Institute and Faculty of Actuaries

Calibration of VaR models with Overlapping Data

by the Extreme Events Working Party:

Ralph Frankland, Andrew D Smith, James Sharpe, Rishi Bhatia, Stuart Jarvis, Parit Jakhria and Gaurang Mehta

25 March 2019
Edinburgh

Presented to the Institute & Faculty of Actuaries

Calibration of VaR models with Overlapping Data

Ralph Frankland, Andrew D Smith, James Sharpe, Rishi Bhatia, Stuart Jarvis, Parit Jakhria and Gaurang Mehta

Contents

1 Abstract

Under the European Union's Solvency II regulations, insurance firms are required to use a one-year VaR (Value at Risk) approach. This involves a one-year projection of the balance sheet and requires sufficient capital to be solvent in 99.5% of outcomes. The Solvency II Internal Model risk calibrations require annual changes in market indices / term structure for the estimation of the risk distribution for each of the Internal Model risk drivers. This presents a significant challenge for calibrators in terms of:

- Robustness of the calibration that is relevant to the current market regimes and at the same time able to represent the historically observed worst crisis;
- Stability of the calibration model year on year with arrival of new information;

The above points need careful consideration to avoid credibility issues with the SCR calculation, in that the results are subject to high levels of uncertainty.

For market risks, common industry practice to compensate for the limited number of historic annual data points is to use overlapping annual changes. Overlapping changes are dependent on each other and this dependence can cause issues in estimation, statistical testing and communication of the uncertainty levels around the risk calibrations.

This paper discusses the issues with the use of overlapping data when producing risk calibrations for an Internal Model. A comparison of the overlapping data approach with the alternative non-overlapping data approach is presented. The comparison is made by comparing the bias and mean square error of the first four cumulants under four different statistical models. For some statistical models it is found that overlapping data can be used with bias corrections to give similarly unbiased results as non-overlapping data; but with significantly lower mean square errors. For more complex statistical models (e.g. GARCH) it is found that published bias corrections for non-overlapping and overlapping data sets do not result in unbiased cumulant estimates and/or leads to increased variance of the process.

In order to test the goodness of fit of probability distributions to data sets it is common to use statistical tests. Most such tests do not function when using overlapping data as overlapping data breaches the independence assumption underlying most statistical tests. We present and test an adjustment to one of the statistical tests (the Kolmogorov Smirnov goodness-of-fit test) to allow for overlapping data.

Finally, we explore methods of converting "high" frequency (e.g. monthly data) to "low" frequency data (e.g. annual data). This is an alternative methodology to using overlapping data and the approach of fitting a statistical model to monthly data and then using the monthly model aggregated over twelve-time steps to model annual returns is explored. There are a number of methods available for this approach. We explore two of the widely used approaches for aggregating the time series.

2 Executive Summary

2.1 Overview

Under the European Solvency II regulations insurance firms are required to calculate a one-year Value at Risk (VaR) of their balance sheet to a 1 in 200 level. This involves a one-year projection of a market consistent balance sheet and requires sufficient capital to be solvent in 99.5% of outcomes. In order to calculate 1-year 99.5th percentile VaR, a significant volume of 1-year non-overlapping data is needed. In practice there is often a limited amount of relevant market data for market risk calibrations and an even more limited reliable and relevant data history for insurance / operational risks.

Two of the key issues with the available market data are:

- The dataset available may be relatively longer (e.g. for corporate credit spread risk, Moody's Default and Downgrade data are available from 1919[1]), but data may not be directly relevant or not granular enough for risk calibration.
- Dataset may be very relevant to the risk exposure and granular as required, but data length is not sufficient, e.g. for corporate credit spread risk, Merrill Lynch or Iboxx data are available from 1996 or 2006 respectively.

As a consequence, practitioners need to make expert judgements about whether to:

- use overlapping data or non-overlapping data. If overlapping data is used then is there any adjustment that can be made to the probability distribution calibrations and statistical tests to ensure that the calibration is still fit for purpose; or
- use non-overlapping data with higher frequency than annual (e.g. monthly) and extract the statistical properties of this data which can allow us to aggregate the time series to lower frequency (e.g. annual) time series.

In section 4 of this paper we consider adjustments to correct for bias in probability distributions calibrated using overlapping data. In section 5 adjustments to statistical tests are defined and tested. In section 6, the issues with using data periods shorter than a year and then aggregating to produce annualised calibrations are considered.

2.2 Calibrating probability distributions using overlapping data

Section 4 discusses the issues of probability distribution calibration using overlapping data. Adjustments to probability distribution calibrations using overlapping data in academic literature are presented and tested in a simulation study. We analysed the impact on cumulant bias and mean squared (MSE) for some of the well-known statistical processes, namely Brownian, Normal Inverse Gaussian (special case of Levy process), ARMA and GARCH processes under both overlapping and non-overlapping data approaches. We have analysed the impact on cumulants after applying corrections outlined by Heng/Sun/Nelken (2009) and Cochrane (1988). The simulation study involves producing computer generated data and comparing the different approaches to estimating the known values of the cumulants.

Cumulants are similar to moments and are properties of random variables. The first three cumulants: mean, variance and skewness are well-known and the same as the first three central moments. The fourth cumulant is the fourth central moment minus 3*variance^2. The cumulants (and derived moments) are widely used in calibrating the probability distribution using Method of Moments style calibration approaches. As with moments, the cumulants

[1] Note: Many datasets may be available for circa 100 years which could be considered sufficient, however it is possible that there exists materially wide confidence interval in the 1-in-200 point.

uniquely define the calibration of a probability distribution within particular parameterised distribution families.

The key conclusions from the simulation study for the processes outlined above are:

- Using published bias adjustments both overlapping and non-overlapping data can be used to give unbiased estimates of statistical models where monthly returns are not auto correlated. Where returns are auto correlated bias are more complex for both overlapping and non-overlapping data.
- In general, overlapping data is more likely to be closer to the exact answer than non-overlapping data. By using more of the available data, overlapping data generally gives cumulant estimates with lower mean square error than using non-overlapping data.

2.3 Statistical tests using overlapping data

In section 5, we have defined and tested an adjustment to a statistical test to allow for overlapping data.

To test whether a probability distribution fitted to a data set is a good fit to the data, it is common to apply a statistical test. Many statistical tests have an assumption that the underlying data is independent, which is clearly not the case for overlapping data.

Using the Kolmogorov Smirnov (KS) statistical test an adjustment is proposed to this test to allow for overlapping data. This adjustment is tested using simulated data and the results presented.

The results of the testing indicate the proposed adjustment for overlapping data to the KS test has a rejection rate consistent with the test functioning as intended.

2.4 Alternative to annual data

An alternative to using annual data is to use a higher frequency "monthly" data and then "annualise" it (i.e. convert the results from the monthly data into annual data results). The issues with this approach are considered in section 6. Higher frequency data has the advantages of having more data points and no issues with overlapping data. The main disadvantage is the non-annual data needs to be annualised, which comes with its own limitations. We have considered the following possible solutions:

- Use of non-overlapping monthly data and annualise using empirical correlation that is present in the time series (see section 6.2 for further details). The key points to note from the use of annualisation are:
  - This technique involves:
    - fitting a probability distribution to monthly data;
    - simulating a large computer-generated data set from this fitted model/distribution; and
    - aggregating the simulated monthly returns into annual returns using a copula or other relevant techniques.
  - It utilises all the data points and therefore would not miss any information that is present in the data and in absence of information on the future data trends would lead to a more stable calibration overall.

    - In the dataset we explored, it improves the fits considerably in comparison to non-overlapping data or monthly annual overlapping data because of the large simulated data used in the calibration.
    - However, it does not remove the autocorrelation issue completely (as monthly non-overlapping data or for that matter any "high" frequency data could be autocorrelated) and does not handle the issues around volatility clustering.
- Use of statistical techniques such as "temporal aggregation" (section 6.4). The key points to note from the use of temporal aggregation are:
    - Temporal aggregation involves fitting a time series model to monthly data; then using this time series model to model annual data;
    - It utilises as much data as possible without any key events being missed;
    - It improves the fit to the empirical data and leads to a stable calibration;
    - It can handle data with volatility clustering and autocorrelation;
    - However, it suffers from issues such as possible loss of information during the increased number of data transformations and is complex to understand and communicate to stakeholders.
- Use of autocorrelation adjustment (or "de-smoothing" the data). This technique is not covered here as this is a widely researched topic (Marcatoo, 2003). However, a similar technique by (Heng Sun Nelken, 2009) has been used in section 4which corrects for bias in the estimate of the variance of the data.

2.5 Conclusions

The key messages concluded from this paper are:

- There is a constant struggle between finding relevant data for risk calibration and sufficient data for a robust calibration;
- Using overlapping data is acceptable for Internal Model calibration, however communication of uncertainty in the model and parameters to the stakeholder is important;
- There are some credible alternatives to using overlapping data such as temporal aggregation and annualization, however these alternatives bring their own limitations and understanding of these limitations is key to using these alternatives. We recommend to considering the comparison of the calibration using both non-overlapping monthly data annualised with overlapping annual data and discuss the advantages, robustness and limitations of both the approaches with stakeholders before finalising the calibration approach.

2.6 Future Work

Further efforts are required in the following areas:

- Diversification benefit using internal models is one of the key discussion topics amongst industry participants. So far, we have only analysed univariate time series. Further efforts are required in terms of analysing the impact of overlapping data on covariance and correlation properties between two time-series.

- Similarly, the impact on statistical techniques such as dimension reduction techniques (e.g. PCA) needs investigation. Initial efforts can be made in terms of treating each dimension as a single univariate time series and apply various techniques such a temporal aggregation or annualisation and apply dimension reduction techniques on both overlapping and non-overlapping transformed data sets to understand the impact.
- The impact on statistical tests other than the KS test have not been investigated. Nor has using different probability distributions than the Normal distribution for the KS test. Both of these areas could be investigated further using the methods covered in this paper.
- Measurement of parameter and model uncertainty in the light of new information has not been investigated either for “annualization” method or for “temporal aggregation” method.

3 Overlapping Data: Econometric Literature Survey

Within the finance literature, many authors have confronted the issue of having a scarcity of data with which to calibrate a multi-period econometric model. Several approaches have been developed which justify the use of historic observation periods that are overlapping. These approaches extend classical statistical theory, which often presumes that the various observations are independent of each other. It is often the case that the naïve statistics (constructed ignoring the dependence structure) are consistent (asymptotically tend towards the true parameters) like their classical counterparts but the standard errors are larger.

Hansen & Hodrick (1980) examined the predictive power of 6-month forward FX rates. The period over which a regression is conducted is 6 months - yet monthly observations are readily available but clearly dependent. They derive the asymptotic distribution of the regression statistics using the Generalized Method of Moments (GMM, Hansen (1982)) which does not require independent errors. The regression statistics are consistent and GMM provides a formula for the standard error. This approach has proved influential, and several estimators have been developed for the resulting standard error: Hansen-Hodrick's original, Newey & West (1987) and Hodrick (1992) being prominent examples. Newey & West errors have become the most commonly used in practice.

However, the derived distribution of fitted statistics is only true asymptotically and the small-sample behaviour is often unknown. Many authors use bootstrapping or Monte Carlo simulation, to help assess the degree of confidence to attach to a specific statistical solution. For example, one prominent strand of the finance literature has examined the power of current dividend yields to predict future equity returns. Ang & Bekaert (2006) and Wei & Weight (2013) show using Monte Carlo simulation that the standard approach of Newey & West errors produces a test size (i.e. probability of a Type I error) which is much worse than when using Hodrick (1992) errors.

In addition to the asymptotic theory, there has also been work on small-sample behaviour. Cochrane (1988) examines the multi-year behaviour of a time series (GNP) for which quarterly data is available. He calculates the variance of this time series using overlapping time periods and computes the adjustment factor required to make this calculation unbiased in the case of a random walk. This adjustment factor generalises the n-1 denominator Bessel correction in the non-overlapping case. Kiesel, Perraudin & Taylor (2001) extend this approach to third and fourth cumulants.

Müller (1993) conducts a theoretical investigation into the use of overlapping data to estimate statistics from time series. He concludes that while estimation of the sample mean is not improved by using overlapping rather than non-overlapping data, if the mean is known then the standard error of sample variance can be reduced by about 1/3 when using overlapping data. His analysis of sample variance is extended to the case of unknown mean, again with improvements of about 1/3, by Heng, Sun, Nelken et al (2009). Heng, Sun, Nelken et al also suggests an alternative approach of using the average of non-overlapping estimates. Like Cochrane and Müller, this leads to a reduction in variance of about 1/3 compared to using just non-overlapping data drawn from the full sample.

