University of Illinois, Urbana-Champaign *Professor Tim Johnson*
FIN 512 - Spring 2015 *tcj@illinois.edu*

Lecture Note 7.2: Model-Free Implied Volatility

This note tells the story of one of the most useful and creative achievement in financial engineering in the last 20 years.

As we have discussed many times, there is a fatal flaw in Black-Scholes implied volatility: it has no consistent interpretation in terms of the *true* volatility of the returns of the underlying variable unless that variable is constant – which it never is.

Yet it would be extremely useful for investors, risk manager, and regulators to be able to use the information in option prices to gauge current uncertainty. However, as we have seen, different models of uncertainty (i.e., different specifications of stochatic volatility) will lead to different options prices. Therefore the same option prices (from the market) coupled with different models will imply different levels of current volatility and expected uncertainty.

The beautiful discovery is that there is indeed a way to measure expected uncertainty for options prices that *does not depend on any stochastic assumptions!*

Outline:

I. The Log Contract
II. VIX
III. Example
IV. Extensions

I. The Log Contract.

- The story begins with an observation due to Anthony Neuberger in 1994.

- He asked the following question:
  - ▶ Suppose there is an asset whose current price is S_0 and whose time-T forward price is $F_{0,T}$ and we are short a derivative contract that will require us to pay $\log(S_T/F_{0,T})$ at time T.
  - ▶ Further suppose we choose to engage in a dynamic hedging strategy between now and T in which every Δt (e.g. every day) we buy $1/F_{t,T}$ forwards.
  - ▶ *How much money will we make?*

- This seems like a funny question. But it is pretty easy to answer. Let's just do the accounting.

- Here are our trades.

  Day 0: Buy $\frac{1}{F_{0,T}}$ forwards at price $F_{0,T}$.

  Day 1: Buy $\frac{1}{F_{1,T}} - \frac{1}{F_{0,T}}$ forwards at price $F_{1,T}$.

  $\vdots$ $\qquad\qquad\qquad \vdots$

  Day $\mathbf{T-1}$: Buy $\frac{1}{F_{T-1,T}} - \frac{1}{F_{T-2,T}}$ forwards at price $F_{T-1,T}$.

- When we get to time T all of these trades settle. Our profits from each trade are:

Day 0: $[\frac{1}{F_{0,T}}] \;\; (F_{T,T} - F_{0,T}) +$

Day 1: $[\frac{1}{F_{1,T}} - \frac{1}{F_{0,T}}] \;\; (F_{T,T} - F_{1,T}) +$

$\vdots \qquad\qquad\qquad \vdots$

Day $\mathbf{T-1}$: $[\frac{1}{F_{T-1,T}} - \frac{1}{F_{T-2,T}}] \;\; (F_{T,T} - F_{T-1,T})$.

- Please verify that adding all these terms up gives

$$\sum_{t=0}^{T-1} \left(\frac{F_{t+1,T}}{F_{t,T}} - 1 \right)$$

or

$$\sum_{t=0}^{T-1} \left(\frac{\Delta F_{t+1,T}}{F_{t,T}} \right).$$

- Now, what is the difference between the quantity in parentheses $\Delta F / F$ and $\Delta \log F$?
  - ▶ If we let Δt get very small so that we are talking about infinitessimals, then, for any diffusion process, we have

$$d \log(F) = \frac{1}{F}\, dF - \frac{1}{2}\sigma^2\, dt$$

from Ito's lemma.

▶ So, for small time intervals, we can replace $\Delta F/F$ by $\Delta \log F + \frac{1}{2}\sigma^2 \; \Delta t$.

▶ Notice that this is the *model free* step: I did not assume anything about the σ here.

  $*$ It could be stochastic – obeying any specification at all.

  $*$ It could even have jumps.

