

Introdução a Superfície de Volatilidade

Rafael Felipe Bressan

2019-01-26

Conteúdo

Prefácio	5
Sobre os Autores	7
Rafael Felipe Bressan	7
Erik Naoki Kawano	7
Glauber Naue	7
1 Introdução as opções	9
1.1 Conceitos in the money, at the money e out the money	10
1.2 Modelos americano e europeu de opções	10
1.3 Hedge	12
1.4 Travas	12
2 Processos Estocásticos em Finanças	13
2.1 Processos de Markov	13
2.2 Movimento Browniano	14
2.3 Definição	15
2.4 Movimento Browniano Geométrico	16
3 Modelo Black-Scholes-Merton	19
3.1 Portfólio de replicação	19
3.2 Precificação neutra ao risco	20
3.3 Encontrando a equação de Black-Scholes	20
3.4 Solução analítica	21
3.5 Paridade compra e venda	22
3.6 As Gregas	23
4 Smile de Volatilidade	25
4.1 Reparametrizando B&S e definição de moneyness	25
4.2 Características de smiles de volatilidade	26
4.3 Smile como forma de precificação	28
4.4 Estrutura a termo	29
4.5 Arbitragem estática	29
4.6 Limites de inclinação	33
4.7 Distribuição implícita	34
4.8 Conclusão	35
5 Métodos de Calibração	37
5.1 Calibração versus Interpolação	37
5.2 Spline cúbica	38
5.3 Suavização	39
5.4 Parametrização	41
5.5 Conclusão	42

6	Calibrando uma SVI	45
6.1	Modelo SVI	45
6.2	Restrições de não-arbitragem	46
6.3	Reparametrização Quasi-explicit	47
6.4	Algoritmo	49
6.5	Resultados	49
6.6	Conclusão	54

Prefácio

Este é um pequeno resumo, elaborado na forma de livro, sobre os estudos realizados pelo núcleo de derivativos e riscos do Clube de finanças ESAG.

Os estudos realizados pelos membros do núcleo foram sendo apresentados na forma de artigos no blog do Clube. A partir destes artigos foi feita esta coletânea de forma a apresentar todo o conteúdo em local único para facilitar a assimilação dos membros futuros e leitores de nosso blog.

Sobre os Autores

Rafael Felipe Bressan

Formado em Engenharia de Controle e Automação Industrial pela UFSC e aluno de graduação do curso de Ciências Econômicas na UDESC/Esag. Membro do Clube de Finanças Esag e gerente do núcleo de pesquisa em riscos e derivativos.

Se interessa por finanças quantitativas, modelagem e controle de riscos e desenvolveu, durante a elaboração deste livro, grande curiosidade sobre precificação de derivativos. Gosta de programar em R, linguagem com a qual elaborou este próprio livro e está aprendendo Python.

Erik Naoki Kawano

Aluno de Engenharia de Produção Elétrica na UFSC. Membro do Clube de Finanças Esag e participante do núcleo de pesquisa em riscos e derivativos. Está atualmente aprendendo a programar em Python e em R.

Glauber Naue

Aluno de Ciências Econômicas na UDESC/Esag. Membro do Clube de Finanças ESAG, participa do núcleo de pesquisa em riscos e derivativos. Atualmente realizando um estudo sobre modelo de fatores aplicado ao mercado brasileiro.

Capítulo 1

Introdução as opções

Neste Capítulo¹ iremos apresentar os instrumentos financeiros conhecidos como opções. Existem dois tipos de opção. Uma opção de compra dá ao detentor o direito de comprar o ativo subjacente até uma determinada data por um determinado preço. Uma opção de venda dá ao titular o direito de vender o ativo subjacente até uma determinada data por um determinado preço.

Alguns termos importantes para o melhor entendimento deste Capítulo são: “Ativo subjacente”, que é o ativo negociado no contrato, “data de vencimento”, no caso do modelo americano é até a data limite para exercer a opção de compra e no modelo europeu é na data final em que a opção de compra pode ou não ser exercida (serão discutidos mais detalhes sobre estas modalidades posteriormente) e “preço de exercício” (strike), é o valor a ser pago pelo ativo de acordo com o contrato. As opções americanas podem ser exercidas a qualquer momento até a data de vencimento. Por sua vez, as opções europeias podem ser exercidas somente na própria data de vencimento.

As opções europeias são geralmente mais fáceis de analisar do que as opções americanas, e algumas das propriedades de uma opção americana são frequentemente deduzidas daquelas de sua contraparte europeia.

A opção de compra (call) dá ao comprador da opção o **direito de comprar** o ativo subjacente pelo preço de exercício. A opção de venda (put) dá ao comprador da opção o **direito de vender** o ativo subjacente pelo preço combinado no contrato na data futura.

O preço de uma opção de compra diminui à medida que o preço de exercício aumenta, enquanto o preço de uma opção de venda aumenta à medida que o preço de exercício aumenta. Ambos os tipos de opção tendem a se tornar mais valiosas à medida que seu tempo até o vencimento aumenta. Na verdade existem seis fatores que afetam o preço de uma opção de ação, por exemplo:

1. O preço atual da ação, S_0
2. O preço de exercício, K
3. O tempo para expiração, τ
4. A volatilidade do preço das ações, σ
5. A taxa de juros livre de risco, r
6. Os dividendos que se espera sejam pagos, q

Observamos que existem quatro tipos de participantes nos mercados de opções:

- Compradores de calls
- Vendedores de calls
- Compradores de puts
- Vendedores de puts

¹Artigo originalmente escrito por Glauber Naue e Erik Kawano, adaptado por Rafael Bressan para este livro.

Tabela 1.1: Exemplo: Comprando call de X com o preço de exercício (strike) de 10.000

Ação	Prêmio Pago	Opção no Exercício	Lucro	Moneyness
9000	200	0	-200	OTM
9500	200	0	-200	OTM
10000	200	0	-200	ATM
10200	200	200	0	ITM
10500	200	500	300	ITM
11000	200	1000	800	ITM

Os compradores são referidos como tendo posições *long*, vendedores são referidos como tendo posições *short*. Vender uma opção também é conhecido como lançar a opção ou subscrever.

Exemplificando uma operação de compra de call, caso o preço do ativo tenha subido acima do preço de strike o comprador pode exercer sua opção de compra e ele lucrará a partir do momento em que o valor da ação for maior que o strike mais o valor pago pela opção (chamado de prêmio). Caso o preço do ativo tenha caído abaixo do strike, o comprador simplesmente não exerce sua opção, limitando sua perda nessa operação ao prêmio pago.

Usando a tabela 1.1 como exemplo é possível ver que o resultado final será maior que R\$ 0 quando o valor do ativo subjacente é maior que R\$10.200 (R\$10.000 de strike + R\$200 de prêmio), e o prejuízo final está limitado ao valor de R\$200. A figura 1.1 abaixo mostra o perfil de lucro da operação exemplificada, típico de uma compra de call.

No caso de uma operação de uma compra de put, caso o preço do ativo tenha subido acima do strike, não faz sentido o detentor da opção exercer seu direito, assim sua perda será apenas o valor pago pela opção. Caso o preço do ativo tenha descido abaixo do strike, o comprador da put pode realizar a venda e começará a lucrar a partir do momento em que o strike fique acima do valor do ativo somado ao valor do prêmio pago pela opção. Usando a tabela 1.2 como exemplo é possível ver que o resultado final será de no mínimo -R\$600 caso o valor do ativo subjacente seja igual ou maior que R\$15.000, e o resultado final aumenta conforme o ativo perde o valor, sendo positivo a partir de quando seu valor é de R\$14.400.

1.1 Conceitos in the money, at the money e out the money

Estes termos são usados para se referir a opções que estão com o preço de exercício (strike) do ativo abaixo, acima ou igual ao valor atual do ativo.

- Out the money: Strike do ativo subjacente está abaixo do valor de mercado no caso de calls ou quando o strike do ativo está acima do valor de mercado no caso de puts.
- At the money: Strike do ativo subjacente é o igual ao valor de mercado.
- In the money: Strike do ativo subjacente está acima do valor de mercado no caso de calls ou quando o strike do ativo está abaixo do valor de mercado no caso de puts.

1.2 Modelos americano e europeu de opções

No modelo americano de opções o comprador pode exercer seu direito de compra ou venda do ativo subjacente a qualquer momento entre o início do contrato e o vencimento dele, enquanto isso no modelo europeu a transação só pode ser realizada na data de vencimento.

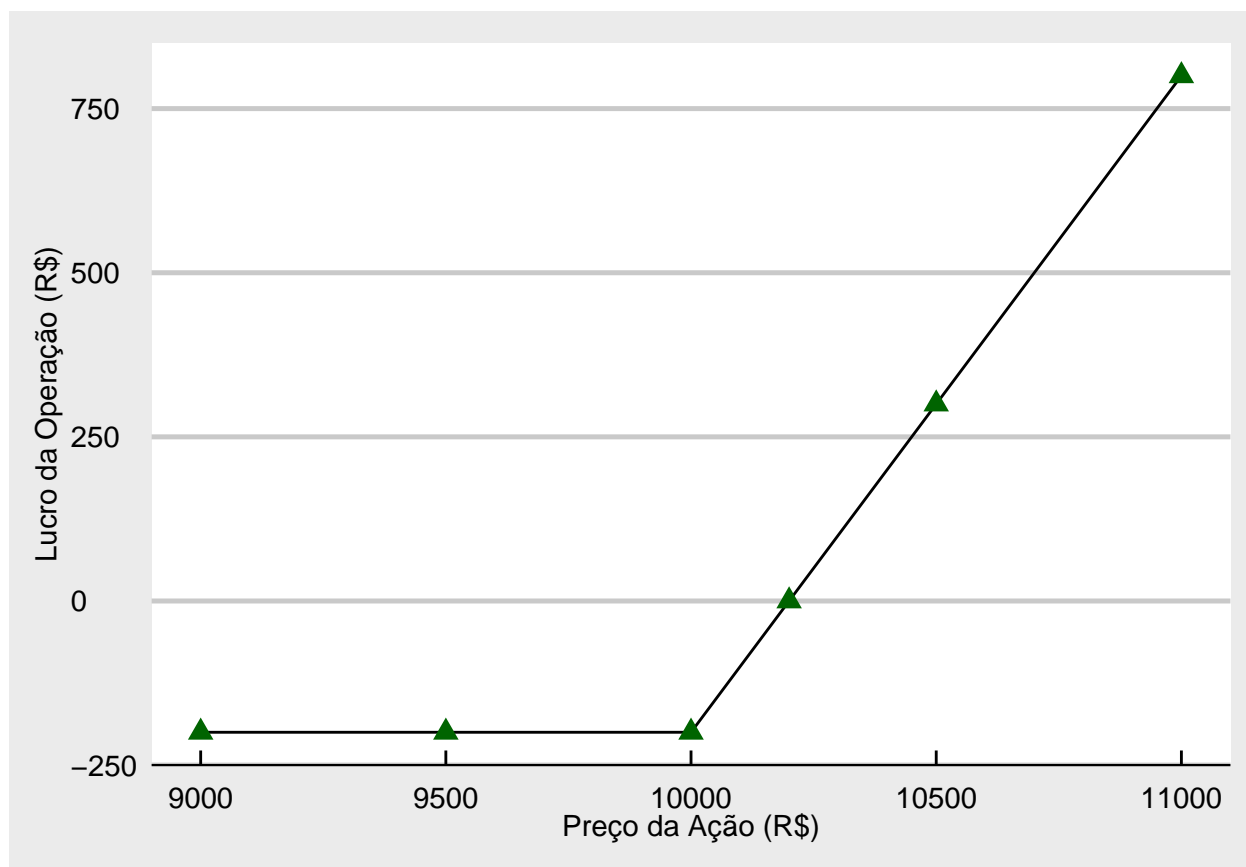


Figura 1.1: Perfil de lucro típico de uma compra de call.

1.3 Hedge

O mercado de opções pode ser usado tanto para hedge (proteção) quanto para especulação. O hedge é feito para limitar as possíveis perdas que um investidor pode ter ao estar com seu patrimônio atrelado a determinado ativo, por exemplo, para um acionista que possui ações de determinada empresa se proteger contra uma possível queda no valor de suas ações, ele pode comprar opções de venda at the money de suas ações para que seu prejuízo máximo seja o prêmio.

1.4 Travas

Devido ao mercado de opções nos oferecer diversas possibilidades entre call e put onde você pode estar comprado e/ou vendido irá surgir várias posições a serem assumidas, para nos adequarmos ao quanto estamos dispostos a encarar o risco para atingirmos o retorno desejado. Essas posições são conhecidas como “Travas”.

Entendendo as travas, existem diversas estratégias, como por exemplo: Trava de alta, trava de baixa, Long Straddle, Short Straddle, entre outras. Mas afinal qual é o funcionamento delas? Supondo que o leitor espere uma alta do mercado, no entanto acredite que não irá superar determinado ponto ele poderá realizar uma Trava de alta. Onde comprará uma opção de Call a um preço de strike X e vender outra Call com o preço de strike Y, onde obrigatoriamente $Y > X$. Nesta operação limitaremos o nosso ganho caso o mercado supere as nossas expectativas, no entanto diminuiremos o custo da operação, o custo será o prêmio pago pela opção X menos o prêmio recebido pela opção Y, para facilitar a compreensão observemos o gráfico a seguir:

Nesse caso o valor do prêmio da compra foi de R\$30,00 enquanto a da venda foi R\$10,00, assim limitamos nossa perda em R\$20,00, enquanto os preços de strike da compra e da venda da call foram respectivamente R\$250,00 e R\$300,00, fazendo o retorno máximo ser R\$30,00 que é a diferença entre os valores de strike a serem realizados e descontado o valor pago pelo prêmio.

Agora que o leitor já entendeu melhor o conceito da trava, vamos explorar uma mais complexa a Long Butterfly. Aqui é realizado a compra de uma put e call com preços de strike iguais, vendesse uma put com preço inferior e vende alguma call com preço superior as iniciais. Observe que pelo fato de contar com a venda de duas opções nessa estratégia tem um custo de operação reduzido, no entanto o ideal é utilizar em um mercado de pouca volatilidade, dado que se a volatilidade ser alta perdesse a possibilidade de ter um ganho maior, nesse caso recomenda o uso por exemplo de uma Long Straddle. Enfim vamos ao gráfico para facilitar a compreensão da estratégia:

Teremos então a compra de uma put e call de strike iguais de R\$150,00 a venda de uma put com strike inferior de R\$80, e a venda de uma call com preço superior R\$220. Os valores exatos dos prêmios não nos interessam no momento, porém é importante entender que teremos dois com saldos positivos referente a nossa venda e dois negativos que advém das compras, o resultado será nosso prejuízo máximo, olhando o gráfico nesse caso é de R\$20,00. Pela área de retorno do gráfico podemos ver que nosso risco está reduzido. Onde o pior cenário possível se encontra em o preço do ativo-objeto se aproximar do valor do R\$150,00, que é onde os contratos adquiridos não serão vantajosos em nenhuma ponta. No entanto se o preço se aproximar de R\$220 poderemos exercer nosso direito da compra da call inicial por R\$150,00(você terá o direito de comprar a um preço inferior), o mesmo será valido caso haja uma queda do preço se aproximado do valor de R\$80,00 onde a logica será a mesma só que aqui será o usado o direito da compra da put por R\$150,00(Você terá o direito de vender a um preço superior). Observe que apesar do nosso risco ser reduzido, limita os nossos ganhos, com a venda da call e da put com preços superior e inferior respectivamente.

Espero que o leitor tenha despertado interesse no assunto, com esse conteúdo dominado já saberá o básico sobre opções, fique atento a novas postagens em breve iremos mais a fundo explicando por exemplo o modelo Black Scholes, como as opções são precificadas entre outros materiais.

Capítulo 2

Processos Estocásticos em Finanças

Neste capítulo abordaremos um assunto técnico, mas muito utilizado e de fundamental importância para a precificação de instrumentos derivativos. Será apresentado o conceito de processos estocásticos - PE, e sua aplicação no mundo das finanças.

Um processo estocástico é a evolução temporal de uma determinada variável de interesse que pode assumir valores aleatórios em cada ponto no tempo. Em outras palavras, o caminho que a variável segue ao longo do tempo evolui de maneira incerta. Estes processos podem se dar em tempo discreto ou em tempo contínuo. Processo em tempo discreto são aqueles onde o valor da variável pode se alterar somente em intervalos pré-definidos de tempo, por exemplo ao final do dia. Em processos em tempo contínuo, o valor de nossa variável está constantemente em mudança, de forma aleatória seguindo alguma distribuição de probabilidades.

Estes processos são muito importantes em finanças pois, é amplamente aceito que a evolução do preço de ativos financeiros pode ser modelado por um PE em tempo contínuo, sendo este modelo portanto, a base para a teoria de precificação de ativos e da qual os derivativos fazem extenso uso. Aprender sobre a evolução temporal do preço de uma ação através de um processo estocástico é o primeiro passo para saber como atribuir um preço a uma opção sobre esta ação, por exemplo.

