

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS**

Humberto Ribeiro de Souza

**Reproduzindo equações diferenciais estocásticas com
redes neurais**

São Carlos

2022

Humberto Ribeiro de Souza

**Reproduzindo equações diferenciais estocásticas com
redes neurais**

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Básica

Orientador: Prof. Dr. Leonardo Paulo Maia

Versão corrigida
(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2022

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Souza, Humberto Ribeiro

Reproduzindo equações diferenciais estocásticas com redes neurais / Humberto Ribeiro Souza; orientador Leonardo Paulo Maia - versão corrigida -- São Carlos, 2022.

71 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Básica) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2022.

1. Econofísica. 2. Volatilidade estocástica. 3. Redes neurais. 4. Método do estado adjunto. 5. Equações diferenciais estocásticas. I. Maia, Leonardo Paulo, orient. II. Título.

Este trabalho é dedicado aos meus pais, aos meus irmãos, a Mariella e ao meu notebook que parou de funcionar 1 semana antes da entrega deste trabalho.

AGRADECIMENTOS

Primeiro de tudo, eu preciso agradecer meus pais por todo o suporte e confiança que vocês me deram por toda minha vida. Eu não estaria aonde estou se vocês não acreditassem em mim.

A todas as pessoas da comissão da SIFSC: sou muito grato por ter feito parte deste time e de todos os eventos que construímos. Os conhecimentos que tive aqui serão importantes para a minha vida.

Ao Diogo, Ricardo e a Cheng: sou grato por ter sido aluno de vocês durante minha pós-graduação, este projeto teve muita colaboração de vocês e do aprendizado que tive em suas disciplinas.

Ao pessoal da Rede Análise: eu também sou muito grato de ter me juntado a vocês durante este tempo. Foram um grande apoio para mim em muitos momentos e por causa de vocês, eu sei como agir daqui pra frente para fazer um grupo se manter.

Para a Neusa e a Cristina da biblioteca: eu não conseguirei agradecer vocês o suficiente. Me abrigaram nos últimos dias de entrega e aceitaram receber inúmeras correções que eu mandei de última hora. Obrigado mesmo por todo esse apoio.

Ao Leo: obrigado por aceitar esse doido perdido. Este tempo de mestrado foi essencial para eu traçar muitos passos relevantes para a minha vida e descobrir quem eu sou, apenas me arrependo de não ter compartilhado tantos com você. Pode ter certeza que vou sair daqui muito menos perdido do que jamais tive (não garanto nada sobre a doideira).

Ao Scar e ao Mufasa: obrigado por estarem sempre ao meu lado. Vocês me ensinam muito sobre amor e cuidado. O carinho de vocês foi e é muito mais importante pra mim do que vocês jamais conseguiriam imaginar.

A Mariella: obrigado por ser minha parceira. Saber que eu tenho o seu apoio sempre me ajuda muito a continuar. Você sempre me instiga a querer ser alguém melhor e mais gentil. Sou muito sortudo por poder dividir a minha vida com você. Não terei palavras o suficiente para dizer o quão especial você é pra mim no tempo restante que tenho para entregar esta dissertação.

E por último, eu preciso agradecer a possibilidade de realizar todas essas conquistas graças a educação pública. Neste tempo de mestrado, um dos aprendizados mais relevantes que tive foi em entender a importância que isso teve em minha vida e que eu não teria as mesmas oportunidades na minha vida sem isso.

“Todos os modelos estão errados, mas alguns são úteis.”
George Edward Pelham Box

RESUMO

SOUZA, H. R. **Reproduzindo equações diferenciais estocásticas com redes neurais**. 2022. 71p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2022.

Aproveitando o extenso ferramentário técnico de físicos, a econofísica é um campo de estudo que deseja entender fenômenos econômicos utilizando este conhecimento. Desde abordagens mais simples, como o modelo padrão de finanças, que considera o retorno de um ativo uma variável aleatória, até em propostas mais sofisticadas, como a volatilidade estocástica, que considera que a volatilidade de um ativo financeiro também será uma variável aleatória, objeto de estudo deste trabalho por meio do modelo de Heston.¹ Nesta dissertação, estudo se é possível reproduzir a equação diferencial estocástica do modelo de Heston, por meio de amostras dos processos estocásticos de retorno e volatilidade, utilizando redes neurais para aproximar as funções $a(X, t)$ e $b(X, t)$, termo determinístico e de ruído da equação. O processo de otimização das redes neurais ocorre por meio do método do estado adjunto, que auxilia no cálculo do gradiente. Os processos estocásticos resultantes deste gerador encontram-se nas últimas figuras deste trabalho, sendo possível observar a dinâmica do retorno financeiro de um ativo está, visualmente bem representada, exceto pelo caráter oscilatório observado no modelo original de Heston. Para a volatilidade, observamos que este gerador não é capaz de reproduzir sua dinâmica de maneira satisfatória, sendo um indicador de que o modelo de volatilidade estocástica não será o mais adequado para esta nova abordagem. Para realizar uma análise mais robusta, emprego uma aproximação para a estimação dos parâmetros relevantes ao modelo de Heston para a equação geradora e comparo com a amostra original dada para a realização deste processo, onde é possível observar que há grande diferença entre as distribuições resultantes para os parâmetros relacionados a volatilidade, sendo mais um indicador que corrobora a necessidade de modelos mais sofisticados que possam ser representados fielmente pela equação geradora estudada neste trabalho.

Palavras-chave: Econofísica. Volatilidade estocástica. Redes Neurais. Método do Estado Adjunto. Equações diferenciais estocásticas.

ABSTRACT

SOUZA, H. R. **Reproducing stochastic differential equations with neural networks**. 2022. 71p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2022.

Taking advantage of the extensive technical tools of physicists, econophysics is a field of study that aims to understand economic phenomena using this knowledge. From simpler approaches, such as the standard finance model, which considers the return on an asset a random variable, to more sophisticated proposals, such as stochastic volatility, which considers that the volatility of a financial asset will also be a random variable, the object of study of this work through Heston's model.¹ In this dissertation, I study if it is possible to reproduce the stochastic differential equation of Heston's model, through samples simulated by me, using neural networks to approximate the functions $a(X, t)$ and $b(X, t)$, the deterministic and noise term of the equation. The process of optimizing the neural networks occurs through the adjoint state method, which aids in the computation of the gradient. The stochastic processes resulting from this generator are shown in the last figures of this paper, and it is possible to observe that the dynamics of the financial return is visually well represented, except for the oscillatory character observed in the original Heston's model. For the volatility, we observe that this generator is not able to reproduce its dynamics in a satisfactory manner, being an indicator that the stochastic volatility model will not be the most adequate for this new approach. To perform a more robust analysis, I use an approximation for estimating the relevant parameters of Heston's model for the generating equation and compare it with the original sample given for this process, where it is possible to observe that there is a big difference between the resulting distributions for the parameters related to volatility, being one more indicator that corroborates the need for more sophisticated models that can be faithfully represented by the generating equation studied in this work.

Keywords: Econophysics. Stochastic Volatility. Neural Networks. Adjoint State Method. Stochastic differential equations.

LISTA DE ILUSTRAÇÕES

Figura 1 – Volatilidade do índice Ibovespa pelos pregões de 2012 a 2021	41
Figura 2 – Fonte: Elaborado pelo autor	41
Figura 3 – Processo iterativo de otimização de uma rede neural	43
Figura 4 – Processo iterativo de otimização utilizando o método do estado adjunto	44
Figura 5 – Processo iterativo de otimização utilizando o método do estado adjunto estocástico	45
Figura 6 – Representação gráfica de um perceptron multicamadas.	47
Figura 7 – Caminho estocástico para o log-retorno utilizando: (7a) o modelo de Heston e suas reproduções utilizando o método da SDE Neural com seu (7b) chute inicial e após (7c) 1000 e (7d) 5000 iterações.	49
Figura 8 – Caminho estocástico para a volatilidade utilizando: (8a) o modelo de Heston e suas reproduções utilizando o modelo da SDE Neural com seu (8b) chute inicial e após (8c) 1000 e (8d) 5000 iterações.	50
Figura 9 – Parâmetros μ , α e θ estimados pela verossimilhança do modelo de Heston comparado ao modelo da SDE Neural gerado após: o chute inicial (9a), 1000 iterações (9b) e 5000 iterações(9c).	54

LISTA DE ABREVIATURAS E SIGLAS

EDO	Equação Diferencial Ordinária
SDE	<i>Stochastic Differential Equation</i> (Equação Diferencial Estocástica)

SUMÁRIO

1	INTRODUÇÃO	19
2	FUNDAMENTOS DA DINÂMICA DE PREÇOS E DO CÁLCULO ESTOCÁSTICO	21
2.1	Ativos e seus preços variáveis	21
2.2	Processo de Wiener e o Movimento Browniano	24
2.3	Cálculo de Itô	27
3	DERIVATIVOS FINANCEIROS E A EQUAÇÃO DE BLACK-SCHOLES	33
3.1	Mercado eficiente	33
3.2	Opções	34
3.3	Demonstração de Black-Scholes	36
4	VOLATILIDADE ESTOCÁSTICA	39
4.1	Limitações de Black-Scholes	39
4.2	Volatilidade estocástica	40
4.3	Modelo de Heston	41
5	MÉTODO DO ESTADO ADJUNTO	43
5.1	Método do Estado Adjunto para Equações Diferenciais Ordinárias .	45
5.2	Método do Estado Adjunto para Equações Diferenciais Estocásticas	46
5.3	Equações Diferenciais Estocásticas compostas por redes neurais . .	46
5.4	Descrição do código	48
5.5	Resultados	49
6	ESTIMAÇÃO DE PARÂMETROS	53
6.1	Estimação da máxima verossimilhança	53
6.2	Resultados	54
7	CONCLUSÃO	57
	REFERÊNCIAS	59
	APÊNDICES	61
	APÊNDICE A – DISTRIBUIÇÃO NORMAL (GAUSSIANA) E O TEOREMA CENTRAL DO LIMITE	63

APÊNDICE B – DISTRIBUIÇÃO LOGNORMAL	65
APÊNDICE C – MÉDIA DE χ^2	67
APÊNDICE D – FÓRMULA DE ITÔ	69
APÊNDICE E – GRADIENTES DO MÉTODO ESTOCÁSTICO ADJUNTO	71

1 INTRODUÇÃO

A interdisciplinaridade entre grandes áreas de pesquisa se mostra cada vez imprescindível para o progresso da ciência e se dá através do intercâmbio constante de conhecimentos, metodologias e ferramentas.

Dentre as áreas que surgiram organicamente desta combinação temos a econofísica, que envolve a interação entre físicos e as ciências econômicas ao se aproveitar de suas experiências prévias com modelagens matemáticas complexas para o desenvolvimento de novos estudos.

Um dos campos disseminados entre pesquisadores da área é a precificação de opções, ferramenta utilizada pelo mercado financeiro visando garantir a possibilidade de compra e venda de determinados ativos financeiros em uma data futura.

Para trazer as fundamentações básicas de precificação de opções, foi feito um capítulo introdutório sobre a dinâmica de preços de um ativo e como temos presentes elementos estocásticos para este movimento, também é necessário uma introdução ao cálculo estocástico. Ele segue com uma fundamentação no funcionamento de derivativos, como é o caso das opções, e as assunções postas para obter a fórmula de Black-Scholes² para precificar este derivativo no terceiro capítulo.

Uma tentativa de aperfeiçoamento propostas a este modelo de precificação é a volatilidade estocástica, tema deste projeto, que propõe a volatilidade de uma ação como variável aleatória. Tal ramificação proporcionou o desenvolvimento de diversos modelos de sistemas de equações diferenciais estocásticas que relacionam o preço de uma ação e sua volatilidade.

Assim sendo, esta pesquisa tem como foco o modelo de Heston,¹ que utiliza uma distribuição Gamma generalizada para a volatilidade, modificando assim a equação de Black-Scholes para um preço de uma ação. Os conceitos mais importantes deste modelo e sua pertinência para o desenvolvimento deste trabalho serão expostos no quarto capítulo desta dissertação.

A intenção ao longo desta pesquisa de mestrado é reproduzir o comportamento estocástico do referido modelo utilizando apenas uma amostra de dados obtida por simulações do mesmo.

Para tanto empregado um método proposto recentemente³ que se utiliza de redes neurais para aproximar as funções $\mu(X_t, t)$ e $\sigma(X_t, t)$ termos determinístico e de ruído, respectivamente, de equações diferenciais estocásticas com otimização de seus parâmetros feita pelo Método do Estado Adjunto,⁴ o qual será apresentado brevemente no quinto

capítulo deste trabalho, em conjunto com seus resultados e analisando se é possível obter os parâmetros atribuídos ao modelo de Heston pela estimativa da máxima verossimilhança das amostras reais e geradas no sexto capítulo.

2 FUNDAMENTOS DA DINÂMICA DE PREÇOS E DO CÁLCULO ESTOCÁSTICO

Este capítulo é baseado em notas de aula ainda não publicadas ^{*}, que, por sua vez, foram inspiradas nos livros de Wilmott⁵ e Luenberger.⁶

2.1 Ativos e seus preços variáveis

Um **agente** é qualquer unidade funcional (não necessariamente um indivíduo, pode ser uma organização/instituição) que interage com outros agentes em um sistema. Quando essa interação se manifesta mediante trocas de vários tipos de bens, é natural que exista alguma moeda, cuja posse indica o poder de troca de cada agente. Nessas condições, tal comunidade é denominada um **mercado financeiro**, onde agentes negociam **ativos** (*assets*).

Porém, nesta dissertação, não haverá nenhuma descrição explícita dos fluxos de bens/ativos e capital entre os agentes. Embora distribuições heterogêneas de riqueza entre os agentes financeiros certamente influenciem suas atuações em qualquer mercado real e, em última instância, afetam todas as características adaptativas do sistema, o foco deste estudo é um problema que costuma ser estudado de forma bem satisfatória mediante modelos quantitativos que dependem apenas de alguns poucos parâmetros fenomenológicos, e não dos detalhes das trocas de capital e sua posse. Assim, no presente contexto, o principal efeito da existência de uma moeda é que ela define uma escala de utilidade universal pela qual é expresso o **preço** (valor) de cada um dos ativos negociados em um mercado. Empiricamente, é notório que preços são temporalmente dinâmicos, mas quasi-estáticos, no sentido em que evoluem em uma escala de tempo bem maior do que aquela das transações entre os agentes. O **problema da precificação** consiste na determinação desses preços de quase-equilíbrio e sua lenta dinâmica [†]. Nesta dissertação, serão estudados modelos de precificação baseados em hipóteses simplificadoras sobre o comportamento do mercado, dependentes de algumas poucas variáveis descritivas de um ambiente microeconômico *que devem ser inferidas*.