- When we make the substitution for $\Delta F/F$ in our hedging profit expression, our total P&L is

$$\sum_{t=0}^{T-1} \left(\Delta \log F_{t+1,T} + \frac{1}{2}\sigma_t^2 \; \Delta t \right)$$

which reduces to

$$\log(F_{T,T}) - \log(F_{0,T}) + \frac{1}{2} \sum_{t=0}^{T-1} \sigma_t^2 \; \Delta t$$

since all the intermediate log terms cancel out.

- Now the first terms here are the same as $\log(F_{T,T}/F_{0,T}) = \log(S_T/F_{0,T})$. And this is exactly the payout that we are short.

- So the total payoff at T from our position is just the second term:

$$\frac{1}{2} \sum_{t=0}^{T-1} \sigma_t^2 \; \Delta t \qquad \text{or} \qquad \frac{1}{2} \int_0^T \sigma_t^2 \, dt.$$

- Notice also that the $1/F$ hedging strategy that we engaged in was self-financing. In fact, it was costless.

- Thus, a short position in the log contract plus a costless strategy generates a payoff that is equal to (half) the realized variance between now and T.

- **Neuberger's conclusion:** The market value of the log contract must be equal to the minus one half times the market value of a claim to the realized total variance.

- Or the value of the log contract must be $-T/2$ times the market value of a claim to the realized average variance.

- The (forward) value of a claim to the realized average variance <u>is</u> what we mean by model free implied volatility (squared).

- Since volatility risk can be hedged, we can also express this definition as

$$MFIV^2 = \frac{1}{T}\, \mathrm{E}_0^Q \left[\int_0^T \sigma_t^2 \, dt \right]$$

  where $\mathrm{E}^Q[\;]$ denotes expectation taken with respect to the risk-neutral measure.

- If L_0 denotes the price of the log contract, we have deduced

$$L_0 = -\frac{T}{2}\, B_{0,T}\, MFIV^2.$$

- Thus, if we could get market participants to tell us the value of the log contract, we would have measured (up to a multiplicative constant) the quantity we were after.

II. VIX

- The most important volatility that everyone wants to measure is that of "the market portfolio".
  - ▶ For a well-diversified investor, in fact, this quantity is the ONLY notion of risk that matters.

- Since the early 1990s, the CBOE had tried to provide the public with summary statistics about implied volatility of options on the S&P 500.

- But they faced a host of problems, especially *which options*.
  - ▶ As we've discussed, every strike and maturity would give them a different answer.
  - ▶ They were tempted to use the most liquid ones, which are at the money. But those clearly understate the average beliefs expressed in out-of-the-money puts.
  - ▶ And it is very difficult to convince anybody that you are telling them anything very meaningful when you have to go into a long explanation to justify the fact that someone else could look at the same data as you and give a completely different answer.

- So they were very interested when they heard about Neuberger's idea.

- Then came another problem: *How can we implement it?*

- These log contracts are not actually traded and no dealer quotes prices in them.

- There was no reason to think if they were listed that any institutions would actually be interested in them.

- If we can't observe prices for the log contract, we've just replaced one measurement problem with another.

- But Neuberger (together with Mark Britten-Jones) had an answer.
  - ▶ Butterflies!

- We learned early on in the course that if you have prices for a lot of ordinary european calls (or puts) we can infer the prices of butterflies, and then construct arbitrary payoff functions from them.
  - ▶ Well, the CBOE has as much data on calls and puts on the S&P500 as anybody.
  - ▶ So they should be able to synthesize something very close to a log contract's payoff by appropriate weighting of butterflies.
  - ▶ Or, since the butterflies are themselves just made up of calls and puts, we should be able to construct a weighting scheme that just sums over all of them.

- Thus the modern VIX was born.

- In fact, the weighting scheme for the log contract is not too hard.

- Consider the idealized case with infinite strikes available:

  ▶ If $b(k)$ denotes the price of a butterfly centered at strike price k, then

$$L_0 = \sum_{k=0}^{\infty} \log(k/F_{0,T})\, b(k) = \sum_{k=0}^{\infty} \log(k)\, b(k) - \log(F_{0,T}) B_{0,T}$$

  where the second line uses the fact that the butterflies sum up to a riskless bond.

