



## Processos de Hawkes: Uma Modelagem do Book de Oferta no Mercado Acionário Brasileiro

### Hawkes Processes: A Modeling of the Offer Book in the Brazilian Stock Market

YURI SAMPAIO MALUF \* AND CIRA E. GUEVARA OTINIANO \*\*

Received, Feb. 10, 2017

Accepted, May. 12, 2017

#### Resumen

Neste artigo é abordado o processo de Hawkes na modelagem do Book de oferta, em especial o fundo de índice ETF iShare Ibovespa. O estudo teve como objetivo investigar a dinâmica das influências das ofertas em relação às ordens passadas, além da interação das taxas de ordens dos lados opostos do Book. Outro ponto tratado é a verificação da estratégia de operação que capte a dinâmica estudada. No primeiro momento, foi realizada a abordagem univariada e no segundo a multivariada do processo de Hawkes. Os resultados mostram que em ambos os casos os dados se ajustaram bem ao processo de Hawkes. Os modelo ajustado indica que os agentes possuem comportamentos similares quando atuam como compradores ou como vendedores de ativos. Quanto a estratégia, não foi possível estabelecer ganhos sobressalentes a partir das mudanças nas taxas de ofertas. No entanto, estes resultados se devem mais ao bid-ask spread do que a capacidade preditiva das funções intensidades.

**Palabras clave.** Processos Pontuais, Processo de Hawkes, Book de Oferta, ETF.

#### Abstract

This paper discusses the Hawkes process to modeling of the offering book, especially the ETF iShare Ibovespa index fund. The study aimed to investigate the dynamics of the influences of the offers in relation to the past orders. We study too to the interaction of the order rates of the opposing sides of the Book. Another point addressed is the verification of the operating strategy that captures the dynamics studied. At the first moment, the univariate approach was performed and in the second moment the multivariate of the Hawkes process. The results show that in both cases the Hawkes process it fits well to the data. The fit model indicates that agents have similar behaviors when they act as buyers or sellers of assets. As for the strategy, it was not possible to establish spare gains from changes in bid rates. However, these results are due to the it bid-ask spread more that the predictive capacity of the functions.

**Keywords.** Point processes, Hawkes Process, Book of Offer, ETF.

**1. Introducción.** Um processo de Hawkes univariado é um caso particular de uma classe de processos pontuais, essa classe é denominada de processo linear de auto excitação.

Por Daley Vere-Jones (2002), um processo estocástico pontual na reta é caracterizado por uma sequência de variáveis aleatórias,  $\{t_i\}_{i=1}^K$ , em que  $0 < t_i < t_{i+1}$  representam os pontos ou eventos ocorridos em determinado instante de tempo. Os pontos podem representar as ocorrências de um fenômeno específico ao longo do tempo. Os processos pontuais são utilizados em várias áreas como: física quântica, engenharia elétrica, pesquisa operacional, engenharia de tráfego, finanças, etc. Em particular, em finanças utilizam-se na modelagem do Book de ofertas de ativos financeiros. O

\*Departamento de Estatística, Universidade de Brasília - UnB, Brazil.  
Corresponding author [yurimaluf@gmail.com](mailto:yurimaluf@gmail.com).

\*\*Departamento de Estatística, Universidade de Brasília - UnB, Brazil.  
Corresponding author [ciragotiniano@gmail.com](mailto:ciragotiniano@gmail.com).

conjunto de pontos servem para representar os tempos específicos em que as ordens de compradores e vendedores chegam ao Book, isto é o instante em que as ordens são efetivadas. Propriedades probabilísticas e aplicações desses processos são analisadas em Cox e Isham (1980), já o estudo das propriedades estatísticas com ênfase no processo de estimação da taxa de intensidade dos eventos é abordado em Diggle (2003).

A taxa ou função de intensidade,  $\lambda$ , que é uma medida para as ocorrências dos eventos por intervalos de tempo descreve a taxa de chegadas dos eventos adaptada a uma filtração  $\mathcal{F}_t$ , definida por,

$$\lambda(t|\mathcal{F}_t) = \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} \mathbb{E}[N_t^m - N_{\delta t}^m | \mathcal{F}_t],$$

sendo  $\{N_t^m\}$  o processo de contagem diretamente ligado ao processo pontual e  $\mathbb{E}$  o operador expectância.  $N_t^m$  representa o número de pontos do tipo  $m$  ocorridos até o tempo  $t$ .

