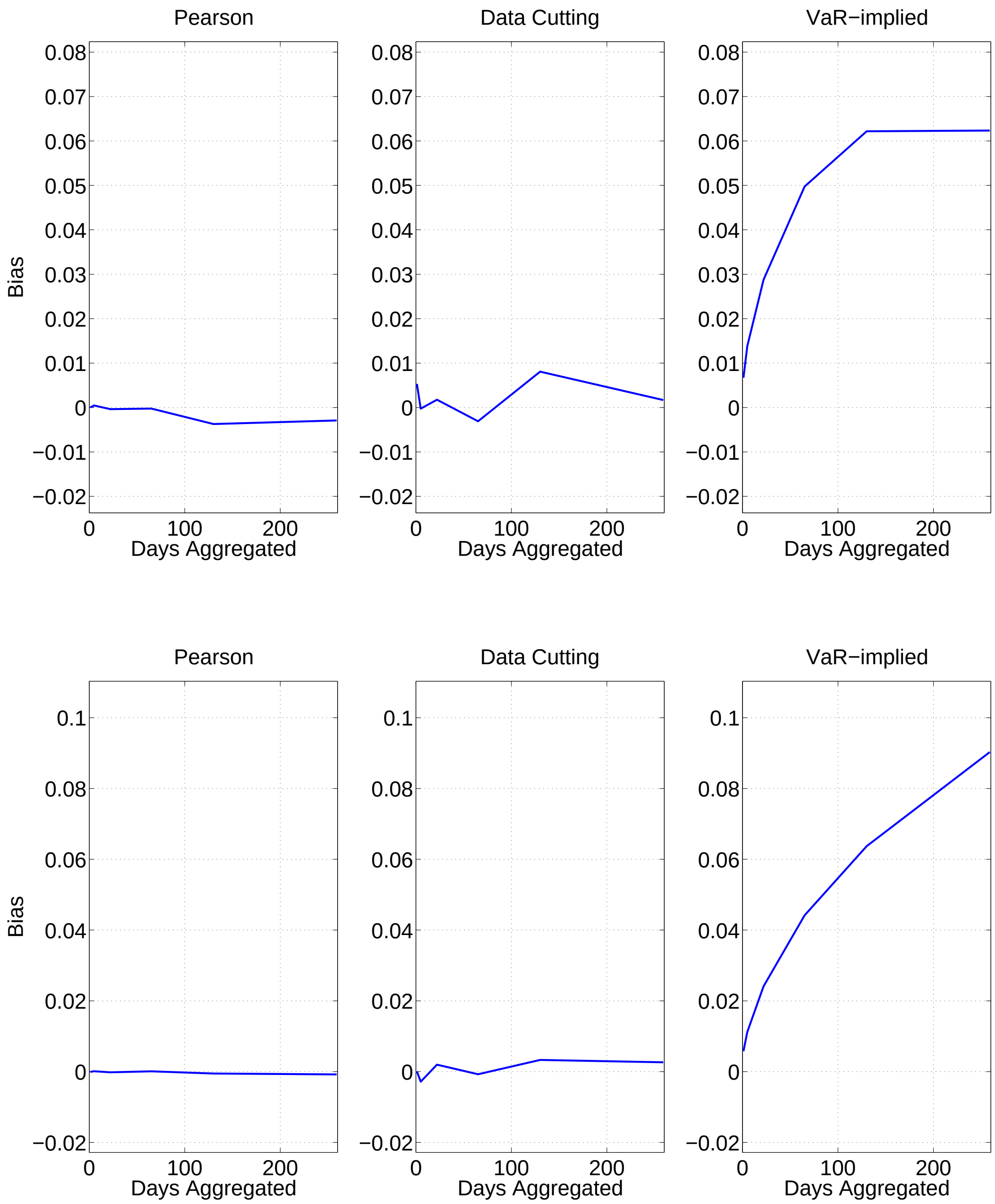


Figure 13: Bias in VaR–implied tail–correlation estimates due to rolling–window aggregation, $\rho = 0$, 10–year (top) and 40–year (bottom) samples

