Economically justifiable dividend modelling

Peter Jäckel

VTB Capital

October 2015

Why model dividends?

- Hedge behaviour: response of option prices (=volatility) to spot moves.
- Vanilla price continuity along adjusted strike line:
  calendar arbitrage across ex-dividend dates.
- Non-vanilla prices are affected by the dividend model. There's a lot of literature on barrier options with discrete dividends.
- Choice of dividend model affects wings of (Black) volatility surface!
- A good part of the literature on the effect of dividends on non-vanillas ignores the effect of calibration to the same (Black) volatility surface.

Common dividend modelling approaches

For simplicity, we assume dividends are all intended to be *absolute* and *discrete*, and interest rates are zero[1].

1. **Escrowed dividends** [HJ88]. For some (pricing) horizon T, the spot $S(t)$ is a log-normal variate *plus all dividends yet to be paid until T*:

$$\mathrm{d}\tilde{S} = \sigma\tilde{S}\mathrm{d}W \qquad\qquad S(t) = \tilde{S}(t) + \sum_{t\leq\tau_i<T} D_i \tag{3.1}$$

2. **Log-normal minus already paid dividends** [Bla75]. The spot is a log-normal variate *minus all dividends already paid*:

$$\mathrm{d}\tilde{S} = \sigma\tilde{S}\mathrm{d}W \qquad\qquad S(t) = \tilde{S}(t) - \sum_{0\leq\tau_i<t} D_i \tag{3.2}$$

[1] Non-zero (deterministic) interest rates, dividend yields, repo rates, and proportional dividends can all be included by minor modifications and/or transformations.

3. **Piecewise log-normal diffusion**. The spot is log-normal diffusion in-between dividends and jumps at the ex-dividend date:

$$\mathrm{d}S = \sigma S \mathrm{d}W + \sum_i D_i \cdot \delta(t - \tau_i)\mathrm{d}t \qquad (3.3)$$

4. **Mixing ideas of 1 and 2 to emulate 3** (at least near the money) [BV02]:

$$\mathrm{d}\tilde{X} = \sigma \tilde{X} \mathrm{d}W \qquad S(t) = [S_0 - D_{\mathrm{near}}(t)] \cdot \tilde{X} - D_{\mathrm{far}}(t) \qquad (3.4)$$

with

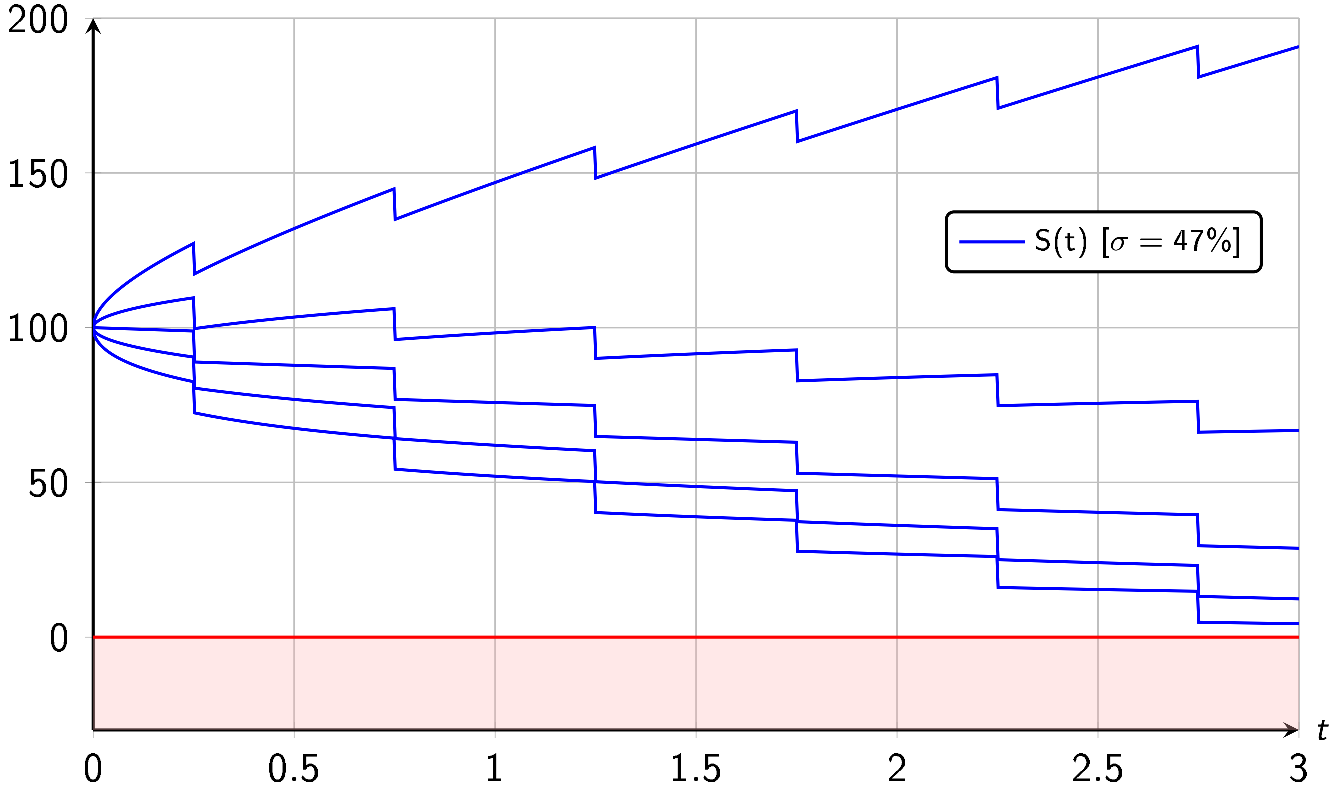
$$D_{\mathrm{near}}(t) = \sum_{0 \le \tau_i < t} \left(1 - \frac{\tau_i}{T}\right) \cdot D_i \qquad D_{\mathrm{far}}(t) = \sum_{0 \le \tau_i < t} \frac{\tau_i}{T} \cdot D_i \qquad (3.5)$$

5. **Proportional dividends**. Last but not least:

$$\mathrm{d}S = \sigma S \mathrm{d}W + \sum_i D_i \cdot \frac{S}{S_0} \cdot \delta(t - \tau_i)\mathrm{d}t \qquad (3.6)$$

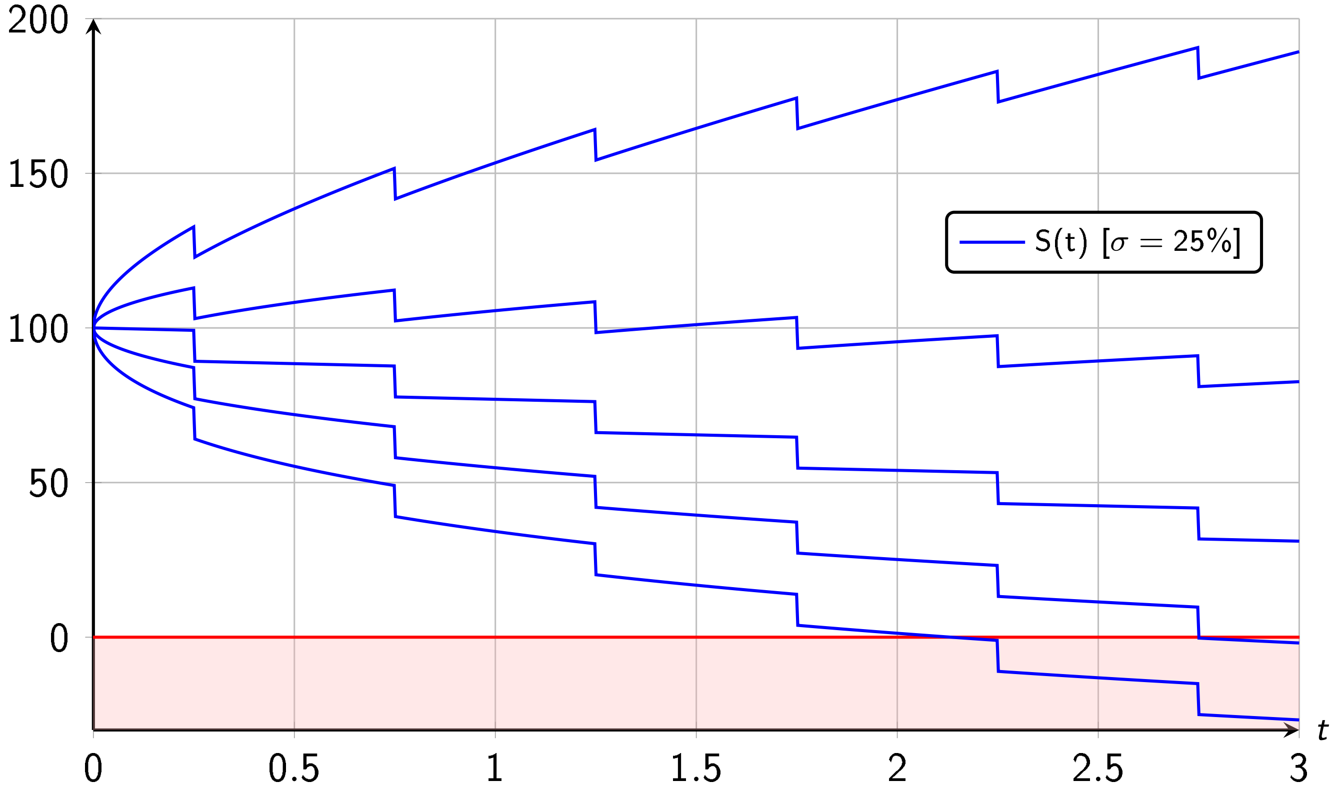
What do these approaches *look* like?

1. Escrowed dividends:



Top to bottom: 99%, 85%, 50%, 15%, and 1% quantile curves (single Karhunen-Loeve factor paths).

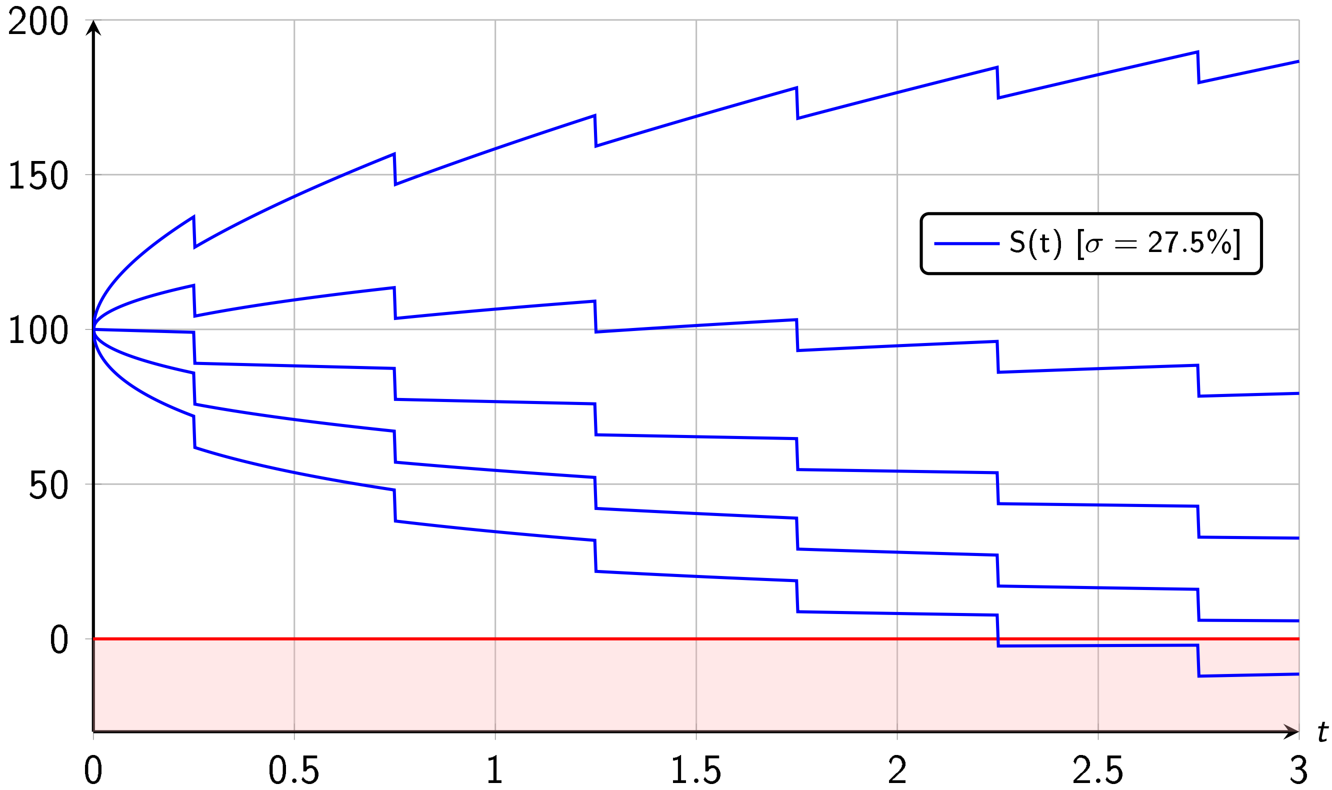
2. Log-normal minus paid dividends:



Top to bottom: 99%, 85%, 50%, 15%, and 1% quantile curves (single Karhunen-Loeve factor paths).

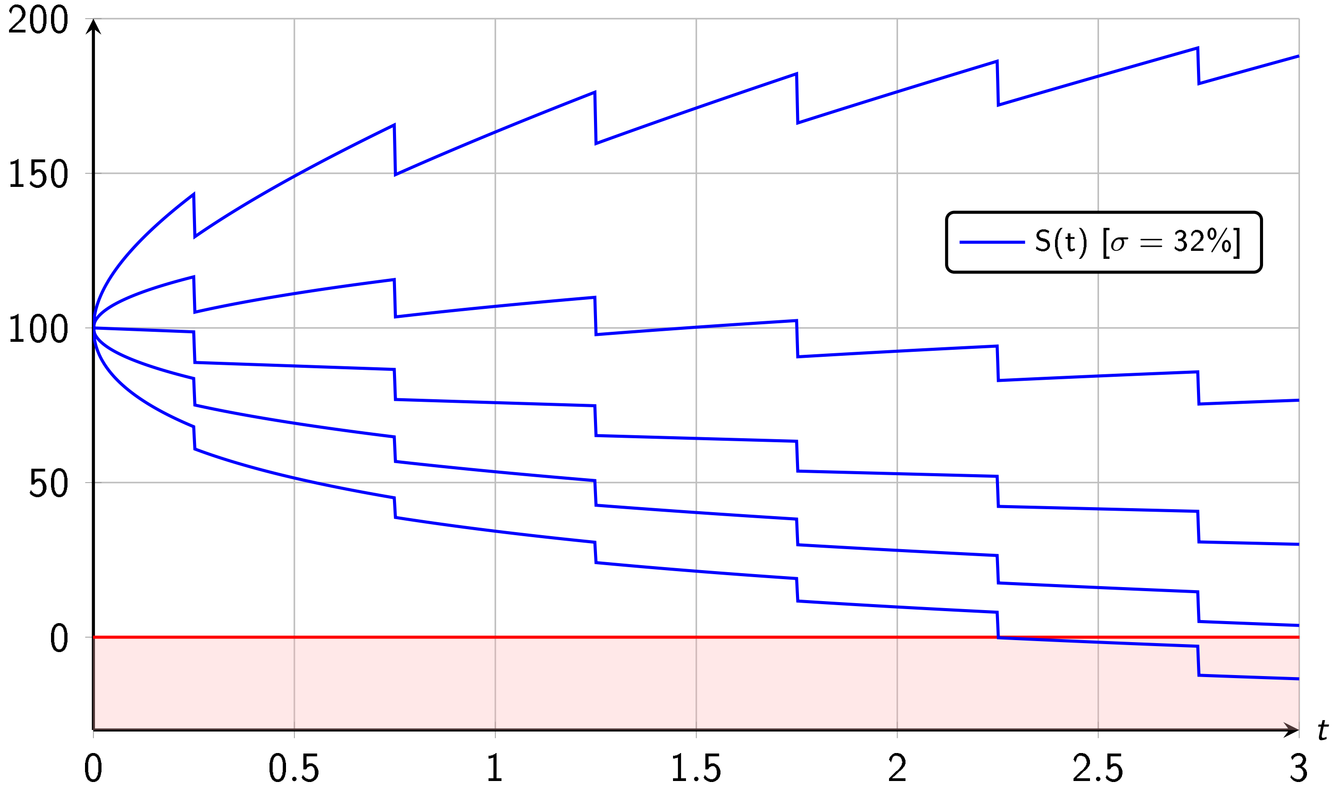
2 Common dividend models

3. Piecewise log-normal diffusion:



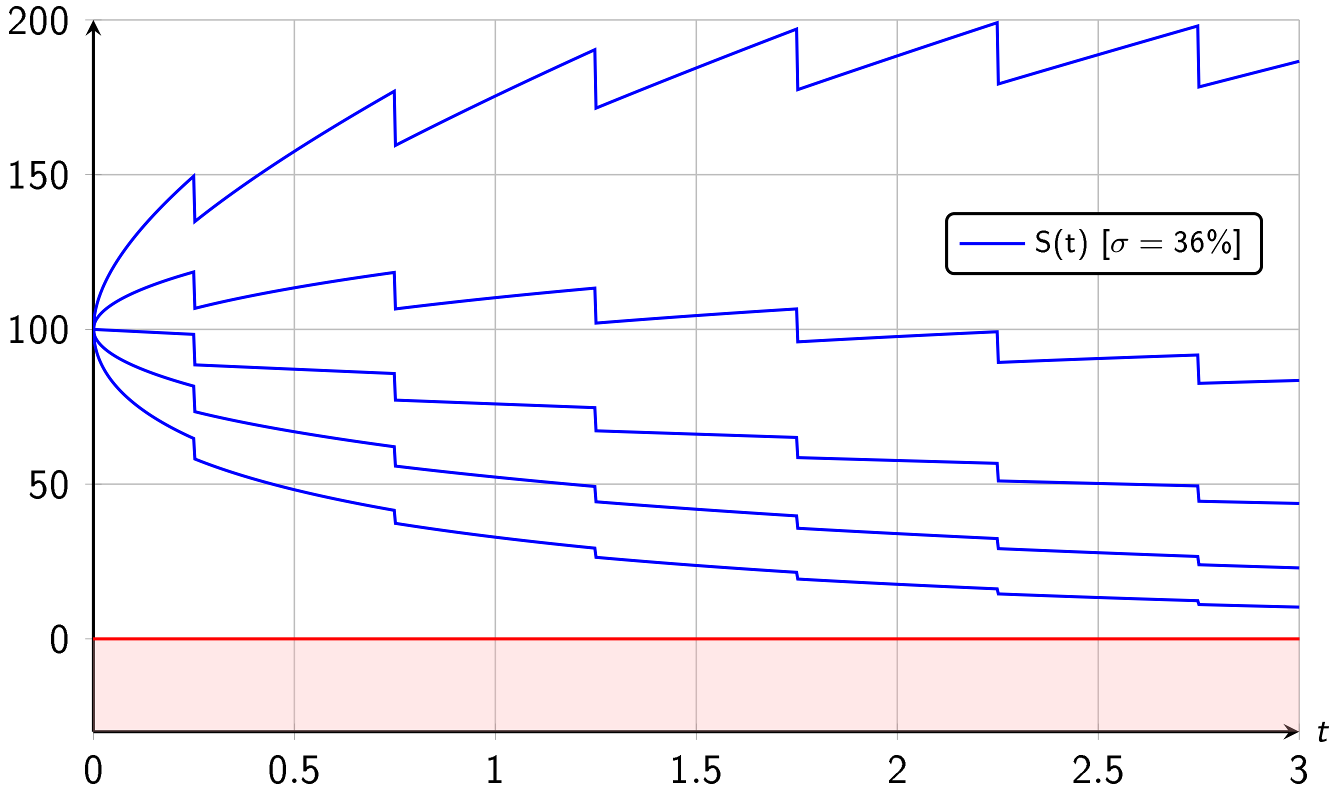
Top to bottom: 99%, 85%, 50%, 15%, and 1% quantile curves (single Karhunen-Loeve factor paths).

