Pricing Vanilla Options with Cash Dividends

Timothy R. Klassen, PFT Analytics

Abstract

The pricing of vanilla options on underliers with cash dividends is a surprisingly contentious and active research subject, for both European or American exercise style. Neither on the listed options side (calls and puts) nor on the flow/structured side of longer-term vanillas or light exotics are market participants in agreement on what model to use, nor on what an efficient practical implementation of the chosen model would be. The modeling problem boils down to the question of what a proper generalization of the Black-Scholes model to the case of cash dividends is, i.e. what should replace simple geometric Brownian motion (GBM).

We discuss this question with the aim of taking a first step towards a rationalization and normalization of the equity volatility market. We compare the two main classes of models in use, namely the "spot model" (piecewise GBM) and several "hybrid models" (shifted GBM). We are interested in consistency, simplicity, speed, and generality (covering all traded vanilla options, dividend and borrow rate assumptions, as well as easy modeling of business time, events, term-structure, credit, light exotics, etc). We also discuss the calibration problems that market participants face in some detail.

We show that: (i) all hybrid models are closely related on a mathematical level – despite qualitatively different financial properties – with simple and accurate relationships between calibrated parameters (borrow costs and volatilities) for both European and American options with cash dividends; (ii) all hybrid models allow accurate and very fast pricing of vanilla options using fine-tuned tree methods; (iii) some hybrid models have essentially all the desired properties outlined above; in particular, we describe a hybrid model closely related to the spot model, motivated by the spot-strike adjustment idea of Bos and Vandermark.

First Version: July 21, 2015

1 Introduction

It might surprise many, even professionals in the financial industry, that more than 40 years after the publication of the Black-Scholes-Merton model, there is no consensus on what its generalization to the case of underliers with cash dividends should be. The question is, basically, what should replace geometric Brownian motion (GBM) when the underlier pays discrete cash dividends?

The reason there is still no consensus on this issue is that there is essentially a "no-go" theorem saying that such a model can not have all the properties one might put on a naive wish list.

Either we can think of the traded stock as being the sum of a "pure" stock that follows GBM, and a risk-less piece that is the present value (PV) of *all* future cash dividends.[1] This model does not seem to have a commonly accepted name; we will refer to as the *(full) hybrid model*. It has been most forcefully advocated in [1, 2]. In this model we have simple analytic formulas for forwards and European option prices, strictly non-arbitrageable prices, fast pricing of any American vanilla option (as we will see), and an easy way to allow the pricing and risk-management of light exotics

[1] Where, however, we should only count (the part of the) cash dividends we are reasonably sure about.

in the same framework. *But*, in this model the pricing of an option of any maturity will depend on all cash dividends, even those after maturity.

Or, if we do not want pricing to depend on dividends after maturity, we *either* have to give up on GBM (leading to what we will refer to as the "spot model") with various concomitant complications and speed issues, *or* we have to accept that there is no consistent stochastic process across maturities and hence option prices might be arbitrageable in certain situations (if we don't give volatilities a suitable term-structure and skew, with jumps across dividend dates).

In the spot model the stock follows GBM between ex-dividend dates, with jumps at the dividend dates (which is why it is sometimes known as the "piecewise log-normal model"). As the stock can, in principle, drop below the next dividend, one has to specify what happens in this case. The spot model is sometimes assumed, explicitly or implicitly, to be *the* correct model to use, and the only reason one might not, is because of its computational inefficiencies. One can argue with this, as we will discuss.

In any case, the non-existence of an obviously perfect model has lead to a plethora of approximations and "recipes" for the pricing of options with cash dividends; see e.g. [3, 4, 5, 6, 7, 8, 9] and references therein. Some of these approaches are restricted to European options, i.e. they do not specify a stochastic process for the underlier, and/or only cover the case of one dividend (or are computationally untenable for more than one or two dividends), and/or only work for calls but not puts, and/or assume the borrow cost is zero or non-positive.

Such approaches are not useful in practice (except perhaps for testing). They only serve to litter the menus of vendor systems and confuse the innocent. Even if we can't have it all, we would like to have "as much as possible". Our wish list for an acceptable model is that it should be general, simple, fast, not exhibit arbitrage in practice, and have some sensible financial motivation. By general we mean that it should cover, say, all listed stock, ETF, futures, and index options trading in the US, most of which are American exercise style (except for index options).[2] This includes many liquid options with multiple dividends: the DIA ETF, for example, pays monthly dividends; six month SPY options might involve three dividends. There are also liquid options on many underliers with large borrow rates: the IWM ETF has had an implied borrow cost of about 140 bps (with flat term-structure) for many years; many liquid (leveraged) ETFs have even higher borrow costs; in times of crises the borrow costs of any number of names can shoot up and stay high for a long time, as witnessed during the financial crisis; names with recent IPOs will also have large borrows.

Generality also means that it should be easy to model business time and events, add default risk, and, ideally, integrate at least light exotics into the valuation and risk infra-structure.

The wish for simplicity and speed hardly needs to be emphasized. The main long-term problem for any sufficiently large (software) system is the management of complexity. If the core pricing infra-structure would use, say, different approximations and approaches in different situations, one is bound to run into nasty surprises from time to time. The need for speed is also obvious in this age of high-frequency trading. There are more than 500,000 options trading in the US alone. Everyone wants risk and scenario reports to be available in real-time. Robust calibration results, e.g. for fitted volatility surfaces should also be fast, so that one can quickly and confidently react to news, a dividend amount or ex-date change, or an earnings date move, for example.

[2] If we can handle all these cases we should also be able to handle other listed options traded around the world, with perhaps simple modifications to take into account local idiosynchracies, as e.g. the tax treatment of NIFTY options in India.

Our contribution here is to point out that the (full) hybrid model is part of a wider class of hybrid models that share many desirable, practical features: (i) They have exact formulas for European options, and the same, simple formula for forwards. (ii) When one of them can fit the market, so can any of the others, with slight caveats to be discussed. (iii) They can all be implemented together via very fast and accurate tuned tree methods. (iv) Implied borrows and volatilities in the various hybrid models are very accurately related via formulas inherited from the European case, even for American options with very large early-exercise premia, and even on small pricing trees. (v) We describe a hybrid model closely related to the spot model, that might represent the best practical compromise of all the features desired in a dividend pricing model, for a vanilla options market maker at least.

Another contribution is to clearly separate modeling, implementation, and calibration issues, which is often not the case in the existing literature. In our discussion, we will usually take the view of an options market maker (OMM), but our point of view should be easy to adapt to that of other market participants.

To put the rest of this paper in context we should briefly sketch some of the pricing related problems that (equity) vanilla options market participants will encounter. First of all, for listed vanilla options, most participants, certainly most market makers, use the Black-Scholes "language" to summarize the prices of all traded options in a more compact and intuitive manner in the form of an implied borrow curve and an implied volatility surface (note that market makers usually assume one borrow rate per term, i.e. that it does not depend on strike, unlike implied vols). This also provides a way of smoothing information from related options. For a "normal" market (i.e ignoring take-over and other special situations), if one just implies borrows and volatilities from the market with a sufficiently accurate pricer, the only way one can fail to match (a smoothed version of) the market, is for underliers with cash dividends, where in-the-money options might be mispriced.[3] Once the price of an option is matched with an implied borrow and vol, one assumes that the Black-Scholes-like model used for pricing and calibrating can be used to calculate greeks. The common belief, borne out in practice, is that with prudent risk management, i.e. hedging options as much as possible with closely related options, in addition to the underlier, this approach will not get one into too much trouble, despite the well-known issues of the Black-Scholes model.

All of the above is just necessary to keep one from losing money if one wants to be an active trader, though with good underlier valuation and speed it might be sufficient to make money off the bid-ask spread in classic market maker fashion. To make money in other ways one needs an edge. One way is to trade implied versus some forecast of realized volatility. Another is to have one's own view and forecast on borrow curves and volatility surfaces, related to any number of other factors (spot-vol correlations, behavior of related options and names, delta and vega flow signals, vol surface shape moves, etc). And one can of course use the options market to make leveraged bets on fundamental factors, from the underlier itself to events, dividends, etc.

In all of this a precise understanding of volatility is very important. Different dividend pricing models give markedly different values and even shapes, as we will see, for implied volatilities and

[3] Out-of-the money options are mainly determined by their implied volatility, so only a not sufficiently flexible vol curve could fail to match a reasonable market. At-the-money options should be matched by a combination of correctly implied vols and borrows. If one uses the correct dividend amount(s) and date(s), one can then only fail to match the in-the-money options if one uses a wrong or at least not "market-consensus" dividend model or business time and hence misprices relative early exercise premia. To be sure, only for names with a very liquid options market, like the SPY ETF, would one have a chance of disentangling whether one has a wrong cash dividend, borrow rate, dividend pricing model, or business time. There is also the question whether one should price with rate and/or volatility term-structure.

their term-structure and skews. If one wants to take a view, form hypotheses about relevant relationships based on historical data, etc, it is important to have a clear understanding of this issue. For the volatility market to become anywhere as efficient as statistical arbitrage has made the equity markets in the last decade, a rationalizing and normalizing of the meaning and calculation of implied volatility will be crucial.

In section 2 we give more background on the dividend modeling problem and the various approaches that have been tried. We also describe a hybrid model that combines the best properties of all the models, with little downside, at least for an OMM. In section 3 we summarize and expand on the general hybrid model framework, where the underlier is driven by shifted GBM, with various possible shifts related to cash dividends. We describe general properties and relationships between all hybrid models. In section 4 we sketch an accurate and fast implementation of general vanilla pricing for all the hybrid models. We also discuss several examples of calibrating the borrow rates and volatilities of different hybrid models to the same "market", showing in detail how they are related. Finally, we discuss calendar arbitrage in the hybrid models, including arbitrage due to discretization errors on small pricing grids. Section 5 provides our conclusions and outlook.

2 Dividend Modeling

2.1 Types of Dividends

As far as derivatives pricing is concerned, dividends come in three forms:

- A dividend yield, which is used to model the borrow cost of an underlier.[4]
- A cash dividend, which is how most dividends are actually paid. The dividend gets effectively paid before the open on the ex-dividend date. In markets where some large (usually institutional) players do not pay taxes on dividends, the expected move in the underlier should match the paid dividend (otherwise, a fixed fraction of the dividend).
- A discrete proportional dividend (aka discrete dividend yield), which pays a fraction of the underlier at a particular date.

Actual cash dividends are hardest to model and price with, and in the early days of derivatives modeling a dividend yield was often used instead. This is in general not very accurate. The next, and much better approximation is the replace a cash dividend by a discrete proportional dividend. Even this will nowadays not allow one to match all listed options on a dividend-paying underlier with a sufficiently liquid options market, at least not with one strike-independent implied borrow cost per maturity.[5] But it still provides a useful reference point for various comparisons.

The main use of proportional dividends is that they provide a better way to model dividends in the medium to distant future. Empirically it is very clear that with some lag dividends tend to go up and down with the stock price, as would be expected intuitively. Most OMMs and equity derivatives desks use some kind of *blending scheme* to smoothly transition from a cash dividend assumption on the short end, to a proportional dividend in the long term.

[4] In practice, when implied from the market, it is also partially used as a "fudge factor" to absorb uncertainties or disagreements about what discount rates to use as well as pricing model differences between market makers.

[5] Amusingly though, there was a period during the financial crisis when options on most underliers, and especially banks, where apparently mainly priced with proportional rather than cash dividends, due to extreme dividend uncertainty.

Once this has been decided upon, the modeling problem with cash dividends amounts, in the simplest situation, to the question what stochastic process should be used instead of GBM. Before we go through various answers one might to provide to this question, let us prepare by discussing what we know about the forward of the underlier with the three types of dividends discussed above.

