Malz, Allan M.

Working Paper

A simple and reliable way to compute option-based risk-neutral distributions

Staff Report, No. 677

Provided in Cooperation with:
Federal Reserve Bank of New York

Suggested Citation: Malz, Allan M. (2014) : A simple and reliable way to compute option-based risk-neutral distributions, Staff Report, No. 677, Federal Reserve Bank of New York, New York, NY

This Version is available at:
https://hdl.handle.net/10419/120801

Federal Reserve Bank of New York
Staff Reports

A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions

Allan M. Malz

Staff Report No. 677
June 2014

A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions

Allan M. Malz

Federal Reserve Bank of New York Staff Reports, no. 677
June 2014
JEL classification: G01, G13, G17, G18

Abstract

This paper describes a method for computing risk-neutral density functions based on the option-implied volatility smile. Its aim is to reduce complexity and provide cookbook-style guidance through the estimation process. The technique is robust and avoids violations of option no-arbitrage restrictions that can lead to negative probabilities and other implausible results. I give examples for equities, foreign exchange, and long-term interest rates.

Key words: option pricing, risk-neutral distributions

Malz: Federal Reserve Bank of New York (e-mail: amalz@nyc.rr.com). The author thanks Sirio Aramonte, Bhupinder Bahra, Benson Durham, Stephen Figlewski, Will Melick, Carlo Rosa, Joshua Rosenberg, Ernst Schaumburg, and seminar participants at the Board of Governors of the Federal Reserve System for comments. Juan Navarro-Staicos, Kale Smimmo, and Steven Burnett have collaborated on the implementation of the techniques described here. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

Contents

1 Introduction

Risk-neutral probability distributions (RNDs) of future asset returns based on the option-implied volatility smile have been available to researchers in finance for decades. These techniques, however, are difficult to implement, because rendering some option data suitable for this purpose requires a great deal of processing, and because the algorithms that compute the RNDs are complex and hard to automate. This is perhaps a major reason that option-based RNDs have been less widely applied and become less standard than might have been expected given their potential value.

This paper describes a simple technique for computing RNDs given suitable input data, requiring relatively straightforward programming. While most elements of the technique have been employed in earlier work, their combination and sequencing as described here greatly reduce the effort required to obtain results. The aim of the technique is to reduce complexity and the aim of the paper is to provide cookbook-style guidance through the estimation process. We give examples for different types of assets: equities, foreign exchange, and long-term interest rates.[1]

Methods for computing RNDs from option prices are inspired by the Breeden and Litzenberger (1978) statement of the relationship between market prices of European call options and the RND: In the absence of arbitrage, the mathematical derivative of the call option value with respect to the exercise price is closely related to the risk-neutral probability that the future asset price will be no higher than the exercise price at option maturity.[2]

The payoff at maturity to a European call option maturing at time T, with an exercise price X, is $\max(S_T - X, 0)$, with S_T representing the terminal underlying price. We denote the observed time-t market value of a European call struck at X and with a tenor of $\tau = T - t$ by $c(t, X, \tau)$. Absent arbitrage, therefore, the option value is equal to the present expected value of the terminal payoff under the risk-neutral distribution:

$$c(t, X, \tau) = e^{-r_t \tau} \tilde{\mathbf{E}}_t \left[\max(S_T - X, 0)\right] = e^{-r_t \tau} \int_X^\infty (s - X) \tilde{\pi}_t(s) ds,$$

where

$S_t \equiv$ time-t underlying price
$r_t \equiv$ time-t continuously compounded financing rate
$\tilde{\mathbf{E}}_t[\cdot] \equiv$ an expectation taken under the time-t risk-neutral probability measure
$\tilde{\pi}_t(\cdot) \equiv$ time-t risk-neutral probability density of S_T

[1] The technique set out here is applied to the measurement of systemic risk in Malz (2013). It is also used in the Federal Reserve Bank of New York's market monitoring.

[2] Surveys of techniques for extracting RNDs from option prices include Jackwerth (1999) and (2004), and Mandler (2003). More recent approaches are cited later in this paper. The Breeden-Litzenberger theorem was first stated in Breeden and Litzenberger (1978) and Banz and Miller (1978).

Differentiate the market call price with respect to the exercise price X to get the "exercise-price delta"

$$\frac{\partial}{\partial X}c(t,X,\tau) = e^{-r_t\tau}\left[\int_0^X \tilde{\pi}_t(s)ds - 1\right]. \tag{1}$$

This result implies that the time-t risk-neutral cumulative distribution function $\tilde{\Pi}_t(X)$ of the future asset price—the probability that the terminal underlying price will be X or lower—is equal to one plus the future value of the exercise-price delta of a European call struck at X:

$$\tilde{\Pi}_t(X) \equiv \int_0^X \tilde{\pi}_t(s)ds = 1 + e^{r_t\tau}\frac{\partial}{\partial X}c(t,\tau,X). \tag{2}$$

Differentiate again to see that the time-t risk-neutral probability density function is the future value of the second derivative of the call price with respect to the exercise price:

$$\tilde{\pi}_t(X) = e^{r_t\tau}\frac{\partial^2}{\partial X^2}c(t,X,\tau). \tag{3}$$

Though we'll describe our technique in terms of the market's pricing schedule for call options, the put price schedule offers a more direct and intuitive way to state the relationship between option prices and risk-neutral probabilities:

$$\tilde{\Pi}_t(X) = e^{r_t\tau}\frac{\partial}{\partial X}p(t,\tau,X),$$

where $p(t,X,\tau)$ represents the time-t value of a European put struck at X and with a tenor of τ.

Figlewski (2010) provides some nice intuition for this statement. Consider the increasing value of a put option, for a given current market price of the underlying, as the exercise price varies from low to high. At very low exercise prices this function has a slope and value near zero, and at very high exercise prices a slope equal to $e^{r\tau}$ and a value near its intrinsic value. As we increase the exercise price from X to a nearby point $X+\Delta$, the risk-neutral expected future value of the payoff of the option increases by Δ times the risk-neutral probability that the option expires in-the-money, that is, $\tilde{\Pi}(X+\Delta)$:

$$\begin{aligned} & \Delta \times \tilde{\Pi}(X+\Delta) \approx e^{r_t\tau}[p(t,\tau,X+\Delta) - p(t,\tau,X)]. \\ \Rightarrow \quad & \tilde{\Pi}(X+\Delta) \approx \frac{1}{\Delta}e^{r_t\tau}[p(t,\tau,X+\Delta) - p(t,\tau,X)] \end{aligned}$$

It's well known, but worth reiterating, that RNDs are not the same as real-world probabilities, or the ones in market participants' heads, but are influenced, perhaps heavily, by risk preferences. A change in risk-neutral probabilities can be due to changes in real-world probabilities, or risk preferences, or both.[3]

2 Overview of the technique

The technique we present here works, in principle, for any asset type, provided data of acceptable quality are available. We'll sketch the approach here and give detail on how it's applied to different asset classes, as well as examples of the results, in subsequent sections.

2.1 Implied volatility data

The approach requires data of reasonably good quality on the Black-Scholes implied volatility smile. The data, that is, are Black-Scholes volatilities for European options of a given tenor τ, but with a range of exercise prices. The volatility smile changes over time and for varying tenors, and can be thought of as a slice through the maturity axis of a time-t Black-Scholes volatility surface $\sigma(t, X, \tau)$. We focus here on a single tenor, rather than the entire surface.

Although Black-Scholes volatilities are expressed in a metric drawn from a particular option pricing model, they are associated with market- rather than model-based prices. Denote the time-t Black-Scholes model value of a European call as

$$v(S_t, X, \tau, \sigma, r_t, q_t) = S_t e^{-q_t \tau} \Phi \left[\frac{\log\left(\frac{S_t}{X}\right) + \left(r_t - q_t + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right] - X e^{-r_t \tau} \Phi \left[\frac{\log\left(\frac{S_t}{X}\right) + \left(r_t - q_t - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right]$$

where

[3] The work surveyed in Garcia, Ghysels and Renault (2010) uses historical data on underlying asset prices as well as contemporaneous option price data to simultaneously estimate both the risk-neutral and real-world probability distributions. Ross (2013) presents a technique that, with suitable assumptions, identifies both the risk-neutral and real-world probabilities of discrete price outcomes from option prices alone.

$\sigma \equiv$ a Black-Scholes implied volatility
$q_t \equiv$ time-t continuously compounded cash flow yielded by the underlying asset

The volatility surface translates into the time-t market price schedule of European calls with different tenors and exercise prices via the relationship

$$c(t, X, \tau) = v[S_t, X, \tau, \sigma(t, X, \tau), r_t, q_t]. \tag{4}$$

We refer to the right-hand side of (4) as the call valuation function. This function is a standard Black-Scholes formula taking as its implied volatility argument the interpolated volatility corresponding to the given exercise price. It takes an observed or estimated market-adjusted Black-Scholes volatility, and returns an estimated market call price. We can view $c(t, X, \tau)$ and $\sigma(t, X, \tau)$ as simply two different metrics for expressing the market values of options.

The implied volatilities can be expressed in various other units, such as Black or normalized volatilities. The exercise prices can also be expressed in different ways, such as the ratio or spread to the current spot or forward price, or the option delta. But under all these conventions, implied volatilities can be transformed into option prices in currency units for given exercise prices.

One of the main challenges in fitting RNDs is the diversity of option data and the difficulty of working with it. That's not the problem we're solving here. Rather, we're attempting to find an easier way to process the option data into an estimated RND and minimizing the extent to which we add assumptions to the information contained in the data. The data we use for this paper are obtained from Bloomberg Financial LP, which aggregates and processes quotes, end-of-day prices, and indicative prices from a range of dealers and exchanges. As we'll describe in a moment, we subject the data to a set of quality diagnostics. While flaws do occasionally appear in the data, the overall quality is good.

The approach here can be applied to a wide range of data types. We've developed the technique for three input data structures. In each, the data on each date consist of two columns/rows, one containing implied volatilities and the other the associated exercise prices:

Asset class	Volatility type	Units	Exercise price metric
Exchange-traded	Black-Scholes volatilities	Pct. p.a.	Ratio to spot
Currencies and gold	Black-Scholes volatilities	Pct. p.a.	Spot delta
Swaptions	Black volatilities	Pct. p.a.	Bps from forward

We'll provide more detail on the data in a subsequent section on each structure.[4]

[4]Intraday data can be displayed for a given asset using the function OVDV. The documentation for

2.2 The technique in brief

The steps in the computation of the RND are:

- Interpolate and extrapolate the volatility smile data using a cubic spline function that is "clamped" at the endpoints. This is tantamount to assuming that implied volatilities for very deep out-of-the-money calls and puts are identical to those for the furthest in- and out-of-the-money strikes in the input data.
- Apply the call valuation function (4), taking the interpolated Black-Scholes volatilities and other inputs called for by the Black-Scholes formula as arguments, and returning an option value in currency units.
- Numerically difference the call valuation function with respect to the exercise price to approximate the risk-neutral cumulative probability and probability density functions. The step size for this differentiation is set so that the density function is non-negative.

The probability distribution and density functions are estimated by taking finite differences in exercise-price space of the call valuation function. Discretized versions of the option-based estimate of the risk-neutral cumulative probability distribution and density functions (2)–(3) for a step size Δ are given by

$$\tilde{\Pi}_t(X) \approx 1 + e^{r_t \tau} \frac{1}{\Delta} \left[c\left(t, X + \frac{\Delta}{2}, \tau\right) - c\left(t, X - \frac{\Delta}{2}, \tau\right) \right]$$

and

$$\begin{aligned} \tilde{\pi}_t(X) &\approx \frac{1}{\Delta} \left[\tilde{\Pi}_t\left(X + \frac{\Delta}{2}\right) - \tilde{\Pi}_t\left(X - \frac{\Delta}{2}\right) \right] \\ &= e^{r_t \tau} \frac{1}{\Delta^2} [c(t, X + \Delta, \tau) + c(t, X - \Delta, \tau) - 2c(t, X, \tau)]. \end{aligned}$$

As $\Delta \to 0$, these expressions converge to the risk-neutral distribution functions, but the propensity for negative probabilities increases.

While fairly standard, two key features in combination simplify the computation process without generating anomalies: the use of a clamped cubic spline to interpolate—and, more importantly, extrapolate—the volatility smile, and treating the differencing step size as a user setting. Both are intended primarily to avoid processing-induced violations of no-arbitrage restrictions. We'll discuss these problems in detail just below. But in a nutshell, if the input implied volatility data don't violate no-arbitrage restrictions, why should the interpolating function?

Bloomberg's implied volatility data is a bit sparse. Some notes and white papers can be downloaded via the function DRVD (Derivatives Documentation Center).

2.3 The volatility interpolating function

A cubic spline is constructed to have continuous first and second derivatives at all its knot points. The construction of a cubic spline involves solving a set of linear equations for the coefficients that impose continuity of the first and second derivatives. To complete the algorithm, additional conditions are imposed on the equations corresponding to the first and last, or boundary, knot points. A natural cubic spline is constructed so that the second derivatives at the boundary knot points are equal to zero. As a result, extrapolation beyond the boundary knot points is linear, but generally with a non-zero slope. This is precisely the behavior that may induce violations of the no-arbitrage bounds on the volatility smile.