❹ Mixing ideas of ❶ and ❷ to emulate ❸ [BV02]:



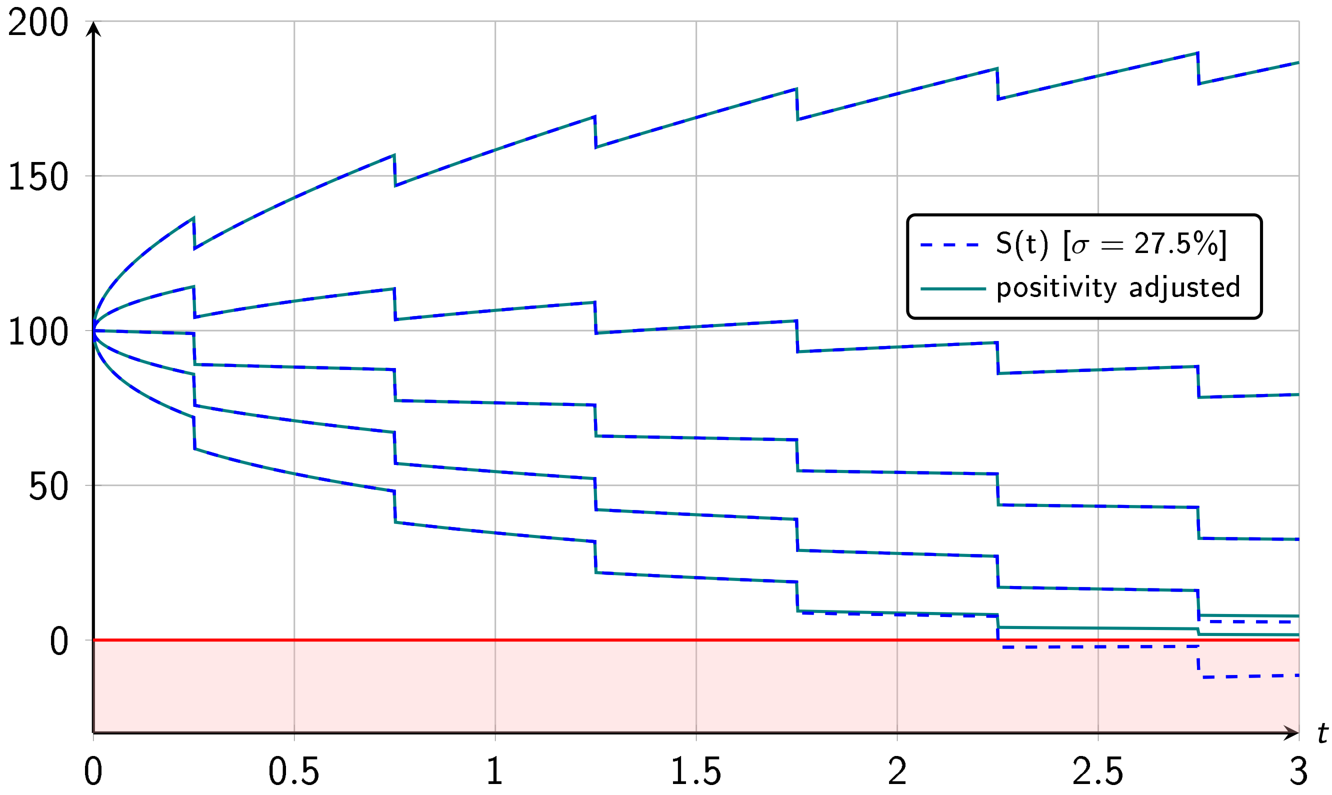
Top to bottom: 99%, 85%, 50%, 15%, and 1% quantile curves (single Karhunen-Loeve factor paths).

❺ Proportional dividends:



Top to bottom: 99%, 85%, 50%, 15%, and 1% quantile curves (single Karhunen-Loeve factor paths).

Ideally, we want piecewise log-normal diffusion, adjusted to remain positive:



Top to bottom: 99%, 85%, 50%, 15%, and 1% quantile curves (single Karhunen-Loeve factor paths).

Drawbacks of the respective approaches

- Subtracting from the spot lowers effective instantaneous volatility at the front (small t).

  Affects: 1 and 4

- Reducing the subtraction as $t \to T$ raises effective instantaneous volatility at the back (large t).

  Affects: 1

- Short term dividends should be absolute cash (no measurable effective proportional component).

  Affects: 4 and 5

- **Short term put options for low strikes** (at the escrow amount) ***should not* have zero value**.

  Affects: 1

- **Put options at zero strike *should* have zero value**.

  Affects: 2and 3 and 4

- There are **legal restrictions on the maximum dividend** any company can pay depending on their equity.

  Dividends cannot bankrupt a company or deplete equity to zero.

  Affects: 1 and 2 and 3 and 4

All common (cash) dividend models violate economic fundamentals.

V. Frishling [Fri02]:

"*The answer to the question of which model reflects reality better and provides consistent pricing across a range of options is almost self-evident. It seems that the third model is more in agreement with the actual evolution of the price process and should be used, particularly if pricing of exotics is required. It should be pointed out that the three models compared in this note are well known, although the second one, which models the accumu- lation process, is less familiar than the others. However, no source known to us analyses the differences between these models or recommends one in preference to another. Moreover,* ***it appears that some commercial systems use or at least provide these models quite arbitrarily and inconsistently****, thus leading to potentially dangerous mispricing and mishedging of the portfo- lios.* ***We have no doubt that many practitioners and quantitative analysts grappled with this problem*** *at some time in their careers and may have arrived at conclusions similar to those outlined. If this article makes the quantitative analyst community more aware of these pitfalls, and helps them to price their positions correctly, then it will have fulfilled its purpose.*"

UK Companies Act 2006[2]

830. Distributions to be made only out of profits available for the purpose

(1) A company may only make a distribution out of profits available for the purpose.

(2) A company's profits available for distribution are its accumulated, realised profits, so far as not previously utilised by distribution or capitalisation, less its accumulated, realised losses, so far as not previously written off in a reduction or reorganisation of capital duly made.

The details of the legislation are intricate and vary by jurisdiction.

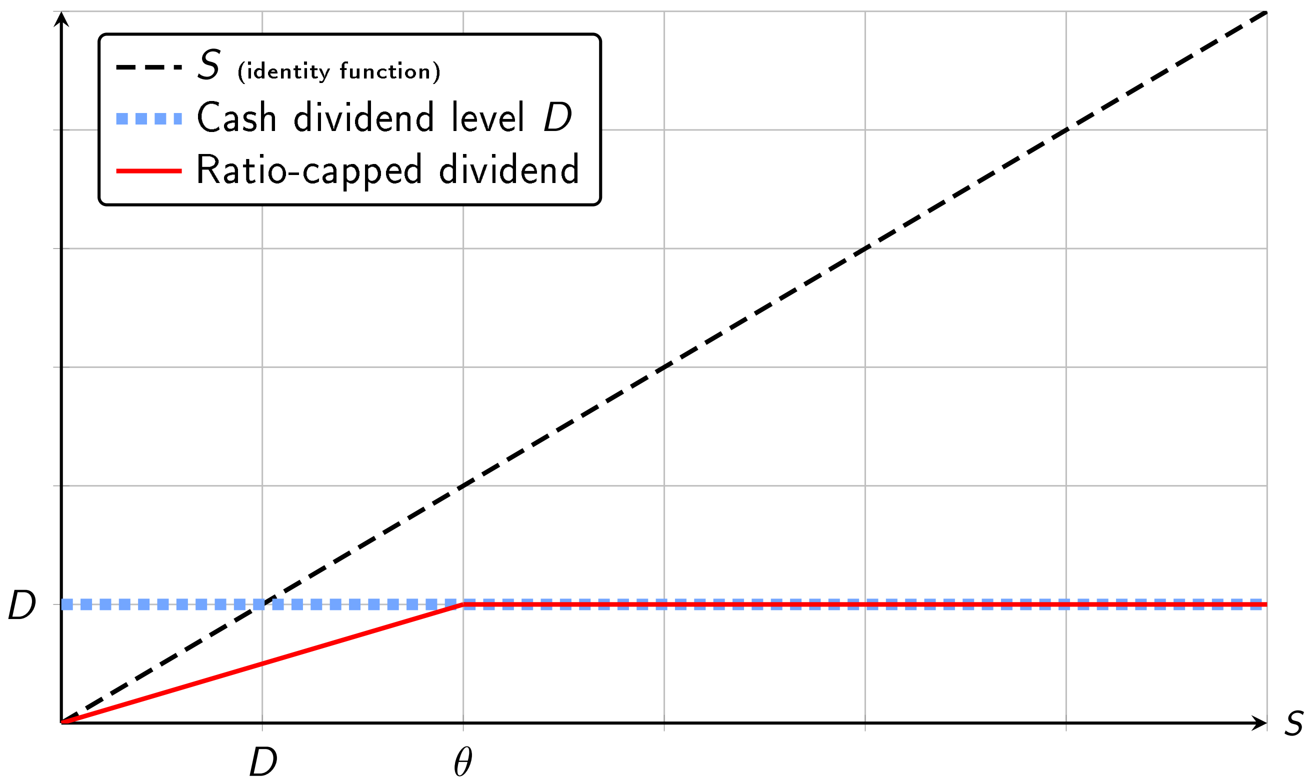
It is fairly safe to assume as a general guideline that **a dividend cannot exceed a certain percentage of the current equity price**.

In a simplistic interpretation, this means **if the spot drops below a certain threshold θ, the dividend *must* be reduced**.

Even if this is not *exactly* in line with actual legislation, it serves as an **excellent, realistic, and economically justifiable dividend model**.

[2] Dividends belong to the category "distributions".

Assume the maximum dividend-to-spot ratio is 50%, i.e., $\theta = 2 \cdot D$.



Now for the details. We use the following simple ***dividend process*** model:-

- Define the cash dividend forecast D as the drop in the forward curve across the ex-dividend date T as seen out of today.

$$D := \underbrace{F(T_-)}_{F_-} - \underbrace{F(T_+)}_{F_+} \tag{4.1}$$

- Define the jump of the spot across the ex-dividend date from $S_- := S(T_-)$ to $S_+ := S(T_+)$ given by a chosen transition function

$$S_+ = f(S_-) . \tag{4.2}$$

- As a balanced choice between matching the reality of dividend cuts for collapsed spots, and simplicity, we choose a continuous piecewise affine function $f(\cdot)$ that comprises:
  - an outright downwards jump by $D^* \approx D$, i.e., an actual cash dividend,
  - *unless* the spot S is below some threshold $\theta \approx 2D$,
  - with $f(0) = 0$.

This gives us

$$S_+ = f(S^-) = \begin{cases} \left(1 - \dfrac{D^*}{\theta}\right) \cdot S_- & \text{if } S_- \leq \theta \\ S_- - D^* & \text{if } S_- > \theta \end{cases} \tag{4.3}$$

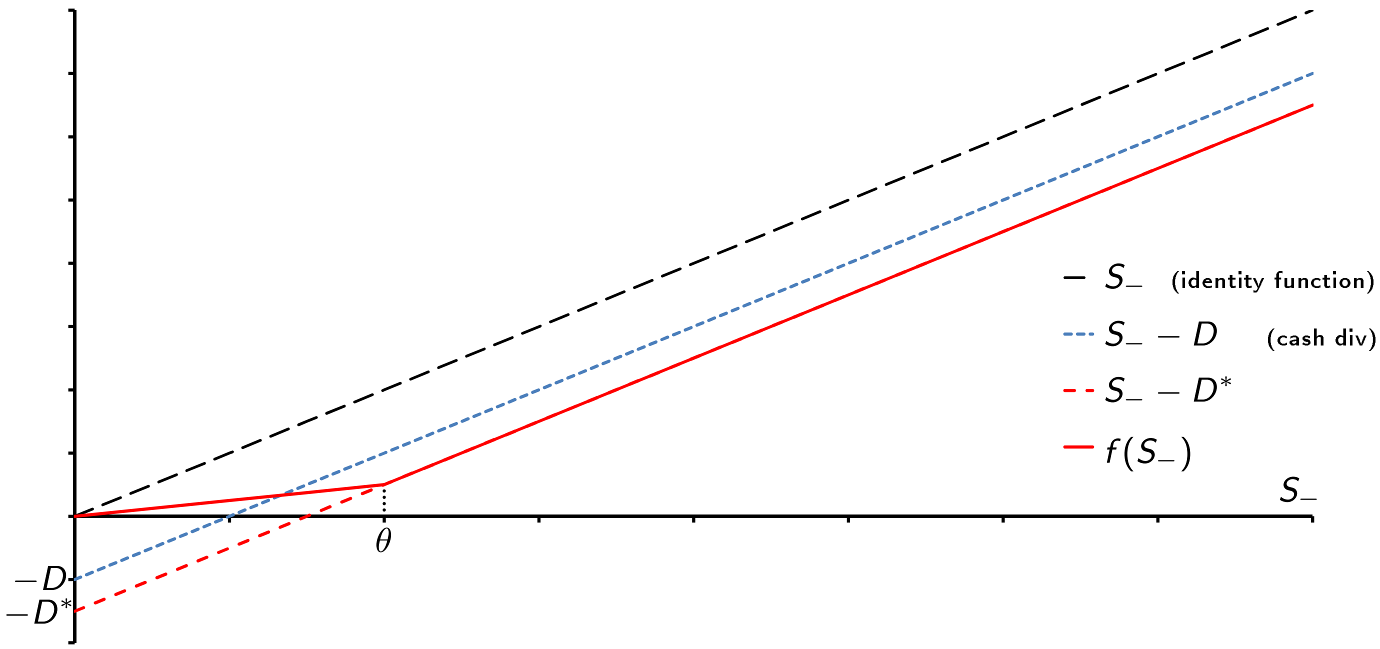
We observe that, unless volatility is zero, we must have

$$D^* \neq D \tag{4.4}$$

but very, very, very close, especially for near-dated dividends.

- **NOTE** that the above piecewise affine modelling is **not to be confused** with what is sometimes referred to as *affine dividends*.
- Those “affine dividends” are in fact a split into absolute and proportional parts of the dividend.
- We will add proportional contributions soon.

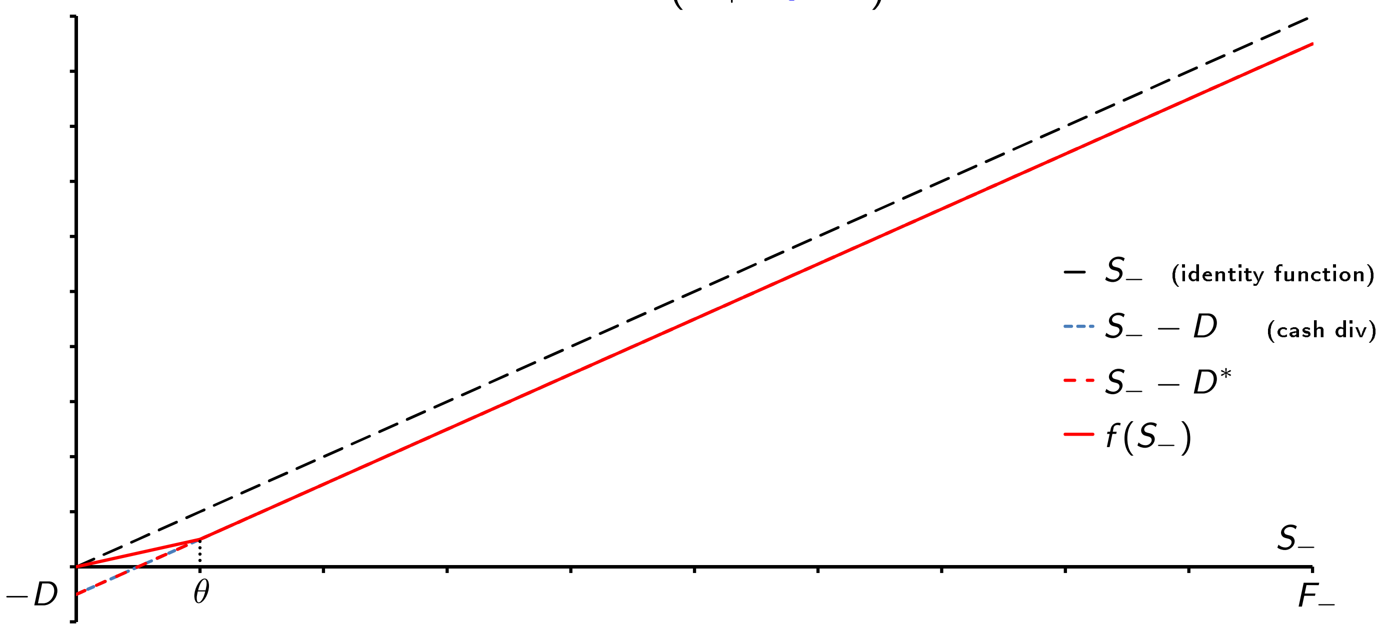
Schematically:



The actual difference between D and D^* is ***tiny for real data***.

For $T = 1$, $\sigma_{\text{ATF}} = 15\%$, $F_- = 100$, $D = 5$, we have (smile data to follow):

$$D^* \approx D \cdot (1 + \mathbf{10^{-11}})$$



Assuming θ is given, D^* is uniquely determined by (4.1).
Rewriting the spot transition function $f(\cdot)$ given in (4.3):

$$f(S_-) \;=\; S_- - D^* + \frac{D^*}{\theta}(\theta - S_-)_+ \tag{4.5}$$

and taking expecations, we obtain

$$F_+ - F_- \;=\; -D \;=\; -D_* + \frac{D^*}{\theta}\cdot p_-(\theta)\ , \tag{4.6}$$

and thus

$$D^* \;=\; \frac{D}{1-\frac{p_-(\theta)}{\theta}} \tag{4.7}$$

where $p_-(\theta)$ is the **price of a T_--expiry put option struck at θ**.

We obtain D^* *analytically* from D and the smile.

The option price transition rules for puts and calls pre- $[p_-(K)$ and $c_-(K)]$ and post-dividend $[p_+(K)$ and $c_+(K)]$ can be readily derived:-

$$p_+(S_+) = \begin{cases} \left(1-\frac{D^*}{\theta}\right)\cdot p_-(S_-) & \text{if } S_- \le \theta \\ p_-(S_-) - \frac{D^*}{\theta}\cdot p_-(\theta) & \text{if } S_- > \theta \end{cases} \tag{4.8}$$

$$c_+(S_+) = \begin{cases} \left(1-\frac{D^*}{\theta}\right)\cdot c_-(S_-) + \frac{D^*}{\theta}\cdot c_-(\theta) & \text{if } S_- \le \theta \\ c_-(S_-) & \text{if } S_- > \theta \end{cases} \tag{4.9}$$

where $c_-(K)$ is the price of a T_--call struck at K.

If we include a proportional dividend component δ such that

$$F_+ \; = \; (1-\delta)\cdot F_- - D\;, \tag{4.10}$$

we have the spot transition rule

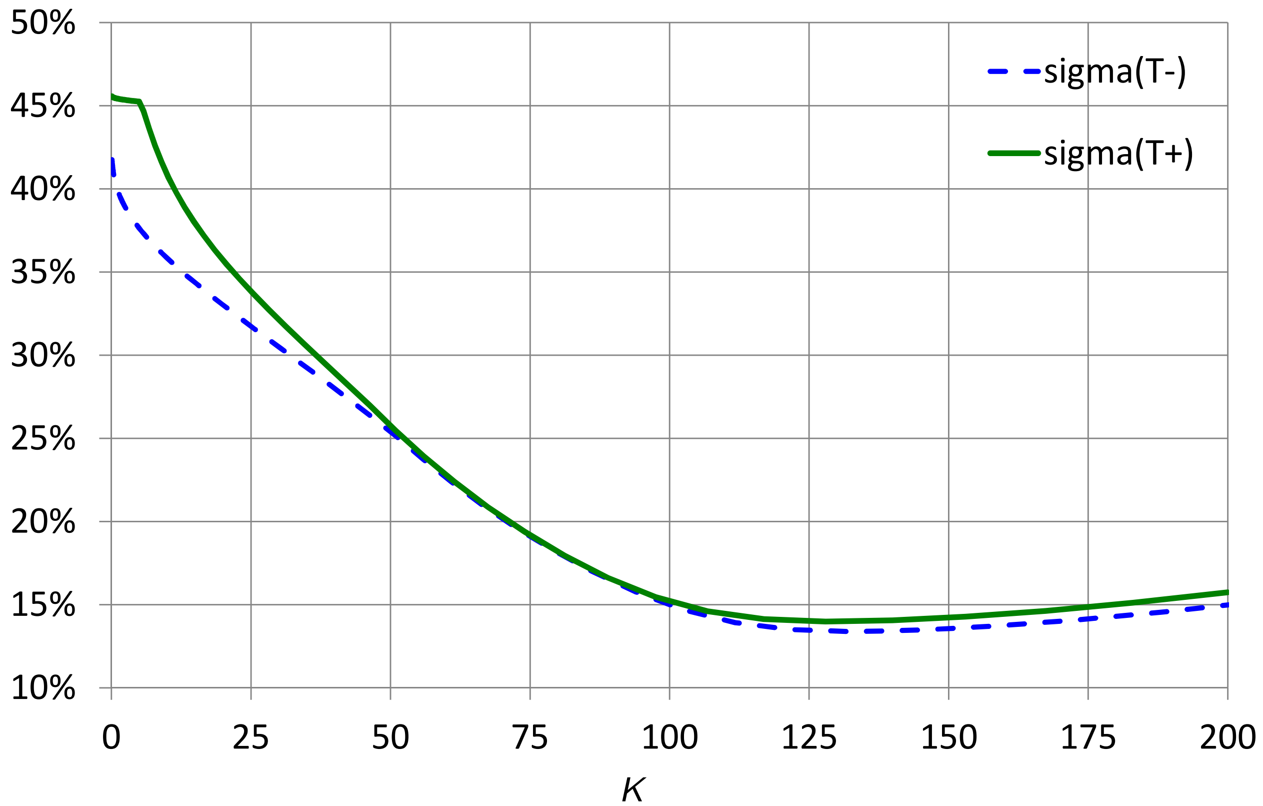
$$S_+ \; = \; f(S_-) \; = \; \begin{cases} \left((1-\delta) - \frac{D^*}{\theta}\right)\cdot S_- & \text{if } S_- \leq \theta \\ (1-\delta)\cdot S^- - D^* & \text{if } S^- > \theta \end{cases} . \tag{4.11}$$

