

Estimação de Variância Latente no Modelo de Heston via Filtragem Gaussiana

Luiz Eduardo Tojal e Marcio Eisencraft

Resumo—Propõe-se um filtro gaussiano para estimar a variância latente no modelo de Heston discretizado pelo método de Euler-Maruyama. A abordagem explora expressões analíticas para os momentos necessários à filtragem, substituindo a propagação de pontos sigma empregada no filtro de Kalman Unscented e reduzindo o custo aritmético por iteração. Em simulações, obtém-se erro quadrático médio inferior ao de um estimador por média móvel exponencialmente ponderada. As estimativas de variância são então convertidas em volatilidade absoluta do preço e empregadas no cálculo de cotações de compra e venda no modelo de Avellaneda-Stoikov, com dados reais da ação Amazon extraídos da base LOBSTER.

Palavras-Chave—Filtragem gaussiana, estimação de estados, modelo de Heston, volatilidade estocástica, formação de mercado.

Abstract—This paper proposes a Gaussian filter to estimate the latent variance in the Heston model discretized by the Euler-Maruyama method. The approach uses analytical expressions for the moments required by the filtering recursion, replacing the sigma-point propagation employed by the Unscented Kalman Filter and reducing the arithmetic cost per iteration. In simulations, the proposed method achieves a lower mean squared error than an exponentially weighted moving average estimator. The variance estimates are then converted into absolute price volatility and used to compute bid and ask quotes in the Avellaneda-Stoikov model using real Amazon stock data from the LOBSTER database.

Keywords—Gaussian filtering, state estimation, Heston model, stochastic volatility, market making.

I. INTRODUÇÃO

A estimação recursiva de estados latentes em sistemas dinâmicos não lineares é um problema clássico de filtragem Bayesiana [1]. Em finanças, um exemplo relevante é a estimação da variância instantânea em modelos de volatilidade estocástica.

Neste artigo, propõe-se um filtro gaussiano para estimar o processo oculto v_k , associado à variância instantânea no modelo de Heston discretizado. Para esse problema, uma alternativa é o filtro de Kalman Unscented (UKF, do inglês *Unscented Kalman Filter*) [2], que aproxima os momentos necessários à filtragem por meio da propagação determinística de pontos sigma. O filtro proposto difere do UKF na etapa de cálculo desses momentos: em vez da propagação de pontos sigma, utilizam-se expressões analíticas específicas do modelo discretizado, reduzindo o custo aritmético por iteração.

Departamento de Engenharia de Telecomunicações e Controle, Escola Politécnica da Universidade de São Paulo, São Paulo-SP, e-mails: {luizeduardotojal, marciof}@usp.br. Este trabalho foi parcialmente financiado pelo CNPq, processo 404081/2023-1.

O modelo de Heston foi proposto em [3] para descrever preços de ativos com volatilidade estocástica. Nesse modelo, v_t representa a variância instantânea do retorno, de modo que $\sqrt{v_t}$ corresponde à volatilidade instantânea. Em tempo contínuo, a dinâmica do preço S_t e da variância v_t é descrita por

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_1(t), \\ dv_t &= \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_2(t), \\ \rho dt &= \mathbb{E}[dW_1(t) dW_2(t)] \end{aligned} \quad (1)$$

em que μ é o parâmetro de deriva, κ é a taxa de reversão à média, θ é a variância de longo prazo, σ é a volatilidade da variância, ρ é o coeficiente de correlação entre os incrementos dos movimentos brownianos $W_1(t)$ e $W_2(t)$.

O modelo é discretizado pelo método de Euler-Maruyama para derivar o filtro gaussiano. Seu desempenho é comparado ao de um estimador por média móvel exponencialmente ponderada e as estimativas obtidas são aplicadas, de forma exploratória, ao modelo de formação de mercado de Avellaneda-Stoikov [4].

II. DISCRETIZAÇÃO DO MODELO DE HESTON

Para aplicar a filtragem recursiva ao modelo de Heston, é necessário obter uma representação em tempo discreto das equações diferenciais estocásticas. O método de Euler-Maruyama é intuitivo se partirmos da representação integral do sistema.

$$X_t = X_0 + \int_0^t a(X_s) ds + \int_0^t b(X_s) dW_s \quad (2)$$

Assumindo X_s constante em cada intervalo de integração, obtém-se

$$X_t = X_0 + a(X_0)\Delta t + b(X_0) \Delta W_t \quad (3)$$

Pela definição de movimento browniano, a distribuição de ΔW_t é conhecida [5].