Efforts have been made to understand the statistical properties and / or behaviour of the "high" frequency (e.g. monthly or daily data points) time series data to transform them into "low" frequency time series data (i.e. annual data points) via statistical techniques such as temporal aggregation. Initial efforts were made to understand the temporal aggregation of ARIMA processes and (Amemiya&Wu, 1972) lead the research in this area. (Feike_Drost_Nijman, 1993) developed the closed form solutions for temporal aggregation of GARCH processes and described relationships between various ARIMA processes under "high" frequency and their transformation under “low” frequency time series. (WSChan_etal, 2008) shows various aggregation techniques using equity returns (S&P500 data) and its impact on real-life situations.

The Method of moments is not the only, and not necessarily the best, method for fitting distributions to data, with maximum likelihood being an alternative. There are some comparisons within the literature; we note the following points:

- Maximum likelihood produces asymptotically efficient (lowest mean-squared error) parameter estimates, while in general the method of moments is less efficient.
- Model mis-specification is a constant challenge, whatever method is used. Within a chosen distribution family, the moments may determine a distribution, but other distributions with the same moments, from a different family, may have different tail behaviour. For moment-based estimates, Bhattacharya’s inequality constrains the difference between two distributions with shared fourth moments, while as far as we know there are no corresponding results bounding mis-specification error for maximum likelihood estimates.
- The method of moments often has the advantage of simpler calculation, and easy verification that a fitted distribution indeed replicates sample properties.
- The adaptation of the maximum likelihood method to overlapping data does not seem to have been widely explored in the literature, while (as we have seen) various overlapping corrections have been published for method-of-moments estimates. For this reason, in the current paper, we have focused on moments / cumulants.

4 Simulation study – overlapping versus non-overlapping

4.1 Background

In this section, the results from a simulation study of the bias and mean square error present in the cumulant estimation using annual overlapping and non-overlapping data are presented. Cumulants are similar to moments (the first three cumulants are the mean, variance and skewness and are exactly the same for moments); further information about cumulants is given in the appendix.

Using a methodology outlined in Jarvis et al 2016 monthly timeseries data are simulated from a known distribution (Reference model) for a given number of years data. The first four cumulants based on annual data are then calculated by considering both non-overlapping annual returns as well as overlapping annual returns (overlapping by 11 out of 12 months). By comparing the results of these with the known cumulant values of the Reference model and averaging across 1000 simulations, the bias and mean square errors of the estimates can be compared. This study is carried out using four different Reference models:

- Brownian Process
- Normal Inverse Gaussian Process
- ARMA Process
- Garch[2] Process

A high-level description of this process is:

- Simulate a monthly time series of n years data from one of the four processes above;
- Calculate annual returns using overlapping and non-overlapping data;
- Calculate the first four cumulants of the annual returns (for overlapping and non-overlapping data);
- Compare the estimated cumulants with the known cumulants;
- Repeat 1000 times to estimate the bias and mean square error of both overlapping and non-overlapping data.

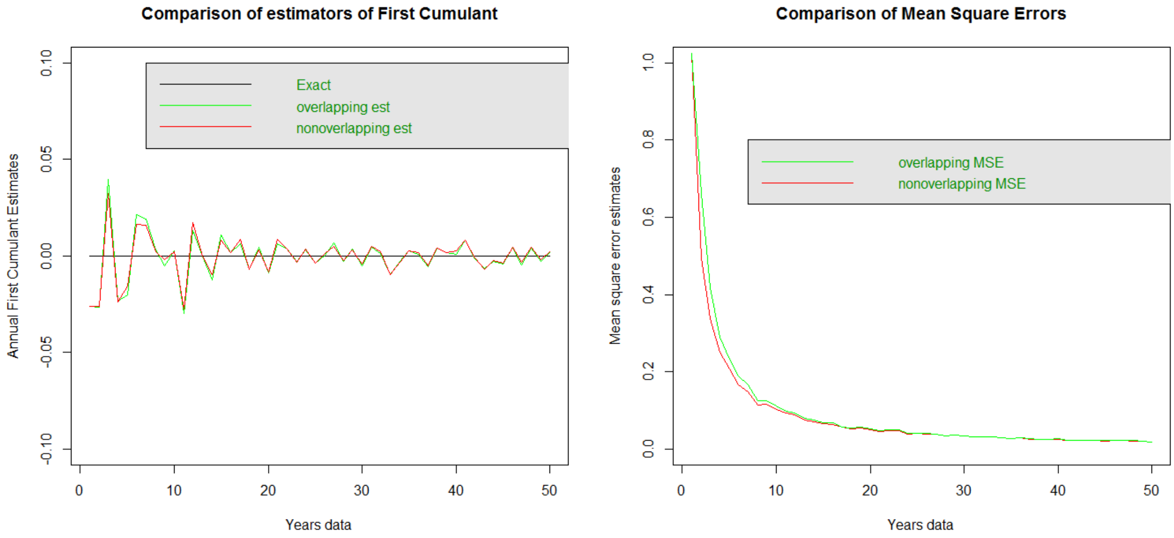
The analysis has been carried out for all years up to year 50 and the results are shown below. The results for the ARMA and Normal Inverse Gaussian are in Appendix B.

4.2 Brownian process results

This section shows the bias and mean square errors for the first four cumulants.

[2] Garch (p,q) model specification is calibrated by making sure that |p+q|<1to ensure the time series remains stable.

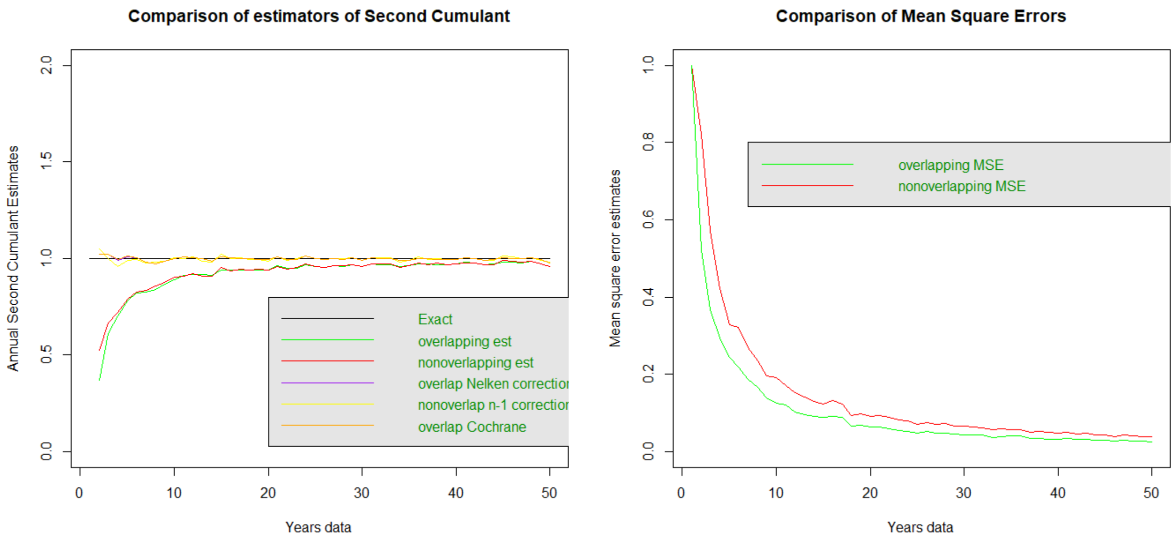
4.2.1 First cumulant – the mean



The plots above show the bias in the plot on the left and the mean square error on the plot on the right. The overlapping and non-overlapping data estimates of the mean appear very similar and not obviously biased. They also have very similar mean square errors across all years.

4.2.2 Second cumulant – variance

The second cumulant is the variance (with divisor n).

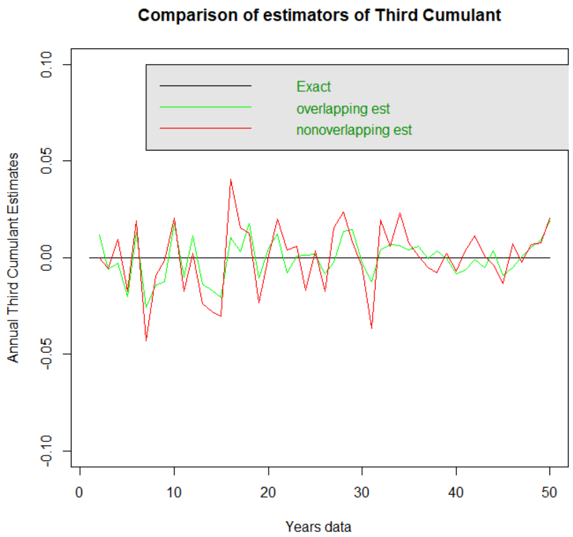


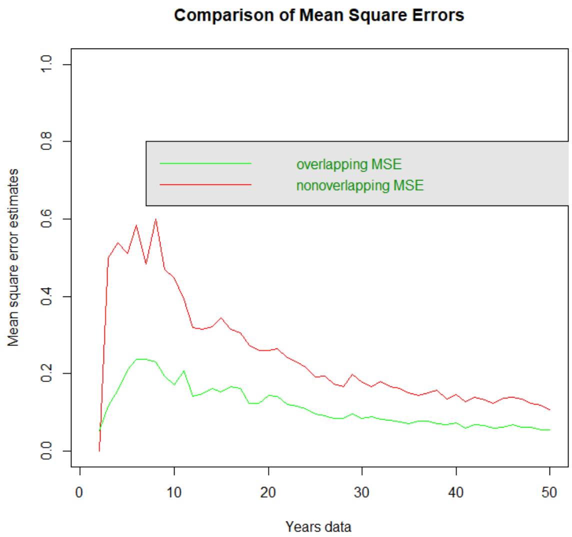
The plot above on the left shows that the overlapping and non-overlapping estimates of the variance (with divisor n) are too low with similar bias levels for all terms. This is more marked the lower the number of years data, and the bias appears to disappear as n gets larger.

The plot on the left also shows the second cumulant but bias corrected, using a divisor (n-1) instead of n for the non-overlapping data and using the formula in (Heng, Sun, Nelken et al) for the overlapping data, as well as the Cochrane adjustment (Cochrane 1988) for overlapping data. Both of these corrections appear to have removed the bias across all terms for overlapping and non-overlapping data.

The plot on the right shows the mean square errors for the two approaches, with overlapping data appearing to have lower mean square errors for all terms.

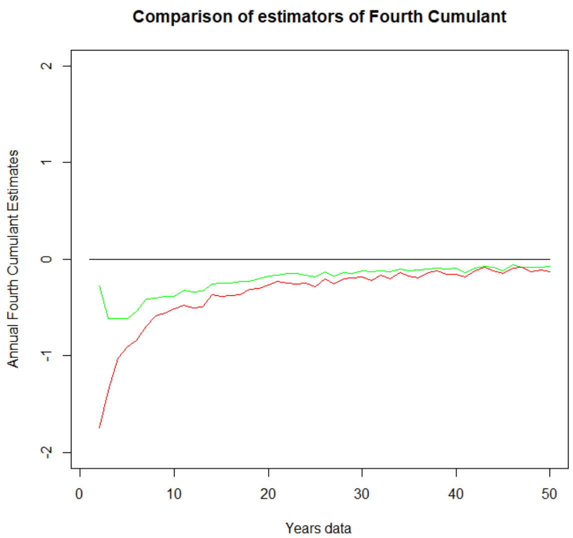
4.2.3 Third cumulant





Neither approach appears to have any systemic bias for the third cumulant. The mean square error is significantly higher for non-overlapping data than overlapping data.

