OpenGamma Quantitative Research

Eight ways to strip your caplets: An introduction to caplet stripping

Richard White & Yukinori Iwashita
Richard@opengamma.com
yuki@opengamma.com

OpenGamma Quantitative Research n. 24 | First version: September 26, 2014; this version October 2, 2014

Abstract

This paper is neither a primer on vanilla interest rate derivatives or numerical optimisation. It is, as its name suggests, an introduction to caplet stripping, i.e. inferring caplet/floorlet prices from the market prices of interest rate caps and floors. This of course required some numerical optimisation.

We present several techniques (of different levels of sophistication) to achieve our goal. These fall roughly into two camps: producing caplet volatilities such that the market (cap/floor) prices are recovered exactly (or to some high tolerance), with little regard for whether there is any arbitrage between caplets at different strikes; producing arbitrage free caplet volatility surfaces, which do not fully recover the market prices.

Contents

1 Introduction **1**

2 Pricing Caps and Caplets **1**
2.1 Pricing Caplets 1
2.2 Pricing Caps 2
2.3 Cap Volatility 2
2.4 Cap Strikes 3

3 General Calibration **3**

4 Term Structure Stripping Methods **4**
4.1 Bootstrapping 5
4.2 Interpolated Curve 5
4.2.1 Using different numbers of knots 7
4.3 Penalty methods 9
4.3.1 Direct Fitting 9
4.3.2 P-Splines 9

5 Global Stripping Methods **10**
5.1 Smile Model Based Strippers 10
5.1.1 SABR with interpolated parameter term structures 11
5.1.2 Smile Based Stripper using P-Splines 14
5.2 Penalty Function based Strippers 14
5.2.1 Direct Fitting 14
5.2.2 P-Splines again 15

6 Conclusion **16**

A Market data **17**

B Difference, Derivative and Penalty Matrices **17**
B.1 Difference Matrix 17
B.2 Derivative Matrix 20

1 Introduction

This paper is concerned purely with the discussing of caplet stripping - that is, the inferring of the values of caplets and floorlets (call and put options respectively on a Libor rate) from the market quoted values of interest rate caps and floors. It is not a primer on these instruments as such, and only discusses their mechanics insofar as it is required to understand the stripping algorithms.[1]

Section 2 discusses the pricing of caplets/floorlets and caps/floors. Section 3 discusses calibration in an abstract sense, while sections 4 and 5 give concrete examples of caplet stripping algorithms with results.

For brevity we will often refer to caps and caplets only, with the understanding that this also applies to floors and floorlets.

2 Pricing Caps and Caplets

In the following section we talk exclusively about caps and caplets; floors and floorlets are completely analogous (i.e. caplets are call options on the underlying Libor rate while floorlets are put options on the same rates), and do not require a separate treatment.

2.1 Pricing Caplets

We consider a Libor rate covering the period T_i to T_{i+1}. The Libor rate fixes at T_i^f,[2] with a value $L(T_i, T_{i+1}) \equiv L_i$. The corresponding forward rate (for this Libor) at some time $t \leq T_i^f$ is given by $F(t, T_i, T_{i+1}) \equiv F_i(t)$.

A caplet (floorlet) is essentially a call (put) option on a Libor rate; for the above Libor rate, the payoff of a caplet/floorlet is given by

$$\text{caplet/floorlet payoff} = \tau_i(\chi[L_i - K])^+ \tag{1}$$

where τ_i is the year fraction for the period[3] and χ is $+1$ for caplets and -1 for floorlets.

The actual payment is at T_{i+1}^p, which for a standard caplet is approximately equal to T_{i+1}.[4] Since the Libor rate fixes at T_i, this is effectively the option expiry,[5] and the value at this time is

$$V_{\substack{\text{caplet}\\\text{floorlet}}}(T_i; K, T_i) = P(T_i, T_{i+1}^p)\tau_i(\chi[L_i - K])^+ \tag{2}$$

where $P(T_i, T_{i+1}^p)$ is the discount factor from the payment time.[6] The fair value at some time $t < T_i$ is given by

$$V_{\substack{\text{caplet}\\\text{floorlet}}}(t; K, T_i) = P(t, T_{i+1}^p)\tau_i\mathbb{E}_t^{\mathbb{T}_{i+1}}\left[(\chi[L_i - K])^+\right] \tag{3}$$

[1] For that consult a interest rate derivatives text book, e.g. [BM06, Reb02, AP10a]

[2] This is usually two days before T_i.

[3] This is determined as $DCC(T_i, T_{i+1})$, where DCC is the relevant *Day Count Convention*.

[4] A difference of a few days may arise from the treatment of none-business days, see [Res13] for details on conventions.

[5] The option payment is deterministic from this point.

[6] While we do not wish to explicitly discuss multi-curves, it should be noted that the discount factor comes from the discount curve, while the forward Libor rate will come from the (relevant) index or Libor projection curve.

where $\mathbb{E}_t^{\mathbb{T}_{i+1}}[\cdot]$ is the expectation (at time t) in the T_{i+1} forward measure. In this measure we have

$$F_i(t) = \mathbb{E}_t^{\mathbb{T}_{i+1}}[L_i] \tag{4}$$

and it is market standard to price caplets/floorlets with the Black formula[Bla76] and implied volatilities, so their price is given by

$$V_{\substack{\text{caplet}\\\text{floorlet}}}(t; K, T_i) = P(t, T_{i+1}^p)\tau_i \text{Black}(F_i(t), K, T_i - t, \sigma(K, T_i), \chi) \tag{5}$$

where $\sigma(K, T_i)$ is the volatility for a particular strike and expiry, and Black's formula is

$$\begin{aligned} \text{Black}(F, K, T, \sigma, \chi) &= \chi\left(F\Phi(\chi d_1) - K\Phi(\chi d_2)\right) \\ d_1 = \frac{\ln\left(\frac{F}{K}\right) + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}} &\quad d_2 = d_1 - \sigma\sqrt{T} \end{aligned} \tag{6}$$

There is no loss in generality in doing this;[7] it simply says that the value of a caplet/floorlet is expressed as a Black volatility - to obtain a price we simply use equation 5.

2.2 Pricing Caps

A cap (floor) is simply a set of spanning caplets (floorlets) with a common strike, so the value of a cap (floor) is simply the sum of the values of the constituent caplets (floorlets). The fair value of the cap covering the period T_α to T_β (so payments are (nominally) at $T_{\alpha+1}, T_{\alpha+1}, \ldots, T_{\beta-1}$ and T_β) is

$$V_{\substack{\text{cap}\\\text{floor}}}(t, K, T_\alpha, T_\beta) = \sum_{i=\alpha}^{\beta-1} V_{\substack{\text{caplet}\\\text{floorlet}}}(t; K, T_i) = \sum_{i=\alpha}^{\beta-1} P(t, T_{i+1}^p)\tau_i \text{Black}(F_i(t), K, T_i, \sigma(K, T_i), \chi) \tag{7}$$

All that is required to price the cap are the volatilities of the constituent caplets.

2.3 Cap Volatility

The price of caps is market information. The cap (implied) volatility is then the single volatility given to all the constituent caplets (recall they have a common strike), such that the cap price is recovered by equation 7. That is, it is the single value, $\sigma(K, T_\alpha, T_\beta)$, that solves the equation

$$V_{\substack{\text{cap}\\\text{floor}}}(t, K, T_\alpha, T_\beta) = \sum_{i=\alpha}^{\beta-1} P(t, T_{i+1}^p)\tau_i \text{Black}(F_i(t), K, T_i, \sigma(K, T_\alpha, T_\beta), \chi) \tag{8}$$

A one year cap will generally have a different volatility from a two year cap (even for the same strike), so this gives different volatilities (and hence prices) to the same caplets depending on whether they are being considered in the one or two year cap. Hence cap volatility should be seen simply as an alternative quoting convention to price.

[7] We are now in a world where negative forward rates (and negative strikes) are appearing for some currencies, in particular EUR. The Black model cannot handle this situation. Alternative strategies include applying an arbitrary shift (100bps say, so the new lower limit is a rate of -1%), or switching to normal (rather than log-normal) dynamics, i.e. Bachelier's formula. This is outside the scope of this paper.

2.4 Cap Strikes

Caps are usually quoted with absolute strike, so there will be 0.5%, 1%, 2% etc strikes. This can be useful for caplet stripping as it guarantees a set of different length caps with a common strike.