Deve ser notado também que apesar de o preço dos ativos serem **observados** apenas em intervalos discretos de tempo (apenas quando existe transação) e assumirem valores também discretos (múltiplos de um centavo), o preço e sua evolução estão ocorrendo continuamente, nossas observações que são discretas. Desta forma os processos em tempo contínuos são ideais para este tipo de modelagem.

2.1 Processos de Markov

Uma primeira definição de deve-se fazer para estudar PE aplicados a evolução do preço de ações é o conceito de processo de Markov. Este tipo de processo é tal que o histórico do processo que o levou até seu estado atual, é **irrelevante** para a previsão de seu estado futuro. Ou seja, toda a informação da história do processo já está contida no seu valor atual. Quando consideramos que preços de ativos seguem um processo de Markov, estamos assumindo válida pelo menos a forma fraca de mercados eficientes.

Uma implicação desta suposição, verificada empiricamente, é que não se pode obter lucros apenas seguindo padrões históricos do preço e extrapolando-os no futuro. Outra, mais importante para nossos processos, é que as distribuições de probabilidade que a variável aleatória segue em cada ponto no tempo são **independentes**.

2.2 Movimento Browniano

Suponha um processo de Markov, que para fins de simplificação consideraremos em tempo discretos. Se a distribuição de probabilidade para o próximo incremento no valor do processo for uma Normal com média zero e variância unitária, podemos representar este incremento por $\phi(0, 1)$. Como este é um processo de Markov, o segundo incremento será independente do primeiro e terá novamente a mesma distribuição de probabilidade. Qual seria então, a partir do período inicial até o segundo período, a distribuição de probabilidade dos possíveis valores de nosso hipotético processo? A resposta é a soma de duas normais $\phi(0, 1)$ que resulta em $\phi(0, 2)$. Se assim continuarmos a fazer previsões para T períodos a frente, nossa distribuição terá densidade $\phi(0, T)$.

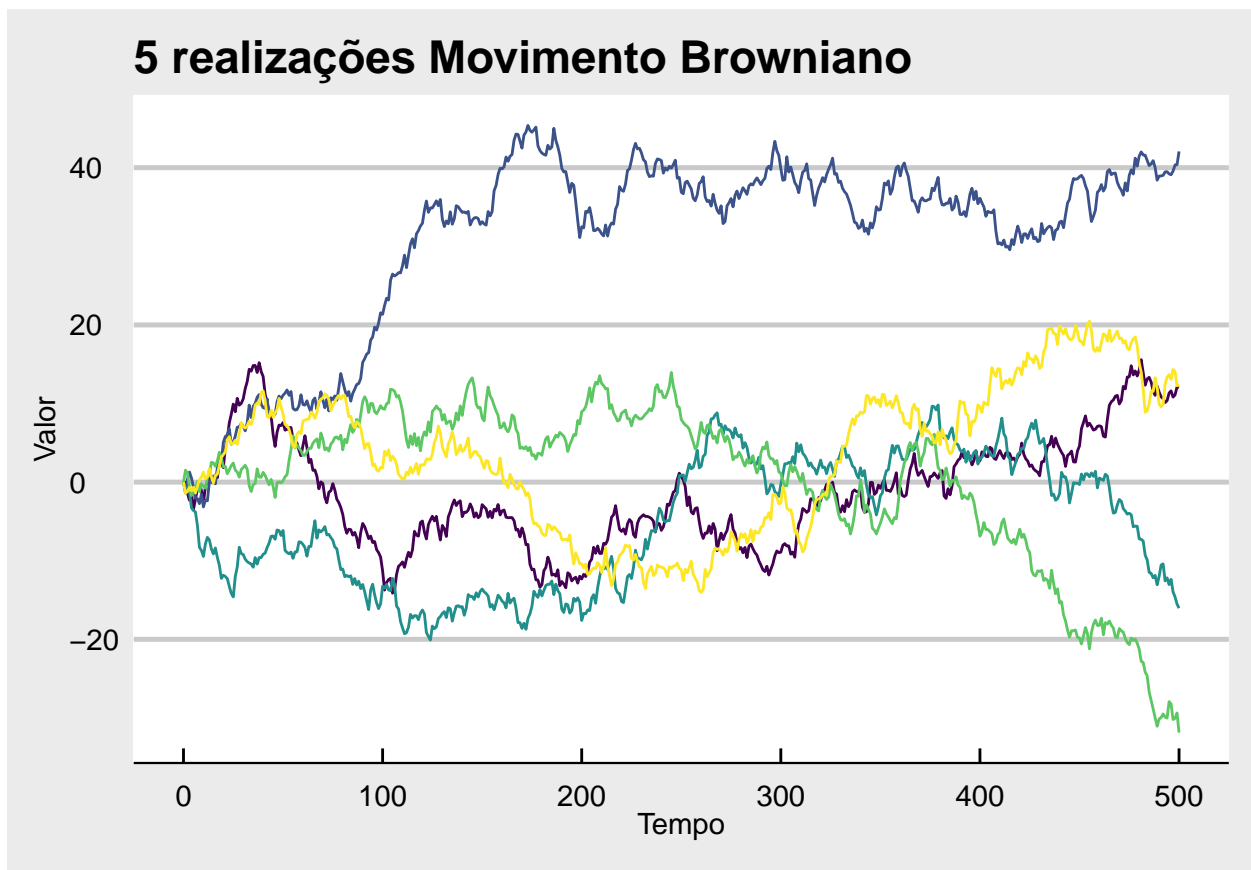
Para tempos discretos, $T \in \mathbb{Z}$ este é o processo do passeio aleatório (*Random Walk*), entretanto para tempo contínuo quando $T \in \mathbb{R}$ com incrementos acontecendo em intervalos de tempo infinitesimalmente pequenos, este é o Movimento Browniano - MB, que também é largamente conhecido como processo de Wiener.

```
t <- 0:500
m <- 5 # Numero de simulacoes
mc_names <- paste0("Sim", seq_len(m)) # Nomes das simulacoes
sigma2 <- 1
mu <- 0
brown_t <- matrix(nrow = length(t), ncol = m)
# MC Simulation
set.seed(543210)
for (i in seq_len(m)) {
  increments <- rnorm(length(t) - 1, mu, sqrt(sigma2))
  brown_t[, i] <- c(0, cumsum(increments))
}

colnames(brown_t) <- mc_names
mb <- as.tibble(cbind(t, brown_t)) %>%
  gather(key = sim, value = value, -t)
```

```
## Warning: `as.tibble()` is deprecated, use `as_tibble()` (but mind the new semantics).
## This warning is displayed once per session.
```

```
ggplot(mb, aes(x = t, y = value, color = sim)) +
  geom_line() +
  # geom_point(data = dens_tbl, aes(x = x, y = y), size = 1) +
  labs(title = "5 realizações Movimento Browniano",
       x = "Tempo",
       y = "Valor") +
  guides(color = FALSE) +
  scale_color_viridis_d() +
  theme_economist_white()
```



A figura acima mostra 5 realizações de um *mesmo* processo estocástico com média zero e variância unitária. É importante frisar que o processo que gerou as cinco séries é exatamente o mesmo, sendo elas tão distintas umas das outras ou não. Esta é uma importante característica dos processos estocásticos nas aplicações reais, o que nós observamos é apenas uma realização do processo, dentre as infinitas possíveis.

2.3 Definição

Agora que já foi passada a intuição sobre processos estocásticos, pode-se partir para definições mais formais sobre estes processos. Vamos adotar a notação do cálculo para tanto, e generalizar nosso MB possibilitando-o que tenha média diferente de zero (μ) e variância qualquer (σ^2), mantendo estas constantes ao longo do tempo, entretanto. Desta forma um movimento Browniano com deriva pode ser descrito através da seguinte equação diferencial estocástica - EDE:

$$dX_t = \mu dt + \sigma dB_t \quad (2.1)$$

onde dB_t é nosso MB padrão em um intervalo de tempo infinitesimal, dt .

O processo X_t possui uma taxa de deriva (média instantânea) igual a μ e volatilidade instantânea igual a σ . Quando um PE possui deriva igual a zero, como nosso MB padrão, o valor esperado deste processo para qualquer período futuro será zero. Este fato deixa de ser verdade no processo generalizado, com taxa de deriva diferente de zero. Neste caso o processo evoluirá seguindo uma taxa crescente (se $\mu > 0$) ou decrescente (se $\mu < 0$). Assim é possível, a partir de um MB padrão, modelar outros PE que possuam tendência temporal e variâncias diferentes.

2.4 Movimento Browniano Geométrico

Apesar de o processo X_t ser bastante flexível e cobrir uma grande gama de usos, ele ainda não é adequado para modelar o preço de ativos, e isto se dá em função de o processo de Wiener, mesmo com deriva positiva, poder atingir valores negativos com probabilidade maior que zero. Isto implicaria na possibilidade do preço de uma ação ser negativo, algo que obviamente não ocorre. Além desta impossibilidade, existe um outro empecilho para se utilizar o MB para modelar o processo de preços, e este é a deriva constante μ com relação ao preço da ação.

A deriva pode ser interpretada como o valor esperado do retorno da ação em um dado período de tempo. Este retorno esperado ele é pode ser constante em termos percentuais (em um modelo simplificado), mas não em termos absolutos! Ou seja, dependendo do preço da ação, R\$ 1,00 ou R\$ 100,00, a deriva μ deve ser diferente para que em termos percentuais a relação seja constante.

A solução para estes dois problemas é modelar o preço como um processo estocástico conhecido como Movimento Browniano Geométrico. Ele difere do MB padrão pois assume que o **logaritmo** da variável aleatória possui distribuição Normal. O MBG é a resolução para a seguinte EDE:

$$dX_t = \mu X_t dt + \sigma X_t dB_t \quad (2.2)$$

Veja que este é basicamente o mesmo processo MB, porém a deriva, termo que multiplica dt , varia linearmente com o valor do processo (μX_t) assim como a volatilidade instantânea (σX_t).

A solução para esta EDE, para um valor inicial qualquer de X ($X_0 > 0$) é dada por:

$$X_t = X_0 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t \right) \quad (2.3)$$

A variável aleatória X segue um MB ao longo de uma trajetória exponencial. É fácil verificar que, por ser exponencial, X_t nunca terá valor negativo.

Esta é uma forma conveniente de representar a evolução de preços de um ativo pois naturalmente surge o conceito de retornos logarítmicos. O log-retorno de X é dado por $r_t = \ln(X_t/X_0)$ de onde inferimos que se o processo de formação de preço de um ativo segue um MBG, então seus log-retornos serão normalmente distribuídos com média $\mu - \frac{\sigma^2}{2}$ e volatilidade σ em uma unidade de período considerado. Se escalarmos o período de tempo considerado para T , temos então que os retornos logarítmicos do ativo X seguem a seguinte distribuição normal:

$$r_T \sim \phi \left(\left(\mu - \frac{\sigma^2}{2} \right) T, \sigma^2 T \right) \quad (2.4)$$

Abaixo apresentamos 5 realizações de um MBG com valor de deriva $\mu = 0,6\%a.p.$ e variância $\sigma^2 = 1\%a.p.$.

```
t <- 0:500
m <- 5 # Numero de simulacoes
mc_names <- paste0("Sim", seq_len(m)) # Nomes das simulacoes
sigma2 <- 0.01
mu <- 0.006 # c(0.098, 0.099, 0.1, 0.101, 0.102) / 10
brown_t <- matrix(nrow = length(t), ncol = m)
# MC Simulation
set.seed(1234567)
for (i in seq_len(m)) {
  log_ret <- rnorm(length(t) - 1, mu - (sigma2 / 2), sqrt(sigma2))
  brown_t[, i] <- c(1, cumprod(exp(log_ret)))
}
```



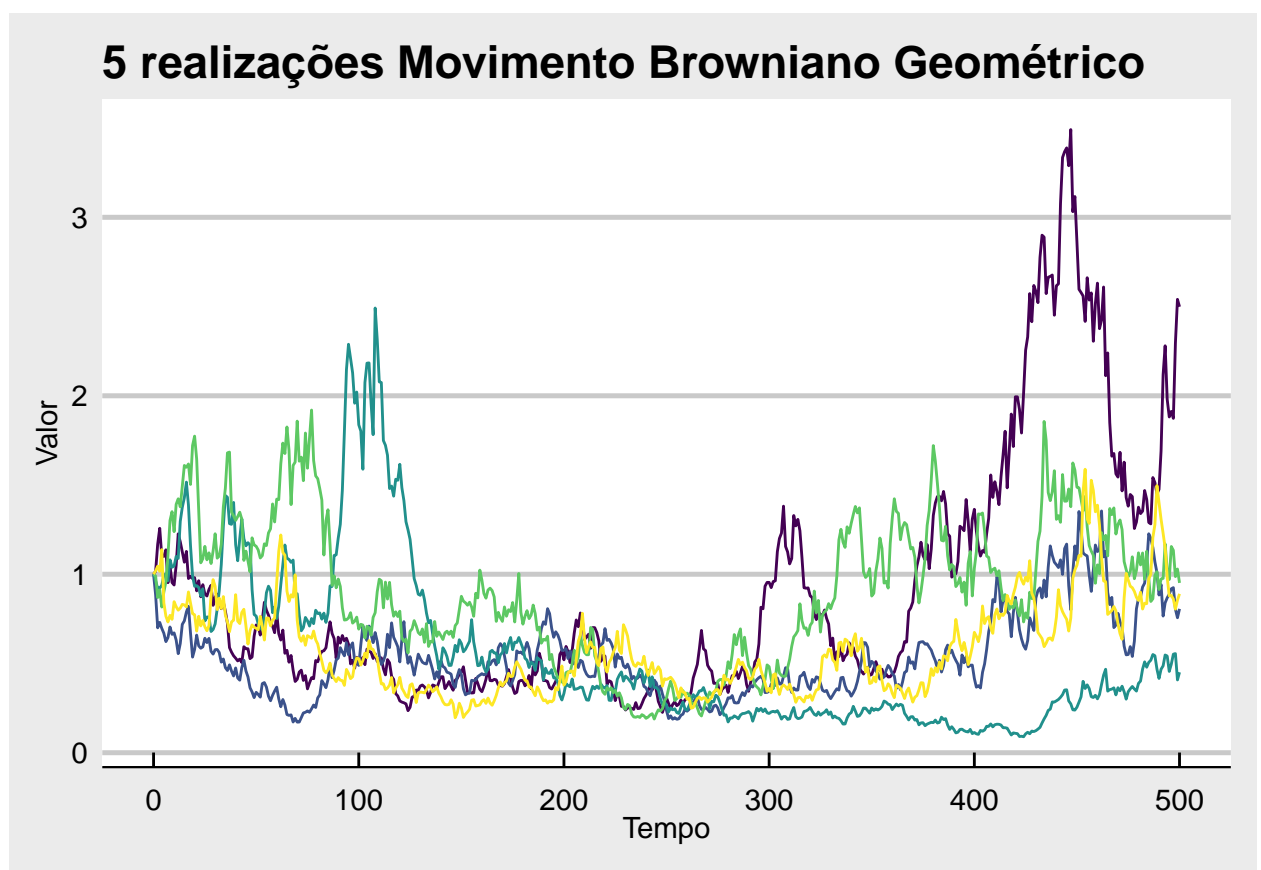
```

}

colnames(brown_t) <- mc_names
mb <- as.tibble(cbind(t, brown_t)) %>%
  gather(key = sim, value = value, -t)

ggplot(mb, aes(x = t, y = value, color = sim)) +
  geom_line() +
  labs(title = "5 realizações Movimento Browniano Geométrico",
       x = "Tempo",
       y = "Valor") +
  guides(color = FALSE) +
  scale_y_continuous(breaks = 0:10) +
  scale_color_viridis_d() +
  theme_economist_white()

```



O Movimento Browniano Geométrico aqui demonstrado serve de base para o famoso modelo **Black & Scholes** de precificação de opções, o qual assume que o ativo subjacente à opção (por exemplo, a ação de uma empresa) tem seu preço formado por um processo MBG.

Capítulo 3

Modelo Black-Scholes-Merton

Neste capítulo desenvolveremos o modelo para precificação de opções do tipo europeias proposto Black and Scholes (1973) e posteriormente expandido por Merton (1976). A derivação deste modelo se baseia nos conceitos apresentados sobre processos estocásticos do Capítulo 2.

Antes de entrar na modelagem desenvolvida pelos autores acima citados, iremos tratar de algumas definições essenciais, como por exemplo a precificação de ativos em um mundo neutro ao risco (*risk neutral valuation*) e portfólio de replicação (*replicating portfolio*), que são largamente utilizados na precificação de quaisquer derivativos, e não somente opções.

3.1 Portfólio de replicação

Suponha que se deseja precificar uma opção de compra sobre uma ação, vamos denotar o preço desta *call* de f_t . Sabemos que na data de expiração, T , da opção de compra seu preço será: $f_T = \max(S_T - K, 0)$, onde S_T é o preço da ação subjacente na data T e K é o preço de exercício da opção. Podemos criar um portfólio que envolva um ativo livre de risco, com preço B_t e a ação objeto do derivativo, S_t , que recrie o mesmo valor de pagamento que a opção na data de expiração. Ou seja, criamos o portfólio de replicação $P_t = \Delta S_t + bB_t$, no qual devemos escolher os valores de Δ e b de tal forma que $f_T = P_T$.