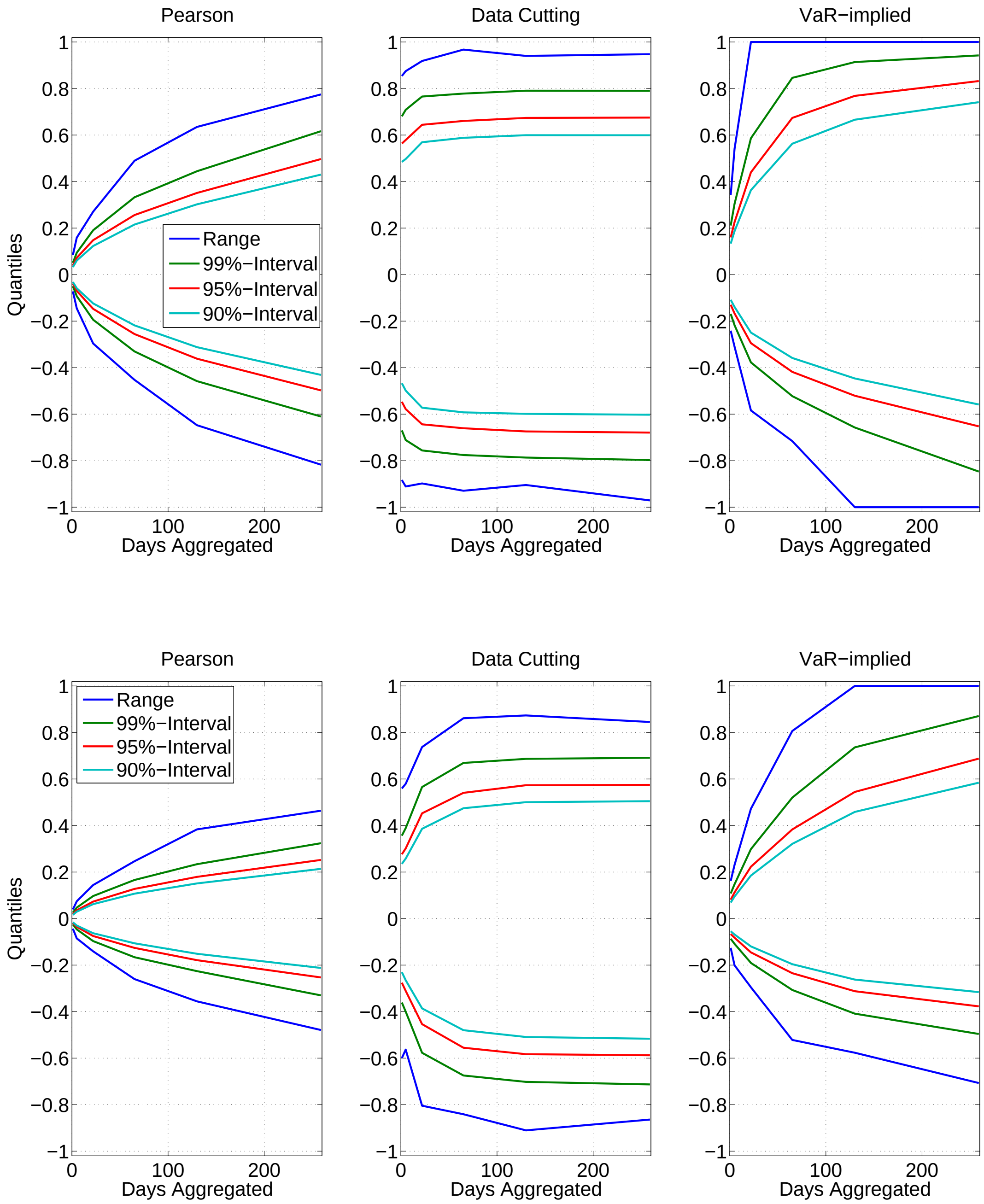


Figure 14: Confidence intervals of correlation estimates and rolling–window aggregation, $\rho = 0$, 10–year (top) and 40–year (bottom) samples

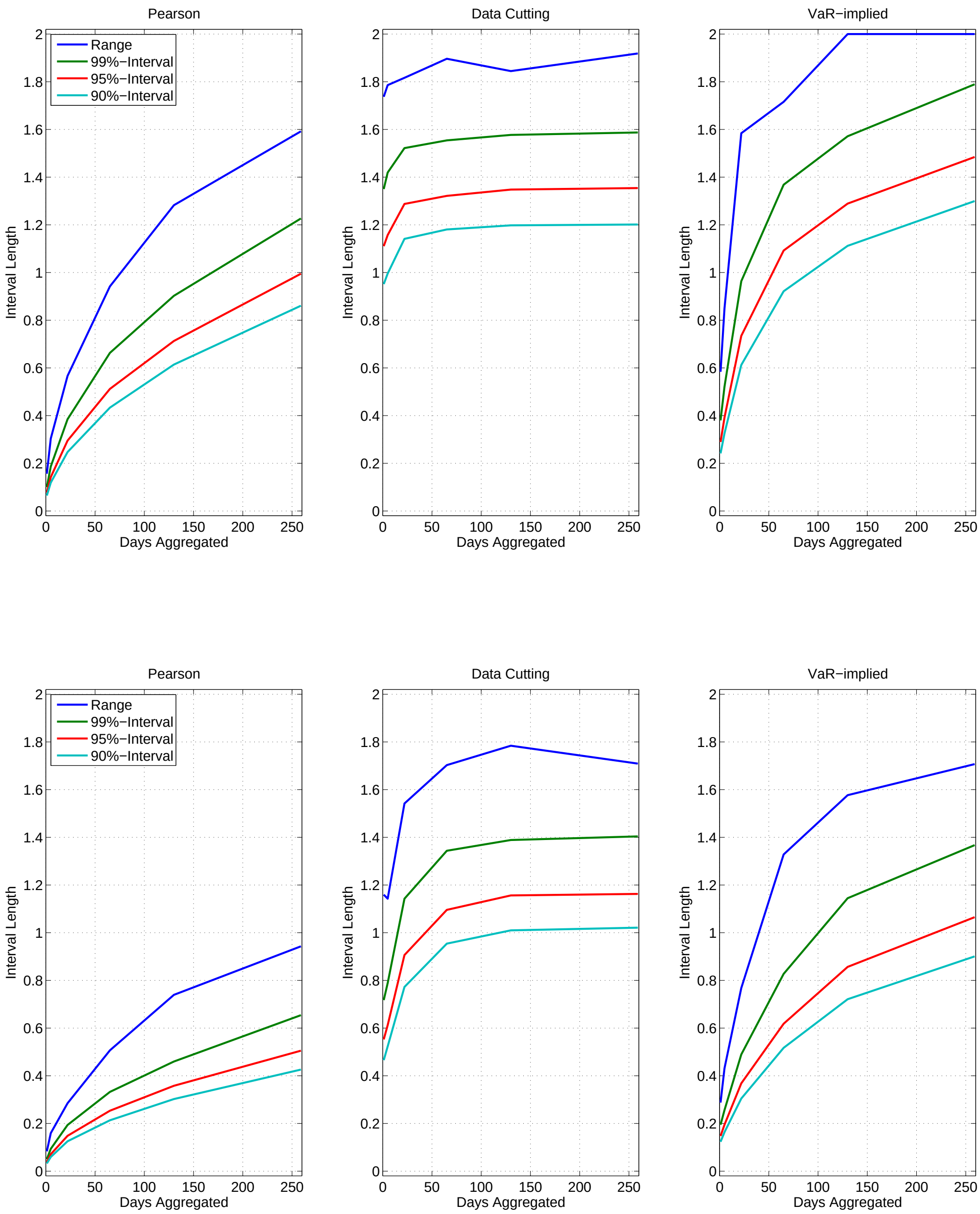


Figure 15: Length of confidence intervals of correlation estimates and rolling–window aggregation, $\rho = 0$, 10–year (top) and 40–year (bottom) samples

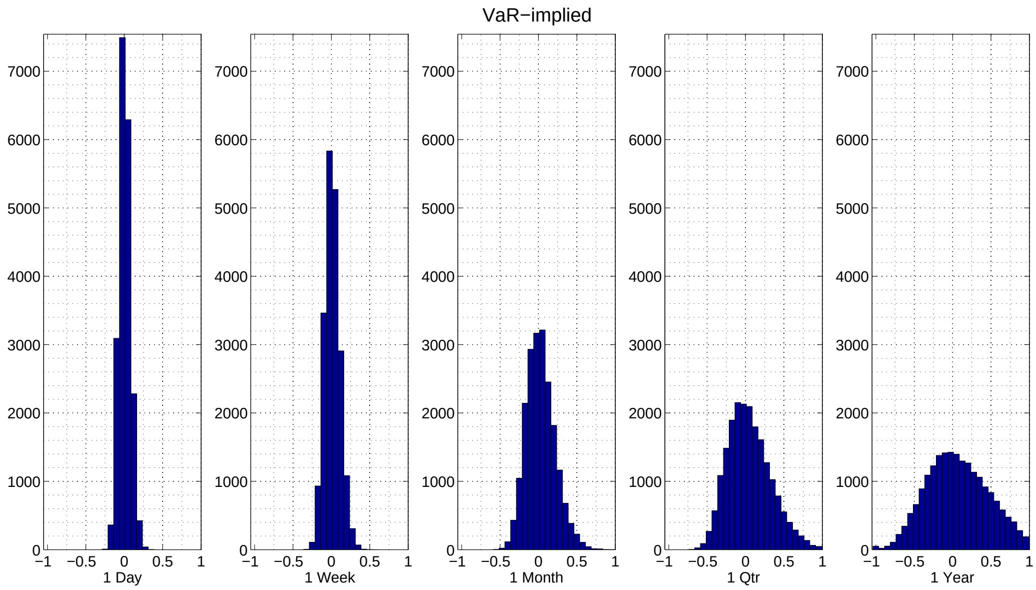


Figure 16: Histogram of VaR-implied tail correlations, $\rho = 0$ and 10–year samples

the aggregation level grows. The modes of the histograms remain more or less at zero. However, as the aggregation length rises, so does the right–skewness, inducing the upward bias in the tail–correlation estimates.