2.2 The Forward

Let us first establish our notation. Relative to today, $t_0 = 0$, let us denote by $t_i > 0$ the ex-dividend dates on which the underlier S_t drops by some amount $\delta_i S_{t_i^-} + d_i$. Here the d_i are the cash dividends, the δ_i are the discrete proportional dividends, assumed to be paid first, and t^- denotes an infinitesimal instant before time t (at $t = t_i$ the dividend jump has already occurred). Between dividend dates the underlier has a deterministic drift $\mu_t = r_t - q_t$, where r_t is the discount rate and q_t the borrow cost. We will use standard continuously compounded and annualized rates and and volatilities.

As mentioned above, in practice, the simplest reasonable assumption about the mix of cash and proportional dividends is to assume a continuous and monotonic transition from pure cash dividends, $d_i > 0, \delta_i = 0$ for some period, to purely proportional dividends, $d_i = 0, \delta_i > 0$ after the transition period.[6] Our discussion below allows an arbitrary mix of all the dividend types.

The assumption of "affine" dividends is simple enough to allow for an analytic expression for forward prices $F_t = E[S_t]$.[7] Our assumptions imply that across a dividend date the forward behaves as

$$F_{t_i} \;=\; (1-\delta_i)\; F_{t_i^-} \;-\; d_i \;. \tag{1}$$

Between dividend dates the forward simply grows with the risk-neutral drift μ_t, so that we can also write

$$F_{t_i} \;=\; \exp\left(\int_{t_{i-1}}^{t_i} \mu_t\, dt\right)\; (1-\delta_i)\; F_{t_{i-1}} \;-\; d_i \;. \tag{2}$$

These recursions have the solution

$$F_T \;=\; f_p(T)\, \left(S_0 \;-\; \sum_{i:t_i \le T} \frac{d_i}{f_p(t_i)}\right) \tag{3}$$

where we introduced the *proportional growth factor*

$$f_p(t) \;:=\; \exp(\int_0^t \mu_t dt)\; \prod_{i:t_i \le t} (1-\delta_i) \tag{4}$$

which is the proper discount (growth) factor to use for the cash-dividends (stock).[8] Note that it is also the delta of the forward. It depends on the proportional dividend fractions δ_i, which suggests one way of empirically estimating the transition from cash to proportional dividends.

[6] For future reference, our recipe is as follows: a (B, E) blending scheme means that we use cash dividends out to B years, and purely discrete proportional dividends after E years, with a linear interpolation in between. The transformation of a cash dividend d into a proportional one is accomplished via $\delta = d/S_0$, or a linearly interpolated fraction thereof. Any change in forwards this might entail will in practice be absorbed into the implied borrow costs.

[7] The material to follow about the forward has previously been covered in e.g. [1, 2], and in the same notation as here in [10].

[8] We should mention that many trading firms' and vendors' quant libraries used to, and perhaps often still, have a "dividend discounting bug", where dividends are only discounted with the discount rate, rather than with the full proportional growth factor.

The only assumption in the above is that the stock can actually pay the cash part of the dividend. This means, in particular, that the stock should not be below the next cash dividend when it is due. Pondering this assumption leads to the heart of the modeling issues surrounding cash dividends.

2.3 Dividend Pricing Models

2.3.1 The Spot Model

In [7] it is argued that one should actually allow the stock to potentially drop below the upcoming cash dividend, in which case the stochastic process for the stock is only well-defined if one specifies a *dividend policy* as to what happens in this case. These arguments are in the context of the spot model (SM), where one assumes, in the simplest case, that the stock follows GBM with some fixed volatility between dividend dates, and jumps by an affine amount at each ex-dividend date.[9] Note that in the spot model the underlier does not follow GBM, but piecewise GBM. This means that even for European-style options there is in general no closed-form formula for vanilla options prices. It also means that there is no simple formula for the forward, in general (and the forward will depend on the chosen dividend policy).

In most practical situations the deviation of the exact forward from eq. (3) is small, especially if one has a blending scheme that cuts off the cash dividends relatively quickly. But if reasonably high precision is required, not having a simple, exact formula for the forward is a significant drawback in a number of problems (e.g. the vanilla market calibration problem for an OMM).

The arguments we have heard in favor of the spot model are: (i) It is arbitrage-free. (ii) Its implied volatility is comparable to standard realized volatility calculations. (iii) It is "intuitively" the right thing to do.

A reply to these arguments would be (i) that it is not the only arbitrage-free model, witness e.g. the full hybrid model to be discussed momentarily, (ii) that it is easy to relate the implied volatilities from other models to standard realized volatility calculations to a sufficient accuracy given the various uncertainties in the implied-to-realized volatility relationship (or one can adjust the calculation of the realized volatility). Finally, concerning (iii), it is true that the spot model has an intuitive and sensible motivation – but, again, so does the (full) hybrid model. It is not *a priori* clear which is "more sensible". As an example of a qualitative question on which the spot and hybrid models differ, consider the behavior of the (local) vol of a stock right after a dividend. With a smooth vol term-structure for the relevant object one is modeling, the local vol of the stock will jump up (perhaps temporarily, depending on the pure stock term-structure) in a hybrid model but not in the spot model. Given that the underlying corporate entity has become more leveraged, the former might be considered more reasonable.

Of course, there is some fuzziness in all these arguments, since we are not really trying to fundamentally model the underlying entities,[10] but just come up with an analog of the Black-Scholes model for the cash dividend case. Nevertheless, given the significant practical drawbacks of the spot model, it should have some serious conceptual and financial advantages over competing

[9] We take the name *spot model* to derive from the fact that it is the full, observed stock price, the "spot", that follows GBM between dividend dates in this model. This is in contrast to the hybrid models discussed below, where only part of observed stock price should be considered to be randomly fluctuating.

[10] In fact, the argument we just made for the behavior of the (local) vol of a stock right after a dividend would not apply e.g. for certain leveraged and other "exotic" ETFs, where no underlying entity becomes more leveraged due to a dividend or capital gains distribution (instead some swap might got unwound, or such). Understanding such underliers in detail can lead to options trading opportunities.

models before one should consider it the gold-standard of dividend modeling.

2.3.2 The Hybrid Model

Another point of view, argued in e.g. [1, 2], is that one should design the stock process to never allow the underlier to drop below some bound related to the cash dividends. If one is really worried about modeling the stock dropping below the (next) dividend, one should perhaps think about credit modeling, i.e. the possibility of default (clearly, the greeks coming out of the spot model do not somehow magically protect one from default risk; one needs credit derivatives for that). Even if one is not inclined to model default risk, one should perhaps simply use more proportional instead of cash dividends.

Thinking about designing such a process more explicitly, given the formula (3) for the forward, one can argue as follows [2].[11] As seen from time t, the forward for maturity T is given by

$$F_{t,T} \;=\; f_p(T) \left(\frac{S_t}{f_p(t)} \;-\; \sum_{i:t<t_i\leq T} \frac{d_i}{f_p(t_i)} \right) . \tag{5}$$

By no-arbitrage the forward can never go negative, which implies the lower bound

$$S_t \;\geq\; f_p(t) \sum_{i:t<t_i\leq T} \frac{d_i}{f_p(t_i)} \tag{6}$$

for any T, i.e. that $S_t \;\geq\; D_t$, where

$$D_t \;:=\; f_p(t) \sum_{i:t<t_i<\infty} \frac{d_i}{f_p(t_i)} . \tag{7}$$

In other words, the properly discounted PV of all future cash dividends provides a lower bound on the stock price. Clearly, for this formula to be sensible/plausible one should convert dividends further out more and more into proportional ones, i.e. use a blending scheme as discussed earlier.

This leads to *the hybrid model* defined by writing the stock process as

$$S_t \;=\; (F_t - D_t)X_t \;+\; D_t , \tag{8}$$

in terms of a positive martingale X_t with expected value $E[X_t] = 1$ for all t. We will refer to X_t, and often $(F_t - D_t)X_t$, by a slight abuse of notation, as the *pure stock* process.

This is the unique general framework that leads to a class of perfectly well-defined stochastic processes for the underlier, while preserving the exact formula (3) for the forward. If the pure stock follows GBM one has simple closed-form solutions for the price of European vanilla options.

It has many other advantages, especially if one wants to handle vanilla options consistently with other products on a flow/exotics desk. As already mentioned, it is easy to add default risk. Local volatilities have to be only calculated for the pure stock process, which has no cash dividends, making the calibration problem much simpler and numerically robust.[12]

---

[11] We will in the following ignore default risk; see [2] for a detailed discussion of how it should be treated.

[12] In fact, to emphasize the issues with local volatilities, we mention that on some desks that use the spot model to price vanillas, local volatilities are calculated by effectively assuming proportional rather than cash dividends; i.e. vanillas and products priced off local volatilities will not be quite consistent.

We are not sure, however, how many OMMs are using the hybrid model in this form. The main objection is that in this model the implied volatility (which is that of the pure stock) depends on dividends after maturity. If ones changes any dividend (before the transition to purely proportional dividends), than implied volatilities for all terms have to be recalculated.

Note that this is a practical problem, not a conceptual flaw of the hybrid model. It is perfectly consistent with the logic of the hybrid model that pricing depends on all dividends, including dividends after whatever maturity one might be considering. If "the market" were using the hybrid model, then a market consensus shift about the value of a dividend would lead to prices getting remarked in line with unchanged (hybrid model) implied volatilitities. The acceptance or not of the hybrid model is clearly somewhat of a chicken-and-egg – or education – problem.

2.3.3 The Partial Hybrid Model

Besides the spot model, it seems that the most popular dividend pricing model used by OMMs nowadays is a bastardized version of the hybrid model, that we will refer to as the *partial hybrid model* (sometimes known as the *escrowed dividend model*). Namely, instead of considering all dividends after t in D_t, one considers only dividends up to T when pricing an option of maturity T. In other words, one uses

$$D_t \;:=\; D_t(T) \;:=\; f_p(t) \sum_{i:t<t_i\leq T} \frac{d_i}{f_p(t_i)} \tag{9}$$

in the stock process (8).[13]

Note that the T-dependence means that we are not using a consistent stochastic process; we have a different one for each term (it is not simply a matter of giving some parameters term-structure, as one can do with rates and volatility). As is well-known, this can lead to arbitrage if one wants implied volatility to vary smoothly through ex-dividend dates.

To be sure, for a listed options market maker using the partial hybrid model matters are not as bad as they might first appear. The situations where this arbitrage is largest have large cash dividends, large maturities, and large volatilities, and one has to look at terms bracketing an ex-dividend date. We will discuss examples in section 4.2.5. OMMs using this model have presumably adapted by using suitable jumps in the vol term-structure across dividend dates (or by avoiding trading problematic options altogether).

Still, there are at least two types of traders for which using the partial hybrid model is problematic. One of course is a desk that trades more than just vanilla options; the partial hybrid model simply can not be integrated in a consistent pricing framework beyond vanillas. The other is an OMM that wants to have a framework to express sensible views on details of the volatility term-structure. As we will see in detail in section 4.2, it is (almost) always possible to calibrate both the partial and full hybrid models to the same sensible market, but their implied volatility term-structures will be quite different. In particular, only one of them can be smooth. Which is "right"?

[13] Just like with the full hybrid model, we are not sure who first thought of the partial hybrid model. But it goes back at least to early editions of John Hull's textbook [11]. One could argue it is implicit in the work of Roll, Geske and Whaley [3] on the pricing of an American call with one cash dividend.

2.3.4 The SKA Model

The fact that neither of the models discussed so far is completely satisfactory from both a theoretical and practical point of view has lead to the proposal of numerous variations and approximations, as mentioned in the introduction. Most of them are not of interest to us, because they are not general (or make no financial sense), but there is one exception. Namely, the spot-strike adjustment proposed by Bos and Vandermark [5] as an approximate solution to the spot model. It allows for the analytic pricing of European-style options, and has a very simple and intuitive motivation, viz, that in the spot model a dividend close to the valuation date essentially acts like a spot adjustment, and a dividend close to expiry acts like a strike adjustment (this is not true in the hybrid model).