De fato, para apenas um período a frente, podemos tomar emprestado a taxa de juros livre de risco¹, r um valor igual ao preço corrente da ação, S_0 e comprá-la, ao mesmo tempo em que se “trava” (faz *hedge*) deste portfólio vendendo a opção. O lucro desta operação em T deve ser zero, pois é um portfólio travado. Com estas premissas é possível calcular o preço do prêmio da opção a ser cobrado no período inicial para que o valor esperado da operação como um todo seja zero.

Este portfólio formado pela venda do ativo livre de risco e compra do ativo objeto é denominado *portfólio de replicação*, pois, ele fornece um fluxo de caixa igual ao derivativo que buscamos precificar. Ao adicionarmos ao portfólio de replicação a venda (ou seja, o negativo) do derivativo, temos um portfólio *hedgado*, onde não existe mais incerteza com relação ao seu retorno esperado.

¹Esta é uma suposição do modelo, pode não ser verdade para algum investidor qualquer mas se for para algum outro investidor representativo, o princípio de ausência de arbitragem passa a valer, uma vez que este segundo investidor explorará o mercado e levará o preço do derivativo para o resultado requerido.

3.2 Precificação neutra ao risco

A metodologia de precificação de derivativos dentro do *mundo neutro ao risco*² é o carro-chefe das metodologias para se avaliar estes instrumentos. O princípio de precificação neutra ao risco afirma que um derivativo pode ser valorado através das seguintes suposições:

- o retorno esperado do ativo objeto é a taxa livre de risco, e
- o valor esperado do derivativo na expiração deve ser descontado pela taxa livre de risco para trazê-lo a valor presente.

É claro que o mundo real não é neutro ao risco, entretanto uma das provas que a metodologia faz é que, ao precificarmos um derivativo de forma relativa ao preço do ativo objeto, a precificação neutra ao risco encontra o mesmo valor para o derivativo que uma metodologia que leve em conta as preferências ao risco dos investidores. Entretanto, precificar um derivativo assumindo neutralidade ao risco é muito mais simples que encarar um modelo baseado no mundo real.

De fato a neutralidade ao risco soa um tanto quanto estranha a primeira vez. Porém, ela tem uma explicação lógica de sua validade nas circunstâncias em que é desenvolvida. Se os investidores forem avessos ao risco, por exemplo, os retornos esperados para o ativo objeto terão embutidos um prêmio pelo risco. Acontece que este prêmio pelo risco também deverá estar na taxa de desconto do valor esperado de pagamento do derivativo, de forma que este prêmio ao risco é cancelado.

É comum na literatura de derivativos encontrarmos os termos mundo-P e mundo-Q. O mundo-P se refere ao mundo real, com probabilidades P reais de ocorrência de eventos, enquanto que o mundo-Q é o mundo neutro ao risco, onde as probabilidades Q são ajustadas (tecnicamente suas medidas são alteradas) para refletir esta neutralidade. No mundo-Q é comum denotarmos o valor esperado de alguma variável aleatória com a seguinte notação: $\mathbb{E}_Q[\cdot]$.

Assim, a precificação de derivativos supondo um mundo neutro ao risco chega no preço correto para todos os mundos.

3.3 Encontrando a equação de Black-Scholes

Vamos partir do princípio que nossa ação, objeto da opção que desejamos precificar, siga um MBG conforme descrito no capítulo anterior. Portanto, o preço de nossa ação no período t deve observar a seguinte equação diferencial estocástica:

$$dS_t = \mu S_t + \sigma S_t dW_t \quad (3.1)$$

onde dW_t é o movimento Browniano, ou processo de Wiener.

Esta equação resume as principais hipóteses do modelo Black&Scholes de precificação de opções, são elas:

- O preço do ativo objeto é um processo estocástico e segue uma distribuição log-normal;
- O retorno esperado (μ) e a volatilidade (σ) deste ativo são **constantes**, tanto no tempo quanto com relação ao próprio nível de preço do ativo objeto.

Ademais destas hipóteses, temos aquelas relacionadas a racionalidade dos mercados e ao princípio de ausência de oportunidades de arbitragem. Estas hipóteses nos levam a validade do mundo neutro ao risco e portanto, a resolução do modelo da forma como descreveremos abaixo.

Nossa opção será descrita por um portfólio de replicação, e se tomarmos seu preço no instante t , então a opção também deve seguir uma equação diferencial estocástica da forma:

²A precificação dentro do mundo neutro ao risco mereceria um, ou mais, *posts* por si só. Aqui lançaremos apenas os principais resultados que nos permitem encontrar o modelo de B&S.

$$df_t = \Delta dS_t + b dB_t \quad (3.2)$$

aqui dB_t representa a variação do ativo livre de risco (dinheiro) dentro de um período de tempo dt . Sabemos que o valor do ativo livre de risco não envolve incerteza alguma, é determinístico, e seu rendimento é a taxa de juros livre de risco. Assim, para uma taxa continuamente composta r uma unidade de B evolui através de $B_t = e^{rt}$, logo:

$$dB_t = rBdt \quad (3.3)$$

De acordo com o lema de Ito o diferencial de uma função que dependa do tempo e de um processo estocástico pode ser encontrado da seguinte forma:

$$df_t = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial S} dS + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} dS^2 \quad (3.4)$$

onde dS^2 é a variação quadrática de nosso processo S_t que é um movimento browniano geométrico, logo $dS^2 = \sigma^2 S^2 dt$.

Igualando as equações (3.2) e (3.4), fazendo as devidas substituições trazidas pelas equações (3.1) e (3.3) e por fim rearranjando os termos, chegamos na seguinte relação:

$$\left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial S} \mu S + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 - \Delta \mu S - rbB \right) dt + \left(\frac{\partial f}{\partial S} \sigma S - \Delta \sigma S \right) dW = 0 \quad (3.5)$$

que para ser válida para todo t , cada termo entre parênteses deve ser igual a zero simultaneamente. Rapidamente chegamos aos valores necessários de nosso portfólio de replicação.

$$\Delta = \frac{\partial f}{\partial S} \quad (3.6)$$

$$rbB = \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 \quad (3.7)$$

Através da equação (3.2), integrando-a, chegamos na relação $bB = f - \Delta S$ e pré-multiplicando ambos os lados por r temos então que:

$$rbB = r(f - \Delta S) \quad (3.8)$$

Finalmente, igualando as equações (3.7) e (3.8) e substituindo o valor de Δ encontramos a famigerada equação diferencial parcial - EDP - de Black&Scholes:

$$\frac{\partial f}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - rf = 0 \quad (3.9)$$

3.4 Solução analítica

A equação possui diferentes formas de resolução³, a precificação do derivativo irá depender da forma que resolvermos a equação (3.9). O Modelo Black&Scholes é famoso por conseguir precificar opções call e put europeias, onde a resolução da equação fará uso dos *payoffs*. Para uma call:

³Uma demonstração completa de como encontrar a solução para a EDP de Black&Schole pode ser encontrada em <https://planetmath.org/AnalyticSolutionOfBlackScholesPDE>

$$f(S, T) = \max(S - K, 0) \quad (3.10)$$

Já no caso de uma put:

$$f(S, T) = \max(K - S, 0) \quad (3.11)$$

Onde:

- T é a data de vencimento da opção,
- K é o preço de exercício, e
- S é o preço do ativo subjacente.

Utilizando os *payoffs* dados nas equações (3.10) e (3.11), e resolvendo a EDP de Black&Scholes (3.9), iremos obter o modelo para se precificar os derivativos citados.

Para uma Call temos que:

$$C(S, t) = SN(d_1) - Ke^{-r(T-t)}N(d_2) \quad (3.12)$$

Já para um Put chegamos a:

$$P(S, t) = Ke^{-r(T-t)}N(-d_2) - SN(-d_1) \quad (3.13)$$

onde:

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \quad (3.14)$$

$$d_2 = \frac{\ln(S/K) + (r - \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} = d_1 - \sigma\sqrt{T - t} \quad (3.15)$$

É a partir das equações (3.12) e (3.13) que se obtém as chamadas gregas, que são as sensibilidades do preço do derivativo em relação a alterações nos parâmetros do modelo. Explicaremos as gregas em maiores detalhes mais adiante.

3.5 Paridade compra e venda

Agora vamos falar um pouco sobre a paridade entre opções de compra e venda, algo que nos ajuda a precificar algum derivativo quando já conhecemos o preço de outro derivativo com especificações semelhantes.

Assumindo ausência de oportunidade de arbitragem, com um ativo subjacente que tenha liquidez, iremos verificar a paridade Call-Put que define uma relação entre os preços de uma call e put do tipo europeu, desde que possuam o **mesmo tempo de maturidade** e **preço de strike**. Aqui possuiremos menos premissas que o modelo Black&Scholes e premissas mais simples, tal relação poderá ser utilizada para encontrar o valor justo de uma opção. Com a ausência de arbitragem observe que dois portfólios que sempre geram o mesmo payoff em um instante T devem ter o mesmo valor em qualquer instante intermediário.

Imagine um primeiro portfólio onde o investidor compre uma opção de compra C, a qual possui tempo de maturidade T e preço de strike K, sobre algum ativo subjacente que chamaremos de S e compre um título B que seu valor no período T seja de \$30. E um segundo portfólio onde tenha o próprio ativo subjacente S que esteja sendo negociado a \$25 e compre uma put P com um preço de strike K e maturidade T.

Estes são portfólios de replicação entre si, eles terão exatamente o mesmo *payoff* no período T, e na ausência de arbitragem, podemos calcular seus retornos através da precificação neutra ao risco. Teremos uma relação onde os valores dos nossos portfólios serão:

$$\mathbb{E}_{\mathbb{Q}}[S_t + P_t] = \mathbb{E}_{\mathbb{Q}}[C_t + B_t]; \quad t \leq T \quad (3.16)$$

Como estamos em um mundo neutro ao risco, P_t e C_t são os preços da put e da call dados pelo modelo B&S, enquanto $\mathbb{E}_{\mathbb{Q}}[B_t]$ se resume a $Ke^{-r(T-t)}$, ou seja, a posição atual em ativo livre de risco deve ser o valor presente do strike das opções. O preço do subjacente é o próprio preço atualmente observado.

Como S_t e $B_t = Ke^{-r(T-t)}$ são conhecidos, se no mercado existir um preço C_t então podemos calcular P_t e vice-versa. Esta é a essência da paridade compra e venda.

Observe que se essa relação não for mantida, teremos arbitragem: Suponhamos que $S_t > B_t$ e que a put esteja com um prêmio mais alto que a call no entanto ambos possuem um preço de strike \$30 e maturidade T, que os prêmios sejam \$20 e \$15 respectivamente teremos então: $30 + 20 = 15 + 25$, nesse caso teremos $50 \neq 40$, então vende-se o que está mais caro, no caso a ação e a put e compraria o título e a call, chegando ao um lucro de $50-40=10$, de \$10 sem risco. Observe que independentemente de a ação subir acima ou cair abaixo do strike, o lucro obtido nesta operação será sempre de \$10.

3.6 As Gregas

As letras gregas utilizadas no mercado de opções são usadas para denotar as sensibilidades do preço da opção com relação a variação de alguma das variáveis do modelo, a seguir entraremos em detalhes sobre as principais gregas usadas para a análise de opções.

3.6.1 Delta

Já havíamos definido o delta anteriormente ao encontrarmos a EDP de Black&Scholes, isto é:

$$\Delta = \frac{\partial f}{\partial S} \quad (3.17)$$

O delta mede a taxa com que o preço da opção muda conforme o valor do ativo subjacente oscila, seu valor pode variar entre 1 e 0 para call e entre 0 e -1 para put.

Exemplo: imagine uma opção de compra de 100 ações que possui um delta de 0,72, caso o ativo subjacente aumente \$1 o valor da opção aumentará \$72 (pois representa 100 ações), e caso o valor do ativo subjacente diminua o preço da opção diminuirá \$72.

O mesmo acontece com opções de venda, só que de uma forma inversamente proporcional pelo seu delta ser negativo. Uma opção de venda de 100 ações que possui um delta de -0,40 vai valer menos \$40 caso o preço do seu ativo subjacente aumente \$1, e vai valer mais \$40 caso o preço do seu ativo subjacente diminua \$1.

O strike da opção influencia o delta diretamente, quanto mais in-the-money for uma opção, call ou put, maior será o seu delta em valor absoluto, e quanto mais out-the-money, menor será o módulo de seu delta.

3.6.2 Gamma

$$\Gamma = \frac{\partial \Delta}{\partial S} = \frac{\partial^2 f}{\partial S^2} \quad (3.18)$$

O gamma mede a taxa com que o delta muda a cada oscilação do preço do ativo subjacente, seu impacto aumenta conforme o valor atual do ativo se aproxima do strike, sendo responsável pela convexidade do valor da opção. Opções de alto gamma são chamadas de explosivas, uma vez que mesmo pequenas variações no preço do ativo se traduzem em grandes oscilações no preço da opção.

3.6.3 Theta

$$\Theta = -\frac{\partial f}{\partial \tau}, \quad \tau = T - t \quad (3.19)$$

O Theta é uma taxa que mede o efeito do tempo sobre o preço da opção. Como as opções tem um maior valor de acordo com a quantidade de tempo até a data de vencimento, seu valor também diminui conforme o tempo passa, theta é a letra que mede essa variação (que sempre é negativa). O valor de theta representa a quantidade de dinheiro perdida no prêmio da opção a cada dia que passa.

3.6.4 Vega

$$\mathcal{V} = \frac{\partial f}{\partial \sigma} \quad (3.20)$$

O vega é uma taxa que mede o efeito da mudança da volatilidade no preço da opção. Seu valor é praticamente constante em opções com a mesma data de vencimento, esse valor aumenta em datas de vencimento mais distantes devido ao maior espaço de tempo em que a volatilidade atua, devido ao maior intervalo de tempo que há para ocorrer mudanças nos preços.

3.6.5 Rho

$$\rho = \frac{\partial f}{\partial r} \quad (3.21)$$

O rho é uma taxa que mede a sensibilidade do preço da ação em relação à taxa livre de risco, caso o valor de rho determinada opção seja 0,7, para cada aumento de 1% da taxa livre de risco o valor da opção aumentará 0,7%. Essa taxa influi principalmente no preço de opções com uma data de vencimento extremamente distante, não afetando muito o preço de opções cuja data de vencimento é próxima.

Capítulo 4

Smile de Volatilidade

A volatilidade instantânea, σ , do ativo subjacente é a única variável no modelo B&S que não pode ser diretamente observada. De fato, a volatilidade (ou equivalentemente a variância) de um ativo é dita uma variável **latente**. Sabemos que ela existe e possui algum valor no processo gerador, o processo pelo qual os preços são formados, porém não conseguimos observá-la diretamente, apenas estimá-la. Uma das formas de estimação de volatilidade pode ser a partir de dados históricos, mas várias outras formas existem, entre elas processos GARCH, volatilidade realizada, volatilidade estocástica, etc.

Uma vez que a volatilidade não pode ser diretamente observada, a prática comum no mercado é fazer o caminho inverso. Considerar os preços de mercado para as opções como dado, e a partir do modelo B&S inverter a equação de preço da Call ou Put para encontrar a volatilidade deste modelo que é compatível com os preços de mercado. A esta volatilidade encontrada damos o nome de **volatilidade implícita**.

Portanto, o smile de volatilidade que tratamos neste post é na verdade um gráfico entre a volatilidade implícita, retirada de opções Européias (baunilhas, do inglês vanilla options) a partir do modelo B&S, contra os *strikes* destas opções.

4.1 Reparametrizando B&S e definição de moneyness

Nem sempre é interessante plotar o smile contra os *strikes* propriamente ditos, uma forma de avaliar o quanto uma opção está “dentro, fora ou no dinheiro” pode ser a grega Delta ou então o chamado *moneyness* (por favor, se alguém tiver uma boa tradução para este termo, deixe nos comentários). Tradicionalmente a medida de *moneyness* é a relação K/S , ou seja o strike contra o preço corrente do subjacente. Porém existem outras definições mais interessantes para se trabalhar, entretanto, antes devemos fazer uso de algumas definições e vamos reparametrizar as expressões $d1$ e $d2$ do modelo B&S.

Lembrando que em precificação de opções estamos no mundo neutro ao risco, vamos definir o valor *forward*, F do subjacente como o valor corrente composto pela taxa livre de risco até a maturidade da opção, ou seja:

$$F = e^{r\tau} S \quad (4.1)$$

A **volatilidade (implícita) total** pode ser definida como a volatilidade reescalada pela raiz do tempo, que nos dá uma informação da volatilidade esperada para o subjacente do período corrente até a maturidade. Da mesma forma, a **variância total**. Denotando a volatilidade total por θ e a variância total por w , temos:

$$\theta = \sigma_{imp} \cdot \sqrt{\tau} \quad (4.2)$$

e

$$w = \sigma_{imp}^2 \cdot \tau \quad (4.3)$$

Vamos também definir a medida *forward log-moneyness* e denotá-la por k . Esta será a medida de *moneyness* que iremos utilizar ao longo deste e de outros artigos, portanto iremos utilizar este termo para designar o *forward log-moneyness* a não ser que expresse de forma contrária no texto.

$$k = \ln\left(\frac{K}{S}\right) - r\tau = \ln\left(\frac{K}{F}\right) \quad (4.4)$$

Logo, o *strike* está relacionado ao *moneyness* de forma que: $K = Fe^k$.