A clamped cubic spline, in contrast, is constructed so that its slope takes on specific values at the boundary knot points.[5] The interpolated volatility function we use is constructed as a clamped cubic spline, with a slope of zero at the boundary knot points. We use the input data on the implied volatility smile as the knot points of the spline. The slope of the fitted spline is thus zero at the highest and lowest exercise prices in the data. The spline is smooth at those transitions, since continuity of the second derivatives is still imposed. The extrapolated spline values beyond those points are then equal to the observed implied volatilities for the highest and lowest exercise prices.[6]

Let $\{(x_1, \sigma_1), \ldots, (x_n, \sigma_n)\}$ represent the input data set or knot points, ordered so $x_i > x_{i-1}, i = 2, \ldots, n$, and let $f(x)$ represent the fitted clamped cubic spline. (As noted, the units of the x_i are different for different asset types.). The interpolating function with its flat-line extensions is defined as the piecewise function

$$\sigma(x) = \begin{cases} \sigma_1 & \text{for} \quad x < x_1 \\ f(x) & \text{for} \quad x_1 \leq x < x_n \\ \sigma_n & \text{for} \quad x \geq x_n \end{cases}$$

[5] Klugman, Panjer and Willmot (2008), pp. 534ff., provides the recipe for constructing a clamped cubic spline, as well those for natural and other types of cubic splines. The recipes are also contained in many other numerical math and computing books. Neuberger (2012) applies a cubic spline to interpolate volatility data; Neuberger (2012) and Carr and Wu (2009) extrapolate the implied volatilities of the boundary points, but without incorporating the clamping into the spline fitting procedure. Bliss and Panigirtzoglou (2002) and (2004) implement flat-line extrapolation of a natural spline by introducing additional synthetic data points with implied volatilities equal to those observed for the highest and lowest exercise prices and with exercise prices outside that range.

[6] Our interpolating function is not a smoothing spline, as employed for example by Bliss and Panigirtzoglou (2002) and (2004); it passes through, rather than close to, all the knot points. A smoothing spline requires additional procedures to ensure that no-arbitrage conditions are preserved after fitting the interpolating function, and introduces additional concerns about inferring rather than observing data.

The highest and lowest exercise prices for which implied volatility data are observed are generally not quite extreme enough to set the estimated risk-neutral probabilities equal to 0 or 1. It is therefore necessary to extrapolate the interpolated smile, and thus the estimated call valuation function, beyond those strikes to obtain a complete RND. Clamping the interpolated smile so that the extrapolated segments are parallel to the x-axis at the extreme implied vols ensures that the call valuation function is monotonic and convex to the origin in the exercise price, avoiding violations of no-arbitrage restrictions. Volatility smiles are typically U-shaped or L-shaped. Extrapolating a steep slope out to high or low exercise prices can cause, for example, a call to have a higher value than a call with a higher exercise price.

Flat-line extrapolation gives the tails of the fitted RND a lognormal shape beyond the highest and lowest exercise prices in the input data. Figlewski (2010) proposes the alternative of first estimating the central portion of the RND using the available input data, and then grafting tails onto it that follow a generalized extreme value (GEV) distribution. The GEV distribution has better empirical support than the lognormal as a description of extreme return behavior. The parameters of the GEV distribution for each tail are estimated by having it coincide with a "penultimate" tail segment of the observable data-based portion of the RND. However, if observable option price inputs are available for exercise prices deep in the tails, there is likely to be only a small impact on estimated probabilities, as these will be already very close to zero or one. If observable option prices do not extend far into the tails, the GEV distribution-based tails will be estimated from less suitable data closer to the center of the distribution.

Extrapolation raises an uncomfortable question: Are we just inventing the risk-neutral tail behavior our procedure will later appear to infer from the data? To some extent, the answer is yes. The input data have to be far enough out-of-the-money for the risk-neutral distributions to be accurate, and we shouldn't be too trustful of statements about outcomes far beyond the exercise prices in the input data. But it's unrealistic to expect data of acceptable quality to typically extend to the points on exercise price axis at which the risk-neutral density is very close to zero. The choice therefore is not whether to extrapolate, but how to extrapolate while adding as little assumed behavior as possible to the available data.

2.4 Addressing violations of no-arbitrage restrictions on the call valuation function

The key model-free arbitrage condition is that the European call valuation function is decreasing and convex with respect to the exercise price. These basic no-arbitrage restrictions imply corresponding restrictions or bounds on the shape of the volatility

smile.[7]

Since the typical volatility smile is U-shaped, flat-line extrapolation seems at first less accurate than continuing the up- or downward sloping behavior. But keeping the slope constant over the extrapolated intervals will at least sometimes lead to arbitrage violations. It makes sense in some contexts, such as the study of market liquidity, to admit the possibility that they occur, but construction of risk-neutral densities is not one of them.

(i) Violations of the slope restrictions

The first slope restriction states that the call value can't rise as the exercise price rises, that is, the exercise-price delta can't be positive:

$$\frac{\partial}{\partial X} c(t, X, \tau) \leq 0. \tag{5}$$

A related restriction pertains to put values, namely, that they are increasing in the exercise price. We can express the put restriction in terms of the exercise-price delta of a call by invoking put-call parity.[8] It states that the absolute value of the call's negative slope with respect to the exercise price can't exceed the risk-free discount factor:

$$\frac{\partial}{\partial X} c(t, X, \tau) \geq -e^{-r_t \tau}. \tag{6}$$

The validity of these restrictions can also be seen from (1), which shows the consequences of violating them: the risk-neutral cumulative probabilities will not tend toward zero (unity) for very low (high) terminal underlying prices, and will therefore not meet the definition of a probability distribution function.

Each of these restrictions leads to a restriction on the slope of the volatility smile. Differentiate (4) to express the slope of the call valuation function in terms of the

[7] The no-arbitrage restrictions on option values are laid out in many option-pricing textbooks, e.g. Cox and Rubinstein (1985), ch. 4. No-arbitrage restrictions on volatility smiles are laid out in Hodges (1996). Aït-Sahalia and Duarte (2003) discuss the no-arbitrage conditions on volatility smiles in relation to estimation of RNDs.

[8] To re-express this condition, differentiate the statement of put-call parity

$$p(t, X, \tau) = c(t, X, \tau) + X e^{-r_t \tau} - S_t$$

where $p(t, X, \tau)$ represents the put value, with respect to X to get

$$\frac{\partial}{\partial X} p(t, X, \tau) = \frac{\partial}{\partial X} c(t, X, \tau) + e^{-r_t \tau}.$$

slope of the volatility smile and of the Black-Scholes sensitivities with respect to the exercise price and volatility, denoted by argument subscripts:

$$\frac{\partial}{\partial X}c(t,X,\tau) = v_X(\cdot) + v_\sigma(\cdot)\frac{\partial}{\partial X}\sigma(t,X,\tau).$$

Substituting into the no-arbitrage restrictions (5)–(6) gives us

$$v_X(\cdot) + v_\sigma(\cdot)\frac{\partial}{\partial X}\sigma(t,X,\tau) \leq 0$$
$$v_X(\cdot) + v_\sigma(\cdot)\frac{\partial}{\partial X}\sigma(t,X,\tau) \geq -e^{-r_t\tau},$$

in turn implying an upper and a lower bound on the slope of the volatility smile:

$$\frac{\partial}{\partial X}\sigma(t,X,\tau) \leq -\frac{v_X(\cdot)}{v_\sigma(\cdot)} > 0 \tag{7}$$

$$\frac{\partial}{\partial X}\sigma(t,X,\tau) \geq -\frac{v_X(\cdot) + e^{-r_t\tau}}{v_\sigma(\cdot)} < 0. \tag{8}$$

The sign and magnitude of each of the bounds stated in (7)–(8) derives from those of the Black-Scholes sensitivities. The Black-Scholes exercise-price delta $v_X(\cdot)$, like that of the call valuation function, is negative and obeys $v_X(\cdot) > e^{-r_t\tau}$. For low exercise prices, $v_X(\cdot)$ is close to $-e^{-r_t\tau}$, that is, slightly flatter than -1 for short tenors and typical interest rates. For high exercise prices, it flattens toward a slope of zero. The Black-Scholes vega $v_\sigma(\cdot)$ is always positive, and bell-curve shaped for varying exercise prices. The upper bound (7) is thus positive, tending to zero for very high exercise prices, while the lower bound (8) is negative, tending to zero for very low exercise prices.

The Black-Scholes sensitivities also vary with the general level of implied volatility. For higher volatilities, $v_X(\cdot)$ rises more gradually toward zero as the exercise price rises, and $v_\sigma(\cdot)$ is higher for any exercise price. When volatility is high, the absolute values of the bounds are low and thus more constraining, since the denominator of (7)–(8) is large. A scheme for interpolating the volatility smile is therefore most apt to violate the restrictions on the slope of the volatility smile if the general level of volatility is high, and then only for very high or low exercise prices.

The extent to which the no-arbitrage constraints bind thus depends on the second-order Black-Scholes sensitivities with respect to implied volatility, which are widely used in option risk management. The most important are vanna, the sensitivity of vega to changes in the spot price, and volga, the sensitivity of vega to changes in volatility.

The no-arbitrage constraints (7)–(8) depend in part on volga and the "exercise-price vanna."[9]

We can think of the bounds in terms of typical U-shaped volatility smile behavior. The bounds permit both upward- and downward-sloping volatility smiles, so a U shape does not *per se* violate them. But the bounds also state that the volatility smile can't still be upward-sloping at very high exercise prices, and can't still be downward-sloping at very low exercise prices, unless vega has become exceptionally low in those intervals.

Figure 1 compares the results of the clamped cubic spline with flat-line extrapolation to an alternative polynomial interpolation and extrapolation scheme that adheres more closely to intuition about typical U-shaped smile behavior, using implied volatilities of 3-month options on the S&P 500 index for two dates. On the earlier date, Feb. 25, 2009, at the height of the post-Lehman financial panic, the general level of S&P 500 implied volatility was extremely high by historical standards. The flat-line extrapolation prevents the slope of the call valuation function from falling below $-e^{-r_t\tau}$ (sloping more steeply downward) for very low exercise prices, and from turning positive for high exercise prices. If you look closely, even on the later date, Dec. 21, 2012, although vol is much lower, the slope of the call function becomes a bit steeper than $-e^{-r_t\tau}$ for low exercise prices and positive for high exercise prices when the extrapolated volatilities are not clamped.[10]

There are infinite ways to interpolate the volatility smile that will not violate (7)–(8). The clamped cubic spline approaching we propose has the conceptual advantage that it adheres to the observable data, and adds little in the way of assumed RND behavior to the data. It has the practical advantages that is simple, and appears to work in all cases, making it suitable for software-like implementations requiring frequent or routinized calculations.

(ii) Violations of the convexity restrictions

The call valuation function must be convex to the origin. The convexity restriction can be written as

$$\frac{\partial^2}{\partial X^2} c(t, X, \tau) \leq 0.$$

If this restriction is violated over some range of exercise prices, it is possible to construct a butterfly consisting of long positions in the relatively cheap pair of options

[9] Castagna and Mercurio (2007) use vanna and volga to find the coefficients of a no-arbitrage implied volatility interpolating function in a stochastic-volatility model.

[10] Note also that our interpolation technique can induce concave "sneering" or "frowning" intervals into the generally "smirking" interpolated smile.

struck at the ends of the range and short positions in the relatively dear option struck at the middle of the range that brings in net premium now and can't lose money at maturity. Violations imply that the risk-neutral cumulative probability distribution is falling and that the probability density function is negative over at least some part of that range.

Even when the volatility smile appears to the eye to be quite smooth, it may still generate nonconvexities in the call valuation function over small exercise-price intervals, particularly near knot or inflection points. A good deal of smoothing of the call valuation function is accomplished by spline interpolation of the volatilities. Permitting users to vary the differencing step size Δ further smooths the interpolated volatility smile and avoids intervals over which the density function is negative.

If Δ is set low enough, negative densities result. We've constructed the algorithm so that the user can vary Δ to find a low value that nonetheless keeps the density positive everywhere on most days. Some experimentation shows that the estimated risk-neutral probabilities are not terribly sensitive to variations in Δ. That is, if Δ is set high to be confident that no negative densities are generated, or Δ is set low enough to induce negative densities over some exercise price intervals on some days, the estimated probabilities and quantiles are not drastically changed. We'll present an example in the next section.

The propensity to generate negative densities, not surprisingly, is greatest when the general level of volatility is high. A practical way to find a suitable Δ for a given asset is to plot the density function for a date on which implied volatility is relatively high. These dates are almost invariably in late 2008 and implied volatility is generally a multiple of the high volatilities observed in other subperiods of the time series. A minimum Δ can be readily found that does not induce negative densities, or induces only slightly negative densities on a handful of extreme-volatility dates. That Δ can be used to compute time series of tail probabilities, moments or quantiles. A procedure could be added to the technique to find a value of Δ that avoids negative densities for each asset on each day, though at the cost of longer computation time.

2.5 Diagnostic analysis of the technique

Diagnostics on the input data are useful to help users understand better how well the interpolation is working, how far the extrapolation might be straying from the unobservable market reality, and assess the potential for estimation error. We'll provide such a table for each of the three asset classes we cover. Among the key diagnostics:

- The option deltas tell us how far into the tails the observed data penetrate.