Esses modelos matemáticos serão formas elaboradas de equações diferenciais estocásticas, em tempo contínuo, mas é conveniente introduzir terminologia e conceitos básicos em uma representação da evolução temporal do preço de um ativo em tempo discreto. Em um único passo temporal, se o ativo tiver sido adquirido pelo preço inicial S_i

^{*} Notas de aula de Mecânica Estatística, de Leonardo P. Maia.

[†] É irresistível a tentação de chamar esse cenário de física estatística de quasi-equilíbrio, no sentido em que, dados horizontes temporais mesoscópicos (escalas temporais bem maiores do que a escala de ação dos agentes, mas ainda pequenas o suficiente para os preços serem quasi-estáticos), cada estado de equilíbrio seria descrito por poucos parâmetros fenomenológicos.

e posteriormente vendido pelo preço final S_f , seu **retorno total** é definido como

$$R \equiv \frac{S_f}{S_i}. \quad (2.1)$$

A taxa de retorno, ou **retorno relativo**,

$$r \equiv \frac{S_f - S_i}{S_i} = R - 1 \quad (2.2)$$

muitas vezes é denominada simplesmente **retorno** no cotidiano. Ao longo da ininterrupta evolução temporal de um preço S_t , tanto R quanto r são variáveis, e, na verdade, as definições mais adequadas para descrever a sequência temporal de evolução dos preços do ativo são aquelas dos retornos instantâneos,

$$R_t \equiv \frac{S_{t+1}}{S_t}, \quad (2.3)$$

e

$$r_t \equiv \frac{S_{t+1} - S_t}{S_t} = R_t - 1 \quad (2.4)$$

total e relativo, respectivamente.

Neste momento, há uma importante questão de modelagem matemática a ser resolvida, que envolve tanto considerações empíricas quanto teóricas. Certamente, há elementos estocásticos na dinâmica realística de preços e uma proposta aceitável *a priori* consistiria em descrever cada **incremento** de preço $\Delta S_t \equiv S_{t+1} - S_t$ como uma variável aleatória Gaussiana independente dos demais incrementos. No entanto, essa é quase sempre uma má escolha [‡]. Isso levaria a um modelo *aditivo* de efeitos estocásticos (em cada passo temporal, o novo preço é dado pelo anterior adicionado de um fator estocástico) e a uma distribuição também Gaussiana para S_t (soma de Gaussianas independentes é novamente uma Gaussiana), que apresentaria pelo menos duas características incompatíveis com a descrição de mercados reais, (a) a possibilidade de preços negativos e (b) uma baixa probabilidade de valores extremos de preços (distantes do valor médio), como enfatizado por Benoît Mandelbrot.⁸ Além disso, variações *absolutas* de diferentes instrumentos financeiros são descorrelacionadas entre si, enquanto variações *relativas* muito frequentemente [§] exibem comportamentos semelhantes entre si, no sentido de estarem conectadas por uma relação de escala.⁹

Seria então mais conveniente introduzir a estocasticidade diretamente em algum dos *retornos*, tratando-os como Gaussianos, ao invés dos incrementos, que apenas herdariam aquela estocasticidade primária? Se sim, qual retorno seria a variável aleatória fundamental,

[‡] Embora Louis Bachelier⁷ mereça elogios por esse trabalho pioneiro.

[§] Em curtíssimas escalas de tempo, *misturas* de modelos aditivos e multiplicativos revelam-se necessárias.⁹

o total ou o relativo? À primeira vista, essa especulação poderia parecer completamente fora de propósito, haja vista que os retornos total e relativo diferem apenas por uma constante aditiva (ver 2.4). Se algum deles for Gaussiano, o outro também o será, e ambos os problemas (a) e (b) supracitados continuam a existir. Contudo, de forma absolutamente surpreendente, essas dificuldades podem ser contornadas no limite de tempo contínuo, que nunca deixou de ser o cenário final concebido desde o início da modelagem.

Como os preços devem variar muito pouco em um pequeno passo temporal, em qualquer instante R_t sempre será próximo de 1 (e, portanto, nunca negativo), enquanto r_t pode exibir qualquer sinal e é pequeno, $|r_t| \ll 1$. Dessa forma,

$$\log R_t = \log(1 + r_t) \approx r_t. \quad (2.5)$$

Parece claro agora que os retornos relativos são os candidatos naturais a serem modelados como Gaussianas, haja vista que, para pequenos passos temporais e certamente também no limite de tempo contínuo, r_t e $\log R_t$ podem ser negativos enquanto R_t permanece positivo e, conseqüentemente, como se vê pela Eq. (2.1), o mesmo ocorre com os preços do ativo [¶]. Ainda por cima, se r_t tem distribuição normal, R_t exibe uma **distribuição lognormal**, conhecida exatamente por exibir valores extremos (distantes da média) mais prováveis em comparação com o caso Gaussiano.⁹ O apêndice B apresenta algumas propriedades básicas de uma distribuição lognormal.

Reorganizando a Eq. (2.3) como

$$S_t = R_{t-1}S_{t-1}, \quad (2.6)$$

agora é óbvio que, a cada instante, um fator estocástico multiplica o preço instantâneo para atualizá-lo, motivo pelo qual o modelo adotado é *multiplicativo*. Quando logaritmos são aplicados à equação acima,

$$\log S_t = \log S_{t-1} + \log R_{t-1}. \quad (2.7)$$

Essa expressão sugere que, quando algum limite de tempo contínuo for considerado, diferenciais de logaritmos de preços serão Gaussianas. Ora, mas o que seria uma *diferencial* de uma variável aleatória?! E como isso poderia ser igual a uma normal? A recorrência resultante pode ser resolvida,

$$\log S_t = \log S_0 + \sum_{i=0}^{t-1} \log R_i, \quad (2.8)$$

e logaritmos de preços serão Gaussianas. Afinal, cada logaritmo de retorno total é uma Gaussiana, e somas de Gaussianas independentes são Gaussianas (ver apêndice A para

[¶] A situação é análoga àquela em que o tamanho de uma população pode permanecer positivo mesmo que a *taxa* do seu crescimento seja negativa em alguns intervalos temporais.

mais detalhes). Mas isso não resolve as dúvidas acerca das diferenciais e do limite de tempo contínuo.

Essas questões demandam a análise cuidadosa que virá logo em seguida, mas ainda neste momento de informalidade é conveniente adiantar que, em algum sentido a distribuição daquela diferencial será uma normal diferencial e surgirá a expressão

$$d(\log S) = \nu dt + \sigma dW, \quad (2.9)$$

que expressa a superposição de uma componente determinística νdt , que representará os efeitos estáveis do ambiente econômico na dinâmica de preços, com uma componente estocástica de média nula e caráter normal. Os parâmetros ν e σ modulam a intensidade desses efeitos.

2.2 Processo de Wiener e o Movimento Browniano

A Eq. (2.7) descreve um processo em que preços revelam-se Gaussianos pela sucessiva adição de retornos (relativos) Gaussianos. Revela-se conveniente adotar, mesmo que apenas provisoriamente, uma representação simplificada daquela equação. Se X_t for a variável normal que se deseja modelar, a partir de impulsos normais ζ_t , a dinâmica estocástica está qualitativamente preservada na relação

$$X_{t+1} = X_t + \zeta_t \quad (2.10)$$

cujas parcelas são normais. O modelo é multiplicativo para as variáveis lognormais mas é aditivo para seus logaritmos - afinal, o logaritmo de um produto é a soma dos logaritmos das parcelas, não?

Porém, como essa relação exhibe um passo temporal unitário, ela também precisa ser modificada, para viabilizar o surgimento de um passo temporal arbitrário Δt que possa tender a zero no limite do contínuo. Qualquer instante t_k considerado agora envolve k passos,

$$t_k = k\Delta t, \quad (2.11)$$

mas é importante perceber que qualquer tempo macroscópico t , finito, emerge de um limite $k \rightarrow \infty$, $\Delta t \rightarrow 0$, $k\Delta t = t_k \rightarrow t$. Em particular, há infinitos passos temporais Δt separando quaisquer dois instantes macroscópicos t e s , $t > s$, por exemplo. Equivalentemente,

$$t_{k+1} = t_k + \Delta t \quad (2.12)$$

e a Eq. (2.10) pode ser convenientemente transformada em

$$X(t_{k+1}) = X(t_k) + (\Delta t)^{1/\alpha} \xi(t_k), \quad (2.13)$$

onde $\xi(t_k) \sim N(0, 1)$ e α é um parâmetro que calibra a ordem de anulamento dos impulsos estocásticos. Uma forma alternativa é

$$X(t_k) = X(t_j) + (\Delta t)^{1/\alpha} \sum_{i=j}^{k-1} \xi(t_i). \quad (2.14)$$

O que houve aqui? Nada havia sido dito sobre ζ_t na Eq. (2.10) além dela ser uma variável Gaussiana. Porém, o termo correspondente àquela variável na Eq. (2.13) certamente precisa anular-se quando Δt tender a zero. Omitindo-se neste momento um possível fator de escala (mas cuja inserção será perfeitamente possível posteriormente, como mostra o fator σ na Eq. (2.9)), é conveniente concentrar esse comportamento infinitesimal em um fator multiplicativo $(\Delta t)^{1/\alpha}$ e incorporar o caráter Gaussiano na variável normal *padrão* $\xi(t_k)$.

O parâmetro α pode, à primeira vista, parecer uma complicação desnecessária. Não é óbvio que $\alpha = 1$? Nada poderia ser mais falso. A situação **não** é análoga à do cálculo usual, onde qualquer discretização para a solução numérica de uma equação diferencial ordinária envolveria simplesmente um fator linear Δt . Neste contexto, são variáveis aleatórias (infinitas, no limite) sendo somadas, e não incrementos escalares, constantes. Se $\alpha \neq 2$, a Eq. (2.13) leva somente a soluções degeneradas para X_t (ou a variância diverge ou se anula). Somente quando $\alpha = 2$ emergem Gaussianas de variâncias finitas ^{||}.

Para subsidiar essas afirmações, é importante calcular a média e a variância de $X(t_k) - X(t_j)$ na Eq. (2.14). Como a média é distributiva em somas e $\langle \xi(t_i) \rangle = 0$,

$$\langle X(t_k) - X(t_j) \rangle = 0. \quad (2.15)$$

O estudo da variância revela que há uma hipótese de modelagem que ainda estava indefinida. Como discutido no apêndice A, fatores multiplicativos afetam quadraticamente a variância de uma normal e a variância da soma de normais *independentes* é a soma das variâncias das parcelas. Assim, sob a hipótese de que $\xi(t_k)$ e $\xi(t_j)$ sejam independentes sempre que

^{||} Na verdade, o termo $\sqrt{\Delta t}$ que decorre de $\alpha = 2$ encontra claros paralelos conceituais com o fator \sqrt{n} que aparece no teorema central do limite (ver apêndice A) e com o processo de difusão regular (não anômala), com variância linear no tempo, tanto na versão contínua da equação de calor quanto no *random walk* em tempo discreto.

$k \neq j$,

$$\begin{aligned}
\text{var}[X(t_k) - X(t_j)] &= \text{var} \left[(\Delta t)^{1/\alpha} \sum_{i=j}^{k-1} \xi(t_i) \right] = \\
&= (\Delta t)^{2/\alpha} \text{var} \left[\sum_{i=j}^{k-1} \xi(t_i) \right] = \\
&= (\Delta t)^{2/\alpha} \sum_{i=j}^{k-1} \text{var}[\xi(t_i)] = \\
&= (\Delta t)^{2/\alpha} \sum_{i=j}^{k-1} 1 = \\
&= (\Delta t)^{2/\alpha} (k - j). \tag{2.16}
\end{aligned}$$

De acordo com a discussão sobre a natureza do limite do contínuo logo após a Eq. (2.11), fica claro que o resultado acima só não será degenerado (zero ou infinito) se $\alpha = 2$, como afirmado antes destes últimos cálculos. O termo $k - j$ tende a infinito com a mesma ordem do número total de passos para criar o intervalo temporal macroscópico $t - s$, e ele gera um resultado finito somente quando contra-balançado por Δt elevado ao expoente 1.

Nessa condição particular, $\alpha = 2$, é tradicional usar a notação $W(t)$ ao invés de $X(t)$, e o processo estocástico compreendido como o limite de tempo contínuo da recorrência

$$W(t_{k+1}) = W(t_k) + \sqrt{\Delta t} \xi(t_k), \tag{2.17}$$

é simbolicamente denotado como

$$dW(t) = \sqrt{dt} \xi(t), \tag{2.18}$$

ou mais concisamente como $\boxed{dW = \sqrt{dt} \xi}$ e é conhecido como **processo de Wiener**,¹⁰ em reconhecimento às descobertas de Norbert Wiener (1894-1964). Como $\xi(t) \sim N(0, 1)$, as representações heurísticas $dW(t) \sim N(0, dt)$ e $\sigma dW(t) \sim N(0, \sigma^2 dt)$ também são de uso comum. Para matemáticos e probabilistas, o processo de Wiener é o que se chama de **movimento Browniano** (às vezes, o adjetivo adicional *padrão* é utilizado), embora físicos reservem esse termo para outro processo estocástico.

Por fim, convém ressaltar que, quando se busca uma definição consistente, muitas vezes o caminho é *definir* o processo de Wiener como o processo estocástico que satisfaz as condições

1. se $s < t$, $W(t) - W(s) \sim N(0, t - s)$;
2. se $0 \leq t_1 < t_2 \leq t_3 < t_4$, $\text{cov}[W(t_2) - W(t_1), W(t_4) - W(t_3)] = 0$, onde cov indica a covariância. ;
3. $\text{prob}[W(0) = 0] = 1$.

