MODELLING EXTREME MARKET EVENTS

A REPORT OF THE BENCHMARKING STOCHASTIC MODELS WORKING PARTY

By R. Frankland, A. D. Smith, T. Wilkins, E. Varnell, A. Holtham, E. Biffis, S. Eshun and D. Dullaway

[Presented to the Institute of Actuaries, 3 November 2008,
and to the Faculty of Actuaries, 19 January 2009]

ABSTRACT

This paper focusses on some practical issues that can arise when developing methodologies for calculating benchmark figures for extreme market events, particularly in the context of the Financial Services Authority's ICAS regime. The paper limits discussion to equity and interest rate risks. Whilst not intended to constitute formal guidance, it is hoped that the material contained within the paper will be useful to practitioners. The paper acknowledges the role of prior beliefs in the choice of data to be used for modelling and its influence upon the ensuing results.

KEYWORDS

Modelling Economic Capital; Extreme Market Events; ICA; Capital Requirements; Market Conditions; Equity Modelling; Yield Curve Modelling; Interest Rates; Prior Beliefs

CONTACT ADDRESS

Ralph Frankland, Norwich Union, 2 Rougier Street, York YO90 1UU.
Tel: +44(0) 1904 452246; E-mail: ralph.frankland@norwich-union.co.uk

1. Introduction

1.1 *Terms of Reference*

1.1.1 As part of their capital modelling, insurance companies need to make assumptions regarding the distributions of extreme market events, such as falls in equity markets or changes in interest rates.

1.1.2 The Benchmarking Stochastic Models working party was set up with the aim of gaining a better and wider understanding of the various methodologies that could be used for developing benchmark figures for these extreme falls, particularly in the context of the FSA's ICAS regime. This paper sets out to describe the practical issues that can arise in such an exercise, the type of data available, what methods can be used and the benchmarks that result when a selection of distributions are fitted to a typical data set for equity returns. In this paper we have limited ourselves to discussing equity and interest rate risks as it is common for these to be material for insurance firms and the data available is comparatively rich.

1.1.3 The results set out in Section 6, and the material described in the rest of this paper, do not constitute formal guidance of any kind. However we hope the material described will be helpful to actuaries and other professionals involved in this type of work.

1.2 *The Role of Prior Beliefs*

1.2.1 Any statistical estimate is a combination of data analysis, prior beliefs (see Rebonato (2007)) and an estimate of the uncertainty in the estimate. The prior beliefs are traditionally expressed in the choice of models, and the data they are fitted to. For example, fitting techniques may reflect beliefs that interest rates are stationary processes, that equity data can be pooled so that losses in one economy provides information about possible falls in another economy, that shifts over one year can be constructed from twelve independent shifts over each month, or that the tails of a risk distribution follow asymptotic power laws.

1.2.2 Although in theory statistical estimates can be made without prior beliefs (a pure 'frequentist' approach), in practice there is rarely sufficient data for the type of market investigations we are concerned with. The relative importance of prior beliefs varies with the availability of data and the object of the analysis. In the context of the FSA's ICAS regime, we are faced with estimating probability distributions that give meaningful results at the 0.5 percentile for movements in the market, from data series that, at best, contain only one hundred non-overlapping observations, dating back to times when the global economy was very different from today. In such circumstances the role of prior beliefs is significant.

1.2.3 This situation can be contrasted to that prevalent in the banking sector, where the calculation of 10-day value-at-risk is standard. There are two hundred and fifty trading days in a year, so even ten years' history will yield two hundred and fifty non-overlapping observations. In such a circumstance analysis can be driven more substantially by the data available and the role of prior beliefs is more limited.

1.2.4 Different analysts adopt different prior beliefs, and for this reason, a range of estimates may be obtained even based on a single set of economic data.

1.2.5 We could, in principle, formulate the effect of prior beliefs using Bayesian statistics. Inference uses the posterior model, that is, a re-weighting of the prior distribution of models conditional on the data. Where relevant data is plentiful, the choice of prior distribution is relatively unimportant. The estimation of rare percentiles, on the other hand, relies on only a handful of data points. As a result, the choice of prior distribution is critical, and consequently, estimates vary wildly between market participants. Any estimate is substantially a reflection of prior views, rather than objective data.

1.3 *Arguments For and Against Standardisation*

1.3.1 There is much to be encouraged in allowing a diversity of views. Insights available from one perspective may be invisible when seen from another point of view. Competition between models allows techniques to adapt and improve over time. The occurrence of a freak event may discredit one particular model, in which case it is healthy to have in reserve a population of alternative models that can still account for emerging data. Furthermore herding around one particular set of assumptions may leave companies vulnerable to the same market shock.

1.3.2 However where many firms have to make an estimate of the same extreme market event (such as the 0.5 percentile equity return over one year[1]) there are some clear arguments for standardisation:

— It facilitates comparison between firms
— It prevents one firm seeking or gaining a competitive advantage by adopting weaker assumptions.

1.3.3 Standards need careful articulation. Standardisation is a useful pragmatic device to assist comparisons and reduce the scope for gaming a regulatory system, but it is not pure science. Intellectual honesty forbids us from denying the role of subjective prior views in formulating any standard. The fact that one particular set of views underlies an emerging standard is no excuse for impugning the scientific credentials of those with competing prior views.

1.3.4 In this paper we do not attempt to answer the question of whether standardisation is necessary or desirable. Either way actuaries (and other finance professionals) are faced with the problem of calibrating (more or less) the same extreme market events from (more or less) the same historic data and the aim of this paper is to present relevant material to assist in that process.

1.4 *Use of Historic Data*

1.4.1 In this paper we have taken the approach of fitting parametric models to the data in order to derive estimates of extreme percentile events. The choice of data, particularly the historic time period used, can have a significant impact on the results.

[1] We could have expressed this equivalently as the 99.5 percentile worst outcome.

1.4.2 Other approaches are possible, such as the use of expert opinion,[2] market implied volatilities or the use of a real-world calibrated Economic Scenario Generator. We include the latter as we understand that it is a commonly adopted approach, although we note that the providers of Economic Scenario Generators must address the same problems set out in this paper in deciding upon the asset models to implement and the calibration parameters to choose. If they prefer, practitioners can, of course, adjust the results from the empirical approach set out in this paper to take account of these other sources of information.

1.5 *What is New in this Paper?*

1.5.1 There is a very significant existing literature on the distribution of extreme market events. This stems largely from developments in the finance industry, and in particular the banking sector, where the calculation of 10-day value-at-risk is a common requirement. In this context, the volume of observations available supports a data driven approach to analysis

1.5.2 Within the insurance sector the focus is on market movements over longer time periods, typically one year. There is relatively little analysis in the literature that considers such long periods, and the 'stylized facts' commonly found in considering shorter periods may no longer apply.

1.5.3 This paper focuses on one-year movements. It considers whether these 'stylized facts' do in practice hold for a period of one year, provides an analysis of data on the distributions of market movements over such a period and, perhaps most importantly, highlights the wide range of possible conclusions that can be drawn from the same observations, given the paucity of data.

1.6 *Structure of the Rest of the Paper*

1.6.1 Section 2 contains a brief discussion of modelling economic capital, explaining why, in the rest of the paper, we focus on looking at the 0.5 percentile outcome in risk factors over a one year horizon.

1.6.2 Section 3 discusses percentile estimates based on conditional versus unconditional distributions; why it is important to be clear about the approach taken, and explaining the choices we make for this paper.

1.6.3 Sections 4 to 6 look at the issues surrounding equity modelling. Section 4 analyses MSCI data to study the mean, standard deviation, skewness and kurtosis of historic equity returns and how this varies between countries (showing a worrying lack of consistency). By looking at how these statistics vary across different time horizons we conclude that analysing more frequent data does not significantly improve our ability to estimate extreme

[2] By 'expert opinion' we are referring to the opinion of respected market observers rather than, say, expert mathematical analysis of historic data.

percentiles of annual returns. This section also calculates skewness and kurtosis from the longer Dimson, Marsh and Staunton (DMS) data set, and then goes on to use bootstrapping techniques to examines the estimation error involved in these calculations. This demonstrates that for those data sets which show extreme skew or kurtosis (including the U.K.), the confidence intervals around the estimates are often very large, questioning whether they are statistically significant.

1.6.4 Section 5 then looks at the estimation error involved in the derivation of extreme percentiles, using bootstrapping techniques to calculate 95% confidence intervals. Although bootstrapping in this way is a standard technique, we are not aware of it being used for this application before. The results show just how significant sampling error can be, even for a relatively data rich risk factor like equities.

1.6.5 Section 6 then fits a variety of distributions to the MSCI data to make estimates of the 0.5 percentile equity return over one year, comparing the results.

1.6.6 Section 7 looks at interest rates and yield curve modelling. We investigate correlations between different points on the yield curve and propose some formulas to capture the important features of historical data. Yield curve movements are a promising candidate for dimension reduction, and we investigate a number of techniques for reducing complex models of yield curve movements down to three factors.

1.6.7 Section 8 provides some concluding remarks and then additional information is contained in the appendices.

2. Modelling Economic Capital

2.1 *What is an Economic Capital Measure and What is its Purpose?*

2.1.1 Economic capital (or risk-based capital as it is also often called) is a concept that is widely recognised but not well defined. Hairs *et al* (2002) described it as:

> *Capital requirement determined in a (more or less) scientific way, having regard to the risks to which the business is exposed.*

2.1.2 While economic capital measures may vary in detail, they share the following core features:

— They specify a certain level of security. This is often expressed in terms of a confidence level or the strength of the stress tests.

— They are risk sensitive. That is to say that the framework captures the material risks to which the firm is exposed and the capital assessment increases as the firm takes more risk and decreases as mitigation measures are put in place.

— They are suitable for the management of risk. To meet this purpose, the model should identify and quantify the materiality of risks which give rise to the capital. The weights given to different types of risk (for example, equity risk and interest rate risk) should be proportionate to the risks they pose to the company.

2.1.3 This means that economic capital is an important tool in understanding and managing an insurance business. However the growth of economic capital measures has also been stimulated by developments in regulation. The FSA's ICA framework already demands an economic capital approach in the U.K.; Solvency II will offer the option of using internal economic capital models for solvency purposes; and even the standard approach under Solvency II is an example of a simple approach to economic capital as currently drafted.

2.2 *Alternative Measures of Economic Capital*

2.2.1 There are many different ways to construct an economic capital measure including choices of:

— *Time horizon* over which the range of risk outcomes are considered: for instance we can project the balance sheet development over one year, or project the run-off of the liabilities.
— *Measurement of assets and liabilities*: we can project assets and liabilities on a realistic (market consistent) basis, an accounting basis or a regulatory basis.
— *Definition of ruin*: we can measure solvency continuously, at a fixed horizon or some other set of times.
— *Risk measure*: for instance value-at-risk measures look at the capital required to meet solvency in x% of outcomes; tail value-at-risk measures look at capital sufficient to meet the average deficit in the worst y% of outcomes.
— *Confidence level or ruin probability*: that is the choice of x% or y% in the above examples.

2.2.2 The ICA submitted to the FSA needs to include an assessment "comparable to a 99.5% confidence level over a one year timeframe".[3] Other confidence levels and timeframes may be appropriate — indeed the FSA requirements imply that another confidence level should be chosen if that better reflects the firm's own risk appetite. However, a 99.5% confidence level over one year represents the easiest way to demonstrate compliance with FSA rules. Furthermore, the vast majority of the U.K. industry has chosen an ICA methodology based on one-year stress and scenario tests; and

[3] INSPRU 7.1.42R.

a one year, 99.5% approach is also aligned with current proposals for Solvency II.

2.2.3 So for the purposes of this paper we have chosen to restrict ourselves to the following framework: a one year projection of a market consistent balance sheet requiring capital sufficient to be solvent in 99.5% of outcomes.

2.2.4 As a consequence, in looking at the results presented in this paper we pay particular attention to the 0.5 percentile worst movements in equities and interest rates over one year. This is not to say that other percentiles are not important; the calculation of economic capital when there are exposures to multiple risk factors involves the convolution of several distributions and the characteristics of the entire distribution will play a role in the level of economic capital required. Nevertheless, the 0.5 percentile provides a simple benchmark with which readers are likely to be familiar.

2.2.5 In theory the same techniques could be used to make estimates for longer timescales and other confidence levels, although in practice it will be difficult to find data sets large enough to give reliable results. Although the techniques may be valid, the results in this paper should only be extrapolated to longer time horizons with great caution.

2.3 *The Need for Continuous Monitoring*

2.3.1 The ICA is the firm's own assessment of the amount (and quality) of capital sufficient to ensure that the firm maintains adequate capital resources *at all times*. We believe that this requirement is commonly interpreted to mean that:

— the ICA should be calibrated to the potential change in conditions that may occur from one point to a point one year ahead (discrete points) rather than the largest change that may occur at any point over one year (a continuous test).
— the systems and controls around the ICA should be used to make sure that the firm has adequate capital against this standard on a continuous basis. This second point is beyond the scope of this paper.

2.3.2 We have adopted this interpretation in this paper. This means that when considering extreme market movements over one year we are concerned with observations one year apart rather than the distribution of the lowest value during the year. It should be noted that this interpretation does imply that we need to consider the largest change that can occur from any point in time, rather than the largest change from, say, December to December. This motivates our choice of overlapping periods in the analysis set out in section 6, despite the statistical complexity this brings.

2.4 *How Should Capital Requirements Respond to Changing Market Conditions?*

2.4.1 When adverse market movements take place, available capital is used up to the extent that there is a reduction in asset values greater than any reduction in liabilities. Available capital reduces, but what, if any, impact should there be on capital requirements following such an event? For instance, in the case of equities, how should a sharp fall in equities over a recent past period affect our estimate of the 0.5 percentile outcome in the next twelve months?

2.4.2 One aspect of this question is to consider whether we are making estimates of the conditional or unconditional distributions of risk factors. This is discussed further in Section 3.

2.4.3 However, there is also a practical and commercial dimension to this question (especially where we are talking about the regulatory framework): is there a willingness to reduce capital requirements after a market stress in the interests of market stability? For example, to avoid the pro-cyclical effect of many financial institutions de-risking their balance sheets by selling significant portions of their equity portfolios.

2.4.4 One example of a rule that attempts to reduce capital requirements in times of market stress is the equity fall tested as part of the resilence capital requirement (previously the resilience reserve). The rule adjusts the equity fall tested to between 25% and 10% depending on the 3 month average level of the FTSE All-Share and the price-earnings ratio compared with 75% of the inverse of the 15-year gilt yield. Figure 1 shows the FTSE index daily since June 2001, the resilience equity fall and the stressed index level.

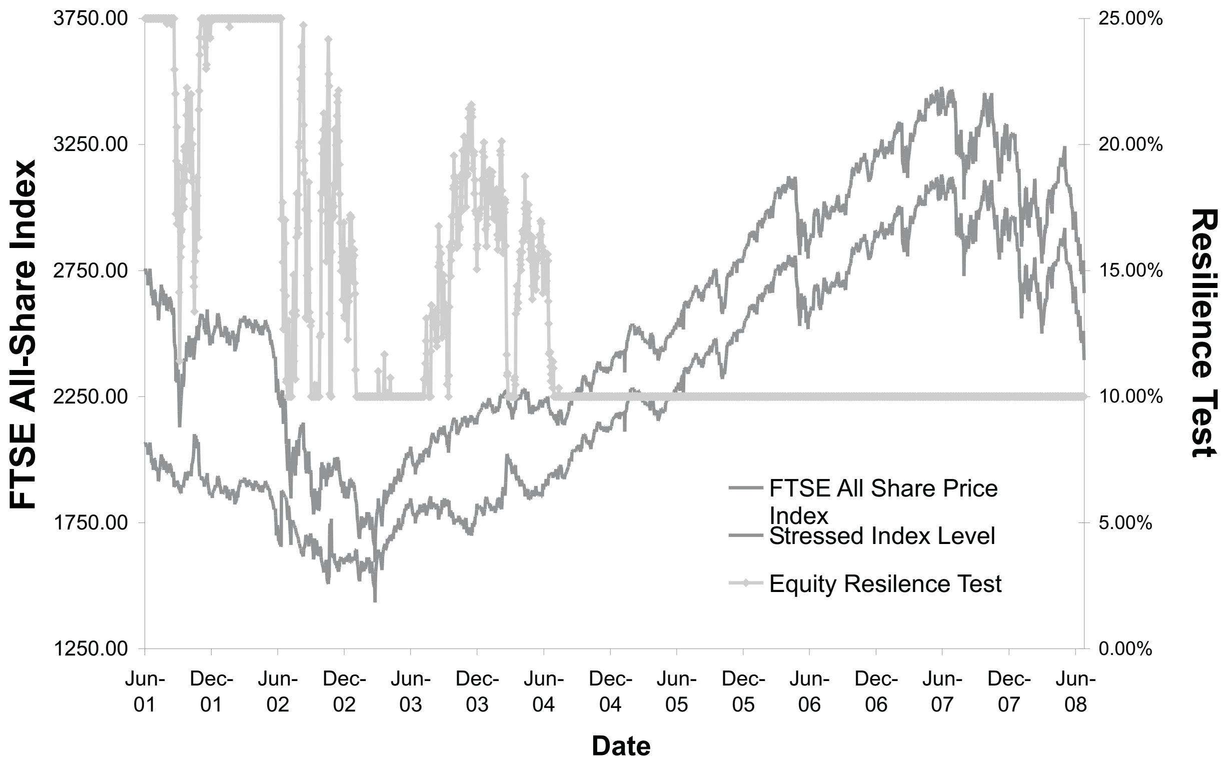

Figure 1

The rule appeared to cope with the market fluctuations over 2001-2003 but since 2004 the equity fall has been at the minimum 10% level because of a low P/E ratio. This illustrates the difficulty of constructing a rule that will cope with changing market conditions. This aspect of the question is beyond the scope of this paper.

2.5 *Stress and Scenario Testing*

2.5.1 In this paper we are concerned with estimating extreme market events at a particular confidence level or percentile, in order to then use the results to calculate economic capital requirements for an insurance company. Often we refer to these extreme events as stress tests.

2.5.2 Increasingly financial institutions are also expected to "think outside the box" in risk management, and to think up extreme scenarios and test what impact they would have on the company. Here the focus is not on calculating a capital requirement at a particular confidence level but to better understand the company's risk exposure. The scenarios tested may not be explicitly ranked at any particular probability of occurring, and some of them might be expected to lead to extreme difficulty or ruin for the company. This exercise may also be referred to as stress testing, although the phrase 'scenario testing' is more common. These stress tests are different in nature to those we are trying to calibrate in this paper, and are outside its scope.

2.6 *The Use of Economic Scenario Generators (ESGs)*

2.6.1 Within the economic capital framework we have chosen, ESGs have two potential uses:

(a) To simulate the change in market conditions over one year.
(b) To construct market consistent balance sheets at the outset, and after one year allowing for changes in market conditions.

2.6.2 These uses are quite different: in the case of (a) we want an ESG that captures extreme market movements over one year; in the case of (b) we want an ESG capable of a market consistent valuation. In this paper we are solely concerned with benchmarks for (a).

2.6.3 Within this framework the ideal theoretical approach would involve nested Monte Carlo simulations (at least for products with embedded options and guarantees): the first level of projection would involve simulating market conditions after one year; then for each such simulation there would be projections from that point onwards to value liabilities on a market consistent basis and construct market consistent balance sheets.

2.6.4 Although it may be possible to adopt this approach in some instances, the processing times are generally prohibitive at the moment, forcing a simpler approach to be taken. This can be either:

(a) simulating market conditions after one year using Monte Carlo simulation, then constructing market consistent balance sheets without e.g. by using closed form solutions to value options and guarantees; or

(b) applying a modest number of stress and scenario tests to determine adverse market movements over one year, then using Monte Carlo simulation to value options and guarantees and determine market consistent balance sheets.

2.6.5 In *either* case we need to measure the net assets of the firm on an economic basis (that is the economic value of assets less liabilities) as a function of various risk factors. To determine our capital requirement we need to make assumptions about the distributions of these risk factors over one year. Deriving these assumptions for equities and interest rates is the central theme of this paper.

3. Conditional and Unconditional Estimates

3.1 *Introduction*

3.1.1 A key question is whether we are trying to model what may happen over the next year (i.e. conditional on some or all the information we know about the current market) or an "average" year (unconditional on current market data).

3.1.2 If we do decide to condition on current data we have choices over what data to include in the conditioning, e.g.

— Starting price or rate data (e.g. the starting yield curve).
— Recent price or rate data (e.g. recent volatility).
— Prices of options and other market instruments (e.g. implied volatilities).

3.1.3 If we condition on a large amount of data then intuitively we may get a more "relevant" stress test. However, we have less data to work with as there will be fewer points in the historical data with the same conditions. If we use a "pure frequentist" approach to derive such a conditional estimate then our confidence interval will increase as the volume of data decreases. An alternative is to fit a model, such as a time series, that allows us to use more of the available data to derive the conditional estimate. However in this case the role of prior beliefs is perhaps greater than in the unconditional case since we are making assumptions about how returns (or yield curve movements) in certain market conditions are related to returns in other market conditions.

3.1.4 In the extreme we could condition on all market data up to the valuation date; in this case we would have no past data with the same conditions so a statistical approach would be useless. It could be argued that information on this conditional distribution is contained in implied volatilities. In theory these reflect the market's view of future volatility, but in practice the information on the "real world" return distribution may be distorted by supply and demand factors in options markets. This approach is beyond the scope of this paper.

3.2 *Conditional vs. Unconditional Distributions — Interest Rates*

3.2.1 Some of the issues can be illustrated by looking at a simple model for interest rates (ignoring some of the complexities that we would be concerned with in a more realistic exercise).

3.2.2 *Assumed interest rate process*

Let us suppose we have a universe where insurers are exposed only to the short term interest rate, R. Let us suppose also that R follows a Cox-Ingersoll-Ross process, so that in continuous time, under the real world probability law, R satisfies the stochastic differential equation:

$$dR_t = \alpha(\mu - R_t)dt + \sigma\sqrt{R}dZ_t.$$

Possible parameters are: $\alpha = 0.05$, $\mu = 0.05$ and $\sigma = 0.044$.

3.2.3 *Solvency regulation and assumed firm structure*

Let us suppose also that we operate in a world where all insurers are *long* interest rates, that is, whose net assets are an increasing function of R. All firms are therefore exposed to a fall in interest rates. This is a convention to illustrate a principle; we could alternatively have written this example with reference to insurers who are exposed to a rise.

3.2.4 Solvency regulation takes the form of a single stress test. The interest rate at time t is R_t, and firms are required to test a fall in interest rates at time $t+1$, to some lower value $L(R_t)$. We will explore different ways of choosing the function L.

3.2.5 We assume that a single new insurer is set up each year, and capitalised in such a way as to meet precisely the regulatory capital requirement. This means that the insurer established at time t is solvent at $t+1$ if $R_{t+1} \geq L(R_t)$ but fails if $R_{t+1} < L(R_t)$. At the end of the year, the insurer is wound up, whether solvent or not, and a new insurer established, capitalised in order to withstand an interest fall to $L(R_{t+1})$.

3.2.6 The regulatory objective for choosing L is that, measured over a very long time period, the proportion of insolvencies tends to 0.5% that is $0.005 = 1/200$. We will consider four algorithms that fulfil this regulatory objective.

3.2.7 *Four possible algorithms*

We consider four possible algorithms for specifying the function L:

Method	Formula for L	Definition of constants
1. Conditional percentile	$\text{Prob}\{R_{t+1} < L_1(R_t) \mid R_t\} = 0.005$	N/A
2. Unconditional rate	$L_2(R) = c_2$	$c_2 =$ 0.5 percentile of the stationary distribution of R_t
3. Unconditional difference	$L_3(R) = R + c_3$	$c_3 =$ 0.5 percentile of the stationary distribution of $R_{t+1} - R_t$
4. Unconditional ratio	$L_4(R) = c_4 R$	$c_4 =$ 0.5 percentile of the stationary distribution of R_{t+1}/R_t

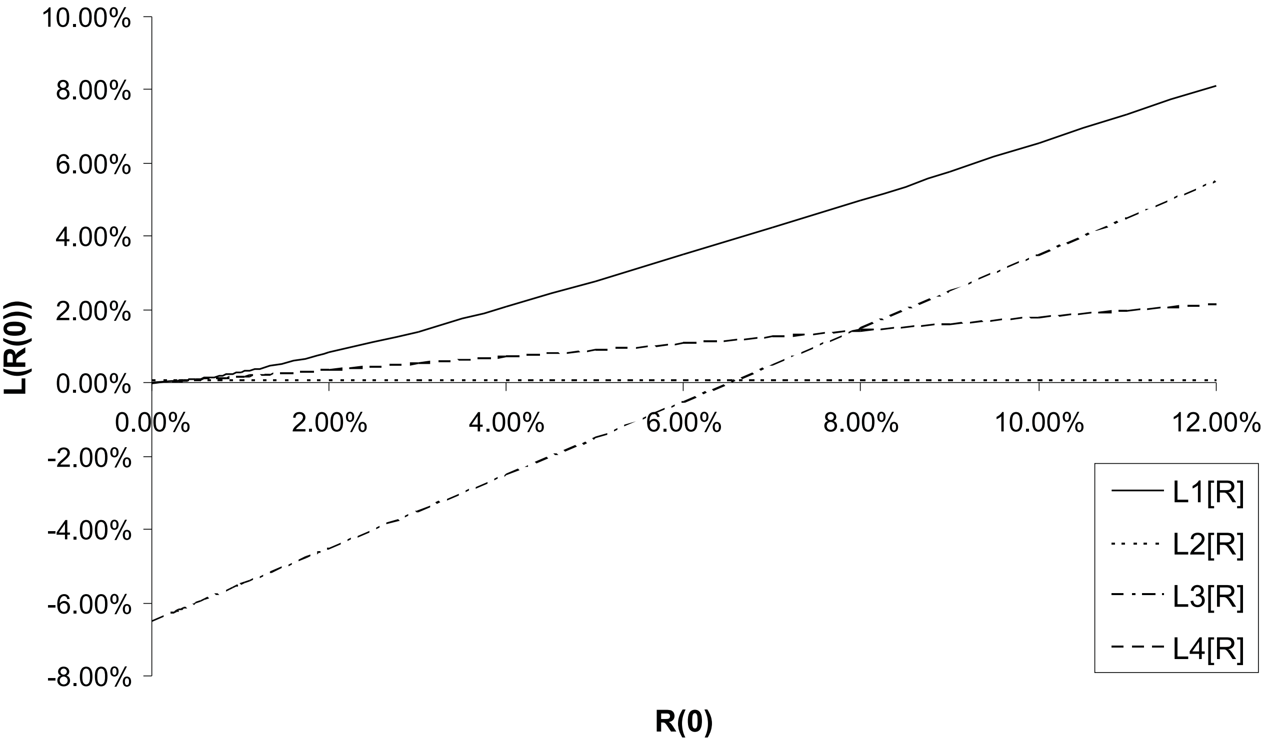

Figure 2

3.2.8 Figure 2 illustrates these four possible stress test constructions. The x-axis shows the starting short rate of interest and the y-axis the 0.5 percentile down-stress interest rate under each of the four possible stress text constructions.

3.2.9 In this chart you would not expect to see one test definition stronger than another in all cases — indeed the curves for L_1 and L_3 do cross but only at much higher interest rates (exceeding 20%).

3.2.10 We can see that, although in theory all four functions fulfil the regulatory objective of 0.5% failure over long time periods, the tests based on unconditional estimates may produce very strange results depending on the starting interest rate. L_2 produces a very extreme stress unless the starting rate is very low. L_3 can produce a stress down to negative rates (despite the fact that negative rates are impossible under the CIR model). Given these results L_4 seems a natural alternative to try, and it does indeed produce a more plausible shape of stresses, but they are still very severe. The underlying reason for this is that extremely high and low rates occur with much higher frequency in the unconditional distribution for this model (given the parameters we have used) than one would intuitively expect based on post-war U.K. experience — or perhaps any historic period measured in tens of years.

3.2.11 Of course these results are a function of the underlying CIR model we have chosen. Any estimate of unconditional distributions from historic data would not be so extreme — in large part due to the limited amount of data available. Nevertheless it illustrates the potential dangers of using a pure unconditional approach, and in this paper we choose to condition on the opening level of the yield curve. We do not, however, condition on any other market data.

3.3 *Conditional vs. Unconditional Distributions — Equities*

3.3.1 The position for equity return distributions is somewhat different. There is some temptation to believe that any significant fall in equity values begins to limit the scope for future falls so that the 0.5 percentile equity return over a year will tend to decrease following a particularly bad fall.

3.3.2 On the other hand it is common for equity falls to be associated with short periods where market volatility increases significantly (whether it is the actual observed volatility that is measured or the market implied volatility in option prices). This *volatility clustering* is a well-established market feature (at least for frequent data such as daily) and is discussed further in section 4.4. This suggests that it may be necessary to *increase* our estimate of the conditional 0.5 percentile equity return following a market fall.

3.3.3 Attempting to model equity returns (or indeed interest rates) conditional on recent volatility may reduce even further the amount of relevant historic data and therefore makes estimation error even more serious a problem (Section 5). Alternatively it may introduce the need to estimate even more parameters from the data available when fitting distributions.

3.3.4 Taking into account the additional data issues involved in estimating a conditional distribution for equities, whether conditioning on implied market volatility or any other measure, as well as the time and resources available to the working party, we have decided not to investigate fitting models (such as generalised autoregressive conditional heteroskedasticity (GARCH) models) that capture any time-varying volatility (heteroscedasticity) in the data. As a possible further justification for this approach, we note that some studies have concluded that volatility is not forecastable beyond a time horizon of a few weeks (see for example Christoffersen Diebold and Schuermann (1998)). However we have attempted to capture the fat-tailed property of returns. These features and their relationship are discussed further in section 4.

3.3.5 Although there may seem to be some inconsistency here in our treatment of equities and interest rates we believe that we have taken a reasonable approach, and one that is common market practice when setting ICA assumptions.

4. Equity Modelling — Analysis of Moments

4.1 *Introduction*

This section starts by considering some broad features of equity market returns.

We next describe a data set of equity market returns. We tabulate means, standard deviations and higher moments for a 1-year holding period.

We then investigate the effect of different horizons, tabulating moments and observing consistent patterns across markets.

Section 6 below describes ways of fitting distributions to those moments.

4.2 *Returns — Historic Data*

4.2.1 Historic equity data usually consists of indices, typically published daily. Indices are traditionally computed as the price of a share portfolio, whose constituents are adjusted from time to time. Other indices may be *total return*, that is, including an allowance for additional shares purchased by the reinvestment of dividend income. Compilers of indices also take account of a number of other transactions such as rights issues and share splits. For this reason, it is usually more straightforward to analyse indices than to analyse individual share prices, provided the analyst is confident in the quality of index compilers' work.

4.2.2 For an index P_t, we might consider a new series P_{t+h}/P_t as a series in t, for some fixed holding period h. It is at least plausible to consider a distribution of monthly returns, which might be broadly constant over time. In contrast, the distribution of an index level expands with longer time horizons.

4.2.3 Statistical analysis may be applied to *simple returns*, *log returns* or *annualised returns*. Where a holding period h is measured in years, these are defined as as:

$$\text{Simple return} = P_{t+h}/P_t - 1$$

$$\text{Annualised return} = [P_{t+h}/P_t]^{1/h} - 1$$

$$\text{Log return} = \ln(P_{t+h}/P_t).$$

4.2.4 Simple returns are most relevant for understanding portfolio constructions and risk-return tradeoffs because they combine linearly across portfolios. Simple returns may be difficult to model, however, because the relevant distributions must be supported on the range $(-1, \infty)$. The analysis of logarithmic returns overcomes this problem because log returns lie on the whole of the real line $(-\infty, \infty)$. There is a wide choice of classical distributions on the real line, potentially applicable to models of log returns.

4.2.5 For solvency purposes, regulation dictates we pay special attention to returns with $h = 1$ year. In this case, the annualised return is the same as the simple return.

4.2.6 The conversion of period-h returns to annual returns is performed by assuming a constant compound rate of investment growth. The calculation gives us ways of comparing returns over different periods on a common basis, and does not require us to believe that all investments grow at constant compound rates. By extension, we can also consider convolution of probability distributions. For example, given an assumed distribution for one month holding periods, we can "annualise" the distribution under the hypothesis that returns over different months are independent and identically distributed. We can also annualise particular statistical properties, such as means and variances. Annualisation does not require us to believe monthly returns are independent (empirically, we will see they are not), but rather offers a simple way to compare distributions over various periods in common terms.

4.2.7 Probably the simplest model of equity markets is to assume that returns are independent and lognormally distributed, i.e. that log-returns have a normal distribution. However the statistical evidence suggests this is not an ideal model in several ways.

4.2.8 Figure 3 plots daily price changes for the U.K. equity market since 1935. The frequency of daily absolute returns in excess of 5% is plotted using bars at the bottom of the chart. It can be seen that, for the U.K. market, most of the extreme returns are bunched together in the 1970s.

4.3 *Asymmetry and Fat Tails*

4.3.1 When looking at empirical log-return distributions for equity markets there are two noticeable differences to the normal distribution.

(a) There is negative skewness, i.e. the distribution is skewed to the left. There is a tendency for there to be larger downward falls in the equity market than rises.
(b) There is positive kurtosis, i.e. the tails of the distribution are significantly fatter than one would expect from a Normal distribution. The probability of extreme moves in the equity index is higher than would be predicted from a lognormal model. These large moves are often caused by (downward) jumps in the equity index at times of significant market turbulence.

4.3.2 Clearly these properties, particularly fat tails, are highly significant

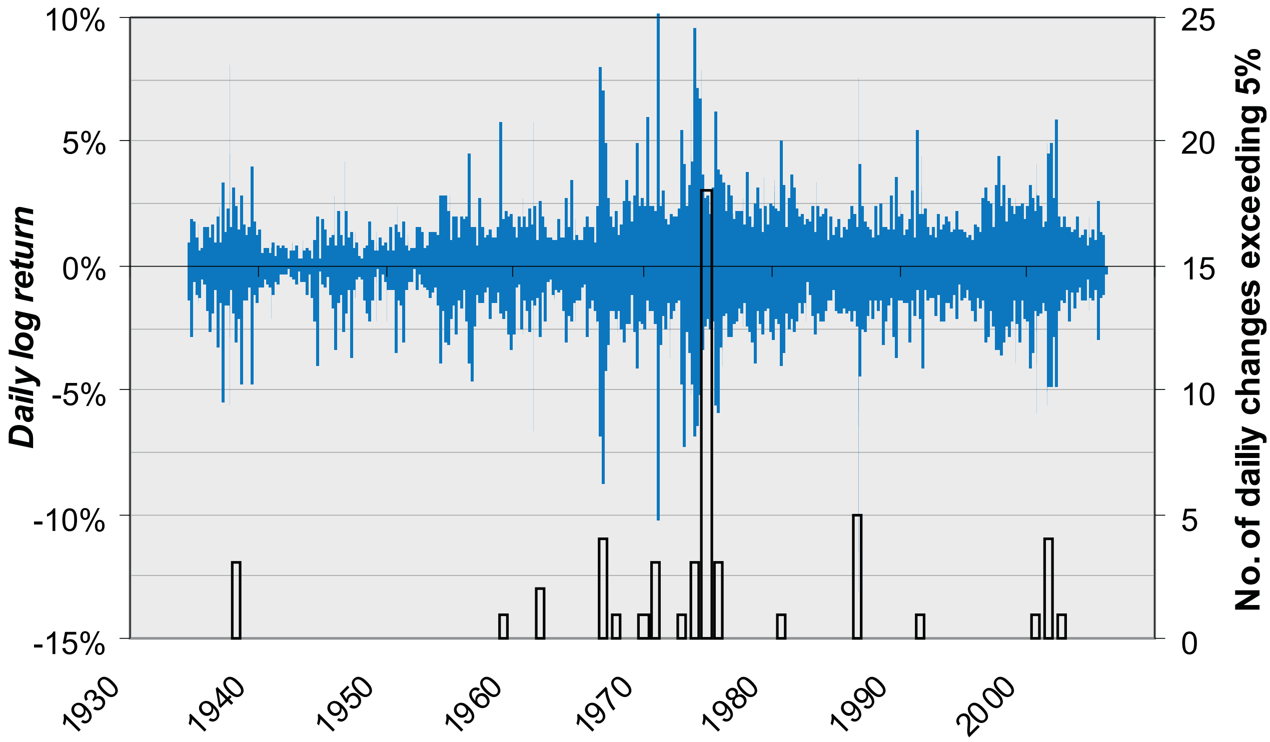

Figure 3

for an analysis of extreme events and lead us to consider using other classes of distributions that can reproduce these features. However the statistical evidence is mixed. Excess skewness and kurtosis are significant for returns over relatively short periods, for example we can see in Figure 3 a number of daily returns that are large multiples of the standard deviation, which would have a very low probability under a lognormal model. For annual returns the evidence is more mixed. The analysis in section 4.12 suggests that excess kurtosis can be observed at longer timescales, but in section 4.15 we show that the estimation error is significant and indeed that the excess kurtosis may not be statistically significant in most economies.

4.4 *Volatility Clustering*

4.4.1 Figure 3 illustrates the presence of periods of relatively low volatility interspersed with periods of relatively high volatility. This is usually described as volatility clustering or heteroskedasticity.

4.4.2 Several models have been proposed to capture this time-dependent feature of volatility. One popular class of such models is GARCH (generalised autoregressive conditional heteroskedasticity), in which volatility is modelled as a moving average of past volatility. As noted above, we have not attempted to capture time-varying features of equity returns in this paper so have not considered these models further.

4.4.3 Volatility clustering has been suggested as a possible explanation for fat-tails in equity returns. Equity returns that follow a log-normal distribution but with volatility that varies over time may appear to have excess kurtosis if the unconditional distribution is estimated over a long period. Kemp (2008) examined this effect in daily returns on selected equity indices and showed that time-varying volatility may account for most of the excess kurtosis in the upside tail but not in the downside tail. This suggests that large negative returns exhibit some "genuine" fat-tailedness arising from unexpected extreme outcomes, at least for high frequency data.

4.5 *Mean Reversion*

4.5.1 As noted by Exley *et al.* (2004) (EMS), this term is used to describe a number of different features. Simplistically the idea is that large negative returns will tend to be followed by positive returns and vice versa. EMS find some weak evidence for mean reversion in equity markets over longer time horizons but note that the effect is highly dependent on the choice of holding period.

4.5.2 Bouchaud & Potters (2003) find that price returns show strong autocorrelations over periods less than a few tens of minutes but little correlation for longer periods. They also find a negative return-volatility correlation over a time scale of the order of a week.

4.5.3 Combining these observations with the volatility clustering feature suggests a complex structure of behaviour over different time scales. Large negative returns will increase realised volatility and, because of the clustering effect, may increase future volatility in the short term. This suggests a higher probability of further large falls initially, but possibly higher expected returns over a longer period depending on the extent to which mean reversion holds.