- The option vega is directly related to the no-arbitrage restrictions. If vega is high at the extremes of the input data, then the choice of extrapolation technique has greater potential to influence the shape of the distribution. The focus here is on how far the vega has fallen at the highest and lowest exercise prices, so we'll express the vega for each strike as its ratio to the vega of the at-the-money (ATM) option.

- A version of the risk-neutral distribution based only on the input data provides rough bounds for the risk-neutral distribution and gives us a sense of how much estimation error there might be. Rather than fixing Δ for the entire RND, we use the successive differences between the exercise prices of the options in the raw data. Let X_{i-1} and X_i be two of the exercise prices in the data set, ordered so $X_i > X_{i-1}$. Then

$$1 + e^{r_t \tau}(X_i - X_{i-1})^{-1} \left[c(t, X_i, \tau) - c(t, X_{i-1}, \tau)\right], \qquad i = 2, \ldots, n,$$

  is an upper bound on $\tilde{\Pi}_t(X_{i-1})$ and a lower bound on $\tilde{\Pi}_t(X_i)$. The upper bound on $\tilde{\Pi}_t(X_n)$ is 1 and the lower bound on $\tilde{\Pi}_t(X_1)$ is zero. Based purely on the observed data, the true values of the $\tilde{\Pi}_t(X_i)$ can be anywhere in between the upper and lower bounds.

3 Application to exchange-traded products

3.1 Data and computation

Options on exchange-traded products, primarily single stocks, indexes and futures, trade on many exchanges and thousands of assets world-wide. The exchanges generate raw option price data in currency terms. Processed implied volatility data are provided by Bloomberg, as fields pertaining to a ticker. Time series history is typically available, though how far back varies widely. "Moneyness" in the data is expressed as a ratio to the current cash price. An example are data for 3-month options on the S&P 500 index, ticker SPX Index, as of Dec. 21, 2012. The data for SPX and other U.S. indexes and single stocks are based on prices of CBOE options on the index.[11]

[11] The Bloomberg data for each ticker are constructed by filtering the raw end-of-day data, extracting European option implied volatilities from the American option prices, and interpolating the results across exercise price and tenor. The resulting surfaces are close to the intraday volatility surfaces displayed on the OVDV screen. Some of the latter data is identified by tickers, but a field search indicates there is no history.

Bloomberg field mnemonic	moneyness	implied vol
3MTH_IMPVOL_80%MNY_DF	80.0	23.95
3MTH_IMPVOL_90.0%MNY_DF	90.0	21.71
3MTH_IMPVOL_95.0%MNY_DF	95.0	18.81
3MTH_IMPVOL_97.5%MNY_DF	97.5	17.40
3MTH_IMPVOL_100.0%MNY_DF	100.0	16.09
3MTH_IMPVOL_102.5%MNY_DF	102.5	14.88
3MTH_IMPVOL_105.0%MNY_DF	105.0	13.84
3MTH_IMPVOL_110.0%MNY_DF	110.0	12.48
3MTH_IMPVOL_120%MNY_DF	120.0	12.34

Computations using these data are illustrated in Figure 2 for two dates, Aug. 7, 2008, just after the first major overt symptoms of the global financial crisis emerged,[12] and Dec. 21, 2012. The upper left panel displays the Bloomberg data and the interpolated volatility smile. The x-axis in this and the other panels in the figure is expressed as the proportional difference between the exercise price and the current forward index level.

The upper right panel of Figure 2 displays the call valuation function, evaluated for each exercise price using the interpolated smile for each date. The call prices are expressed as a fraction of the current forward index level, calculated as $F_{t,\tau} = e^{(r_t - q_t)\tau}$. Option prices for the S&P 500, forward index levels, and the diagnostics in Table 1, are calculated using 3-month T-bill yields as a financing rate and trailing (rather than estimated forward) 12-month dividend yields as the underlying cash flow rate.

The bottom panels of Figure 2 display the risk-neutral distribution and density functions. The finite differences are calculated setting $\Delta = 0.025$ (as a fraction of current forward index level). For any point on the x-axis, the plot in the bottom left panel can be read as giving the probability that the price return of the S&P 500 vis-à-vis the current forward index level over the subsequent 3 months will be that level or less.

Table 1 displays diagnostics for the computations. The deltas of the input options extend close to zero and unity on both dates, and the vegas are reasonably small at the endpoints.

The distributions are typically multi-modal for SPX Index, with a left-tail hump particularly pronounced. Multi-modal behavior is both an authentic result and an artifact of the technique. Take, for example, the left-tail hump for Aug. 7, 2008. It appears for exercise prices roughly 10 to 20 percent below the current forward index. These

[12] The "quant event," in which algorithmic equity-trading programs abruptly began experiencing losses far in excess of prior extremes, began on Aug. 6. Paribas halted redemptions from three subprime-focused hedge funds it managed on Aug. 9. The Federal Reserve introduced its first policy measures to address the crisis the next day.

are the highest implied vols on the smile. In that interval, the call valuation function declines less slowly than it would if those low-strike implied vols were closer to the ATM vol. Hence the risk-neutral density is high. But a spline knot point imposes an inflection point in the smile at an exercise price equal to 1139.46. From that point, the slope of the volatility smile goes rapidly from steep to flat. Although it is impossible to discern in the graph, at that point the decline in the call valuation function decelerates, inducing a small region in which the density is close to zero.

The hump behavior is the feature most directly affected by the smoothing parameter Δ. For example, if Δ were set higher than the value of 0.025 used in the lower-right plots of Figure 2, the density would be estimated by bridging across wider intervals of the interpolated smile, reducing the variations in the convexity of the call valuation function, and thus the propensity of the estimated risk-neutral density to rise and fall. If Δ is set high enough, the additional mode can be eliminated, without drastically changing the probabilities of returns of specific magnitudes.

Data on options on money-market futures are also available, but these present particular difficulties, especially in the current low-rate environment, as the actively traded exercise-price range is highly compressed against the zero bound. For this group, however, it is relatively straightforward to construct a cruder estimate of the RND along the lines of the diagnostic table.[13]

3.2 Time series of tail risk estimates

The results can be used to compute time series of statistics of interest, including moments, quantiles and the probabilities of returns of specified sizes. For example, we can represent risk-neutral tail risk as the probability of a decline in the S&P of a specific large magnitude. Determining a magnitude to focus on raises similar issues to stress testing in risk management, namely, finding a shock that qualifies as very severe, but is nonetheless plausible and in the realm of possibility. If we choose a very high shock, its risk-neutral probability will almost always be zero. If we choose too small a shock, its risk-neutral probability will almost always be very high. Either way, little insight is gained.

One way to find a useful shock magnitude is through this back-of-the-envelope calculation: If returns were normally distributed, a decline (or runup) of about 2.33 standard deviations would have a probability of one percent. The long-term average annualized implied as well as realized volatility of S&P 500 price returns is roughly

[13] The Bloomberg data for EDA Comdty contain only three distinct values for the 3-month tenor, and it is unclear if the interpolation technique they apply generally to exchange-traded options is well-suited to money-market futures.

20 percent. A rough estimate of the first percentile of 3-month returns is therefore $-20 \times 2.33 \times \sqrt{0.25} = -23.3$ percent. Avoiding exact numbers, so as not suggest that this is a precise estimate, the risk-neutral probability of a 20 or 25 percent decline in the S&P 500 is a reasonable representation of tail risk. We have a mild preference for 20 percent, since it is the lowest observed exercise price in the data and reduces reliance on extrapolation.

The results are displayed Figure 3, covering the period since end-Nov. 2005. The upper panel displays the probability of a three-month decline in the S&P 500 of at least 20 percent. The lower panel displays the first percentile of the S&P 500 price return, displayed as a positive number in percent, in other words, the value-at-risk (VaR) of a long S&P 500 position, expressed in return terms, at a 99-percent confidence level.

Risk-neutral tail risk was low prior to the crisis, apart from a brief but sharp increase in mid-2006. At the end of Feb. 2007, tail risk increased sharply, and again after the quant event of August 2007. Tail risk peaked following the Lehman bankruptcy at a probability near 35 percent of a *further* decline of the S&P 500 in excess of 20 percent over the subsequent quarter. The tail probability is low at the time of writing, just a few percent, but remains generally higher than pre-crisis and fluctuates quite a bit more than pre-crisis. The extreme quantile or VaR of the distribution tracks the probability closely, ranging from about 20 percent before and after the crisis to about 60 percent at its peak in late 2008.

To gain some insight on the the effect of different settings for Δ, Figure 4 compares the estimated tail risk time series for two values, $\Delta = 0.025$ and $\Delta = 0.100$, each held constant over the entire observation interval. The time series are very close to one another. The correlation of the two probability series is 0.997 and the correlation of their daily first differences is 0.977.

As an example of how the techniques can be applied to single stocks, and perhaps interesting in its own right, Figure 5 displays equity tail risk for American International Group, Inc. from late 2007 until the Friday preceding the Lehman bankruptcy filing, Sep. 12, 2008. Tail risk is measured by the risk-neutral probability of a decline of 50 percent or more in the stock price, which can be plausibly said to represent the risk of a corporate bankruptcy. It is somewhat uncomfortable far from the observed data, but that far in the tails, the vega is likely very low even for high volatility levels, and the exercise-price delta very close to $-e^{-r_t \tau}$. If there is significant error in the extrapolation, relative to the unobserved "true" market volatility levels, there will be more (or less) probability mass between -50 and -20 percent, and less (or more) between -100 and -50 percent.

The probability is close to zero for most of the period, rising a bit during periods of fear near the end-2007 and Bear Stearns. The "failure probability" began to rise rapidly during July 2008, as market concerns about losses at Fannie Mae and Freddie

Mac intensified rapidly. By Sep. 12, the Friday before the Lehman bankruptcy filing, the probability reached 40 percent, but most of that runup had taken place during the previous few days.

A characteristic of risk-neutral tail risk behavior that appears clearly in Figures 3 and 5 is its propensity to have risen very abruptly when it is high. Tail risk measures tend to decline gradually from these peaks—unless, as in the AIG case, the peak proves to be terminal. Peaks in tail risk are associated with and subsequent to an event, but occur when market-adjusted tail risk has been relatively low. These characteristics seem to indicate that high tail risk estimates do not provide reliable early warning signals of risk events.

But periods of low tail risk estimates, especially if interrupted by sudden transitory spikes in tail risk unaccompanied by major events, such as those of June 2006 and February 27, 2007, may indicate unease in markets that can lead to future risk events. This observation is closely related to the "paradox of volatility," in which low volatility is associated with the buildup of financial imbalances, rising leverage and higher financial stability risk.

4 Application to currencies

4.1 Data and computation

Prices of options on currencies and precious metals are typically expressed by traders as Black-Scholes implied volatilities. The exercise price of an at-the-money option is generally understood to be equal to the current forward rather than spot exchange rate with a time to settlement equal to the option tenor, and the option is called at-the-money forward (ATMF).

The exercise prices of in- and out-of-the-money currency options are typically expressed in terms of the Black-Scholes delta

$$v_S(\cdot) \equiv \frac{\partial}{\partial S_t} v(S_t, \tau, X, \sigma, r_t, q_t). \tag{9}$$

For this data structure, therefore, it is most convenient to think of the Black-Scholes volatility surface as a function $\sigma(t, \delta, \tau)$ of the date, tenor and delta rather than exercise price. Computation of prices of options in currency units for trade-settlement purposes is easy via the Black-Scholes formula.

Currency options are typically traded as combinations: straddles, strangles and risk reversals. Strangles and risk reversals, which are combinations of out-of-the-money

options, typically have a delta of 0.10 or 0.25. These combinations can be readily converted into prices of individual options with the specified deltas. For example, consider a 25-delta one-month strangle. Its price is quoted as the implied vol spread or difference between the average implied vols of the 25-delta put and call, which are not directly observed, and the ATMF put or call vol.

$$\text{strangle price} = \frac{1}{2}\left[\sigma\left(t, 0.25, \frac{1}{12}\right) + \sigma\left(t, 0.75, \frac{1}{12}\right)\right] - \text{ATMF vol},$$

The risk reversal quote is the implied vol spread between the two "wing" options:

$$\text{risk reversal price} = \sigma\left(t, 0.25, \frac{1}{12}\right) - \sigma\left(t, 0.75, \frac{1}{12}\right).$$

Note that strangle and risk reversal are quoted as vol spreads, while the ATMF is a vol level. Using these definitions, the vol levels of the wing options can be inferred from the strangle, risk reversal, and ATMF quotes:

$$\sigma\left(t, 0.25, \frac{1}{12}\right) = \text{ATMF vol} + \text{strangle price} + \frac{1}{2} \times \text{risk reversal price}$$
$$\sigma\left(t, 0.75, \frac{1}{12}\right) = \text{ATMF vol} + \text{strangle price} - \frac{1}{2} \times \text{risk reversal price}$$

Analogous formulas describe the 10-delta versions of these standard option combinations, and versions for other tenors. From them, we can obtain the 10-, 25-, 75-, and 90-delta implied volatilities. The ATM and ATMF options have deltas close to, but not exactly, equal to 0.50. We obtain an option with a delta near 50 from the ATMF option, using (9) to compute the exact delta.