The other liquid strike is at-the-money (ATM); this the strike that equals the swap rate for the period covered by the cap, that is

$$K^{ATM}_{\alpha,\beta} = \frac{\sum_{i=\alpha}^{\beta-1} \tau_i P(t, T^p_{i+1}) F_i(t)}{\sum_{i=\alpha}^{\beta-1} \tau_i P(t, T^p_{i+1})} \tag{9}$$

Of course, different length caps will have different values of the ATM strike. Caps can also be quoted with relative strikes, e.g. ATM±100bps, ATM±200bps, etc.

3 General Calibration

In reality caplets are not traded instruments, so there are no market prices (or volatilities) of individual caplets, and instead they must be inferred from cap prices (or equivalently cap volatilities). Since there are many more caplets than caps, there is no unique solution, and we wish to choose, in some sense, the *best* solution from the universe of solutions.

Before we get into specific models for caplet stripping, we present a brief overview of different classes of calibration. In the context of caplet stripping, *observables* mean cap/floor prices or alternatively (Black) volatilities.[8]

Given a set of n observables $\mathbf{y}$ (as a vector) and a model that has a set of m parameters, $\mathbf{x}$, which produces a set of model values, $\mathbf{y}_{\text{mdl}}$, according to

$$G : \mathbb{R}^m \to \mathbb{R}^n \quad \mathbf{x} \mapsto G(\mathbf{x}) = \mathbf{y}_{\text{mdl}},$$

then model calibration falls into three distinct cases:

- $m < n$: The number of model parameters is less than the number of observables. Here the problem is under specified, and it is generally not possible to find a set of model parameters such that $\mathbf{y}_{\text{mdl}} = \mathbf{y}$. Instead, one solves for $\mathbf{x}$ in a least squares sense, i.e.

$$\text{minimise} \quad (\mathbf{y} - G(\mathbf{x}))^T \mathbf{W} (\mathbf{y} - G(\mathbf{x})) \tag{10}$$

  where $\mathbf{W}$ is a diagonal weighting matrix, which can be used to give more importance to fitting some observables over others.[9] This type of problem can be solved with the Levenberg-Marquardt algorithm[Lev44] (see [PTVF07] for a discussion).

- $m = n$: The number of model parameters equals number of observables. In principle we can perform a multidimensional root find to obtain

$$\mathbf{x} = G^{-1}(\mathbf{y})$$

[8] Generally different results will be obtained from fitting to prices vs fitting to volatilities; the exception is when the market observables can be recovered exactly, then (up to numerical accuracy) the two are equivalent.

[9] If solving for price, it is common practise to use the (inverse) option vega as the weight - this is a reasonably good proxy to fitting for volatility. Of course, the identity matrix can be used to give equal weighting.

which gives model values than match the observed values exactly. This can be solved by quasi-Newton methods such as Broyden's method[Bro65] (again discussed here [PTVF07]). In general there is no guarantee that such a solution exists, in which case least squares, as above, can be used.

- $m > n$: The number of model parameters exceeds the number of observables. Here the problem is over specified, and there is generally no unique solution. To impose a unique solution, some other constraint is required. One possibility is to introduce some penalty function, $P(\cdot)$, on the model parameters, and solve the constrained optimisation problem

$$\text{minimise} \quad P(\mathbf{x}) \quad \text{subject to} \quad \mathbf{y} = G(\mathbf{x}) \tag{11}$$

  that is, our unique solution has the lowest value of $P(\mathbf{x})$ subject to the calibrated model recovering the observed values. A proxy for the constrained optimisation problem, that has a simple implementation, is

$$\text{minimise} \quad (\mathbf{y} - G(\mathbf{x}))^T \mathbf{W} (\mathbf{y} - G(\mathbf{x}))) + \lambda \mathbf{x}^T \mathbf{P} \mathbf{x} \tag{12}$$

  where $\mathbf{P}$ is some penalty matrix[10] and λ controls the balance between having the model exactly recover the observed values, and minimising the penalty. This type of problem can be solved using a modification of the Levenberg-Marquardt algorithm.[11]

All these methods benefit greatly from having the Jacobian matrix,[12] $\mathbf{J}(\mathbf{x})$, available analytically (as opposed to having to estimate it using finite difference).

In the following sections, we will present caplet stripping methods belonging to all three classes of calibration we have identified above. These methods are all applied to the same cap volatility data (shown in appendix A), to provide a consistent comparison.

4 Term Structure Stripping Methods

The first set of methods we will look at assume that caplet volatilities are functions of time-to-expiry only, i.e. there is a volatility term structure, but no strike dependence. This is applicable to absolute strike quotes, where a set of caps (of different lengths) exist with a common strike; different strikes are treated as a separate problem, so we can build up the caplet volatilities across strikes, by joining these separate solutions together. This approach will not work when the market quotes relative strikes (e.g. ATM±100bps, ATM±200bps, etc); in this case, one must use one of the global methods from the next section.

We assume that for each strike a caplet volatility curve, $\sigma_K(T)$, exists that is a function of the caplet expiry only. Caplet volatilities for all the underlying caplets in a cap are taken from this curve, and used in equation 7 to give the cap price (which may then be converted to a cap volatility).

Of course, our aim is to infer this caplet volatility curve from the market values of caps. We present several methods to do this below.

[10] $\mathbf{x}^T \mathbf{P} \mathbf{x}$ is a quadratic approximation of $P(\mathbf{x})$.

[11] This is implemented in OpenGamma as *NonLinearLeastSquareWithPenalty*.

[12] The Jacobian is given by $J_{i,j} = \frac{\partial (y_{\text{mdl}})_i}{\partial x_j}$

4.1 Bootstrapping

This is the most basic of all the caplet stripping algorithms. It is very quick and guaranteed to work provided a solution exists.[13] It works as follows:

- Put the caps in ascending order of maturity.
- Find the price difference series for the caps (the first entry is just the price of the shortest cap.
- Partition the caplets so they are assigned to the relevant price difference (again the first set of caplets correspond to the caplets in the shortest cap)
- For each caplet partition, assign a common volatility, and solve by 1-D root finding, for the price difference.

There are other ways the same result could be achieved, but this way decouples the problem and allows the root finding at different expiries to be done in parallel.[14] This algorithm will only fail if a particular price difference is less than the intrinsic value of the constitute caplets. A modification to the partitioning would allow caps with different start times to be used, but we do not consider that here.

The result of this procedure on our sample data is shown in figures 1 and 2. Viewed against time for fixed strikes,[15] and within the limitations of piecewise constant volatility, the result seems plausible, in that there are no large jumps at the cap expiries. However viewed as a volatility smile (i.e. volatility against strike at fixed expiries), the picture is less good, with the 6M smile in particular showing a great deal of oscillation.

This method, like the two below, makes 18 separate fits to the 18 absolute strikes. There is no constrain on smoothness across these separate fits, so it is hardly surprising that the smiles that emerge can be far from smooth and potentially admit arbitrage.

4.2 Interpolated Curve

If the volatility curve is represented as a 1-D interpolated curve with the same number of knots as there are cap prices, then in principle this can be solved to give a curve that will exactly reprice all the caps. The choice of knot positions and interpolator will have a large effect on the quality of the resulting curve - it is very easy to produce a wildly oscillating curve that nevertheless perfectly reprices all the caps.

If a local interpolator (i.e. piecewise constant or linear) is used, and the knots are placed at the expiries of the caps, then the system can be solved left-to-right (i.e. bootstrapped) by a series of 1-D root findings.[16] In general we must solve for all the knot values together using a multi-dimensional root finder [PTVF07]. In the example below, we use a double-quadratic

[13] For a solution to exist the cap price differences must not be less than the sum of the intrinsic value of the caplets; if this is the case, there is an arbitrage.

[14] Normal bootstrapping is a series of 1-D root finding problems, that must be solved in order.

[15] We have shown 3 strikes out of the 18 that are fitted.

[16] Of course for a piecewise constant interpolator, the result will be identical to the direct bootstrapping discussed in the previous section.

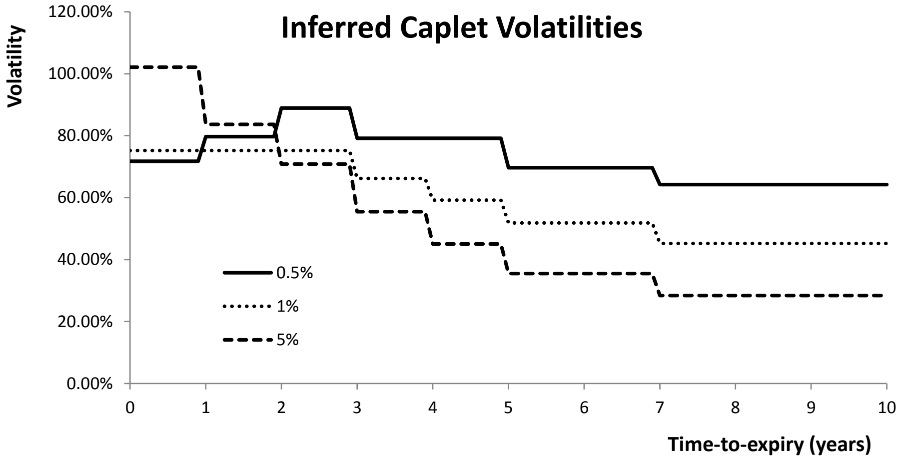

Figure 1: The result of the bootstrap procedure for strikes of 0.5%, 1% and 5%. The slope (rather than a jump) is an artefact of how we have plotted the data.

