A Cost of Capital Approach to Extrapolating an Implied Volatility Surface

B. John Manistre, FSA, CERA, FCIA, MAAA

Abstract[1]

This paper develops an option pricing model that takes cost of capital concepts as its foundation rather than dynamic replication. The resulting model, called the '*C*' measure in this paper, is related to the family of Affine Jump Diffusion models that are well known in the finance literature, so it is fairly easy to understand and implement. We argue that this is a reasonable model to use for estimating equity implied volatilities that are beyond the five- to 10-year horizon that can typically be observed in today's capital markets. The paper concludes with a short discussion of how to grade from observable market data to the '*C*' measure.

[1] The views and opinions expressed in this paper are those of the author and not the author's employer AEGON NV.

1. Introduction and Summary of Results

A problem faced by many life insurers today is that of putting a market-consistent value on liability instruments that are longer than any comparable assets for which market data is available. This raises the issue of extrapolating yield curves, equity implied volatilities, correlations and other data beyond the available information. This paper focuses on the implied volatility problem, but many of the ideas could be applied to the other problems as well.

One approach that immediately comes to mind is to take an industry standard model, a Heston model or a Heston model with jumps, for example, and use the available market data to calibrate the model's parameters. Having calibrated the first 10 or so years of the model we simply assume the model applies in the later years and use it. We will call this approach "simple extrapolation" in this paper.

This extrapolation method has the advantages of being simple to understand and easy to implement. A potential disadvantage is that the parameters that produce a reasonable fit to near-term market data may produce an unreasonable extrapolation.

The cost of capital method described here is a practical alternative to simple extrapolation. It is based on an approach endorsed by the CRO Forum for valuing non-hedgeable risk. The basic idea is that if a risk cannot be hedged in the capital markets, it should be valued using best estimate assumptions together with sufficient risk margins to pay for the cost of holding an appropriate amount of economic capital for the risk.

This new approach essentially derives a long-term implied volatility assumption from first principles. In most situations it will do a better job of fitting observed volatility surfaces than the Black-Scholes model but, only by chance, will it produce a good fit to observed market data. Additional work is then needed to engineer a model that fits both the current market and then grades into the longer-term implied volatilities described here.

This paper will use a three-step process to develop a cost of capital model, which we illustrate here using both a "normal" insurance risk, mortality and option pricing risk.

1) Develop a best estimate model. In the case of mortality risk this could be a deterministic or stochastic mortality table. For the option pricing problem, we will take best estimate to mean the P, or real-world, measure. Let σ be our best estimate of long-term volatility and let μ be our best estimate of long-term drift. These might come from a simple lognormal model or something more sophisticated.

2) Acknowledge the fact that even if the best estimate model is correct we can still have bad experience in any given year. The occurrence of a serious flu pandemic would be an example in the mortality case. Economic capital and margins need to be established to cover this risk. We will refer to this as contagion risk.

   For the option pricing problem, we will take the analog of a contagion event to be an instantaneous jump in the equity markets $S \to JS$. Here J would be 60 percent if we were holding capital cover a 40 percent drop in the equity markets. One of this paper's key results is that the impact of adding contagion risk to the option pricing model is to raise long-term implied volatilities from the best estimate: σ to $\sqrt{\sigma^2 + 2(\mu - r)(J - 1 - \ln J)/(1 - J)}$. Here r is a long-term interest rate so $\mu - r$ is the long-run equity premium.

3) As time evolves new information can arrive, which causes us to revise one or more of the parameters in our best estimate model. We must hold economic capital and margins to cover a plausible shock to the best estimate model. Using mortality as an example, this could be a shock to the mortality level or improvement trend. This will be called parameter risk in this paper.

   If our best estimate equity volatility were σ, then this might entail holding capital to cover a jump in this parameter to something like $\sqrt{\sigma^2 + \Delta\sigma^2}$ or higher. Adding parameter risk to the model increases long-term implied volatilities over and above the result obtained from contagion risk, although we argue that contagion risk is the more material issue.

The main goal of this paper is to show how this three-step approach can be applied to develop an option pricing model. Once the model is developed, and its properties understood, we argue that it is a reasonable approach to valuing options that can't be hedged directly in the capital markets. The final step of the paper is to consider the practical problem of grading from a market-calibrated model to a cost of capital model.

Before going into the option pricing issues in more detail, it is appropriate to point out that all cost of capital models are vulnerable to the criticism that it is not always clear how a given parameter or assumption should be shocked . For simple life insurance or annuity products, it is clear that mortality rates should be shocked up or down as the case may be. However, it is possible to engineer a product with a mix of mortality/longevity issues such that the nature of the risk varies by contract duration or possibly even market conditions. In this more general situation, rigorous application of the cost of capital principles can lead to problems that require stochastic control concepts for their solution.

For the option pricing problem, the analog of the mortality/longevity conundrum described above is whether we have a long/short equity exposure or whether the volatility exposure is convex/concave. Since most life insurers are long equity exposure and have convex liabilities, there is a wide range of practical applications where this fundamental conundrum is not an issue. Problems requiring stochastic control concepts are therefore outside the scope of this paper.

2. Step 1: The Best Estimate Model

We take the *P*, or real-world, measure to be the analog of a best estimate mortality table. This could be a very simple model such as the standard lognormal stock process $dS = \mu S dt + \sigma S dz$ or something more sophisticated such as an affine jump diffusion model. For simplicity of exposition, we will use the the standard lognormal model as a starting point in the examples that follow.

A high-level formula for a best estimate value could be written as:

$$V_0(t,S) = E_P[PV \text{ "}Cash\ Flows\text{"}]$$

where present values are calculated using an appropriate risk-free rate and *Cash Flows* are the projected cash flows of the instrument being valued.

For a vanilla put or call option, a more mathematical formulation of the above idea is that the value $V_0(t,S)$ satisfies the partial differential equation:

$$\frac{\partial V_0}{\partial t} + \mu S \frac{\partial V_0}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_0}{\partial S^2} = rV_0 . \tag{1}$$

Here *r* is the risk-free rate, and appropriate boundary conditions at maturity of the option must be specified.

3. Step 2: Contagion Risk

The analog of a contagion event for the option pricing problem is a finite jump $S \to JS$ where the jump factor *J* would be 60 percent if one wanted to hold capital for a 40 percent drop in the equity markets.

The Responsible Speculator

Let's start by taking the point of view of a speculator who doesn't delta hedge. A responsible speculator should hold enough economic capital to cover the loss that would occur if such a market jump occurred. A high-level formula that captures this idea is:

$$\begin{aligned} V\ (t,S) &= E_P[PV \text{ "}Cash\ Flows\text{"+"}Cost\ of\ \ Capital\text{"}] \\ &= E_P[PV \text{ "}Cash\ Flows\text{"+}\pi\{V(t,JS) - V(t,S\}] \end{aligned}$$

This value differs from the best estimate in that it adds in the cost of holding capital for the jump risk. Here π is the cost of capital or risk premium that an investor expects to receive for putting up the risk capital. We can think of this value as a sum of the best estimate value $V_0(t,S)$ plus risk margins.

The formula is intended to capture the idea that the cost of holding capital is being captured at all future points in time and market conditions. If the assumptions underlying the *P* measure model come true, then an investor putting up the risk capital would earn, on average, the risk-free rate plus the cost of capital π .