Foreign-exchange option price data is available from a number of data providers and dealers. The data used here are downloaded from Bloomberg, which stores implied volatility histories for each point on the volatility surface—tenor and exercise price—for each currency pair, as a distinct ticker. The data are aggregated, filtered and, possibly, interpolated from a number of dealer quotes. Bloomberg's currency option data appear generally to be the highest quality of the three structures discussed here.

The data structure is illustrated here using 1-month options on EUR-USD, the price of a Euro in dollars, as of Dec. 31, 2012.[14]

[14] Data are also available for the 1-week, 3-, 6-, and 12-month, and 10-year tenors.

Bloomberg ticker	description	implied vol/spread
EURUSDV1M Curncy	EUR-USD OPT VOL 1M	8.2200
EURUSD25R1M Curncy	EUR-USD RR 25D 1M	-0.3025
EURUSD25B1M Curncy	EUR-USD BFY 25D 1M	0.1050
EURUSD10R1M Curncy	EUR-USD RR 10D 1M	-0.4875
EURUSD10B1M Curncy	EUR-USD BFY 10D 1M	0.2875

Transformed into a volatility smile in (δ, σ)-space, the data become

delta	implied vol
0.1000	8.26375
0.2500	8.17375
0.5015	8.22000
0.7500	8.47625
0.9000	8.75125

Once the input data has been prepared, the volatility smile can be interpolated. We carry out the interpolation via a clamped cubic spline, but in (δ, σ)- rather than (X, σ)-space. The x-axis values $0.10, 0.25, 0.75$, and 0.90 are the same on each date, but the center knot point has a slightly different x-axis value near 0.50 each day. Options with deltas below 0.10 are assigned the 10-delta volatility and options with deltas above 0.90 are assigned the 90-delta volatility.

For this data structure, there is an additional step following interpolation, by which the smile in (δ, σ)-space is transformed into one in (X, σ)-space. This is slightly less simple than it might seem, as we can't map directly from exercise price to delta via (9), and then to the smile in (δ, σ)-space. The reason is that the volatility argument in (9) is not constant, but itself varies with delta.[15]

The computation is as follows: Substitute the expression for the Black-Scholes delta into the interpolated smile $\sigma(t, \delta, \tau)$. For any stipulated X°, and for fixed values of the other arguments, we can solve

$$\sigma^\circ = \sigma\left[t, v_S(S_t, \tau, X^\circ, \sigma^\circ, r_t, q_t), \tau\right)]$$

numerically for σ°.[16]

[15] We don't have that problem when calculating the delta of the ATMF option because we have a fixed exercise price and volatility.

[16] In one approach to RND construction from data on exchange-traded options, implied volatilities initially associated with exercise prices are converted to volatilities associated with the corresponding

This transformation is illustrated in Figure 6 for two dates, May 22, 2009 and Nov. 18, 2011. The input data and the initial smile interpolation, carried out via a clamped cubic spline, are displayed in the left panel. The x-axis is in delta units. The volatility smiles in the right panel are computed from those in the left panel. They are not derived by a fresh interpolation but rather functionally, from the interpolated smile in (δ, σ)-space, via the numerical procedure described in the previous paragraph. Note that the direction of the x-axis is reversed between the two graphs. On the later date, options with especially high payoffs if the dollar appreciates sharply vis-à-vis the euro have high implied volatilities. These correspond to low exercise prices in currency units but high call deltas.

Computations using these data are illustrated in Figure 7 for the same two dates as in Figure 6, May 22, 2009 and Nov. 18, 2011. In all four panels, the x-axis is expressed as the proportional difference from the 3-month forward rate (USD per EUR). The RND estimates are computed using $\Delta = 0.005$ (as a fraction of the forward rate). Option prices for EUR-USD, forward exchange rates, and the diagnostics in Table 2, are calculated using 1-month U.S. dollar and euro Libor rates as the financing and underlying cash flow rates.

The two dates display a sharp contrast in the direction of skewness of the risk-neutral distribution. On the earlier date, there is a sharp skew toward a weaker dollar, while on the later date there is a skew toward a stronger dollar.

Diagnostics for the data and computations are shown in Table 2. The deltas of the input options, naturally, extend exactly from 0.10 to 0.90, but the vegas are reasonably small at the endpoints. The data are somewhat better-behaved than the S&P 500 option data; the foreign-exchange option data permit a smaller step size in differencing without encountering non-convexities.

4.2 Time series of tail risk estimates

An example of how the results can be applied is displayed in Figure 8. The upper panel plots time series of the risk-neutral probabilities of the dollar appreciating and depreciating by 7.5 percent or more over the subsequent month.[17] The lower panel plots the

deltas using (9). Interpolation is then carried out in (δ, σ)-space. The conversion to deltas may be done using the same at-the-money volatility for all strikes (so-called "point conversion") or using each strike's volatility ("smile conversion") to avoid cases in which segments of the volatility smile are so steep that an option may have a lower call delta than another with a higher exercise price. Bu and Hadri (2007) discuss the phenomenon, which intuitively seems likely to be due to no-arbitrage violations in the data. The issue doesn't arise with our technique because we are going from input data sets in (δ, σ)-space to (δ, X)-space rather than vice versa.

[17] This seems like a reasonable threshold: volatility for EUR-USD is typically in the neighborhood of 10 percent. If exchange rate returns were normally distributed, the first and last percentiles of 1-month

difference between these probabilities, and highlights the direction and magnitude of the skew in tail risk estimates. In contrast to the S&P 500 and other equity indexes, the tail risk skew for major currency pairs can and does change direction.

Tail risk first began to rise sharply around the time of the Bear Stearns failure and spiked following the Lehman filing. Since Lehman, tail risk has often been very high, and the risk-neutral probability of a sharp dollar appreciation has generally been much higher than that of a depreciation. This pattern likely reflects safe-haven positioning, as it began well before the European debt crisis, but was reinforced as the latter played out.

Both the level of risk-neutral tail risk and its skew to a weaker euro rose steadily through 2011, but dropped abruptly following the announcement by the European Central Bank of its longer-term refinancing operations (LTROs) on December 8, 2011. Tail risk has most recently dropped back to pre-2008 levels, and the directional difference between dollar appreciation and depreciation is near zero, in spite of a steady appreciation of the euro vis-à-vis dollar amounting to 15 percent since mid-2012.

5 Application to swaptions

5.1 Data and computation

Standard swaptions are options that exercise into a payer or receiver position in a LIBOR interest-rate swap. They are one of the two more-liquid types of markets in which exposures to longer-term interest rates are traded.[18] The other type is options on government bond futures. Swaption data are better suited than implied volatilities derived from bond futures options prices for computing interest-rate RNDs:

- Swaptions have a fixed term to maturity rather than a fixed maturity date, generating a time series of expectations measures with a fixed horizon without requiring interpolation across maturities.
- Swaption prices map directly into interest-rate expectations, rather than indirectly via bond prices.
- Prices of options on bond futures include compensation for the delivery option, and switches in the cheapest-to-deliver can distort their signals of interest-rate prospects.

returns would be about $\pm 10 \times 2.33 \times \sqrt{0.08\overline{33}} = \pm 6.73$ percent.

[18] Breeden and Litzenberger (2013) describe a technique for extracting RNDs of shorter-term rates from implied volatilities of caps and floors.

One disadvantage of swaption data should also be mentioned: The underlying price of a swaption is the LIBOR swap rate, rather than the risk-free rate, which may differ from the risk-free rate for a number of risk- and liquidity-based reasons.

Swaption implied volatility data are available on Bloomberg. They are expressed as Black or lognormal vols, that is, as the standard deviation of logarithmic changes in the forward swap rate for the given swaption "tail" (swap maturity) and tenor (option maturity), expressed in percent units at an annual rate. The data are based on quotes aggregated by Bloomberg from submissions by several contributing dealers. Bloomberg interpolates across strikes when data is missing. The data appear to be of reasonably good quality from early 2013 on.

A wide range of tails and tenors are priced. Option tenors range from 3 months to 20 years and underlying swap tails from 2 to 30 years. Exercise prices range from 200 basis points below to 200 above the current forward swap rate for the given tail and tenor. For tenors and tails with forward swap rates that are close to the zero bound, there are no recent data for exercise prices 200 basis points below the forward swap rate, as these would be exercisable only if longer-term rates turned negative.[19]

As with other types of options, expressing the value of a swaption in terms of an implied volatility based on a particular model of interest-rate behavior does not mean the market believes in that model. Rather, it represents a convenient unit for expressing the value or market price of the swaption.

Black vols fit without much further ado into our RND computation scheme. The data structure on Sep. 5, 2013 for "2-year into 10-year" swaptions—2-year options on 10-year swaps—was

Strike	Bloomberg ticker	description	Black vol
-200	USPAV07C Curncy	USD BVOL SWPT-200 2Y10Y	32.5790
-100	USPAV04K Curncy	USD BVOL SWPT-100 2Y10Y	28.9314
-50	USPAV036 Curncy	USD BVOL SWPT-50 2Y10Y	27.8261
-25	USPAV02H Curncy	USD BVOL SWPT-25 2Y10Y	27.3975
0	USSV0210 BBIR Curncy	USD SWPT BVOL ATM 2Y10Y	27.0250
25	USPAUZA1 Curncy	USD BVOL SWPT 25 2Y10Y	26.7361
50	USPAUZAQ Curncy	USD BVOL SWPT 50 2Y10Y	26.4866
100	USPAUZC4 Curncy	USD BVOL SWPT 100 2Y10Y	26.1151
200	USPAUZEW Curncy	USD BVOL SWPT 200 2Y10Y	25.7388

[19]The available Bloomberg tickers and data can be identified by configuring the VCUB or interest rate vol cube function. The configuration tab enables the user to select and display contributed Black vols for OTM swaptions.

The exercise prices are equal to the 10-year swap rate 2 years forward on Sep. 5, 2013, 4.0888, less the stipulated moneyness, expressed in basis points in the first column. The forward swap rate is today's market assessment of the fixed rate that sets to zero the net present value of a 10-year fixed-for-floating swap initiated 2 years hence. The Black vols (percent per annum) in the last column are the input data provided by Bloomberg.

The Black formula for the price of a swaption in currency units is the product of three terms: (i) the notional amount, (ii) the "bps running" or annuity or present value per basis point of the payments by the fixed leg of the swap, and (iii) the Black-Scholes option value formula applied to the current swap rate as though it were a proper asset, and with the risk-free or financing rate set to zero. We can ignore the first two terms, which are invariant across exercise prices. The last component can be written as $e^{r_t \tau} v[F_{t,\tau}, X, \tau, \sigma(t, X, \tau), 0, 0]$ for a payer swaption, where $F_{t,\tau}$ is the current forward swap rate for a swap initiated τ years hence.[20] A payer swaption gives its owner the right to enter into a swap at a fixed rate X, and is analogous to a put on a bond, and to a call in interest-rate terms.

In essence, the swaption valuation formula has a component containing the expected value of changes in the swap rate vis-à-vis the current forward value in excess of a given strike rate, and a component expressing how much that expected value is worth. The Black formula gives the value of the option in interest-rate terms. It is converted into currency units using the notional amount and the annuity value.

With these modifications, the same calculation procedure as for exchange-traded products can be used to compute the RND. The computations are illustrated in Figure 9 for two dates, May 1, 2013 and Sep. 5, 2013. We use a small $\Delta = 0.0001$ (1 basis point), so this data structure can be said to be relatively cooperative with our technique. The x axis in the upper panels of the charts is expressed as differences from the forward swap rate in basis points, analogous to the previous examples. In the lower panels, the distribution and density are represented as functions of the terminal 10-year swap rate.

Diagnostics for the computations are displayed in Table 3. We see that the data extend far enough above and below the forward swap rate that the deltas cover much of the interval $(0, 1)$. The vegas for the highest and lowest exercise prices are fairly low. We are applying a version of the Black formula that isn't discounted to the present by the risk-free rate, so low-strike call deltas can be very close to unity.

The volatility smile and the implied RNDs are heavily influenced by the proximity of spot and forward swap rates to the zero bound. On the earlier date, the implied RND is skewed quite strongly to higher rates, and on the later date, much less so. But

[20] The term of the swap isn't displayed in the notation.

on both dates, implied volatilities of low strike options close to the zero bound are higher, not lower, than those of high-rate strikes. A distribution skewed to the left is incompatible with low rates.

5.2 Time series of tail risk estimates

As we did for other asset classes, we'll illustrate the results with time series of tail risk estimates. We use changes in basis points vis-à-vis the current forward swap rate rather than proportional changes to represent extreme moves. In Figure 10, the top two panels display the risk-neutral probabilities of specific changes in rates, while the lower panel displays the probabilities of rates reaching specific levels.

The upper panel displays probabilities of changes of at least 200 basis points. From the beginning of May 2013, the probabilities both of very large decreases and increases in rates, as well as the forward rates themselves, began to rise. The probability of a sharp drop in rates rose faster, but the probability of a rate rise accelerated following the Chairman's May 22 Joint Economic Committee testimony. As forward rates rose, these probabilities drew closer together. By the time rates peaked in early September 2013, the tail probabilities were nearly equal. More recently, a skew to sharply higher has been re-established, but it is less pronounced than in early 2013.