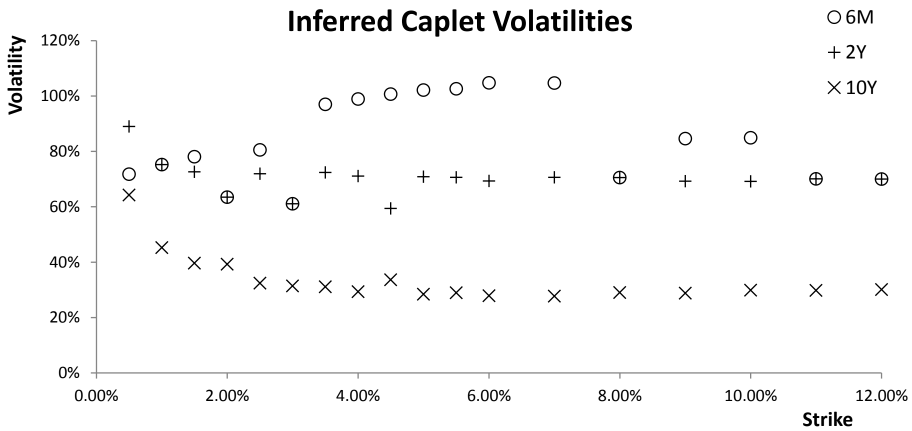

Figure 2: The result of the bootstrap procedure (i.e. volatility smiles) for maturities of 6M, 2Y and 10Y.

interpolator[17] (and a linear extrapolator) with knots at the co-start of the caps, and all but the last cap expiry.[18]

Figure 3 shows the resultant caplet volatility surface[19] from joining together the 18 separate calibrations to the absolute strikes. These caplet volatilities do indeed produce cap prices that exactly match the market, however one would expect the surface to be 'smoother'. This can be seen more clearly by taking slices through it: Figure 4 shows slices in the time direction (at strikes of 0.5%, 1% and 5%), which are smooth term structures of volatilities; however slices in the strike direction (smiles shown in figure 5) look completely implausible. The explanation is the same - there is no coupling between the solutions across strikes to impose any smoothness in that direction.

4.2.1 Using different numbers of knots

Fewer knots than caps could be used, with the system solved in a least-squares sense. However, since this will not recover the market values exactly while conferring no obvious advantages, there seems little point in exploring this further.

We could also use more knots than market cap values, with an additional penalty on the *smoothness* of the resultant curve. This is discussed below in the context of representing the curve by B-splines[dB78].

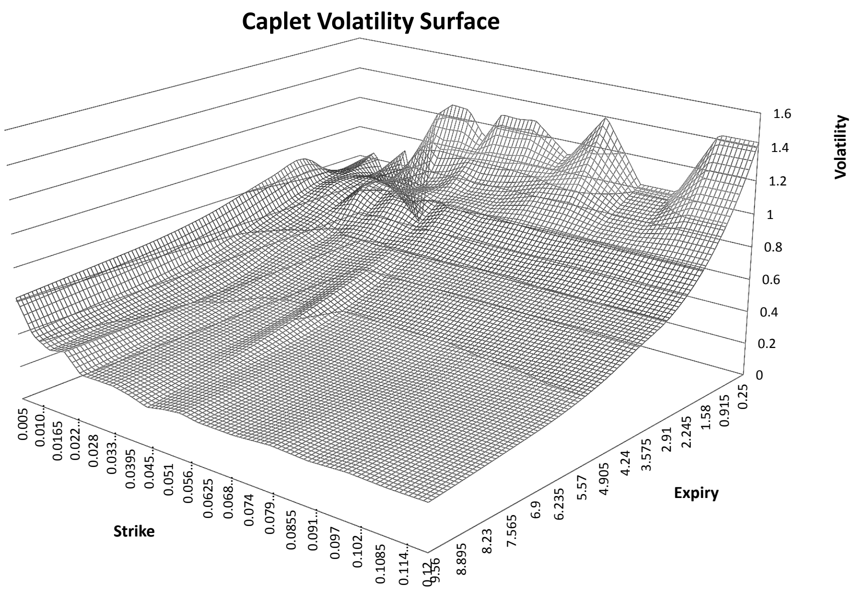

Figure 3: The caplet volatility surface from a series of independent calibrations to separate strikes using interpolated volatility term structures.

[17] Two quadratic fits are made: one involving the previous two knots and the next; and the other involving the previous knot and the next two. A linear weighting, based on position between the two adjacent knots, is then taken between the two curves. The result is a semi-local third-order polynomial interpolation.

[18] We found this gave a more stable curve.

[19] A linear 2-D interpolator has been used to fill in between the caplet volatilities.

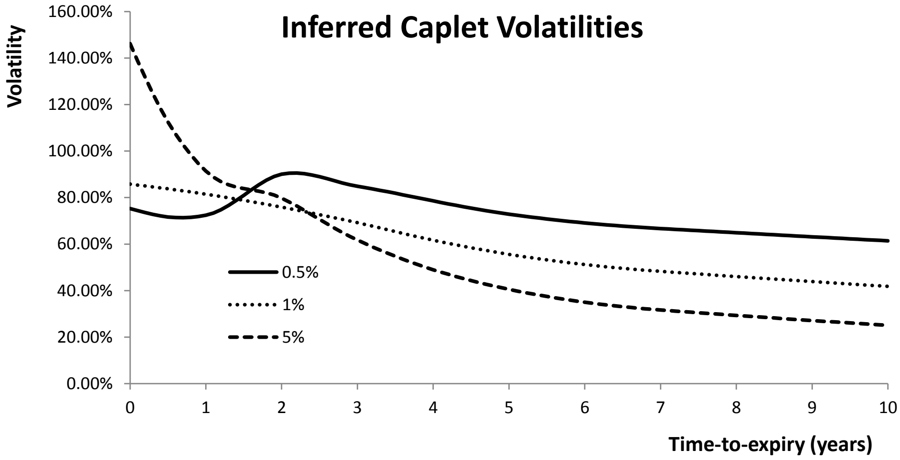

Figure 4: The result of the interpolation procedure for strikes of 0.5%, 1% and 5%.

4.3 Penalty methods

4.3.1 Direct Fitting

The problem with an interpolated term structure is that it requires a choice of knots and interpolator: a poor choice of either can give an oscillating volatility curve, that nevertheless reprices all the caps. An alternative is to fit directly for all the underlying caplet volatilities (of a given strike), so these volatilities are themselves the model parameters.

We order the unique set of underlying caplets by their expiries and let the vector of caplet volatilities be $\boldsymbol{\sigma}$. The vector of cap values (either prices or volatilities) we denote by $\mathbf{v}$, then equation 7[20] can be expressed in a compact vector notation as

$$G : \mathbb{R}^m \to \mathbb{R}^n; \boldsymbol{\sigma} \mapsto G(\boldsymbol{\sigma}) = \mathbf{v} \tag{13}$$

That is, $G(\cdot)$ is the mapping from the vector of m caplet volatilities to the vector of n cap values. Since $n > m$, there is no unique solution, and we must impose some additional penalty.

Let $\mathbf{P}$ be some penalty matrix (which we detail below), such that the quantity

$$\boldsymbol{\sigma}^T \mathbf{P} \boldsymbol{\sigma}$$

is zero only if $\boldsymbol{\sigma}$ is perfectly smooth and increases the less smooth $\boldsymbol{\sigma}$ is. We now have a penalty method,

$$\text{minimise} \quad (\mathbf{v} - G(\boldsymbol{\sigma}))^T (\mathbf{v} - G(\boldsymbol{\sigma})) + \lambda \boldsymbol{\sigma}^T \mathbf{P} \boldsymbol{\sigma} \tag{14}$$

The penalty matrix, $\mathbf{P}$ is constructed as $\mathbf{P} = \boldsymbol{\mathcal{D}}^T \boldsymbol{\mathcal{D}}$ where $\boldsymbol{\mathcal{D}}$ is the k^{th} order differentiation matrix such that $\boldsymbol{\mathcal{D}} \boldsymbol{\sigma}$ is the finite difference estimate of $\frac{\partial^k \sigma}{\partial t^k}$. Further details of the penalty matrix are given in appendix B. We use $k = 2$ which penalises the curvature of the volatility term structure.