Podemos agora reparametrizar as expressões $d1$ e $d2$ do modelo B&S de forma que serão mais facilmente trabalhadas em modelos de volatilidade. Lembrando destas expressões que já foram apresentadas no Capítulo 3:

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(\tau)}{\sigma\sqrt{\tau}} \quad (4.5)$$

$$d_2 = \frac{\ln(S/K) + (r - \sigma^2/2)(\tau)}{\sigma\sqrt{\tau}} = d_1 - \sigma\sqrt{\tau} \quad (4.6)$$

Substituindo as expressões para *forward log-moneyness* e volatilidade total nas definições acima temos as novas parametrizações para $d1$ e $d2$:

$$d_1 = -\frac{k}{\theta} + \frac{\theta}{2} \quad (4.7)$$

$$d_2 = -\frac{k}{\theta} - \frac{\theta}{2} = d_1 - \theta \quad (4.8)$$

Retomando o valor da opção do tipo Call no modelo B&S, podemos reescrever sua fórmula de apreamento da seguinte forma:

$$\begin{aligned} C(K, \tau) &= SN(d1) - Ke^{-r\tau}N(d2) \\ e^{r\tau}C(K, \tau) &= FN(d1) - KN(d2) \\ C_B &= F[N(d1) - e^k N(d2)] \end{aligned} \quad (4.9)$$

Esta equação é conhecida como a forma de Black de precificação (*Black Call price formula*), que relaciona os valores *forward* da opção (também conhecido como valor não descontado), do subjacente e do *strike*. Esta formulação é particularmente útil quando formos extrair a distribuição neutra ao risco do subjacente que está implícita nos preços de mercado das opções.

4.2 Características de smiles de volatilidade

Caso o modelo de Black, Schole e Merton estivesse em acordo com a realidade, e os ativos tivessem seus preços formados a partir de um verdadeiro MBG, a volatilidade implícita seria uma constante. O gráfico do smile de volatilidade seria uma reta horizontal, com a mesma volatilidade para qualquer nível de *moneyness* e se considerarmos a superfície toda (que leva em conta os diversos tempos para expiração) esta seria paralela ao domínio (k, τ) . Não estaríamos escrevendo (e você lendo) este artigo se este fosse o caso...

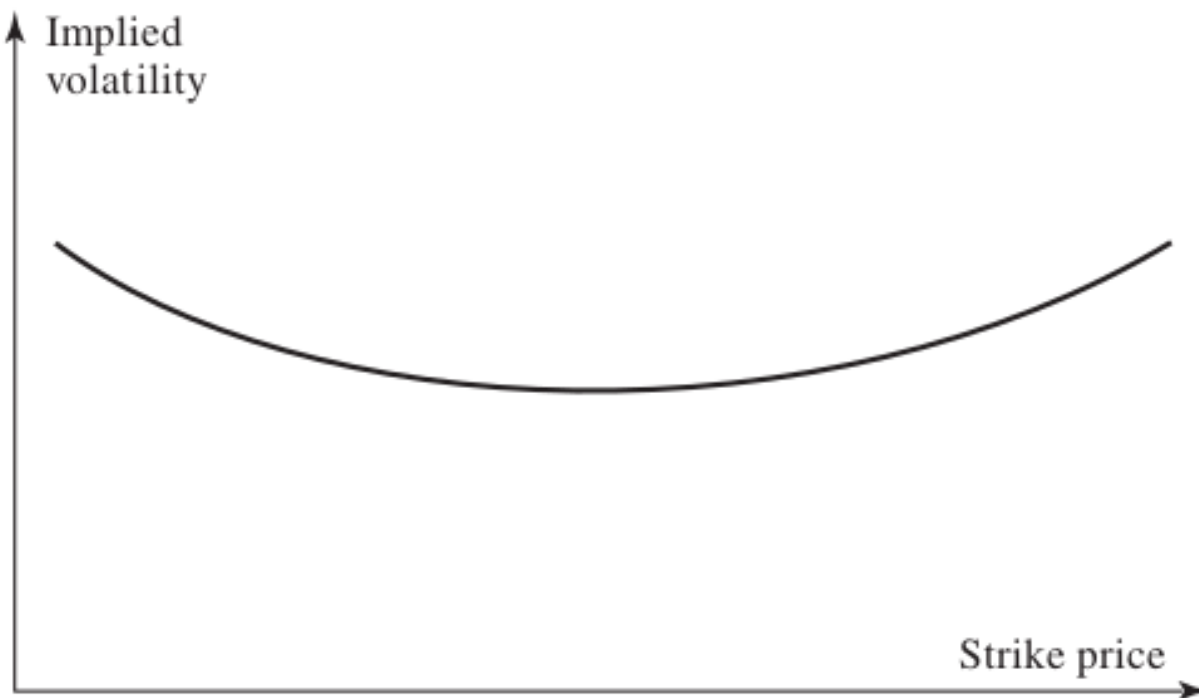


Figura 4.1: Smile típico de um mercado cambial.

O fato é que o modelo B&S é um modelo muito restritivo, com inúmeras suposições que não se verificam no mundo real e que por conseguinte, tornam os resultados do modelo pouco acurados. Entretanto este é um modelo muito conhecido, de fácil assimilação por parte dos agentes de mercado e que virou a língua franca nos mercados de derivativos. Se todos os *traders* conversarem em termos do modelo B&S, todos se entenderão, mesmo que internamente cada um possua seu próprio modelo de apreçamento.

Entre as características tipicamente observadas em smiles (e superfícies) de volatilidades pode-se citar:

- As volatilidades implícitas variam conforme o *strike* e prazo de expiração
- Smiles apresentam *skew*. Maior inclinação em uma das asas, representando uma maior probabilidade daqueles *strikes* acontecerem
- Smiles de equity tipicamente são negativos
- Mercados diferentes apresentam padrões de smile diferentes

4.2.1 Mercados cambiais

Opções sobre moedas possuem tipicamente um smile de volatilidade conforme mostrado na figura 4.1 abaixo. A volatilidade implícita é relativamente baixa para opções ATM. Esta torna-se progressivamente maior quando a opção se move para dentro do dinheiro ou para fora.

Caso a distribuição dos preços do ativo subjacente, neste caso uma taxa de câmbio fosse perfeitamente log-normal como no modelo B&S, o smile não teria esta curvatura. Desta forma podemos afirmar que o mercado, ao precificar as opções, acredita que a distribuição deste ativo possui caudas com maior densidade que supõe a log-normal, existem maiores probabilidades de retornos muito baixos ou muito altos.

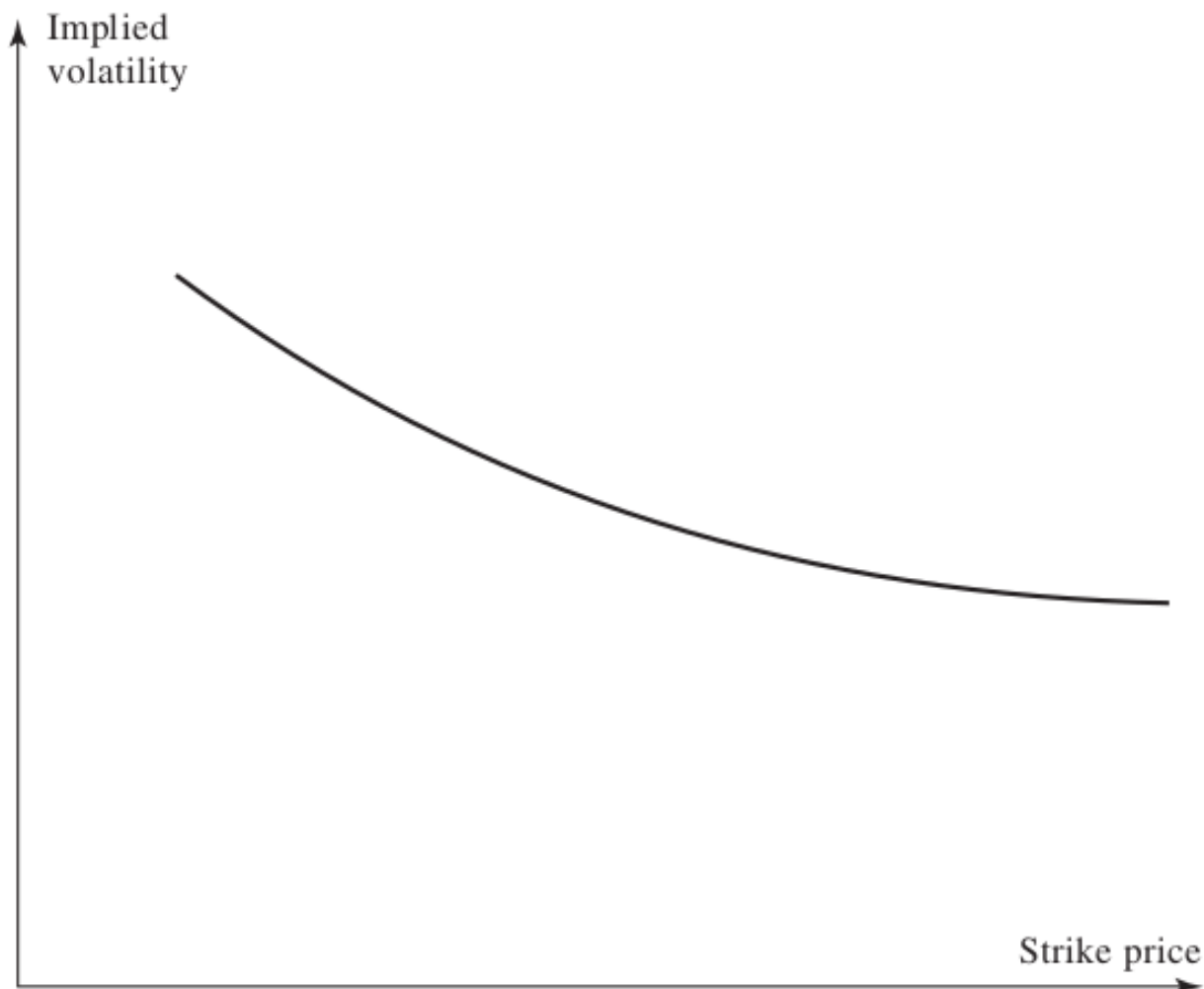


Figura 4.2: Smile típico de uma ação ou índice de ações.

4.2.2 Mercados de *equities*

Nos mercados de *equities*, ações, índices de ações e ETFs, por exemplo, o smile apresenta uma característica de assimetria (skew, em inglês) negativa. A asa esquerda (parte onde as puts estão fora do dinheiro) apresenta valores de volatilidade implícita muito maiores que suas contrapartes no lado das calls. Este comportamento reflete a percepção de mercado de uma maior probabilidade de grandes perdas nas ações que altos ganhos, gerando portanto, uma distribuição de preços assimétrica. Como existe uma maior probabilidade de perdas extremas, o seguro para estas, ou seja, uma put é relativamente mais cara que uma call.

4.3 Smile como forma de precificação

Analisando a equação de B&S com a parametrização para $d1$ e $d2$ dada no início deste artigo é possível verificar que existe uma relação direta entre volatilidade implícita e preço de uma opção, seja esta uma call ou put.

Como $d1$ é estritamente crescente em θ e $d2$ é estritamente decrescente e ao mesmo tempo o preço de uma opção é crescente em $d1$ e decrescente em $d2$, logo, temos uma relação direta entre o preço de uma opção e sua volatilidade implícita para uma dada maturidade. Em outras palavras, em um smile, tudo o mais constante, quanto maior a volatilidade implícita maior o preço da opção naquele *strike*.

Outra forma de verificar esta relação é perceber que a grega Vega, que é calculada da mesma forma para calls e puts, é sempre positiva. Ou seja, um aumento no valor da volatilidade sempre leva a elevações no preço de uma opção.

Desta forma é normal entre os praticantes de mercado fazer a precificação de opções em termos de “pontos de volatilidade” e não em valores monetários propriamente ditos. Isto porque o modelo B&S, apesar de não ser o modelo correto (nenhum é) para a precificação de opções, é conhecido e de fácil entendimento para todos. Então todos os praticantes podem fazer suas cotações em termos de volatilidades implícitas, que são extraídas de opções baunilhas com o modelo B&S, e somente na hora de fechar um negócio e liquidar o pagamento, o preço efetivo a ser pago é acordado entre as partes.

4.4 Estrutura a termo

O mercado precifica a volatilidade implícita de forma que esta dependa também do tempo até expiração, bem como do preço de exercício, agregando uma segunda dimensão ao smile e transformando-o na famigerada **superfície de volatilidade implícita**.

A volatilidade implícita tende a ser uma função crescente da maturidade quando as volatilidades de curto prazo são historicamente baixas e função decrescente da maturidade quando as volatilidades de curto prazo são historicamente altas. Isso porque existe uma expectativa de reversão a uma média de longo prazo embutida na volatilidade. Esta característica é explorada explicitamente por alguns modelos de volatilidade, como em Heston (1993).

As superfícies de volatilidade combinam smiles com a estrutura a termo de volatilidade para tabular valores apropriados para precificar uma opção com qualquer preço de exercício e prazo de expiração.

Da mesma forma como a curva de juros em um dado momento é uma descrição concisa dos preços dos títulos negociados naquele mercado, assim, para um ativo subjacente em particular em determinado momento, a superfície de volatilidade implícita fornece uma descrição resumida de seu mercado de opções. Considerando que os rendimentos dos títulos são diferenciados pelo seu tempo até o vencimento, as opções são diferenciadas por seu tempo até a expiração e o *strike*, logo requerem uma superfície ao invés de uma curva.

A figura 4.3 demonstra uma superfície de volatilidade implícita do SPX em 15/09/2005, conforme apresentado em Gatheral (2011).

4.5 Arbitragem estática

Antes de definir o que é arbitragem estática que pode estar presente em uma superfície de volatilidade (ou na superfície de preço de opções), vamos partir para a intuição por trás desta definição.

O princípio de ausência de arbitragem é dominante na teoria financeira. Este princípio nos informa que não deve existir lucro sem que se incorra em algum tipo de risco, o lucro sempre é a remuneração do investidor que aceitou carregar alguma forma de risco durante o investimento. Portanto, não devem existir perfis de lucro acima da taxa livre de risco (*payoffs* positivos) com probabilidade de 100%.

Primeiro consideramos uma trava de alta com opções do tipo call. Excluindo-se os custos de transação, esta operação sempre oferece um retorno positivo ou zero, conforme a figura 4.4. Por mais que esta estratégia esteja montada fora do dinheiro, sempre existe uma possibilidade de ela ter lucro, $S_T > K$ e portanto seu preço deve ser sempre maior que zero.

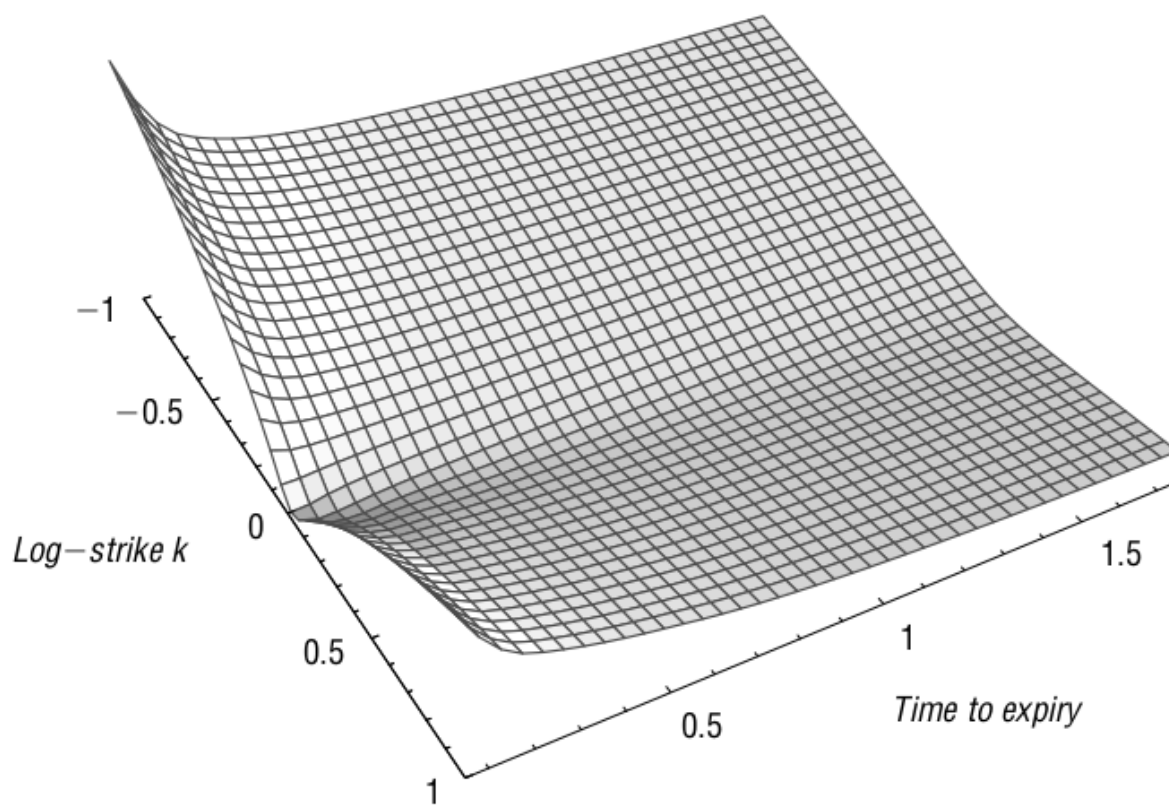


Figura 4.3: Superfície de volatilidade implícita.