Next, we investigate the performance of the correlation estimators in the presence of nonzero correlations between assets. We repeat the Monte Carlo experiment, but now specify different levels of correlation for the daily returns, namely, $\rho = 0.2, 0.4, 0.6, 0.8$. Figure 17 shows the histograms of the VaR–implied correlations obtained from 20,000 Monte Carlo replications for each of the four correlations specified, assuming ten years of daily data. In each case, the VaR–implied correlation estimates from annualized returns exhibit an upward biased. Even more of a concern is the fact that there is a serious pile–up of estimates near or at $+1$ when daily correlations reach $\rho = 0.4$. For $\rho > 0.4$, the mode of the distribution lies near $+1$, so that, if daily correlations exceed 0.4, there seems to be an excessively high probability that VaR–implied correlation estimates based on annualized returns assume values that are near or exactly $+1$.

The median tail–correlation estimates for the cases $\rho = 0.2, 0.4, 0.6, 0.8$ are 0.2619, 0.4794, 0.6860, and 0.8751, respectively. Thus, if the true correlation is 0.4, we have a 50% probability that the annualized data will produce an estimate above 0.48. Table 2 summarizes selected probabilities for VaR–implied tail–correlation estimates to exceed certain thresholds. For example, if the correlation of the underlying daily data is 0.2, in Solvency II calibration produces with a probability of 25%, a tail–correlation estimate above 0.56 and, with a probability of 10%, above 0.77. If the underlying correlation is 0.6, there is a 50% probability that the estimate will

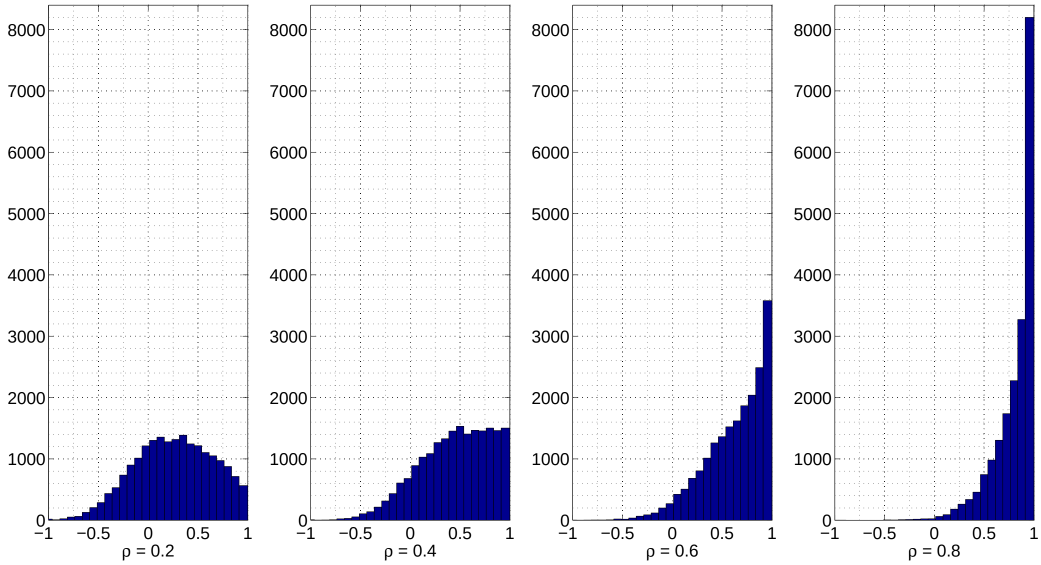


Figure 17: Histogram of VaR-implied tail correlations from annualized returns for differently correlated daily returns

exceed 0.67, and 25% probability it will lie above 0.87. Thus, there is a rather large probability of ending up with grossly overstated tail–correlation estimates.

For $\rho = 0.4$, the simulated confidence intervals, shown in Figure 18, reveal that the upper boundaries of the intervals for the VaR–implied correlation move extremely close to +1—even that of the 90% confidence level. Thus, it is very likely to encounter tail–correlation estimates close to unity, even though the true correlation is 0.4. For $\rho > 0$, the widths of the confidence bands of the VaR–impled estimates become somewhat shorter relative to the uncorrelated case. This, again, is due to the fact that correlation estimates have an upper bound of +1. However, for the annual—and for Solvency II relevant—aggregation level, the range still covers the maximum possible interval [−1, +1].

The simulation experiments reconfirm that, regardless of the level of the underlying correlation, VaR–implied tail–correlation estimates derived from overlapping rolling–window returns behave extremely erratic.

4.3 Heavy Tails

To assess the consequences of moving from a normal distribution to a fat–tailed—but still elliptical—bivariate t–distribution, we repeat the Monte Carlo experiment and generate vectors $r_t = (r_{1t}, r_{2t})'$ from a joint t–distribution with $\nu = 1, 2, 3, 4$

Table 2: Bias in VaR–implied correlation estimates due to rolling–window annualization; sample size 10 years.

The entries represent exceedance probabilities. For example, the entry 0.63 in the last row of Column 2 states that there is 10% probability that the estimated $\text{VaR}_{99.5}$–implied correlation is higher than 0.63, although the correlation of the underlying data is 0.0.

	Daily ρ				
Exceedance Probability	0.0	0.2	0.4	0.6	0.8
50%	0.09	0.26	0.48	0.67	0.88
25%	0.37	0.56	0.74	0.87	0.96
10%	0.63	0.77	0.89	0.96	0.99

degrees of freedom and correlation $\rho = 0$, i.e.,

$$r_t = \begin{pmatrix} r_{1t} \\ r_{2t} \end{pmatrix} \overset{iid}{\sim} t(\mu, \Sigma, \nu), \quad \text{with } \mu = 0,\ \nu = 1, 2, 3, 4 \text{ and } \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{14}$$

Altogether, we performed 20,000 simulation runs, in each of which we generated 40 years of daily data and aggregate again over 5, 22, 65, 130, and 259 trading days.