4.6 *Tail-Correlation*

4.6.1 It tends to be the case that the correlation between short term moves in equity markets and individual stocks is elevated when prices fall sharply, i.e. in the far downside tail of the equity return distribution. In extreme market conditions "herding" behaviour may take over, so that in a stock market crash all stocks tend to fall together. This suggests that the diversification benefit from holding a portfolio of exposures to market indices and/or individual stocks is much reduced at just the point when it is needed.

4.6.2 This is an example of possible differences in the behaviour of equity returns between normal and stressed market conditions. Thus when we consider the distribution of returns we may expect differences between the central part of the distribution and the tails. Shiryaev (1999) suggests that the presence of multiple traders and investors with varying interests and time horizons implies that it is unlikely that a single "standard" distribution will capture all features of financial indices and that a better approach might be to use different standard distributions for different regions of their values. However for returns over longer time horizons this approach would make the amount of relevant data too small to perform any meaningful analysis, so it seems preferable to use data from all parts of the distribution. Kemp (2008) suggests an approach that involves fitting the observed distribution directly but giving greater weight to the observations of most interest.

4.7 *Stationary Time Series*

4.7.1 There is a large body of statistical theory in relation to *stationary* time series. Stationarity means that the distribution of observations, or clusters of observations spaced at fixed intervals, have a common distribution that does not depend on the time of the first observation.

4.7.2 Stationarity is not a credible property of equity indices. Equity indices tend to increase in time, and price uncertainty becomes ever greater with longer term horizons. There is no limiting distribution. To make the process stationary, and so bring to bear the associated statistical machinery, we convert the index into returns.

4.7.3 The best known stationary return model is the geometric random walk where returns over different years are independent and have a common distribution (so called "iid" — independent and identically distributed). The class of stationary processes includes iid models, but also many others. Stationarity can permit dependence between returns. Where such dependencies exist, a distinction emerges between the *unconditional* distribution of a return compared to the *conditional* distribution given returns in previous periods, or given information from other sources. For example, we could investigate effects such as *mean reversion* or *volatility clustering*, as described above. Wilkie (1986, 1995) describes many examples of such models in some detail.

4.7.4 We have taken the position of assuming that returns follow a stationary distribution. If we were to reject stationarity of returns, what might we have instead? We might have seasonality — it would not be surprising if returns in January were more volatile than in February, because February has fewer days. Or we could have an underlying long term trend, perhaps relating to growing expected human life time or to productive industrial inventions. However we do not pursue this idea and proceed with our analysis on the basis that the returns are in fact stationary. We do this because there are a much richer set of quantitative tools at our disposal for stationary series.

4.7.5 In what follows, we shall apply the concept of stationary returns to equity markets. This appears to be a majority view, although by no means unanimous, among econometric modellers. The assumption of stationarity is an example of a prior view, which inevitably affects any forecasts we make.

4.8 *Estimation of Stationary Models*

4.8.1 The estimation theory for stationary models draws on *ergodic theory*. A stationary model prescribes a common distribution for X_t which applies for any one value of t. This is a *cross-sectional* distribution, as it applies to all possible outcomes at a single point in time. *Ergodic theorems* clarify conditions under which the sample distribution of X_t converges to the underlying distribution as the number of counted time points increases. An estimate taken in this way, over one sample path but many time points is a *longitudinal distribution*.

4.8.2 This suggests that the empirical distribution of values of X_t over a time interval may approximate the stationary distribution of the underlying process. In the case where X_t is an iid sequence, this result follows from the strong law of large numbers applied to indicator functions. The extension of the iid case to more general stationary processes is helpful because it broadens the class of models for which we may plausibly measure longitudinal distributions and treat the result as approximate cross sectional distributions, provided the observation period is long enough.

4.8.3 The class of stationary processes is a very broad one. Because the class is so broad it is difficult to quantify the model error because so many different models could be selected. Possible problems include bias from taking too short a series, as well as variability from having selected a particular outcome. In practice, wishful thinking comes into play, with "long enough" conveniently assumed equal to the available data period. Our analysis uses data from 18 different countries. A comparison between the results for different countries can give some indication of the estimation error. The estimation error is understated, firstly because each country is analysed over the same period and so subject to the same world events, and secondly because of survivorship bias: we include in our sample only those countries with a long history, excluding such newcomers as Brazil, Russia, India and China.

4.8.4 By specifying a specific stationary model (for example, GARCH), much more can be said about estimation uncertainty, all predicated on the prior view that GARCH is the underlying model. The problem is that this fails to capture the uncertainty associated with the possibility that GARCH is inappropriate.

4.9 *Distribution Properties: Mean and Standard Deviation*

4.9.1 Our starting point in the analysis is a consideration of the sample mean and standard deviation of returns for each country.

4.9.2 We calculate these both for simple and logarithmic returns. These are shown in Figure 4 for all countries considered. For each line segment, the estimates based on logarithmic returns are at the bottom left and those based on simple returns are at the top right.

4.9.3 In Figures 4 and 5 we have used data from MSCI indices (See

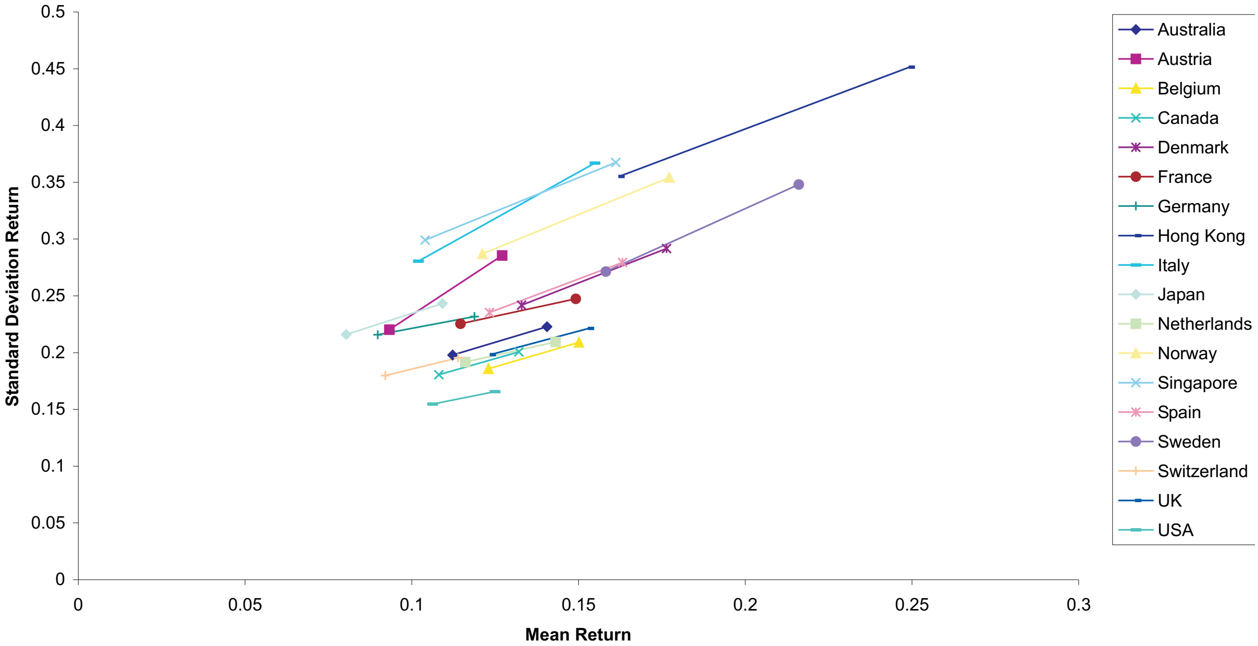

Figure 4. Annual return by country: simple and logarithmic

Figure 5. 1-in-200 year equity market falls based on normal distributions

Appendix A for more details). We have used 480 annual returns based on overlapping time periods separated by one month. This approach does lead to autocorrelation in the returns as each 12 month period will share 11 months of its return with both the subsequent and previous annual returns. This introduces a theoretical bias into estimates of means and other moments, but this effect is small relative to the underlying uncertainty in the estimates.

4.9.4 We can see a close relation between the statistics calculated on simple or logarithmic returns. The similarity reflects the first order expansion: $\log(1+x) \approx x$.

4.9.5 Fitted means and standard deviations do not determine the values of extreme percentiles. To estimate percentiles, we must fit a distribution. The choice of a distribution is where the prior views enter the calculation. One distribution that is well-known, simple to use and probably inappropriate is the Normal distribution. The 1-in-200 year event, or 0.5 percentile, is calculated as the mean minus 2.57583 standard deviations.

4.9.6 Figure 5 shows the computed 1-in-200 equity market falls by country. These are obtained either by fitting a log-normal distribution or a normal distribution to simple returns. We can see that the normal distribution fitted to simple returns produces larger estimated extreme events; in some cases much larger. This is unsurprising, and reflects the fact that the normal distribution has a fatter left tail than the lognormal. If we looked far enough into the tail, the normal fit to simple returns could produce infeasible values below the physical constraint of -100%, while this could never happen fitting distributions to log returns. This does not imply that lognormal distributions are the right answer, but powerfully illustrates the importance of prior distributional assumptions on stress test construction.

4.9.7 A fundamental question for consideration is the extent to which the different parameter estimates for different economies genuinely reflect different risks, or whether the differences may simply be due to sampling error. The answer must be a mixture of the two. There are good reasons to expect the United States market to be less risky than Hong Kong — as the world's largest market, the U.S. has more stocks in the index and benefits more from diversification than anywhere else. On the other hand, the U.S. has also had a run of good luck, and we should not discount the possibility that disastrous returns seen elsewhere may in future replicate in the U.S.

4.10 *Distribution Properties: Higher Cumulants*

4.10.1 To move beyond Normal distributions, we need to extract more information about distribution shape. One approach to this is to use cumulants.

4.10.2 The first five cumulants (or semi-invariants) of a random variable X are defined as follows:

$$\kappa_1(X) = \mu = \mathbf{E}(X)$$
$$\kappa_2(X) = \sigma^2 = \mathbf{E}(X-\mu)^2$$
$$\kappa_3(X) = \mathbf{E}(X-\mu)^3$$
$$\kappa_4(X) = \mathbf{E}(X-\mu)^4 - 3\kappa_2^2$$
$$\kappa_5(X) = \mathbf{E}(X-\mu)^5 - 10\kappa_2\kappa_3.$$

The cumulants have the property of being additive for independent random variables, which we find convenient when annualising distributions. Simpler moment definitions, for example omitting the $-3\kappa_2^2$ term from the definition of κ_4, would not satisfy the additive property. For more information on cumulants, see Kendall & Stuart (1979).

4.10.3 The first of these are the familiar mean and variance. The series continues beyond the fifth cumulant, but we do not need higher cumulants for the purpose of this paper. If X is normal then $\kappa_j(X) = 0$ for $j \geq 3$.

4.10.4 We define the *skewness* and *kurtosis* of a distribution to be the third and fourth cumulants, scaled by a power of the second cumulant in order to derive a dimensionless quantity, invariant under translation and scaling by a positive factor. We denote these by $sk(X)$ and $ku(X)$ respectively. This definition of kurtosis corresponds to what is sometimes referred to as excess kurtosis.

$$sk(X) = \frac{\kappa_3(X)}{\kappa_2(X)^{3/2}} = \frac{\mathbf{E}(X-\mu)^3}{\left\{\mathbf{E}(X-\mu)^2\right\}^{3/2}}$$
$$ku(X) = \frac{\kappa_4(X)}{\kappa_2(X)^2} = \frac{\mathbf{E}(X-\mu)^4}{\left\{\mathbf{E}(X-\mu)^2\right\}^2} - 3.$$

4.10.5 Given an extract $X_1, \ldots X_T$ from a stationary series, we can estimate these cumulants from the observed longitudinal distribution. The first two are:

$$\hat{\kappa}_1 = \frac{1}{T}\sum_{i=1}^{T} X_i$$
$$\hat{\kappa}_2(X) = \frac{1}{X}\sum_{i=1}^{T}(X_i - \hat{\kappa}_1)^2.$$

4.10.6 For this purpose, we calculate variance based on the empirical distribution. Our formula differs from the more conventional unbiased population estimate, where our $1/T$ is replaced by $1/(T-1)$. The bias correction fails for general stationary processes, being appropriate only when the X_i are iid. For this reason, we persist with the more intuitive $1/T$ version. The two versions converge, if T is large enough.

4.11 *Skewness and Kurtosis by Country*

4.11.1 Skewness is a measure of the asymmetry of a distribution. Positive skewness means that the left tail is thinner than the right tail. Positive kurtosis means thicker tails and a sharper peak at the mode, compared to a normal distribution.

4.11.2 Provided a distribution has a finite fourth moment, we can consider the shape of that distribution in terms of its skewness and kurtosis. We can plot these in a chart. These are shown in Figures 6 and 7, for both simple and logarithmic returns. They have been calculated using the 480 overlapping annual returns from the MSCI data described above.

4.11.3 Here, the pictures are very different according to whether we analyse simple or log returns. A model that is fat tailed in one space may be thin tailed in the other. The majority of countries have positively skewed simple returns but negative skewed log returns.

4.11.4 Once again, we can ask whether differences between different countries reflect genuine differences in economic risk or whether they may simply be an artefact of random sampling. The effect of sampling error on skewness and kurtosis is explored later in this section.

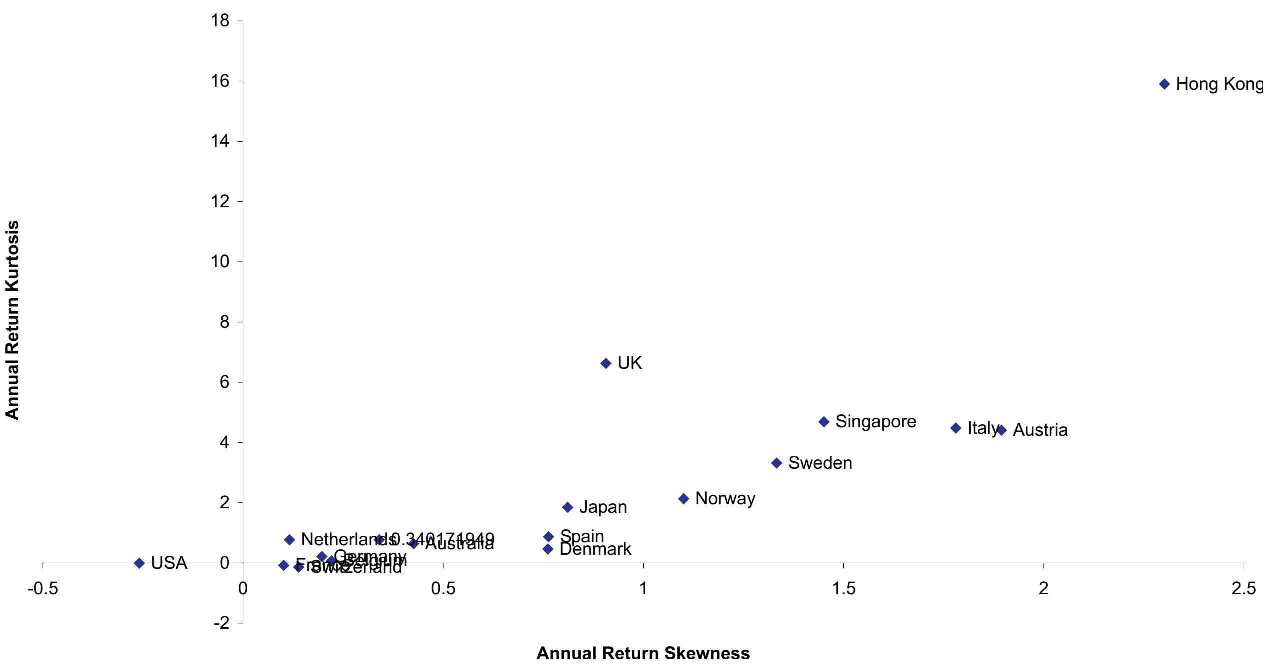

Figure 6. Simple annual returns: skewness and kurtosis

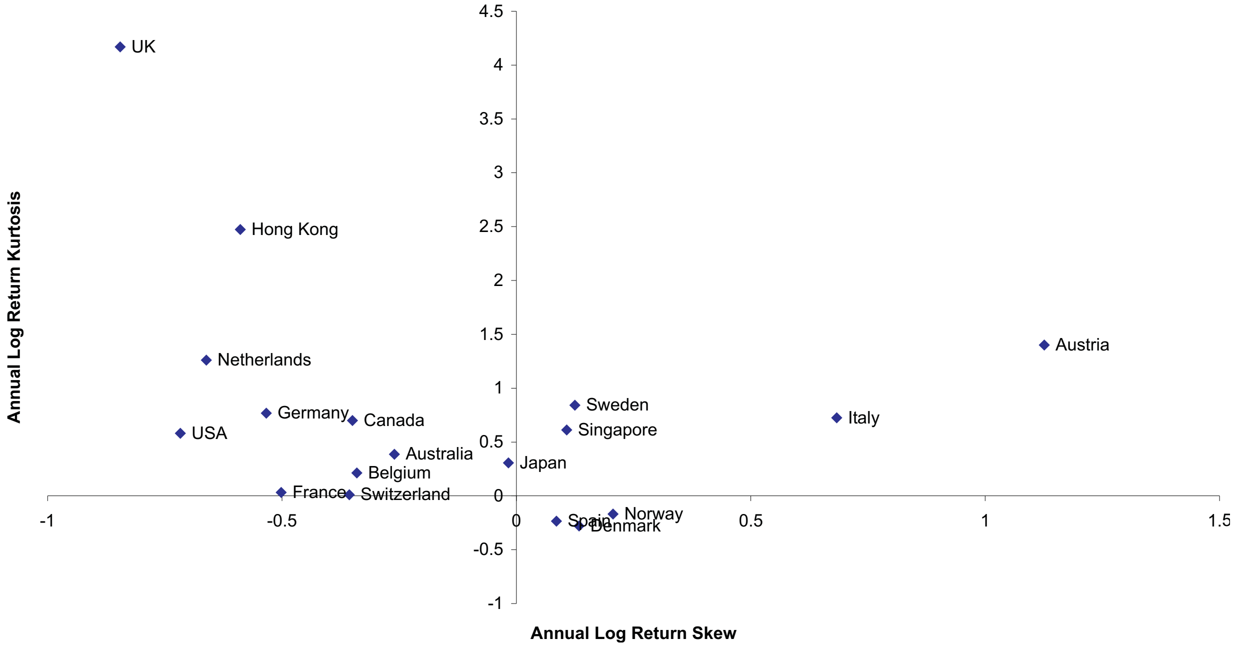

Figure 7. Log annual returns skewness and kurtosis

4.12 *How Cumulants Vary by Holding Period*

4.12.1 We have measured moments for annual returns. We now consider returns over other holding periods. One reason for doing this is that using more frequent observations increases the available number of data points. In a 40 year period, there are 480 months of monthly data. We might think there are 468 annual returns, and indeed there are, but they overlap and so a count of datapoints overstates the quantity of available information.

4.12.2 If we had a reliable model of monthly returns, and a reliable way to aggregate that model to annual holding periods, then we would have a recipe for constructing a model of annual returns making full use of 480 non-overlapping data points.

4.12.3 With the aim of constructing such an aggregation model, we start by investigating moments for different time horizons. Our data sets are then as follows:

Choice to make	Selected for this paper
Data sampling frequency	Monthly, based in index values from 31/12/1969 to 31/12/2007. This corresponds to 456 months
Country and reporting currency	We consider the 18 countries for which data is available, values reported in local currency
Return modelled	Simple return and log return
Holding period	1 month to 24 months

4.12.4 The distribution of 24-month returns is much more dispersed than 1-month returns. To make a fair comparison, we seek to compare parameters on an annualised basis. Therefore, we restate the observed parameters as equivalent annual parameters. For the purpose of annualisation, we make the assumption that the returns over non-overlapping intervals are independent, that is, the random walk assumption. If this assumption holds, then we should find that annualised parameter estimates take the same value, except for the effect of sampling error, irrespective of holding periods. To the extent that annualised parameters do appear to vary with holding period, this may be due to sampling error or a fault in the independence assumption underlying our annualisation. A plot of annualised parameters against holding period therefore provides a useful indication of the extent to which the independence assumption is borne out in the data sets.

4.12.5 In general, annualisation requires further distributional assumptions. A statement of the 0.5 percentile of monthly returns, together with an assumption of independence, does not allow us to construct the 0.5 percentile of annual returns. To do this calculation requires knowledge of the full monthly return distribution. However, there is a major simplification when we instead examine cumulants. It turns out that the aggregation formulas for cumulants, assuming independent returns, are generic and apply across all possible distributions. This fact holds whether we look at simple or logarithmic returns, although the aggregation rules are different in each case, dealing respectively with products or sums of random variables.

4.12.6 We consider first the mean returns. These are shown in Figures 8 and 9 for both simple and log returns.

4.12.7 By eye, at least, this offers little challenge to our aggregation method. The high returning countries are Hong Kong followed by Sweden. The worst performer is Japan, followed by Switzerland based on simple returns and by Germany based on log returns.

4.12.8 We now move on to an analysis of standard deviations as shown in Figures 10 and 11. The familiar process for historic volatility estimates involves taking logs of an index, then examining the standard deviation of

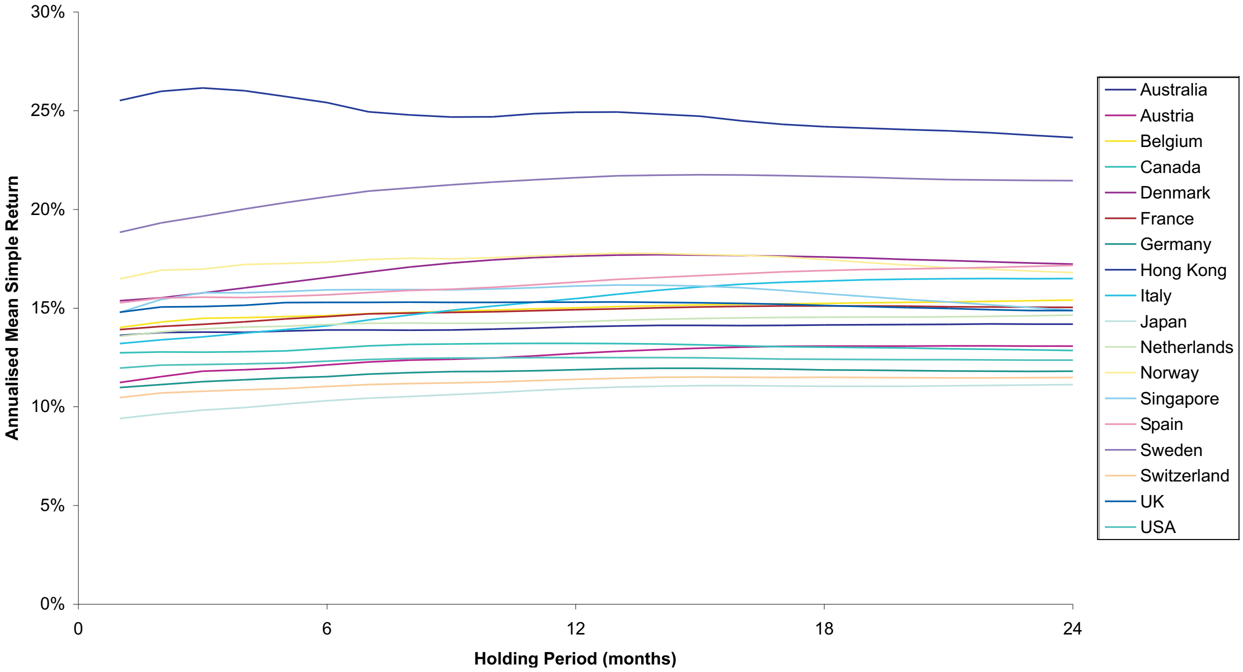

Figure 8

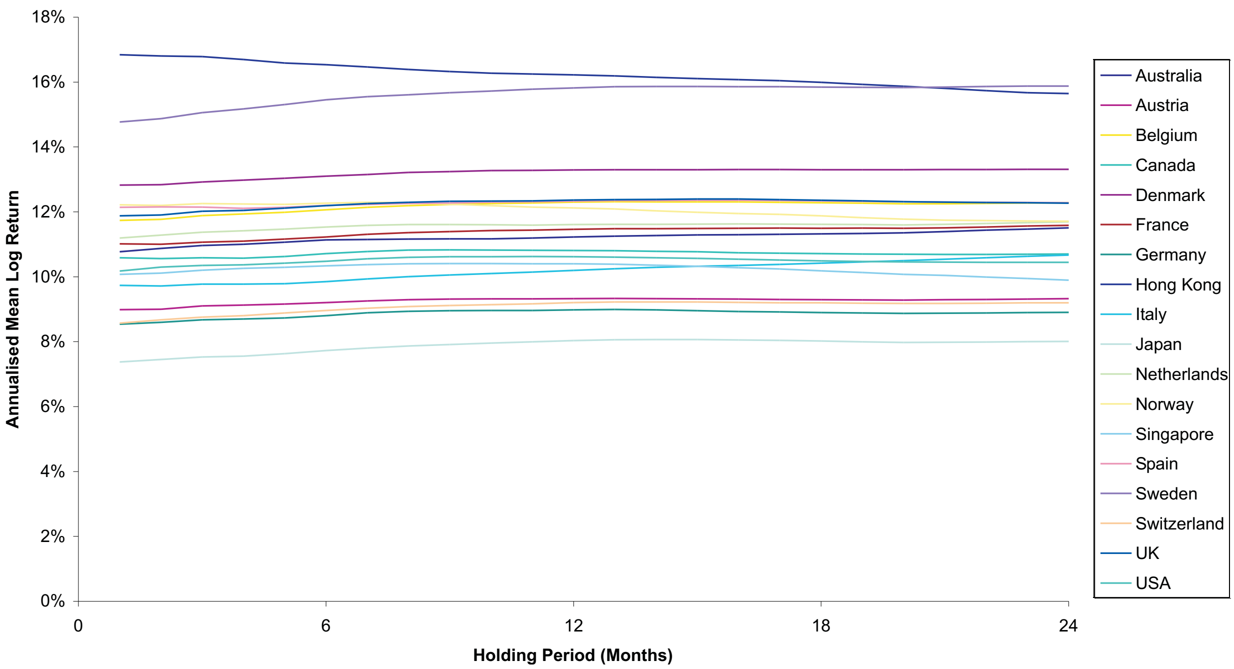

Figure 9

price changes. To annualise the standard deviation, it is customary to divide by the square root of the holding period. This is equivalent to the independence assumption in the case of log returns. For simple returns, the algorithm is slightly different: the expected $(1 + \text{year return})^2$ over the whole year is equal to the product of the expected $(1 + \text{month return})^2$ over each month, assuming independence. In terms of an equation we can write:

$$(1 + R_A)^2 = \prod_{i=1}^{12} (1 + R_{M_i})^2,$$

where R_A is the annual return and R_{M_i} is the return in each month i.

4.12.9 Here, it is less clear that Figures 10 and 11 represent horizontal lines together with sampling error. While most of these lines look reasonably flat, some seem to show an upward slope. Italy is the most conspicuous here, with Austria and Spain some way behind. An upward slope means that the standard deviation of annual returns is higher than we would expect from compounding monthly returns independently. This implies positive correlations between returns over shorter time periods; a rise one month is likely followed by a rise the next month. This phenomenon is sometimes called *mean aversion*. The U.K. is unique in showing the reverse effect, that a good return is more likely followed by poor returns, that is, *mean reversion*.

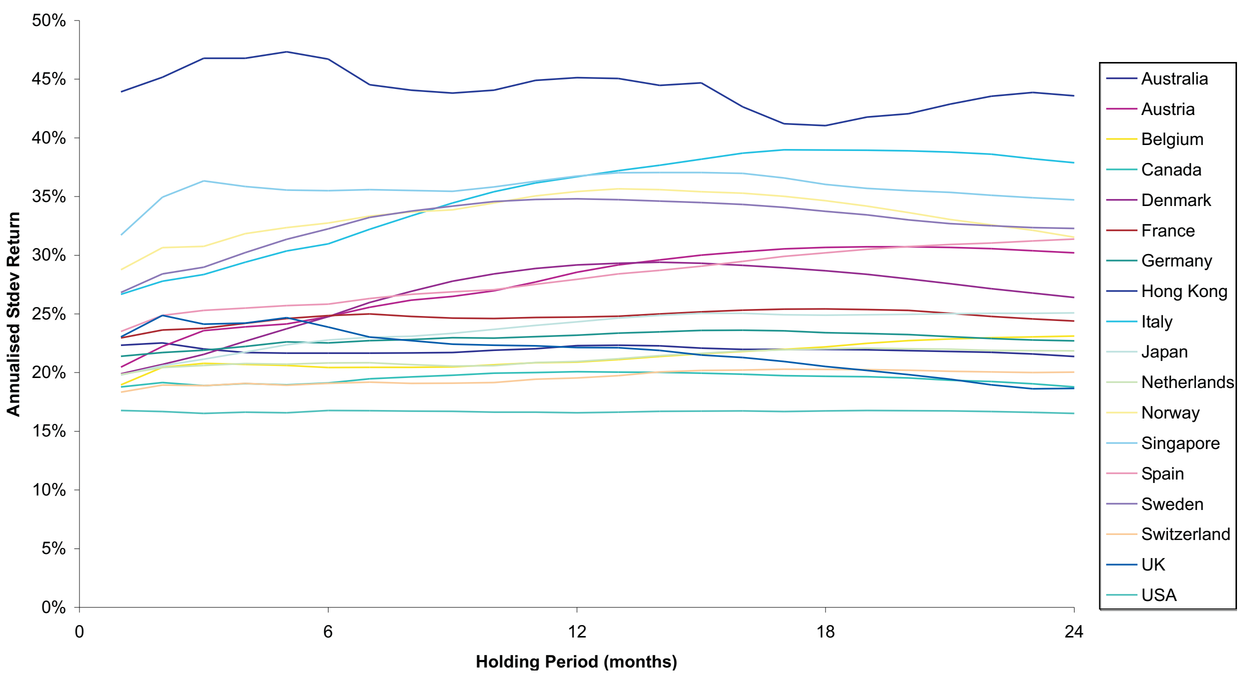

Figure 10

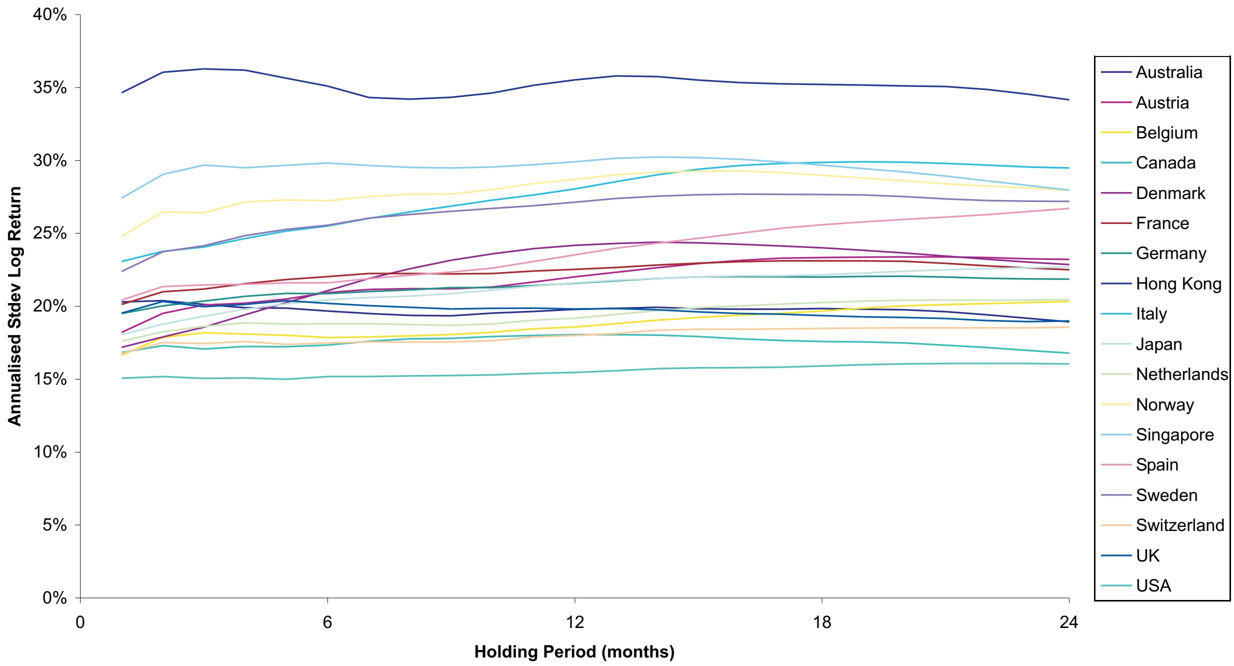

Figure 11

However, the mean reversion is visible only in the simple returns and not in the log returns. This suggests a more subtle effect than classical mean reversion may be at work.

4.12.10 If the annualised volatility is flat, that means that returns are uncorrelated. The fact of a good return last month gives us no guidance on the direction of next month's move. This provides some level of support for the (weak form of) the *efficient markets hypothesis*, that current prices already contain all the information in past returns. Therefore past returns provide no useful guide to future movements.

4.12.11 We note in passing that market efficiency may not always imply uncorrelated returns. Many analysts view annual equity returns as being composed of a one-year interest rate, plus a constant expected risk premium, plus additional noise. If one noise observation does not predict the next, then this implies the noise terms are uncorrelated. However, the equity returns themselves may remain correlated because of the effect of one-year interest rates. Historic data clearly shows a positive autocorrelation in one-year rates, which would also imply positive autocorrelation in equity returns, which implies a positive volatility slope as a function of holding period. However, the numerical effect is small, and negligible relative to the noise in our data series.

4.12.12 We now move on to annualised skewness. These are shown, both for simple and log returns, in Figures 12 and 13.

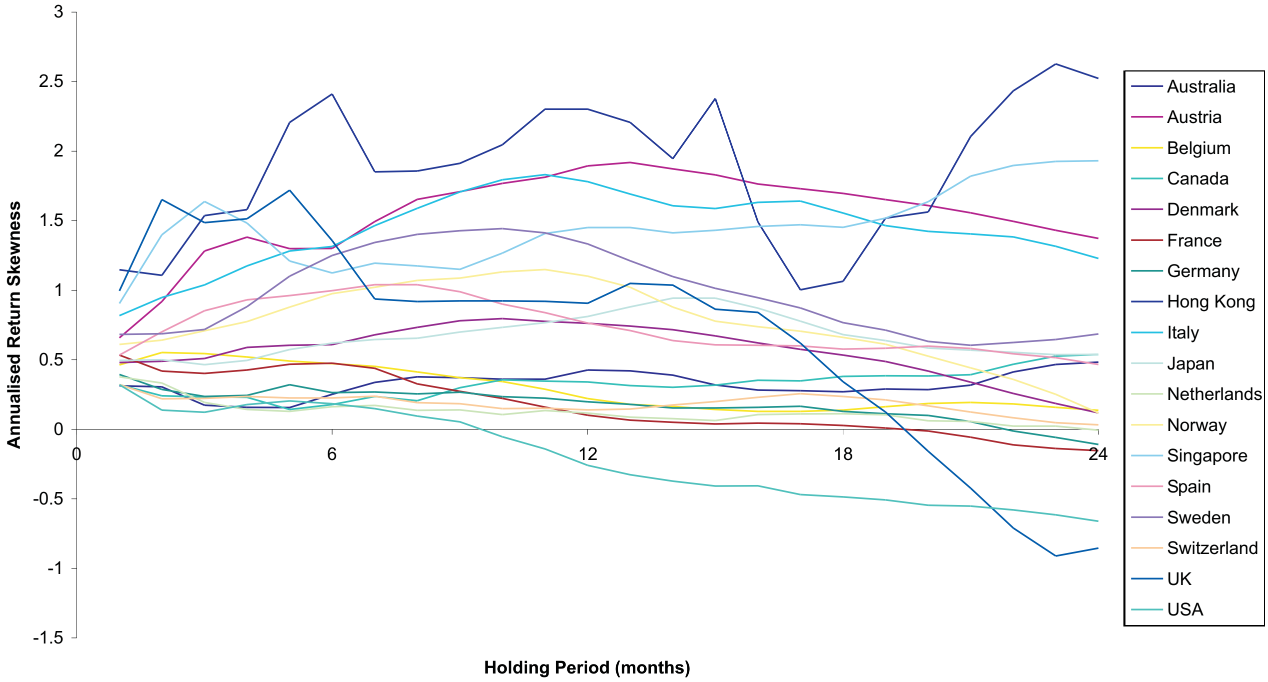

Figure 12

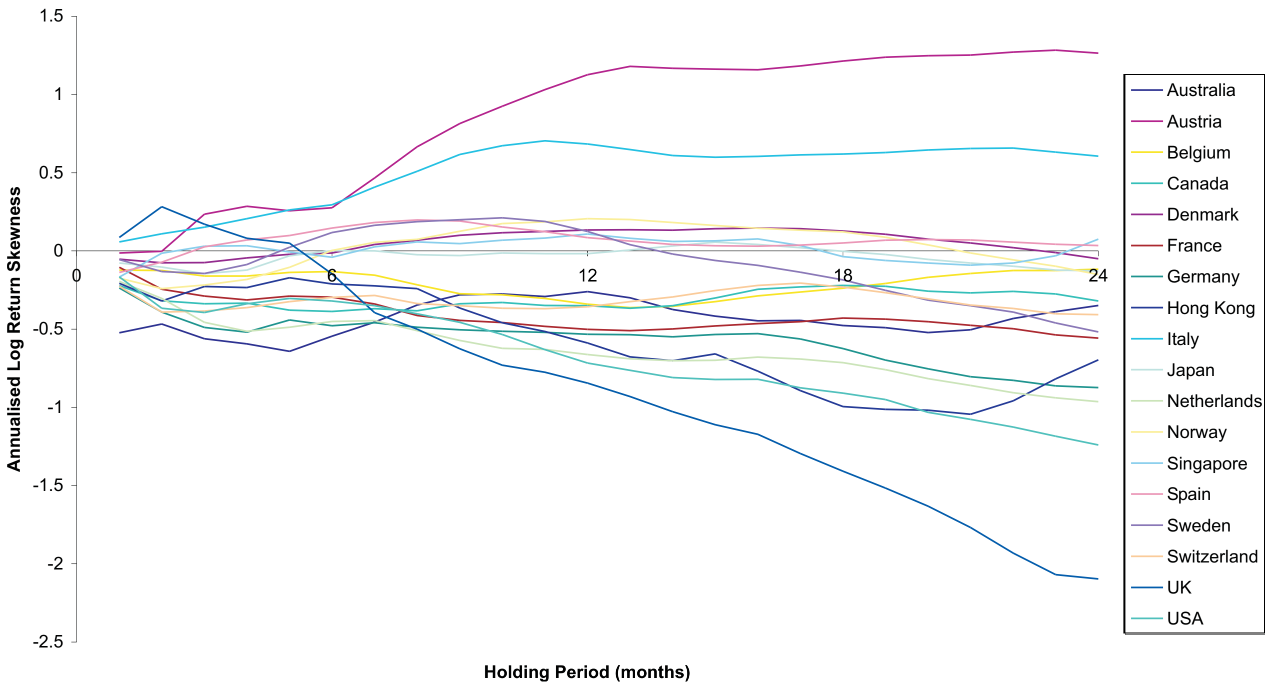

Figure 13

4.12.13 These results may be something of a shock. These charts do not appear to be horizontal lines. With a holding period of one month, the skewness for each country lies within a narrow band. These bands fan out for large time periods, with some appearing to increase with holding period, and others appearing to decrease. The fan is particularly marked in the case of log returns.