[20]If we are using cap volatilities, then equation 8 is also required to imply the volatility from the price.

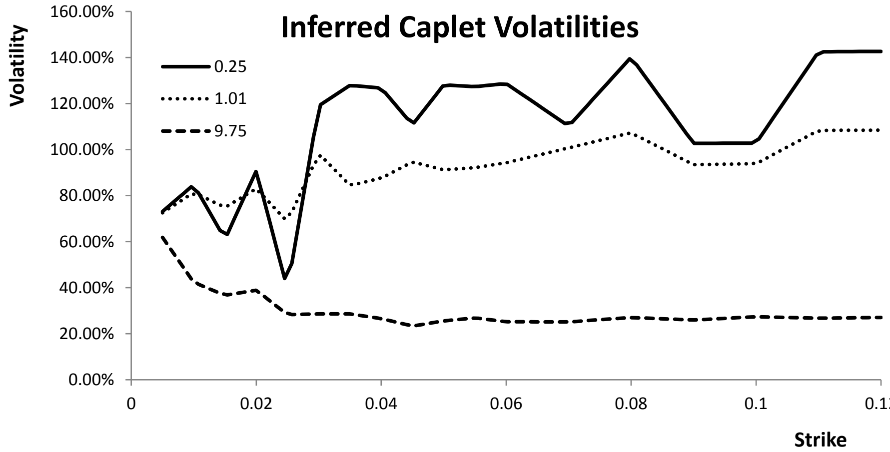

Figure 5: The result of the interpolation procedure (i.e. volatility smiles) for maturities of 3M, 1Y and 9Y9M.

Figure 6 shows the resultant volatility surface for $\lambda = 0.1$ - this recovers all the cap volatilities to less than 1bps. As in the interpolated term structure case, this 'surface' comes from joining the results from the 18 independent calibrations.

4.3.2 P-Splines

The volatility term structure can be expressed in terms of basis functions (B-splines) as [dB78]

$$\sigma(T) = \sum_{i=1}^{N} w_i f(T - T_i) \tag{15}$$

where the basis function, $f(x)$, is an even function made from n^{th} order polynomial pieces spanning $n+2$ evenly spaced knots: a first order B-spline is made from two linear parts to form a triangle; while a second order consists of three quadratic parts, which are zero at the two end knots and match value and first derivative at the two internal knots [EM96].

If we partition the time between the first and last caplet expiry into M evenly spaced knots, then we will have $N = M + n - 1$ basis functions, with the vector of weights, $\mathbf{w}$ controlling the shape of the volatility curve. To have enough flexibility in the curve, we would have more weights than there are caps (at a particular strike). Similar to the direct method, the system is solved by applying a penalty to the smoothness of the weights.[21]

If the knots are placed at the caplet expiry, this method is effectively identical to the direct method. With fewer knots (which makes the system smaller, thus quicker to calibrate), the results are still very similar to the direct method, and we do not show them here.

[21] We generate the penalty matrix from the second order difference in the weights. Again see appendix B.

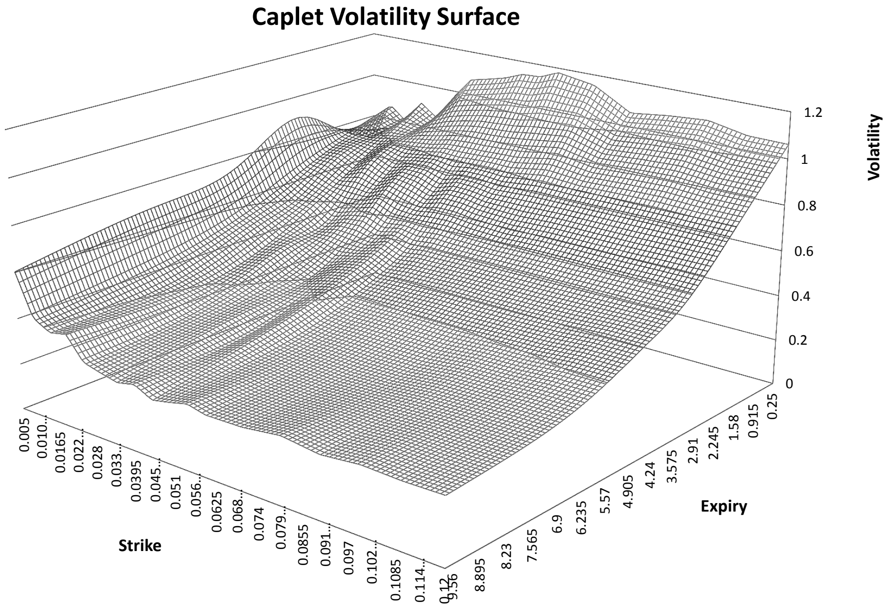

Figure 6: The caplet volatility surface from a series of independent calibrations to separate strikes using direct caplet volatility fits with penalty.

5 Global Stripping Methods

To impose smoothness on the smiles and to incorporate ATM and relative (strike) market quotes, we need some sort of global caplet volatility fitter.

If the caplet volatility surface, $\sigma(K, T)$, or simply the discrete set of caplet volatilities are described by a set of parameters $\boldsymbol{\theta}$, then there is a mapping from these parameters to the cap values which result from pricing the underlying caplets, which may be expressed generically in our usual form as

$$\mathbf{v} = G(\boldsymbol{\theta})$$

Below we discuss some models of this form.

5.1 Smile Model Based Strippers

In this section we discuss using smile models with expiry dependent parameters to form a caplet volatility surface. This approach is advocated in [AP10b].

A smile model is any model that produces a volatility smile - SABR[HKLW02], Heston[Hes93] and SVI[Gat06] are all examples. We do not concern ourselves here with the assumed dynamics of the forward rate that lead to the (implied) volatility smiles.[22] We simply treat them as a mapping from a set of parameters to a (hopefully arbitrage free) volatility smile.

[22]SABR and Heston are stochastic volatility models and SVI (stochastic volatility inspired) is simply a formula that gives plausible smiles.

The (Black) volatility produced by a model can be expressed as

$$\sigma(K) = s(K; T, F; \boldsymbol{\phi}) \tag{16}$$

This notation indicates that the smile (a function of strike, K) in general depends on state variables (the forward rate, F, and the expiry, T) and model parameters, $\boldsymbol{\phi}$. For the case of SABR, there are four model parameters, which are usual given the symbols, α, β, ρ and ν.

By making the model parameters expiry dependent, we may extend the smile to a volatility surface (function of strike and expiry), i.e.

$$\sigma(K, T) = s(K, T; F(T); \boldsymbol{\phi}(T)) \tag{17}$$

Again for SABR, there would be four parameter term structures $\alpha(T)$, $\beta(T)$, $\rho(T)$ and $\nu(T)$.[23]

The problem is then to find the parameters term structure, $\boldsymbol{\phi}$, that gives a (caplet) volatility surface that will minimise the error between the model and market cap values in a least square sense.

5.1.1 SABR with interpolated parameter term structures

We now give a concrete example using the SABR model. There is somewhat of a trade off between β and ρ (they both affect the skew), so we choose to make β a single global parameter; α and ν are represented by double-quadratic interpolated curves with knots at 1, 2, 3, 5, 7 and 10 years; ρ is also represented by double-quadratic interpolated curve with knots at 1, 3 and 7 years.

We calibrate to the cap volatilities using an error of 100bps. This results in a chi-square of 93.6 (or 0.86 per cap), so the RMS difference between model and market volatility is 86bps - this is shown graphically in figure 7.

It is a feature of this approach that the market values are not fully recovered; SABR (with a fixed β) has three parameters, so cannot in general recover option prices at more than three strikes. We have 18 absolute strikes, plus the ATM, so even allowing for an arbitrarily complex term structure, full recovery is not possible.[24] The calibrated parameter term structures are shown in figure 8. While these are relatively smooth, we had to experiment with the choice of knot positions to obtain this. In particular, ρ has a small number of knots to suppress oscillations which occur with more knots. This manual turning is not practical, and we would prefer a methodology that automatically guaranteed smoothness of the term structure.

The smooth term structures of parameters does in turn lead to a smooth (caplet) volatility surface. This is shown in figure 9. Whether the relatively poor recovery of market values is acceptable depends on what exactly the inferred caplet volatilities will be used for. One advantage of this approach (other than producing smooth volatility surfaces) is that smile model parameters exist for every expiry (in the range from the first to last caplet expiry), so extrapolation to higher or lower strikes using the model is trivial.

