



## ABSTRACT OF THE EDINBURGH DISCUSSION

# Calibration of VaR models with overlapping data

[Extreme Events Working Party, Institute and Faculty of Actuaries, Sessional Research Event, 25 March 2019, Edinburgh]

### Abstract

This abstract relates to the following paper:

Frankland, R., Smith, A. D., Sharpe, J., Bhatia, R., Jarvis, S., Jakhria, P. and Mehta, G. (2019) Calibration of VaR models with overlapping data. *British Actuarial Journal*, 24, e23. doi: [10.1017/S1357321719000151](https://doi.org/10.1017/S1357321719000151)

**The Chairman (Professor A. J. G. Cairns, F.F.A.):** Tonight, the Extreme Events Working Party presents their paper: Calibration of VaR models with overlapping data.

This is the latest in a series of papers by the Extreme Events Working Party. The Party has been going for 10+ years: our speaker, Gaurang Mehta, has been a member for the past 2 or 3 years. The party has produced several papers and is one of the most productive working parties that the profession has had.

In the paper, the authors compare the different ways in which one can obtain an estimate of the whole of the distribution of a quantity of interest with a 1-year time horizon with Solvency II in mind. Once a distribution has been found, it is possible to pick off the 99.5% value at risk (VaR). The starting point for this is a set of data, which is required for any piece of statistical modelling. The challenge is that this data set might not go back many years. Simple approaches to identify the 1-year-ahead distribution are susceptible to significant levels of sampling variation or parameter estimation error.

Is there a clever way to exploit the higher-frequency data, since it is monthly, to get an improved estimate of the 1-year-ahead distribution? This leads to the use of overlapping data, which is one of our main topics tonight.

The authors discuss the different ways that overlapping data can be used and compare the pros and cons of each leading to full stochastic modelling of the monthly data and generating forecasts for the 1-year return and so on. Full modelling has its own strengths and weaknesses, which the authors also discuss.

With that overview, our speaker is Gaurang Mehta, a member of the working party and one of the authors of the paper.

Gaurang is a consulting actuary and currently works for Eva Actuarial and Accounting Consultants Ltd. He works as a subject matter expert on market risks and internal model methodology and risk scenario generators. He is qualified as a Fellow of the IFoA and has a financial risk manager qualification from Global Association of Risk Professionals (GARP).

**Mr G. J. Mehta, F.I.A. (introducing the paper):** The paper was originally presented at the Life Conference in 2017 in Birmingham. Further progress has been made on various areas and this is an update.

First, I will describe the problem statement with which we started doing further work on the concept of overlapping data and treatment of overlapping data.

After that, I will give you the key conclusions that we have drawn from our analysis. Then, I will talk about cumulant estimation and the pros and cons of overlapping versus non-overlapping data. Next is the simulation study that we have carried out and the conclusions we have drawn from it.

We also explore the areas where we can use statistical tests for calibration purposes. Is there any smarter way of exploiting the annual data to ensure that we can get more information out of it, solving solutions to all living data that we have tried and explored?

So, what is the problem statement? Solvency II regulations prescribe that you should have a calibration that is sufficient for a 1-in-200-year event over a 1-year period.

In the UK, we believe that most firms have adopted a 1-year VaR approach. They are trying to estimate the calibrations that can produce a 1-in-200-year event on a marginal basis.

Just to give an idea of the problem, to do that level of statistically credible calibration, we need at least 200 years' worth of data. We do not have that in most of the market risks. Forget about insurance calibrations.

Insufficiency of the data is at the core of this problem. Therefore, the general market practice for calibration is to use annual overlapping data. To give you an idea of overlapping data, when you are calculating the annual change in overlapping data, 11 months are in common. The time series for 11 months is virtually the same. Only 1 month differs at every time step.

Therefore, the key question or dilemma that any calibrator faces is: what should I do?

When you go for longer series or longer historical data, that data are no longer relevant to the current context or the timeframe.

For example, a lot of liquidity data are available from government financial data series or from the Financial Times Stock Exchange (FTSE) indices. But, the market constituents are different and the structure of the market is different. Industries are different. Therefore, in the current context, historical data are no longer relevant.

You need to make a choice whether to go for 11 months' data, which is shorter, or longer data, which is not relevant. It is a choice. We are setting no hard and fast rules. Most probably, more relevant data are being used or retrofitted.

