

# Análise de Séries Temporais

## Capítulo VI. Modelos de Espaço de Estados

Fernando Lucambio

Departamento de Estatística  
Universidade Federal do Paraná

Novembro, 2023

Os modelos de volatilidade estocástica (SV) são uma alternativa aos modelos do tipo GARCH apresentados na **Seção V.3**.

Ao longo desta seção,  $r_t$  vai denotar os retornos de algum ativo financeiro,

$$r_t = \sigma_t \epsilon_t,$$

onde  $\{\epsilon_t\}$  é uma sequência independente identicamente distribuída e o processo de volatilidade  $\sigma_t$ , é um processo estocástico não negativo tal que  $\epsilon_t$  é independente de  $\sigma_s$  para todo  $s \leq t$ .

Frequentemente, assume-se que  $\epsilon_t$  tem média zero e variância unitária.

Nestes modelos, a volatilidade é uma transformação não linear de um processo auto-regressivo linear oculto, onde o processo de volatilidade oculta  $X_t = \log(\sigma_t^2)$ , segue um processo autoregressivo de primeira ordem,

$$\begin{aligned}X_t &= \phi X_{t-1} + W_t, \\r_t &= \beta \exp\left(\frac{1}{2}X_t\right)\epsilon_t,\end{aligned}$$

onde  $W_t \sim N(0, \sigma_w^2)$  são independentes identicamente distribuídos e  $\epsilon_t$  é o ruído, sendo também independentes identicamente distribuídos com momentos finitos.

Os processos de erro  $W_t$  e  $\epsilon_t$  são considerados mutuamente independentes e  $|\phi| < 1$ .

Como  $W_t$  é normal,  $X_t$  também é normal. Todos os momentos de  $\epsilon_t$  existem, de modo que todos os momentos de  $r_t$ , definido acima, existem também. Supondo que  $X_0 \sim N(0, \sigma_w^2/(1 - \phi^2))$ , a distribuição estacionária, a curtose de  $r_t$  é dada por

$$\kappa_4(r_t) = \kappa_4(\epsilon_t) \exp(\sigma_x^2),$$

onde  $\sigma_x^2 = \sigma_w^2/(1 - \phi^2)$  é a variância estacionária de  $X_t$ .

Lembremos que, para um inteiro  $m$  e uma variável aleatória  $U$ ,

$$\kappa_m(U) = \frac{E(|U|^m)}{\sqrt{E(|U|^2)^m}}.$$

Tipicamente,  $\kappa_3$  é chamado de coeficiente de skewness ou coeficiente assimetria e  $\kappa_4$  é chamado de coeficiente de kurtosis ou coeficiente de curtose.

Assim,  $\kappa_4(t_t) > \kappa_4(\epsilon_t)$ , de modo que se  $\epsilon_t \sim N(0, 1)$  independentes identicamente distribuídos, a distribuição de  $r_t$  é leptocúrtica. As distribuições leptocúrticas são distribuições estatísticas com curtose maior que três.

Podem ser descritas como tendo uma forma mais larga ou mais plana com caudas mais grossas, resultando em uma maior chance de eventos positivos ou negativos extremos. As distribuições leptocúrticas são uma das três categorias principais encontradas na análise da curtose. Suas outras duas contrapartes são as distribuições mesocúrticas, que não apresentam curtose e estão associadas à distribuição normal e as platicúrticas, que apresentam caudas mais finas e menos curtose.

A função de autocorrelação de  $\{r_t^{2m}\}_{t=1,2,\dots}$ , para qualquer inteiro  $m$ , é dada por

$$\text{Corr}(r_{t+h}^{2m}, r_t^{2m}) = \frac{\exp(m^2 \sigma_x^2 \phi^h) - 1}{\kappa_{4m}(\epsilon_t) \exp(m^2 \sigma_x^2) - 1}.$$

A taxa de decaimento da função de autocorrelação é mais rápida do que a exponencial em pequenos intervalos de tempo e, em seguida, estabiliza para  $\phi$  em grandes intervalos.

