 doi:10.20944/preprints201607.0021.v1

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 München

This version: June 2016

Abstract

The Solvency II regulatory framework specifies procedures and parameters for determining solvency capital requirements (SCRs) for insurance companies. The proposed standard SCR calculations involve two steps. The Value–at–Risk (VaR) of each risk driver is measured and, in a second step, all components are aggregated to the company's overall SCR, using the *Standard Formula*. This formula has two inputs: the VaRs of the individual risk drivers 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.

This paper demonstrates that the parameter calibration for the equity–risk module—overall, with about 25%, the most significant risk component—is seriously flawed, giving rise to spurious and highly erratic parameter values. As a consequence, an implementation of the Standard Formula with the currently proposed calibration settings for equity–risk is likely to produce inaccurate, biased and, over time, highly erratic capital requirements.

*I would like to thank Christian Groll, Elmar Helten, Sandra Paterlini, Joseph Reiter, participants at the Financial Econometrics Seminar at Ludwig–Maximilians–Universität, the Quantitative Finance Seminar at Stony Brook University, the Economics Research Seminar at WU Vienna University of Economics and Business, and the 2011 CEQURA Conference on Advances in Financial and Insurance Risk Management in Munich for constructive comments. They are, however, not responsible for any errors or omissions. Support from the *Society for Financial and Insurance Econometrics* for this research project is gratefully acknowledged.

Address for correspondence: Center for Quantitative Risk Analysis, Ludwig–Maximilians–Universität, Akademiestrasse 1/I, 80799 Munich, Germany; Tel.: +49 (0)89 2180-3224; Email: finmetrics@stat.uni-muenchen.de

 doi:10.20944/preprints201607.0021.v1

1 Introduction

In 2007, the 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 policy-holders (consumers, businesses) and the stability of the financial system as a whole.*

To achieve this, the Solvency II Directive (see European Parliament (2009)) aims at linking regulatory and economic capital more closely and improving risk management practices. In addition to pure insurance risks, Solvency II also includes Solvency Capital Requirements for market, credit and operational risks. The EU Directive specifies in detail the kind of losses the capital requirements should be able to absorb:[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, occurring with a probability of 0.5% or less in a one–year period. This definition equates the SCR directly to the Value–at–Risk (VaR) risk measure at the 99.5% confidence level and a one–year holding period. Moreover, the Directive requires that SCR calculations take risk reduction due to diversification effects into account.

To determine its SCR, an insurer can use the *Standard Formula* with the parameters provided by the regulator, use its own internal model, or use a combination of the two. The Standard Formula has a modular structure and is to be applied in a stepwise, bottom–up fashion. First, capital charges are derived for each risk (sub–)module and then, level by level, aggregated to the overall SCR, with correlations entering the calculations to allow for diversification effects among the risk components. At the top level, the main risk modules are:

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

[1]Paragraph 1, European Commission (2007b).

[2]Paragraph 65, European Parliament (2009)

The *Basic Solvency Capital Requirement* (BSCR) comprises these five main modules. 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 and is given by the 99.5% VaR associated with that module; and $\rho_{i,j}$ denotes the correlation between the risk modules i and j. If a module consists of several submodules, they are aggregated analogous to the Standard Formula.

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),[3] the market–risk module—with a weight of more than 60% of overall SCR—is the most important module. It consists of several submodules, of which equity risk is the largest.[4] It makes up about 40% of market risk and, thus, contributes about 25% to the overall SCR.[5] To compute the standard capital charge, equities are divided into "global equities," defined as equities on exchanges listed in countries belonging to the European Economic Area (EEA) or the OECD, and "other equities," which include

- equities listed in countries not belonging to the EEA or OECD
- non-listed and private equities
- hedge funds
- commodities
- other alternative investments

In the analysis below, we focus exclusively on equity risk. However, it is to be expected that the findings also apply to most other submodules within the market–risk module—specifically, currency risk, property risk, spread risk, and concentration risk—as their calibration appears to suffer from the same critical annualization procedure discussed below.[6]

The Standard Formula will play a crucial role in future regulation and management of insurers' risk, as it is likely to be—fully or partially—adopted by most insurance companies. Only for large and/or "sophisticated" companies will it be economical to develop their own internal model. But even then, the Standard Formula will, in one way or another, be a kind of anchor for any (partial) internal

[3]See CEIOPS (2010).

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

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

[6]Unfortunately, not adhering to conventional standard towards reproducibility of empirical analysis, the QIS5 Calibration Paper CEIOPS (2010) lacks necessary information about how data were analyzed and handled before calibrating the various submodules.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

model. Therefore, a proper calibration of the input parameters entering the Standard Formula, i.e., risk–specific SCR factors and correlations, are crucial to ensure a sound regulatory framework.

In the following, focusing on the equity–risk submodule, we will demonstrate that the QIS5 calibration procedure leads to SCR factors and correlations that are "spurious" and far from reliable. Here, the term "spurious correlation" refers to the situation, where the observed correlation between two variables is not genuine, but "...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 a certain annualization procedure, which transforms daily return data into annual returns, causes the calibration parameters to be severely distorted. The chosen annualization strategy has serious implications, as it affects dispersion and dependence structures in the data used for calibration. On the one hand, it induces spurious dependence patterns, which are not genuinely present in the observed data and, on the other hand, may eliminate dependence existing in the data.

Two specific 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 several 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. Unfortunately, the currently proposed Solvency II calibrations for equity risk hamper understanding and modeling of risk and rather tend to obfuscate insurers' risk assessment efforts.

It should be emphasized that the issues raised here differ from the criticism against specific calibration choices that has been voiced before.[7] They are more fundamental and call the calibration procedure as such and, therefore, virtually all parameters derived for the equity–risk module into question, as they turn out to be largely a product of chance.

The organization of this paper follows the two dimensions in which dependencies can take affect: in the temporal and the cross–sectional direction. After reviewing next the annualization procedure chosen for QIS calibrations, Section 3 investigates consequences for return and risk dynamics that arise from this procedure, and Section 4 those for dependencies between asset classes. Section 5 discusses the implications of our findings. Some technicalities are treated in the appendix.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

2 Rolling–window Annualization

Solvency II calibrations for equity–risk assume a one–year holding period for investments. All SCR or, for that matter, VaR–calibrations are tuned to that horizon. Therefore, all inputs for the Standard Formula need to be VaRs and correlations associated with annual returns. A straightforward calibration strategy would rely on annual return data for deriving the input parameters. For most asset classes in the equity–risk module, there exist only rather short histories, however, so that the analysis would rest on very few annual return observations. Specifically, having daily data with histories ranging from about 12 to close 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 12 to 40 non–overlapping annual return observations, we cannot directly derive empirical VaR–estimates at the 99.5% confidence level nor the type of correlation, i.e., tail correlation, adopted in QIS calibrations.

To still make use of historical market data, QIS calibrations employ a rolling–window approach to obtain annual returns at a daily frequency. Let 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, i.e.,

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

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, annual returns 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 observations. Clearly, the use of *non–overlapping* annual return data is preferable, because only they represent independent pieces of information. 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 quote may suggest.

[8]To compute correlations among the indices used to proxy the asset–classes, CEIOPS used about 40 years of daily observations for the asset pair Global Equities/Commodities, about 15 for the pairs Global Equities/Private Equities and Global Equities/Emerging–Markets Equities, and about 12 for the pair Global Equities/Hedge Funds.

[9]In the simulations discussed below, we choose window lengths ranging from $w = 1$ (i.e., no temporal aggregation) and $w = 259$ (the average annual number of trading days recorded for the MSCI World index) representing an aggregation over one calendar year.

[10]See Paragraph 3.56 in CEIOPS (2010).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

The most damaging implication is that the annualization tends to induce spurious dependence patterns, both over time and across assets, which, in turn, produce artifactual risk structures.

Before continuing, it should be noted that returns calculated via (2) will be referred to below as *discrete* returns. For reasons of analytical tractability, empirical and theoretical analyses in finance typically employ approximations in form of *continuous* returns, denoted by lower–case r_t^w and defined by $r_t^w = \log P_t - \log P_{t-w}$.[11]

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 annualization step for the dynamic properties of the *returns* themselves and then of the *volatility* of returns. Finally, we examine the implications of the calibration of the VaR stress factors entering the Standard Formula.

3.1 Return Dynamics

The determination of VaR parameters 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 multi–period returns, r_t^w and r_{t-1}^w, becomes stronger as the length of the rolling window, w, grows, 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 approaches a random–walk–type process and, thus, approaches nonstationarity. A random–walk, say x_t, in its purest form is generated by the stochastic first–order difference recursion

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

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

[11]For analytical analyses, we will resort to continuous approximation, r_t. All simulations, however, are conducted with exact, discrete returns, R_t, because continuous returns are typically poor approximations in case of—typically larger—annual returns. See Appendix A for a discussion on this issue.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

$s \neq t$. Process (4) with $a = 1$ is also referred to as a *unit-root process*.[12] Expressing rolling–window returns, r_t^w, $t = 1, 2, \ldots, T$, in terms of the first–order recursion

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

the ordinary least–squares (OLS) estimator of the autoregressive (AR) 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 pseudo–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 return series, r_t, $t = 1, 2, \ldots, T$, of length $T =$ 2,590 and $T =$ 5,180, with returns being iid and 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 problem increases as the aggregation level increases. For each aggregation window, we estimate the first–order AR 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 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 do not 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, grows. Weekly aggregation produces a value of about 0.80 and monthly aggregation of about to 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 do not reject the

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

null hypothesis of a unit root at any conventional significance level. For the 20–year sample, we can reject at the 90% and 95% levels, but not at the 99% level.

These findings suggest that for large annualized 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 implied AR coefficient or, for that matter, the value of w.

An aggregation window of $w = 259$ turns out to 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., $\text{Corr}(R_t, R_{t-k})$ and $\text{Corr}(R_t^{259}, R_{t-k}^{259})$, $k = 1, 2, \dots, 259$. The SACF for daily returns (blue) 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 (red) resembles that of a unit–root series. It starts near one, decays in a very slow and almost 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), when plotting the daily (blue) and annualized (red) returns of day t against those at $t-1$.

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 issue, but that it has practical consequences for Solvency II.

3.2 Volatility Dynamics

Rolling–window annualization not only affects the dynamics of the return series in terms of autocorrelations, the volatility or risk dynamics will be affected as well.

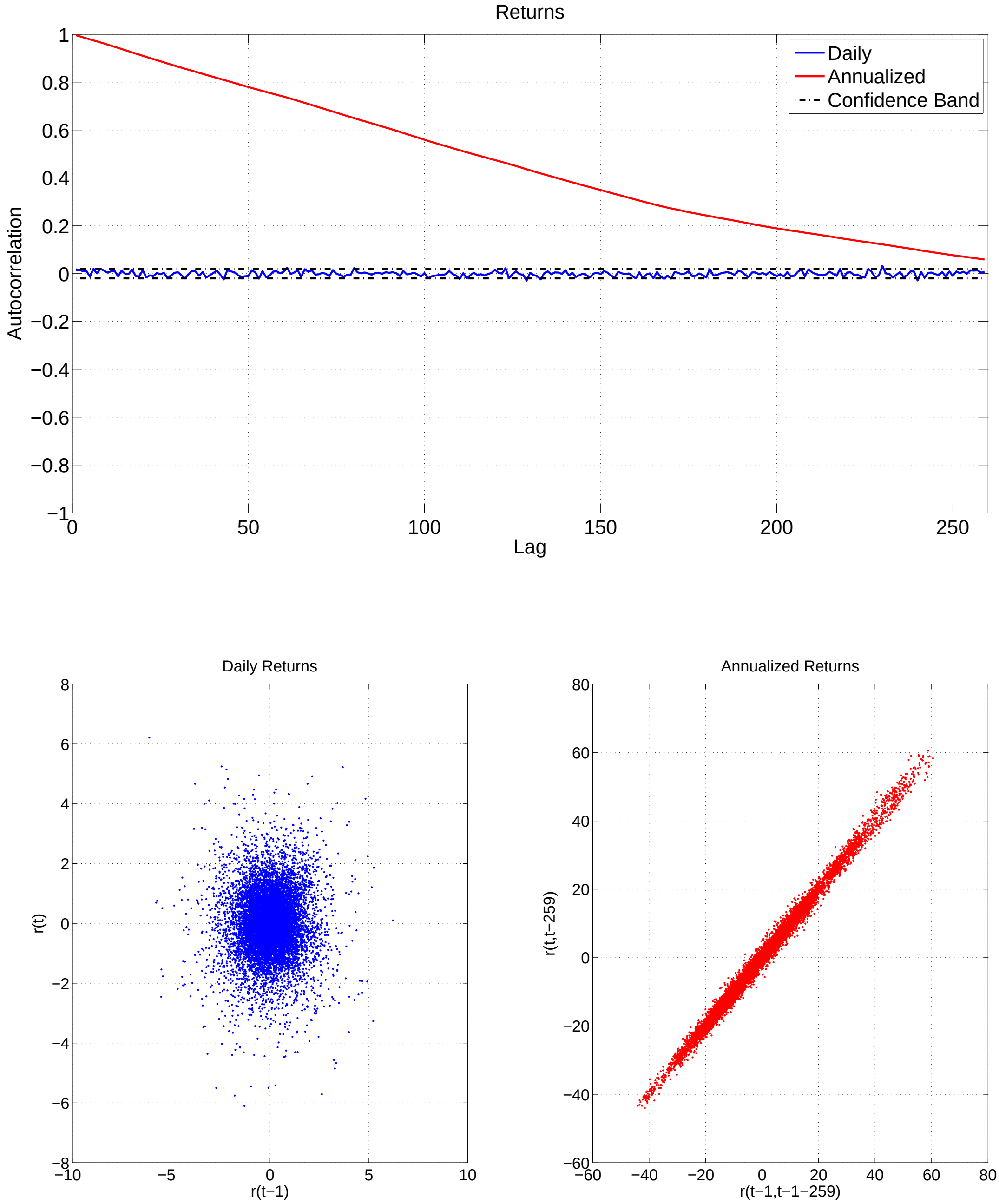

Figure 1: Sample autocorrelations (top) and scatter plots of simulated daily (bottom left) and annualized (bottom right) returns

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 of certain data transformations rather than a genuine property of the underlying return process, all risk–management efforts will be seriously undermined.