Aplicando esse esquema ao modelo de Heston e definindo o retorno simples $r_k \triangleq \frac{S_k - S_{k-1}}{S_{k-1}}$, obtém-se

$$\begin{aligned} r_k &= \mu \Delta t + \sqrt{v_{k-1}} \sqrt{\Delta t} \varepsilon_{1,k}, \\ v_k &= v_{k-1} + \kappa(\theta - v_{k-1}) \Delta t + \sigma \sqrt{v_{k-1}} \sqrt{\Delta t} \varepsilon_{2,k}. \end{aligned} \quad (4)$$

As variáveis aleatórias $\varepsilon_{1,k}$ e $\varepsilon_{2,k}$ são gaussianas, têm média nula, variância unitária e coeficiente de correlação ρ . Além disso, os pares $(\varepsilon_{1,k}, \varepsilon_{2,k})$ são independentes para instantes distintos.

A partir de (4), considera-se que a observação disponível no instante k é o retorno simples r_k . Para aplicar a filtragem gaussiana, definem-se o vetor de estados e de observação

$$\mathbf{x}_k \triangleq \begin{bmatrix} r_k \\ v_k \end{bmatrix}, \quad y_k \triangleq r_k = [1 \quad 0] \mathbf{x}_k. \quad (5)$$

Assim, r_k é observado diretamente, enquanto v_k é a componente latente de interesse, estimada a partir das observações $y_{1:k}$.

III. FILTRO GAUSSIANO

A filtragem Bayesiana obtém recursivamente $p(\mathbf{x}_k | y_{1:k})$ a partir de $p(\mathbf{x}_{k-1} | y_{1:k-1})$, do modelo dinâmico e da nova observação y_k [1]. Para o modelo não linear de Heston, adota-se uma aproximação gaussiana, representando as distribuições envolvidas por suas médias e covariâncias.

Com essa aproximação, a atualização pode ser realizada a partir dos momentos de uma distribuição normal conjunta. Sejam μ_x , μ_y , Σ_{xx} , Σ_{xy} , Σ_{yx} e Σ_{yy} os momentos da distribuição conjunta aproximada de $(\mathbf{x}_k, y_k)$, condicionada às observações anteriores $y_{1:k-1}$:

$$\begin{bmatrix} \mathbf{x}_k \\ y_k \end{bmatrix} \Big| y_{1:k-1} \sim \mathcal{N} \left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right). \quad (6)$$

Nesse caso, os momentos da distribuição posterior $p(\mathbf{x}_k | y_{1:k})$ são dados por

$$\begin{aligned} \mathbb{E}[\mathbf{x}_k | y_{1:k}] &= \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y_k - \mu_y), \\ \text{Cov}(\mathbf{x}_k | y_{1:k}) &= \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}. \end{aligned} \quad (7)$$

Para aplicar (7) ao modelo de Heston discretizado, é necessário calcular os momentos de primeira e segunda ordem de $(\mathbf{x}_k, y_k) | y_{1:k-1}$. Como $y_k = r_k$, esses momentos são obtidos a partir das expressões de r_k e v_k em (4). Denotando $\mu_{v,k-1|k-1} \triangleq \mathbb{E}[v_{k-1} | y_{1:k-1}]$ e $P_{v,k-1|k-1} \triangleq \text{Var}(v_{k-1} | y_{1:k-1})$, obtêm-se os momentos de primeira ordem

$$\begin{aligned} \mathbb{E}[r_k | y_{1:k-1}] &= \mu \Delta t, \\ \mathbb{E}[v_k | y_{1:k-1}] &= \kappa \theta \Delta t + (1 - \kappa \Delta t) \mu_{v,k-1|k-1}. \end{aligned} \quad (8)$$