Figure 19 illustrates that, for heavy–tailed, t–distributed data, the bias of the VaR–implied correlation[31] becomes more severe than in the normal case. The bias in $\hat{\rho}^{VaR}_{99.5}$ increases as the degrees of freedom decrease, that is, as the heavy–tailedness increases. In the case of a one–year rolling window and $\rho = 0$, temporal aggregation with normally distributed daily returns produces a bias of +0.09 (see bottom right plot in Figure 13). When daily returns come from a t–distribution with $\nu = 4$ degrees of freedom, the bias in the VaR–implied correlation for annualized returns rises to 0.12; it increases to 0.15 for $\nu = 3$ and jumps to 0.26 for $\nu = 2$.

Completely different from the above bias pattern is the case of $\nu = 1$, with a bias of –0.06. A t–distribution with $\nu = 1$, which corresponds to a Cauchy distribution, is extremely fat–tailed so that even the mean of the distribution is infinite.[32] The histograms of the VaR–implied correlation estimates for all four degrees–of–freedom values (see Figure 20) show that the pile–up of the estimates at +1 starts with $\nu = 3$ and becomes serious for the infinite variance case of $\nu = 2$. Note that the pile–ups observed above for normally distributed returns occurred for correlations exceeding 0.4. With t–distributed returns, the pile–up happens even for $\rho = 0$ when the degree–of–freedom parameter gets sufficiently small. For $\nu = 1$ extreme pile–ups occur at both +1 and –1, with the remaining estimates being more or less evenly distributed in–between.

[31] In view of the the dismal performance of the data–cutting correlation estimates in the normal case, we focus only on VaR–implied correlations here.

[32] Due to the size of the draws from a bivariate t–distribution with $\nu = 1$, we set both dispersion parameters to 0.1 rather than 1, as was the case in all other simulations. In general, the scaling of the variables should not affect the results. However, we did not investigate this issue.

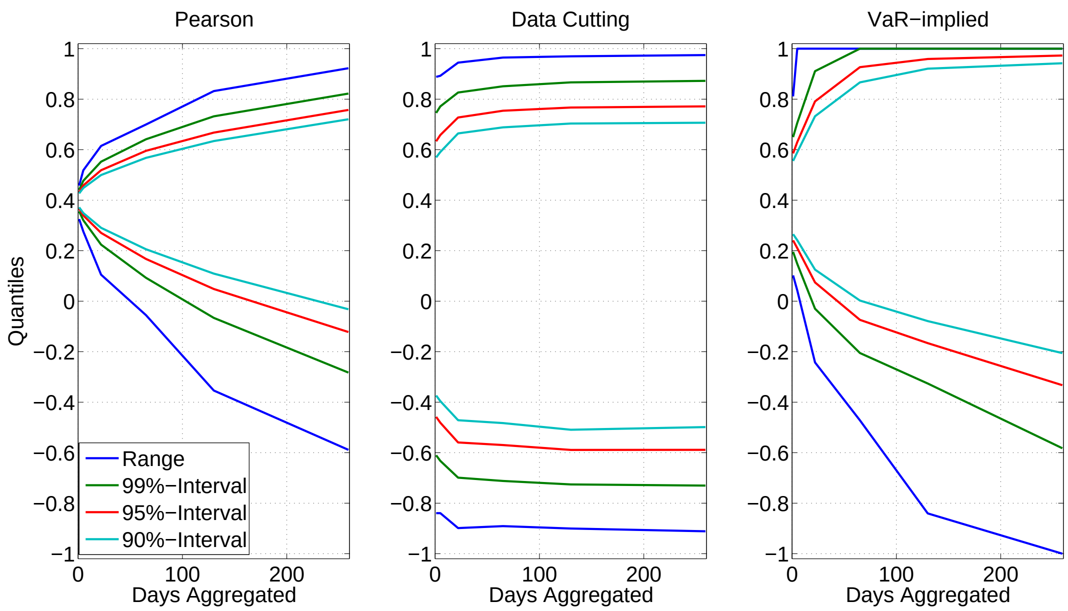


Figure 18: Confidence intervals and temporal aggregation, $\rho = 0.4$; sample size 10 years

The pile–up problem of the VaR–implied correlation estimates indicates that this dependence measure is unsuitable for the situation at hand. The seriousness of the problem for $\nu = 2$ and $\nu = 1$ could largely be due to the nonexistence of variances (for $\nu = 2$) or the lack of finite means and variances (for $\nu = 1$). However, if that was the case, the VaR–implied correlation estimator should already break down when applied to daily returns, not subjected to any temporal aggregation. The histograms in Figure 21 illustrate the behavior of the estimates as the aggregation window increases. For non–aggregated, daily data, the estimator produces a somewhat dispersed but "reasonable" histogram without any pile–ups. The pile–up problem arises already at weekly aggregation and worsens dramatically for higher aggregation levels with pile–ups occurring at both −1 and +1.

The simulations demonstrate that the performance of the VaR–implied correlation estimator applied to annualized returns becomes even worse when the underlying returns are fat–tailed, but still, as is required for the estimator, elliptical, distributed.

4.4 Tail Dependence

The nature of the comovements of risk factors is essential when assessing diversification benefits. If the focus is on extreme risks, we have to be interested in tail dependence, which can be measured by the *coefficient of tail dependence*, denoted here by λ. Let r_i and r_j be the returns of two risk components with marginal dis-

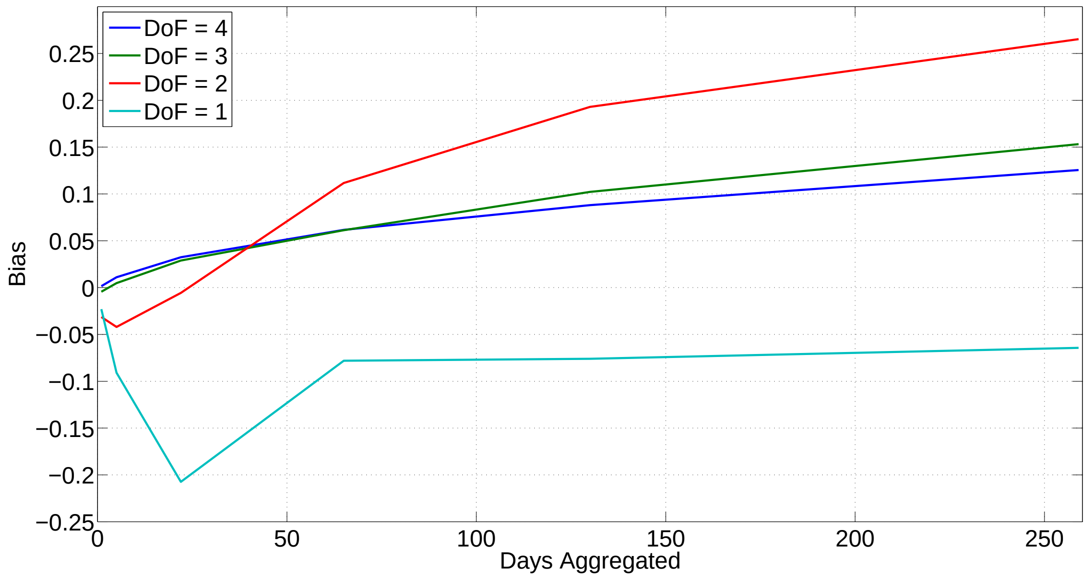


Figure 19: Bias due to rolling–window aggregation for t–distributed daily returns with different degrees of freedom and $\rho = 0.0$

tributions F_i and F_j, respectively. Then, the coefficient of lower tail dependence is defined by[33]

$$\lambda = \lim_{q \to 0} \mathrm{P}\left(r_i \leq F_i^{-1}(q) \,|\, r_j \leq F_j^{-1}(q)\right) \quad \in \quad [0,1]. \tag{15}$$