4.2.4 Fourth cumulant

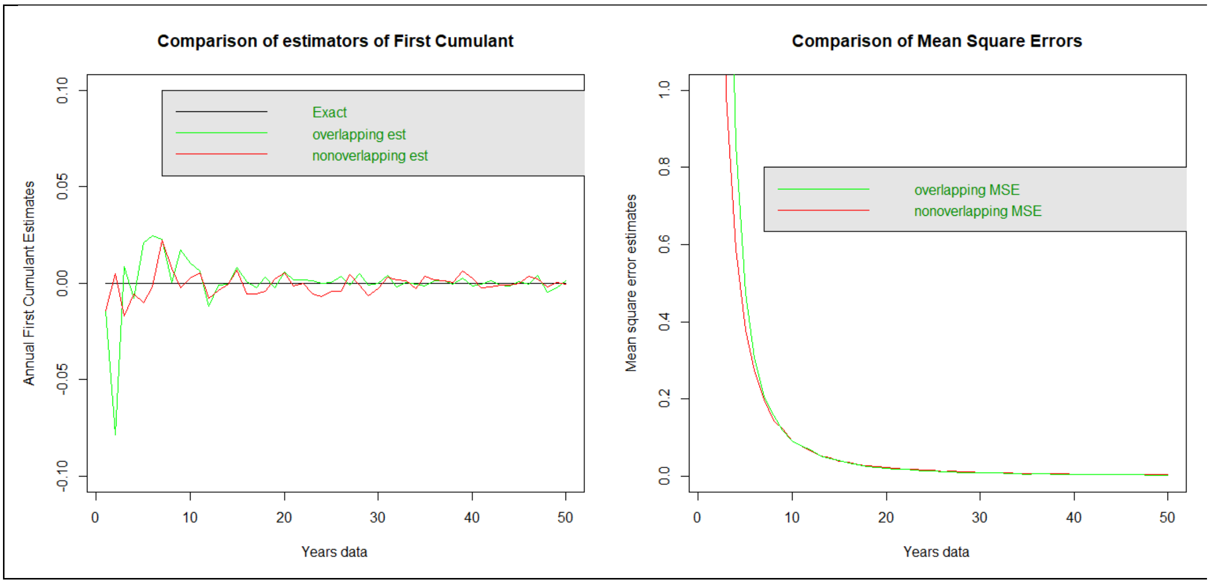


In this case the non-overlapping data appears to have a higher downward bias than overlapping data at all terms; both estimates appear biased. The non-overlapping data has higher mean square error than the overlapping data.

Plots of the bias and mean square error for the Normal Inverse Gaussian are given in Appendix B. They are very similar to those of the Brownian process.

4.3 Garch results

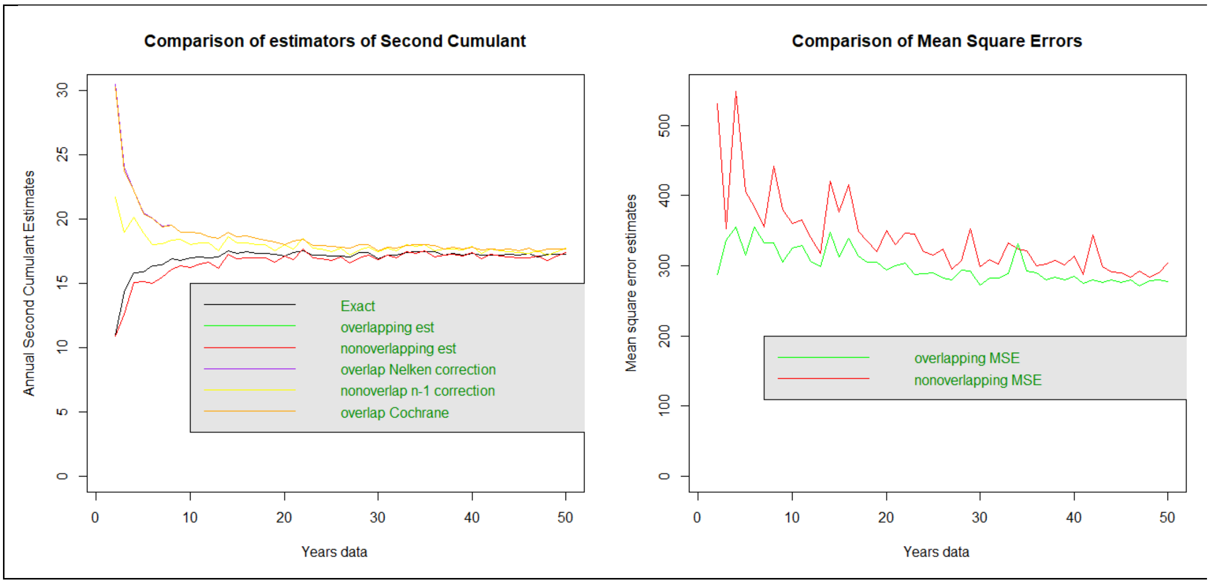
4.3.1 First cumulant – the mean



The plots above show the bias in the plot on the left and the mean square error on the plot on the right. The overlapping and non-overlapping data estimates of the mean appear very similar after 20 years. Below 20 years, the data does show some bias under both overlapping and non-overlapping. They have very similar mean square errors after 10 years and below 10 years non-overlapping data has comparatively marginally lower MSE.

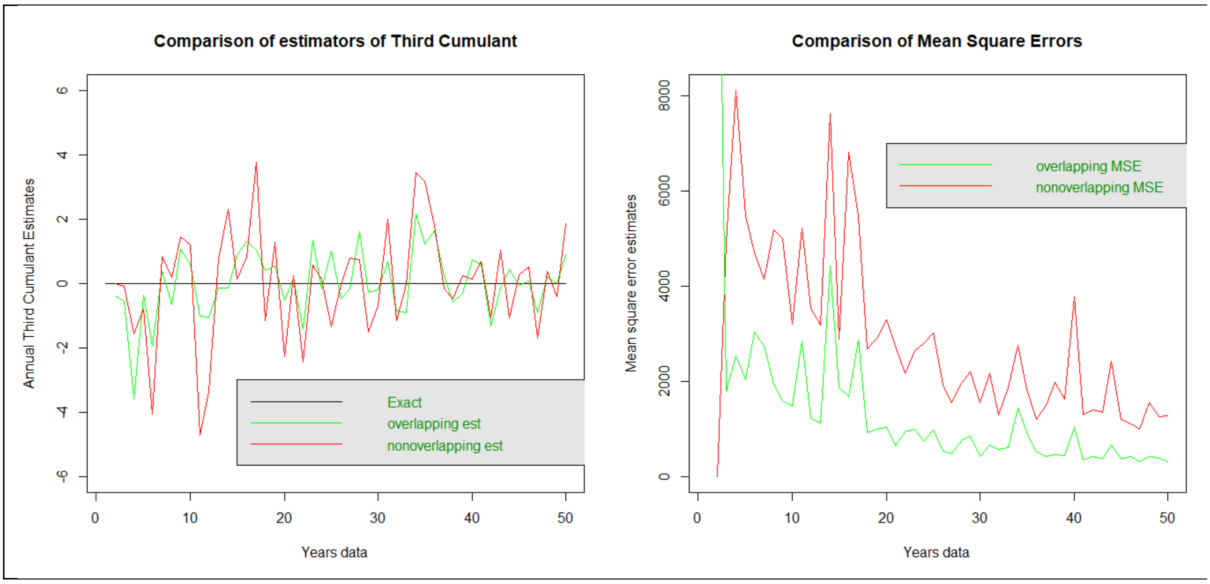
4.3.2 Second cumulant – variance

The second cumulant is the variance (with divisor n).



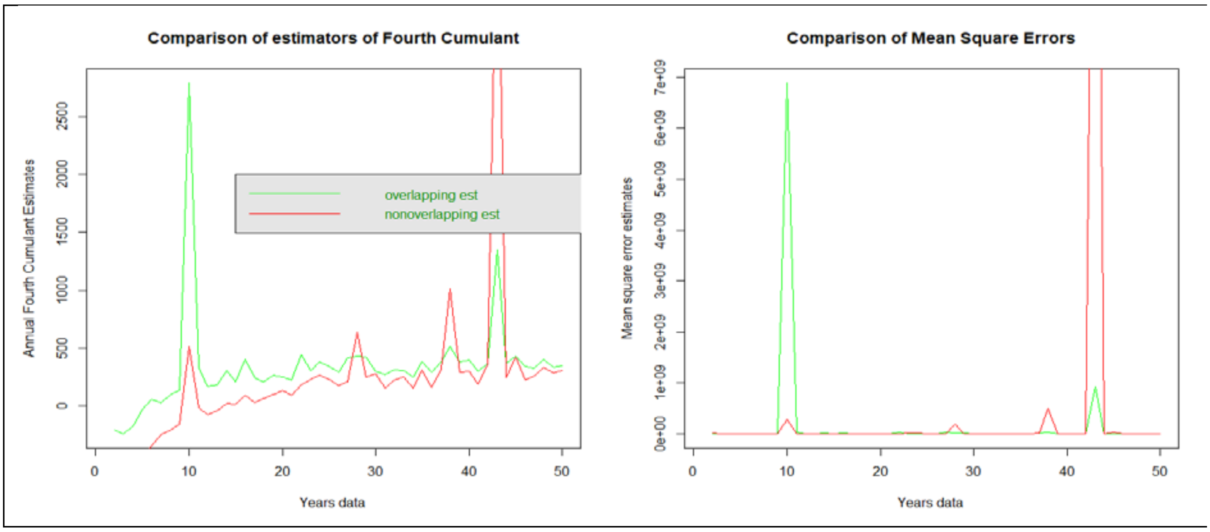
Both approaches have similar level of bias, particularly when available data is limited. The bias corrections now both overstate the variance particularly strongly for data sets with less than ten years data. The MSE for overlapping data appears to be materially lower than non-overlapping data.

4.3.3 Third cumulant



The non-overlapping data has both higher bias and MSE compared to overlapping data across all years.

4.3.4 Fourth cumulant



Non-overlapping data has lower bias compared to overlapping data, however overlapping data has lower MSE.

4.4 Discussion of the simulation results

The results above show the bias and mean square errors for the first four cumulants of each of the reference distributions.

For the first cumulant the results are similar for all four reference distributions tested. There is no obvious bias for either the non-overlapping or the overlapping data series. The mean square error of the overlapping and non-overlapping series is at a similar level for both. These results show that for estimation of the first cumulant both approaches perform similarly on the bias and mean square error tests and there is no need for any correction for bias.

For the second cumulant the results vary for different reference models.

- For the Brownian and Normal Inverse Gaussian reference models, the non-overlapping and overlapping are both downwardly biased to a similar extent. In both cases the bias can be corrected for by using the Bessel correction for the non-overlapping data, and the Cochrane (1988) or Heng Sun Nelken (2009) corrections for overlapping data. The mean square error is lower for the overlapping data (due to the additional data included). These results suggest that for the estimation of second cumulant the overlapping data performs better due to the lower mean square error, and so greater likelihood to be nearer to the true answer.

5 Statistical tests using overlapping data

In this section the use of statistical tests with overlapping data is discussed. An adjustment to a statistical test to allow for the use of overlapping data is proposed. This adjustment is then tested using computer generated data.

5.1 Statistical tests

When fitting a probability distribution to a data set, it is common practice to assess its goodness of fit using a statistical test, such as the Chi Squared test (Bain et al 1992 p453), Anderson Darling test (Bain et al 1992 p458) or in the case of this paper the Kolmogorov Smirnov test (Bain et al 1992 p460).

The Kolmogorov Smirnov test is based on calculating the largest Kolmogorov distance, which is the distance between a single data point and the points' projected position on the probability distribution fitted to the data.

The Kolmogorov Smirnov test is intended to compare the underlying data against the distribution the data came from where the parameters are known. If the parameters themselves tested against in the KS test have been estimated from the data this introduces sample error. This sample error is not allowed for in the standard KS test and a sample error adjustment is required which in the case of the Normal distribution is known as the Lillifors adjustment (Conover 1999).

A description of this adjustment for sampling error for data from a Normal distribution is given below:

1. Fit a Normal distribution to the data set of n data points and calculate the parameters for the Normal distribution;
2. Measure the Kolmogorov distance for the fitted distribution and the data set, call this D;
3. Simulate n data points from a Normal distribution with the same parameters as found in step 1. Re-fit another Normal distribution and calculate the Kolmogorov distance between this newly fitted Normal and the simulated data;
4. Repeat step 3 1000 (or suitably large) times to generate a distribution of Kolmogorov distances;
5. Calculate the percentile the distance D is on the probability distribution calculated in step 4.;
6. If the distance D is greater than the 95th percentile of the probability distribution calculated in step 4, then it is rejected at the 5% level.

The reason this approach works is because the Kolmogorov Smirnov distance is calculated between the data and a fitted distribution and then compared with 1000 randomly generated such distances. If the distance between the data and the fitted distribution is greater than 95% of the randomly generated distances, then there is statistically significant evidence against the hypothesis that the data is from the fitted distribution.

