4
Modelo STAR-Tree Heterocedástico

4.1.
Introdução

Um dos tópicos mais estudado em econometria financeira é a modelagem e previsão da variância condicional, ou volatilidade, de séries temporais de finanças, pois desempenham um importante papel na avaliação dos ativos financeiros. A volatilidade prevista é usada, por exemplo, em seleção de portfólio, controle de risco, apreçamento de derivativos, etc.

Alguns modelos não-lineares surgiram para modelar a variância da série ao invés do nível. Os mais populares, o ARCH (Autoregressive Conditional Heteroskedasticity), proposto por Engle (1982) e o GARCH (Generalized Autoregressive Conditional Heteroskedasticity), apresentado por Bollerslev (1986), estão, hoje, completamente incorporados nas práticas econométricas de finanças.

O modelo STAR-Tree, proposto por Rosa, Veiga e Medeiros (2008) é um modelo para a média de séries homocedásticas, mas para a aplicação em séries financeiras, essa hipótese não é real, uma vez que as séries de retorno não apresentam variância constante. Neste capítulo apresentaremos uma metodologia para a estimação de séries temporais heterocedásticas usando a modelagem STAR-Tree apresentada no Capítulo (3).

4.2.
Volatilidade Realizada

Com o rápido crescimento do mercado financeiro, e a dificuldade em se modelar e prever retornos diários de ativos financeiros, especialmente de ações, há uma crescente necessidade de conhecimento teórico e empírico sobre a volatilidade nas séries financeiras. Como a volatilidade dos retornos apresenta correlação, é mais fácil de prever do que a série de retornos em si. Por esses motivos, a econometria financeira, em particular a modelagem da volatilidade financeira, vem representando um importante papel nas teorias de apreçamento e controle de risco.

No entanto, o uso desses modelos para a volatilidade apresenta um problema inerente, a volatilidade é latente, e não diretamente observável. A volatilidade pode ser estimada por modelos de variância, como mencionado na seção anterior, mas a maioria dos modelos de volatilidade latente falha em descrever, de maneira satisfatória, diversos fatos estilizados observados nas séries de finanças.

Alguns fatos estilizados da volatilidade são bastante estudados, como auto-correlações fortemente persistentes, documentado inicialmente por Taylor (1986) para valores absolutos de retornos de ações, e assimetria na resposta da variância condicional das séries a choques (notícias inesperadas). Black (1976) observou que o mercado se torna mais volátil em resposta a choques negativos (más notícias) do que a choques positivos (boas notícias).

A busca por um arcabouço adequado para a estimação e previsão da variância condicional de retornos de ativos financeiros, levou a uma análise de dados intra-diários de alta freqüência. Merton (1980) notou que a variância de um intervalo fixo pode ser estimada arbitrariamente, porém de maneira acurada, como a soma dos quadrados das realizações disponíveis em uma base de dados de alta freqüência. A partir deste momento, as técnicas na literatura para medir a volatilidade de séries financeiras em dados intra-diários se multiplicaram. Com isso, a volatilidade de séries financeiras de realizações passadas se torna, essencialmente, “observável”.

Ao se tornar “observável”, a volatilidade pode ser modelada diretamente, e não mais ser tratada como uma variável latente. Baseados nestes resultados, diversos estudos recentes documentaram as propriedades de volatilidades realizadas construídas a partir de dados de alta frequência. A volatilidade realizada herdou todos os fatos estilizados que foram estabelecidos para a volatilidade nas especificações de variável latente. Recentemente, Andersen et al. (2001) estabeleceu o conceito de memória longa, ou dependência de longo prazo, como essencial para o estudo da variância de ativos financeiros.

Alguns problemas na estimação da volatilidade diária, como efeitos de microestrutura, são discutidos na literatura e soluções são propostas, mas não faz parte do escopo deste trabalho.

Supondo que, em um dado dia de transação *t*, os preços são observados em certa freqüência, suficientemente alta[2], onde n_t é o número total de observações no dia *t*, a variância realizada é definida como a soma de todos os retornos quadráticos intra-diários de alta freqüência disponíveis da seguinte forma:

$$RV_t^{(all)} = \sum_{i=0}^{n_t} r_{t,i}^2 \tag{4.1}$$

onde $r_{t,i}^2$ é o quadrado do retorno da observação *i* do dia *t*, $i = 1, \dots, n_t$.