and equation (4.7) for D^* still holds. The price transitions become:-

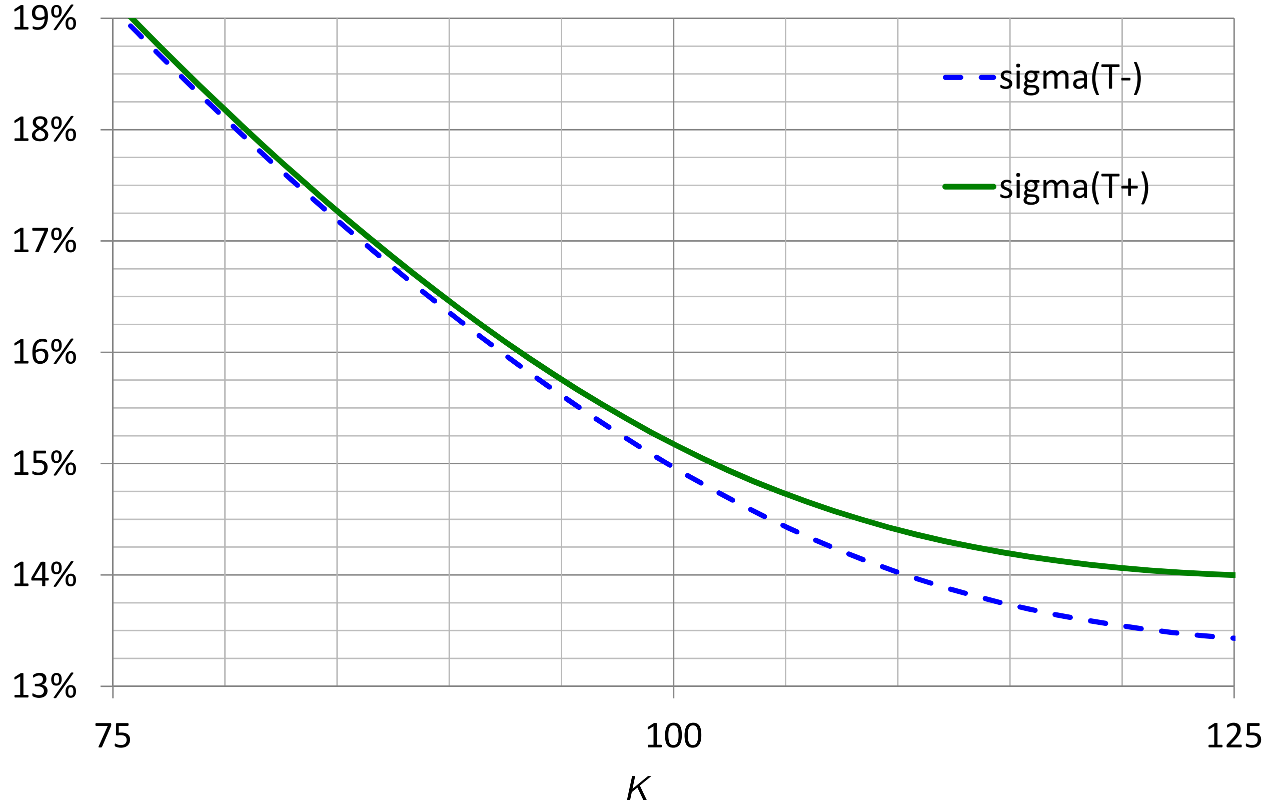
$$p_+(S_+) \; = \; \begin{cases} \left((1-\delta) - \frac{D^*}{\theta}\right)\cdot p_-(S_-) & \text{if } S_- \leq \theta \\ (1-\delta)\cdot p_-(S_-) - \frac{D^*}{\theta}\cdot p_-(\theta) & \text{if } S_- > \theta \end{cases} \tag{4.12}$$

$$c_+(S_+) \; = \; \begin{cases} \left((1-\delta) - \frac{D^*}{\theta}\right)\cdot c_-(S_-) + \frac{D^*}{\theta}\cdot c_-(\theta) & \text{if } S_- \leq \theta \\ (1-\delta)\cdot c_-(S_-) & \text{if } S_- > \theta \end{cases} \tag{4.13}$$

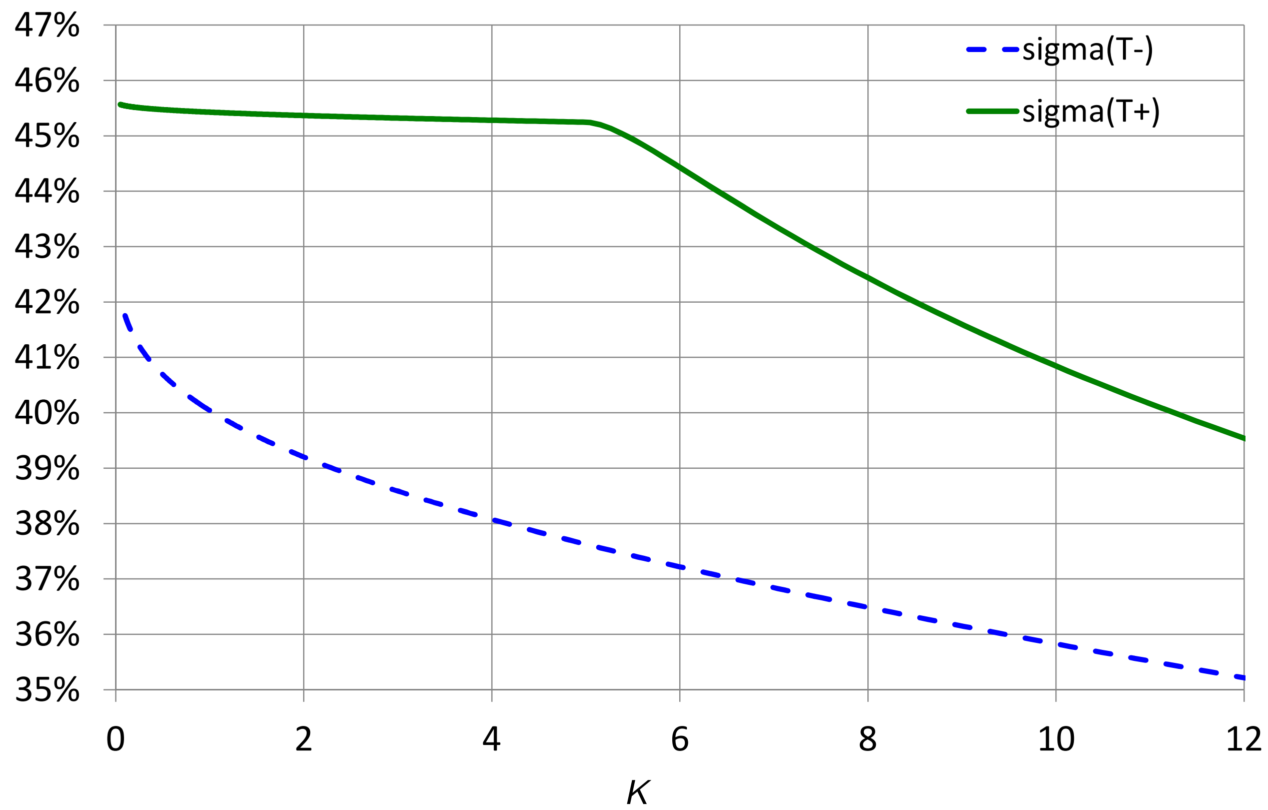
As before, $T = 1$, $F_- = 100$, $D = 5$ (all volatilities are Black):-



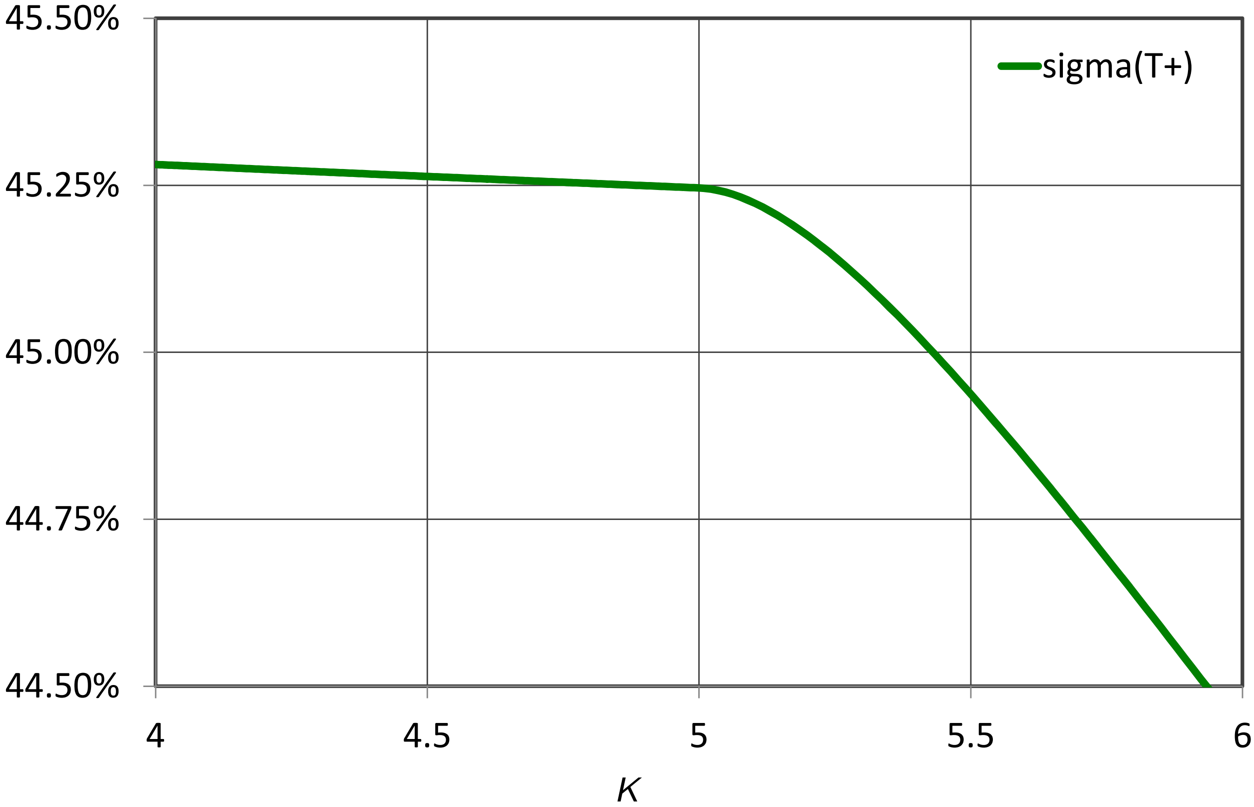
Near the money:-



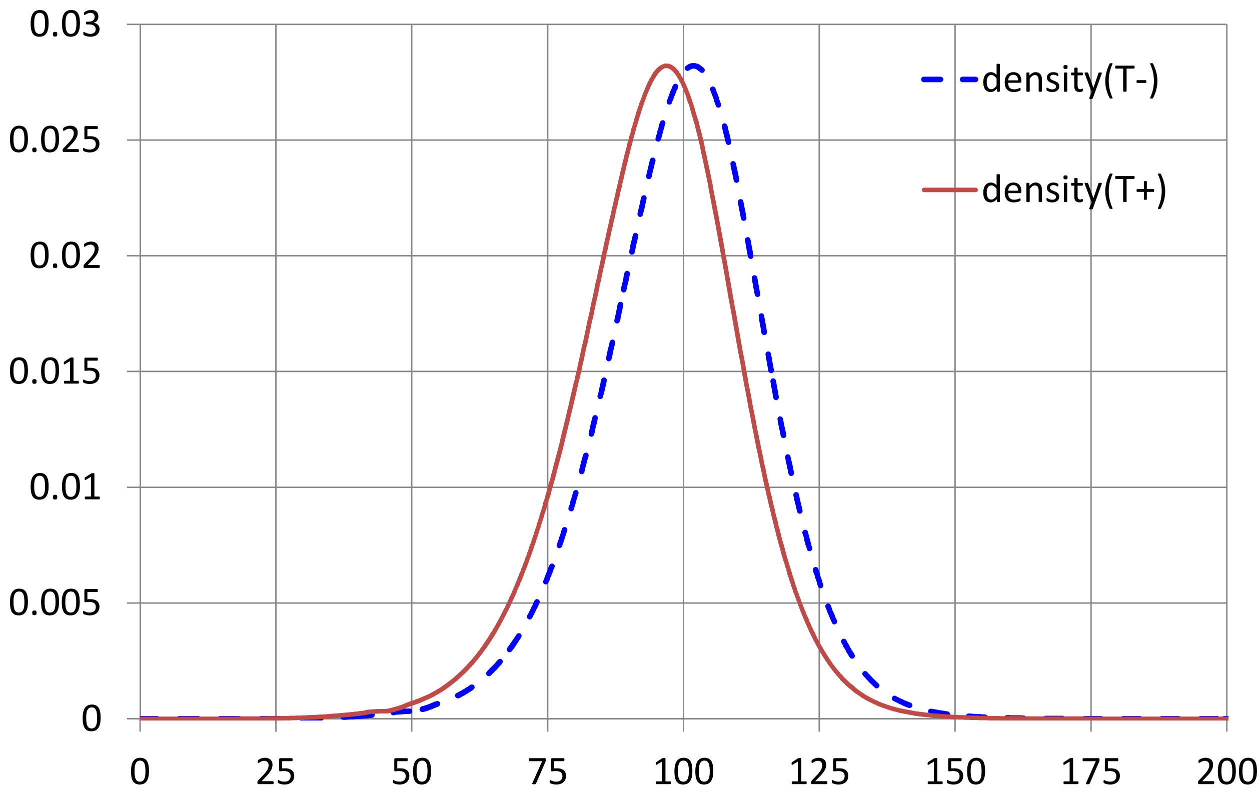
Near zero:-



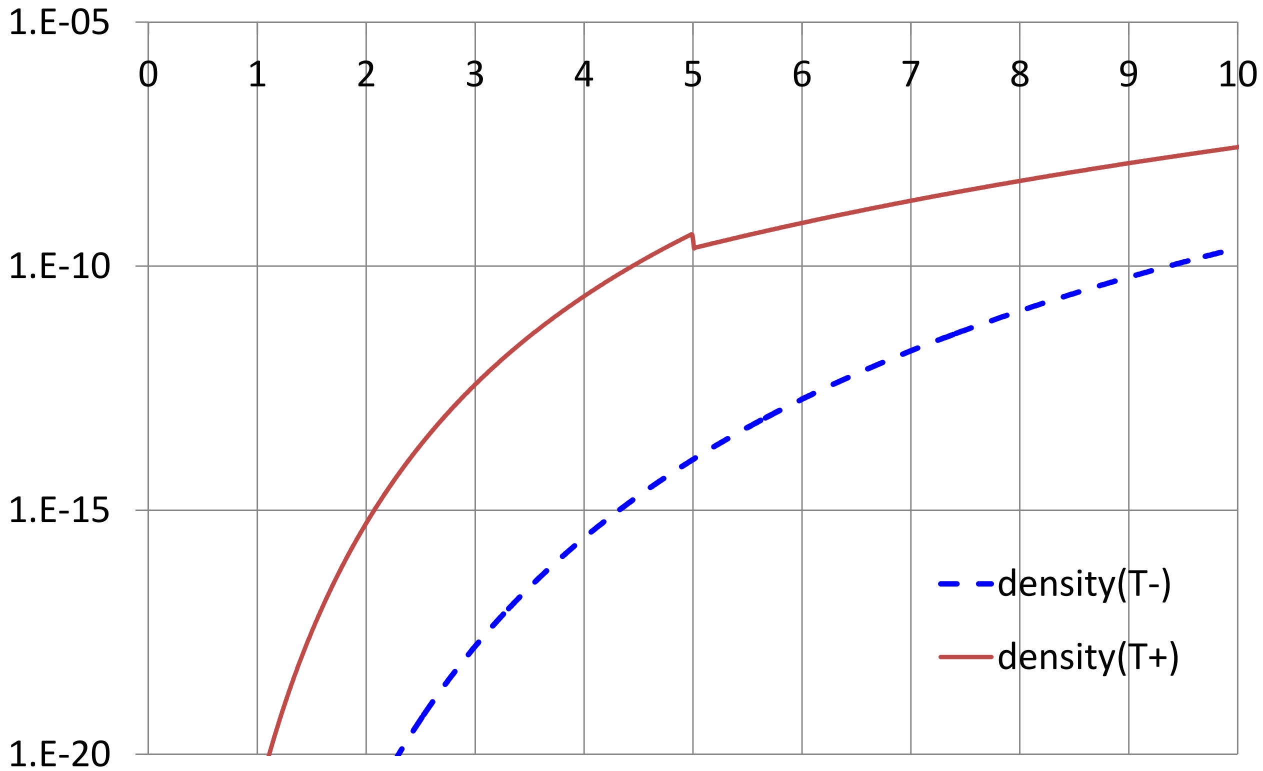
Near $D = 5$, it's a *bend* not a *kink*:-



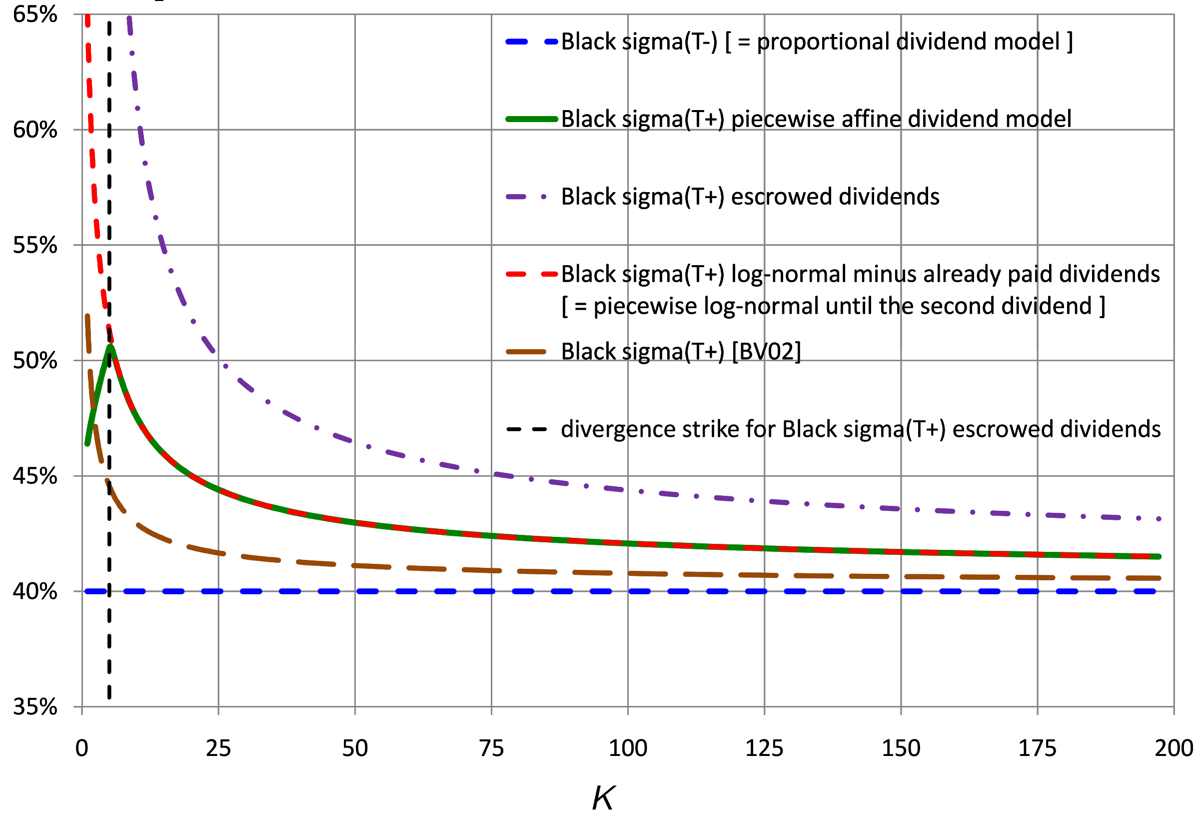
The risk-neutral density does what we expect, *a macroscopic translation*:-



At $D = 5$ it incurs an upwards shift due to the compression towards zero:-



Now: $T_- \equiv T_1 = 0.75$, Black $\hat{\sigma}(T_-) = 40\%$ for all strikes, $F(T_-) = 100$, $D(T_1) = 5$
For the escrowed dividend model, assume: $T_2 = 1.75$, $D(T_2) = 5$, $T = 2$.
We have at T_1:



Parametric volatility with cash dividends

Incorporating this translation mechanism into any simple parametric formula is inherently doomed.

No matter what *smart* analytics are being put in place, somewhere, it'll creak at the seams, foiling all the *"clever"* effort.

Life isn't simple.
Cars aren't simple, but perfectly workable engineering problems.
Why should dividends be simple to the point of being a one line formula?