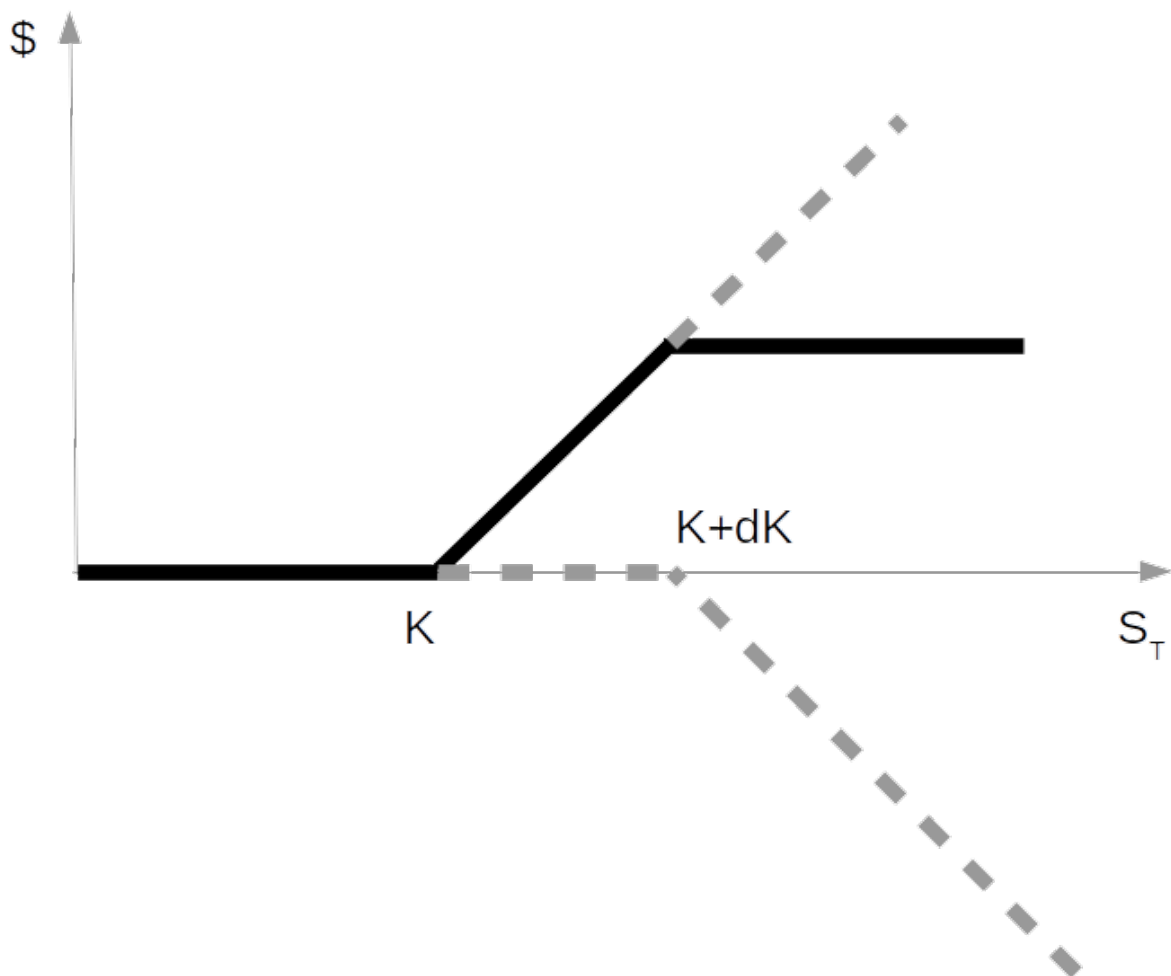


Figura 4.4: Perfil de lucro de uma trava de alta.

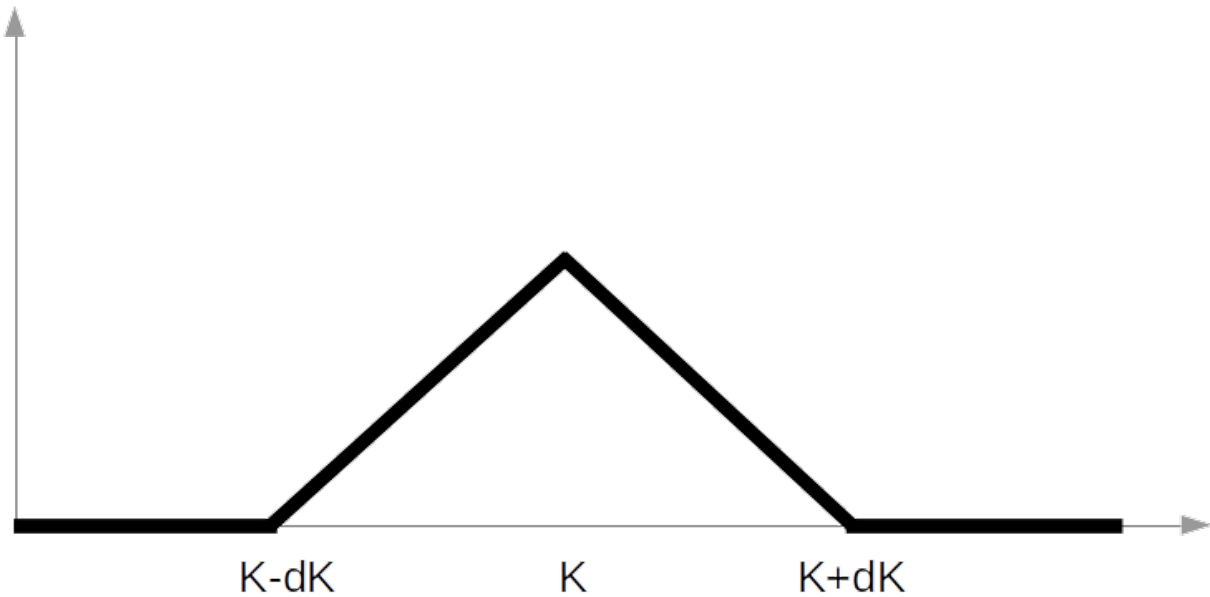


Figura 4.5: Borboleta realizada com calls.

É claro que quanto mais ITM estejam as opções, maior seu preço e quanto mais fora do dinheiro menor será seu valor até o limite inferior zero. Se levarmos a diferença entre os *strikes*, dK a zero temos que:

$$\frac{\partial C}{\partial K} \leq 0 \quad (4.10)$$

Este é o limite de arbitragem para travas de alta ou, mais conhecido pelo termo em inglês *call spread no-arbitrage* e impõe que os preços das calls devem ser uma função decrescente no *strike*. De forma equivalente e através da paridade compra-venda este limite de arbitragem para as puts é:

$$\frac{\partial P}{\partial K} \geq 0 \quad (4.11)$$

4.5.1 Arbitragem de borboleta

Também deve ser imposta uma restrição na segunda derivada do preço das opções em relação ao *strike*, e esta é conhecida como limite de arbitragem para borboletas. Vejamos porquê.

Considere uma estratégia do tipo borboleta, onde se compra uma quantia de calls no *strike* $K - dK$, vende-se duas vezes esta quantia em K e compra-se novamente um lote em $K + dK$, o perfil de lucro desta operação no vencimento está representado na figura 4.5.

Seguindo a mesma linha de raciocínio anterior, como o *payoff* da borboleta é sempre não negativo também deve ser o seu valor para qualquer período anterior a expiração. Se denotarmos π_B o valor da borboleta, então $\pi_B \geq 0$.

Agora imagine que escalamos a estratégia de forma que um lote de compras (na venda são dois lotes) seja de tamanho $1/dK^2$, o valor para a montagem desta operação deve ser, portanto:

$$\pi_B = \frac{C(K - dK) - 2C(K) + C(K + dK)}{dK^2} \quad (4.12)$$

E se levarmos ao limite em que $dK \rightarrow 0$, a equação acima torna-se justamente a segunda derivada do preço da call no *strike* K .

$$\begin{aligned} \frac{\partial^2 C(K)}{\partial K^2} &= \pi_B \\ &\geq 0 \end{aligned} \quad (4.13)$$

Ou seja, os preços das calls são uma função **convexa** nos *strikes*. O mesmo raciocínio pode ser feito para uma borboleta com puts e o resultado será equivalente, o preço das puts também deve ser uma função convexa nos *strikes*.

4.5.2 Arbitragem de calendário

Passamos agora a analisar os limites de arbitragem na estrutura a termo da superfície de volatilidade. A arbitragem de calendário normalmente é expressa em termos de monotonicidade dos preços em relação ao período para expiração. Ou seja, quanto maior o prazo de maturidade de uma opção para um mesmo preço de exercício, maior deve ser seu valor.

É fácil de entender este limite com base nas probabilidades de exercício. Como sabemos, em um processo estocástico do tipo MBG a variância do processo cresce conforme a **raiz do tempo**, $\sqrt{\tau}$. Quanto maior a variância do ativo subjacente, maior a probabilidade deste alcançar um determinado preço, mais elevado ou não. Assim, seja uma call ou put OTM quanto mais distante estiver seu prazo de maturidade, maior a probabilidade de exercício e portanto, maior seu valor.

Dado que a relação de **volatilidade total** implícita e preço de uma opção também é direta e positiva, conforme demonstrado na Seção 4.3, segue que a volatilidade total deve ser não decrescente no tempo para expiração.

Esta relação pode ser expressa através da seguinte equação para uma call precificada através de B&S:

$$\frac{\partial C_{BS}(k, \theta(\tau))}{\partial \tau} = \partial_\theta C_{BS} \cdot \partial_\tau \theta \geq 0 \quad (4.14)$$

onde $\partial_\theta C_{BS}$ é a derivada parcial do preço da call em relação a volatilidade total implícita, que já demonstramos ser positiva e $\partial_\tau \theta$ é a derivada parcial da volatilidade total implícita em relação ao tempo para maturidade que, portanto, deve ser maior ou igual a zero para obedecer a restrição imposta ao preço da call.

4.6 Limites de inclinação

Se mantivermos a volatilidade implícita constante para todos os *strikes*, os preços das calls no modelo B&S devem ser decrescentes. Por outro lado, para um *strike* fixo, o preço de uma call se eleva à medida que a volatilidade implícita aumenta. Suponha por um momento que a volatilidade implícita varia com o *strike* como é o caso nos smiles. À medida que o *strike* aumenta, se a volatilidade implícita aumentar muito rapidamente, seu efeito sobre o preço da call pode mais que compensar o declínio no preço devido a elevação do preço de exercício e, assim, levar a um aumento líquido no preço da opção. Isso violaria o requisito de que $\partial C / \partial K \leq 0$ e, portanto, leva a um limite superior na taxa em que a volatilidade implícita pode aumentar com o *strike*.

Novamente, o mesmo raciocínio pode ser imposto para o lado das puts. A volatilidade implícita não pode se elevar tão rapidamente quando os *strikes* se reduzem de forma que uma put de *strike* menor tenha valor mais elevado que outra que esteja mais próxima do dinheiro.

Finalmente, um sumário dos limites impostos a uma superfície de preços de opções (calls no caso apresentado), que implicam em limites para a superfície de volatilidade é apresentado abaixo¹:

1. $\partial_\tau C \geq 0$
2. $\lim_{K \rightarrow \infty} C(K, \tau) = 0$
3. $\lim_{K \rightarrow -\infty} C(K, \tau) + K = a, \quad a \in \mathbb{R}$
4. $C(K, \tau)$ é convexa em K
5. $C(K, \tau)$ é não-negativa

4.7 Distribuição implícita

O modelo B&S é baseado na suposição que o ativo subjacente segue uma distribuição log-normal em seus preços. Caso esta suposição fosse de fato realizada no mercado, o smile de volatilidade seria uma reta completamente horizontal, não haveria variação na volatilidade implícita conforme o preço de exercício. Entretanto, esta não é a realidade dos smiles e podemos fazer a pergunta inversa portanto, qual a distribuição neutra ao risco que está implícita no smile de volatilidade?

Certamente não é uma log-normal. Na verdade, a densidade da distribuição que está implícita em um smile nada mais é que a convexidade deste smile, ou seja, sua segunda derivada em relação ao *strike*. Esta distribuição implícita também é por vezes chamada de RND (*risk neutral density*) e é muito útil para fazer a precificação de outras opções que não são observadas no smile ou extrair probabilidades de ocorrência de eventos precificadas pelo mercado.

Pode-se obter este resultado a partir da definição do valor de uma call e é conhecido como a fórmula de Breeden and Litzenberger (1978). O valor de uma call é o valor esperado do *payoff* terminal desta call ponderado pela densidade neutra ao risco do subjacente. Ou seja:

$$C(S, t) = e^{-r\tau} \int_0^\infty p(S, t, S_T, T) \max\{S_T - K, 0\} dS_T \quad (4.15)$$

onde $p(\cdot)$ é a densidade neutra ao risco e estamos supondo uma taxa de juros livre de risco constante durante o período de vida da opção. Como o *payoff* da call é não linear, sendo zero para qualquer valor de $S_T \leq K$ e igual a $S_T - K$ quando $S_T > K$, podemos escrever esta equação como:

$$C(S, t) = e^{-r\tau} \int_K^\infty p(S, t, S_T, T) (S_T - K) dS_T \quad (4.16)$$

que pode ser rearranjada, com alguma simplificação na notação, da seguinte forma.

¹Retirado de Aurell (2014), p. 25.

$$\begin{aligned}
\frac{\partial C}{\partial K} &= -e^{-r\tau} \int_K^\infty p(S_T) dS_T \\
e^{r\tau} \frac{\partial C}{\partial K} &= \int_{-\infty}^K p(S_T) dS_T \\
e^{r\tau} \frac{\partial^2 C}{\partial K^2} &= p(K) \\
\frac{\partial^2 C_B}{\partial K^2} &= p(K)
\end{aligned} \tag{4.17}$$

Onde usou-se a notação C_B para denotar a formulação de Black para o preço de uma call. Ou seja, a segunda derivada em relação ao strike do preço não descontado de uma call é a distribuição neutra ao risco do ativo subjacente, e é válida para todos preços de exercício.

Portanto, se desejarmos saber qual a distribuição de probabilidades de preços do ativo subjacente em uma data futura que possua vencimento de opções, basta encontrarmos a convexidade do smile dos preços *forward* daquele vencimento².

4.8 Conclusão

O modelo de Black-Scholes-Merton, pode ser considerado a pedra fundamental para a precificação de opções. Entretanto, este modelo apresenta uma série de limitações que fazem com que os praticantes de mercado utilizem outras técnicas neste mercado. Uma destas é o uso do smile de volatilidade e sua interpretação como forma de precificar opções e extrair informações implícitas nos preços.

A assimetria do smile e suas asas informam que as distribuições de probabilidades para o ativo subjacente não são exatamente log-normais, e podem apresentar discrepâncias significativas, especialmente nas caudas da distribuição que muito interessam a gestão de risco, por exemplo.

Este foi um artigo denso, porém com vários conceitos importantes para a compreensão do comportamento da superfície de volatilidade. A estrutura a termo também é existente na volatilidade implícita e está limitada pela ausência de arbitragem do tipo calendário. O smile de volatilidade, que é uma fatia da superfície com prazo de expiração constante, possui suas próprias limitações de forma, com a ausência de arbitragem do tipo borboleta e limitações quanto a inclinação.

Por fim, foi demonstrado como a convexidade do smile de preços fornece a distribuição implícita para os preços do ativo subjacente para a data de expiração das opções.

²Simples em teoria, muito mais complicado na prática, com diversos problemas para a extrapolação do smile para *strikes* extremos.

Capítulo 5

Métodos de Calibração

Neste post iremos mostrar as diferenças existentes entre “interpolação”, “suavização” e “parametrização” de superfícies de volatilidade.

Como já apresentado em posts anteriores, existem diversas formas de interpolar, extrapolar, parametrizar e calibrar smiles de volatilidade. Existem vantagens e desvantagens para cada método.