Once you have decided what data you want to go for, the next question is: will you use overlapping data or non-overlapping data? Overlapping data certainly give you more data points and therefore more data to play with. As a calibrator, you always like more data to play with rather than less. Also, it does help in the stability of the calibration in terms of justifying it to external auditors.

If you have used overlapping time series, then definitely you are introducing autocorrelation in your time series and other problems. For example, your data are no longer independent and identically distributed (IID). What can we do about it? Can we solve that problem? Can we do some adjustment: statistical testing, for example, or goodness of fit tests?

What alternatives can we explore? Can we explore something statistically more advanced? Can we make sure that, although we use high-frequency data, we can convert it into low-frequency data and use it? What are the implications for validation of the results in terms of the mean square error (MSE) for the estimate, or the bias for the estimate that we are trying to achieve?

A summary of our key conclusions is distribution fitting, and use of cumulants is common approach in industry; many practitioners use method of moments, but many use cumulants as well and then convert into method of moments to fit the data.

Members of the working party have different views on which is the best approach. Some believe that method of moments is superior for a certain range of distributions.

The key to using data is data points and stability of calibration. There is no problem of window selection for overlapping data. Non-overlapping data are IID, and therefore, the standard statistical tests are all valid. But then, there is a window selection problem for non-overlapping data.

The two possible solutions for correction of the bias in the data for overlapping data are Nelken/Sun and a Cochrane adjustment. We can also use temporal aggregation or annualisation

processes. Although they are useful, they have their own limitations. They make the entire calibration process more complex than it is already. It is hard to communicate to senior stakeholders what we are trying to communicate without any formula. Adjustments to standard statistical tests are also difficult to maintain. All the processes we are describing are good as a technical discussion, but when you are in a work situation where your calibrations have to be done within 3 months and be approved by the board, it is hard to maintain these processes, where you are changing your testing approaches, your calibration approaches and justifying them at the same time.

Starting with cumulants, these are used to do our simulation studies, to analyse the bias and the variance for both overlapping and non-overlapping data. They are very similar to moments – at least the first three moments are the same up to the third cumulant. The fourth cumulant slightly differs from the central moment. But, the first three are virtually the same.

It is not difficult to convert the fourth moment to a cumulant. The cumulants have nice properties because they are random variables. Therefore, we use cumulants in our simulation study.

There are two standard approaches to fitting the distribution. One is the method of moments or the cumulants. The second one is maximum likelihood estimate (MLE) approach. We believe that the MLE approach is statistically stronger, and therefore, it is a preferred approach.

In terms of analysing the distribution generally, we are interested in the 1-in-200 point. As the emphasis moved in Solvency II, for the entire risk distribution, we are interested in the body of the distribution as well as in the tail of the distribution. Therefore, not only the first two moments but also the skewness and kurtosis are equally as important to understand, and we need to make sure that we are always getting them correct.

Overlapping data have more data points, and there are problems of window selection. For example, in 1974–1975, the oil shock event happened over a span of just 4–5 months. If you analyse only January to January data on a non-overlapping time series, you might miss the oil shock event and the stress can vary by 10%, plus or minus.

When you are using overlapping data, all the information is used in the data set. Therefore, you are calculating changes from January to January and February to February, and there is no way you can miss any event that has persisted for more than a month. That is a very important feature of the data where you are trying to capture certain extreme events that have happened over a short time.

Therefore, the expert judgement about window selection that is there in the case of non-overlapping data is gone if you are using overlapping data. That is very important, particularly for credit risk, where the 2008 event is the only tactical event that we observed in the data. The 1932 crisis was a default crisis not a spread crisis. The 2008 crisis was a spread crisis. That is the only crisis that is there in the data.

Therefore, it is important to ensure that the window selection problem is not there. All living data help with that.

In terms of disadvantages, 11 months is in overlapping data, and therefore, you can say that they are autocorrelated. We also observe bias in the data when using overlapping data.

The standard statistical tests are no longer valid when you are using overlapping data. There are two options. Either you correct or adjust your standard statistical tests or you just use them as indicators and not as a decision-making test and use other statistics to accept or reject a distribution.

On the other side, you are directly calculating your annual changes, so there is no need to adjust to a 1-year VaR framework. Theoretically, it is more accurate. Estimates are biased in both cases whether you use overlapping or non-overlapping data. The main problem is there is not enough data to do statistically stable non-overlapping risk calibration.