2.3 Cálculo de Itô

A equação heurística Eq. (2.18) é sugestiva do natural desejo que o processo de Wiener e outros processos mais complexos nele baseados sejam estudados da forma mais efetiva possível **. De fato, tornou-se popular um formalismo baseado em uma notação similar àquela do cálculo usual, que facilita as manipulações algébricas envolvendo o movimento Browniano padrão. Esse formalismo é denominado **cálculo estocástico** ou **cálculo de Itô**, e foi desenvolvido por Kiyosi Itô (1915-2008). Porém, é fundamental ressaltar de partida que sua aplicação inspira cuidados: embora a notação do cálculo de Itô tenha sido propositalmente escolhida para inspirar familiaridade com técnicas já conhecidas, ele descreve objetos matemáticos distintos do cálculo usual e, inevitavelmente, há contextos em que *diferenciais estocásticas como dW comportam-se de forma distinta das diferenciais usuais*.

O que está sendo denominado uma diferencial estocástica? Ora, trata-se exatamente de uma expressão simbólica como dW , que é uma variável aleatória com variância infinitesimal, no caso, dt . Assim como uma diferencial do cálculo usual não admite uma interpretação precisa por si só mas indica o que está sendo integrado, uma diferencial estocástica é um símbolo formal para indicar que será realizada uma soma de infinitas variáveis aleatórias de variância infinitesimal para resultar em uma variável aleatória finita, um procedimento denominado **integração estocástica**.

A partir deste ponto, o cálculo de Itô exibe uma complexidade técnica que transcende o escopo desta dissertação. Embora o processo de Wiener seja tão fundamental que em qualquer teoria de integração estocástica ††

$$\int_0^t dW = W(t) - W(0) = W(t) \sim N(0, t), \quad (2.19)$$

existem *diversas* teorias matematicamente válidas de integração estocástica, que podem gerar resultados distintos para processos estocásticos mais complexos, e nem sempre é claro qual delas é a mais apropriada para descrever um dado cenário físico ou de modelagem financeira. Deste ponto em diante, diversas sutilezas não serão sequer mencionadas e a apresentação focará nos aspectos que são universalmente válidos no cálculo estocástico.

As diferenças entre o cálculo usual e qualquer versão de cálculo estocástico são dramaticamente ilustrados pelo fato de que *uma diferencial estocástica de 2ª ordem equivale a uma diferencial determinística de 1ª ordem*. Ainda informalmente, ao multiplicar $\xi(t)$ na Eq. (2.18), o fator \sqrt{dt} não afeta a média (pois $\langle \xi(t) \rangle = 0$) e multiplica a variância de

** Depois de que são aprendidas as regras básicas de derivação, raramente é preciso tratar derivadas como limites de diferenças discretas em um limite de contínuo.

†† A ideia aqui é somar infinitas normais independentes uma da outra, todas de média nula e mesma variância dt . O resultado é uma normal de média nula e cuja variância é a soma das variâncias das parcelas. Mas a soma é contínua, uma integral, de todos os dt : o resultado é o tempo macroscópico t .

$\xi(t)$ como dt , de modo que na realidade nunca surge a dificuldade de uma soma/integral envolvendo uma raiz de incremento. Por outro lado, o quadrado de uma diferencial estocástica (que é o que se denominou diferencial estocástica de 2ª ordem) tem média não nula (suas possíveis realizações são todas não negativas), que é *afetada* por um fator dt . Sua variância, contudo, sente $(dt)^2$, que anula-se rápido demais e não contribui no limite.

Para ilustrar de forma bem mais concreta essas considerações informais, é conveniente apresentar a importante identidade

$$(dW)^2 = dt \quad (2.20)$$

e demonstrar sua validade. Ora,

$$dW = \sqrt{dt} \xi \implies (dW)^2 = dt \xi^2, \quad (2.21)$$

de modo que

$$\langle (dW)^2 \rangle = dt \langle \xi^2 \rangle \quad (2.22)$$

e

$$\text{var}[(dW)^2] = (dt)^2 \text{var}[\xi^2]. \quad (2.23)$$

O quadrado de uma normal padrão é uma variável aleatória muito bem conhecida, denominada χ^2 . Como apresentado no apêndice C, sua média vale 1, de modo que $\langle (dW)^2 \rangle = dt$. Sem variabilidade, $(dW)^2$ é essencialmente igual á sua média em qualquer cenário de tempo contínuo ^{‡‡}.

Antes que a identidade $(dW)^2 = dt$ possa ser produtivamente utilizada, é conveniente apresentar o cenário mais simples de construção abstrata de processos estocásticos baseados no processo de Wiener. Heuristicamente, mas com uma interpretação microscópica análoga àquela da Eq. (2.17), o incremento de um novo processo X é representado como

$$dX = a(X, t) dt + b(X, t) dW, \quad (2.24)$$

ou, equivalentemente,

$$X(t) = X(0) + \int_0^t a(X, s) ds + \int_0^t b(X, s) dW(s), \quad (2.25)$$

segundo *algum* sentido de integração estocástica. Destaca-se que mesmo a parcela do incremento de natureza determinística carrega, em geral, um caráter estocástico, mediante

^{‡‡} A ideia que uma certa função - como tomar o quadrado - de uma variável aleatória poderia torná-la determinística nada tem de exótica, nem requer qualquer cenário infinitesimal. Uma moeda, justa ou não, pode ser descrita probabilisticamente por uma variável aleatória binária X , com possíveis resultados $+1$ e -1 . A variável aleatória X^2 é evidentemente determinística, pois só gera o resultado $+1$.

a possível dependência em X do coeficiente $a(X, t)$. Mesmo que isso ocorra, se $b(X, t)$ não depender de X , X é um processo de **ruído aditivo**. Em contraste, quando $b(X, t)$ efetivamente depende de X , X é um processo de **ruído multiplicativo**, muito mais complexo, em geral. Mas um certo caso de ruído multiplicativo não apenas é tratável como também revela-se o modelo básico para o preço de um ativo financeiro! Tal caso será descrito após a apresentação de alguns casos relevantes de ruído aditivo.

Note que, nesta discussão, nunca faria sentido considerar $b(X, t) \equiv 0$, pois isso eliminaria qualquer estocasticidade. O caso ainda geral $a(X, t) \rightarrow f(t)$, $b(X, t) \rightarrow g(t)$ é solúvel, em princípio. Se $\xi(t)$ for, novamente, uma normal padrão,

$$\begin{aligned} X(t) &= X(0) + \int_0^t f(s) ds + \int_0^t g(s) dW(s) \implies \\ \implies X(t) &= X(0) + \int_0^t f(s) ds + \left\{ \sqrt{\int_0^t g^2(s) ds} \right\} \xi(t), \end{aligned} \quad (2.26)$$

pois a integral estocástica é a soma de infinitas normais padrão, cada uma multiplicada por $g(s)\sqrt{ds}$ e, conseqüentemente, de variância $g^2(s)ds$. Quando $a(X, t) \equiv 0$, $b(X, t) = \sigma$, X é um processo de Wiener reescalado, que se reduz ao caso canônico quando $\sigma = 1$.

Um caso diferente e bem importante é

$$dX = \lambda X(t) dt + \sigma dW. \quad (2.27)$$

Um fator integrante $\exp(-\lambda t)$, inspirado na solução determinística $\exp(+\lambda t)$, multiplicando todas as parcelas da Eq. (2.27) revela que, com a mudança de variável estocástica $X(t) \rightarrow Y(t)$ dada por

$$Y(t) = X(t) \exp(-\lambda t), \quad (2.28)$$

vale a relação

$$dY = \exp(-\lambda t) \sigma dW, \quad (2.29)$$

cuja solução é dada pela Eq. (2.26) com $f(s) \equiv 0$ e $g(s) = \sigma \exp(-\lambda s)$. Feitos os cálculos necessários,

$$X(t) = X(0) \exp(\lambda t) + \left\{ \sqrt{\frac{\sigma^2}{2\lambda} (\exp(2\lambda t) - 1)} \right\} \xi(t), \quad (2.30)$$

ou seja, $X(t)$ é uma Gaussiana de média $X(0)e^{\lambda t}$ e variância $\frac{\sigma^2}{2\lambda}[e^{2\lambda t} - 1]$. Por que esse caso é relevante? Os matemáticos e probabilistas o denominam **processo de Ornstein-Uhlenbeck**,¹¹ mas os físicos o chamam de movimento Browniano! De fato, se $X(t) \rightarrow v(t)$ for a velocidade aleatória de uma partícula, $\lambda \rightarrow -\gamma$ for uma constante negativa associada

à dissipação da energia da partícula pelo meio fluido onde ela se encontra e $\sigma^2 \rightarrow \Gamma$ calibrar as flutuações na velocidade da partícula induzidas pelo meio, a **equação diferencial estocástica** formal que surge quando a relação diferencial Eq. (2.27) é manipulada para parecer uma equação diferencial *mesmo que não faça sentido a noção de derivada de uma variável aleatória* é a **equação de Langevin**¹²

$$\frac{dv}{dt} = -\gamma v + \zeta(t), \quad (2.31)$$

convenientemente interpretada como a aceleração resulta de dissipação e um impulso estocástico $\zeta(t)$, mesmo que tal impulso, que, a menos de uma constante, pareceria corresponder à expressão dW/dt , não faça nenhum sentido matemático.

Por mais que sejam interessantes, esses casos vistos até agora (primeiro $a(X, t) \rightarrow f(t)$ e $b(X, t) \rightarrow g(t)$, depois Ornstein-Uhlenbeck, $a(X, t) \rightarrow \lambda X(t)$ e $b(X, t) \rightarrow \sigma$) ainda são simples demais para ilustrar o cálculo de Itô e sua efetiva aplicação na modelagem de preços. Embora uma mudança de variável $X \rightarrow Y = f(X, t)$ só tenha sido explicitada no caso Ornstein-Uhlenbeck, quando isso era indispensável na busca de uma solução, em todos os casos anteriores havia alguma mudança de variável que levava a uma relação do tipo $dY = g(t)dW$, como a Eq. (2.29). Isso é insuficiente para casos mais complexos, já no âmbito deste trabalho.

Em geral, se vale a Eq. (2.24), repetida aqui por conveniência,

$$dX = a(X, t) dt + b(X, t) dW,$$

e um novo processo de difusão Y depende de X mediante uma certa forma funcional f , $Y = f(X, t)$, é a **fórmula de Itô**

$$dY = \left(\frac{\partial f}{\partial t} + a \frac{\partial f}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2 f}{\partial x^2} \right) dt + b \frac{\partial f}{\partial x} dW \quad (2.32)$$

demonstrada no apêndice D com o auxílio de $(dW)^2 = dt$, que expressa corretamente dY .

Somente agora estão disponíveis todos os conceitos para que o modelo básico para que a evolução temporal do preço de um ativo estocástico, Eq. (2.9), reproduzido aqui por conveniência,

$$d(\log S) = \nu dt + \sigma dW,$$

seja realmente compreendida. Nesse modelo, o logaritmo do preço do ativo sofre incrementos estocásticos Gaussianos enquanto segue um viés determinístico, representativo da taxa de juros ν que caracterizaria um ambiente econômico estável (pelo menos em um horizonte temporal de duração bem mais longa do que a escala temporal da mudança de preços).

Os físicos chamam de **ruído branco** esse processo estocástico $\{\xi(t)\}$ caracterizado por $\langle \xi(t) \rangle = 0$ e $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$, por sua transformada de Fourier exibir todas as frequências (cores) com as mesmas amplitudes.

É tentador lembrar do cálculo usual, onde $d(\log x) = dx/x$, e imaginar que

$$\frac{dS}{S} = \nu dt + \sigma dW \implies dS = \nu S dt + \sigma S dW,$$

gerando um processo de ruído *multiplicativo* para S . Assim, S é regido por um ruído multiplicativo, mas não este acima. As expressões logo acima estão incorretas, razão pela qual não foram sequer numeradas. Por quê? Note que $\log S$ é uma variável aleatória, concebida como uma função do ativo estocástico de preço S . Dessa forma, dada uma dinâmica estocástica como o **movimento Browniano geométrico**

$$dS = \mu S dt + \sigma S dW \quad (2.33)$$

para S , a diferencial de $\log S$ *precisa* ser descrita pela fórmula de Itô, e, em claro conflito com a intuição pré-estocástica,

$$d(\log S) \neq \frac{dS}{S} \quad (2.34)$$

e o resultado *não é* $d(\log S) = \mu dt + \sigma dW$. De fato, pela regra de Itô,

$$\begin{aligned} d(\log S) &= \left[\mu S \frac{1}{S} + \frac{1}{2} \sigma^2 S^2 \left(-\frac{1}{S^2} \right) \right] dt + \sigma S \frac{1}{S} dW = \\ &= \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dW. \end{aligned} \quad (2.35)$$

Portanto, embora exista uma constante

$$\nu \equiv \mu - \frac{\sigma^2}{2} \quad (2.36)$$

tal que a Eq. (2.33) leva ao modelo fundamental da Eq. (2.9),

$$d(\log S) = \nu dt + \sigma dW,$$

definitivamente ν não é a constante μ na Eq. (2.33). A **volatilidade** σ no movimento Browniano geométrico original, que calibra sua variância, contribui para um viés *na média* do logaritmo do preço do ativo.

Como o modelo lognormal acima admite a integração

$$\log S(t) - \log S(0) = \nu t + \sigma W(t), \quad (2.37)$$

$$\log S(t) \sim N(\log S(0) + \nu t, \sigma^2 t) \quad (2.38)$$

e, de acordo com os resultados do apêndice B, $S(t)$ é lognormal de média

$$\langle S(t) \rangle = \exp \left\{ \left[\log S(0) + \nu t \right] + \frac{\sigma^2 t}{2} \right\} = S(0) \exp \left(\nu t + \frac{\sigma^2 t}{2} \right) \quad (2.39)$$

e desvio padrão

$$\sqrt{\text{var}[S(t)]} = S(0) \exp\left(\nu t + \frac{\sigma^2 t}{2}\right) \sqrt{e^{\sigma^2 t} - 1}. \quad (2.40)$$

Fica claro que o estudo de dinâmicas estocásticas definidas como funções de outras dinâmicas já estabelecidas requerem muito cuidado. Essa constatação tem enorme relevância prática, haja vista que um **derivativo financeiro** é exatamente um instrumento financeiro cujo valor é atrelado a um outro ativo mais simples/fundamental. Esta dissertação tem como objetivo exatamente investigar derivativos financeiros, cuja teoria elementar está descrita no próximo capítulo.