The probabilities of changes of at least 100 basis points, displayed in the center panel of Figure 10, also rose in 2013. These probabilities are more nearly equal to each other than those of more extreme rate moves, as one would expect of events closer to the center of the distribution.

Proximity to the zero bound makes it more difficult to interpret risk-neutral interest-rate distributions, because it is hard to distinguish between the effects of movement away from or toward the zero bound from other influences on the shape of the distribution. The impact of proximity to zero is similar to the pattern seen in the lower panel of Figure 10, which displays the risk-neutral probabilities of the rate ending at 5 percent or higher, or at 2 percent or lower. These probabilities are driven in large part by how close to these thresholds the current forward rate happens to be.

Similarly, when rates are close to zero, the probability of a large decline cannot be high, because there is nowhere for rates to go but up. When the forward swap rate is relatively low, it is more strongly correlated with the risk of sharply lower rates. When the swap rate is relatively high, it moves more closely with the risk of a drastic rise in rates. The level of rates, however, is not the only determinant of rate RNDs. Since their early September peak, 10-year swap rates 2 years forward have fluctuated in a range between about $3\frac{1}{2}$ and 4 percent. During that time, overall rate volatility has declined, and the probability of a decline in rates of at least 200 basis points has fallen relative to that of a rise in rates of the same magnitude.

6 Conclusion

The technique for estimating risk-neutral RNDs described here appears to work well with several different data structures, and is relatively easy to program and use. There is considerable demand, particularly in central banks, to apply risk-neutral probabilities in market monitoring and policy work, and our technique should make it possible to take some of the effort out of creating the RNDs.

That effort would be better focused on other aspects of RNDs. As far as the quality and reliability of the results is concerned, assembling and filtering better-quality data sets is one challenge. But perhaps the most important open task with respect to risk-neutral RNDs remains how to use and interpret them.

References

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions, *Journal Of Econometrics* **116**(1/2): 9–47.

Banz, R. W. and Miller, M. H. (1978). Prices for state-contingent claims: some estimates and applications, *Journal of Business* **51**(4): 653–672.

Bliss, R. R. and Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions, *Journal of Banking and Finance* **26**(2–3): 381–422.

Bliss, R. R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates, *Journal of Finance* **59**(1): 407–446.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices, *Journal of Business* **51**(4): 621–651.

Breeden, D. T. and Litzenberger, R. H. (2013). Central bank policy impacts on the distribution of future interest rates. Available at `http://www.dougbreeden.net/uploads/Breeden_Litzenberger_with_Postscript_Central_Bank_Policy_Impacts_9_20_2013.pdf`.

Bu, R. and Hadri, K. (2007). Estimating option implied risk-neutral densities using spline and hypergeometric functions, *Econometrics Journal* **10**(2): 216–244.

Carr, P. and Wu, L. (2009). Variance risk premiums, *Review of Financial Studies* **22**(3): 1311–1341.

Castagna, A. and Mercurio, F. (2007). The vanna-volga method for implied volatilities, *Risk* pp. 106–111.

Cox, J. C. and Rubinstein, M. (1985). *Options markets*, Prentice–Hall, Englewood Cliffs, NJ.

Figlewski, S. (2010). Estimating the implied risk-neutral density for the U.S. market portfolio, *in* T. Bollerslev, J. Russell and M. Watson (eds), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford University Press, Oxford and New York, pp. 323–353.

Garcia, R., Ghysels, E. and Renault, E. (2010). The econometrics of option pricing, *in* Y. Aït-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics Tools and Techniques*, Vol. 1, Elsevier, Amsterdam, pp. 479–552.

Hodges, H. M. (1996). Arbitrage bounds on the implied volatility strike and term structures of European-style options, *Journal of Derivatives* **3**(4): 23–35.

Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees: a literature review, *Journal of Derivatives* **7**(2): 66–82.

Jackwerth, J. C. (2004). Option-implied risk-neutral distributions and risk aversion, *Monograph*, Research Foundation of CFA Institute. `http://www.cfapubs.org/doi/pdf/10.2470/rf.v2004.n1.3925`.

Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2008). *Loss models: from data to decisions*, 3rd edn, John Wiley & Sons, Hoboken, NJ.

Malz, A. M. (2013). Risk-neutral systemic risk indicators, *Staff Reports 607*, Federal Reserve Bank of New York. Available at `http://www.newyorkfed.org/research/staff_reports/sr607.pdf`.

Mandler, M. (2003). *Market expectations and option prices: techniques and applications*, Physica-Verlag, Heidelberg and New York.

Neuberger, A. (2012). Realized skewness, *Review of Financial Studies* **25**(11): 3423–3455.

Ross, S. A. (2013). The Recovery Theorem, *Journal of Finance* . Forthcoming, available at `http://onlinelibrary.wiley.com/doi/10.1111/jofi.12092/pdf`.

Table 1: **Data and diagnostics for S&P 500 index**

07Aug2008

$\frac{X}{S}$	X	Volatility	Call value	Delta	Vega	Lower bound	Upper bound	$\Pi_t(X)$
0.800	1012.86	25.0589	252.3658	0.9609	0.1872	0.0000	0.1156	0.0440
0.900	1139.46	24.8552	140.8629	0.8106	0.6690	0.1156	0.2237	0.1857
0.950	1202.77	22.9498	91.9227	0.6843	0.8872	0.2237	0.3226	0.2816
0.975	1234.42	21.9857	70.5727	0.6034	0.9646	0.3226	0.4050	0.3630
1.000	1266.07	21.0202	51.8199	0.5121	1.0000	0.4050	0.4988	0.4497
1.025	1297.72	20.0641	36.0214	0.4138	0.9786	0.4988	0.5994	0.5496
1.050	1329.37	19.1198	23.3948	0.3142	0.8923	0.5994	0.7485	0.6468
1.100	1392.68	17.3066	7.5434	0.1401	0.5608	0.7485	0.9460	0.8520
1.200	1519.28	17.3481	0.7322	0.0188	0.1159	0.9460	1.0000	0.9847

04Apr2014

$\frac{X}{S}$	X	Volatility	Call value	Delta	Vega	Lower bound	Upper bound	$\Pi_t(X)$
0.800	1492.07	20.3093	364.9525	0.9813	0.0889	0.0000	0.0511	0.0179
0.900	1678.58	17.9679	187.9768	0.8735	0.5085	0.0511	0.1429	0.0992
0.950	1771.84	15.5268	108.0520	0.7344	0.8170	0.1429	0.2395	0.1972
0.975	1818.46	14.0873	72.5961	0.6247	0.9492	0.2395	0.3612	0.2914
1.000	1865.09	12.7135	42.8133	0.4800	1.0000	0.3612	0.5285	0.4384
1.025	1911.72	11.4576	20.8310	0.3117	0.8892	0.5285	0.7199	0.6230
1.050	1958.34	10.3961	7.7694	0.1565	0.6037	0.7199	0.9225	0.8127
1.100	2051.60	9.4013	0.5452	0.0175	0.1088	0.9225	0.9971	0.9826
1.200	2238.11	9.4095	0.0008	0.0000	0.0004	0.9971	1.0000	1.0000

Delta and vega are the Black-Scholes sensitivities for an option with the indicated exercise price. Delta is the derivative of the call value with respect to the underlying price. Vega is the derivative of the call or put value with respect to the implied volatility, measured as the value response to an increase of vol. The table displays the ratio of the vega for each strike to the vega of the at-the-money option. The lower and upper bounds on $\Pi_t(X)$ are derived from the call value changes in the raw data points, as described in the text. Other inputs and intermediate results (percent where applicable, r_t and q_t refer to USD and EUR Libor):

Date	S_t	r_t	q_t	F_t	$\frac{S_t}{F_t} - 1$
07Aug2008	1266.07	1.6750	2.4377	1263.71	0.19
21Dec2012	1430.15	0.0580	2.2306	1422.49	0.54

Table 2: **Data and diagnostics for EUR-USD**

22May2009

X	$\frac{X}{F}-1$	Volatility	Call value	Delta	Vega	Lower bound	Upper bound	$\Pi_t(X)$
1.3228	-0.055	15.3750	0.07924	0.9000	0.4378	0.0000	0.1700	0.1071
1.3598	-0.028	15.0800	0.04855	0.7500	0.7957	0.1700	0.3905	0.2522
1.3994	0.000	15.2525	0.02444	0.5067	1.0000	0.3905	0.6769	0.5258
1.4455	0.033	16.2000	0.00954	0.2500	0.7970	0.6769	0.8696	0.8127
1.4937	0.067	17.3550	0.00325	0.1000	0.4402	0.8696	1.0000	0.9089

18Nov2011

X	$\frac{X}{F}-1$	Volatility	Call value	Delta	Vega	Lower bound	Upper bound	$\Pi_t(X)$
1.2600	-0.068	19.2800	0.09503	0.9000	0.4370	0.0000	0.1377	0.1083
1.3083	-0.033	17.2513	0.05341	0.7500	0.7953	0.1377	0.3136	0.1818
1.3525	0.000	15.1550	0.02308	0.5012	1.0000	0.3136	0.5899	0.4684
1.3895	0.027	13.8538	0.00791	0.2500	0.7970	0.5899	0.8191	0.7125
1.4199	0.050	13.1750	0.00240	0.1000	0.4402	0.8191	1.0000	0.9058

See the footnote to Table 1. Other inputs and intermediate results (percent where applicable):

Date	S_t	r_t	q_t	F_t	$\frac{S_t}{F_t}-1$
22May2009	1.3998	0.3131	0.9060	1.3994	0.031
18Nov2011	1.3525	0.2566	1.1990	1.3525	0.002

Table 3: **Data and diagnostics for 2-year into 10-year swaptions**

01May2013

$X-F$	X	Volatility	Call value	Delta	Vega	Lower bound	Upper bound	$\Pi_t(X)$
-200	0.48	50.5574	0.0200	0.9961	0.0301	0.0000	0.0734	0.0261
-100	1.48	36.8030	0.0108	0.8948	0.4686	0.0734	0.2698	0.1612
-50	1.98	34.1660	0.0072	0.7605	0.7993	0.2698	0.4123	0.3768
-25	2.23	33.2844	0.0057	0.6777	0.9232	0.4123	0.5061	0.4599
0	2.48	32.5500	0.0045	0.5910	1.0000	0.5061	0.6034	0.5598
25	2.73	32.0970	0.0035	0.5061	1.0267	0.6034	0.6806	0.6473
50	2.98	31.7132	0.0027	0.4265	1.0093	0.6806	0.7770	0.7180
100	3.48	31.2299	0.0016	0.2924	0.8844	0.7770	0.8949	0.8306
200	4.48	30.9423	0.0006	0.1286	0.5405	0.8949	1.0000	0.9419