Every new data point you add is only one data point every year and that can change the calibration year-on-year significantly, depending on what timeframe window you are selecting and how typical that event is for the entire risk distribution.

Information is possibly missing from the existing data so you are not utilising the entire data set.

We have tried three bias corrections. One is Bessel's  $n-1$  correction. There are two other adjustments that are proposed by Cochrane and by Nelken & Sun. They give very similar results. None of these bias corrections removes arbitrary dependence, for example, mean reversion characteristics of the data. They do not remove that type of bias. They can remove only statistical bias that is there in the data.

No matter which bias correction method you apply, we observe that it increases the variance of the time series. When it increases the variance of the time series, the stresses that you get overestimate the actual event and therefore the stresses will be higher when you apply these corrections.

You can argue whether it is the correct approach or not but, generally, that is the behaviour that we have seen.

We did a simulation study using four processes: Brownian motion, Normal Inverse Gaussian, autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH). The Brownian motion and the Normal Inverse Gaussian processes give similar results. ARIMA and GARCH are slightly different in terms of results. The methodology of the simulation study was described in the Ersatz Model paper.

I will give you a brief overview of our processes. We selected any known distribution or process, for example, a Brownian process. You perform arbitrary simulations of the data. From the arbitrary simulated data, the monthly data, you calculate the overlapping and the non-overlapping time series that you are interested in. Based on that, you calculate all the four cumulants of the data and then calculate the bias and the variance with the true values that you know from the start and see which one is giving you a better result. You do this for several years and over a number of time periods.

We have represented this for up to 50 years. You take  $N$  years of monthly simulated data, calculate the overlapping and non-overlapping time series, calculate the first four cumulants and compare the bias and MSE for both.

Whether you use overlapping data or non-overlapping data, the bias is there in the initial years. But as the number of years of data goes up, the bias is gone from the data.

Similarly, for the Brownian 1st cumulant, the MSE is similar whether you use overlapping data or non-overlapping data. The non-overlapping data are slightly lower than the overlapping data.

For calibration purposes, the first cumulant is not that important. The most important cumulant is the second cumulant, which in most cases has the value of the stress, particularly the 99.5 percentile in which you are interested.

Looking at second cumulant charts, we tried Neiken, Cochran and Bessel corrections for the bias. The time series in its ordinary form, whether it is overlapping or non-overlapping data, leads to virtually zero bias in proper time series. It works for the second cumulant only for the bias correction.

When we compare the MSE, the overlapping data MSE is considerably lower than the non-overlapping data MSE. That tells you that using overlapping data is closer to the known answer or the correct answer. That was one of the key conclusions that we drew from this.

This is similar to all the processes that we have tried, whether it is a GARCH process, ARIMA process, Normal Inverse Gaussian or Brownian motion process.

For the third cumulant, both overlapping and non-overlapping data, without doing any sort of correction, have bias in the data. But, they are in the same direction, so it does not matter materially.

The important factor again here is the MSE value. Again, for overlapping data, the MSE is considerably lower. Therefore, use of overlapping data is closer to the right answer.

Finally, for the fourth cumulant in the Brownian process, bias is present. As you increase the length of the data, it reduces. Again, overlapping data have a lower MSE and are therefore closer to the right answer for the Brownian process.

Similarly, for the GARCH process, other than the third cumulant, the MSE is slightly better, particularly for earlier years. Bias corrections sometimes work; sometimes do not. Even for the second cumulant when the process is complex. Here, you can say that they are overestimating the biases when you apply bias corrections.

Again, overlapping data MSE is lower even for a complex process like the GARCH process as compared to non-overlapping data.

The third cumulant MSE is lower for overlapping data in comparison to non-overlapping data. Even the bias is slightly lower over a time period.

For the fourth cumulant, under the GARCH process, the results may or may not be reliable. We have tried ourselves to reduce them. What we are seeing is, even for the fourth cumulant, using annual overlapping data is more helpful than using non-overlapping time series.

The conclusion is that the bias corrections do work but not always for all the processes. We only tried second cumulant bias corrections. Bias corrections are available for the third and fourth cumulants as well, but they are not included.

Applying bias corrections, we have observed, generally increases the variance and therefore the stresses. Bias corrections do not work for mean reversion processes that are arbitrary.

The ARIMA model has similar results to the Normal Inverse Gaussian process and the Brownian process.

