Lecture Quantitative Finance Spring Term 2015

Prof. Dr. Erich Walter Farkas

Lecture 11: May 7, 2015

Motivation and setup

- the goal of this chapter is to treat the implied volatility which requires an algorithm for solving a nonlinear equation
- the general problem is
  - given a function $F : \mathbb{R} \to \mathbb{R}$, find an $x^* \in \mathbb{R}$ such that $F(x^*) = 0$
- in general, of course, we cannot find an x^* analytically, and must therefore content ourselves with an approximation via a computational method
- it is worth keeping in mind that, depending on the nature of F, there may be no suitable x^*, exactly one x^* or many x^* values
- we introduce two algorithms for solving a nonlinear equation
  - the bisection method
  - Newton's method (also called Newton-Raphson method)

The bisection method

- is based on the observation that if a continuous function changes sign, then it must pass through zero, that is
  - for continuous functions F, if $x_a < x_b$ with $F(x_a)F(x_b) < 0$ then there exists some x^* with $x_a < x^* < x_b$ with $F(x^*) = 0$
- having found x_a and x_b with $F(x_a)F(x_b) < 0$ we could evaluate F at the mid-point $x_{mid} := (x_a + x_b)/2$
- the sign of $F(x_{mid})$ must match either the sign of $F(x_a)$ or $F(x_b)$; this means that one of the intervals $[x_a, x_{mid}]$ or $[x_{mid}, x_b]$ must contain an x^*
- by repeating this process we can construct an arbitrarily small interval in which an x^* must lie, hence we can find an x^* to any level of accuracy

The bisection method: algorithm

- **Step 1**: find x_a and x_b with $x_a < x_b$ such that $F(x_a)F(x_b) \leq 0$
- **Step 2**: set $x_{mid} := (x_a + x_b)/2$ and evaluate $F(x_{mid})$
- **Step 3**: if $F(x_a)F(x_{mid}) < 0$ then reset $x_b = x_{mid}$. Otherwise reset $x_a = x_{mid}$
- **Step 4**: if $x_b - x_a < \varepsilon$ then stop. Use $(x_a + x_b)/2$ as the approximation to x^*. Otherwise return to Step 2.

- note that we must choose a value $\varepsilon > 0$ for our stopping criterion $x_b - x_a < \varepsilon$
- it is easy to see that the value $(x_a + x_b)/2$ on termination is no more than a distance $\varepsilon/2$ from a solution x^* (hence ε controls the accuracy of the process)
- because the bisection method halves the length of the interval $[x_a, x_b]$ on each iteration, we may bound the error at the kth iteration by $L/2^{k+1}$ where L is the length of the original interval, $x_b - x_a$
- this is referred to as a linear convergence bound because the error decreases by a linear factor (in this case $1/2$) on each iteration

Newton's method

- is faster than the bisection method
- can be derived in a number of ways: here we will use a Taylor series approach
- suppose we wish to compute a sequence $x_0, x_1, x_2, \dots$ that converges to a solution x^*
- we may expand $F(x+\delta)$ for small δ by

$$F(x_n+\delta) = F(x_n) + \delta F'(x_n) + O(\delta^2)$$

- ignoring $O(\delta^2)$ and setting $F(x_n) + \delta F'(x_n) = 0$ gives $\delta = -F(x_n)/F'(x_n)$
- it follows that if x_n is close to a solution x^* then

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}$$

  should be even closer
- given a starting value, x_0, the last iteration defines Newton's method

Newton's method

- since we discarded an $O(\delta^2)$ term in Taylor's approximation we may expect that the error $x_n - x^*$ squares as n increases to $n+1$: that is if $x_n - x^* = O(\delta)$ then $x_{n+1} - x^* = O(\delta^2)$
- to see this more clearly, note that using $F(x^*) = 0$ and assuming $F'(x_n) \neq 0$ a Taylor series gives

$$
\begin{aligned}
x_{n+1} - x^* &= x_n - x^* - \left(\frac{F(x_n) - F(x^*)}{F'(x_n)} \right) \\
&= x_n - x^* \\
&\quad - \frac{(x_n - x^*)F'(x_n) + O((x_n - x^*)^2)}{F'(x_n)} \\
&= O((x_n - x^*)^2)
\end{aligned}
$$

- this type of analysis can be formalised in a theorem

Newton's method: Theorem

Suppose

- F has a continuous second derivative
- $x^* \in \mathbb{R}$ satisfies $F(x^*) = 0$ and $F'(x^*) \neq 0$

Then

- there exists a $\delta > 0$ such that for $| x_0 - x^* | < \delta$ the sequence given by