5.2 Adjustment for overlapping data

If overlapping data is used instead of non-overlapping data then even if the non-overlapping data is independent and identically distributed, the overlapping data will not be, as each adjacent overlapping data point will be correlated. This means overlapping data will not satisfy the assumptions required of most statistical tests, such as the KS test.

However, it is possible to adjust most statistical tests to allow for the use of overlapping data. A method for doing so is shown here for the KS test. The approach used is similar to that described above to correct for sampling error; except that both the data being tested and the data simulated as part of the test are overlapping data. The steps are:

1. Fit a Normal distribution to the data set of n overlapping data points and calculate the parameters for the Normal distribution;
2. Measure the Kolmogorov distance for the fitted distribution and the data set, call this D;
3. Simulate n overlapping data points from a Normal distribution with the same parameters as found in step 1. Re-fit another Normal distribution and calculate the Kolmogorov distance between this newly fitted Normal and the simulated data;
4. Repeat step 3 1000 times to generate a distribution of Kolmogorov distances;
5. Calculate the percentile the distance D is on the probability distribution calculated in step 4;
6. If the distance D is greater than the 95th percentile of the probability distribution calculated in step 4, then it is rejected at the 5% level.

A key question is how to simulate the overlapping data in step 3 in the list above. For levy stable processes such as the Normal distribution this can be done by simulating from the Normal distribution at a monthly timeframe and then calculating the annual overlapping data directly from the monthly simulated data. For processes which are not levy stable, an alternative is to directly simulate annual data and then aggregate into overlapping data using a gaussian copula with a correlation matrix which gives the theoretical correlation between adjacent overlapping data points, where the non-overlapping data is independent. This approach generates correlated data from the non-levy stable distribution, where the correlations between adjacent data points are in line with theoretical correlations for overlapping data. (this last approach is not tested below).

This adjustment works for the same reason as the adjustment described in section 5.1. The Kolmogorov Smirnov distance is generated between the data and the fitted distribution. This distance is then compared with 1000 randomly generated distances, except this time using overlapping data. If the distance between the data and the fitted distribution is greater than 95% of the randomly generated distances, then there is statistically significant evidence against the hypothesis that the data is from the fitted distribution.

5.3 Testing the adjustments to the KS test using simulated data

Using the same testing approach applied in section 4, defined in Jarvis et al (2016) the KS test and the adjustments described above have been assessed. This approach to testing involves simulating data from known distributions, then fitting a distribution to the data, carrying out a statistical test and then assessing the result of the test against the known correct answer.

The testing carried out has been:

1. Test of the standard KS test. This is done using non-overlapping simulated data from a Normal distribution with mean 0 and standard deviation 1.
    a. 100 data points are simulated from this Normal distribution.
    b. The KS test is carried out between this simulated data and the Normal distribution with parameters 0 for mean and 1 for standard deviation.
    c. The p-value is calculated from this KS test.
    d. Steps a, b and c are repeated 1000 times and the number of p-values lower than 5% is calculated and divided by 1000.
2. Test of the KS test with sample error. This test is done using non-overlapping simulated data from a Normal distribution with mean 0 and standard deviation 1. The

difference between this test and test 1 is that step b in test 1 is using known parameter values, whereas this test uses parameters from a distribution fitted to the data.

   a. 100 data points are simulated from this Normal distribution
   b. The Normal distribution is fitted to the data using the Maximum Likelihood Estimate (MLE).
   c. The KS test is carried out between the simulated data and the fitted Normal distribution
   d. The p-value is calculated for this KS test
   e. Steps a, b, c and d repeated 1000 times and the number of p-values lower than 5% is calculated and divided by 1000

3. Test of the KS test with correction for sample error. This test is done using non-overlapping simulated data from a Normal distribution with mean 0 and standard deviation 1. The difference between this test and test 2 is that step c is carried out using the KS test adjusted for sample error.
   a. 100 data points are simulated from this Normal distribution
   b. The Normal distribution is fitted to the data using the Maximum Likelihood Estimate (MLE).
   c. The KS test adjusted for sample error (as described in section 5.1) is carried out between the simulated data and the fitted Normal distribution
   d. The p-value is calculated for this KS test
   e. Steps a, b, c and d repeated 1000 times and the number of p-values lower than 5% is calculated and divided by 1000
4. Test of the KS test with correction for sample error applied to overlapping data. This test is done using overlapping simulated data from a Normal distribution with mean 0 and standard deviation 1. The difference between this test and test 3 is that the simulated data in this test is from an overlapping data set.
   a. 100 data points are simulated from this Normal distribution
   b. The Normal distribution is fitted to the data using the Maximum Likelihood Estimate (MLE).
   c. The KS test adjusted for sample error (as described in section 5.1) is carried out between the simulated data and the fitted Normal distribution
   d. The p-value is calculated for this KS test
   e. Steps a, b, c and d repeated 1000 times and the number of p-values lower than 5% is calculated and divided by 1000
5. Test of the KS test with correction for sample error applied to overlapping data; and correction for overlapping data (as described in 5.2). This test is done using overlapping simulated data from a Normal distribution with mean 0 and standard deviation 1. The difference between this test and test 4 is that the KS test corrects for overlapping data as well as sample error.
   a. 100 data points are simulated from this Normal distribution
   b. The Normal distribution is fitted to the data using the Maximum Likelihood Estimate (MLE).
   c. The KS test adjusted for sample error (as described in section 5.1) is carried out between the simulated data and the fitted Normal distribution. This was done with a reduced sample size of 500 in the KS test to improve run times.
   d. The p-value is calculated for this KS test
   e. Steps a, b, c and d repeated 500 times and the number of p-values lower than 5% is calculated and divided by 500

5.4 Results of the simulation study on the KS test

The results of the simulation study are the rejection rate for each statistical test. For data generated randomly from a known distribution tested against a 5% level, we would expect a 5% rejection rate. The results from each of the tests described in 5.3 are:

Test (as described in 5.3)	**Result**
1	4.3%
2	0%
3	5.0%
4	44%
5	5.3%

5.5 Discussion of the results

This section discusses each of the test results presented in section 5.4.

For test 1, the test assesses the rejection rate for the standard KS test applied as it is intended to be applied (i.e. compared against known parameter values). The result of 4.3% compares to an expected result of 5%. This may indicate the standard KS has a degree of bias.

For test 2, the test assesses the rejection rate for the KS test applied using the sample fitted parameters with no allowance for sample error. The rejection rate of 0% indicates that if sample error is not corrected for there is almost no chance of rejecting a fitted distribution.

For test 3, the KS test is now corrected for sample error and the rejection rate of close to 5% indicates the KS test with sample error correction is working as intended.

For test 4, the KS test with the sample error correction, applied to overlapping data. The rejection rate is very high at 44% relative to an expected 5% level. This indicates that applying the KS test with sample error correction to overlapping data will have a much higher rejection rate than expected.

For test 5, the KS test with sample error and overlapping error correction is applied to overlapping data. The 5.3% result of this test (closely in line with expected rate of 5%) indicates the overlapping error correction is working as expected.

This test shows it is possible to achieve a rejection rate in line with expectations by adjusting the KS test for overlapping data as described in section 5.2.

6 Using periods shorter than annual data

So far, we have discussed the issues with using overlapping data for the purpose of risk calibration and possible methods of correcting for the overlapping data including the adjustments made to the data as covered in sections 4 and 5.

Alternatively, the industry participants have tried to use “high” frequency e.g. monthly or quarterly data to get to “low” frequency data, e.g. annual to meet the solvency II requirement of performing a 1-in-200 year calibration over a 1 year period.

In this section, we consider the issues around these alternative approaches where time periods shorter than 1 year are used to derive the annual calibration. This avoids some of the problems with using overlapping data directly, considered in sections 4 and 5. An example of this approach is to fit a model to monthly data, then extend this same model to also model annual returns.

The approaches discussed in this section can be considered as possible alternatives to using annual non-overlapping and / or monthly annual overlapping data. The uncertainties present in the approaches discussed in this section are also considered.

6.1 Approaches using data periods shorter than annual

Three possible approaches to using data periods shorter than a year for the calibration of VaR at an annual time frame are:

- Use of non-overlapping monthly data but annualise them using autocorrelation that is present in the time series (section 6.2);
  - This technique involves fitting a probability distribution to monthly data. Simulation from a large computer-generated data set from this fitted distribution. Aggregating the simulated monthly returns into annual returns using a copula and the correlation;
  - It utilises all the data points and leads to a stable calibration;
  - It improves the fits considerably in comparison to non-overlapping data or monthly annual overlapping data;
  - However, it does not remove the autocorrelation issue completely and does not handle the issues around volatility clustering.
- Use of statistical techniques such as “temporal aggregation” (section 6.3)
  - It involves fitting a time series model to monthly data; then using this time series model to model annual data;
  - Annualises the monthly data systematically in line with the monthly time series model fitted to the monthly data;
  - Utilises as much data as possible without any key events being missed; and
  - Improves the fit to the empirical data and leads to a stable calibration.
  - It can handle the data with volatility clustering and avoids the issue of autocorrelation.
- Use of autocorrelation adjustment (or “de-smoothing” the data). This technique is not covered here as this is a widely researched topic (Marcatoo, 2003). However, we have tried using a similar technique by (HengSun_Nelken, 2009) which tries to correct the bias in the overlapping variance of the data. We have analysed the impacts of using this adjustment in section 4 of the paper and have not discussed further in this section.
- The testing carried out in section 6 is based on empirical data where the underlying model driving the data is unknown. As the model is unknown the bias and mean square error tests carried out in section 4 are not possible (as these require the model parameters to be known).

6.2 Annualisation Method

Under this approach, we analyse the data points using monthly non-overlapping time steps but utilise the correlation present in the monthly time series data to create a large data set to perform an annual non-overlapping calibration.

The key data analysis steps are as follows:

- Calculate the monthly changes in the time series;
- Calculate the empirical correlation between each 12 calendar month period by arranging all January changes in one column and February change in the next one and so and so forth and calculate the correlations;
- Apply this correlation to generate a large number of monthly steps (e.g. 100k) and aggregate monthly steps to come up with annualised simulations depending upon whether we modelling the time series multiplicatively or additively.
- The annualisation is performed using empirical marginal distributions and Gaussian (or even empirical copula) copula using an autocorrelation matrix for each of the time series to avoid any information loss due to fitting errors.
- This technique has the advantage of fitting distributions based on a large sample leading to more stable results, however suffers from the fact that it still uses monthly data which may be auto correlated.

We fit distributions to these annualised simulations. We present the use of this technique using Merrill Lynch (ML) credit data in this section where we compare the results of using annual overlapping data (without any aggregation approach) and using the above autocorrelation aggregation approach.

6.2.1 Dataset Used

Although methodology used for annualisation process is quite generic in nature and can be used for a wide range of datasets, we have used ML credit indices because of the following peculiarities of this dataset:

- The dataset is limited (starting 1996) and therefore utilisation of information available in each of the data points is important;
- This dataset has a single extreme market event (2008-09 Global credit crisis) and rest of the data is relatively benign.
- Two significant challenges for calibrating to this data set are:
    - If we use annual non-overlapping dataset, we may lose the key events of 2008-09 Global credit crisis where the extreme movements in spreads happened during June 2008 to March 2009 (9 month period);
    - If we use annual overlapping dataset, the data points used in the fitting process are more than the data points using annual non-overlapping dataset, however, not sufficient for generating a credible and robust fit at the 99.5th percentile point.