The model also assumes that gains and losses are continuously being trued up as time evolves. Gains are immediately paid out to the investor, and losses are immediately replaced. The investor is willing to replace losses, or put up additional required capital, because there is always sufficient margin left on the balance sheet to guarantee a reasonable future return on the newly invested capital.

As written, the above formula is not very practical because it defines value in a circular fashion. It turns out that this valuation problem has a very practical solution. We will define a new risk-adjusted process, called the *C* measure here, which is the *P* measure process augmented by a jump process $S \rightarrow JS$ where the instantaneous probability of a jump occurring in a small time interval *dt* is equal to the cost of capital rate π multiplied by the time interval.

A more mathematical way of seeing how the *C* measure concept comes about is to write down the basic valuation equation analogous to equation (1) above:

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

and then simply move all capital terms to the left-hand side to get:

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

The way to interpret equation (2) is to say that, in the real world, we expect the quantity *V* to grow at the risk-free rate while releasing sufficient margin to pay for the cost of capital. The mathematically equivalent formula (3) says that the quantity *V* has an expected rate of change equal to the risk-free interest rate in the *C* measure world where the dynamics of the stock price are given by

$$dS = \mu S dt + \sigma S dz + (J-1)dN, \quad \Pr[dN = 1] = \pi dt.$$

In terms of the *C* measure, the value *V* defined above can then calculated as the expected present value:

$$V\ (t, S) = E_C[PV\ "Cash\ Flows"].$$

A minimal requirement for such a model to be market-consistent is that it price the stock process *S* back to itself. Mathematically, this means that the function $V(t,S) = S$ must be a solution of (3) above. Going through the mechanics we find

$$\mu = r - \pi(J-1),$$
$$\Rightarrow \pi = (\mu - r)/(1-J).$$

This return makes sense from the perspective of investors putting up the risk capital since they have a leveraged exposure to the equity risk. The investor puts up a fraction $(1-J)$ of the equity position but takes 100 percent of the risk associated with that position.

As a simple example, assume the equity premium is $(\mu - r) = 4\%$ and $J = 60\%$, then the cost of capital for this risk would be $\pi = (\mu - r)/(1-J) = .04/(1-.6) = 10\%$.

The C measure model described above is actually a special case of Merton's (1973) Jump Diffusion model. There are many well documented[2] technical tools available for working with this model. In particular, for a vanilla put or call option that expires at time T, the relationship between the value $V(t,S)$ described here and the best estimate value $V_0(t,S)$ is given by:

$$V(t,S) = \exp[-\pi(T-t)] \sum_{n=0}^{\infty} \frac{[\pi(T-t)]^n}{n!} V_0(t, J^n S) . \tag{4}$$

This formula is easy to program, so it is not hard to generate useful examples once a P measure model is chosen. We'll get more insight into what this means in the next section.

A reasonable criticism of the model described above is that it only appears to make sense for an instrument where $V(t,JS) - V(t,S) > 0$, i.e., a put option or a long cash position. The next section will show why this model actually makes sense for any instrument that is convex.

Some additional insight into the model can be gained by considering the limit in which the jump size goes to zero i.e. $J \to 1$. In this limit the cost of capital rate $\pi = (\mu - r)/(1-J)$ goes to infinity while the economic capital amount $[V(t,JS) - V(t,S)]$ itself goes to zero. The limiting dollar cost of capital does have a finite limit since

$$\begin{aligned} \frac{\mu - r}{1-J}[V(t,JS) - V(t,S)] &= -(\mu - r) S \frac{V(t,JS) - V(t,S)}{JS - S}, \\ &\to -(\mu - r) S \frac{\partial V}{\partial S}, \, as \; J \to 1. \end{aligned}$$

Putting this result in to the fundamental valuation equation (3) we find

$$\frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - (\mu - r) S \frac{\partial V}{\partial S} = rV .$$

Simplifying this equation shows that the no jump limit is just the Black-Scholes model

$$\frac{\partial V}{\partial t} + rS \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV ,$$

which we have just derived without making any assumptions about hedging.

The Responsible Hedger

We now take the point of view of someone who chooses to delta hedge an obligation. A delta hedger can hold less economic capital because a portion of the risk is hedged. However, the act of hedging puts the company in the Q measure, effectively changing the expected return from μ to the risk-free rate r. A high-level formula for this new situation would be:

[2] See, for example, Haug, E.G. 1997. *The Complete Guide to Option Pricing Formulas*. McGraw-Hill.

$$V(t,S) = E_Q[PV\text{ "}Cash\ Flows\text{"} + \hat{\pi}\{V(t,JS) - V(t,S) - (J-1)S\partial V/\partial S\}].$$

The valuation equation is now given by:

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = rV - \hat{\pi}[V(t,JS) - V(t,S) - (J-1)S\frac{\partial V}{\partial S}]. \qquad (5)$$

The responsible delta hedgers using this model are saying that their *Q* measure calculation is basically right but they still hold enough capital to cover the un-hedged loss that would occur if a single large movement actually occurred. Adding this term addresses the very common criticism of the Black-Scholes model that it ignores the possibility of large price movements. We also note that the adjustment makes the model more conservative as long as the instrument is convex.

We have used the symbol $\hat{\pi}$ for the cost of capital here because this risk is technically different from the one faced by the responsible speculator. In particular, this model prices the stock process *S* back to itself no matter what we assume for $\hat{\pi}$.

As before, there is a new risk-adjusted measure that can be used to solve the valuation problem posed above. The solution is to start with the *Q* measure $dS = rSdt + \sigma Sdz$ and make two adjustments:

1. Change the drift from the risk free rate r to $r + \hat{\pi}(1-J)$.
2. Add jumps $S \to JS$ with intensity $\hat{\pi}$ as before.

The risk-adjusted stock process is now:

$$dS = (r + \hat{\pi}(1-J))Sdt + \sigma Sdz + (J-1)dN, \quad \Pr[dN = 1] = \hat{\pi}dt.$$

Again this is easily justified by rewriting equation (5) above with all capital terms on the left.

We now note that this model will agree with the *C* measure derived for the responsible speculator provided we assume the same cost of capital $\hat{\pi} = \pi = (\mu - r)/(1-J)$. We conclude that speculators and hedgers can agree on value if they each hold capital for their respective un-hedged risk and they agree that the cost of capital is $\pi = (\mu - r)/(1-J)$. In the author's opinion this is a compelling argument for the use of the *C* measure as defined here because it gives us a value which is independent of the assumed risk management strategy.

The chart below shows three simulated price processes. The first is a standard lognormal *P* measure scenario with $\mu = .08, \sigma = .15$. The second series is the corresponding Black-Scholes *Q* measure scenario assuming the risk-free rate is $r = .04$. The third series is an example of a *C* measure scenario, which mimics *P* measure changes at all points in time except when a jump occurs. In this particular scenario, jumps occur near years 8 and 21 with the result that the *Q* and *C* measures end up at very similar points after 25 years.

With the parameter choices made above, the cost of capital is $\pi = (.08-.04)/(1-.6) = 10\%$ and the expected number of jumps over a 25-year period is then 2.5. Observing exactly two jumps is therefore a relatively high-probability outcome.

