CRO Forum Best Practice Paper - Extrapolation of Market Data

August 2010

CRO FORUM

Table of contents

1. Executive summary

The CRO Forum has identified 4 principles for the extrapolation of market data for market consistent valuation purposes:

1. Use all relevant observed market data where available
2. Extrapolated market data should be arbitrage-free
3. The extrapolation method should be theoretically sound
4. The extrapolation should follow a smooth path from the entry point to the unconditional ultimate long-term level.

This best practice paper details considerations and best practices on how these principles can be applied to the extrapolation of interest rates, equity and interest rate[1] implied volatility and in situation in which an option has been written on a security for which no liquidly traded options exist at all. We keep the extrapolation method as simple and pragmatic as possible while remaining within the principles.

For the extrapolation of interest rate curves 5 additional principles have been developed by a Solvency II taskforce[2] on the illiquidity premium. These principles are supported by the CRO Forum and are also discussed in this paper.

[1] Although we include interest rate implied volatility, we only make some high level remarks to clarify why interest rate implied volatility is not so much a matter of extrapolation, but a matter of calibration and this paper does not intend to provide best practice on calibration to market data.

[2] The Solvency II taskforce on the illiquidity premium consists of CEIOPS, EC, CRO Forum, CFO Forum, CEA, Group Consultativ and academic representation. This group is tasked to develop principles on illiquidity premium, risk-free interest rate curve and interest rate extrapolation.

2. Introduction

This paper is part of the "Best Risk Management Practices" series of the CRO Forum. The paper outlines principles and considerations that should be taken into account for a best practice method on extrapolation of market data. We hope this paper will facilitate the implementation of best risk management practices and will help in achieving consistency for regulatory purposes and external disclosures.

Extrapolation of market data is often necessary for the market consistent valuation of insurance liabilities. We refer to the CRO Forum publication on valuation of insurance liabilities[3] for an introduction of the basic concepts of a market consistent valuation. The graph below summarizes the main components of an insurance company's balance sheet and capital requirements under Solvency II.

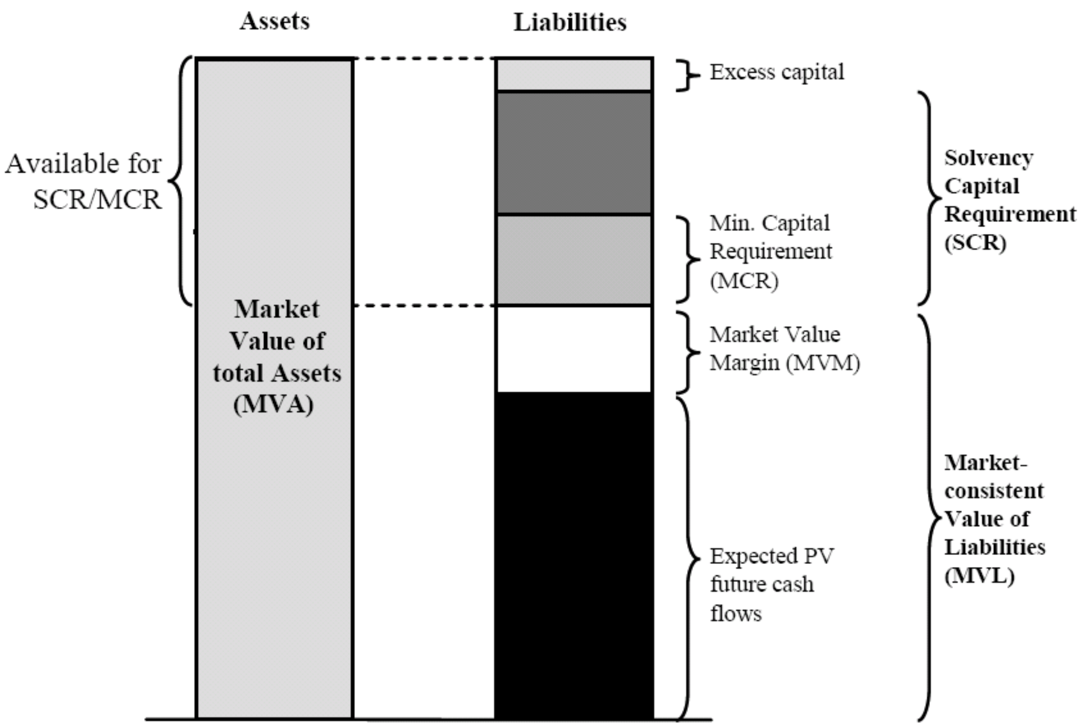


Figure 1 Main components of the balance sheet of an insurance company

2.1 Why is extrapolation important?

The market consistent valuation of insurance liabilities requires the use of market data like interest rates and option prices. Insurance liabilities often have characteristics that cannot be found exactly in financial instruments with observable market prices. Examples are:

- An insurance product with a guaranteed benefit payment up to 40 years from the valuation date, while the longest observable fixed rate risk-free bond matures in 25 years.
- An equity linked insurance product with a return of premium guarantee in year 20, while the longest observable equity option has a term of only 5 years.

Extrapolation of market data is required in both examples to be able to calculate the market consistent value of the insurance liability.

[3] "Market Value of Liabilities for Insurance Firms", CRO Forum, 28 July 2008

The choice of extrapolation methodology will impact the insurer in numerous ways:

- It impacts the Market Value of the Liability (MVL) and hence the solvency position of the company.
- The extrapolation methodology will impact the sensitivity of the MVL to changes in the observable market data. This will have consequences for the amount of Solvency Capital Requirement (SCR) and risk management practices like hedge programs.

From a financial stability perspective extrapolation methodology is important as it can exuberate or dampen volatility in the financial markets into the entire industry.

The impact on MVL and SCR can be very significant, particularly if the financial market has limited depth and the insurance liability includes long term guarantees. The examples on the following page show an extreme – yet realistic - case of the impact on MVL for two (simplified) extrapolation methods.

2..2 How is this paper organized?

First we will develop a set of principles that apply to the extrapolation of all sorts of market data in any currency. From these principles we will derive choices and consequences for the practical implementation of a best practice extrapolation method. After that we will give implementation guidance on how the principles can be applied to market data that is most relevant for the insurance industry. Types of market data we will explicitly cover are:

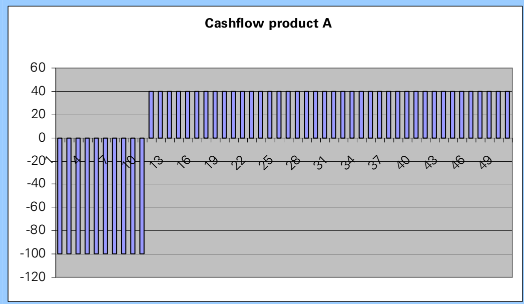
- Interest Rates[4]
- Equity implied volatilities
- Interest rate implied volatilities
- Options on securities for which no liquid traded market exist

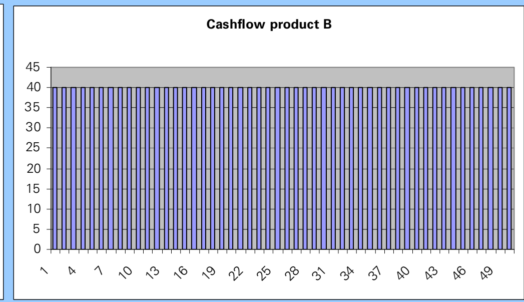
[4] In this paper we will not discuss what the most appropriate discount rate for insurance liabilities is. The extrapolation covers the basis risk free interest rate term-structure and leaves the liquidity premium and its extrapolation outside scope. The paper also does not cover any adjustments for credit risk in the market data.

EXAMPLE OF IMPACT OF DIFFERENT EXTRAPOLATION METHODS

Product

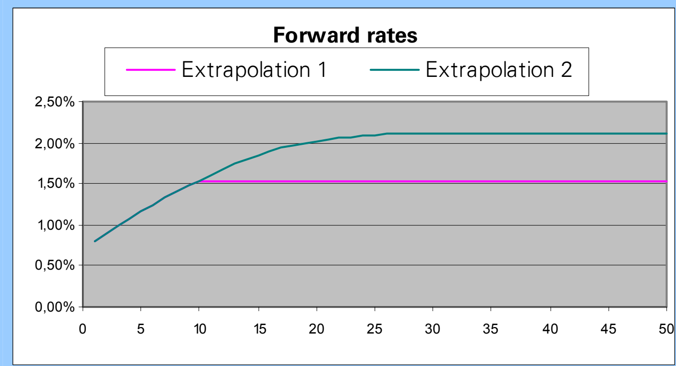
We will consider two simplified products. Product A is product with premium payments in the first 10 years and benefit payments in the remaining 40 years. The second product is a single premium product with benefit payments throughout the coverage period of 50 years.





Interest rates

The risk-free rate is observable for fixed interest bonds with maturity up to 10 years. Beyond 10 years we have applied two different extrapolation methods.



Impact on MVL

The impact on MVL is summarized in the table below.

	MVL method 1	MVL method 2	% change
Product A	88	-47	-153%
Product B	1392	1258	-10%

The (relative) interest rate sensitivity of the regular premium product is much higher than that of the single premium product. Even for the single premium product the impact on MVL is 10%. This illustrates the potential importance of the choice of extrapolation method.

3. Principles of Extrapolation

The following four principles apply to the extrapolation of all types of market data in scope.

3.1. Principle 1: Use all relevant observed market data where available

This principle is fundamental to any market consistent valuation. If relevant observed market data would be ignored then objectivity of the valuation would be compromised and the cost of the replicating portfolio would differ from the calculated MVL. Market data is considered relevant when

a. the market instrument has characteristics that are similar to a component of the liability that is being valued; and

b. the data is from a liquid market; i.e. market participants can rapidly execute large-volume transactions with little impact on the prices of the financial instruments used in the replication; or – based on expert judgment – allow to include market data that does not meet the strict liquidity requirement, but is still considered to be a good reflection of the cost of the replicating portfolio; and

c. the market data is reliable; trade and quote information of these prices is supplied by third parties and accessible and verifiable (e.g. multiple providers with similar price) to market participants.

3.1.1. Considerations when determining the last observed liquid data point

The liquidity and reliability of market data is typically lowest for market data points with the longest tenor. Hence it is critical to determine the last observed liquid data point. A number of considerations should be included in deciding where to stop using market observed data and start the extrapolation.

Consider the actual investments of insurance companies

Especially in less developed markets there is often no deep and liquid market by European standard. However, many insurance companies have the available longer tenor local government bonds in their asset portfolio. These bonds, while typically not very liquid, can give a good reflection of the cost of the replicating portfolio and should therefore be considered in building the risk free curve.

Consider quantitative measures

Quantitative measures, like bid-ask spread, trading volume or frequency can be used to gain insights in the liquidity of a market data point. However it is not straightforward to arrive at a consistent rule-based approach that works for all economies. In the appendix we show a table with bid-ask spread of swaps. It is not feasible to say that e.g. only points within a spread of 5bp should be included as this would exclude all swap points in some currencies. The other way around one could consider developing criteria like: all swap point with a spread of 5bp or less should be included, which would not necessarily exclude points that do not meet this criteria.

Data can be particularly helpful to detect trends in liquidity for a data series in a specific economy. Data can also be used as a warning limit. For example, review would be warranted if trades in a specific data point occur with less than daily frequency.

Another example is shown below. The observed spread between the 50 year and 30 year swap in the EUR market became extremely volatile during Q4 2008 and Q1 2009 which could point to decreased liquidity in the EUR markets, which for example was not observed in the USD market.

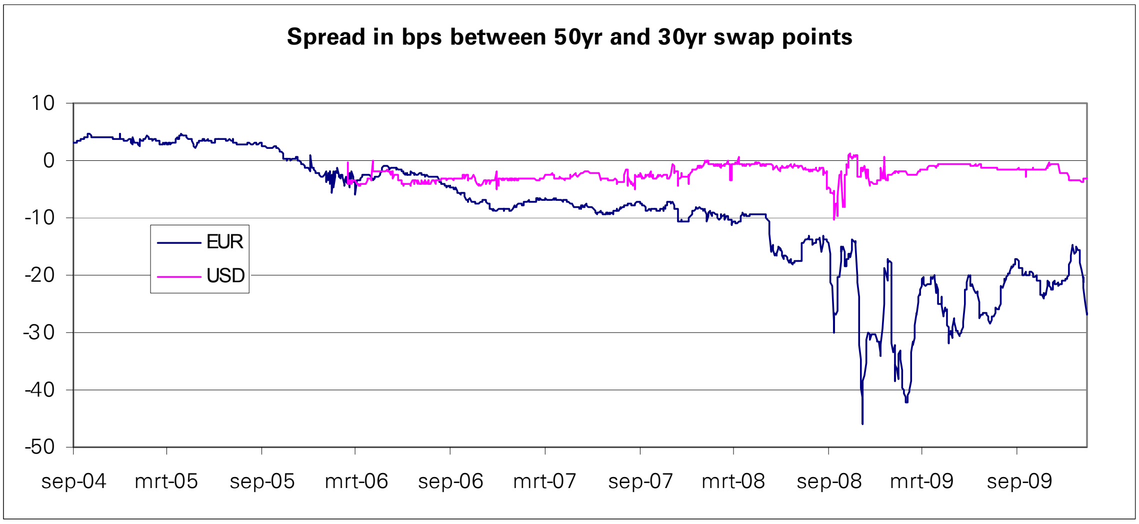


Figure 2 Spread in bps between 50 year and 30 year swap rates

Consider the circumstance under which the market price was arrived at

Market data that is biased because of a forced liquidation or distress sale should not be directly used. We refer to – and support – the work [5] by the IASB Expert Advisory Panel in this regard. Below we include a direct copy of some of their advice.

"Even when a market has become inactive, it is not appropriate to conclude that all market activity represents forced transactions. However, as noted above, an entity does not conclude automatically that any transaction price is determinative of fair value. An entity considers all available information, but does not use a transaction price when there is evidence that the transaction was forced. Determining fair value in a market that has become inactive depends on the facts and circumstances and may require the use of significant judgement about whether individual transactions are forced. Any transaction determined to be forced does not form part of a fair value measurement.

An imbalance between supply and demand (for example, fewer buyers than sellers) is not always a determinant of a forced transaction. A seller might be under financial pressure to sell, but it is still able to sell at a market price if there is more than one potential buyer in the market and a reasonable amount of time is available to market the instrument.

Indicators of a forced transaction might include, for example:

a. a legal requirement to transact, for example a regulatory mandate.

b. a necessity to dispose of an asset immediately and there is insufficient time to market the asset to be sold.

c. the existence of a single potential buyer as a result of the legal or time restrictions imposed.

[5] See section 17 to 25 of "Measuring and disclosing the fair value of financial instruments in markets that are no longer active", IASB Expert Advisory Panel, October 2008.

However, if an entity sells assets to market participants to meet regulatory requirements, the regulator does not establish the transaction price and the entity has a reasonable amount of time to market the assets, the transaction price provides evidence of fair value. Similarly, transactions initiated during bankruptcy should not automatically be assumed to be forced. The determination of whether a transaction is forced requires a thorough understanding of the facts and circumstances of the transaction."

Consider market best practices

The CRO Forum has conducted a survey among its members to identify market best practices in determining the last market data point to be used. The results, based on year end 2009 data, show that member companies have made rather consistent choices on where extrapolation should start. We do also point out that liquidity was significantly lower in a number of currencies during the 2008-2009 crisis period. Industry consensus of liquid market data throughout the crisis shows e.g. that market data in EUR and USD was only liquid up to 30 years. The graph below summarizes the findings for interest rates in some key currencies. A summary for more currencies can be found in appendix 2, which also includes an overview of industry consensus on liquid tenors throughout the 2008-2009 crisis period (based on QIS5 submission to European Commission).

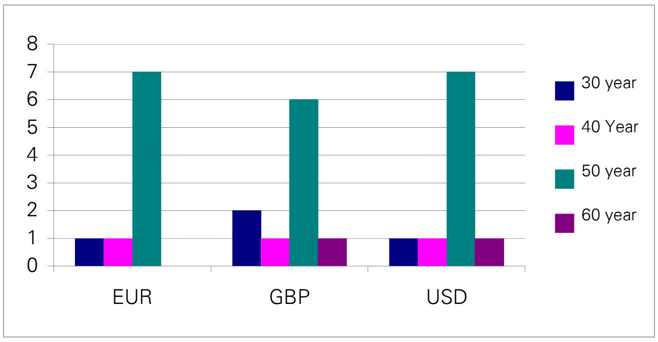


Figure 3 Number of CRO Forum members using a specific maturity as the last observed liquid market data of the interest rate curve per 31 December 2009[6].