  ▶ Now consider the summation term. Put back in the definition of the butterflies.

$$\sum_{k=0}^{\infty} \log(k)\, b(k) = \sum_{k=0}^{\infty} \log(k)\, [c(k-\Delta k) - 2c(k) + c(k+\Delta k)]/\Delta k$$

  ▶ Rewrite this sum by grouping together terms involving the same option.

  * For example, if $\Delta k = 0.01$ then the terms involving calls with $k = 100.00$ are

$$c(100.00) * [\log(99.99) - 2\log(100.00) + \log(100.01)]/0.01$$

- Then, our summation should be the same as

$$\sum_k c(k) \; \frac{[\log(k-\Delta k) - 2\log(k) + \log(k+\Delta k)]}{\Delta k^2} \; \Delta k$$

  (I am being intentionally imprecise about what happens to the $k = 0$ term in this expression.)

- As the strikes become infinitely close together, you will recognize that the fractional term here is just going to become a second derivative operator.

- And we know from calculus that $\frac{d^2 \log(x)}{dx^2} = -\frac{1}{x^2}$.

- Putting this in our summation, as $\Delta t \to 0$ it should reduce to

$$-\int_0^\infty \frac{c(k)}{k^2} \, dk.$$

- So the basic idea is going to be to apply $1/k^2$ weightings to all the call prices.

- That derivation was not very rigorous because it ignored some important technicalities.
  - In particular, this last integral is not convergent!
  - We have to be careful about what happens near $k = 0$. The $c()$s approach a non-zero limit, but the weights explode.
  - (Formally, I applied integration-by-parts twice, but without being careful about the boundary terms.)

- We can make our summation better behaved by a change of variables.

- Define $c_0(k) = B_{0,T} \ \max[F_{0,T} - k, 0]$ to be each option's intrinsic (present) value.

- Now let $b_0(k)$ be the second-difference of these intrinsic values: $[c_0(k - \Delta k) - 2c_0(k) + c_0(k + \Delta k)]/\Delta k$.
  - Notice that these would be butterfly prices for calls on F_T expiring today.

- Suppose we add and subtract the sum of $b_0(k)$ from our original summation above

$$\sum_{k=0}^{\infty} \log(k)\, b(k) = \sum_{k=0}^{\infty} \log(k)\, [b(k) - b_0(k)] + \sum_{k=0}^{\infty} \log(k)\, b_0(k).$$

- What happens to the last term as Δk gets small?
  - Each one of the expiring butterflies is worthless except for the one centered around the price $F_{0,T}$. In other words, they collapse to delta functions.
  - Thus only one term in that summation is non-zero: $B_{0,T} \ \log(F_{0,T})$.
  - This exactly cancels out the second term in our expression for L above. So we can forget about both.

- Now consider the first term, and define $\hat{c}(k)$ to be $c(k) - c_0(k)$.

- Then our summation is

$$\sum_{k=0}^{\infty} \log(k) \, \frac{[\hat{c}(k-\Delta k) - 2\hat{c}(k) + \hat{c}(k+\Delta k)]}{\Delta k}.$$

- Now we can (correctly) apply our integration-by-parts trick because as k goes to zero $c(\hat{k})$ goes to zero, unlike $c(k)$.

- I will leave it as an exercise for you to verify that

$$\lim \underline{k} \to 0 \sum_{k=\underline{k}}^{\infty} \log(k) \, \frac{[\hat{c}(k-\Delta k) - 2\hat{c}(k) + \hat{c}(k+\Delta k)]}{\Delta k}$$

  is indeed

$$\sum_{k} \hat{c}(k) \, \frac{[\log(k-\Delta k) - 2\log(k) + \log(k+\Delta k)]}{\Delta k^2} \, \Delta k$$

  - ▶ (In fact, this requires that two different things go to zero: $\hat{c}/k$ and $\log(k) d\hat{c}/dk$.)