We can get a high-level sense of how this model differs from the standard Black-Scholes model by starting with equation (5) above and then approximating the cost of capital by a Taylor series:

$$\pi[V(t,JS)-V(t,S)-(J-1)S\frac{\partial V}{\partial S}] \approx \frac{1}{2}\pi(J-1)^2 S^2 \frac{\partial^2 V}{\partial S^2}.$$

The valuation equation can then be rewritten as:

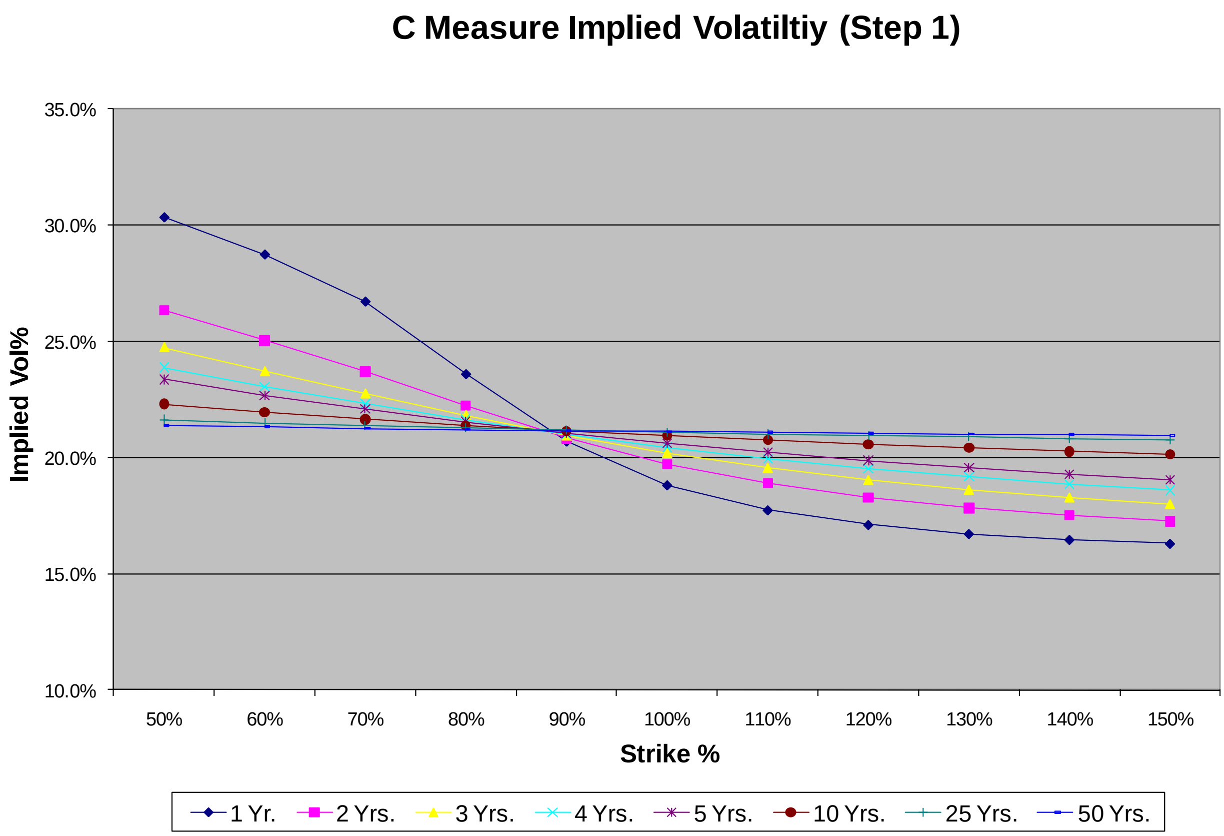
$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}[\sigma^2 + \pi(J-1)^2]S^2\frac{\partial^2 V}{\partial S^2} - rV \approx 0.$$

Adding this cost of capital to the Black-Scholes equation is therefore roughly equivalent to using the Black-Scholes model with an implied volatility given by the simple formula:

$$\begin{aligned} \sigma_{imp}^2 &\approx \sigma^2 + \pi(1-J)^2, \\ &= \sigma^2 + (\mu - r)(1-J). \end{aligned}$$

In practice, this approximation is not very good at short durations, and it slightly underestimates implied volatility at longer durations. A more detailed argument outlined in the appendix shows that $\sigma_{imp}^2 \approx \sigma^2 + 2\pi(J - 1 - \ln J)$ is a better formula approximation.

The graph below shows the entire volatility surface for the C measure model using the yield curve at Dec. 31, 2008, a dividend rate of 2 percent, a jump factor of 60 percent and a cost of capital equal to 10 percent. For P measure, the graph assumes the standard lognormal model with $\mu = .08, \sigma = .15$.



For each maturity and strike, put option values were calculated using the series expansion (4). We then solved for the volatility assumption that would produce the same value using the Black-Scholes model.

Three points worth noting at this stage of the model's development are:

- The model exhibits the phenomenon of "skew" at shorter maturities where implied volatilities are a decreasing function of the strike price when the option is close to being at the money.

- The surface is almost flat by the time we are out 50 years. The implied volatility varies from 21.4 percent to 21.0 percent as the strike price ranges from 50 percent to 150 percent. This is consistent with the formula approximation

$$\sigma_{imp}^2 \approx (.15)^2 + .10 \times 2(J - 1 - \ln J) \approx (.21)^2$$

described earlier. The model has added a risk premium of about 600 vol points.

- The model does not exhibit the kind of "smile" that is observed in actual volatility surfaces. One reason for this is the use of the lognormal model as our best estimate. Had we started with a stochastic volatility model—the Heston model, for example—the resulting *C* measure model would exhibit more smile at shorter durations.[3]

4. Step 3: Parameter Risk

The model summarized briefly above has two key parameters: the best estimate volatility σ and the long run equity premium $\mu - r$. If there were no uncertainty in the best estimate model, or its parameters, then $\sigma_{imp}^2 \approx \sigma^2 + 2(\mu - r)(J - 1 - \ln J)/(1 - J)$ might be an appropriate long-run implied volatility assumption as determined by the cost of capital method.

A wide range of models have been proposed to explain real-world volatility. In the finance literature, the Heston model is well-known; while regime-switching models[4] are common in the actuarial literature. No matter what model we pick, we have to estimate parameters, and, as time evolves, new information can arrive which causes us to change our parameter estimates.

The issue is illustrated by Figure 1, which shows realized volatility by calendar year for the S&P 500 for the 60-year period beginning in 1950. Daily data was used to calculate the realized volatilities.

---

[3] For more background on the concepts of skew and smile, see Gatheral, J. 2006. *The Volatility Surface*: *A Practitioner's Guide*. Wiley.
[4] Hardy, M.R. 2001. "A Regime-Switching Model of Long-Term Stock Returns." North American Actuarial Journal 5(2): 41–53.