6.2.2 Empirical Data Analysis

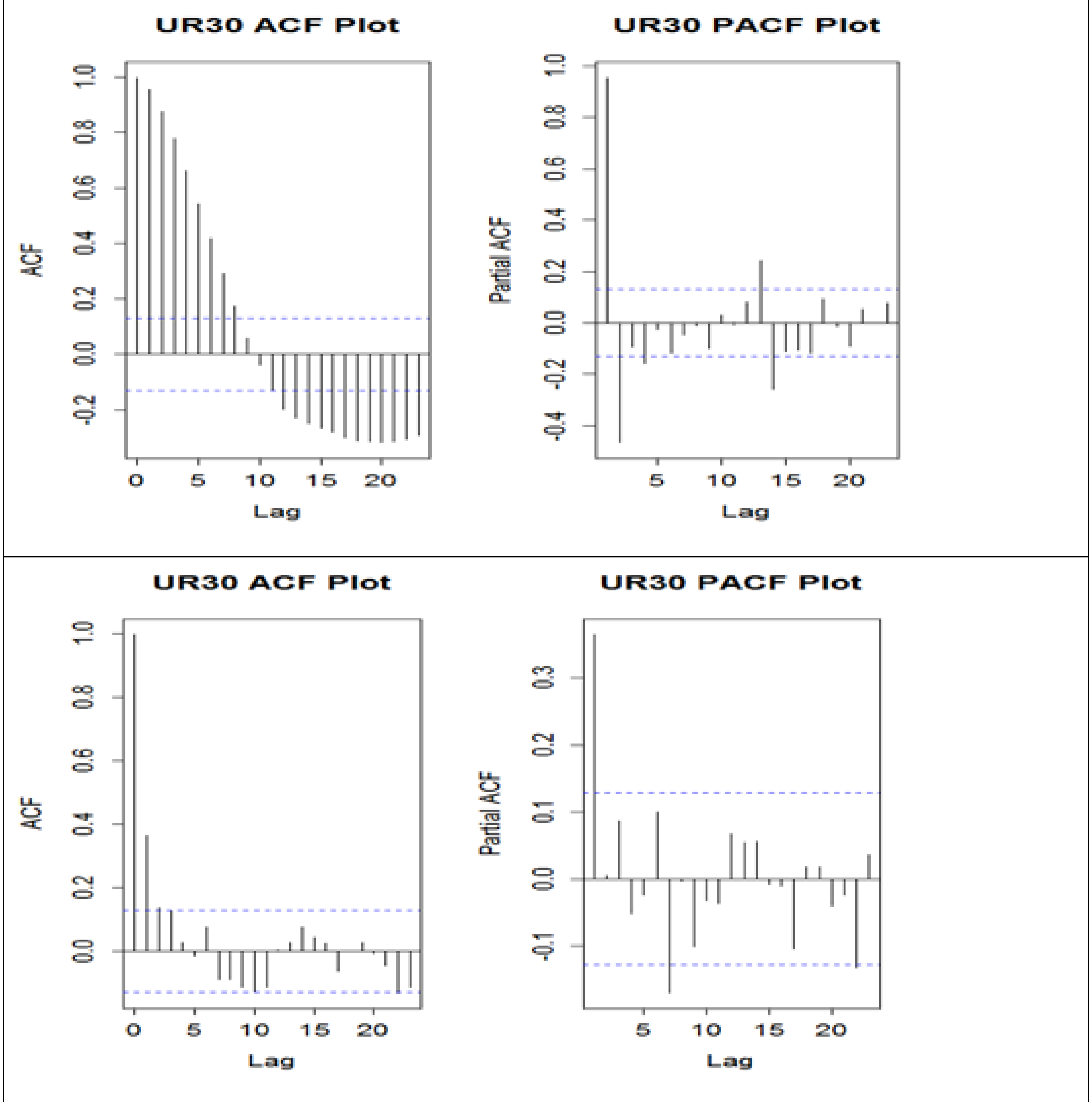
In this section, the main purpose is to compare the results of some of the general tests applied to both annual overlapping and monthly non-overlapping data to check whether using monthly non-overlapping time series is more conducive to risk calibration or not.

We consider a practical example of the approach described in the previous section based on credit spread data. We first look at auto correlation function (ACF) and partial auto correlation function (PACF) plots[3] using two of the ML credit indices UR30 (ML A rated index – all maturities) and UR40 (ML BBB rated index – all maturities). The term Annual Overlapping is used in this section to mean annual data overlapping by 11 out of 12 months of the year.

The auto-correlation plots show the correlation between data points with different lags on the x-axis. Similarly, for the partial auto-correlation plots. Note that stationarity tests have been carried out in appendix C, section 11.

[3] Autocorrelation Function (ACF) and Partial Autocorrelation (PACF) are standard techniques used for determining the order of ARIMA process and provides indications for stationarity property of the time series amongst their other uses.

Figure 1: Annual Overlapping vs. Monthly Non-Overlapping – A rating – All Maturities



Explanation:

Under the annual overlapping time series(top left diagram) the ACF starts at 1, slowly converges to 0 (slower decay) and then becomes negative and exceeds the 95% confidence level for the first 9 lags.

Under the monthly non-overlapping time series (bottom left diagram) the ACF quickly falls to a very low number and beyond lag 2 for most time lags the autocorrelations are within the 95% confidence interval, for all practical purposes we can ignore the ACF after time lag 2. The suggests that using monthly non-overlapping time-series is less auto-correlated than the annual overlapping time series..

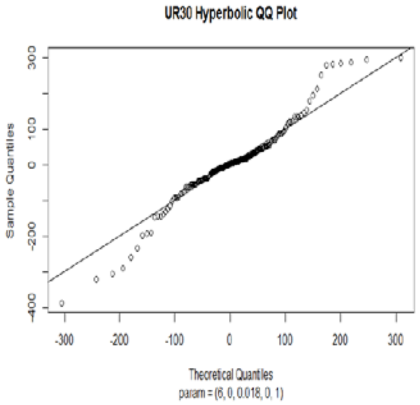
Similarly the PACF for monthly non-overlapping data (bottom right) shows more time steps where autocorrelations beyond lag 2 are within the 95% confidence interval in comparison to the annual overlapping data (top right).

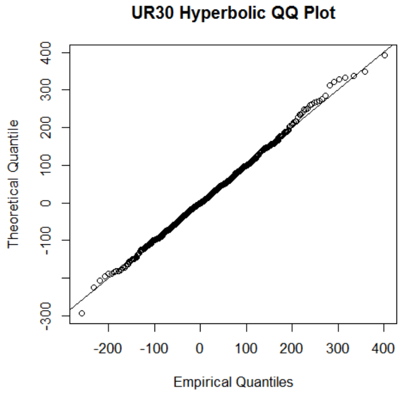
The purpose of performing these tests is to show if using monthly non-overlapping time series is more conducive to modelling or not.

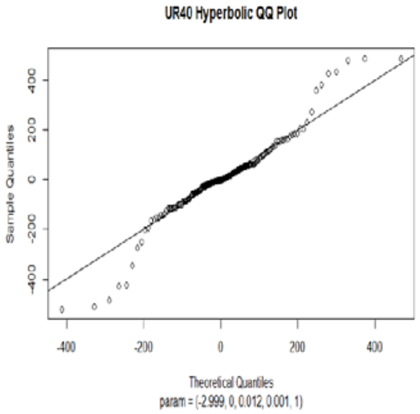
6.2.2.1 Fitting Results – QQ Plots – Hyperbolic Distribution

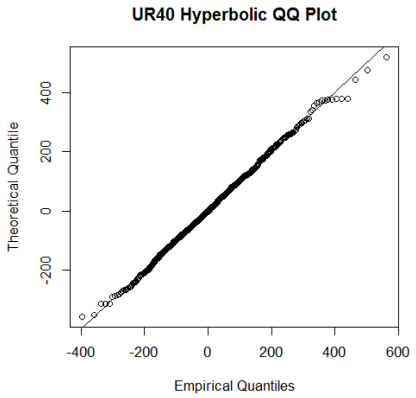
We present the QQ plots for annual overlapping vs. monthly non-overlapping with autocorrelation using hyperbolic distribution.

Figure 2: Annual Overlapping vs. Monthly Non-overlapping with Autocorrelation

<table>
<tr><td>QQ Plots using Monthly Annual Overlapping</td><td>QQ Plots using Monthly Non-overlapping with Annualisation using Autocorrelation</td></tr>
<tr><td>


</td><td>


</td></tr>
<tr><td>


</td><td>


</td></tr>
<tr><td colspan="2">Explanation: From the QQ plots (both using Hyperbolic distribution) between monthly annual overlapping and monthly non-overlapping with annualisation above, it is clear that using monthly non-overlapping data with autocorrelation appears to improve the fits in the body as well in the tails. This is because the QQ plots show a much closer fit to the diagonal for the monthly non-overlapping data with annualisation.

Note: We have used hyperbolic distribution as it is considereed one of the most sophisticated distributions. Similar conclusions can be drawn using more simpler distributions such as Normal distribution.</td></tr>
</table>

6.2.2.2 Conclusions

Annualising monthly non-overlapping data using monthly auto-correlations can be a versatile alternative to annual overlapping data, particularly where data is limited.

It is important to note that annualisation process is not always the ideal solution because the annualisation process also introduces uncertainty depending upon the aggregation approach used. However, this uncertainty can be reduced by using empirical distributions for the high frequency data (i.e. monthly process) where possible and use empirical or Gaussian copula where minimum number of parameter estimations are required in the annualisation process.

6.3 Temporal Aggregation Methods

Another alternative approach to using overlapping annual data is to use temporal aggregation. This is an approach where we construct a low frequency series (e.g. annual series) from a high frequency series (e.g. monthly / daily series). This is done by fitting a time series model e.g. Auto regressive, GARCH etc…) to the monthly data; which then gives all the information required to model the annual time series.

Temporal aggregation can be very useful in the cases where we have limited relevant market data available for calibration and we want to infer the annual process from the monthly/daily process.

6.3.1 Introduction

Under the temporal aggregation technique, the low frequency data series is called the aggregate series, e.g. annual series. The high frequency data series is called the disaggregate series, e.g. monthly series. Deriving a low frequency model from the high frequency model is a two stage procedure:

- ARMA-GARCH models are given in terms of lag polynomials, where it is necessary to choose the polynomials orders. Temporal aggregation allows us to infer the orders of the low frequency model from those of high frequency.
- After inferring the orders, the parameters of the low frequency model should be recovered from the high frequency ones, rather than estimating them. Hence, the low frequency model parameters incorporate all the economic information from the high frequency data.

$$Y_t^* = W(L)y_t = \sum_{j=0}^{A} w_j y_{t-j} = \sum_{j=0}^{k-1} L^j y_t$$

Where W (L) is the lag polynomial of order A. W (L) = 1+L+..+L^ (k-1) where k represents the order of aggregation.

If the disaggregate time series y_t were to follow a model of the following type

$$\emptyset(L)y_t = \theta(L)\varepsilon_t$$

Where $\emptyset(L)$ and $\theta(L)$ are lag polynomials and ε_t is an error term. Then the temporally aggregated time series can be described by

$$\beta(B)y_t^* = \varphi(B)\varepsilon_t^*$$

- We perform a time series regression model to estimate the coefficients of an ARMA or ARIMA model on monthly non-overlapping data;

- Annual estimates are constructed out of non-overlapping monthly observations.
- Autocorrelation in the data is accounted for to make sure the estimates are valid; and
- The standard goodness-of-fit techniques are valid

The key limitations of temporal aggregation are as follows:

- Temporal aggregation leads to loss of information in the data due to information that will be lost along the way during performing various data transformations. However, empirical work done using equity risk data shows that this loss of information has not been materially significant based on the quantile results observed under various approaches in section 6.3.2.
- Rigorous testing and validation of the behaviour of the residuals will be necessary;
- It is complex to understand and communicate.

The main complication with using temporal aggregation technique is the fact that it involves solving algebraic system of equation which can get complex for complex time series models as models have higher orders, e.g. ARIMA (p,d,q) where p, d and /or exceed 3.