Os momentos de segunda ordem necessários à matriz de covariância são

$$\begin{aligned} \text{Var}(r_k | y_{1:k-1}) &= \Delta t \mu_{v,k-1|k-1} \\ \text{Var}(v_k | y_{1:k-1}) &= (1 - \kappa \Delta t)^2 P_{v,k-1|k-1} \\ &\quad + \sigma^2 \Delta t \mu_{v,k-1|k-1} \\ \text{Cov}(r_k, v_k | y_{1:k-1}) &= \sigma \rho \Delta t \mu_{v,k-1|k-1}. \end{aligned} \quad (9)$$

Os momentos em (8) e (9) determinam μ_x , μ_y , Σ_{xx} , Σ_{xy} , Σ_{yx} e Σ_{yy} , que são então substituídos em (7). A estimativa de interesse é a segunda componente de $\mathbf{x}_k$, correspondente a v_k . A média e a covariância são atualizadas recursivamente, enquanto os parâmetros $\mu, \kappa, \theta, \sigma, \rho$ permanecem fixos durante a filtragem; sua estimação conjunta é deixada para trabalhos futuros.

IV. AVALIAÇÃO NUMÉRICA

Para o UKF baseado no vetor aumentado $\xi_k = [v_{k-1}, \varepsilon_{1,k}, \varepsilon_{2,k}]^T$, a Transformada Unscented utiliza $2L+1 = 7$ pontos sigma. Uma contagem direta das operações da implementação, desconsiderando o custo de geração dos pontos sigma, resulta em aproximadamente 91 multiplicações e 70 adições por passo, contra 5 multiplicações e 2 adições para as expressões analíticas propostas, considerando Δt constante e constantes auxiliares pré-computadas.

Antes da comparação de desempenho, verificou-se a relação entre o filtro gaussiano e o UKF aumentado. Para a mesma inicialização e os mesmos parâmetros, as estimativas produzidas pelos dois métodos são equivalentes. Essa igualdade não constitui uma propriedade geral do UKF, mas decorre da estrutura particular do modelo em (4). De fato, v_{k-1} é independente das inovações gaussianas, enquanto a correlação entre $\varepsilon_{1,k}$ e $\varepsilon_{2,k}$ é inteiramente descrita por ρ . Assim, os momentos necessários à atualização dependem apenas de $\mathbb{E}[v_{k-1}]$, $\text{Var}(v_{k-1})$, $\mathbb{E}[\varepsilon_{i,k}^2] = 1$ e $\mathbb{E}[\varepsilon_{1,k} \varepsilon_{2,k}] = \rho$.

Os pontos sigma simétricos preservam a média e a covariância de ξ_k [2]. Além disso, com a fatoração em blocos usada no UKF aumentado, sua propagação pelos termos $\sqrt{v_{k-1}} \varepsilon_{i,k}$ reproduz exatamente os momentos em (8) e (9). Consequentemente, ambos os filtros empregam o mesmo ganho e, para $\hat{v}_{k-1|k-1} > 0$, resultam na recursão

$$\begin{aligned} \hat{v}_{k|k} &= \kappa \theta \Delta t + (1 - \kappa \Delta t) \hat{v}_{k-1|k-1} + \sigma \rho (r_k - \mu \Delta t), \\ P_{v,k|k} &= (1 - \kappa \Delta t)^2 P_{v,k-1|k-1} + \sigma^2 (1 - \rho^2) \Delta t \hat{v}_{k-1|k-1}. \end{aligned} \quad (10)$$

Logo, se as médias e covariâncias coincidem no instante anterior, elas também coincidem no instante seguinte. Em aritmética exata, as trajetórias estimadas e seus EQMs. A equivalência pressupõe também o mesmo tratamento de eventuais valores negativos, a mesma regularização e os mesmos parâmetros. Por esse motivo, incluir o UKF como referência de acurácia apenas duplicaria a estimativa do filtro gaussiano. A avaliação a seguir utiliza o EWMA, que possui uma recursão distinta e constitui uma referência independente, permitindo comparar efetivamente os erros de estimação.