3 DERIVATIVOS FINANCEIROS E A EQUAÇÃO DE BLACK-SCHOLES

O objetivo deste capítulo é apresentar e deduzir a **equação de Black-Scholes**,²

$$\frac{\partial f}{\partial t} + rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = r f, \quad (3.1)$$

que rege a evolução temporal do preço $Y = f(S, t)$ de um ativo derivado de um outro mais fundamental, de preço S dado por um movimento Browniano geométrico,

$$dS = \mu S dt + \sigma S dW, \quad (3.2)$$

dado que os ativos derivado e original são negociados em um ambiente econômico estável que oferece uma taxa de juros r para a evolução *livre de risco* de um capital $B(t)$,

$$dB = r B dt. \quad (3.3)$$

Note que a equação de Black-Scholes depende de r e da volatilidade σ , mas não do parâmetro μ na Eq. (3.2).

Como será discutido neste capítulo, ela indica o preço que minimiza os riscos das transações com o derivativo $f(S, t)$ e revolucionou o mercado de opções, que são os únicos tipos de derivativos que serão discutidos neste trabalho. Ela foi concebida e popularizada por Fischer Black e Myron Scholes,² mas, como Black já havia falecido em 1997, foi apenas Robert Merton quem recebeu o Prêmio Nobel em Economia com Scholes naquele ano. Merton foi o primeiro a publicar um artigo técnico expondo em detalhes os fundamentos matemáticos da equação.

A demonstração desse importante resultado será apresentada após uma breve apresentação dos diversos conceitos relevantes para uma adequada compreensão das hipóteses sob as quais espera-se que a equação de Black-Scholes seja válida.

3.1 Mercado eficiente

O problema da precificação foi mencionado no capítulo 2, mas nada foi dito sobre alguma razão pela qual preços atingem valores de equilíbrio, pelo menos durante certos horizontes temporais. Mesmo que detalhes das trocas de capital e sua posse pelos agentes sejam intratáveis, há princípios gerais que lançam luz sobre essa questão.

Os preços de ativos rapidamente alcançam valores de equilíbrio devido à ação de agentes racionais que sempre evitariam prejuízos óbvios e/ou nunca perderiam oportunidades de obter o lucro livre de risco que tornar-se-ia possível na presença de desequilíbrios. Se, em certo instante, houvesse dois preços distintos para um mesmo ativo, compradores racionais interessados no ativo sempre optariam pelo menor preço. Também fariam o

mesmo compradores de ocasião, que tivessem interesse apenas em pagar pouco pelo ativo para depois revendê-lo pelo maior preço, apropriando-se da diferença. Nesse cenário, vale a **lei da oferta e demanda**: o agente que vende barato o ativo observa muita procura e naturalmente eleva o preço do seu produto para aumentar sua margem de lucro; em contraste, a baixa procura pelos ativos caros pressiona um vendedor a abaixar seu preço, para aumentar suas vendas. Dessa forma, preços altos tendem a cair, enquanto os baixos tendem a elevarem-se, resultando em um preço de quasi-equilíbrio.

O modelo microdinâmico extremamente idealizado que viabiliza a construção de modelos matemáticos tratáveis em finanças é a **hipótese do mercado eficiente**, pela qual os agentes do mercado apresentariam percepção de oportunidades e ações *instantâneas*.

Mercados sempre incluem agentes econômicos diferenciados, como governos e bancos, que influenciam fortemente o ambiente econômico. Tais agentes podem ofertar instrumentos financeiros elaborados, como títulos do Tesouro Nacional e diversos tipos de derivativos. Mas oferecem também uma taxa básica de juros essencialmente livre de risco, como forma de atrair capital. Em um mercado eficiente, as taxas de juros cobradas por empréstimos funcionariam como ativos, e deveriam convergir para um valor comum.

No sentido mais elementar possível, o **risco** de um investimento pode ser entendido como a variabilidade do seu retorno. Dadas alternativas de investimento de mesmo risco, agentes racionais sempre escolhem aquelas com os maiores retornos médios. Porém, o cenário é bem mais nebuloso na comparação entre investimentos de riscos distintos: alguns agentes podem, ainda racionalmente, estarem dispostos a correr maior risco desde que eventuais retornos positivos sejam maiores do que seriam em outras alternativas.

Mesmo assim, em um cenário de racionalidade, espera-se que um agente racional não concentre todos os seus investimentos em poucas opções. Uma estratégia básica para minimizar risco consiste na chamada diversificação de risco, que nada mais do que a distribuição do capital investido em uma **carteira de investimentos** (ou **portfólio**). Surpreendentemente, o conceito de portfólio é útil também na análise teórica de um único derivativo, e não apenas no cenário mais geral, que transcende o escopo desta dissertação, de um agente distribuindo seus investimentos entre diversos ativos financeiros.

3.2 Opções

Uma **opção de compra europeia** é um contrato firmado entre um autor/proponente e um comprador/contratante, estabelecendo certos direitos e obrigações. O contratante paga * um valor $C(S(0), 0)$ pelo direito de adquirir do autor um dado montante do ativo fundamental de preço $S(0)$ pelo **preço de exercício** † K (para cada unidade do ativo básico) depois de um intervalo de tempo T após a contratação, na **data de expiração ou**

* A letra C é devida à expressão em inglês *call option*.

† Em inglês, *strike price*.

de maturidade. O comprador pode optar por não exercer seu direito de compra [‡], mas o autor é obrigado a efetuar a venda pelo preço de exercício, se assim quiser o contratante.

O contratante faz esse investimento porque acredita que, na maturidade, será verdade que $S > K$ (aqui, trata-se de $S(T)$). Se isso realmente ocorrer, ele estaria adquirindo pelo preço baixo K um ativo que ele pode imediatamente revender no mercado por S , lucrando [§] a diferença, $S - K$. Porém, se $S < K$ na maturidade, o comprador não exercitará seu direito e amargará o prejuízo da compra da opção: por que agora ele pagaria caro, K , por um ativo disponível no mercado por um preço menor, S ? Por outro lado, o autor tem interesses diametralmente opostos, ele deseja que $S(T) < K$.

Mas um fato crucial ainda não discutido é que o contrato de opção tem valor e pode ser novamente negociado a qualquer momento! Admite-se que sempre há um comprador disponível (mercado eficiente). Em qualquer instante t entre $[0, T]$, a opção vale $C(S(t), t)$, e é esse preço que pode fazer o papel de f na equação de Black-Scholes (3.1). À medida que $S(t)$ flutua estocasticamente no tempo, o mesmo ocorre com $C(S(t), t)$. Haja vista a discussão do parágrafo acima,

$$C(S(T), T) = \max(S(T) - K, 0). \quad (3.4)$$

Além disso, quanto mais baixo for o valor atual do ativo básico, menor valor terá um contrato baseado na fé que ele atingirá o preço de exercício K na maturidade, de modo que

$$C(0, t) = 0. \quad (3.5)$$

As duas condições de contorno acima são essenciais à dinâmica da opção e devem ser associadas à equação de Black-Scholes na busca de uma solução para o preço dinâmico de uma opção de compra europeia.

Este preço tende a acompanhar o preço do ativo básico subjacente. De fato, em qualquer momento t específico entre a contratação e a maturidade, a opção de compra é tão mais valiosa em uma eventual nova transação quanto maior for $S(t)$, pois o interesse do portador final do contrato é que $S(T)$ seja o maior possível. Contudo, a correlação entre os preços de um ativo e de um derivativo nele baseado não precisa ser positiva. Para ilustrar essa possibilidade, é conveniente, antes da demonstração da equação de Black-Scholes ser discutida, apresentar pelo menos um outro tipo de derivativo.

Uma **opção de venda europeia**, cujo preço será denotado por $P(S(t), t)$ devido ao termo em inglês *put option*, difere de uma de compra no fato que o contratante passa a

[‡] Se o comprador decidir não exercer sua opção, o mais comum é que ele precise pagar uma pequena taxa para o autor, conhecida como **preço da opção**.

[§] Para ser preciso, é preciso lembrar do investimento inicial, $C(S(0), 0)$. Só haverá lucro de verdade se $S(T) - K > C(S(0), 0)$.

ter o direito de vender ao autor o pacote do ativo básico definido na contratação. Como no caso anterior, são definidos um preço de exercício K e um prazo de expiração T . O autor é obrigado a efetuar a compra na maturidade, se o contratante quiser exercer sua opção. Nesta nova situação, o contratante aposta/deseja que o preço do ativo básico caia, para poder vender caro um produto barato na data de expiração. Ao contrário do que ocorria com uma opção de compra, o preço $P(S(t), t)$ de uma opção de venda cai quando $S(t)$ sobe.

Uma possível ação dos agentes investidores que é absolutamente essencial tanto para a prática das transações com derivativos quanto para sua descrição teórica é a eventual compra de uma quantidade *negativa* de ativos. Um ambiente econômico que oferece uma taxa de juros r livre de risco possibilita que um agente contraia um empréstimo para investir nos ativos de risco, enquanto acredita que seu lucro futuro excederá, com folgas, aquela dívida que viabilizou seu investimento.

3.3 Demonstração de Black-Scholes

Agentes racionais negociam, contínua e simultaneamente, tanto um derivativo quanto seu ativo subjacente! As negociações envolvendo qualquer um desses instrumentos financeiros afetam os preços do outro. Mas as flutuações estocásticas dos preços estão correlacionadas. É esperado que, pelo menos aproximadamente, o mercado se auto-organize, com os agentes, racional e espontaneamente, negociando o ativo e o derivativo em proporções ótimas, maximizando o retorno esperado e minimizando riscos. O grande mérito de Black e Scholes foi elaborar um modelo que descreve essa situação, com base na hipótese do mercado eficiente.

Dessa forma, a ideia básica de Black e Scholes é considerar o *hedging*, a formulação de um portfólio de ativos de risco correlacionados (como evidentemente o são um ativo e seu derivativo, mesmo que a correlação não tenha um único sinal definido) que seja (localmente no tempo) *livre de risco*. A eficiência do mercado faz com que o rendimento desse portfólio ótimo, livre de risco pelos efeitos estocásticos serem cancelados pelo balanceamento apropriado do portfólio, equivalha à taxa básica de juros do ambiente econômico.

A demonstração está fundamentada em 4 elementos,

1. o uso da fórmula de Itô,
2. a construção racional de um portfólio elementar com um derivativo e um pacote $(-\Delta)$ de ativos básicos,
3. a escolha de Δ por *hedging* para eliminar o risco e

4. a equivalência racional entre a taxa de juros básica e a taxa de retorno do portfólio ótimo.

Cada um desses elementos corresponde a expressões quantitativas.

1. Para o movimento Browniano geométrico, Eq. (3.2),

$$dS = \mu S dt + \sigma S dW,$$

a fórmula de Itô Eq. (2.32), com $Y = f(X, t)$ mas $X \rightarrow S$, $a(X, t) \rightarrow \mu S$, $b(X, t) \rightarrow \sigma S$, $f(X, t) \rightarrow f(S, t)$ e $dY \rightarrow df$, adquire a forma

$$df = \left(\frac{\partial f}{\partial t} + \mu S \frac{\partial f}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2} \right) dt + \sigma S \frac{\partial f}{\partial S} dW. \quad (3.6)$$

2. O valor $\Pi(S, t)$ de um portfólio com um derivativo de preço $f(S, t)$ e $-\Delta$ unidades do ativo básico de preço unitário S é

$$\Pi = f + (-\Delta)S, \quad (3.7)$$

com diferencial

$$d\Pi = df + (-\Delta)dS, \quad (3.8)$$

pois, embora a escolha do Δ varie no tempo (o balanceamento ótimo dos investimentos dependerá dos preços instantâneos do ativo e do seu derivativo), seu valor é constante entre passos temporais.

3. Se df for eliminado na conjunção das Eqs. (3.6) e (3.8), e a Eq. (3.2) for utilizada na expressão dali resultante, a escolha de *hedging*

$$\Delta^* = \frac{\partial f}{\partial S} \quad (3.9)$$

faz com dW desapareça e com que o incremento diferencial do portfólio ótimo, livre de risco, seja

$$d\Pi^* = \left(\frac{\partial f}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 f}{\partial S^2} \right) dt. \quad (3.10)$$

4. Finalmente, pela eficiência do mercado, $d\Pi^*$ deve ser equivalente ao incremento diferencial $d\Pi$ também livre de risco que caracteriza o ambiente econômico, obtido se $B \rightarrow \Pi$ na Eq. (3.3),

$$d\Pi = r \Pi dt. \quad (3.11)$$

A igualdade

$$d\Pi = d\Pi^* \quad (3.12)$$

juntamente com a Eq. (3.7) para eliminar Π , resulta na equação de Black-Scholes.

4 VOLATILIDADE ESTOCÁSTICA

Neste capítulo, são discutidos temas mais voltados à prática efetiva das transações com opções. Mesmo nos cenários mais simples, é muito difícil obter soluções da equação de Black-Scholes. O próprio modelo é inapropriado para descrever várias situações empíricas relevantes.

Em particular, a volatilidade de um ativo nunca é constante em horizontes temporais típicos, e tratá-la como um processo estocástico é um passo natural como uma tentativa de aperfeiçoamento dos métodos de precificação. Nessa direção, apresento o Modelo de Heston para a volatilidade estocástica, um dos mais utilizados para este problema.

4.1 Limitações de Black-Scholes

As opções europeias de compra ou venda descritas no capítulo anterior estão entre os mais simples tipos de derivativos. Elas constituem praticamente os únicos casos que admitem soluções analíticas, que, mesmo assim, ainda não são expressas pura e diretamente em termos de funções elementares.