4.12.14 The most obvious explanation of these effects is sampling error. Sampling error is lower for shorter holding periods where data is more frequent. So we would expect to see a greater dispersion of results for larger holding periods, and that is exactly what we see. However, this is unlikely to be the whole explanation, as this dispersion effect is not evident in the mean and variance estimates.

4.12.15 The question still arises whether, behind the apparently diverse behaviour between countries, there lies some underlying pattern. A possible explanation for the pattern is that a market fall heralds increased volatility in the next period.

4.12.16 Let $R(t)$ denote the log return in a year, and $R(t+1)$ the log return in the subsequent year, adjusted to have zero mean. Then we might consider the terms in the identity:

$$\mathbf{E}(R(t)+R(t+1))^3 = \mathbf{E}(R(t)^3) + 3\mathbf{E}(R(t)^2 R(t+1)) + 3\mathbf{E}(R(t)R(t+1)^2) + \mathbf{E}(R(t+1)^3).$$

4.12.17 Our hypothesis implies that:

$$\mathbf{E}(R(t)R(t+1)^2) < 0.$$

4.12.18 On the other hand, if, consistent with market efficiency, we cannot use any power of $R(t)$ to predict $R(t+1)$, we must have:

$$\mathbf{E}(R(t)^2 R(t+1)) = 0.$$

4.12.19 Putting these together, we should have:

$$\tfrac{1}{2}\mathbf{E}(R(t)+R(t+1))^3 \le \tfrac{1}{2}\mathbf{E}(R(t)^3) + \tfrac{1}{2}\mathbf{E}(R(t+1)^3).$$

4.12.20 The left hand side of Figure 14 is the annualised skewness from two-year holding periods, while the right hand side is the average of two estimates of one-year skewness, the first omitting the last year and the final term omitting the first year. If you look at the plot of log return annualised skewness, you might just be able to discern a downward slope, and the average of all 18 countries is indeed a decreasing function.

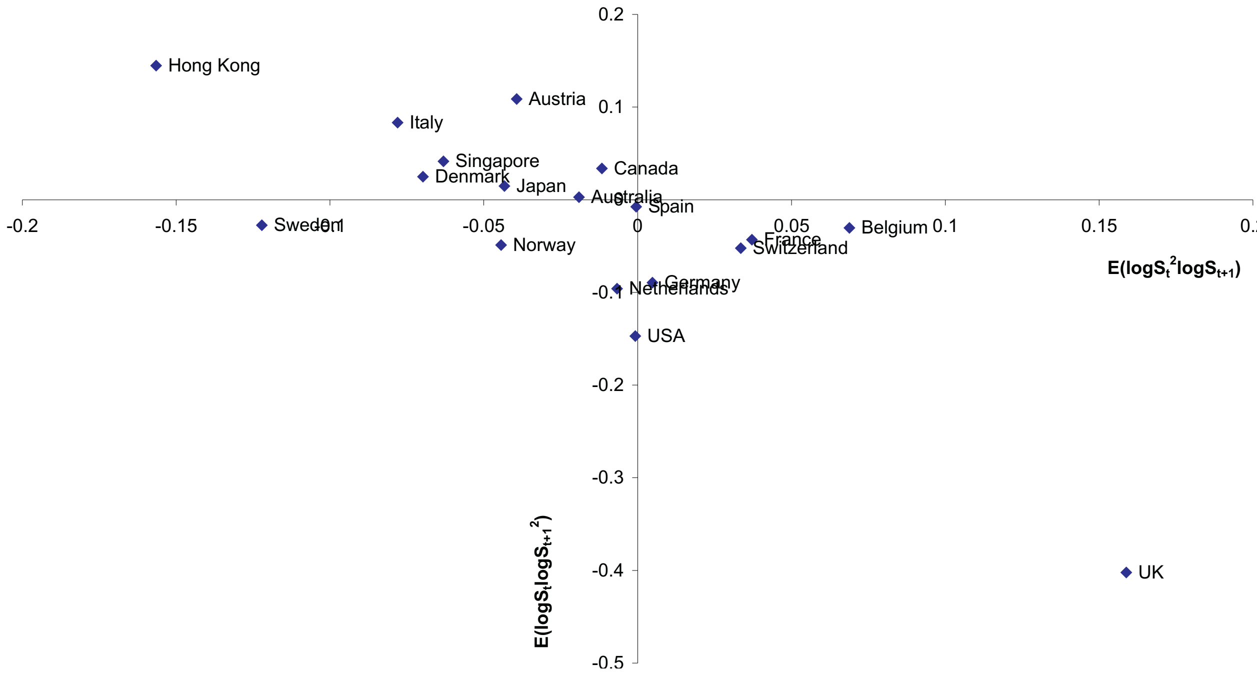

Figure 14. Dependent Skewness components

4.12.21 If we look in more detail, we can decompose the slope (of annualised skewness against holding periods) into two components: $\mathbf{E}(R(t)^2R(t+1))$ and $\mathbf{E}(R(t)R(t+1)^2)$. According to our hypothesis of higher volatility following a fall, the first of these terms should be around zero, while the second term should be negative. A scatter plot of these data points by country should be clustered around the negative Y axis (Figure 14).

4.12.22 Unfortunately, the data does not support our hypothesis. Indeed, it suggests the reverse, that a large move in either direction tends to precede a fall. This is hard to rationalise economically. The apparent downward slope could be mere noise. We note that the U.K. is an outlier in this plot, the only country where large historic moves have been followed by a rise in the following year.

4.12.23 Annualised kurtosis shows a clear upward trend as a function of holding period. This once again contradicts the prior hypothesis of independent returns (Figure 15).

4.12.24 One explanation for upward sloping annualised kurtosis is *volatility clustering*, that is, a tendency for a large move (of either sign) in one period to be followed by a large move (again of either sign) in the next. While this effect is well attested for high frequency data (daily, for example), it does appear to persist with less force at the scale we seek to measure.

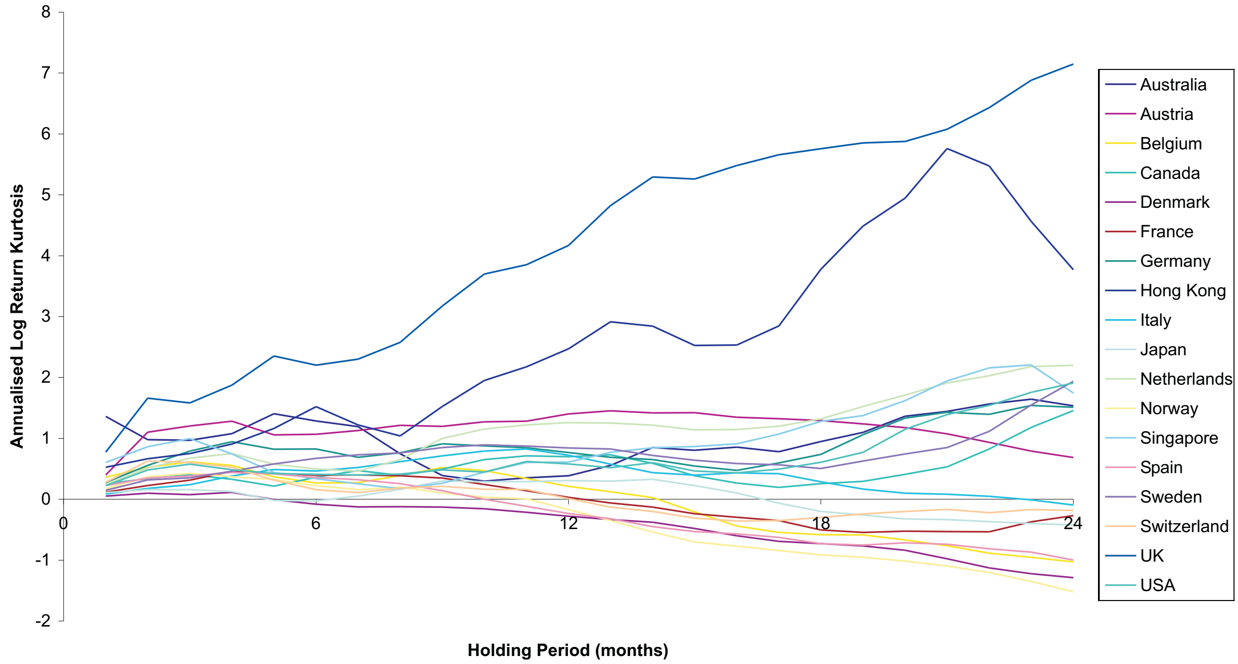

Figure 15

4.13 *Estimation Error in Skewness and Kurtosis*[4]

4.13.1 In the following subsections we pause to consider the estimation error inherent in estimating the skewness and kurtosis from a data series.

4.13.2 In the previous sub-sections the skewness and kurtosis of various markets were calculated from recent market data — specifically overlapping annual data from around 1970. In this section we use a different data set that is non-overlapping. The data is from Dimson Marsh Staunton from 1/1/1900 to 31/12/2002 and from the MSCI Global Equity Indices from 1/1/2003 to 31/12/2007. The data is used to calculate the skew and kurtosis of various markets before going on to examine the estimation error in these calculations. The DMS data comprises annual returns so the analysis that follows assumes a 1 year time horizon.

4.13.3 In the analysis we have excluded Germany due to the disproportionate effect of the hyperinflation years which leads to very extreme (positive) skew and kurtosis. A more sophisticated analysis could include calculations from Germany with the hyperinflation years removed.

4.13.4 In the analysis we calculated statistics based on log returns. We have also used absolute returns. An alternative could have been to use returns in excess of cash.

[4] In the text we refer to Kurtosis. For avoidance of doubt we are always referring to Excess Kurtosis and use the term Kurtosis for brevity.

4.13.5 We have also both here and in section 5 used some abbreviations for various country names. The following table can be used as a reference if any of the labels are not obvious:

Country	Country code
United Kingdom	UK
USA	US
Germany	GER
Japan	JAP
Netherlands	NET
France	FRA
Italy	ITA
Switzerland	SWZ
Australia	AUD
Canada	CAN
Sweden	SWE
Denmark	DEN
Spain	SPA
Belgium	BEL
Ireland	IRE
South Africa	SAF

4.14 *Measuring Skewness and Kurtosis on the DMS Data*

4.14.1 In Figure 16 we show the skewness and excess kurtosis of the absolute log total returns on equity markets for 15 different economies.

4.14.2 It is interesting to see that there are two obvious outliers; U.K. and Denmark. The high excess kurtosis of the U.K. market is due to the events of the early 1970s when the market fell and rose dramatically over a two year period. In 1974 the U.K. Equity Market fell by 48.9% in simple returns (a fall of 67.0% in log returns). While in the next year, 1975, it rose by 146% in simple returns (a rise of 90% in log returns).

4.14.3 The Danish market has a moderately high excess kurtosis but is noticeable for having a high positive skewness. This is mainly due to a significant rise in the stock market following economic reforms announced in the early 1980s. This culminated in a 1983 rise of 120% in simple returns (a rise of 79% in log returns). Meanwhile Denmark appears to have remained relatively untroubled by large stock market falls over the period 1900-2007. The worst return came in 1992 when the equity market fell by just 25% measured in simple returns (or 28% in log returns). This was the most moderate "worst year" fall of any of the 16 economies studied and helps explain the positive skew in the Danish data.

4.14.4 We can see that the higher moments of these two outlier economies are mainly determined by 1 or 2 extreme observations. In Figure 17 we have

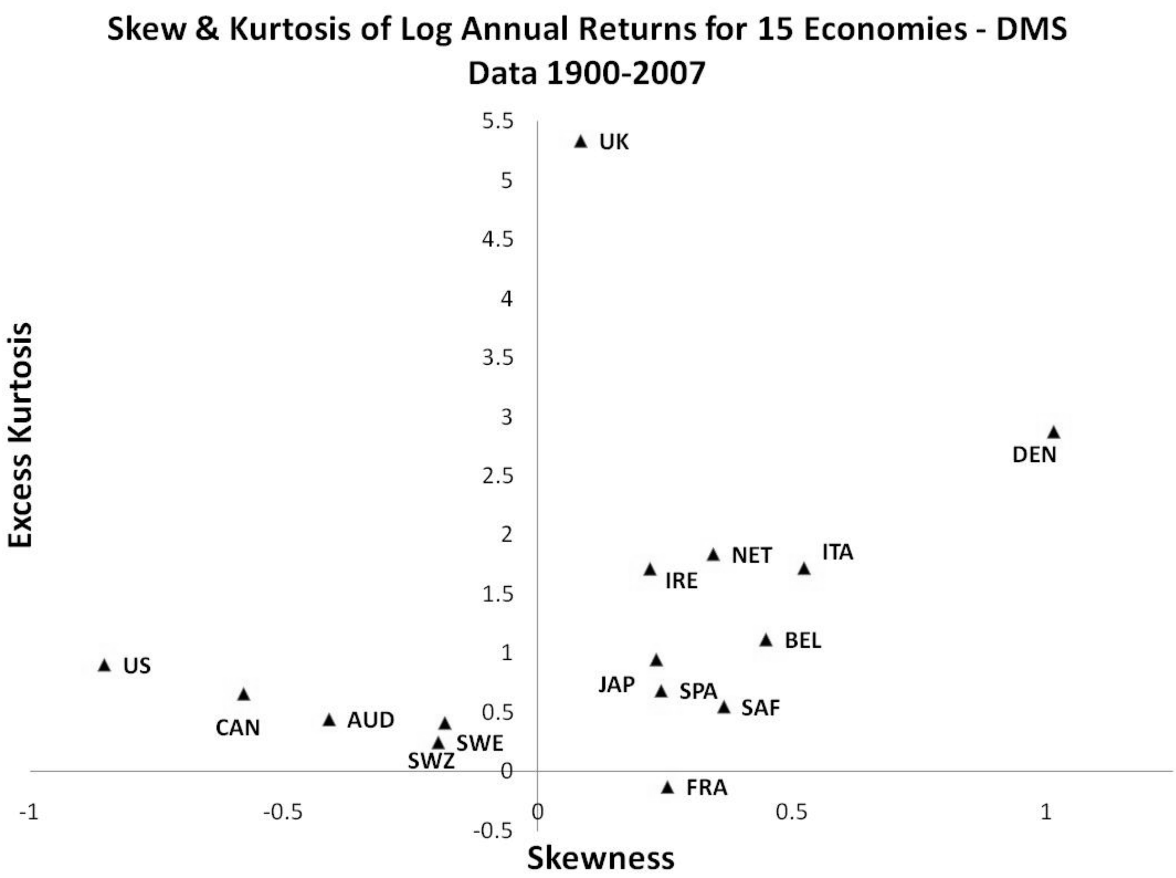

Figure 16

excluded the two U.K. extreme returns and the single extreme Danish return described above. Without these data the skew/kurtosis pairing for the U.K. and Denmark look very similar to those of other countries. Noticeably the U.K. returns appear very nearly normally distributed when the two 1970s observations are removed. We are not advocating the removal of outliers in doing this analysis as it is just these events that are of most interest to our analysis. However it is our intention to illustrate how measures of skew and especially kurtosis can be significantly influenced by a single event over the course of a 100 year period.

4.15 *Measuring Estimation Error on Skewness and Kurtosis on the DMS Data*

4.15.1 We estimate the sampling error for the skewness and kurtosis of these log returns. To do this we use a formula approach to construct a 95% confidence interval around the estimates of the skewness and kurtosis; Scherer (2007).

4.15.2 The results are displayed in Figure 18.

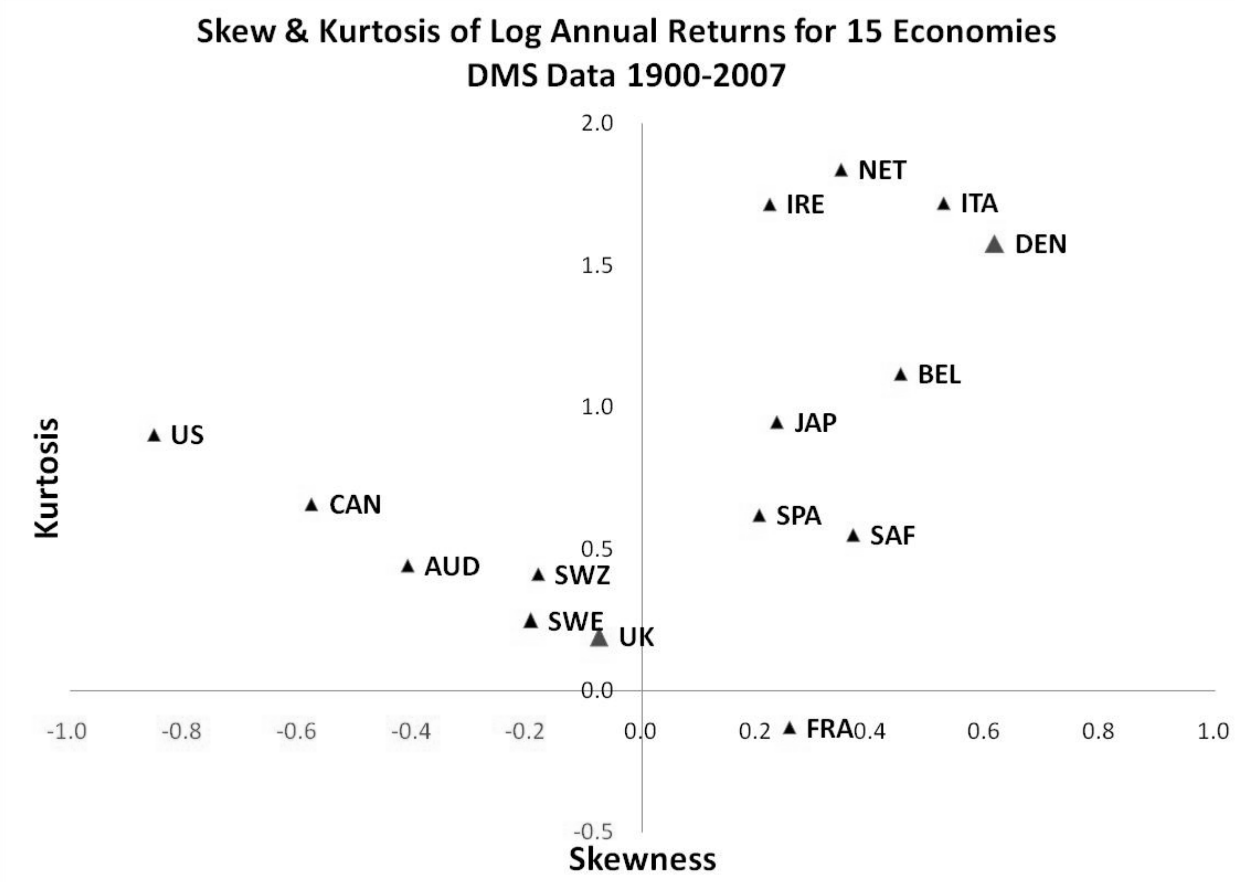

Figure 17

4.15.3 In Figure 19 we zoom in to give more detail in the centre of the chart shown at Figure 18.

4.15.4 Interestingly these confidence intervals suggest that there are several countries with significant skewness or kurtosis in their log returns over the period 1900-2007. In the following table we display the raw skewness and kurtosis figures and indicate whether the skewness or kurtosis is:

— significantly positive at the 95% level (+),
— significantly negative at the 95% level (−), and
— not significant at the 95% level (NO).

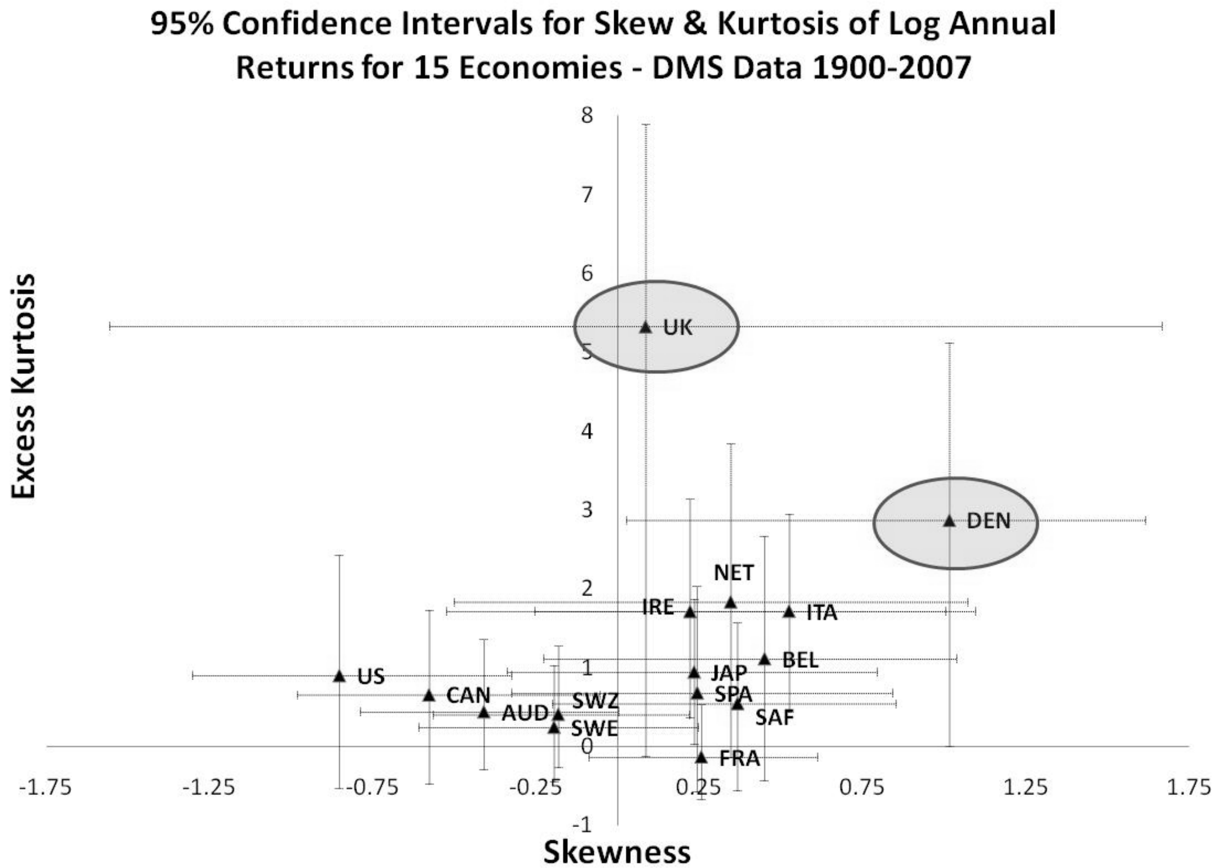

Figure 18

4.15.5 We have included the raw figures for Germany too for completeness.

Country	Country code	Skew	Kurt	Significant Skew	Significant Kurt
United Kingdom	UK	0.0850	5.3372	NO	+
U.S.A.	US	−0.8528	0.9037	−	NO
Germany	GER	9.8261	100.1024	NO	+
Japan	JAP	0.2338	0.9487	NO	+
Netherlands	NET	0.3458	1.8404	NO	+
France	FRA	0.2559	−0.1303	NO	NO
Italy	ITA	0.5247	1.7213	+	+
Switzerland	SWZ	−0.1825	0.4116	NO	NO
Australia	AUD	−0.4102	0.4420	NO	NO
Canada	CAN	−0.5785	0.6584	−	NO
Sweden	SWE	−0.1954	0.2479	NO	NO
Denmark	DEN	1.0161	2.8768	+	+
Spain	SPA	0.2434	0.6838	NO	NO
Belgium	BEL	0.4497	1.1188	NO	+
Ireland	IRE	0.2214	1.7172	+	+
South Africa	SAF	0.3669	0.5500	NO	NO

4.15.6 Only Canada and the U.S.A. showed significant negative skewness, while Denmark showed significant positive skewness. When considering excess kurtosis we find that Italy, Denmark, Ireland, Japan, Belgium, Netherlands and the U.K. showed significantly positive excess kurtosis. Several countries showed neither significant skew nor kurtosis. These were Japan, France, Switzerland, Australia, Sweden, Spain and South Africa.

4.15.7 The formula approach results in a 95% confidence interval width for skew of 0.92 and a 95% confidence interval width for kurtosis of 1.5 based on 108 overlapping data points in the sample.

4.15.8 Before we conclude this section it is interesting to compare the skewness and kurtosis observed from the 1970-2008 MSCI data with overlapping time periods and the 1900-2007 DMS data with non-overlapping time periods. There are 14 countries for which we had data from both series.

4.15.9 In Figure 20 we have plotted the MSCI skewness estimate vs. the DMS skewness estimates. We have also plotted the $x = y$ line so we can see the degree of variation between the two data series. We can see that there are

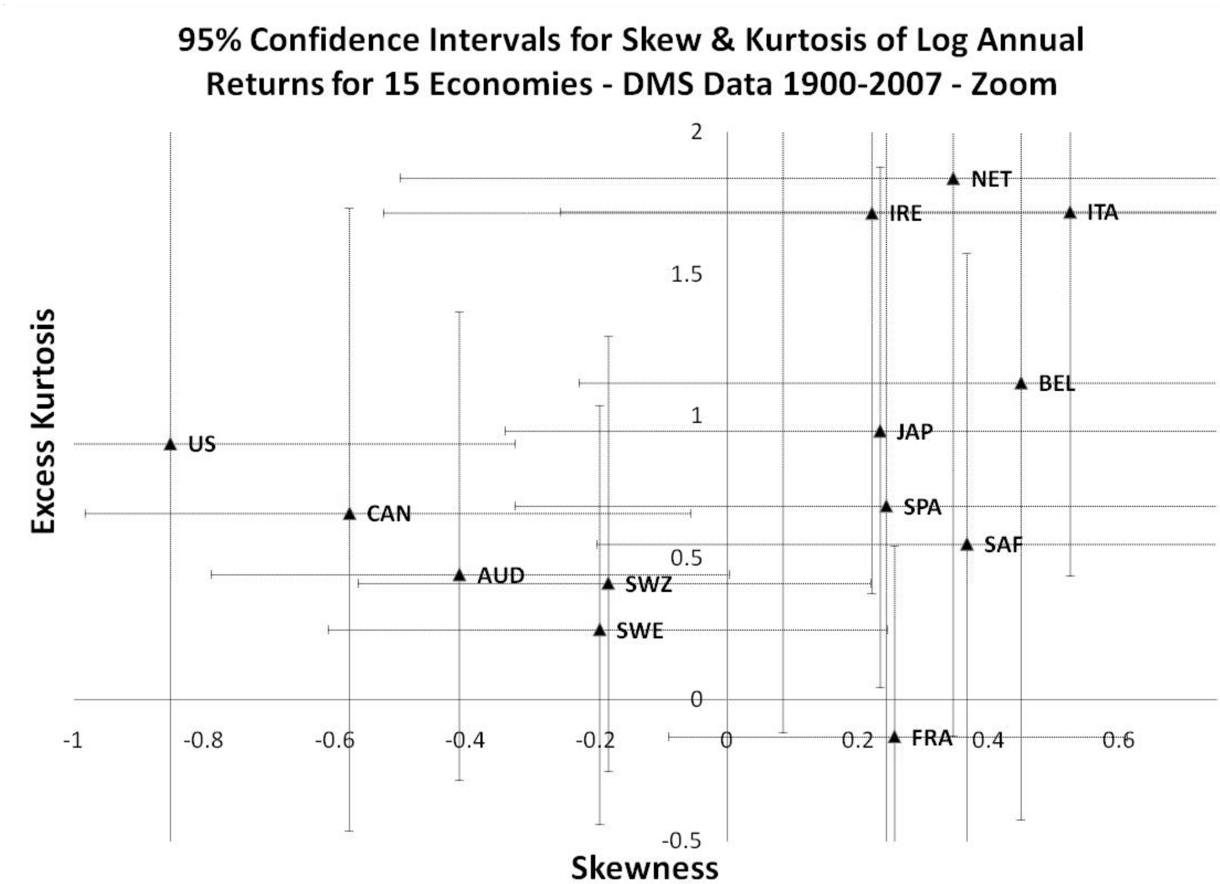

Figure 19

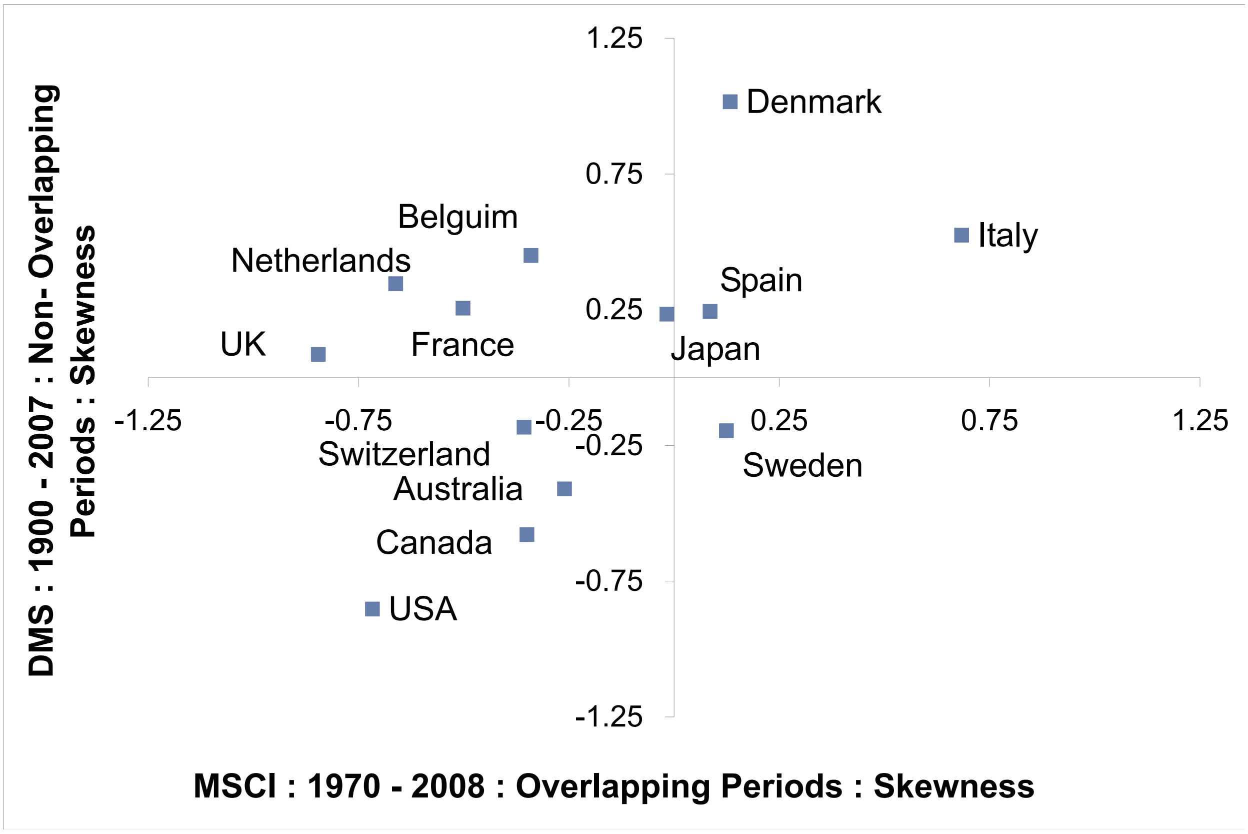

Figure 20

a number of countries where the two estimates are close. However there is also a cluster of Western European countries where the skewness is noticeably more positive in the DMS sample than the MSCI sample.

4.15.10 In Figure 21 we have plotted the MSCI kurtosis estimate vs. the DMS kurtosis estimates. We have also plotted the $x = y$ line so we can see the degree of variation between the two data series.

4.15.11 We can see again that there are a number of countries where the two estimates are close. However for all but three countries, the DMS estimates of kurtosis are higher than those for the MSCI data. This is especially noticeable for Denmark.

4.16 *Conclusion*

4.16.1 Our analysis in this section has revealed that equity returns are, for most countries, uncorrelated but not independent, assuming that the effects we see in the analysis are not due to sampling error.

4.16.2 The idea of constructing annual returns from more frequent data is appealing, because frequent data gives us more points to investigate. Unfortunately, to annualise return distributions, we need to make assumptions about volatility clustering and the correlation of volatility with market

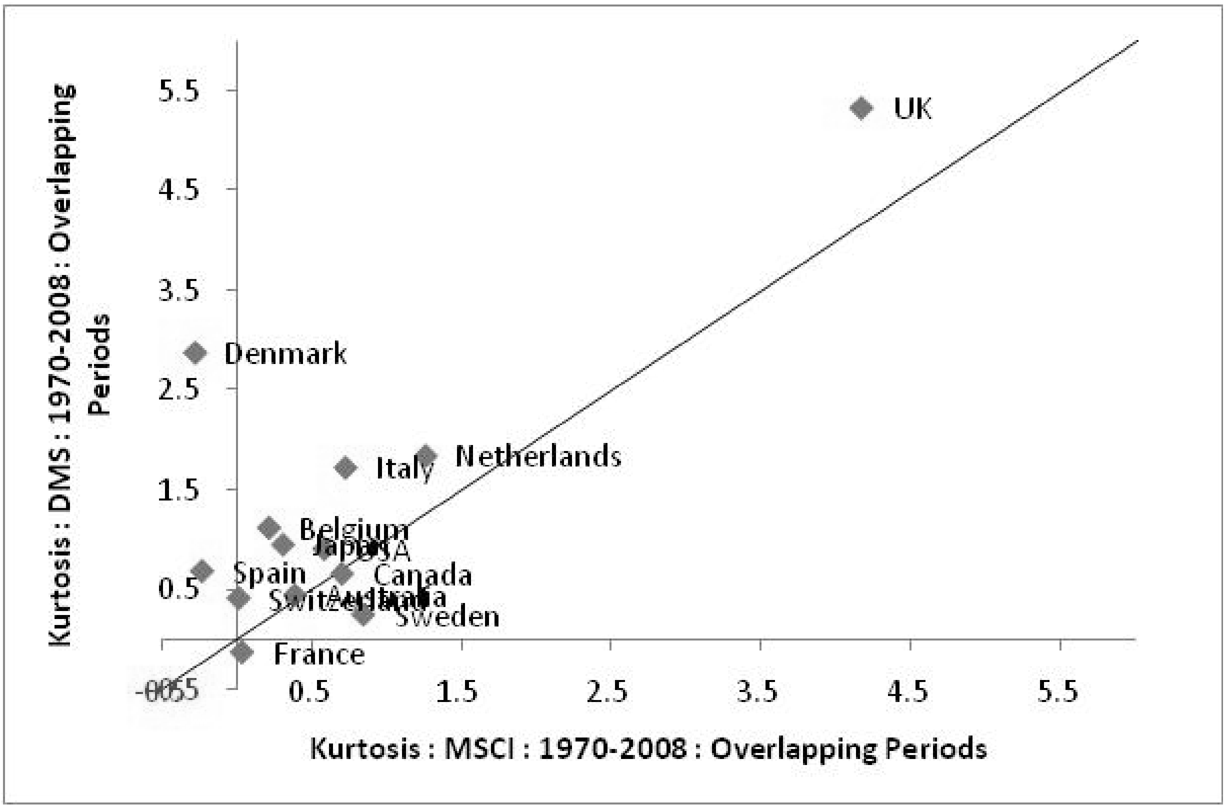

Figure 21

levels. The simplest assumption, that these effects are negligible and returns are independent, is untenable based on the data we have investigated. Instead, we have to look at returns over annual periods to estimate the best way to compound distributions — which unfortunately is what we wanted to estimate in the first place.

4.16.3 Therefore, estimated annual percentile estimates are not significantly improved by analysing more frequent data. We can estimate more accurately the distribution of high frequency returns, but uncertainty over the aggregation procedure undermines confidence we might have in the distribution of annual returns derived from overlapping time periods.

4.16.4 We then went on to look at a non-overlapping data set of annual returns. We measured the skewness and kurtosis on this data and estimated the sampling error for each of these statistics. We found that the sampling error was very significant especially for countries that had experienced a significant positive and/or negative return.

5. Equity Modelling — 20th Century Data — Estimation Error of Extreme Percentiles

5.1 *Introduction*

5.1.1 Calculation of any statistic should be coupled with some measure of the estimation error in calculating the statistic. We believe that actuaries should be aware of the uncertainty inherent in the estimation of the extreme percentiles of financial asset returns for economic capital calculations. There are a number of factors that contribute to this uncertainty:

- The limited volume of historical data.
- The complex underlying behaviour (probability distribution) of the underlying asset. For example the well-behaved normal distribution may not be a reliable description of the behaviour of the asset.
- The non-constant nature of financial markets.

5.1.2 Financial markets are socio-economic rather than physical systems whose characteristics might be expected to change over time as the world changes. This challenges the standard working assumption of statistical analysis that data is identical and independently distributed (iid) and more generally that the time series is stationary.

5.2 *Data*

5.2.1 In this section we have focussed on simple returns as we believe they are easier to interpret. We have also elected to use absolute returns rather than excess of cash returns. Our reasons are that actuaries will need to consider the percentile of absolute returns in their capital calculations. An analysis based on excess of cash returns would also have value because it would remove some of the serial correlation we might expect to see from the cash return component. However the results would require additional assumptions (about extreme cash returns and their tail correlation with the extreme excess cash equity returns) to convert the results into an absolute return stress test. For simplicity we have limited our attention to absolute returns for this analysis.

5.2.2 The data used for the analysis that follows was the data series used by Dimson, Marsh and Staunton (DMS) for their research into the long term behaviour of markets. In their book Triumph of the Optimists there were some summary statistics given for the data series to end 2000. We have used the DMS data set of annual U.K. Equity Total Returns from 1/1/1900 to 31/12/2002. The full data set is available on a subscription from Morningstar. We have supplemented the DMS data with MSCI data from 1/1/2003 to 31/12/2007. Several of the DMS data series are based on the MSCI indices in the later years of their survey.

We should point out to the reader that the statistics calculated on the full data set contain some discrepancies compared to those cited in the book. Some sample discrepancies are:

5.2.3 Germany: For their book DMS remove the years 1922 and 1923 due to the hyperinflation in these years.

5.2.4 Switzerland: For their book DMS elect to only calculate statistics from 1911. No obvious explanation is given for this editing.

5.2.5 We have not carried out an exhaustive comparison of the difference between the summary statistics given in the book and the original DMS data set available for purchase online. The data was collected from Morningstar after the end of 2002 while the book was published on 1 January 2002. Because the book does not provide the actual returns from the data set and the online data set was more recent, we consider it reasonable to rely on the data purchased online.