This leads to the idea that one should decompose any dividends into a suitably defined "near" and a "far" term piece, subtract the near-term dividends from the spot, and add the far-term dividends to the strike. This is indeed the recipe to price European-style options; but this idea can actually be used to define a stochastic process for the stock (albeit maturity-dependent), that neatly fits into the hybrid model framework, and can therefore be used to price American-style vanilla (and other) options. Namely, we will define this model (SKA) via the shift

$$D_t \ := \ D_t(T) \ := \ f_p(t) \sum_{i:t<t_i\leq T} \frac{d_i}{f_p(t_i)} \ - \ D_t^{(f)}(T) \tag{10}$$

where the first term is the partial hybrid model shift, and the second the "far-term" dividend component

$$D_t^{(f)}(T) \ := \ f_p(t) \sum_{i:0<t_i\leq T} \frac{t_i}{T} \frac{d_i}{f_p(t_i)} \tag{11}$$

Although the shift $D_t(T)$ is maturity-dependent, as in the partial hybrid model, the SKA model seems superior to it. In particular it has much better no-arbitrage properties, see section 4.2.5. We will describe this and much else in more detail as we now summarize the various hybrid models we discussed, and compare their common and contrasting properties.

3 The Hybrid Model Framework

3.1 General Properties

Let us summarize and provide some notation about what we have learned so far. By *a hybrid model* we mean a stock price process written in terms of the forward F_t, a deterministic time-dependent shift D_t, and a martingale X_t satisfying $E[X_t] = 1$, as $S_t = (F_t - D_t)X_t \ + \ D_t$. We will also write this as

$$S_t \ = \ \tilde{S}_t + D_t \tag{12}$$

and refer to $\tilde{S}$ as the *pure stock*. If we have a good handle on how to price with the pure stock $\tilde{S}$ then this carries over to pricing with the shifted stock S. For example, European options on S can be priced in terms of shifts on the inputs to the pricing formula for $\tilde{S}$ (which can involve any process, not just GBM).

Explicitly, in terms of a strike K, a vanilla option on S involves $S_t - K = \tilde{S}_t - \tilde{K}_t$, where we defined the adjusted time-dependent strike $\tilde{K}_t$ as

$$\tilde{K}_t \ := \ K - D_t \ . \tag{13}$$

So any vanilla option, with European or American exercise style, can be priced in terms of an adjusted stock price process $\tilde{S}_t$ and a time-dependent strike $\tilde{K}_t$. Considering European options, the quantities relevant for pricing are D_0, which can be thought of as a spot-adjustment, and D_T, which one can think of as a strike adjustment. Equivalently, if we think in terms of the Black formula for European options, where the forward rather than the spot is used as input, D_T is both the strike and forward adjustment (as required by European put-call parity).

In all hybrid models we will consider, $F_t - D_t$ has no cash dividend related jumps, i.e. the pure stock $\tilde{S}_t = (F_t - D_t)X_t$ follows a process with no cash dividend jumps (for fixed term T at least). For the rest of this paper we will assume that X follows GBM. Then the pure stock $\tilde{S}$ also follows GBM with perhaps occasional proportional jumps due to discrete proportional dividends.

For completeness and future reference let us explicitly give the European pricing formulas for a hybrid model under Black-Scholes assumptions (with term-structure if desired, with the obvious re-interpretation of symbols). We can write the Black formula for an *un-discounted* call or put, $\hat{C}$ or $\hat{P}$ (denoted generically as $\hat{V}$), of maturity T and strike K as

$$\hat{V}(T,K) \;=\; \hat{V}(F_T,K,\hat{\sigma}) \;=\; F_T\, f_\pm(k,\hat{\sigma}) \tag{14}$$

in terms of two dimensionless functions $f_\pm$, for call and put respectively, of two dimensionless variables $k := K/F_T$, $\hat{\sigma} := \sigma\sqrt{T}$ given by

$$f_\pm(k,\hat{\sigma}) \;=\; \pm\, N(\pm d_+(k,\hat{\sigma})) \;\mp\; k\, N(\pm d_-(k,\hat{\sigma}))\ . \tag{15}$$

Here $N(x)$ is the cumulative normal distribution function and

$$d_\pm(k,\hat{\sigma}) \;:=\; \frac{-\ln k}{\hat{\sigma}} \pm \tfrac{1}{2}\hat{\sigma}\ . \tag{16}$$

The un-discounted price of a European vanilla in a hybrid model with shift D at maturity is then

$$\hat{V}_D(T,K) \;=\; \hat{V}(F_T - D, K - D, \hat{\sigma})\ . \tag{17}$$

Note that for strikes below the shift, $K \leq D$, which the stock can never reach, we have $\hat{C} = F_T - K$, $\hat{P} = 0$.[14] All formulas are of course consistent with put-call parity with the "usual" formula for the forward, eq. (3).

3.2 The Hybrid Models Compared

In the original hybrid model we discussed, the shift D_t is given by the PV, suitably discounted to t, of all cash dividends after t. This makes financial and intuitive sense, and options prices calculated in this model will obviously be arbitrage-free (with sensible, smooth borrow and vol curves at least). This is the *(full) hybrid model*; we will denote it as HM2 or FHM.

The *partial hybrid model* discussed next considers only dividends up to maturity T in D_t when pricing options of term T. It is not really a consistent model, as the stock process is different for different maturities (i.e. D_t is really a function of T too). It is however commonly used by OMMs. We will denote it as HM1 or PHM.

The *spot-strike adjustment* approximation of [5] led us to define another hybrid model, which we will refer to as SKA or HM3. Its shift $D_t(T)$ is also term-dependent, and at maturity $D_{t=T}(T)$

[14] This means that the Black vol of options managed with a (smooth) hybrid model vol surface will vanish as $K \searrow D$, and start to "keel over" before that.

is negative, so D_t can not be interpreted as any kind of "PV of cash dividends" anymore. Instead, as in the original motivation for this approach, we think of D_t as interpolating between a spot adjustment for small t and a *positive* strike adjustment for t close to maturity T. Nevertheless, on a purely mathematical level it fits just as well into our hybrid model framework as HM2 and HM1.

For reference, let us repeat the formulas for the shifts in the various models:

$$
\begin{aligned}
D_t^{(2)} &= f_p(t) \sum_{i:t<t_i<\infty} \frac{d_i}{f_p(t_i)} && \text{HM2} \\
D_t^{(1)}(T) &= f_p(t) \sum_{i:t<t_i\le T} \frac{d_i}{f_p(t_i)} && \text{HM1} \\
D_t^{(3)}(T) &= f_p(t) \sum_{i:t<t_i\le T} \frac{d_i}{f_p(t_i)} \;-\; f_p(t) \sum_{i:0<t_i\le T} \frac{t_i}{T} \frac{d_i}{f_p(t_i)} && \text{HM3}
\end{aligned}
\tag{18}
$$

The spot adjustments D_0 are given by

$$
\begin{aligned}
D_0^{(2)} &= \sum_{i:0<t_i<\infty} \frac{d_i}{f_p(t_i)} && \text{HM2} \\
D_0^{(1)}(T) &= \sum_{i:0<t_i\le T} \frac{d_i}{f_p(t_i)} && \text{HM1} \\
D_0^{(3)}(T) &= \sum_{i:0<t_i\le T} \left(1 - \frac{t_i}{T}\right) \frac{d_i}{f_p(t_i)} && \text{HM3}
\end{aligned}
\tag{19}
$$

Finally, the shifts at maturity T, aka the strike or forward adjustments, are

$$
\begin{aligned}
D_T^{(2)} &= f_p(T) \sum_{i:T<t_i<\infty} \frac{d_i}{f_p(t_i)} && \text{HM2} \\
D_T^{(1)}(T) &= 0 && \text{HM1} \\
D_T^{(3)}(T) &= -f_p(T) \sum_{i:0<t_i\le T} \frac{t_i}{T} \frac{d_i}{f_p(t_i)} && \text{HM3}
\end{aligned}
\tag{20}
$$

Let us summarize some properties of the shifts. First of all, we have

$$
D_t^{(3)}(T) \le D_t^{(1)}(T) \le D_t^{(2)} \; . \tag{21}
$$

Qualitatively their t-dependence is very similar in that just like the forward and the stock they all jump across dividend dates like

$$
D_{t_i^+}^{(k)}(T) \;=\; (1-\delta_i)\, D_{t_i^-}^{(k)}(T) \;-\; d_i \; . \qquad\qquad k = 1, 2, 3 \tag{22}
$$

This is of course by design. Between dividend dates they vary in a similar manner, simply due to the dependence on the proportional growth factors $f_p(t)$, i.e. completely smoothly. We show an example of the time-dependence of the various shift functions in figure 1.

At maturity, there is no forward or strike adjustment in the partial hybrid model, $D_T^{(1)}(T) = 0$. This means that, as long as the forward is the same, European implied vols in this model would be the same as without cash dividends. Another way of saying this, its European implied vol is

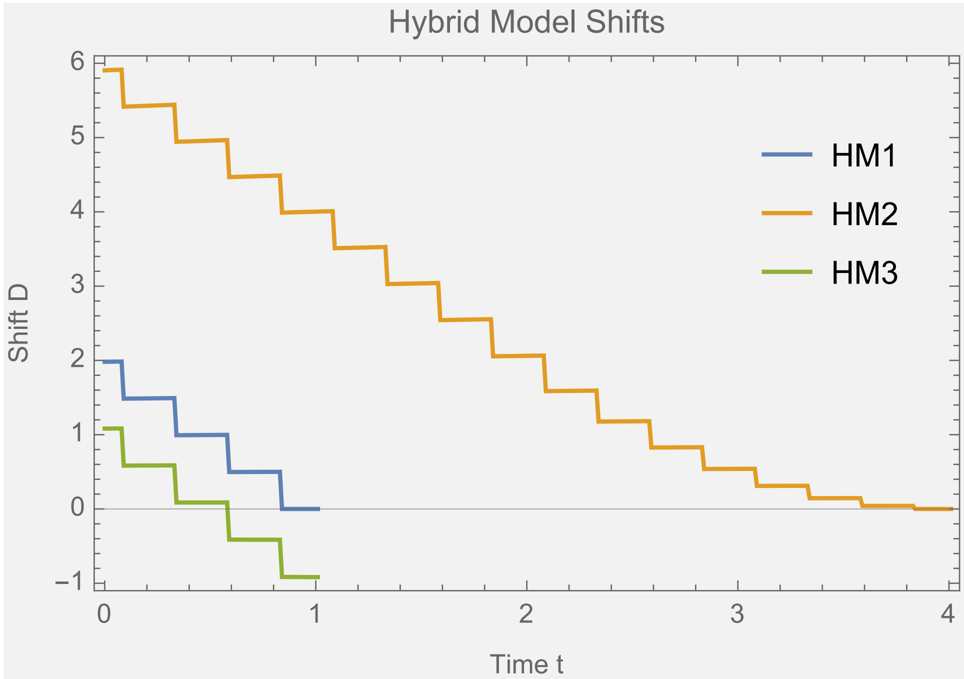

Figure 1: The shifts D_t for various hybrid models with $r = 3\%, q = 1\%$, and a quarterly cash dividend of 0.5 first paid after about one month, $t_1 = 0.085$. We assume a blending scheme with cash dividends transitioning linearly to zero between 2 and 4 years. For HM1 (aka PHM) and HM3 (aka SKA) the shift functions depend on maturity, which here is taken as $T = 1.01$.

that of the forward, aka the "Black vol". This might be one of the reasons this model is popular, despite its obvious drawbacks. Note that in the American case its non-zero shift function D_t does of course affect pricing.

For the other two hybrid models D_T has the same behavior when T passes through a dividend date, namely

$$D^{(k)}_{t_i^+}(t_i^+) \;=\; (1-\delta_i)\, D^{(k)}_{t_i^-}(t_i^-) \;-\; d_i\;. \qquad\qquad k = 2, 3 \tag{23}$$

For the spot adjustment D_0 we note that in the full hybrid model it is a constant independent of T. In the partial hybrid model it jumps up by the (discounted) cash dividend amount each time T passes through a dividend date. This is the source of the obvious arbitrage that exists in this model when one tries to maintain smooth implied vols through dividend dates. Finally, in the SKA model the spot adjustment involves only the near-term piece of the dividends up to maturity, which increases continuously, rather than with a jump, as T goes through a dividend date. That is why it is much harder to find arbitrage in the SKA model than in the partial hybrid model (see section 4.2.5 for details).