Com uma mudança apropriada de variáveis, a equação de Black-Scholes pode ser transformado em uma equação de calor simples, cuja solução depende das condições de contorno do problema. A solução para a opção de compra, sujeita às Eqs. (3.4) e (3.5), é

Call:

$$C(S, t) = Ke^{-r(T-t)}\mathcal{N}(d_2) + S\mathcal{N}(d_1), \quad (4.1)$$

enquanto o preço da opção de venda é

Put:

$$P(S, t) = Ke^{-r(T-t)}\mathcal{N}(-d_2) - S\mathcal{N}(-d_1), \quad (4.2)$$

com as condições de contorno

$$P(S, T) = \max(0, K - S) \quad (4.3)$$

e

$$P(\infty, t) = 0. \quad (4.4)$$

Nos resultados acima, $\mathcal{N}(x)$ a distribuição cumulativa normal

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz \quad (4.5)$$

e os parâmetros d_1 e d_2 são

$$\begin{aligned} d_1 &= \frac{\ln(S/K) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned} \quad (4.6)$$

Esse resultado nos diz que o preço da opção, por este modelo, deve ser a diferença entre o preço atual da ação S e o preço de exercício K , mas também inclui um peso para ambos os preços. O peso $\mathcal{N}(d_2)$ é a probabilidade de que o preço da ação, no vencimento, seja superior ao preço de exercício, enquanto $\mathcal{N}(d_1)$ é a probabilidade condicional de que, se d_2 for satisfeito, também irá considerar o quão distante do preço de exercício o preço da ação está. Esse é um termo importante, pois se a oferta do lançador para o exercício for muito menor do que o valor esperado no vencimento T , ele deveria pelo menos estar oferecendo um preço para o contrato de opção mais alto para compensar.

Observe que também temos um termo $e^{r(T-t)}$ multiplicado pelo preço de exercício K , considerando o fato de que o dinheiro gasto pelo comprador pagando o preço da opção antecipadamente ao lançador poderia ser usado de forma livre de risco investimento, com retorno de r . Essa consideração é um dos principais avanços neste modelo.

Mas este modelo funciona sob o axioma de que a volatilidade de um ativo financeiro mantém-se constante sobre todo o intervalo desejado para a análise. Existem tentativas de correções desta suposição que consideram que este parâmetro também pode ser considerado como uma variável aleatória. Na próxima seção, apresentarei essa classe de modelos e quais parâmetros adicionais são necessários para precificar uma ação.

4.2 Volatilidade estocástica

A volatilidade do índice Ibovespa* (Figura 1) exibe um ruído considerável, com um pico ocorrendo no início da pandemia, mostrando que uma abordagem estocástica desta variável poderia apresentar resultados mais fiéis à realidade.

Um modelo que descreve esse comportamento em algum grau pode ser estruturado com *acoplamento de diferentes processos de Wiener*,

$$\begin{cases} dx &= \mu_1(x, y, t)dt + \sigma_1(x, y, t)dW_1 \\ dy &= \mu_2(y, t)dt + \sigma_2(y, t)dW_2 \end{cases} \quad (4.7)$$

onde x será o log-retorno de uma ação e y uma variável aleatória relacionada à volatilidade. A equação de Fokker-Planck⁹ relacionada a ela será

* A base de dados utilizada para elaborar este gráfico pode ser encontrada no link: <https://br.financas.yahoo.com/quote/%5EBVSP?p=%5EBVSP>

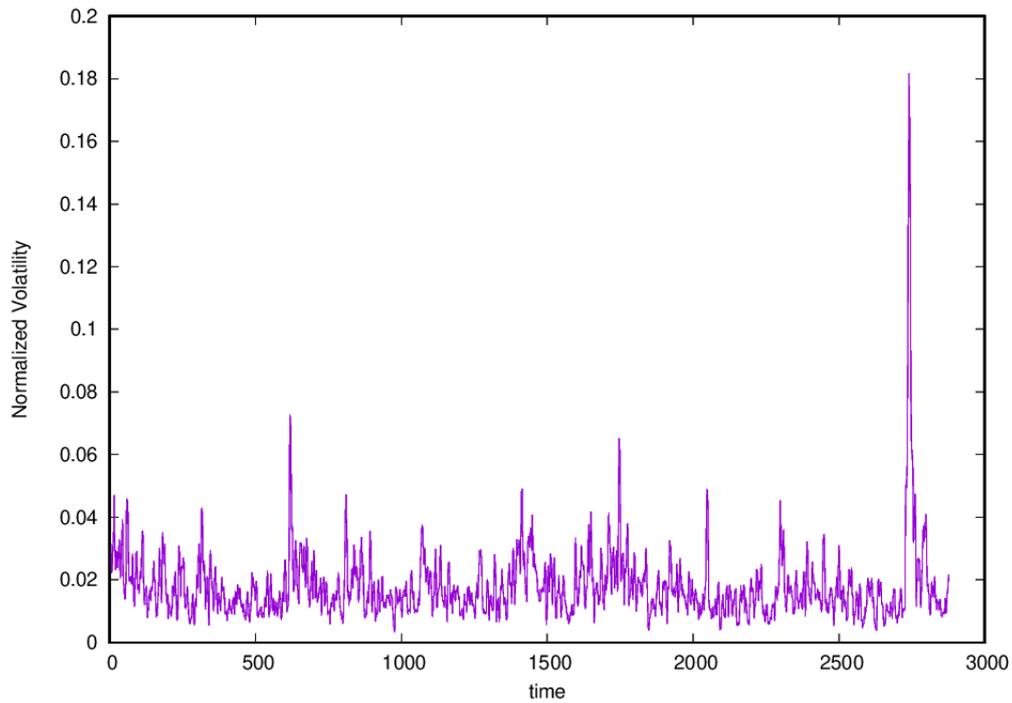


Figura 1 – Volatilidade do índice Ibovespa pelos pregões de 2012 a 2021

Figura 2 – Fonte: Elaborado pelo autor

$$\partial_t \Pi = -\partial_y(\mu_2 \Pi) + \frac{1}{2} \partial_y^2(\sigma_2^2 \Pi). \quad (4.8)$$

Se Π for a distribuição de probabilidade da volatilidade, sua solução estacionária ($\partial_t \Pi = 0$) é

$$\Pi(y) = \frac{N}{\sigma_2^2(y)} \exp\left(-2 \int^y \frac{\mu_2(y')}{\sigma_2^2(y')} dy'\right). \quad (4.9)$$

Neste trabalho, analisarei o comportamento desta abordagem utilizando o Modelo de Heston, um dos primeiros e mais consolidados modelos de volatilidade estocástica. Na próxima seção, discutirei suas características mais relevantes para meu estudo.

4.3 Modelo de Heston

Para o Modelo de Heston, o sistema de Eqs. (4.7) será dado pelas relações

$$\begin{cases} dx &= \mu dt + \sqrt{v} dW_1 \\ dv &= \alpha(\theta - v)dt + k\sqrt{v} dW_2 \end{cases} \quad (4.10)$$

onde a volatilidade e a variável estocástica y são equivalentes ($\sigma^2 = y$) e se a condição $2\alpha\theta > k$ for satisfeita, temos a garantia de que a volatilidade não será negativa.¹³ Conhecendo esses parâmetros, podemos encontrar sua distribuição de probabilidade estacionária, resultando

na seguinte distribuição gama generalizada

$$\Pi(v) = \frac{\gamma^\gamma}{\theta^\gamma \Gamma(\gamma)} \exp\left(-\frac{\gamma v}{\theta}\right) v^{\alpha-1} \quad (4.11)$$

com: $\gamma = 2\frac{\alpha\theta}{k^2}$.

Com estas devidas correções, a equação diferencial parcial que determina o preço de uma ação⁷ deste modelo será

$$\frac{\partial V}{\partial t} + \frac{1}{2} \left(vS^2 \frac{\partial^2 V}{\partial^2 S} + vk^2 \frac{\partial^2 V}{\partial^2 v} + 2\rho S kv \frac{\partial^2 V}{\partial S \partial v} \right) + rS \frac{\partial V}{\partial S} + [\alpha(\theta - v) - k\sqrt{v}] \frac{\partial V}{\partial v} - rV = 0 \quad (4.12)$$

que depende dos valores dos parâmetros α , θ e k para obter a solução da equação, sendo este o objetivo deste trabalho. Foram propostas diversas maneiras para a determinação destes valores, como a própria estimativa de máxima verossimilhança do modelo.¹⁴

5 MÉTODO DO ESTADO ADJUNTO

Esta classe de métodos deseja otimizar um conjunto de parâmetros dados para sistemas lineares, não-lineares ou equação diferencial ordinárias, parciais e, mais recentemente, estocásticas.

Uma abordagem moderna utiliza modelos altamente parametrizáveis, como as redes neurais, que requerem uma quantidade considerável de dados para seu processo de otimização. Em suma, o intuito é propor que a função $f = dy/dt$ para uma EDO, ou, no caso estocástico, propor que as funções $a(X, t)$ e $b(X, t)$ de um processo X possam ser aproximadas por redes neurais, utilizando o método do estado adjunto para o cálculo de seus gradientes.

Os conceitos teóricos necessários para a utilização deste método estão sumarizados nas Figuras 3, 4 e 5 abaixo.

A Figura 3 representa o processo de otimização usual de uma rede neural, com processo de propagação dado pelo termo $\mathcal{NN}(x, \theta)$, onde θ são os parâmetros a serem otimizados e escolhidos arbitrariamente por um chute inicial. Então é atribuída uma função de perda ℓ , que guiará o processo de otimização dos parâmetros ao comparar o resultado obtido por $y = \mathcal{NN}(x, \theta)$ com um conjunto de dados reais (x, y) e sua escolha deverá ser condizente com o problema ao qual ela será aplicada. Para otimizar, calcula-se o gradiente de ℓ em relação aos parâmetros θ que então são atualizados pelo método dos gradientes descendentes $\theta^{j+1} = \theta^j - \alpha \nabla_{\theta} \ell$, repetindo todo este processo pelo número desejado de

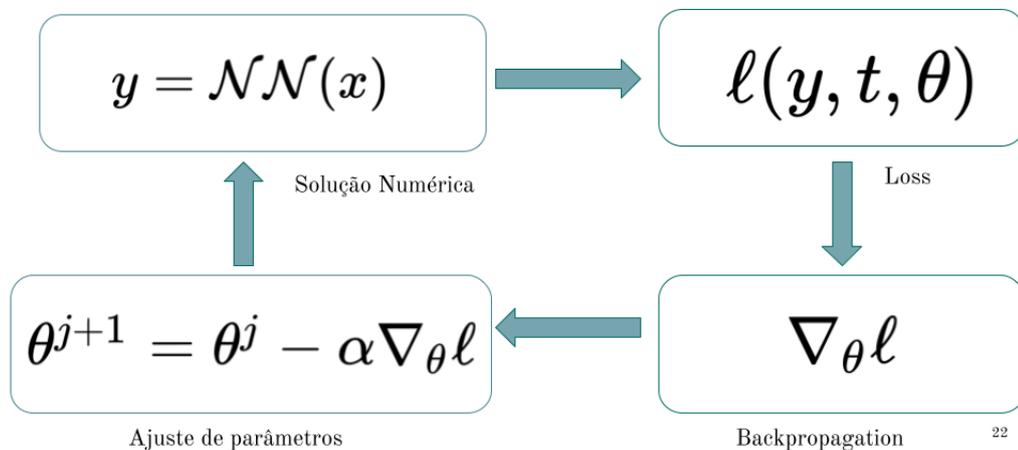


Figura 3 – Processo iterativo de otimização de uma rede neural

Fonte: Elaborada pelo autor

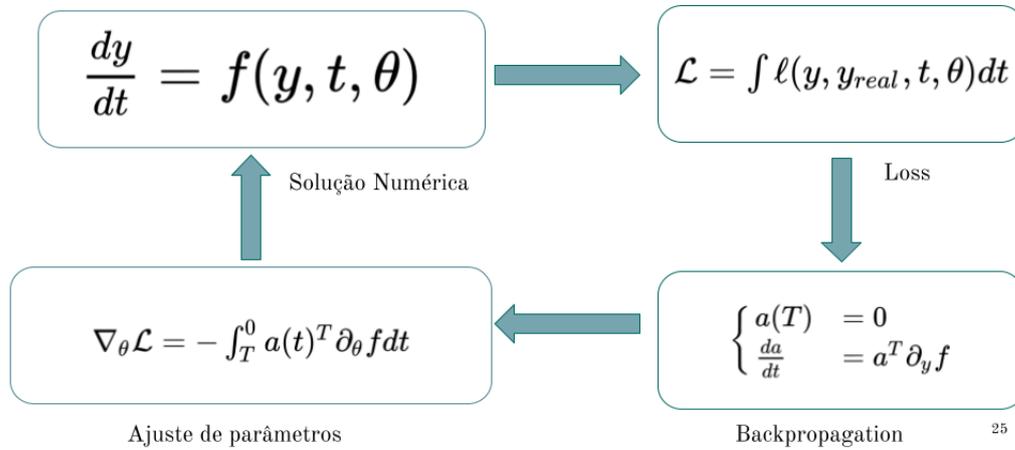


Figura 4 – Processo iterativo de otimização utilizando o método do estado adjunto

Fonte: Elaborada pelo autor

iterações. Os detalhes de como será construída a função $\mathcal{NN}(x)$ estão apresentados nas seções abaixo deste capítulo.

A Figura 4 representa o processo de otimização usual do Método do Estado Adjunto, com processo de propagação dado pelo termo $y' = f(y, t, \theta)$, solucionado numericamente, onde θ são os parâmetros a serem otimizados e escolhidos arbitrariamente por um chute inicial. Então é atribuído um funcional de perda \mathcal{L} , que guiará o processo de otimização dos parâmetros ao comparar o resultado obtido por $y' = f(y, t, \theta)$ com um conjunto de dados reais (y, t) e sua escolha deverá ser condizente com o problema ao qual ela será aplicada. Para otimizar, calcula-se o gradiente de \mathcal{L} em relação aos parâmetros θ , com apoio da equação diferencial adjunta, apresentada na figura pela variável $a(t)$, que então são atualizados pelo método dos gradientes descendentes, repetindo todo este processo pelo número desejado de iterações. Aplicações mais recentes⁴ utilizam uma rede neural para aproximar f , ou seja, $f(y, t, \theta) = \mathcal{NN}(y, t, \theta)$, método denominado como *Neural ODE*, que será apresentado na próxima seção deste capítulo.