5.3 *Percentile Calculation*

5.3.1 Before analysing the estimation error we calculate the sample statistic on which the error will be estimated. The statistic is the 0.5 percentile return from the U.K. total equity return. We calculate this to be a simple return of −35.7% (or −46.1%[5] as a log return). This statistic was calculated using the percentile function in Excel applied to the U.K. Equity Total Return data from 01 January 1900 to 31 December 2007. We note that the Excel percentile function uses linear interpolation to estimate percentile values when the percentile requested is between data observations. This causes some issues which are explored later in this section

5.3.2 We now move on to measuring the estimation error of the 0.5 percentile return statistic. Before we do we introduce the bootstrapping technique used to calculate the estimation error.

5.4 *The Bootstrapping Process*

5.4.1 The impact of the limited volume of historical data can be quantified using standard statistical techniques such as bootstrapping. Bootstrapping allows us to develop estimates of a lower bound for the uncertainty in extreme percentile estimates.

5.4.2 There has been a lot of work on the banking sector on back-testing which should be mentioned. Much of the literature we have seen

[5] It may be noted that $\ln(1 - 0.357) = -0.442$ and not -0.461. The figures in the paper where calculated by applying the Excel Percentile function to the full U.K. DMS dataset for both log and simple returns. Because Excel Percentile always uses linear interpolation for interim percentiles the interpolation for log returns differs from simple returns. We have not explored the difference in interpolation for log returns and simple returns further in this paper as the main focus is estimation error and both simple and log returns have only been shown for completeness.

focuses on the back-testing of a VaR model to see if the number of VaR limit breaches is suggesting the model is invalid. For more details on methods used in the banking sector see the Journal of Risk reference at the end of this paper. There is also ongoing work in academia to try and quantify estimation error on extreme percentile returns.[6]

5.4.3 A paper that takes a very similar approach to the bootstrap approach taken here is Dowd (2002). The analysis below extends the bootstrap approach to the long term DMS dataset. It also looks at the analysis on several different economies and investigates the biases that can be found in common spreadsheet functions that might be used for this type of work.

5.4.4 The bootstrap technique is not being proposed as the only way to measure estimation error of percentiles. However we believe that it is relatively easy to explain to a less technical audience and relatively easy for a practitioner to implement.

5.4.5 When we describe bootstrapping we refer to a technique that constructs an empirical distribution function from a data set and then generates new data samples from this empirical distribution. Typically we might generate 10,000 data samples from the empirical distribution and calculate the statistic of interest in each of the 10,000 samples. We can then calculate a 95% confidence interval (for example) using the 10,000 samples of the statistic.[7]

5.4.6 This bootstrapping technique should not be confused with the identically named method for constructing a zero coupon yield curve from coupon bond yields.

5.5 *Applying Bootstrapping to U.K. Data*

5.5.1 In this sub-section we run the bootstrapping analysis on the DMS U.K. Equity data series. As described above the bootstrapping method relies

[6] One approach is to assume a distribution of the returns, determine the Maximum Likelihood Estimator (MLE) for the parameter(s) of the return distribution. The distribution of the estimator itself can be determined, e.g. computationally or via asymptotic theory. The return distribution and the parameter distribution can then be combined to produce a more volatile return distribution taking account of uncertainty in the parameter estimation. Finally the extreme percentile return from that distribution is calculated to arrive at an estimate of the extreme percentile return taking account of parameter uncertainty. One of the issues with this approach is that where data is limited — as is typically the case when estimating a 1 in 200 year equity return from historical data — we will struggle to estimate the distribution of the parameters. However, research currently being undertaken in academia by Russell Gerrard and Andreas Tsanakas at CASS Business School suggests that for some choices of return distribution the probability of default taking account of parameter uncertainty depends only on the number of observations used to estimate the MLE.

[7] More details can be found in Ross (1997).

on constructing an empirical distribution function which we will then sample from. The empirical distribution function calculated from the DMS U.K. equity return data 1900-2007 is shown in Figure 22.

5.5.2 We can see the empirical distribution function is smoother in the main body of the distribution but is more jagged in the tail. This is what we would expect as there are more data points in centre of the distribution.

5.5.3 We ran the bootstrapping method on this data set to investigate the uncertainty in the 0.5 percentile return. We decided to calculate the 95% confidence interval of the estimator — that is the range in which we could be 95% confident that the true 0.5 percentile equity return would lie.

5.5.4 We have taken two approaches to estimating the 95 percent confidence interval. In the first approach — which we will label PF — we calculate the 2.5 percentile and the 97.5 percentile from the 10,000 estimates of the 0.5 percentile return.

5.5.5 After running 10,000 trials of independent bootstrapping samples we found the central 95% confidence interval of the 0.5 percentile return to be [−18.7%, −48.8%]. This is a range of 30%. This confidence interval has been calculated by using the underlying data without making any distributional prior assumptions.

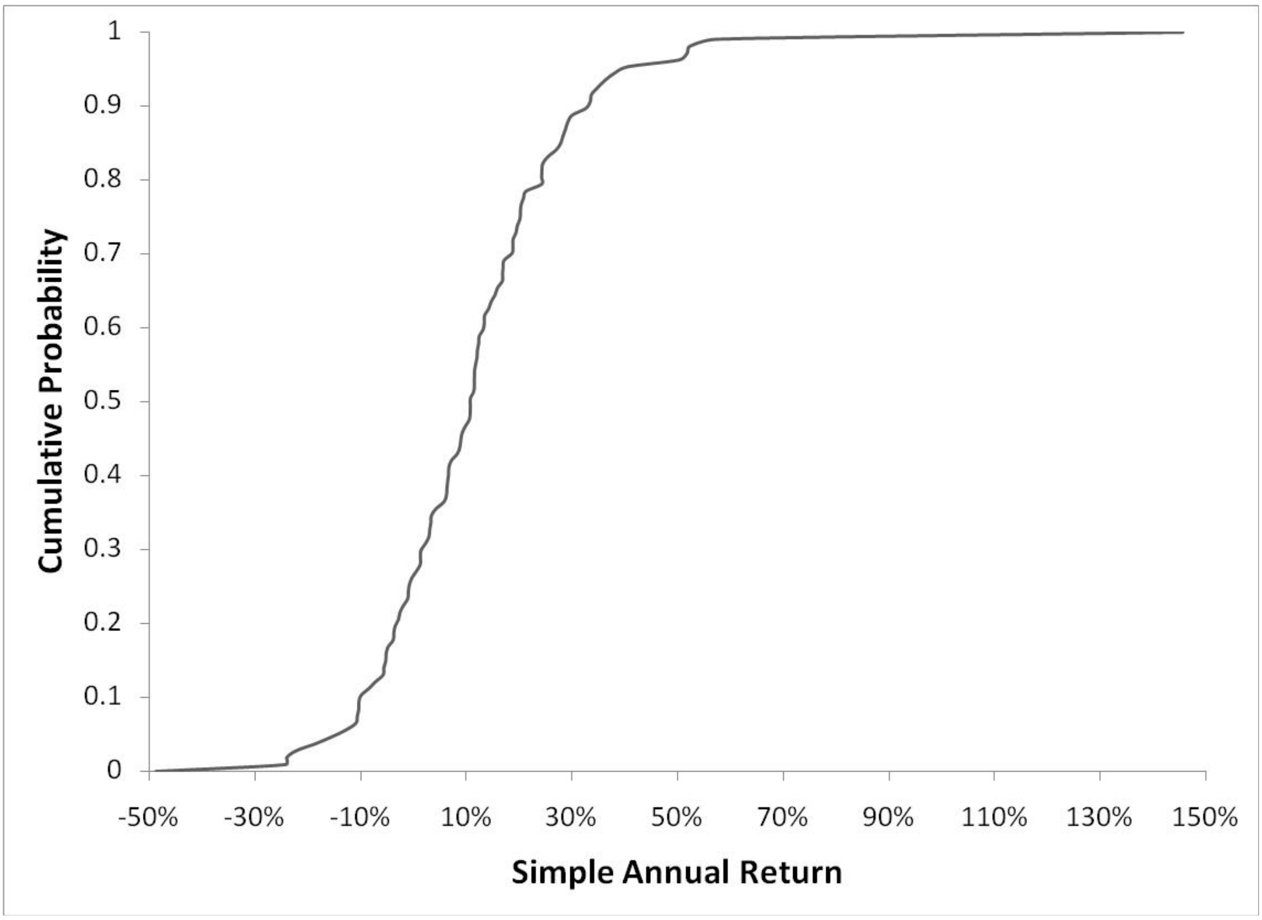

Figure 22

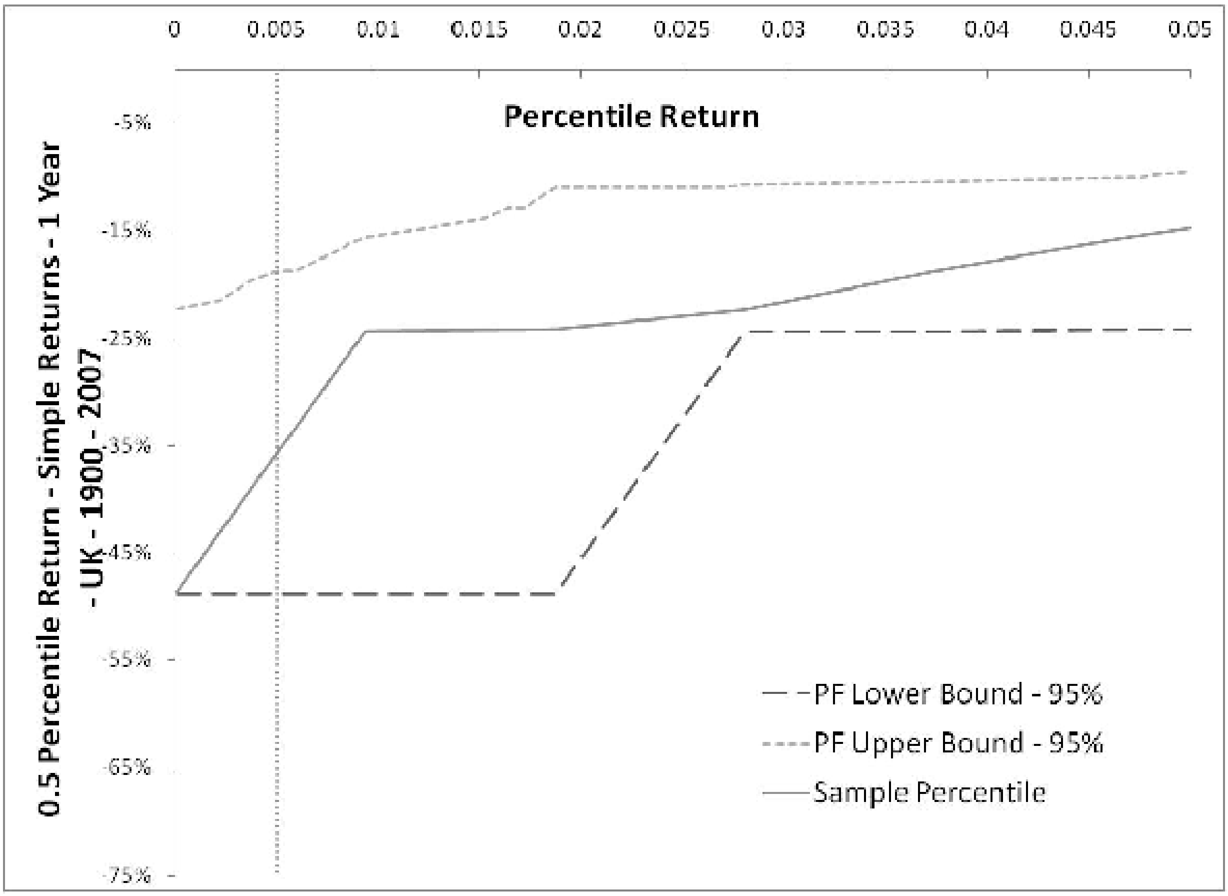

Figure 23

5.5.6 In Figure 23[8] we show the 95% confidence interval for a range of equity return percentiles from 5% down to 0.01% using the *PF* approach.

5.5.7 Investigating the distribution of the 0.5 percentile return estimator we find that it tends to have a particularly skewed distribution as we estimate the far tail events because the minimum return is floored at −48.8%. We could try to get around this limitation by relaxing our *pure frequentist* approach and assuming that we believe the estimates of the 0.5 percentile return should have a particular distribution. For simplicity we assume the estimator has a normal distribution in the analysis that follows.

5.5.8 We now calculate the Mean Squared Error (MSE) of our 0.5 percentile equity return estimates and, assuming a normal distribution for the MSE, calculate the 95% confidence interval. We will label this the *MSE-N* approach.

[8] We have set up these figures with the percentile return increasing along the x-axis. This means that we end up with more extreme percentile stresses nearer to the origin. We mention this in case it causes any confusion to the reader.

5.5.9 The MSE for the 0.5 percentile return is 10.3% which gives a 95% confidence interval for the 0.5 percentile return of [−15.5%, −55.8%]. This range is just over 40% wide; around a third wider than the range we calculated in 5.5.5 using the *PF* approach. This should not surprise us as the *MSE-N* approach permits equity returns lower than those seen in the data sample.

5.5.10 Below, in Figure 24, we show the size of the MSE calculated for a range of percentile equity returns from 0.001% to 5%. The 0.5 percentile return is marked by a vertical dotted line.

5.5.11 We notice that the 10% MSE that we calculated for the 0.5 percentile occurs at the bottom of the most prominent dip in the curve. For example both the 0.01 and the 1 percentile points have a MSE close to 12%. All along the curve there appear to be dips — which are due to the linear interpolation in the percentile calculations. We will label this effect the *Percentile Interpolation Effect*.

5.5.12 Because we are dealing with a relatively small amount of data and we are measuring the percentiles in the tail we often need to rely on the linear interpolation (in the Excel function *percentile*) to derive the percentile

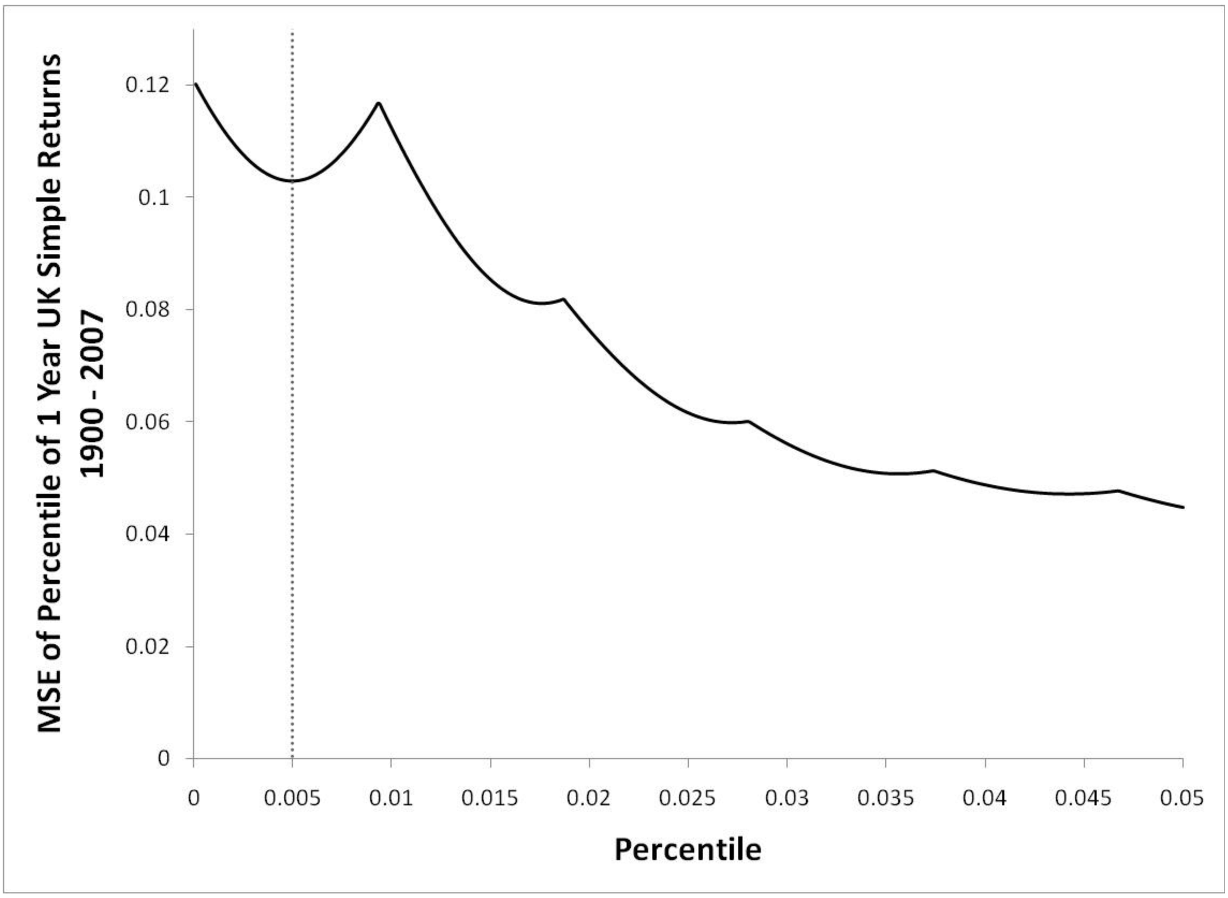

Figure 24

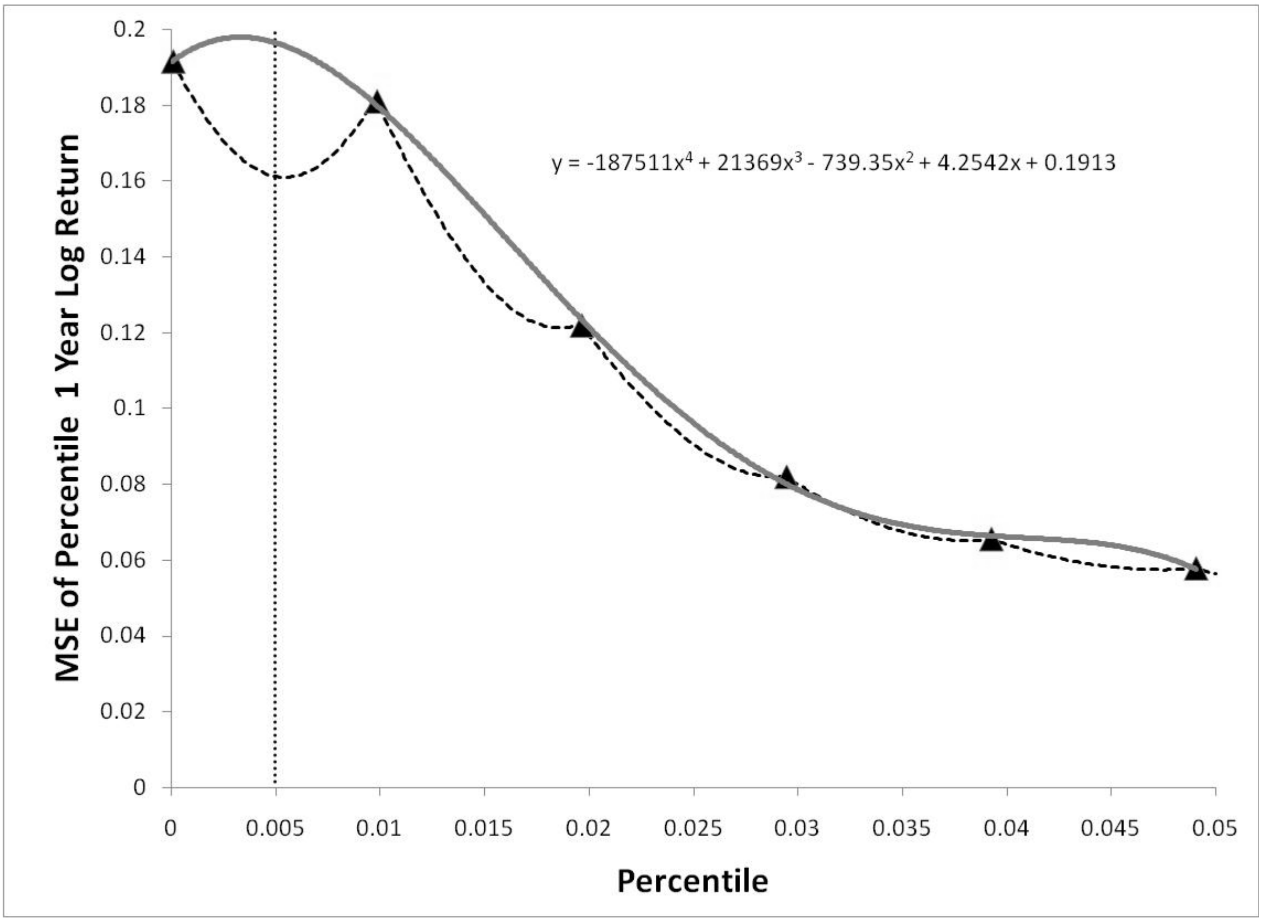

Figure 25

value. It is this linear interpolation that causes the *Percentile Interpolation Effect* by artificially reducing the MSE between the actual data points.

5.5.13 By taking the local maxima from Figure 24 above and constructing a smooth curve through them, we can try to recover what the MSE would be for a percentile function with a more sophisticated interpolation method. The original MSE line and a polynominal interpolation of maxima are shown in Figure 25. We used the smallest order polynomial that would give a good fit to data points.[9]

5.5.14 We can see there is an uplift to the MSE from using a polynomial to fit through the maxima — particularly in the region of the 0.5 percentile return. Focussing on the region up to 1% return we show in the Figure 26

[9] It may be considered unreasonable that the standard error rises as the percentile return increases away from 0. The choice of interpolation function is arbitrary and was chosen the easiest (and quickest method) to get a smooth line through most of the points. If we had taken more time to choose a function which had the desired property of being monotonic we do not believe that the message of our paper would have been significantly different.

that the maximum uplift from using the best fit polynomial actually occurs near to the 0.5 percentile return that we are particularly interested in. The uplift is a factor of 1.23.

5.5.15 Figure 26 suggests the MSE for the 0.5 percentile return should be around 12.5% — rather than the 10% calculated in section 5.5.9 above. If we use an uplifted MSE of 12.5% the 95% confidence interval for the 0.5 percentile equity return would have been [−11.3%, −60.1%] — a range of almost 50%.

5.5.16 It is interesting to superimpose the *MSE-N* confidence interval to the PF confidence interval shown earlier. This shows the effect of assuming the percentile estimator is normally distributed.

5.5.17 We can see in Figure 27 that the *MSE-N* lower bound appears to smooth out the much rougher lower bound of the *PF* approach — especially as we enter the far tail of the distribution. It also enables us to attain a lower bound that extends beyond the maximum observed loss seen in the data. It is not obvious that the 0.5 percentile simple return estimator of should have a normal distribution — especially as the minimum return is possible, −100%, but we feel this is a reasonable first

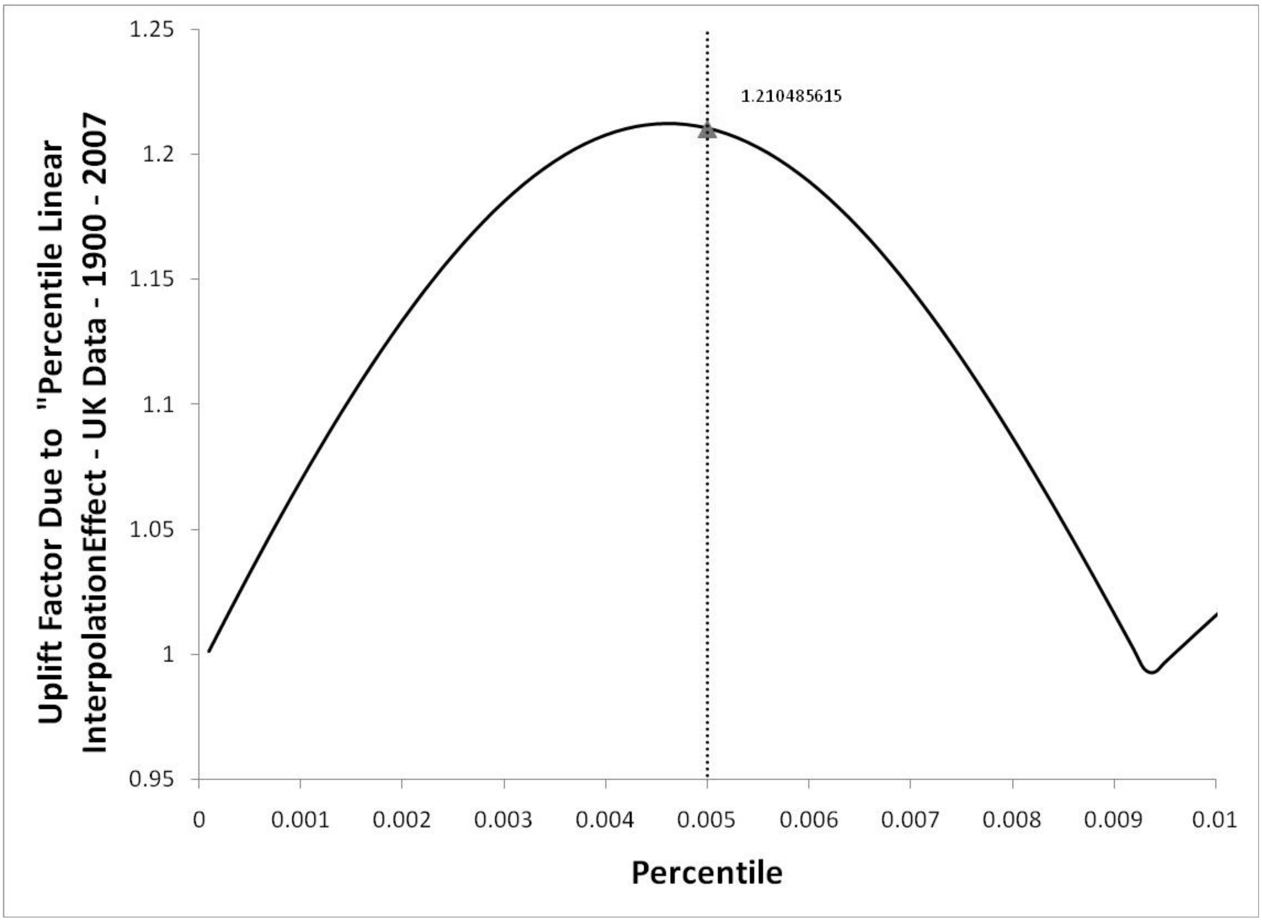

Figure 26

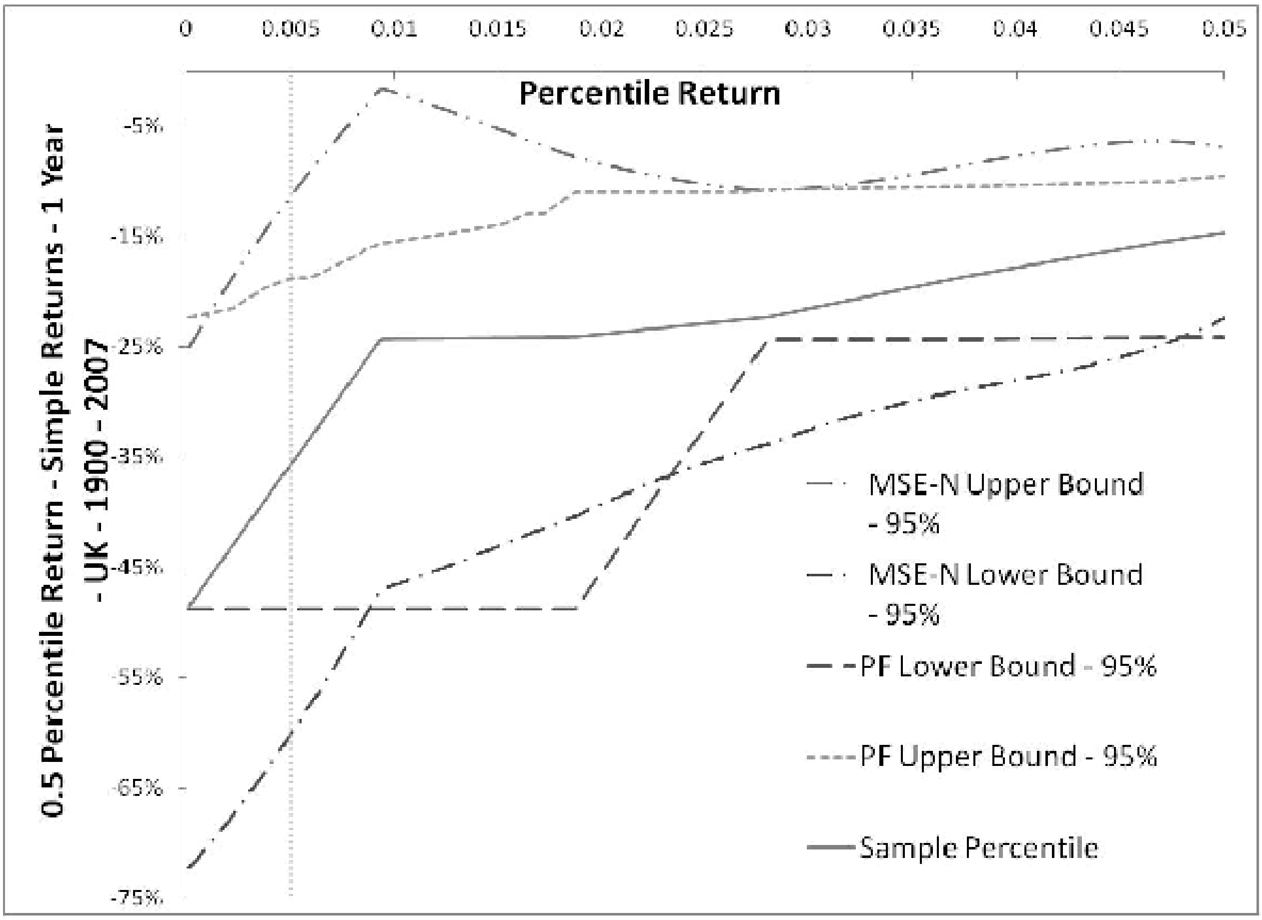

Figure 27

approximation given our percentile estimate is over 60% away from the −100% minimum.[10]

5.6 *Applying Bootstrapping to Global Data*

5.6.1 It is also interesting to consider the PF approach to the 95% confidence interval for the 0.5 percentile equity return, for each of the other economies in the DMS sample. This is illustrated in Figure 28. We have overlaid the CEIOPS QIS-4 equity stress test for comparison.

5.6.2 We can see that Germany has the largest uncertainty in the 0.5 percentile equity return due to a −80% fall in 1948. After Germany the U.K. has the widest uncertainty for the 0.5 percentile return. The width of the confidence intervals of Denmark and Spain are particularly narrow being less than 10%.

[10] The reader may notice that the MSE lines are relatively smooth until just below the 1 percentile point. This is due to the very sharp inflexion in the MSE at this value as can be seen in 5.5.13. combined with the much steeper fall in percentile return at then low percentile return levels.

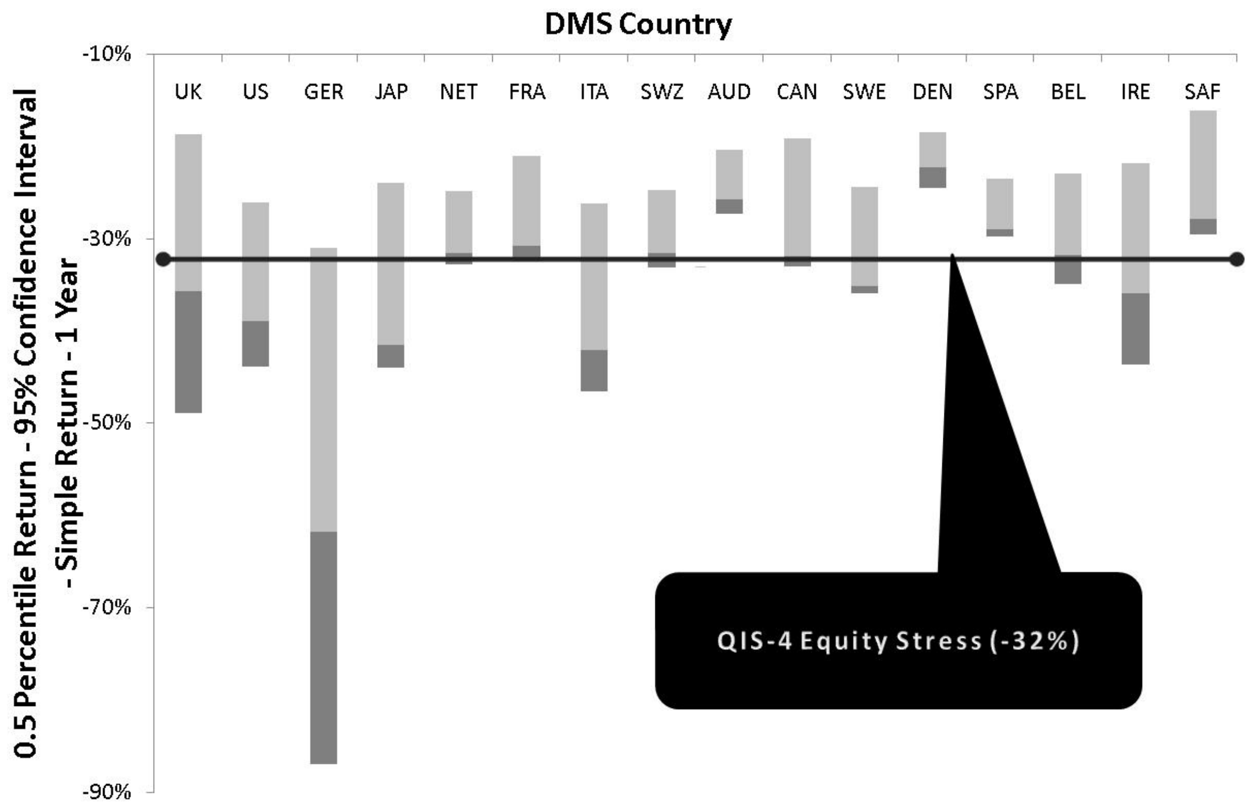

Figure 28

5.7 *Applying Bootstrapping to Aggregate Global Data*

5.7.1 We could make a prior assumption that data from outside the U.K. could also be used in order to estimate the 0.5 percentile return. For example we could assume that the DMS data for 16 countries was independent and therefore that we had 1,648 independent observations of annual equity return data. Many of the calculation issues we experienced due to there being only around 108 data points are eased by using all 1,648 data points from the 16 different economies. This assumption is clearly not true but allows us to reduce the sampling error and to develop a lower bound for the estimation error of the 0.5 percentile return taking account of the returns in other countries.

5.7.2 By re-running the bootstrapping analysis on the full data set for all 16 economies we find that the estimate of the 0.5 percentile equity simple return over one year is −38.2%. The 95% confidence interval using the *PF* method for this estimate is [−33.1%, −43.9%], a range of just 10%. The equivalent confidence interval using just the 108 data points fro the U.K. data set was [−18.7%, −48.8%]; a range of 30%. The assumption that the data from all 16 economies represents 16 centuries of independent returns results in a significantly lower confidence interval than we have seen when using only the U.K. data.

5.8 *Interdependence between Economies*

5.8.1 We need to bear in mind that the 16 equity markets will have reacted to similar economic events over the period under observation. Therefore the equity returns in different countries cannot be considered truly independent.

5.8.2 To get some sense of the level of independence we can measure the correlation over the whole distribution of returns for the various countries in the set. The average of the pair-wise correlation between the returns is around 35%. This suggests a reasonable level of co-dependence but not so much that we should not be able to get some useful reduction in the uncertainty of the statistical estimates by using data from other economies.

5.8.3 However the whole distribution is not our focus in this exercise. Of more importance is the downside tail correlation between the various economies which generally tends to be a lot stronger than the correlation over the whole distribution.

5.8.4 As an example of this elevated tail correlation we can look at the worst simple return loss in the 20th century for all 16 economies covered by DMS data. This is shown in Figure 29.