If large losses in asset i tend to coincide with large losses in asset j, the coefficient of (lower) tail dependence will be close to 1; if there is no such joint tendency, it will be close to 0. Thus, the coefficient of tail dependence conveys important information when, as in Solvency II calibrations, assessing the consequences of extreme losses.

To investigate the implications of rolling–window annualization of daily returns on the joint tail behavior, we simulate 5,000 bivariate time series of lengths 40 and 4,000 years, respectively, with daily returns drawn from the bivariate t–distribution (14), with $\rho = 0.5$ and $\nu = 4$. The estimated coefficients of tail dependence for quantiles $1-\alpha$ range from 0.001% to 2.5%. The results are shown in Figure 22. The horizontal line in these plots indicates the theoretical value[34] of the tail dependence coefficient

[33] In the case of asymmetric distributions, we distinguish between upper and lower tail dependence. The coefficient of upper tail dependence is defined by simply reverting the inequalities in (15). Below, we will only focus on lower tail dependence and use λ to denote the coefficient of *lower* tail dependence.

[34] The analytic expression for the coefficient of tail dependence is given by

$$\lambda^* = 2F_{t_{\nu+1}}\left(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right),$$

where $F_{t_{\nu+1}}(\cdot)$ denotes the cumulative distribution function of the standard t–distribution with $\nu+1$ degrees of freedom; see Embrechts et al. (2002).

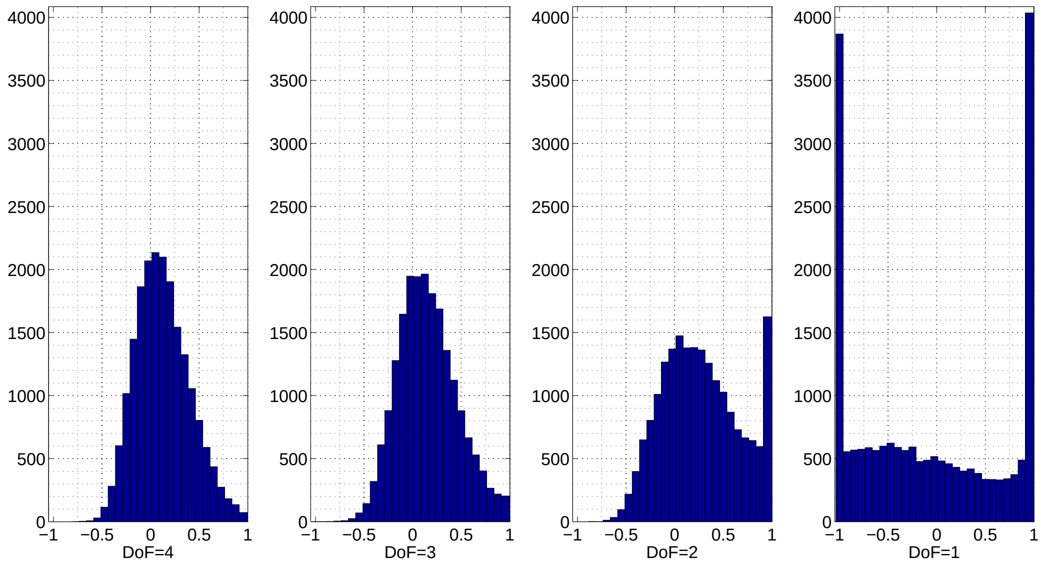


Figure 20: Histograms of VaR–implied tail–correlations estimates for t–distributed daily returns with different degrees of freedom and $\rho = 0.0$

for the bivariate t–distribution with $\nu = 4$ and $\rho = 0.5$, namely $\lambda^* = 0.2532$; and the dashed lines indicate the (bootstrapped) 95%–confidence bands.

The estimates from the daily data slightly overestimate the theoretical value of λ^*, but they approach it very closely the further we move into the tail. The λ–estimates from the annualized data behave very differently. Throughout the range, they underestimate the theoretical value and approach zero the further we get into the tails, suggesting absence of tail dependence. The confidence band for the 40–year samples turns out to be extremely wide. Throughout the tail area considered, the band includes zero, suggesting that the absence of tail dependence is not rejected. The upper limit of the band hovers mostly around 0.6—except for the extreme tail area, i.e., $q \leq 0.1\%$, when the upper limit of the confidence band quickly drops to zero, implying the certain absence of tail dependence. The reason for both the point estimates and the confidence intervals collapsing to zero is that rolling–window annualization annualization not only scrambles *linear* dependence structures between the assets, but also annihilates their joint tail behavior, so that there are virtually no common tail observations left for a sample size of "only" 40 years.

The scatter plots for a typical realization of daily and annualized returns from a 40–year sample in Figure 23 illustrate this phenomenon. For the simulated daily returns (left graph) we observe, in accordance with the quantile considered, a fairly large number of common tail observations and that there is ellipticity. For the annualized data (right graph), although visual inspection suggest some form of negative dependence, both the ellipticity and the common tail behavior disappear. For both