Figure 1

S&P 500 Total Return Realized Volatility 1950-2009

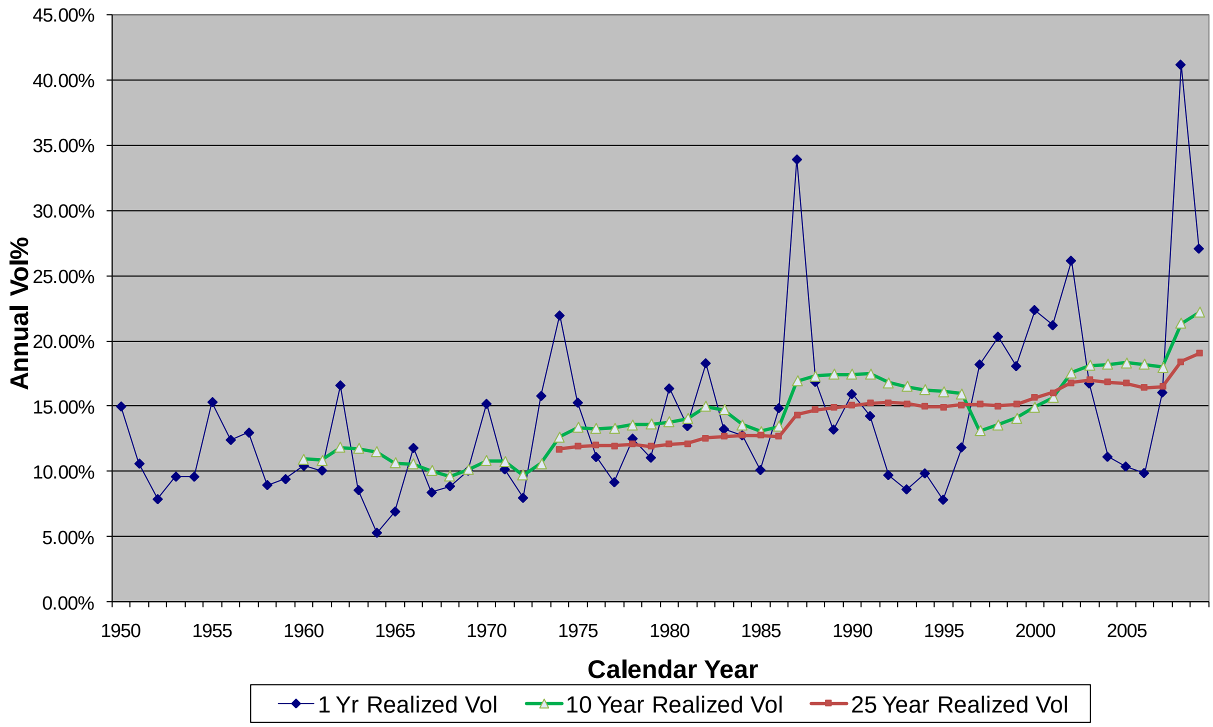


Figure 1 also shows the trailing 10 and 25-year realized volatilities. If we take the midpoint of the 10 and 25 year average numbers to be our "best estimate" of long run volatility then we could justify a value of $\sigma = .20$ at the end of 2009.

While there is no unique methodology for turning historical data into a best estimate assumption, it is clear that any reasonable combination of *P* measure model and parameter estimation method would have resulted in revisions to the long-term parameter assumptions as time evolved. Ten years ago $\sigma = .15$ would have been a very defensible number but the experience of the last few years would have forced us to revise that assumption.

One way to deal with this is to hold sufficient economic capital that we can cover the loss that would occur if the liability were revalued using a revised set of parameter values. We'll illustrate this idea in the case where the *P* measure model is just the standard lognormal. We have a best estimate long-term volatility assumption of σ^2, and we want to hold capital and margins to cover a shocked assumption $\hat{\sigma}^2 = \sigma^2 + \Delta\sigma^2$. Here $\Delta\sigma^2$ is a plausible (99.5 percent) shock to the volatility assumption that could result from new information arriving in the course of the next year.

As an example, suppose we are at the end of 2009 and our best estimate of $\sigma = .20$ is based on the prior 10-25 years of history. Assume 2010 were another year like the financial crisis, in 2008, when realized volatility jumped to 41 percent. A revised best estimate might then be something like:

$$\hat{\sigma}^2 = \frac{14(.20)^2 + (.41)^2}{15} \approx (.22)^2$$
$$\approx (.20)^2 + (.09)^2$$

This kind of analysis suggests that a plausible one-year shock in the range of $(.05)^2 < \Delta\sigma^2 < (.10)^2$ would be reasonable. For definiteness, we will take the 1 year 99.5% shock $\Delta\sigma^2$ to be such that $\hat{\sigma}^2 = \sigma^2 + \Delta\sigma^2 = (.225)^2 \Rightarrow \Delta\sigma = .103$.

Taking the responsible speculator's point of view, a high-level formula that captures this idea would be to write

$$V\ (t,S) = E_P[PV\ "Cash\ Flows" + \pi\{V(t,JS) - V(t,S\} + \tilde{\pi}\{\hat{V}(t,S) - V(t,S)\}].$$

Here $\hat{V}(t,S)$ is a value calculated using shocked volatility and $\tilde{\pi}$ is a new cost of capital appropriate for this new risk. If $\hat{P}$ is the shocked *P* measure, then we would compute the shocked value $\hat{V}(t,S)$ using

$$\hat{V}\ (t,S) = E_{\hat{P}}[PV\ "Cash\ Flows" + \pi\{\hat{V}(t,JS) - \hat{V}(t,S\} + \tilde{\pi}\{\hat{\hat{V}}(t,S) - \hat{V}(t,S)\}].$$

This formula assumes we take the same approach to contagion risk when calculating $\hat{V}(t,S)$ that we did in the base case. The formula also indicates that we would continue to hold capital $\{\hat{\hat{V}}(t,S) - \hat{V}(t,S)\}$ for parameter risk in the shocked world. The only reason to omit such a term from the calculation of $\hat{V}(t,S)$ would be because we thought the shocked parameter value $\hat{\sigma}^2 = \sigma^2 + \Delta\sigma^2$ could never get any worse. If this is not the case, then double, triple, …,*n* times shocked worlds need to be considered, in theory.

If this seems like over-engineering, we would agree, and we will shortly show how the technicalities can be simplified significantly. However, it is worth emphasizing why such a model structure makes sense. If the shocked value $\hat{V}(t,S)$ contains no margins, then we are not in a position to attract new capital if we found ourselves in the shocked world and the previous capital $\{\hat{V}(t,S) - V(t,S)\}$ was used up by an assumption change. In order to attract a capital infusion of $\{\hat{\hat{V}}(t,S) - \hat{V}(t,S)\}$, the shocked value $\hat{V}(t,S)$ must contain sufficient margin to compensate an investor for taking the risk.

The technical solution to the valuation problem we have just defined is a regime-switching model where volatility starts out equal to the best estimate value and then randomly jumps from one level to the next with a transition rate equal to the cost of holding capital for parameter risk.

In theory, this requires us to specify an infinite hierarchy of volatility levels $\sigma^2 \to \sigma^2 + \Delta\sigma^2 \to ... \to \hat{\sigma}^2_\infty$ along with a cost of capital $\tilde{\pi}$. The *C* measure introduced earlier is now augmented by allowing the volatility parameter to jump randomly up the shock hierarchy with transition intensity $\tilde{\pi}$.