Generally, the bias correction errors, even for GARCH processes, led to higher stresses overall.

The Kolmogorov–Smirnov (KS) test assumes that the data are independent and the distribution is a known distribution. Those are theoretically sound assumptions, but we are facing the problem that the data are overlapping and the assumptions for independence are not valid.

There is a correction to the KS test, which is what we have tried to implement and which is based on something called a KS distance: this is the largest distance between your fitted distribution and any of the data points.

The five steps are too technical to describe here. They are in the paper. Essentially, you fit any known distribution to start with. You simulate the data from a known distribution for several times on random trials. Based on that simulated data, you refit the standard distributions to it and calculate the KS distances. You derive the distribution of the KS distances. If the final distribution that you are fitting is higher than the KS distance, the 95th percentile, you say that the model is not good enough. That is the test structure that we have used.

We applied a similar structure to a total of five tests. In the base case, the result is 4.3%. There is some bias as it is closer to 5%.

Then, we applied the simple correction to the KS test. You are creating multiple samples of the known distributions and calculating KS distances.

In the third test, when you apply the corrections to the sampling error, using those five steps, you can see that it is about to reject up to 5% of the tests where the distribution is not good enough.

When you apply the same test without doing any correction of data problems, you could see that the rejection rate is much higher at 44%. In the third test, when you apply the correction to the sampling error, using those five steps, it is about to reject up to 5% of the tests where the distribution is not good enough.

Then, you apply where the test data and the sample data both are overlapping and make sure that it is on a consistent basis. Then, you can see that you are getting back to 5.3%, which is closer to 5%.

So, linear adjustment does work for both sample error correction as well as for the overlapping data problems. We only tried the stable processes here. We have not tried the non-stable processes. Consequently, it may or may not be valid for complex linear processes.

The hypothesis testing ultimately cannot describe whether the fit is to a better distribution or not a better distribution; it is just a yes or no answer.

Can we see some alternative smarter ways of utilising the same data set? The first thing that we tried was annualisation transformation. The first step is to take the monthly non-overlapping time series. Then, we try to fit a distribution to the time series. There are various ways. You can calculate the monthly changes and find out the correlation in the monthly time, so you get all the changes from, let us say, January to January, February to February, and try to find the empirical correlation in that time series; or try and use a copula to make sure that you create distributions of the monthly time steps or create a large number of time steps and then calculate the annual changes from the data set.

Here you can use an empirical copula, but then you need to estimate an extra parameter. That is a hard process and probably spuriously accurate.

Depending on what complexity you are interested in, you try to measure it to a copula and then simulate the time series. Based on that you can do the fitting of the distribution. That has a key advantage as you utilise all the information that is there in your data and you are not missing any information.

You have created a large sample, and therefore, you give this to any standard statistical package like R and it will give you much better stable results and generally no errors.

At the same time, because you have used a large sample of the data, the stability of the calibration year-on-year is going to improve. There are 11 points adding to the process, and therefore, unless there is a massive crash or something like that, the results are going to be stable year-on-year through this process.

A key disadvantage is that you are not avoiding autocorrelation and, therefore, the standard classical tests, Akaike information, the criterion, will not be valid for this way of doing things.

For this, we have tried it on credit risk. UR30 is a Merrill Lynch single A index. Using the overlapping data, there is very gradual decay in the autocorrelation function. There are similar aspects in the partial autocorrelation function (PACF) where it is sinusoidal and sometimes goes above the 5% line.

The PACF, when you apply the transformation, reduces suddenly and remains within the 5% range.

You compare the two blocks. If we use the overlapping data as it is, the tails are not good enough; the body is good enough, but the tails are not a good enough shape.

When you apply the process, you do get a distribution that is a reasonably good fit to your data. That is what we have seen particularly for credit risk. We have shown here one index, but we have tried multiple indices from the Merrill Lynch data set and we drew similar observations.

The second alternative, which is statistically more complex but a credible approach, is temporal aggregation. The main problem with doing temporal aggregation is that it becomes very complex very easily.

It is hard to use the formulas where you carry out calibrations on a quarterly or six-monthly basis.

The example we covered is for equity risk. You use the same monthly time series. The stressed values will be higher on the 12-month line as compared to any lower frequency line. This is a feature of the way that you do temporal aggregation. This is a formulaic approach.

You are using monthly annual non-overlapping data, and therefore, all those criteria for the standard statistical tests are valid and you do not need to do any corrections.