A volatilidade realizada é a raiz quadrada da variância realizada.

4.3. Metodologia

Como foi falado no Capítulo (3), Seção (3.4), a estimação do modelo STAR-Tree leva em conta a hipótese de que o erro do modelo é homocedástico, apresenta variância constante.

Para modelarmos séries com heterocedasticidade, variância não constante, definimos o seguinte modelo:

$$y_t = H_{\mathbb{JT}}(\mathrm{x}_t; \psi) + \varepsilon_t = \sum_{i \in \mathbb{T}} \beta_i' \tilde{\mathrm{z}}_t B_{\mathbb{J}i}(\mathrm{x}_t; \theta_i) + \varepsilon_t \tag{4.2}$$

tal que,

$$\varepsilon_t = h_t^{1/2} \eta_t$$

onde $\eta_t \sim \mathrm{NID}(0,1)$ e $h_t^{1/2}$ é a volatilidade da série. Os parâmetros de $H_{\mathbb{JT}}$ foram definidos previamente no Capítulo (3). As funções de pertinência $B_{\mathbb{J}i}$ são definidas pelas equações (3.4) e (3.5).

Define-se $\mathfrak{F}_{t-1}$ como o conjunto das variáveis explicativas em *t*-1 e parâmetros do modelo, ou seja $\mathfrak{F}_{t-1} = (\mathrm{x}_t; \psi)$, onde x_t pode conter lags de y_t funções de lags de y_t, variáveis exógenas e tendências lineares. Deste modo, calcula-se o valor esperado e a variância condicionais da variável y_t como

[2] Andersen e Bollerslev (1998) sugerem a frequência de 5 minutos para uma boa medida, por exemplo.

$$\mathrm{E}\left[{y_t}/{\mathfrak{F}_{t-1}}\right] = H_{\mathbb{J}\mathbb{T}}(\mathfrak{F}_{t-1}) \tag{4.3}$$

$$\mathrm{VAR}\left[{y_t}/{\mathfrak{F}_{t-1}}\right] = \mathrm{E}\left[{h_t\eta_t^2}/{\mathfrak{F}_{t-1}}\right] = \mathrm{E}\left[{h_t}/{\mathfrak{F}_{t-1}}\right] \tag{4.4}$$

Observa-se que para o cálculo da variância condicional da série modelada é necessário calcular o valor esperado condicional de h_t. Como $h_t^{1/2}$ é a volatilidade diária de y_t, podemos tratar h_t como a variância realizada da série.

Para modelar e estimar a volatilidade realizada, Scharth, M. e Medeiros, M. (2006) propuseram um modelo STAR-Tree com a forma

$$\log(h_t^{1/2}) = H_{\mathbb{J}\mathbb{T}}(\mathrm{x}_t, \mathrm{w}_t; \psi) + \xi_t = \alpha'\mathrm{w}_t + \sum_{i\in\mathbb{T}} \beta_i'\tilde{\mathrm{z}}_t B_{\mathbb{J}i}(\mathrm{x}_t; \theta_i) + \xi_t \tag{4.5}$$

onde $\log(h_t^{1/2})$ é o logaritmo da volatilidade realizada diária, $\mathrm{w}_t \in \mathbb{R}^d$ é um vetor de regressores lineares, tal que $\mathrm{w}_t \not\subseteq \mathrm{x}_t$, e $\xi_t \sim \mathrm{NID}(0, \sigma_\xi^2)$. O resto dos parâmetros segue as definições da Seção (3.2) e as funções $B_{\mathbb{J}\mathrm{i}}$ são definidas pelas equações (3.4) e (3.5). A partir da equação (4.5), obtém-se

$$h_t^{1/2} = \exp\left(H_{\mathbb{J}\mathbb{T}}(\mathrm{x}_t, \mathrm{w}_t; \psi) + \xi_t\right) \tag{4.6}$$

Assim, temos a equação para a modelagem da variância realizada:

$$h_t = \exp\big(2H_{\mathbb{J}\mathbb{T}}(\mathrm{x}_t, \mathrm{w}_t; \psi) + 2\xi_t\big) = e^{2H_{\mathbb{J}\mathbb{T}}(\mathrm{x}_t, \mathrm{w}_t; \psi)} e^{2\xi_t} \tag{4.7}$$