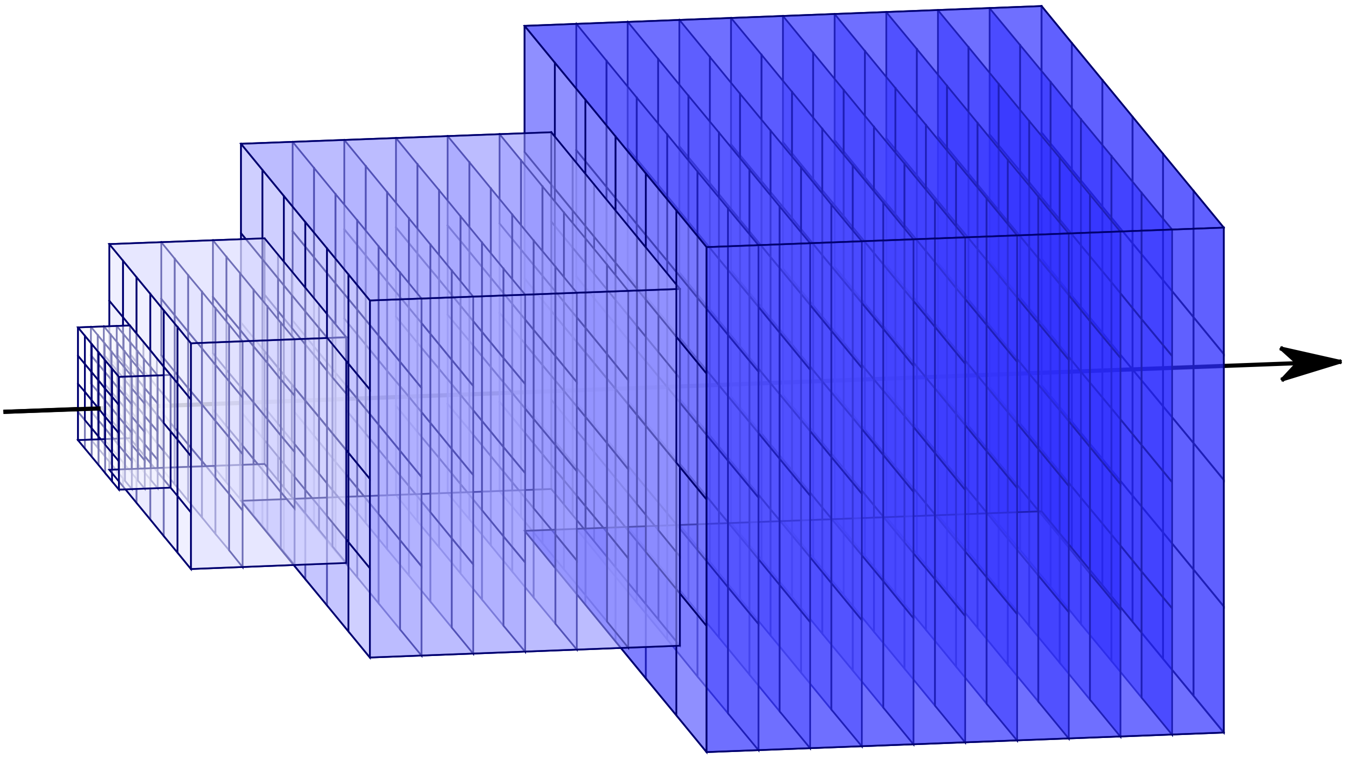
A. Lipton [2006] :

"The hunt for closed form solutions is ultimately nothing but the pursuit of fool's gold."

The days of simple formulae are gone.

The age of engineering solutions has come.

- In the *"DHI framework"* [JAL14], we introduced the construction of a parametric implied volatility surface on a sparse lattice in 2D.
- This approach is robust and fast, and gives us full flexibility and control over any further modelling aspects.
- Incorporating a different dividend model is almost trivial.
- Here is why:
- As we propagate in time, it is necessary to resize the 2D lattice every now and then to dimensions of local relevance at the given time horizon.
- We refer to a sequence of steps with constant lattice layout as a *box*.



Lattice *boxes*

- At *box transition points*, we need to redistribute the discrete probability masses from the previous lattice to the new lattice.
- This is not the same task as interpolating a continuous value function such as a contract price or a payoff!
- **The conditions for a meaningful transfer are subtle.**
- Conventional concepts of "interpolation" are simply not applicable to the probability translation problem.
- The most important requirement is: **continuity of option prices**!
- *We use an arbitrage-free implied volatility interpolator [Jäc14] constructed from the previous lattice nodes to infer the distribution on the new lattice.*

- First, we compute out-of-the money option prices $\{v_i\}$, $i = 1...n_z$, struck at the new lattice's spot levels, *using the interpolator's implied volatilities!*

- Dropping the leftmost and rightmost nodes, we bootstrap a set of discrete z-marginal probabilities $\tilde{p}'(z'_i)$ such that

$$\begin{aligned}
v'_k &= \sum_{i=1}^{i_{\min}(\mathrm{e}^{z'_k})} \tilde{p}'(z'_i) \cdot (\mathrm{e}^{z'_k} - \mathrm{e}^{z'_i}) \qquad \text{for } 1 < k < \tfrac{n_z+1}{2} \\
v'_k &= \sum_{i=i_{\min}(\mathrm{e}^{z'_k})}^{n_z} \tilde{p}'(z'_i) \cdot (\mathrm{e}^{z'_i} - \mathrm{e}^{z'_k}) \qquad \text{for } \tfrac{n_z+1}{2} < k < n_z \qquad (5.1) \\
\tilde{p}'(z'_{\frac{n_z+1}{2}}) &= 1 - \sum_{i \neq \frac{n_z+1}{2}} \tilde{p}'(z'_i)
\end{aligned}$$

where we have assumed that n_z is odd and greater than four.

4 Cash dividends in (*not*) SABR? | Transfer via Implied Volatility | From prices back to probabilities

- We then redistribute p'_i in the y direction by building a two-dimensional interpolator $Q(z, y)$ of discrete probabilities from $\tilde{p}(z_i, y_j)$, i.e., from the data of the earlier box's lattice.

- We use this two-dimensional interpolator for the purpose of interpolation in the y-direction, *conditional on a given z-level*, and so to distribute the z-marginal probability mass at some level z_i in the y-direction.

- After flooring and conditioning, the discrete Green's function on the new lattice is given by

$$\tilde{p}(z'_i, y'_j) = \tilde{p}'(z_i) \cdot \frac{\left(Q(z'_i, y'_j)\right)_+}{\sum_l \left(Q(z'_i, y'_l)\right)_+} . \qquad (5.2)$$

This procedure is a generic methodology to redistribute discrete probabilities from one set of discrete nodes to another, whilst preserving the quantities that are of most importance in our context, namely:-

- the sum of all probabilities
- the expectation of the underlying, i.e., the forward,
- the prices of vanilla options on the new lattice's nodes as implied by the earlier lattice's probability distribution.

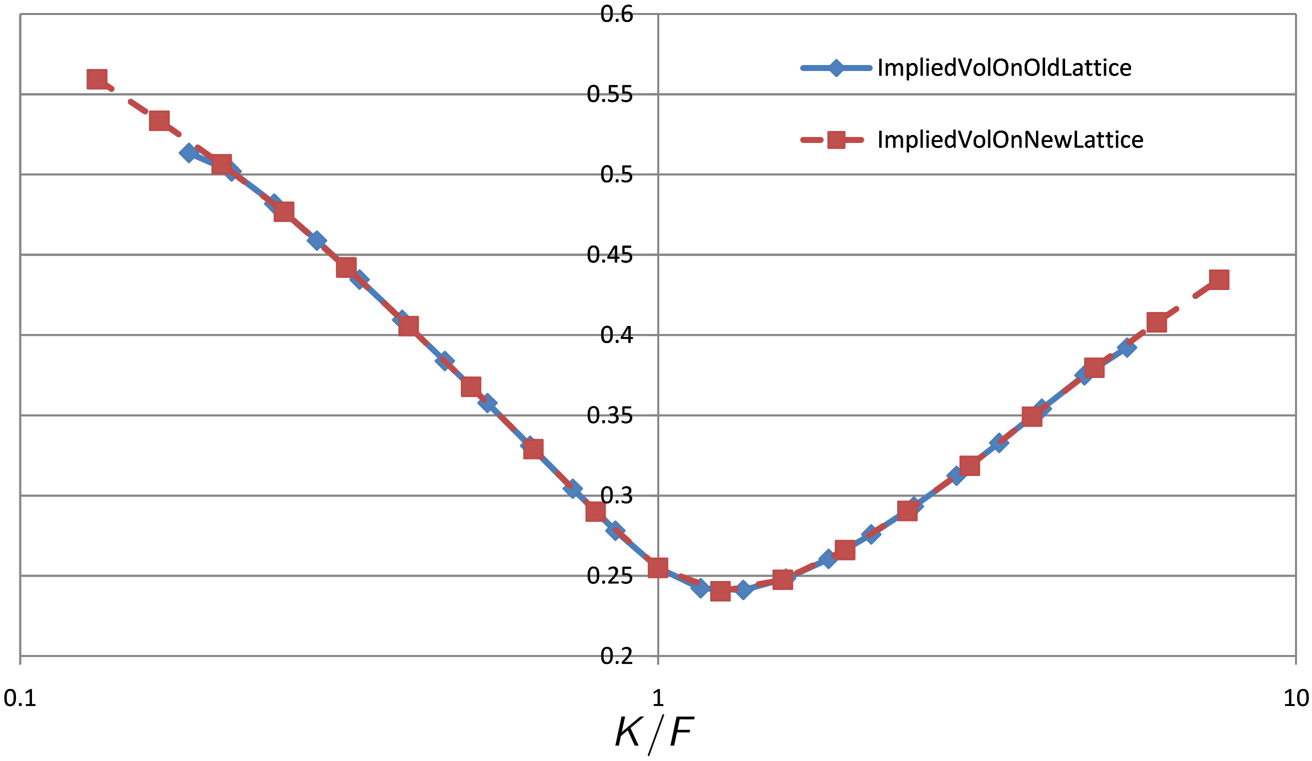
This method of translating a set of discrete probabilities from one lattice discretisation to another is in its own right a subject that is little documented in the literature and can be of use in other contexts.

For the DHI parameters

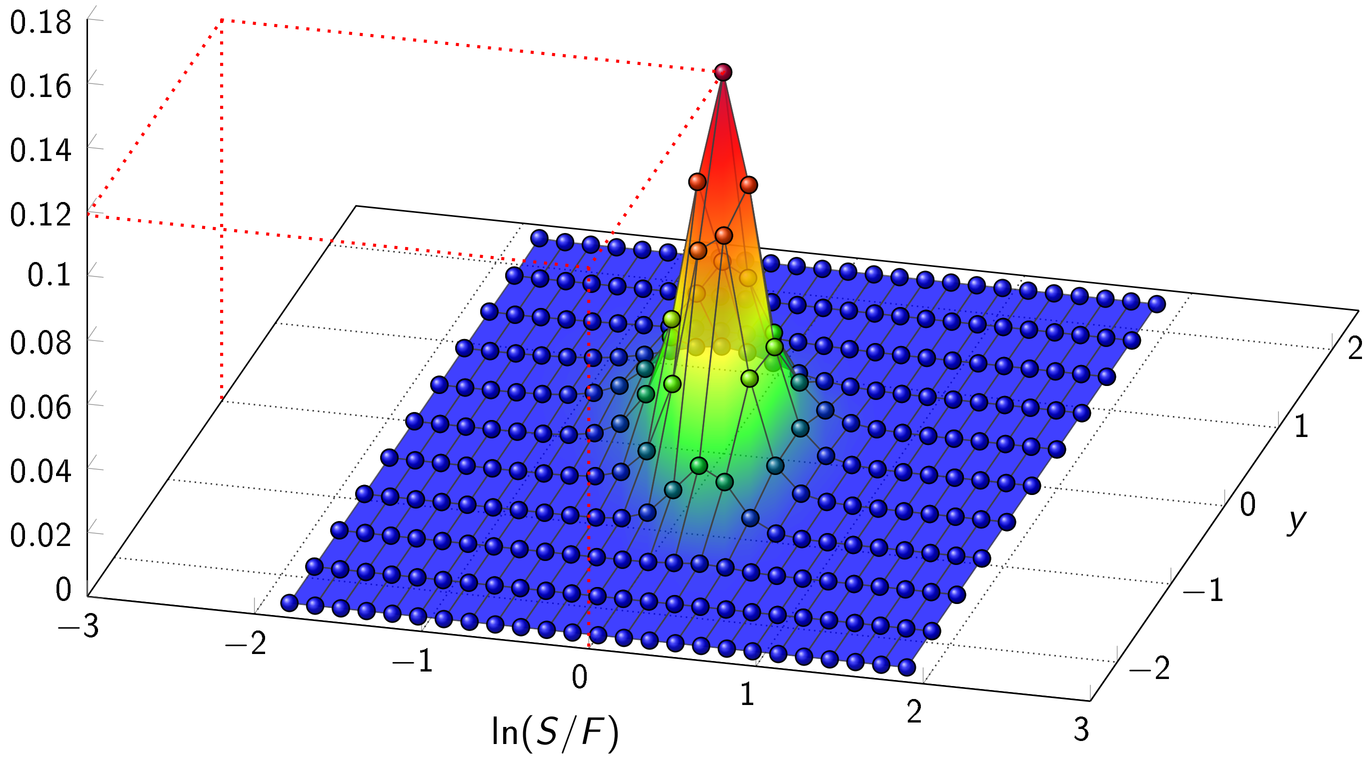
β	κ	σ	α	ρ
1	1	25%	50%	-50%

, at $T = 1$,

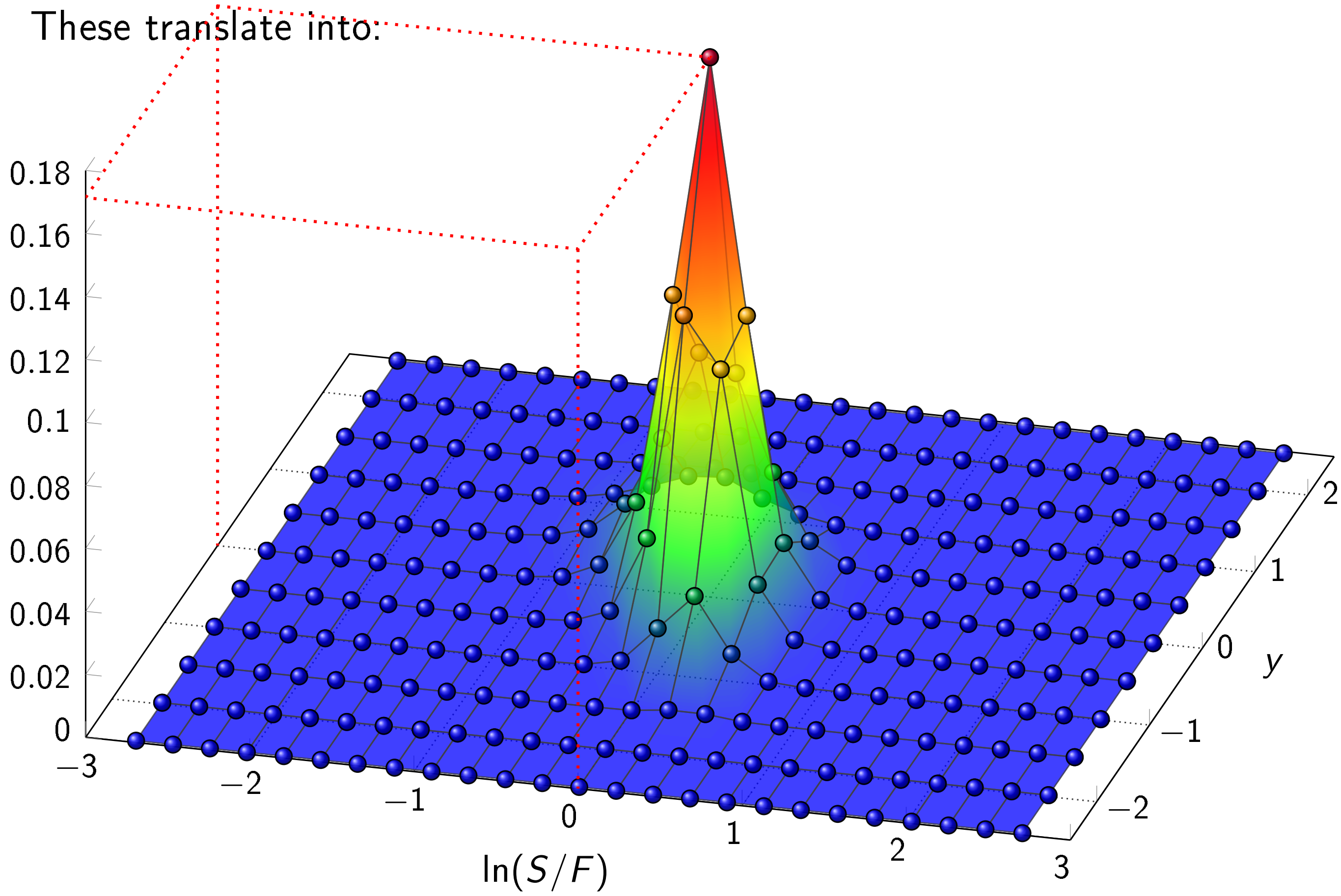
the implied volatility transition nodes are:



The discrete probabilities before the transition are:



These translate into:



Within the DHI framework, we can include this piecewise affine dividend model with ease using the techniques developed for box transitions!

- We only need *a box transition, with or without change of the lattice layout*.
- **We populate the probability levels at T_+ from option prices struck at the T_+ lattice node levels**, via the price transition rules (4.12) and (4.13).
- This of course requires option prices at T_- at strikes at which there are no lattice nodes. For this, ***we have the implied volatility interpolator at T_- as at any other box transition.***
- **All else is already in place! And not a shred of arbitrage!**
- Between dividend dates, the dynamic evolution is unchanged. *We only need the probability transitions at the dividend date.*
- This is another example where the probability transfer method via transformation to an implied volatility smile, and back, is of great use.

When applying the piecewise affine dividend process modelling ideas to the local volatility framework, we have two major issues to address:-

1. *Local* versus *Global* method of *"calibration"* of local volatility.

   More about this later, time permitting.

2. *In finite differencing*:

   Node-to-node transfer probability assignment across dividend events.

   We need this for the backward induction of (deflated) net present values.

We first address 2. We discuss two methods how this can be done.

In all approaches, we keep the x-coordinates of the lattice nodes constant between ex-dividend times. These relate to strikes in spot space via

$$K_i(T) \;=\; F(T) \cdot \mathrm{e}^{x_i(T)} \,. \tag{6.1}$$

Approach #1: one-to-one node transitions

Intuitively appealing, we connect via

$$\begin{aligned} K_i(T_+) &= f\left(K_i(T_-)\right) \\ x_i(T_+) &= \xi\left(x_i(T_-)\right) \end{aligned} \tag{6.2}$$

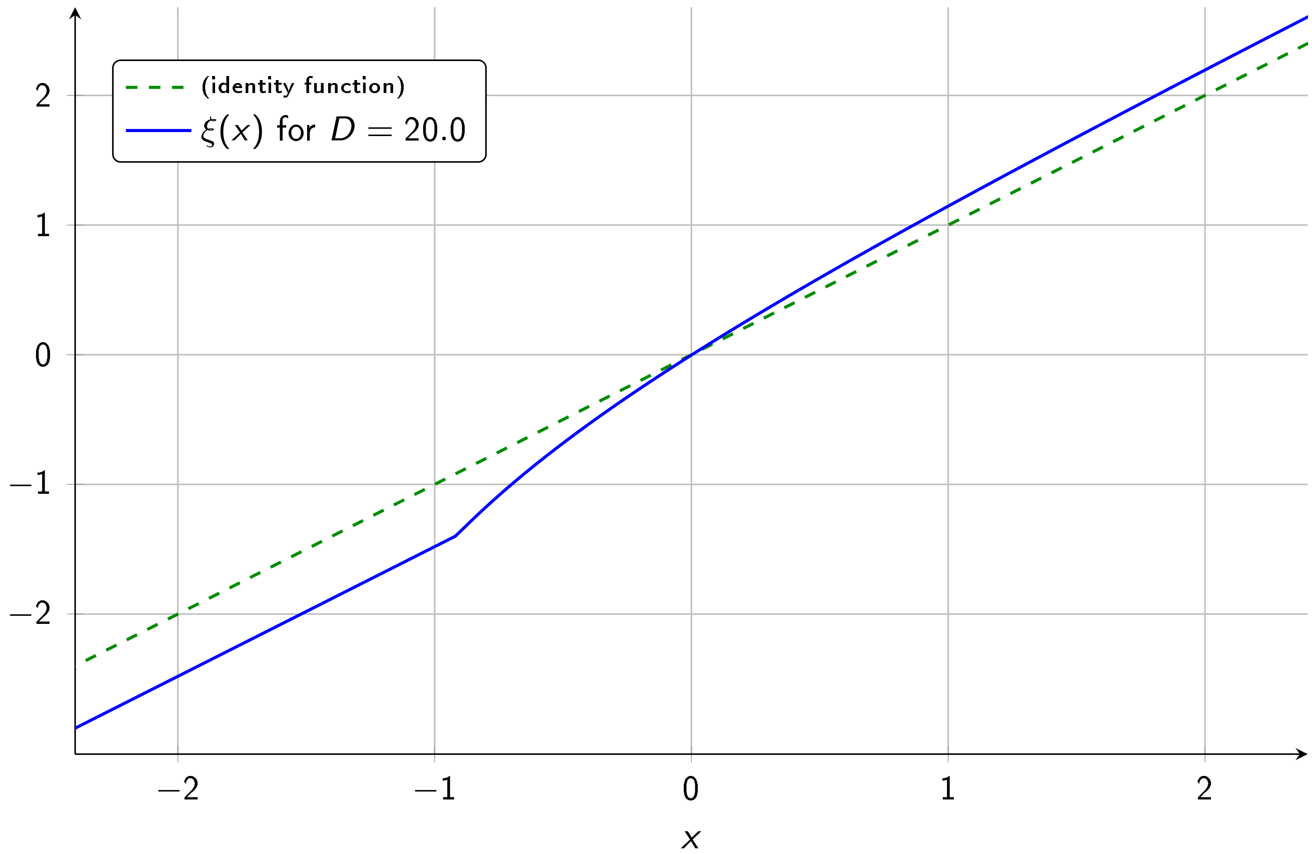
with

$$\xi(x) := \ln\left(\frac{f(\,F_- \cdot \mathrm{e}^x\,)}{F_+}\right) \tag{6.3}$$

where we dropped the dependence of $f(\cdot)$ on $D^*(T_-)$ and $\theta(T_-)$.