Mathematically, this is equivalent to an infinite system of equations of the form

$$\frac{\partial V}{\partial t} + (r-q)S\frac{\partial V}{\partial S} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2} - rV = -\tilde{\pi}\left[\hat{V}(t,S) - V(t,S)\right],$$
$$\frac{\partial \hat{V}}{\partial t} + (r-q)S\frac{\partial \hat{V}}{\partial S} + \frac{\hat{\sigma}_1^2 S^2}{2}\frac{\partial^2 \hat{V}}{\partial S^2} - r\hat{V} = -\tilde{\pi}\left[\hat{V}^{(2)}(t,S) - \hat{V}(t,S)\right],$$
$$\frac{\partial \hat{V}^{(2)}}{\partial t} + \ldots$$

To calculate the shocked value $\hat{V}(t,S)$, we need a shocked *C* measure $\hat{C}$, which is the same as the *C* measure except the volatility assumption starts out in the first shocked level $\hat{\sigma}^2 = \sigma^2 + \Delta\sigma^2$ and then jumps randomly up the hierarchy from there.

In theory, the discussion above completes the description of the *C* measure. Had we started with a more sophisticated *P* measure model, the same basic ideas could have been developed except that the specific parameters subject to regime switching might be different from what makes sense in the standard lognormal model.

As a practical matter, we need a way to justify a set of assumptions and, if possible, simplify the actual calculations. Most practitioners would consider implementing the regime-switching model, as specified, to be over-engineering.

To simplify the regime-switching model, we make a new assumption.

> The shock hierarchy is assumed to have a simple geometric structure governed by a parameter $0 < \alpha < 1$. In terms of this parameter the *n*'th level in the volatility hierarchy is given by:
>
> $$\hat{\sigma}_n{}^2 = \sigma^2 + \Delta\sigma^2 \frac{1-\alpha^n}{1-\alpha} \ .$$

The parameter α is then chosen so that the ultimate volatility level $\sigma_\infty^2 = \sigma^2 + \Delta\sigma^2 /(1-\alpha)$ makes sense. The historical evidence shown earlier indicates that it is very difficult to get S&P 500 realized volatility over a 10-year, or longer, period to exceed 25 percent. This supports an alpha factor of about 50 percent. For definiteness, we set $\alpha \approx .53$ so that $\sigma_\infty^2 = (.25)^2$.

If the cost of capital is $\tilde{\pi}$, then the *C* measure expected squared forward volatility, *T* years from the valuation date, is[5]:

---

[5] This follows from the fact that number of regime changes *N*(*T*) has a Poisson distribution with mean $\tilde{\pi}T$ so that $E[\alpha^{N(T)}] = \exp[-\tilde{\pi}(1-\alpha)T]$. This is one reason for choosing the geometric hierarchy.

$$\bar{\sigma}^2(T) = \sigma^2 + \frac{\Delta\sigma^2}{1-\alpha}[1-\exp(-\tilde{\pi}T(1-\alpha))]$$

which corresponds to the spot volatility:

$$s^2(T) = \frac{1}{T}\int_0^T \bar{\sigma}^2(v)dv = \sigma^2 + \frac{\Delta\sigma^2}{1-\alpha}[1-\frac{1-\exp(-\tilde{\pi}T(1-\alpha))}{\tilde{\pi}T(1-\alpha)}].$$

This is a good approximation to the regime-switching model's implied volatility for all but very short maturities.

A more formal way to see why this approximation works is to imagine that, in the *C* measure, the variance process $v = \sigma^2$ follows a mean reverting deterministic process of the form

$$dv = \tilde{\pi}(1-\alpha)(v_\infty - v)dt, v(t) = \sigma^2.$$

Here $v_\infty = \sigma^2 + \frac{\Delta\sigma^2}{1-\alpha}$ is the ultimate variance level under the regime switching model. The valuation equation corresponding to this deterministic volatility assumption is

$$\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \tilde{\pi}(1-\alpha)(v_\infty - v)\frac{\partial V}{\partial v} + \frac{1}{2}vS^2\frac{\partial^2 V}{\partial S^2} = rV,$$

which can also be written as

$$\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}vS^2\frac{\partial^2 V}{\partial S^2} = rV - \tilde{\pi}(1-\alpha)(v_\infty - v)\frac{\partial V}{\partial v}.$$

In the *P* measure, where $v = \sigma^2$, we can interpret the last term as a cost of capital. In particular,

$$\begin{aligned}
\tilde{\pi}(1-\alpha)(v_\infty - v)\frac{\partial V}{\partial v}\bigg|_{v=\sigma^2} &= \tilde{\pi}(1-\alpha)[\sigma^2 + \frac{\Delta\sigma^2}{1-\alpha} - \sigma^2]\frac{\partial V}{\partial v}\bigg|_{v=\sigma^2}, \\
&= \tilde{\pi}\Delta\sigma^2\frac{\partial V}{\partial v}\bigg|_{v=\sigma^2}, \\
&\approx \tilde{\pi}[V(t,S,\sigma^2+\Delta\sigma^2) - V(t,S,\sigma^2)].
\end{aligned}$$

Our short cut deterministic volatility model can therefore be viewed as an approximation to the exact regime switching model, or as an exact solution to a model where the economic capital is approximated by $\Delta\sigma^2\frac{\partial V}{\partial v}$. As noted earlier, the approximation is very good for the longer option maturities of interest to us in this paper and is very easy to implement.

If we now put the contagion and parameter risk pieces together, we find that, for long-dated options, the cost of capital model suggests the following approximate formula for implied volatility:

$$\begin{aligned}\sigma_{imp}^2 &\approx \sigma^2 + 2\pi(J-1-\ln J) + \frac{\Delta\sigma^2}{1-\alpha}, \\ &= \sigma_\infty^2 + 2\pi(J-1-\ln J) \\ &= (.25)^2 + 2(.10)(.60-1-\ln(.60)) \\ &= (.29)^2\end{aligned}$$

The end result of the process would appear to add roughly 900 vol points to the best estimate volatility of 20 percent. As Table 1 shows, it takes more than 50 years for this long-term implied volatility to be reached if we assume a reasonable cost of capital for parameter risk.

Each line in Table 1 is designed to illustrate the impact of a specific set of *C* measure parameters on at-the-money (ATM) implied volatilities for maturities of 10, 25 and 50 years.

Table 1

	Contagion Shock			Parameter Shock			At the Money Implied Vol %		
	σ	*J*	*π*	*Δσ*	*α*	*π~*	10	25	50
1	20.0%	60.0%	10.0%	10.3%	53.0%	6.0%	25.3%	26.0%	26.8%
2	20.0%	60.0%	10.0%	0.0%	53.0%	6.0%	24.7%	24.8%	24.8%
3	22.5%	60.0%	10.0%	10.3%	53.0%	6.0%	27.3%	28.0%	28.7%
4	20.0%	50.0%	8.0%	10.3%	53.0%	6.0%	26.7%	27.5%	28.2%
5	20.0%	60.0%	15.0%	10.3%	53.0%	6.0%	27.2%	28.0%	28.7%
6	20.0%	60.0%	10.0%	12.0%	53.0%	6.0%	25.5%	26.5%	27.5%
7	20.0%	60.0%	10.0%	10.3%	75.0%	6.0%	25.3%	26.2%	27.2%
8	20.0%	60.0%	10.0%	10.3%	53.0%	10.0%	25.6%	26.6%	27.4%

The first line of Table 1 shows what happens if we assume that a risk premium of 6 percent is appropriate for parameter risk. This rate seems reasonable given that parameter risk is not leveraged like contagion risk, and so it is more like an un-hedgeable insurance risk.