Às vezes é mais fácil trabalhar com a forma linear do modelo onde definimos

$$Y_t = \log(r_t^2) \quad \text{e} \quad \nu_t = \log(\epsilon_t^2),$$

nesse caso, podemos escrever  $Y_t = \alpha + X_t + \nu_t$ .

A constante é geralmente necessária na equação de estado ou na equação de observação, mas não normalmente em ambas, então escrevemos a equação de estado como

$$X_t = \phi_0 + \phi_1 X_{t-1} + W_t,$$

onde  $W_t$  é o ruído branco Gaussiano com variância  $\sigma_w^2$ . A constante  $\phi_0$  às vezes é chamada de efeito de alavancagem. As duas expressões acima juntas compõem o modelo de volatilidade estocástica de Taylor (1982).

Se  $\epsilon_t^2$  tivesse uma distribuição log-normal,

$$Y_t = \alpha + X_t + \nu_t \quad \text{e} \quad X_t = \phi_0 + \phi_1 X_{t-1} + W_t,$$

formariam um modelo de espaço de estados Gaussiano e poderíamos usar os resultados do modelo dinâmico linear padrão para ajustar o modelo aos dados.

Infelizmente, essa suposição não parece funcionar bem. Em vez disso, geralmente mantemos a suposição de normalidade do ARCH em  $\epsilon_t \sim N(0, 1)$  independentes, caso em que  $\nu_t = \log(\epsilon_t^2)$  é distribuído como o logaritmo de uma variável aleatória qui-quadrada com um grau de liberdade.

Esta densidade é dada por

$$f(\nu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(e^\nu - \nu)\right), \quad -\infty < \nu < \infty.$$

A média da distribuição é  $-(\gamma + \log(2))$ , onde  $\gamma \approx 0.5772$  é a constante de Euler e a variância da distribuição é  $\pi^2/2$ . A suposição do modelo básico de  $\epsilon_t$  ser gaussiano é irreal para a maioria das aplicações. Em um esforço para manter as coisas simples, mas mais gerais no sentido de que permitimos que a dinâmica do erro observacional dependa dos parâmetros que serão ajustados, nosso método de ajuste de modelos de volatilidade estocástica é manter a equação de estado Gaussiana, mas escrever a equação de observação, como

$$Y_t = \alpha + X_t + \eta_t,$$

onde  $\eta_t$  é o ruído branco, cuja distribuição é uma mistura de duas normais, uma centrada em zero.

Em particular, escrevemos

$$\eta_t = I_t Z_{t0} + (1 - I_t) Z_{t1},$$

onde  $I_t$  é um processo Bernoulli independentes,  $P(I_t = 0) = \pi_0$ ,  $P(I_t = 1) = \pi_1$ , sob a condição  $\pi_0 + \pi_1 = 1$ ,  $Z_{t0} \sim N(0, \sigma_0^2)$  independentes identicamente distribuídas e  $Z_{t1} \sim N(\mu_1 \sigma_1^2)$  independentes identicamente distribuídas também.

A vantagem desse modelo é que ele é fácil de ajustar porque usa a normalidade. De fato, as equações que definem do modelo são semelhantes às apresentadas em Peña and Guttman (1988), que utilizaram a ideia para obter um filtro de Kalman robusto e, como mencionado anteriormente, em Kim et al. (1998).

O material apresentado na Seção VI.10 se aplica aqui e, em particular, as equações de filtragem para este modelo são

$$X_{t+1}^t = \phi_0 + \phi_1 X_t^{t-1} + \sum_{j=0}^1 \pi_{tj} K_{ij} \epsilon_{ij},$$

$$P_{t+1}^t = \phi_1^2 P_t^{t-1} + \sigma_w^2 - \sum_{j=0}^1 \pi_{tj} K_{ij}^2 \Sigma_{tj}$$