6.3.1.1 Technical Details for AR(1) Process

We study this technique using a simple example Auto Regressive AR(1) process. Assume that the monthly log- return r_t follows an AR(1) process. (WSChan_etal, 2008)

$$r_t = \emptyset\, r_{t-1} + a_t\, , a_t \sim N(0, \sigma_a^2\,)$$

The annual returns are noted as R_T and frequency is defined as m (where m=12 for annual aggregation). The lag-s auto-covariance functions of the m-period aggregated log return variable.

$$Cov[R_T, R_{T+s}] = [m + 2(m-1) + 2(m-2)\emptyset^2 +..+ 2\emptyset^{m-1}]\, \frac{\sigma_a^2}{1-\emptyset^2}\ if\ s = 0$$

$$Cov[R_T, R_{T+s}] = [1 + \emptyset + \emptyset^2 +..+ \emptyset^{m-1}]\left[\frac{\emptyset^{m(|s|-1)+1}}{1-\emptyset^2}\right]\sigma_a^2\ if\ s = \pm 1, \pm 2.$$

$$Var[R_T] = [12 + 22 + 20\emptyset^2 +..+ 2\emptyset^{11}]\, \frac{\sigma_a^2}{1-\emptyset^2}\ when\ s = 0\ and\ m = 12$$

$$(1 - \emptyset^* L)R_T = \ (1 - \theta^* L)a_t^*\ a_t^* \sim N(0, \sigma_{a^*}^2\,)$$

$\emptyset^* = \emptyset^m$ (for real life applications where for annualisation we use $\emptyset^{12}$ it will be close to zero and therefore the process essentially becomes an MA (1) process). For $|\emptyset^*| < 1$,

$$(\emptyset^m - \theta^*)\frac{(1 - \emptyset^m \theta^*)}{1 - 2\emptyset^m \theta^* + \theta^{*2}} = \emptyset[1 + \emptyset + \emptyset^2 +..+ \emptyset^{m-1}]^2 / [m + 2(m-1) + 2(m-2)\emptyset^2 +..+ 2\emptyset^{m-1}]$$

6.3.1.2 Technical Details for GARCH (1, 1) Process

Let $a_t = (r_t - \mu)$ is a mean corrected log return and follows Garch (1,1) process, then

$$\varepsilon_t = \ a_t / h_t^{0.5}$$

$$h_t = \omega + \beta h_{t-1} + \alpha a_{t-1}^2$$

The m-month non-overlapping period can be "weakly" approximated by Garch (1,1) process with corresponding parameters:

$$\mu^* = m\mu$$

$$\omega^* = m\omega \left\{ \frac{1-(\alpha+\beta)^m}{1-(\alpha+\beta)} \right\}$$

$$\alpha^* = (\alpha+\beta)^m - \beta^*$$

$|\beta^*| < 1$ is the solution of the following quadratic equation:

$$\frac{\beta^*}{1+\beta^{*2}} = \frac{(\Theta(\alpha+\beta)^m - \Lambda)}{(\Theta(1+(\alpha+\beta)^{2m}) - 2\Lambda)}$$

$$\Lambda = \left(\frac{\big(\alpha - \alpha\beta(\alpha+\beta)\big)(1-(\alpha+\beta)^{2m})}{1-(\alpha+\beta)^2} \right)$$

$$\Theta = m(1-\beta)^2 + \left\{ \frac{2m(m-1)(1-\alpha-\beta)^2(1-2\alpha\beta-\beta^2)}{(\kappa-1)(1-(\alpha+\beta)^2)} \right\} + 4\left\{ \frac{(m-1-m(\alpha+\beta)+(\alpha+\beta)^m)\big(\alpha-\alpha\beta(\alpha+\beta)\big)}{1-(\alpha+\beta)^2} \right\}$$

Where $\kappa$is the unconditional kurtosis of the data.

$$\kappa^* = 3 + \frac{\kappa-3}{m} + 6(\kappa-1)\left\{ \frac{\big(\alpha-\alpha\beta(\alpha+\beta)\big)(m-1-m(\alpha+\beta)-(\alpha+\beta)^m)}{m^2(1-\alpha-\beta)^2(1-2\alpha\beta-\beta^2)} \right\}$$

6.3.2 UK Equity Case Study Temporal Aggregation – Garch (1,1)

In this section, we present and example of the temporal aggregation method applied to UK (FTSE All Share Total Return) index data using the Garch model fitted to monthly data.

The calculation steps applied are as follows:

- We calculate excess of mean log monthly non-overlapping returns of the data;
- We fit a Garch (1,1) model these excess of mean log returns and derive the fitted parameters of Garch model;
- Calculate the temporally aggregated parameters for the annual time series;
- Compare the (simple) quantiles of empirical annual-non-overlapping, empirical annual overlapping and Temporally aggregated Garch (1,1) process.

Table 1: Key Quantiles – Monthly Annualised vs. Monthly Annual Overlapping Data

Percentiles	Empirical Non-overlapping	Empirical Monthly Annual Overlapping	Temporally Aggregated Garch (1,1)
99.9%	99.9%	119.8%	121.8%
99.5%	97.9%	84.3%	85.6%
99.0%	73.5%	59.2%	66.0%
98.0%	33.4%	39.3%	50.8%
97.5%	27.1%	35.3%	46.4%
95.0%	24.7%	26.9%	34.5%
90.0%	18.7%	20.8%	24.0%
80.0%	14.8%	14.8%	14.2%
75.0%	11.3%	11.8%	11.0%
50.0%	3.2%	2.6%	0.0%
25.0%	-8.0%	-8.7%	-10.0%
20.0%	-11.9%	-11.7%	-12.5%
10.0%	-18.4%	-19.5%	-19.4%
5.0%	-32.2%	-29.2%	-25.7%
2.5%	-35.2%	-35.9%	-31.9%
2.0%	-37.5%	-37.6%	-33.8%
1.0%	-47.6%	-45.0%	-39.9%
0.5%	-52.0%	-56.4%	-46.1%
0.1%	-55.3%	-58.1%	-60.5%

From the comparison of the key quantiles in the table above, we conclude:

- On the extreme downside and upside, temporally aggregated Garch (1,1) process leads to stronger quantiles in comparison to annual non-overlapping and annual overlapping time series;
- In the "body" of the distribution temporally aggregated Garch (1,1) process leads to weaker quantiles in comparison to annual non-overlapping and monthly annual overlapping time series;

The calibration parameters of Garch (1, 1) process fitted to monthly non-overlapping and temporally aggregated Garch (1, 1) are outlined in *Table 5*.

Table 2: Table of parameters

Parameter	Monthly Non-overlapping	Temporally Aggregated parameter (Annual, m=12)
Mu	0	0
Omega	0.00014	0.01591
Alpha	0.1475	0.1656
Beta	0.8071	0.4070
Dof	5.954	4.945

7 Conclusions

This paper has considered some of the main issues with using overlapping data as well as looking at the alternatives.

Section 4 presented the results of a simulation study designed to test whether overlapping or non-overlapping data is better for distribution fitting. For the models tested, overlapping data appears to be better as biases can be removed (in a similar way to non-overlapping data), but overlapping makes a greater use of the data, meaning it has a lower mean square error. A lower mean square error suggests distributions fitted with overlapping data are more likely to be closer to the correct answer.

Section 5 discussed the issues of statistical tests using overlapping data as well as presenting a methodology for using statistical tests with overlapping data. This methodology was tested and the adjustment for overlapping data was found to correct the statistical tests in line with expectations.

Section 6 presented alternative methods for model fitting, by fitting the model to shorter time frame data and then aggregating the monthly model into an annual model. This approach was successfully tested in a practical example.

The overall conclusions from this paper are:

- Overlapping data can be used to calibrate probability distributions and is expected to be a better approach than using non-overlapping data, particularly when there is a constant struggle between finding relevant data for risk calibration and maximising the use of data for a robust calibration. However, communication of the uncertainty in the model and / or parameter uncertainty to the stakeholder is equally important.
- Some credible alternatives exists to using overlapping data such as temporal aggregation and annualization, however these alternatives bring their own limitations and understanding of these limitations is key to using these alternatives. We recommend to considering the comparison of the calibration using both non-overlapping monthly data annualised with overlapping annual data and discuss the advantages, robustness and limitations of both the approaches with stakeholders before finalising the calibration approach.

8 References

AMEMIYA&WU. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of American Statistical Association*, 628-632.

Bain and Engelhardt (1992). Introduction to Probability and Mathematical Statistics ISBN 0-534-92930-3 Duxbury

Conover, W. J. (1999). "Practical nonparametric statistics", 3rd ed. Wiley: New York.

Cont, R. (2005). Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. In A. K. Teyssiere, *Long Memory in Economics.* Springer.

D. Kwiatkowski, P. C. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics*, 1590178.

Feike_Drost_Nijman. (1993). Temporal Aggregation of Garch Processes. *Econometrica*, 909-927.

Frankland, R (chair), Biffis E, Dullaway D, Eshun S, Holtham A, Smith A, Varnell, E and Wilkins T (2008). The Modelling of Extreme Events. British Actuarial Journal.

HengSun_Nelken. (2009, April). Error of VaR by overlapping intervals. *Risk Management*.

Ljung&Box. (1978). On a measure of lack of fit in time series models. *Biometrika*, 297-303.

Marcatoo, P. B. (2003). *THE MEASUREMENT AND MODELLING OF COMMERCIAL REAL ESTATE PERFORMANCE.* Presented to IOFA.

Mathworks. (2017, 10 9). *uk.mathworks.com*. Retrieved from https://uk.mathworks.com/help/econ/kpsstest.html

Mandelbrot. (n.d.). The Variation of Certain Speculative Prices. *Journal of Business*.

PhillipsPerron. (1988). Testing for a unit root in time series regression. *Biometrica*, 335-346.

WSChan_etal. (2008, July). Temporal aggregation of equity return time-series models. *Mathematics and Computers in Simulation*.

9 Appendix A

In this section, we provide the mathematical definitions and descriptions of the technical terms used in the paper[4].

9.1 What Are Cumulants of a Random Variable?

9.1.1 Definitions

Cumulants are properties of random variables. The first two cumulants: mean and variance, are well-known. The third cumulant is also the third central moment. For a random variable *X* with mean μ, the first four cumulants are:

$$\kappa_1 = \mu = \mathbb{E}(X)$$

$$\kappa_2 = \mathbb{E}(X-\mu)^2$$

$$\kappa_3 = \mathbb{E}(X-\mu)^3$$

$$\kappa_4 = \mathbb{E}(X-\mu)^4 - 3\kappa_2^2$$

Higher cumulants theoretically exist but are less often encountered. We restrict our discussion to first four cumulants only.

9.2 Statistical Properties of Cumulants

9.2.1 Additive Property

The cumulants satisfy an additive property for independent random variables. If *X* and *Y* are statistically independent and $n \geq 1$ then

$$\kappa_n(X+Y) = \kappa_n(X) + \kappa_n(Y)$$

For a Normal distribution, the third and subsequent cumulants are zero.

9.2.2 Skewness and Kurtosis

The *skewness* and *kurtosis* of a random variable are defined in terms of the cumulants, as follows:

Skewness = $\frac{\kappa_3}{\kappa_2^{3/2}}$

Kurtosis =$\frac{\kappa_4}{\kappa_2^2}$

Skewness and kurtosis are both shape attributes, which are unchanged when a random variable is shifted or scaled by a positive multiple.

It is a consequence of the additive property that, for sums of independent identically distributed random variables, the skewness and kurtosis tend to zero as the number of observations in the sum tends to infinity. This observation is consistent with the central limit theorem.

9.3 Using Cumulants to Estimate Distributions

9.3.1 Empirical Cumulants

Given a number *n* of data points, the *empirical* distribution puts a mass of n^{-1} on each observation.

The *empirical* cumulants are the cumulant estimates based on the empirical distribution, which we will denote with a tilde (~). The first empirical cumulant $\tilde{\kappa}_1$ is the sample average. Other

[4] Source: http://mondi.web.elte.hu/spssdoku/algoritmusok/acf_pacf.pdf

empirical cumulants are defined similarly. For example, the empirical variance (second cumulant) is the average squared deviation between each observation and that sample average. The empirical fourth moment $\tilde{\kappa}_4$ is the average fourth power of deviations, minus three times the squared empirical variance.

9.3.2 Distribution fitting with Cumulants

We can use the empirical cumulants, or modifications thereof, to estimate distributions. The methodology is to find a distribution whose cumulants match the cumulants estimated from a data sample (i.e. as with the method of moments a probability distribution is uniquely defined by its cumulants).

Common practice (see EEWP 2008 and Willis Towers Watson (WTW)[5] risk calibration survey 2016) for market risk models is to pick a four-parameter distribution family, closed under shifting and scaling. EEWP 2008 showed distributions from the Pearson IV family and the hyperbolic family. In this paper we show examples based the NIG (Normal Inverse Gaussian) family. In each case, the procedure is the same:

- Estimate the mean, variance, skewness and kurtosis from the historical data;
- Pick a four-parameter distribution family;
- Evaluate whether the estimated (skew, kurtosis) combination is feasible for the chosen family. If not, adjust the historical values by projecting onto the boundary of the feasible region;
- Find the distribution matching the adjusted historical skewness and kurtosis;
- Match the mean and variance by shifting and scaling;
- Compare the fitted distribution to the historic data, either by inspection of histograms or more formal statistical tests. If the fit is not good enough then think of another four-parameter family and repeat from the third step above.

[5] WTW risk calibration survey 2016 suggests 4 parameter distributions such as Hyperbolic and EGB2 are widely used by UK insurers. Please note WTW risk calibration survey 2016 is not a publicly available document, however, can be made available if requested after permission from Wills Towers Watson.