With a 6 percent probability of stepping up the shock hierarchy each year, we have experienced, on average, only three regime changes in 50 years. This would lead to an expected forward volatility of about 24 percent and a spot volatility of 22.5 percent under the regime-switching model. A formula estimate for the full model's implied volatility at the 50 year point is therefore $\sqrt{(.225)^2 + 2(.10)(.6-1-\ln(.6))} = .270$ which is close to the more accurate value of .268 in line 1 of the table.

The last line in Table 1 shows what happens if we increase the cost of parameter risk capital from 6 percent to 10 percent. The effect is not huge. In fact, an important high-level conclusion is that contagion risk issues are more important than parameter risk issues.

Line 2 in the table simply turns off the parameter shock. This shows that the impact on 50-year implied volatility of adding parameter risk is about 200 vol points which is less than the 250 vol point shock to best estimate volatility. The reason we get a smaller number is because there is a diversification benefit when we aggregate parameter and contagion risk.

Line 3 shows what happens if we change the underlying best estimate volatility from 20 percent to 22.5 percent. Fifty-year implied volatilities go up by about 190 vol points. There is a diversification effect at work here, which is why the implied volatility did not go up by 250 vol points.

Line 4 shows what happens if the contagion capital shock is changed from 40 percent to 50 percent while we leave the equity risk premium fixed at 4 percent. The leveraged cost of capital is now 8 percent. This does increase implied volatilities in a material way though not as much as in Line 3.

Line 5 shows what happens if we change the cost of contagion risk capital while leaving the capital shock at 40 percent. This is equivalent to assuming that the equity risk premium is raised from 4 percent to 9 percent. This produces results very similar to Line 3.

Lines 6,7 and 8 test the model's sensitivity to changes in the parameters controlling the cost of parameter risk. Not surprisingly, the first level shock in Line 6 is the most significant issue.

5. Step 4: The Case for the *C* Measure

The first argument to support the use of the *C* measure is that it represents a cost of manufacturing an option that depends only on a small number of fundamental economic assumptions. All of these assumptions are fairly transparent and easily subjected to scrutiny. Furthermore, this has been done within the context of high-level principles already accepted by the CRO Forum as reasonable for valuing non-hedgeable risk.

The model does not assume delta hedging, but it is not inconsistent with delta hedging either. In fact, the model addresses one of the principal criticisms of the Black-Scholes model in that it recognizes that delta hedging, if implemented, is an imperfect process.

The technicalities of the model are not onerous. Anyone familiar with basic Black-Scholes concepts and Merton's Jump Diffusion extension can come to grips with the technical details at whatever level is necessary.

It is not necessary to take the standard lognormal model as the starting point. This paper takes this approach to simplify the presentation of the relevant new ideas. Once can easily start with a more sophisticated *P* measure model and then adjust it for contagion and parameter risk as we have done here.

6. Step 5: Grading from a Market-Calibrated Model to the *C* Measure

The model described in this paper can be used in a number of different ways. The simplest way is to use the ideas developed here to justify the choice of parameters in some other model. How this works in practice will depend on the other model being used, so we can't comment further here.

At the other end of the spectrum, one can imagine trying to come up with a version of the *C* measure model that can be calibrated to market data. Some of the challenges that must be addressed can be seen in the short history of 10-year ATM options in Table 2.

Table 2
S&P 500 10 Year
ATM Implied Vol

6/30/2010	35.6%
6/30/2009	31.1%
12/31/2008	34.9%
12/31/2007	27.5%
12/31/2006	22.4%
12/30/2005	24.0%
12/31/2004	19.7%
12/31/2003	18.0%

These volatilities have not borne much resemblance to the *C* measure estimate of 25.3 percent since 2007. More generally, the volatility surfaces that have been observed in the market during the last few years exhibit more skew and smile than can be explained by the simple *C* measure model discussed in this paper.

If a high degree of fit is required, then we would need to add elements such as stochastic volatility and/or jumps to the *P* measure model before we started adjusting for the cost of capital. See Chapter 5 of Gatheral[6] for more details of how this might work.

If we have a more limited objective, matching ATM implied volatilities only, for example, then a simpler approach is possible. One idea is to start with a slightly more sophisticated *P* measure model and then go through the same cost of capital risk adjustment steps described earlier. We will use the continuous time limit of a GARCH stochastic volatility model, as described by Heston & Nandi (2000)[7]. In this model the variance $v = \sigma^2$ follows a mean reverting stochastic process. The stock price and variance processes are assumed to be

$$dS = S(r + \lambda v)dt + \sqrt{v}Sdz,$$
$$dv = \kappa(\bar{v} - v)dt + \xi\sqrt{v}dw,\ dwdz = \rho dt.$$

In this model the quantities $\bar{v}, \kappa, \lambda, \xi, \rho$ are assumed to be constants that must be estimated from empirical data. Notice that the equity premium is now given by $\mu - r = \lambda v = \lambda\sigma^2$.

A contagion event in this model would consist of a market jump $(S, v) \to (JS, v + \Delta v)$ and theory suggests we should consider subjecting all of the model's parameters to some kind of regime switching process. In practice, this will be simplified by considering separate jumps $S \to JS$,

[6] Gatheral, J. 2006. *The Volatility Surface A Practitioner's Guide*. Wiley.
[7] Heston, N.L., Nandi,S., "*A Closed-Form GARCH Option Valuation Model*", Review of Financial Studies (2000).

$v \to v + \Delta v$ and, since $\bar{v}$ is clearly the most important parameter, it is the only one that will be dynamic in the risk adjusted *C* measure. The speculator's form of the valuation equation for $V = V(t, S, v, \bar{v})$ is then

$$\frac{\partial V}{\partial t} + (r + \lambda v) S \frac{\partial V}{\partial S} + \kappa(\bar{v} - v)\frac{\partial V}{\partial v} + \frac{vS^2}{2}\frac{\partial^2 V}{\partial S^2} + \rho \xi v \frac{\partial^2 V}{\partial v \partial S} + \frac{\xi^2 v}{2}\frac{\partial^2 V}{\partial v^2} - rV$$
$$= -\tilde{\pi}(1-\alpha)(\sigma_\infty^2 - \bar{v})\frac{\partial V}{\partial \bar{v}} - \frac{\lambda v}{1-J}[V(t, JS, v, \bar{v}) - V(t, S, v, \bar{v})] - \gamma v \frac{\partial V}{\partial v}\Delta v$$

The first two terms on the right hand side of the equation above are directly analogous to what was done for the simple lognormal model. The cost of capital for the equity jump must be $\pi = \lambda v/(1-J)$ in order for the model to re-price the stock.