$$\epsilon_{t0} = Y_t - \alpha - X_t^{t-1},$$

$$\Sigma_{t0} = P_t^{t-1} + \sigma_0^2,$$

$$K_{t0} = \phi_1 P_t^{t-1} / \Sigma_{t0},$$

$$\epsilon_{t1} = Y_t - \alpha - X_t^{t-1} - \mu_1,$$

$$\Sigma_{t1} = P_t^{t-1} + \sigma_1^2,$$

$$K_{t1} = \phi_1 P_t^{t-1} / \Sigma_{t1}.$$

Para completar a filtragem, devemos ser capazes de avaliar as probabilidades  $\pi_{t1} = P(I_t = 1 | Y_{1:t})$ , para  $t = 1, \dots, n$ ; é claro que,  $\pi_{t0} = 1 - \pi_{t1}$ . Denotemos por  $p_j(t | t - 1)$  a densidade condicional de  $Y_t$  dado o passado  $Y_{1:t-1}$  e  $I_t = j$  para  $j = 0, 1$ .

Então,

$$\pi_{t1} = \frac{\pi_1 p_1(t | t - 1)}{\pi_0 p_0(t | t - 1) + \pi_1 p_1(t | t - 1)},$$

onde assumimos que a distribuição  $\pi_j$ , para  $j = 0, 1$  foi especificada a priori.

Se o investigador não tiver razão para preferir um estado a outro, a escolha a priori uniformes  $\pi_1 = 1/2$ , será suficiente. Infelizmente, é computacionalmente difícil obter os valores exatos de  $p_j(t | t - 1)$ ; embora possamos dar uma expressão explícita de  $p_j(t | t - 1)$ , o cálculo real da densidade condicional é proibitivo.

A estimação dos parâmetros  $\Theta = (\phi_0, \phi_1, \sigma_0^2, \mu_1, \sigma_1^2, \sigma_w^2)^\top$  é realizada via MLE com base na verossimilhança dada por

$$\ln(L_Y(\Theta)) = \sum_{t=1}^n \ln \left( \sum_{j=0}^1 \pi_j p_j(t | t-1) \right),$$

onde a densidade  $p_j(t | t-1)$  é aproximada pela densidade normal  $N(X_t^{t-1} + \mu_j, \sigma_j^2)$ , mencionada anteriormente.

Podemos considerar a maximização da verossimilhança acima diretamente como uma função dos parâmetros usando o método de Newton ou podemos considerar a aplicação do algoritmo EM para a verossimilhança completa dos dados.

## Exemplo VI.24. Produto Interno Bruto (PIB) dos EUA.

Vamos agora considerar algumas abordagens Bayesianas para ajustar modelos de espaço de estado lineares gaussianos via métodos de Monte Carlo por Cadeias de Markov (MCMC).

Assumimos que o modelo é dado na forma da Definição VI.1 quando  $\Upsilon = \Gamma = 0$ . Nesse caso, Frühwirth-Schnatter (1994) e Carter and Kohn (1994) estabeleceram o procedimento MCMC que discutiremos aqui.

Um texto abrangente que recomendamos fortemente para este caso é Petris et al. (2009) e o pacote R correspondente **dilm**. Para modelos não lineares e não gaussianos, o leitor deve consultar Douc, Moulines and Stoffer (2014).

Como nas seções anteriores, temos  $n$  observações denotadas por  $Y_{1:n} = \{Y_1, \dots, Y_n\}$ , enquanto os estados são denotados como  $X_{0:n} = \{X_0, X_1, \dots, X_n\}$ , com  $X_0$  sendo o estado inicial.

Os métodos MCMC referem-se aos métodos de integração de Monte Carlo que usam um esquema de atualização Markoviano para obter amostras de distribuições a posteriori intratáveis.

O método MCMC mais comum é o amostrador de Gibbs, que é essencialmente uma modificação do algoritmo Metropolis desenvolvido por Hastings (1970) no cenário estatístico e por Geman and Geman (1984) no contexto de restauração de imagem.

Mais tarde, Tanner and Wong (1987) usaram as ideias em sua abordagem de amostragem por substituição e Gelfand and Smith (1990) desenvolveram o amostrador de Gibbs para uma ampla classe de modelos paramétricos.