Os processos de Poisson homogêneos são uma classe particular dos processos de contagem, eles possuem incrementos independentes e estacionários e sua função de intensidade é constante,  $\lambda = 1$ .

Outros processos pontuais,  $\{t_i\}_{i=1}^K$ , no qual a intensidade é estocástica e é regida pelos eventos passados são chamados de *processos lineares de auto-excitação* e possuem a seguinte função de intensidade,

$$\lambda(t) = \lambda_0(t) + \int_0^t \nu(t-s) dN_s,$$

em que  $\lambda_0(t)$  é determinística,  $\nu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  uma função núcleo real não negativa e  $\{N_s\}$  é um processo de contagem.

Um processo Hawkes  $\{t_i\}_{i=1}^K$  é um *processo linear de auto-excitação* cuja função núcleo é dada pela função,  $\nu(t) = \sum_{j=1}^P \alpha_j e^{-\beta_j t}$  e  $\lambda_0(t) = \lambda_0$ . O valor  $P$  remete a ordem do modelo, o caso mais simples do processo de Hawkes é quando  $P = 1$ . Assim, a configuração da função intensidade usada neste artigo é dada pela função de intensidade,

$$\begin{aligned} \lambda(t) &= \lambda_0 + \int_0^t \sum_{j=1}^P \alpha_j e^{-\beta_j(t-s)} dN_s \\ (1.1) \quad &= \lambda_0(t) + \sum_{j=1}^P \sum_{k=0}^{N_i-1} \alpha_j e^{-\beta_j(t-t_k)}. \end{aligned}$$

Na modelagem do book de oferta de ativos financeiros, a função  $\nu$  descreve a propagação da influência de uma oferta na taxa de chegada futura de ordens,  $\lambda_0$  também chamada de função base determinística, é interpretado como a taxa de chegada de oferta no longo prazo.

No processo de Hawkes a medida que o tempo passa a função intensidade das chegadas de ofertas vão sendo alteradas e sofrem aumentos instantâneos com a ocorrência de um evento e decaem exponencialmente até a chegada de um novo evento. O gráfico abaixo ilustra o comportamento da função intensidade quanto a ocorrência de inserções de oferta no book.

Outra função que desempenha um papel importante nos processos Hawkes é a função denominada de  $\mathcal{F}_t$ -compensador, definida por

$$\begin{aligned} \Lambda(s_i, s_{i+1}) &= \int_{s_i}^{s_{i+1}} \lambda(t|\mathcal{F}_t) dt, \quad \forall i \in \mathbb{N}^* \\ (1.2) \quad &= \int_{t_{i-1}}^{t_i} \lambda_0(t) dt + \int_{t_{i-1}}^{t_i} \sum_{j=1}^P \sum_{k=0}^{i-1} \alpha_j e^{-\beta_j(t-t_k)} dt. \end{aligned}$$

Na modelagem do book de ofertas, a versão multivariada do processo Hawkes é imprescindível para analisar a dinâmica do ponto de vista da interação entre as partes compradoras e vendedoras de ativos. Nesta circunstância, o comportamento da intensidade de um processo de chegada é descrito mediante a realizações pretéritas de mais de um processo. No caso de um processo de Hawkes

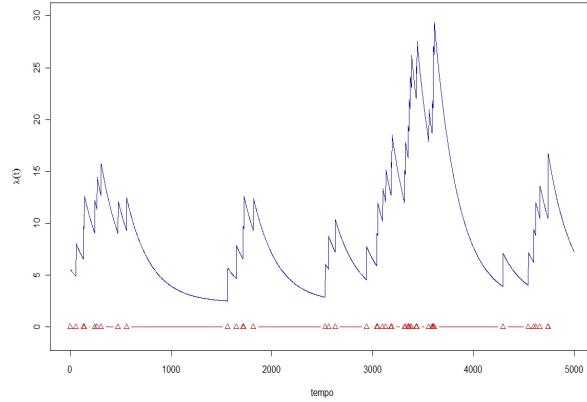


FIGURA 1.1. Função Intensidade com a chegada de ofertas

multivariado  $M$ -dimensional a função intensidade é dada por,

$$\begin{aligned}
 \lambda_m(t) &= \mu_m + \int_0^t \sum_{n=1}^M \sum_{j=1}^P \alpha_{j,m,n} e^{-\beta_{j,m,n}(t-s)} dN_s^n, \\
 (1.3) \quad &= \mu_m + \sum_{n=1}^M \sum_{j=1}^P \sum_{t_k < t} \alpha_{j,m,n} e^{-\beta_{j,m,n}(t-t_k^n)},
 \end{aligned}$$