The class of Generalized Autoregressive Conditional Heteroskedasticty (GARCH) models, introduced by Engle (1982) and Bollerslev (1986), is the most common approach to 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 normal iid process with $\mathrm{E}(u_t) = 0$ and $\mathrm{Var}(u_t) = 1$, for all t. For the simulation, we use the GARCH parameters obtained by fitting (7)–(8) to the daily returns on the MSCI World index,[16] the index employed in QIS5 to calibrate the asset class "global equities" 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 (blue) 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 to become more or less insignificant after a lag of about 80 days. Thus, a (negative or positive) return–shock carries over to next period's volatility with a correlation of 0.2. The impact gradually vanishes for higher lags. For absolute annualized returns (red), 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.

A comparison of the two SACFs shows that rolling–window annualization not only affects the temporal correlation of an individual return series, but it also alters the risk structures by inducing much stronger and more persistent temporal risk dynamics. As a result, and shown next, the calibration of stress factors of individual equity classes can produce extremely misleading results.

[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 to proxy unobserved conditional volatility, because absolute returns tend to exhibit superior forecastability; see Granger and Sin (2000).

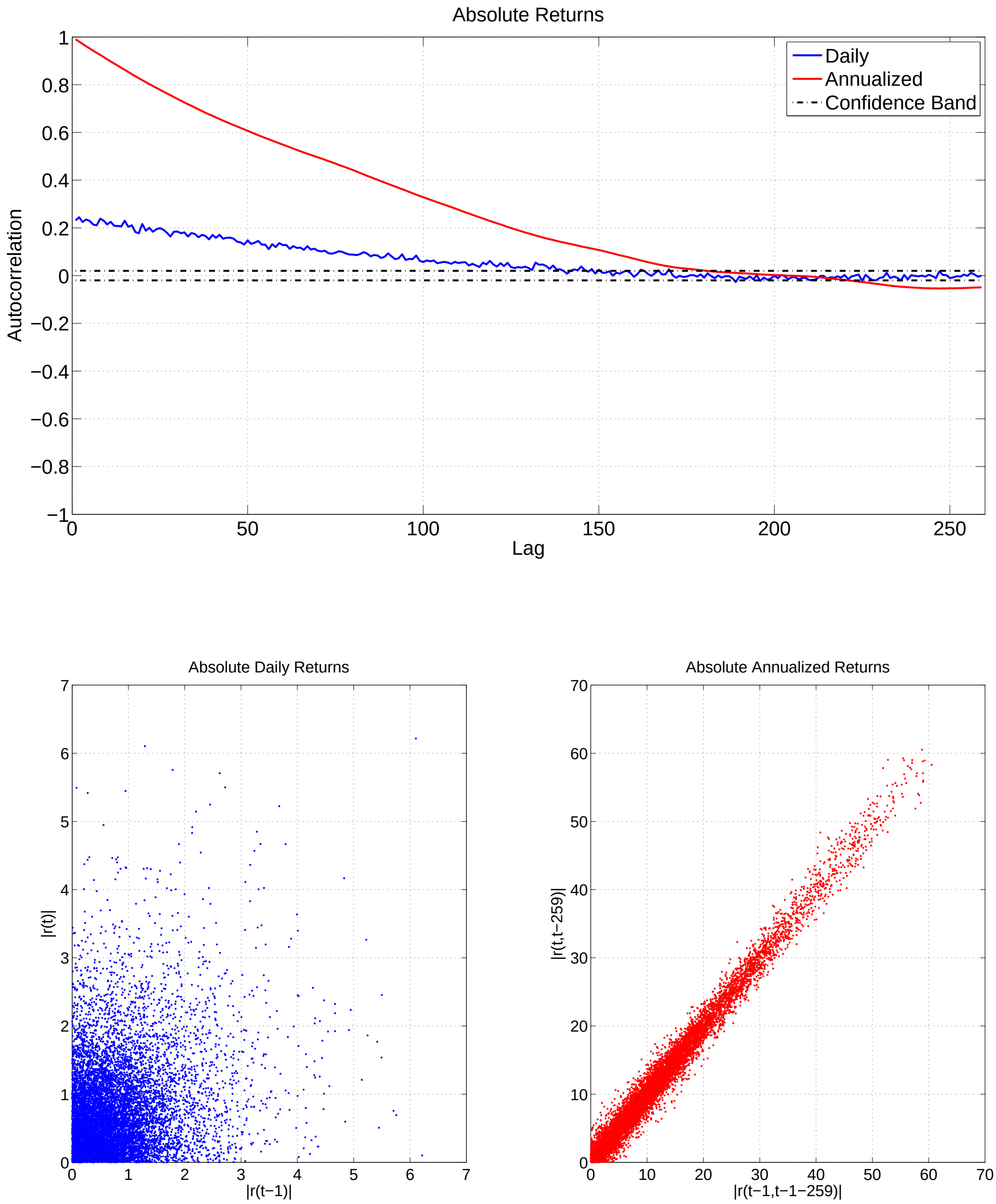

Figure 2: Sample autocorrelations (top) and scatter plots of GARCH–simulated absolute daily (bottom left) and absolute annualized (bottom right) returns

3.3 Consequences for Stress Factors

The presence of unit roots or near–unit roots has implications for risk assessment. If a time series is nonstationary, past behavior will be a poor indicator for 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 potential future losses. To illustrate this, we conduct a Monte Carlo experiment, generating iid data from a standard normal distribution, i.e., $r_t \overset{iid}{\sim} \mathrm{N}(0,1)$. 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 observations at a daily frequency.

Figure 3 plots the daily and annualized returns of the two simulated risk factors.[18] As is to be expected, being generated by the same process, the two *daily* series look very similar. They move closely around zero, roughly within the ± 3 interval, without displaying any temporal patterns. In contrast, the *annualized* versions appear to fluctuate in a cyclical pattern with extended up– and down–swings, varying between –50% and +60%. Though being generated by identical processes, the locations of their peaks and troughs differ considerably.

Next, we estimate SCR stress factors for the simulated return series. 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 two daily return series are rather stable; they hover around the expected value (solid horizontal line) and, ranging from 2.3% to 2.8%, stay about 95% of the time within the 95% confidence bounds.[19] Compared to this, the VaR estimates from annualized returns vary dramatically. They assume values between 16% and 46% during the 89 years[20] sampled and deviate considerably from the expected value. They deviate by more the than ten standard deviations in either direction and stay for long periods far away from the expected value. It is the exception rather than the rule that the estimates fall inside the confidence band.

Given that the data were generated by iid white noise processes, i.e., processes without any temporal dependence structure, the VaR sequences from annualized returns appear to exhibit distinct temporal patterns, which, in practice, may easily be mistaken for structurally inherent properties. Such SCR patterns may trigger

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

[19] Note that, against common convention but in line with CEIOPS' usage, Figure 4 plots negative VaR–values.

[20] We obtain estimates for 89 years because we lose the initial 11 years of the sample: one year due to the annualization and ten years to calculate VaRs from ten–year histories.

 doi:10.20944/preprints201607.0021.v1

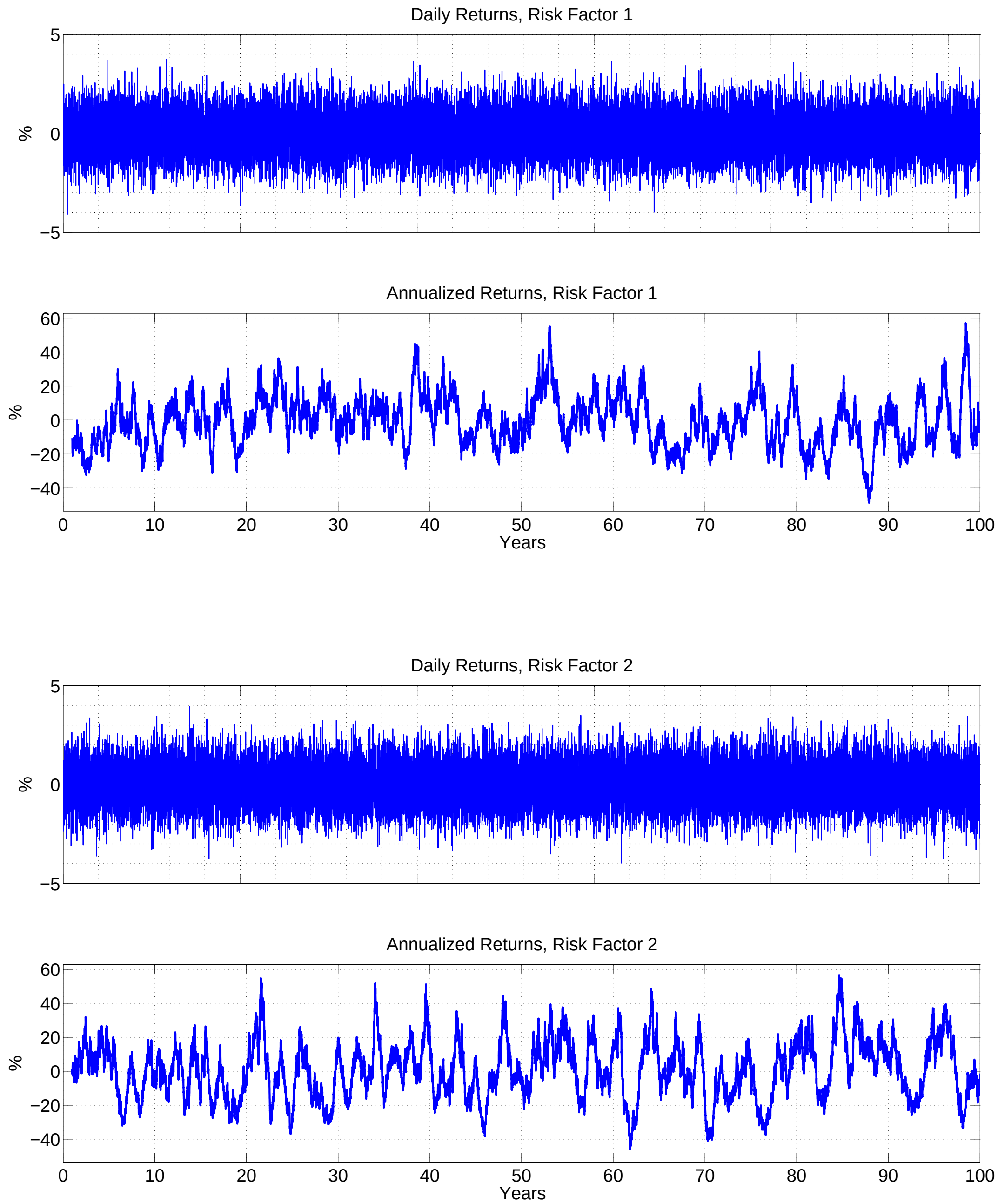

Figure 3: Time series plots of two simulated independent daily return series and corresponding annual rolling–window returns