Consider the market data beyond the last observed liquid market data point

Once a preliminary choice has been made on what the last liquid market data point is, and the extrapolation has been performed, it is advised to compare the extrapolated data points to the observed data points that did not get included. For example, if a data point was not included because it was deemed to be biased because of a forced liquidation or distress sale then it would be expected that the extrapolated data point would reflect a higher price. If this is not the case then one should re-asses the decision to exclude the data point or review the extrapolation parameterization. Another example would be if a data point is excluded because the bid-ask spread is too wide. If the extrapolated data point would not fall in between the bid-ask spread one should try to understand why this is the case and potentially revise the decision to exclude the data point or review the extrapolation parameterization.

[6] One CRO Forum member has lowered the last liquid interest rate data point for EUR, GBP and USD to 30 years after the survey was conducted.

Consider extrapolation using multiple data sources

Another issue related to the relevance of observed data is when the situation occurs in which two different instruments are considered relevant. In particular for interest rate extrapolation one could be in a situation in which the preferred instrument (e.g. swap) is no longer liquid for a certain term, but that another instrument (e.g. government) is still considered relevant. We argue that in these situations multiple instruments could be used to develop the observable part of the market data. For the case of interest rates we will give practical consideration in section 4.3 on how this could be done.

Consider stressed market conditions

The considerations to identify the last liquid market data point would not change under stressed market conditions. However, the number of data points that would meet the conditions set out above would typically be smaller.

Consider OTC markets

Expert judgment is especially relevant for option markets where option prices and implied volatilities are often based on OTC (Over-The-Counter) prices. OTC markets can be very liquid, but it is difficult to objectively determine to what tenor such markets are considered liquid.

3.2. Principle 2: Extrapolated market data should be arbitrage-free

Arbitrage opportunities should be avoided to ensure that the best-estimate is appropriate. For example, when the interest rate curve is extrapolated based on the last observed liquid spot rate this implies in most cases that you observe an abrupt change in slope of the curve, which implies a jump in the instantaneous forward interest rate. Any arbitrage free extrapolation therefore should have a continuous curve plus a continuous forward curve. Extrapolating based on forward rate achieves this by construction and is therefore a good starting point for any extrapolation method.

In this context the interpolation method used in bootstrapping the observed part of the swap curve is also very relevant as this can have significant impact on the value of the last observed liquid forward rate. Therefore, it is important to have consistency between interpolation and extrapolation when moving from observed market data into the extrapolated part of the curve.

When combining multiple sources of market data to construct a risk free curve it is also important to ensure this is done in an arbitrage-free manner. When there is a basis spread between such sources of market data (for example government swap spread) then the extrapolation method should also clarify how this basis spread is extrapolated.

Arbitrage-free extrapolation is also very relevant for volatilities. A simple method that extrapolates linearly from the last observed liquid spot market implied volatility towards a long-term volatility level can result in illogical patterns of implied forward volatilities and even negative forward volatilities in some circumstances. For example combining a 5-year spot volatility of 20% with a 10-year spot volatility of 10% would result in a negative forward volatility between year 5 and 10.

Well-known models to fit market implied volatilities either model volatility as a stochastic process (e.g. Heston model) or apply parametric functions to describe the volatility process (e.g. SABR) assuming in both cases continuous forward volatilities. Extrapolating forward volatility results by construction in arbitrage free market data and is therefore a good starting point for any extrapolation method.

Arbitrage-free should also be interpolated in the sense that there should be consistency in extrapolation between highly similar markets. For example, if the observed market data for implied volatility in a particular index (e.g. Nasdaq 100) is consistently higher than the observed market data in another index (e.g. S&P500) and both indices are highly correlated, then you would also expect that the extrapolation of both indices to ensure this property is maintained in the extrapolation.

3.3. Principle 3: The extrapolation method should be theoretically sound

It is important that the chosen method is theoretically sound. Extrapolation methods that are in violation of generally accepted economic theory or clearly observed historical patterns should be avoided.

The extrapolation method should not result in spurious volatility. Volatility arises from genuine changes in expected levels of long-term best-estimates and changes in risk premia. Thus the results of the extrapolation method should be consistent with the observed patterns of market data. For example, realized interest rate volatility is typically declining with the tenor of the rate and the result of the extrapolation method should be consistent with that.

In general implied volatilities are used to calibrate theoretical models for the underlying market risk factors. E.g. swaption volatilities are used to calibrate stochastic interest rate models (e.g. Libor market model or extended Vasicek). These models are then used to value embedded insurance guarantees and options based on these market risk factors. Such models typically try to capture well described properties of market data such as mean reverting behavior. Therefore such theoretical models provide a solid basis for any extrapolation method. Moreover, by construction such models are arbitrage-free.

Theoretical models also imply certain features for long-term levels. There is for example a lot of literature showing that impact of convexity in long-term forward interest rates [7]. Another example is that for long-term equity options it can be shown that interest rate volatility can have a significant impact on the pricing and therefore the implied volatility (which is typically used based in combination with the assumption of deterministic interest rates).

[7] A well known article in this context is “Interest rate volatility and the shape of the term structure [and Discussion]” by Brown, Schaefer, Rogers, Metha and Pezier, 1994 The Royal Society.

3.4. Principle 4: The extrapolation should follow a smooth path from the entry point to the unconditional ultimate long-term level

The view on the long-term level should be forward looking, reflecting market analysis, future expectations, economic theory and historic experience. This view should be documented appropriately and regularly reviewed.

Two extrapolation methods can be distinguished:

1. Grading from the last observed liquid market forward data point to a long-term ultimate level;
2. Keeping the last observed liquid market forward data point constant for the entire extrapolation.

In the first method a long-term ultimate level of the market forward data is set and an appropriate grading method is chosen for the area in between the last observed liquid market forward data point and the ultimate level. This method requires three key assumptions:

a. **What is the ultimate forward level?** The view on long-term level should be forward looking, and based on market analysis, future expectations, economic theory and historic experience. This view should be documented appropriately and regularly reviewed.

b. **When is the ultimate forward level reached?** This should be based on an analysis of when the predictive power of the currently observed market data on the future level of the market data becomes negligible. The forward level is kept constant once the ultimate level has been reached.

c. **What is the grading process between the last observed liquid forward data and the ultimate level?** The shape and speed of the grading process to the long-term ultimate level should comply with principle 2 and 3; i.e. it should be arbitrage-free and based on sound economic theory and relevant historic data, e.g. from other currencies with a deeper market.

In the second method the simplifying assumption is made that the ultimate forward level is equal to the current level. In general this method can work fine when the extrapolation is limited in length. This method may not be appropriate in cases where the liabilities are long and the observable market is much shorter.

Principle #4 requires extrapolation to an ultimate long-term level. The observed market data which forms the basis for the extrapolation is regarded as including an implicit risk margin. This is particularly the case where there is a natural mismatch in supply-demand as for the longer tenors in the swap curve. Consequently, the extrapolated part of the curve that results from it will also include an implicit risk margin. When the ultimate long-term level is based on market data and economic theory as set out in this paper, the extrapolation is consistent with the observed data. If a risk margin were to be excluded from the ultimate long-term level, then the extrapolation would have only a partial risk margin embedded in the extrapolation. Excluding the risk margin completely from the extrapolation is not possible as it would violate principle 4 (smooth path). Adding an explicit risk margin (i.e. first calculate present value of cashflows, next calculate a risk margin) would represent double counting of risk and be highly impractical. In the box on page 15 the relationship between extrapolation and risk margin is described.

We will describe these issues and implementation of the principles in more detail for each of the types of data in the next chapters.

RELATIONSHIP BETWEEN EXTRAPOLATION AND RISK MARGINS

One can distinguish three types of extrapolation:

1 A simple direct extrapolation of the market price; it is assumed that risk margins embedded in the market price stay constant in the extrapolation.

2 An advanced direct extrapolation of the market price; in this method the components of the market price are extrapolated based on a macro-economic view of the long-term market price (i.e. not necessarily the same as in the last observed liquid market data).

3 Start with an extrapolation to a best-estimate long-term level; i.e. expected “real world” realization of interest rates or volatilities. This does not include compensation for all risk; a risk margin needs to be calculated separately and included in the MVL.

The first method can be comfortably applied as long as the extrapolation is of limited length. Say, the market is liquid up to 40 years, but some of the liability cashflows go out to 45 years. Major deviations between the forward rate in year 40 and 45 are highly unlikely and would have a negligible impact on the MVL anyway. This is also evidenced by the results in Figure 4. Investment banks would typically apply such an approach to make a market. However, unlike insurance companies, they tend not to take material amounts of non-hedgeable financial risk and hence this method suffices for them. Even if one would try to build in a risk margin, this would turn out to be negligible since one could hedge the position quite accurately with the 40 year instrument and the little remaining risk would be drastically reduced due to the discounting effect over 40+ years.

Now consider the same product in a situation in which we have observable interest rate forwards for only the first 5 years. This creates a number of complications:

- Yield curves tend to be steep on the shorter end of the curve. This could result in extrapolation to extraordinary levels
- Yield curves tend to be more volatile in the shorter end of the curve. This could result in extraordinary volatility in the extrapolated curve.
- The fact that the liquid curve ends at 5 years indicates that there is no natural supply of long dated fixed income. Any rational player who would offer a 45 year guarantee would be exposed to a very large amount of risk and would want to be compensated for this in excess of the best estimate level.

For this type of situations it is desirable to use the second method, as it allows to take into account the long term nature of the product.

The third method is extremely difficult, if not impossible, to apply accurately. Since the last observed liquid data point includes a risk margin - but the long-term level does not - the extrapolation will consist of a partial risk margin. The remaining risk margin will need to be captured separately. To make this method work would require a significant amount of assumptions that will be difficult to verify.

Where should one draw the line where the first method can still be applied and where the second method above should be applied? In practice, this question may not be so relevant. Once a more sophisticated method (i.e. method 2) is developed it can just as well be applied to liabilities that have limited exposure to extrapolated market data.

4. Extrapolation of interest rate curve

4.1. Solvency II taskforce

For the extrapolation of interest rate curve 5 additional principles have been developed by the Solvency II taskforce on the illiquidity premium. These principles are supported by the CRO Forum:

a. The extrapolated part of the basis risk free interest rate curve should be calculated and published by a central EU institution, based on transparent procedures and methodologies, with the same frequency and according to the same procedures as the non extrapolated part.

b. Extrapolation should be based on forward rates converging from one or a set of last observed liquid market data points to an unconditional ultimate long-term forward rate to be determined for each currency by macro-economic methods. Methods can take differences between currencies into account. The principles used to determine the macro-economic long-term forward rate should be explicitly communicated.

c. Criteria should be developed to determine the last observed liquid market data points which serve as entry point into the extrapolated part of the interest curve and for the pace of convergence of extrapolation with the unconditional ultimate long-term forward rate.

d. Techniques should be developed regarding the consideration to be given to observed market data points situated in the extrapolated part of the interest curve.

e. The calibration of the shock to the risk free interest rate term structure used for the calculation of the SCR should be reviewed in order to be compatible with the relative invariance of the unconditional ultimate long-term forward rate.

The benefit of principle A is consistency and comparability between companies; see section 4.5.5. Guidance on principle B can be found in section 4.5. Considerations on C and D are included in section 3.1.1. We believe that it is difficult to develop easy to apply “criteria” (C) and “techniques” (D) and rather refer to “considerations” that need to be taken into account for determining the last observed liquid data point. Some level of expert judgment will ultimately be required to determine the last liquid market data point based on such considerations. Key is that the market data used should be “a good reflection of the cost of the replicating portfolio”. Principle E is discussed in section 4.5.5.

4.2. Extrapolation methods

From the principles it is firstly derived that any extrapolation method should be based on extrapolating forward rates.

> **BEST PRACTICE**
>
> It is best practice to apply the extrapolation to continuously-compounded forward interest rates.

From the last observed liquid forward rate, the extrapolation should follow a path towards a robust long-term level. We consider two methods for setting this long-term best-estimate level.

More precisely the second method is a special case of the first method and can be appropriate in certain circumstances. The two methods for setting the long-term forward rate level are:

- Macro-economic method: A long-term market forward rate is set based on market analysis, future expectations, economic theory and historical experience. Below the components of this long-term best estimate forward rate are discussed in detail.
- Constant forward rate: The last observed liquid forward rate derived from market data is assumed to be kept constant beyond the last observed liquid market data point. Below it is discussed how this last observed liquid forward rate should be determined. This method is a special case of the first method with the long-term market forward rate set equal to the last observed liquid forward rate.

The second method is only appropriate when the available market data is sufficiently long compared to the cash flows that need to be discounted. We performed tests for a number of currencies and actual liability portfolios. To illustrate that this ultimate level is especially relevant for markets where limited market data is available we show the absolute valuation difference in the Expected Present Value of Liability Cash Flows for actual portfolios of CRO Forum companies based on the above two methods for a number of currencies. It can be clearly seen that the range in valuation start to deviate significantly when the available market data becomes shorter. For most markets, it seems that 30 years of market data is sufficient for using the second method.

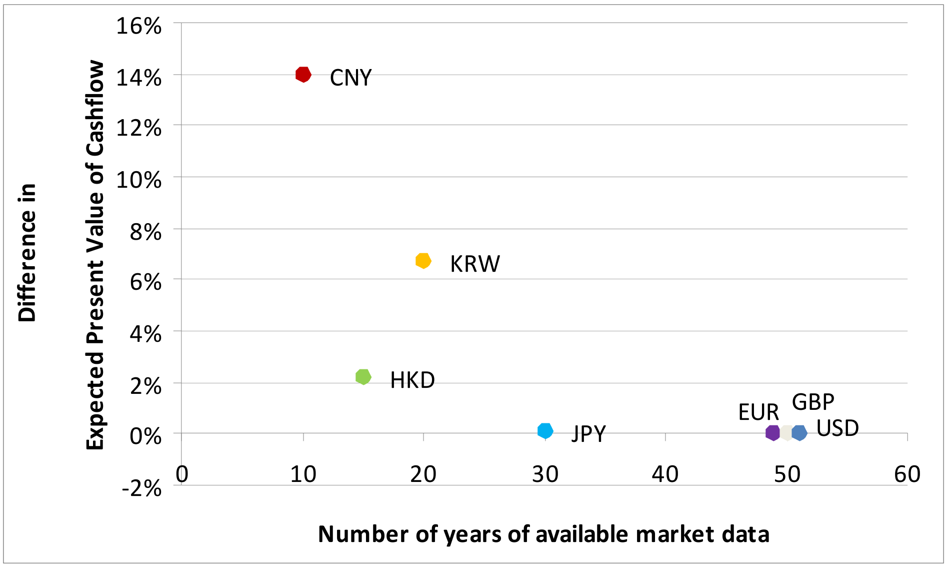


Figure 4[8] Absolute difference in Expected PV of Liability Cash Flows between a simple (constant forward rate) and advance (macro-economic) extrapolation. Information provided by CRO Forum working group members

We also note that there is a spectrum of feasible approaches in between the two methods described above. A moving average of long-term forward rates could for instance be used as a practical basis for estimating the ultimate rate in the macro-economic method.

Both methods start their extrapolation based on the last observed liquid market forward rate and therefore the implication for risk management is that interest rate risk can be neutralized through positions mimicking this last observed liquid forward rate. However, we would like to stress that there remains exposure to the forward rates beyond the last observed liquid market data point that cannot be fully hedged.

[8] Points depicted in diagram depend on chosen portfolio

BEST PRACTICE

We consider the macro-economic method "best practice", as it is appropriate irrespective of the amount of market data available.

4.3. Determining the last observed liquid interest rate data point

The general considerations as laid out in section 3.1.1 apply for determining the last observed liquid interest rate data point. Specific for interest rates we will discuss in more detail the use of multiple data sources and what to do with market data points that are beyond the last liquid market data point.

4.3.1. Using multiple data sources

The CRO Forum recommends the use of swap rates. However, if other relevant data (typically government curve) has availability beyond swap then this should also be used.

In most markets the swap market is available for the same or longer tenors than government bonds. E.g. in the EUR and USD market the swap market currently goes up to 50 years where government bonds are mostly only available up to 30 years. However, in less developed markets the swap market is not that developed yet. For example Taiwan (swap up to 10 years, government bonds up to 20 years), Malaysia (swap up to 15 years, government bonds up to 20 years) and Thailand (swap up to 15 years, government bonds up to 30 years). The question is how these extra market data points can be included in the risk free yield curve while maintaining consistency with the principles of yield curve extension; In particular avoiding discontinuities in the curve that could result in arbitrage opportunities. One should also take into account that both sources of market data (e.g. swap and government bonds) show a difference (spread) for the observed part of the curve.

The best practice would be to add additional raw points to the curve before a continuous compounded yield curve is bootstrapped. An example of how such an extra point could be created (using Taiwan swap curve as an example):

TWD 20yr swap point = TWD 20yr Gov. Bond yield + 10 yr Swap-Gov Spread.

= TWD 10yr swap point + TWD 20yr Gov. Bond yield – TWD 10yr Gov. Bond yield

The new swap point is therefore equal to the last observed liquid swap point plus the observed slope in the Government bond curve. After constructing this additional swap point, the continuously compounded yield curve can be bootstrapped in a standard manner including this additional constructed swap point.