The properties of the different models are summarized in table 1. For completeness we also added the model "HM0", where cash dividends are converted to discrete proportional dividends. The underlier follows GBM with proportional jumps at the dividend dates. It is a useful reference model with appealing theoretical and numerical properties – but unfortunately not realistic and not what leading OMMs in the US equity markets use nowadays.

Feature	HM0	HM1=PHM	HM2=FHM	HM3=SKA	SM
SDE	GBM	shiftedGBM	shiftedGBM	shiftedGBM	pwGBM
Depends on small S dividend policy	No	No	No	No	Yes
Consistent SDE across terms	Yes	No	Yes	Almost	Yes
Arbitrage with accu implementation	No	Yes	No	Little	No
Usual, exact forward formula	Yes	Yes	Yes	Yes	No
Exact, fast Euro prices	Yes	Yes	Yes	Yes	No
Pricing depends on dividends after T	No	No	Yes	No	No
Spot adjustment at T: $D_0(T)$	0	> 0	> 0	> 0	0
Strike adjustment at T: $D_{t=T}(T)$	0	0	≥ 0	< 0	0
Spot adjustment $D_0(T)$ thru t_d	0	Jumps up	Constant	Continuous	0
Strike adjustment $D_{t=T}(T)$ thru t_d	0	0	Jumps down	Jumps down	0

Table 1: Comparison of qualitative features of the various hybrid models and the spot model in the presence of cash dividends. See the main text for details and explanations.

3.3 Calibration Properties

We would now like to gain some understanding of the problem that market markets and other market participants face in practice, namely how to calibrate one's pricing model to the market. Here this means calibrating implied borrow rates and volatilities. A market maker would like to match – or make – the market with as little term-structure in these parameters as possible.[15]

To make this concrete and precise we can think of "the market" as given by the full hybrid model, which is a sensible model that we know has no arbitrage. We can use constant values for rates and vols, as an extreme case of smooth parameters. Qualitative relationships we find in such cases, will largely carry over to more realistic ones. In the next section we will explore examples of this setup numerically, here we will discuss what can be said from first principles and analytically.

Imagine picking a set of maturities and strikes and calculating the prices of calls and puts in "the market", i.e. the hybrid model with our chosen "input" rates and vols. First, recall that by design all hybrid models have exactly the same forwards for given rates and dividends. This means that the implied borrow rates will be very similar if we calibrate HM1 and HM3 to the HM2 market. In fact, for European options the implied borrows will be *exactly* the same as the input one, as follows from put-call parity. For American options they will be extremely close due to the very similar qualitative properties of the shifts D_t in these models, which therefore lead to very similar early-exercise premia in corresponding options across these models. We will see this numerically in the next section.

Having the same or very close implied rates (and forwards) is very nice, and means that we can also gain some intuition about the relationships of implied vols in the these models. First, consider at-the-forward (ATF) European options. Since these are priced with the Black formula for the pure stock, we know that the price of an undiscounted ATF call or put is given by

$$\hat{C} \;=\; \hat{P} \;=\; F\,\left(2N(\tfrac{1}{2}\hat{\sigma})-1\right) \;=\; \frac{1}{\sqrt{2\pi}}\,F\,\hat{\sigma}\,\left(1-\frac{\hat{\sigma}^2}{24}+\frac{\hat{\sigma}^4}{640}-\ldots\right) \tag{24}$$

where $\hat{\sigma} := \sigma\sqrt{T}$. Note that the notion of "ATF-ness" is the same in all these models, since $\tilde{F}_T - \tilde{K}_T = F_T - K$ in all cases.

[15]Unless there is a good reason to believe there should be term-structure, as with ongoing or anticipated "events" of various kinds. In that case we are here referring to the "background" or "clean" volatility term-structure.

Using the appropriate pure forwards $\tilde{F}_T^{(k)} = F_T - D_T^{(k)}$ we can write the relationship of the calibrated ATF vols of any two hybrid models, labelled '1' and '2' here, as

$$\sigma_T^{(1)} \; = \; \sigma_T^{(2)} \, \frac{\tilde{F}_T^{(2)}}{\tilde{F}_T^{(1)}} \left(1 \; - \; \frac{{\sigma_T^{(2)}}^2 T}{24} \left(1 - \left(\tilde{F}_T^{(2)} / \tilde{F}_T^{(1)} \right)^2 \right) \; + \; \mathcal{O}({\sigma_T^{(2)}}^4 T^2) \right) \tag{25}$$

This formula is extremely accurate in the European case, and also very accurate for American-style options, in all realistic scenarios where accuracy would be required. What is even more amazing is that the exact relationship between the hybrid model volatilities in the European case for generic strikes, away from ATF, which can numerically be implemented at the cost of an inversion of the Black-formula, also holds extremely accurately in the American case. We will give details in section 4.2.4.

Note that eq. (21) implies that for the implied vols in different calibrated hybrid models we have the same ordering,

$$\sigma^{(3)}(T,K) \;\le\; \sigma^{(1)}(T,K) \;\le\; \sigma^{(2)}(T,K)\ . \tag{26}$$

In the European case, around ATF, this is obvious from eq. (24). For general strikes this follows from the Black formula. Namely, it is easy to show that in terms of our earlier notation

$$\partial_D \hat{V}(F-D, K-D, \hat{\sigma}) \; = \; N(d_-(\tilde{k}, \hat{\sigma})) \; - \; N(d_+(\tilde{k}, \hat{\sigma}))\ , \tag{27}$$

where $\tilde{k} := (K-D)/(F-D)$. The rhs of the above is strictly negative if $\hat{\sigma} > 0$, $K > D$, and zero otherwise. Since the Black-formula is stricly increasing in $\hat{\sigma}$, the ordering in eq. (26) follows. In the American case, a similar claim presumably holds, but is harder to prove; remember that in this case even the calibrated borrows will be (slightly) different.

4 Implementation and Numerical Results

4.1 Implementation

It is sometimes stated that accurate and fast pricing of generic American vanilla options with cash dividends requires the full arsenal of modern finite difference methods, i.e. that simple binomial (or trinomial) tree pricing methods will not be adequate. If one were to use the spot model, where one really has to "put the cash dividends on the grid", this might be the case (although we are not aware of an explicit study of this question). It is certainly the case once one has enough exercise and other discrete features that have to be treated on the grid, as in convertible bonds, for example. It is presumably also the case if very high accuracy is required (recent work of Andersen et al [12] raises the prospect that for certain cases there might be even much faster customized methods).

But using any of our hybrid models, the cash dividends are largely taken out of the pricing problem for vanilla options; they only appear as a time-dependent strike. Other discrete features like events are customarily treated via a rescaling of the time axis in the core pricing code, and so do not create problems.

In this situation carefully tuned tree methods provide very fast algorithms to achieve the kind of accuracy required in practice, relative to the bid-ask spreads one finds in the market. Here we will use Leisen-Reimer (LR) binomial trees [13] to implement pricing of all hybrid models within our general dividend framework. In the European case without discrete dividends it has been rigorously proved that for vanilla calls and puts LR trees have second-order convergence in the number of steps in the tree, even without Richardson extrapolation [14].

Several numerical studies, often for the case of American puts with $q=0, \delta_i=0, d_i=0$, indicate that they are significantly more accurate than CRR and various other trees, as well as various finite difference implementations, for a given CPU time budget [13, 15, 16, 17].

Many of their good properties carry over to generic American vanilla options, with some modifications even to the case of discrete dividends, cash or proportional. A serious study of this question involves a discussion of which metrics to use when evaluating a pricer for practical use, in addition to speed, accuracy and convergence tests in the much larger parameter universe that presents itself when considering discrete dividends. This, and the details of our implementation, which involves several tricks to improve speed and convergence, are beyond the scope of the current work. They will be discussed in [18]. Here we just present a few remarks concerning the correctness and speed of our implementation.

We tested our code against most results available in the literature [7, 12, 16, 19, 20], including greeks when available.[16] High precision results are not commonly quoted, especially in cases with discrete dividends. We can test to high precision against analytical results for the European case, and e.g. against the exact integral formula for an American call option with one discrete dividend at time t,

$$C_0^{\mathrm{1div}}(T,K) \;=\; \mathrm{e}^{-rt}\int_0^\infty dS_{t^-}\;\rho(S_0,0|S_{t^-},t)\;\max\big((S_{t^-}-K)_+, C_t^E(S_{t^+},T,K)\big) \tag{28}$$

which holds as long as exercise is optimal only just before the dividend (which is guaranteed to be the case if $r \geq$ and $q \leq 0$).[17]

A reasonably efficient C++ implementation allows us to price one option on an $N=101$ step tree in about 12 microseconds on a 3.5GHz Intel Core i5 (iMac Retina 5K, late 2014); an $N=65$ tree takes about 6 microseconds. If this pricer is used for both calibration and pricing of, say, the US equity options universe, we believe that $N=65$ or 101 steps are adequate in practice (assuming one handles various corner cases, e.g. when an uncertain dividend date is modeled by "splitting" the dividend across closely neighboring dates).

Our pricer is faster (by factors of 100s or even 1000s) than all other generic vanilla pricers with discrete dividends we are aware of. It rivals many "fast, analytical" pricers, e.g. if they involve numerical integration or root finding steps. In fact, using this pricer one can dispense with another layer of complexity that most OMM firms use in practice, namely to save prices and greeks from a "slow" finite difference or tree pricer in tables whose axes are spot and volatility, and then interpolate these tables during the trading day. These tables will have to be re-generated if e.g. a dividend amount or date changes. Instead, we can just use our pricer and its greeks for small spot or vol moves, and easily handle any intra-day changes in dividend schedules, etc.

---

[16] The only results we disagree with are for some of the American calls with discrete proportional dividends quoted in [20]. For e.g. $S_0 = 100, \sigma = 30\%, r = 4\%, q = 0, T = 1$ with $\delta_1 = 5\%, d_1 = 0$ at $t_1 = 0.5$, we find at $K = 80, 120$ prices of 23.6712, 5.0295, respectively, rather than 26.086, 8.609.

[17] Here $C_t^E(S_{t^+},T,K)$ is the price of a European call of strike K at time t with remaining time-to-maturity $T-t$. ρ is the transition density for the underlier (in a hopefully obvious notation), and S_{t^+} refers to the stock right after the dividend, S_{t^-} just before. Any dividend model can be used, as long as the dividend is a function of the stock price at (or up to) t. In particular, an affine dividend is allowed, where $S_{t^+} = S_{t^-}(1-\delta) - d$. Note that we can allow vol term-structure in the above, because we can use a different vol for the Black-Scholes transition density ρ from time 0 to t, and the European call $C_t^E(S_{t^+},T,K)$ that is relevant thereafter. At the cost of requiring multi-dimensional integrals, this formula can be generalized to multiple dividends (one integral per dividend). The integrals can be expressed in terms of the standard or higher-dimensional cumulative normal distributions; see e.g. [21] and references therein.

4.2 Calibration Results

Since none of the models used in practice by various market participants is perfect we would like to understand the relationship between them better, in addition to understanding what properties they have in common and where they differ. In particular, we want to show, that if one of the models HM1, HM2, HM3 matches the market – i.e. might have been used to *make* the market! – then any of the other two will match the market too, though perhaps not with a smooth term-structure/surface for borrow rates and volatilities.

To do so, we will consider prices from the full hybrid model HM2 with flat borrow rates and vols as the reference "market". It does not matter much which model we choose as reference, since we mainly care about the *relationship* between calibrated versions of the models. But HM2 is a good choice, since we know it to be arbitrage-free with constant input parameters.