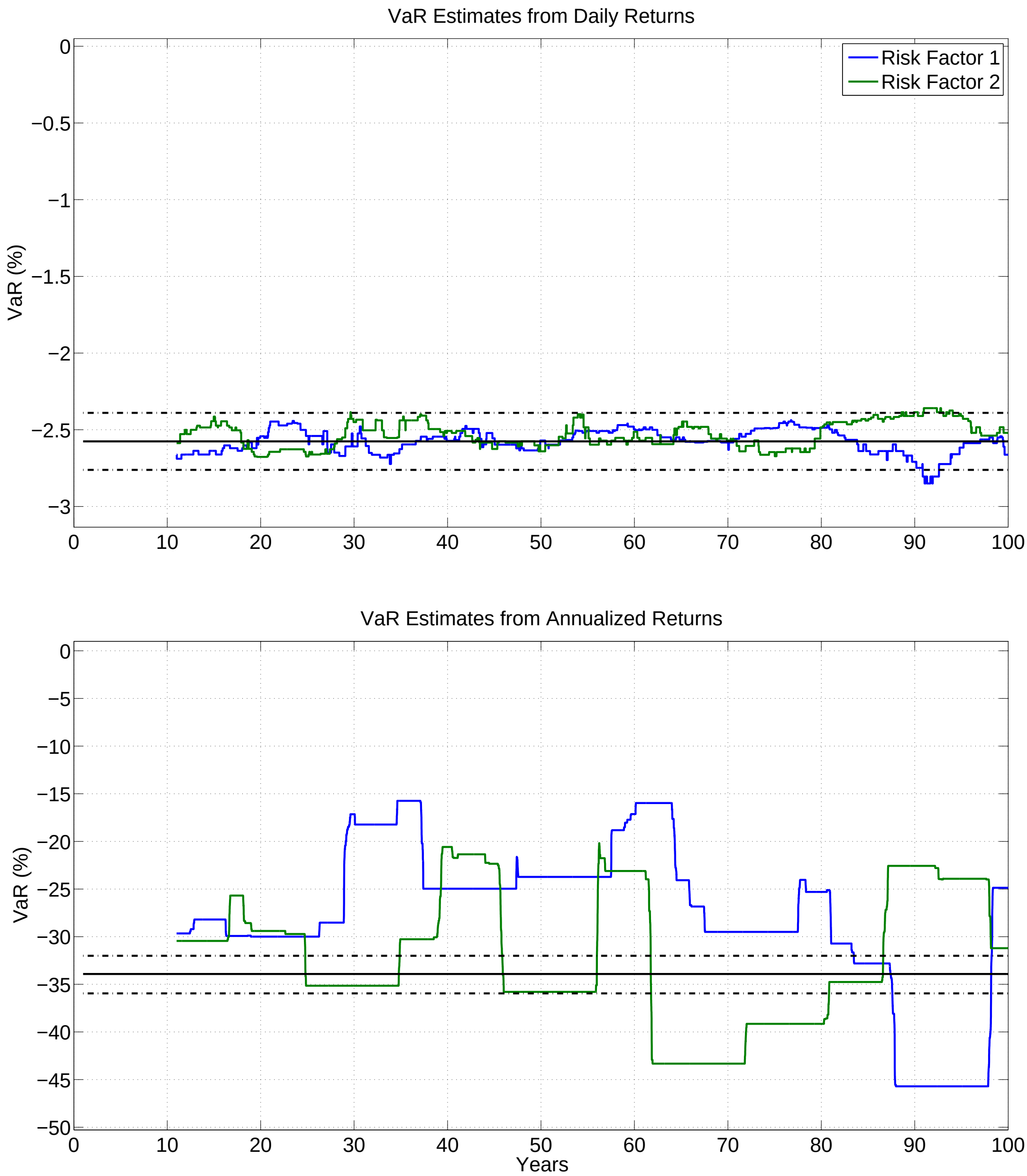

Figure 4: Historical VaRs for daily (top) and annualized (bottom) returns with theoretical VaRs (solid horizontal line) and 95% confidence intervals

specific regulatory actions, as they suggest structural changes in the riskiness of asset classes. 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.

A disturbing fact is that, although annual–return VaRs exhibit strong persistence, they can change very abruptly. An insurance company's reliance on annualized–return VaRs is bound to make sudden, erratic and costly portfolio adjustments, even though there are no changes in the underlying market processes.

From all this, it follows that the use of VaR estimates derived from one–year rolling–window returns in either regulatory or insurers' risk management processes will produce highly 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 use the conventional Pearson correlation coefficient. Not only is it easily computed, Pearson correlation is also the cornerstone 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 framework. 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 can assume values far away from zero, despite the two series being completely independent. Clearly, if this is the case, any correlation estimate between two nonstationary time series will be unreliable.

Figure 5 indicates the potential problem when assessing the dependence structure between risk factors after rolling–window annualization. The graph in the top half overlays the two independently simulated annualized return series plotted in Figure 3. We observe periods where both series seem to run pretty much in a synchronous

[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).

fashion 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 patterns of the original and the annualized 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 splattered. This spottiness arises from the fact that the joint behavior appears to change in long swings.

An illustrative selection of three 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 the two series exhibit over fairly long periods 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 correlation estimates of independent (near–)unit root process.

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

4.1 Correlation Concepts

QIS calibrations for equity risk are based on "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 sets, the number of data points 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

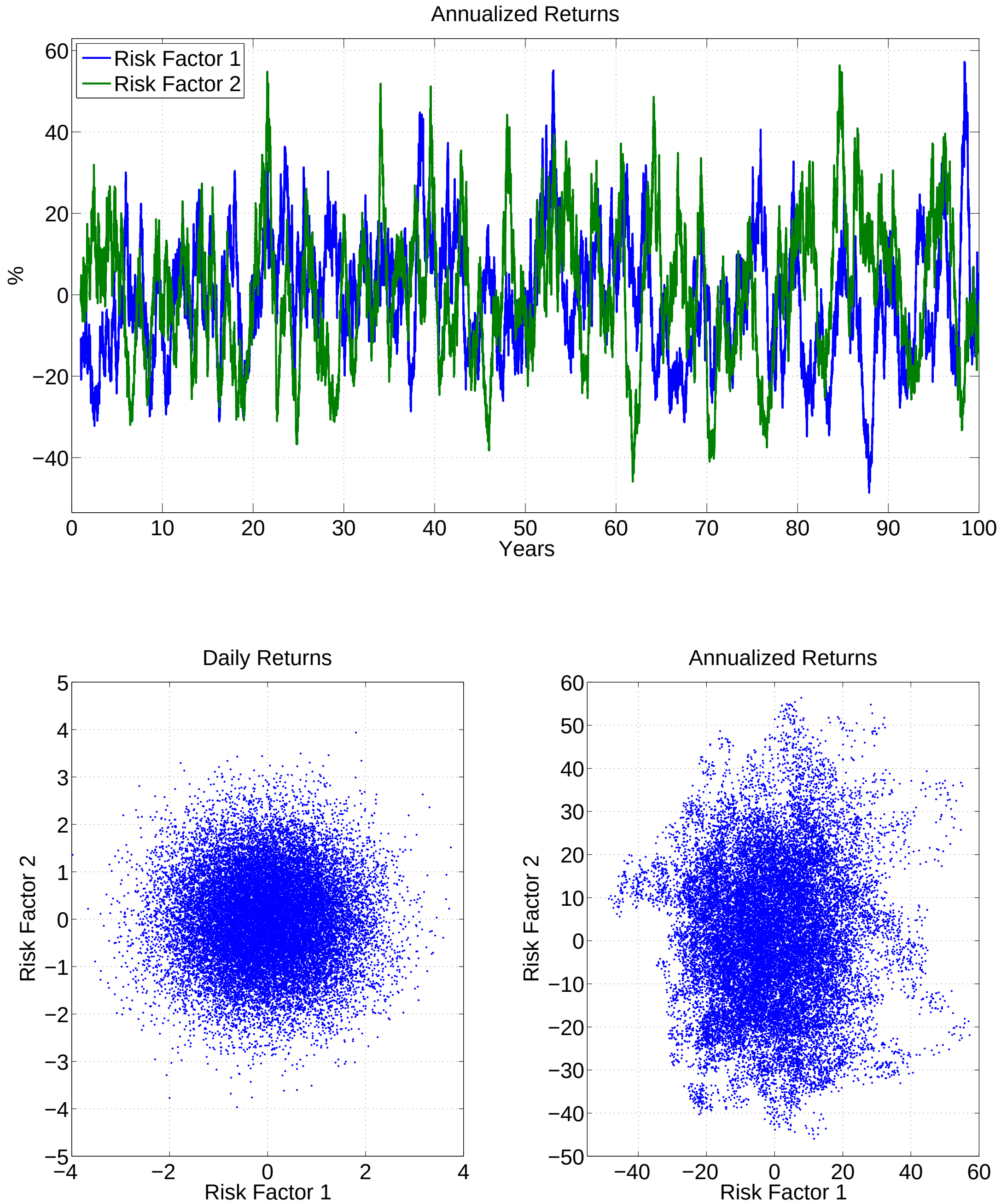

Figure 5: Time series plots of simulated annualized returns (top) and scatter plots of daily (bottom left) and annualized (bottom right) return series

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

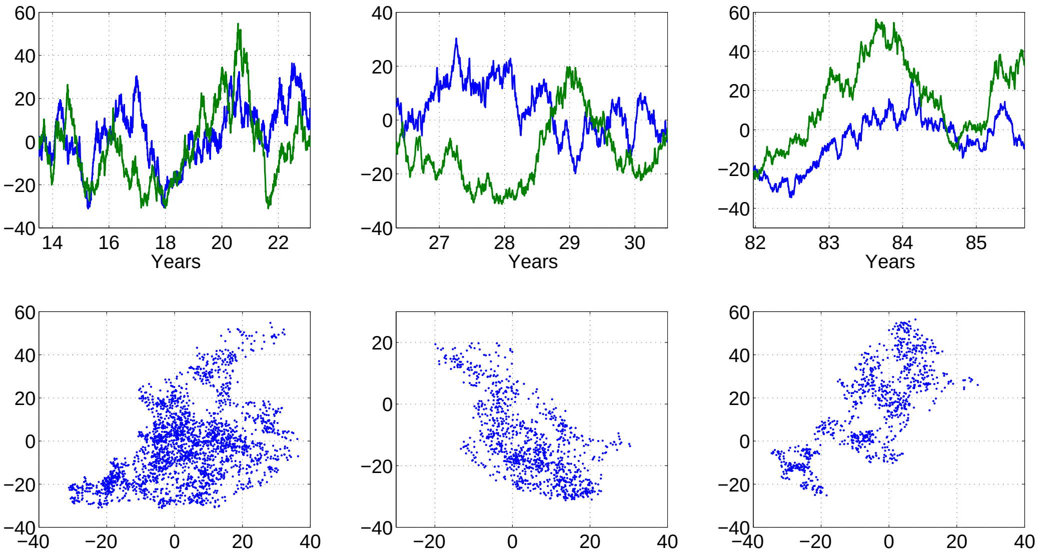

Figure 6: Selected subsets of the annualized return series showing time series plots (top) and corresponding scatter plots (bottom) with sample correlations $\hat{\rho} = 0.42$, $\hat{\rho} = -0.65$ and $\hat{\rho} = 0.75$ (from left to right)

0.5%–quantile matter. This leaves us with 50 tail–observations for each individual asset. The intersection of these two subsets, i.e., data pairs where *both* components simultaneously fall below the 0.5%–quantile, defines the set of *joint* tail observations. Depending on the degree of dependence, this will generally leave us with much fewer than 50 observations.

Figure 8 illustrates for a bivariate normal distribution how the portion of common tail observations quickly drops as we move away from perfect positive correlation. For example, given a correlation of, say, $\rho = 0.75$ and having observations on 10,000 return pairs, we can expect to have only 14 joint tail observations. So that even for large data sets, tail–correlation estimates via data–cutting 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 and especially on far–tail correlations may give 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 will approach zero as we go further into the tails,[22] suggesting the absence of dependence.[23]

To avoid all these problem, a different data–cutting strategy could be adopted. Rather than computing correlations from *joint* tail observations, we can condition

[22]See Rosenbaum (1961).

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

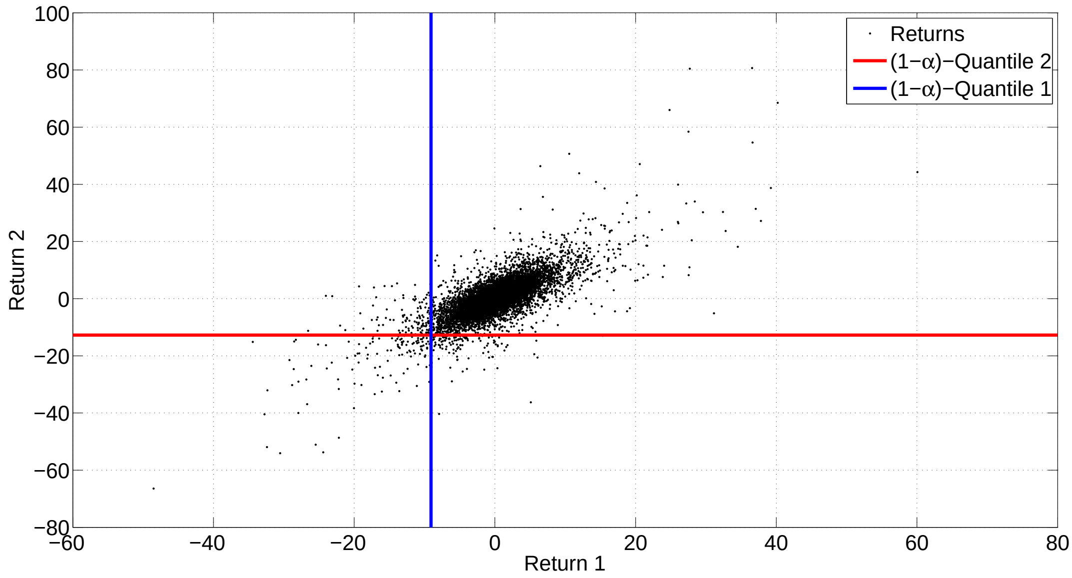

Figure 7: Tail correlation via data–cutting approach

on only one risk factor and compute

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

This segments the two–dimensional return plane into half–planes rather than quadrants and ensures that we do not end up with an insufficient number of tail observations, as the resulting effective sample size corresponds to the chosen VaR–quantile. The use of (10) is particularly appropriate when the asset conditioned on is regarded as the underlying risk driver.