5.1 Calibração versus Interpolação

Uma forma simples de gerar um smile de volatilidade a partir de dados observados no mercado é a **interpolação** destes dados. Diversas formas de interpolação existem, sendo talvez a mais conhecida a spline cúbica. Não é a proposta deste artigo detalhar os procedimentos de interpolação, restando saber que em tal procedimento é gerada uma função contínua em partes (piecewise) que **passa** por todos os pontos observados.

Uma interpolação força a passagem da função interpolada em todos os seus pontos de referência, como se estivesse ligando os pontos em um desenho a mão livre. Portanto, nestes pontos o erro da interpolação é zero por definição, entretanto em pontos intermediários podem surgir erros, inclusive aqueles que possibilitam a existência de arbitragem entre strikes de um mesmo smile¹.

Em contraposição a métodos de interpolação, podem ser derivados métodos de suavização (smoothing) ou então a parametrização do smile de volatilidade. Seja suavização, ou parametrização, estes métodos não forçam a passagem da função que representa o smile pelos pontos de mercado, mas buscam minimizar alguma função perda com relação aos desvios em relação a estes pontos ao mesmo tempo em que buscam “suavizar” o smile, para que este não apresente variações bruscas entre os strikes ou alterações de convexidade na curva de preços, que não são condizentes com a teoria de precificação de derivativos.

Um método paramétrico, como o SVI, Heston, SABR ou Volatilidade Local, busca ajustar às volatilidades implícitas observadas através dos preços das opções sendo praticados no mercado uma determinada função, que possui parâmetros em sua definição que por sua vez determinam a forma desta função. Ao se ajustar os parâmetros, pode-se adequar a função para ficar “o mais próxima possível” dos dados observados, sem necessariamente, no entanto, passar por todos estes pontos.

A figura abaixo tenta mostrar as diferenças entre uma interpolação spline cúbica, uma suavização e uma parametrização SVI. Enquanto que a interpolação liga todos os pontos marcados, a suavização e a parametrização não necessariamente passam sobre estes pontos mas fornecem uma curva mais “suave”, sem trocas de convexidade, o que gera oportunidades de arbitragem e probabilidades negativas de ocorrência de determinados preços para o ativo subjacente, que ferem os princípios de precificação de opções. Os dados

¹Veja mais detalhes no Capítulo Smile de Volatilidade

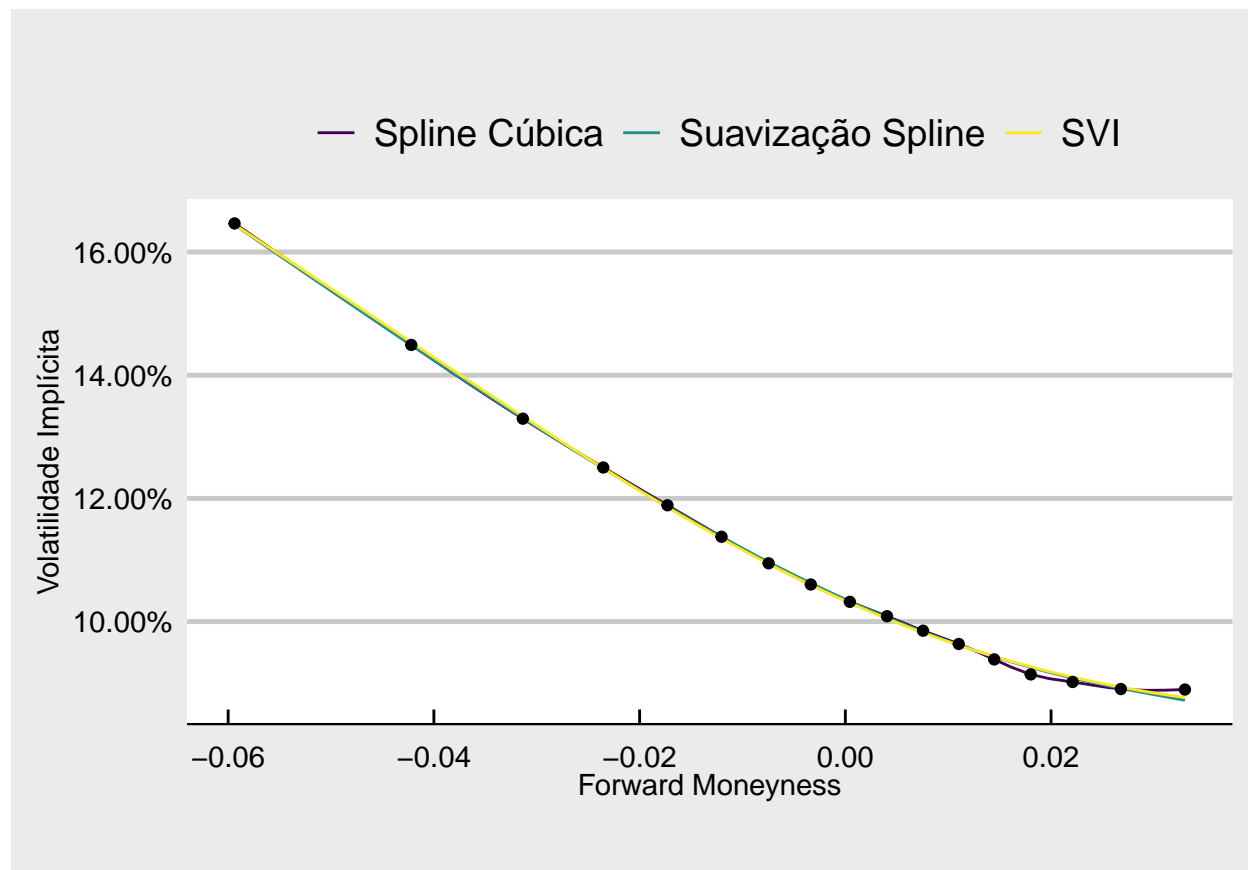


Figura 5.1: Diferentes métodos de ajuste de dados a um smile.

utilizados neste e nos próximos artigos sobre superfícies de volatilidade foram obtidos do site ivolatility.com na forma de amostra gratuita fornecida livremente. O ativo subjacente é o ETF IWM para a data de 21/09/2017.

Pode-se verificar como os métodos SVI e a suavização não passam sobre todos os pontos marcados, com a suavização tendo dificuldade com a curvatura nos valores mais altos de *moneyness* e a SVI possuindo uma inclinação mais branda na asa esquerda do smile.

5.2 Spline cúbica

Este método é possivelmente um dos mais flexíveis e conhecidos de interpolação de dados univariados existente, embora também exista sua versão bi-dimensional. Uma spline nada mais é que “uma curva definida matematicamente por dois ou mais pontos de controle”².

No caso da spline cúbica, esta é uma função polinomial de grau 3 definida em cada subintervalo demarcados pelos pontos de controle, no caso de interpolação são todos nós. Ou seja, considere um segmento entre dois pontos consecutivos $[c, d] \in S$ a spline é uma função cúbica com seus parâmetros calculados pelo algoritmo de ajuste. Para o próximo intervalo de pontos dentro do domínio da função, um novo polinômio de grau 3 é ajustado, sendo que nos pontos de nós uma restrição de igualdade entre as derivadas nos dois segmentos é aplicada para garantir a suavidade da função interpolada como um todo.

²Definição retirada de <https://pt.wikipedia.org/wiki/Spline>

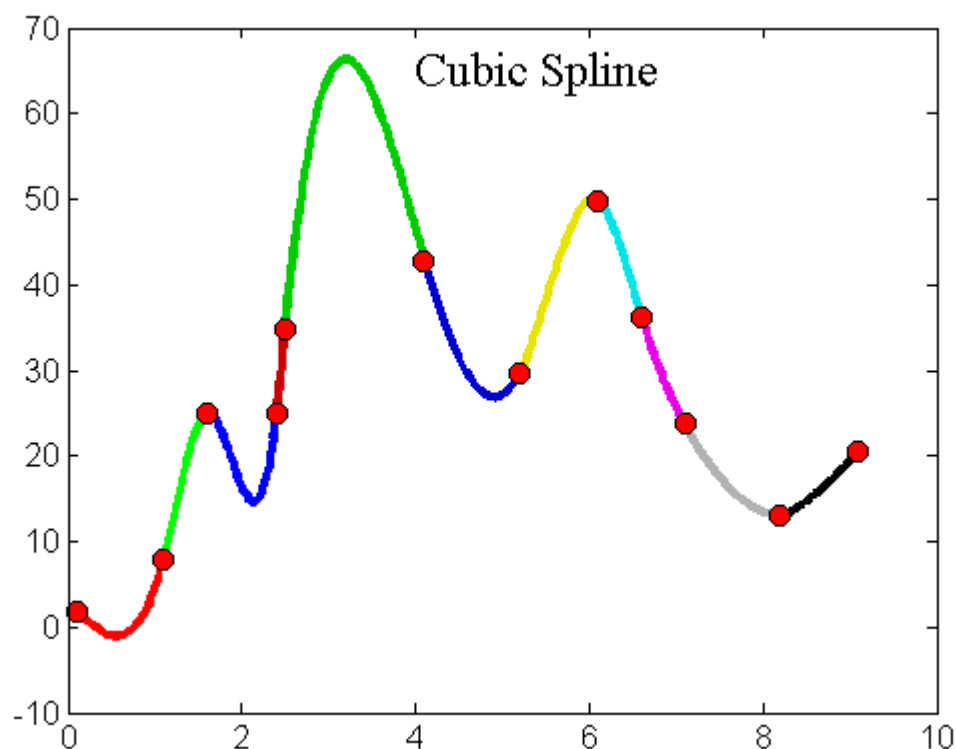


Figura 5.2: Cada segmento de uma spline cúbica é um polinômio de grau 3 diferente.

Assim, uma spline cúbica é uma função contínua, suave e diferenciável até a segunda ordem. Entretanto, suas derivadas, apesar de contínuas, podem não ser suaves, especialmente aquela de segunda ordem que pode apresentar pontos de “ruptura”. Esta característica de uma spline cúbica a torna pouco atrativa para a inferência de distribuições de probabilidade a partir de dados de volatilidade ou mesmo dos preços de opções.

5.3 Suavização

A técnica de suavização é muito semelhante a interpolação, inclusive o método spline também é aplicado, com algumas modificações de forma que nem todos os pontos fornecidos serão nós.

Na spline de suavização (ou aproximação), os pontos fornecidos são separados entre os nós, onde a função deve passar e pontos de controle, que são utilizados para controlar a curvatura da função nestes pontos.

Estas suavizações são principalmente utilizadas quando se possui muitas observações sujeitas a ruídos, de forma que uma interpolação entre todos os pontos seria tanto impraticável quanto sem sentido. O que se deseja, portanto, é uma função **aproximada** que melhor descreva o processo sob análise.

Um ponto em comum entre estas técnicas é o parâmetro de suavização, ausente, na interpolação, que controla a “suavidade” da função estimada.

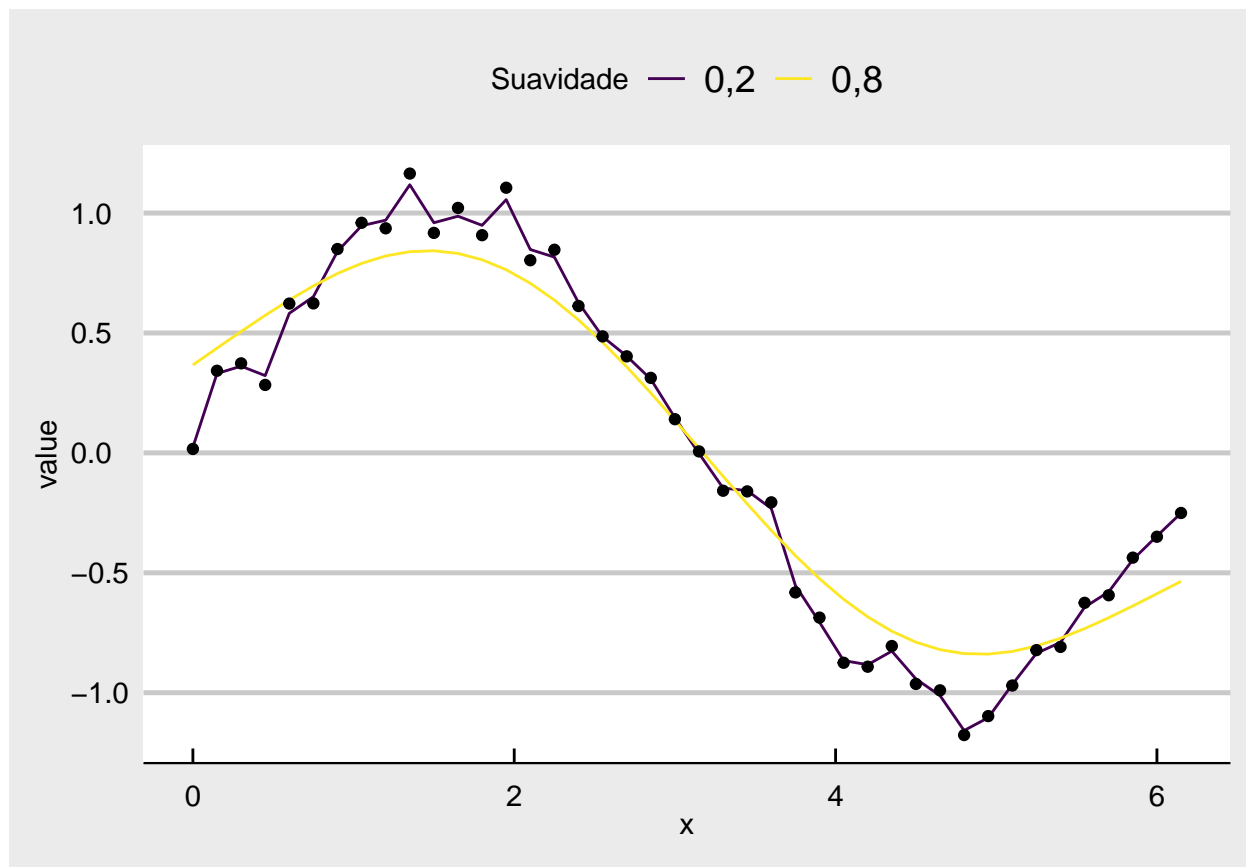


Figura 5.3: Menor parâmetro de suavização gera granularidade na curva.

5.4 Parametrização

E por fim as técnicas de parametrização. Nesta categoria estão diversos conhecidos modelos de superfícies de volatilidade implícita, dentre eles os modelos de Heston (1993), Volatilidade Local de Dupire (1994) e SVI de Gatheral (2004).

Em comum, estes modelos tentam parametrizar a superfície, e por conseguinte o smile de volatilidade, de acordo com alguma função, em geral não-linear, que possui características condizentes com a teoria de precificação de derivativos e também a observação empírica das superfícies.

Por exemplo, a parametrização *raw* da SVI possui a seguinte forma para a **variância total**³ :

$$w(k) = a + b \left(\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right)$$

que fornece um espaço de cinco parâmetros $\chi_B = \{a, b, \rho, m, \sigma\}$ que definem o smile e devem, portanto, serem calibrados a partir de dados observados no mercado.

O procedimento de calibração consiste em encontrar o conjunto de parâmetros que minimizam uma função perda entre a volatilidade prevista pelo modelo e os dados de mercado, enquanto satisfazem algumas restrições adicionais, como “ausência de arbitragem”, suavidade, etc. Trata-se, via de regra, de problemas de otimização não-linear com restrições de desigualdade também não-lineares.

5.4.1 Função perda

A função perda, ou função de calibração pode ser definida de diversas maneiras, de forma geral, para uma determinada maturidade, ela toma a forma:

$$L = \sum_{i=1}^n \lambda_i \|\hat{w}(k_i) - w_{imp}(k_i)\|$$

onde $\|\cdot\|$ é alguma medida de norma, sendo a mais conhecida o quadrado das diferenças, dando origem a minimização do erro quadrático médio (RMSE). Para este smile sendo calibrado existem n strikes (k_i) e suas volatilidades implícitas observadas são $w_{imp}(k_i)$. A resposta do modelo para um determinado strike é $\hat{w}(k_i)$ e λ_i são os pesos dados na função perda para cada um destes strikes.

Os pesos λ_i são utilizados para ponderar as observações das volatilidades mais importantes para o cálculo, onde se deseja que a curva ajustada possua um menor erro. Em geral, estes pesos são calculado como inversamente proporcionais:

- ao quadrado dos *spreads bid-ask*, para dar mais importância às opções mais líquidas
- ao quadrado da grega vega calculada a partir do modelo BSM

5.4.2 Otimizadores

Os otimizadores são os algoritmos pelos quais o problema de minimização posto é resolvido. Se a função perda é convexa, e ela deve ser construída de forma a ser, mesmo que não estritamente, então ela possui um ou mais pontos de mínimo onde o gradiente desta função é igual a zero. O que os otimizadores fazem é buscar o conjunto de parâmetros que minimizam a função perda e atendem as restrições impostas simultaneamente. Os otimizadores podem ser classificados em dois grandes grupos, globais e locais.