What we will do is to take the HM2 prices and for each pair of a call and put at a given term and strike, imply what borrow and volatility is required to match them in any of the other models. This is almost always possible; the only generic exception seems to be very low strikes close to the HM2 shift D_t. This of course makes sense, since the put prices drop to 0 there. For "reasonable" values of terms and strikes does the calibration process for borrow and vol succeed.[18]

To illustrate that the close relationship of the hybrid models is not in some sense automatic or trivial, we will also sometimes show calibration results for two other models, that do not fall in the hybrid framework. They are, in fact, the two models that were often used in the past. The first is simply the "DivYield" model, where we ignore all discrete dividends and match the market prices for a call and a put with an implied dividend yield (which includes the borrow cost) and an implied vol. In the second model we replace any cash dividend with a discrete proportional dividend.[19] We will refer to this model as HM0, as mentioned earlier, by a little bit of abuse of notation.

We have studied many examples of this calibration process and will here present results for a relatively "hard" case. Namely, our reference market will be HM2 with $S_0 = 100$, $\sigma = 30\%$, $r = 3\%$, $q = 1\%$, $\delta_i = 0$, and quarterly cash dividends, which unless otherwise indicated are taken to be $d_i = 2$, with the first one paid at $t_1 = 0.085$ (we also looked at $t_1 = 0.005$, which leads to similar results and the same conclusions). We will assume a linear transition form pure cash to discrete proportional dividends between 2 and 4 years, i.e. a $(2, 4)$ blending scheme (so, for maturities $T \geq 4$ the HM1 and HM2 models will be identical).

Note that this example corresponds to a cash dividend yield of about 8%, which is much larger than that of the names in the US equity universe with the highest accuracy requirements, namely ETFs like SPY and single stocks like AAPL.

4.2.1 Implied Borrow Rates

We first present results for the implied borrow term-structure in various model obtained from strike $K = F_T$ (with the forward from HM2). For all the models mentioned this is shown in figure 2; leaving out the DivYield model we can zoom in for higher resolution in the remaining models in

[18]We have taken pains to insure that our calibration algorithm is accurate and all implied borrows and vols shown below should be accurate to $\mathcal{O}(10^{-7})$ or better.

[19]There are actually two "modes" in which this model could be used: Either replacing the cash dividend d with a δ that gives the same forward, and then calibrating borrow and vol. This assumes one knows the borrow and/or forward to start with, which in practice might not be the case; it is also maturity-dependent. So a more realistic and simpler way of using this model would be to set $\delta_i \to \delta_i + d_i/S_0$, $d_i \to 0$, and then imply borrow and vol. This is what we will do.

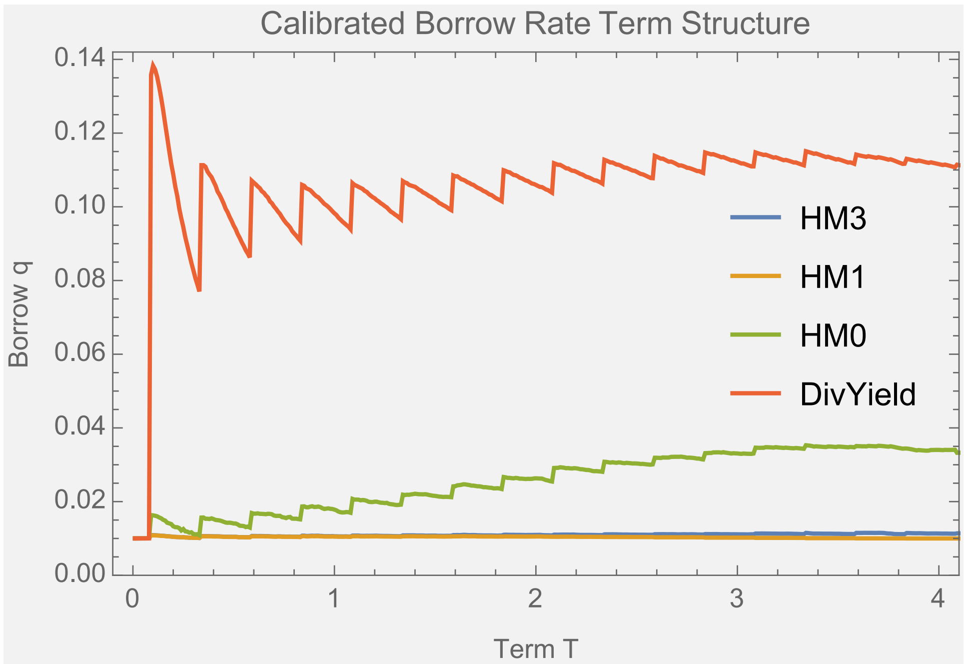

Figure 2: Implied borrow rate term-structure of various models calibrated to reference market HM2 with $r = 3\%, q = 1\%, \sigma = 30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1 = 0.085$. Exercise-style is American and $N = 65$.

figure 3. For the HM0 and especially DivYield models we see that the implied borrow is very different from the input one, with an erratic term-structure. HM1 and HM3, on the other hand, have a borrow cost that is much closer to the input HM2 one, with only slight bumps at the dividend dates. Recall that in the European case they would have, in fact, *exactly* the same implied borrow as the input HM2 one, since all these models have the same formula for the forward, and put-call parity for a call and a put gives us the forward. If we had used a cash dividend yield of around 2% rather than 8%, than the implied HM1, HM3 borrow would be virtually indistinguishable from the input HM2 one, even in the American case (the deviations seem to scale roughly quadratically in dividend and linear in vol).

Note that the bit of "noise" seen in figure 3 is not due to our calibration algorithm, but purely due to discretization errors of the early exercise boundary for the small number of tree steps ($N = 65$) used in this figure. This noise disappears with a larger number of steps, as seen in figure 4.

4.2.2 Implied Volatilities

Next, let's show the same plots for the calibrated volatility term-structure along $K = F_T$, starting with figure 5. Now we see some more interesting phenomena. First of all, note the extremely zig-zaggy term-structure in the DivYield model. To put this in context, recall that in the European case the vol term-structure of the HM1, HM0 and DivYield models would be *exactly* the same, and very closely match that of HM1 shown in the figure. This illustrates that (a) early-exercise premia are *not* small in this example, and (b) that their values behave qualitatively very differently in the DivYield and HM0 models relative to HM1 and HM2. As alluded to earlier, the relationship between HM1 and HM2 vols is very much the same in the American and European cases, and very accurately given by (25). The same is true for the HM3 and HM2 vols (and hence for the HM3

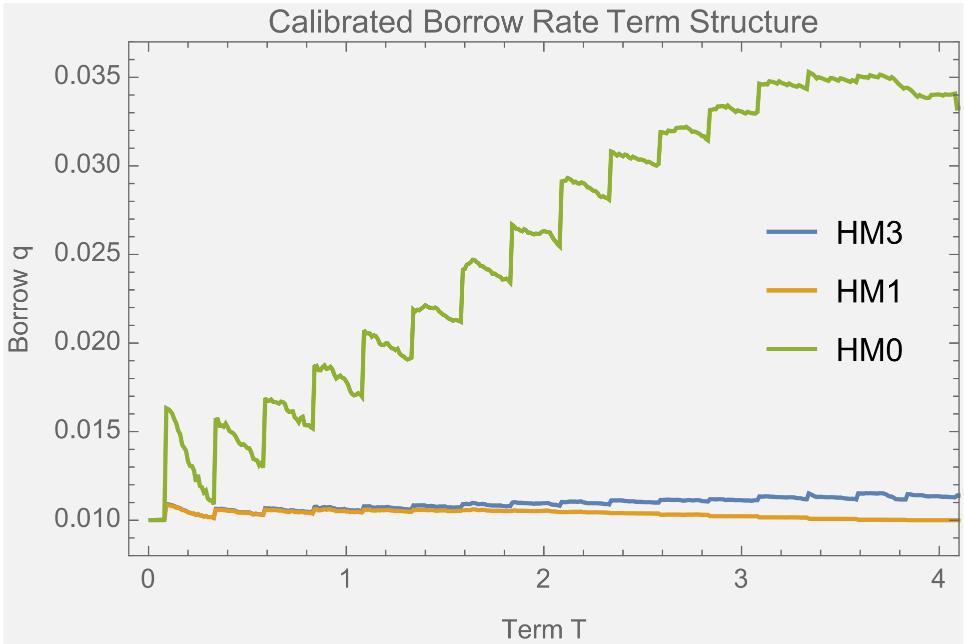

Figure 3: Implied borrow rate term-structure of various models calibrated to reference market HM2 with $r = 3\%, q = 1\%, \sigma = 30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1 = 0.085$. Exercise-style is American and $N = 65$.

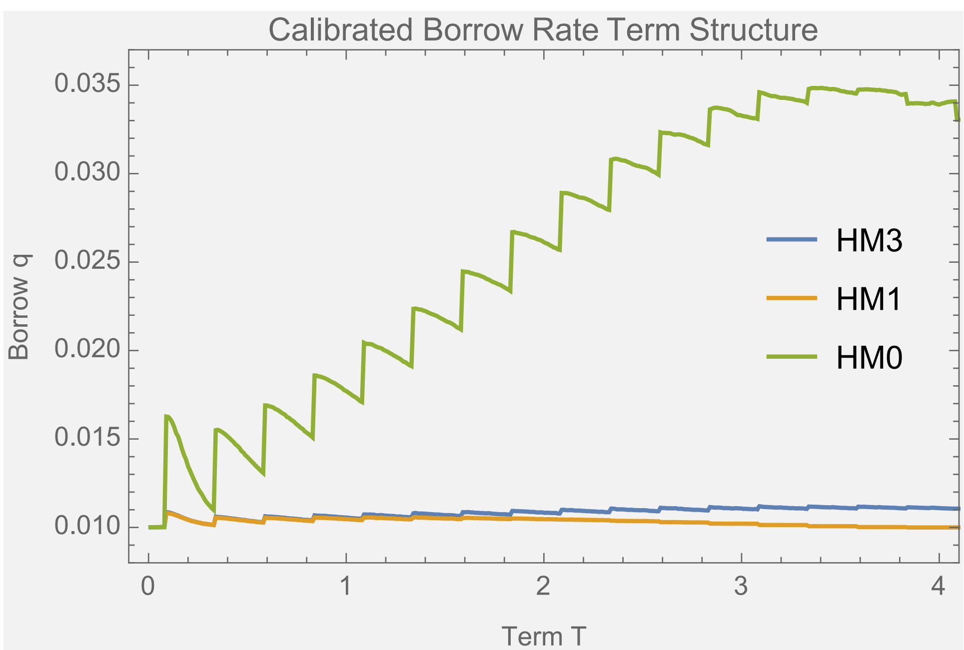

Figure 4: Implied borrow rate term-structure of various models calibrated to reference market HM2 with $r = 3\%, q = 1\%, \sigma = 30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1 = 0.085$. Exercise-style is American and $N = 651$.