Because of the small number of data points for computing tail correlations—even in the presence of large data sets—, QIS calibrators do not, or not exclusively rely on the data–cutting method (9). They also seem to adopt what we refer to as *VaR–implied correlations*,[24] which simply result from an inversion of the Standard Formula. For two risk components, the Standard Formula reduces to[25]

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

CEIOPS (2010)[26] suggests to use that value for ρ which minimizes the "aggregation

[24]Ultimately, it is not clear which particular method has been used to derive the equity–risk correlations reported in CEIOPS (2010).

[25]Equation (11) assumes that both return series have mean zero. In practice, this assumption is violated. Ignoring this fact, use of (11) will 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 means, i.e., expectations should come from.

[26]See Paragraph 3.1251 in CEIOPS (2010).

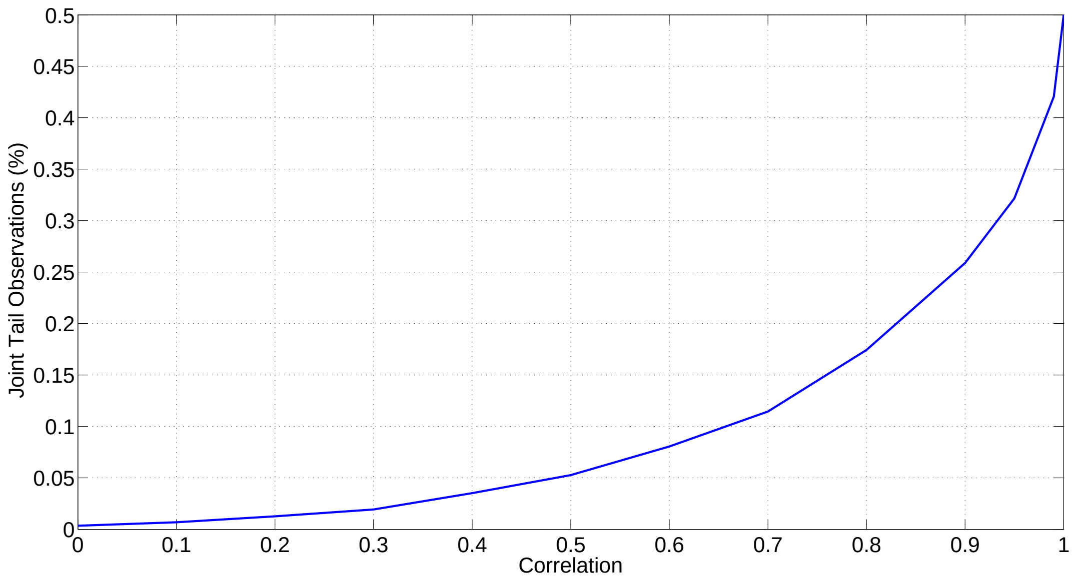

Figure 8: Percentage of joint tail observations in data–cutting approach

error"

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

Having empirical VaR–estimates, denoted by $\widehat{\mathrm{VaR}}_\alpha(\cdot)$, for the returns on the individual assets i and j as well as on their the sum, the solution to CEIOPS' minimization problem becomes

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

The first condition in (12) arises in the presence of superadditivity, i.e., when subadditivity[27] fails. The second condition could be referred to as "superdiversification," i.e., the (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 applies, the VaR–implied–correlation estimate will be strictly between ± 1. Although superadditivity and superdiversification may be rarely encountered with equity returns, the coarseness of extreme–quantile estimates may, in empirical analysis, lead to such pathological situations.[28]

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

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

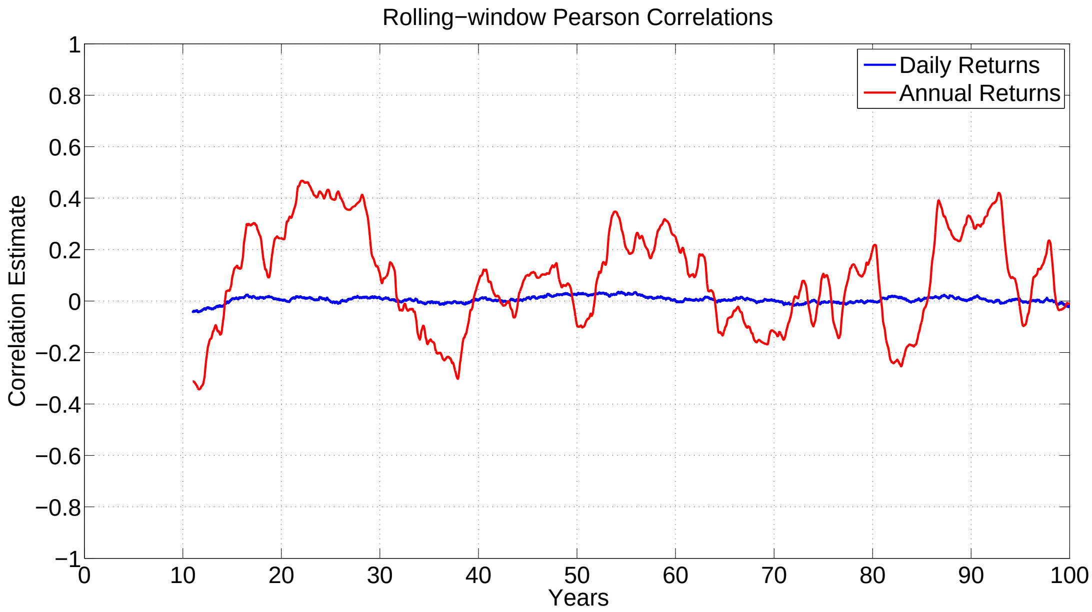

Figure 9: Pearson–correlation estimates for daily and annual returns

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, a 10–year rolling window and derive correlation estimates for each day in the 100–year period, starting at the beginning of 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, within a range of ±0.05. The estimates derived from the one–year rolling–window returns behave very differently. They vary erratically, assuming values between about –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 usual Pearson correlations, we also compute 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

Figure 7, to contain any data, so that tail correlations cannot be computed.

If the data–cutting approach is to be adopted, it is unlikely that one can 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. CEIOPS analysts also experimented with alternative confidence levels.[29] Analyzing the dependence between equity and fixed income, CEIOPS considers confidence levels from 99% down to 80%. As the 99% confidence level is, in general, still too ambitious to obtain sufficient joint tail observations, we compute data–cutting tail correlations for the 95% and 80% confidence levels.[30]

The number of available joint tail observations (top) and the corresponding 95%–level tail–correlation estimates (bottom) for both daily and annualized returns 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 (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, it is not clear what performance is to be expected from the data–cutting approach, when applied to risk aggregation in the Standard Formula. When using annualized returns, the problem will not vanish, even when working with much longer than 10–year samples. One option could be to substantially lower the confidence level. But even for the 80%–level, the number of observations can be insufficient. 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, even at the 80%–level, the tail–correlation estimates jump erratically and assume values between –1 and +0.7.

Clearly, in view of these problems, the DCQ–correlation approach does not qualify for regulatory purposes, unless more observations from the center of the distribution

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

[30]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.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

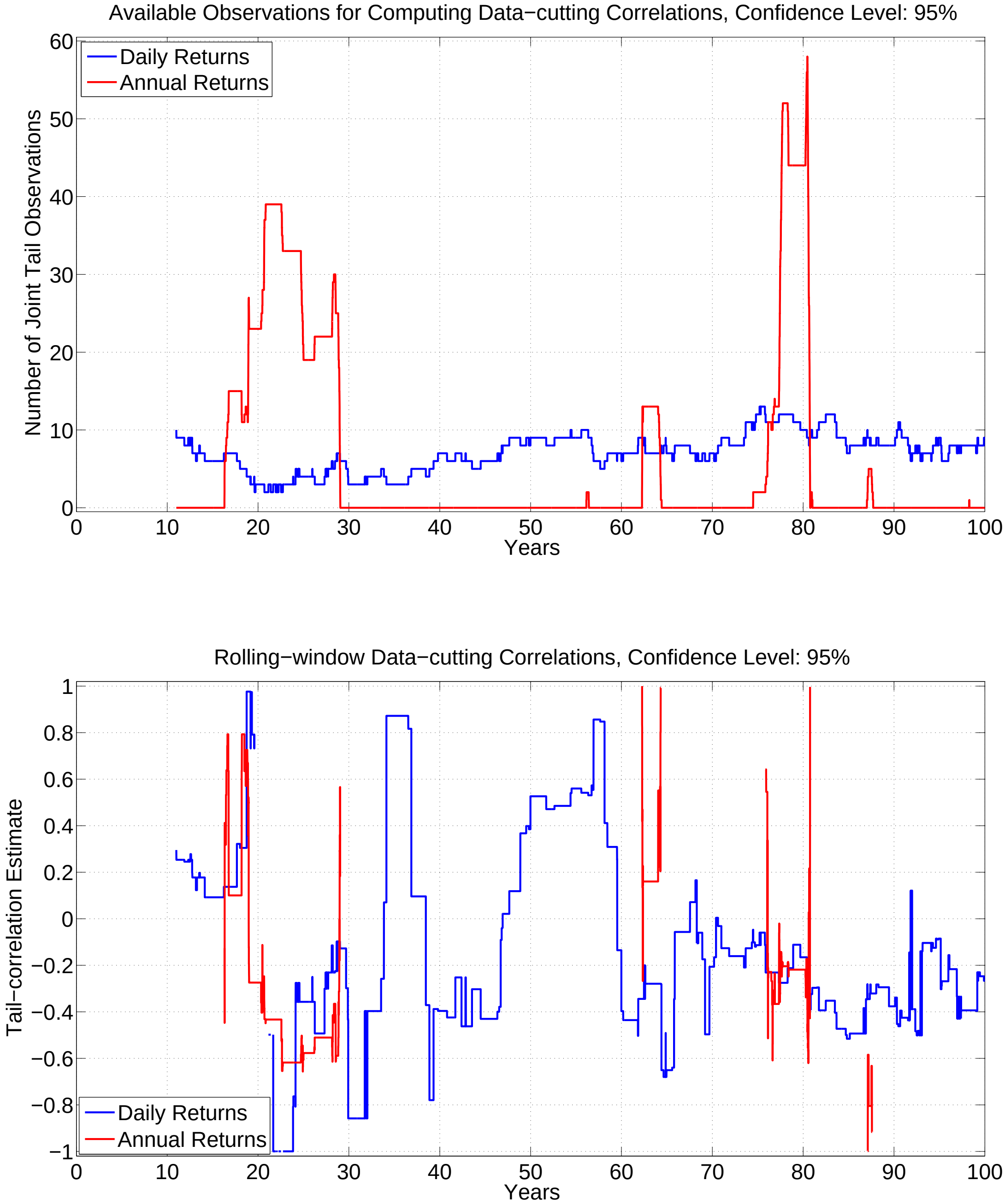

Figure 10: Available number of observations (top) and data–cutting tail–correlation estimates (bottom), $\hat{\rho}_{95}^{DCQ}$, applied to daily and annualized returns; 95% confidence level and 10–year estimation window