5.1.2 Smile Based Stripper using P-Splines

The need to fine tune the knots to avoid (smile model) parameter oscillations is a major drawback to representing the parameters as interpolated curves. An alternative representation of the

[23] This is not the same as some extensions of SABR which make the parameters appearing in the stochastic differential equation (SDE) time dependent. Our approach essentially has separate dynamics for every forward Libor rate.

[24] This is also true for Heston and SVI which have five parameters.

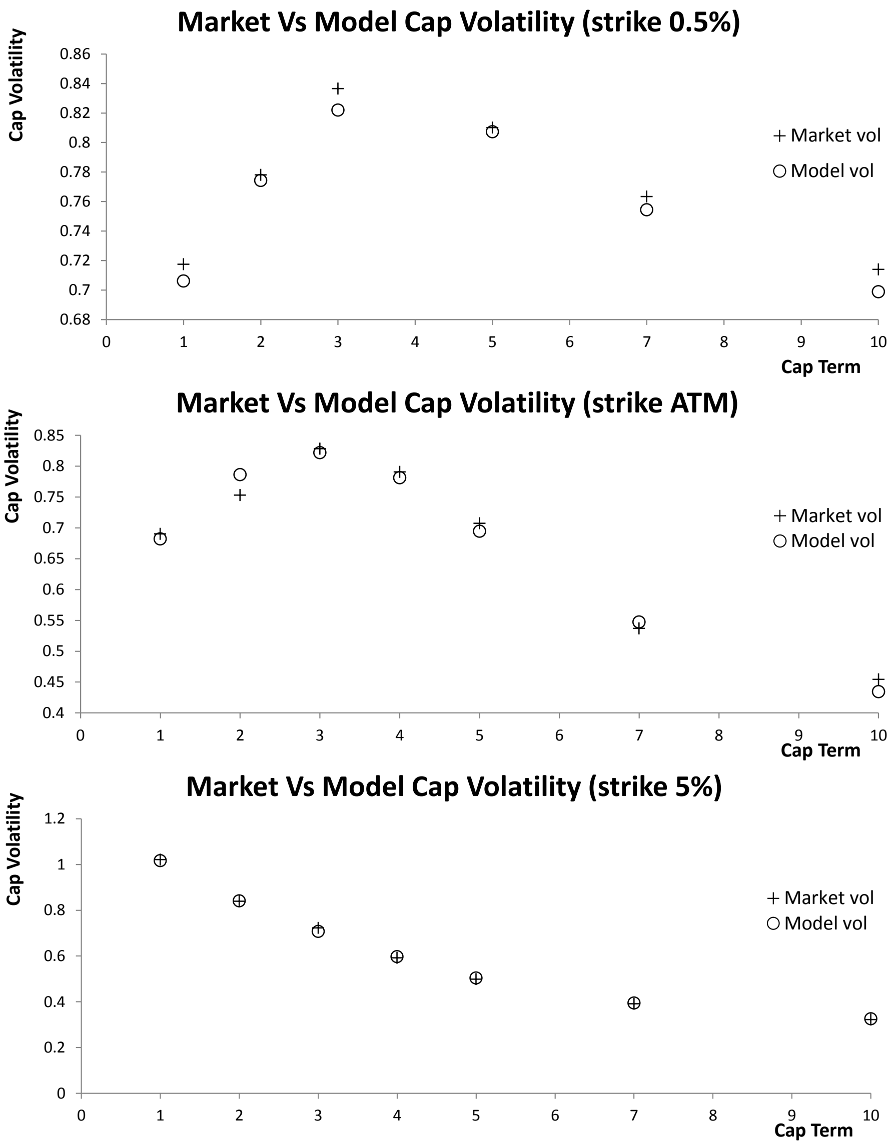

Figure 7: The market vs model cap prices (expressed as volatilities) at three strikes (absolute strikes of 0.5% and 5%, and ATM) for a SABR based caplet stripper.

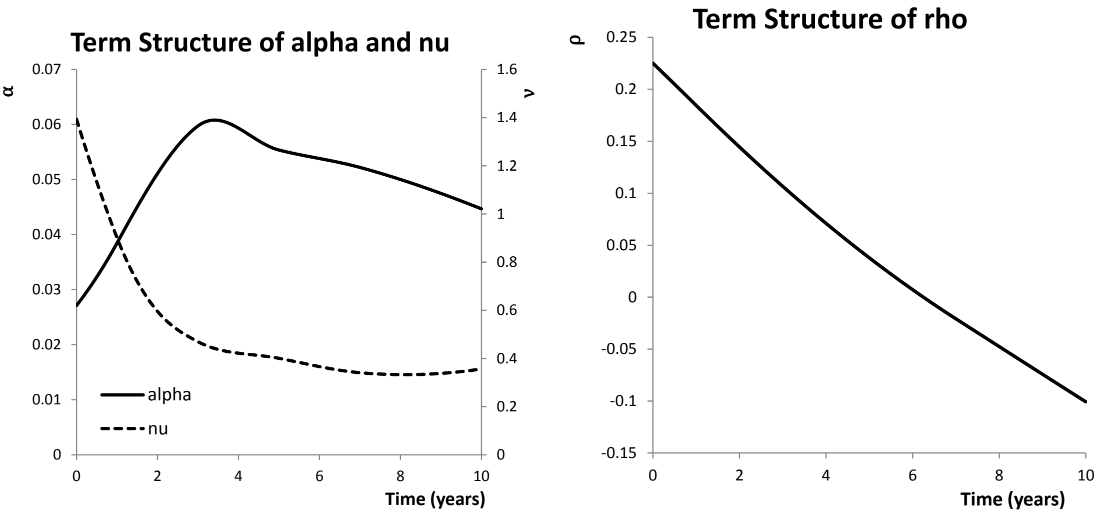

Figure 8: The calibrated SABR parameters. We treat β as a global parameter (i.e. no time dependance), and its calibrated value is 0.479.

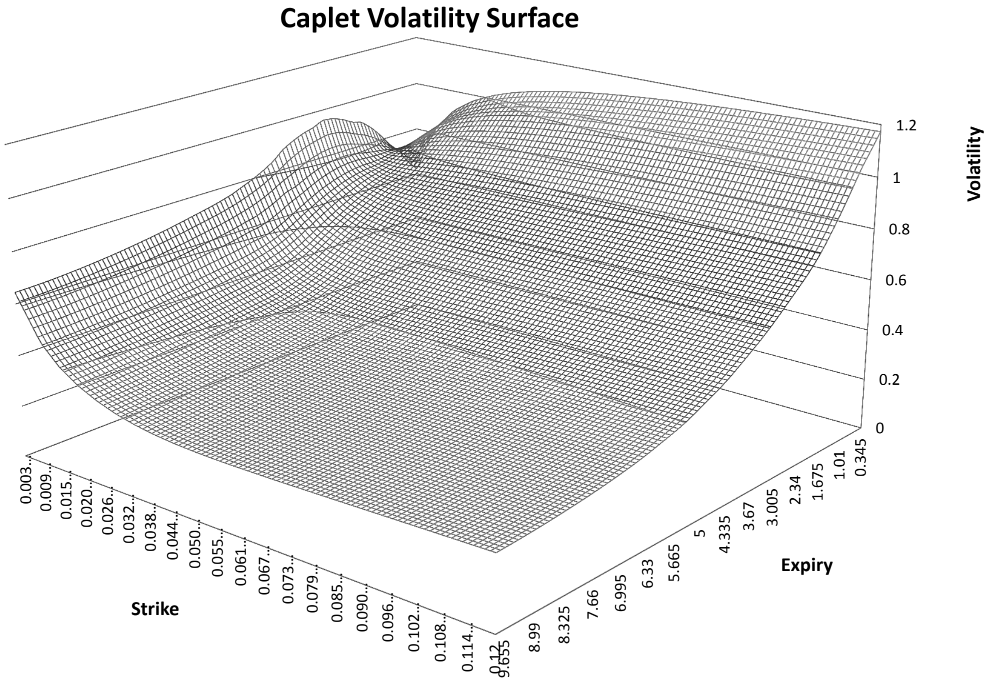

Figure 9: A smooth caplet volatility surface resulting from the SABR based caplet stripping (with interpolated parameter term structures) discussed in the main text.

curves is to use P-splines (as in section 4.3.2) with a (separate) penalty on the curvature of each parameter curve. While this can produce plausible parameter term structures, it is the stiffness of the smile model (in the strike direction) that prevents a close fit to the market - this is true regardless of how the parameter term structures are represented.