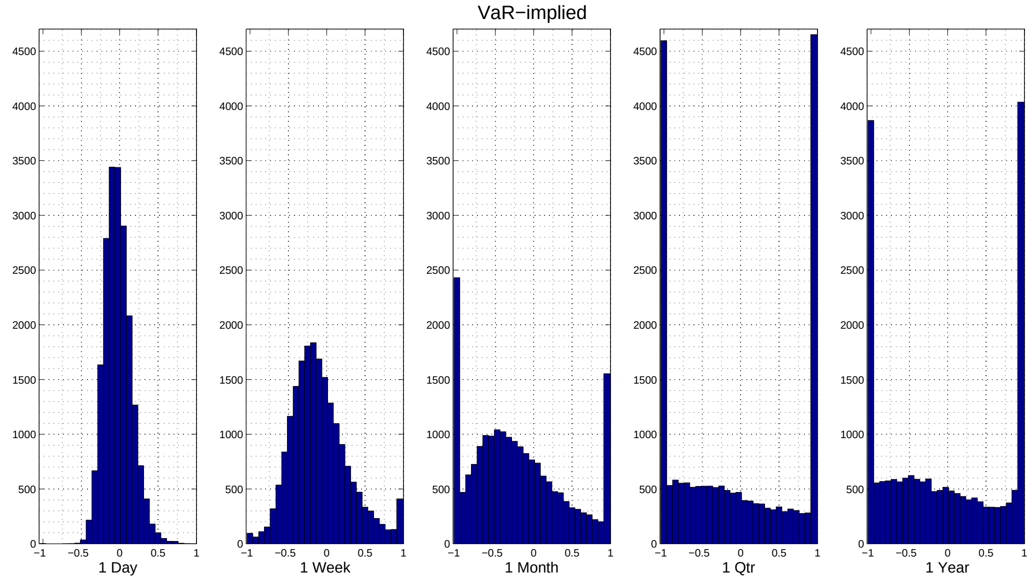


Figure 21: Histograms of VaR–implied tail–correlation estimates with growing aggregation windows, w, for t–distributed daily returns with $\nu = 1$ degrees of freedom and $\rho = 0.0$

risk factors, the annualized returns exhibit maximum losses to about –45%, however, there are no observations in the joint tail region $\{R_1, R_2 \colon R_1 \leq -30\%, R_2 \leq -30\%\}$.

The bias for annualized returns remains even when having 4,000 years of data. The λ–curve shift slightly upward, but stays well below the theoretical value of of 0.2532—especially, in the far–tail with an estimate under 0.05. The confidence band narrows substantially and includes zero only in the far tail.

In view of the results of the simulation experiment. It is evident that the rolling–window annualization more or less wipes out any tail dependence that is present in the original data. Therefore, the hope to more adequately capture the dependence between non–normally distributed asset classes by estimating tail–dependence coefficients, as expressed in Paragraphs 3.1255 and 3.1256 in CEIOPS (2010), will not be fulfilled when the analysis is based on data that have been subjected to a rolling–window annualization.

5 Conclusions

Given the significant role the insurance industry has in its own right as well as its relevance for both the financial and the real sector of the economy, prudent risk–assessment processes, ensuring insurers' solvency, are of paramount importance. As EIOPA (2011) (p. 5) states:

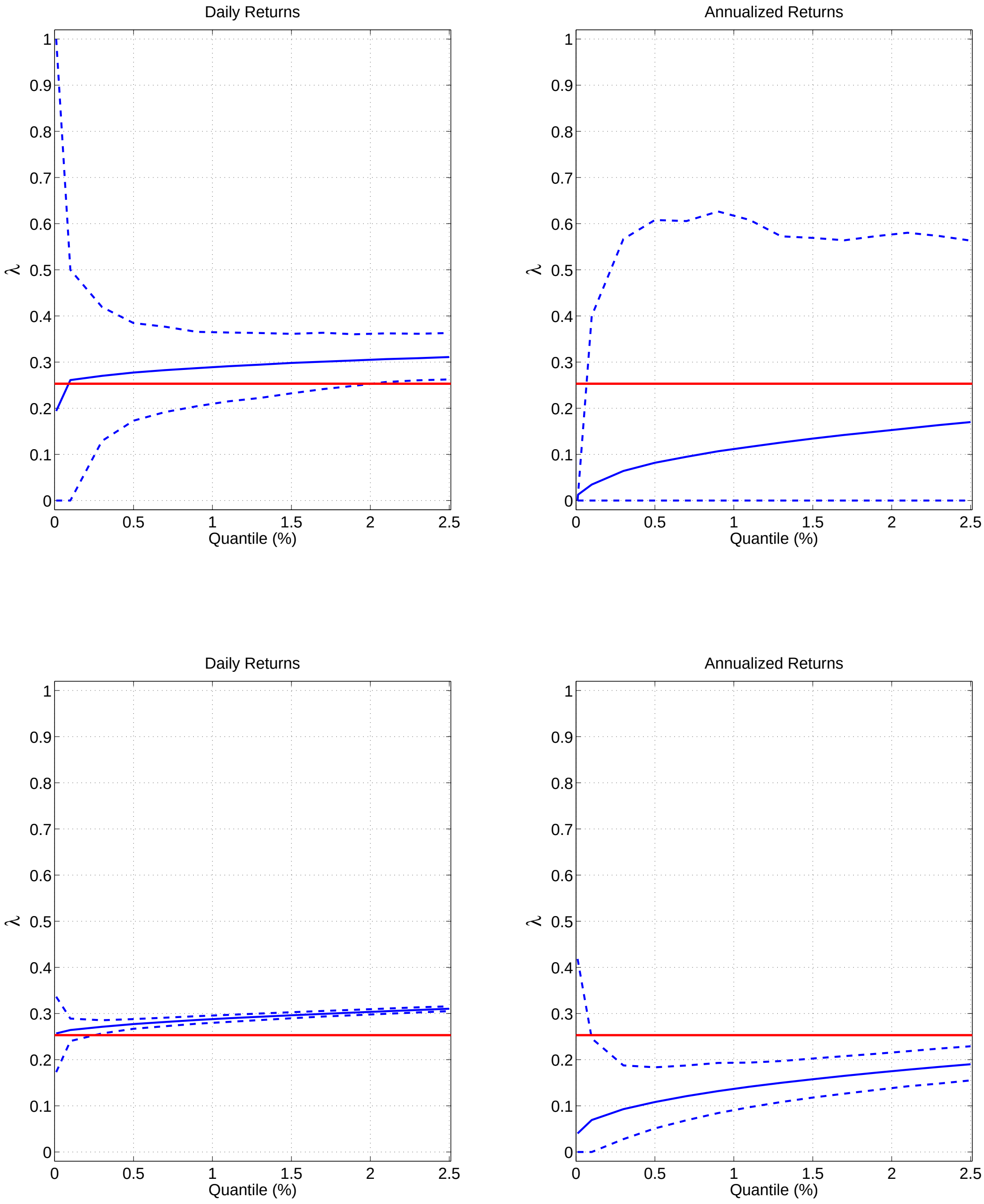


Figure 22: Mean of estimated tail–dependence coefficients (solid curves) from daily (left) and annualized (right) returns generated from bivariate t–distribution ($\rho = 0.5$ and $\nu = 4$) from a 40–year sample (top) and a 4,000–year sample (bottom) with 95% confidence bands (dashed)