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}$$

  is well-defined for all $n > 0$,
- with

$$\lim_{n\to\infty} | x_n - x^* | = 0,$$

- and there exists a constant $C > 0$ such that

$$| x_{n+1} - x^* | \leq C \ | x_n - x^* |^2 .$$

Newton's method: comments

- the last inequality shows that Newton's method has a quadratic (or second order) convergence
- this result requires the starting value x_0 to be chosen sufficiently close to x^*; in practice Newton's method works very well when a suitable x_0 is found, but may fail to converge otherwise

Newton's method: computational example

- suppose we wish to find the value of x^* such that $\mathbb{P}(X \leq x^*) = \frac{2}{3}$ where $X \sim N(0,1)$
- essentially we want to solve $F(x) = 0$, where $F(x) := N(x) - \frac{2}{3}$ with

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{s^2}{2}}\, ds$$

- it follows from the definition of N that F is an increasing function and $F(0) = \frac{1}{2} - \frac{2}{3} < 0$ and $\lim_{x \to \infty} F(x) = 1 - \frac{2}{3} > 0$
- hence we may immediately conclude that the equation $F(x) = 0$ has a unique solution $0 < x^* < \infty$

Newton's method: computational example (cont'd)

- we may apply the bisection method with $x_a = 0$ and with x_b sufficiently large such that $F(x_b) > 0$
- for the choice $x_b = 10$ and a tolerance of $\varepsilon = 10^{-5}$ in the stopping criterion the bisection method needs 20 iterations
- setting $x_0 = 1$ and stopping with Newton's method when $\mid x_{n+1} - x_n \mid < 10^{-5}$ only four iterations are needed to produce an error of around 10^{-12} and the error roughly squares from one step to the next
- repeating Newton's method with $x_0 = 2$ however, results in a sequence that blows-up

Motivation

- the Black-Scholes call and put values depend on S, K, r, $T-t$ and σ^2
- of these five quantities, only the asset volatility cannot be observed directly; how do we find a suitable value for σ?
- approach: extract the volatility from the observed market data - given a quoted option value, and knowing S, t, K, r and T find the σ that leads to this value
- having found σ, we may use Black-Scholes formula to value other options on the same asset
- a σ computed this way is known as an *implied volatility*; the name indicated that σ is implied by option value data in the market
- this is a totally different way to get σ compared with the historical volatility

Option value as a function of volatility

- we focus here on the case of extracting σ from a European call option quote
- an analogous treatment can be given for a put, or alternatively, the put quote could be converted into a call quote via put-call parity
- we assume that the parameters K, r and T and the asset price S and time t are known
- in practice we will typically be interested in the time-zero case, $t = 0$ and $S = S_0$
- we thus treat the option value as function of σ only, and, from now on, denote it by $C(\sigma)$
- given a quoted value C^*, our task is to find the implied volatility σ^* that solves $C(\sigma^*) = C^*$
- it is possible to exploit the special form of the nonlinear equation arising in this context

Option value as a function of volatility:

$\sigma \to \infty$

- since volatility is non-negative, only values $\sigma \in [0, \infty)$ are of interest
- let us look at $C(\sigma)$ in the case of large or small volatility
- first assume $\sigma \to \infty$
  - recall
    $$d_1 = \frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$
    so that $d_1 \to \infty$ and hence $N(d_1) \to 1$
  - similarly $d_2 = d_1 - \sigma\sqrt{T-t}$ so that $d_2 \to -\infty$ and hence $N(d_2) \to 0$
  - using Black-Scholes formula
    $$C(\sigma) = S \cdot N(d_1) - K \cdot e^{-r(T-t)} \cdot N(d_2)$$
    it follows that
    $$\lim_{\sigma\to\infty} C(\sigma) = S$$

Option value as a function of volatility: $\sigma \to 0^+$

- next we look at $\sigma \to 0^+$ and distinguish three cases
  1. $S - Ke^{-r(T-t)} > 0$; in this case $\log(S/K) + r(T-t) > 0$ so that if $\sigma \to 0^+$ we have $d_1 \to \infty$, $N(d_1) \to 1$, $d_2 \to \infty$ and $N(d_2) \to 1$.
     Hence, $C \to S - Ke^{-r(T-t)}$.
  2. $S - Ke^{-r(T-t)} < 0$; in this case $\log(S/K) + r(T-t) < 0$ so that if $\sigma \to 0^+$ we have $d_1 \to -\infty$, $N(d_1) \to 0$, $d_2 \to -\infty$ and $N(d_2) \to 0$.
     Hence, $C \to 0$.
  3. $S - Ke^{-r(T-t)} = 0$; in this case $\log(S/K) + r(T-t) = 0$ so that if $\sigma \to 0^+$ we have $d_1 \to 0$, $N(d_1) \to 1/2$, $d_2 \to 0$ and $N(d_2) \to 1/2$.
     Hence, $C \to \frac{1}{2}(S - Ke^{-r(T-t)}) = 0$.
- these three cases are summarized neatly by the formula