Given the importance of the slope of the forward curve as the starting point for the extrapolation one has to be careful in fitting the slope between two market data points based on a different source, e.g. swap vs government bond (see section 3.1.1.2).

As long as the swap and government curves themselves are arbitrage free this combination will also result in an arbitrage free continuous yield curve.

Note that in the example we have chosen to keep the last observed spread between swap and governments constant. A more refined method could be developed in which the extrapolated spread increases or decreases with maturity. However, often there is not sufficient market data to support a more refined method.

4.3.2. Considerations for using interest rate data points that are not fully incorporated

It is important to realize that there is not a very clear line between market data points that are liquid and should be fully incorporated in the observed part of the curve and market data points that are not liquid and should not be taken into account. Below a few considerations on how to deal with data points before and after the last liquid market data point.

The slope of the last observed liquid forward rate should be relatively stable

The last observed liquid market data points are relatively important as they determine the shape of the forward curve which is the starting point for the extrapolation. A last liquid data point that is trading slightly out of sync with its previous market data point can result in spurious volatility in the extrapolated part of the curve. Therefore it should be feasible to deviate from the last liquid market data point in fitting the curve to ensure a stable slope of the forward curve. Such deviations should be typically limited to the bid-ask spread of such points to avoid arbitrage opportunities.

Market data points not included in the observed part of the curve.

There might be swap points or government bond yields that are not (yet) considered to be liquid enough to be included in the observed part of the curve. Such point should give a sensibility check on the slope of the extrapolated part of the curve. While it is not necessary that the extrapolation stays within the bid-ask spread of such points, they should not be substantially out of sync with each other. Ideally it should be understood why they deviate. Lastly, once such points become more liquid they should ideally converge such that including these points in the observed part of the curve should not result in a big change in the curve.

Importance of market data cut-off tenor in stressed market conditions

The extrapolation of market data has an important role in avoiding too much pro-cyclicality in the Solvency II framework. A solvency regime based on market values, already has build in mechanisms that in times of stressed market conditions there is an extra tendency to de-risk and therefore put extra pressure on financial markets, which could worsen again the solvency on a market value basis. The nature of in particular life insurance companies is such that in general their liabilities have longer durations than the available assets in the markets.

As already referred to this results in supply-demand pressure on interest rates for longer tenors. In a crisis situation where both insurance companies and pension funds try to de-risk such pressure can result in unbalanced markets. A very clear example of this is the period end-2008 to mid-2009 in the EUR interest rate market, where interest rate forward rates plummeted. Similar issues also arose in the equity implied volatility market with forward volatilities shooting sky-high. The role of extrapolation in such situations is to ignore market data points that are subject to significant unbalanced market and promote stability in liability valuation to avoid that additional pro-cyclical effects worsen the solvency position of insurance companies.

How can this be achieved? Firstly, the fact that the long-term unconditional forward rate is set in a stable manner and is not impacted by economic cycles is a good basis condition. However, extrapolation starts from the last observed liquid market data forward rate. So any instability in such point is extended to longer tenors. The mechanism that provides stability is that in a crisis situation the transition point of where market data is used and where extrapolation starts is moved to an earlier point. This is also in line with the principles set out in this paper as long-tenor swap points in such market conditions and unbalanced demand/supply would be assessed as being illiquid. So while in liquid times the EUR swap market would be considered somewhat liquid up to a long term, in a crisis such as end-2008 this would reduce to 30 years. The stable long-term forward rate will then ensure enough stability in the remaining tenors.

How can such an unbalance be observed? There are two elements that could indicate an unbalanced market. Firstly an interest rate forward rate that is significantly steeper downward sloping then pre-crisis and secondly, the fact that long-term swap rates dropped significantly below traded AAA government bonds in times of crisis.

Special attention has to be paid in times of changes in regulation. In particular markets have to be watched closely around the introduction of SII and IFRS phase 2 whether short term hedging activities of insurers impact the long end of the curve temporarily and therefore markets have to be regarded as distorted.

4.4. Calculation of the last observed liquid forward rate

It is considered best practice to base the last observed liquid forward rate on the last two liquid market data points. It is of crucial importance how the yield curve is being bootstrapped as this impacts the forward rate between the last two observed liquid market data points. This is shown with below simplified yet realistic example.

Assume that the yield curve is based on below five swap points. These points are quoted on an actual/actual basis with annual payment frequency.

1	5	10	15	20
3.00%	3.40%	3.80%	4.10%	4.30%

We bootstrap this swap curve based on two methods:

1 Assuming a piece-wise linear zero rate curve: Zero rates are fitted for the 1, 5, 10, 15 and 20 year tenors and rates in between are found with linear interpolation.

2 Assuming a piece-wise linear forward rate curve: 1 year forward rates are fitted for the 0, 4, 9, 14 and 19 year points and forward rates in between are found with linear interpolation.

Both methods result in nearly identical zero curves with differences of below 1bp for all tenors. However, both methods result in significant differences in 1 year forward rates as shown from below graph. Moreover, the ultimate forward rate differs by 14 bps. The second method is considered best practice.

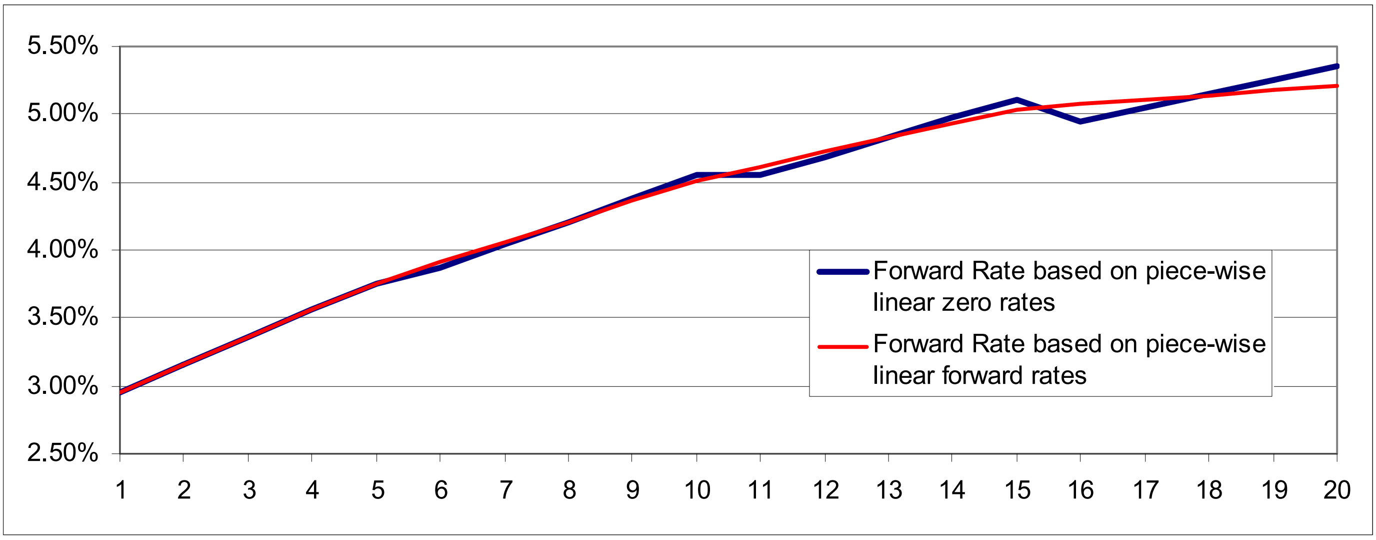


Figure 5 Two methods to interpolate the yield curve

An alternative method for fitting the existing data points as seen in practice is fitting a spline curve which simultaneously fits the data points. This method is also considered best practice, but is more difficult to implement and can give weird results if data points are not fully consistent. We therefore prefer the simpler linear piece-wise forward rate method as it can be implemented more easily.

4.5. Extrapolation towards ultimate forward rate

Best practice method is to set ultimate forward rate based on an estimate of long-term economic environment. Forward rates will be interpolated between last observed liquid forward rate and ultimate forward rate. Two principles have been developed on how this ultimate long-term forward rate should be set:

1 The unconditional long-term forward rate should be relatively stable over time and only change due to fundamental changes in long-term expectations and should not be affected materially by short term economic changes.

2 The unconditional long-term forward rate is the sum of forward looking expectations of the:

- i. Long-term Inflation Level
- ii. Long-term Real Rate
- iii. Long-term Term Premium
- iv. Long-term Convexity Adjustment

The estimation of these components should be forward looking, and based on market analysis, future expectations, economic theory, historical experience and should reflect a risk margin already.

In case one component is not reflected in the extrapolation it should be at least reflected within an additional risk margin - as this introduces additional complexity in some cases it's not our recommendation but feasible as long as the risk margin is set in a consistent manner with the long-term forward rate.

4.5.1 Components of ultimate rate

These components of the unconditional long-term forward rate defined by above principle can be estimated individually or can be estimated together. The following sections describe best-practice methods to estimate these components.

4.5.1.1 Long-term real forward interest rate

We consider 3 methods for determining the real (forward) interest rates:

1) Historical average
2) Derived from nominal forward rates and inflation swaps
3) Directly observed in the market

Historical Average

This method for determining real interest rate levels is to look at historical averages for real cash returns constructed as the yield on a short term interest rate less CPI inflation. Work by B&H[9] shows that this results in a range of outcomes for various developed countries (e.g. 0.6% for Japan versus 1.4% for US and 2.2% for Germany) and the median estimate is set at 1.8%. These estimates are based on a 1970-2007 reference period. Main drawback of this method is that is it backward looking rather than forward looking and assumes that historical observations are representative for the (far) future.

Derived from nominal forward rates and inflation swaps

This method looks at real forward interest rates implied by nominal forward interest rates and inflation swap prices. Since inflation swaps are available up to 30 years, it is feasible to estimate the forward real rate at the 30 year point. Looking at the reference period 2006-2008 this method would have resulted in real rates around 2.5% for both EUR and USD. During the 2009 crisis the implied real rates for both currencies dropped significantly to around 1.0% and since then recovered to around 2.0%.

The advantage is that this method is forward looking. However, the cost of this is more instability. This could be solved by applying some averaging over time.

Directly observed in the market

A third alternative is available for some countries only. In many Latin-American countries, both nominal and real interest rate curves are directly quoted in the market. So the real forward rate can be directly observed in the market for significantly long tenors. Examples include Mexico (UDI curve up to 30 years), Chile (CLF curve up to 20 years) and Colombia (UVR curve up to 20 years). The real rates in these markets are typically somewhat higher than those seen for EUR and USD, at around 3.0% to 4.0% as investors demand a higher real return in these markets to compensate for more inflation uncertainty.

---

[9] "Interest rate calibration, how to set long-term interest rates in the absence of market prices", Steffen Sorensen, B&H, Financial Economic Research, Version 1, September 2008, Pg 6.

The choice for one of these three methods (or a combination) will depend on the availability of data. If sufficient data is available then all three methods should give a roughly similar extrapolation of the forward rate. Whether this will be the case in practice depends on the speed of convergence to the long-term expected forward rate.

BEST PRACTICE

Both "market implied" (method 2 & 3) and historical (method 1) should be evaluated in estimating the real forward rate. These methods are not mutually exclusive since historical data of "market implied" real rates is available. Key is that enough historical data (at least one economic cycle) is used to ensure stability in the long-term forward real rate and avoiding that the short-term economic cycle impacts the estimation process. It should also be avoided that too much historical data is used to avoid that structural changes in the real rates are not reflected in the estimation.

4.5.1.2. Long-term expected Inflation

We consider 3 methods for determining the long-term expected inflation:

1 Weighted average of historical inflation and inflation target

2 Weighted average of economic forecast estimation per currency

3 Derived from inflation swaps/bonds.

Weighted average of historical inflation and inflation target

Historical analysis as shown by B&H [10] highlights a large discrepancy in average inflation between different countries. To combat the high level of inflation in the 1970s and 1980s, a number of central banks now operate with an explicit target for inflation. The level of such inflation targets tends to be around 2%. Historical inflation averages are rather high compared to this target level, while more recent history is quite close to this level. The question is therefore whether we believe that central banks would credibly succeed in keeping inflation at low and stable levels in the very long-term. B&H explicitly summarizes this dilemma in three ways to set the long-term expected inflation:

	Method	Description
1	No prior on inflation target credibility	Compute an estimate for long-term inflation expectations using the exponentially weighted average of CPI inflation
2	Strong prior on credible inflation	Attach a weight of x to an inflation target of 2% for all economies and (1-x) to the exponentially weighted average of historical CPI inflation. The closer x is to one, the higher a prior on the inflation target.
3	Credible inflation target	Impose long run inflation expectations of 2% at any point in time across each of the economies.

[10] "Interest rate calibration, how to set long-term interest rates in the absence of market prices", Steffen Sorensen, B&H, Financial Economic Research, Version 1, September 2008, Pg 8.

B&H opt for the second method and apply an 80% weight on the 2% inflation target and a 20% weight on the exponentially weighted historical average. This results in a 2.4% long-term expected inflation for all countries. We believe it makes sense to differentiate the inflation target by currency. For example the Norwegian central bank has an inflation target of 2.5% while most other central banks have a lower target. In setting the inflation target one should also consider the credibility of such a target if the actual inflation level consistently over or undershoots the target set by the central bank.

Weighted average of economic forecast estimation per currency

The disadvantage of above estimate is that it assumes that inflation expectations will be the same for all countries. Historically there are significant differences over longer periods in inflation levels in different countries. Therefore it can be argued to use a different inflation forecast for different (blocks of) countries. One good source of inflation expectations is a semi-annual survey of Consensus Economics which includes a wide range of countries and sources (both banks and economic institutions) and predicts inflation forward up to 10 years. The question is how representative this inflation forecast is for the very far future.

One could use this forecast to derive the expected inflation levels going forward. Below table shows for some example currencies the average 10 year inflation forecast over the last 2 years (average last four surveys). The 2.0% inflation target is credible for hard currencies such as EUR, SEK and CAD. Some countries with less explicit inflation targets have higher inflation and for countries like Japan and Taiwan the forecast is persistently below 2.0%. We note that forecasts are broadly consistent with inflation targets set by central banks (e.g. EUR 2.0% and NOK 2.5%).

Europe		Americas		Asia Pacific	
EUR	2.0	USD	2.3	JPY	1.2
GBP	2.3	CAD	2.0	KRW	2.5
NOK	2.4	MXN	3.8	TWD	1.8
SEK	2.0	BRL	4.1	AUD	2.6

In summary, this method applies a separate inflation forecast for each currency based on a moving average of Consensus Economics forecasts. One exception might be for the currencies in Central Europe. Since these currencies are expected to sooner or later join the EUR, it makes sense to apply the EUR inflation forecast to set the long-term forward rate.

Compared to the previous method, which assumes one inflation target for all currencies based on historical data, this method takes the other extreme by setting a country specific inflation target based on (relatively) short-term projections. The ideal set up might be in the middle taking a weighted average of a generic long-term target inflation and a country specific projection.

Derived from inflation swaps/bonds.

Inflation forecast can also be derived from traded instruments. Most specifically one can use inflation swaps which exchange the actual observed inflation against a fixed coupon. Based on such instruments one can predict the inflation up to 30 years. Drawback is that such inflation swaps are only traded of a limited number of currencies such as EUR, USD, GBP and JPY. An alternative is to derive inflation from inflation linked bonds or the difference between nominal and real interest rates such as quoted in many Latin-American countries (e.g. MXN versus UDI rates).

The current levels of the 30-year inflation swap quotes for EUR and USD are respectively 2.4% and 2.8%. There is evidence that inflation swaps are trading at a premium of about 40 bps over actual inflation. This is consistent with the 'difference' versus the 'Economic Consensus' forecasts. This method can be used in combination with the previous method and can be used instead of a Consensus Economics forecast.

E.g. a weight of 50% to the inflation target and 50% to an average of 2 years of survey data would results in below long-term inflation levels for some example currencies.

	Inflation targe	Avg Survey	Weight	LT inflation
EUR	2.0%	2.0%	50.0%	2.0%
GBP	2.0%	2.3%	50.0%	2.2%
JPY	2.0%	1.2%	50.0%	1.6%
HKD	2.0%	2.6%	50.0%	2.3%

BEST PRACTICE

We believe it is important to keep a balance between on one hand a generic inflation level for all countries and on the other hand the current inflation credibility of a country. Therefore we propose to calculate for each country a weighted average between the currency specific inflation target (weight X%) and a historical average of long-term forecasts (based on either survey data or market traded inflation) (weight 1-X%). The weight can differ by currency.

4.5.1.3. Term premium

The component of the long-term forward rate that is perhaps the most difficult to estimate is the long-term term premium. The basic theory of the term structure of interest rates is the expectations hypothesis. According to this hypothesis, the expected return from holding a long bond until maturity is the same as the expected return from rolling over a series of short bonds with a total maturity equal to that of the long bond. That is, the long bond yield is the average of the expected short-term rates. Though the expectations hypothesis provides a simple and intuitive appealing interpretation of the yield curve, it ignores interest rate risk and investors may require compensation for this risk. The "term premium" refers to such compensation. There are also other factors influencing term premia such as liquidity considerations and investor habitats. One example is the "flight to quality" effect in some major government securities markets at times of extreme volatility. Demand for government securities from large institution such as pension funds and insurance companies in certain market segments is an example of investor habitats.