For any (single-boxed) finite differencing calculation from 0 to some time horizon T, we can choose the lattice geometry $\boldsymbol{x}(T^*)$ at one time horizon T^*, typically in the interval $(0, T]$, and the geometry in all other inter-dividend intervals is determined by a chained iteration of equation (6.2). Since $f(\cdot)$ is guaranteed to be monotonic and invertible, we can construct the lattices in both directions from T^*.

$T = 1$, $\hat{\sigma}(\theta) = 50\%$, $F_- = 100$:



Across a dividend transition, going forward in time:-

- Proportional dividends have no impact on a logarithmic lattice layout.
- Cash dividends *widen* a logarithmic lattice *in both directions*.

Assuming the forward is met exactly to T_- by the probabilities p_i (which don't change in the transition) on the nodes situated at T_- at the locations $x_i(T_-)$ by

$$F_- = \sum_i p_i \cdot K_i(T_-) \qquad \text{with} \qquad K_i(T_-) = F_- \cdot \mathrm{e}^{x_i(T_-)}\,, \tag{6.4}$$

we preserve the relationship $D = (1-\delta)\cdot F_- - F_+$ in (4.10) when

$$\begin{aligned} F_+ &= \sum_i p_i \cdot K_i(T_+) \\ &= \sum_i p_i \cdot f(K_i(T_-)) \end{aligned} \tag{6.5}$$

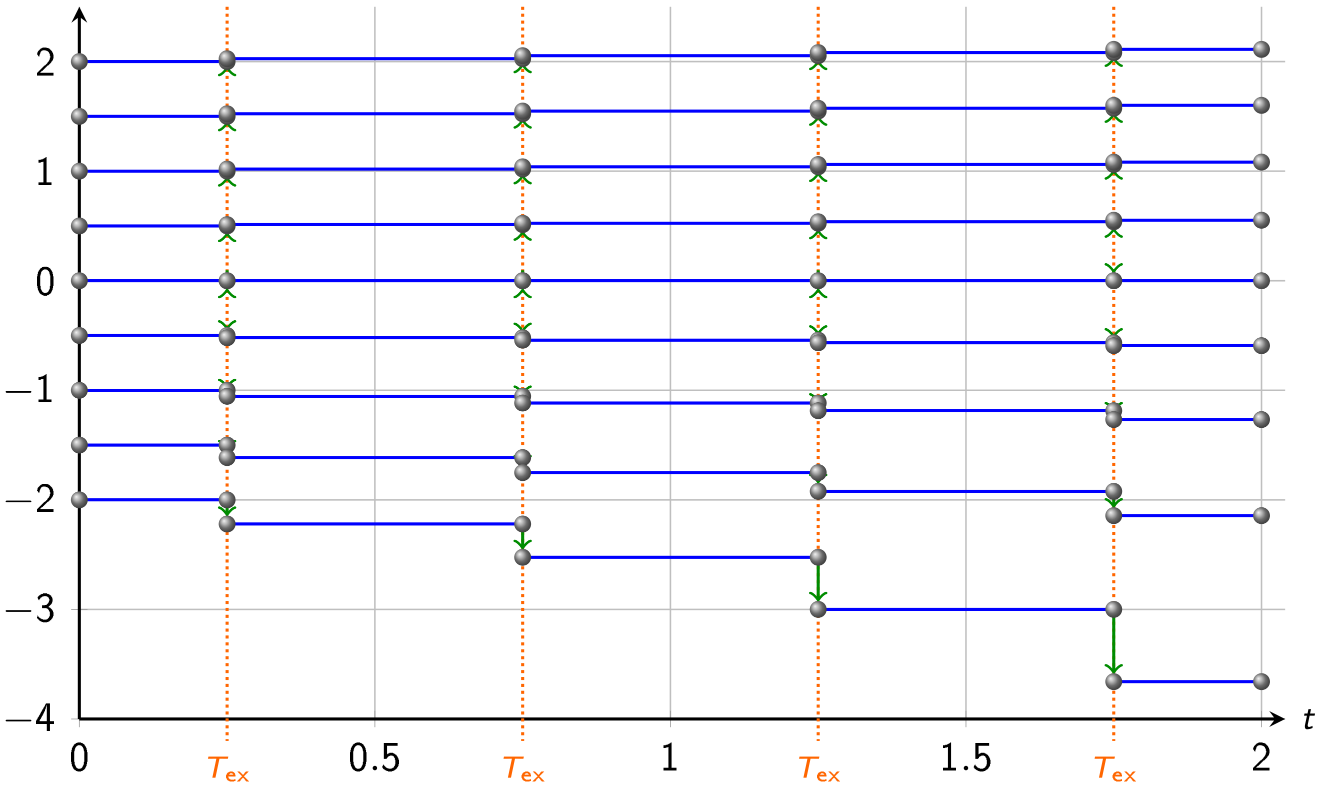
Expanding $f(\cdot)$ via (4.11), we obtain

$$\begin{aligned} F_+ &= \sum_i p_i \cdot \left((1-\delta)\cdot K_i(T_-) - D^* \cdot \left[1 - \left(1 - \frac{K_i(T_-)}{\theta} \right)_+ \right] \right) \\ F_+ &= (1-\delta)\cdot F_- - D^* \cdot \sum_i p_i \cdot \left[1 - \left(1 - \frac{K_i(T_-)}{\theta} \right)_+ \right] \end{aligned} \tag{6.6}$$

and thus the *analytical calibration of D^* on the lattice*:

$$\boxed{D^* = \frac{D}{\sum_i p_i \cdot \min\left(1, \frac{K_i(T_-)}{\theta}\right)}} \tag{6.7}$$

$D = 3$, $\hat{\sigma}(\theta) = 50\%$, $S_0 = 100$:



Logarithmic lattice evolution with approach #1: *one-to-one node transitions*.

Approach #2: Probability re-distribution

We lay out the lattice to propagate from T_+ to $T_+ + \Delta T$

any way we see fit

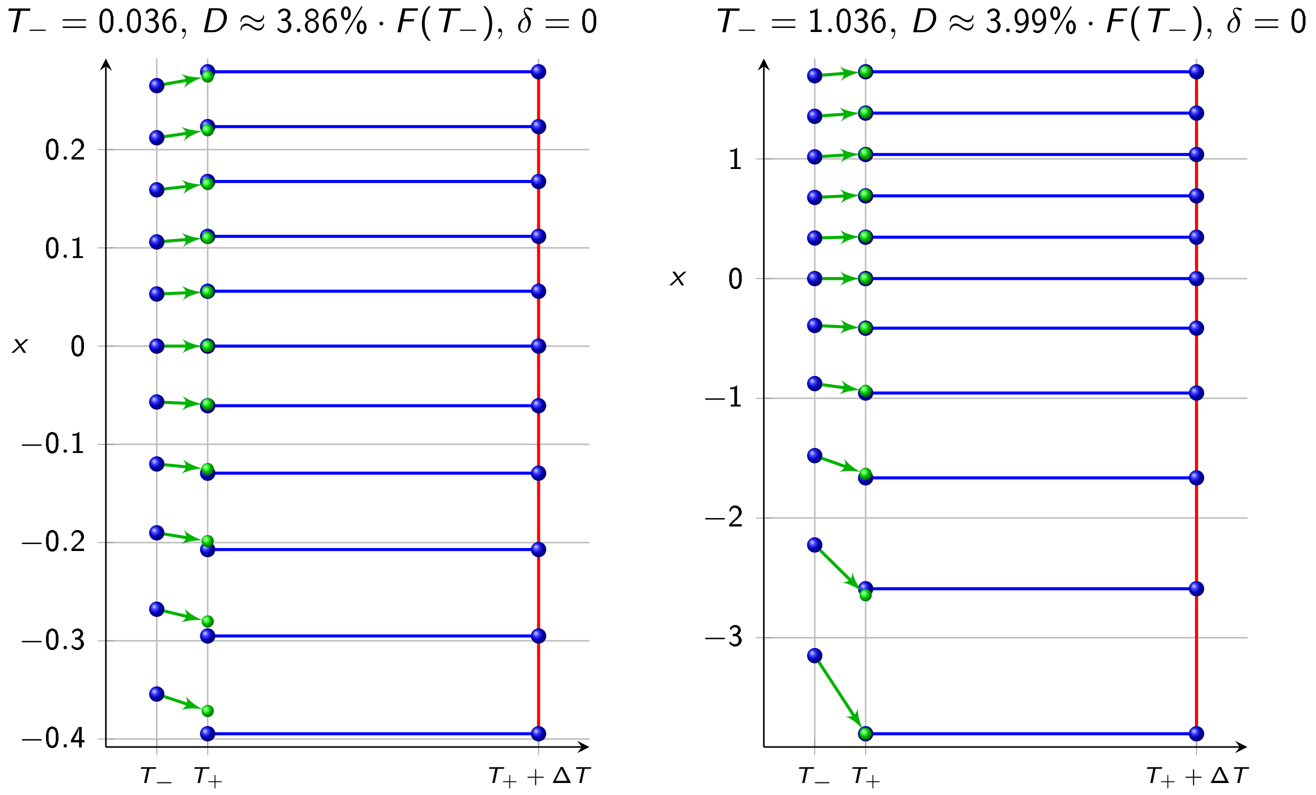
for the target horizon $T_+ + \Delta T$, e.g., asymmetrically widening according to the respective skew and smile.

As usual (laying out a lattice governed by the later time horizons), since we assume the lattice in logarithmic coordinates to be constant when there are no cash dividends, we set

$$x_i(T_+) \;:=\; x_i(T_+ + \Delta T)\,. \tag{6.8}$$

This means the lattice is not in line with the end-of-dividend-jump locations $\check{x}_i(T_+)$ we define as

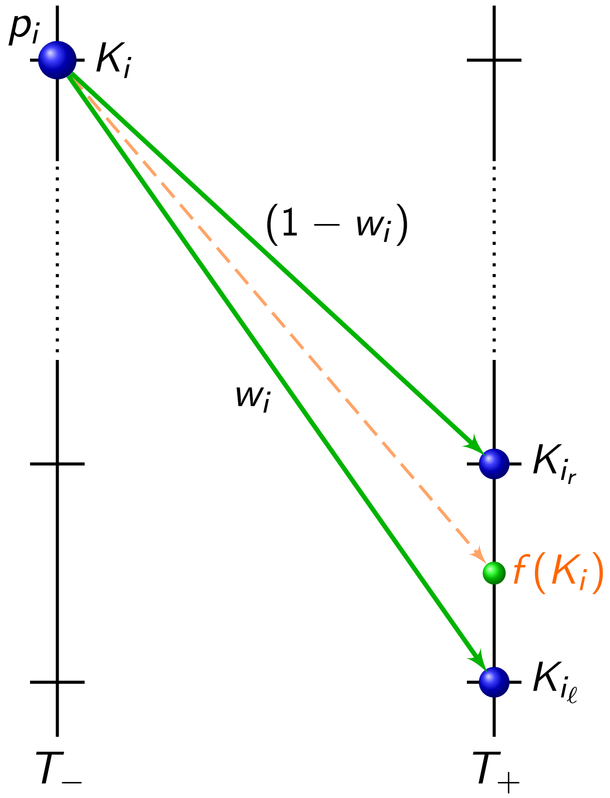
$$\check{x}_i(T_+) \;:=\; \xi\left(x_i\left(T_-\right)\right)\,. \tag{6.9}$$



Schematic of the jump from $\boldsymbol{x}(T_-)$ to the unoccupied locations $\check{\boldsymbol{x}}(T_+)$. Green arrows: dividend jump. Blue lines: independently chosen lattice node levels from T_+ to $T_+ + \Delta T$.

- Instead of jumping from $x_i(T_-)$ to $\check{x}_i(T_+)$,
- we split the probabilities over the nearest lattice nodes i_ℓ to the left and i_r to the right of $x_i(T_+ + \Delta T)$
- with weights w_i and $(1 - w_i)$
- such that the probability mass p_i arrives, *in spot coordinates*,
- at the unattainable location $f(K_i(T_-))$

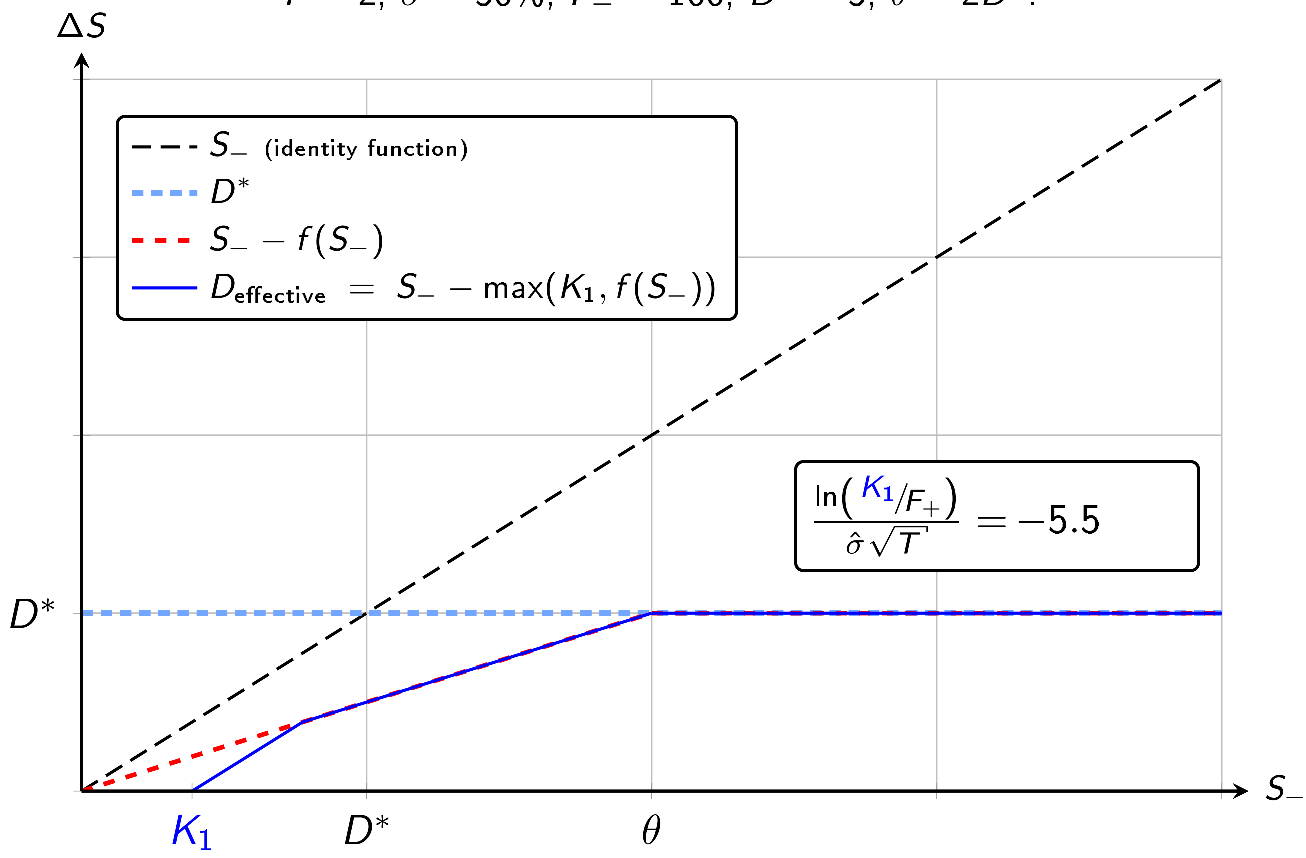
in expectation.



Schematic of probability split during the ex-dividend jump on a spot-fixed lattice layout.

- We do not necessarily have to widen the lattice at the lower end —
- we can simply absorb jumps beyond the bottom node!
- This is merely a small modification to our dividend process on the lattice.
- The cash dividend is cut down more rapidly when $f(S_-) < K_1$ where K_1 is the lowest lattice node.
- Unless a lattice is laid out for very short maturities, K_1 is typically orders of magnitude smaller than the forward.

$T = 2,\ \hat{\sigma} = 50\%,\ F_- = 100,\ D^* = 5,\ \theta = 2D^*$:



This of course means that we must adjust how we compute D^* !

Putting this together:

$$K_i^*(D^*) := \max\Big(K_1(T_+), f(K_i(T_-))\Big) \tag{6.10}$$

$$i_\ell = \max\Big(\, j \leq n-1 \,\Big|\, K_j(T_+) \leq K_i^*(D^*) \,\Big) \tag{6.11}$$

$$i_r = i_\ell + 1 \tag{6.12}$$

$$w_i = \frac{K_{i_r}(T_+) - K_i^*(D^*)}{K_{i_r}(T_+) - K_{i_\ell}(T_+)} \,. \tag{6.13}$$

In order to compute D^*, we demand the preservation of the forward F_+ as the probability-weighted average over the lattice nodes:

$$F_+ = \sum_i p_i \cdot K_i^*(D^*) \tag{6.14}$$

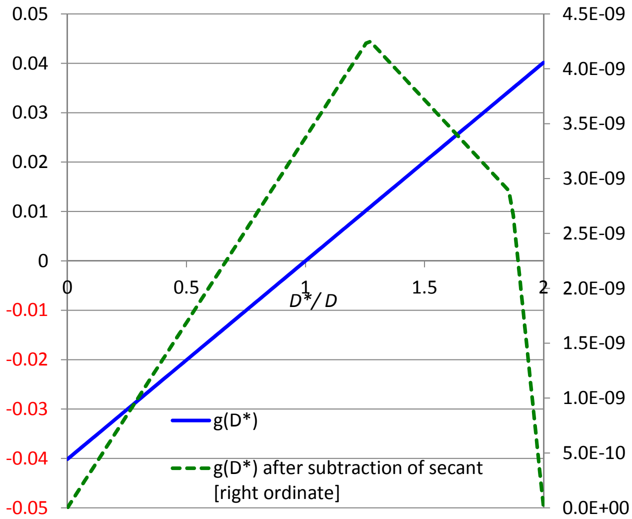
Note: the right hand side is continuous and piecewise affine in D^*.

D^* is the root of

$$g(D^*) \;:=\; 1 - \sum_i p_i \cdot K_i^*(D^*)/F_+ \; . \qquad (6.15)$$

Here are two examples:

$T_- = 0.036,\ D \approx 3.86\% \cdot F(T_-),\ \delta = 0$

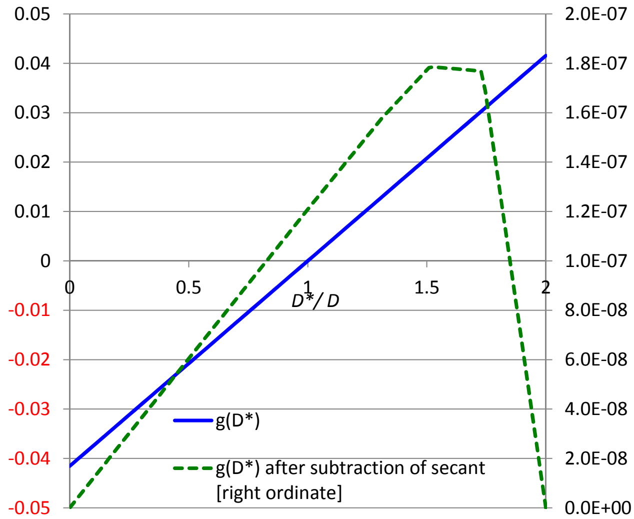


$\hat\sigma(F_-, T_-) \approx 26\%,\ \hat\sigma(\theta, T_-) \approx 39\%$

31 nodes, ∼5.35 ATF std. deviations

$^{D^*}/_{D} - 1 \;\equiv\; 0$

$T_- = 1.036,\ D \approx 3.99\% \cdot F(T_-),\ \delta = 0$

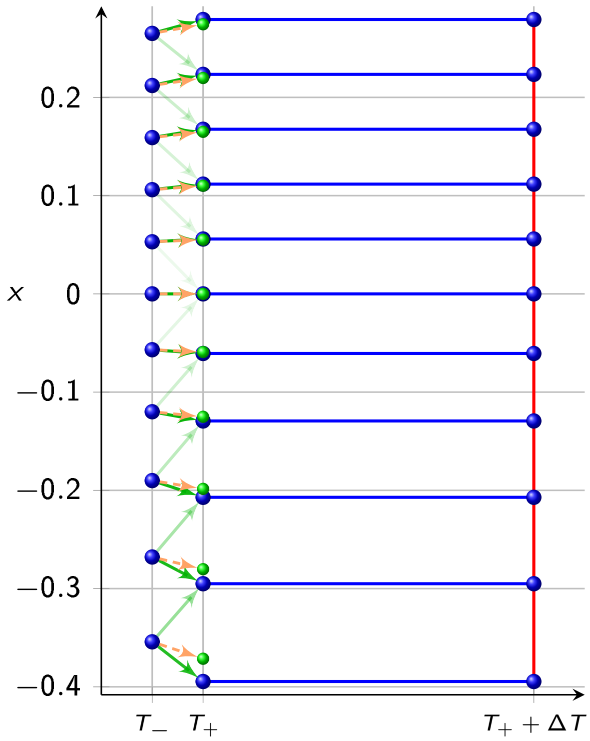


$\hat\sigma(F_-, T_-) \approx 29\%,\ \hat\sigma(\theta, T_-) \approx 55\%$

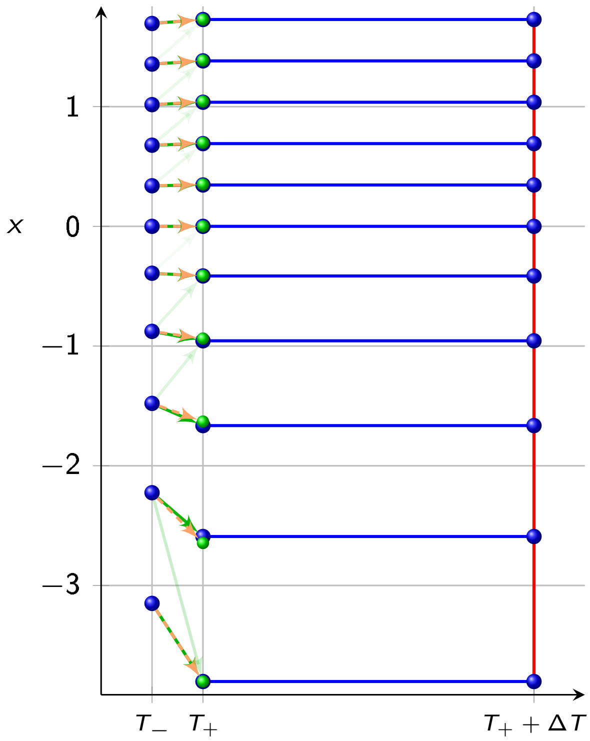
31 nodes, ∼5.35 ATF std. deviations

$^{D^*}/_{D} - 1 \;\approx 1.2\text{E-}6$

$T_- = 0.036,\ D \approx 3.86\% \cdot F(T_-),\ \delta = 0$



$T_- = 1.036,\ D \approx 3.99\% \cdot F(T_-),\ \delta = 0$



Schematic of the jump-and-split from $\boldsymbol{x}(T_-)$ to $\boldsymbol{x}(T_+)$. Orange dashed arrow: dividend jump *in expectation*. Green arrows: split jump. Blue lines: lattice levels from T_+ to $T_+ + \Delta T$.