A Figura 5 representa o processo de otimização do Método do Estado Adjunto Estocástico, com processo de propagação dado pelo termo $Z(t, \theta)$, solucionado numericamente, onde θ são os parâmetros a serem otimizados e escolhidos arbitrariamente por um chute inicial. Então é atribuído um funcional de perda \mathcal{L} , que guiará o processo de otimização dos parâmetros ao comparar o resultado obtido por Z com um conjunto de dados reais (Z, t) e sua escolha deverá ser condizente com o problema ao qual ela será aplicada. Para otimizar, calcula-se o gradiente de \mathcal{L} em relação aos parâmetros θ , com apoio da equação diferencial adjunta, apresentada na figura pela variável $A(t)$, que então são atualizados pelo método dos gradientes descendentes, repetindo todo este processo

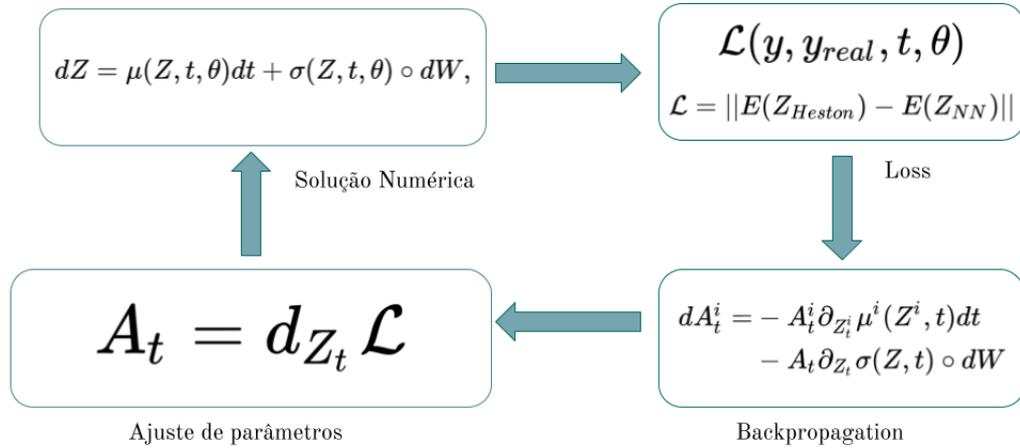


Figura 5 – Processo iterativo de otimização utilizando o método do estado adjunto estocástico

Fonte: Elaborada pelo autor

pelo número desejado de iterações. Aplicações mais recentes⁴ utilizam uma rede neural para aproximar μ e σ , método denominado como *Neural SDE*, que será apresentado na próxima seção deste capítulo.

5.1 Método do Estado Adjunto para Equações Diferenciais Ordinárias

Para o caso de uma EDO da forma

$$f(t, y, \theta) - \frac{dy}{dt} = g(t, y, \theta) = 0, \quad (5.1)$$

sendo a equação 5.1 denominada como **equação de estado**, y as **variáveis de estado** de um sistema e θ os parâmetros referentes a este sistema de equações. O objetivo do método do estado adjunto é determinar quais valores para θ serão os mais apropriados em uma equação para um conjunto de dados observados para as variáveis de estado, minimizados em torno de um funcional $\mathcal{L}(y, y_{real}, t)$, por meio do método dos gradientes descendentes.

O processo para obter o gradiente deste funcional em relação aos parâmetros θ encontra-se de maneira mais detalhada, para uma abordagem mais completa, no apêndice E. Para este trabalho, utilizaremos a equação adjunta⁴ para o gradiente

$$\nabla_{\theta} \mathcal{L} = - \int_T^0 a(t)^T \partial_{\theta} f dt \quad (5.2)$$

$$a(T) = 0 \quad (5.3)$$

$$\frac{da}{dt} = a^T \partial_y f, \quad (5.4)$$

onde $a(t)$ é denominado como termo adjunto e deverá ser solucionado até o tempo inicial $a(0)$ para obter o gradiente.

Esta abordagem é considerada parte da classe de métodos denominada *backpropagation*, que otimiza os parâmetros θ de um problema baseado em resultados reais y_{real} por meio do funcional \mathcal{L} . Devido a proximidade entre este conceito e o modelo de redes neurais, por causa da otimização por meio do gradiente, surge uma nova proposta de algoritmo, denominada EDO's Neurais⁴ (*Neural ODE's*), que utiliza o modelo de redes neurais para aproximar a função f da equação 5.1 e as equações 5.2-5.4 são utilizadas para realizar o processo de *backpropagation* para atualizar os parâmetros θ desta rede.

Avanços recentes na área propõe a extensão deste método para as equações diferenciais estocásticas, objeto de estudo neste meu trabalho. Na próxima seção, apresentarei brevemente as equações desta extensão e posteriormente contextualizarei como ela será aplicada em meu caso, apresentando por fim os resultados que obtive com ela.

5.2 Método do Estado Adjunto para Equações Diferenciais Estocásticas

Compactando a equação (4.10) na forma

$$dZ = \mu(Z, t, \theta)dt + \sigma(Z, t, \theta) \circ dW, \quad (5.5)$$

onde $\mu(Z, t)$ é um vetor e $\sigma(Z, t)$ uma matriz quadrada, que para nosso caso será diagonal, pois considere que não há correlação entre os processos do log-retorno e da volatilidade, ambos com dimensão equivalente a Z .

Então, utilizando o modelo proposto por,³ podemos determinar um processo estocástico A_t adjunto, necessário para determinar o gradiente de \mathcal{L} dado por

$$\begin{cases} dA_t^i &= -A_t^i \partial_{Z^i} \mu^i(Z^i, t)dt - A_t^i \partial_{Z^i} \sigma(Z, t) \circ dW \\ A_t &= d_{Z^i} \mathcal{L} \end{cases} \quad (5.6)$$

com a condição final $A_T = 0$.

Assim como no caso das EDO's, novos avanços na área também propõem a utilização de redes neurais para aproximar as funções $\mu(Z, t, \theta)$ e $\sigma(Z, t, \theta)$, utilizando o conjunto de equações 5.5-5.6 para otimizar os parâmetros θ da função. Esta abordagem de aproximação das funções de uma equação diferencial estocástica é denominada SDE's Neural (*Neural SDE's*).

Na próxima seção, abordarei como o método das SDEs Neurais será utilizado para reproduzir o processo dado pelo modelo de Heston, utilizando um conjunto de simulações Z_{real} do mesmo como entrada para realizar esta reprodução.

5.3 Equações Diferenciais Estocásticas compostas por redes neurais

Além a abordagem clássica, que utiliza equações diferenciais estocásticas para as distribuições conhecidas, contendo poucos parâmetros, está surgindo uma nova abordagem

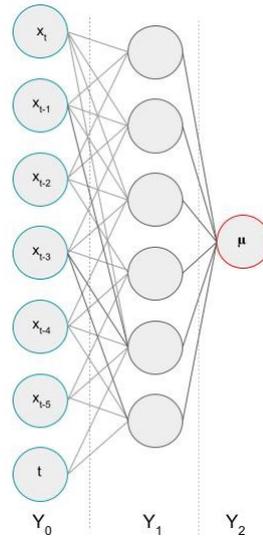


Figura 6 – Representação gráfica de um perceptron multicamadas.

Fonte: Elaborada pelo autor

que considera o uso de redes neurais para encontrar aproximações as funções $\mu(X_t, V_t, t)$ e $\sigma(X_t, V_t, t)$ das equações (4.7), onde a quantidade de parâmetros são ordens de grandezas superiores ao caso clássico.

Neste trabalho, as redes dadas por seu modelo não-linear mais simples, o perceptron multicamadas, que em inglês tem a sigla MLP (*multi-layer perceptron*). O processo iterativo para este modelo é dado pela relação:

$$Y_{i+1} \leftarrow f(W_i Y_i) \quad (5.7)$$

Onde Y_i será a entrada para cada camada de uma rede neural, Y_{i+1} a saída de cada camada e Y_0 serão os vetores de entrada X_t e V_t . A matriz W_i é composta pelos parâmetros a serem otimizados durante o processo, enquanto que a função f é denominada função de ativação e para este trabalho será utilizada a função tangente hiperbólica.

A Figura 6 mostra uma representação gráfica de uma MLP. Onde cada uma das conexões entre as camadas Y_i são os parâmetros da matriz W_i , que serão otimizados pelo método do estado adjunto apresentado anteriormente.

As características das redes neurais que utilizei para aproximar as funções $\mu(Z, t)$ e $\sigma(Z, t)$ são:

- Quantidade de camadas internas: 1
- Quantidade de neurônios em cada camada: 8
- Quantidade de valores de entrada a serem considerados: 16

- Quantidade total de parâmetros: 128

Na próxima seção, será apresentado o resultado da utilização deste modelo para os parâmetros das equações (4.10) de Heston:

- $\mu = 0.2$
- $\alpha = 0.4$
- $\theta = 0.6$
- $k = 0.01$
- $\rho = 0$

A implementação para este trabalho com as especificações acima está disponível no Github*, desenvolvido utilizando a biblioteca Diffrax,¹⁵ criada para providenciar solucionadores numéricos para equações diferenciais, e baseada em um dos exemplos fornecido por seus desenvolvedores.³

5.4 Descrição do código

O código utilizado para este trabalho[†] está disponível por meio de um *notebook* dividido em 7 blocos, sendo que os 2 primeiros deste são apenas para chamar as bibliotecas necessárias e construir a função de ativação da rede neural.

O terceiro bloco é responsável por criar as classes redes neurais que aproximarão $\mu(X, V, t)$ e $\sigma(X, V, t)$, que irão aproximar o modelo de Heston, com μ sendo representada pela classe *Vector Field*, enquanto que a classe *Controlled Vector Field* construirá a rede neural para σ .

O quarto bloco será responsável por construir a *Neural SDE* por uma classe com este mesmo nome, utilizando as classes *Vector Field* e *Controlled Vector Field* para compor seus termos. A classe *Neural CDE* será utilizada de apoio para o cálculo da função de perda a ser otimizada. Para ambas as classes, já está embutida uma função que calcula a solução da *Neural SDE*.

O quinto bloco cria os dados simulados para X e V utilizando o modelo de Heston por meio da função *get_data*.

O sexto bloco cria funções necessárias para o processo de otimização, como a função de perda *loss*, o cálculo do gradiente *grad_loss* e a função *make_step* que realiza uma iteração do Método do Estado Adjunto.

* Repositório com o código utilizado para este trabalho <https://github.com/1beto/Mestrado>

† Repositório com o código utilizado para este trabalho <https://github.com/1beto/Mestrado>

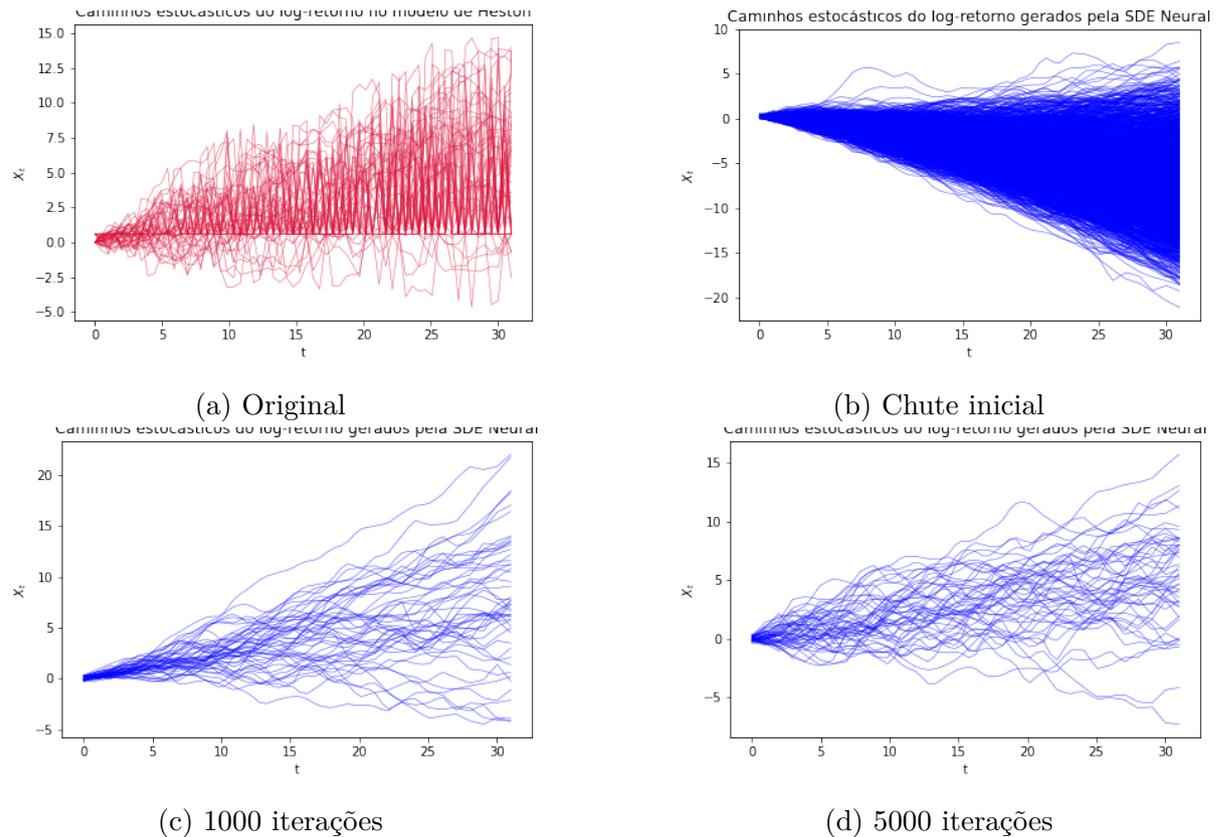


Figura 7 – Caminho estocástico para o log-retorno utilizando: (7a) o modelo de Heston e suas reproduções utilizando o método da SDE Neural com seu (7b) chute inicial e após (7c) 1000 e (7d) 5000 iterações.

Fonte: Elaborada pelo autor

O sétimo e último bloco é o que operacionaliza o código, chamando todas as classes e funções utilizadas previamente para realizar o processo de otimização dos parâmetros e obter os gráficos apresentados na próxima seção deste trabalho.