- This then gives us

$$-\sum_{k} \frac{\hat{c}(k)}{k^2} \Delta k \qquad \text{or} \qquad -\int_0^{\infty} \frac{\hat{c}(k)}{k^2} \, dk$$

$$= -\int_0^{\infty} \frac{c(k) - B_{0,T} \max[F_{0,T} - k, 0]}{k^2} \, dk.$$

- So thus far we have deduced that this integral gives us the value, L_0, of the log contract.

- Since we deduced above that $L_0 = -\frac{T}{2}\, B_{0,T}\, VIX^2$, we conclude
$$VIX^2 = \frac{2}{T}\, B_{0,T}^{-1} \int_0^\infty \frac{c(k) - B_{0,T}\max[F_{0,T} - K, 0]}{k^2}\, dk.$$

- Lastly, we can break the integral up into $k \le F_{0,T}$ and $k > F_{0,T}$
$$\frac{2}{T}\, B_{0,T}^{-1} \left(\int_0^{F_{0,T}} \frac{c(k) - B_{0,T}[F_{0,T} - K]}{k^2}\, dk + \int_{F_{0,T}}^\infty \frac{c(k)}{k^2}\, dk \right).$$

- Now observe that the numerator in the first integral is just $p(k)$ by put-call-parity. (These are european options, so it is exact).

- This gives us
$$VIX^2 = \frac{2}{T}\, B_{0,T}^{-1} \left(\int_0^{F_{0,T}} \frac{p(k)}{k^2}\, dk + \int_{F_{0,T}}^\infty \frac{c(k)}{k^2}\, dk \right).$$

- That equation is in fact what the CBOE's measure tries to approximate.

- Also the CBOE tries to keep the time horizon of VIX fixed by linearly interpolating the (squared) VIXes computed from two expirations on either side of 30 days.
  - ▶ And notice, the formula assumes the futures and options have the same T

- As a practical matter, they are limited by only having finitely many strikes.

- The first integral is the numerically sensitive part because it involves both numerator and denominator going to zero as k goes to zero.
  - The exchange has listed some puts with extremely low strike prices to help in this direction.
- Exact details of the biases due to (a) upper and lower limits not being 0 and infinity and (b) finite Δk have been analyzed by Jiang and Tian (2007) and (2009).
  - They also suggested some numerical corrections.
  - As far as I know, CBOE has not adopted them yet.
- Still it is remarkable how much you can do with butterflies.
  - It is fascinating that a contract whose payoff depends on the realized, random path of volatility can be replicated by a *static* portfolio of puts and calls.

III. Example Computations.

- We have seen that VIX (squared) measures model-free risk-neutral expected integrated variance for the S&P 500 for a fixed time horizon.

- You might be asking yourself some questions about this quantity, such as:

(A) How does the average variance relate to the *current* instantaneous variance? (And how do their dynamics relate to each other?)

(B) How does the risk-neutral expected variance relate to the *true* expected variance?

- I thought it would be helpful to do an explicit computation (NOT model-free) to address these.

- Let's assume the true instantaneous variance $v_t = \sigma_t^2$ obeys

$$dv = \kappa(\bar{v} - v)dt + svdW^v$$

This is a little different from assuming that σ obeys the exact same specification, but it's the same idea.

- Now I will just state a handy fact that can be proved with a little stochastic analysis. For this process,

$$\mathrm{E}[v_u | v_t] = v_t \, e^{-\kappa(u-t)} + \bar{v}(1 - e^{-\kappa(u-t)}).$$

- ▶ This is a very intuitive result that says that your forecast for volatility at a time horizon of $(u-t)$ in the future is a linearly weighted average of today's value and the long-run mean.
- ▶ And the weights decline exponentially at rate given by κ.