Even though various research has been done there is little consensus in the literature on the empirical properties of the term premium. The discussion is further complicated by the existence of multiple definitions of the term premium. Three commonly used definitions are:[11]

(1) The expected return of holding a multi-period zero coupon bond for one period minus the one-period yield (short rate).
(2) The forward rate minus the expected future spot rate.
(3) The yield on a zero coupon bond minus the average expected short rates from the present to the maturity of the bond.

Most studies are based on 10-15 year bond data (mostly government bonds) and indeed show a significant positive term premium in the order of 1% to 2%, although with a big standard deviation. Work by B&H [12] estimates this term premium to be 1.7% for swap rates. Some caveats need to be made. Firstly, that term premiums show a declining trend over time as shown by Orphanides. Secondly, the question arises on how to extend the term premium beyond the 10-15 years as used in those studies. There is also literature that clearly shows that forward interest rate curve consistently decline beyond 15-20 years as shown by Brown and Schaefer (2000). We indeed observe this behaviour also consistently for all major currencies over the last 6 years with 30 year forward rates in the order of 1% lower than 20 year forward rates. The demand supply mismatch in longer tenor bonds plays definitely a role in this observed decline in forward rates and therefore term premium for those tenors. In summary, literature does not give us conclusive information on what the long-term term premium should be.

When the real interest rate level is based on the observed cash real rate then a term premium will be required to estimate the ultimate forward rate. When real rates are derived from forward looking long-term nominal rates and inflation swaps then this is not applicable.

The term premium can be thought of as a risk premium. If there would be a natural balance between demand and supply then the term premium could be close to zero. If there is no natural balance then the term premium gives the additional incentive on top of the best-estimate to either buyer or seller to reach a mutually acceptable price. In observed bond prices this is typically thought of the additional premium required to compensate an investor for locking in the money for a long time and taking the associated interest rate risk. The same argument would hold for a policyholder buying a guaranteed insurance product with term beyond what is available in the fixed income market [13]. However, if one would look at it from the perspective from the insurance company then it is the other way around. The insurance company would want to be compensated for offering a long-term guarantee it cannot hedge. This would argue for a negative adjustment to the best-estimate. Basically, one can see the bid-ask spread widening. In this framework the size – and sign – of the term premium depends on the perspective one takes: bid, ask or mid.

While using mid prices has the advantage that assets and liabilities with exactly the same cashflow profile have the same value, it does create the problem that the insurance company would not be able to earn a return on the capital it would need to hold for the non-hedgeable interest rate mismatch risk; i.e. it would value the product at a price below what it would cost to manufacture it.

[11] In this text we use multiple reference from the BIS paper of Orphanides, "The bond market term premium: what is it, and how can we measure it?"
[12] "Interest rate calibration, how to set long-term interest rates in the absence of market prices", Steffen Sorensen, B&H, Financial Economic Research, Version 1, September 2008, Pg 3
[13] This would also explain why the investor is willing to pay extra (distribution cost, profit margin) on top of the cost to manufacture an insurance product.

The impact on the risk margin calculation is another important consequence of the approach taken in the extrapolation of the term premium. One would be double counting if a bid price is modeled and in addition a risk margin would be calculated. Vice versa, if mid or ask price is modeled then an additional risk margin would be required.

The critical point here is that the risk margin for non-hedgeable market risk should be set in a consistent manner with the term premium in the ultimate long-term forward rate. Below we illustrate this based on two examples.

Consider two nearly identical extrapolations based on the principles and proposals in the best practice paper. The only difference is the assumption in the term premium for which we applied either 1.7% or 0.0%. The goal of these examples is not to claim that either value is right or wrong, but to show the need for a consistent risk margin to arrive at market consistent value of insurance liabilities.

We look at three currencies: TWD, PLN and EUR. The first two have limited market data (we use respectively 10 and 20 years of market data), while we use 50 years of market data for EUR. The graphs below show the average forward curve over the past five years (excluding Q4 2008 to Q1 2009 [14]) for TWD and PLN versus EUR based on above two extrapolations.

The first graph shows PLN, which is a candidate currency to enter the EUR at some point in the relatively nearby future.

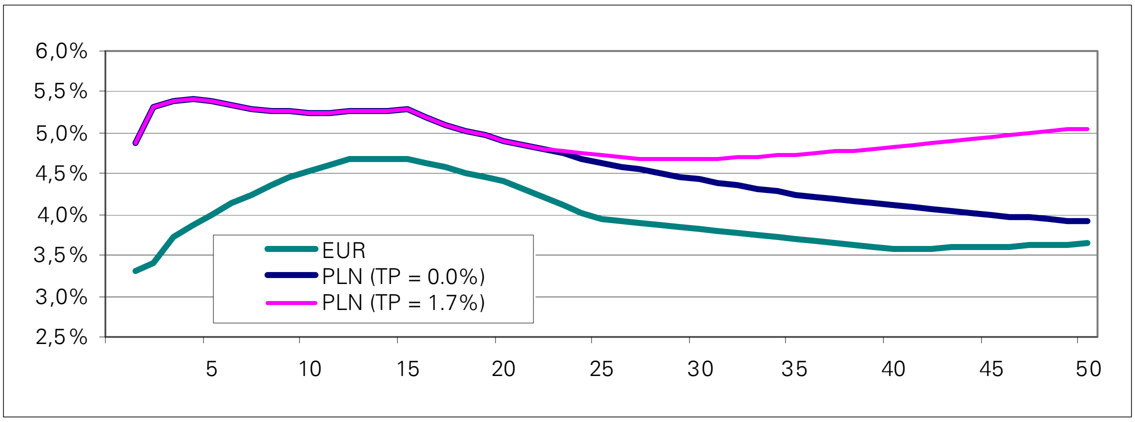


Figure 6 Average forward rate curves for EUR and PLN implied by swap curves over the last 5 years (excl Q4 2008 – Q1 2009) based on quarterly observation points.

Key point: Assuming a higher term premium in the extrapolation as implied from observed market data (i.e. EUR in above example) implies that forward rates in related markets with less market data not necessary have convergence in forward rates. To get a consistent MVL the risk margin would need to be set consistent to ensure such convergence after applying the risk margin for un-hedgeable market risk.

[14] During this period we noted that EUR forwards beyond 20 years deviated significantly from the period before and after these dates due to the credit crisis and reduced liquidity in such points.

The second graph shows TWD where forward rates have been consistently below the level seen in EUR over the last five years.

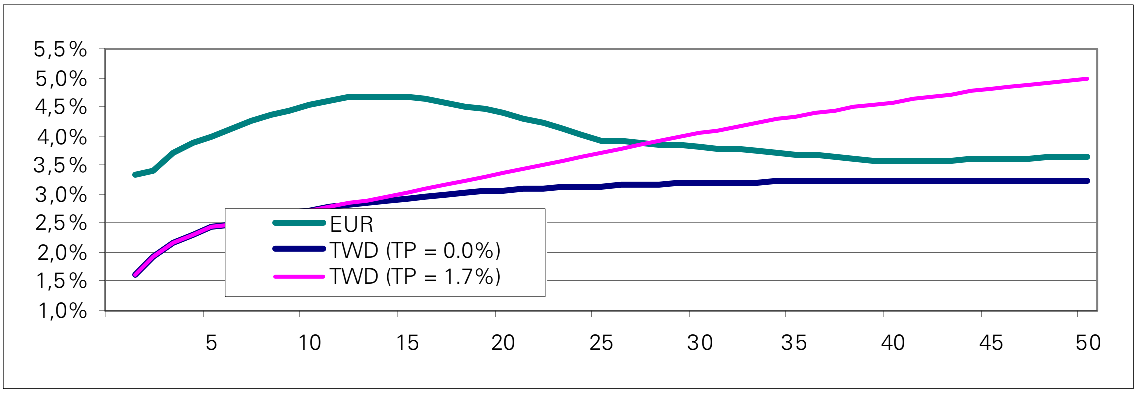


Figure 7 Average forward rate curves for EUR and TWD implied by swap curves over the last 5 years (excl Q4 2008 – Q1 2009) based on quarterly observation points.

Key Point: One would expect to see some level of consistency between the difference and shape of the observed part the curve and the extrapolated part. Assuming a higher risk premium in the extrapolation as implied from observed market data (i.e. EUR in above example) implies that forward rates can rise above the levels seen in observed markets even though the observed part of the curve has consistently been lower. To get a consistent MVL the risk margin would need to be set consistently to ensure that it is NOT more attractive to write long-term guarantees in currencies where we CANNOT hedge than in currencies where we CAN hedge our exposure.

The CRO Forum prefers to embed the risk margin for non-hedgeable market risk in the yield curve extrapolation. The size of this embedded risk margin depends on the currency, and how effective the exposure in the extrapolated part of the curve can be hedged with available instruments. This also to achieve consistency with market data point that lie in the extrapolated part of the curve, but are still relevant for risk management purposes. Refer to section 4.5.5 for a discussion on risk management considerations.

In the context of Solvency II directive, it is important to note that as the "market" term premium already includes an allowance for the risk margin for non-hedgeable market risk it is not necessary to separate the calibration between a best-estimate and charge a separate risk margin for non-hedgeable market risk. This is equally relevant for implied volatilities, which also include a risk margin at any tenor.

BEST PRACTICE

The CRO Forum prefers a term premium derived from observed market data and economic theory that therefore embeds the risk margin implicitly in the extrapolation. This allows for a natural extension from observed market data and would ensure a consistent MVL calculation across companies.

4.5.1.4. Convexity

From extrapolation principles #1 and #3 one can derive the existence of convexity of interest rates. Convexity arises when there is uncertainty about what interest rates will be in the future. The return to the investor in a fixed income security is higher when rates go down with a certain amount compared to the situation in which rates go up with the same amount. This phenomenon is called convexity.

A simplified example can help understand the convexity adjustment. Consider an investor that knows the current short-term interest is 5%. The investor knows rates in the future are uncertain, but expects rates on average to remain 5%. Furthermore, for illustration purpose we assume the investor is indifferent between an investment with interest rate risk (fixed income) and without interest rate risk (cash). The investor has a choice between investing in a 5% short-term asset and rolling it forward every time, or to invest in a fixed long-term rate. The long-term rate at which the investor is indifferent between the short-term strategy and the long-term strategy must be lower than the short-term rate, even though one expects interest rates to stay on average the same. The volatility in the rates ensures that the expected return to the investor is the same. In this example the convexity adjustment is the difference in the short-term and long-term rate. Of course, in reality risk aversion and expectations of changes in the average level of rates have an impact on the shape of the curve, but the concept of convexity adjustment remains in tact under those circumstances[15].

Depending on the selection of the short rate model, the convexity adjustment is somewhat different. Typically the size of the adjustment depends on the volatility and tenor of the forward rate.

- The higher the volatility, the higher the adjustment;
- The longer the tenor, the higher the adjustment.

The estimation based on the real interest rate, the inflation and the term premium is still an expected rate. We need to apply the convexity adjustment to get the forward rate.

The observed swap curves in major markets such as EUR, GBP, JPY and USD support the existence of the convexity adjustment too. They point to consistently downward sloping forward interest rate curves for tenors beyond 20 years.

Methods to estimate the convexity adjustments

1) Derived from the variance of a 10 year Government bond
2) Derived based on short rate model
3) Observed from long-term swap rates.

[15] For empirical evidence see Brown and Schaefer, "Why Long Term Forward Interest Rates (Almost) Always Slope Downwards", 1 May 2010.

Derived from the variance of a 10 year Government bond

From finance theory it can be found as shown by B&H[16] that the expected return of holding a long-term bond, which equals the unconditional nominal forward rate is equal to the one period rate, the one period expected inflation, a holding period risk premium and a convexity adjustment. This convexity adjustment is equal to minus half the variance of the excess return of the bond.

B&H use a 10-year nominal Government bond to estimate this convexity adjustment. Their analysis estimates this adjustment as -40 bps.

Derived from Short rate models

Based on the Ho Lee model, the convexity adjustment is one half of volatility squared times the beginning and ending tenor. It does not converge as the tenor goes to infinity.

Based on the Hull White model, the convexity adjustment is more complicated. Depending on the parameter, it does converge to a specific level; see figure 8 for the convexity adjustment at different maturities. If the estimation of the real rate is based on the short-term rate the ultimate convexity adjustment would be slightly more than -30 bps. If the estimation of the real rate is based on the 30 year data point, then those rates would already include a convexity adjustment of around -30 bps. Based on the chosen parameters of the Hull-White model there is hardly any additional increase in convexity adjustment that should be included in the extrapolation (i.e. -1 bps at 100 years).[17]

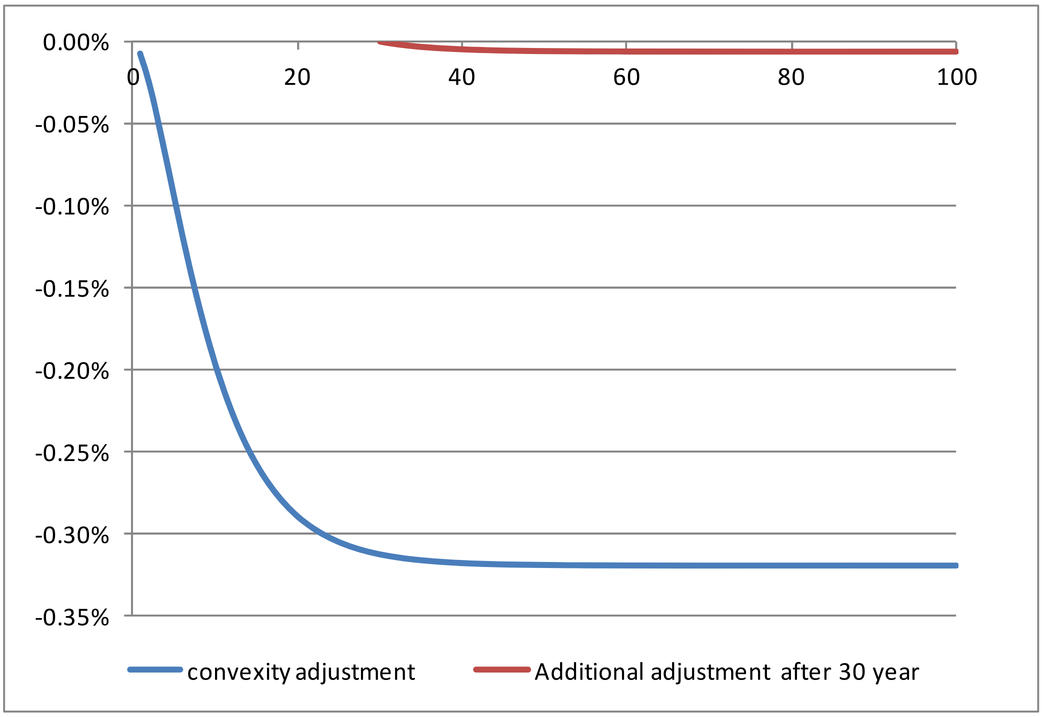


Figure 8 Convexity adjustment using Hull White interest rate model.

[16] "Interest rate calibration, how to set long-term interest rates in the absence of market prices", Steffen Sorensen, B&H, Financial Economic Research, Version 1, September 2008, Appendix.
[17] We have assumed a volatility of 1.2% and speed of the mean reversion factor of 15%.

Observed from long-term swap rates

One can also observe convexity directly in swap rates in various markets. Continuously compounded forward rates derived from swap rates typically are downward sloping in all markets. Moreover this convexity adjustment also seems to increase for 40 and 50 year tenors in EUR and USD markets. However, this "market" implied convexity adjustment combines two elements: 1) The theoretically justified convexity adjustment as shown above and 2) a risk margin as there are no 40 and 50 year bonds available to hedge the exposure that traders in such transactions take. Therefore a "risk premium" for such non-hedgeable risk is included.

The increase in the observed spreads between Q3 2008 and Q1 2009 could be interpreted as an increase in the risk premium added by market participants but could be misleading due to a drop in liquidity in those tenors.