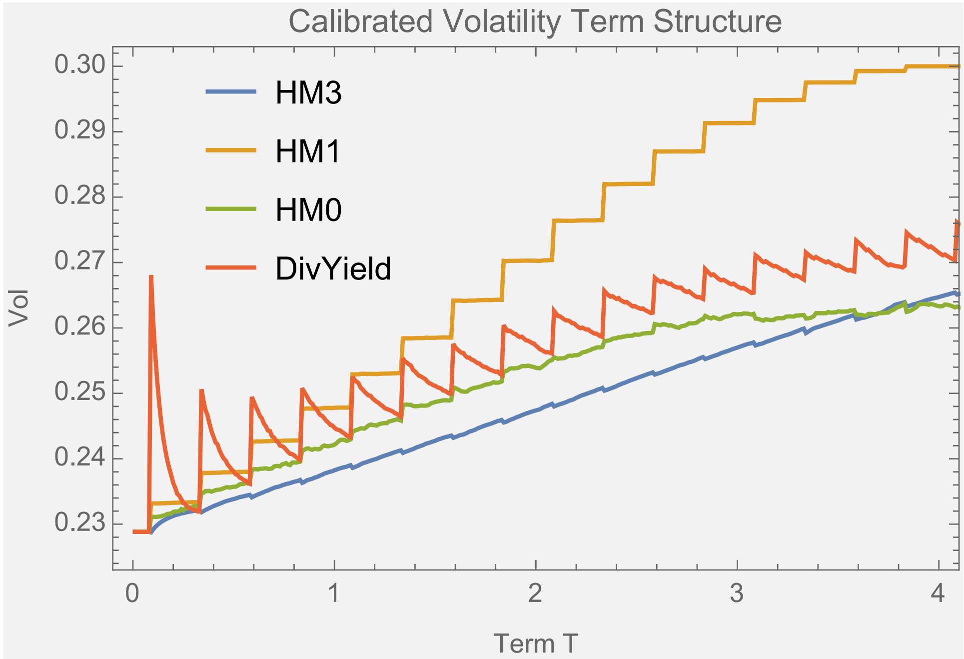

Figure 5: Implied volatility term-structure of various models calibrated to reference market HM2 with $r = 3\%, q = 1\%, \sigma = 30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1 = 0.085$. Exercise-style is American and $N = 65$.

and HM1 vols).

We can now also see differences between HM1 and HM3 that were not apparent in the implied borrow term-structure. Namely, whereas implied HM3 vols have an essentially smooth relationship with HM2 ones (certainly for lower div yields), HM1 vols have discrete up-jumps at each dividend date. This is of course to avoid arbitrage, since at each dividend date the spot-adjustment in HM1 jumps up, as we saw earlier.[20]

4.2.3 Implied Skews

So far we discussed the implied parameter term-structure, along the ATF "back bone". But what about the "skew", i.e. the strike dependence of the implied parameters for fixed term T? In figure 8 we show the implied borrow skew of HM0, HM1, HM3 for one maturity just before, and one just after a dividend date. Whereas HM1 and HM3 have an essentially flat implied borrow, close to the input HM2 one, there is a steep skew for HM0 (with a large jump across the dividend date). This means that if "the market" is using HM1, HM2 or HM3, we could not match it with one borrow per term if we were using the HM0 model. We would need a "borrow skew", which OMMs are not using.[21]

We see a similar if less dramatic picture in the implied vol skew in figure 9. Here the most remarkable feature is perhaps how close the HM3 skews before and after the dividend date are, generalizing what we saw for the ATF term-structure of vols. Ignoring the term-structure jumps

[20] To be precise, in HM1 the spot of the pure stocks jumps down while its strike (at maturity) does not change, whereas for the pure stock in HM2 the spot does not change, but the strike jumps up. However, for American options, lowering the initial spot lowers the price of a call much more, usually, than increasing the strike at maturity.

[21] Though conceptually there is nothing wrong with it – it can be justified with the same reasoning as the implied vol skew – and it is sometimes used in the pricing of structured products.

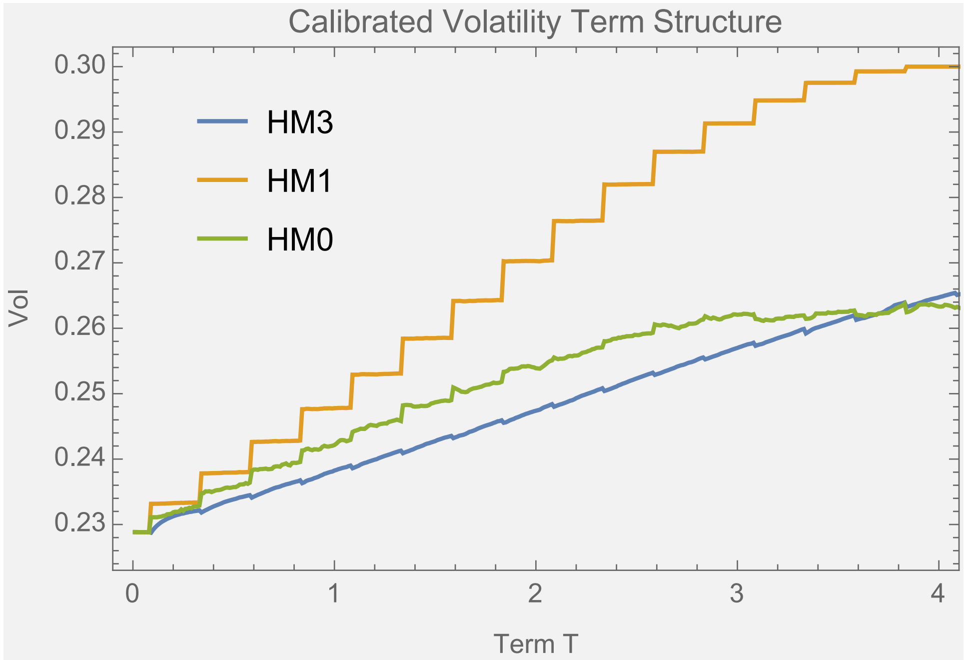

Figure 6: Implied volatility term-structure of various models calibrated to reference market HM2 with $r = 3\%, q = 1\%, \sigma = 30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1 = 0.085$. Exercise-style is American and $N = 65$.

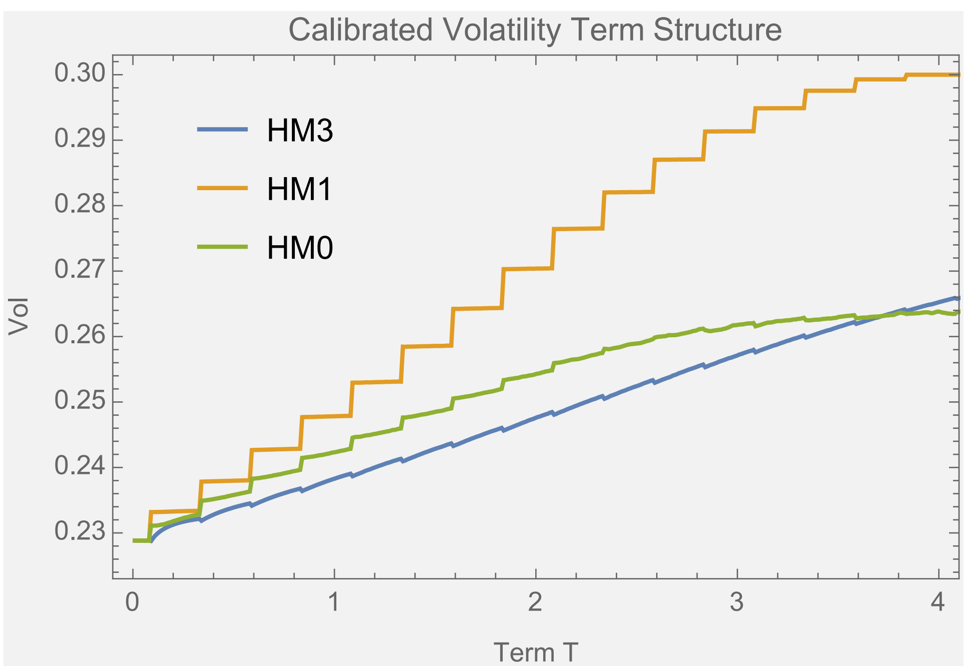

Figure 7: Implied volatility term-structure of various models calibrated to reference market HM2 with $r = 3\%, q = 1\%, \sigma = 30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1 = 0.085$. Exercise-style is American and $N = 651$.

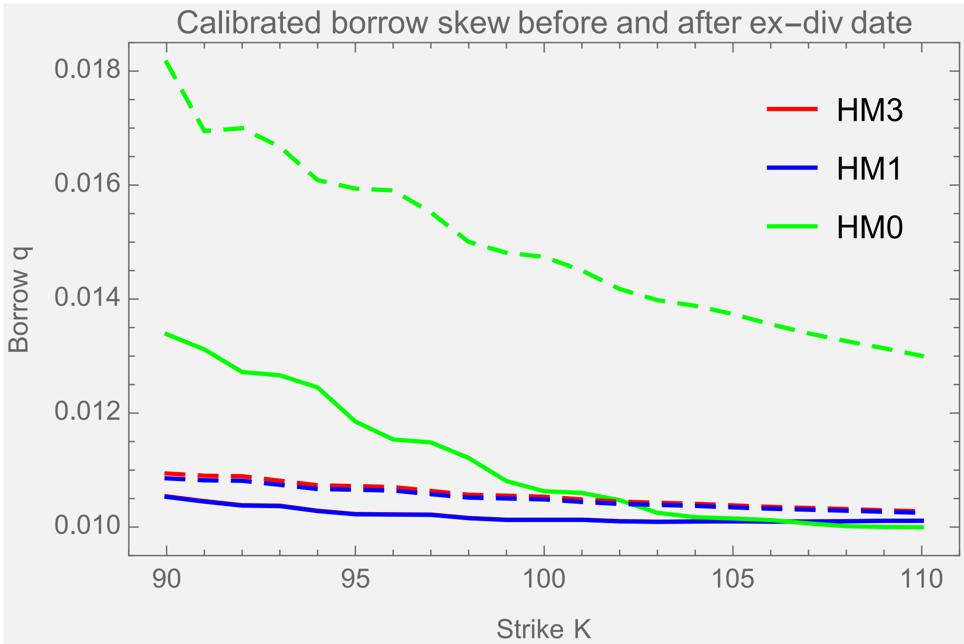

Figure 8: Implied borrow skew of various models calibrated to reference market HM2 with $r=3\%, q=1\%, \sigma=30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1=0.085$. We show the skew just before, $T=0.33$ (solid), and after, $T=0.34$ (dashed), the second dividend. Exercise-style is American and $N=65$. With larger N the HM0 lines becomes smoother, but keep the same shape.

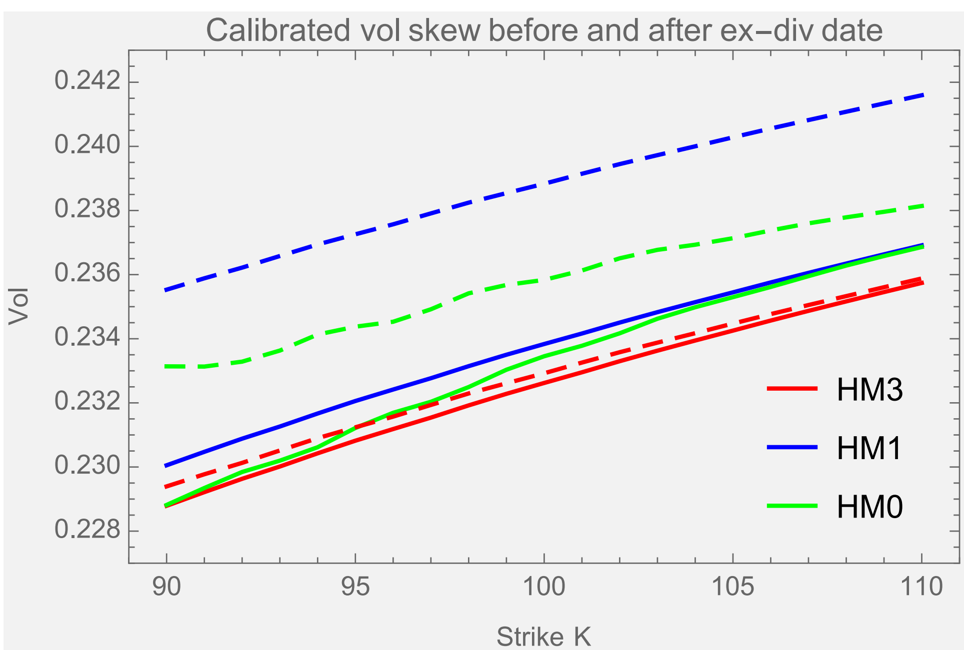

Figure 9: Implied volatility skew of various models calibrated to reference market HM2 with $r=3\%, q=1\%, \sigma=30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1=0.085$. We show the skew just before, $T=0.33$ (solid), and after, $T=0.34$ (dashed), the second dividend. Exercise-style is American and $N=65$.

in HM1 and HM0, the relatively flat implied vol skew in these models can be easily absorbed in a somewhat different fitted vol curve than would be used for HM2 (if we were using a volsurface rather than a flat vol for HM2).