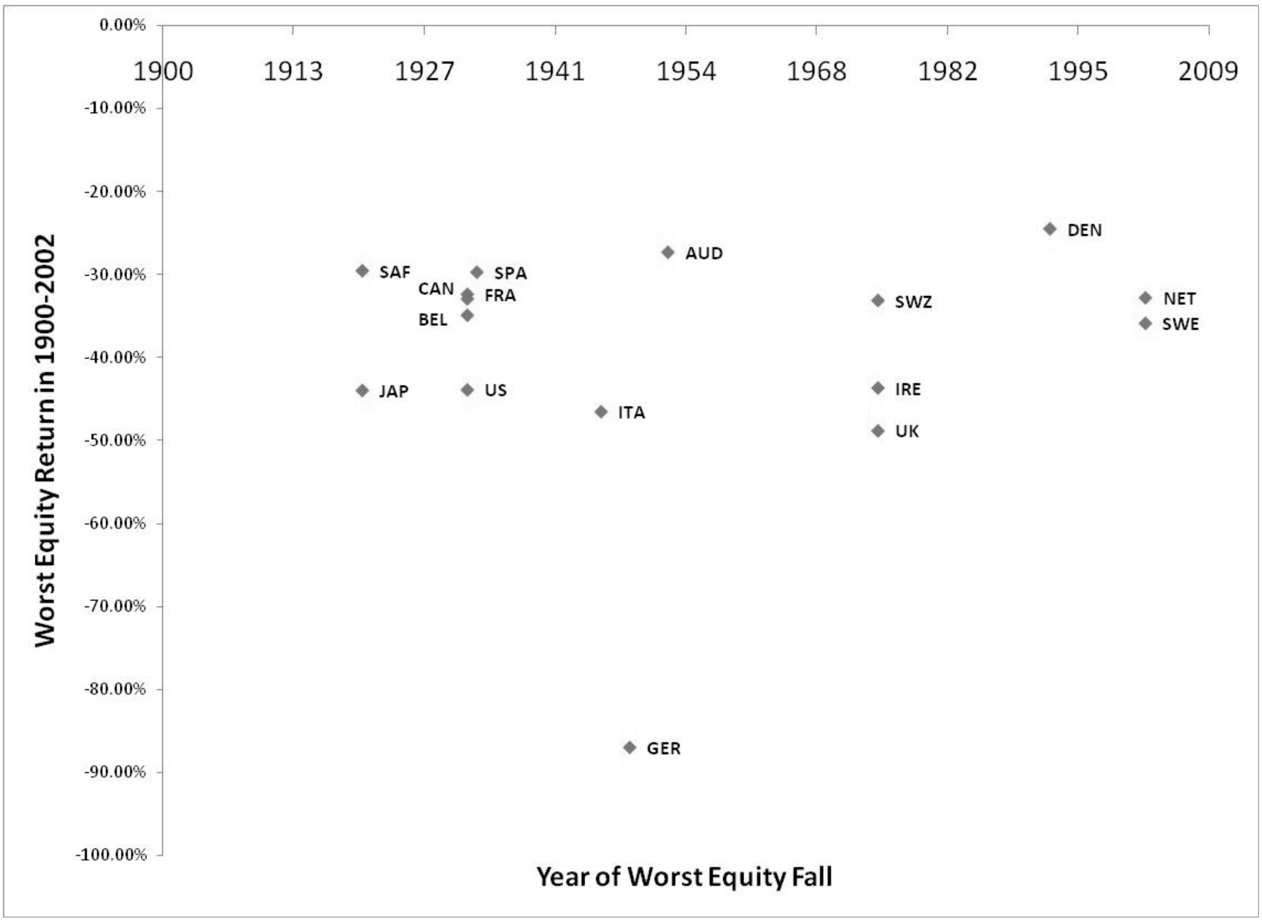

Figure 29

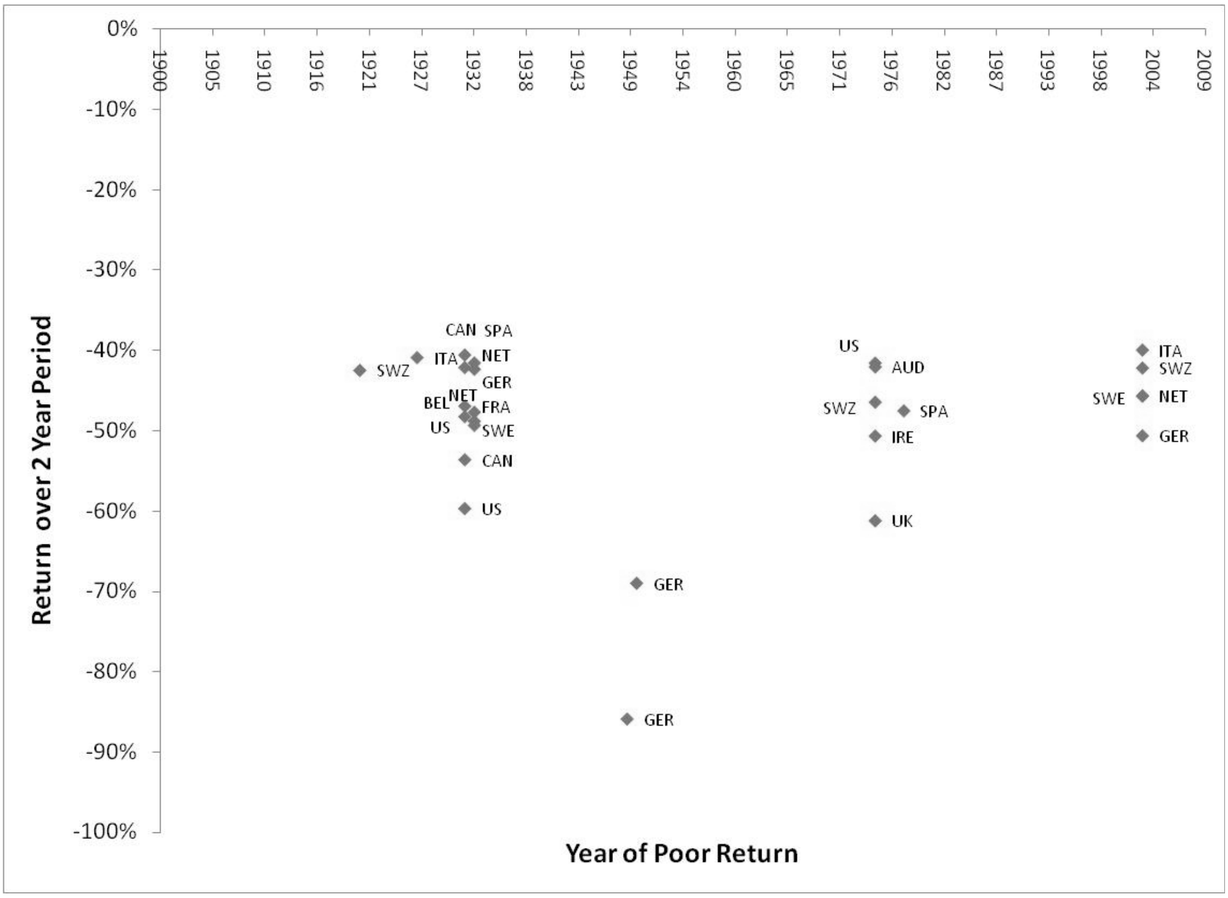

Figure 30

5.8.5 We see that there is clustering with several countries experiencing their worst 20C falls in the same year — notably in the depression of the early 1930s and the economic crisis of the mid 1970s.

5.8.6 In Figure 29 we have only considered returns over the course of a single calendar year. Some of the worst bear markets have occurred over multiple years. In Figure 30 we have considered the two year simple returns for all 16 countries over the period 1990-2007.

5.8.7 Figure 30 shows more clearly how poor returns occur in several markets at the same time sometimes with a slight delay. It is this tendency for poor returns to cluster that leads to elevated tail correlation between equity returns in poor economic conditions.

5.8.8 It is clear from Figure 30 that the depression of the early 1930s caused sustained bear markets in many economies; for example all of U.S.A., Canada, Spain, Netherlands, Germany, France, Sweden and Belgium experienced a fall in their equity market of at least 40% over a 2 year period. Similar clustering occurs in the early 1970s due to the oil crisis and in the early 2000s as a result of the dot com crash. The final cluster around 1948 is due to economic crisis in Germany during the aftermath of the 1939-1945 war.

5.9 *Summary*

5.9.1 Using the DMS data, we estimated the 0.5 percentile 1 year U.K. Equity Total Return to be around 37% measured in simple returns.

5.9.2 We have explored various approaches to determining the 95% confidence interval for a 0.5 percentile equity return. Using historical data from the U.K. only we could make a case for a confidence interval as wide as 50%. On the other hand making the assumption that data from 16 different countries could represent consecutive returns in a single economy, we could make a case for a confidence interval just 10% wide.

5.9.3 Perhaps a prudent approach would be to choose a confidence interval of between 30%-40%, accepting some narrowing of the confidence interval in the light of the other economies, but erring on the side of caution due to strong tail dependence between equity markets and the fact we are only considering equity market responses to the natural, political and economic events of a single century.

6. Equity Modelling — Fitting Distributions

6.1 *Empirical Deviations from Normality*

6.1.1 Although the normal distribution curve is broadly the right shape for many financial return data sets, there is some evidence that the real data diverges from the normal distribution on several points of detail. We have already discussed the presence of *fat tails*. A normal distribution fitted to the mean and standard deviation of a sample tends to underestimate the number of extreme outcomes. A second property is thin peaks. Historical return data shows a greater concentration of returns close to the mean, compared to a normal distribution. These two properties offset a lower than predicted density of returns at moderate distance from the mean.

6.1.2 We have also examined evidence for asymmetry or skew in returns. A typical pattern in equity markets, especially in log returns, is a left tail that is fatter than the right tail. This implies that the median is greater than the mean, because the median calculation is less sensitive to magnitude of extreme events in the left tail.

6.2 *From Moments to Percentiles*

6.2.1 In section 4, we examined equity markets by calculating moments. To construct stress tests, it is necessary to re-express these fitted moments in terms of percentiles.

6.2.2 There is no unique way to calculate percentiles given the first four moments. When skew and kurtosis are both zero, a normal distribution is the natural choice. In the case of non-zero skew and kurtosis, there has been some work on limiting expansions. For example, Cornish and Fisher (1937) consider distributions obtained as the sum of independent random variables.

They derive an approximation to the p^{th} quantile x_p of a random variable X. The first few terms are:

$$x_p = \mu + \sigma z + \tfrac{1}{6}\sigma^{-2}\kappa_3(z^2 - 1) + \tfrac{1}{24}\sigma^{-3}\kappa_4(z^3 - 3z) - \tfrac{1}{36}\sigma^{-5}\kappa_3^2(2z^3 - 5z) + \ldots$$

6.2.3 In this expression, X has mean μ, standard deviation σ, third and fourth cumulants κ_3 and κ_4, respectively. The variable z is the p^{th} quantile of the standard normal distribution.

6.2.4 Focusing on distributions with mean 0 and unit variance, we can consider how the 0.5 percentile, $x_{0.005}$ varies with skewness κ_3 and kurtosis κ_4. Under the normal distribution, the 0.5 percentile is $z = -2.5758293$. In Figure 31 we have plotted the function in 6.2.2 above. The chart shows the factor by which the 0.5$^{\text{th}}$ percentile of a standard normal distribution should be multiplied to allow for non-zero skewness and non-zero kurtosis. For example at the origin, where skewness = kurtosis = 0, we see that the function has a value of 1. This indicates that the 0.5 percentile would be represented by the 0.5$^{\text{th}}$ percentile of a standard normal distribution. If we move up from the origin retaining zero skewness but increasing kurtosis the

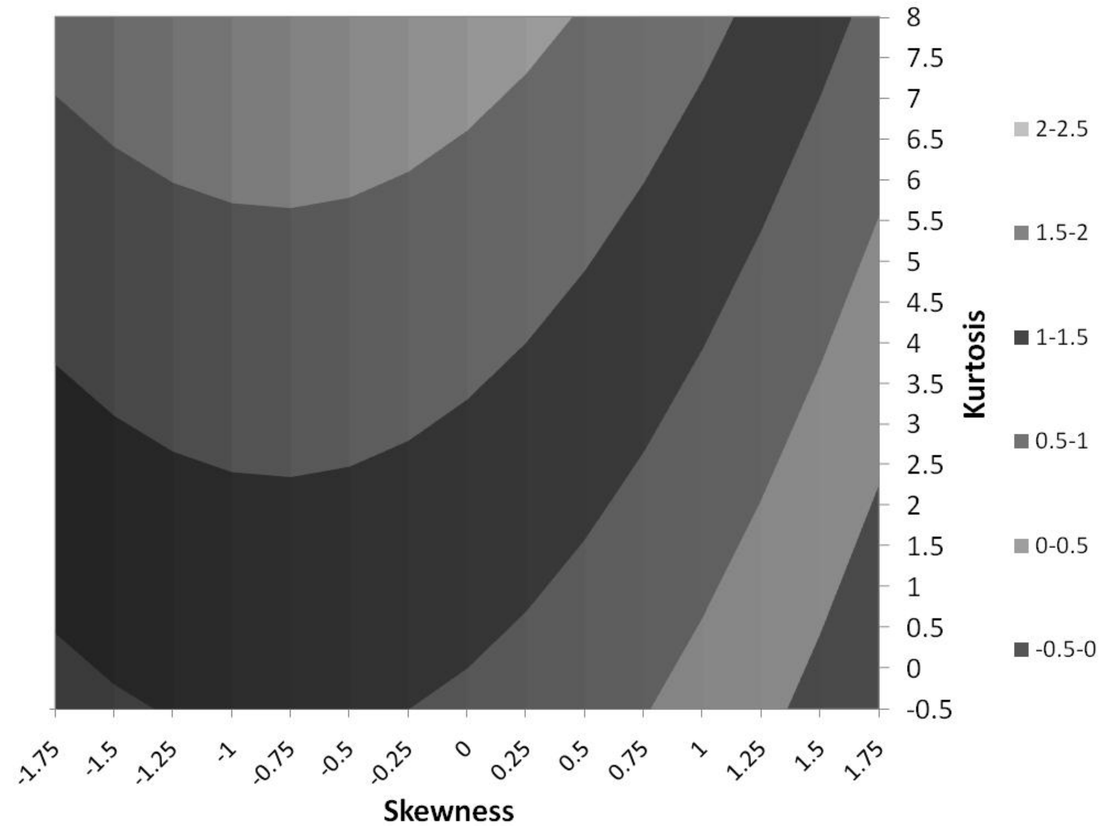

Figure 31

multiplicative factor gets larger.[11] It is interesting to note the way that as the negative skewness becomes more pronounced there is an initial fall in the 0.5^{th} percentile. However beyond a certain point making the skewness more negative actually increases the 0.5^{th} percentile (all other things equal).

6.2.5 It should be noted that there is an identity that constrains the skewness and the kurtosis; $k_4 \geq k_3^2 - 2$. This means that a small area in both the bottom right and the bottom left of Figure 31 are infeasible.

6.3 *Families of Fat Tailed Distributions*

6.3.1 An alternative to asymptotic expansions is explicit calculation using distributions. We consider two families of symmetric distributions, and two asymmetric distributions. These all take values anywhere on the real line.

6.3.2 Our first symmetric family, the Student-T, has polynomial tails, which means that the density behaves asymptotically like some negative power of x for large absolute values of x. This distribution currently underlies the RiskMetrics value-at-risk methodology (see www.riskmetrics.com). William Gossett first published the Student-T distribution in 1908. The density is given by the formula

6.3.3

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}.$$

6.3.4 Smith (2005) proposes investigation of a four-parameter generalistion of the Student-T distribution. This is the Pearson Type IV distribution, defined by the following density:

$$f(x) = \frac{2^{2m-2}\Gamma(m + \frac{1}{2}iv)\Gamma(m - \frac{1}{2}iv)}{\pi a \Gamma(2m-1)} k \left[1 + \left(\frac{x-\lambda}{a}\right)^2\right]^{-m} \exp\left[-v \tan^{-1}\left(\frac{x-\lambda}{a}\right)\right].$$

6.3.5 Eberlein and von Hammerstein (2002) have extensively studied Barndorff-Nielsen's (1977) hyperbolic distribution for modelling market returns. Unlike Student-T, the hyperbolic distribution has exponential tails and is defined by the density

[11] To create this chart we used the function in 6.2.2 with $\mu = 0$, $\sigma = 1$, $z = -2.57$. The result of the function in 6.2.2 was then divided by $z = -2.57$ to get the multiplicative factor illustrated as a function of skewness, k_3, and excess kurtosis, k_4.

$$f(x;\mu,\alpha,\beta,\delta)=\frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)\right\}.$$

6.3.6 The special case of a symmetric hyperbolic distribution arises when $\beta=0$:

$$f(x;\mu,\alpha,\delta)=\frac{1}{2\delta K_1(\alpha\delta)}\exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2}\right\}.$$

6.3.7 Several other families have been proposed in the literature. One popular family, arising from extreme value theory (see Embrechts, Mikosch & Kuppelberg, 2008) is the generalised Pareto distribution. These distributions are usually fitted only to the tail of a data set, which requires the selection of a cut-off point to define where the tail starts. There is currently little consensus on how best to choose the cutoff point objectively and consistently, so we have not followed this route in the current paper. However, to the extent that generalised Pareto distributions exhibit power law or exponential tails, we expect the results of these distributions to approximate our figures for the Pearson IV and hyperbolic distributions, respectively.

6.3.8 Another widely cited distribution family is the stable family proposed by Levy (1953) and popularised for stock market modelling by Mandelbrot (1963). Walter (1989) offers some fitted parameters to various markets, and Finkelstein (1997) examines this distribution in a U.K. context. This family contains the normal distribution, but otherwise produces distributions of infinite variance. Parameter estimation is a significant technical challenge, as Levy stable distributions are defined by their characteristic function. Analytical expressions for the density are available only in special cases which, sadly, do not appear to correspond to financial data. If we had managed to fit Levy stable distributions, it is likely that the results would be even more extreme than those we obtain for the power law tail using the Pearson IV family.

6.4 *Fitted Distributions by Method of Moments or Maximum Likelihood*

6.4.1 There are two established methods for fitting statistical distributions to data samples: maximum likelihood and the method of moments. Maximum likelihod chooses parameters to maximise the product of the probability density over all the observations. The method of moments chooses parameters to equate the first few sample moments of the data to their theoretical values according to the distribution.

6.4.2 There is much theory about the properties of maximum likelihood parameter estimates. In particular, as the sample size increases, the distribution of maximum likelihood estimates tends to a multivariate normal distribution, whose variance-covariance matrix is related to the second derivative of the density at the optimum (the first derivative being zero, of course). This theory allows us to estimate the effect of parameter error — provided we believe the data set is "sufficiently" large for the limiting regime to apply. It also provides a demonstration (the famous Cramer Rao Lower Bound) that no other method could produce more accurate (less variable) parameter estimates than maximum likelihood. Finally, the maximum achieved provides a goodness of fit statistic which permits comparison of fit between different distribution families.

6.4.3 We have examined annual returns sampled from the U.K. MSCI index from 31/12/1969 to 31/08/2008. Given the fatness of observed tails and the influence of a small number of data points on our results, it is difficult to defend our 38 years of market prices as "sufficiently large" to rely on asymptotic normality theorems. In addition, the theory requires the data to be a random sample, which does not apply in our case because of overlapping time intervals and volatility clustering. The asymptotic result also relies on the hypothesis that the true underlying distribution lies in the chosen family; a hypothesis that cannot simultaneously be true for all the families we investigate! We also constantly struggle with more mundane matters of convergence — the likelihood maximisation suffers from multiple local maxima which pose a challenge for optimisation algorithms. We have resorted to subjective inspections of fits by eye to convince ourselves that our fits obtained are global and not only local maxima.

6.4.4 The method of moments is computationally simpler, because the estimation of moments is a mechanical calculation requiring no optimisation. Furthermore, the back-solving to find parameters from stated moments is well understood and produces a unique distribution for the families we investigate. The Pearson IV is particularly well-behaved in this regard, as the moments can be derived analytically and the complex Gamma functions cancel out. To fit hyperbolic distributions is more challenging because of the appearance of Bessel functions, but robust approximations are available (Abramowitz and Stegun, 1972). When we fit symmetric distributions by the method of moments, we focus on the first, second and fourth moments (because the theoretical skew of a symmetric distribution is zero). The method of moments has the further advantage of relating fitted percentiles directly to tabulated moments, facilitating comparison between economies. This method suffers from a major disadvantage in the case of the Pearson IV distribution, that some parameter values result in infinite moments, and these values are inaccessible by method of moments fitting.

6.5 *Graphical Analysis of U.K. Data*

6.5.1 Figure 32 shows a Q-Q plot for U.K. equity returns. The data are annual returns, sampled from the U.K. MSCI index, starting 31/12/1969

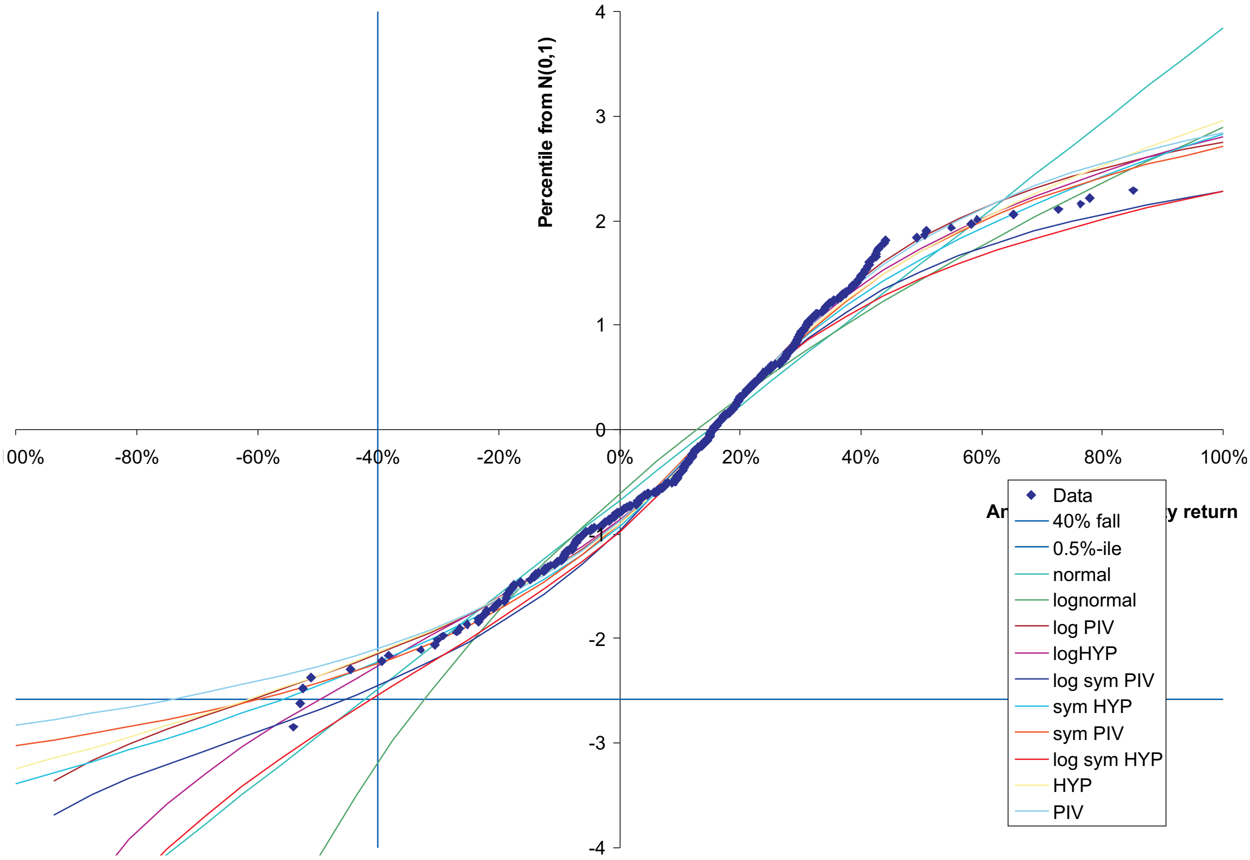

Figure 32

through to 31/08/2008. We use monthly starting dates, so the calculated returns overlap to some degree. This is evident at the lower end of the scale, where a cluster of extreme negative returns are close because these data points all refer to the same crash event in 1974.

6.5.2 On the horizontal axis is the total return achieved over one year. These are sorted into ascending order. On the vertical axis are the corresponding percentiles of the standard $N(0, 1)$ distribution. Of particular interest are the blue line, $x = -40\%$, which corresponds to a widely used stress test, and $y = -2.58$ corresponds to a 1-in-200 event (because the cumulative normal probability function is 0.005 at $y = -2.58$). To justify a 40% test, we need to fit a curve to the blue data points that also passes through the intersection of the blue lines at $(-40\%, -2.58)$.

6.5.3 If the original data were drawn from a normal distribution, then (apart from sampling error) the data should lie on a straight line. The x-intercept would be the mean, and the slope would be the reciprocal of the standard deviation. The pale blue line is a fitted normal distribution. The intercept with $y = -2.58$ is slightly to the left of -40%, so a fitted normal distribution would produce a stress test slightly more onerous than 40%.

6.5.4 A fitted lognormal distribution (for 1 + return) corresponds to the green curve, which is essentially a logarithmic function. This produces a stress test less severe than 40%, but the data fit is poor. Despite the widespread use of lognormal return distributions in finance, the historical returns do not support its use for this application.

6.5.5 Figure 32 shows best fit distributions (by maximum likelihood) to return and log(1+return) data. The distributions are Pearson Type IV (PIV) and hyperbolic (HYP). Both of these distributions are four-parameter families, allowing calibration of location, scale, asymmetry and tail fatness. The PIV has power law tails and the HYP has exponential tails, although our log transform to the data prevents predicted returns below −100%. We have also looked at the subsets of these which are symmetrical. We can see that allowing for the fatter tails produces significantly more onerous stress tests at the 1-in-200 level, in the PIV case, a fall in excess of 70%.

6.6 *Commentary on the U.K. Results*

6.6.1 In fitting these data, we must not lose sight of one simple fact. From the end of November 1973 to the end of November 1974, the U.K. market (measured on a total return basis) fell by 54%. As far as we are aware, this is the most severe one-year stock market loss in U.K. recorded history. Attempts to fit distributions usually result in fitted distribution functions passing near that point. This means, for example, that if we use 40 years of data, then a 54% fall looks like a 1-in-40 year event. If we use 100 years of data, then a 54% fall looks like a 1-in-100 year event. We do not have 1000 years of data, but hypothetically, if we did, and 1974 was still the worst year in that sample, we would estimate 54% as a 1-in-1000 year event. At the other extreme, if we use fewer than 34 years' data, then this extreme event is excluded from the data set — we could easily convince ourselves, or even "prove" statistically, that a 54% fall is far more extreme than 1-in-200. This simple arithmetic substantially explains the reasons for different results of analysis based on different U.K. data sets. The most onerous fitted stress test arises from the use of a 40 year data period, with either longer or shorter data periods implying less onerous stress. It so happens that 40 years is the period for which the MSCI indices are available. Although these indices are widely used, and recognised for their international consistency (see Wilkie, 1995) they do have this disadvantage of producing large stress tests. The moral of this story, for would-be data producers seeking widespread citation, is to avoid a data start date immediately prior to the largest market crash of the century.

6.6.2 Even after settling on a single data set, the fitted curves for U.K. produce a wide range of values for the 1-in-200 fall. The most extreme results are from a Pearson Type IV, applied to simple returns, which implies a fall of 75% at the 1-in-200 probability level. At the other extreme is the lognormal distribution, with a fit implying that even a 35%

fall would be more extreme than 1-in-200. Other distributions produce intermediate results.

6.6.3 The use of overlapping data intervals frustrates our efforts to test statistical significance of the fit. However, a visual inspection of the QQ plot suggests that the lognormal distribution is a particularly poor fit to the data.

6.6.4 In theory, the maximum likelihood technique also allows us to estimate confidence intervals for fitted percentiles. Unfortunately, our use of overlapping intervals once again invalidates the classical formulas for confidence intervals. If we argued, implausibly, that the overlapping intervals contain as much information as independent observations, then our 95% confidence interval for the 0.5 percentile is around 5% wide, varying from one distribution family to another. If we argue, equally implausibly, that the overlapping data has no more information than annual data without the overlap, then the width of the 95% confidence interval is around 15%. Even with this wider confidence interval, we can see that 95% confidence intervals from different models fail to overlap. Naively, this seems wrong, as two non-overlapping sets cannot both contain 95% of outcomes. The explanation is that each confidence interval is valid only when the chosen distribution family contains the "true" model. This cannot hold simultaneously for all our models, so not all of the confidence intervals are valid. This suggests we treat any claimed confidence interval with caution.

6.6.5 If overlapping intervals cause us so much trouble, why do we not consider only annual intervals, for example December to December? There are two reasons for preferring overlapping data. Firstly, the use of December data would overlook some key historic events, in particular the crash of October 1987 which recovered by the end of the year. So annual returns starting at points other than year-end, do possess some information, even if we are unsure exactly how much. Secondly, even if we restrict ourselves to December data, we still do not have independent samples, because of the effect of volatility clustering, and so the classical tests still do not strictly apply.

6.7 *International Comparisons*

6.7.1 We have encountered difficulties in the construction of sampling errors. One pragmatic way to evaluate the robustness of apparent patterns is to test them on international data. A pattern that applies consistently across many countries is more likely to be robust. We therefore repeated the analysis on the other 17 countries for which MSCI index data is available.

6.7.2 The wide spread of results from different models is consistently seen internationally. There is no country for which the models agree on a 1-in-200 year stress test. In fact, there is no country for which the most extreme model (usually the Pearson IV) implied a fall of less than double the fall implied by the mildest model (always lognormal).

6.7.3 We can, however, often justify rejection of these two extreme cases. The lognormal is a consistently poor fit, failing to capture observed tail behaviour in every country. In some countries we can also dismiss the Pearson IV applied to simple returns, because it implausibly implies a material probability of a simple return below −100%. This would require equities to flip over from assets to liabilities, such an outcome being implausible, on the grounds of limited liability. Where this flip has a fitted probability in excess of 0.5%, the implied stress test is a fall exceeding 100%, which is economically nonsensical.

6.7.4 There is no consistent pattern in the goodness of fit between Pearson IV and hyperbolic distributions. The maximised likelihoods are usually close, with the apparent better fit varying from one country to another. This suggests there is insufficient data to determine whether the tails follow asymptotic power laws or have exponential tails. There is also insufficient data to reject both of these alternatives in favour of another form of tail behaviour. As the two families are not nested, that is, neither contains the other, the use of statistical tests to distinguish them is technically questionable in any case. We are left in the uncomfortable situation of being unable to distinguish statistically between competing models that nevertheless produce dramatically different results. The Pearson IV, with power law tails, usually produces the more onerous stress test at the 1-in-200 level.

6.7.5 There is also little consistent evidence as to whether taking logs improves the fit or not. Fitting to log returns removes the problem of fitted stress tests below −100%. Even where this issue does not arise, the use of log distributions usually produces less onerous stress tests.

6.7.6 A restriction to symmetric distributions has only a small effect on the fitting simple returns. The distributions appear approximately symmetric anyway. Imposition of symmetry causes a modest reduction in maximised likelihood — there must always be some reduction because we are forcing the asymmetry parameter to be zero. In the fit of asymmetric distributions, there is no consistency to whether the fitted distribution is left or right skewed. On the other hand, log returns are consistently left-skewed, and the inclusion of asymmetry parameters has a large effect on the goodness of fit.

6.7.7 While the use of fatter-tailed distributions often results in larger fitted stress tests, this is not always the case. For some countries, an allowance for distribution shape implies a less onerous 1-in-200 test than a simple fit of a normal distribution. This occurs when, for that particular country, there have been no extreme falls observed in the chosen data period. The lack of extreme events in the history does not exclude their future occurrence. However, these differences serve to highlight the importance of international comparisons in deriving stress tests, as well as the dependence of fitted results on the peculiarities of a particular chosen data set.

6.7.8 By way of some international comparisons we now present some

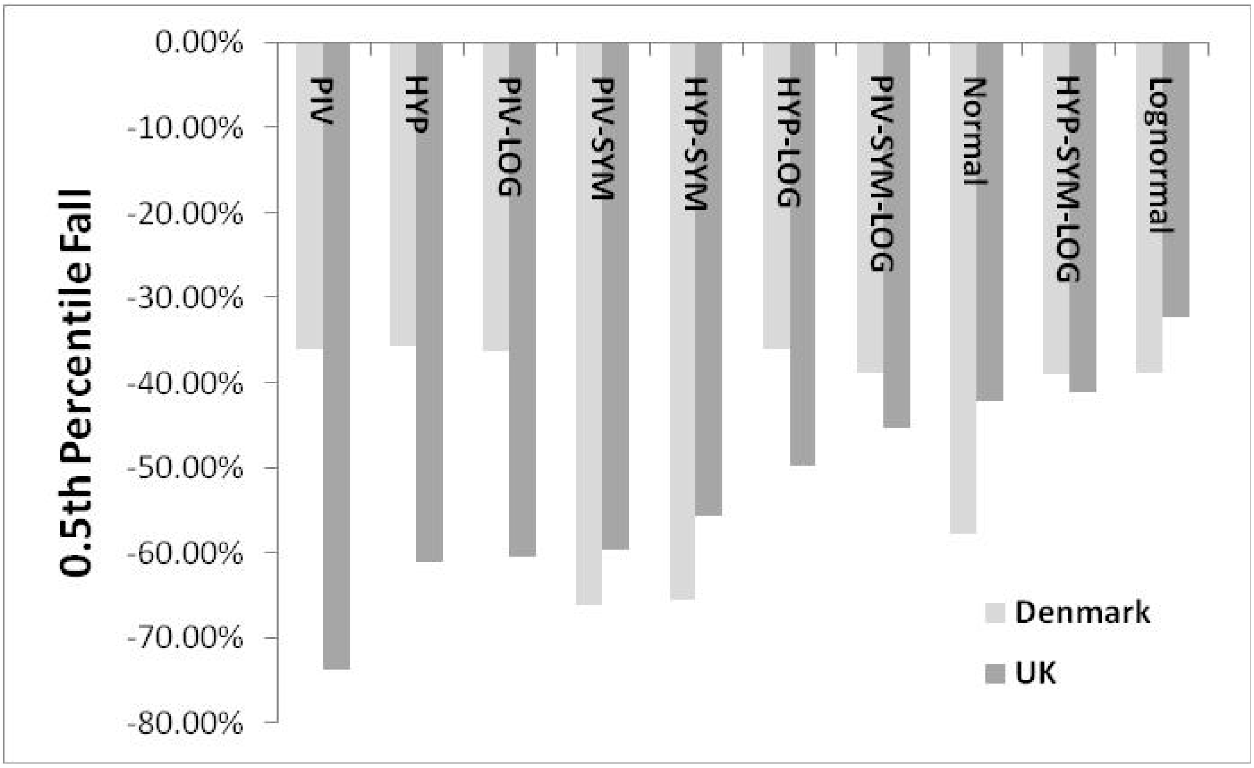

Figure 33

figures showing how the different models perform for different countries. Figure 33 shows the critical 0.5th percentile return for each distribution in respect of two contrasting economies; U.K. and Denmark. The models are ordered in worst to best order for the U.K. It can be seen that this ordering is quite different from Denmark which has experienced much more benign equity market falls.

6.7.9 Figure 34 shows the results for all the economies covered using a single model; the Pearson IV distribution (not taking logs and non-symmetric). The data are ordered from worst to best 0.5th percentile return. We note that there is a wide dispersion in the results from around −30% for Austria to below −70% for Hong Kong and the U.K.

6.7.10 Finally, in Figure 35, we look at the worst 0.5th percentile return for each country and noted the model that produced this worst return. We plotted these in worst to best order. We can see that many of the worst results are caused by the symmetric PIV model.

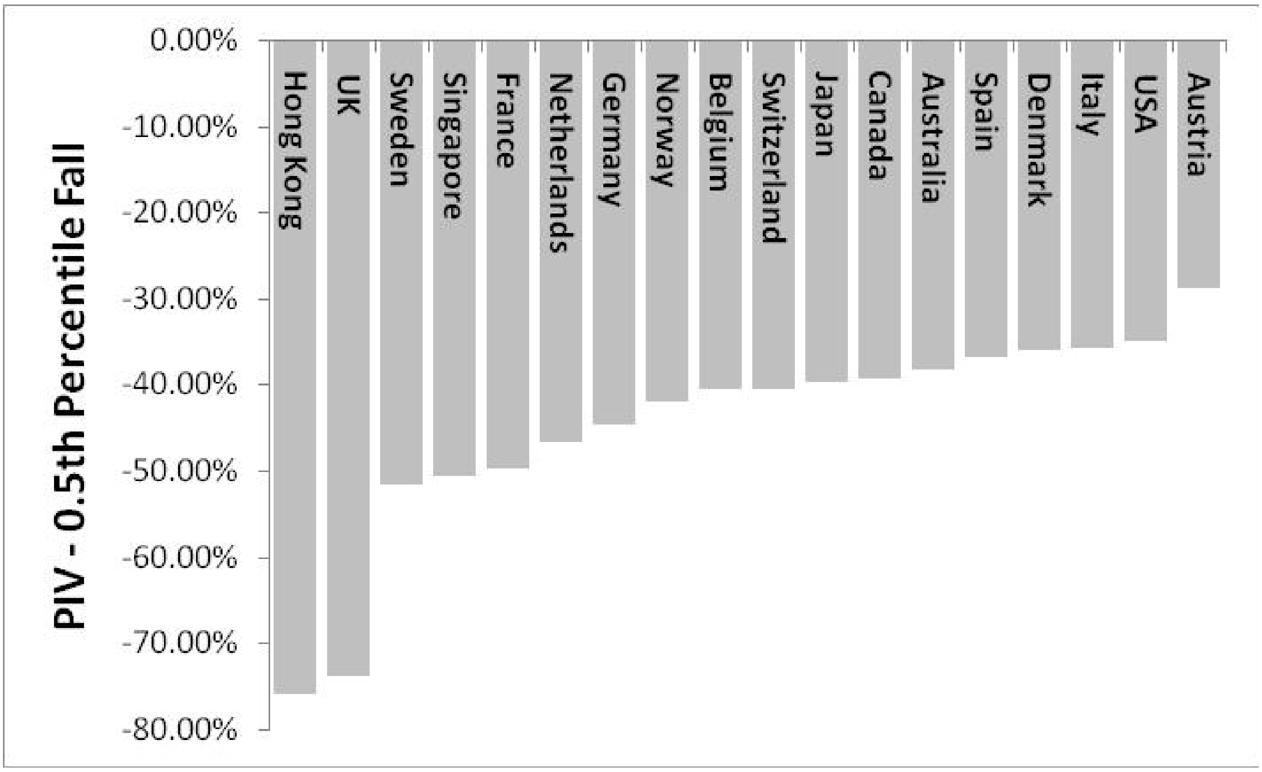

Figure 34

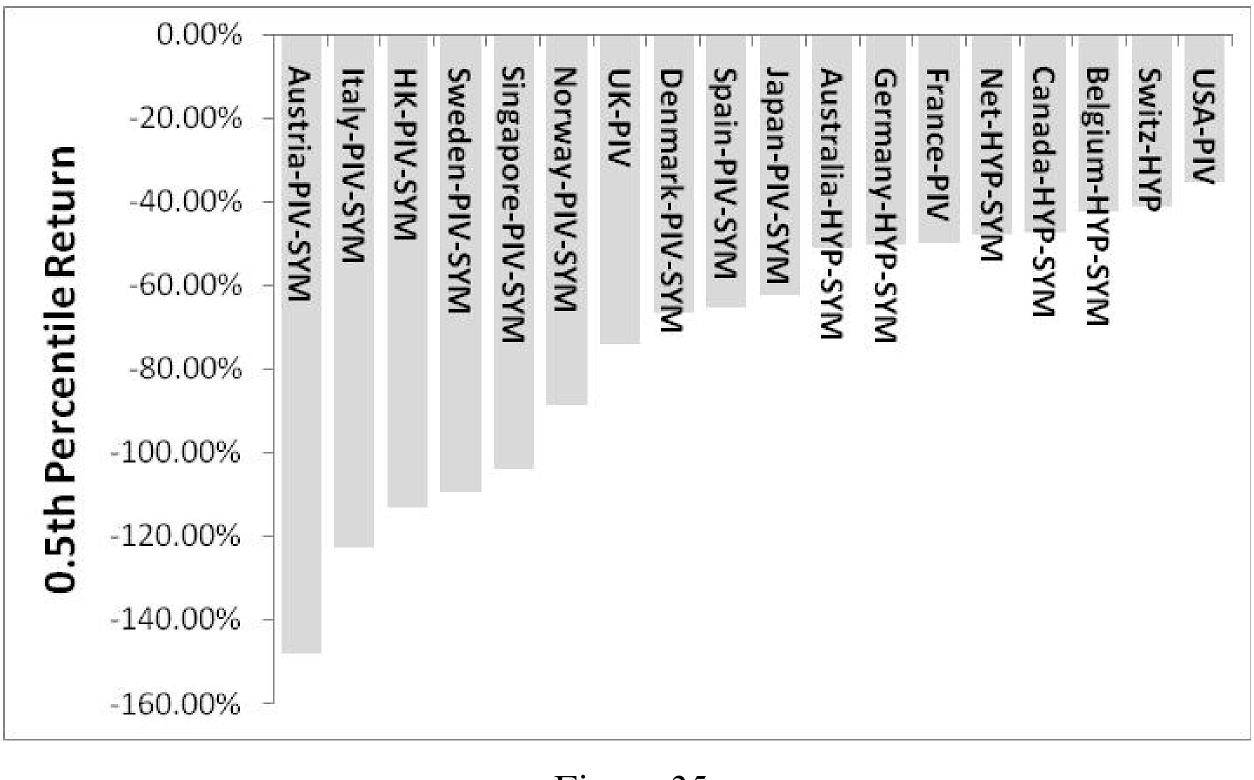

Figure 35

7. Yield Curve Modelling

7.1 *Introduction*

7.1.1 Unlike the analysis of equity returns, where returns on individual stocks are compressed into a single index, the analysis of interest rates tends to use the whole term structure of the yield curve, adding extra complexity to the analysis.