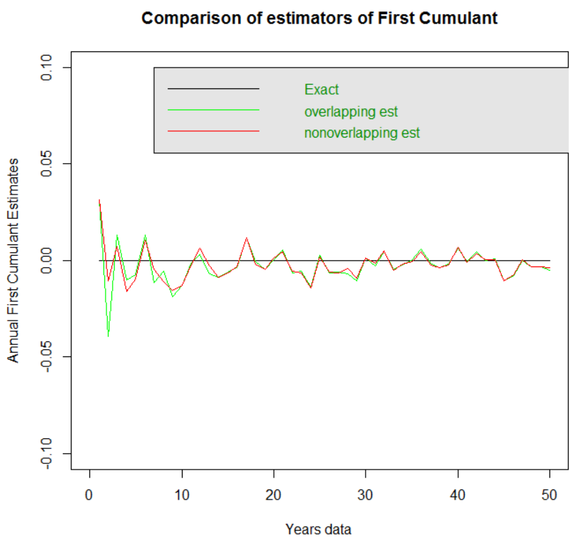
10 Appendix B – Simulation study – additional results

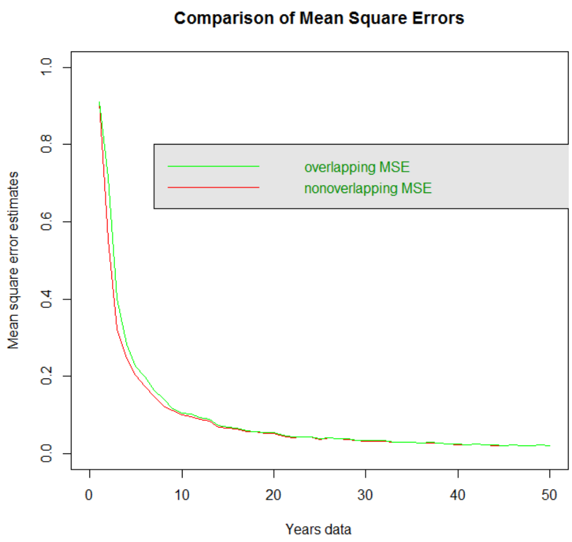
This section shows additional results from the simulation study in section 4.

10.1 Normal Inverse Gaussian results

The results for the Normal Inverse Gaussian Reference model are shown below. These results are very similar to the Brownian case.

10.1.1 First Moment – the mean

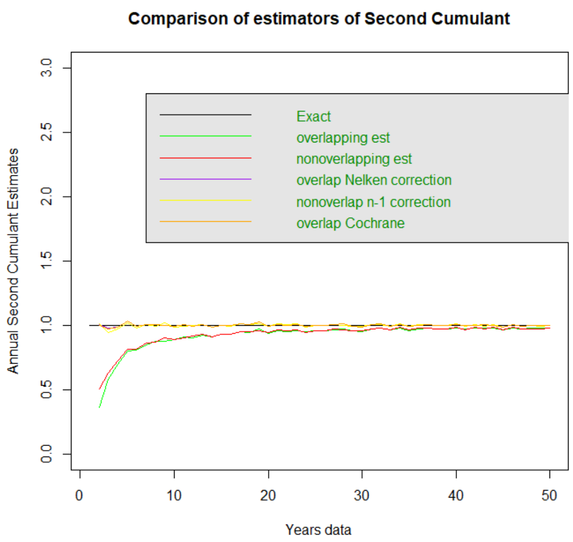




The plots above show the bias in the plot on the left and the mean square error on the plot on the right. The overlapping and non-overlapping data estimates of the mean appear very similar and not obviously biased. They also have very similar mean square errors across all years. Very similar conclusions to the Brownian case.

10.1.2 Second cumulant – variance

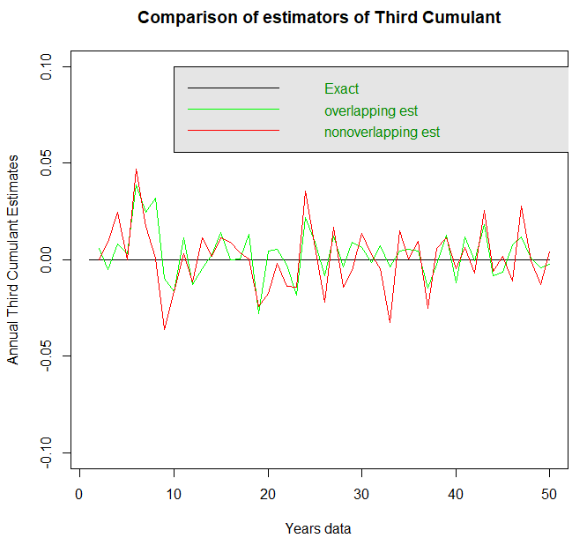
The second cumulant is the variance (with divisor n).



Very similar conclusions as for the Brownian case.

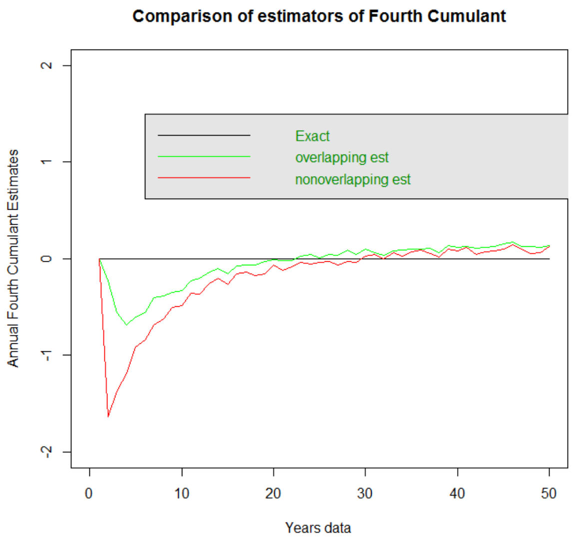
- Overlapping and non-overlapping data both give biased estimates of the second cumulant to a similar extent across all terms
- The bias correction factors (using divisor n-1 for non-overlapping variance and the Nelken formula for overlapping variance) appear to remove the bias. This is evidence the Nelken bias correction factor works for other process than just Brownian motion.
- The plot on the right shows the mean square errors for the two approaches, with overlapping data appearing to have lower mean square errors for all terms.

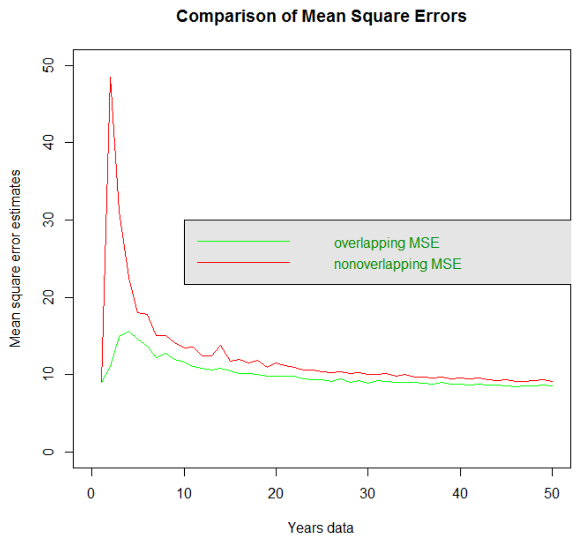
10.1.3 Third cumulant



Neither approach appears to have any systemic bias for the mean. The mean square error is significantly higher for non-overlapping data than overlapping data.

10.1.4 Fourth cumulant

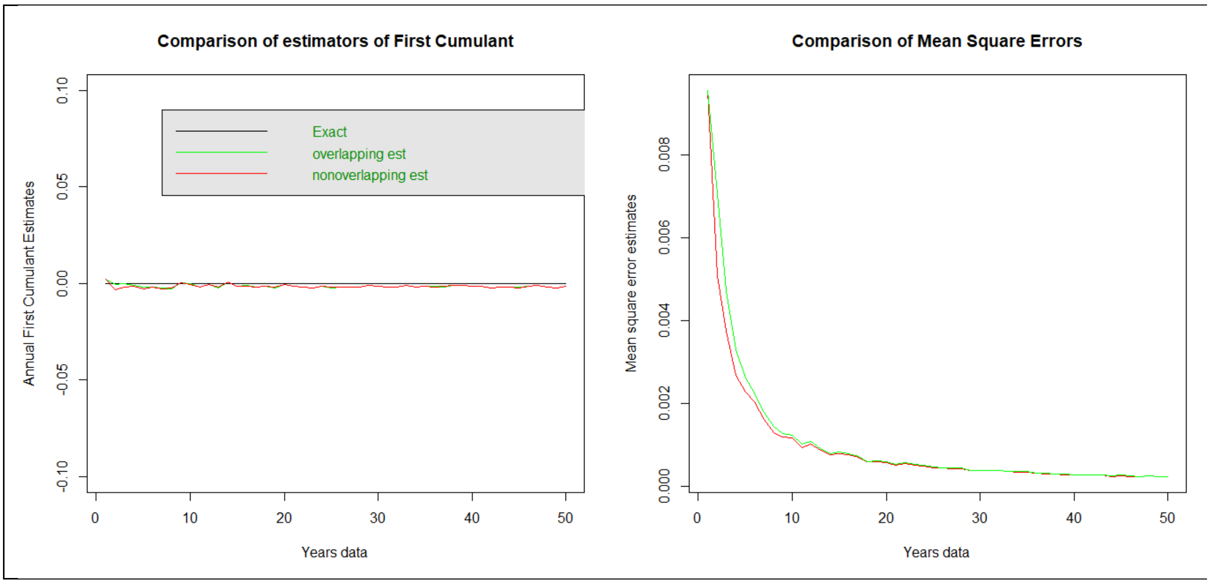




In this case the non-overlapping data appears to have a higher downward bias than overlapping data at all terms; both estimates appear biased. The bias does not appear to tend to zero as the number of years increases, but it rises above the known value. The non-overlapping data has higher mean square error than the overlapping data.

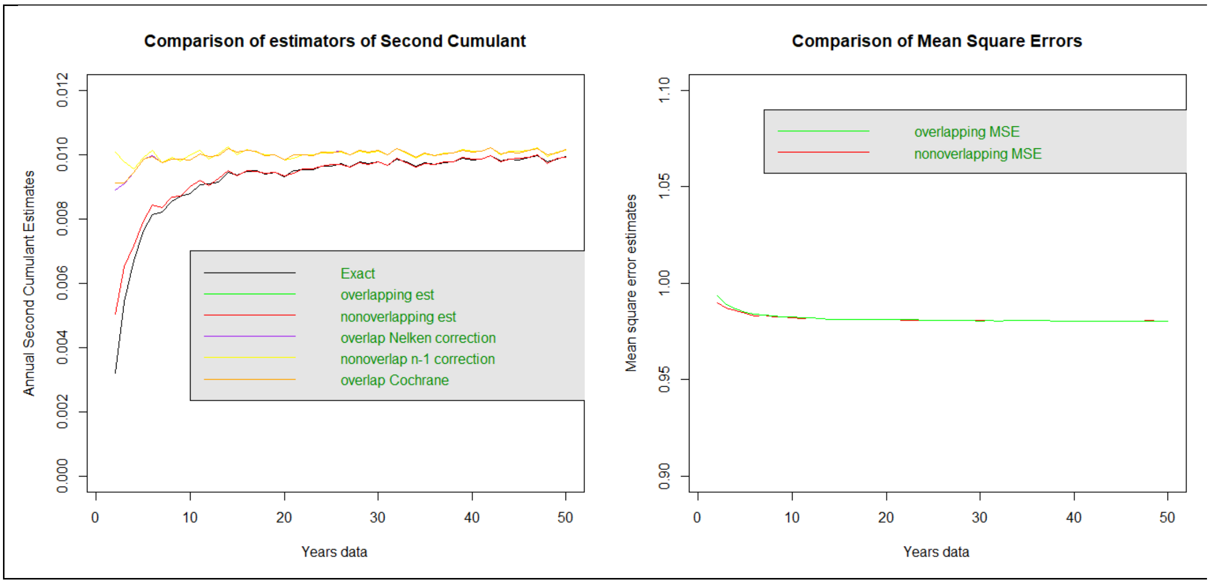
10.2 ARIMA results

10.2.1 First cumulant – the mean



The plots above show the bias in the plot on the left and the mean square error on the plot on the right. The overlapping and non-overlapping data estimates of the mean appear very similar and unbiased. They also have very similar mean square errors after 10 years but overlapping data appears to have marginally higher MSE below 10 years.