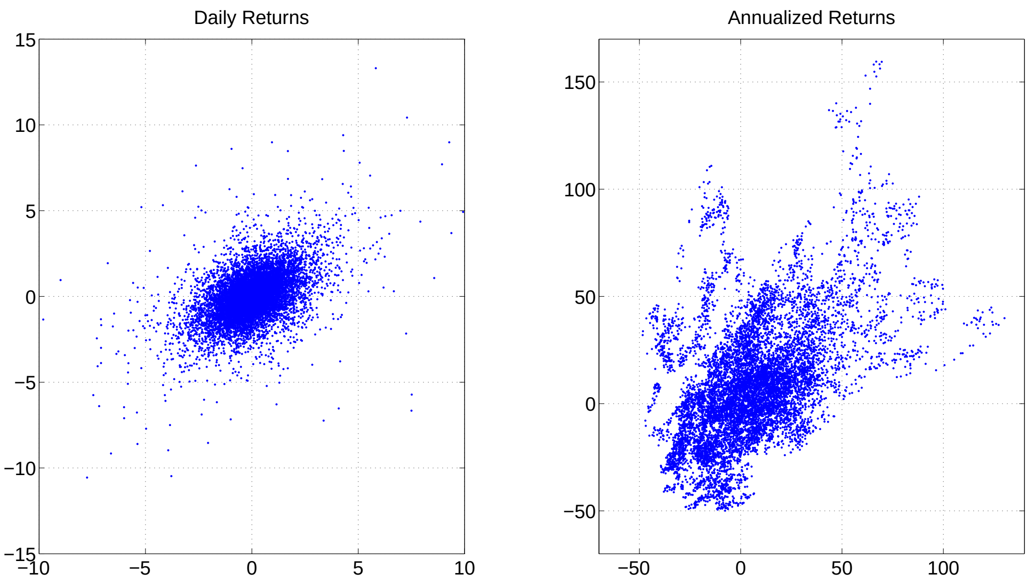


Figure 23: Scatter plots of daily (left) returns generated from bivariate t–distribution ($\rho = 0.5$ and $\nu = 4$) and annualized (right) returns from a 40–year sample

> *QIS exercises are crucial to the development of EU regulation. … [They] are essential to strive to ensure that Solvency II is designed in the most appropriate manner …*

CEIOPS has set up systematically structured procedures for measuring and aggregating the risk components faced by insurance companies. Clearly, designing a regulatory framework of this complexity is a lengthy, if not never–ending process, and the implementation cannot wait until the "most appropriate" design has been achieved. The question, however, are: Does the Solvency II framework, as currently proposed, represent an overall improvement towards a prudential regulation of the insurance industry? Or are there parts or modules whose implementation would be premature?

Criticism against Solvency II calibrations has been raised before, arguing, for example, that indices chosen to represent particular equity classes are inappropriate (Aria et al., 2010), or that the Standard Formula is unstable with respect to distributional settings (Pfeifer and Strassburger, 2006). The problems detailed here are, however, more fundamental. They strongly suggest that, by subjecting historical market data to a particular annualization procedure prior to performing the calibration exercises, much of the QIS equity–risk calibrations may be rendered meaningless.

Specifically, an analysis of the consequences for assessing equity–risk (generally, the most significant risk driver for insurers), when daily return data are annualized via overlapping rolling–window annualization, shows that this transformation has a number of harmful implications. The main findings reported here can be summarized as follows:

1. Altogether, rolling–window annualization leads to highly unreliable and erratic VaR and correlation estimates, which are the sole input parameters of the Standard Formula. As a result, both equity–specific as well as aggregate capital–requirement estimates derived with the Standard Formula may be misleading indicators of true risk exposures.

2. The annualization leads to strong temporal return and risk dependencies in the data, as it induces near–unit–root characteristics, which are responsible for SCR estimates being highly unstable over time and assuming more or less arbitrary values.

3. The annualization also produces arbitrary contemporaneous dependence structures between asset classes. This affects the conventional Pearson correlation and, more so, the data–cutting and VaR–implied tail–correlations favored in the QIS calibrations.

4. A disturbing result is that, if the original data are weakly positively correlation, tail–correlation estimates from annualized data may often be at or near +1 and, thus, greatly exaggerate the presence of dependencies. This pile–up problem may very well be the reason that QIS calibrations specify perfect positive correlation among "other equities."

5. On the other hand, the annualization wipes out any tail dependence that may be present in the original data. This problem results from the fact that rolling–window annualization destroys any (near-)ellipticity in the data, a property the Standard Formula requires.

The argument that tail correlation is a more appropriate dependence measure than conventional Pearson correlation is appealing, given that asset returns often exhibit asymmetries. The assumption of asymmetry contradicts, however, the use of the Standard Formula, which is only valid for elliptical and, thus, symmetric return distributions. If, on the other hand, we assume symmetry, there is no point in using downside–risk and downside–dependence measures, such as VaR and lower–tail correlation.

As they stand, QIS5 equity–risk calibrations fall far short of the goals EU legislators strive for. Their application may do more harm than good. Setting, for example, the correlations among all "other–equity" types—comprised of diverse asset classes, such as private equity, commodities, hedge fonds, and emerging–market stocks—equal to +1 generates a serious disincentive to risk diversification. This may severely affect investment activities that are essential for the development of emerging and, in the case of private equity, developed economies.

In view of these findings, the implementation of the Solvency II framework with the currently proposed equity–risk calibrations seems far from prudent, if not irresponsible.

Calibration results involving data subjected to rolling–window annualization need to be reexamined using non–overlapping—weekly or daily—return data, as they produce more reliable parameter estimates. To derive annualized SCRs, the calibration needs to capture temporal dependence structures as well, so that risk can be aggregated over time. Clearly, the latter will be an additional, nontrivial task requiring serious efforts. But given the stakes involved, the necessary resources appear negligible in comparison.

References

Aria, L., El Hedi Arouri, M., Foulquier, P., Gregoir, S., 2010. On the Suitability of the Calibration of Private Equity Risk in the Solvency II Standard Formula. Working Paper, EDHEC Business School.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent Measures of Risk. Mathematical Finance 9, 203–228.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307–327.

CEIOPS, 2010. QIS5 Calibration Paper. CEIOPS-SEC-40-10, Committee of the European Insurance and Occupational Pension Supervisors, 15 April 2010.

Dickey, D., Fuller, W., 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association 74, 427–431.

EIOPA, 2011. EIOPA Report on the fifth Quantitative Impact Study (QIS5) for Solvency II. EIOPA-TFQIS5-11/001, 14 March 2011, European Insurance and Occupational Pensions Authority.

Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M. (Ed.), Risk Management: Value at Risk and Beyond. Cambridge University Press.

Engle, R., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. Econometrica 50, 987–1008.

European Commission, July 2007a. Proposal for a Directive of the European Parliment and of the the Council on the taking–up and pursuit of the business of Insurance and Reinsurance – Solvency II. 2007/0143(COD), Commission of the European Communities.

European Commission, 2007b. 'Solvency II': Frequently Asked Quations (FAQs). Memo/07/286, 10 July 2007, Commission of the European Communities.