7.1.2 A number of methods have been used to make the analysis more tractable. A crude approach is to consider only parallel shifts in the yield curve, perhaps based on analysis of yields for a single term. A slightly more sophisticated approach is to observe that the yield curve can be considered as a family of random variables to which we can apply principal components analysis (see Appendix C for a description of this approach). As yields of similar terms are highly correlated it is not surprising to find that the first component corresponds roughly to a parallel shift. However there is also a tendency for short-term yields to have higher volatility than long-term yields, so the shift is not quite parallel.

7.1.3 A potential drawback of a model with a limited number of factors is that it does not allow for the full range of possible changes in the shape of yield curves, since a curve contains an infinite number of points. Infinite factor models are discussed briefly later in this section. In practice, only a finite number of points on a yield curve correspond to bonds that are traded in the market and the rest of the curve is filled in by interpolation. The implications of this are discussed further in Appendix C.

7.1.4 One noticeable feature of recent U.K. interest rate data is the significant and sustained falls in general yield levels from around 1997, coinciding with the point at which the Monetary Policy Committee of the Bank of England took responsibility for setting base rates. Current yield levels are therefore low compared with a longer-term historical average. This can cause a problem when estimating extreme percentiles for falls in the yield curve, since an extreme fall applied to an initial yield curve that is well below the long-term average may lead to very low or negative yields. One common approach to get around this is to consider relative rather than absolute changes in yields, which typically leads to stresses expressed as ratios of the starting yield instead of a movement by a fixed amount.

7.2 *Choice of Yield Curve*

7.2.1 The yield on a bond is defined as the internal rate of return, at which the market price of the bond price is equal to the present value of its cash flows.

7.2.2 Empirically, bond yields depend on many factors. Most important are the term of the bond, its coupon level and timings, and the extent of any credit risk. Also of importance are effects due to liquidity and whether bid or offer prices are used. There may be a further inconvenience premium due to adverse tax, regulatory or accounting treatment.

7.2.3 The “yield curve” refers to yields as a function of time. This requires standardisation of the other factors affecting yields. Our analysis uses yield curves derived from interest swaps based on 3-month or 6-month LIBOR. In that case, maturities are a whole number of periods from the valuation dates, and credit risk reflects the banks submitting data to the LIBOR panel, refreshed throughout the term of the swap. Published swap yields are equivalent to par yields. Swap data has the advantage of covering a range of terms, with broadly homogenous contract design, tax and regulatory treatment over across a range of economies and contract terms. Unfortunately, historic interest rate swap data are available for major world currencies only from the mid 1990s, with further limitations on data availability for long terms or minor currencies.

7.2.4 We define a yield curve in terms of the prices of zero coupon bonds. We use P_t to denote the price of a zero coupon bond of term t. We expect $P_0 = 1$, and usually P_t should be a continuous decreasing function of t. We distinguish a starting curve P_t^{base} from a stressed curve P_t^{stress} after a possible market move.

7.2.5 A capital assessment requires consideration of all risks to bond values, whether caused by moves in a reference yield curve, credit events, or changes in spreads due to liquidity, credit or convenience. In this paper, however, we focus only on moves related to yield curves.

7.3 *Factor Models of the Yield Curve*

7.3.1 Stress test constructions for yield curves must start from a study of probable yield curve moves. The first step is to quantify the volatility of yield moves as a function of term. This measurement may refer either to absolute changes in yields, or to proportional changes, the latter being broadly equivalent to changes in log yields. In this paper, we have analysed absolute changes in yields. The table below shows the historic volatility of yield changes, measured for sterling and Euro over the last ten years. For the Euro figures, we have used Deutsche Mark swaps prior to to the introduction of the Euro.

Term	Anualised standard deviation of changes in GBP swap yields	Anualised standard deviation of changes in EUR swap yields
1	0.70%	0.54%
2	0.78%	0.67%
3	0.79%	0.69%
5	0.75%	0.67%
10	0.69%	0.61%
20	0.65%	0.57%

7.3.2 One way to think about yield moves is in terms of factors. There are several different representations of the yield curve, and factor models can be applied in different ways according to the chosen representation. Possible factor models are shown in the table below.

Representation	
Zero coupon bonds	$P_t^{stress} = P_t^{base} * \left\{1 + Z_1\zeta_1(t) + Z_2\zeta_2(t)\right\}$
Annualised spot yields	$(P_t^{stress})^{-1/t} = 1 + \left\{(P_t^{base})^{-1/t} - 1\right\} * \left\{1 + Z_1\zeta_1(t) + Z_2\zeta_2(t)\right\}$
Continuously compounded spot yields	$-\ln P_t^{stress} = -\ln P_t^{base} * \left\{1 + Z_1\zeta_1(t) + Z_2\zeta_2(t)\right\}$
Continuously compounded forward yields	$-\frac{d}{dt}\ln P_t^{stress} = -\frac{d}{dt}\ln P_t^{base} * \left\{1 + Z_1\zeta_1(t) + Z_2\zeta_2(t)\right\}$

7.3.3 European regulators, in QIS3 and QIS4, focus on the annual spot yield representation. Many models used for option pricing, including the famous Heath-Jarrow-Morton framework, use the forward yield representation. In this paper, we use the zero coupon bond representation, because it allows exact value at risk calculations in the important special case when cash flows are fixed.

7.3.4 In a factor model construction, Z_1 and Z_2 are independent standard normal variables. The functions $\zeta_1(t)$, $\zeta_2(t)$ represent different possible shocks to the shape of the curve. For example, $\zeta_1(t)$ might capture changes to the level of the yield curve, and $\zeta_2(t)$ the slope. These functions must satisfy some constraints. For example, the $\zeta_i(t)$ must be smooth functions in order to produce a suitable degree of smoothness in the stressed bond prices. In the zero coupon bond representation, we must also have $\zeta_i(0) = 0$.

7.3.5 We showed an example of a 2-factor model above. If we know two points on the stressed yield curve, we can (by linear elimination) determine the values at all points. Models with three or more factors are constructed analogously by adding the obvious further terms.

7.4 *Infinite Factor Models*

7.4.1 Calibration of factor models for risk purposes is often problematic. Difficulties arise because, empirically, factors seem to come and go over different periods. One period may be dominated by changes in level, while over a different period the short end of the curve may fluctuate while the long end remains more or less fixed. We cannot be confident that moves in past yield curves are representative of future changes. The calibration problem is exacerbated by the large number of parameters to be estimated — in the three factor case, we must calibrate three functions of time. There is a danger that calibration simply wraps the model around historic data.

7.4.2 An alternative approach is to model yield curve changes using a small number of parameters but a very large number of factors. The small parameter count has the effect of smoothing out any wrinkles from historic data sets, potentially reducing sampling error and more robust for projection. One such approach is the IOU (integrated Ornstein-Uhlenbeck) model.

7.4.3 The integrated Ornstein-Uhlenbeck process takes the form of

$$P_t^{stress} = P_t^{base} * \left\{ 1 + \int_0^t OU_s ds \right\},$$

where OU_s is an Ornstein–Uhlenbeck (OU) process. The OU process is the continuous version of a first order autoregressive Gaussian time series, with parameters corresponding to (σ_S) standard deviation at time 0, (σ_L) the standard deviation for large t, and (α) the speed of mean reversion. In that case, the covariance structure is

$$\mathrm{Cov}(OU_s, OU_t) = \sigma_S^2 e^{-\alpha(s+t)} + \sigma_L^2 e^{-\alpha(s+t)}(e^{2\alpha \min\{s,t\}} - 1).$$

7.4.4 This is an infinite factor model because the entire path of a process is required just to model the changes in yield curves from one period to the next. The construction means that even if we knew 100 points on the yield curve, some uncertainty still remains about the remaining points. This gives the model a richness that is useful for understanding risks. For example, under finite factor models it is possible to find a set of assets that appear to replicate liabilities exactly. Under the IOU model this can only be achieved if all asset and liability cash flows are perfectly matched. Otherwise, the model always captures some degree of mismatch risk.

7.4.5 Infinite factor models (also called random field or stochastic string models) are commonly used to describe the evolution of the entire surface of forward rates parameterized by time and time-to-maturity. The IOU model captures the uncertainty generated by such models over a fixed time-horizon only. Evolution of the IOU model in the time direction is of limited relevance for stress-testing purposes. Additional details and discussion are provided by Santa-Clara and Sornette (2001) for example.

7.4.6 The IOU model describes zero coupon bond prices. Observed interest rate volatilities and correlations are in terms of par yields. An application of Ito's formula expresses par yield volatilities in terms of the volatilities of zero coupon bonds, given a starting yield curve which we take as being flat at 5%. We can fit three parameters, and in this paper we have chosen these to capture the observed historic volatilities of 2 and 10 year par yields, as well as the correlation between them. Possible fitted parameters are as follows:

Parameter	Fitted to GBP swap yields	Fitted to EUR swap yields
σ_S	0.75%	0.65%
σ_L	0.91%	0.75%
α	18.0%	12.0%

7.4.7 We have fitted to two points on the volatility curve, and to a correlation between them. Figure 36 compares observed volatilities by term to the fitted model. The correpsondence is close, and exact for the 2 and 10 year volatilities we have chosen to fit. However, there are systematic deviations, including an over-fitting of volatilities at the short end of the curve and under-fitting at the long end. Both of these suggest the model may be improved, and calibration complicated, by the inclusion of further parameters.

7.5 *Variance Matching*

7.5.1 We now consider how to translate a model of yield curve moves into a series of stress tests for capital calculations. Given that stress tests are

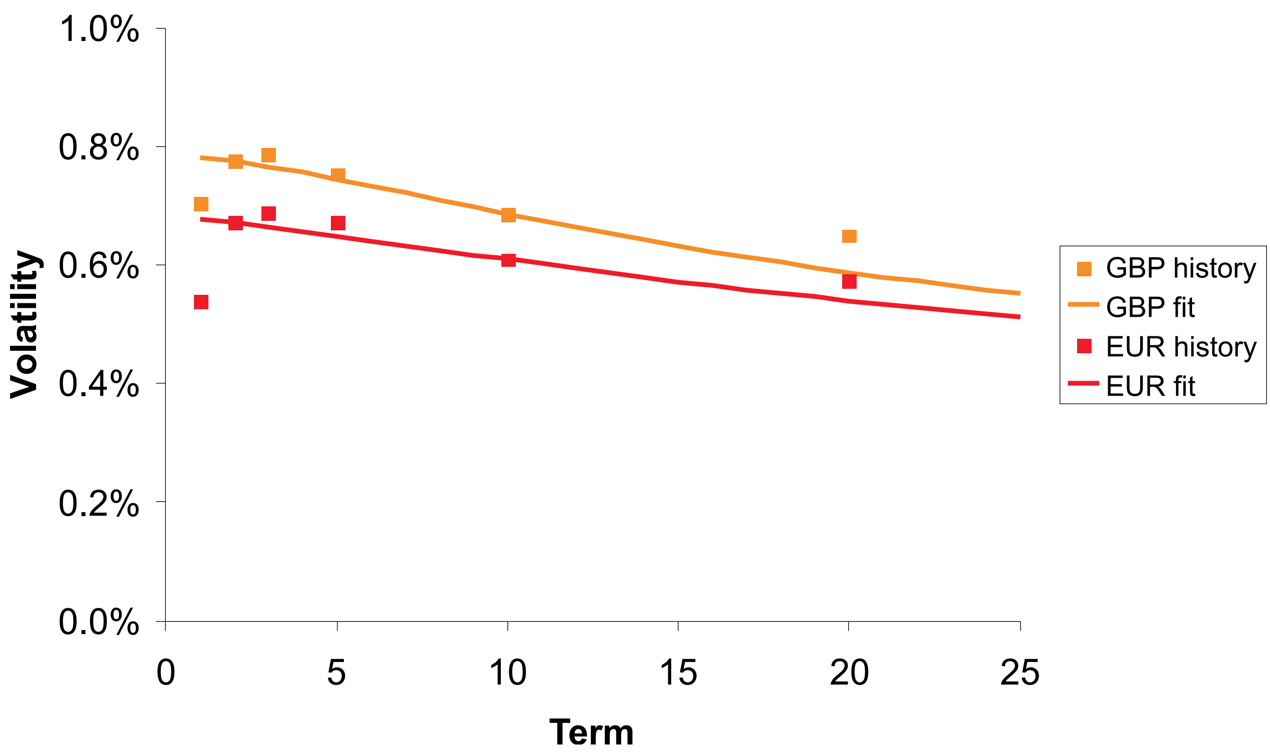

Figure 36

finite in nature, this necessarily involves condensing historic data, or a fitted model, into a finite number of factors.

7.5.2 As part of the development of Solvency II, European regulators have published a series of quantitative impact studies. Their yield curve stresses are calibrated using a method which we will call *variance matching*. They construct a one factor model with the same variance as the full model.

7.5.3 Let us write the general multiplicative model for bond prices in the form:

$$P_t^{stress} = P_t^{base} * H_t$$
$$\mathrm{Var}(H_t) = \sigma_t^2.$$

7.5.4 We previously took the example of a two-factor model:

$$P_t^{stress} = P_t^{base} * \left\{1 + Z_1\zeta_1(t) + Z_2\zeta_2(t)\right\}.$$

The variance-matched one-factor version is then obtained with a root-sum-of-squares formula, representing the standard deviation of a sum of uncorrelated variables.

$$\begin{aligned} P_t^{stress} &= P_t^{base} * \{1 + Z_4\sigma_t\} \\ &= P_t^{base} * \left\{1 + Z_4\sqrt{\zeta_1^2(t) + \zeta_2^2(t)}\right\}. \end{aligned}$$

7.5.5 Under the variance matching method, a pair of stress tests (up and down) is based on historic variability of yields by each term. In the context of the IOU model, based on normal distributions (for non-normal distributions, replace 2.58 with the appropriat percentile) the corresponding 1-in-200 year stresses are:

$$P_t^{stress} = P_t^{base} * \{1 \pm 2.58\sigma_t\}.$$

7.5.6 This correctly calculates the required capital when there is one fixed future cash flow. In other cases, the use of up and down stress tests implicitly assumes that yields of all terms are 100% correlated with each other. Empirically, we know this is false, but there is no alternative within the constraints of a pair of opposite stress tests.

7.5.7 The effect of overstated correlations depends on the business cash flows. Suppose first that a firm has a net long position in cash flows of all terms, so is exposed to a rise in interest rates of all terms. In that case, the variance matched stress test overstates the risk, in failing to give credit for diversification between yields of different terms. A multifactor model (with the same volatilities) would give a lower capital requirement. The same applies if the firm has net short cash flow position at all future dates.

7.5.8 On the other hand, consider a firm who hedges a liability outgo at time 11 with a 10-year bond. In that case, the firm is exposed to a rise in 10 year yields and a fall in 11 year yields. A correlation assumption of +100% excludes this most painful combination, and so understates the risk. A multi-factor model would give a higher capital requirement than the statutory stress test. We cannot therefore simply suppose that the chosen single stress test is cautious.

7.5.9 We note in passing that the stresses may give rise to negative bond prices, if t is too large. A possible pragmatic fix for these cases is to set the bond price to zero, or to a small positive value.

7.6 *Spot and Forward Matching*

7.6.1 We noticed a particular concern with the variance matching method, when asset and liability cash flows were of similar magnitude but opposite sign. The variance match method can be improved to cover this situation, by equating not only the variance of ZCB curve but its slope, as well as the covariance between level and slope. To capture this, we need, at least, a two-factor model.

7.6.2 The general two factor model takes the form:

$$\begin{aligned} P_t^{stress} &= P_t^{base} * H_t \\ &= P_t^{base} * \left\{1 + Z_1\zeta_1(t) + Z_2\zeta_2(t)\right\}. \end{aligned}$$

7.6.3 To assist calibration of a spot and forward match, we rearrange this into polar coordinates.

$$\sigma_t = \sqrt{\zeta_1^2(t) + \zeta_2^2(t)}$$
$$\begin{pmatrix} \zeta_1(t) \\ \zeta_2(t) \end{pmatrix} = \begin{pmatrix} \sigma_t \cos(\psi_t) \\ \sigma_t \sin(\psi_t) \end{pmatrix}.$$

7.6.4 Stress tests are constructed in the usual way, by moving each of Z_1 and Z_2 to their 0.5 percentile and 99.5 percentile levels. In some situations, we can interpret this as a classical level and slope decomposition. A stress to the level of the yield curve changes all bond prices except P_0. Our first factor can capture such a stress if ψ_t lies in the limited range $-\pi/2 < \psi_t < \pi/2$, in which case $\cos(\psi_t) > 0$. We might then interpret the second factor as a change in slope, particularly if ψ_t is an increasing function of t, and there is some pivotal t_{piv} for which $\psi(t_{piv}) = 0$. In that case, t_{piv} represents the term which is unaffected by the second stress test. If $Z_2 > 0$, then bonds shorter than t_{piv} become more expensive, while longer bonds become cheaper. This corresponds to an increase in yield curve slope with the yield fixed at t_{piv}. Conversely $Z_2 < 0$ corresponds to a decrease in yield curve slope. Whether we can achieve this interpretation, however, depends on the range of ψ_t. If ψ_t covers too wide a range (exceeding π from the highest to lowest values) then both factors will vary sinusoidally with time.

7.6.5 We use the Box–Muller algorithm to express Z_1 and Z_2 as functions of an exponentially distributed random variable, R, with mean 1, and a uniformly distributed random variable θ on $(-\pi, \pi)$.

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \sqrt{2R}\cos(\theta) \\ \sqrt{2R}\sin(\theta) \end{pmatrix}.$$

7.6.6 We can finally rewrite H_t in a form that is easier to calibrate. The new form also makes clear that only relative values of ψ_t affect the model's behaviour. The addition of a common constant to all values of ψ_t does not affect the distribution of H_t, although it may affect how we decompose a given stress into its two components.

$$\begin{aligned} H_t &= 1 + Z_1\zeta_1(t) + Z_2\zeta_2(t) \\ &= 1 + \sigma_t\sqrt{2R}\cos(\psi_t)\cos(\theta) + \sigma_t\sqrt{2R}\sin(\psi_t)\sin(\theta) \\ &= 1 + \sigma_t\sqrt{2R}\cos(\psi_t - \theta). \end{aligned}$$

7.6.7 Suppose now we want to fit this model as a simplified version of a more complex model. In addition to capturing the variance of H, we wish to capture the variance of H and its slope. That means we need to capture not only $\mathrm{Var}(H_t)$ and $\mathrm{Var}(H_{t-1})$ but also $\mathrm{Var}(H_t - H_{t-1})$. Equivalently, we need to capture $\mathrm{Correl}(H_{t-1}, H_t)$.

7.6.8 A simple calculation under the two-factor model shows that $\mathrm{Correl}(H_{t-1}, H_t) = \cos(\psi_t - \psi_{t-1})$. Therefore, we can calibrate the function ψ_t inductively by:

$$\psi_t = \psi_{t-1} + \cos^{-1}\mathrm{Correl}(H_{t-1}, H_t).$$

7.6.9 Values of σ_t and ψ_t are shown for the U.K. calibration of the IOU model in the chart below, together with the implied stress tests to be applied as multiplicative factors to zero coupon bond prices:

Time horizon t	σ_t	ψ_t	stress 1 $= 2.58 * \sigma_t \sin(\psi_t)$	stress 2 $= 2.58 * \sigma_t \cos(\psi_t)$
0	0.00%	0.00	0.00%	0.00%
1	0.75%	0.00	0.00%	1.93%
2	1.48%	0.28	1.04%	3.68%
3	2.20%	0.49	2.70%	5.00%
4	2.90%	0.68	4.70%	5.82%
5	3.58%	0.84	6.88%	6.15%
6	4.23%	0.99	9.10%	6.01%
7	4.86%	1.12	11.27%	5.46%
8	5.46%	1.24	13.34%	4.55%
9	6.04%	1.36	15.24%	3.33%
10	6.61%	1.46	16.94%	1.84%
11	7.15%	1.56	18.44%	0.13%
12	7.67%	1.66	19.70%	−1.76%
13	8.17%	1.75	20.73%	−3.79%
14	8.65%	1.84	21.53%	−5.92%
15	9.12%	1.92	22.08%	−8.13%
16	9.57%	2.00	22.41%	−10.39%
17	10.01%	2.08	22.52%	−12.66%
18	10.44%	2.16	22.40%	−14.94%
19	10.85%	2.23	22.09%	−17.20%
20	11.25%	2.30	21.57%	−19.42%
21	11.64%	2.37	20.87%	−21.58%
22	12.02%	2.44	20.00%	−23.69%
23	12.38%	2.51	18.96%	−25.72%
24	12.74%	2.57	17.77%	−27.66%
25	13.09%	2.63	16.44%	−29.51%
26	13.44%	2.69	14.98%	−31.26%
27	13.77%	2.75	13.40%	−32.91%
28	14.10%	2.81	11.71%	−34.44%
29	14.42%	2.87	9.92%	−35.85%
30	14.73%	2.93	8.05%	−37.15%
31	15.04%	2.98	6.09%	−38.32%
32	15.34%	3.04	4.07%	−39.37%
33	15.64%	3.09	1.98%	−40.29%
34	15.93%	3.15	−0.15%	−41.09%
35	16.21%	3.20	−2.33%	−41.76%
36	16.49%	3.25	−4.54%	−42.31%
37	16.77%	3.30	−6.78%	−42.73%
38	17.04%	3.35	−9.04%	−43.03%
39	17.31%	3.40	−11.31%	−43.20%
40	17.57%	3.45	−13.59%	−43.25%

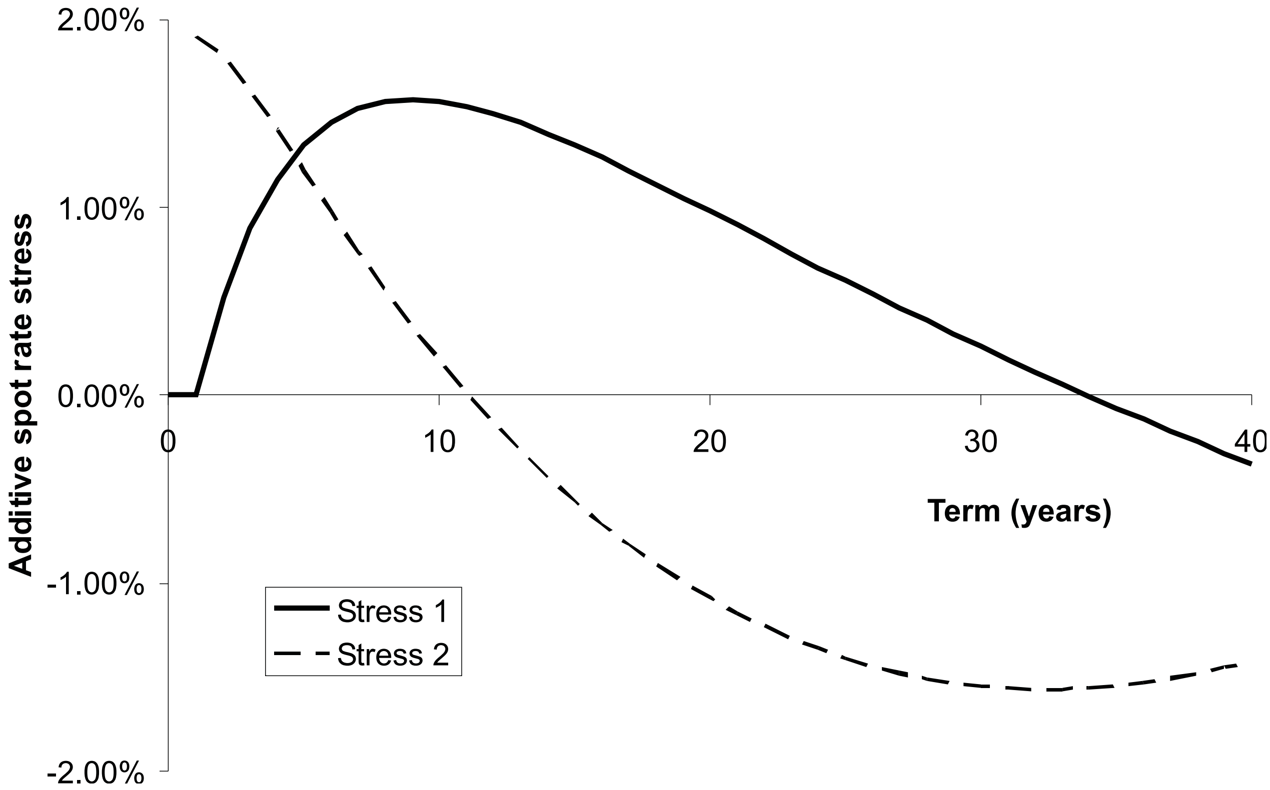

Figure 37

7.6.10 We can express these as stresses to continuously compounded spot rates. These are shown in Figure 37.

7.7 *Addition of Further Factors*

7.7.1 We have taken care to fit an infinite dimensional model, namely the IOU model. We have then distilled this into a two-factor model for the purpose of constructing stress tests. In many ways this is unsatisfactory, as by reducing the model to two factors we have thrown away infinitely many remaining factors.

7.7.2 The number of factors to be considered depends on the extent of existing hedging attempts. Where close interest rate hedging is practiced, it is unlikely that simple factors of the form we have calculated can reveal subtle mismatches. For this purpose, higher factor models are required. The variance matching algorithm extends to produce as many factors as are required.

7.7.3 On the other hand, for firms taking a strong directional position on interest rate moves, for example with liaiblity duration far exceeding asset duration, a low factor model may be adequate to capture the main sources of risk.

8. Conclusions

8.1 *General Comments*

8.1.1 We have chosen to study the 0.5 percentile outcome over one year, in line with current and emerging regulatory practices, primarily focussing on equity returns.

8.1.2 At this time horizon and confidence level prior beliefs play a critical role in determining the outcome of any estimate, since there is not sufficient (relevant) historic data for a pure frequentist approach. Section 5 shows the large estimation error involved when looking at a hundred years of equity data — a much longer series than we have for most asset classes. Such prior beliefs can be obvious, such as the choice of distribution to fit to data, or more subtle, such as the exclusion of a data point as an outlier.

8.1.3 Of course it is often possible to construct more data points by looking at monthly, weekly or even daily data. However the way in which the distribution of annual returns is constructed from the monthly, weekly or daily distribution is itself an assumption founded largely on prior beliefs. In the case of equities, our analysis of skewness and kurtosis in Section 4.13 does not support the hypothesis that monthly equity returns are independent — the most obvious approach to annualisation.

8.1.4 Results can also be highly dependent on the choice of data used. The decision over what data is and is not relevant is itself a form of prior belief.

8.2 *Equity Returns*

8.2.1 It is a common belief (especially for high frequency data such as daily) that equity returns show a negative skew (fatter left tail than right) and leptokurtosis (fat tails). Our own analysis of annual log-returns from two data sets shows a scatter of positive and negative skews for different countries. Virtually all countries show positive kurtosis, with the U.K. having by far the greatest due to the influence of 1974 and 1975. However when we try to calculate 95% confidence intervals for our estimates, very few countries show either significant skew or kurtosis (including the U.K.) for annual returns. This is because high skew and kurtosis is often driven by one or two data points and so the sampling error involved is very large. If 1974 and 1975 are excluded from the U.K. data set then the skew and kurtosis become lower than those of most other countries.

8.2.2 The Dimson, Marsh and Staunton data set gives us over 100 years of data, but even if we include all the data as relevant we are still faced with large estimation error for the 0.5 percentile event for annual returns — not surprising as this represents a 1-in-200 year event. We have estimated a 95% confidence interval of 30-40% using bootstrapping.[12]

[12] Just to be clear, if the confidence interval is 30% and our estimate is a 0.5 percentile return of −40%, the range is −25% to −55%.

8.2.3 Taking a parametric approach, we can fit a distribution to our data (a strong example of a prior belief of course) and use that to estimate the 0.5 percentile return. A distribution that fitted the data well and showed some stability in parameters from one country to another and one time period to another would then be a plausible choice. If the 0.5 percentile event was similar between distributions that would be even more encouraging.

8.2.4 We have tried a range of distributions fitted to a variety of countries. Unfortunately the results appear to be unstable by distribution and country with estimates of the 0.5 percentile return varying from −30% to −80%. A simple 2-parameter distribution like the log-normal does not capture the positive kurtosis that most data sets show. However when we fit more complex 3 or 4-parameter distributions the results are very sensitive to the data and we get unstable results from one country to another. More parameters allow us to capture better the distribution shape, but at the risk of overfitting to the data. Once again a few data points can have a large influence on the results — again the U.K. is an outlier in international comparisons because of the events of 1974-5.

8.3 *Interest Rates*

8.3.1 With equity returns we can compress the returns on individual stocks into a single index. In contrast interest rates have a term structure which adds significant extra complexity.

8.3.2 For practical purposes the richness of the full yield curve analysis is often compressed into a small number of factors. A common approach is principal components analysis, and an alternative approach is described in Section 7. However care is needed to ensure that the number of factors, and so the richness of yield curves tested, is suited to the asset/liability position in question. A position where assets and liabilities have been matched against parallel shifts in yields curves may show little capital requirement under a one factor model, but a larger requirement under a two factor model that includes the possibility of a change in slope. More generally once we have immunised against n factors it is the $n+1$th and higher factors that we need to consider.

8.4 *Other Asset Classes and Confidence Levels*

Having concentrated in this paper on equities (a relatively simple, well studied and data rich asset class) it is sobering to reflect that in practice actuaries and other finance professionals have to estimate similar extreme events for more complex asset classes with much less relevant data. In addition, for internal purposes many financial institutions calculate capital at much more extreme percentiles (say the 0.05 percentile instead of the 0.5 percentile). The most we can say at present is that in such cases we should be clear about the role of prior beliefs in forming these estimates, and clear about the likely levels of uncertainty (estimation error) that will be involved.

Acknowledgements

The authors would like to thank the following people for their helpful comments and input during the production of this paper: John Hibbert, Nick Jessop, Craig Turnbull, Russell Gerrard, Amy Williams, Jessica Queree, Andreas Tsanakas and participants at the 2008 Finance and Investment Conference.

References

Abramowitz, M. & Stegun, I.A., eds. (1972). *Handbook of mathematical functions with formulas, figures, and mathematical tables*, New York: Dover Publications, ISBN 978-0-486-61272-0.

Anderson, T.W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, **52**(278), 200-203.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London, Series A*, **353**(1674), 401-419.

Barndorff-Nielsen, O. & Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Probability Theory and Related Fields*, **38**(4), 309-311.

Bouchaud, J.-P. & Potters, M. (2003). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge University Press, ISBN 0521819164.

Cochrane, J.H. (1999). New facts in finance. CRSP Working Paper No. 490.

Cornish, E.A. & Fisher, R.A. (1937). Moments and cumulants in the specification of distributions. *Revue de l'Institute International de Statistique*, **5**, 307-320.

Christoffersen P., Diebold F. & Schuermann T. (1998). Horizon problems and extreme events in financial risk management, Federal Reserve Bank of New York Economic Policy Review.

Christoffersen F. & Diebold. F. (2000). How relevant is volatility forecasting for financial risk management? *The Review of Economics and Statistics*, **82**(1), 12-22.

Dowd, K. (2002). A bootstrap backtest. *Risk*, **25**(10), 93-94.

Eberlein, E. & von Hammerstein, E.A. (2002). Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes, Working Paper No. 80, University of Freiburg.

Eberlein, E. & Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, **1**(3), 281-299.

Embrechts, P., Mikosch, T. & Kuppelberg, C. (2008). *Modelling extremal events for insurance and finance*. Series: Stochastic Modelling and Applied Probability, Vol. 33, Springer Verlag, ISBN: 978-3-540-60931-5.

Exley, J., Meyta, S. & Smith, A.D. (2004). *Mean Reversion*, presented to Finance and Investment Conference 2004.

Finkelstein, G.S. (1997). Maturity guarantees revisited: allowing for extreme stochastic fluctuations using stable distributions. *British Actuarial Journal*, **3**(2), 411-482.

Hairs, C.J., Belsham, D.J., Bryson, N.M., George, C.M., Hare, D.J.P., Smith, D.A. & Thompson, S. (2002). Fair valuation of liabilities. *British Actuarial Journal*, **8**, 203-299.

Heinrich, J. (2004). A guide to the Pearson type IV distribution. Note 6820, Collider Detector at Fermilab, Fermilab, Batavia, Illinois.

Kemp, Malcolm (2008). *Catering for the fat-tailed behaviour of investment returns*", Threadneedle working paper.

KENDALL, M.G. & STUART, A. (1979). *The advanced theory of statistics*. Charles Griffin & Company, London, 4th edition.

LEVY, P. (1953). Random functions: general theory with special reference to Laplacian random functions. *Univ. California Publ. Statist.*, **1**, 331-390.

MANDELBROT, B. (1963). The variation of certain speculative prices. *The Journal of Business*, **36**, 249-219.

REBONATO, R. (2007). *Plight of the fortune tellers: why we need to manage financial risk differently*. Princeton University Press, ISBN 0691133611.

ROSS, S.M. (1997). *Simulation*. Academic Press, 2nd edition, pp121-127.

SANTA-CLARA, P. & SORNETTE, D. (2001). The dynamics of the forward interest rate curve with stochastic string shocks. *The Review of Financial Studies*, **14**, 149-185.

SCHERER, B. (2007). *Portfolio construction and risk budgeting* (3rd edition).

SHIRYAEV, A.N. (1999). *Essentials of stochastic finance: models, theory*. World Scientific, River Edge, NJ.

WALTER, C. (1989). *Mise en evidence de distributions Levy-stables et d'une structure fractable sur le marche de Paris.* Proceedings of the First AFIR Colloquium,
http://www.actuaries.org.uk/AFIR/Colloquia/Paris/Walter.pdf

WILKIE, A.D. (1986). A stochastic investment model for actuarial use. *Transactions of the Faculty of Actuaries*, **39**, 341-381.

WILKIE, A.D. (1995). More on a stochastic investment model for actuarial use. *British Actuarial Journal*, **1**(5), 777-964.

PCA References

ANDERSON, T.W. (1957). *Multivariate analysis*. Wiley. 3rd edition 2003.

GIBSON, M.S. & PRITSKER, M. (2000). *Improving grid-based methods for estimating value at risk of fixed-income portfolios*.
http://www.federalreserve.gov/Pubs/feds/2000/200025/200025pap.pdf

JAMSHIDIAN, F. & ZHU, Y. (1997). Scenario simulation model: theory and methodology. *Finance and Stochastics*, **1**(1).

LITTERMAN, R. & SCHEINKMAN, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income*, **1**.

LORETAN., M. (1997). *Generating market risk scenarios using principal components analysis: methodological and practical considerations*. http://www.bis.org/publ/ecsc07c.pdf

MATZNER-LØBER, E. & VILLA, C. (2004). *Functional principal component analysis of the yield curve*. http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf

THOMAS, M. & Maré, M. (2007). *Long term forecasting and hedging of the South African yield curve*. http://www.actuarialsociety.org.za/scripts/file_build.asp?id=100000794

TOLMASKY, C.F. (2007). *Principal components analysis in yield-curve modelling (slides)*.
http://www.math.umn.edu/finmath/seminar/Materials/tolmasky1.pdf

Useful Journals

Journal of Risk, Vol. 9, No. 2 (winter 2006/07) — Backtesting Techniques.

Useful Links

http://en.wikipedia.org/wiki/Student_t_distribution
http://en.wikipedia.org/wiki/Hyperbolic_distribution
http://www-cdf.fnal.gov/publications/cdf6820_pearson4.pdf (Heinrich (2004))

APPENDIX A

SELECTED SOURCES OF MARKET DATA

A.1 *Introduction*

A.1.1 There are a large number of commercial and free data sources for equity and interest rate data. The quantity and quality of data have increased significantly in recent years and most sources provide a wide range of indices and statistics. However many indices have a relatively short history and it is less easy to find long-term data series. Constructing such a series is a non-trivial task often requiring multiple sources of data.

A.1.2 In this appendix we give some background to one database of long-term returns, the Dimson, Marsh & Staunton dataset, that we have used in sections 4 and 5 of this paper. We also give some brief details of a selection of other commonly-used data sources, although it is important to note that we have made no effort to give a comprehensive list and we do not make any recommendations. Finally, although not a data source, we give some notes on the MSCI indices used in section 4.

A.2 *Dimson, Marsh & Staunton*

A.2.1 Elroy Dimson, Paul Marsh and Mike Staunton (DMS) of the London Business School have compiled a global database containing annual returns on stocks, bonds, bills, inflation and currencies for 17 countries from 1900-2005. DMS comment that, although an earlier start date would be desirable, the poor quality of older data makes the returns calculated before 1900 unreliable. Before this date there are few stock indices so individual stock data must be used. The problems they identify include lack of coverage, omission of dividends, survival bias and a tendency to omit periods of market stress.

A.2.2 The countries included are: Australia, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Japan, The Netherlands, Norway, South Africa, Spain, Sweden, Switzerland, United Kingdom and United States. The U.K. equity return series was constructed from the London Business School's share price database from 1955 onwards and by collecting share prices from old issues of the Financial Times from 1899 to 1954. The database also includes four world indices based on the countries included in the DMS dataset: a World equity index, a World excluding U.S. index, a World bond index and a World ex-U.S. bond index.

A.2.3 The DMS dataset was published and analysed in Triumph of the Optimists: 101 Years of Global Investment Returns. It has been updated annually and published in the ABN AMRO Global Investment Returns Yearbook. The dataset is also distributed commercially by Morningstar. Website: www.morningstar.com

A.2.4 References.

A.2.5 E. Dimson, P. Marsh, and M. Staunton, Triumph of the Optimists: 101 Years of Global Investment Returns. New Jersey: Princeton University Press (2002).

A.2.6 E. Dimson, P. Marsh, and M. Staunton, The Worldwide Equity Premium: A Smaller Puzzle (2006).

A.3 *Bank of England*

A.3.1 The Bank of England is a useful source for U.K. yield curve data. It publishes its own estimates of U.K. government yield data in the statistics section of its website. Government nominal data is available since 1979 daily and since 1970 monthly, while real and inflation data are available since 1985 daily. The data includes instantaneous implied forward rates and implied spot curve.

A.3.2 This information is free but the Bank does not take responsibility for the accuracy of the information.

A.3.3 Website: www.bankofengland.co.uk

A.4 *European Central Bank*

A.4.1 The ECB Statistical Data Warehouse provides free Government Bond yield data for the European area as well as the United States and Japan.

A.4.2 Website: http://sdw.ecb.europa.eu/

A.5 *British Bankers' Association*

A.5.1 The BBA publishes historical LIBOR rates for several currencies starting from January 1986 and repo rates from May 1999.

A.5.2 Website: www.bba.org.uk

A.6 *Euribor*

A.6.1 Euribor is an "international non-profit association" backed by the European Banking Federation (EBF) that calculates the Euro Interbank Offered Rate (Euribor). Historical data from its launch in December 1998 are available from its website.