5.5 Resultados

Utilizando a configuração apresentada para as redes neurais que aproximam as funções $\mu(X_t, V_t, t)$ e $\sigma(X_t, V_t, t)$ em conjunto com o método estocástico adjunto para o ajuste de parâmetros, ambos implementados no código*, temos os resultados das Figuras 7 e 8. Para cada conjunto é exibido o resultado das SDE Neurais após 0 (chute inicial), 1000 e 5000 iterações, onde cada iteração realiza o cálculo do gradiente para a atualização dos parâmetros deste gerador.

O primeiro conjunto de gráficos (Figura 7) apresenta a série temporal de 2048 simulações log-retorno X_t utilizando o modelo de Heston com os parâmetros apresentados na seção acima e sua reprodução utilizando o modelo de SDE Neurais para reproduzi-las. Seu resultado nos mostra que, após uma grande quantidade de iterações, a SDE Neural geradora é capaz de reproduzir visualmente o comportamento dispersivo do log-retorno

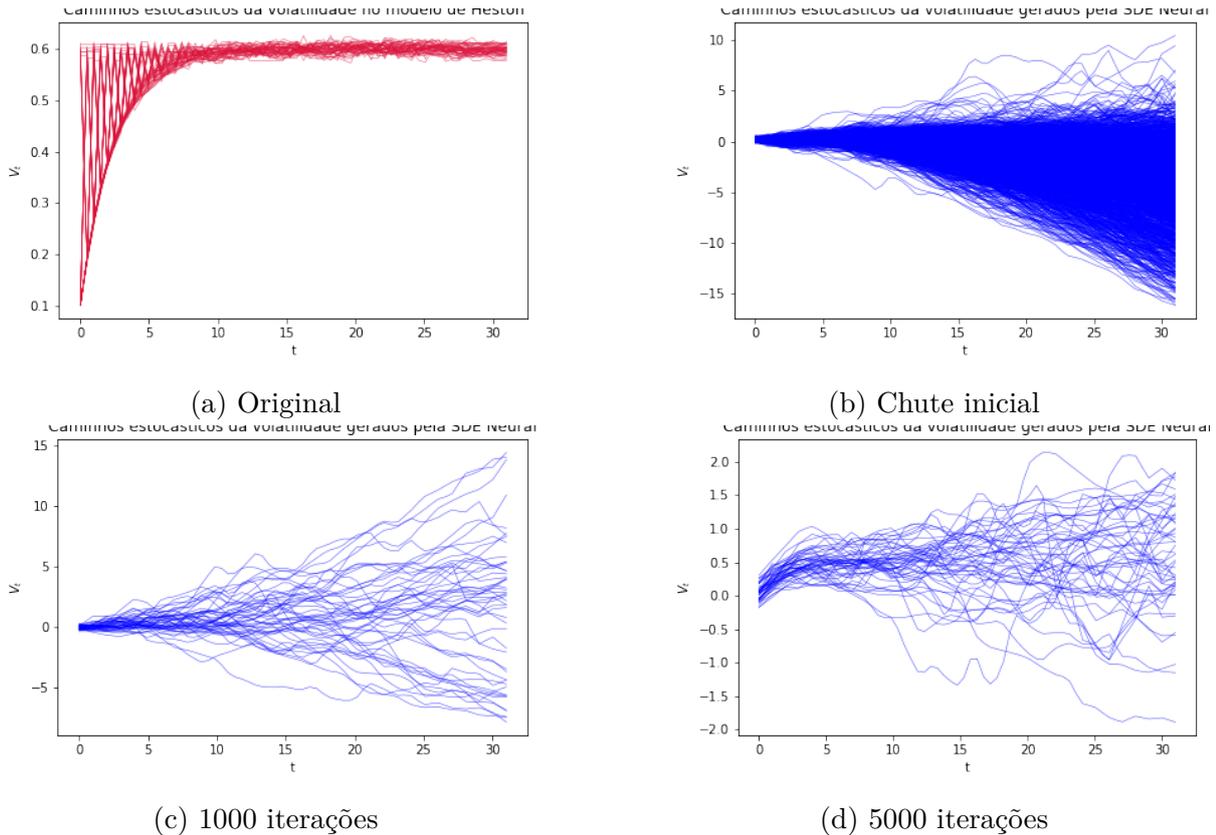


Figura 8 – Caminho estocástico para a volatilidade utilizando: (8a) o modelo de Heston e suas reproduções utilizando o modelo da SDE Neural com seu (8b) chute inicial e após (8c) 1000 e (8d) 5000 iterações.

Fonte: Elaborada pelo autor

para o modelo de Heston, mas sem o caráter oscilatório presente no processo obtido pelo modelo original. Para obter o resultado de 5000 iterações, foram demandadas 10 horas com o código implementado.

O segundo conjunto de gráficos (Figura 8) realiza o mesmo procedimento para a volatilidade V_t . Analisando visualmente o resultado obtido, é possível observar que as distribuições simuladas para volatilidade pelo modelo de SDE Neurais não tem o mesmo comportamento que o modelo de Heston. O principal motivo para este resultado decorre da ordem de grandeza dos valores de V_t ser inferior a X_t , fazendo com que o erro entre o resultado esperado e o resultado obtido para o log-retorno seja muito maior e, por consequência, mais relevante para a atualização dos parâmetros, despriorizando a otimização necessária para reproduzir a dinâmica da volatilidade.

Devido a dificuldade de calibrar seus parâmetros e ao fato da volatilidade neste modelo é uma variável latente, isto é, não há maneira de observá-la empiricamente, mas sim inferida baseada no log-retorno da ação, é possível chegar a conclusão de que o modelo de volatilidade estocástica não é o mais apropriado para a utilização de SDE Neurais, já que o ajuste da distribuição para log-retorno é muito mais relevante para as aplicações

que o mercado financeiro demanda. Os resultados obtidos pela estimação de parâmetros utilizando a máxima verossimilhança apresentados no próximo capítulo corroboram com a conclusão acima.

6 ESTIMAÇÃO DE PARÂMETROS

Pelo resultado do capítulo anterior, observamos visualmente que o comportamento do log-retorno X_t simulado se aproximará consideravelmente das amostras reais, enquanto que na volatilidade V_t é possível observar visualmente que o resultado ainda necessita ser aprimorado. Para realizar uma análise mais robusta destes resultados, neste capítulo será utilizado a estimação de máxima verossimilhança para os dados reais e gerados e estudar os resultados obtidos deste método.

Na primeira seção, será apresentado o método utilizado e as equações necessárias para realizar este cálculo e as características da base de dados para a realização deste estudo, enquanto que a última seção irá apresentar os resultados deste método e as discussões pertinentes relacionados a ele.

6.1 Estimação da máxima verossimilhança

Devido a complexidade envolvida no modelo de Heston, não é possível encontrar uma relação para a função de verossimilhança, surgindo diversas aproximações para utilizar este método. Neste trabalho será utilizado a aproximação para a estimação dos parâmetros μ , α e θ dada por¹⁶

$$\hat{\mu} = \frac{\sum_{n=1}^N \frac{\Delta X_n}{V_n}}{T \sum_{n=1}^N \frac{1}{V_n}}, \quad (6.1)$$

$$\hat{\alpha} = \frac{Ng - \sum_{n=1}^N \Delta V_n}{T \left(\sum_{n=1}^N V_n - \frac{N^2}{\sum_{n=1}^N \frac{1}{V_n}} \right)} \quad (6.2)$$

e

$$\hat{\theta} = \frac{\sum_{n=1}^N \frac{\Delta V_n}{V_n} + \alpha NT}{\alpha T \sum_{n=1}^N \frac{1}{V_n}}. \quad (6.3)$$

onde $\Delta X_n = X_n - X_{n-1}$ e g

$$g = \frac{\sum_{n=1}^N \frac{\Delta V_n}{V_n}}{\sum_{n=1}^N \frac{1}{V_n}}. \quad (6.4)$$

Para realizar este cálculo em meu trabalho, que está disponibilizado por meio da plataforma Github *, utilizo 20480 simulações do modelo de Heston e do modelo simulado para obter os parâmetros estimados. Com os resultados obtidos, realizo um histograma de seus valores em forma de *violin plot* e comparo seus comportamentos na seção abaixo

* Repositório com o código utilizado para este trabalho <https://github.com/1beto/Mestrado>

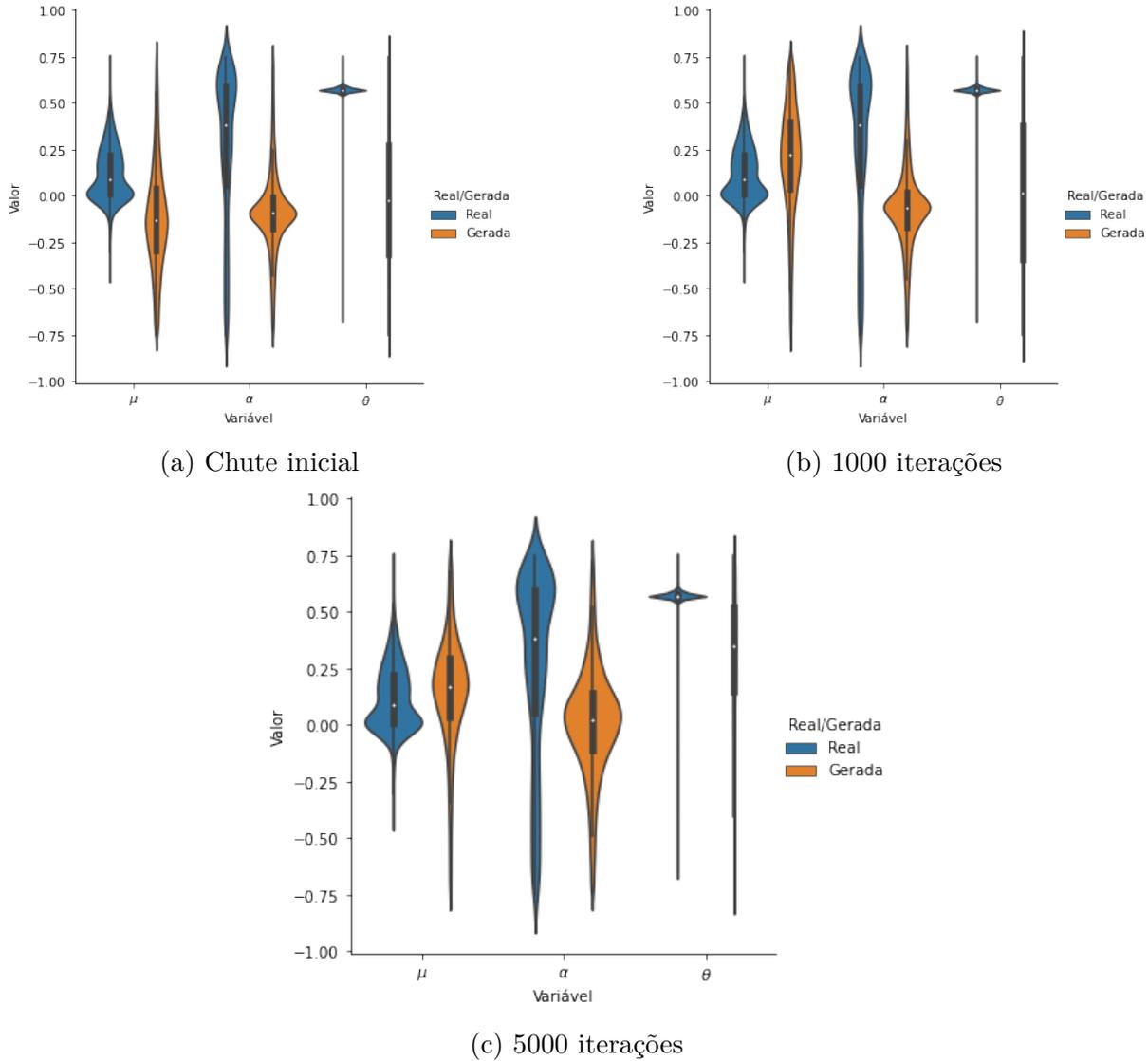


Figura 9 – Parâmetros μ , α e θ estimados pela verossimilhança do modelo de Heston comparado ao modelo da SDE Neural gerado após: o chute inicial (9a), 1000 iterações (9b) e 5000 iterações(9c).

Fonte: Elaborada pelo autor

6.2 Resultados

Utilizando a aproximação para a estimação de parâmetros pelo método da máxima verossimilhança apresentado acima tanto para a série temporal obtida pelo modelo de Heston quanto para a SDE Neural, onde a implementação de seu código pode ser encontrada nas referências deste trabalho (ref código), obtemos os resultados para os parâmetros μ , α e θ , sendo apresentados na forma de histograma no conjunto de figuras 9.

Analisando o histograma das estimativas para os três parâmetros para as 2048 simulações analisadas em cada caso, repara-se que em todos os casos os parâmetros θ e α encontrados após 5000 iterações da SDE Neural não se aproximam dos parâmetros encontrados para o modelo de Heston, enquanto que para μ temos uma concentração de

valores em torno do seu valor real $\mu = 0.2$, mas a distribuição de Heston real também tem uma concentração de valores estimados em torno de $\mu = 0$, que o modelo da SDE Neural não foi capaz de simular. A dificuldade de reproduzir o comportamento da volatilidade se mostra como o grande empecilho para garantir que um resultado que possa ser considerado minimamente satisfatório para essa estimativa.

Assim como o capítulo anterior, os resultados deste capítulo também corroboram com a conclusão de que o modelo de volatilidade estocástica não será o mais apropriado para o método de SDE Neurais no mercado financeiro, já que a adição desta variável complica o processo de otimização e não há vantagens qualitativas nos resultados obtidos ao utilizá-la.

7 CONCLUSÃO

O modelo de volatilidade estocástica surge devido a deficiências observadas no modelo de Black-Scholes para a precificação de opções. Em contrapartida, esta sofisticação faz com que o modelo torne-se muito complexo para certas aplicações, como foi o caso deste estudo. Em minha análise, o método das SDE Neurais foi capaz de reproduzir o comportamento do log-retorno esperado para o modelo de Heston de maneira satisfatória, enquanto que não foi capaz de realizar o mesmo para a volatilidade.

É possível propor técnicas simples para garantir que este método seja capaz de reproduzir o comportamento da volatilidade, como realizar a otimização desta variável de maneira independente do log-retorno, mas como a presença dela não é essencial para a modelagem financeira, realizar este procedimento não terá ganhos em aplicações práticas na área.