Algoritmos locais dependem de uma estimativa inicial dos parâmetros para começarem a busca pelo mínimo. Seguindo uma regra utilizando-se o gradiente da função ou alguma heurística, estes otimizadores

³A variância total é definida pelo produto entre a variância implícita e o tempo para expiração, $w = \sigma_{imp}^2 \cdot \tau$.

caminham em direção ao ponto de mínimo mais próximo da estimativa inicial, daí o nome “local”. Como desvantagem destes otimizadores é a mais evidente é que se a função perda for altamente não-linear, com diversos pontos de mínimo local, este otimizador pode ficar preso em um destes pontos sem nunca, no entanto, encontrar o mínimo global. Eles são, portanto muito sensíveis à estimativa inicial dos parâmetros.

Por sua vez, otimizadores globais buscam mapear todo o espaço factível para os parâmetros e encontrar o ponto mínimo da função perda dentro deste espaço. Estes algoritmos não dependem de estimativas iniciais, uma vez que tentarão avaliar o espaço completo. São utilizados quando o problema de minimização é não-linear e possui múltiplos pontos de mínimo local. Estes algoritmos usam alguma forma de heurística para encontrar a região onde o mínimo global está localizado, mas são, em geral, ineficientes em apontar rapidamente onde este ponto de mínimo se encontra com precisão. Por esta razão, é frequente a utilização de otimizadores globais com um posterior refinamento de sua solução por algum algoritmo local.

Abaixo apresentamos alguns exemplos mais comuns de otimizadores, tanto locais quanto globais:

- **Gauss-Newton:** Este método é utilizado para encontrar as raízes de alguma função. Para encontrar o ponto de mínimo da função perda, precisa-se encontrar as raízes do gradiente desta função, portanto o método de Newton em otimização faz uso da função gradiente. Este é um método de otimização local.
- **Levenberg-Marquardt:** Método muito utilizado para problemas não-lineares, ele parte de uma modificação ao método de Gauss-Newton ao introduzir um fator de amortecimento calculado iterativamente.
- **L-BFGS-B:** BFGS é um método conhecido como quasi-Newton, onde não é necessário calcular a Hessiana do problema, ela é aproximada a partir do próprio gradiente. É bastante utilizado para resolver problemas não-lineares e em sua versão L-BFGS-B pode lidar com restrições do tipo *box*, intervalo dos parâmetros é fixo.
- **Nelder-Mead:** Este é um método livre do uso de gradiente, já que usa uma heurística para construir um simplex e a partir deste “procurar” por um mínimo. Bastante utilizado quando a função objetivo pode não ser diferenciável. Faz uso de um simplex inicial, que pode ser grande o suficiente para encampar o mínimo global, entretanto, não se classifica como um otimizador global.
- **Algoritmo Genético:** Este método utiliza conceitos da seleção natural para gerar os resultados da otimização. É um otimizador global, no sentido que independe de uma estimativa inicial de parâmetros e faz uma busca por todo o espaço factível. Em um algoritmo genético, uma população aleatória inicial de parâmetros é criada e a partir desta, as gerações evoluem conforme mutações e *cross-over* de características e é avaliado o *fitness* de cada conjunto de parâmetros até um deles ser considerado adequado.
- **Evolução Diferencial:** É um método de otimização global, assim como o Algoritmo Genético e o Enxame de Partículas. Sua diferença reside no fato de que sua população inicial é constantemente avaliada e deslocada de posição. Se o agente obtiver uma situação melhor (menor valor para a função perda) na nova posição, esta agora faz parte da população. Desta forma os agentes, antes espalhados pelo espaço factível dos parâmetros, tendem a convergir para um ponto com o menor valor da função perda.
- **Enxame de Partículas:** Do inglês, *Particle Swarm Optimization - PSO* este método é semelhante ao DE (*Differential Evolution*) porém as partículas (o equivalente dos agentes no DE) matém informações sobre a posição da melhor partícula até então, de forma a fazer com que as partículas tendam para a melhor solução.

5.5 Conclusão

Dependendo do objetivo da aplicação, superfícies de volatilidade podem ser interpoladas, suavizadas ou parametrizadas. A parametrização tem recebido especial interesse pois pode, ao mesmo

tempo que garante uma superfície livre de arbitragem estática se devidamente construída, ajustar-se muito bem aos dados observados e gerar distribuições neutras ao risco implícitas factíveis.

Para gerar uma superfície parametrizada, primeiramente é necessário um modelo teórico com propriedades desejáveis e que se ajuste aos dados de mercado quando calibrado. Escolhido este modelo paramétrico, passa-se a calibração do mesmo onde existem diversas opções de escolha entre otimizadores. Ao final do processo teremos um modelo de superfície devidamente parametrizado com valores que melhor se ajustam segundo alguma função perda escolhida.

Com a superfície de volatilidade calibrada, as aplicações possíveis incluem a precificação de derivativos, gerenciamento de risco, simulações de Monte Carlo, análises de stress, entre outras.

Capítulo 6

Calibrando uma SVI

Neste capítulo iremos mostrar como fazer uma calibração de um smile SVI baseado nos trabalhos de (Gatheral, 2004) e (De Marco and Martini, 2009). Escolheremos apenas uma fatia da superfície de volatilidade, fixando o tempo para expiração (maturidade) e coletando as volatilidades implícitas para diversos strikes.

Como já apresentado em posts anteriores, existem diversas formas de interpolar, extrapolar, parametrizar e calibrar smiles de volatilidade. Existem vantagens e desvantagens para cada método. Neste post iremos fixar nossa atenção no modelo paramétrico de smile chamado SVI - Stochastic Volatility Inspired - uma forma que une a “simplicidade” de um modelo paramétrico com o poder de adesão aos dados de mercado dos modelos de volatilidade estocástica (i.e. Heston, SABR e afins).

Chamar o modelo SVI de simples é puro eufemismo, ele é um modelo poderoso, com fundamento teórico avançado e diversos detalhes para sua calibração.

6.1 Modelo SVI

Este modelo foi apresentado por Jim Gatheral na conferência *Global Derivatives & Risk Management 2004* e foi bem recebido pelos profissionais de mercado interessados em superfícies de volatilidade para *equities*, principalmente.

O modelo possui duas propriedades que são as razões para sua popularidade. Ele satisfaz a fórmula do momento de Lee (2004), que é um resultado independente de modelo que especifica os limites assintóticos para um *smile* de volatilidade implícita. Portanto, o modelo SVI é válido para extrapolação além da região central dos dados disponíveis. Além disso, afirma-se que o modelo SVI é relativamente fácil de calibrar para dados de mercado, de modo que a superfície de volatilidade implícita correspondente é livre de arbitragem de calendário. As condições que garantem a ausência de arbitragem de borboleta foram resolvidas em um segundo artigo por Gatheral and Jacquier (2014).

No SVI é possível se estabelecer condições explícitas em seus parâmetros, de modo que o modelo não gere preços onde oportunidades de arbitragem estáticas possam ocorrer. A calibração para dados reais de mercado requer algoritmos de otimização não-linear e pode ser bastante demorada. Mais recentemente, um método para calibração que usa a estrutura inerente do modelo para reduzir as dimensões do problema de otimização foi desenvolvido em De Marco and Martini (2009).

A parametrização conhecida como *RAW* do SVI é apresentada na equação (6.1), seguindo a notação já introduzida anteriormente, portanto, estamos modelando a **variância total** implícita para um determinado prazo. Para diferentes maturidades, teremos diferentes conjuntos de parâmetros.

$$w(k) = a + b \left(\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right) \quad (6.1)$$

onde: $a \in \mathbb{R}$, $b \geq 0$, $|\rho| < 1$, $m \in \mathbb{R}$, $\sigma > 0$, e $a + b\sigma\sqrt{1-\rho^2} \geq 0$ para garantir que $\min w(k) > 0$, $\forall k \in \mathbb{R}$.

6.2 Restrições de não-arbitragem

Antes de demonstrar a restrição imposta aos parâmetros b e ρ em função dos limites de inclinação das asas do *smile*, vamos derivar as expressões para $w'(k)$ e $w''(k)$ que nos serão úteis na demonstração.

A expressão para $w'(k)$ é bastante simples:

$$w'(k) = b \left[\rho + \frac{(k - m)}{\sqrt{(k - m)^2 + \sigma^2}} \right] \quad (6.2)$$

Derivando novamente a equação (6.2) em relação a k teremos uma expressão ainda mais simples, mesmo que após alguma manipulação algébrica um tanto tediosa¹, e resulta em:

$$w''(k) = \frac{b\sigma^2}{[(k - m)^2 + \sigma^2]^{3/2}} \quad (6.3)$$

onde, se considerarmos $b > 0$ temos que $w''(k) > 0$, $\forall k \in \mathbb{R}$, ou seja, o *smile* de volatilidade definido pela equação (6.1) é **estritamente** convexo.

Rogers and Tehranchi (2010) definiram os limites possíveis para a inclinação das asas em função do tempo para expiração, provando que o *smile* tende a ficar mais horizontal a medida que o prazo aumenta. Este limite pode ser escrito da seguinte forma e é uma **condição necessária** para a ausência de arbitragem:

$$|w'(k)| \leq \frac{4}{\tau} \quad \forall k \in \mathbb{R}, \quad \forall \tau \in (0, \infty) \quad (6.4)$$

Sendo o *smile* convexo, suas máximas inclinações ocorrem quando $k \rightarrow \pm\infty$. Portanto, deve-se avaliar a restrição dada pela equação (6.4) nestes limites da seguinte maneira:

$$\lim_{k \rightarrow \infty} w'(k) = b(1 + \rho) \geq 0 \quad (6.5)$$

$$\lim_{k \rightarrow -\infty} w'(k) = -b(1 - \rho) \leq 0 \quad (6.6)$$

que satisfazendo estas duas relações ao mesmo tempo em que se restringe os parâmetros b e ρ através da desigualdade de Rogers e Tehranchi nos garante o seguinte resultado para um SVI **livre de arbitragem de travas**.

$$b(1 + |\rho|) \leq \frac{4}{\tau} \quad (6.7)$$

Para garantir que a superfície gerada está livre de arbitragem do tipo borboleta deve-se primeiramente definir uma função² $g : \mathbb{R} \rightarrow \mathbb{R}$, tal que:

¹A resolução desta derivada é uma simples regra da divisão, entretanto a simplificação do resultado pede alguma manipulação algébrica. É possível utilizar sistemas de computação simbólica, o qual recomendamos o SymPy

²Condições para ausência de arbitragem do tipo borboleta em um SVI estão detalhadas na seção 2.2 do artigo de Gatheral and Jacquier (2014).

$$g(k) = \left(1 - \frac{kw'(k)}{2w(k)}\right)^2 - \frac{w'(k)^2}{4} \left(\frac{1}{w(k)} + \frac{1}{4}\right) + \frac{w''(k)}{2} \quad (6.8)$$

e seguir o lema :

Lema 1 Uma fatia da superfície de volatilidade está livre de arbitragem do tipo borboleta se, e somente se, $g(k) \geq 0$ para todo $k \in \mathbb{R}$ e $\lim_{k \rightarrow +\infty} d_1(k) = -\infty$.

Infelizmente, a natureza altamente não-linear da função $g(k)$ impossibilita a derivação de restrições gerais aos parâmetros do SVI. A forma mais simples de eliminar arbitragem do tipo borboleta é incluir a restrição $g(k) \geq 0$ na função perda e proceder com a calibração dos parâmetros.

6.3 Reparametrização Quasi-explicit

Um dos problemas mais marcantes com a calibração do SVI dado pela equação (6.1) é sua natureza altamente não-linear que gera inúmeros pontos de mínimo locais. Mesmo em um ambiente simulado, um típico otimizador de mínimos quadrados como Levenberg-Marquardt não consegue chegar ao mínimo global, onde a função perda é igual a zero. A solução encontrada é dependente dos valores iniciais inputados ao otimizador e a robustez do conjunto de parâmetros encontrados não é garantida.

Uma forma de contornar este problema pode ser a utilização de otimizadores globais, como algoritmos genéticos, em um primeiro estágio e então o refinamento desta solução através de um otimizador local (LM, por exemplo).

Outra forma, adotada em De Marco and Martini (2009) é a reparametrização da equação (6.1) de forma que esta possa ser tratada como um problema linear. Para tanto, considere a seguinte troca de variáveis:

$$y = \frac{k - m}{\sigma} \quad (6.9)$$

então a parametrização *RAW* do SVI se torna.

$$w(y) = a + b\sigma \left(\rho y + \sqrt{y^2 + 1}\right) \quad (6.10)$$

Definindo agora as seguintes variáveis reparametrizadas é possível reduzir a dimensão de parâmetros de um SVI de 5 para apenas 3:

$$c = b\sigma \quad (6.11)$$

$$d = \rho b\sigma \quad (6.12)$$

$$w(y) = a + dy + c\sqrt{y^2 + 1} \quad (6.13)$$

Portanto, para um par fixo de (m, σ) nosso problema reduzido é:

$$P_{m,\sigma} := \min_{a,c,d \in D} f_y(a, c, d) \quad (6.14)$$

onde $f_y(\cdot)$ é a função objetivo da reparametrização, e é dada pela seguinte equação:

$$f_y(a, c, d) = \sum_{i=1}^n [w(y_i) - \tilde{w}_i]^2 \quad (6.15)$$

onde \tilde{w}_i é a variância total observada correspondente ao *moneyness* k_i .

O domínio D dos parâmetros $\{a, c, d\}$ é encontrado a partir do limite imposto por (6.7).

$$D = \begin{cases} 0 \leq c \leq 4\sigma \\ |d| \leq c \quad \text{e} \quad |d| \leq 4\sigma - c \\ 0 \leq a \leq \max\{\tilde{w}_i\} \end{cases} \quad (6.16)$$

O problema reduzido dado pela equação (6.14), é um típico problema de **mínimos quadrados** com restrições lineares. Este problema, por ser convexo, admite uma única solução interior (se existente) que será o mínimo global para este problema e é encontrada através do gradiente da função objetivo igualando-o a zero, $\nabla f_y = 0$. Esta equação gera um sistema linear nos parâmetros a, c, d que pode ser explicitamente resolvido. Caso a solução encontrada para este problema esteja contida no domínio D , esta solução é interior e é o mínimo desejado, caso contrário, deve-se percorrer o perímetro do domínio e encontrar o menor valor da função objetivo que será uma solução de canto.

Seja (a^*, c^*, d^*) a solução de (6.14) e (a^*, b^*, ρ^*) os correspondentes parâmetros originais recuperados, então o problema completo de calibração é:

$$P := \min_{m, \sigma} \sum_{i=1}^n (w_*(k_i) - \tilde{w}_i)^2 \quad (6.17)$$

$$\text{onde } w_*(k) = a^* + b^* \left(\rho^*(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right).$$

O problema completo, (6.17) é um problema em apenas duas dimensões, (m, σ) e não-linear, que deve ser abordado através de algum tipo de otimizador global.

6.3.1 Solução explícita do problema reduzido

Nesta seção apresentaremos a solução, sem considerar as restrições impostas em (6.16) para o problema reduzido em (6.14), algo omitido em De Marco and Martini (2009). Esta seção é opcional para o leitor atento que já percebeu a semelhança entre o problema reduzido e um típico problema de regressão linear múltipla.

Para encontrar o conjunto de parâmetros (a^*, c^*, d^*) que representam os valores ótimos na equação (6.14), devemos resolver o seguinte sistema de equações:

$$\nabla f_y = \begin{bmatrix} \partial f_y / \partial a \\ \partial f_y / \partial d \\ \partial f_y / \partial c \end{bmatrix} = \mathbf{0} \quad (6.18)$$

Cada uma das derivadas parciais da equação acima quando igualadas a zero dão origem ao sistema linear apresentado abaixo:

$$\begin{bmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n y_i^2 & \sum_{i=1}^n \frac{\sqrt{y_i^2 + 1}}{y_i \sqrt{y_i^2 + 1}} \end{bmatrix} \cdot \begin{bmatrix} a \\ d \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \tilde{w}_i \\ \sum_{i=1}^n \tilde{w}_i y_i \\ \sum_{i=1}^n \tilde{w}_i \sqrt{y_i^2 + 1} \end{bmatrix} \quad (6.19)$$

Portanto, o problema reduzido pode ser resolvido através de um sistema linear de ordem 3 sob restrições também lineares.

6.4 Algoritmo

A otimização para encontrar os parâmetros ótimos de um SVI dadas observações de mercado e a técnica de calibração **Quasi-explicit** de De Marco and Martini (2009) pode ser resumida nos seguintes passos:

1. Definir valores iniciais para os parâmetros (m, σ) ,
2. Iniciar algum otimizador global com estes parâmetros e resolver o problema completo (6.17)
 - 2.1 Dentro da otimização global, resolver o problema reduzido (6.14) para os parâmetros (m, σ) dados,
3. Na convergência do problema completo do passo 2, otimizar uma última vez o problema reduzido,
4. Recuperar os parâmetros (a, b, ρ, m, σ)

A escolha dos otimizadores fica a cargo pessoal, sendo sugerido testar vários para o mesmo problema. Eventualmente, para um determinado *smile* um otimizador pode se mostrar melhor que outro que vinha sendo utilizado em outras ocasiões.