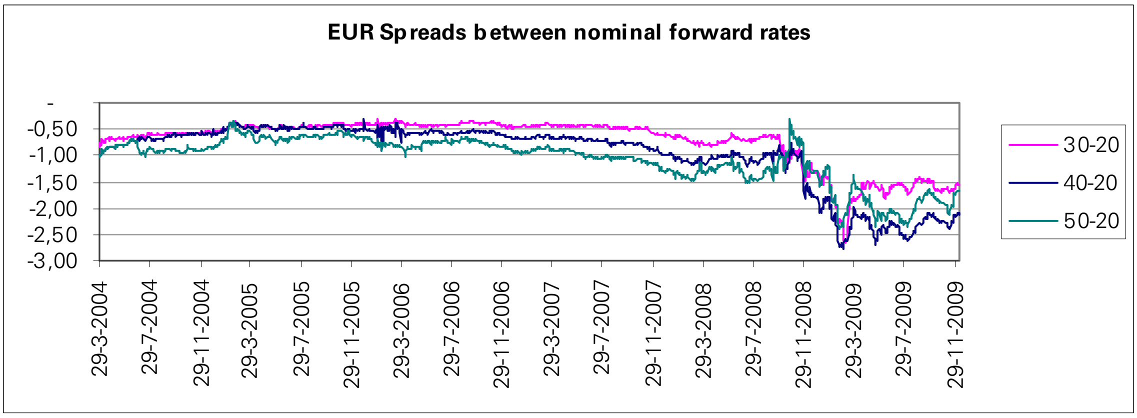


Figure 9 Observed differences between the 50, 40, 30 year 1yr forward rates versus the 20yr 1yr forward rate for EUR during the last 5 years.

The downward sloping forward curves can be interpreted as the combined effect of the term premium and the convexity effect. The drop in the 30, 40 and 50 year forwards versus the 20yr forward between Q4 2008 and Q1 2009 could be thought of as a combined effects from increased volatility (and therefore higher convexity) and a temporary increased risk premia (i.e. lower term premium) due to hedging activities by various market players resulting in more supply-demand mismatch.

BEST PRACTICE

The CRO Forum prefers an estimation of the convexity adjustment using a model approach (method 1 and 2). Both models have some limitations, so it is advised to calculate the convexity adjustment based on both models. Expert judgment will be required to take into account the limitations of these methods in setting the convexity adjustment. Method 3 has significant limitation and should only be used as a reasonableness check.

4.5.2. When is the ultimate rate reached?

The question that automatically arises when setting a long-term unconditional forward rate is when this rate might be reached. Is it reasonable to assume such a limiting rate is reached after 50, 100 or even more years?

Evidence suggests that expectations for real short rates and inflation change only very slowly for horizons beyond 10 years[18]. This is at least for developed markets with high credibility in targeting low inflation. For countries with limited credibility in setting target inflation such expectations can change more rapidly. E.g. forecasted 10 year inflation for Argentina jumped from 3.8% to 7.3% in 2008 when actual inflation increased rapidly.

It seems that variability in required risk premia accounts for a major source of asset price volatility. The question B&H [19] pose is: "how far into the future do these risk premiums variations extend" and "…, how far it is reasonable to adjust the 50-year, 100-year and 1000-year forward rate?"

As long as the ultimate rate is set sufficiently far beyond the last observed liquid market data this choice will have negligible impact on the valuation of insurance liabilities. The more critical decision is the speed of the grading process, which will be discussed in the next section. Ultimately, the point at which the ultimate rate is reached, the speed of convergence in the grading process and market data lying in the extrapolated part of the curve have to be evaluated in combination to achieve the best result from a risk management perspective. We also refer to section 4.5.5 for a discussion on risk management considerations.

BEST PRACTICE

The CRO Forum recommends that the ultimate rate is reached sufficiently far beyond the last observed liquid market data.

4.5.3. Consistency in setting all components of the unconditional long-term forward rate

We would like to emphasize that the four components of the unconditional long-term forward rate need to be set in a consistent manner. Overall a good balance needs to be found between historical observed information and market observations to ensure that on one hand the method results in a stable anchor point for extrapolation and does not create spurious interest rate volatility and on the other hand ensures market consistency. So the estimation cannot look at each component in isolation, but should look at the total framework. An implementation of all components of the unconditional long-term forward rate for end 2008 and end 2009 is shown in Appendix 4.

[18] See for example Kim and Orphanides, 2005: Term Structure Estimation with Survey Data on Interest Rate Forecasts, Finance and Economics Discussion Paper 2005-48, Board of Governors of the Federal Reserve System.

[19] "Market-consistent valuation of ultra long-term cash flows", John Hibbert, B&H.

4.5.4. How to grade from last observed liquid forward rate to the ultimate rate?

Defining an ultimate rate translates the extrapolation problem of yield curves into an interpolation ("grading") problem so that it can be treated with known methods as described in the following. The grading methodology chosen has a large effect on the variation in yields beyond the last observed liquid market data point. Spurious variation in these yields can feed through to give excessive balance sheet and valuation variation that should be avoided. With this in mind it becomes important to develop a robust extrapolation method that reflects both current market conditions and empirical views of long rate volatility, while simultaneously displaying adequate stability. One method that combines these properties very well is the Nelson-Siegel method. Below we extensively leverage work by B&H on the Nelson-Siegel (N-S) method [20] and consider their work on fitting this method best practice.

This N-S parametric form of the forward curve is described by the following formula:

$$F(t) = \beta_1 + (\beta_2 + \beta_3(t - t_{max}))\exp(-\lambda(t - t_{max}))$$

We show in the appendix how to derive this representation of N-S from other commonly used representations.

Here β_1 corresponds to the unconditional long-term forward rate, λ is a free parameter which we can use to control the speed of convergence and β_2 and β_3 can be specified by matching the value and derivative of the curve at the last observed liquid forward rate. The maturity of the last observed liquid interest rate is denoted t_{max}.

The N-S method has a number of desirable properties. At long maturities it becomes flat, and as we approach our unconditional target its second derivatives is already negative (and increasing) if approaching from below and positive (and decreasing) if approaching from above reflecting prevailing market conditions. Because forward rates are described using a very straightforward form we can analytically match the level and gradient of the curve at the terminus of the market data, and setting the unconditional level of rates is trivial. Finally, speed of convergence can be directly and intuitively controlled using the λ parameter, giving us an easy handle on long rate volatility and correlation avoiding spurious variation in yields.

The term $\exp(-\lambda(t - t_{max}))$ fades the information on β_2 and β_3 out and can be translated in a (intuitively more accessible) half life of $\ln(2)/\lambda$ – i.e. the recommended value of $\lambda=0.06$ by B&H translates into a half life of 11.6 years fading out half of the value of ($\beta_2 + \beta_3(t - t_{max})$) every 11.6 years (please be aware that $(t - t_{max})$ works in the opposite direction but with rising t $\exp(-\lambda(t - t_{max}))$ will outweigh this effect).

[20] "Fitting the yield curve, Spline interpolation and Nelson-Siegel extrapolation", David Antonio, David Roseburgh, B&H, Version 2.0, September 2008.

This approach leads implicitly to falling historic volatilities with increasing maturities. This is clearly preferred over the yield curve construction methods - used by many practitioners and central banks - that introduce spurious variation in the long-term nominal forward rates. The N-S method does not introduce the problem of some other methods in which small changes due to measurement/pricing errors (for long maturity financial instruments) can induce large variation in the long end of the forward rate curve [see B&H: "How to construct a volatility term-structure of interest rates in the absence of market prices"].

The following is a chart of the realized swap coupon rate volatility for major currencies from May 2000 to September 2008. The numbers are calculated based on the absolute change of the swap rates. 1% means that the one standard deviation of the interest rate movement over the next year is either up or down 1%. The swap rates beyond 30 years are calculated using the N-S extension. This method extends the decline of realized volatility beyond the observed period quite nicely.

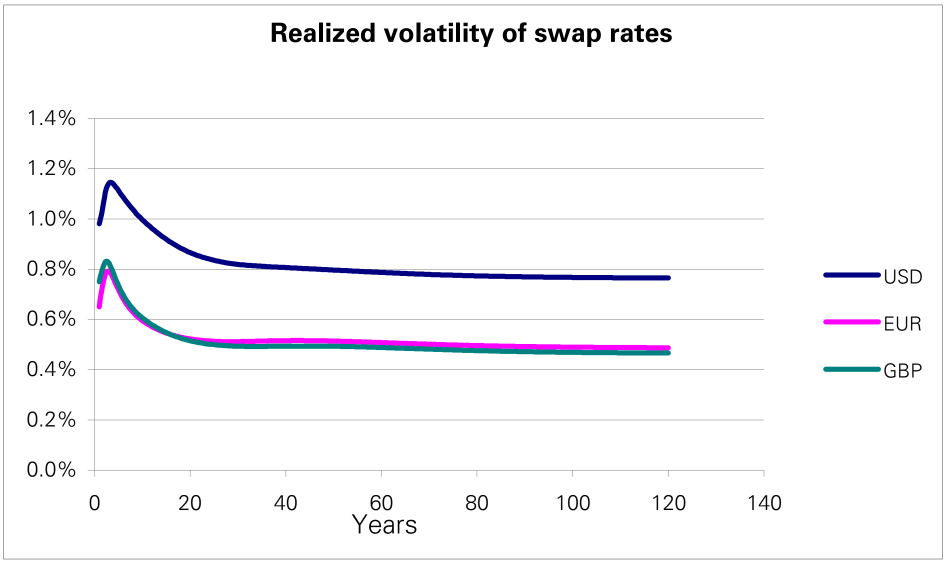


Figure 10 Realized volatility of swap rates since 2000 (including the extrapolation)

Note that the realized volatility in the extrapolated part of the curve is rather sensitive to the last observed liquid market data points especially with regards to parameter beta3 which might lead to reducing the impact of beta3 in an implementation. Another approach would be applying a (typically small) adjustment to this part of the curve to ensure that the volatility decay is in line with the observed part of the curve. The ultimate test for any grading method is to back-test the method for a range of currencies and over a significant time horizon (e.g. last 5 years) to see that the method achieves it goal of a stable curve with declining forward volatility over time and broad consistency with market data lying in the extrapolated part of the curve (e.g. within a certain band). In this area the B&H method has proven itself in practice, although alternatives can be evaluated to come up with the most appropriate method. If multiple methods work well then the most simple method would have the preference.

BEST PRACTICE

The CRO Forum supports the grading method based on Nelson-Siegel developed by B&H as being appropriate, but alternatives can be evaluated in order to achieve the goal of a curve with the desired properties with less technical efforts.

4.5.5. Risk management considerations

The extrapolation methodology should adequately reflect robust and market consistent valuation of reserves, as well as being risk manageable. A risk manageable extrapolation methodology allows the industry to neutralize effectively the hedge-able part of the interest rate exposure, which implies that the behavior of the curve given the observed part of the curve should be predictable given the observed market data and that there should be no excess volatility in the extrapolated part of the curve.

With predictable we imply that the shape of the curve cannot change overnight. As the last observed liquid forward rate is the starting point of the extrapolation, such forward rate has a high impact on the start and direction of the forward curve. Therefore, it is key that in fitting the curve to the observed market data that there is some smoothing mechanism that avoids that small inconsistencies in the last observed liquid market data points can result in big swings in the level and the shape of the last observed liquid forward rates.

Furthermore, the extrapolation method should not result in abrupt changes in the shape of the forward curve. This would result in spurious volatility in the extrapolated part of the curve. Any method used should be back-tested to show that the volatility in the extrapolated part of the curve shows the same declining pattern as observed in the observed part of the curve.

Lastly, there are often market instruments available beyond the last liquid market data point. Therefore, such instruments lie in the extrapolated part of the curve. Although these instruments are not extremely liquid, they can still be use in effectively managing the risks. Therefore, the extrapolation should be fitted such that the valuation the liabilities lie within a reasonable boundary of such instruments to avoid that using such instruments for risk management purposes does not produce counter-intuitive results.

4.5.6. Solvency II: adjustment for credit risk and illiquidity premium

In the context of Solvency II the swap curve is proposed to be adjusted to reflect

- the credit risk associated with earning the floating part of a interest rate swap transaction
- the illiquidity premium associated with the "predictability" of the liability cashflows.

Extrapolation of these adjustments is straightforward.

One of the questions that arises is how the illiquidity premium and the extrapolation of this premium interferes with the extrapolation of the base risk free curve. One of the principles for applying the illiquidity premium is there should be available assets to lock in the illiquidity premium at the valuation date[21]. Corporate bonds and covered bonds are typically referred to as the instruments for locking in the illiquidity premium. The swap curve can have liquidity beyond the corporate and covered bond market as there might be risk-free assets that are liquid and hence do not generate an illiquidity premiums. So the cut-off tenor until which the full illiquidity premium can be earned is not necessary equal to the cut-off point of liquidity for the base risk free curve, although in many cases it is. The question is whether there are instances of when the illiquidity premium can be earned beyond the last observable liquid point for the base risk free curve.

[21] Illiquidity task force principle #3 "*The liquidity premium applicable to a liability should not exceed the extra return which can be earned by the insurer by holding illiquid assets free of credit risk, available in the financial markets and matching the cash flows of the liability.*"

While there maybe circumstances in which for example illiquid bonds with a maturity beyond the last liquid risk-free instrument are available, it would be practically impossible to quantify the illiquidity premium that far out, as one would need the risk free rate to determine the size of the liquidity premium.

The extrapolation of the illiquidity premium can be applied separately from the extrapolation of the base risk free curve. The same principle is applied that this extrapolation requires a continuous forward curve. Based on the principle that the illiquidity premium can only be applied when assets are available to lock it in, the extrapolation of the forward liquidity premium goes to zero. In practice it is suggested to grade the forward liquidity premium to zero in five years. It is no problem when this period overlaps with the extrapolated part of the base risk free curve. Liquidity in the base risk free curve also does not stop immediately, so the cut-off tenor of the base risk free curve should be interpreted as the maximum tenor on which the full illiquidity premium can be earned.

4.5.7. Solvency II: extrapolation by an EU institution

The CRO Forum is supportive of CEIOPS proposal that a central EU institution shall calculate and publish the extrapolated part of the basis risk free interest rate curve based on transparent procedures and methodologies. However, we note that expert judgement plays a crucial role in setting the unconditional ultimate long term forward rate. It is therefore vital that the industry has a good understanding of how this expert judgement is determined and used to set the long term rate.

To monitor solvency position on an ongoing basis, as required by the Solvency II directive, (re)insurers would need to be able to fully understand and replicate the extrapolation process internally. It would defeat the purpose of Solvency II if the entire industry cannot monitor its solvency position because they can't predict the calculations from the central EU body. One specific concern the industry has relates to how the risk free curve is adjusted in situations where market liquidity drops and long-term observed market is no longer a good reflection of the value of risk free assets and extrapolation should start at an earlier point. Therefore it would be good to pre-define upfront triggers for an adjustment in the starting point of the extrapolated part of the curve.

Insurers may also want to perform projections and stress testing on the risk free rate for internal purposes when making business and strategic decisions. Therefore the ability to accurately and independently replicate the extrapolated risk free interest rate would be important for sound risk taking, as promoted by the Solvency II Directive.

In section 4.2 we discussed two interest rate extrapolation methods: macro-economic and constant forward rate. In the light of the principles that the risk free curve should be calculated and published by a central EU institution, based on transparent procedure and methodology, it seems logical to apply the macro-economic method for all currencies.

4.5.8. Relationship between extrapolation and calculation of SCR

The Solvency Capital Requirement for interest rate risk (in Solvency II context) is calculated by applying shocks to the interest rate environment and revaluing assets and liabilities in the shocked environment. For the observed part of the curve the shocks will be derived from historical volatility. The extrapolation of the shocked curve should be based the same best practice methods described in this paper. The calibration of the long-term forward rate should take into account the shocked market environment. However, since the long-term forward rate by design is rather stable and the shock only has a one year horizon, a shock – if any - to this rate should be minor. The path towards the long-term forward rate would change as a result of the change in the observed part of the curve.

5. Extrapolation of equity implied volatility

5.1. Importance of Extrapolation for Equity Implied Volatility

Many types of Insurance contracts, mainly life insurance, have options and guarantees linked to equity prices. These options typically have quite long durations which can be significantly beyond options traded on exchanges and in OTC markets. Moreover, these options are not held for trading, but embedded in our liabilities and therefore held until maturity or when clients surrender their policies. Given the long tenor of such options their market consistent value is very sensitive to the volatility used in the calculation.

Market equity implied volatilities are very liquid for shorter tenors, but liquidity typically declines rapidly beyond 5 years and there is often no liquidity at all beyond 10 years. During a crisis market liquidity can even evaporate as we have witnessed at the end of 2008. It is also very difficult to objectively measure liquidity as most equity options beyond a 2-year tenor are traded in OTC markets.

One important element of equity implied volatility that is relevant for its extrapolation is mean reversion. After a period of high implied volatility levels, typically pared with large drops in equity levels, the volatility level returns to normalized levels. This feature is also very clearly visible in the term-structure of implied volatility. During periods of high market volatility the term-structure of declining as longer tenor implied volatilities incorporate an expectation that volatility is returning to a long-term level and visa versa during periods of low market volatility. Therefore the extrapolation of equity implied volatility should have a built in feature to incorporate mean reversion in volatility levels.