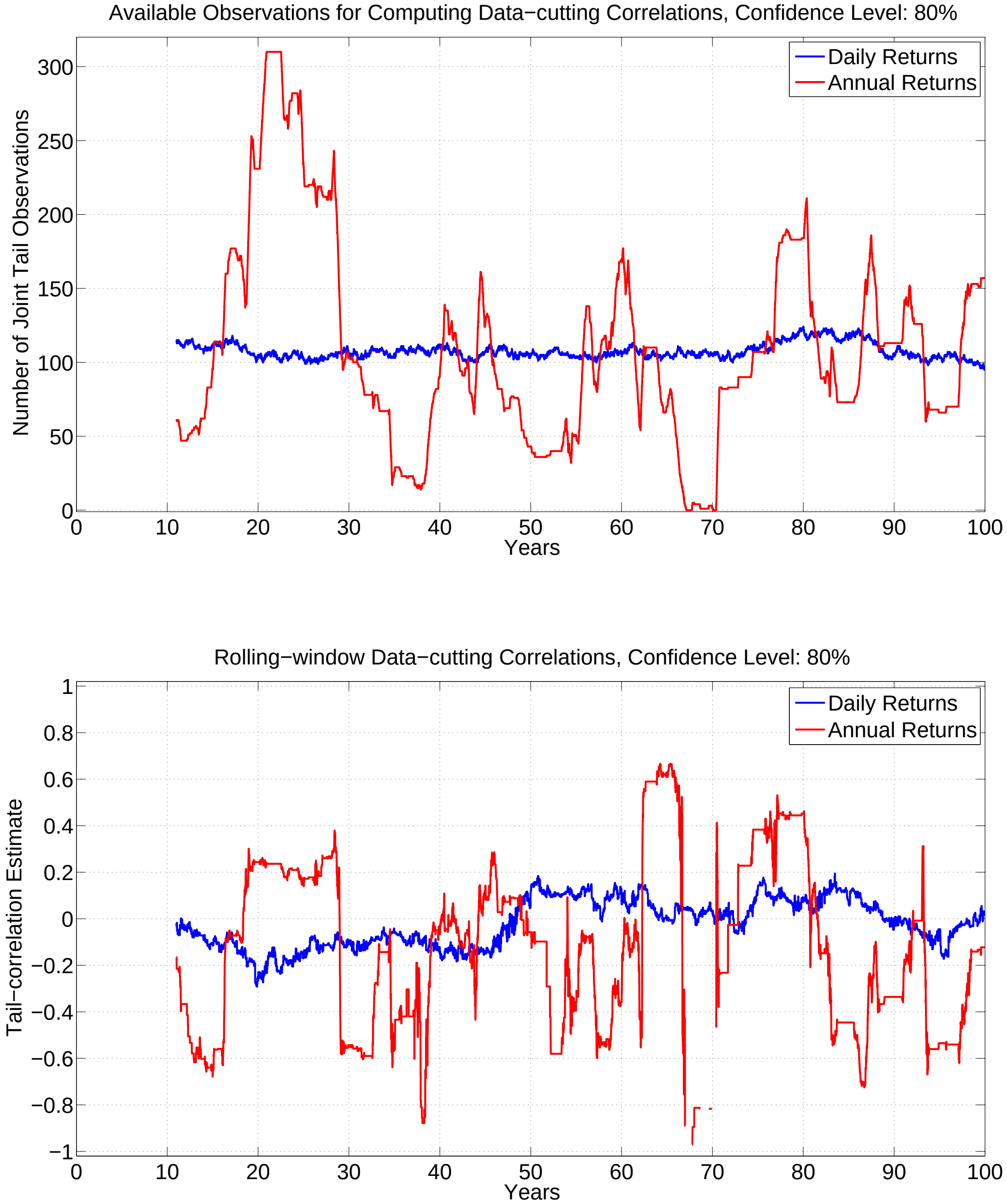

Figure 11: Available number of observations (top) and data–cutting tail–correlation estimates (bottom), $\hat{\rho}_{80}^{DCQ}$, applied to daily and annualized returns; 80% confidence level and 10–year estimation window

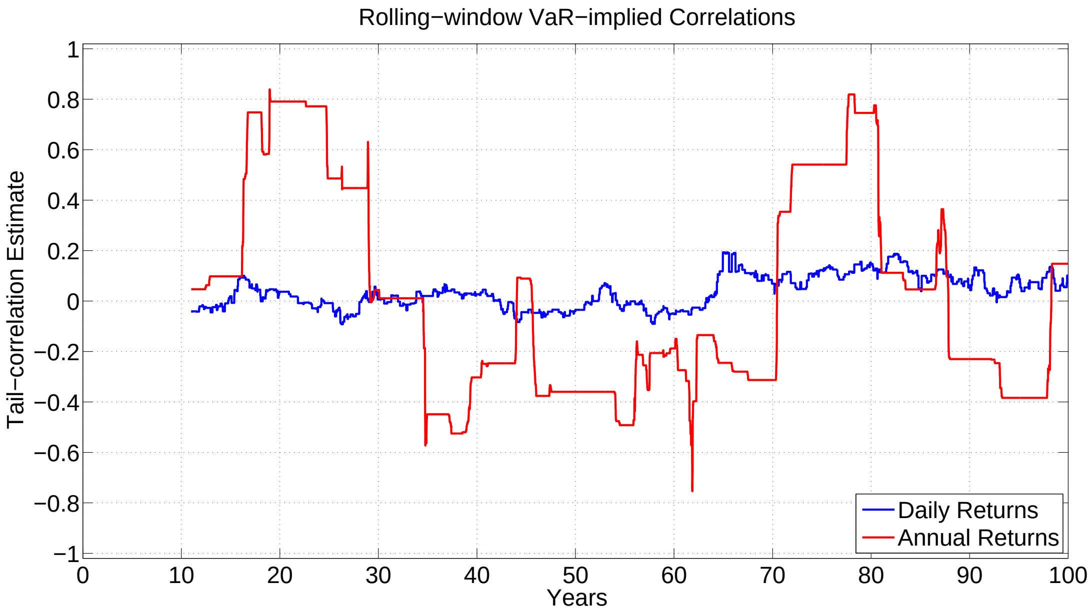

Figure 12: VaR–implied tail correlations for daily and annual returns

are included. Then, however, we can no longer speak of *tail*–correlations estimates.[31]

We also compute $\text{VaR}_{99.5}$–implied tail correlations from the simulated data (Figure 12). For daily returns, the estimates lie consistently between –0.1 and +0.2. For annualized–returns, however, we obtain extremely erratic results. Although the tail–correlation estimates should all be zero, they assume values from about –0.75 to +0.85; they exhibit sudden jumps and sign switches; and they are hardly ever close to zero.

To summarize, the simulation results for data–cutting and VaR–implied correlations strongly indicate that overlapping annual rolling–window returns will prevent a meaningful calibration of the correlational input parameters for the Standard Formula.

4.2.2 Bias and Efficiency

In empirical analysis, it is commonly desired to work with unbiased and efficient estimators. That is, the estimator should, on average, produce accurate estimates; and it should do so with little uncertainty—meaning that the intervals of uncertainty around these 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 simulation analyses to investigate both the bias

[31] It should be noted that the problem of insufficient joint tail observations for computing data–cutting correlations may be less dramatic when returns are very heavy–tailed. We will address this issue in Section 4.3.

and the efficiency of correlation estimates as the window lengths increase.

To do so, we draw the daily 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 length $259 \times n$, 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 correlation based on all data; the half–plane data–cutting correlation, $\rho^{DCH}_{99.5}$, based on the 0,5%–portion of the largest losses; and the VaR–implied correlation at the 99.5% level, $\rho^{VaR}_{99.5}$.

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, w, varies. Whereas both the conventional Pearson and the data–cutting correlation remain unbiased, the VaR–implied correlation estimates exhibit a systematic upward bias that grows 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 0.1, wrongly suggesting a positive dependence.

Turning to the efficiency of the correlation estimators, Figure 14 reveals that the confidence intervals 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 lengths 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 of ±1 for annual aggregation. The maximum range could be due to a couple of extreme outliers. But even the 90%–confidence interval runs from –0.5 to about +0.8, indicating that, apart from being biased, VaR–implied correlation estimates from rolling–window returns can be virtually all over the place. They provide practically no information about the dependencies governing the underlying data process.

The seriousness of the problem is especially evident from the plots in Figure 15. They show how the width of the confidence intervals grows as the window length increases. Clearly, debating whether two particular asset classes have a tail correlation of, say, -0.5 or 0.9 is rather meaningless, given the blatant instability of data–cutting and VaR–implied correlation estimates 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 rapid increase of the estimates' dispersion as the aggregation level grows. The modes of the histograms remain more or less at zero. However, as the aggregation length rises, so does right–skewness, which goes

doi:10.20944/preprints201607.0021.v1

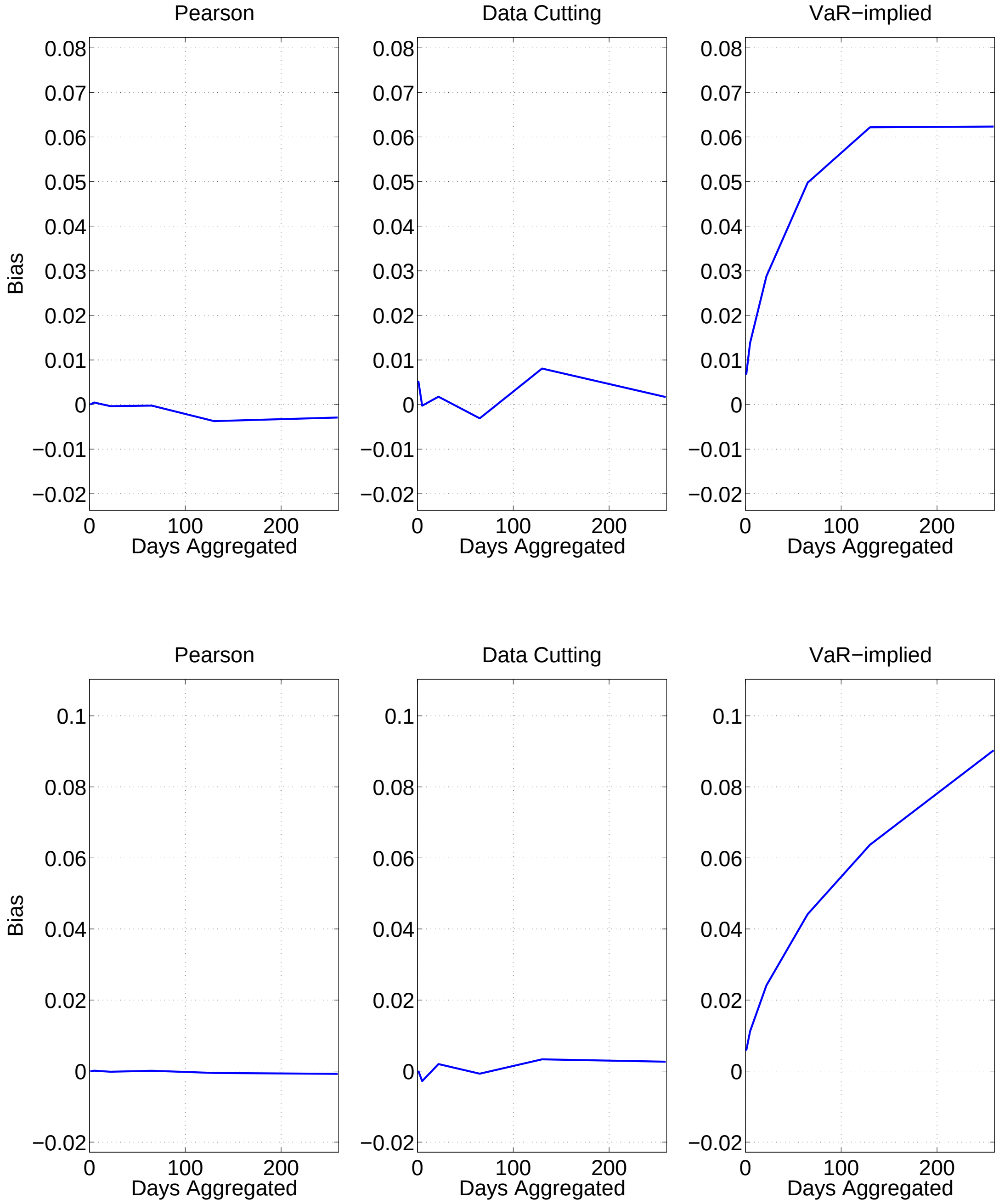

Figure 13: Bias in VaR–implied tail–correlation estimates due to rolling–window aggregation, $\rho = 0$, 10–year (top) and 40–year (bottom) samples