5.2 Penalty Function based Strippers

5.2.1 Direct Fitting

An alternative global method is based on a nonparametric approach in which constituent caplet volatilities are fitted directly such that market cap values (prices or volatilities) are recovered and smoothness across the caplet volatilities is achieved. In other words, the caplet volatilities are our model parameters. Let σ be the vector of caplet volatilities ordered firstly by strike then by expiry (so caplets with the same strikes will appear next to each other), then we wish to solve

$$\text{minimise} \quad (\mathbf{v} - G(\boldsymbol{\sigma}))^T (\mathbf{v} - G(\boldsymbol{\sigma})) + \boldsymbol{\sigma}^T \mathbf{P} \boldsymbol{\sigma} \tag{18}$$

which is just equation 14, except now we have all the caplet volatilities, not just those at a particular strike. The form of the penalty matrix is also more complex; it must impose smoothness in both the expiry and strike directions.

We first consider the case without the ATM caps. Since the longest dated cap (10 years) has quotes for all 18 absolute strikes, the caplets lie on a 18 by 39 grid in strike-expiry space with no gaps. If we let $\mathbf{P}_K$ and $\mathbf{P}_T$ be the penalty matrices that act in the strike and expiry direction respectively[25] (see appendix B), then the full penalty matrix that acts on the vector $\boldsymbol{\sigma}$ is given by

$$\mathbf{P} = \lambda_K \mathbf{I}_{n_T} \otimes \mathbf{P}_K + \lambda_T \mathbf{P}_T \otimes \mathbf{I}_{n_K} \tag{19}$$

where $\mathbf{I}_{n_T}$ and $\mathbf{I}_{n_K}$ are respectively size 39 and size 18 identity matrices, and λ_K and λ_T respectively the strength of the smoothness penalty in the strike and expiry direction.

The fit is made directly to the cap volatilities with an error of 1bps and a common λ of 0.03.[26] With this set up we achieve an RMS match to the market of just under 1bps. By reducing λ further we can increase the accuracy of the fit, but 1bps should be well within the bid-offer, so we consider this an acceptable recovery of the market. Figure 10 shows the resultant (caplet) volatility surface. This should be compared with the equivalent figure for the SABR model (fig. 9). Globally they have very similar shapes; the difference is that the direct fit is less smooth in the strike direction, and this is the difference between fitting the market cap volatilities to 1bps vs 100bps.

We excluded the ATM quotes so that the remaining caplets lay on a grid with no gaps - this allowed the penalty matrix to be expressed as a Kronecker product between (a suitably sized) identity matrix and a 1-D penalty matrix. Including the 7 ATM quotes produces gaps in the grid. We could craft a penalty matrix that handles this; however, in this case, a simpler approach is to introduce *phantom* caplets to fill in the gaps. These have no effect on the model cap values (since they do not actually form part of any caps), and since we are penalising curvature, they will end up with values that are linearly interpolated from the actual caplet volatilities.

There is some inconstancy in our market data between the ATM quotes and the absolute strike quotes (see appendix A). For this reason we fit to the ATM quotes with an error of 1bps and the rest with an error of 10bps, and use $\lambda_T = 0.01$ and $\lambda_K = 0.0002$. This achieves a

[25] They are 18 by 18 and 39 by 39 respectively.

[26] If we changed the error and scaled λ by the square of this, we will obtain the same result.

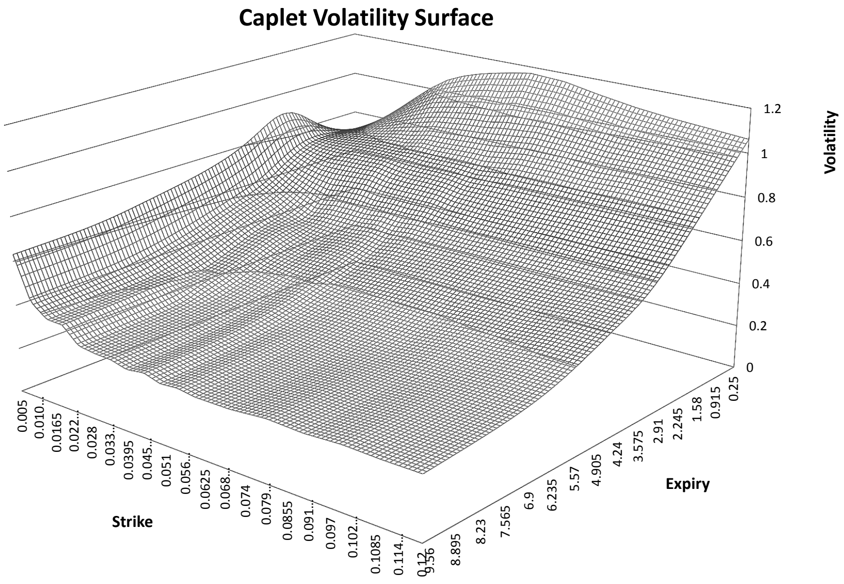

Figure 10: A smooth caplet volatility surface resulting from a direct fit to the caplet volatility with a penalty on the curvature in both the strike and expiry directions. Here the ATM cap quotes have been excluded.

chi-squared per cap of 1.2. Figure 11 shows the resultant caplet volatility surface. Since we have reduced λ_K by two orders of magnitude, the surface is inevitably less smooth in the strike direction.

Without the ATM caps we have 702 caplets, thus 702 variables to solve for. With the inclusion of the ATM quotes, we have 823 caplets, but using phantom caplets this jumps to 975 (25×39). These are all relatively large systems, which makes calibration slow.

If caps are quoted with relative strikes, the situation is worse. For 1, 2, 3, 4, 5, 7, and 10 years caps with, say, 9 relative strikes, we have 1089 unique caplets (assuming 3M caplets with the first period not paid). If we used the method above of adding phantom caplets, we would have a total of 2457. At this point it may be better to look for an alternative approach.

5.2.2 P-Splines again

A surface can be represented using B-Splines; the 2-D basis functions are just the Kronecker product of their 1-D counterparts. The volatility at some point on the surface is given by

$$\sigma(K,T) = \sum_{i=1}^{N_K}\sum_{j=1}^{N_T} w_{i,j} f_{i,j}(K,T) \tag{20}$$

where N_k and N_T are the number of basis functions in the strike and expiry directions. If the weights matrix is flattened row-wise then we can generate a penalty matrix (on the curvature of the weights) as above. This gives us the problem where the weights are the model parameters,

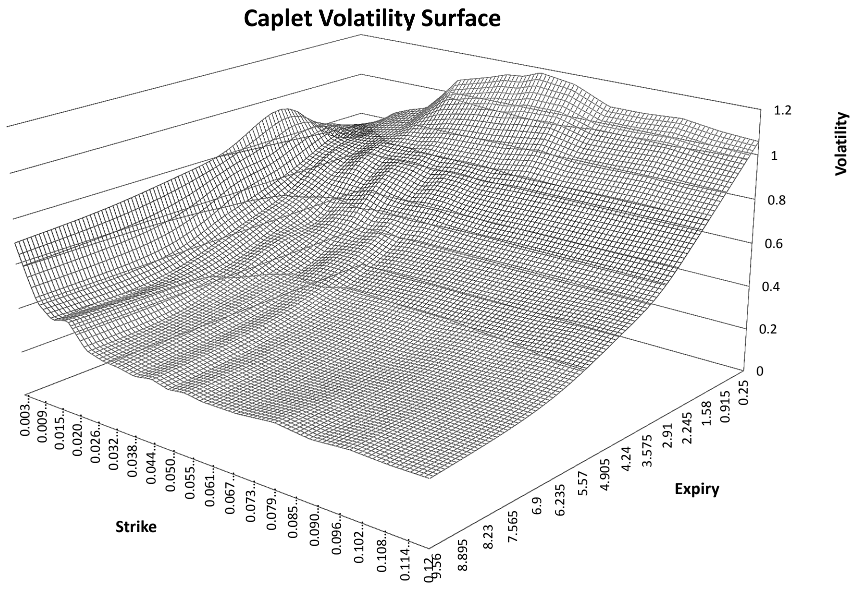

Figure 11: A smooth caplet volatility surface resulting from a direct fit to the caplet volatility with a penalty on the curvature in both the strike and expiry directions. This includes the ATM quotes.

i.e.

$$\text{minimise} \quad (\mathbf{v} - G(\mathbf{w}))^T (\mathbf{v} - G(\mathbf{w})) + \mathbf{w}^T \mathbf{P} \mathbf{w} \tag{21}$$

The advantage of this method, over directly fitting the caplet volatilities, is firstly there is no restriction on the strike type (absolute or relative), and secondly we can make do with a much smaller system.