Os resultados obtidos neste trabalho também evidenciam o potencial do método das SDE Neurais tem como aproximadores universais de equações diferenciais estocásticas. Foi utilizado a mesma estrutura para todas as distribuições de probabilidade abordadas neste estudo, que foi capaz de simular o comportamento de todos os testes feitos exceto a volatilidade. Assim, é possível concluir que este método abre portas para a utilização de uma nova forma de se abordar este tema, sendo necessário entender quais as vantagens e restrições do mesmo.

Para conseguir realizar este projeto, foi necessário adquirir conhecimento nas áreas de Equações diferenciais estocásticas e Inteligência Artificial, que não eram de meu domínio no início desta pesquisa e estar informado sobre os avanços recentes no segundo destes temas mostra-se cada vez mais estratégico para a compreensão de nossa realidade e tomar uma posição crítica sobre ela, como também me possibilitou o envolvimento em uma das propostas mais promissoras para esta área.

Minha participação em outros projetos relatados no início desta dissertação também merecem ser ressaltadas na conclusão deste trabalho, pois as experiências e conhecimentos obtidos por eles também serão essenciais para minha formação e carreira e os resultados destes projetos também terão impactos de longo prazo mais expressivos que este trabalho que acaba de ser apresentado a você.

REFERÊNCIAS

- 1 HESTON, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. **The Review of Financial Studies**, Oxford University Press, v. 6, n. 2, p. 327–343, 1993.
- 2 BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. **Journal of Political Economy**, The University of Chicago Press, v. 81, n. 3, p. 637–654, 1973.
- 3 KIDGER, P. *et al.* **Efficient and accurate gradients for neural SDEs**. 2021. Disponível em: <https://arxiv.org/pdf/2105.13493>. Acesso em: 07 jul. 2022.
- 4 CHEN, R. T. *et al.* **Neural ordinary differential equations**. 2018. Disponível em: <https://arxiv.org/pdf/1806.07366>. Acesso em: 07 jul. 2022.
- 5 WILMOTT, P. *et al.* **The mathematics of financial derivatives: a student introduction**. Cambridge: Cambridge University Press, 1995.
- 6 LUENBERGER, D. G. *et al.* **Investment science**. Oxford: Oxford University Press, 1997.
- 7 BACHELIER, L. **Louis bachelier’s theory of speculation**. Princeton: Princeton University Press, 2011.
- 8 MANDELBROT, B. B. **Fractals and scaling in finance discontinuity, concentration, risk**. New York: Springer, 1997.
- 9 BOUCHAUD, J.-P.; POTTERS, M. *et al.* **Theory of financial risk and derivative pricing: from statistical physics to risk management**. Cambridge: Cambridge University Press, 2003.
- 10 WIENER, N. **Collected works**. Cambridge: MIT Press, 1976.
- 11 UHLENBECK, G. E.; ORNSTEIN, L. S. On the theory of the brownian motion. **Physical Review**, APS, v. 36, n. 5, p. 823, 1930.
- 12 LANGEVIN, P. Sur la théorie du mouvement brownien. **Comptes-Rendus**, v. 146, p. 530–533, 1908.
- 13 ALBRECHER, H. *et al.* The little Heston trap. **Wilmott**, n. 1, p. 83–92, 2007.
- 14 DUNN, R. *et al.* **Estimating option prices with Heston’s stochastic volatility model**. 2014. Disponível em: <https://www.valpo.edu/mathematics-statistics/files/2015/07/Estimating-Option-Prices-with-Heston%E2%80%99s-Stochastic-Volatility-Model.pdf>. Acesso em: 30 mar. 2022.
- 15 KIDGER, P. **On neural differential equations**. 2022. Disponível em: <https://arxiv.org/pdf/2202.02435>. Acesso em: 07 jul. 2022.
- 16 AZENCOTT, R.; GADHYAN, Y. **Accurate parameter estimation for coupled stochastic dynamics**. n. Special, p. 44, 2009. Disponível em: <https://www.aims sciences.org/article/doi/10.3934/proc.2009.2009.44>. Acesso em: 30 mar. 2022.

Apêndices

APÊNDICE A – DISTRIBUIÇÃO NORMAL (GAUSSIANA) E O TEOREMA CENTRAL DO LIMITE

Uma variável aleatória X é denominada normal ou Gaussiana de média μ e variância σ^2 , o que se denota por $X \sim N(\mu, \sigma^2)$, se ela for descrita pela densidade de probabilidade

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \quad (\text{A.1})$$

Cálculos triviais e tradicionais, encontrados em textos introdutórios de Probabilidade ou Física Estatística, mostram facilmente que os parâmetros μ e σ^2 realmente correspondem às definições de média e variância. Se a média for nula e a variância, unitária, trata-se de uma distribuição Gaussiana **padrão**.

Gaussianas exibem diversas propriedades simples, mas úteis. Por exemplo, se $X \sim N(\mu, \sigma^2)$, as novas variáveis aleatórias aX e $b + X$ obtidas, respectivamente, por uma transformação de escala e uma translação por constantes a e b , são tais que

$$aX \sim N(a\mu, a^2\sigma^2) \quad (\text{A.2})$$

e

$$b + X \sim N(b + \mu, \sigma^2). \quad (\text{A.3})$$

Embora haja formas “menos abstratas” de demonstrar tais propriedades, o formalismo de funções características revela-se particularmente informativo. A função característica $\phi_X(k)$ de uma variável aleatória X é a transformada de Fourier (segundo uma das várias convenções possíveis) da densidade de probabilidade de X ,

$$\phi_X(k) \equiv \int_{-\infty}^{+\infty} \rho_X(x) e^{+ikx} dx = \langle e^{+ikX} \rangle \quad (\text{A.4})$$

Em particular, é possível mostrar que, para $X \sim N(\mu, \sigma^2)$,

$$\phi_X(k) = \exp\left\{ik\mu - \frac{\sigma^2 k^2}{2}\right\}, \quad (\text{A.5})$$

ou seja, novamente uma Gaussiana. Essa forma analítica “ajusta-se” muito bem às propriedades gerais das funções características, como

$$\phi_{aX}(k) = \phi_X(ak) \quad (\text{A.6})$$

e

$$\phi_{b+X}(k) = e^{ikb} \phi_X(k), \quad (\text{A.7})$$

e levam às propriedades supracitadas quando X for Gaussiana. Além disso, em geral, se X_1 e X_2 foram variáveis aleatórias quaisquer, mas independentes,

$$\phi_{X_1+X_2} = \phi_{X_1} \cdot \phi_{X_2}, \quad (\text{A.8})$$

o que, no caso Gaussiano, leva a

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2), \quad (\text{A.9})$$

se $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Esta última propriedade é a razão pela qual a distribuição Gaussiana é ubíqua na natureza e na Matemática: “somadas de efeitos aleatórios são Gaussianos”. Essa expressão informal, contudo, pode levar a enganos. Precisamente, o que pode ser dito?

Uma pedra fundamental da Teoria da Probabilidade é o Teorema Central do Limite (TCL): somas de variáveis aleatórias *independentes* entre si e de variâncias *finitas*, devidamente transladadas e reescaladas, convergem em distribuição para uma distribuição Gaussiana padrão com o aumento do número de parcelas. No caso de n variáveis $\{X_i\}$ independentes e identicamente distribuídas (i.i.d.), $i = 1, \dots, n$, tais que $\langle X_i \rangle = \mu$ e $\text{var}(X_i) = \sigma^2$, define-se

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \quad (\text{A.10})$$

e a convergência descrita no TCL pode ser expressa concisamente como

$$Z_n \xrightarrow{d} Z \quad (\text{A.11})$$

quando $n \rightarrow \infty$, onde $Z \sim N(0, 1)$, ou seja, sua densidade de probabilidade é

$$\rho_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \quad (\text{A.12})$$

As distribuições Gaussianas exibem “caudas leves”, no sentido em que seus decaimentos assintóticos (quando $x \rightarrow \pm\infty$) são exponenciais, mais rápidos do que qualquer potência de x .

APÊNDICE B – DISTRIBUIÇÃO LOGNORMAL

Uma variável aleatória Y é chamada de **lognormal** se o seu logaritmo for uma outra variável aleatória X , com distribuição normal, $X \sim N(\mu, \sigma^2)$. Formalmente,

$$Y = e^X. \quad (\text{B.1})$$

Qual é a densidade de probabilidade de uma variável Y lognormal?

$$\begin{aligned} \rho_Y(y) &= \int dx \rho_X(x) \delta(y - e^x) = \\ &= \frac{[\rho_X(x)]_{x=\log y}}{|[(e^x)']_{x=\log y}|} = \\ &= \frac{1}{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log y - \mu)^2}{2\sigma^2}\right] = \\ &= \frac{1}{y} \rho_X(\log y). \end{aligned} \quad (\text{B.2})$$

Quem são a média e a variância de Y ? Como

$$\langle Y \rangle_{\rho_Y} = \langle e^X \rangle_{\rho_X} \quad (\text{B.3})$$

e

$$\langle Y^2 \rangle_{\rho_Y} = \langle e^{2X} \rangle_{\rho_X}, \quad (\text{B.4})$$

tais grandezas podem ser obtidas “completando quadrados” em cálculos simples com a distribuição Gaussiana, mas que serão omitidos por serem longos. Os resultados são

$$\langle Y \rangle = e^{\mu + \sigma^2/2} \quad (\text{B.5})$$

e

$$\text{var}(Y) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1). \quad (\text{B.6})$$

APÊNDICE C – MÉDIA DE χ^2

Se $\xi \sim N(0, 1)$,

$$\begin{aligned}
 \langle \xi^2 \rangle &= \int dx x \rho_{\xi^2}(x) = \\
 &= \int dx x \left[\int dz \rho_{\xi}(z) \delta(x - z^2) \right] = \\
 &= \int dz \rho_{\xi}(z) \left[\int dx x \delta(x - z^2) \right] = \\
 &= \int dz \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) [z^2] = \\
 &= \frac{1}{\sqrt{2\pi}} (-1) \left\{ \frac{d}{d\alpha} \int dz \exp(-\alpha z^2) \right\}_{\alpha=1/2} = \\
 &= \frac{1}{\sqrt{2\pi}} (-1) \left\{ \frac{d}{d\alpha} \sqrt{\frac{\pi}{\alpha}} \right\}_{\alpha=1/2} = \\
 &= \frac{1}{\sqrt{2\pi}} (-1) \sqrt{\pi} \left(-\frac{1}{2}\right) \left(\frac{1}{2}\right)^{-3/2} = \\
 &= 1.
 \end{aligned} \tag{C.1}$$

APÊNDICE D – FÓRMULA DE ITÔ

A fórmula de Itô é a “expansão de Taylor” para funções de variáveis aleatórias. Reproduzindo mais uma vez a Eq. (2.24) por conveniência,

$$dX = a(X, t) dt + b(X, t) dW,$$

e lembrando que há um \sqrt{dt} em $dW = \sqrt{dt}\xi$, é fácil ver que $dX dt = 0$ “no limite”, pois todos os seus termos apresentam dt elevado a expoentes maiores do que um. Porém, $(dX)^2 \neq 0$, pois $(dW)^2 = dt$. Na verdade, no limite,

$$(dX)^2 = b^2 dt. \tag{D.1}$$

Explicitamente,

$$\begin{aligned} (dX)^2 &= (a dt + b dW)^2 = a^2 (dt)^2 + 2ab dt dW + b^2 (dW)^2 = \\ &= b^2 (dW)^2 + \mathcal{O}[(dt)^2] = b^2 dt. \end{aligned} \tag{D.2}$$

Assim, embora a expressão

$$dy = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dx + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (dx)^2 + \mathcal{O}[(dt)^2] \tag{D.3}$$

seja inconsistente para uma relação $y = f(x, t)$ entre funções comuns, é exatamente ela que descreve adequadamente a relação $Y = f(X, t)$ entre as variáveis aleatórias X e Y .

Dessa forma, utilizando a Eq. (D.3) com as diferenciais estocásticas, $dy \rightarrow dY$, $dx \rightarrow dX$, considerando que $dX = a dt + b dW$, e, claro, usando a Eq. (D.1), segue naturalmente a fórmula de Itô,

$$dY = \left(\frac{\partial f}{\partial t} + a \frac{\partial f}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2 f}{\partial x^2} \right) dt + b \frac{\partial f}{\partial x} dW.$$

APÊNDICE E – GRADIENTES DO MÉTODO ESTOCÁSTICO ADJUNTO

Apresentado pela primeira vez na metade do século anterior, este método deseja determinar, para uma equação diferencial ordinária da forma:

Sendo a equação 5.1 denominada como **equação de estado** e y as **variáveis de estado** de um sistema. A tarefa de determinar quais valores para θ serão apropriados em uma equação para um conjunto de dados observados para as variáveis de estado, como o método dos gradientes descendentes, sendo necessário obter os gradientes de um escalar em relação aos parâmetros θ a ser minimizado.

Escolhendo o funcional:

$$\mathcal{L} = \int_{t_1}^{t_2} \ell(y(t)) dt \quad (\text{E.1})$$

Para ser minimizado, como o desvio quadrático médio em comparação a solução. Utilizando o método dos multiplicadores de lagrange, adicionamos o vínculo:

$$\mathcal{L} = \int_{t_1}^{t_2} [\ell - \lambda g] dt \quad (\text{E.2})$$

Com gradiente em relação a θ resultante:

$$\nabla_{\theta} \mathcal{L} = \lambda(0)^T \frac{dy}{dt}(0) + \int_0^T \lambda \partial_{\theta} f dt + \underbrace{\left(\partial_y \ell + \lambda^T \partial_y f + \frac{d\lambda^T}{dt} \right)}_{(I)} \frac{dy}{d\theta} dt \quad (\text{E.3})$$

Devido a dificuldade de calcular o jacobiano de y em relação a θ é imposto que o termo (I) da equação acima seja nulo, com o a condição final $\lambda(T) = 0$. Obtendo a **equação adjunta**:

$$\begin{cases} 0 &= \lambda(T) \\ 0 &= \partial_y \ell + \lambda^T \partial_y f + \frac{d\lambda^T}{dt} \end{cases} \quad (\text{E.4})$$

Que deverá ser solucionada até o tempo inicial $\lambda(0) = 0$ para obter o gradiente.