5.2. Extrapolation method

Before discussing specific extrapolation methods, it is also important to realize how equity implied volatility is used to value embedded insurance options and guarantees linked to equity price levels. Typically options are priced with Monte Carlo simulation based on an underlying stochastic model. This model is often a Black-Scholes model which directly uses the at-the-money implied volatility term-structure or more advanced models such as the Heston or Bates stochastic volatility models which are calibrated to the at-the-money implied volatility term-structure and often additionally to the implied volatility skew. The focus here is to provide a best-practice methodology for the extrapolation of the **at-the-money implied volatility term-structure**. The complexity of the stochastic model used to value embedded options and guarantees is directly linked to the complexity of the underlying insurance contract and out of scope in this paper. We also need to clarify how at-the-money is interpreted. The most widely followed definition of at-the-money option for a time T in the future is the option with a strike level equal to the **forward equity price level** for time T. We follow this definition.

Key elements of equity implied extrapolation:

1) Determining the last liquid market data point

2) Functional form of grading from the last observed liquid forward volatility towards the long-term forward volatility

3) Long-term forward volatility

4) Adjustment for stochastic interest rates

As noted above the functional form used to grade towards the long-term level should incorporate that implied volatility follows a mean reverting process and how quickly such mean-reversion takes place. The long-term forward volatility is capturing both the long-term best estimate volatility and an adjustment for non-hedgeable market risk to capture that implied volatility is structurally higher than historical volatility due to the capital cost involved in hedging options. Lastly, implied volatility is typically quoted in the market based on the Black & Scholes formula which assumes constant interest rates. As interest rates are in fact stochastic, this needs to be taken into account when pricing long-term options & guarantees.

BEST PRACTICE

It is best practice to apply the extrapolated equity implied volatility from the last observed liquid forward volatility towards the long-term forward volatility. The functional form of grading takes into account that volatility is following a mean reverting process.

5.2. Determining the last available liquid market data point

Determining the liquidity of equity option markets is very difficult. For most indices exchange traded options only have tenors up to two years. Most liquidity is coming from OTC markets, but such liquidity is difficult to observe. Early in 2009 a CRO Forum Survey on market data indicated that companies use between 5 and 10 years of market data before starting to extrapolate. However, liquidity is not as deep for each market. Key indices such as DJ EuroStoxx50 and S&P500 have a higher liquidity than smaller indices such as IBEX or DAX for example. Furthermore, liquidity can dry up during a crisis as we have witnessed end 2008 when long-term equity implied volatilities where driven by liquidity rather than by fundamental expectations.

Quantifying specific criteria for determining the last liquid market data point is difficult and subject to expert judgment. Some elements to consider include:

- Market depth: open interest for option contracts traded in exchange markets versus historical levels. Such information is not easily obtainable and does not give information about the OTC market.
- Observed forward implied volatility levels and their volatility. One should expect that these mean revert towards a long-term level and that the volatility of implied volatility declines with the tenor. Analysis can point our where such trends are broken. E.g. at the end of 2008 forward implied volatilities beyond 2yrs did not show any mean reversion anymore.
- Market surveys among large insurance companies that participate in option hedging of their long-term options. Here we can refer to the above mentioned CRO Forum survey and for example a 2008 US survey among variable annuity writers who are active in hedging programs.
- Abnormalities in the market: Large shocks to the market typically impact the liquidity of the market.

5.3. Functional form of grading from the last observed liquid forward volatility towards the long-term forward volatility

Functional form of grading from the last observed liquid forward volatility towards the long-term forward volatility has to incorporate that volatility follows a mean reverting process. A logical starting point therefore is to look at the Heston model, which is widely used in financial markets to price equity implied volatilities. The Heston model describes the volatility as a mean-reverting stochastic process, where the process reverts back to its long-term levels with a certain speed. The model is very simple as it has very limited number of parameters and gives a good fit to both the term-structure of implied volatility and the implied volatility skew.

The basic Heston model[22] assumes that both St, the price of the asset, and vt, the assets instantaneous variance, are determined by stochastic processes.

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^S$$
$$dv_t = \kappa(\theta - v_t) + \xi \sqrt{v_t} dW_t^v$$

Where dW_t^S and dW_t^v are Wiener processes with correlation □. In this format one can see clearly that the variance follows a mean reverting process in which □ is the long-term volatility and □ the rate at which v_t reverts to □.

From the Heston model it is feasible to derive a simple and intuitive functional form for the forward implied volatility or more precisely the forward variance as the Heston model describes the variance process.

$\text{Forward IV}(t)^2 = \text{Mkt Forward IV}(t)^2$	for $t \leq T$
$\text{Forward IV}(t)^2 = \text{Mkt Forward IV}(T)^2\, w(t) + \text{LT Forward IV}^2\, [1 - w(t)]$	for $t > T$

With

$$w(t) = e^{-K(t-T)}$$

and where

T	: Start point of extrapolation, so cut-off point of market data used.
LT Forward IV	: Long-term forward implied volatility level
K	: Mean reversion speed towards long-term level
Mkt Forward IV(t)	: Forward implied volatility from observed market prices at time t

The forward implied volatility is in this functional form simply a weighted average of the last observed liquid forward variance and the long-term implied variance level. The mean reversion speed K can easily be implied from observed market data. This method is in line with the methodology applied by B&H. In the Appendix it is shown how K could be estimated from observed variation in implied volatility markets.

[22] See http://en.wikipedia.org/wiki/Heston_model for more details on the Heston model.

5.4. Long-term forward implied volatility

Implied volatility has some different features than just the equity price level volatility observed in the market. A market participant that wants to replicate option prices by a dynamic rebalancing strategy (also known as delta-hedging) is exposed to non-linear equity price movements (gamma exposure). There are capital costs associated with such risk. In particular there is the risk of a jump in equity markets which can result in losses for market participants following a dynamic rebalancing strategy. Furthermore, there is uncertainty in the long-term implied volatility level which cannot be hedged in the market. We note that such capital costs for non-hedgeable market risk is not only related to the extrapolated part of the term-structure, but already is incorporated in the observed part as well. When comparing short-term (very liquid) implied volatility levels with short-term realized historical volatility then one can observe a consistent ratio between implied volatility and historical volatility of 110% to 120% for most indices. In the appendix an analysis is included for several indices to show this difference. For longer tenor implied volatilities there is the additional uncertainty of the long-term implied volatility level. By calibrating this long-term implied volatility level from market data one can observe that market participants do not view this long-term level as fully certain.

Therefore the long-term forward implied volatility (in fact its square, the variance) can be described as:

$$\text{Long-term Forward IV}^2 = \text{Long-term Best Estimate Volatility}^2 + \text{Capital Cost Non-hedgeable Market Risk} + \text{Uncertainty LT Best Estimate Volatility}$$

In the following section we will discuss each of these components. However, we should stress that they cannot be determined fully independently as they are related.

5.4.1 Long-term Best Estimate Volatility

The first question is what historical volatility is representative for the long-term best estimate volatility. The balance has to be found between using a long enough historical period to warrant stability in the estimate and using a relevant enough period as the volatility is used for projecting forward which implies that more recent information is more relevant. One observation when looking at current historical volatilities for several tenors is that beyond 7 years the volatility level stabilizes.

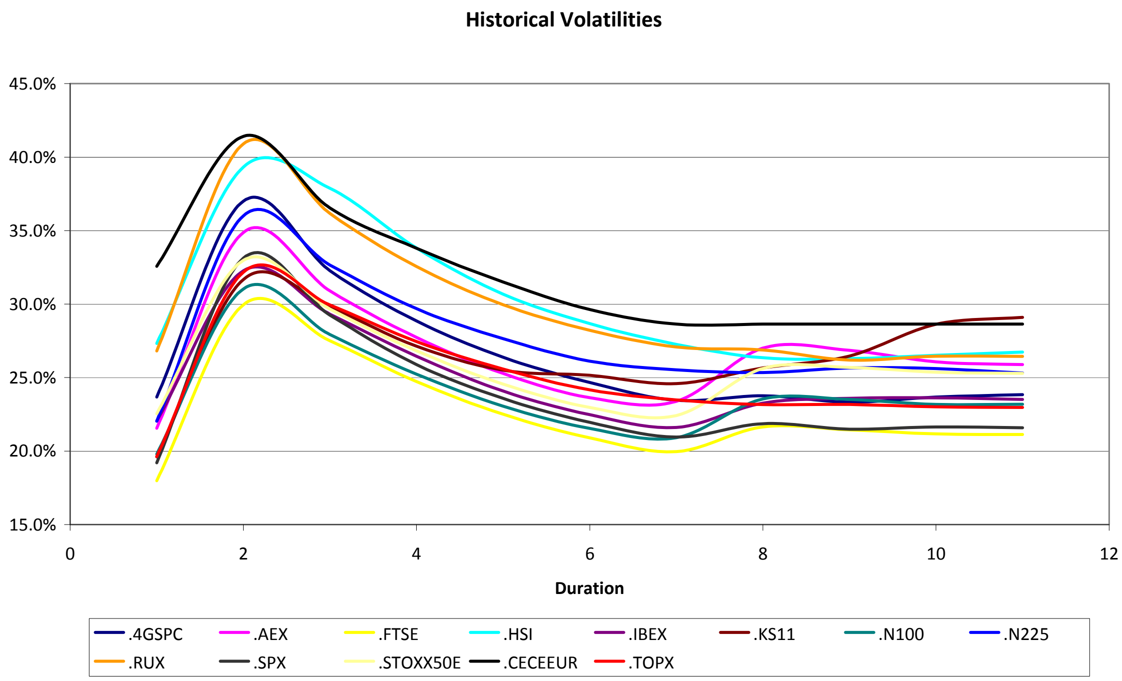


Figure 11 Historical volatilities of selected equity indices based on different length of historical period

Another observation is that the volatility of the last 10 years has been higher than over the previous 10 years. So when looking at very long historical data one might not capture that historical volatility has gone up over time. Below the 10yr volatility of different indices over the three blocks of 10yr daily return data from 1/1/1978 to 1/1/2008.

	DJ EuroStoxx 50	S&P 500	Nikkei 225
1998 - 2008	23.06%	17.97%	22.02%
1988 - 1998	14.30%	13.12%	21.88%
1978 - 1988		17.00%	13.20%

The second question is whether one single historical volatility can be used for all indices or whether it is index specific. Our view is that there can be significant differences between indices due to the number of stocks in an index, E.g. 50 stocks in DJ EuroStoxx50 versus 500 in the S&P500; The correlation of stocks within an index, e.g. a world index is more diversified than a country index; And the type of stocks of which the index is composed of e.g. the Nasdaq 100 was much more volatile than the S&P500 in the 1990s during the internet-bubble. We therefore prefer a separate estimate of the long-term best estimate volatility per index.

Lastly, one should consider consistency between the long-term forward implied volatility and the long-term level implied from the market. When calibrating the functional form by estimating the long-term level and the mean reversion speed by fitting the functional form to the market, one gets an estimate of the long-term level implied by the market. The long-term best estimate volatility and the adjustments discussed in the next section should be set such that they are broadly consistent with the market.

5.4.2 Capital Cost of Non-Hedgeable Market Risk

As explained before it can be observed that implied volatility is consistently higher than historical volatility over time. It can also be theoretically explained why this is observed in the market. The market investor who holds an equity option cannot fully offset the risk by a static replication with the underlying stock or index. A common hedging strategy is to regularly rebalance the underlying position such that the linear risk is neutralized. Such a strategy is called delta hedging. In such a delta-hedging strategy the investor is subject to the risk of a big overnight shock in equity prices resulting in a loss due to the non-linearity of options. Therefore, the investor will need to hold capital for this risk. In the appendix it is shown that a simple model can be constructed in which the capital that the investor needs to hold is a direct add-on to the long-term implied volatility based on just a few parameters.

Capital cost Non-Hedgeable Market Risk = $2 \cdot \pi \cdot [\mathrm{J} - 1 - \ln(\mathrm{J})]$,

where

π is the cost-of-capital and

J the jump risk in equity markets to which the investor is sensitive (J = 70% implies a 30% jump in markets).

For example applying a cost-of-capital of 6% and a jump of 30% in markets (J=70%) and a best estimate volatility of 25% results in an add-on for implied volatility of 1.3%. We also note that the method implies a constant add-on in variance resulting to a higher add-on in standard deviation for when the best estimate volatility is lower as can be seen from below table.

Best estimate volatility	15.0%	20.0%	25.0%	30.0%
Best estimate variance	0.0225	0.04	0.0625	0.09
Non-hedgeable Market Risk	0.0068	0.0068	0.0068	0.0068
Total variance	0.0293	0.0468	0.0693	0.0968
Long-term implied volatility	17.1%	21.6%	26.3%	31.1%
Implied add-on for non-hedgeable Market Risk	2.1%	1.6%	1.3%	1.1%

5.4.3 Uncertainty Long-term Best Estimate Volatility

There is uncertainty in the long-term best estimate volatility. As this long-term best estimate volatility cannot be locked in effectively this implies that capital is required to capture the uncertainty in this long-term best estimate volatility. This can already be observed to some extent in observed implied volatilities for longer tenors, although it is hard to measure given other elements (less liquidity, demand/supply mismatch) that start to play are role. However, technically it can be shown as by Manistre (2010)[23] that the add-on for the uncertainty in the long-term best-estimate volatility can be written as

Uncertainty Long-term Best Estimate Volatility = $\dfrac{(\Delta\sigma)^2}{1-\alpha}$

where

$\Delta\sigma$ a shock in the best-estimate volatility over a 1 year period.

α a parameter that governs potential future shocks over time.

[23] Manistre, B.J. "A Cost of Capital Approach to Extrapolating an Implied Volatility Surface". Proceedings of the 2010 ERM Symposium (Chicago). http://www.ermsymposium.org/2010/pdf/2010-erm-chicago-manistre.pdf

The shock is indirectly related to the level of the historical best estimate volatility. When this level is already near an historical high then even a shock in the market will not result in a big jump in the long-term estimate, while the shock is bigger when the historical volatility is at a low level. In fact this becomes a kind of stabilizing component in the overall long-term volatility that offsets actual changes in the best-estimate volatility to some degree. Manistre argues an of around 50%.

Below is graph showing how the 7yr and 10yr absolute historical volatilities behave on a data set of 30yrs of daily data of the S&P 500. It shows that annual movements in the historical volatility moved by up to 4% for the 10yr best estimate volatility and up to 6% in the 7yr best estimate volatility.

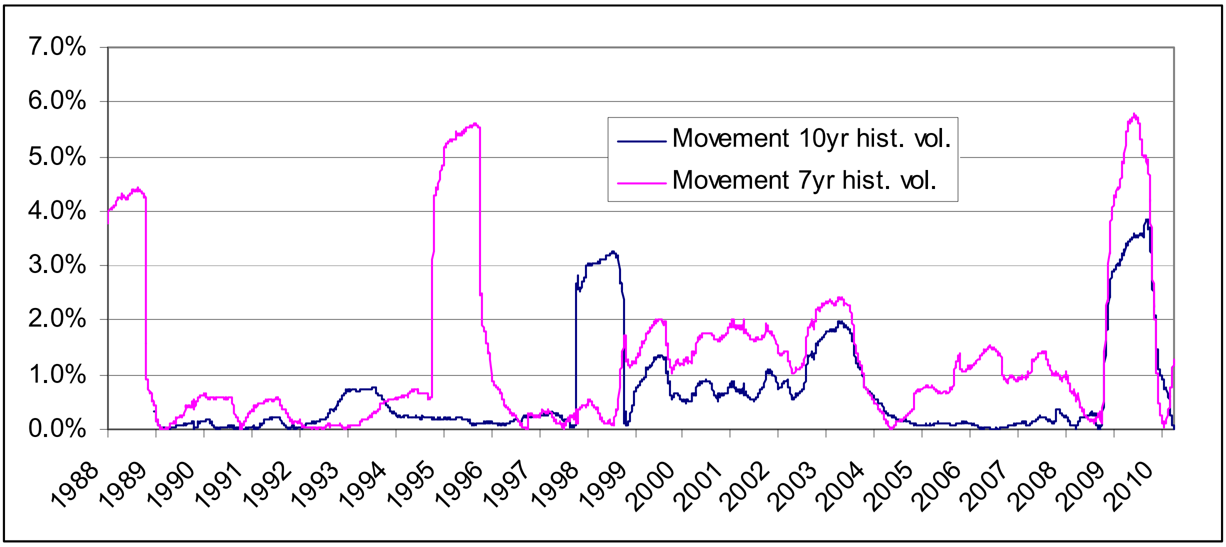


Figure 12 Absolute change in historical volatility of S&P 500 based on 10 and 7 year period to calculate the historical volatility.

Applying the 4% (based on shocks of 10yr historical tenor) and an alpha factor of 50% results in an add-on of about 0.6% on the long-term implied volatility based on a best-estimate volatility of 25%. See below the impact for various levels of the best estimate volatility.

Best estimate volatility	15.0%	20.0%	25.0%	30.0%
Best estimate variance	0.0225	0.04	0.0625	0.09
Uncertainty Long-term best estimate	0.0032	0.0032	0.0032	0.0032
Total variance	0.0257	0.0432	0.0657	0.0932
Long-term implied volatility	16.0%	20.8%	25.6%	30.5%
Implied add-on for uncertainty LT best estimate	1.0%	0.8%	0.6%	0.5%

Combining the impact the impact of non-hedgeable market risk and uncertainty in the long-term best estimate one can calculate the long-term implied volatility for different levels of the best estimate volatility.