4.2.4 Relationship between Hybrid Model Parameters

In sum, these figures show that the different hybrid models HM1, HM2 and HM3 are very closely related. If one can be calibrated to the market, so can any of the other ones – except that for HM1 we should expect the volatility to jump up across dividend dates to avoid arbitrage – as long as strikes stay above the floor provided by the dividend-related shifts D_t. The fact that their implied borrows are exactly the same in the European case, carries over with high accuracy to the American case. We mention again that the extremely accurate relationships (25) of the ATF vols for the European case hold with very high accuracy also in the American case. As an example, we note that for a SPY-like cash dividend yield of 2% the HM1 and HM3 ATF vols can be calculated from the HM2 vols to a fraction of a basis point out to at least $T = 2$ from eq. (25) in the American case; much more accurately still in the European case.

In fact these relationships even hold away from ATF: In the European case the exact relationship between different hybrid model volatilities for any term and strike is implicitly defined via the Black formulas with adjusted forward and strike inputs, as appropriate for the different models. Away from ATF these relationships between implied volatilities are only slightly less accurate than what we found ATF. They are at least 100 times more accurate than the types of relationships between dividend models discussed in [6]. For (most) practical purposes these can be considered to be exact relationships. Note that they even hold for the very small ($N = 65, 101$, etc) tuned trees that we are usually considering here. It is pretty amazing to find such relationships for *anything* involving American options that have, as generically here, large early exercise premia.

All of this is testament to the fact that the "fine-structure" of early exercise premia are very similar in these models. This can be traced to the similar qualitative features that the shifts D_t have in the different models (as functions of t for fixed T). Up to factors with smooth time-dependence they are just shifted versions of each other, as we saw earlier.

Note that these accurate relationships can be used to solve various problems arising in practice. For example, if an OMM insists on using the partial hybrid model HM1, one can use its relationship to the full hybrid model HM2 to estimate the vol jumps across dividend dates necessary to avoid arbitrage; we assume this would have been done for a while using the well-known approximate relationship of ATF vols (of course, it still seems easier and more sensible to just manage an HM2 or HM3 vol surface directly). Similar, if one is using an HM1 or HM3 model one can transform their vol surface to that of an HM2 model and then use HM2 local vols, or a more suitable microscopic model calibrated to the HM2 surface, to manage (light) exotics consistent with vanillas.

Another question we should briefly address here is the relationship of implied parameters when using different cash dividend blending schemes, for a given hybrid model. In general, the relationship between the implied borrows and vols for the American case will not be as close to the European case as we saw when using the same blending scheme for different hybrid models discussed above. Intuitively this makes sense, since the the shifts D_t in different blending schemes are *not* essentially just shifted versions of each other (as for $t \leq T$ in figure 1). However, note that when comparing two blending schemes there will be no difference at all between implied parameters for terms before the onset of blending in either scheme, as long as pricing does not depend on cash dividends after maturity, i.e. for HM1 and HM3. For HM2 there will be differences, but they will be very accurately predicted by the European case, similar to the situation when using different hybrid models (and for the same intuitive reason).

Empirically, for the US market, the onset of blending seems to be quite far out, at least $2-3$ years for names where it matters, i.e. names where bid-ask spreads are tight enough to see differences that remain after calibration with different blending schemes (chiefly, with too early blending one will not be able to match in-the-money options for larger maturities). In other words, the onset of blending seems to be not earlier than the longest-dated listed maturities, currently (the OTC market might give more information). Differences in the the details of the blending scheme thereafter will not effect the ability of one OMM versus another to match or make the listed market.

In terms of our efforts to normalize the equity volatility market, agreeing to quote all volatilities in terms of, say, a $(2,4)$ blending scheme should therefore not create practical problems. If we ever get into a situation like the 2008 financial crisis again, where apparently many, perhaps most, options chains were priced for a while with mainly discrete proportional dividends, this issue will need some more thought.

4.2.5 Calendar Arbitrage

We now come to the question of arbitrage. Arbitrage that arises from "bad" volatility surfaces (or "bad" rate term-structures) will be discussed elsewhere. Here we restrict ourselves to discussing calendar arbitrage for American call options, i.e. whether the prices of American calls at fixed strike are sometimes decreasing as maturity T increases. We will consider all hybrid models, with flat rate and vol inputs (so in contrast to above we are not calibrating the models to each other). We know that if we could price with infinite accuracy there would be no arbitrage in HM2 (and HM0) in this situation. The partial hybrid model HM1 is well-known to have arbitrage across dividend dates, and for HM3 there might be arbitrage, since it does not come from a consistent SDE across terms.

However, in practice we are forced to price with finite precision and then even HM2 and HM0 might have calendar arbitrage. To study this question we looked at the "pretty hard" example from above, and calculated the maximum amount of calendar arbitrage across a wide range of strikes and closely-spaced maturities, and plot it as a function of the number of steps N in our tree. The results are shown in figure 10, where we used $N = 41, 65, 81, 101, 201, 401, 801, 1601$.

For HM1 the maximum arbitrage is large, about \$0.30, and hardly dependent on N, as expected. All other models have arbitrage that is at least a factor of 10 smaller, even for small N, and vanishes with increasing N. In more realistic examples, dividends will be smaller and we will not care about as large maturities (where terms closely bracketing a dividend date do not exist in the listed market). This is illustrated in figures 11 and 12. We see that with a lower dividend or a smaller maximum maturity of $T = 0.6$ there is absolutely no arbitrage in all the non-HM1 models for $N > 100$ or so; whereas for HM1 the arbitrage only decreases by a factor of about 2.

Another way to illustrate the problems with HM1 is to look directly at the prices of American calls at fixed strike, as a function of time to maturity. This is illustrated in figure 13 for the at-the-money strike $K = 100$.

We should point out, as should intuitively be clear, that the presence of calendar arbitrage in HM1 with constant volatility input depends strongly on what maturities are available. In the above examples we assumed all maturities with a separation of 0.01 years exist, out to some maximum maturity. If the maximum maturity allows up to three quarterly dividends, then the maximum calendar arbitrage is about 14 cents; cf. figure 12. If maturities are separated by 0.02, 0.04, or 0.08

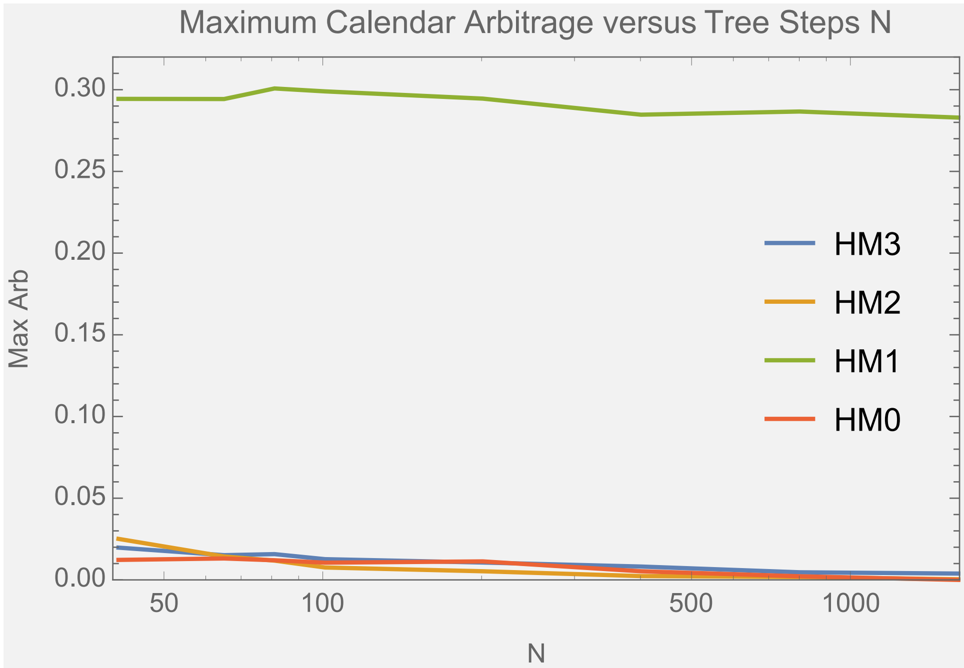

Figure 10: Maximum calendar arbitrage for American calls as a function of the number of steps in our tree, in various models, with $r=3\%, q=1\%, \sigma=30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1=0.085$. The maximum is calculated over strikes $K=80\ldots100$, with maturities sampled every 0.01 years up to $T=4.10$.

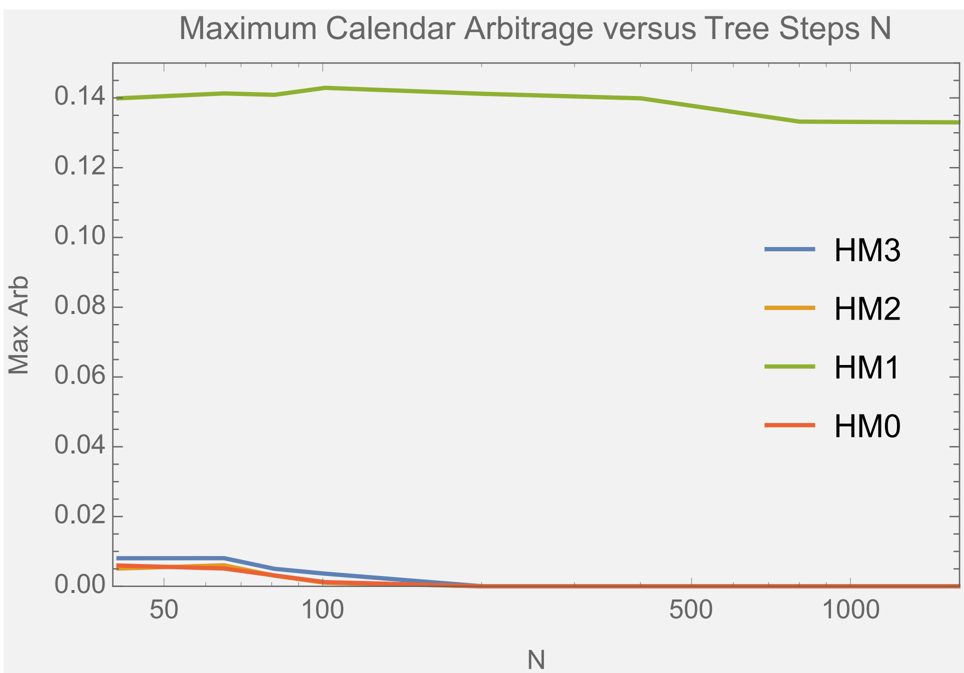

Figure 11: Maximum calendar arbitrage for American calls as a function of the number of steps in our tree, in various models, with $r=3\%, q=1\%, \sigma=30\%$, and a quarterly cash dividend of 1 first paid after about one month, $t_1=0.085$. The maximum is calculated over strikes $K=80\ldots100$, with maturities sampled every 0.01 years up to $T=4.10$.