Para avaliar as equações de filtragem obtidas, simulou-se o modelo de Heston discretizado em (4) e comparou-se sua estimativa com um estimador por média móvel exponencialmente ponderada (EWMA, do inglês *exponentially weighted moving average*), definido por

$$\hat{v}_k^{\text{EWMA}} = \alpha \hat{v}_{k-1}^{\text{EWMA}} + (1 - \alpha) \frac{(r_k - \mu \Delta t)^2}{\Delta t}. \quad (11)$$

Na geração da trajetória simulada, após cada passo do esquema de Euler-Maruyama, aplicou-se a operação $v_k \leftarrow \max\{0, v_k\}$ para preservar a não negatividade da variância. Como v_k não é observável em dados reais, a simulação permite comparar diretamente $\hat{v}_k$ com sua trajetória verdadeira.

Os parâmetros usados na simulação foram $\mu = 0.05$, $\kappa = 2.0$, $\theta = 0.4$, $\sigma = 0.2$, $\rho = -0.7$, $v_0 = 0.1$ e $\Delta t = 4/252$. Foram testados $\alpha \in \{0.50, 0.70, 0.85, 0.90, 0.95, 0.97, 0.99\}$, sendo o menor EQM igual a 1.06×10^{-2} , obtido para $\alpha = 0.97$. Mesmo considerando o melhor desempenho obtido entre esses valores, esse erro foi superior ao valor 3.3×10^{-3} obtido

com o filtro gaussiano. A Figura 1 ilustra as estimativas obtidas pelos dois métodos.

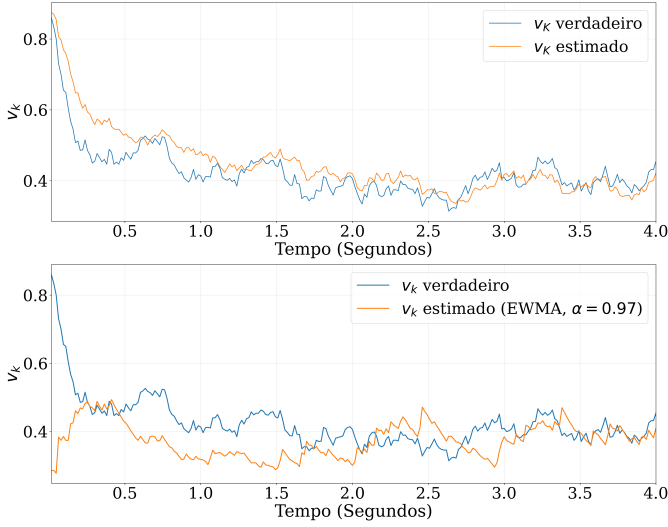


Fig. 1. Comparação entre a variância instantânea simulada v_k e as estimativas obtidas: (a) filtro gaussiano; (b) estimador EWMA com $\alpha = 0.97$.

V. APLICAÇÃO COM DADOS REAIS

Um formador de mercado atua inserindo continuamente ordens de compra e venda no livro de ordens de uma bolsa de valores. A cada instante, esse agente deve escolher cotações suficientemente competitivas para serem executadas, mas que também limitem o acúmulo de inventário.

Esse equilíbrio entre probabilidade de execução e controle de inventário é formalizado no modelo de Avellaneda-Stoikov [4]. Nesse modelo, o preço médio, definido como a média entre a melhor oferta de compra e a melhor oferta de venda, S_t segue um movimento browniano, $dS_t = \sigma dW_t$, e o inventário q_t varia conforme a execução das ordens de compra e venda. As intensidades de execução são modeladas por processos de Poisson com taxas $\lambda^a(\delta^a)$ e $\lambda^b(\delta^b)$, dependentes das distâncias δ^a e δ^b entre as cotações ofertadas e o preço médio. A rapidez com que essas taxas decaem com a distância é controlada pelo parâmetro η .

O objetivo do formador de mercado é maximizar a utilidade esperada da riqueza terminal, composta pelo caixa acumulado e pelo valor do inventário em T . O parâmetro γ controla a aversão ao risco, penalizando a exposição ao inventário.