A.6.2 Website: www.euribor.org

A.7 *David Wilkie — Andrew Cairns Database*

A.7.1 David Wilkie and Andrew Cairns have created the Herriot-Watt/Faculty and Institute of Actuaries Gilt Database. This database is available for free and is limited to U.K. gilts, covering U.K. gilts indices, gilts prices and other technical data. Yields and yield indices are available monthly since November 1998.

A.7.2 There are a number of commercial data providers offering historical equity and/or fixed income data. A selection of the major providers is listed below.

A.8 *Global Financial Data*

A.8.1 Global Financial Data, as the name suggests, is a database providing financial information for global markets. It contains some long-term series for both U.K. gilt yields and equity returns. Some examples are given below.

A.8.2 Government Bond Yields

— United Kingdom 20-year Government Bond Yield (monthly from 1933 to 1985, daily from 1986).
— United Kingdom 30-year Government Bond Yield (weekly for 2003 and 2004, daily from 2004).
— United Kingdom 2 1/2% Consol Yield (Monthly from 1800 to 1980, weekly from 1880 to 1915, monthly from 1915).

A.8.3 Equity returns

— FTSE 100 Total Return Index (available daily from June 1994).
— U.K. FTSE All-Share Return Index (extended by Global Financial Data and available monthly from 1694 to 1964 and daily from 1964).

A.8.4 Older values of the FTSE All-Share index have been compiled from sources such as the Banker's Magazine, the Investor's Chronicle, the London and Cambridge Economic Service and the Financial Times-Actuaries Indices.

A.8.5 Website: www.globalfinancialdata.com

A.9 *FTSE*

A.9.1 FTSE calculates and publishes the well-known FTSE indices, including the FTSE U.K. Index Series for equities and the FTSE Actuaries U.K. Gilts Index Series. The FTSE All-Share has been calculated since July 1962 and the FTSE 100 since January 1984.

A.9.2 Website: www.ftse.com

A.10 *Bloomberg*

A.10.1 Bloomberg is a large database containing information on the majority of existing indices as well as its own calculations and indices.

A.10.2 For U.K. interest rates, the data series are generally high frequency but relatively short. For example, FTSE Actuaries U.K. Gilts Yield 5/10/15 Years indices are available daily since March 1998. However "OECD United Kingdom Interest Rates 10 Year Government bond" is available monthly from 1960.

A.10.3 For U.K. equity returns the FTSE 100/250/350 and All Share indices are available daily from 1985.

A.10.4 Website: www.bloomberg.com

A.11 *Thomson DataStream*

A.11.1 DataStream is a database of company, financial and economic data. The data is updated daily with some series dating back as far as 1970.

A.11.2 Website: www.datastream.com

A.12 *Reuters*

A.12.1 Thomson Reuters was recently purchased by The Thomson Corporation.

A.12.2 Reuters does not offer real time data but sells datasets of historic data. The fixed income data includes 54 global government yield curves with histories dating back to 1993. The equity data provides a minimum of 20 years' pricing for G7 countries.

A.12.3 Website: www.reuters.com

A.13 *MSCI Indices*

Morgan Stanley Capital International (MSCI) indices are published starting from 31/12/1969. Series since then are available for 18 countries: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Italy, Japan, Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, U.K. and U.S.A., although indices are now available for a much wider range of countries. They are available both in local currency terms and U.S. Dollars.

APPENDIX B

THE STRESS TEST PLUS CORRELATIONS METHOD FOR MODELLING ECONOMIC CAPITAL

B1 *The Approach Step by Step*

B1.1 A typical procedure used by firms to calculate economic capital for ICA requirements is the "stress test plus correlation" method. When estimating extreme market percentiles we need to bear in mind how they will be used, so it is useful to look at the steps involved in this method.

B.1.2 Step 1. Identify the standard deviation of each of a small number (<30) of risk drivers, and define stress tests to be ± 2.58 standard deviations from the current value. These stress tests are called "up" and "down" stresses, respectively. Each driver is stressed individually, with the other drivers retaining their current values. These are called "one way stresses", as opposed to "combined stresses" under which two or more drivers move simultaneously from their current values.

B.1.3 Step 2. For each risk driver, identify which of {current value, up stress, down stress} produces the lowest net assets. Define the signed capital as follows.

— If current value is the worst then
— signed capital $=$ 0.
— If up stress is the worst then
— signed capital $=$ up stress net assets $-$ current net assets.
— If down stress is the worst then
— signed capital $=$ current net assets $-$ down stress.

B.1.4 Step 3. Add up the sum of the squares of the signed capital for each driver. Then for each pair of drivers, add to this twice the product of the signed capital amounts, multiplied by an assumed correlation. Finally, take the square root of the total. This is the aggregate capital.

B.1.5 Step 4. The estimated 0.5 percentile of net assets is the starting value minus the aggregate capital.

B.2 *Possible Assumptions Underlying Stress Tests plus Correlations*

B.2.1 This procedure can be justified under the following assumptions

— Assumption 1. The drivers are small in number.
— Assumption 2. Each driver has a normal distribution with mean equal to its current value.
— Assumption 3. The net assets are a monotone (increasing or decreasing) function of each driver.
— Assumption 4. The drivers do not interact — for example, the profit or loss arising from a fall in interest rates is unaffected by the level of equity markets.

— Assumption 5. The profits or losses from each driver follow a jointly elliptically contoured distribution.

B.2.2 Where these assumptions hold, the estimation of stress tests reduces to the estimation of the standard deviation of each driver.

B.2.3 If the assumptions do not hold, we may turn to more generic approaches, such as Monte Carlo simulation. Monte Carlo is a generally applicable method, but is considerably more onerous than stress tests and correlations, both from a coding perspective and also in terms of computer run time.

B.3 *Approximations and Adjustments*

B.3.1 Instead, where the assumptions are approximately true, it may be possible to apply the stress test and correlation method, but with some adjustment to reflect ways in which the assumptions are not borne out. Examples of possible adjustments are as follows. These are numbered in order to correspond with the assumptions to which they relate.

B.3.2 Adjustment 1. If the number of factors is large in number, dimension reduction techniques may be applied. For example, rather than modelling each equity holding and each corporate bond individually, these may be assumed to follow a particular index. Where yield curves may in principle change shape in a variety of ways, it may be assumed that yields of all terms move in parallel. More advanced techniques such as principle components analysis or variance matching, may be brought to bear.

B.3.3 Adjustment 2. To reflect distributions with fatter tails than normal, use a stress test larger than 2.58 standard deviations — for example, 2.8 standard distributions instead.

B.3.4 Adjustment 3. To allow for non monotonicity, take the worst net assets for any driver value between the up and down stresses. In the case of monotone functions, this minimum will correspond to either the up or down stress, but the formulation in terms of worst case also captures the situation where the most painful situation does not correspond to the most extreme driver values. This situation usually arises when the original stress tests are used as input to the construction of a hedging programme.

B.3.5 Adjustment 4. To capture interactions, consider combined events where more than one stress test occurs simultaneously. These tests need to allow for diversification, so typically the combined stress multiplies each factor stress by a shrinkage factor si where the index i runs through the risks modelled. To ensure this is indeed a shrinkage, we insist $|si| \leq 1$, with positive shrinkage implying up stresses and negative values down stresses. For example, the medium bang approach considers shrinkage where the si all take the same absolute value. Another approach defines a shrinkage vector s as likely if $sT\rho - 1s \leq 1$, where ρ is the correlation matrix. Keen readers may enjoy demonstrating that this implies $|si| \leq 1$. Capital can then be set according to the scenario that minimises net assets subject to this constraint. This approach combines the effects of Adjustment 3 and Adjustment 4.

B.3.6 Adjustment 5. It is sometimes suggested that correlations always serve to increase risk in extreme events, even if there are offsets in normal conditions. To reflect this, any negative contributions may be excluded from the sum in Step 3. A more extreme adjustment is to replace any negative contributions with a positive contribution of the same magnitude. This is equivalent to using absolute capital values in place of signed capital values, as is required for the standard formula under current drafts of Solvency II. In addition, or instead, firms may choose prudent correlation assumptions in order to reflect more extreme correlations which may apply in stressed situations. More sophisticated approaches involve copulas to reflect non-ellipticity.

APPENDIX C

COMPONENTS OF YIELD CURVE MOVEMENTS — ILLUSTRATIVE WORKED EXAMPLES

Introduction

C.1 *Variables Modelled*

We consider the movement in the yield curve from now until a point in one year's time. We denote this movement by X, a random vector. For the purposes of this section, it does not matter whether these are spot yields, par yields or forward yields.

We use a deliberately unsophisticated model, so we can easily interpret outputs and so readers can readily replicate our results. Suppose X has zero mean, so the expected yield curve in a year's time is exactly where it is now. We suppose X is modelled at certain *key maturities*, these being maturities: $\{1, 2, 3, 5, 10, 20\}$.

C.2 *Components*

We seek to express X in the form:

$$X = BZ.$$

Here, Z is a vector of independent random variables; for this section we use normal variables with mean zero and unit variance. The matrix B holds the components, with each column of B corresponding to one component. The element B_{ij} is the value of the j^{th} component evaluated at the i^{th} key maturity.

From standard matrix theory (see, for example, the book by Anderson (1957)), the variance-covariance matrix of X is BB^{T}, where a superscript T denotes a matrix transpose. We therefore want the decomposition to reproduce the variance-covariance matrix of X. This means that

$$BB^{\mathrm{T}} = V.$$

We can choose B with fewer columns than the number of elements of X, i.e. we have fewer components than points on the yield curve. In this case we are unlikely to match V exactly but instead seek B such that BB^{T} is as close as possible to V.

C.3 *Why Reduce the Number of Components?*

When all components are considered, any yield curve decomposition explains 100% of the variance. If some components are excluded, then principal components analysis converges best, in terms of average variance explained.

Why, then might we want to truncate the decomposition, and exclude some components? The most important reason arises in the context of Value-at-Risk calculations.

Let us suppose the net assets of a firm are a smooth function of the yield curve shift X. Then, for small X, the net assets are approximately of the form $a_0 + g^{\mathrm{T}} X$, where a_0 is the starting net assets before the yield shift and g is the vector gradient. With a variance-covariance matrix V, the variance of net assets is then $g^{\mathrm{T}} Vg$. Assuming multivariate normal distributions and a specified confidence level α (for example $\alpha = 0.995$), the value at risk is $\sqrt{g^{\mathrm{T}} Vg}\Phi^{-1}(\alpha)$.

To evaluate this expression, we need to compute the gradient g. This is commonly estimated using central finite difference methods. With our model, based on 6 points of the yield curve, the estimation of g requires 12 net asset calculations, also known as "stress tests".

The calculation of stress tests may be an easy task, for example if all future cash flows are fixed. But financial firms' cash flows are typically variable, depending not only on market conditions but also the actions of customers and management. In this case, net asset calculation may be an onerous task. A request to recalculate 12 stress tests could have important operational implications. The burden gets worse if more than 6 points on the curve are modelled.

A constraint on the number of components can reduce the effort required in value-at-risk calculations. As an intermediate step, we need to compute the variance $g^{\mathrm{T}} Vg$.

Now suppose we can approximate $V \approx BB^{\mathrm{T}}$. Then we can approximate the variance $g^{\mathrm{T}} Vg \approx g^{\mathrm{T}} BB^{\mathrm{T}} g = (B^{\mathrm{T}} g)^{\mathrm{T}} (B^{\mathrm{T}} g)$. A saving arises because we can compute $B^{\mathrm{T}} g$ with fewer stress tests than to estimate g. The number of stress tests required is twice the number of components.

This is how to estimate $B^{\mathrm{T}} g$ without knowing g. Let us consider the first column of B, that is, the first component, a vector, b say. Then the first element of $B^{\mathrm{T}} g$ is estimated as

$$(2h)^{-1}[\text{net assets}(X = hb) - \text{net assets}(X = -hb)].$$

We need one of these calculations for each component, not for each point on the yield curve.

Other contexts may also show an advantage from needing fewer components. For example, Monte Carlo work involves simulating the components of Z and computing the matrix product BZ. Both of these operations are quicker if the number of components is reduced.

The pricing of some financial products involves optimal stopping problems. Examples include estimating the best time to pre-pay a fixed-rate loan or to cash in a fixed-rate deposit. The analysis of such products involves movable boundary problems, whose solution is only straightforward for low dimensional problems. For this reason, consideration of such products is usually in the context of 1 or 2-component models.

In each of these cases, neglecting higher components introduces errors, because the correlation structure is modelled inaccurately. The errors may still be considered a price worth paying in order to benefit from the run time advantages of a low dimensional interest rate model.

C.4 *Solution Rotation*

Solutions B to the equations $X = BZ$ and $BB^{\mathrm{T}} = V$ are not unique and we can ask how the different solutions are related.

The answer lies in rotations. We have the model $X = BZ$. Here, the random vector Z consists of independent identically distributed $N(0, 1)$ variables. Contours of equal density are concentric hyper-spheres around the origin.

Rotations can be written as matrix multiplication, by some matrix Ω. A matrix Ω corresponds to a rotation if and only if $\Omega - 1 = \Omega^{\mathrm{T}}$. If Z is a vector of independent identically distributed $N(0, 1)$ variables, then so is ΩZ. Thus, a model $X = BZ$ produces the same distribution for X as a model $B\Omega Z$. This means that if B is one decomposition, then $B\Omega$ is another.

C.5 *Variance-Covariance Matrices*

At each key rate, we assume the standard deviation of X is 1%. The correlation matrix between yields at different maturities is assumed to take the following simple form:

Correlations	$t = 1$	$t = 2$	$t = 3$	$t = 5$	$t = 10$	$t = 20$
$t = 1$	1	0.9	0.8	0.7	0.6	0.5
$t = 2$	0.9	1	0.9	0.8	0.7	0.6
$t = 3$	0.8	0.9	1	0.9	0.8	0.7
$t = 5$	0.7	0.8	0.9	1	0.9	0.8
$t = 10$	0.6	0.7	0.8	0.9	1	0.9
$t = 20$	0.5	0.6	0.7	0.8	0.9	1

The standard deviations are all 1%. Therefore, measured in percentage terms, the covariance matrix is the same as the correlation matrix, which is convenient for our example. We denote this matrix by V.

The appropriate assumptions to use for a particular yield curve, a particular financial entity and a particular point in time are all subject to debate. Our assumptions for the worked example are not derived from any particular data set, but they are broadly representative of moves in yields for developed economies in recent times. By choosing simple assumptions we aim to make it easy for anyone else to verify our numbers.

C.6 *Principal Components*

One of the best known methods of choosing components is *principal component analysis*, or PCA (Anderson (1957)), devised so that the early components (Z_1 and Z_2 for example) explain as much as possible of the variability in the rates, minimising the role of later components Z_3 to Z_6. This equates to maximising the sum of the squares of elements in the first two columns. It can be shown that this is equivalent to choosing the components as the eigenvectors of V sorted in descending order of eigenvalue.

The table shows principal components of what we will later call "Model 6".

Principal components: Model 6

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
$t=1$	0.833278	−0.481806	0.222651	0.129099	0.077942	0.034592
$t=2$	0.909996	−0.352706	0.014383	−0.129099	−0.147311	−0.094507
$t=3$	0.949163	−0.129099	−0.209257	−0.129099	0.072806	0.129099
$t=5$	0.949163	0.129099	−0.209257	0.129099	0.072806	−0.129099
$t=10$	0.909996	0.352706	0.014383	0.129099	−0.147311	0.094507
$t=20$	0.833278	0.481806	0.222651	−0.129099	0.077942	−0.034592

The columns of the matrix represent the principal components. As an example, if we use just one component then we would reduce the vector Z to $(Z_1, 0, 0, 0, 0, 0)^{\mathrm{T}}$ and our estimate BZ of X would be the one factor model $(0.83, 0.91, 0.95, 0.95, 0.91, 0.83)^{\mathrm{T}} * Z_1$.

Each component is independent of (or at least uncorrelated with) the other components. Traditionally, the components are interpreted so that the first represents the level of the curve, the second component the slope and the third component the curvature.

C.7 *Cholesky Decomposition*

A *Cholesky decomposition* is another text-book solution to finding a matrix B such that $BB^{\mathrm{T}} = V$. The decomposition works provided that V is symmetric and positive semi-definite. These are precisely the conditions that V is a valid variance-covariance matrix.

The Cholesky method produces a matrix B which is *lower triangular*, that is, so that all elements above and to the right of the main diagonal are zero. It also ensures that the diagonal elements of B are non-negative. Within these constraints, B is uniquely specified.

The table shows a Cholesky decomposition for our correlation matrix:

Cholesky decomposition	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
$t = 1$	1.0	0	0	0	0	0
$t = 2$	0.9	0.435890	0	0	0	0
$t = 3$	0.8	0.412948	0.435286	0	0	0
$t = 5$	0.7	0.390007	0.411103	0.434613	0	0
$t = 10$	0.6	0.367065	0.386921	0.409048	0.433861	0
$t = 20$	0.5	0.344124	0.362738	0.383482	0.406745	0.433013

We can see that the first component captures the first column of the correlation matrix. The second component starts at zero for $t = 1$. This is because the first component already explains the yield shift at $t = 1$. And so the pattern continues. Figure 38 shows the components.

Although these components do reproduce the desired correlation matrix, they do not appear natural or intuitive. For example, if we recomputed the Cholesky calculation with the time points in reverse order, starting from $t = 20$, the components look different — in fact, given the symmetry of the correlation matrix, reversing the time axis has the effect of a vertical reflection of the whole chart.

This lack of intuition is not a problem for some applications. For example, Cholesky decomposition is the preferred algorithm in Monte Carlo work, on account of its simplicity. On the other hand, for communicating capital requirements, more intuitive decompositions are helpful.

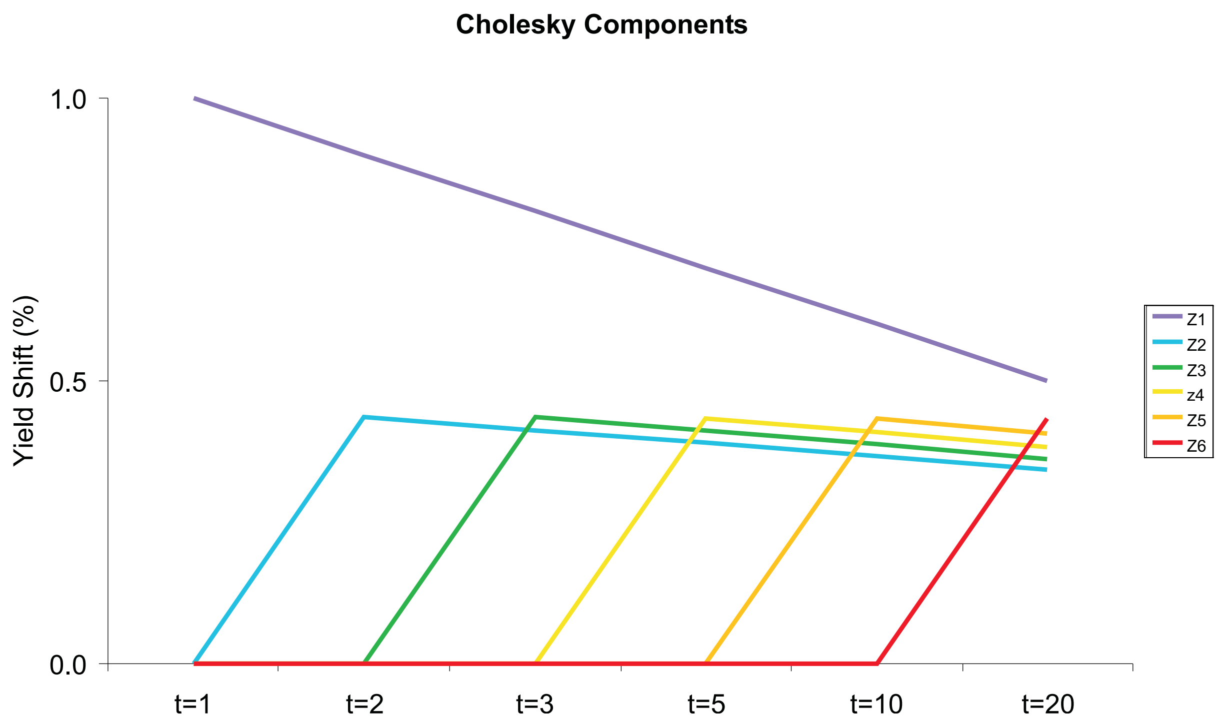

Figure 38

In our example, we can convert a Cholesky decomposition to polynomial components using the following rotation matrix:

Rotation to convert Cholesky decomposition to polynomial components	Original Z_1	Original Z_2	Original Z_3	Original Z_4	Original Z_5	Original Z_6
New Z_1	0.866025	−0.482214	0.017380	0.065366	−0.113533	−0.002384
New Z_2	0.198680	0.338353	−0.688167	−0.501604	−0.312521	−0.152294
New Z_3	0.209427	0.266734	−0.338611	0.200056	0.513594	0.683033
New Z_4	0.221404	0.440493	−0.104847	0.606656	0.163600	−0.592584
New Z_5	0.234834	0.568667	0.480340	0.047983	−0.529753	0.328335
New Z_6	0.250000	0.253819	0.412004	−0.577723	0.564144	−0.226510

C.8 *Polynomial Components*

Until now, our analysis has treated the yields at $t = \{1, 2, 3, 5, 10, 20\}$ as six distinct random variables. We have not used the associated time values in our models.

However, it is seen empirically that yield curves are often smooth functions of time t. This suggests we look to smooth sample functions to build up components. One possible choice is the family of polynomials in t. A disadvantage of polynomials is their tendency to infinity for large t, while yield curves in practice tend to flatten out. A solution to this problem is use polynomials, not in t itself, but in A^t for some $0 < A < 1$. In our examples, we have selected $A = 0.8$.

The table shows the coefficients of the polynomial decomposition:

Polynomial coefficients	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
1	0.8660	0.4963	0.0091	−0.1128	0.1816	−0.0776
A^t	0	−1.2232	−2.3339	4.2749	−6.3338	7.9003
A^{2t}	0	0	2.9302	−13.7102	37.9863	−87.5140
A^{3t}	0	0	0	10.8064	−74.8113	328.9030
A^{4t}	0	0	0	0	45.8107	−493.1470
A^{5t}	0	0	0	0	0	254.3899

The upper diagonal form gives a hint that these coefficients are also obtained from a form of Cholesky decomposition.

The coefficients of Z_n in the n^{th} column is a polynomial of order $n-1$ in A^t. Evaluating this gives the corresponding components:

Polynomial components	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
$t = 1$	0.866025	−0.482214	0.017380	0.065366	−0.113533	−0.002384
$t = 2$	0.866025	−0.286508	−0.284323	−0.159815	−0.238404	−0.068529
$t = 3$	0.866025	−0.129943	−0.417666	−0.067763	0.003679	0.232518
$t = 5$	0.866025	0.095510	−0.440996	0.196032	0.080885	−0.037812
$t = 10$	0.866025	0.364980	−0.207677	0.201473	−0.147034	0.107004
$t = 20$	0.866025	0.482214	−0.017380	−0.065366	0.113533	0.002384

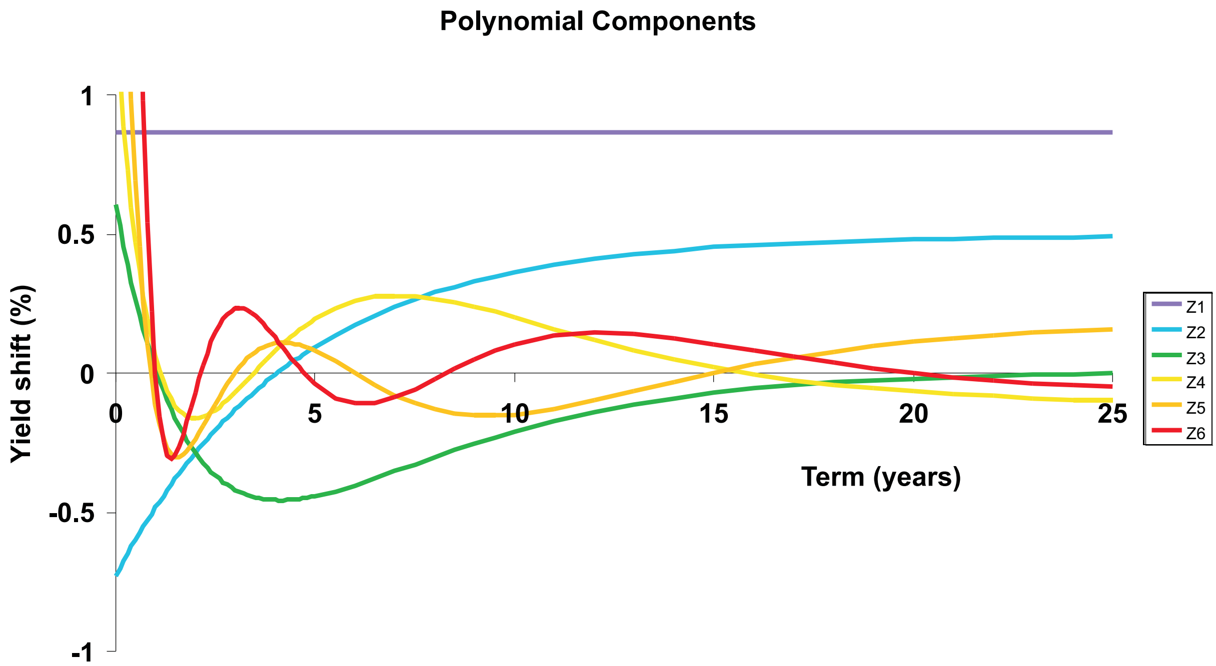

Figure 39

One advantage of using polynomials is that they interpolate naturally for other values of t. Figure 39 shows the polynomial components for a range of t values. We see that successive components become smaller but also more wiggly — a well-known feature of higher order polynomials.

The behaviour for $t < 1$ is a possible cause for concern. It reflects the tendency of polynomials to wiggle. Our choice of $A = 0.8$ produces only minor wiggle problems with our chosen correlation structure. An informal test of examining charts by eye, suggests that most other choices of A produce more pronounced wiggles, which is how we chose $A = 0.8$ in this example.

C.9 *Proportions Explained*

Let us consider the variance of the rate at a particular term as components are added.

The first component necessarily understates the variance. For the yield curve as a whole, the contribution of the first component can be measured as the sum of the squares of the first column, which is the sum of the squares of the first noise term Z_1. The inclusion of each new component increases the variance, until all the components are incorporated and the sum of the squares of the entire matrix is 1. We can investigate the proportion of variance explained by each component, split according to the time point, that is, by rows of the original matrix. Figure 40 shows the results for polynomial components. In order to make the higher components more visible, the vertical axis starts at 60%.

We also show the corresponding figure for Cholesky decomposition (Figure 41).

Here we see that the first component explains all of the variance for the 1 year rate. However, converge is much slower for longer term rates, and indeed, is so bad for $t \geq 5$ as to fall below our vertical scale.

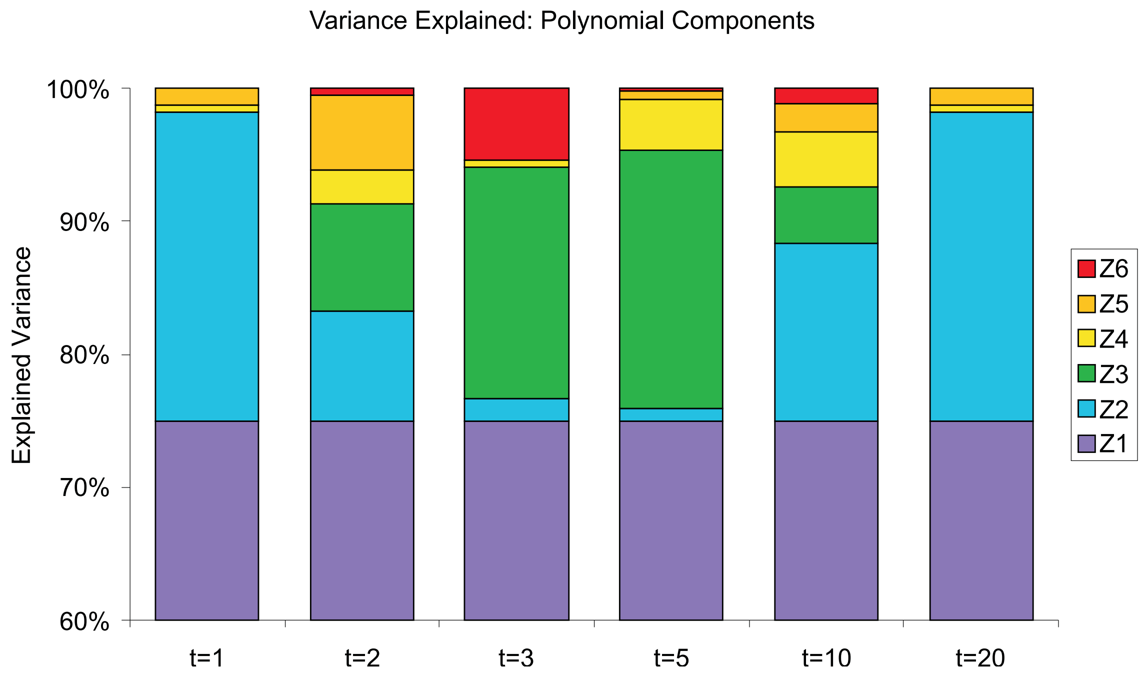

Figure 40

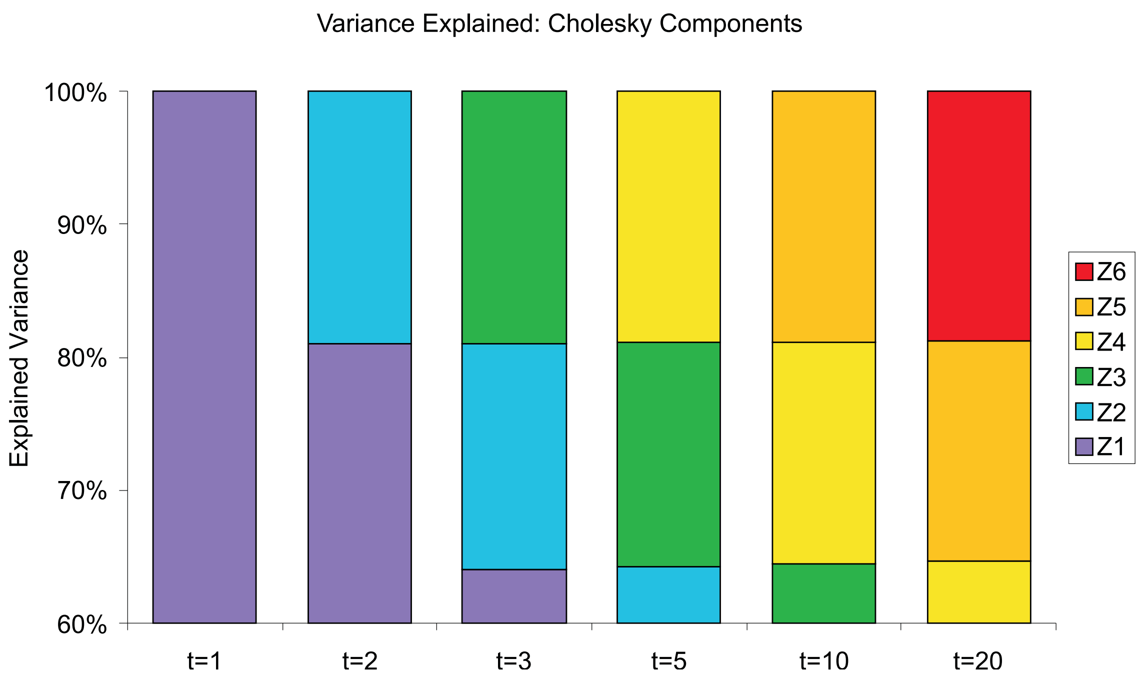

Figure 41

We can show the same chart for principal components (Figure 42).

We can consider how to choose the components to maximise the variance that the early components explain. There is clearly a trade-off here. Cholesky decomposition does a good job at $t = 1$ but a terrible job at $t = 20$. Polynomials and principal components are more consistently convergent across a range of terms.

Taking the average across the six terms considered, we can measure the speed of convergence as more components are added. Figure 43 shows the results (note that in order to make the differences more visible, the vertical axis starts at 60% explanation and not 0%).

We can see that the Cholesky approach converges slowest of all. The polynomial decomposition is much better, but not quite as fast as principal components analysis.

In fact, principal components are defined in order to maximise the speed of convergence, averaged across the key terms, that is to maximise the height of each green bar in Figure 43. Therefore, the principal components represent the best possible convergence outcome, which cannot be beaten. This is encouraging, because it gives us a basis for claiming that some decompositions are "better" than others, and in particular that principal components analysis is "best possible". Furthermore, this analysis has assumed nothing about the underlying financial business. This raises the hope of component analysis that can be performed once and is then valid for multiple applications.

Figure 44 shows the principal components resulting from the analysis.

We can see that the first three components capture aspects of level, slope and curvature, with higher components being wigglier. This is similar to the situation for polynomial components.

C.10 *Alternative Choices of Weights*

We have analysed principal components based on speed of convergence at key time points $t = \{1, 2, 3, 5, 10, 20\}$.

We could consider alternative weights. For example, a "Model 20" that gives equal weight to all time points between $t = 1$ and $t = 20$ inclusive. Or a "Model 50" based on a 50-year yield curve. The respective weights are shown in Figure 45.