$$\lim_{\sigma \to 0^+} C(\sigma) = \max(S - Ke^{-r(T-t)}, 0)$$

Bounds for the option value as a function of volatility

- now we recall from previous lectures that the derivative of C with respect to σ, that is the vega, is given by

$$\text{vega} = S\sqrt{T-t}\,N'(d_1)$$

  and in particular we know that $\partial C/\partial\sigma > 0$
- since $C = C(\sigma)$ is continuous with a positive first derivative, we conclude that C is monotonically increasing on $[0,\infty)$
- from

$$\lim_{\sigma\to 0^+} C(\sigma) = \max(S - Ke^{-r(T-t)}, 0)$$

  and from

$$\lim_{\sigma\to\infty} C(\sigma) = S$$

  the values of $C(\sigma)$ must lie between $\max(S - Ke^{-r(T-t)}, 0)$ and S
- consequently the equation $C(\sigma) = C^*$ has a solution if, and only if,

$$\max(S - Ke^{-r(T-t)}, 0) \leq C^* \leq S$$

The second derivative of $C(\sigma)$

- for later use we will calculate the second derivative
- differentiating

$$\text{vega} := \frac{\partial C}{\partial \sigma} S\sqrt{T-t}\, N'(d_1)$$

we get

$$\frac{\partial^2 C}{\partial \sigma^2} = -\frac{S\sqrt{T-t}}{\sqrt{2\pi}}\, e^{-\frac{1}{2}d_1^2}\, d_1 \frac{\partial d_1}{\partial \sigma}$$

- using

$$d_1 = \frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

we have

$$\begin{aligned}\frac{\partial d_1}{\partial \sigma} &= -\frac{\log(S/K) + r(T-t)}{\sigma^2\sqrt{T-t}} + \frac{1}{2}\sqrt{T-t} \\ &= -\frac{\log(S/K) + (r - \sigma^2/2)(T-t)}{\sigma^2\sqrt{T-t}} = -\frac{d_2}{\sigma}\end{aligned}$$

The second derivative of $C(\sigma)$ (cont'd)

- consequently

$$\frac{\partial^2 C}{\partial \sigma^2} = \frac{S\sqrt{T-t}}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} \cdot \frac{d_1 d_2}{\sigma} = \frac{d_1 d_2}{\sigma} \frac{\partial C}{\partial \sigma}$$

- from the last equation it follows that $\partial C / \partial \sigma$ has its maximum over $[0, \infty)$ at $\sigma = \widehat{\sigma}$ given by

$$\widehat{\sigma} := \sqrt{2 \left| \frac{\log(S/K) + r(T-t)}{T-t} \right|}$$

Exercise Prove that $\partial C / \partial \sigma$ has a unique maximum over $[0, \infty)$ at $\sigma = \widehat{\sigma}$ defined above.

- moreover

$$\frac{\partial^2 C}{\partial \sigma^2} = \frac{T-t}{4\sigma^3}(\widehat{\sigma}^4 - \sigma^4)\frac{\partial C}{\partial \sigma}$$

Bisection for computing the implied volatility

- we will write our nonlinear equation for σ^* in the form $F(\sigma) = 0$ where $F(\sigma) = C(\sigma) - C^*$
- to apply the bisection method, we require an interval $[\sigma_a, \sigma_b]$ over which $F(\sigma)$ changes its sign
- it follows from

$$\lim_{\sigma \to \infty} C(\sigma) = S$$

and from

$$\lim_{\sigma \to 0^+} C(\sigma) = \max(S - Ke^{-r(T-t)}, 0)$$

and the monotonicity of $C(\sigma)$ that this can be done by fixing K (say $K = 0.05$) and trying $[0, K]$, $[K, 2K]$, $[2K, 3K]$,...