A escolha das cotações de compra p_t^b e venda p_t^a pode ser formulada como um problema de controle estocástico. Sob as hipóteses do modelo de Avellaneda-Stoikov, obtêm-se as aproximações

$$\begin{aligned} p_t^b &= p_t^r - \delta_t, & p_t^a &= p_t^r + \delta_t, \\ p_t^r &= S_t - q_t \gamma \sigma^2 (T - t), \\ \delta_t &= \frac{1}{\gamma} \ln \left(1 + \frac{\gamma}{\eta} \right) + \frac{\gamma \sigma^2 (T - t)}{2}. \end{aligned} \quad (12)$$

Como v_k representa a variância dos retornos relativos, a variância instantânea absoluta do preço é aproximada por $\sigma_{S,k}^2 \approx S_k^2 \hat{v}_k$. Substitui-se, portanto, σ_S^2 por $S_k^2 \hat{v}_k$ nas expressões anteriores. Dessa forma, as cotações passam a incorporar

variações temporais da variância estimada a partir dos retornos observados.

O filtro foi aplicado a dados da ação Amazon (AMZN) da base LOBSTER [6], referentes ao período das 9h30 às 16h00 de 21 de junho de 2012. Os primeiros 30 minutos foram reservados à calibração de $\mu, \theta, \sigma, \rho$, enquanto $\kappa = \log(2)/300$ foi fixado, correspondendo a uma meia-vida de 5 minutos. Utilizaram-se $\mu = 1.13 \times 10^{-6}$, $\kappa = 2.31 \times 10^{-3}$, $\theta = 3.77 \times 10^{-8}$, $\sigma = 4.41 \times 10^{-5}$, $\rho = 6.57 \times 10^{-2}$ e $v_0 = \theta$. Para Avellaneda-Stoikov, usaram-se $\gamma = 1.46 \times 10^{-5}$ e $\eta = 15.38$. Os resultados têm caráter exploratório, pois não foram utilizados dados de opções na calibração.

Como a estimação usa preços de execução observados em instantes irregulares, Δt é atualizado a cada etapa da predição.

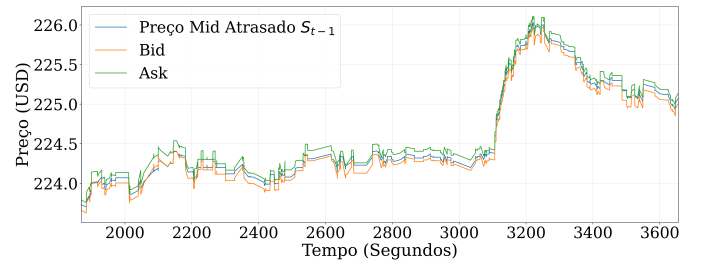


Fig. 2. Cotações de compra e venda dependentes do inventário, obtidas pela adaptação do modelo de Avellaneda-Stoikov usando a variância estimada pelo filtro gaussiano.

Partindo de $q_0 = 100$ e admitindo ordens unitárias, uma compra é considerada executada quando $p_k^b \geq P_k^{\text{exec}}$, incrementando q_k , e uma venda quando $p_k^a \leq P_k^{\text{exec}}$, decrementando-o.

VI. CONCLUSÕES E PRÓXIMOS PASSOS

Propôs-se um filtro gaussiano para estimar a variância latente no modelo de Heston discretizado. O cálculo analítico dos momentos reduz o custo aritmético frente à propagação de pontos sigma do UKF, e, em simulação, obteve menor EQM que o estimador EWMA. As estimativas foram ainda utilizadas em uma aplicação exploratória de formação de mercado com dados reais. Como trabalhos futuros, pretende-se estimar conjuntamente os parâmetros do modelo e estender a avaliação a outros ativos e métricas de formação de mercado.

REFERÊNCIAS

- [1] S. SÄRKKÄ, *Bayesian filtering and smoothing*. Cambridge: Cambridge University Press, 2013.
- [2] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [3] S. L. HESTON, “A closed-form solution for options with stochastic volatility with applications to bond and currency options,” *The Review of Financial Studies*, vol. 6, no. 2, pp. 327–343, 1993.
- [4] M. AVELLANEDA and S. STOIKOV, “High-frequency trading in a limit order book,” *Quantitative Finance*, vol. 8, no. 3, pp. 217–224, 2008.
- [5] P. E. KLOEDEN and E. PLATEN, *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics*. Berlin: Springer, 2013.
- [6] LOBSTER, “Lobster: Limit order book system.” <https://lobsterdata.com/>, 2025. Acesso em: 22 jun. 2025.