As we shall see, it is an unhelpful feature of principal components analysis that the choice of weights is a major determinant of the calculated components. The need to choose weights does not arise for our other methods of component construction:polynomials and variance matching. When using principal components analysis, we might (for example) construct a system of weights to reflect the relative size of cash flows at different terms. The ideal choice of weights might therefore vary from one business to another. On the other hand, it may be better to use a common compromise

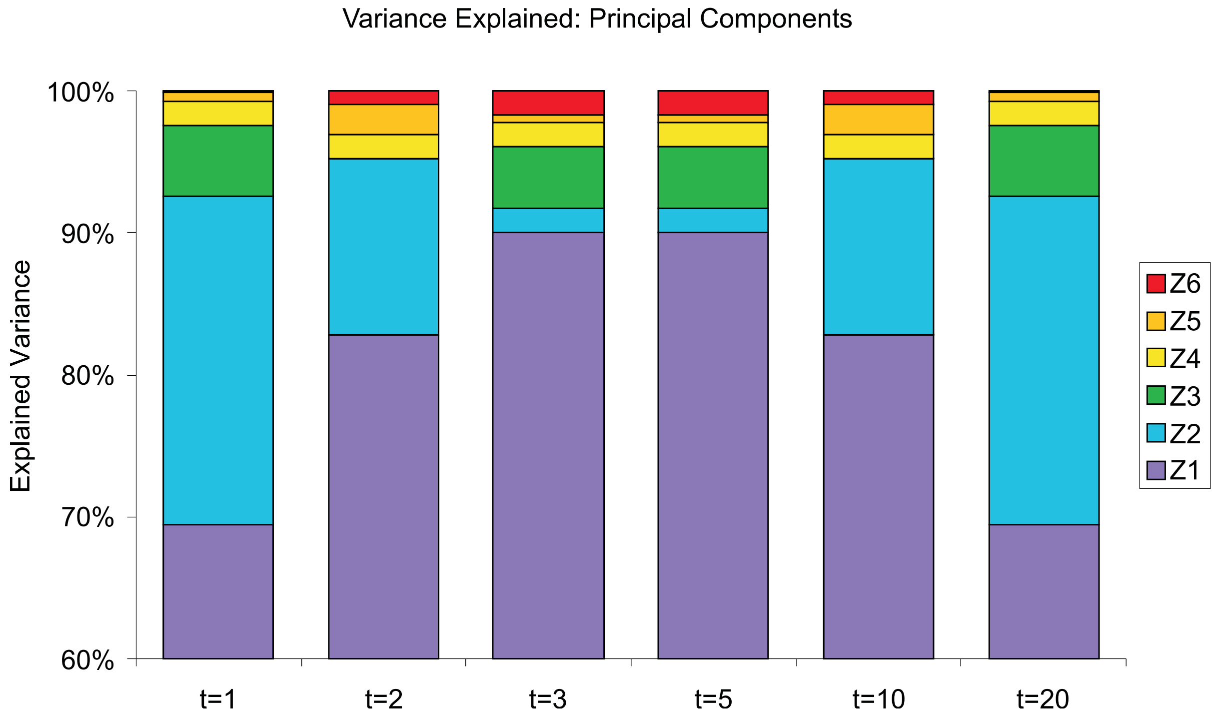

Figure 42

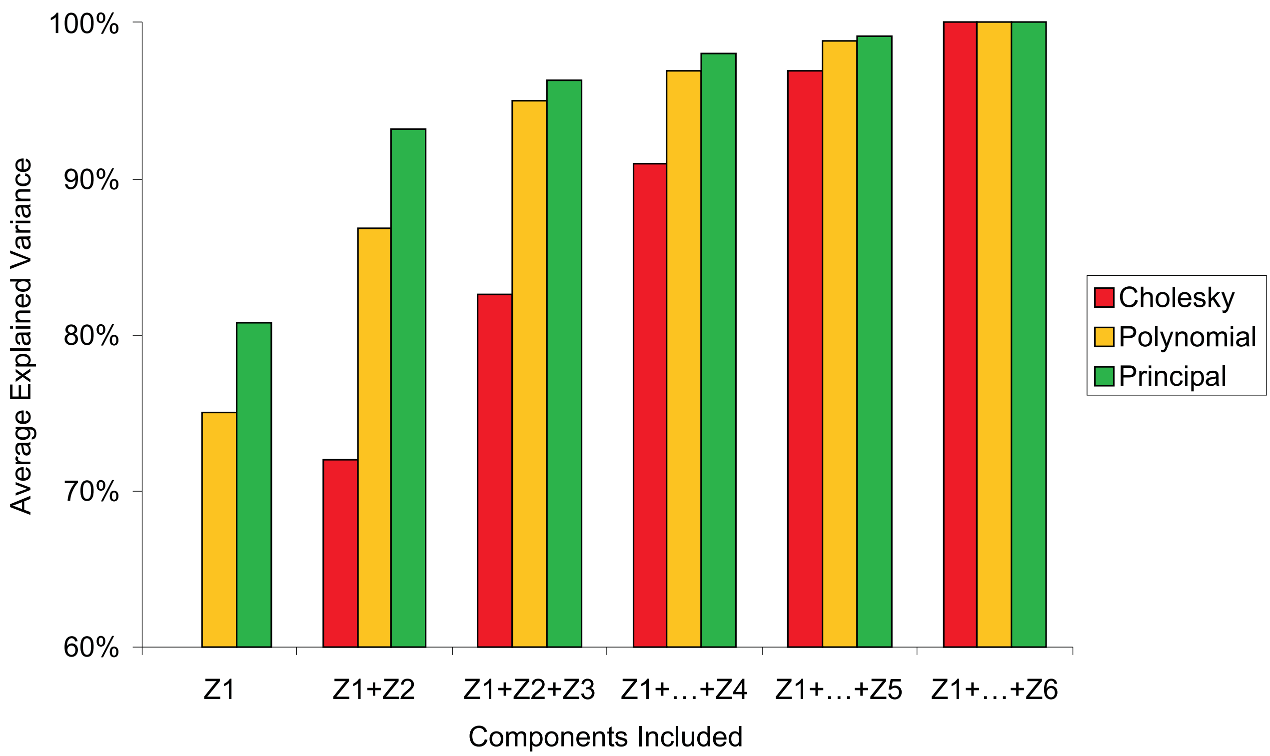

Figure 43

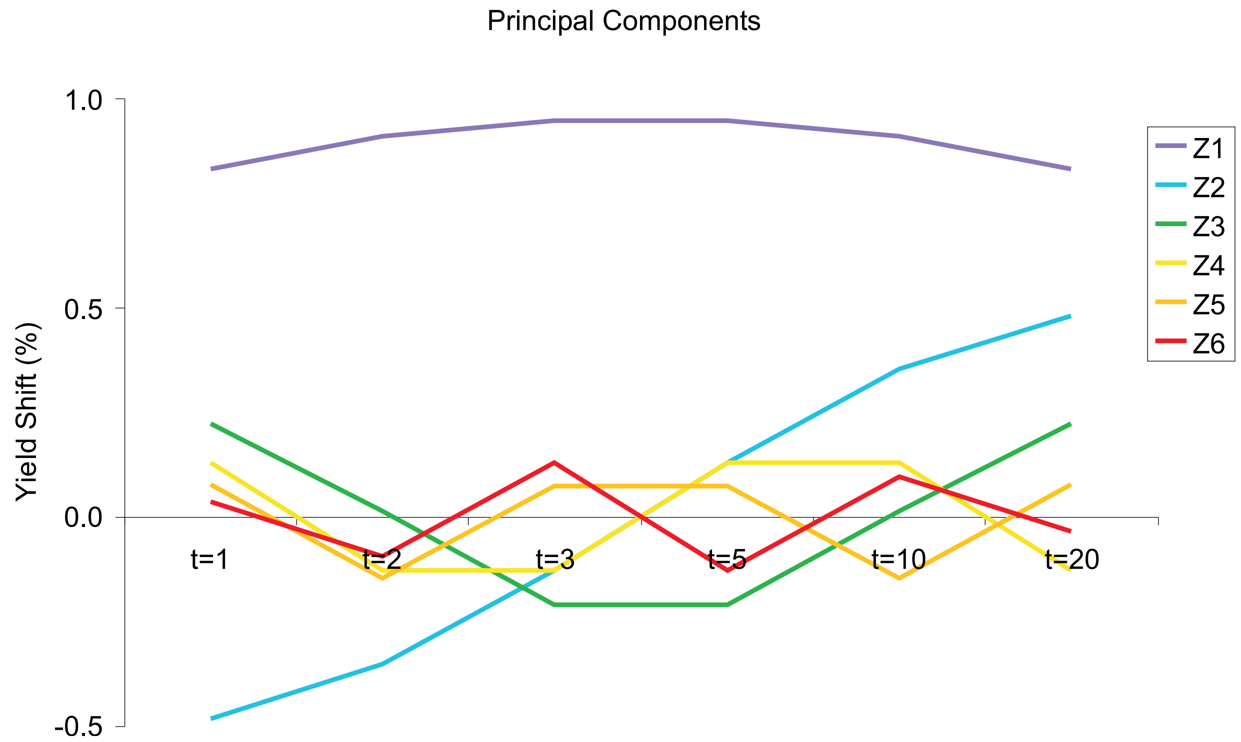

Figure 44

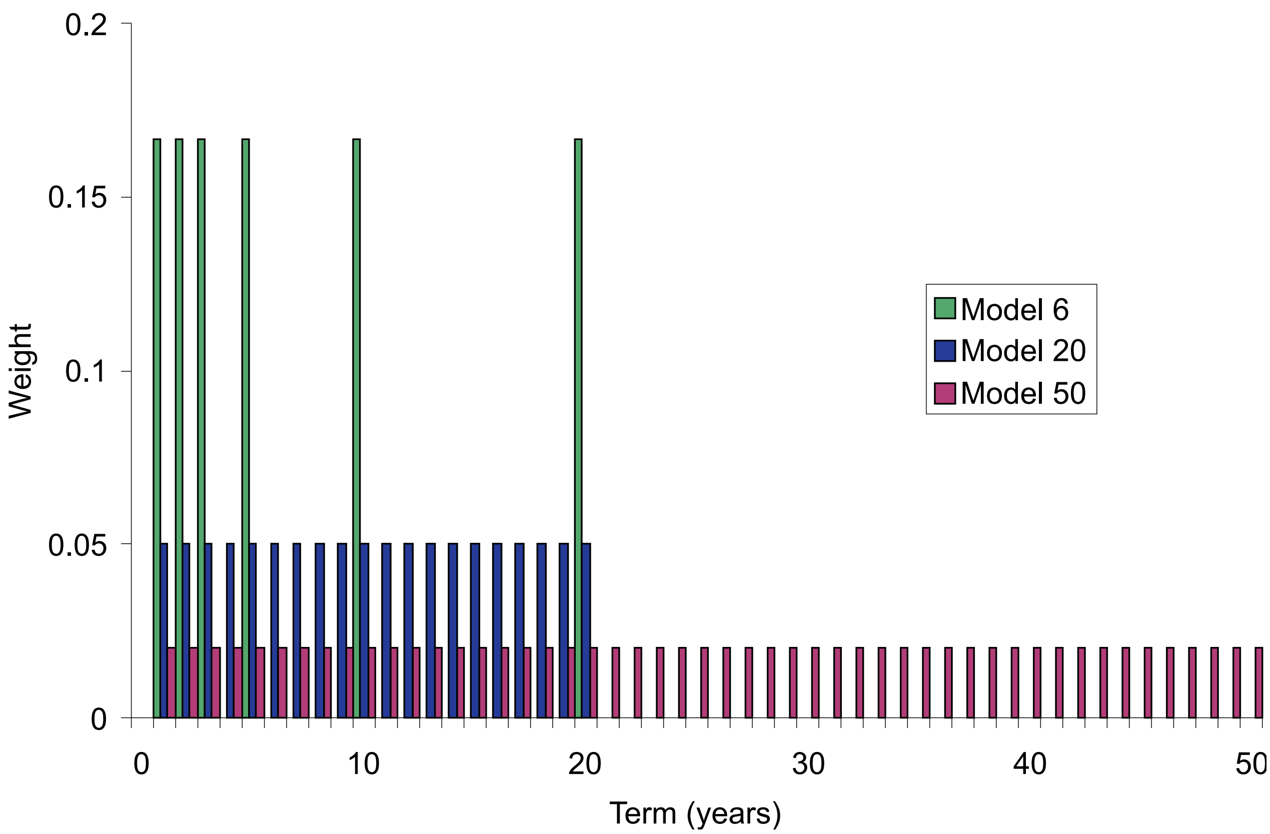

Figure 45

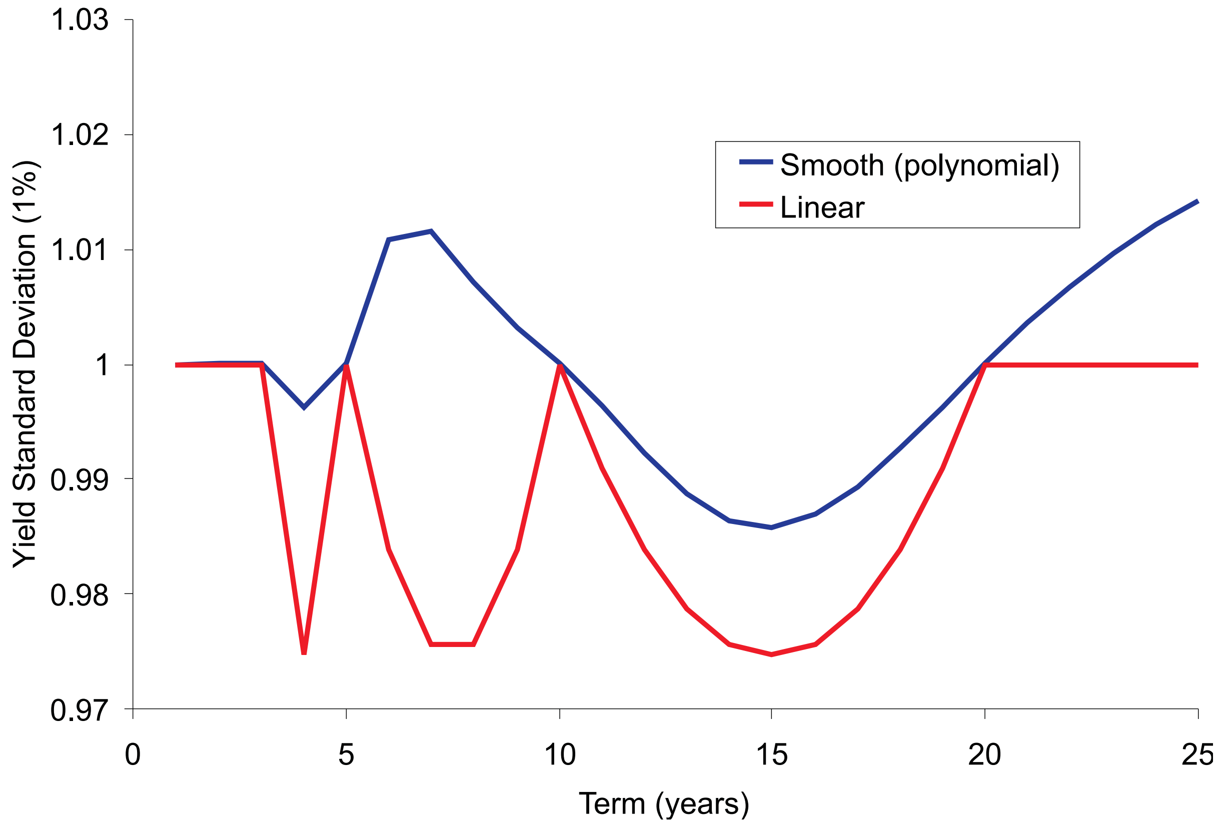

Figure 46

set of weights for all businesses in order to simplify aggregation calculations across multiple lines of business.

If we are to use intermediate time points between those originally modelled, we must also specify a model for yields at those points. In this note we consider two alternatives. The linear model "L" uses linear interpolation, with flat extrapolation beyond the 20 year point. The smooth model "S" fits a polynomial of order 5 to the six observed points.

The possible combinations of weights and interpolations give us five models, which we will denote as Model 6, 20L, 20S, 50L and 50S.

Our original model had yield standard deviations of 1% at all terms. However, these standard deviations do not necessarily apply at other, interpolated terms. Figure 46 compares standard deviations for both smooth and linear interpolation. Note the vertical axis — the standard deviations are close to 1, in fact, as near to 1 as makes little practical difference.

Linear interpolation always gives a standard deviation of 1% or less, with the reduction due to diversification in the interpolation between points.

C.11 *Principal Component Comparisons*

We now compare principal components under our five models: 6, 20L, 20S, 50L and 50S (Figure 47).

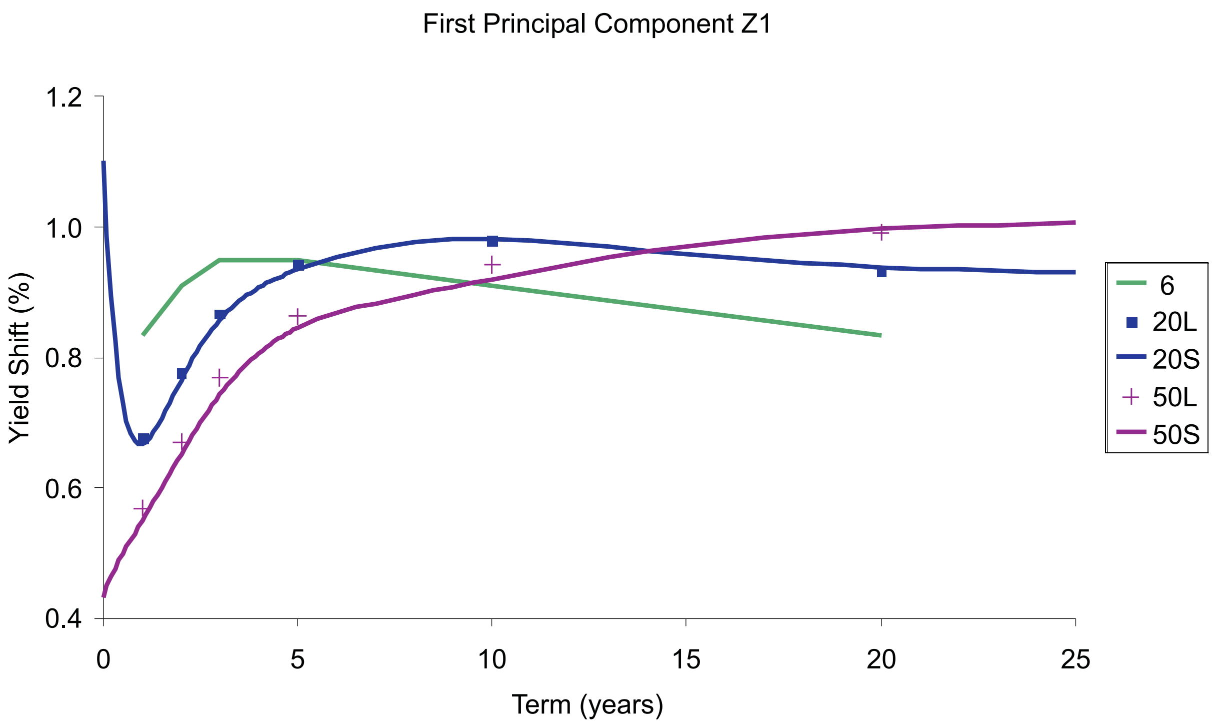

Figure 47

In each case, the first principal component can reasonably be interpreted as a measure of "level". We see differences according to where the weights lie. The first component is largest where the weights are largest. Thus, model 6 produces a maximum at $t = 3$, while model 50 has its maximum at $t = 25$. The choice of interpolation method (L or S) has little effect.

The second component captures the slope, or twist, of a yield curve move. One important question in value at risk calculation is the choice of pivot, that is, the term of interest rates which is unchanged by the twist stress test. This corresponds to the Y intercept of the second principal component.

Figure 48 shows that the pivot is not an inherent property of past yield curve moves. On the contrary, it depends on the weights chosen for principal components analysis. If the weights extend a long way into the future, then the pivot occurs at a large t value. The choice of weights is not a purely technical decision with limited impact. On the contrary, a consideration of the second principal component reveals the importance of the choice of weights. Given the potential difference in calculated value at risk, a rigorous motivation for the choice of weights is important, in place of the heuristic reasoning we are able here to provide. The choice of smoothing method, however, remains unimportant.

We now jump to the last principal component (Figure 49).

It is unlikely that this component is much used in practice. However, the comparison with previous figures is interesting. We see a modest effect of the choice of weights, with the 20 and 50 year models looking very similar.

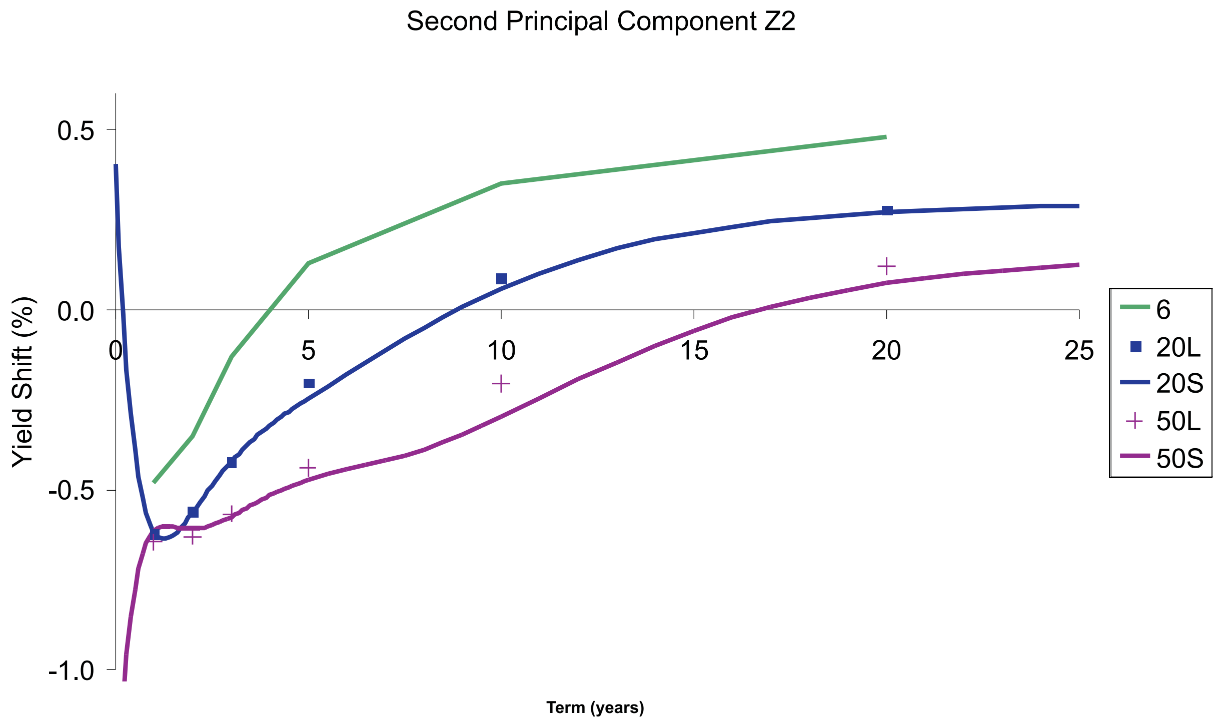

Figure 48

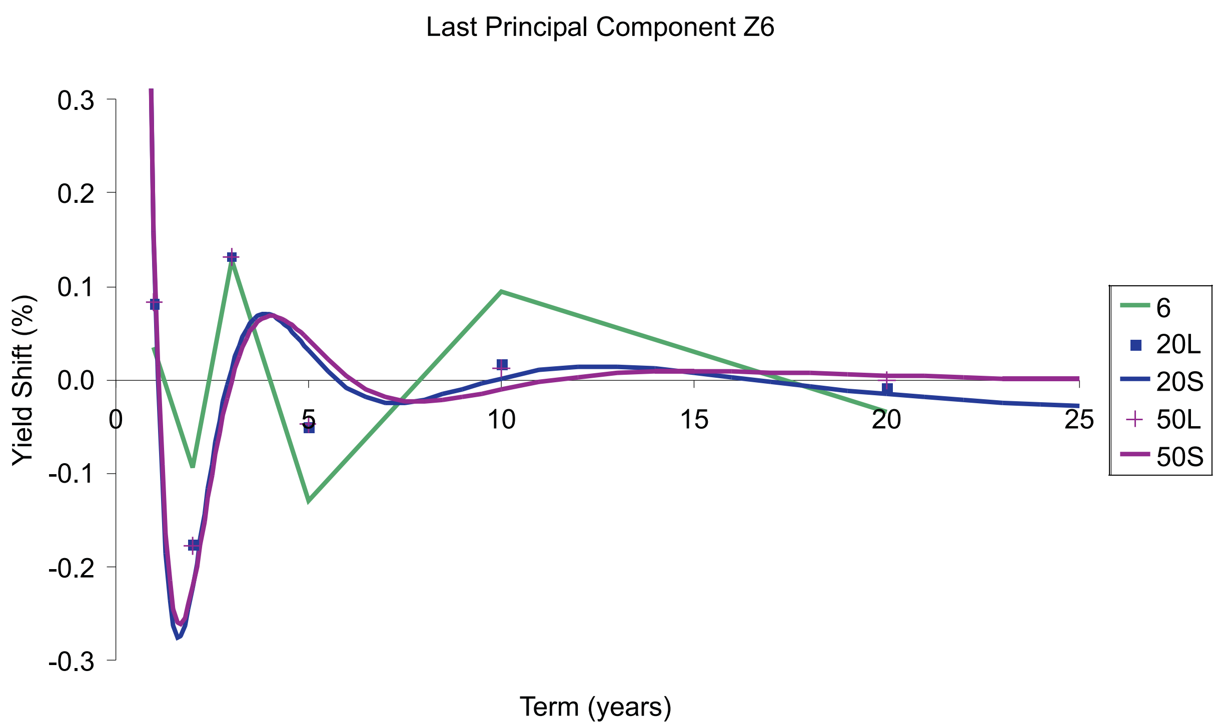

Figure 49

What the last component picks up is the smoothing algorithm. The wiggles from the "S" models come through clearly, in contrast to the simpler shape under linear interpolation.

C.12 *Implicit Data Enhancement*

Ideally, interest rate data is observed from actual trades or bid/offer quotes in a deep liquid market. In many markets, however, trades may be infrequent or bid-offer spreads wide, especially for long dated cash flows.

Given these difficulties, we might expect consequences for yield curve data quality and availability. Online sources, however, provide apparently complete information extending far back in time. They can do this because of substantial investment in data cleaning. Cleaning methods include interpolation and extrapolation to infer missing data points or to adjust out-of-date price information. The data collection may be a many stage process: individual banks apply their own cleaning algorithms to data made even cleaner by commercial data vendors' systems.

Clean data does not easily reveal which data points are real market information and which are filled in by algorithm. However, principal component analysis may reveal this information. And you could test this in a real market by trying to transact at prices posted on the system.

For an example of the power of PCA, suppose a user downloads yield curve data out to 50 years maturity. Theoretically, this data set may require 50 components for a full explanation of the correlation structure. Maybe, analysis of the data shows 6 significant components, with the higher components accounting for a negligible proportion of total variability.

One possible explanation is that there really are only 6 components in the economy. Another explanation is that we are dealing with a 50L or 50S model, with 6 real data points and 44 points constructed by interpolation. The shape of the 6th component reveals more about the type of interpolation used. In other words, the higher components tell us about the process of data collection and cleaning, rather then about risk in the financial markets.

C.13 *Value-at-Risk for Hedging*

Economic capital is an important application of yield curve models. Economic capital is often regarded as having a cost, and so firms try to trade to minimise stated capital requirements.

One important tool for this is *hedging*. For example, if a firm has a long exposure to an 11 year interest rate, they may trade in the interest rate markets to acquire a corresponding short exposure to the same interest rate. The net effect is immunisation — the firm is protected against moves in either direction, at least against small moves.

Hedging is often less exact than this. A firm may decide to hedge the 11-year exposure with a 10-year trade, for example because the 10-year instrument is more liquid than an exact match. This hedge should still be effective — because the 10 year rate and 11 year rate are strongly correlated, but not quite as good as hedging the same rate as the original exposure.

In this case, PCA converge more slowly than we hope. PCA finds components that well explain movements in the yield curve, but that is not the same as explaining movements in my portfolio. PCA may ensure that the variances of 1 and 2 year rates are well explained. But to assess hedge effectiveness you also need to know the correlation between them. PCA solves a particular, objective function, but that objective function takes no account of how fast correlations converge. Technically, we could include covariances in the weights for evaluating convergence, but the optimisation would then trade off covariances against variances elsewhere.

For example, let us suppose an investor has an exposure of £1 per 1% move in 1 year rates, which they have partially hedged with an equal and opposite exposure to the two year rate. From our assumed correlation matrix, we can calculate the variance of profit to be $1 + 1 - 2 * 0.9 = 0.2$. We can investigate how quickly the different decompositions converge to the true value. The principal components calculation refers to Model 6 (see Figure 50).

As we would expect, the Cholesky method scores well, as it builds up the volatility by iteration starting at the short end of the curve. With cash flows only at $t = 1$ and $t = 2$ in our example, the Cholesky method has captured all the variance in the first two terms.

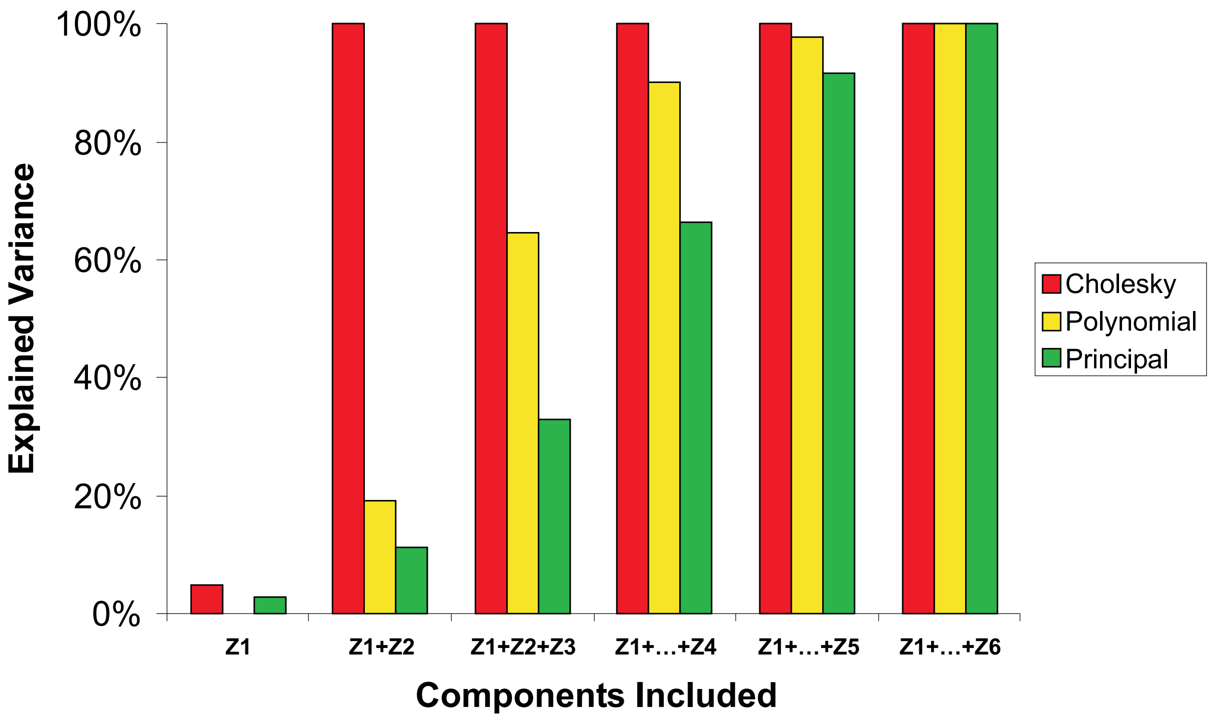

Figure 50

The principal components method performs surprisingly badly. Section 3.1 suggests that the first three principal components explain more than 95% of the yield curve variability. However, in our example, only 33% is explained by the first three components. Arguably, this is an unfair comparison; had we known that the cash flows stopped at time 2, we would have applied more weight to the early years when calculating principal components, and so obtain faster convergence for those flows.

In this particular example, the Cholesky method gives good convergence, while Principal Components has the slowest convergence. This will not always be the case. We could construct examples illustrating any of the six possible orderings. Instead, our example is intended to illustrate the potential gap (in either direction) between an advertised "percentage explained" for the yield curve as a whole, compared to the actual explanatory power for a particular set of cash flows. By construction, PCA maximises the advertised percentage explained, and therefore carries the greatest potential to disappoint.

In general, the problem of slow convergence is particularly acute if yield decomposition into components is used for constructing a hedge in the first place. For example, given a particular definition of level, slope and curvature, it is easy to find a hedge portfolio that immunises all three. Tested against that decomposition, it appears that risk is eliminated. What has really happened is that risk is concentrated in the fourth and higher components, which have been discarded on order to speed up the value-at-risk calculation. There are two solutions to this problem. One is to use more components for analysing risk than are used for building the hedge in the first place. The second solution is to use a full model, including all components rather than truncating.

C.14 *A One-Factor Model*

Suppose we are constrained to use only a single component. Then interest rates moves at different terms must be 100% correlated.

A simple solution is to set the interest rate standard deviation to 1% at all the modelled terms. We then have a solution that explains 100% of the variance for each key rate, although clearly the correlations are overstated. We call this "variance matching". This is essentially the test which CEIOPS have calibrated for solvency II, based on the volatility of yields at various terms.

A theoretically more sophisticated solution is to use the first component from PCA. Figure 51 shows a comparison.

The PCA is supposed to maximise the variance explained by each component. In our example, the first component explains 80.8% of the yield curve variance.

Yet, the variance match explains 100% of the variances. The correlations are equally wrong in both one-component models. So it is difficult to

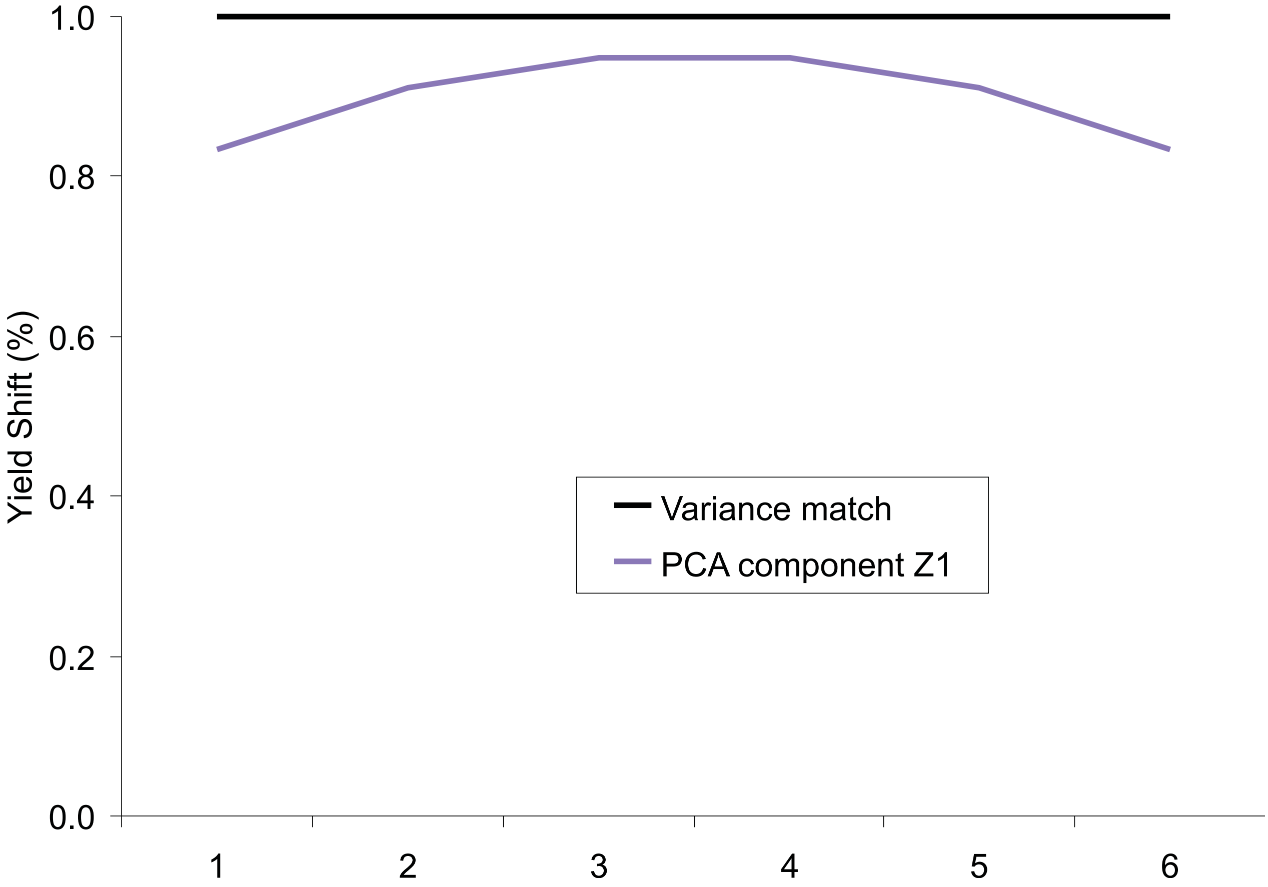

Figure 51

describe any sense in which the theoretical superior PCA is better in practice. Indeed, we might wonder how we ever convinced ourselves that 80.8% is the best possible, given that variance matching explains 100% of the variance.

The answer to these points is subtle. The advantage of PCA is that the first component is one of a series. By adding more and more terms, we can get closer to the true yield curve distribution. This is useful if we can somehow test convergence, optionally adding higher terms only when necessary.

In contrast, the variance match might produce a good first guess, but we cannot refine that guess by adding more components. If we add further components, we might get the correlations more accurate, but those extra turns will also increase the variances, making them too large.

If we know at the outset that we might want to explore further stress tests, then the PCA makes sense. The first component is one step on a longer path. On the other hand, if more than one stress is excluded, for example by computation costs, then the variance match could be a better approach.

C.15 *Three Factor Variance Match*

We have discussed a variance match for a single factor model. We now consider extending this idea to three factors.

We assume the importance of capturing the variance of yields at each term. With three factors, we are able to impose additional constraints. Knowing that hedging often involves offsetting risks at adjacent terms, we can ask that our three factors correctly replicate the correlations between adjacent terms. This is equivalent to reproducing the variance of the yield curve slope between terms.

These are still too few constraints to determine a three factor model. We can insist also on capturing the correlations between rates that are next-but-one to each other. Equivalently, we reproduce the variance of the second differences in yield slope. These differences are relevant to "barbell" hedging strategies that (for example) seek to hedge a 2 year exposure with an average of 1 and 3 year exposures.

We have articulated some constraints that are relevant to common business strategies. Now all we need is to solve the equations. The solution is:

Variance matching components	Z_1	Z_2	Z_3
$t = 1$	0.974679	−0.223607	0.000000
$t = 2$	0.974679	0.223607	0.000000
$t = 3$	0.872082	0.223607	−0.435286
$t = 5$	0.872082	−0.223607	−0.435286
$t = 10$	0.974679	−0.223607	0.000000
$t = 20$	0.974679	0.223607	0.000000

We can use these three components to reconstruct the following implied correlation matrix:

Correlation matrix implied by Variance Matching Components	$t = 1$	$t = 2$	$t = 3$	$t = 5$	$t = 10$	$t = 20$
$t = 1$	1	0.9	0.8	0.9	1	0.9
$t = 2$	0.9	1	0.9	0.8	0.9	1
$t = 3$	0.8	0.9	1	0.9	0.8	0.9
$t = 5$	0.9	0.8	0.9	1	0.9	0.8
$t = 10$	1	0.9	0.8	0.9	1	0.9
$t = 20$	0.9	1	0.9	0.8	0.9	1

We can see that the shaded elements: the main diagonal and two above and below it, are replicated exactly, as required. However, the variance matching components overstate the elements on the bottom left and top right of the matrix. In particular, the variance matching components force the rates at $t = 1$ and $t = 10$ to be equal. It also forces equality between rates at $t = 2$ and $t = 20$. This is an undesirable side-effect of using only three components when six are required.

C.16 *Three Factor Models: Alternative Weightings*

We have constructed a three factor variance-match solution for a 6-point model. We now consider how that solution extends to models with more points on the yield curve.

The variance match for model 6 also works for model 20L and 50L. The reason is that the variance match fits the variance of the value and the first and second differences. The second differences of a linearly interpolated curve, are zero, except at the data points. This means that solving for the three components commutes with interpolation. It does not matter whether we interpolate first, then construct components, or if we solve first for the components and then interpolate.

The situation is different for the models 20S and 50S. The non-linear interpolation means that fitting the variance of second differences is not trivial. Nevertheless, it can be done. Furthermore, the solution for model 20S is simply the decomposition for 50S, restricted to the first 20 years (Figure 52).

As with the original decomposition, these solutions are unique only up to rotation of the underlying normal variables. In this example, we chose a rotation so that the second component vanishes at $t = 4$, while the third vanishes at $t = 2$ and $t = 10$. These properties are then shared with the Model 6 components.

The variance matching solution has many advantages, including replication of yield variances, and also the variances of first and second

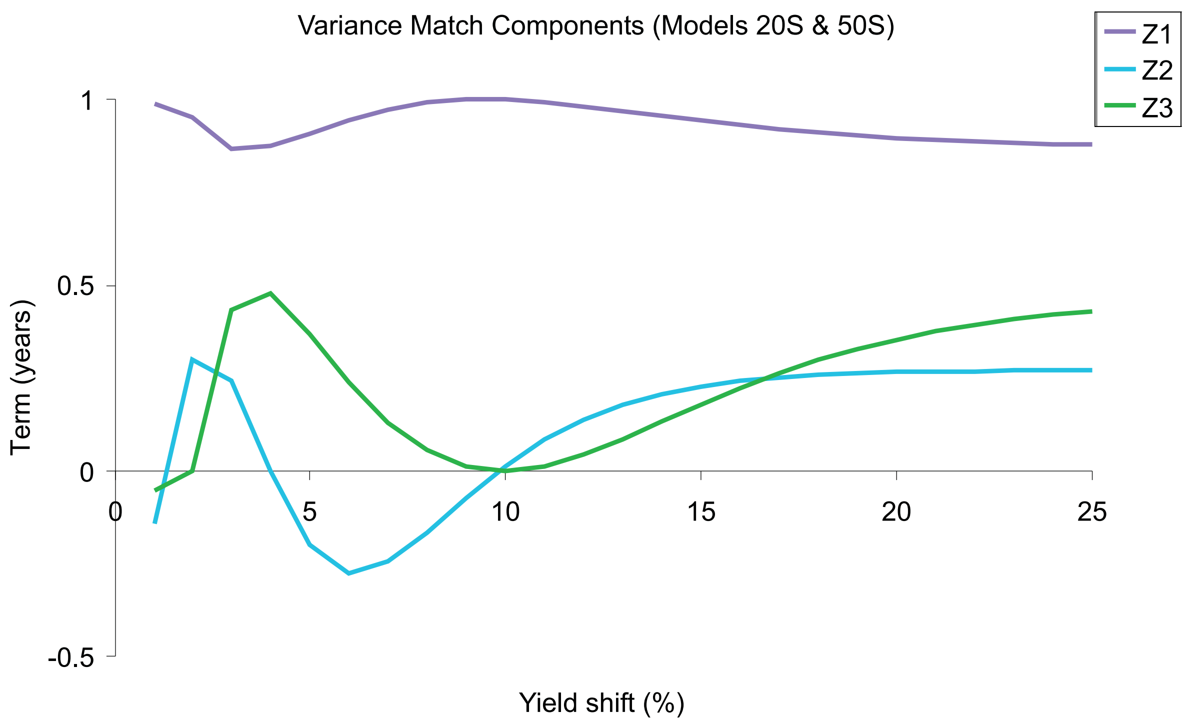

Figure 52

differences, as well as avoiding the need to specify time weights. There are also a few disadvantages. One disadvantage is the inability to build on this solution by adding more factors — if we want to add a fourth factor we have to go back to the beginning and build all four from scratch. In that sense, the PCA approach is better, as fourth factor can be added without disrupting the previous three. The second disadvantage of the variance match approach is the tendency to wiggle. This is necessary in order to capture the variability of yield curve slopes, but at the same time negates the intuitive appeal of a second factor relating to "slope" and a third to "curvature".

Printed by Bell & Bain Ltd., Glasgow