onde  $m = 1, 2, \dots, M$ . Assim, como na abordagem univariada, neste artigo utilizamos a configuração  $P = 1$  para as aplicações.

O  $\mathcal{F}_t$ -compensador da  $m$ -ésima coordenada do processo é computado por,

$$\begin{aligned}
 \Lambda_m(t_{i-1}^m, t_i^m) &= \int_{t_{i-1}^m}^{t_i^m} \mu_m ds + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^M \sum_{j=1}^P \sum_{t_k^n < t_i^m} \alpha_{j,m,n} e^{-\beta_{j,m,n}(s-t_k^n)} ds \\
 &= \int_{t_{i-1}^m}^{t_i^m} \mu_m ds + \sum_{n=1}^M \sum_{t_k^n < t_{i-1}^m} \sum_{j=1}^P \frac{\alpha_{j,m,n}}{\beta_{j,m,n}} \left( e^{-\beta_{j,m,n}(t_{i-1}^m - t_k^n)} - e^{-\beta_{j,m,n}(t_i^m - t_k^n)} \right) \\
 (1.4) \quad &+ \sum_{n=1}^M \sum_{t_{i-1}^m < t_k^n < t_i^m} \sum_{j=1}^P \frac{\alpha_{j,m,n}}{\beta_{j,m,n}} \left( 1 - e^{-\beta_{j,m,n}(t_i^m - t_k^n)} \right),
 \end{aligned}$$

Os valores do  $F$ -compensador são importantes sobre tudo para realizar um diagnóstico do ajuste da série. Caso a realização do processo siga um processo de Hawkes os  $\Lambda_m(t_{i-1}^m, t_i^m)$  devem possuir uma distribuição exponencial ( $\lambda = 1$ ).

Na abordagem multivariada leva-se em conta a relação mútua entre compradores e vendedores. Desta maneira, é possível verificar o grau de influência das chegadas de ordens de compra passadas (venda) e a interação de ordens de venda (compra) sobre a taxa de chegada de ofertas de compra (venda) no book. Esta dinâmica pode ser expressa a partir da Equação (1.4), com  $M = 2$ , através das equações:

$$(1.5) \quad \lambda_C(t) = \mu_c + \int_0^t \alpha_{cc} e^{-\beta_{cc}(t-s)} dN_c(s) + \int_0^t \alpha_{cv} e^{-\beta_{cv}(t-s)} dN_v(s),$$

$$(1.6) \quad \lambda_V(t) = \mu_v + \int_0^t \alpha_{vc} e^{-\beta_{vc}(t-s)} dN_c(s) + \int_0^t \alpha_{vv} e^{-\beta_{vv}(t-s)} dN_v(s).$$

No sistema de equações (1.5) e (1.6) os índices  $c$  e  $v$  dizem respeito a ordem de compra e venda respectivamente. O par de constantes  $\mu_c$  e  $\mu_v$  representam a intensidade de chegada de ordens a longo prazo. Nesta estrutura, cada parâmetro possui uma interpretação econômica. O primeiro indexador, de cada parâmetro, indica o processo que sofre a ação e o segundo diz respeito ao processo que origina a influência.

Os parâmetros  $\alpha_{cc}$ ,  $\alpha_{cv}$ ,  $\alpha_{vc}$  e  $\alpha_{vv}$  mostram o impacto instantâneo na função intensidade de acordo com a natureza da ordem. Pelo modelo, a inserção de uma oferta provoca uma tendência maior de movimentação no book. Já os parâmetros  $\beta_{cc}$ ,  $\beta_{cv}$ ,  $\beta_{vc}$  e  $\beta_{vv}$  indicam como ocorre o decaimento da influência de chegada de oferta ao longo do tempo. Quanto maior o valor de  $\beta$  menor a propagação da influência dos eventos passados.