A estratégia básica é usar distribuições condicionais para configurar uma Cadeia de Markov para obter amostras de uma distribuição conjunta. O seguinte caso simples demonstra essa ideia.

### Exemplo VI.25. Amostrador de Gibbs.

Suponha que desejamos obter amostras de uma distribuição normal bivariada,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

onde  $|\rho| < 1$ , mas somente podemos gerar amostras de uma normal univariada.

As distribuições condicionais são

$$(X | Y = y) \sim N(\rho y, 1 - \rho^2) \quad \text{e} \quad (Y | X = x) \sim N(\rho x, 1 - \rho^2),$$

e podemos simular a partir dessas distribuições. Construir uma Cadeia de Markov: Escolhendo  $X^{(0)} = x_0$  e então iterando o processo

$$X^{(0)} = x_0 \mapsto Y^{(0)} \mapsto X^{(1)} \mapsto Y^{(1)} \mapsto \dots X^{(k)} \mapsto Y^{(k)} \mapsto \dots,$$

onde

$$\begin{aligned} (Y^{(k)} | X^{(k)} = x_k) &\sim N(\rho x_k, 1 - \rho^2), \\ (X^{(k)} | Y^{(k-1)} = y_{k-1}) &\sim N(\rho y_{k-1}, 1 - \rho^2). \end{aligned}$$

A distribuição conjunta de  $(X^{(k)}, Y^{(k)})$  é

$$\begin{pmatrix} X^{(k)} \\ Y^{(k)} \end{pmatrix} \sim N \left( \begin{pmatrix} \rho^{2k} x_0 \\ \rho^{2k+1} x_0 \end{pmatrix}, \begin{pmatrix} 1 - \rho^{4k} & \rho(1 - \rho^{4k}) \\ \rho(1 - \rho^{4k}) & 1 - \rho^{4k+2} \end{pmatrix} \right).$$

Assim, para qualquer valor inicial  $x_0$ ,  $(X^{(k)}, Y^{(k)})$  converge em distribuição ao vetor  $(X, Y)$  quando  $k \rightarrow \infty$ ; a velocidade depende de  $\rho$ .

Então, alguém executaria a cadeia e jogaria fora os  $n_0$  valores iniciais não amostrados (burnin) e reteria o resto.

Para modelos de espaço de estados, o objetivo principal é obter a densidade a posteriori dos parâmetros

$$p(\Theta | Y_{1:n}) \quad \text{ou} \quad p(X_{0:n} | Y_{1:n})$$

se os estados forem significativos. Por exemplo, os estados não têm nenhum significado para um modelo ARMA, mas são importantes para um modelo de volatilidade estocástica.

Geralmente é mais fácil obter amostras da distribuição a posteriori completa  $p(\Theta, X_{0:n} | Y_{1:n})$  e então marginalizar para obter  $p(\Theta | Y_{1:n})$  ou  $p(X_{0:n} | Y_{1:n})$ . Como mencionado anteriormente, o método mais popular é executar um amostrador de Gibbs completo, alternando entre os parâmetros do modelo amostral e as sequências de estado latente de suas respectivas distribuições condicionais completas.

## Procedimento VI.1. Amostrador de Gibbs para modelos de espaço de estado

(i) Sortear  $\Theta' \sim p(\Theta | X_{0:n}, Y_{1:n})$

(ii) Sortear  $X'_{0:n} \sim p(X_{0:n} | \Theta', Y_{1:n})$

O Procedimento VI.1- (i) é geralmente muito mais fácil porque condiciona nos dados completos  $\{X_{0:n}, Y_{1:n}\}$ . O Procedimento VI.1- (ii) equivale à amostragem da distribuição conjunta de sua-  
vização da sequência de estado latente e é geralmente difícil.

Para modelos lineares Gaussianos, entretanto, ambas as partes do Procedimento VI.1 são relativamente fáceis de executar.

## Exemplo VI.26. Modelo de nível local.