Remark 6.1. The probability splitting explained here in a forward induction setting is *exactly equivalent to linear interpolation* of function values with weights w_i and $(1 - w_i)$ in backward induction calculations.

Remark 6.2. The splitting of probabilities has an impact on the net local variance that is required by the subsequent diffusion step to match the target probabilities at $T_+ + \Delta T$, though this effect will only apply to the time step immediately following the ex-dividend event.

In other words, **the local variance immediately following an ex-dividend event is affected by the spatially discretised dividend model we choose**.

This brings us to the choice of

Local versus *Global* *"calibration"* of local volatility

(time permitting).

Calibrating local variance coefficients

Define the normalized spot process $X(t)$

$$X(t) \ := \ S(t)/F(t) \tag{7.1}$$

where $S(t)$ is the observable spot at time t and $F(t)$ is the forward as seen at time $t_0 = 0$ for expiry t. $\boldsymbol{X(t)}$ **is a martingale**.

In the absence of cash dividends, the local volatility model is then given by the stochastic differential equation

$$\mathrm{d}X(t) = \sigma(X(t), t)\mathrm{d}W \ . \tag{7.2}$$

The time-continuous and space-continuous infinitesimal generator of this process has the form

$$\mathcal{L}(X, t) \ = \ \tfrac{1}{2}v(X, t)\partial_X^2 \tag{7.3}$$

with the **local variance $\boldsymbol{v(X, t)}$** obviously defined as $v(X, t) := \sigma^2(X, t)$.

Still in continuous time,

$$L \quad := \quad \text{a matrix comprising a spatial discretisation of } \mathcal{L}$$

that

- preserves positivity of probabilities
- ensures that the process X is a martingale under L.

It follows that:-

- the top and bottom row of L contain only zeros since the lowest and highest nodes must be absorbing for L to represent a martingale in X.
- All rows of L must sum up to zero.
- L must have only non-negative elements away from the diagonal (as is the case for all spatially discrete Markov generators).
- In fact, with $\hat{\boldsymbol{X}}$ being the vector of lattice node levels, we must have

$$L \cdot \hat{\boldsymbol{X}} = \mathbf{0} \tag{7.4}$$

for L to be a local martingale in X in all lattice node levels.

For the sake of efficient spanning of spot space, we now change spatial coordinates to

$$x := \ln(X) \, . \tag{7.5}$$

In spatially continuous calculations, this transformation introduces a drift term of magnitude $-\frac{1}{2}v$ in order to maintain the martingale condition, i.e., we obtain

$$\mathcal{L}(x,t) = \tfrac{1}{2}v\left(\partial_x^2 - \partial_x\right) \, . \tag{7.6}$$

However, ***in discrete space***,

$$L(x,t) \neq \tfrac{1}{2}v\left(\hat{\partial}_x^2 - \hat{\partial}_x\right) \tag{7.7}$$

where $\hat{\partial}_x$ symbolically represents suitably chosen finite differencing stencils.

We mention that conditions for X to be an exact martingale, not just in the limit of $\Delta x \to 0$, were already given in [ABR97, section 3.3 "Fitting of Asset Forward"]!

Conveniently, irrespective of the chosen finite differencing stencils, we can always put L into the form

$$L = \tfrac{1}{2} \cdot V \cdot M \tag{7.8}$$

where V is a diagonal matrix of the local variances $\{v_1(t), v_2(t), \ldots, v_n(t)\}$ with n being the number of spatial nodes.

M is a matrix of coefficients that are determined by:-

- the spatial discretisation
- the requirement that each row comprises a finite differencing stencil for the second derivative in x, which we denote as $\hat{\partial}_x^2$
- X (not x) is a martingale under L and thus under M, i.e.,

$$M \cdot \hat{\mathbf{X}} = \mathbf{0} \,. \tag{7.9}$$

We meet (7.9) by adding advection using a finite differencing stencil $\hat{\partial}_x$.

This is the *exact* application of Itō's lemma to spatially discrete Markov chains.

When the lattice is homogeneous in x with

$$\Delta := x_i - x_{i-1} \,, \tag{7.10}$$

we can set

$$M = \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ a & b & c & 0 & & & \cdot \\ 0 & a & b & c & 0 & & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot \\ \cdot & & 0 & a & b & c & 0 \\ \cdot & & & 0 & a & b & c \\ 0 & \cdot & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix} , \tag{7.11}$$

with

$$a = \frac{2}{\Delta^2} \cdot \frac{1}{1+\mathrm{e}^{-\Delta}}, \qquad b = -\frac{2}{\Delta^2}, \qquad c = \frac{2}{\Delta^2} \cdot \frac{1}{1+\mathrm{e}^{\Delta}} \tag{7.12}$$

so that

$$a+b+c = 0 \qquad \text{and} \qquad a \cdot \mathrm{e}^{-\Delta} + b + c \cdot \mathrm{e}^{\Delta} = 0 \,. \tag{7.13}$$

- Row #i of M evaluates ∂_x^2 at location x_i accurate to order $o(\Delta x^2)$.
- X is a martingale under M since

$$M \cdot \hat{\boldsymbol{X}} = 0 \,.$$

- M is a **Metzler** matrix.
- $L = \frac{1}{2} \cdot V \cdot M$ is stable and positivity preserving in the spatially discrete forward Kolmogorov equation

$$\partial_t \boldsymbol{p} = L^* \boldsymbol{p} \tag{7.14}$$

in continuous time exactly when

all local variance coefficients are non-negative.

- All of the above is **exact**, even when the lattice is **ultra-sparse** !
- **No need to have many nodes *only because* we have local volatility.**

The only known (two time) scheme that is guaranteed to preserve positivity of probabilities for any $\Delta t \geq 0$ is the first order fully implicit scheme

$$\boldsymbol{p}(t + \Delta t) = (1 - \Delta t \cdot L^*)^{-1} \cdot \boldsymbol{p}(t) \,. \tag{7.15}$$

Since L is (in exact arithmetic) a perfect local martingale, this scheme[3] gives us an *exact preservation of the X-martingale condition in discrete time*.

Transferring the fully implicit operation to the left hand side,

$$(1 - \Delta t \cdot L^*) \cdot \boldsymbol{p}(t + \Delta t) = \boldsymbol{p}(t) \tag{7.16}$$

and using $L = \frac{1}{2} \cdot V \cdot M$, we can turn this into

$$M^\top \cdot V \cdot \boldsymbol{p}(t + \Delta t) = \frac{2}{\Delta t} \cdot [\boldsymbol{p}(t + \Delta t) - \boldsymbol{p}(t)] \tag{7.17}$$

[3] like all Padé schemes

If we already have the discrete Arrow-Debreu probabilities $\boldsymbol{p}(t)$, then we can now back out the local variance coefficients $\boldsymbol{v}$!

Since $V \equiv \mathrm{diag}(\boldsymbol{v})$, we can rewrite

$$V \cdot \boldsymbol{p}(t+\Delta t) \equiv P \cdot \boldsymbol{v} \tag{7.18}$$

with

$$P := \mathrm{diag}(\boldsymbol{p}(t+\Delta t)) \,. \tag{7.19}$$

Further denoting

$$\Delta\boldsymbol{p} := \boldsymbol{p}(t+\Delta t) - \boldsymbol{p}(t) \tag{7.20}$$

we now have

$$\boxed{M^\top \cdot P \cdot \boldsymbol{v} = \frac{2}{\Delta t} \cdot \Delta\boldsymbol{p}} \tag{7.21}$$

which is a linear system for the local variance vector $\boldsymbol{v}$.

Due to the structure of M, this linear system is twofold over-determined in the coefficients $v_2, \ldots, v_{n-1}$ and provides no equations for v_1 and v_n.

This is the result of the fact that on the first and on the last node we have absorption and thus v_1 and v_n are undefined by construction.

The system has a solution exactly if $p_i(t+\Delta t) \neq 0$ and

$$\sum_i p_i(t) = \sum_i p_i(t+\Delta t) \qquad \text{[probability preservation]} \tag{7.22}$$

$$\sum_i \hat{X}_i \cdot p_i(t) = \sum_i \hat{X}_i \cdot p_i(t+\Delta t) \qquad \text{[forward preservation]} \tag{7.23}$$

both of which we had guaranteed from the start.

We therefore drop the top and bottom equations and indicate this by a tilde.

With this in mind, we define

$$\tilde{\boldsymbol{y}} := \widetilde{M}^{\top\,-1} \cdot \frac{2}{\Delta t} \cdot \Delta\tilde{\boldsymbol{p}} \tag{7.24}$$

which always exists due to the well-conditioned special structure of $\tilde{M}$.

We now only have to solve
$$\tilde{P} \cdot \tilde{\boldsymbol{v}} = \tilde{\boldsymbol{y}} \,. \tag{7.25}$$

The matrix $\tilde{P}$ is diagonal, so this should be trivial, right?

Alas,

$\tilde{P}$ can (*and sometimes will*) have zero entries on the diagonal!

What's more, we can (and sometimes will) end up with **negative** v_i.

We can resolve the underdetermined system with Lagrange multipliers and an objective of minimum variation of neighbouring elements of $\boldsymbol{v}$.

We also need the constraint $v_i \geq 0$.

In practice, this is equivalent to the following very simple logic:

1. Solve for all coefficients v_i for which $\tilde{P}_{ii} \neq 0$.
2. Linearly interpolate between the so obtained v_i (over their location in logarithmic coordinates) to resolve all v_j for which $\tilde{P}_{jj} = 0$.
3. Set
$$\begin{aligned} v_1 &:= v_2 \\ v_n &:= v_{n-1} \end{aligned} \tag{7.26}$$
4. Floor all v_i at zero.

Now that we have computed the local variance coefficients, we satisfy the forward Kolmogorov equation

$$\boldsymbol{p}(t+\Delta t) \;=\; \left(1-\Delta t\cdot L^{*}\right)^{-1}\cdot\boldsymbol{p}(t)\;. \tag{7.15}$$

Unless we had to floor local variances!!!

In that case, vanilla options at $t+\Delta t$ no longer meet the input volatility surface, and as a consequence, if we simply proceed as above, not at $t+2\Delta t$ either, and so on!

The error will gradually (=slowly) average out.

Any additional later flooring will just compound the problem.

Note that this problem is implicitly part of any “local volatility” calculation that uses any form of (local) analytical formulae.

We can, however, **compensate to repair the damage** done by the floored local variances (to some extent)!

Once we have the local variances for the time step from t to $t+\Delta t$, and so populate $L(t,t+\Delta t)$, we numerically compute

$$\hat{\boldsymbol{p}}(t+\Delta t) \;:=\; \left[1-\Delta t\cdot L^{*}(t,t+\Delta t)\right]^{-1}\cdot\boldsymbol{p}(t)\;. \tag{7.27}$$

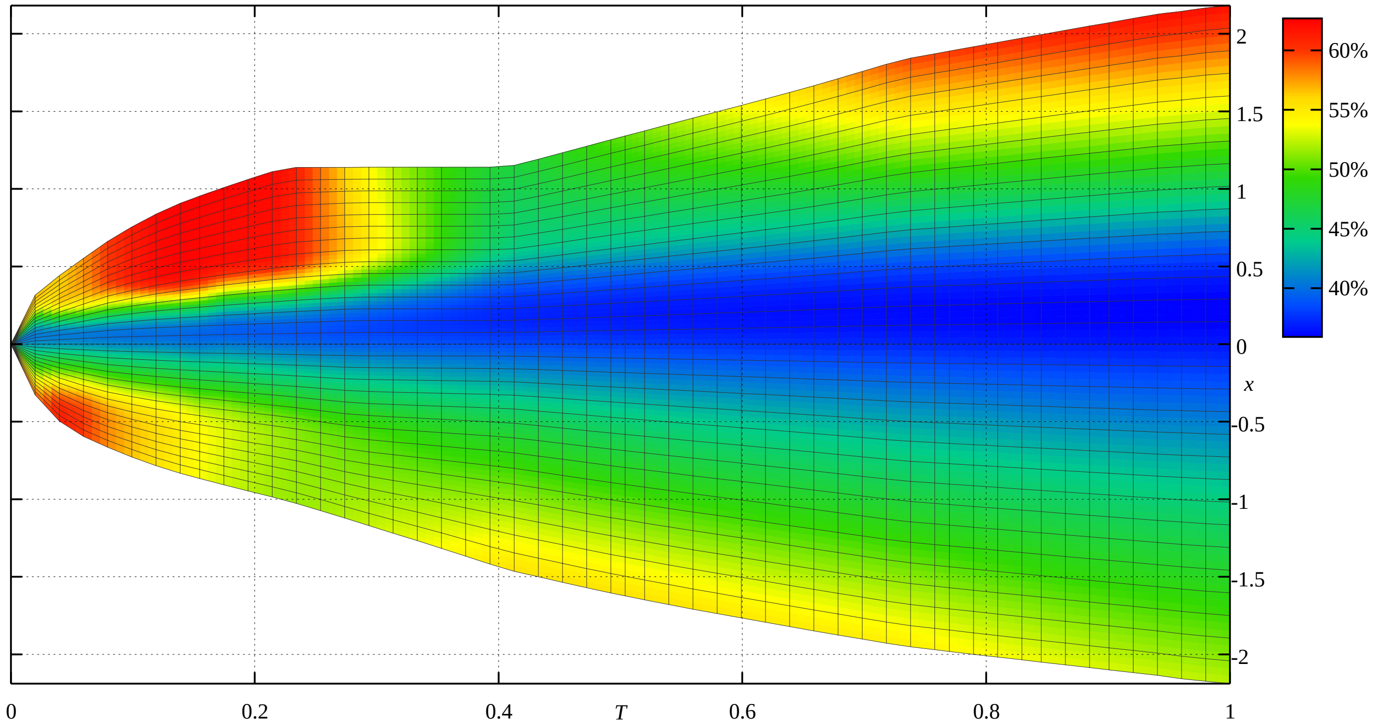
We then use

$$\hat{\boldsymbol{p}}(t+\Delta t) \qquad \text{instead of} \qquad \boldsymbol{p}(t+\Delta t)$$

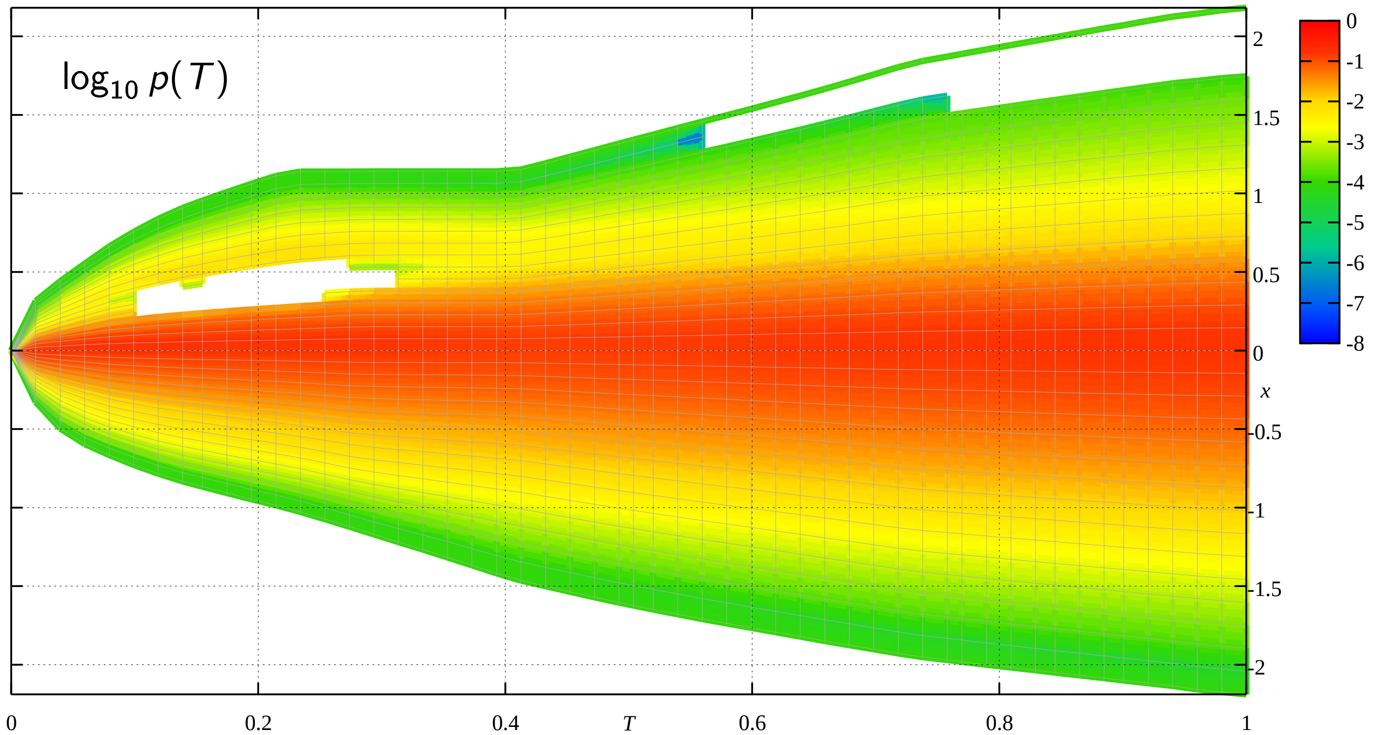
for the calculation of $v(t+\Delta t,t+2\Delta t)$.

Wherever local variances from $t+\Delta t$ to $t+2\Delta t$ would have been positive without compensation, we find that those immediately following floored local variances are depressed if compensation is invoked.