We have simplified the variance jump term by assuming the cost of capital for this risk is γv and by using the approximation

$$V(t, S, v + \Delta v, \bar{v}) - V(t, S, v, \bar{v}) \approx \frac{\partial V}{\partial v}\Delta v.$$

Collecting like terms we get the hedger's form of the equation

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + (\kappa - \gamma \Delta v)\left(\frac{\kappa}{\kappa - \gamma \Delta v}\bar{v} - v\right)\frac{\partial V}{\partial v} + \frac{vS^2}{2}\frac{\partial^2 V}{\partial S^2} + \rho \xi v \frac{\partial^2 V}{\partial v \partial S} + \frac{\xi^2 v}{2}\frac{\partial^2 V}{\partial v^2}$$
$$= rV - \tilde{\pi}(1-\alpha)(\sigma_\infty^2 - \bar{v})\frac{\partial V}{\partial \bar{v}} - \frac{\lambda v}{1-J}[V(t, JS, v, \bar{v}) - V(t, S, v, \bar{v}) - (J-1)S\frac{\partial V}{\partial S}].$$

From the equation above we see that the impact of adding a variance shock to the model is to risk adjust the variance dynamics from the *P* measure $dv = \kappa(\bar{v} - v)dt + \xi\sqrt{v}dw$ to a new process $dv = \kappa'(\bar{v}' - v)dt + \xi\sqrt{v}dw$ where $\kappa' = \kappa - \gamma \Delta v$ and $\bar{v}' = \bar{v}\kappa/(\kappa - \lambda \Delta v)$. We still have a mean reverting variance process but the speed of mean reversion has been reduced and reversion target itself has been grossed up.

The model above is technically more involved than the simpler Black-Scholes-Merton model we have been working with so far. However, the above model falls within the affine jump-diffusion family of models for which there are industry standard techniques[8] available to get numerical results.

The chart that follows in figure 2 is based on the above model using the following specific parameter choices.

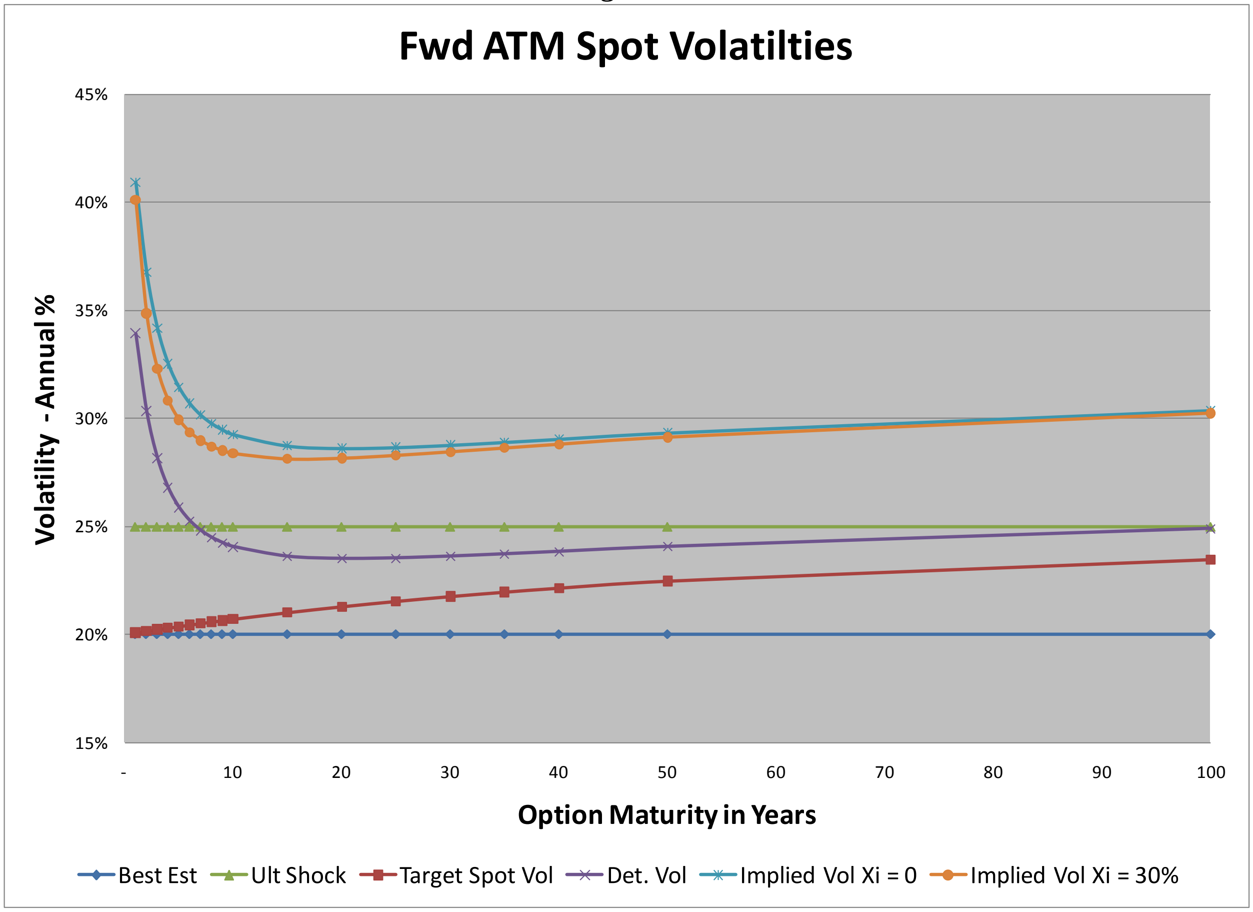
For the P measure we assume:

$r = .04,\ \lambda = 1.00,\ q = .02,\ \kappa = 1.20,\ \xi = .30, v(t) = (.40)^2, \bar{v}(t) = (.20)^2$

and the risk adjustment related parameters are given by;
$\sigma_1 = .225,\ \sigma_\infty = .25, J = .6, \Delta v = (.40)^2 - (.20)^2 = .12, \gamma = 1.0$

[8] This model can be solved using Fourier Transform techniques.