Newton's method for computing the implied volatility

- Newton's method takes the form

$$\sigma_{n+1} = \sigma_n - \frac{F(\sigma_n)}{F'(\sigma_n)}$$

  where $F'(\sigma) = \frac{\partial C}{\partial \sigma}$ is given above
- using $F(\sigma^*) = 0$ and the mean value theorem, we have

$$\begin{aligned} \sigma_{n+1} - \sigma^* &= \sigma_n - \sigma^* - \frac{F(\sigma_n) - F(\sigma^*)}{F'(\sigma_n)} \\ &= \sigma_n - \sigma^* - \frac{(\sigma_n - \sigma^*)F'(\xi_n)}{F'(\sigma_n)} \end{aligned}$$

  for some ξ_n between σ_n and σ^*
- hence

$$\frac{\sigma_{n+1} - \sigma^*}{\sigma_n - \sigma^*} = 1 - \frac{F'(\xi_n)}{F'(\sigma_n)}$$

Newton's method for computing the implied volatility

- we know that $F'(\sigma)$ is positive and takes its maximum at the point $\widehat{\sigma}$ from above
- hence, using the starting value $\sigma_0 = \widehat{\sigma}$ we must have $0 < F'(\xi_0) < F'(\sigma)$ so that the last equality implies

$$0 < \frac{\sigma_1 - \sigma^*}{\sigma_0 - \sigma^*} < 1$$

- this means that the error in σ_1 is smaller than, but has the same sign as, the error in σ_0
- we will distinguish if $\widehat{\sigma} < \sigma^*$ or if $\widehat{\sigma} > \sigma^*$

Newton's method for computing the implied volatility

- to proceed assume first that $\widehat{\sigma} < \sigma^*$
  - then from the last inequalities we have $\sigma_0 < \sigma_1 < \sigma^*$
  - we know $F''(\sigma) < 0$ for all $\sigma > \widehat{\sigma}$ and ξ_1 lies between σ_1 and σ^*
  - hence $0 < F'(\xi_1) < F'(\sigma_1)$ and

$$0 < \frac{\sigma_2 - \sigma^*}{\sigma_1 - \sigma^*} < 1$$

  - repeating this argument we get

$$0 < \frac{\sigma_{n+1} - \sigma^*}{\sigma_n - \sigma^*} < 1 \quad \text{for all} \quad n \geq 0$$

    so the error decreases monotonically as n increases
- in a similar manner one can treat the case $\widehat{\sigma} > \sigma^*$

Newton's method for computing the implied volatility

- overall we conclude that with the choice $\sigma_0 = \widehat{\sigma}$ the error will always decrease monotonically as n increases
- it follows that the error must tend to zero and the previous theory shows that the convergence must be quadratic
- therefore using $\sigma_0 = \widehat{\sigma}$: this is our method for computing the implied volatility

Implied volatility with real data

- we now look at the implied volatility for call options traded at the London International Financial Futures and Options Exchange (LIFFE) as reported in the *Financial Times* on Wednesday, 22 August 2001
- the data is for the FTSE 100 index, which is an average of 100 equity shares quoted on the London Stock Exchange

Exercise price		Option price
——	——	——
5125		475
5225		405
5325		340
5425		280.5
5525		226
5625		179.5
5725		139
5825		105

Implied volatility with real data

- the expiry date for these options was December 2001
- the initial price (on 22 August 2001) was 5420.3
- we take values of $r = 0.05$ for the interest rate and $T = 4/12$ for the duration of the option
- the implied volatility computed for the eight different exercise prices is decreasing (from approx. 0.19 to 0.174)
- of course, if Black-Scholes formula would be valid, the volatility would be the same for each exercise price
- however in this example the implied volatility varies by around 10%

Implied volatility with real data

- note: implied volatility is higher for in-the-money equity call options than for out-of-the-money equity call options
- this behaviour is typical for data arising after the stock market crash of October 1987
- pre-crash plots of implied volatility against exercise price would often produce a convex *smile* shape; more recent data tends to produce more of a *frown*

Implied volatility: some final comments

- the widely reported phenomenon that the implied volatility is not constant as other parameters are varied does, of course, imply that the Black-Scholes formulas fail to describe the option values that arise in the marketplace
- this should be no surprise, given that the theory is based on a number of simplifying assumptions
- despite the disparities, the Black-Scholes theory, and the insights that it provides, continue to be regarded highly by both academics and market traders
- it is common for option values to be quoted in terms of *vol*; rather than giving C^*, the σ^* such that $C(\sigma^*) = C^*$ in the Black-Scholes formula is used to describe the value
- many attempts have been made to fix the nonconstant volatility discrepancy in the Black-Scholes theory; a few of these have met with some success but none lead to the simple formulas and clean interpretation of the original work: see Chapter 17 of Hull (2000)

Answer to the Exercise

Exercise Hint: Discuss the monotonicity of $\partial C/\partial\sigma$ analysing the sign of $\frac{\partial^2 C}{\partial\sigma^2}$

Reference

Higham, Desmond J., "*An Introduction to Financial Option Valuation - Mathematics, Stochastics and Computation*", Cambridge University Press, 2004, Chapter 14: Implied Volatility.