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

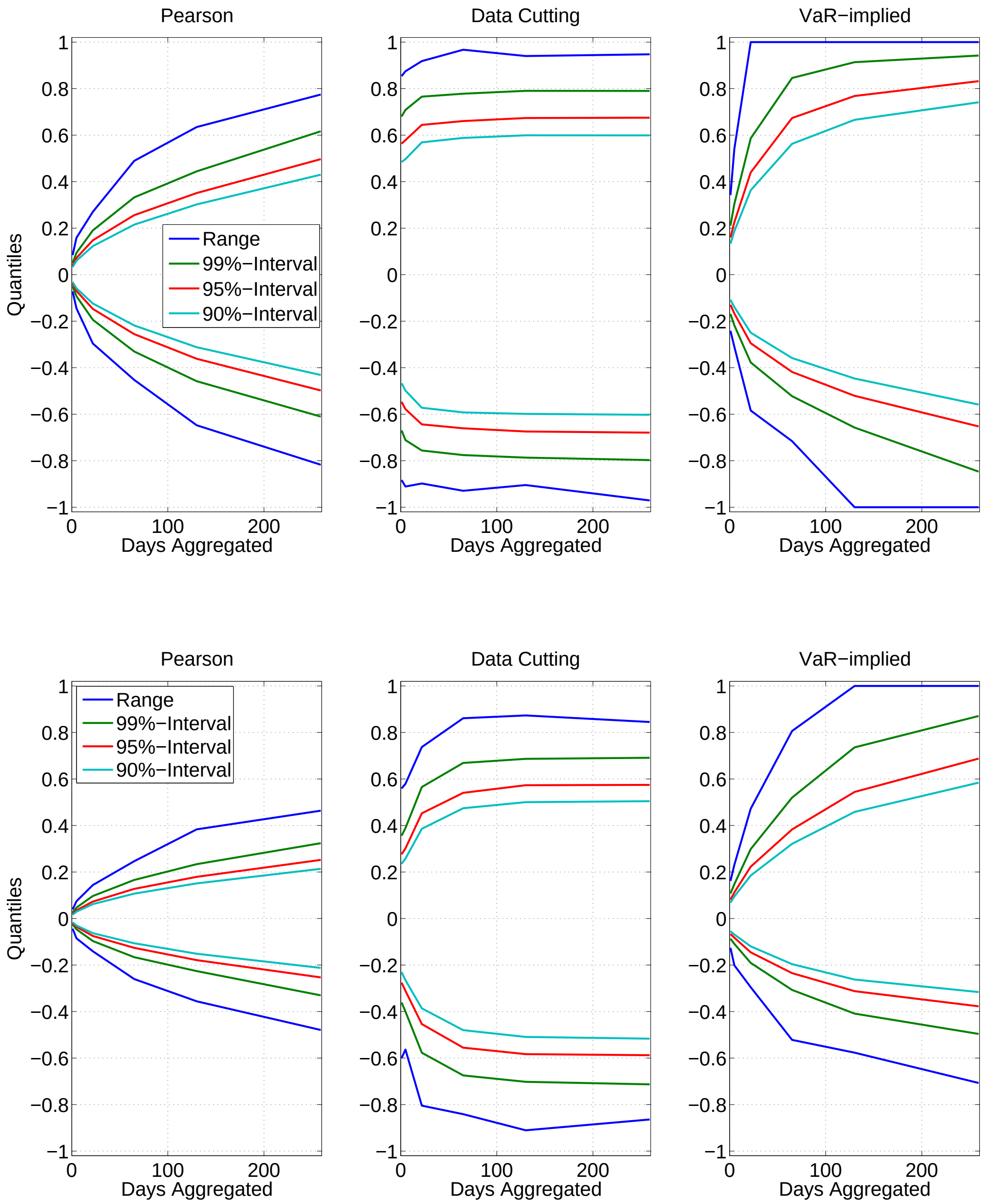

Figure 14: Confidence intervals of correlation estimates and rolling–window aggregation, $\rho = 0$, 10–year (top) and 40–year (bottom) samples

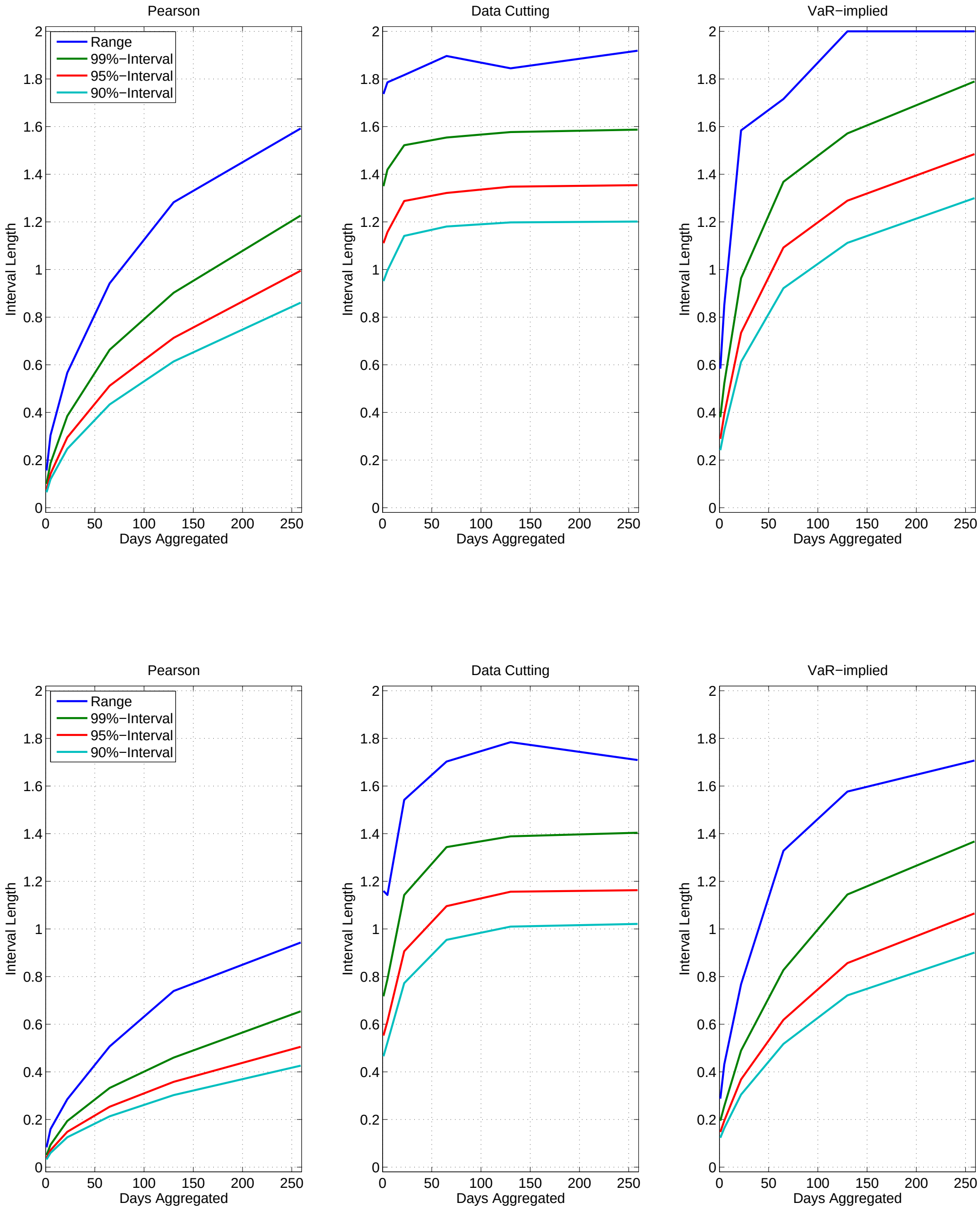

Figure 15: Confidence–interval lenghts of correlation estimates and rolling–window aggregation, $\rho = 0$, 10–year (top) and 40–year (bottom) samples

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

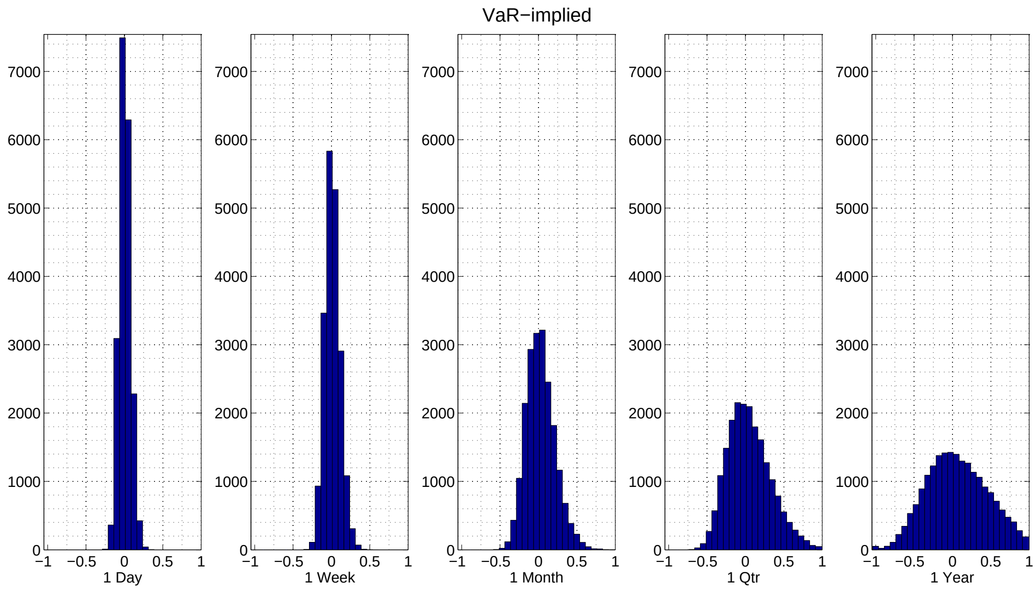

Figure 16: Histograms of VaR-implied tail correlations, $\rho = 0$ and 10–year samples

along with an upward bias in the tail–correlation estimates.

Next, we investigate the performance of the correlation estimators when there is nonzero correlation between assets. We repeat the above 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 correlation levels specified, assuming ten years of daily data. In each case, the VaR–implied correlation estimates from annualized returns exhibit an upward bias. Even more of a concern is the extensive pile–up of estimates near or at $+1$ when daily correlations assume a value of $\rho = 0.4$ or higher. If daily correlations exceed 0.4, the mode of the distribution lies near $+1$, so that there is 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, for example, 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, 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 25% probability that the estimate will lie above 0.87. Thus, tail–correlation estimates tend to be

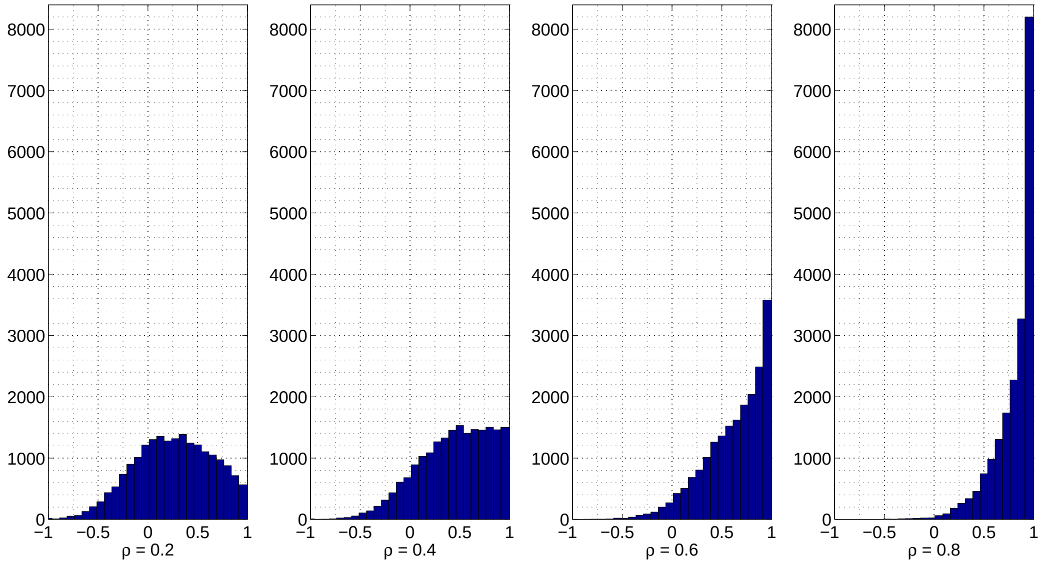

Figure 17: Histograms of VaR-implied tail correlations from annualized returns for differently correlated daily returns

overstated with rather large probabilities.

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 those for the 90% confidence level. Hence, it is very likely to encounter tail–correlation estimates close to unity, even though the true correlation is only 0.4.

Note that for $\rho > 0$, the widths of the confidence bands of the VaR–implied estimates become somewhat shorter relative to the uncorrelated case. Again, this is due to the fact that correlation estimates have the upper bound $+1$. 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 and are practically uninformative.

4.3 Heavy Tails

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2016 doi:10.20944/preprints201607.0021.v1

Table 2: Bias of 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 there is a 10% probability that the estimated $\mathrm{VaR}_{99.5}$–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	0.69	0.88
25%	0.37	0.56	0.74	0.87	0.96
10%	0.63	0.77	0.89	0.96	0.99

freedom and correlation $\rho = 0$, i.e.,[32]

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

Altogether, we performed 20,000 simulation–runs. In each of which we generated 40 years of daily data and aggregated 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 estimate becomes more severe than in the normal case.[33] The bias in $\hat{\rho}_{99.5}^{VaR}$ increases as the degrees of freedom decrease, that is, as tails become heavier. For a one–year rolling window and $\rho = 0$, temporal aggregation with normally distributed daily returns produced 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 VaR–implied correlation estimates from 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 observed pattern is the case of $\nu = 1$, with a bias of –0.06. A t–distribution with $\nu = 1$ corresponds to a Cauchy distribution and is extremely fat–tailed, so that even the mean of the distribution is infinite.[34] The histograms of the VaR–implied correlation estimates for all four degree–of–freedom values (see Figure 20) show that the pile–up of the estimates at +1 starts with $\nu = 3$ and becomes serious for $\nu = 2$. As reported above, for normally distributed returns, the pile–ups occurred for correlations exceeding 0.4. With t–distributed returns, pile–ups happen 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.