- So for this model it is easy to compute the average expectation (which is VIX):

$$\mathrm{E}\left[\frac{1}{\tau}\int_t^{t+\tau} v_u \; du | v_t\right] = \frac{1}{\tau}\int_t^{t+\tau} \mathrm{E}\left[v_u | v_t\right] \; du$$

$$= \frac{1}{\tau}\int_t^{t+\tau} [v_t \; e^{-\kappa(u-t)} \; + \; \bar{v}(1 - e^{-\kappa(u-t)})] \; du$$

$$= \bar{v} \; + \; \frac{1}{\tau}[(v_t - \bar{v}) \int_t^{t+\tau} e^{-\kappa(u-t)} \; du = \bar{v} \; + \; (v_t - \bar{v})\frac{1 - e^{-\kappa\tau}}{\kappa\tau}$$

$$= w(\tau)v_t + (1 - w(\tau))\bar{v}.$$

- ▶ Like the τ-period-ahead expectation, the average expectation is a weighted average of today's value and the long-run value.
- ▶ If we call this weighted expectation V (= VIX), then you can readily verify that

$$dV = (w\kappa + w') \; (\bar{v} - v_t) \; dt + w \; svdW^v$$

$$= (\kappa + \frac{w'}{w}) \; (\bar{v} - V_t) \; dt \; + \; s[V_t - (1 - w)\bar{v}] \; dW^v$$

where $w = w(\tau) = (1 - e^{-\kappa\tau})/\kappa\tau$.

- To compute VIX, however, we were supposed to use risk-neutral expectations.
- Well, we've learned how to change the process to its RN version:

$$
\begin{aligned}
dv &= [\kappa(\bar{v} - v) - \lambda\, s\, v]\; dt \;+\; s\, v\, dW^v \\
&= (\kappa + \lambda\, s)\left(\frac{\kappa\bar{v}}{\kappa + \lambda\, s} - v\right)\; dt \;+\; s\, v\, dW^v \\
&= \kappa^\star(v^\star - v)\; dt \;+\; s\, v\, dW^v.
\end{aligned}
$$

- Very interesting! This tells us that to "risk-neutralize" this model we have to adjust the long-run level of variance and the speed of mean reversion to the new values that I denoted $\kappa^\star$ and $v^\star$.
  - ▶ Recall that under the Vasicek specification we only had the second adjustment.

  Q: If the market price of variance risk is negative, which way do these adjustments go?

  Q: Are they big?

  Q: What happens if $\kappa + \lambda\, s \leq 0$?

- Returning to the earlier calculation, now the VIX index will be a weighted average of v_t and $v^\star$ with weight, $w^\star$, given by

$$w^\star = \frac{1 - e^{-\kappa^\star \tau}}{\kappa^\star \tau} \qquad \text{instead of} \qquad \frac{1 - e^{-\kappa \tau}}{\kappa \tau}.$$

- And the same substitutions go into the formula for dV – the (true) dynamics of VIX.
  - I'll leave it as an exercise to derive its *risk neutral* dynamics.

- The specific calculations here will be different, of course, if you start with a different model of the true variance dynamics.

- As an exercise, you might want to re-do the steps above with a different assumption.
  - For example, let $Z = \log(\sigma) = 0.5\ \log(v)$ and suppose