Best estimate volatility	15.0%	20.0%	25.0%	30.0%
Best estimate variance	0.0225	0.04	0.0625	0.09
Non-hedgeable Market Risk	0.0068	0.0068	0.0068	0.0068
Uncertainty Long-term best estimate	0.0032	0.0032	0.0032	0.0032
Total variance	0.0325	0.0500	0.0725	0.1000
Long-term implied volatility	18.0%	22.4%	26.9%	31.6%
Absolute add-on vs BE volatility	3.0%	2.4%	1.9%	1.6%
Relative add-on vs BE volatility	120.2%	111.8%	107.7%	105.4%

We note that above numbers are to illustrate the methodology and that the parameters used are representative, but not specific recommended values by the CRO Forum. We believe that some more work needs to be done in estimating the parameters. Moreover, the resulting long-term implied volatilities need to be benchmarked with values implied from option prices in the market such that the methodology does not result in a consistently over- or under-pricing of long-term equity guarantees.

5.3. Adjustment for impact of stochastic interest rate volatility

Industry practice is to quote implied volatility as the volatility to plug into the Black & Scholes formula to get the market implied volatility. One of the assumptions in the B&S formula is that interest rates are constant over time. When pricing short-term options this assumption is not material as forward equity prices for shorter tenor are driven by equity volatility rather than interest rate volatility. However, for longer tenor options the impact of stochastic interest rate cannot be ignored. Therefore there are two options in calibrating our stochastic models:

1. We interpret the implied volatility in the extrapolated part of the implied volatility term-structure as a pure equity volatility. When calibrating these volatilities we assume therefore also deterministic interest rates. However, when simulating scenarios for our options and guarantees we combine the stochastic process for equity with a stochastic interest rate model. As a result the total volatility of "forward" equity prices will be higher than under deterministic interest rates. The implied volatilities of the observed part of the curve need to be adjusted down for the stochastic process to reproduce the dollar price of the traded options.
2. We interpret the implied volatility in the extrapolated part of the implied volatility term-structure as the implied volatility that we need to enter into the B&S formula to price options. In such case an adjustment needs to be made to compensate for the fact that the B&S formula assumes deterministic interest rates. This adjustment needs to be consistent with the interest rate model used to simulate interest rate levels in the valuation of embedded options and guarantees in insurance products. In the Appendix this adjustment is worked out for the Hull & White interest rate model.

The two options will give the same result. This adjustment is broadly acknowledged by market participants working with long-term options and therefore any long-term equity guarantee also needs to be valued in combination with a stochastic interest rate model.

5.4. Implied volatility surface

In the above we fully focused on the extrapolation of the at-the-money term-structure of implied volatility. In reality the implied volatility level not only depends on the tenor, but also on the moneyness level of the option (ratio between strike and forward equity level). In practice this aspect should be captured by the stochastic model used to generate equity scenarios. The complexity of the model used should take into account the complexity of the liability and its embedded equity guarantees (e.g. complex knock-in guarantees at different levels are more complex than plain guarantees like return of initial investment). The calibration of any model should at least consider the at-the-money term-structure and for more complex models (e.g. Heston) should also consider the full volatility surface of the liquid part of the volatility market. This then automatically implies extrapolated volatilities for in- and out-the-money volatilities in the extrapolated part of the curve.

6. Extrapolation of interest rate implied volatility

Extrapolation of interest rate implied volatility should follow the principles identified in this paper. However, interest rate volatility cannot be extrapolated in a straightforward way. Contrary to equity implied volatility, interest rate volatility is not typically quoted in term of the volatility of the interest rate term-structure itself (i.e. zero rate volatility), but in terms of swaption volatilities.

Therefore, interest rate volatility extrapolation and calibration of stochastic interest rate models are two very related topics. Although calibration of stochastic models to market data is outside the scope of this paper we would like to illustrate this with an example. Any stochastic interest rate model such as the Libor Market model, Hull & White or Black-Karasinski allow us to describe the term-structure of interest rate volatility in terms of the parameters of the model. Therefore, extrapolating interest rate volatility towards a long-term level would imply putting restrictions on the model parameters.

Another element that needs to be considered is consistency in market data used to calibrate the stochastic interest rate model. When the swap market is only considered to be liquid up to 30yrs then it does not make sense to include swaption in the calibration where the combined option and swap tenor is beyond 30yrs (e.g. 10y x 30yr or 20yr x 20yr swaptions).

Lastly, one important observation is that for many currencies there is no developed interest rate option market yet. Therefore, in line with the next chapter, proxies need to be constructed based on historical data.

Overall, the calibration of stochastic interest rates and the implicit extrapolaton of interest rate volatility through this calibration is an area that requires more research to develop best practices.

7. Valuation of options on instruments with no options traded in the market

7.1. Implied Volatilities for Unobservable Volatility Markets

Implied volatilities are based on option prices, and are indicative of the market's consensus on the future volatility of the underlying price and the risk aversion in the market. Where traded option prices are not accessible or non-existent, alternative approaches to constructing suitable implied volatility surfaces are needed.

Two general approaches can be used to construct implied volatility surface proxies:

- Construction from an implied volatility surface of another index accessible in the market
- Construction from historical price volatility of the target index, adjusted with a risk margin add-on

7.2. First Approach

This first approach is based on the assumption that historical price volatilities are indicative of implied volatilities.
A benchmark index with accessible implied volatility data is first selected. The benchmark and target index prices should ideally be highly correlated, or should share comparable characteristics. Next the numerical ratio is determined based on the indices' price volatilities. This ratio indicates the extent by which the benchmark's implied volatilities are adjusted to construct the target's implied volatilities. This method may be extended to further allow for volatility smiles, through applying this ratio approach on up- /down-side strikes. In case the relative ratio overstates implied volatilities in very skewed areas of the surface one solution is to replace it by an absolute difference.

To illustrate, if the ratio of historical volatilities of the target index to that of the benchmark is 125%, then this approach would estimate the target's implied volatilities to be in the order of 125% of the (implied volatility) benchmark's. It should be noted that the ratio need not necessarily be constant throughout the entire surface. Ratios at different tenors may be determined via different periodic lengths of historical volatility, for example.

Consequences of the first approach:
In this approach, the target volatility surface mimics the benchmark's characteristics. This has implications for volatility arbitrage trades, especially when both target and benchmark indices are involved. Diversification of such option portfolios may be understated, given the assumed perfect correlations underlying the construction of the target's implied volatilities. In such cases, the basis risk should be captured within the risk module[24].

7.3. Second Approach

In the second approach, the implied volatility of the target index is estimated using the historical price volatilities of the underlying, with a risk margin adjustment.

This method is simply a generalization of the equity implied volatility method developed by Manistre described in section 5.4 in which the implied volatility is estimated as the sum of the expected best-estimate volatility plus adjustments for a shock in the underlying investment and to the best-estimate parameter.

Alternatively a combination of the first and second approach can be used in which a marker is first selected whereby the implied volatility will be estimated from historical index price volatilities, adjusted for the difference between implied and historic volatilities. Ideally, the tenor of this marker should be of short-term. The implied volatility surface is then constructed by graduating the estimated short-term volatility with the long-term volatility referred to in section 5.4.

---

[24] Also required by Final Advice of CEIOPS, CP31, on risk mitigation techniques.

This approach captures the variability of short-term implied volatilities and the relative stability of long-term ones; where the graduation of implied volatility term structure can be controlled by a mean-reversion parameter (kappa).

To illustrate, the short (3-Month) marker tenors is first selected. The 3-month implied volatility is estimated as 110% of the historical 3-month index price volatility, where 10% is the assumed risk margin adjustment in this case. The remaining implied volatility term structure can then be determined by interpolating between this point and the long-term volatility from section 5.4.

8. Appendices

Appendix 1. Swap Bid-Ask Spread Overview

	10Y	20Y	30Y	40Y	50Y
AUD	4	5	5		
CAD	3	4	4		
CHF	8	10	10		
EUR	3	3	3	4	4
GBP	5	13	13	17	17
JPY	4	4	4		
KRW	3	4			
MXN	14	14			
MYR	10				
NOK	4	7	8		
PLN	7	7			
SEK	4	8	9		
TWD	6				
USD	3	3	3	3	3

Average bid-ask spread on swaps rates in basis points over 2005 – 2009 period based on broker data (mainly ICAP). Note that spreads might be smaller for some currencies when looking at different brokers.

Appendix 2. CRO Forum Survey on Extrapolation of Market Data

Below table summarizes the CRO Forum survey results for all currencies with at least 4 companies submitting results. The "Most used" column shows the tenor that was most frequently used as the last liquid swap tenor (all companies use the swap curve as their risk free curve with the exception of a few currencies). The table further shows the longest available swap tenor, the number of companies in the survey using this particular currency and the % of these companies using the "most used" tenor (about 65% on average). Data is based on market situation per December 31, 2009. In the last column we also show the industry submission of the cut-off tenor for the QIS5 tenor. This column represents the max tenor of market data points that have been liquid during the 2007-2009 crisis, which indicates that for a number of markets this tenor is (significantly) shorter then most companies used per end-2009.

Nb	Currency	Most used end-09	Max Swap Tenor	Number of Companies	% Equal to "most-used"	QIS5 sub-mission
1	EUR	50	60	12	67%	30
2	GBP		60	12	58%	50
3	USD		60	12	67%	30
4	CHF		50	9	56%	15
5	JPY		50	9	56%	20
6	CAD	30	30	8	100%	30
7	SGD		30	8	88%	30
8	AUD		30	7	86%	15
9	CZK		30	7	57%	15
10	MXN		30	6	50%	30
11	SEK		30	6	83%	10
12	DKK		30	5	80%	30
13	NOK		30	4	50%	10
14	PLN	20	20	8	63%	15
15	HUF		20	7	57%	15
16	MYR *		10	6	50%	20
17	KRW **		20	6	67%	20
18	THB **		20	6	50%	20
19	RON		20	4	50%	10
20	HKD	15	15	9	89%	15
21	TWD *		10	8	50%	15
22	NZD		15	5	40%	NA
23	CNY	10	10	8	63%	10
25	INR **		10	8	50%	10
24	IDR		10	4	75%	NA
26	PHP		10	4	75%	NA
27	RUB		10	4	50%	NA
28	TRY		10	4	50%	10
29	BRL	5	5	5	80%	5

* In some currencies the swap market was extended by government yields. In two currencies the best practice tenor is beyond the last available liquid swap tenor.

** In these currencies a minority of companies extended the curve using Government yields.

Appendix 3. Nelson-Siegel:

As most readers might be more familiar with a different representation of the Nelson-Siegel formula like this:

$$f_t(\tau) = \beta_{1,t} + \beta_{2,t}\exp\left(-\frac{\tau}{\lambda_t}\right) + \beta_{3,t}\left(\frac{\tau}{\lambda_t}\right)\exp\left(-\frac{\tau}{\lambda_t}\right)$$

with $\beta_{1,t}$ being the long term component, $\beta_{2,t}$ being the short term component and $\beta_{3,t}$ being the mid-term component in a traditional application of Nelson-Siegel.

we show shortly how to derive it from the (more convenient) representation we made use of:

$$\begin{aligned} F(t) &= b_1 + (b_2 + b_3(t - t_{\max}))e^{-\lambda(t - t_{\max})} \\ &= b_1 + b_2 e^{-\lambda(t - t_{\max})} + b_3(t - t_{\max})e^{-\lambda(t - t_{\max})} \\ &= b_1 + b_2 e^{\lambda t_{\max}} e^{-\lambda t} + b_3(t - t_{\max})e^{\lambda t_{\max}} e^{-\lambda t} \\ &= b_1 + \left\{b_2 e^{\lambda t_{\max}} - b_3 t_{\max} e^{\lambda t_{\max}}\right\}e^{-\lambda t} + \left\{\frac{b_3 e^{\lambda t_{\max}}}{\lambda}\right\}\lambda t e^{-\lambda t} \end{aligned}$$

Both representations are equivalent as:

$$\begin{aligned} \beta_{2,t} &= \left\{b_2 e^{\lambda t_{\max}} - b_3 t_{\max} e^{\lambda t_{\max}}\right\} \\ \beta_{3,t} &= \left\{\frac{b_3 e^{\lambda t_{\max}}}{\lambda}\right\} \\ \lambda_t &= 1/\lambda \end{aligned}$$

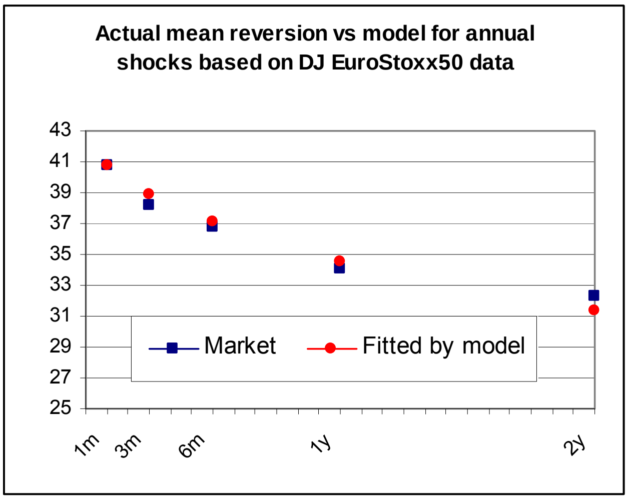
Appendix 5 – Mean Reversion during 2008-2009

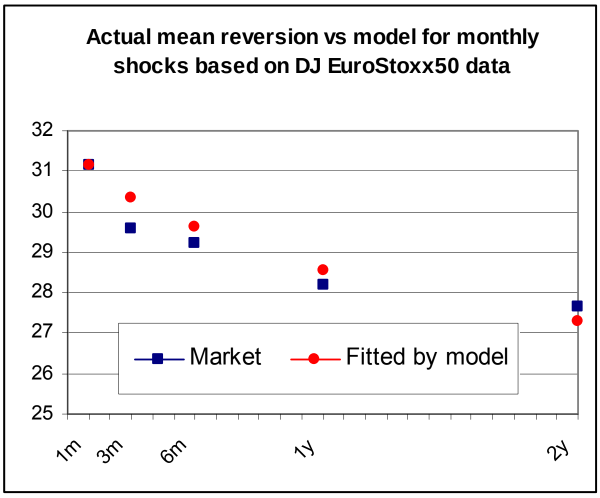
Some analysis was done to show how well the functional form captures the mean reversion observed in the equity implied volatility term structure. We estimated the annual and monthly standard deviation of 1m, 3m, 6m, 1y and 2y at-the-money options for the DJ EuroStoxx 50. This is the most liquid segment of the market. We then applied a 1 standard deviation shock to the implied volatility term structure assuming a flat 25% term-structure to start with. In the graphs below the red dots show the shocked implied volatilities after a 1 year (left graph) or 1 month (right graph) standard deviation shock.

We then applied the functional form with the 1m option as the short-term volatility and the initial 25% as the long-term implied volatility level.

$$\text{Forward IV}(t)^2 = \text{1m shocked IV}^2\, w(t) + \text{LT Forward IV}^2\, [1 - w(t)] \text{ for } t > 1m$$

We then calibrated the optimal mean reversion parameter k to get the optimal fit to the shocked market data for all tenors beyond 1month (by construction the 1 month point is perfectly fitted). It turned out that a K equal to 1.0 for the annual shocks and to 1.1 for the monthly shocks resulted in the most optimal fit. In below graphs we show the modeled equity implied volatilities for K=1.0.





Conclusion is that we indeed see a clear mean reversion implied from the actual movements in implied volatility data and that proposed functional form provides a good fit with this observed behavior.

Appendix 6 – Observed difference between short-term historical volatilities and implied volatilities

In observed market data a clear difference can be observed between historical and implied volatility. A test was performed on daily observed market data for the period 2005 and 2008. Each day the historical 3 month volatility and the observed 3 month implied volatility have been observed. Below table shows the average value of both the historical and the implied volatility for a range of indices. It can be observed that implied volatility is consistently higher by up to 2% points and on a relative basis in the range of 10-20%.

	AEX	IBEX	FTSE 100	DJ EuroStoxx50	S&P 500	Nikkei 225	Hang Seng	Average
Avg. 3M HV	13.3%	13.5%	12.8%	14.1%	11.5%	17.3%	16.7%	14.2%
Avg. 3m IV	14.9%	14.8%	13.7%	15.5%	14.0%	18.1%	18.3%	15.6%
Difference	1.6%	1.3%	0.8%	1.4%	2.5%	0.8%	1.6%	1.4%
Relative Diff	18%	8%	10%	17%	20%	14%	6%	13%

Appendix 7 – Cost of capital approach for extrapolating the implied volatility surface

This section is a short summary of the 2010 paper by Manistre[25] referred to earlier. The basic idea is to start with a best estimate or P measure model of equity dynamics and then consider the problem of valuing an option using only the cost of capital concepts that have been developed to value other forms of non-hedgeable risk. Models that have been calibrated to fit observed market data then grade to this cost of capital model over some reasonable time frame.