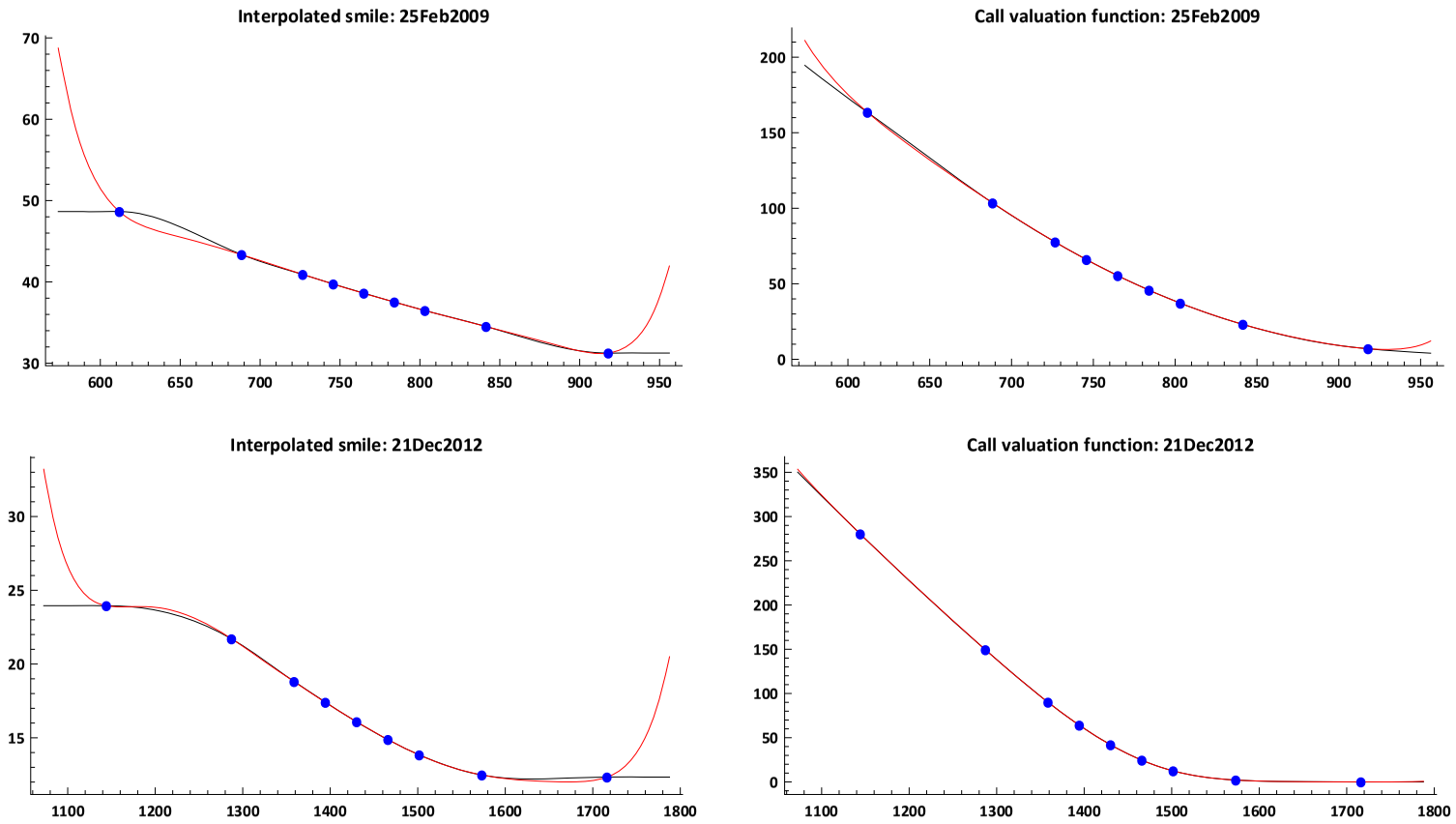
05Sep2013

$X-F$	X	Volatility	Call value	Delta	Vega	Lower bound	Upper bound	$\Pi_t(X)$
-200	2.09	32.5790	0.0204	0.9543	0.2450	0.0000	0.1566	0.1097
-100	3.09	28.9314	0.0121	0.8133	0.6853	0.1566	0.3321	0.2614
-50	3.59	27.8261	0.0088	0.7013	0.8858	0.3321	0.4362	0.4125
-25	3.84	27.3975	0.0074	0.6393	0.9557	0.4362	0.5047	0.4765
0	4.09	27.0250	0.0062	0.5758	1.0000	0.5047	0.5738	0.5462
25	4.34	26.7361	0.0051	0.5128	1.0179	0.5738	0.6352	0.6109
50	4.59	26.4866	0.0042	0.4520	1.0110	0.6352	0.7173	0.6689
100	5.09	26.1151	0.0029	0.3417	0.9372	0.7173	0.8375	0.7681
200	6.09	25.7388	0.0012	0.1809	0.6720	0.8375	1.0000	0.8990

See the footnote to Table 1. The 2- into 10-year forward swap rates are:

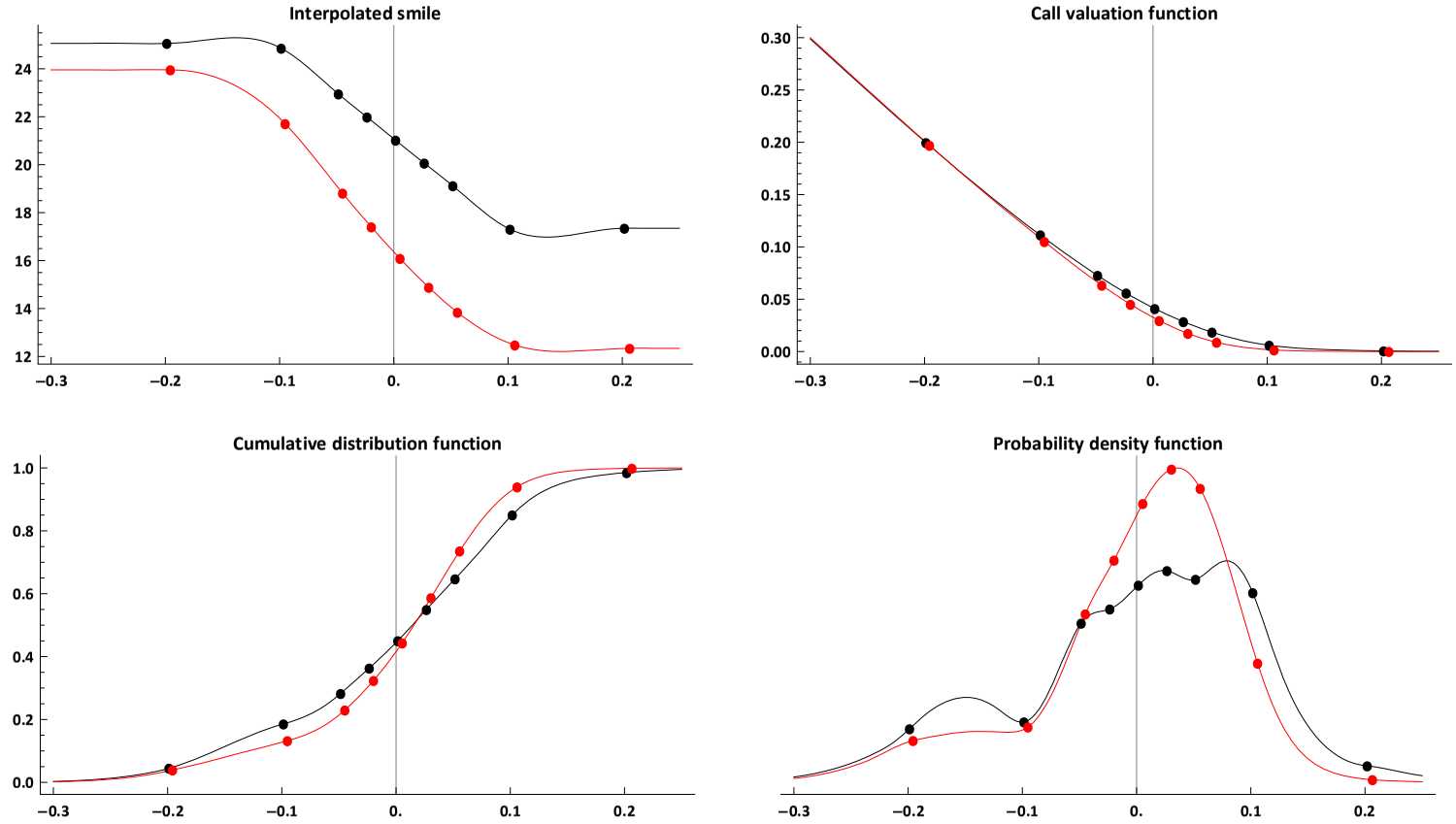
Date	F_t
01May2013	2.4791
05Sep2013	4.0888

Figure 1: **Extrapolation and no-arbitrage restrictions**



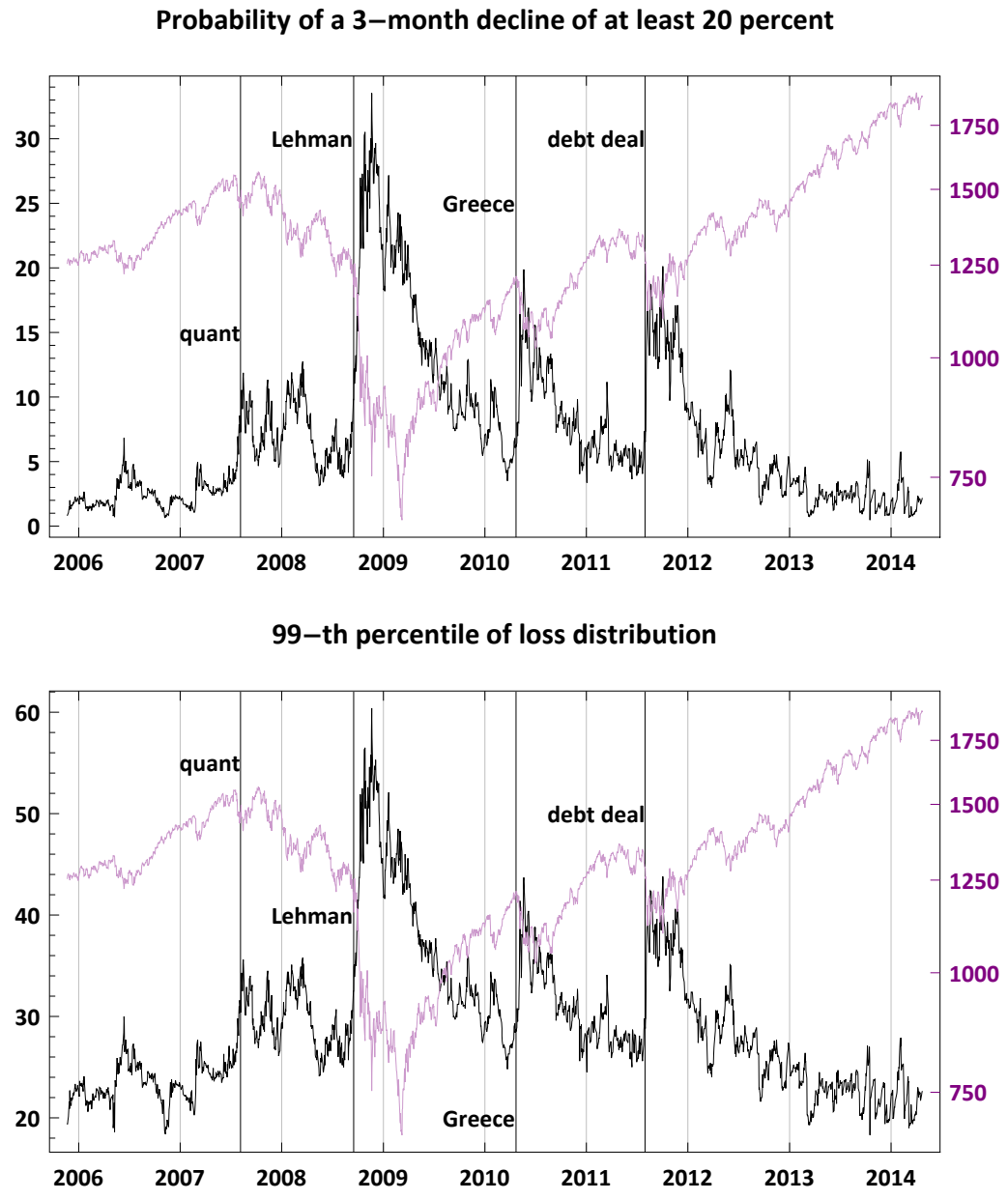
The left panel in each row compares the results of the clamped cubic spline with flat-line extrapolation (black plot) to an alternative (red plot). Blue dots mark the input data/knot points. The right panel in each row compares the call valuation function resulting from each interpolation and extrapolation scheme. Blue dots mark the exercise prices corresponding to the input data/knot points. The data are implied volatilities of 3-month options on the S&P 500 index. The x-axis in each panel is the exercise price in S&P 500 index terms. The y-axes in the left panels are implied volatilities in percent; the y-axes in the right panels are call prices in S&P 500 index units.

Figure 2: **Computation example: 3-month SPX options**



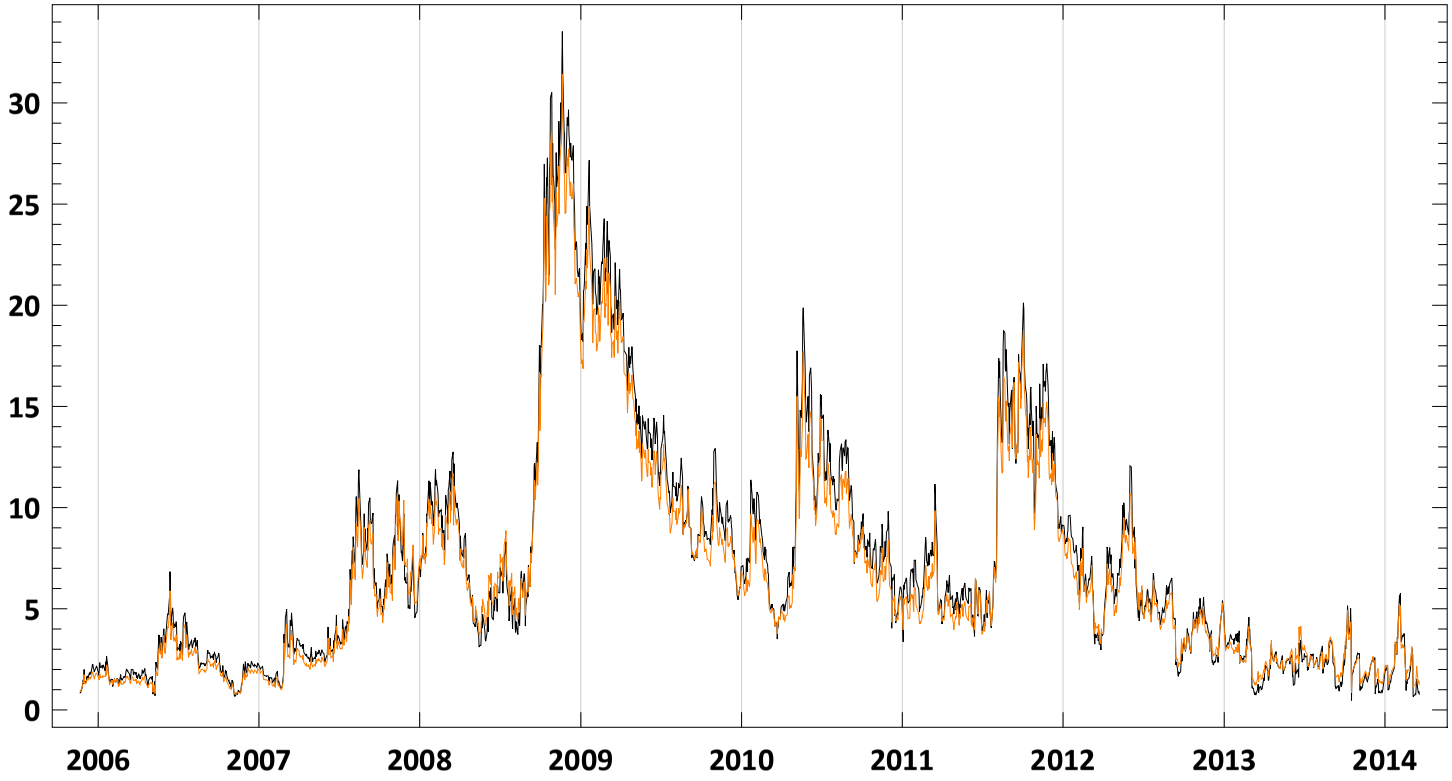
Black plots: Aug. 7, 2008; red plots: Dec. 21, 2012. The units of the x-axes in all four panels are proportional differences between the exercise price or future index level and the forward index level. Input data/knot points are marked by dots and the values (apart from densities) displayed in Table 1. Upper left: input data and the interpolated smile. Implied volatilities on the y-axis are in percent. Upper right: call valuation function, evaluated using the interpolated smile. Call prices on the y-axis are expressed as a fraction of the forward index level. Bottom panels: risk-neutral distribution and density functions. Step size $\Delta = 0.025$ (as a fraction of the forward index level). For any point on the x-axis, the plot in the bottom left panel can be read as the probability the S&P 500 ends at that proportional difference from the forward index level or less in 3 months.