    $$dZ = \kappa(\bar{Z} - Z)\, dt + \ s\, dW^Z$$

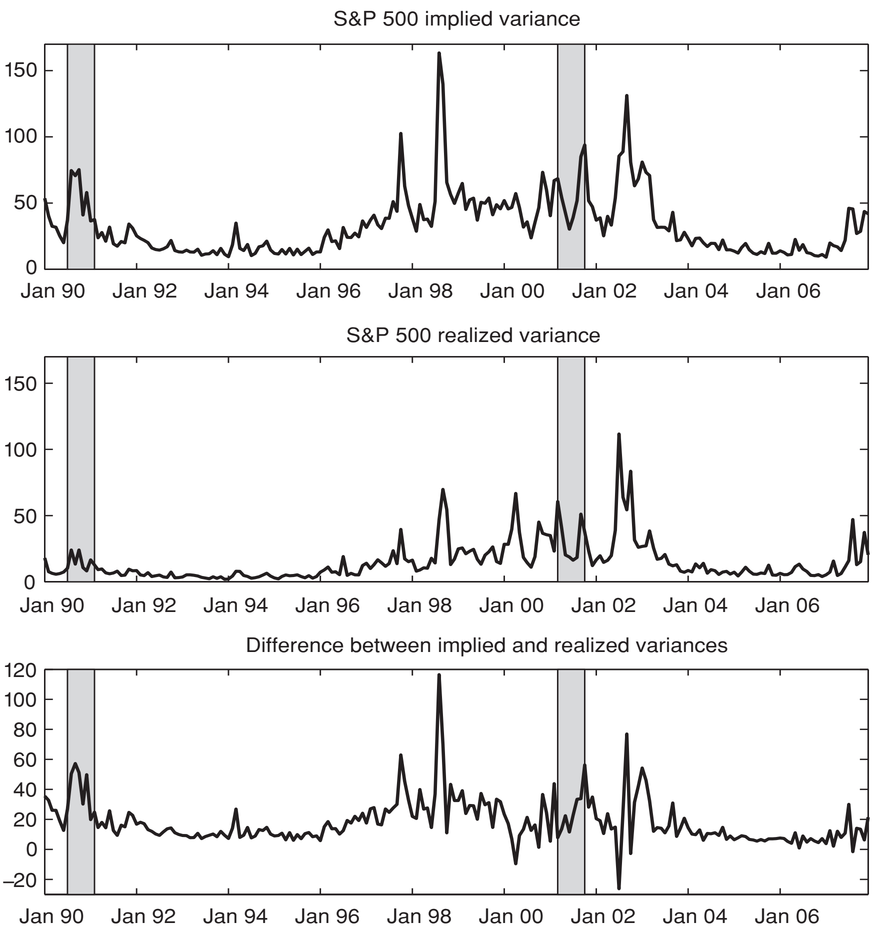
    i.e. Z follows an Ornstein-Uhlenbeck process.

- You will still find the same result from my calculation that *the risk-neutral expectation is biased upward* relative to the true expectation.

- And the bias is proportional to the market price of volatility risk.

- Notice that this bias does <u>not</u> imply that options markets are inefficient or irrational in any sense.

- In fact, there is a lot of on-going research that tries to measure the gap between the best true volatility forecast for the market and the VIX.
  - This is termed the "volatility risk premium".
  - It seems that this quantity is <u>itself</u> time varying.
  - And its variations may capture important information about crash fears in the economy.
- Here's a picture from a paper by Bollerslev, Tauchen and Zhou (2009).



IV. Extensions.

- Model-free VIX has been so successful in gaining widespread acceptance as the correct measure of risk-neutral expected volatility that the CBOE has even introduced cash-settled futures on VIX and options on those futures.
  - Using our earlier results, you could derive expressions for the no-arbitrage futures (or forward) price of VIX.

- They have also begun to compute model-free implied volatilities on other indexes and on commodities like gold and crude oil.

- In fact, CBOE now even computes VIX on VIX !
  - See `http://www.cboe.com/micro/VVIX/`

- Most recently, they have borrowed another idea from Neuberger and started to compute model-free <u>skewness</u>.
  - This is a measure of the asymmetric (risk-neutral) crash risk in the market.
  - I don't have time to explain its construction, but the basic principles are the same.
  - See `http://www.cboe.com/micro/SKEW/`

V. Summary

- It is a very cool fact that we can derive the risk-neutral expected cumulative variance from butterfly prices without ever having to specify what we think the *true* dynamics of volatility are.

- Having done so, we still have to be careful in interpreting what VIX means. It is not the instantaneous volatility of the market. And its dynamics are not the same as the dynamics of σ_t.

- Most important, it is not an unbiased forecast of future market volatility: it embeds a very substantial negative risk premium.
  - ▶ The discovery of the volatility risk premium is one of the most important developments in finance in the last decade.