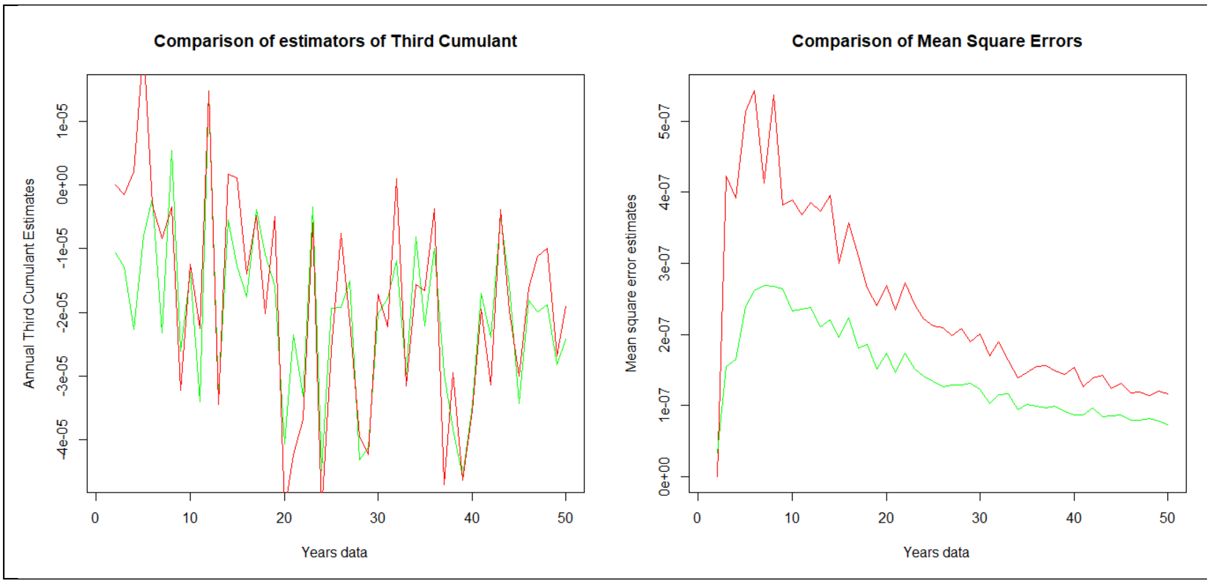
10.2.2 Second cumulant – variance



The plot above on the left shows that the overlapping and non-overlapping estimates of the variance (with divisor n) are too low with similar bias levels for all terms. This is more marked the lower the number of years data, and the bias appears to disappear as n gets larger.

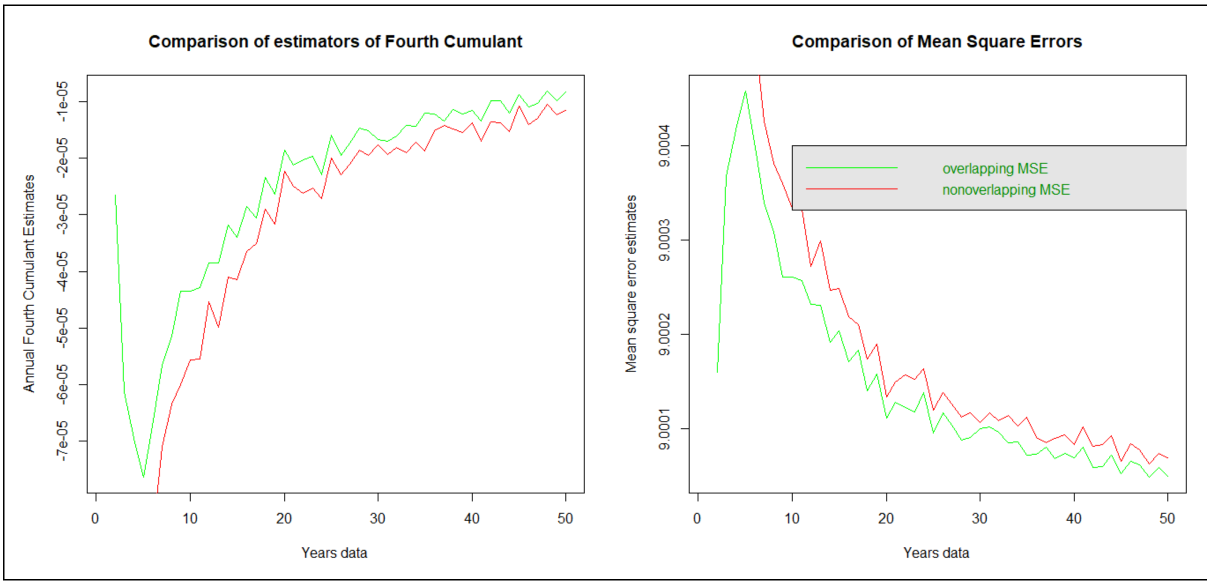
The plot on the left also shows the second cumulant, but bias corrected, using a divisor (n-1) instead of n for the non-overlapping data and using the formula in Heng Sun Nelken et al as well as in Cochrane 1988 for the overlapping data. Both of these corrections appear to have removed the bias across all terms for overlapping and non-overlapping data. The MSE is very similar for both overlapping and non-overlapping data.

10.2.3 Third cumulant



It is important to note that neither approach appears to have any materially different bias. Non-overlapping data has higher MSE compared to overlapping data.

10.2.4 Fourth cumulant



Non-overlapping data has lower bias compared to overlapping data but overlapping data has lower MSE.

11 Appendix C – Stationarity tests

Phillips-Perron Test (PP Test): (PhillipsPerron, 1988)

Phillips-Perron Test involves fitting the following regression model:

$$\boldsymbol{y_t = \alpha + \rho y_{t-1} + \delta t + u_t}$$

The results are used to calculate the test statistics proposed by Phillips and Perron. Phillips and Perron's test statistics can be viewed as Dickey–Fuller statistics that have been made robust to serial correlation by using the Newey–West (1987) heteroskedasticity- and autocorrelation-consistent covariance matrix estimator. Under Phillips-Perron unit root test the hypothesis are as follows:

H null: The time series has unit root (which means it is non-stationary)

H Alternative: The time series does not have unit root (which means it is stationary)

Kwiatkowski-Phillips-Schmidt-Shin Test (KPSS Test): (D. Kwiatkowski, 1992)

The KPSS Test has been developed to complement unit root tests as the last have low power with respect to near unit-root and long-run trend processes. Unlike unit root tests, Kwiatkowski et al. provide straightforward test of the null hypothesis of trend and level stationarity against the alternative of a unit root.

For this, they consider three-component representation of the observed time series ADF Time Series as the sum of a deterministic time trend, a random walk and a stationary residual:

$$Y_t = \beta t + (r_t + \alpha) + e_t$$

$= r_{t-1} + ut$ is a random walk, the initial value $r_0 = \alpha$ serves as an intercept, t is the time index, u_t are independent identically distributed (0,σ_u^2). Under KPSS test the hypothesis are as follows:

H null: The time series is trend/level stationary (which means it does not show trends)

H Alternative: The time series is not trend/ level stationary (which means it does show trends)

Ljung-Box Q Test: (Ljung&Box, 1978)

Ljung-Box Q test is whether any of a group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags, and is therefore a portmanteau[6] test. Under Ljung-Box Q test the hypothesis are as follows:

H null: The time series is independent

H Alternative: The time series is not independent and has positive or negative strong serial correlation

[6] A portmanteau test is a type of statistical hypothesis test in which the null hypothesis is well specified, but the alternative hypothesis is more loosely specified

The statistic under Ljung-Box test is calculated as follows:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

Where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Under null hypothesis, $Q \sim \chi_h^2$ where h degrees of freedom.

These tests have been applied to the corporate bond indices with the results presented below:

The credit spread data is subject to a number of different stationarity tests. If the process is stationary, it is more conducive for a robust calibration because its statistical properties remain constant over time (e.g. the mean, variance, autocorrelation etc. do not change). If the process is not stationary, the variation in the fitting parameters can be significant as the new information emerges in the new data or in some cases the model may no longer remain valid. This point is important for stakeholders because the stability of the SCR depends upon the stability of the risk calibrations.

There are various definitions of stationarity in the literature, we present "weak" stationarity definition here, we believe it is widely used, however, stronger forms may be required, for example, when considering higher moments.

A process is said to be covariance stationary or "weakly stationary", if its first and second moments are time invariant, i.e.

$$E(Y_t) = E(Y_{t-1}) = \mu \qquad \forall t$$

$$Var(Y_t) = \gamma_0 < \infty \qquad \forall t$$

$$Cov(Y_t, Y_{t-k}) = \gamma_k < \infty \qquad \forall t, \forall k$$

The third condition means that the autocovariances only depend on the decay in the time but not in the time itself. Hence, the structure of the series does not change with the time.

A number of statistical tests for stationarity are defined in Appendix C. the results of these tests together with a discussion of these results are presented below.

Table 3: Credit Indices – Monthly non-overlapping data - stationarity and unit-root tests

	Monthly Non-overlapping data annualised		Monthly Annual Overlapping[7]	
P-values[8]	UR30	UR40	UR30	UR40
PP Single Mean Test	1%	1%	1%	1%
PP Trend Test	1%	1%	7%	6%
KPSS Trend	24%	67%	59%	72%
KPSS Level	15%	26%	67%	75%
Ljung-Box	90%	13%	0%	0%

Table 4: PP and KPSS Tests

[7] Note: no adjustment has been applied to this stationarity test for overlapping bias

[8] Note: p-value are used to determine statistical significance in a hypothesis test. Intuitively, Higher p-values than the threshold indicate the data is likely with a true null hypothesis and Lower p-values than a threshold indicate the data is unlikely with a true null hypothesis. Typically, a 5% threshold is used in many applications.

	Result	p-value	Conclusion
Phillips-Perron Test (PP Test)	Stationary	1%	The p-value is less than 5% which suggests that we **reject** the null hypothesis of the time series having a unit-root. This is strong evidence of stationarity in the time series. Both Monthly non-overlapping annualised data and monthly annual overlapping data both have similar results supporting that both time series don't support presence of unit-root.
KPSS Trend Stationarity Test	Stationary	15%-75%	The p-value is greater than 5% which means we are unable to reject the null hypothesis. This means the time series is trend stationary. Both Monthly non-overlapping annualised data and monthly annual overlapping data both have similar results supporting that both time series are trend stationary.
KPSS Level Stationarity Test	Stationary	15%-41%	The p-value is greater than 5% which means we are unable to reject the null hypothesis. This means the time series is level stationary. Both Monthly non-overlapping annualised data and monthly annual overlapping data both have similar results supporting that both time series are level stationary.
Ljung-Box Test	Not Independent	>10%	The p-values are greater than 5%. We are able to reject the null hypothesis and conclude that the time series does not show serial correlation. Monthly non-overlapping annualised data do not show serial correlation, however we are unable to reject the hypothesis for monthly annual overlapping.

The key implications of these tests are:

- The stationarity tests support (or are unable to reject) the hypothesis that the time series are stationary under both monthly non-overlapping annualised data and monthly annual overlapping data.
- It is important to note that the Ljung-Box test suggests that the data has serial correlation for monthly annual overlapping data however; we are able to reject the hypothesis for monthly non-overlapping annualisation approach.

The purpose of doing these tests is to show that using monthly non-overlapping annualised data can be a better alternative if we can annualise it rather than using monthly annual overlapping data.

Institute
and Faculty
of Actuaries

Institute and Faculty of Actuaries

London

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100 · Fax: +44 (0) 20 7632 2111

Edinburgh

Level 2 ·Exchange Crescent · 7 Conference Square · Edinburgh ·EH3 8RA
Tel: +44 (0) 131 240 1300 · Fax +44 (0) 131 240 1311

Oxford

1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD
Tel: +44 (0) 1865 268 200 · Fax: +44 (0) 1865 268 211

Beijing

6/F · Tower 2 · Prosper Centre ·5 Guanghua Road · Chaoyang District · Beijing · China 1000020
Tel: +86 (10) 8573 1000

Hong Kong

2202 Tower Two · Lippo Centre · 89 Queensway · Hong Kong
Tel: +11 (0) 852 2147 9418 · Fax: +11 (0) 852 2147 2497

Singapore

163 Tras Street · #07-05 Lian Huat Building · Singapore · 079024
Tel: +65 6717 2955

www.actuaries.org.uk