European Parliament, 2009. Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking–up and pursuit of the business of Insurance and Reinsurance (Solvency II). Official Journal of the European Union, The European Parliament and the Council of the European Union.

Granger, C., Newbold, P., 1974. Spurious Regressions in Econometrics. Journal of Econometrics 2, 111–120.

Granger, C., Sin, C.-Y., 2000. Modelling the Absolute Returns of Different Stock Indices: Exploring the Forecastability of an Alternative Measure of Risk. Journal of Forecasting 19, 166–187.

Longin, F., Solnik, B., 2001. Extreme Correlation of International Equity Markets. Journal of Finance 56, 649–676.

Mittnik, S., Paterlini, S., Yener, T., 2011. Operational–risk Dependencies and the Determination of Risk Capital. Working Paper, Center for Quantitative Risk Analysis, University of Munich.

Pfeifer, D., Strassburger, D., 2006. Solvency II: Stability problems with the SCR aggregation formula. Scandinavian Actuarial Journal 1, 61–77.

Phillips, P., 1987. Time Series Regressions with a Unit Root. Econometrica 55, 277–301.

Voigt, H., 2005. Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences. Sage Publication, Thousand Oaks.

6 Appendices

Appendix A: Continuous versus Discrete Returns

There are two approaches to calculating returns on financial assets. Practitioners commonly use *discrete* returns, whereas empirical analysts and researchers typically resort to *continuous* returns. The former reflect the true, relative price change, and is used when calculating the return on an investment or measuring the performance of an asset. The latter represent an approximation, which is convenient for empirical or analytical investigations as they can be additively rather than mutiplicatively aggregated over time.

Let P_t and P_{t-1} denote the price of an asset at the end of period t and $t-1$, respectively. Then, the discrete return over the period $(t-1, t]$, denoted by R_t is given by[35]

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1; \tag{16}$$

and the continuous return, denoted by r_t, by

$$r_t = \log P_t - \log P_{t-1} = \log \left(\frac{P_t}{P_{t-1}} \right). \tag{17}$$

If price changes, $P_t - P_{t-1}$, are small, then, discrete returns can be approximated by their continuous counterpart, i.e.,

$$r_t = \log(1 + R_t) \approx R_t, \tag{18}$$

with the approximation following from the fact that, for small x, $\log(1+x) \approx x$. Note that, if continuous returns are normally distributed, gross discrete returns, i.e., $1 + R_t$, are lognormally distributed.

The assumption of small price changes is not unreasonable, when dealing with returns over short holding periods, such as a day or a week. For longer horizons, such as the one–year holding period assumed for Solvency II regulation, approximation (18) can be poor, so that discrete returns should be used for empirical analysis. All simulation results reported here are based on discrete returns. However, due to better tractability, analytical results, involving daily return, rely on continuous returns. The proximity of simulated and analytically derived results, when available, indicates the appropriateness of the approximation.

Continuous and discrete multi–period returns over, say, $w > 0$ periods, given by $r_t^w = \sum_{i=0}^{w-1} r_{t-i}$ and $R_t^w = \prod_{i=0}^{w-1}(1 + R_{t-i}) - 1$, respectively, are related via

$$\frac{P_t}{P_{t-w}} = 1 + R_t^w = \prod_{i=0}^{w-1}(1 + R_{t-i}) = \prod_{i=0}^{w-1} \exp\{r_{t-i}\} = \exp\left\{ \sum_{i=0}^{w-1} r_{t-i} \right\} = \exp\{r_t^w\}. \tag{19}$$

All Monte Carlo simulations reported here are based on discrete returns, which we obtain by drawing continuous daily returns, r_t, from a normal or Student–t distribution (at one occasion “enriched” with GARCH dynamics) and computing multi–period, rolling–wind returns via (19).

Appendix B: Multi–period Rolling–window Returns and Near–unit Roots

Continuous rolling–window returns over horizon w are given by

$$r_t^w = \sum_{i=0}^{w-1} r_{t-i}, \quad w \geq 1, \quad t = 1, 2, \ldots . \tag{20}$$

[35] We abstract here from possible adjustments that arise from dividend payments, splits or other measures.

If daily returns, r_t, are white noise, i.e., $r_t \overset{iid}{\sim} (0, \sigma^2)$, (20) corresponds to a moving–average process of order $w-1$. This process is, in fact, a stationary process[36] and not a nonstationary unit–root process. However, as w increases, the process approaches a nonstationary unit–root process. Process (20) can also be rewritten as

$$r_t^w = r_{t-1}^w + r_t - r_{t-w}. \tag{21}$$

This amounts to a special autoregressive moving–average process with orders 1 and w, but it is only the term r_{t-w} on the right–hand side that distinguishes it from a random walk. As w increases, the influence of r_{t-w} on the variation of r_t^w diminishes, because

$$\frac{\mathrm{Cov}(r_t^w, r_{t-w})}{\mathrm{Var}(r_t^w)} = \frac{1}{w}\ . \tag{22}$$

To demonstrate, as stated in (6), that the ordinary least–squares (OLS) estimator, $\hat{a}_T$, for a in autoregression $r_t^w = a r_{t-1}^w + v_t$, given by

$$\hat{a}_T = \frac{\sum_{t=1}^T r_t^w r_{t-1}^w}{\sum_{t=1}^T \left(r_{t-1}^w\right)^2},$$

approaches $(w-1)/w$ as the sample size, T, grows, we show that $\mathrm{plim}_{T\to\infty}\ \hat{a}_T = (w-1)/w$. Assuming that the one–period returns are white noise, i.e., $r_t \overset{iid}{\sim} (0, \sigma^2)$, we obtain for the numerator and denominator

$$\begin{aligned} \mathrm{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^T r_t^w r_{t-1}^w &= \mathrm{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\left[\left(\sum_{i=0}^{w-1} r_{t-i}\right)\left(\sum_{i=0}^{w-1} r_{t-1-i}\right)\right] \\ &= (w-1)\sigma^2 \end{aligned}$$

and

$$\begin{aligned} \mathrm{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^T \left(r_{t-1}^w\right)^2 &= \mathrm{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\left(\sum_{i=0}^{w-1} r_{t-1-i}\right)^2 \\ &= w\sigma^2, \end{aligned}$$

respectively, so that (6) follows.

Given that the root of a first–order autoregressive process is the reciprocal value of the autoregressive coefficient, i.e., $w/(w-1)$, a rolling–window return series approaches a unit–root process as the window length increases.

[36] We have $\mathrm{E}(r_t^w) = 0$ and $\mathrm{Cov}(r_t^w, r_{t-k}^w) = (w-k)\sigma^2$, for $k = 0, 1, \ldots, w-1$, and $\mathrm{Cov}(r_t^w, r_{t-k}^w) = 0$, for $k \geq w$.