Figure 2



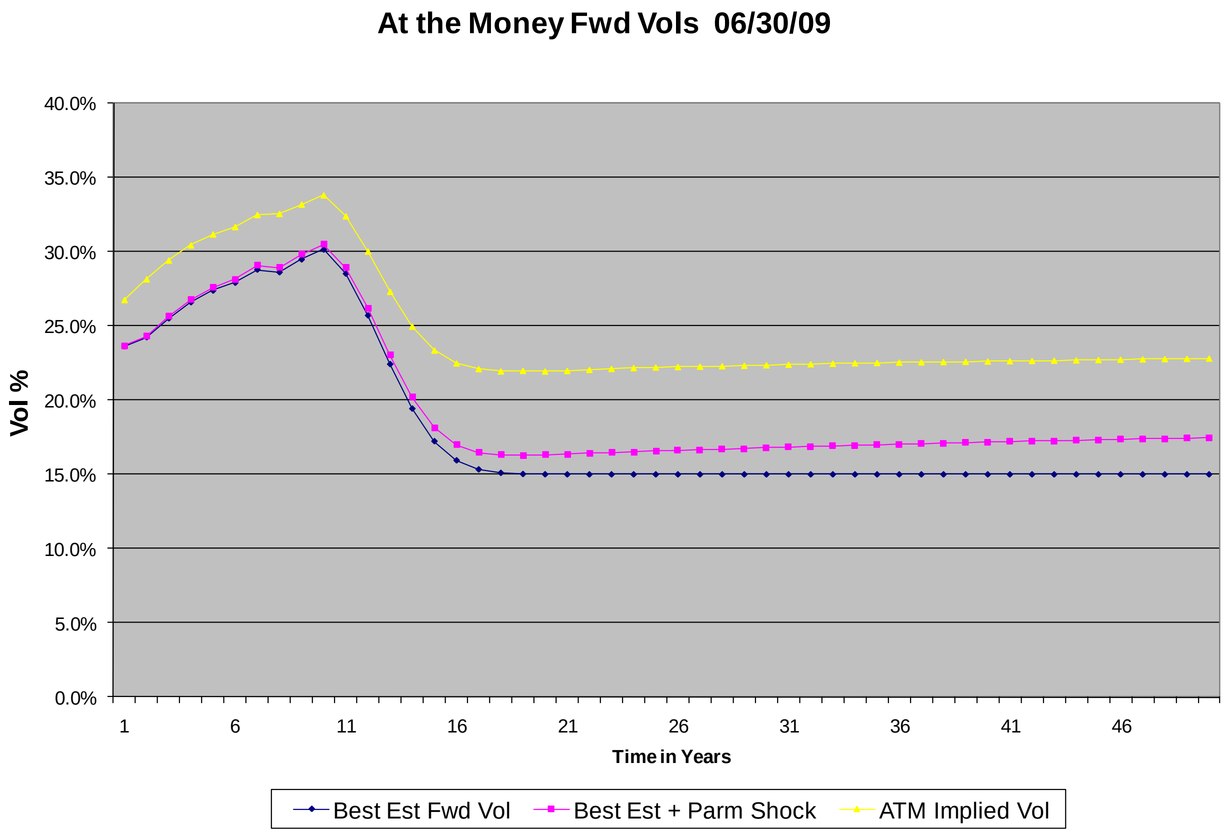
The first two horizontal lines on the chart show the best estimate and ultimate shock assumptions to the $\sigma = .20$ and $\sigma_\infty = .25$. The next series show the target spot volatility which starts at $\sigma = .20$ and then grades toward $\sigma_\infty = .25$. The fourth series is the deterministic spot volatility derived by assuming $\xi = 0$ so that the forward variance evolves according to

$$\begin{aligned} dv &= (\kappa - \gamma\Delta v)\left(\frac{\kappa\bar{v}}{\kappa - \gamma\Delta v} - v\right)dt, \\ &\approx (\kappa - \gamma\Delta v)((1.05)^2\bar{v} - v)dt. \end{aligned}$$

This variance is not grading toward $\sigma_\infty = .25$ but rather $1.05\sigma_\infty = .264$ it just has not got there.

The final two series represent at the money implied volatilities for the full model under the assumptions that $\xi = 0$ and $\xi = .30$ respectively. For short maturities the deterministic variance $\xi = 0$ model is more conservative than the stochastic variance model. As the maturity grows the two models get closer. We conclude that assuming $\xi = 0$ is a reasonable simplification for extrapolation purposes.

Figure 2



There is no real science in the choice of grading scheme, but most people would probably find grading over a five- to 10-year period reasonable.

An alternative is to allow the forward rate assumption to jump discontinuously from the last market determined rate to *C* measure value. This may seem extreme, but it is consistent with the idea of using the *C* measure to put a value, at time 10, on all cash flows beyond the market's horizon. Market discounting is then used to discount those values from time 10 to the valuation date.

While neither of the two approaches outlined above is "right," they do have different risk management implications. Allowing discontinuous forward rates will generally lead to values that are less volatile and more easily hedged with real market instruments.

7. Conclusion

This paper has developed an option pricing model from first principles using cost of capital concepts as a foundation. In the author's opinion, the most useful aspect of the model is its ability to defend a long-term implied volatility assumption. This long-term volatility depends on a small number of fairly transparent macroeconomic inputs and is independent of many modeling details.

We have also shown that modeling parameter risk naturally leads to a regime-switching model for the uncertain parameters, and we have given a simple example.

APPENDIX

The main purpose of this appendix is to derive the formula approximation for long-term implied volatility used in Step 2 of this paper:

$$\sigma_{imp}^2 \approx \sigma^2 + 2\pi(J - 1 - \ln J) = \sigma^2 + (\mu - r)(J - 1 - \ln J)/(1 - J).$$

The key idea is that, over longer periods of time, the law of large numbers allows us to approximate any reasonable *P* measure process by a lognormal model. Adding the simple jumps required by the *C* measure simply gives rise to a different lognormal approximation.

Suppose that, under the *P* measure, we have the approximation:

$$S(t) = S(0)\exp[(\mu - \sigma^2/2)t + \sigma Z(t)].$$

Going to the *C* measure means we now have:

$$\begin{aligned} S(t) &= S(0)\exp[(\mu - \sigma^2/2)t + \sigma Z(t)]J^{N(t)}, \\ &= S(0)\exp[(\mu - \sigma^2/2)t + \sigma Z(t) + \ln(J)N(t)]. \end{aligned}$$

Here *N*(*t*) is a Poisson process with mean $\pi t = (\mu - r)t/(1 - J)$.

Now use $\mu = r + \pi(1 - J)$ to write the stock process as

$$S(t) = S(0)\exp[(r - \sigma^2/2)t + \sigma Z(t)]\exp[\pi(1 - J)t + \ln(J)N(t)].$$

This shows that the *C* measure process is, roughly, the product of a Black-Scholes *Q* measure process with volatility σ and a second, independent, process $\exp[\pi(1 - J)t + \ln(J)N(t)]$. This second process always has a mean of 1, so we expect that, for large enough *t*, we should have an approximation of the form:

$$\exp[\pi(1 - J)t + \ln(J)N(t)] \approx \exp[-\tilde{\sigma}^2 t/2 + \tilde{\sigma}\tilde{Z}(t)], \qquad (*)$$

where $\tilde{Z}(t)$ is a new Weiner process independent of $Z(t)$. For large *t* we know that $N(t) \to \pi t + \sqrt{\pi}\tilde{Z}(t)$, which suggests that $\tilde{\sigma} = \sqrt{\pi}\ln(J)$ might be a good answer. This approximation suggests that $\sigma_{imp}^2 \approx \sigma^2 + \pi \ln(J)^2$ ought to be a reasonable long-term implied volatility. Practical testing shows that this formula tends to overstate the implied volatility for long-dated vanilla options.

A better approximation is obtained by choosing $\tilde{\sigma}$ so that the log means of each term in equation (*) above are equal, i.e., we choose $\tilde{\sigma}$ so that:

$$
\begin{aligned}
&E[\pi(1-J)t+\ln(J)N(t)]=E[-\tilde{\sigma}^2 t/2+\tilde{\sigma}\tilde{Z}(t)],\\
&\Leftrightarrow \pi(1-J)t+\ln(J)\pi t=-\tilde{\sigma}^2 t/2,\\
&\Rightarrow \tilde{\sigma}^2=2\pi(J-1-\ln J).
\end{aligned}
$$

In conclusion, this paper has actually derived three formula approximations for long-term implied volatility. These are:

$$\sigma_{imp}^2 \approx \sigma^2+\pi(1-J)^2 \le \sigma^2+2\pi(J-1-\ln J) \le \sigma^2+\pi\ln(J)^2.$$

Empirical testing shows that the middle formula works best for the implied volatility of vanilla options.

The other two expressions are reasonable approximations for the fair value of a variance swap contract under the *C* measure. Each formula applies to a slightly different definition of realized volatility.