Because the number of transformations involved in doing temporal aggregation is materially more than any annualisation process or using standard processes of overlapping data, you miss a lot of information in between the transformations that are required to do temporal aggregation.

The key advantage of doing temporal aggregation is it handles data with clustering problems or where a big event has happened, and therefore, a similarly big event is likely to happen, very easily. The complexity is the main problem in doing temporal aggregation. It is hard to maintain these models.

The results, using temporal aggregation, are based on liquidity data. For a simple process, the calculation time using a computer is around half a day to do the temporal aggregation approach. For monthly overlapping time series, in the tails the fit is not particularly good. In the body, the fit is acceptable, but not in the tail. With the temporal aggregation approach, the fit is very good, even for the tails, when compared to non-temporal aggregated time series.

We can argue which one is more correct, which one is not. It is an expert judgement and you need to convince the stakeholders which approach to believe in because the temporal aggregation time series generally underestimates the downside risk but overestimates the upside risks. That is what we have seen, both using the auto-regressive process and the GARCH process.

Now, using the same data set but using some empirical GARCH (1,1) process, the autocorrelation function (ACF) is materially improved when you do the temporal aggregation of the time series. You get high quantiles if you use temporal aggregation time series.

In summary, there is a constant struggle between 11-month data and longer data sets, and it is ultimately an expert judgement that is not avoidable. We believe that the bias correction studies do help in terms of correcting for the bias but directly or indirectly increase the variances of the time series and therefore can overestimate the true quantiles.

The MSE that we observed in our simulation study is lower for overlapping time series. Therefore, using overlapping data is acceptable where the data are limited and you are trying to come up with a 1-in-200-year type of calibration approach for insurance purposes.

We have seen possible solutions to the problem using an annualisation process and using the temporal aggregation process. They both lead to materially improved fits whether we use criteria such as ACF, PACF quantile–quantile (QQ) plots or measures such as actual distances between the empirical data and the fitted distribution.

Therefore, you can say that they lead to more stability of the calibration year-on-year. That is what our observation is because we did try removing 1 year's data, 2 years' data and performed the same activities.

But both the methods that we used have their own problems in terms of complexity and communication of the process. Temporal aggregation solves the problem of autocorrelation, but the annualisation process does not solve the problem of autocorrelation in the data.

**Dr D. Stevenson, F.F.A.:** You talked about the temporal aggregation approach, which means setting a time series and projecting a 1-year distribution from that. For these models, there is a wealth of statistical tests for the overlapping data problem.

It seems to me the volume of fitness tests is quite limited. You mentioned the KS test, which has been adapted to deal with overlapping data. However, my understanding is that at present, the critical values that are available are based on normal distributions and more work is required to develop those tests to deal with the typical distributions that are used.

Do you have any comment on the robustness of the statistical tests for use in overlapping data; how might those be seen by regulators or other stakeholders? Is it really just ease of communication that is the driving factor for choices?

**Mr Mehta:** You have correctly identified the problem that is there for correcting any of the hypothesis testing. It is true that the tables are for standard processes. If you are using complex solutions, you need to come up with your own standard tables.

When you choose to change the distribution, you need to change your hypothesis testing standard tables, depending on what distribution you choose. Therefore, it is not easy to maintain these sorts of processes in a regular day-to-day calibrator's life.

The key question that you asked is: what do you do about it? These tests are there for communication purposes.

As a calibrator, as a practitioner, I have not used the results of these tests to decide whether to select a distribution or not.

Ultimately, we present our results through various distributions and let an expert panel make a decision based on what fits well; whether you are concentrating on the body of the distribution or

the tail of the distribution; whether you look at QQ plots as a more reasonable way of assessing the fit of the distribution or whether you look at the stability of the calibrations year-on-year. When you remove 3 or 4 years of data and reapply the same distributions that you have tried, which one is fitting well?

Those are other ways of assessing the goodness of fit, which are more practical than relying on standard statistical tests.

**The Chairman:** I could add to that. Just reading the paper and listening to the presentation, one of the advantages that we have now is we do not need to look at the old-fashioned printed tables.

As you did in the paper, you can simulate the distribution of the statistics under a variety of models and they all produce different results. You need to do enough repeated simulations in order to get a good estimate of, for example, the 95% quantile.

In the paper, you mentioned using 1,000 repeats. Is that enough? Or should it be 10,000, say?