Nos testes realizados pelo Clube de Finanças, entre os otimizadores globais para o problema completo utilizamos **Algoritmos Genéticos**, **Nelder-Mead restrito** e um **Método não-linear generalizado**. Para o problema reduzido, apesar de ser linear, o método de Nelder-Mead restrito se mostrou tão eficiente quanto e de mais fácil implementação. Se o objetivo for fazer uma calibração direta, dos cinco parâmetros ao mesmo tempo, uma combinação de otimizador global em primeiro estágio e o método de Levenberg-Marquardt restrito para refinamento da solução é o ideal.

6.5 Resultados

A seguir apresentamos um *smile* de referência para a calibração, obtido de ivolatility.com e então partimos para diferentes técnicas de calibração de um SVI. Os códigos em R também estão apresentados ao longo do texto para melhor compreensão e estudo do leitor.

Os dados utilizados neste exemplo estão apresentados na tabela 6.1 abaixo. Esta é uma típica apresentação de um *slice* de superfície, ou seja, dados para um *smile* apenas. As principais variáveis são: a data em que os dados foram coletados (*date*), o preço de fechamento do ativo (*stock_price*), o prazo para expiração em dias (*period*), e medidas de moneyness como delta, *strike* e o próprio *moneyness*, além é claro da volatilidade implícita (*iv*) retirada do mercado.

Esta tabela poderia conter (de fato contém no arquivo original) outros períodos de expiração, e neste caso uma das colunas de *moneyness* começa a se repetir, no caso seria o delta pois baixamos uma tabela de volatilidades implícitas por delta. Assim, em uma tabela simples em formato *tidy* é possível armazenar informações de uma superfície inteira, a qual de outra forma necessitaria de um arranjo em 3 dimensões.

Ressaltamos aqui que a unidade de volatilidade implícita está em percentuais **ao ano**, enquanto que nosso período é de dias corridos. É necessário harmonizar estas medidas de forma que, para volatilidades dadas em percentual ao ano, o período também seja dado em **anos**. Logo nosso $\tau = 30/365$, ou seja, 0.08219.

Demonstraremos aqui os resultados para a calibração de uma *RAW SVI* pelos métodos “Direto”, “GA”, “Quasi-NM” e “Quasi-PQ”, abaixo explicados.

O método “Direto” é uma calibração direta através de um algoritmo de **Levenberg-Marquardt** da equação (6.1), ou seja, não existe reparametrização *Quasi-explicit* e o problema resolvido é não-linear em 5 dimensões. São realizadas 10 calibrações com estimativas iniciais dos parâmetros aleatórias, mas dentro

Tabela 6.1: Dados reais para exemplo de calibração de uma SVI.

date	symbol	exchange	stock_price_for_iv	period	delta	moneyness	strike
2017-09-21	IWM	NYSEArca	143.73	30	10	0.03	148.4
2017-09-21	IWM	NYSEArca	143.73	30	15	0.03	147.4
2017-09-21	IWM	NYSEArca	143.73	30	20	0.02	146.8
2017-09-21	IWM	NYSEArca	143.73	30	25	0.02	146.2
2017-09-21	IWM	NYSEArca	143.73	30	30	0.01	145.6
2017-09-21	IWM	NYSEArca	143.73	30	35	0.01	145.1
2017-09-21	IWM	NYSEArca	143.73	30	40	0.01	144.6
2017-09-21	IWM	NYSEArca	143.73	30	45	0.00	144.1
2017-09-21	IWM	NYSEArca	143.73	30	50	0.00	143.6
2017-09-21	IWM	NYSEArca	143.73	30	55	0.00	143.1
2017-09-21	IWM	NYSEArca	143.73	30	60	-0.01	142.5
2017-09-21	IWM	NYSEArca	143.73	30	65	-0.01	141.8
2017-09-21	IWM	NYSEArca	143.73	30	70	-0.02	141.1
2017-09-21	IWM	NYSEArca	143.73	30	75	-0.02	140.2
2017-09-21	IWM	NYSEArca	143.73	30	80	-0.03	139.1
2017-09-21	IWM	NYSEArca	143.73	30	85	-0.04	137.6
2017-09-21	IWM	NYSEArca	143.73	30	90	-0.06	135.3

de seus respectivos domínios. A melhor solução, aquela com o menor valor para a função objetivo, é selecionada. Todos os outros métodos utilizam a reparametrização, ocorrendo variações apenas nos algoritmos de otimização utilizados nos problemas reduzido e completo.

A calibração “GA” faz uso do otimizador global de **algoritmos genéticos** para o problema completo, ou seja, para estimar o par (m, σ) que corresponde ao mínimo global. Após, o problema reduzido é resolvido através do algoritmo de **Nelder-Mead**. Este método é robusto, pois o algoritmo genético tem grande probabilidade de encontrar a região onde se encontra o mínimo global e não ficar preso localmente. Entretanto a robustez ocorre às expensas do tempo de computação.

Os métodos ditos “Quasi” diferem entre si na resolução do problema reduzido. Enquanto “PQ” remete a **programação quadrática** e faz uso da resolução do sistema linear apresentado na equação (6.19) com as restrições impostas por (6.16), o método “Quasi-NM” utiliza o método de **Nelder-Mead** com restrições para a resolução deste mesmo problema reduzido. Em ambos os métodos, o problema completo é resolvido com um algoritmo de Nelder-Mead com 50 reinicializações das estimativas iniciais dos parâmetros (m, σ) , o que causa algum impacto no tempo de computação destes métodos.

```
smile <- dados %>%
  mutate(tau = period / 365) %>%
  select(moneyness, iv, tau)

par_names <- factor(c("a", "b", "\\rho", "m", "\\sigma"),
  levels = c("a", "b", "\\rho", "m", "\\sigma"))
k <- smile$moneyness
w <- smile$iv^2 * smile$tau
```

```

init_direct <- proc.time()
par_direct <- svi_fit_direct(k, w)
end_direct <- proc.time()
time_direct <- end_direct - init_direct

init_ga <- proc.time()
par_ga <- svi_fit_ga(k, w)
end_ga <- proc.time()
time_ga <- end_ga - init_ga

init_quasipq <- proc.time()
par_quasipq <- svi_fit_quasi(k, w, inner = "quadprog")
end_quasipq <- proc.time()
time_quasipq <- end_quasipq - init_quasipq

init_quasinm <- proc.time()
par_quasinm <- svi_fit_quasi(k, w)
end_quasinm <- proc.time()
time_quasinm <- end_quasinm - init_quasinm

iv_direct <- sqrt(svi_fun(par_direct$par[[1]], k) / smile$tau)
iv_ga <- sqrt(svi_fun(par_ga$par[[1]], k) / smile$tau)
iv_quasipq <- sqrt(svi_fun(par_quasipq$par[[1]], k) / smile$tau)
iv_quasinm <- sqrt(svi_fun(par_quasinm$par[[1]], k) / smile$tau)

plot_tbl <- tibble(k = k,
  Direct = iv_direct,
  GA = iv_ga,
  QuasiPQ = iv_quasipq,
  QuasiNM = iv_quasinm,
  observed = smile$iv) %>%
  gather(key = method, value = iv, -c(k, observed))

par_tbl <- bind_rows(par_direct, par_ga, par_quasipq, par_quasinm) %>%
  select(method, par) %>%
  unnest() %>%
  mutate(names = rep(par_names, 4)) %>%
  spread(method, par) %>%
  select(names, Direct, GA, QuasiPQ, QuasiNM) %>%
  mutate(names = as.character(names)) %>%
  mutate_at(vars(Direct:QuasiNM), arred)

rmse_tbl <- bind_rows(par_direct, par_ga, par_quasipq, par_quasinm) %>%
  select(method, par) %>%
  unnest() %>%
  group_by(method) %>%
  summarise(RMSE = rmse(par, k, w)) %>%
  spread(method, RMSE) %>%
  mutate(names = "RMSE") %>%
  select(names, Direct, GA, QuasiPQ, QuasiNM) %>%
  mutate_at(vars(Direct:QuasiNM), format, digits = 3, scientific = TRUE)

time_tbl <- tibble(method = c("Direct", "GA", "QuasiPQ", "QuasiNM"),

```

Tabela 6.2: Parâmetros estimados da calibração, RMSE e tempo de computação em segundos.

Estimativa	Direto	GA	QuasiPQ	QuasiNM
a	0.00000	0.00001	0.00000	0.00000
b	0.01950	0.01954	0.01749	0.08475
ρ	-0.79529	-0.80676	-1.00000	-0.99999
m	-0.00725	-0.00807	-0.01420	-0.07875
σ	0.05026	0.05048	0.06985	0.04228
RMSE	8.67e-06	8.7e-06	9.66e-05	5.01e-05
Tempo	0.17200	29.88700	0.14800	5.19600

```

time = rbind(time_direct, time_ga,
              time_quasipq, time_quasinm)[, 3]) %>%
spread(method, time) %>%
mutate(names = "Tempo") %>%
select(names, Direct, GA, QuasiPQ, QuasiNM) %>%
mutate_at(vars(Direct:QuasiNM), arred)

frame_tbl <- bind_rows(par_tbl, rmse_tbl, time_tbl)

```

Abaixo é apresentada uma tabela com os valores estimados para os parâmetros da SVI, o RMSE (root mean square error) e o tempo total em segundos para a calibração. Aqui o RMSE é definido como $\sqrt{1/n \sum (w(k_i) - \tilde{w}_i)^2}$ e nos fornece um valor típico de erro **na variância**.

```

kable(frame_tbl,
      col.names = c("Estimativa", "Direto", "GA",
                    "QuasiPQ", "QuasiNM"),
      caption = "Parâmetros estimados da calibração, RMSE e tempo de computação em segundos.",
      booktabs = TRUE) %>%
kable_styling(bootstrap_options = "striped",
              font_size = 18,
              full_width = FALSE)

```

O método Direto, com algoritmo de Levenberg-Marquardt se mostrou muito mais rápido que os demais, principalmente com relação ao algoritmo genético, e com um bom ajuste dado o baixo valor de RMSE. O algoritmo genético é consideravelmente mais lento, entretanto durante as várias calibrações realizadas em testes (e que não estão apresentadas na tabela 6.2), este algoritmo sempre se mostrou robusto, com baixo RMSE, diferentemente dos outros métodos que por vezes, dependendo das estimativas iniciais, podem convergir para um mínimo local.

O gráfico com os ajustes realizados pode ser observado abaixo.

```

ggplot(plot_tbl, aes(x = k)) +
  geom_point(aes(y = observed)) +
  geom_line(aes(y = iv, color = method)) +
  guides(color = guide_legend(title = "")) +
  labs(title = "",
       x = "Forward log-moneyness",

```

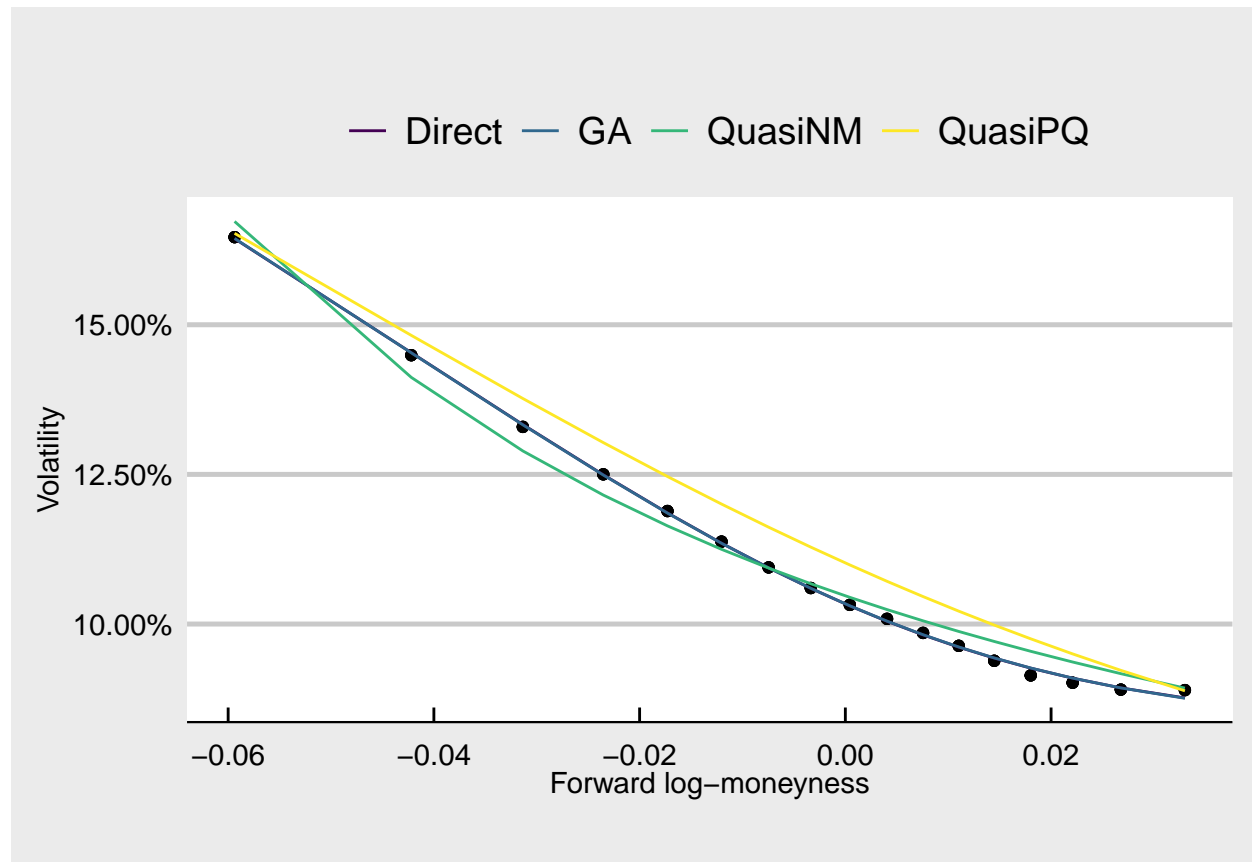


Figura 6.1: Comparação entre diferentes métodos de calibração de uma SVI.

```
y = "Volatility",
caption = "") +
scale_y_continuous(labels = scales::percent) +
scale_color_viridis_d() +
theme_economist_white()
```

De fato o método direto e o método *Quasi-explicit* com otimizador global do tipo algoritmo genético se mostram mais adequados para a calibração de um SVI. Enquanto o método direto é muito mais eficiente em termos computacionais, o método *Quasi-explicit* com GA é mais robusto. Desta forma, deve-se salientar que é necessário que o usuário, ao fazer uma calibração de *smile* de volatilidade, deve dispor de diferentes métodos de fazê-lo, e a inspeção visual do resultado é **obrigatória** para determinar qual método foi mais eficiente em ajustar a curva aos dados.

6.6 Conclusão

Apesar de neste exemplo ter se mostrado um método efetivo, com bom ajuste e baixo tempo de calibração, o método direto é altamente dependente dos valores iniciais dos parâmetros. Para tornar este método mais robusto, um número maior de reinicializações deve ser feita o que penaliza o tempo de calibração. O método *Quasi-explicit* com algoritmo genético para encontrar a região de (m, σ) onde se encontra o mínimo global se mostrou bastante robusta, entretanto, de convergência lenta. Para ajustar

apenas um *smile* alguns segundos a mais não representam problema. Porém, se imaginarmos que em uma grande instituição financeira são necessárias calibrações de, talvez, milhares de *smiles* representando inúmeras superfícies de diversos instrumentos, este método pode se mostrar computacionalmente caro.

Já os métodos *Quasi-explicit* que utilizam um algoritmo de Nelder-Mead para a resolução do problema completo se mostraram muito sensíveis às estimativas iniciais dos parâmetros. Mesmo utilizando 50 reinicializações do método, diversas vezes o ajuste realizado foi insatisfatório. A resolução através de programação quadrática é rápida, se comparada com NM, entretanto, quando as restrições impostas pela equação (6.16) se tornam ativas, este método parece sofrer com algum viés em sua solução.

Bibliografia

- Aurell, A. (2014). The svi implied volatility model and its calibration. Master's thesis, Kungliga Tekniska Högskolan.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.
- Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.
- De Marco, S. and Martini, C. (2009). Quasi-explicit calibration of gatheral's svi model'. *Zeliade White Paper*, pages 1–15.
- Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1):18–20.
- Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives & Risk Management, Madrid*.
- Gatheral, J. (2011). *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons.
- Gatheral, J. and Jacquier, A. (2014). Arbitrage-free svi volatility surfaces. *Quantitative Finance*, 14(1):59–71.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343.
- Lee, R. W. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 14(3):469–480.
- Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125 – 144.
- Rogers, L. C. and Tehranchi, M. (2010). Can the implied volatility surface move by parallel shifts? *Finance and Stochastics*, 14(2):235–248.