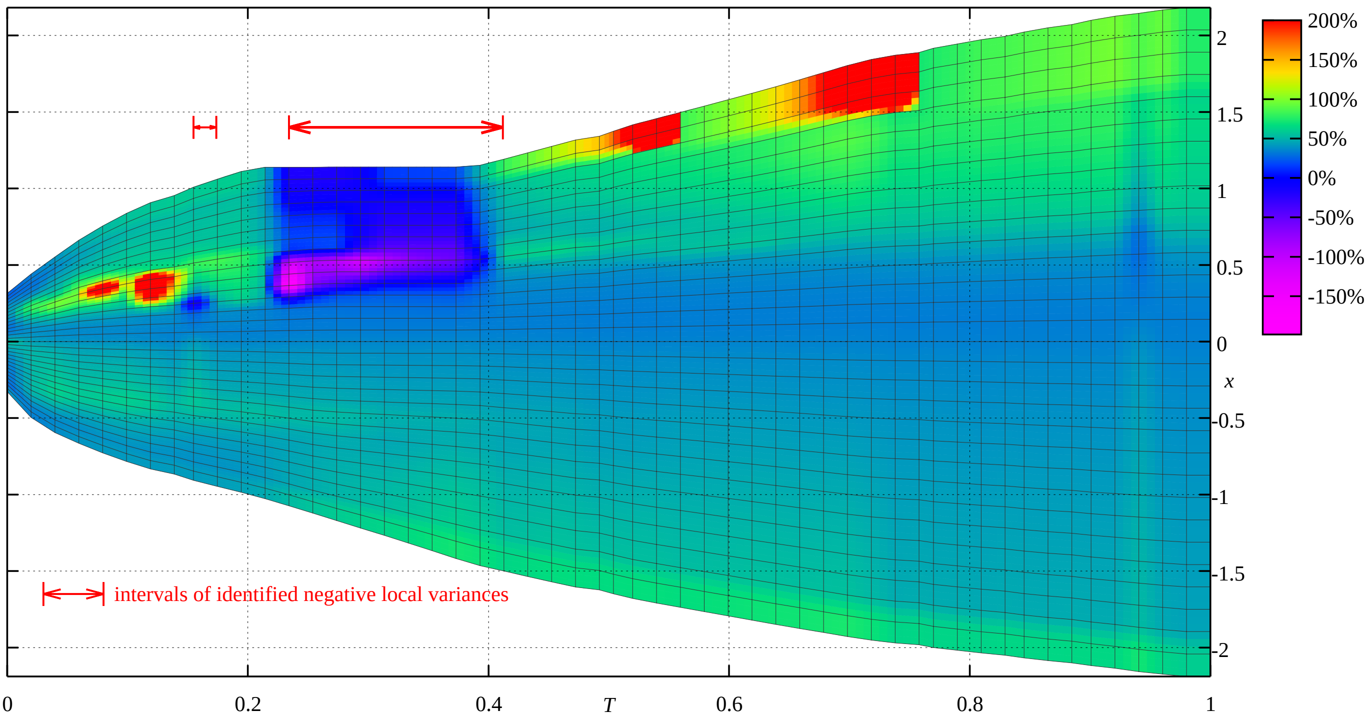
This is intuitively right: *flooring* is akin to *raising* the local variance, so to compensate, subsequent local variance needs to be *reduced*.



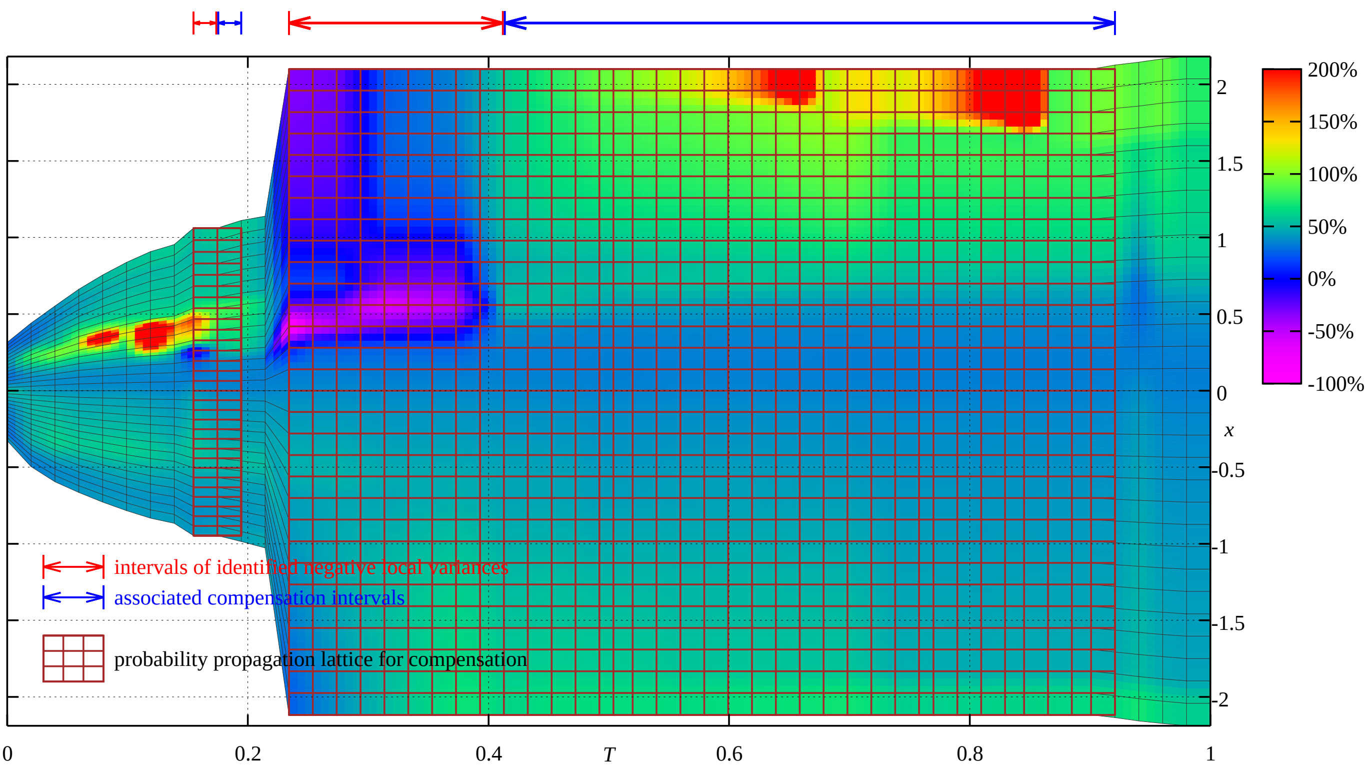
Implied volatility surface as a heat map. We used 31 nodes and an asymmetric lattice layout to cater for smile and skew.



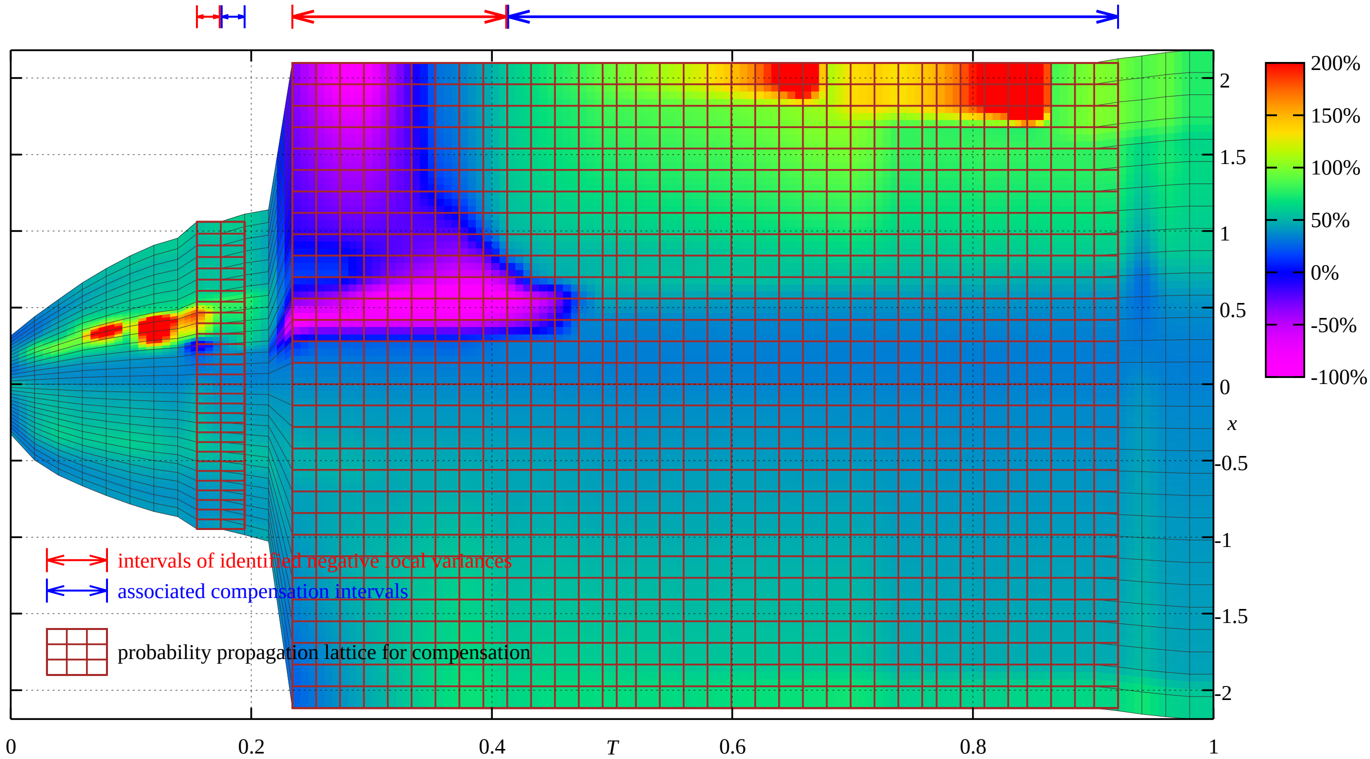
Discrete probabilities bootstrapped from the data in the previous figure on a logarithmic scale. Blank regions correspond to zero-floored probabilities.



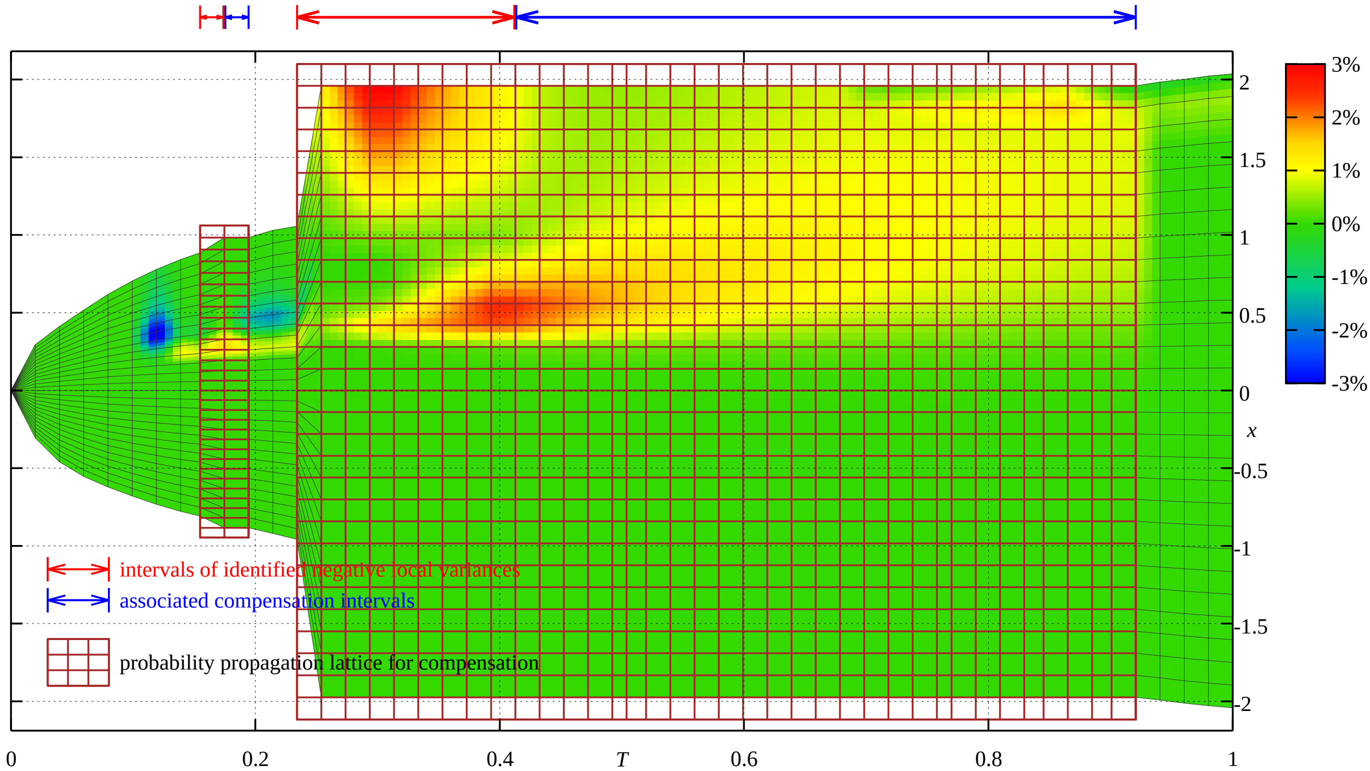
'Signed' local volatility defined as $\mathrm{sign}(v) \cdot \sqrt{|v|}$ (where v is the unfloored local variance) resulting from the implied volatility surface (unfloored data to enhance visualisation).



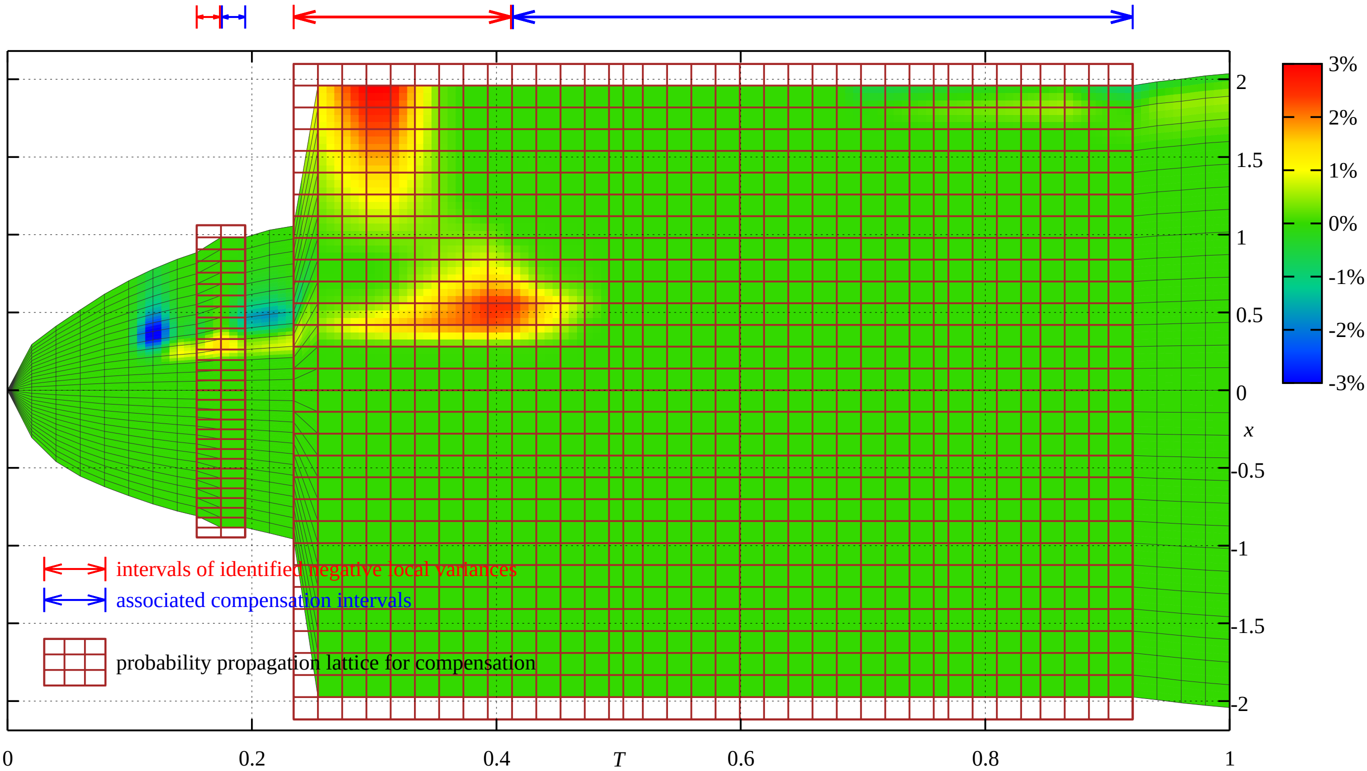
'Signed' local volatility without compensation (unfloored data to enhance visualisation).



'Signed' local volatility with compensation (unfloored data to enhance visualisation).



Residual in implied volatility without compensation when re-pricing vanilla options from (floored) local variances.



Residual in implied volatility with compensation when re-pricing vanilla options from (floored) local variances.

In a similar vein, we ought to be careful how we compute the original target probabilities $\boldsymbol{p}(T)$ for the time stepping horizons.

We recognize two commonly used methods.

First, we define the normalized vanilla option prices as follows:

$$\begin{aligned}\widetilde{\text{Call}}_i &= \text{Black}_{\text{call}}\Big(\text{Forward}=1, \text{strike}=\tilde{K}_i,\, T, \hat{\sigma}=\hat{\sigma}_{\textbf{surface}}(T, \text{strike}=F(T)\cdot\tilde{K}_i)\Big)\\ \widetilde{\text{Put}}_i &= \text{Black}_{\text{put}}\Big(\text{Forward}=1, \text{strike}=\tilde{K}_i,\, T, \hat{\sigma}=\hat{\sigma}_{\textbf{surface}}(T, \text{strike}=F(T)\cdot\tilde{K}_i)\Big)\end{aligned} \tag{7.28}$$

with

$$\tilde{K}_i \;:=\; \mathrm{e}^{x_i}\ . \tag{7.29}$$

1. **Asymmetric Butterflies** (also known as *Arrow-Debreu securities*)

   For all i except the lowest and the highest, we set the probability p_i according to the asymmetric butterfly price

$$p_i = \left(\frac{\omega_{(i-1)} \cdot (\tilde{K}_{i+1} - \tilde{K}_i) - \omega_{(i)} \cdot (\tilde{K}_{i+1} - \tilde{K}_{i-1}) + \omega_{(i+1)} \cdot (\tilde{K}_i - \tilde{K}_{i-1})}{(\tilde{K}_{i+1} - \tilde{K}_i) \cdot (\tilde{K}_i - \tilde{K}_{i-1})} \right)_+ \tag{7.30}$$

   where we use $\omega_{(\cdot)}$ for the precomputed normalized call prices when $x_i \geq 0$ and put prices otherwise.

   At the top and bottom, we set

$$p_1 = \left(\frac{\widetilde{\mathrm{Put}}_2}{\tilde{K}_2 - \tilde{K}_1} \right)_+ \qquad p_n = \left(\frac{\widetilde{\mathrm{Call}}_{n-1}}{\tilde{K}_n - \tilde{K}_{n-1}} \right)_+ . \tag{7.31}$$

Crucially,

all probabilities are individually floored at zero without any interaction.

As a consequence, any occurrences of floored negative probabilities are likely to lead to a violation of probability preservation

$$\sum_i p_i = 1 , \tag{7.32}$$

and a violation of the forward condition

$$\sum_i p_i \tilde{K}_i = 1 . \tag{7.33}$$

Apart from the wing nodes, **this method is equivalent to the calculation of "local volatility" from (local) analytical formulae, particularly if derivatives are replaced with finite differences of prices**.

3 Telescopic bootstrapping

The top and bottom probabilities are also set according to (7.31). Then, in sequence towards the centre,

$$p_i = \begin{cases} \left(\dfrac{\widetilde{\text{Call}}_{i-1} - \left(\sum_{j=i+1}^{n} p_j \tilde{K}_j - \tilde{K}_{i-1} \sum_{j=i+1}^{n} p_j \right)}{\tilde{K}_i - \tilde{K}_{i-1}} \right)_+ & \text{if } x_i \geq 0 \\ \left(\dfrac{\widetilde{\text{Put}}_{i+1} - \left(\tilde{K}_{i+1} \sum_{j=1}^{i-1} p_j - \sum_{j=1}^{i-1} p_j \tilde{K}_j \right)}{\tilde{K}_{i+1} - \tilde{K}_i} \right)_+ & \text{else} \end{cases}, \tag{7.34}$$

for all

$$i \neq k := \frac{n-1}{2}$$

(assuming that n is always odd).

The central probability is finally given by

$$p_k = 1 - \sum_{j \neq k} p_j \,. \tag{7.35}$$

In this method, the calculation of any p_i takes into account the effect of flooring any of its outer predecessors in the sequence!

The occurrence of the flooring of a negative probability at some location $i > k+1$ (or $i < k-1$) does by itself not lead to a violation of the forward preservation condition (7.33) as long as $p_{k+1} \geq 0$ before any flooring (respectively $p_{k-1} \geq 0$).

Probability is always preserved!

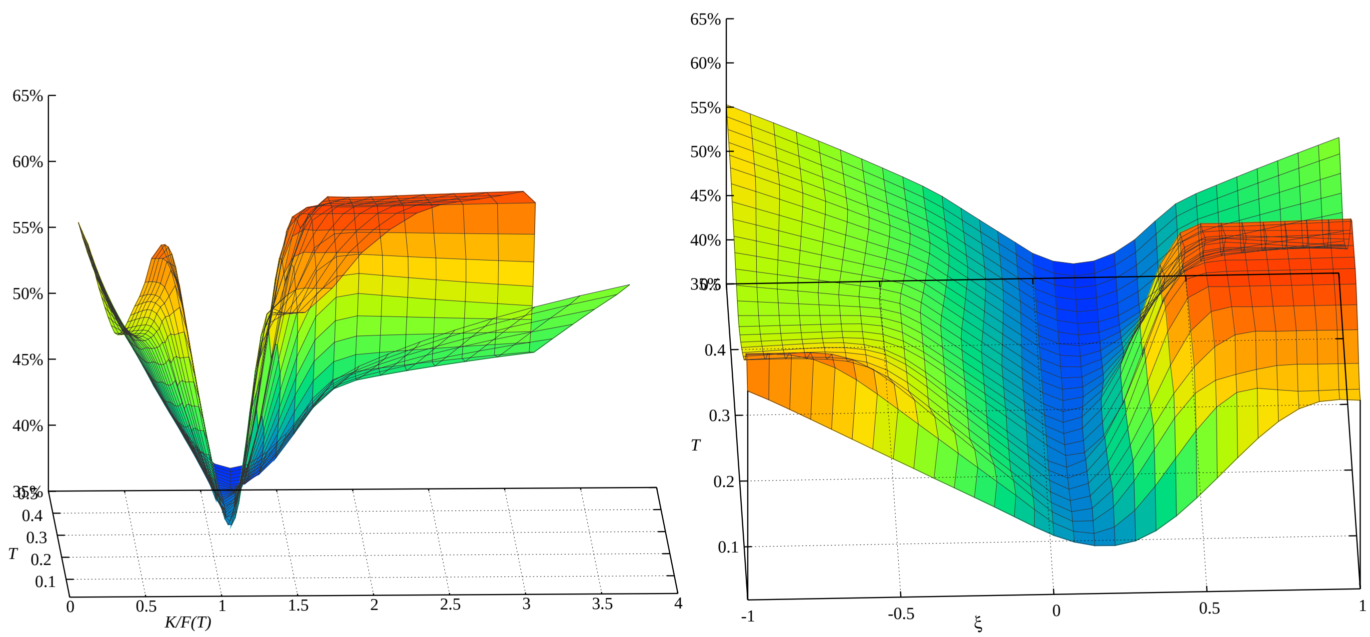
Remark 7.1. The flooring of probabilities in the telescopic bootstrapping method is akin to a modification of the input volatility smile such that it is free of negative-butterfly-arbitrage and such that all data on the outside of the locations of the floored zero probabilities are preserved.