**Mr Mehta:** Certainly, 1,000 is not enough when working for a company in a proper setting and would suggest at least 10,000 samples.

**The Chairman:** I guess that the 1,000 simulations can be used for a few different models just to get a first feel for which model is preferred. Then, you can maybe do some further tests, although I also take your point that ultimately you do not place your faith in just that one test. There are many other, maybe slightly more subjective, elements that go into the discussion. You should never use just one numerical or quantitative feature to decide between models.

Why did you choose cumulants rather than other properties, or distributions, in order to do your fitting?

**Mr Mehta:** For fitting, we have used two approaches, method of moments and MLE. We chose cumulants because in our previous working party paper on the Ersatz model, we used cumulants.

Also, if we use cumulants, there is more statistical evidence on the subject, which might be useful.

We could have used moments. The conclusions would not have differed materially, either way.

**Mr C. J. Turnbull, F.I.A.:** I come to this with a reflex scepticism that overlapping data can tell of anything new that is not already in the non-overlapping information.

To help me explore your argument for why we should use overlapping data, could you take this logic to a limiting place? Let us suppose that we are dealing with, say, equity returns on the FTSE – something where there is lots of data available. An example you gave looked at monthly returns, but for an asset like that you will have time series data which could be available daily, hourly, by the minute and perhaps now even more frequently than that.

If there were minute-by-minute returns available, would you advocate using that in estimating the 1-year distribution? If so, why? If not, why not?

**Mr Mehta:** We did consider this question but we have not explored it practically so I will not be able to tell you the results of the studies. When we explored this, we saw that people are using monthly data as the highest frequency data, and not daily data, on any of the market risks so far.

I do not see any reason why we cannot use them. I cannot tell you upfront what the results would be.

From a theoretical discussion, and from the work that we have done so far, I believe that the higher the frequency of the data you use, the more transformations that are required. From daily data or minute data, you need to convert it into hourly or higher frequency or lower frequency data overall, so you are doing many more transformations in the process. You will lose information at every stage of the transformation.

Therefore, whether more information is gained by going into lower or higher frequency data, we will not know, unless we do the analysis.

**The Chairman:** One of the things that you mentioned was, looking at the credit data, the big shock in 2008. The challenge there is that one big event happened over 20 or 30 years. That 30 years is going to make a big difference to the implied frequency of that type of event.

Do you have any comments in terms of how that might play out? Is there any method that can do a good job of capturing the chances of a similar thing happening in the future?

**Mr Mehta:** I would like to learn about that but, so far, we do not know whether there is any method that can accommodate a 2008 type crisis.

My experience is varied on the subject. Before 2010, we could say that the 2008 type of event is more than a 1-in-200-year event. It is a 1-in-250- or 1-in-300-year event. By the time the European sovereign crisis happened, we all had to change our beliefs and convince ourselves that these events can happen in less than 1-in-200 years. Therefore, it had to be part of the 1-in-200-year calibration.

That is a big shift. We used to previously argue that the default crisis was a liquidity crisis, and therefore, we looked at it in a different way and different forces are affecting them.

For credit risk, the data are scanty. The 2008 event was such a large event that probably no distribution fits well. You cannot compare liquidity data with the credit data because the market structure is totally different. The way things are assigned is constantly changing. After the 2008 event, rating agencies were forced to change their approach to the ratings.

There is a whole host of different arena in which we are working these days.

**The Chairman:** Is there anybody who could comment on how they deal with that period of data?

**Mr P. O. J. Kelliher, F.I.A.:** On the credit data, if you were to look at say Merrill Lynch and iBoxx data, there are significant differences. This is a problem with the data itself. The prices quoted back then did not have a huge amount of credibility.

To clarify my understanding, would it be fair to say that, for non-overlapping data, once you have made that  $n-1$  correction, there is not bias in the variance but there is a much larger degree of error compared to overlapping data, where there is a bias, but there is a much lower error. Is that a fair assessment?

**Mr Mehta:** The sample size in the non-overlapping data is materially smaller than the sample size that you get in the overlapping data. Therefore, for a known distribution, you get a lower MSE for overlapping data.

Credit risk is a problem in multiple dimensions. One of the three most commonly used data sources is Moody's data source, which is mainly based on US data from 1919 onwards. It is not 100% nice data to work with, in my practical experience, and is not directly relevant to the UK experience that we are generally trying to calibrate it to. It is more of a default crisis and the data are totally different in that era.