We find that second order basis functions with 10 internal knots in the time direction, and third order basis functions with 25 internal knots in the strike direction[27] gives sufficient flexibility to fit the market data. Using $\lambda_k = 100$ and $\lambda_T = 1000$, we obtain an RMS fit of 0.4bps when the ATM caps are excluded, which degrades to around 10bps when they are included. The resultant caplet volatility surface is very similar to that obtained with the direct method (i.e. figure 10) - we do not show it for brevity.

6 Conclusion

We have demonstrated several methods of inferring caplet/floorlet volatilities from the market prices of interest rate caps/floors - so called caplet stripping. Some of these methods will recover exactly (or to some high tolerance) the market prices, but provide no guarantee that the resultant caplet smiles are free from arbitrage; other methods, based on smile models, will guarantee

[27]This gives a total of 434 2D basis functions.

arbitrage free caplet smiles,[28] but will not recover the market prices.

It may be possible to build a lack of arbitrage as a constraint into the penalty methods - i.e. find the smoothest arbitrage free (caplet) volatility surface that recovers the market prices. Of course, the market data may be such that this surface does not exist; then there is (again) a choice between exactly fitting the market and having some arbitrage. This, however, is outside the scope of this paper.

A Market data

In the tables below we show the market data used in the examples in this paper. It is from early 2013, and is provided to allow easy reproduction of the results given here. The same data can be found in the OpenGamma test code.[29]

The original data requires some cleaning before it can be used. To highlight problems, in figure 12 we plot the one and ten year cap/floor 'smiles', using all the available volatility data. We see that, especially for the one year data, there are points that are clearly outliers which should be removed.

The cap volatility data (table 1) has been manually cleaned to remove outliers (these are usually stale prices), and is a representative volatility grid suitable for testing the different caplet stripping methodologies discussed in the main paper. We have left the ATM points in our data set, even though they are often inconstant with close-by absolute strike quotes. This inconstancy does affect the fits made in the main paper.

This paper is not concerned with calibration of the funding (or discount) and (relevant) index curves. Details of the approach used by OpenGamma can be found here [Whi12, Hen12]. For reference we provide the knots of the interpolated curves[30] in table 2 and show these curves in figure 13.

B Difference, Derivative and Penalty Matrices

B.1 Difference Matrix

We define the n^{th} order difference matrix acting on a vector of length m, as a $m \times m$ matrix with the first n rows set to zero. The zeroth order matrix is trivially the identity matrix, and the first and second order difference matrices are

$$\mathbf{D_1} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \quad \mathbf{D_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix} \tag{22}$$

with obvious extensions to higher differences.

[28] A caveat is required here: this is only true if the smile model is arbitrage free; our example used Hagan's SABR formula which is not in fact arbitrage free.

[29] com.opengamma.analytics.financial.interestrate.capletstripping

[30] We have used a double-quadratic interpolator in both cases.

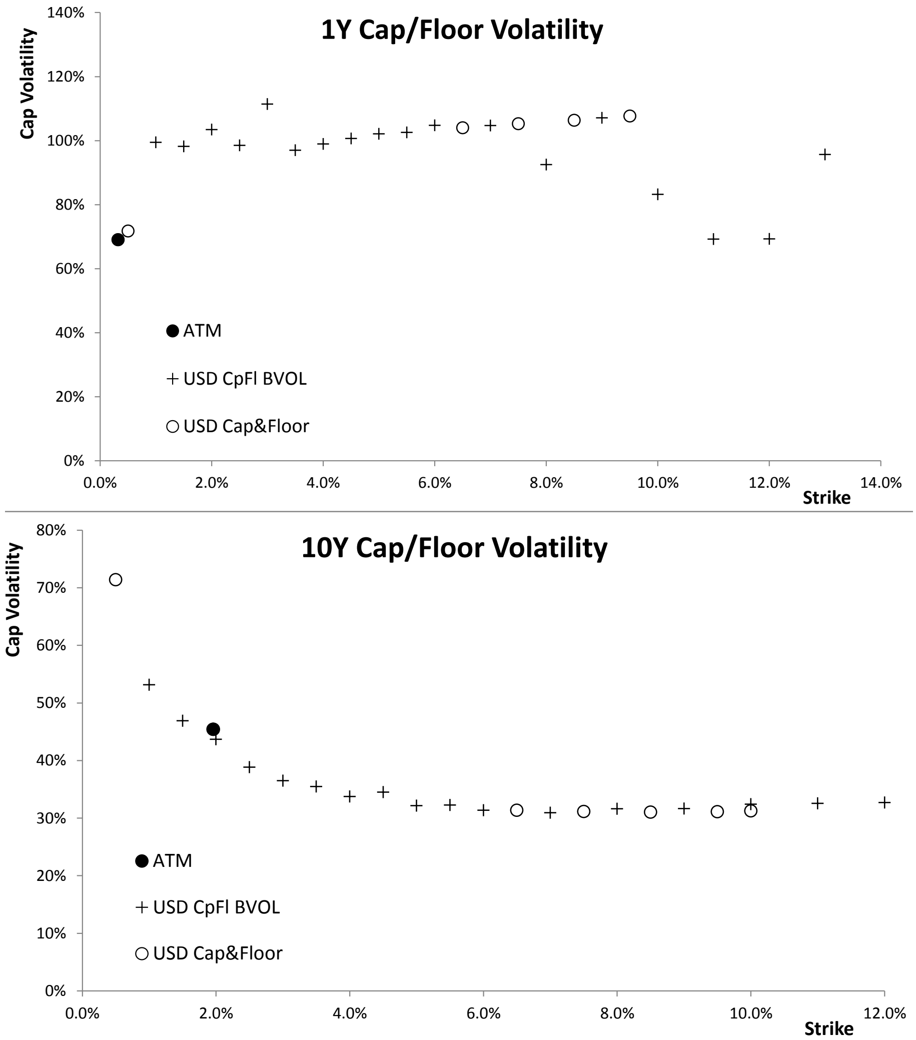

Figure 12: Data on USD cap/floor volatility.

		Cap Length (Years)						
		1	2	3	4	5	7	10
Cap Strikes	0.5%	71.75%	77.81%	83.66%		81.01%	76.33%	71.40%
	1.0%			75.23%	70.56%	66.10%	59.33%	53.13%
	1.5%		78.09%	73.99%	66.70%	61.47%	53.50%	46.91%
	2.0%				63.46%	56.98%	48.24%	43.69%
	2.5%		80.50%	73.41%	60.81%	54.72%	46.45%	38.85%
	3.0%				61.06%	52.87%	43.20%	36.50%
	3.5%	96.98%	82.27%	73.76%		51.68%	41.83%	35.49%
	4.0%	98.93%	83.76%	72.74%	59.83%	51.69%	40.83%	33.75%
	4.5%	100.63%	83.62%		60.03%			34.50%
	5.0%	102.13%	83.91%	72.26%	59.21%	49.84%	39.14%	32.16%
	5.5%	102.55%	84.06%	71.96%	59.62%	50.35%	38.73%	32.27%
	6.0%	104.76%	84.11%	70.72%	58.85%	50.06%	38.56%	31.35%
	7.0%	104.67%		70.66%	58.55%	50.17%	38.24%	30.93%
	8.0%			70.50%	58.46%	49.90%	38.02%	31.60%
	9.0%		84.58%	70.35%	58.34%	49.84%	38.91%	31.64%
	10.0%		84.89%	70.21%	58.25%	49.99%	39.16%	32.39%
	11.0%			70.09%	58.24%	50.59%	40.05%	32.55%
	12.0%			69.97%	58.18%	50.59%	40.14%	32.71%
	ATM	69.03%	75.30%	82.84%	79.07%	70.74%	53.70%	45.42%

Table 1: Volatility grid for USD caps. Outliers have been manually removed from the table.