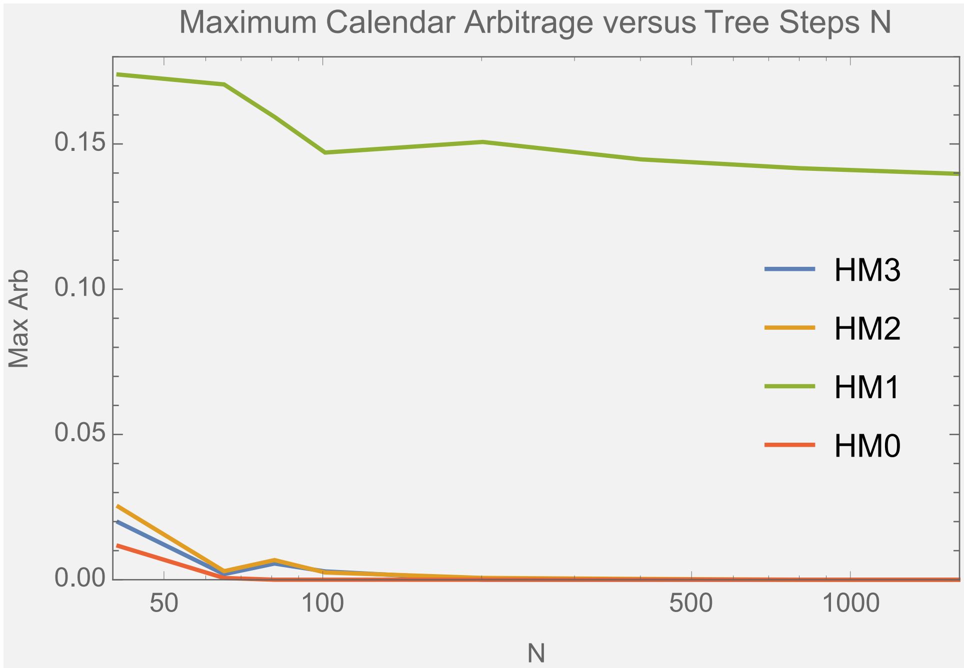

Figure 12: Maximum calendar arbitrage for American calls as a function of the number of steps in our tree, in various models, with $r=3\%, q=1\%, \sigma=30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1=0.085$. The maximum is calculated over strikes $K=80\ldots100$, with maturities sampled every 0.01 years up to $T=0.60$.

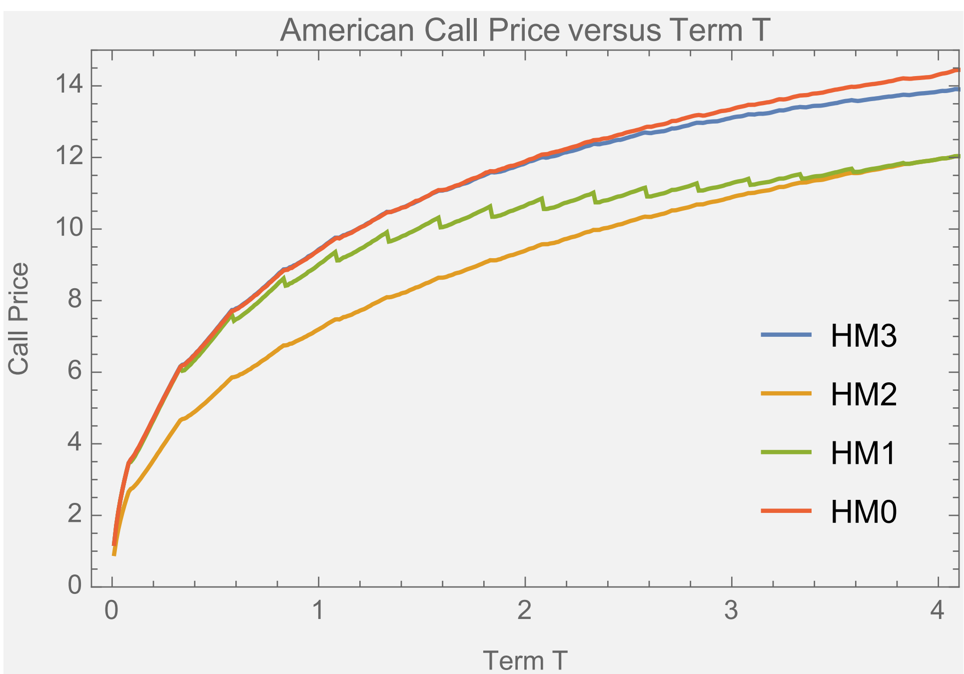

Figure 13: American call price with strike $K=100$ as a function of time to maturity T in various models, with $r=3\%, q=1\%, \sigma=30\%$, and a quarterly cash dividend of 2 first paid after about one month, $t_1=0.085$. The number of steps in our tree is $N=65$.

years, the maximum arbitrage possible would drop to about 12, 7 and 1 cent, respectively. If we allow up to five quarterly dividends, then these numbers increase again to about 20, 17, and 7 cents, respectively. Clearly, the amount of arbitrage one might encounter in practice when managing a smooth volatility surface and prices with the partial hybrid model would depend very sensitively on the precise combination of maturities one trades.[22] But it would be a lot easier to use a more sensible model to manage the volatility term-structure then to always have to worry about this.

One question we still have to answer is whether there is arbitrage in the SKA aka HM3 model. In the above plots, it behaves pretty much like all the other arbitrage-free models. We have numerically proved that it indeed can have calendar arbitrage in certain situations. But one must drag up quite unrealistic cases to find it. For example, if we restrict ourselves to terms $T \leq 1.1$ and sample every 0.01 years for fixed strikes $K = 80 \ldots 120$, then we have to increase the quarterly dividend in our "pretty hard" example to about $d = 3$ to find a maximum arbitrage of \$0.003 for asymptotically large N. The maximum arbitrage in HM1 in this situation is about \$0.32, i.e. about 100 times larger. We conclude that for practical purposes SKA has no significant calendar arbitrage.

5 Conclusion and Outlook

The equity options market does not have a standard for how to price vanilla options with cash dividends, and hence for how to define and calculate implied volatilities. Different market makers use different dividend pricing models, as well as more or less accurate and efficient pricing and calibration methodologies. Recall that this affects some of the most liquid options traded, from options on ETFs like SPY and QQQ to single stocks like AAPL. A similar set of issues surrounds the borrow costs and volatility curves used by different market participants. We believe this lack of consistency and transparency hinders the wider use of options, as well as the efficient transfer of volatility information across related products, and that the equity options market is due for some "RND": *Rationalization, Normalization and Democratization.* Perhaps it's ambitious, but we would like to see the day where equity option prices can be quoted in volatility terms for all underliers.

The current situation is perhaps comparable to that of the VIX before the "new" VIX was introduced. The calculation of the new VIX is based on a clear model-independent methodology involving only quoted prices [22], that anyone with basic financial and mathematical knowledge can reproduce. Since its introduction options and futures based on it have seen a huge growth in volume.

As a first step we have argued here that market participants should converge on using the hybrid model framework (shifted GBM) to price vanilla options with cash dividends (see table 1 for a comparison of the different models). It is technically cleaner than the spot model (piecewise GBM) – the supposedly more realistic features of the stochastic process underlying the spot model are of a rather formal nature (e.g. they don't help hedging against default risk; credit derivatives are needed for that) and not worth the trouble in practice, in our experience. The different hybrid models we discussed are all closely related, essentially analytically, on a mathematical and numerical level, despite their significant financial differences. In particular, they include the full hybrid model that can be made arbitrage-free with a smooth volatility surface, and can be used to consistently price and risk-manage options beyond the listed vanilla market. All the usual bells and whistles that an OMM might desire can easily be added (modeling of business time, parameter term-structure,

[22] In the listed US market there are now weekly, monthly, quarterly, end-of-month, and perhaps more types of option maturities available across various classes of underliers.

events, and even credit risk, which is currently not commonly done by OMMs). Each of the other hybrid models might have some desirable feature, to one market participant or another (it may just be the fact that a particular model was used in the past and traders got used to it). One of them, SKA, is a close cousin of the spot model, and essentially smoothly related to the full hybrid model. In any case, we recall, due to the relationships established here, they can all be accurately transformed to each other, and to an eventual common standard.[23]

This paper is the first in a series where we will try to "RND" various aspects of the equity options world. In one we will describe, in more detail than was possible here, how to implement efficient pricing of all the hybrid models [18]. In others we will discuss e.g. volatility curve fitting.

Acknowledgements: I would like to thank Leif Andersen, Misha Fomytskyi, Jim Gatheral, and Jiri Hoogland for discussions and comments on the manuscript.

References

[1] M. Overhaus et al, *Equity Hybrid Derivatives*, Wiley, 2007.

[2] H. Buehler, *Volatility and Dividends*, various versions, available at ssrn, since 2007.

[3] R. Roll, *An analytic valuation formula for unprotected American call options on stocks with known dividends*, Journal of Financial Economics, **5**, 251–258, 1977.
R. Geske, *A note on an analytic formula for unprotected American call options on stocks with known dividends*, Journal of Financial Economics, **7(A)**, 375–380, 1979.
R. Whaley, *On the valuation of American call options on stocks with known dividends*, Journal of Financial Economics, **9**, 207–211, 1981.

[4] V. Frishling, *A Discrete Question*, Risk, p.115–116, Jan. 2002.

[5] M. Bos, S. Vandermark, *Finessing Fixed Dividends*, Risk, p.157–158, Sep. 2002.

[6] R. Bos, A. Gairat, A. Shepeleva, *Dealing with Discrete Dividends*, Risk, p.109–112, Jan. 2003.

[7] E.G. Haug, J. Haug, A. Lewis, *Back to Basics: A New Approach to the Discrete Dividend Problem*, Wilmott Magazine, 37–47, Sep. 2003.

[8] M.H Vellekoop, J.W. Nieuwenhius, *Efficient Pricing of Derivatives on Assets with Discrete Dividends*, Applied Mathematical Finance, **13(3)**, 265-284, 2006.

[9] A. Buryak, I. Gao, *New Analytic Approach to Address Put-Call Parity Violation due to Discrete Dividends*, Applied Mathematical Finance, **19(1)**, 37–58, 2012 (We have to object to the title of this paper, as it refers to European options, where put-call parity always holds: The authors mean that the "usual" formula for the forward does not hold in the spot model.)

[10] T.R. Klassen, *Pricing Variance Swaps with Cash Dividends*, Wilmott Journal, **1(4)**, 173–177, 2009.

[23] But, to be sure, managing a smooth volatility surface will only be possible and provide an edge if one uses a definition of volatility that makes sense (at least according to the market consensus at the time).

[11] J. Hull, *Options, Futures, and Other Derivatives*, 6th edition, Pearson Prentice Hall, 2006.

[12] L. Andersen, M. Lake, D. Offengenden, *High Performance American Option Pricing*, working paper, 2015.

[13] D.P. Leisen, M. Reimer, *Binomial models for option valuation: examining and improving convergence*, Applied Mathematical Finance, **3**, 319–346, 1996.

[14] M. Joshi, *Achieving Higher Order Convergence for the Prices of European Options in Binomial Trees*, Mathematical Finance, **20(1)**, 89–103, 2010.

[15] M. Staunton, *Efficient estimates for valuing American options*, in: P. Wilmott (ed.) *The Best of Wilmott 2*, Wiley, 2005

[16] C. Wallner, U. Wystrup, *Efficient computation of option price sensitivities for options of American style*, Wilmott Magazine **2(11)**, 72–81, Nov. 2004.

[17] M. Joshi, *The Convergence of Binomial Trees for Pricing the American Put*, Journal of Risk, **11(4)**, 87–108, Summer 2009.

[18] T.R. Klassen, *Efficient Pricing of American Options with Discrete Dividends*, in preparation.

[19] M. Broadie, J. Detemple, *American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods*, Review of Financial Studies, **9(4)**, 1211–1250, 1996.

[20] R. Villiger, *Valuation of American Call Options*, Wilmott Magazine, 64–67, Mar. 2006.

[21] D. Cassimon, P.J. Engelen, L. Thomassen, M. Van Wouwe, *Closed-form valuation of American call options on stocks paying multiple dividends*, Finance Research Letters, **4**, 33–48, 2007.

[22] T.R. Klassen, J. Levin, S. Rattray, D. Shah, W.M. Speth (inventors), *Volatility index and derivative contracts based thereon*, US 20150039532 A1, patent pending, `http://www.google.com.ar/patents/US20150039532`.
See also the CBOE VIX white paper `https://www.cboe.com/micro/vix/vixwhite.pdf`.