In terms of financials and non-financials, the time dimension is not there. Only the rating dimension is there. Therefore, people were using Moody's, and I have seen them adapting the stresses to month 11 data.

The issue is the level of granularity that is available. Most of the data sources have ratings in term dimensions but not in industry dimensions, for example, the automobile industry or the utilities industry. That level of granularity is not available in certain data sets. You might get slight differences using either of the data series as well.

Also, not all the data sets have all the risk measures. That worsens the problem for calibrators.

**The Chairman:** If you had the IID data, clearly higher frequency data will get you better estimates of some parameters, but not all of them.

Is that something that is generally true when you go to models where there is more autocorrelation or, indeed, when you go to IID data? When you look at the other cumulants, do they also become more accurate?

**Mr Mehta:** We experimented only with standard data sets. We have not seen materially higher autocorrelated data. So, we do not know about that. Practically speaking, higher frequencies have better chances of getting you to the right answer than using lower frequencies data. That is what we have observed so far.

**The Chairman:** If you use weekly or daily data, would you get substantially better results or would you just get a little better than you have managed so far with the overlapping data?

**Mr Mehta:** The balance that you need to strike is how much information you are going to lose when you use high-frequency data to convert that into low-frequency data with the amount of information that you are gaining. It depends on how many parameters you are estimating throughout the process and what freedom you are losing on the way.

If you are using simple models, the model parameters you are estimating at every transformation stage are few, so probably you might get better results for using higher frequency data.

**Mr R. D. MacPherson, F.F.A.:** It sounds as if you are looking at equity returns in nominal terms. I am just wondering whether there is any benefit at looking at things in real terms, especially if you are considering a much longer period. Obviously, in the UK, we have been targeting inflation at 2%, maybe, over the past 20 years, say, but we if are looking at much longer data sets, is there any merit in putting inflation into things?

**Mr Mehta:** We have used total returns, but still you can say they are nominal returns. We have not corrected them for inflation. The key issue is that inflation risk itself is a time series that you need first of all to model. In estimating the time series itself and in correcting that for inflation, we might introduce some model error and parameter error.

We have not tried it but it is a good point to take away and think about. What conclusions would we get if we corrected it for inflation? The one thing that we have done is to ensure we have corrected it for the historical mean of the data. We have removed the historical mean from the data to ensure that the bias is not there.

Whether it will impact on any of the cumulants or not, we have not tested.

**The Chairman:** When you say you are taking the historical mean, is that taken as a constant over the whole period?

**Mr Mehta:** Yes, it is a constant you take out from the entire series. Therefore, the mean of that time series is virtually zero.

**The Chairman:** You could equally, with the historical data, either look at the real returns or at equity returns, or whatever, in excess of the risk-free rate of interest as a further variant.

**Mr Mehta:** The spread we have used is over the swap rate or the gilts rate, whichever country you are using it for. It is above the risk-free data. But, it is still nominal data that we are using, not inflation corrected or inflation data.

**The Chairman (closing the discussion):** I presume that this depends on the particular time series or data set you are looking at as to the relevance there. The methods I assume could be easily adapted to extract one thing or another.

I will briefly try to summarise our discussion. We have had a presentation from Gaurang, then a few questions. David Stevenson was talking about the tests for the different approaches and that these might be limited. The point there is that with computing power, we can be much more adaptable. Equally, we also need to be mindful that the formal test results are just one part of a model selection criterion and that so long as you are not miles above or below the critical threshold, then there are other things that you might want to think about.

Craig Turnbull talked about the frequency of the data. Should it be just monthly? Should we use even higher frequency data? Certainly, high-frequency data to me would mean more frequent than daily rather than monthly.

That also links to whether you want to go down this route of using the overlapping data versus a full-blown stochastic model.

We had some discussion about the credibility in the credit data from Patrick Kelliher and about the credit crunch and how it might impact on the analysis.

Lastly, we had some comments about how you might adjust the data before you put it through the modelling process from Roddy MacPherson. Should you use nominal or real returns or returns that are in excess of the risk-free rates?

Of course, there are other data sets where you may be looking at credit spreads and other things, so it is not returns that you are looking at, but maybe other issues.

Finally, it remains for me to express our thanks to the authors and in particular Gaurang for his presentation and for all of those present who participated in the discussion.