Figure 3: **Risk-neutral S&P 500 tail risk**



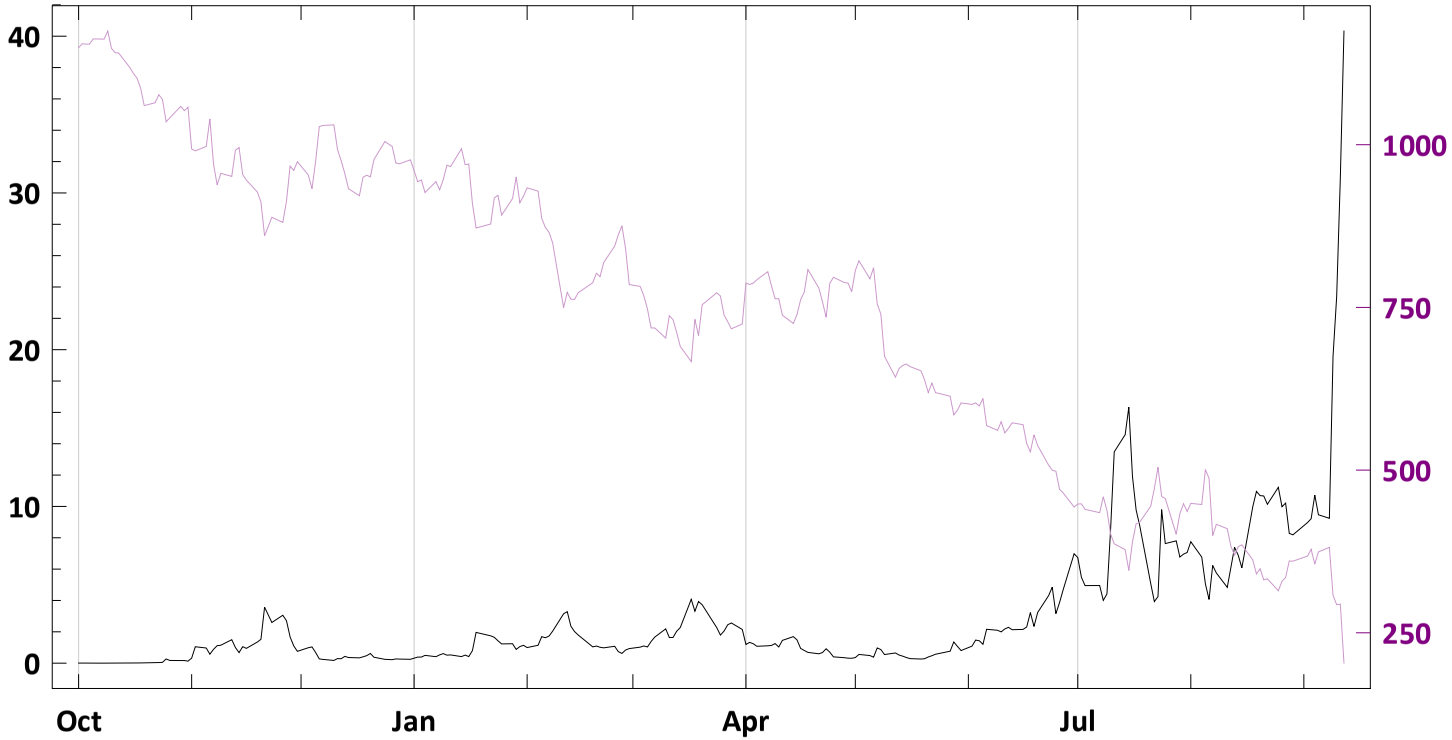
Upper panel:Risk-neutral probability of a 3-month decline of at least 20 percent (black plot, left axis). Lower panel: $(-1)\times$first percentile of the risk-neutral cumulative distribution of 3-month S&P 500 price returns, percent (black plot, left axis). $\Delta = 0.025$, Nov. 21, 2005 to Mar. 20, 2014. Purple plot (right axis): logarithm of the S&P 500 index; axis labels show the index level.

Figure 4: **Effect of varying Δ on estimated tail risk**



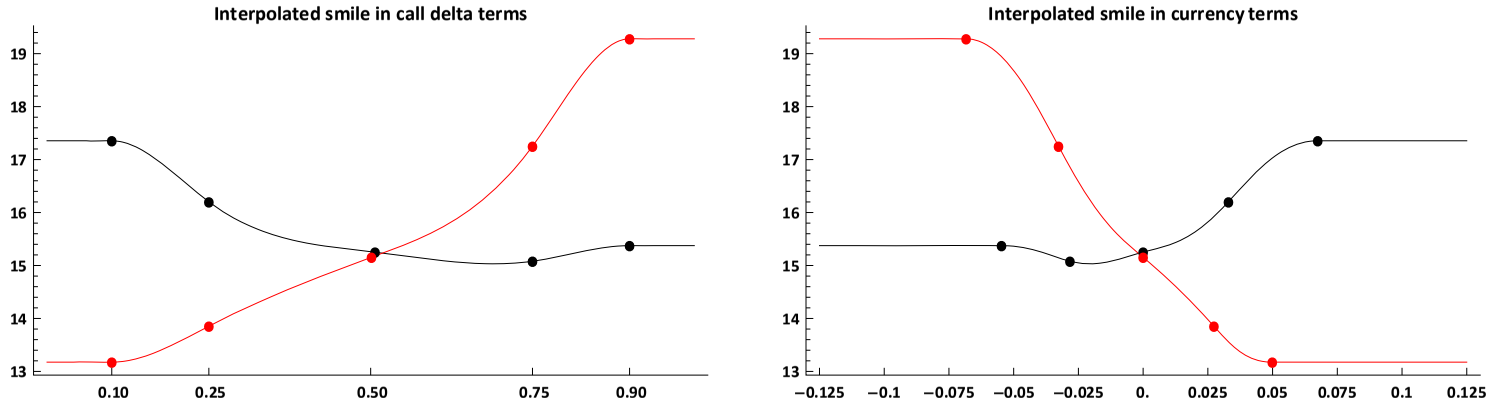
Risk-neutral probability of a 3-month decline in the S&P 500 of at least 20 percent, Nov. 21, 2005 to Mar. 20, 2014. Black plot: $\Delta = 0.025$; orange plot (right axis): $\Delta = 0.100$.

Figure 5: **Risk-neutral AIG tail risk**



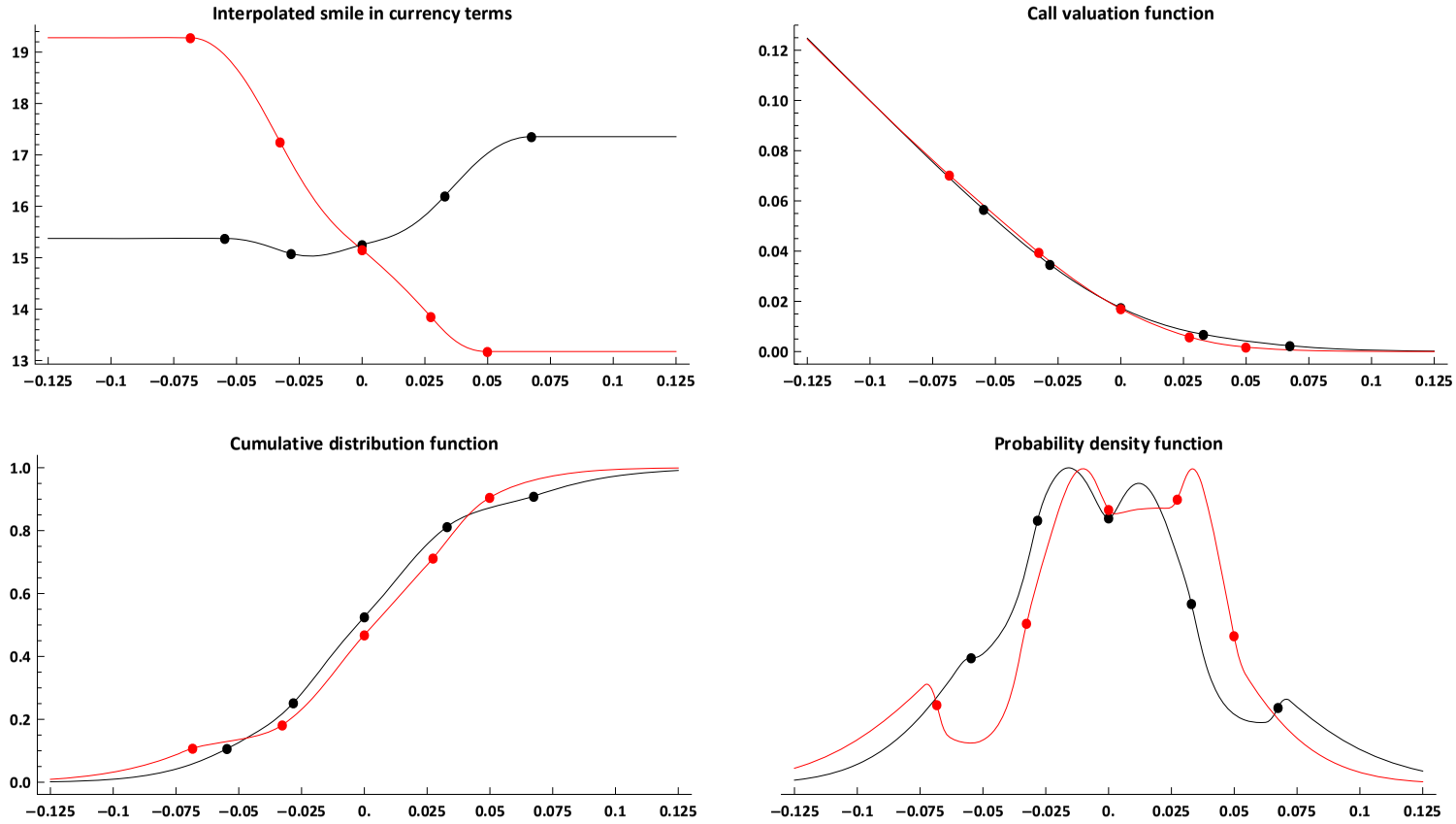
Black plot (left axis): risk-neutral probability of a 3-month decline of at least 20 percent, Oct. 1, 2007 to Sep. 12, 2008, $\Delta = 0.025$.
Purple plot (right axis): split-adjusted AIG equity price.