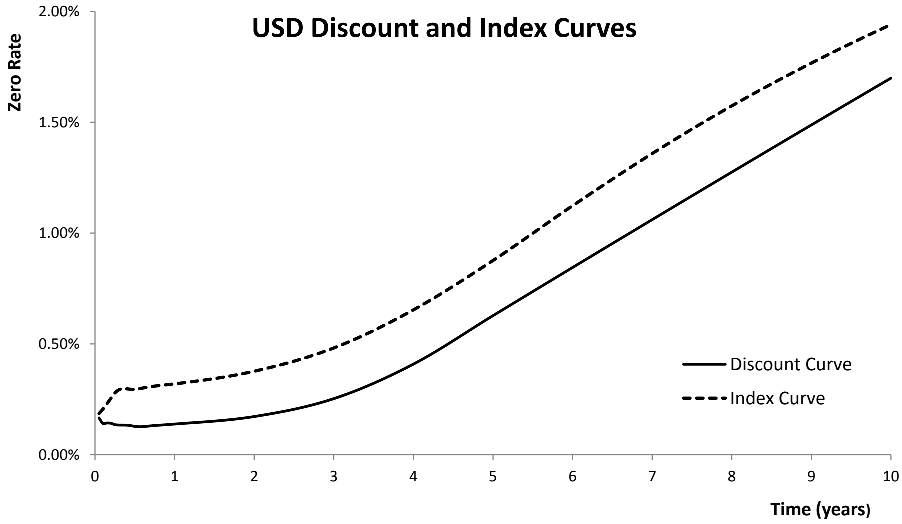

Figure 13: The USD discount (funding) and index (3m Libor) curves used in this paper.

Time	Zero Rate
0.003	0.213%
0.088	0.144%
0.173	0.145%
0.255	0.135%
0.345	0.134%
0.425	0.133%
0.507	0.128%
0.759	0.132%
1.005	0.139%
2.005	0.173%
3.010	0.254%
4.005	0.410%
5.005	0.629%
10.005	1.700%

Time	Zero Rate
0.044	0.184%
0.088	0.201%
0.173	0.241%
0.255	0.281%
0.507	0.295%
0.759	0.310%
1.005	0.320%
2.014	0.378%
3.010	0.483%
4.005	0.655%
5.005	0.878%
6.005	1.125%
7.018	1.363%
8.011	1.576%
9.008	1.768%
10.008	1.942%
15.007	2.540%
20.008	2.825%
25.011	2.986%
30.014	3.112%

Table 2: The knot values for the interpolated discount curve (left) and index curve (right).

The associated penalty matrix is given by

$$\mathbf{P_n} = \mathbf{D_n}^T \mathbf{D_n}$$

B.2 Derivative Matrix

For a (non-uniform) set of points $\mathbf{x} = (x_1, x_2, \ldots, x_i, \ldots, x_m)^T$ and a function evaluated at those points (i.e. $y_i = f(x_i)$), the n^{th} order derivative matrix, $\boldsymbol{\mathcal{D}_n}$, is such that $\boldsymbol{\mathcal{D}_n}\mathbf{y}$ is the finite difference estimate of the n^{th} order derivative of the function at the points, i..e.

$$(\boldsymbol{\mathcal{D}_n}\mathbf{y})_i \approx \left.\frac{d^n y}{dx^n}\right|_{x=x_i}$$

The three-point estimate (suitable for $n = 1$ & 2) can be found here [Whi13].

We then define the penalty matrix as

$$\mathbf{P_n} = \frac{1}{(x_m - x_1)^{2n}} \boldsymbol{\mathcal{D}_n}^T \boldsymbol{\mathcal{D}_n}$$

where the first term is present to make the penalty matrix scale invariant.

To impose some level of smoothness, often one wished to have a penalty of the form

$$\lambda \int_a^b \left(\frac{d^n y}{dx^n}\right)^2 dx$$

The quantity $\mathbf{x}^T\mathbf{P_n}\mathbf{x}$ is an approximation of this (up to a scaling factor) only if the points $\mathbf{x}$ are uniformly spaced. However as the sum of squares of the n^{th} order derivatives, it is sufficient as a smoothness criteria.

References

[AP10a] Leif Andersen and Vladimir Piterbarg. *Interest Rate Modeling – Volume I: Foundations and Vanilla Models*. Atlantic Financial Press, 2010.

[AP10b] Leif Andersen and Vladimir Piterbarg. *Interest Rate Modeling – Volume III: Products and Risk Management*. Atlantic Financial Press, 2010.

[Bla76] Fischer Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3:167–179, 1976.

[BM06] Damiano Brigo and Fabio Mercurio. *Interest Rate Models, Theory and Practice*. Springer Finance. Springer, second edition, 2006.

[Bro65] C G Broyden. A Class of Methods for Solving Nonlinear Simultaneous Equations. *Mathematics of Computation (American Mathematical Society)*, 19 (92):577–593, 1965.

[dB78] Carl de Boor. *A Practical Guide to Splines*. Springer, 1978.

[EM96] Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.

[Gat06] J. Gatheral. *The Volatility Surface: a Practitioner's Guide*. Finance Series. Wiley, 2006.

[Hen12] Marc Henrard. Multi-curves: Variations on a Theme. Quantitative Research 6, OpenGamma, October 2012. Available at docs.opengamma.com.

[Hes93] Steven Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6(6):327–343, September 1993.

[HKLW02] Patrick Hagan, Deep Kumar, Andrew Lesniewski, and Diana Woodward. Managing smile risk. *Wilmott Magazine*, Sep:84–108, 2002.

[Lev44] Kenneth Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.

[PTVF07] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brain P. Flannery. *Numerical Recipes*. Cambridge University Press, 2007.

[Reb02] Riccardo Rebonato. *Modern pricing of interest-rate derivatives: the LIBOR Market Model and Beyond*. Princeton University Press, Princeton and Oxford, 2002.

[Res13] Quantitative Research. *Interest Rate Instruments and Market Conventions Guide*. OpenGamma, 2.0 edition, December 2013. Available at www.opengamma.com.

[Whi12] Richard White. Multiple Curve Construction. Technical report, OpenGamma, 2012. Available at docs.opengamma.com.

[Whi13] Richard White. Numerical Solutions to PDEs with Financial Applications. Technical report, OpenGamma, 2013. developers.opengamma.com/quantitative-research/Numerical-Solutions-to-PDEs-with-Financial-Applications-OpenGamma.pdf.

OpenGamma Quantitative Research

1. Marc Henrard. Adjoint Algorithmic Differentiation: Calibration and implicit function theorem. November 2011.
2. Richard White. Local Volatility. January 2012.
3. Marc Henrard. My future is not convex. May 2012.
4. Richard White. Equity Variance Swap with Dividends. May 2012.
5. Marc Henrard. Deliverable Interest Rate Swap Futures: Pricing in Gaussian HJM Model. September 2012.
6. Marc Henrard. Multi-Curves: Variations on a Theme. October 2012.
7. Richard White. Option pricing with Fourier Methods. April 2012.
8. Richard White. Equity Variance Swap Greeks. August 2012.
9. Richard White. Mixed Log-Normal Volatility Model. August 2012.
10. Richard White. Numerical Solutions to PDEs with Financial Applications. February 2013.
11. Marc Henrard. Multi-curves Framework with Stochastic Spread: A Coherent Approach to STIR Futures and Their Options. March 2013.
12. Marc Henrard. Algorithmic Differentiation in Finance: Root Finding and Least Square Calibration. January 2013.
13. Marc Henrard. Multi-curve Framework with Collateral. May 2013.
14. Yukinori Iwashita. Mixed Bivariate Log-Normal Model for Forex Cross. January 2013.
15. Yukinori Iwashita. Piecewise Polynomial Interpolations. May 2013
16. Richard White. The Pricing and Risk Management of Credit Default Swaps, with a Focus on the ISDA Model. September 2013
17. Yukinori Iwashita. Conventions for Single-Name Credit Default Swaps. December 2013
18. Arroub Zine-eddine. Brazilian Swaps. December 2013.
19. Arroub Zine-eddine. Inflation: Instruments and curve construction. January 2014.
20. Arroub Zine-eddine. Inflation caps and floors. January 2014.
21. Arroub Zine-eddine. Convexity adjustment for inflation derivatives. January 2014.
22. Richard White. Forward CDS, Indices and Options. March 2014
23. Yukinori Iwashita. Tree Option Pricing Models. April 2014

About OpenGamma

OpenGamma helps financial services firms unify their calculation of analytics across the traditional trading and risk management boundaries.

The company's flagship product, the OpenGamma Platform, is a transparent system for front-office and risk calculations for financial services firms. It combines data management, a declarative calculation engine, and analytics in one comprehensive solution. OpenGamma also develops a modern, independently-written quantitative finance library that can be used either as part of the Platform, or separately in its own right.

Released under the open source Apache License 2.0, the OpenGamma Platform covers a range of asset classes and provides a comprehensive set of analytic measures and numerical techniques.

Find out more about OpenGamma
`www.opengamma.com`

Download the OpenGamma Platform
`developers.opengamma.com/downloads`

Europe
OpenGamma
185 Park Street
London SE1 9BL
United Kingdom

North America
OpenGamma
125 Park Avenue
25th Floor, Suite 2525
New York, NY 10017
United States of America

OpenGamma