As long as $p_{k-1} \geq 0$ and $p_{k+1} \geq 0$ (where k is the central node), the forward is also preserved.

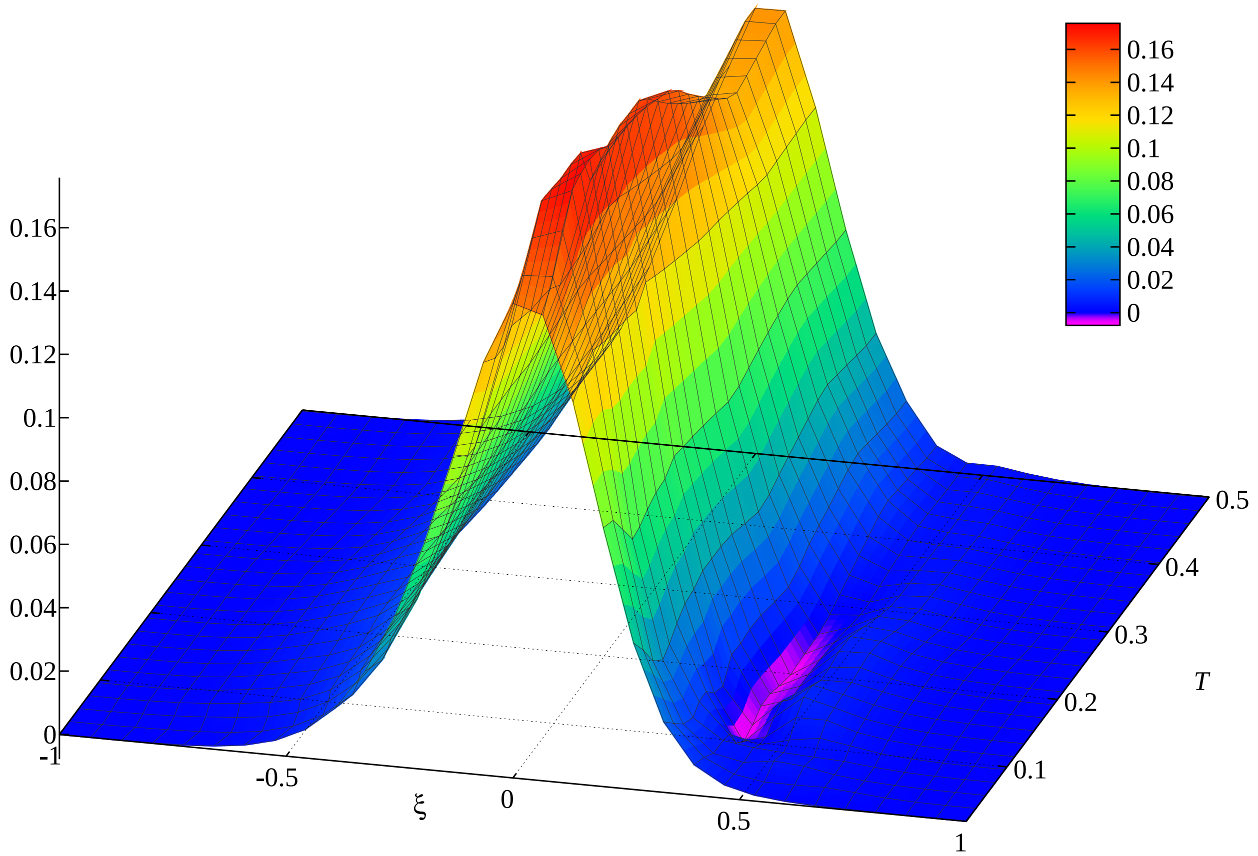
The conditions $p_{k-1} \geq 0$ and $p_{k+1} \geq 0$ are in practice always preserved since their violation would indicate negative butterflies near the money which is something the market just doesn't do.

We now show some examples. In the following diagrams, we use the normalized log-spot ξ given by

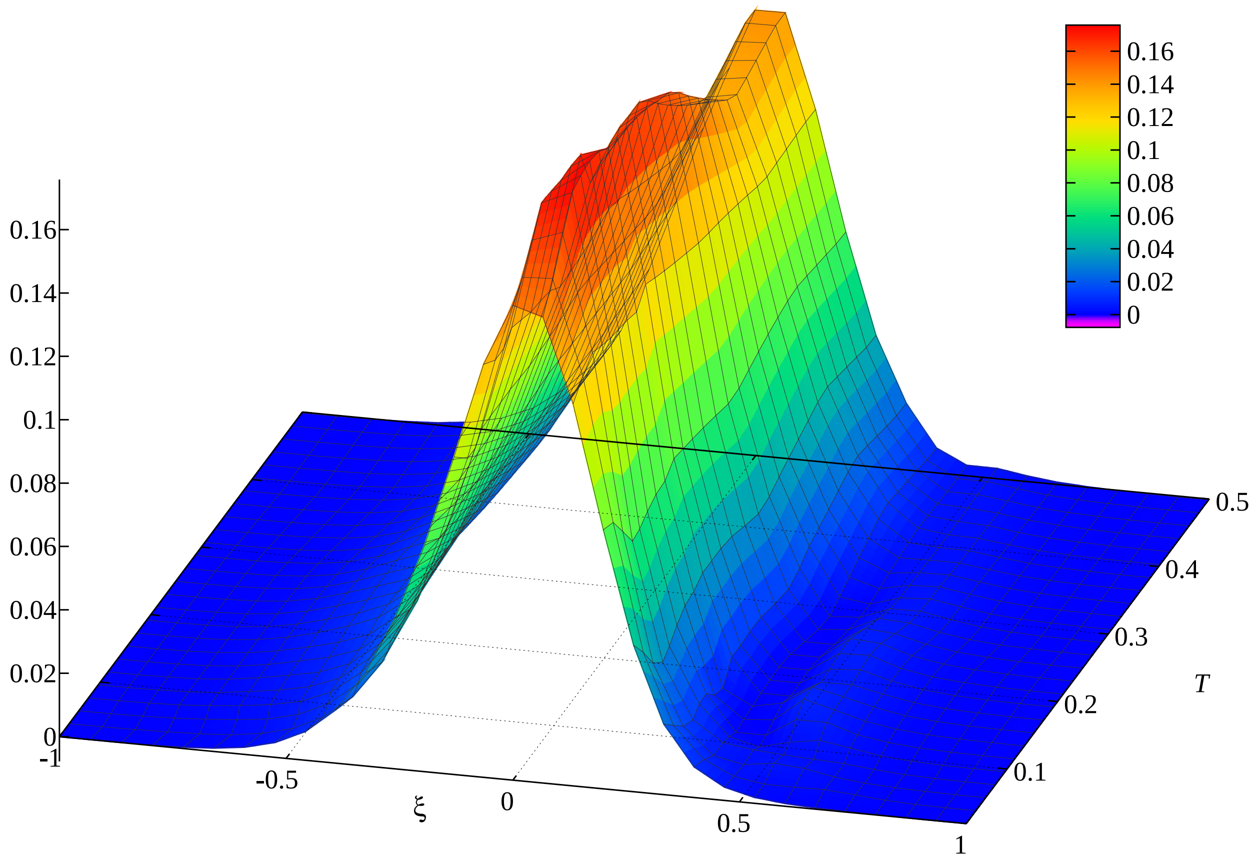
$$\xi(t) := x \Bigg/ \begin{cases} |x_{\min}(t)| & \text{if } x < 0 \\ |x_{\max}(t)| & \text{else} \end{cases} . \tag{7.36}$$



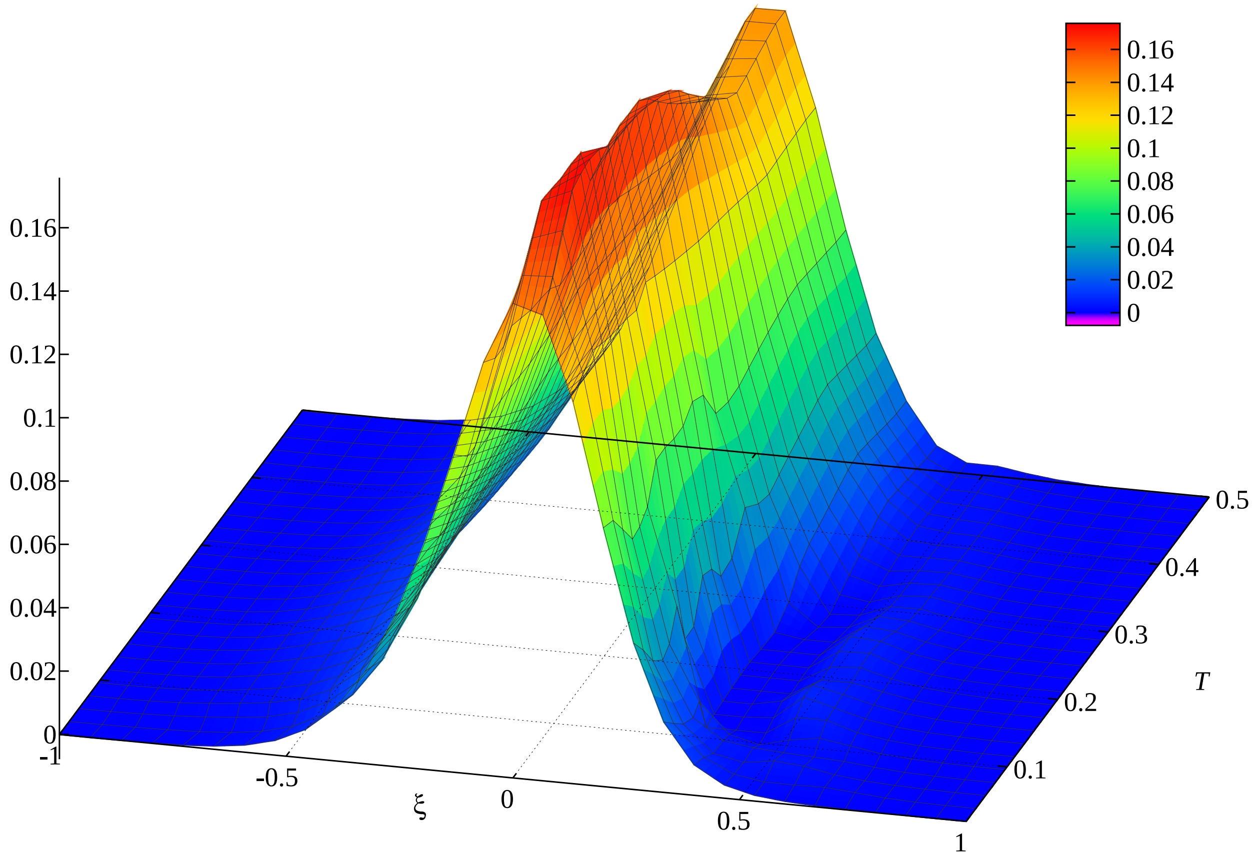
Implied volatilities with arbitrage. To the left over normalized strike, to the right over normalized log-strike ξ as defined in (7.36).



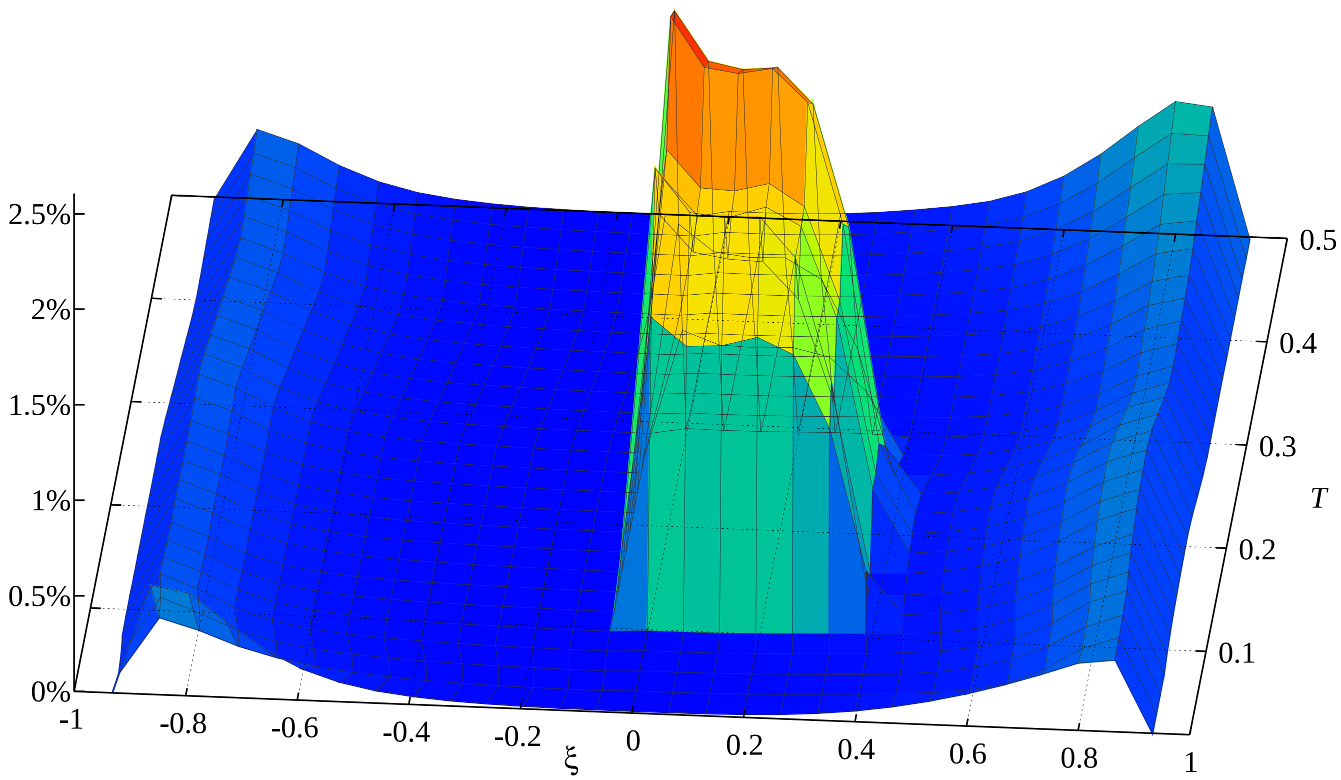
Probabilities from asymmetric butterflies before flooring.



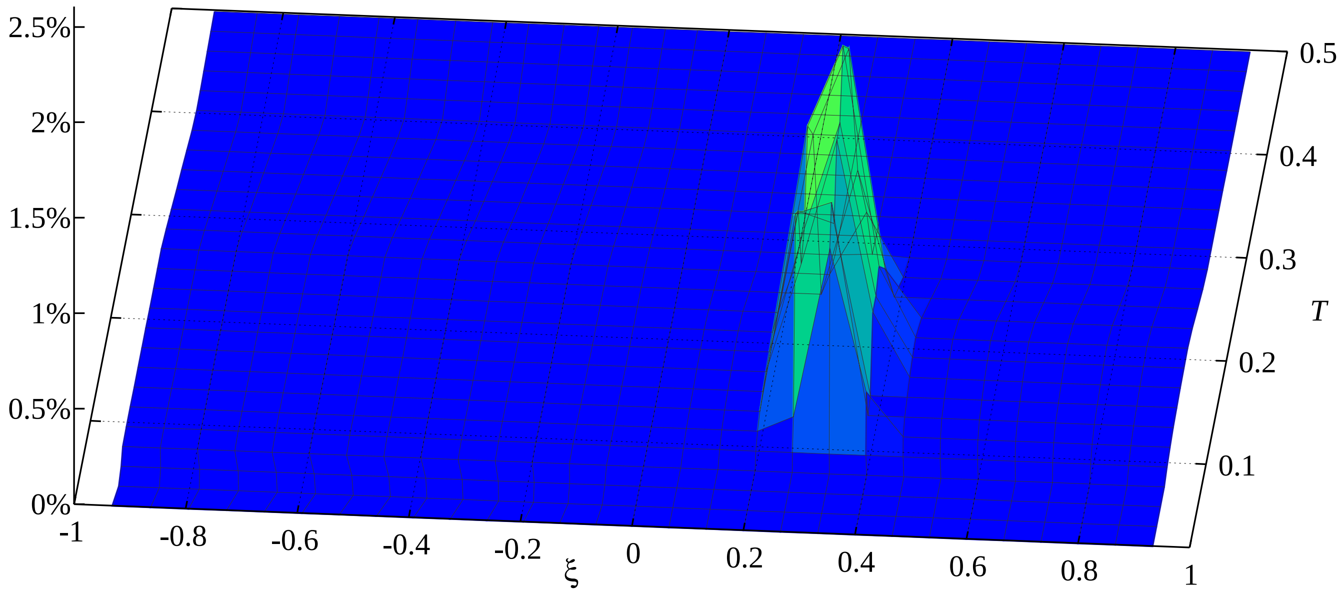
Probabilities from asymmetric butterflies after flooring.



Probabilities from telescopic bootstrapping.



Absolute difference of implied volatilities computed from the probabilities backed out with the asymmetric butterfly method with flooring.



Absolute difference of implied volatilities computed from the probabilities backed out with the telescopic bootstrapping procedure.

- We reviewed commonly used dividend models.
- We designed a dividend process around economic and legal reality.
- We showed that this requires only the slightest adjustment to the assumed cash lump sum dividend for most of the distribution.
- We computed the adjusted cash dividend D^* analytically.
- We showed how this can be incorporated into a parametric implied volatility surface generation model without arbitrage or approximations.

- We showed how this can be incorporated into a local volatility framework, again without arbitrage or approximations[4].
- We showed how we can calibrate local variance coefficients even in the presence of cash dividends, without approximation or arbitrage.
- The solution is ***not*** yet another analytical approximation for a local volatility formula that is adjusted for dividends.

  Instead, we simply solve the linear equations exactly.

  This is efficient, accurate, fast, and safe.
- A key consideration for any local variance logic is that calibration must always take a global view, both in the time and in the strike direction.

  This applies not only to the case when cash dividends are present.

[4] We focussed on finite differencing but this can also be done for Monte Carlo.

[ABR97] L. Andersen and R. Brotherton-Ratcliffe.
The equity option volatility smile: An implicit finite-difference approach.
Journal of Computational Finance, 1:5–38, 1997.

[BGS03] R. Bos, A. Gairat, and A. Shepeleva.
Dealing with discrete dividends.
Risk, January 2003.
www.risk.net/data/Pay_per_view/risk/technical/2003/0103_options.pdf.

[Bla75] F. Black.
Fact and Fantasy in the Use of Options.
Financial Analysts Journal, 31(4):36–41,61–72, Jul.–Aug. 1975.

[BV02] M. Bos and S. Vandermark.
Finessing fixed dividends.
Risk, September 2002.
www.risk.net/data/Pay_per_view/risk/technical/2002/0902_bos.pdf.

[Fri02] V. Frishling.
A discrete question.
Risk, January 2002.
www.risk.net/data/Pay_per_view/risk/technical/2002/0102_frishling.pdf.

[gov06] The UK government.
UK Companies Act, 2006.
www.legislation.gov.uk/ukpga/2006/46/pdfs/ukpga_20060046_en.pdf.

[HJ88] D. Heath and R. Jarrow.
Ex-dividend stock price behaviour and arbitrage opportunities.
Journal of Business, pages 95–108, 1988.
forum.johnson.cornell.edu/faculty/jarrow/020%20ExDividend%20Stock%20JB%201988.pdf.

[Jäc14] P. Jäckel.
Clamping down on arbitrage.
Wilmott Magazine, pages 54–69, May 2014.
www.jaeckel.org/ClampingDownOnArbitrage.pdf.

[JAL14] P. Jäckel and M. Abellan-Lopez.
Ultra-Sparse Finite-Differencing For Arbitrage-Free Volatility Surfaces From Your Favourite Stochastic Volatility Model.
In *WBS Fixed Income Conference, Barcelona*, October 2014.
www.jaeckel.org/DHIPresentation-Handout.pdf.

[MR97] M. Musiela and M. Rutkowski.
Martingale methods in financial modelling.
Springer, 1997.