[32]Note that it is not unusual to obtain ν–estimates from daily stock returns that are between 2 and 4.

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

[34]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 here.

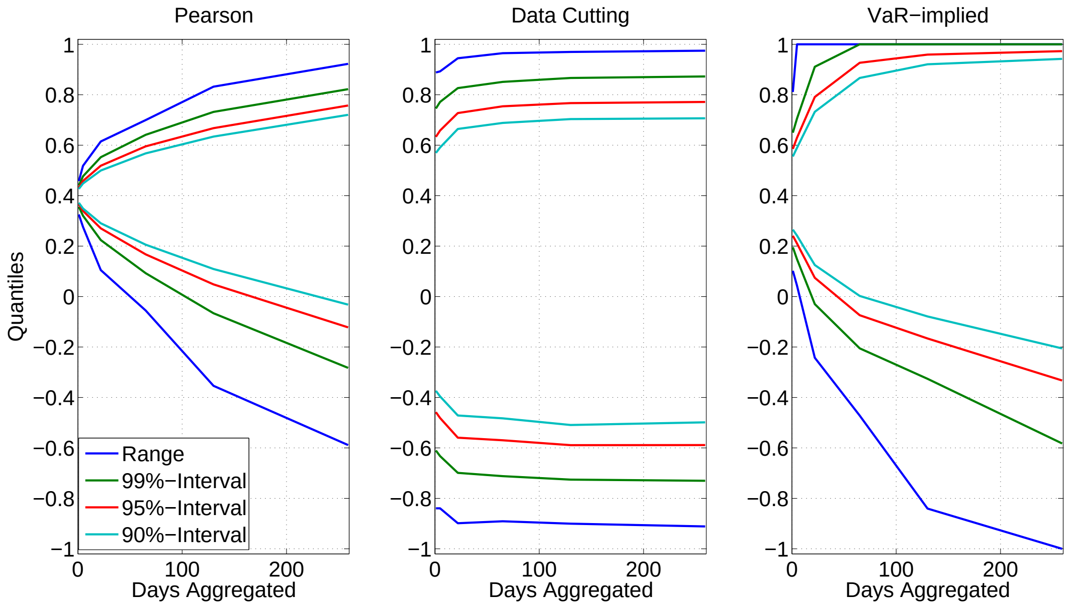

Figure 18: Confidence intervals and temporal aggregation, $\rho = 0.4$ and 10–year sample size

The pile–up problem of the VaR–implied correlation estimates indicates that this dependence measure is not suitable in hte presence of temporally aggregated and overlapping return data. The seriousness of the problem for $\nu = 2$ and $\nu = 1$ could be due to the nonexistence of variances ($\nu = 2$) or the lack of finite means ($\nu = 1$). However, if that was the case, the VaR–implied correlation estimator should already break down when applied to daily returns that are 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, however, already at weekly and worsens dramatically for higher aggregation levels, with pile–ups occurring at both −1 and +1.

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

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 the joint occurrence of losses. This can be measured by the so–called *coefficient of tail depen-*

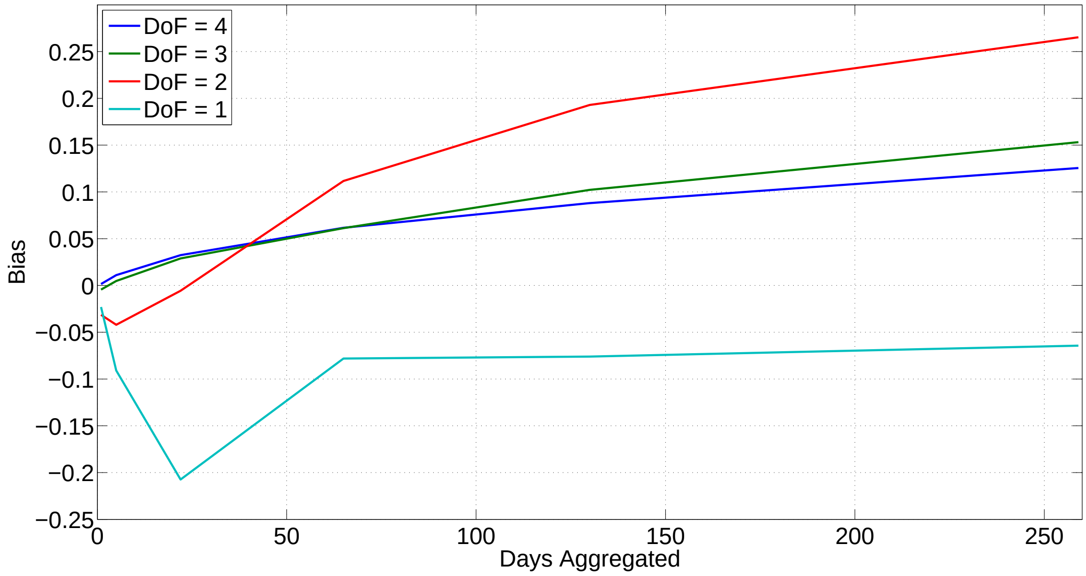

Figure 19: Bias due to rolling–window aggregation for t–distributed daily returns with $\rho = 0.0$ and different degrees of freedom

dence. Let r_i and r_j be the returns of two risk factors with marginal distributions F_i and F_j, respectively. Then, the coefficient of lower tail dependence, denoted by λ, is defined as[35]

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

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 Solvency II regulation intends, assessing the consequences of extreme losses in a given portfolio.

To investigate the implications of rolling–window annualization 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) and $\rho = 0.5$ and $\nu = 4$. The coefficients of tail dependence are estimated for quantiles ranging from 0.001% to 2.5%.[36]

The results are shown in Figure 22. The horizontal line in the plots indicates the

[35]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). Here, we will only focus on lower tail dependence and let λ denote the coefficient of *lower* tail dependence.

[36]We use the regression–type estimator discussed in van Oordt and Zhou (2011).

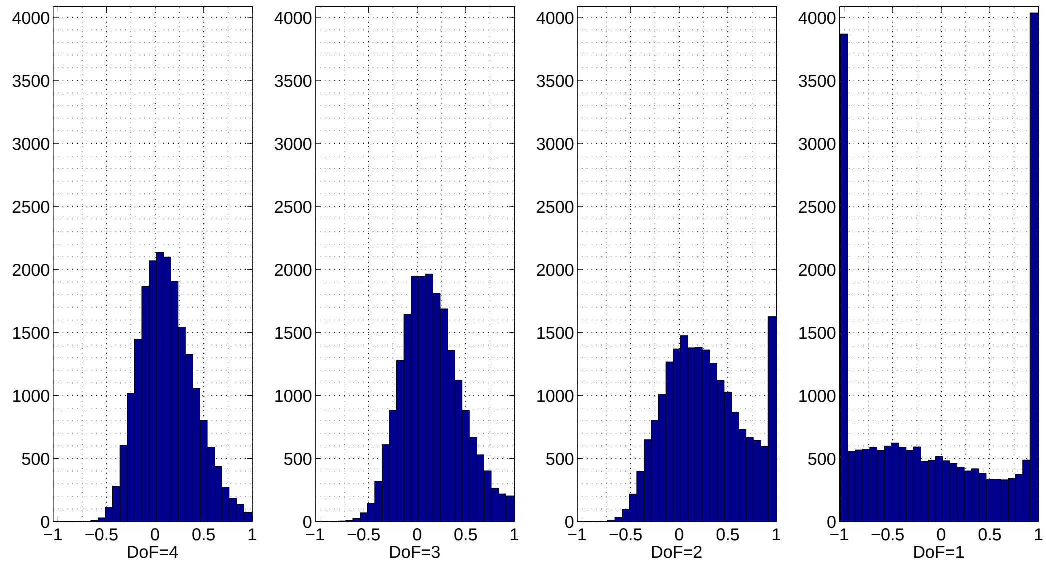

Figure 20: Histograms of VaR–implied tail–correlations estimates for t–distributed daily returns with different degrees of freedom and $\rho = 0.0$

theoretical value[37] of the tail dependence coefficient for the bivariate t–distribution with $\nu = 4$ and $\rho = 0.5$, given by $\lambda^* = 0.2532$; and the dashed lines indicate the (bootstrapped) 95%–confidence bands. The estimates from the daily data slightly overestimate the theoretical value of λ^*, but—as it should be—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. For the 40–year samples, the confidence band becomes extremely wide; and throughout the tail area considered the band includes zero so that the hypothesis of "no of tail dependence" cannot be rejected. The upper limit of the band hovers mostly around 0.6—except for the extreme tail area, i.e., $1 - \alpha \leq 0.1\%$, where the upper limit of the confidence band quickly drops to zero, suggesting 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 not only scrambles *linear* dependence structures between the assets, but it also annihilates their joint tail behavior, so that

[37] 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).

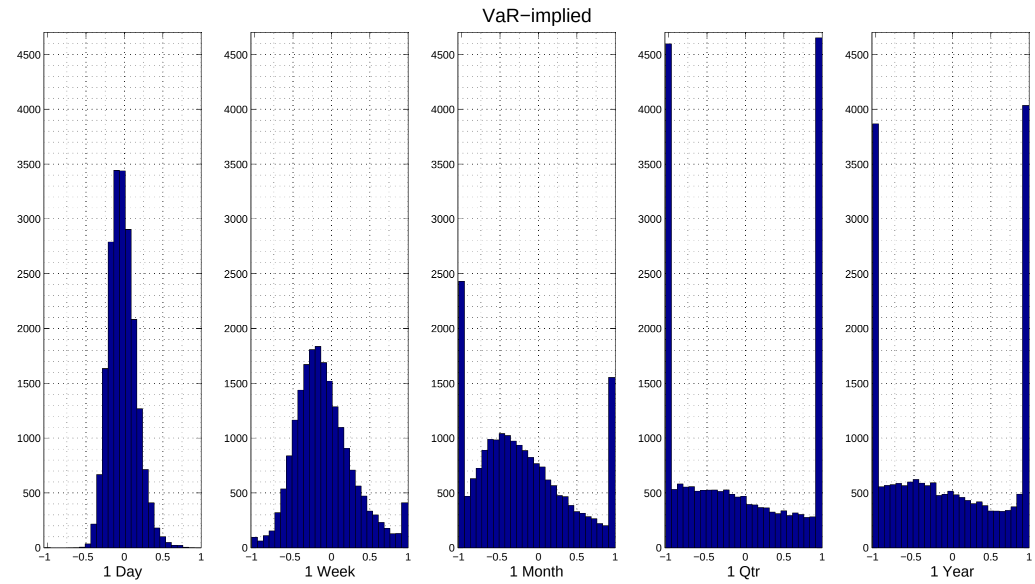

Figure 21: Histograms of VaR–implied tail–correlation estimates with growing aggregation windows, w, for t–distributed daily returns for $\nu = 1$ degrees of freedom and $\rho = 0.0$

there are virtually no common tail observations left given a sample size of "only" 40 years.

The scatter plots for a typical simulation–run for daily and annualized returns and the 40–year sample, shown in Figure 23, illustrate this phenomenon. For the daily returns (left graph) we observe, for a given tail–quantile, a relatively large number of common tail observations and that there is ellipticity. For the annualized data (right graph), although visual inspection suggests some form of negative dependence, both the ellipticity and the common tail behavior disappear. For each risk factor, the annualized returns exhibit maximum losses of about –45%. However, there are no observations in the joint tail region $\{R_1, R_2 \colon R_1 \leq -30\%, R_2 \leq -30\%\}$.

For the annualized returns, the bias remains even when having 4,000 years of data (see bottom Figure 22). The λ–curve shifts slightly upward, but stays well below the theoretical value of 0.2532—especially, in the far–tail with an estimate of about 0.05. The confidence band narrows substantially and includes zero only in the far tail.