Sabendo que

$$2\xi_t \sim \mathrm{NID}\big(0, 4\sigma_\xi^2\big),$$

e utilizando o valor esperado da distribuição log normal temos que,

$$\mathrm{E}\big[e^{2\xi_t}\big] = e^{2\sigma_\xi^2}, \tag{4.8}$$

logo,

$$\mathrm{E}\left[{h_t}/{\mathfrak{F}_{t-1}}\right] = e^{2H_{\mathbb{J}\mathbb{T}}(\mathfrak{F}_{t-1})} e^{2\sigma_\xi^2} = e^{2H_{\mathbb{J}\mathbb{T}}(\mathfrak{F}_{t-1}) + 2\sigma_\xi^2}, \tag{4.9}$$

Substituindo a equação (4.9) na equação (4.4), chegamos à expressão da variância condicional de y_t:

$$\text{VAR}\left[{}^{y_t}/_{\mathfrak{F}_{t-1}}\right] = e^{2H_{\mathbb{JT}}(\mathfrak{F}_{t-1})}e^{2\sigma_\xi^2} \tag{4.10}$$

Nota-se que a variância condicional de y_t não é constante, o que era o nosso objetivo inicial, para desta forma, propor uma metodologia para modelar séries com heterocedasticidade.

Com a disponibilidade de dados de volatilidade realizada diária em séries financeiras, utilizamos o artifício de dividir a equação (4.2) pela volatilidade realizada $h_t^{1/2}$ dos dois lados da igualdade, obtendo

$$\frac{y_t}{h_t^{1/2}} = \frac{H_{\mathbb{JT}}(\mathrm{x}_t;\psi)}{h_t^{1/2}} + \frac{\varepsilon_t}{h_t^{1/2}} = \sum_{i\in\mathbb{T}} \beta_i' \frac{\tilde{\mathrm{z}}_t}{h_t^{1/2}} B_{\mathbb{J}i}(\mathrm{x}_t;\theta_i) + \frac{\varepsilon_t}{h_t^{1/2}}, \tag{4.11}$$

Pode-se escrever a equação (4.11) da seguinte forma:

$$y_t{}^* = H_{\mathbb{JT}}{}^*(\mathrm{x}_t;\psi) + \eta_t = \sum_{i\in\mathbb{T}} \beta_i'\tilde{\mathrm{z}}_t{}^* B_{\mathbb{J}i}(\mathrm{x}_t;\theta_i) + \eta_t \tag{4.12}$$

Desta maneira, temos um modelo homocedástico para a série $y_t{}^*$, pois $\eta_t \sim \text{NID}(0,1)$, e para ajustá-lo, podemos aplicar a metodologia do STAR-Tree apresentada no Capítulo (3), que tinha como hipótese para os testes estatísticos e estimação dos parâmetros, a variância constante dos resíduos.

Ao estimar $y_t{}^*$, obtemos

$$\hat{y}_t{}^* = H_{\mathbb{JT}}{}^*\left(\mathrm{x}_t;\hat{\psi}\right) = \sum_{i\in\mathbb{T}} \hat{\beta}_i'\tilde{\mathrm{z}}_t{}^* B_{\mathbb{J}i}(\mathrm{x}_t;\hat{\theta}_i) \tag{4.13}$$

Se multiplicarmos os dois lados da equação (4.13) por $h_t^{1/2}$, temos

$$\hat{y}_t = H_{\mathbb{JT}}\left(\mathrm{x}_t;\hat{\psi}\right) = \sum_{i\in\mathbb{T}} \hat{\beta}_i'\tilde{\mathrm{z}}_t B_{\mathbb{J}i}(\mathrm{x}_t;\hat{\theta}_i)$$

Ou seja, depois de estimado o modelo da equação (4.13), basta aplicar a estimativa dos parâmetros, estimados utilizando o artifício descrito, nos dados originais, para obter a estimativa da série y_t original.

De maneira genérica, é possível definir um modelo STAR-Tree bi-variado para a média e a variância de uma variável dependente y_t, da seguinte forma:

$$\begin{pmatrix} y_t \\ h_t \end{pmatrix} = \begin{pmatrix} H_{\mathbb{JT}}\left(\mathrm{x}_{yt}, \mathrm{w}_{yt}; \psi_y\right) \\ H_{\mathbb{JT}}(\mathrm{x}_{ht}, \mathrm{w}_{ht}; \psi_h) \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \xi_t \end{pmatrix} \tag{4.14}$$