Figure 6: **Volatility smile interpolation for EUR-USD**



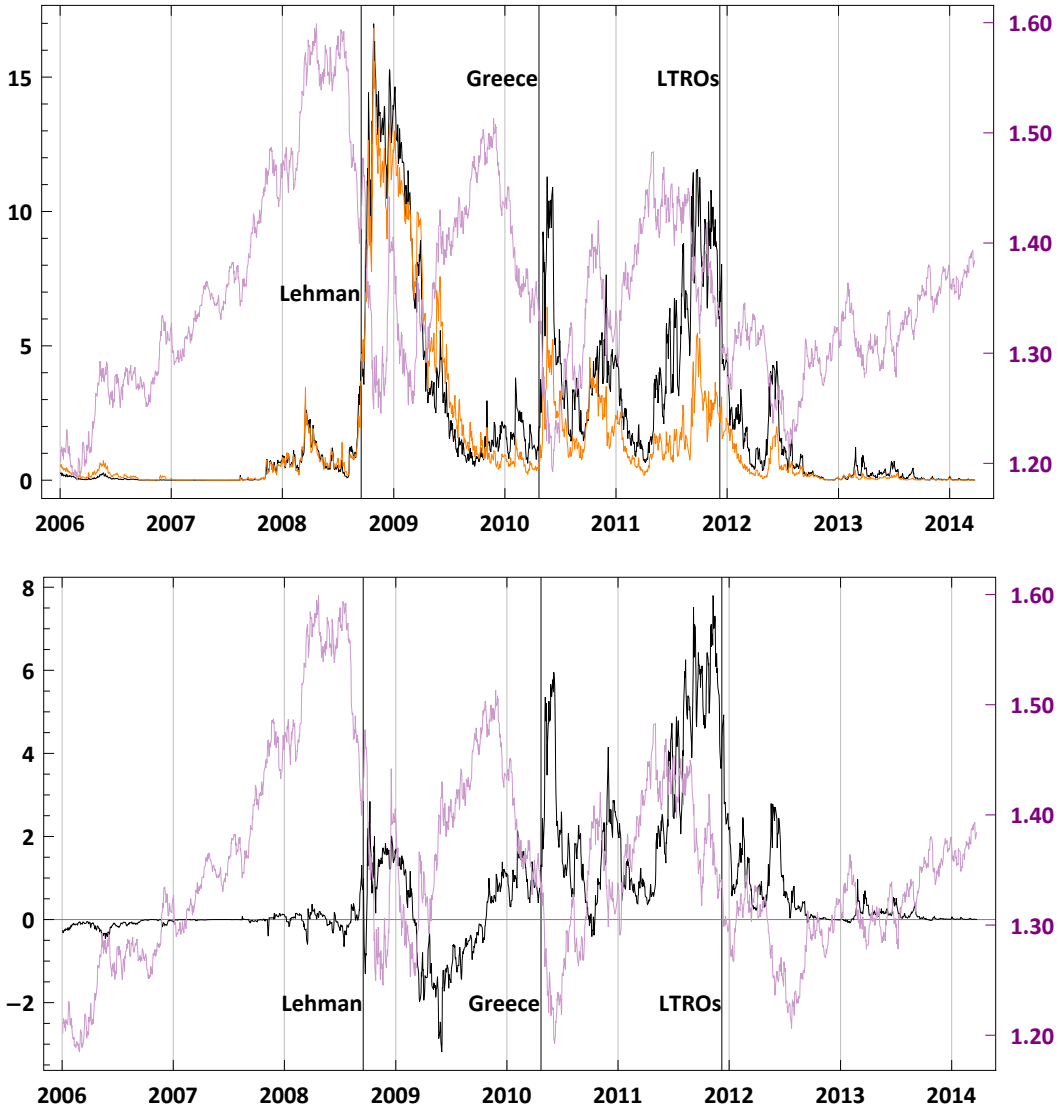
Black plots: May. 22, 2009; red plots: Nov. 18, 2011. Left panel: raw data and the initial smile interpolation, carried out via clamped cubic spline, x-axis is in delta units. Right panel: volatility smile recomputed in (X, σ)-space.

Figure 7: **Computation example: 3-month EUR-USD options**



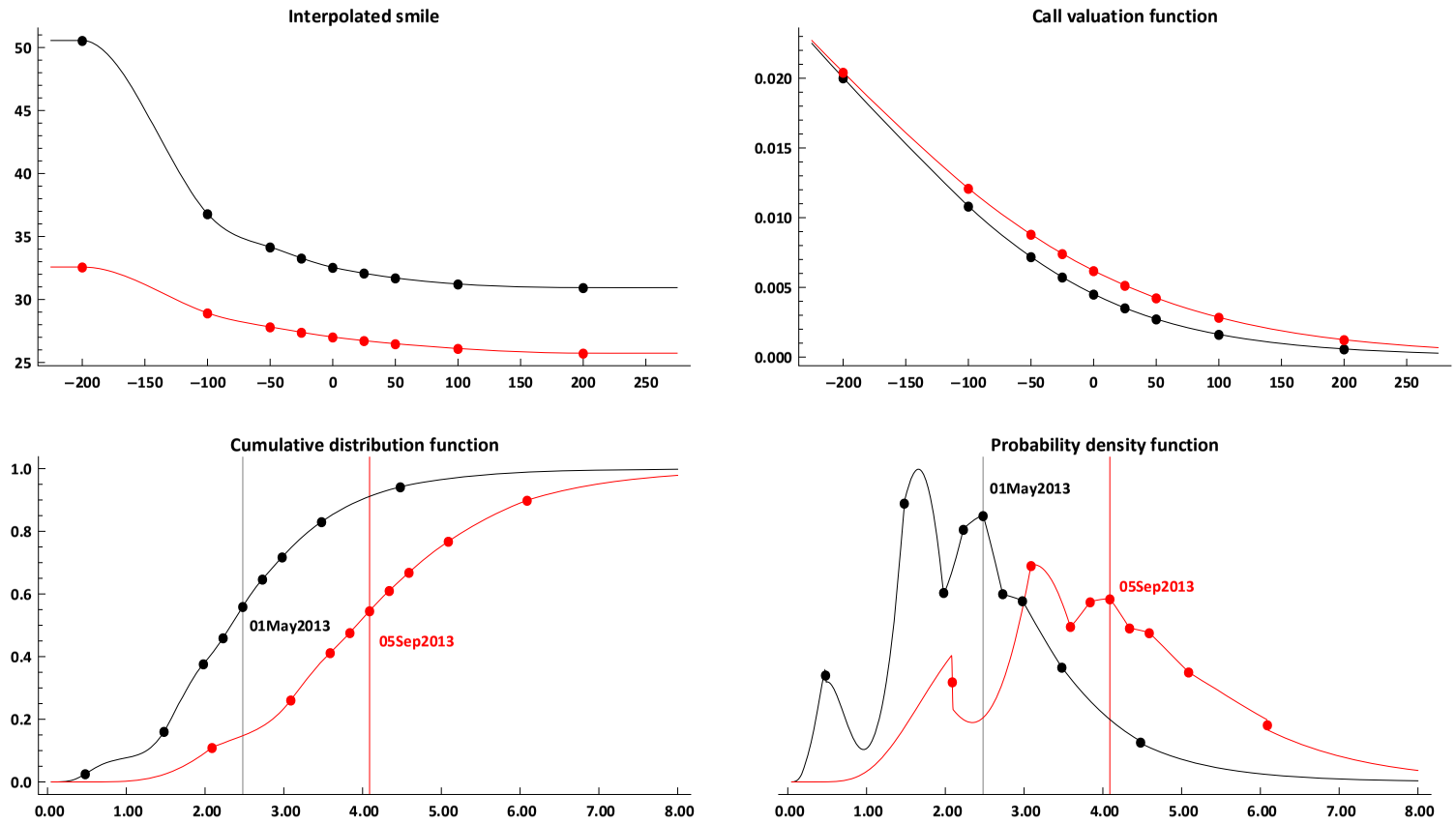
Black plots: May. 22, 2009; red plots: Nov. 18, 2011. x-axes expressed as the proportional difference between the exercise price and the forward rate. Input data/knot points are marked by dots and the values (apart from densities) displayed in Table 2. Upper left: input data and the interpolated smile. Upper right: call valuation function, evaluated using the interpolated smile. Call prices expressed as a fraction of the forward rate. Bottom panels: risk-neutral distribution and density functions. Step size $\Delta = 0.005$ (as a fraction of forward rate). For any point on the x-axis, the plot in the bottom left panel can be read as the probability the euro ends at that proportional difference from the forward rate or less in 1 month.

Figure 8: **Risk-neutral currency tail risk: EUR-USD**



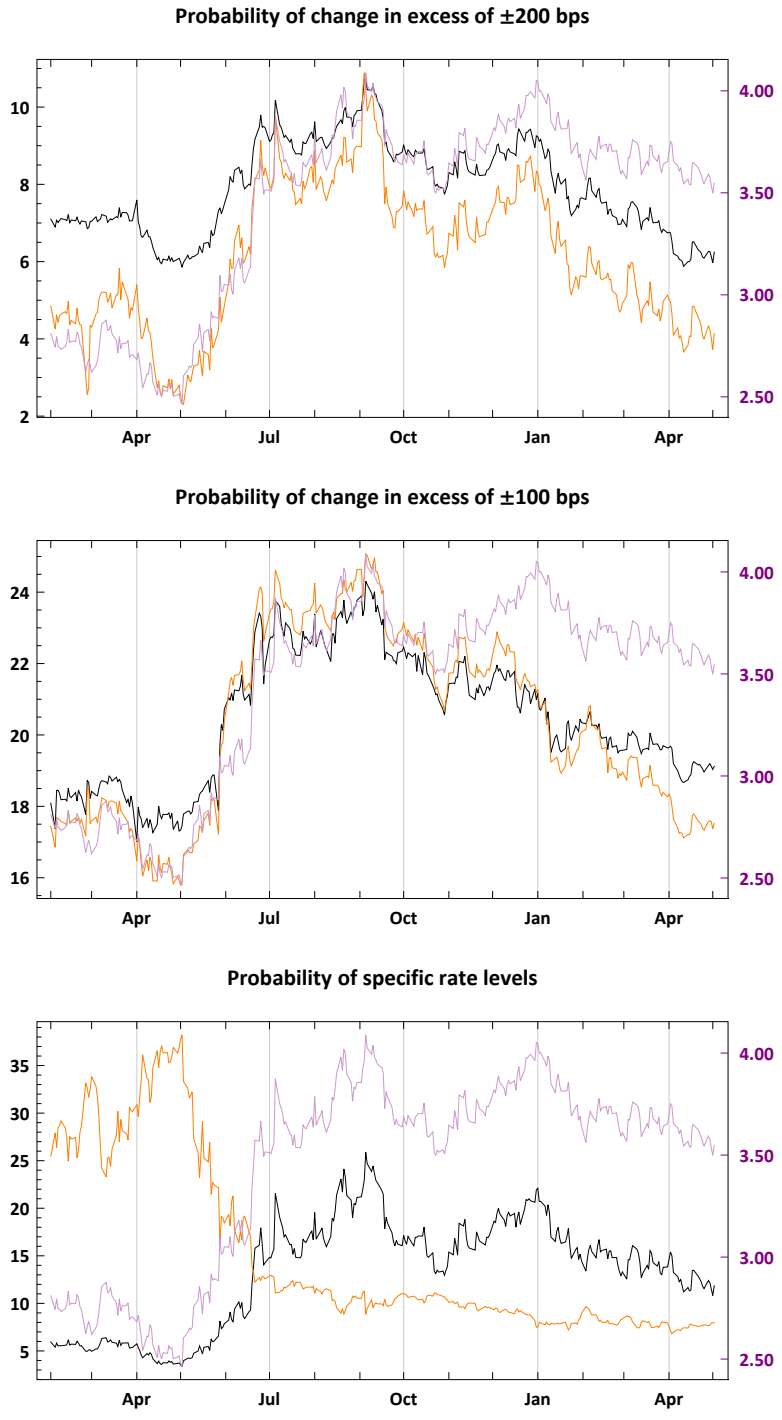
Upper panel: Black (orange) plot (left axis): risk-neutral probability of a 1-month dollar appreciation (depreciation) of at least 7.5 percent vis-à-vis the euro. Lower panel: Difference between risk-neutral probability of a 1-month dollar appreciation of at least 7.5 percent vis-à-vis the euro minus that of depreciation. Purple plot (right axis) in both panels: EUR-USD spot exchange rate. Jan. 3, 2006 to Mar. 25, 2014, $\Delta = 0.005$.

Figure 9: **Computation example: 2-year into 10-year swaptions**



Black plots: May 1, 2013; red plots: Sep. 5, 2013. In the two upper panels, the x-axes are expressed as the difference in bps between the exercise price and the forward swap rate. In the two lower panels, the x-axes are expressed as the forward swap rate. Input data/knot points are marked by dots and the values (apart from densities) displayed in Table 3. Upper left: input data and the interpolated smile. Upper right: call valuation function, evaluated using the interpolated smile. Call prices expressed in interest rate units as a decimal. Bottom panels: risk-neutral distribution and density functions. Step size $\Delta = 0.0001$ (1bp). For any point on the x-axis, the plot in the bottom left panel can be read as the probability the 10-year swap rate ends at the corresponding forward swap rate or less in 2 years.

Figure 10: **Risk-neutral interest-rate tail risk**



Upper panel: black (blue) plot (left axis) risk-neutral probability of a 2-year increase (decline) of at least 200 bps in the 10-year swap rate vis-à-vis the current forward swap rate. Center panel: black (blue) plot (left axis) risk-neutral probability of a 2-year increase (decline) of at least 100 bps in the 10-year swap rate. Lower panel: black (blue) plot (left axis) risk-neutral probability that the 10-year swap rate will be 5 percent or higher (3 percent or lower) in 2 years. Purple plot (right axis): forward swap rate. Feb. 1, 2013 to May 2, 2014; $\Delta = 0.0001$ (1 bp).