In view of the results of this simulation experiment, it is evident that rolling–window annualization can eliminate virtually all tail dependence that is present in the original data. The hope of capturing the dependence between non–normally distributed asset classes more adequately by estimating tail–dependence coefficients, as expressed in Paragraphs 3.1255 and 3.1256 in CEIOPS (2010), is likely to be disappointed when the analysis is based on data subjected to a rolling–window

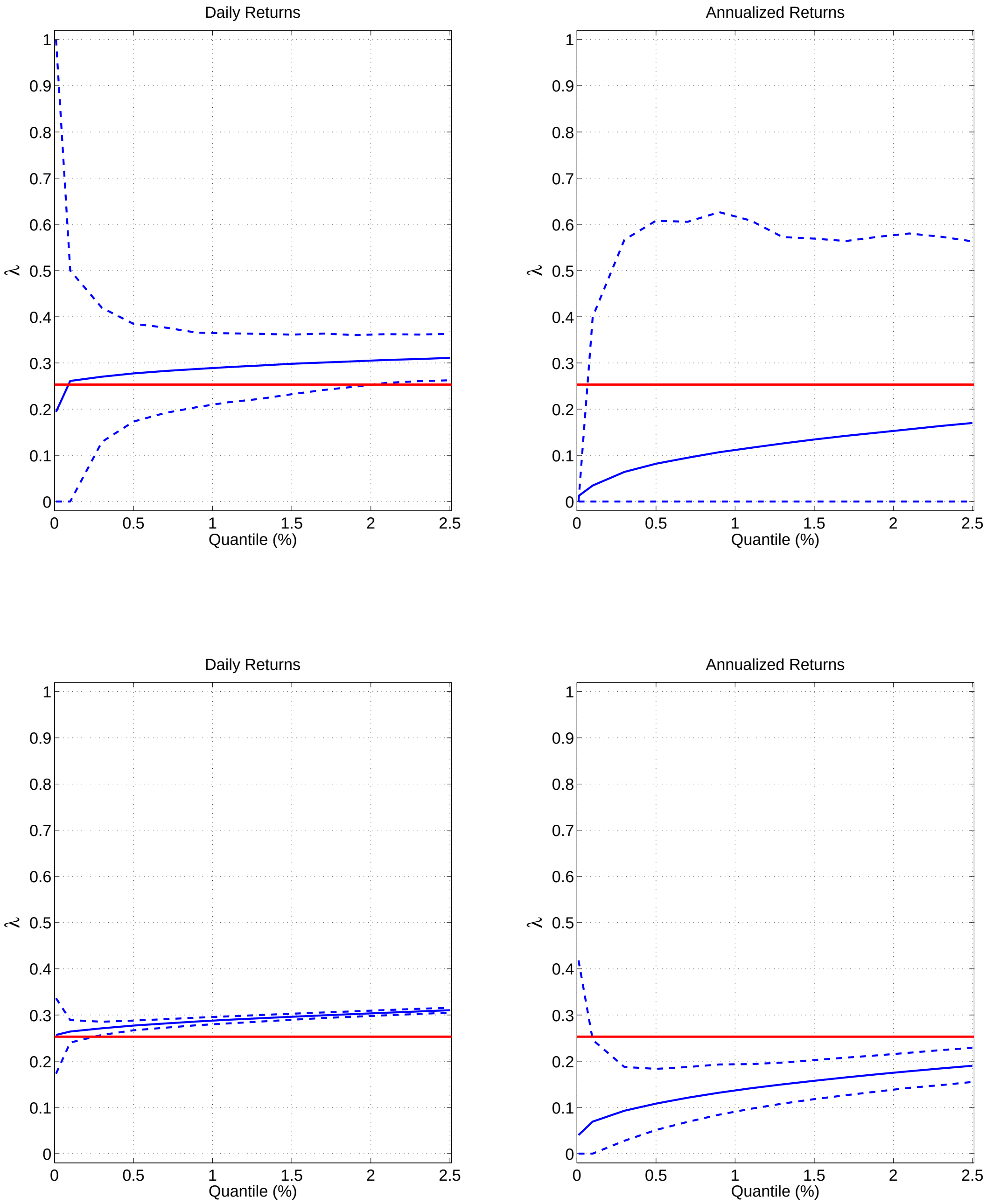

Figure 22: Mean of estimated tail–dependence coefficients (solid curves) from daily (left panel) and annualized (right panel) returns generated from bivariate t–distribution ($\rho = 0.5$ and $\nu = 4$) from a 40–year sample (top) and a 4,000–year sample (bottom) together with 95% confidence bands (dashed)

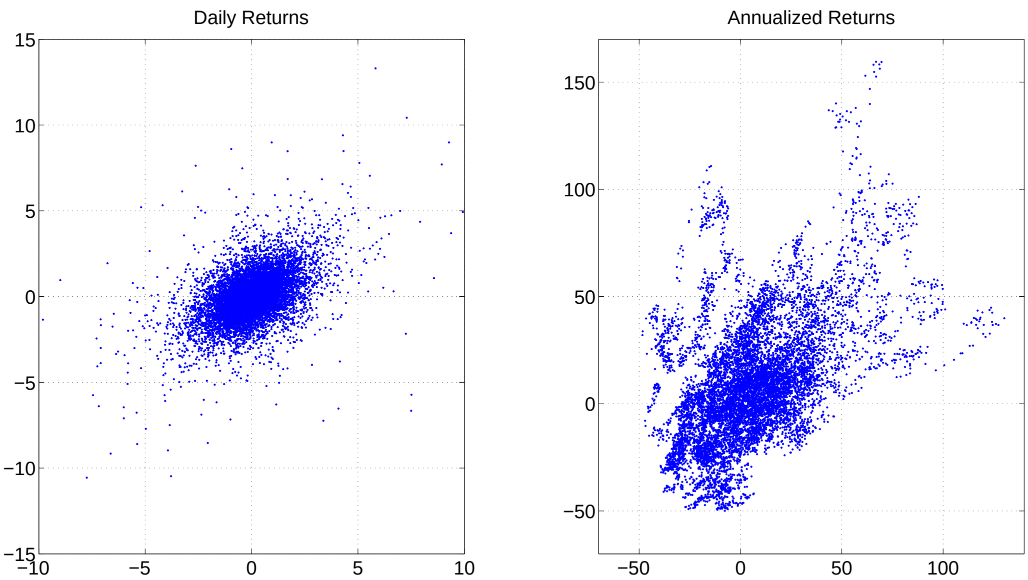

Figure 23: Scatter plots of daily returns (left) generated from bivariate t–distribution ($\rho = 0.5$ and $\nu = 4$) and annualized returns (right) from a 40–year sample

annualization.

5 Conclusions

Given the significant role the insurance industry has in its own right and its relevance for both the financial and the real sector of developed economies, prudent risk–assessment processes that ensure insurers' solvency are of paramount importance. With its Standard Formula, CEIOPS has set up a systematically structured procedure for measuring and aggregating the risks 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 "perfect" design has been achieved. But does the Standard Formula, as currently proposed, represent an overall improvement towards a prudent regulation of the insurance industry? Does it come close to meeting the objective of CEIOPS' successor organization EIOPA, namely[38] "...to ensure that Solvency II is designed in the most appropriate manner..."? The results of this study strongly suggest that an implementation of the Standard Formula with its currently proposed equity–risk calibrations is imprudent, if not irresponsible.

Criticism against Solvency II calibrations has been raised before, arguing, for ex-

[38]See Page 5, EIOPA (2011).

ample, that the Standard Formula is unstable with respect to distributional settings (Pfeifer and Strassburger, 2006) or that indices chosen to represent particular eqity classes are inappropriate (Aria et al., 2010). The problems detailed here are, however, more fundamental. By subjecting historical market data to a rolling–window annualization procedure prior to performing the calibration exercises, virtually all equity–risk calibrations are rendered meaningless.

5.1 Implications for Calibration

The use of overlapping rolling–window return data, when calibrating the equity–risk parameters of the Standard Formula, has serious consequences. The main implications for equity–risk calibrations can be summarized as follows:

1. The annualization induces strong temporal return and risk dependencies in the data, as it induces near–unit–root or random–walk–like characteristics, which are responsible for VaR estimates being highly unstable and erratic over time.
2. The annualization also produces spurious contemporaneous dependence structures between asset classes, substantially deteriorating the accuracy of conventional Pearson–correlation and, even more so, tail–correlation estimates, with the latter playing a prominent role in Solvency II equity–risk calibrations.
3. Altogether, rolling–window annualization leads to highly unreliable input parameters for the Standard Formula, so that capital–requirement estimates for equity risk will be mainly a product of chance rather than a realistic and dependable indication of true risk exposures. This will lead to inefficient and volatile investment decisions.
4. Monitoring market behavior and adapting the input parameters over time, regulators may drastically alter the parameter values and, thereby, cause costly portfolio adjustments, even though the underlying market processes remain unchanged.
5. A disturbing finding is that, if the original data have a weak, positive correlation, tail–correlation estimates from annualized data are likely to be at or near +1 and, thus, tend to greatly exaggerate the presence of dependencies. This pile–up problem at +1 may very well be the reason why QIS calibrations specify a perfect positive correlation between the asset classes within “other equities.”
6. Seemingly contradicting the previous finding, the annualization can also eliminate tail dependence that is present in the data. This results from the fact that rolling–window annualization destroys (near-)ellipticity in the data—a property which defines the joint tail behavior and that both the VaR–implied tail–correlation and the Standard Formula require to justify their use.

This last issue draws attention to a fundamental inconsistency in the Solvency II approach to equity–risk calibration. 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, for risk assessment.

5.2 Implications for Economic Growth and Systemic Risk

The central objective of the Solvency II regulation is that (Paragraph 65, European Parliament (2009))

> *...economic capital should be calculated on the basis of the true risk profile...taking account of the impact of possible...diversification effects.*

As they stand, Solvency II equity–risk calibrations fall critically short of this goal. Their application is likely to have a number of potentially far–reaching implications, affecting not only the insurance industry, but also the real economy—both on the European and the global level. Some of the expected consequences will be:

1. Setting the correlations among all "other equities" equal to +1 completely rules out the possibility of taking diversification effects into account. This will not only overstate risk but also eliminate the incentive to diversify among such heterogeneous asset classes like emerging–market stocks, private equity, hedge fonds, and commodities.
2. By neglecting diversification benefits, the group of "other equities" as a whole becomes less attractive. As a result, long–term growth in Europe will be negatively affected as, for example, lower private–equity investments will reduce the funding of innovative, high–growth firms in the EU. Similarly, a drop in emerging–market investments will hamper economic growth in these countries and, ultimately, feed back negatively to Europe.
3. These negative effects will be further amplified by imposing a rather high correlation of +0.75 between "other" and "global equity." This, together with other, non–equity–risk calibration choices in Solvency II, will divert investments from the private sector to funding EU public debt. This is likely to further reduce long–term growth—unless, of course, the additional funds will be used for public investments that are more productive than those of the private sector.
4. A shift from equity and corporate bonds to EU government bonds will lower an insurer's capital requirements, as the latter are "calibrated" to have zero capital requirements. But it will not necessarily entail an equivalent reduction

in risk exposure. This will only be the case if the default rates of government bonds are in line with the Solvency II framework. In other words, governments also need to apply the EU's Solvency II standards and adopt the $\text{VaR}_{99.5}$ risk concept in their own budget planning "in order to ensure that ... [they] will still be in a position, with a probability 99.5%, to meet their obligations ... over the following 12 months" (Paragraph 65, European Parliament (2009)). As long as this is not the case, prudent regulation has to make sure that the risk of government bonds enters capital–requirement calculations.

5. There will be a substantial increase in systemic risk, if the Standard Formula systematically steers investments to an asset class whose risk parameters are—whether by design or by accident—kept artificially low and where there is only a small number of counterparties.

In view of the calibration deficits presented here and their far–reaching consequences, the implementation of the Solvency II framework in its current form needs to be postponed until the equity–risk calibrations have been fundamentally repaired. In the same vain, there should be no considerations at the moment to extend Solvency II–type regulation to European pension fonds.

Calibration results involving data subjected to rolling–window annualization have to be reexamined with non–overlapping return data at a daily or weekly frequency, in order obtain more reliable parameter estimates. To derive *annualized* SCRs, the calibration approach needs to also capture temporal dependence structures, so that risk can be aggregated over time. Clearly, the latter is not a trivial task and requires additional efforts. But, given the stakes involved, the necessary resources appear negligible in comparison.

Implementing Solvency II without proper equity–risk calibrations and attempting to recalibrate "on the fly," with the regulation already being online, is likely to produce volatile and unreliable SCR estimates and will sooner rather than later trigger calls for Solvency III. If it gets that far, however, other fundamental flaws[39] could be tackled.

[39] One major flaw is the use of $\text{VaR}_{99.5}$ for measuring an insurer's risk exposure and determining the appropriate capital charge. This makes the empirical validation of risk parameters virtually impossible and ignores potential losses that may be less extreme but more frequent than those associated with once–in–two–hundred–years events. Focusing on such extreme risks is like asking a medical doctor to determine the right dosage of treatment and providing her with a thermometer that only indicates temperatures of 42 degrees Celsius and higher.

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. M., 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. F., 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 Parliament and of 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. W. J., Newbold, P., 1974. Spurious Regressions in Econometrics. Journal of Econometrics 2, 111–120.

Granger, C. W. J., 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. C. B., 1987. Time Series Regressions with a Unit Root. Econometrica 55, 277–301.

Rosenbaum, S., 1961. Moments of a Truncated Bivariate Normal Distribution. Journal of the Royal Statistical Society, Series B 23, 405–405.

van Oordt, M. R. C., Zhou, C., 2011. The Simple Econometrics of Tail Dependence. DNB Working Paper, De Nederlandsche Bank, No. 296, May 2011.

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

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 multiplicatively 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. The discrete return over the period $(t-1, t]$, denoted by R_t is given by[40]

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

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

 doi:10.20944/preprints201607.0021.v1

and the continuous return, denoted by r_t, by

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

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.

Assuming 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 $w \geq 1$ 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 are obtained 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–window returns via (19).

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

Continuous rolling–window returns over horizon w are given by

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

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, stationary.[41] However, as w

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

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 of 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.