Starting with the standard lognormal model of equity price movement $dS_t = \mu S_t dt + \sigma S_t dW_t^S$ the paper assumes insurers hold economic capital for two very different risks

1. A contagion shock or jump $S \rightarrow JS$ where the factor J is chosen to be consistent with the economic capital model's VaR level. Most capital models assume a J factor in the 30%-40% range. This is analogous to a life insurer holding capital for a repeat of the 1918 flu epidemic. If $V(t,S)$ is the value of the option we valuing then the model assumes we hold capital equal to $V(t,JS) - V(t,S)$. This ensures that the enterprise can withstand a shock event even if it is not compatible with the best estimate model.
2. The second risk taken into account is the fact that our best estimate P measure parameters μ, σ can be wrong. If $\sigma \rightarrow \hat{\sigma}$ is a plausible shock to the volatility parameter over a 1 year time frame then the insurer needs to hold sufficient capital to cover the increase in value $V(t,S,\hat{\sigma}) - V(t,S,\sigma)$. This is analogous to a life insurer holding capital for a shock to a mortality or lapse assumption.

Having identified the key risks the paper then defines the value of the option to be the expected present value of option payouts plus the cost of holding economic capital for the two risks described above.

Considering the contagion shock first, the principles above tell us that the option value satisfies a differential equation of the form

$$\frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV - \pi[V(t,JS) - V(t,S)]. \qquad (1)$$

This equation says we expect the value of the option to grow with interest while releasing sufficient margin to pay for the cost of capital. The quantity π is the cost of capital rate. An investor who is willing to put up the risk capital can expect to earn the risk free rate r plus π on their investment.

If we want our model to be market consistent, in the sense that it prices the equity index back to itself, then the function $V(t,S) = S$ must be a solution of the equation above. Demanding that this be the case allows us to solve for the cost of capital. We find that the cost of capital is the leveraged equity premium $\pi = (\mu - r)/(1-J)$.

[25] Manistre, B.J. "A Cost of Capital Approach to Extrapolating an Implied Volatility Surface". Proceedings of the 2010 ERM Symposium (Chicago). http://www.ermsymposium.org/2010/pdf/2010-erm-chicago-manistre.pdf

If we assume a typical equity premium of $\mu - r = 3\%$ and $J = 70\%$ then $\pi = 10\%$. The leverage factor makes sense from the perspective of a shareholder who puts up the risk capital. Such an investment has a beta factor of $1/(1-J)$. Note that this is a very different situation from the cost of capital for underwriting risk which is usually considered to be a low beta risk.

If we substitute $\mu = r + \pi(1-J)$ into the valuation equation (1) above we can rewrite it in a way that makes comparison with the standard Black-Scholes model easy. We find, after a simple rearrangement

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV - \pi[V(t, JS) - V(t, S) - (J-1)S\frac{\partial V}{\partial S}].$$

The way to interpret this form of the equation is to say that V is the risk neutral present value of option cash flows plus the cost of holding capital for the loss that would occur if a large jump occurred while delta hedging. The cost of capital is positive as long as the instrument being valued is convex.

We have found a model where hedgers and speculators can agree on value as long they each hold capital for their respective un-hedged risk and they both use $\pi = (\mu - r)/(1-J)$ as the cost of capital. The model derived therefore does not assume delta hedging but is not inconsistent with it either.

One of the paper's more technical results is that effect of using the above model to value long tenor options is equivalent to using the Black-Scholes model with an implied volatility of the form $\sigma_{imp}^2 \approx \sigma^2 + 2\pi(J - 1 - \ln J)$. This is essentially a law of large numbers limiting result which becomes more accurate as the maturity date of the option increases. A detailed derivation of this useful formula is in the paper's appendix.

To put a value on the parameter shock $\sigma \to \hat{\sigma}$ risk we need to add a term to equation (1) to get something like

$$\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV - \pi[V(t, JS) - V(t, S)] - \hat{\pi}[\hat{V}(t, S) - V(T, S)] \qquad (2)$$

The paper argues that the cost of capital for parameter risk $\hat{\pi}$ is not the same as the leveraged equity premium $\pi = (\mu - r)/(1-J)$ because it is not a direct market risk. $\hat{\pi}$ should be similar to rates used to value underwriting risk.

The shocked value $\hat{V}$ is calculated using a similar equation but with shocked volatility

$$\frac{\partial \hat{V}}{\partial t} + \mu S\frac{\partial \hat{V}}{\partial S} + \frac{1}{2}\hat{\sigma}^2 S^2 \frac{\partial^2 \hat{V}}{\partial S^2} = rV - \pi[\hat{V}(t, JS) - \hat{V}(t, S)] - \hat{\pi}[\hat{\hat{V}}(t, S) - \hat{V}(T, S)] \qquad (3)$$

Equation (3) illustrates a conundrum that arises when valuing parameter shocks. In order to compute V we need to know $\hat{V}$ and that requires a double shocked value $\hat{\hat{V}}$ etc.. This is known as the circularity problem and arises when computing market value margins for insurance risks. The paper explains more fully why this model structure makes theoretical sense and, at the same time, finds a very practical way to solve the conundrum without over engineering.

A practical model that captures the essence of the structure outlined above is to use a time dependent, but deterministic, variance $\nu = \sigma^2$ that evolves, in the valuation measure, according to $d\nu = \hat{\pi}(1-\alpha)(\sigma^2 + \frac{\hat{\sigma}^2 - \sigma^2}{1-\alpha} - \nu)dt$. The valuation equation (2) becomes

$$\frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{1}{2}\nu S^2 \frac{\partial^2 V}{\partial S^2} = rV - \pi[V(t, JS, \nu) - V(t, S, \nu)] - \hat{\pi}(1-\alpha)(\sigma^2 + \frac{\hat{\sigma}^2 - \sigma^2}{1-\alpha} - \nu)\frac{\partial V}{\partial \nu}.$$

The way to interpret this equation is to say that, in the P measure, $\nu = \sigma^2$ is constant so that in the real world we get a continuous margin release equal to

$$\hat{\pi}(1-\alpha)(\sigma^2 + \frac{\hat{\sigma}^2 - \sigma^2}{1-\alpha} - \nu)\frac{\partial V}{\partial \nu} = \hat{\pi}(\hat{\sigma}^2 - \sigma^2)\frac{\partial V}{\partial \nu}.$$

The quantity $(\hat{\sigma}^2 - \sigma^2)\left.\frac{\partial V}{\partial \nu}\right|_{\nu=\sigma^2}$ is clearly the model's required parameter risk capital.

The parameter $\alpha \le 1$ governs the model's simplified approach to the shock hierarchy or circularity issue. Given the first shock $\hat{\sigma}^2$ we are basically assuming that the next shock is $\hat{\hat{\sigma}}^2 = \hat{\sigma}^2 + \alpha(\hat{\sigma}^2 - \sigma^2)$ with subsequent shocks similarly scaled by the factor α. If $\alpha < 1$ this sequence of values will converge to a limiting value of $\sigma^2 + \frac{\hat{\sigma}^2 - \sigma^2}{1-\alpha}$ which we can think of as an upper bound on any plausible sequence of variance assumption changes. This is the value to which the models variance is mean reverting towards at the rate $\hat{\pi}(1-\alpha)$.

A plausible value of the rate $\hat{\pi}(1-\alpha)$ is something like $6\%(1-.5) = 3\%$ which means that it can take a long time for the variance to reach its ultimate value.

When we put these two pieces together we conclude that the ultimate long term implied volatility that captures both types of risk for long dated options is

$$\sigma_{imp}^2 \approx \sigma^2 + 2\frac{\mu - r}{1-J}(J - 1 - \ln J) + \frac{\hat{\sigma}^2 - \sigma^2}{1-\alpha}.$$

A final point to note is that we could have started with a more sophisticated P measure model (e.g. a Heston type model) and gone through a similar risk analysis to arrive at very similar conclusions.

Appendix 8 – Adjustment to equity implied volatility term-structure to compensate of stochastic interest rates

In this appendix, we illustrate the impact of stochastic interest rates on the equity implied volatility term structure by way of two popular interest models – the Hull-White 1-factor and Hull-White 2-factor models.

As explained in the main text, extrapolated equity implied volatilities should be adjusted to allow for interest rate volatility. These adjusted implied volatilities can then be directly plugged in to option pricing formulae. This adjustment is based on the principle that interest rate volatilities are not represented in the extrapolated implied volatilities. This adjustment may also be applied to the extrapolated implied volatilities of other market variables such as Foreign Exchange if these extrapolates do not possess any interest rate volatility features.

We commence with a brief theoretical construct of the Hull-White models before providing numerical illustration on how the implied volatilities are adjusted. The adjusted volatilities can then be 'plugged-and-played' in routine option pricing formulae such as the Black-Scholes Merton.

It should be noted that the implied volatilities may be also adjusted via other interest rate models, with the appropriate adjustment formulae; although many of such models are seldom tractable.

We begin with the following definitions:

S	denotes the underlying equity
σ_S	forward implied volatility of an option
σ_{adj}	adjusted forward implied volatility of an option
α and σ	short rate mean-reversion and sigma parameters in the Hull-White 1-factor model, respectively
ρ	correlation between the equity and interest rate processes in the HW-1 model
α_i and σ_i	short rate mean-reversion and sigma parameters in the Hull-White 2-factor model, respectively
ρ_{MN}	correlation between the processes M and N in the HW-2 model; where M,N = stock or interest rate factors

Hull-White 1-factor model

In the Hull-White 1-factor model, the adjusted volatility is given by

$$\sigma_{adj}^2(t).(T-t)=\int_t^T\{\sigma_S^2+2\rho\sigma_S\sigma_y.B(s,T)+\sigma_y^2.B(s,T)^2\}.ds$$

where $$B(s,T)=\frac{1}{\alpha}.(1-e^{-\alpha(T-s)})$$

The above equation can be interpreted as the total forward (adjusted) equity volatility being composed of the sum of:

- the equity volatility, as represented by the forward implied volatility
- the interest rate volatility; and
- the covariance between the equity and interest rate processes

Since σ_S is implicitly assumed as a flat volatility forecast, the integration of the above formula should arrive at

$$\sigma^2_{adj}(t).)(T-t)=\sigma^2_S.(T-t)+2\rho\sigma_S\sigma_y.\int_t^T B(s,T).ds+\sigma^2_y.\int_t^T B^2(s,T).ds$$

the two integrands in the above equation can then be simplified as

$$\int_t^T B(s,T).ds=(T-t)-\frac{1}{\alpha}(1-e^{-\alpha(T-t)})$$

$$\int_t^T B^2(s,T).ds=(T-t)-\frac{1}{2\alpha}(1-e^{-2\alpha(T-t)})-\frac{2}{\alpha}(1-e^{-\alpha(T-t)})$$

such that the adjusted volatility, under the Hull-White 1-factor model, is thus

$$\boxed{\sigma^2_{adj}(t).(T-t)=\sigma^2_S.(T-t)+\frac{\sigma^2_y}{\alpha^2}\left((T-t)+\frac{2}{\alpha}e^{-\alpha(T-t)}-\frac{1}{2\alpha}e^{-2\alpha(T-t)}-\frac{3}{2\alpha}\right)+\frac{2\rho\sigma_S\sigma_y}{\alpha}\left((T-t)-\frac{1}{\alpha}(1-e^{-\alpha(T-t)})\right)}$$

Hull-White 2-factor model

In the Hull-White 2-factor model, the adjusted volatility is similar and given by

$$\sigma^2_{adj}(t).(T-t)=$$
$$\int_t^T\left\{\sigma^2_S+2\rho_{xS}\sigma_S\sigma_x B_x(s,T)+2\rho_{yS}\sigma_S\sigma_y B_y(s,T)+2\rho_{xy}\sigma_x\sigma_y B_x(s,T)B_y(s,T)+\sigma^2_x B_x(s,T)^2+\sigma^2_y B_y(s,T)^2\right\}.ds$$

where the interest rate dynamics are described by the two factors, denoted by x and y:

$$\begin{aligned} r_t &= x_t + y_t + \phi_t \\ dx_t &= -\alpha_1.x_t.dt+\sigma_x.dW_t^X \\ dy_t &= -\alpha_2.y_t.dt+\sigma_y.dW_t^Y \end{aligned}$$

Note that $\square_1$, $\square_2$, $\square_1$ and $\square_2$ are four model parameters derived from calibration. The inter-factor correlations are given by:

$$\begin{aligned} \langle dw_x dw_y\rangle &= \rho_{xy}dt \\ \langle dw_x dw_S\rangle &= \rho_{xS}dt \\ \langle dw_y dw_S\rangle &= \rho_{yS}dt \end{aligned}$$

and

$$\begin{aligned} B_x(s,T) &= \tfrac{1}{\alpha_1}(1-e^{-\alpha_1(T-s)}) \\ B_y(s,T) &= \tfrac{1}{\alpha_2}(1-e^{-\alpha_2(T-s)}) \end{aligned}$$

Working through in similar manner as the Hull-White 1-factor derivation, the adjusted implied volatility under the Hull-White 2-factor model can be presented as:

$$
\begin{aligned}
&\sigma_{adj}^2(t).(T-t) = \\
&\sigma_S^2.(T-t) + \frac{\sigma_x^2}{\alpha_1^2}\left((T-t) + \frac{2}{\alpha_1}e^{-\alpha_1(T-t)} - \frac{1}{2\alpha_1}e^{-2\alpha_1(T-t)} - \frac{3}{2\alpha_1}\right) + \frac{\sigma_y^2}{\alpha_2^2}\left((T-t) + \frac{2}{\alpha_2}e^{-\alpha_2(T-t)} - \frac{1}{2\alpha_2}e^{-2\alpha_2(T-t)} - \frac{3}{2\alpha_2}\right) \\
&+ \frac{2\rho_{xS}\sigma_S\sigma_x}{\alpha_1}\left((T-t) - \frac{1}{\alpha_1}(1-e^{-\alpha_1(T-t)})\right) + \frac{2\rho_{yS}\sigma_S\sigma_y}{\alpha_2}\left((T-t) - \frac{1}{\alpha_2}(1-e^{-\alpha_2(T-t)})\right) \\
&+ \frac{2\rho_{xy}\sigma_x\sigma_y}{\alpha_1\alpha_2}\left((T-t) - \frac{1}{\alpha_1}(1-e^{-\alpha_1(T-t)}) - \frac{1}{\alpha_2}(1-e^{-\alpha_2(T-t)}) + \frac{1}{\alpha_1+\alpha_2}(1-e^{-(\alpha_1+\alpha_2).(T-t)})\right)
\end{aligned}
$$

The following table illustrates the impact of stochastic interest rates on implied volatilities. As one would expect, these adjustments increase with the option lifespan.

HW-1 parameters	
alpha	5.00%
sigma	1.00%
cutoff tenor	5
rho_Sx	15%

HW-2 parameters	
alpha_01	5.00%
sigma_01	1.00%
alpha_02	5.00%
sigma_02	1.00%
cutoff tenor	5
rho_Sx	15%
rho_Sy	15%
rho_xy	-75%

tenor	Actual Market Volatility	Adjusted Vol HW-1 Model	Adjusted Vol HW-2 Model
0.08	16.94%	16.94%	16.94%
0.25	18.95%	18.95%	18.95%
0.50	20.08%	20.08%	20.08%
0.75	20.79%	20.79%	20.79%
1	21.28%	21.28%	21.28%
1.5	21.64%	21.64%	21.64%
2	21.96%	21.96%	21.96%
3	22.39%	22.39%	22.39%
4	23.55%	23.55%	23.55%
5	24.39%	24.39%	24.39%
6	25.02%	25.14%	25.17%
7	25.51%	25.74%	25.80%
8	25.89%	26.24%	26.32%
9	26.20%	26.66%	26.77%
10	26.44%	27.03%	27.16%
12	26.82%	27.65%	27.80%
15	27.20%	28.39%	28.56%
20	27.58%	29.34%	29.52%

The following points are noteworthy for the above illustration:

- the cutoff tenor denotes the tenor at which historical equity-based extrapolation commences. The higher the cutoff tenor, the further the impact of stochastic interest rates is delayed in the implied volatility structure.
- Although the adjustments are to be applied to forward volatilities, we have applied the adjustments to the spot volatilities in this illustration for purposes of simplification.
- This adjustment may also be applied for non-complete markets, where implied volatility surfaces are fabricated from historical equity prices. In the case where the entire implied volatility surface is constructed from equity price movements, the cutoff tenor should be set to 0.

Recent publications of the CRO Forum

Calibration recommendation for the correlations in the Solvency II standard formula

December 2009

Calibration Principles

May 2009

Internal Model Admissibility

April 2009

Insurance Risk Management Response to the Crisis

April 2009

Addressing the pro-cyclical nature of Solvency II

November 2008

Members of the CRO Forum