Um análise de interesse é verificar se  $\beta_{cv} = \beta_{vc}$ , isto é, se os impactos passados de ordens de compra (venda) afetam as ordens de venda (compra) da mesma forma. Da mesma forma outro ponto é examinar se  $\alpha_{cc} = \alpha_{vv}$  e  $\alpha_{vc} = \alpha_{cv}$ . Caso estes pontos se verifiquem, os agentes tanto compradores como vendedores agem de modo similar quanto aos impactos instantâneos com as chegadas de ordens ao book.

Neste artigo, Utilizamos um processo de Hawkes uni e bi-variado para a descrição do comportamento dos compradores e dos vendedores em um book de ofertas de ações do mercado brasileiro. Primeiro, na Seção 2, descrevemos a função de verossimilhança do modelo univariado e em seguida do modelo bivariado para uma abordagem mútua entre compradores e vendedores. Na seção 3, apresentamos a estratégia de operação para testar a capacidade preditiva dos modelos. Finalmente, na Seção 4, mostramos os resultados correspondentes aos dois modelos para uma série de dados do índice ETF iShare. Nesta seção, verificamos também o grau de influência das chegadas de ordens de compra passadas (venda) e a interação de ordens de venda (compra) sobre a taxa de chegada de ofertas de compra (venda) no book, via função de intensidade.

## 2. Estimação.

**2.1. Modelo univariado.** Do ponto de vista computacional, tanto a função intensidade como os *F-Compensators* exigem um esforço considerável. Para auxiliar a construção de algoritmos são utilizadas relações de recorrências.

Em um primeiro momento, é considerado a função intensidade determinística base como uma constante,  $\lambda_0(t) = \mu$  nos mesmos moldes de outros trabalhos na literatura. No entanto, é bem conhecido que a intensidade base, ao longo do pregão, sofre uma sazonalidade. Nas primeiras horas de negociação há uma maior intensidade nas taxas de chegadas de ofertas, reduzindo paulatinamente até o período do meio dia em que é conhecido como efeito almoço (*lunch effect*). As taxas voltam a aumentar apenas nas últimas horas do pregão. O comportamento ao longo do dia tem, portanto, o formato de *U* (*U shape*). Para lidar com este caráter sazonal é empregada uma interpolação por meio de um *spline* cúbico,  $\phi(t_i)$ , com 14 nós espaçados a cada 30 minutos. Cada ponto é calculado por

$$(2.1) \quad P_j = \frac{1}{N_{(zj)} - N_{(z(j-1))}} \sum_{i=N_{(z(j-1))}}^{N_{(z(j))}} \frac{1}{t_i - t_{i-1}}, \quad j = 1, 2, \dots, 14,$$

em que  $z$  é o número de segundos em 30 minutos. Por meio da interpolação temos uma aproximação da função intensidade base determinística,  $\phi(t) \approx \lambda_0(t)$ .

A função verossimilhança da classe de processos pontuais em que a função intensidade condicional é regida em termos das ocorrências passadas, pode ser expressa por,

$$(2.2) \quad \begin{aligned} L_T(N, t_1, t_2, \dots, t_N) &= e^{-\Lambda(0, T]} \prod_{i=1}^N \lambda(t_i | \mathcal{F}_t), \\ &= \exp \left( - \int_0^T \lambda(t | \mathcal{F}_t) dt + \int_0^T \log \lambda(t | \mathcal{F}_t) dN_t \right). \end{aligned}$$

Muitos são os processos pontuais incluindo os estacionário e não estacionários que possuem esta característica como o processo de Poisson, homogêneo e heterogêneo, renovação, processo de Wold e o de Hawkes. A partir da equação (2.2) é possível descrever a função log-verossimilhança do processo de Hawkes univariado por,

$$(2.3) \quad \begin{aligned} \mathcal{L}_T(\Theta | N_T, t_1, \dots, t_{N_T}) &= - \int_0^T \lambda_0(u) du - \sum_{i=1}^{N_T} \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j(t_{N_T} - t_i)}) \\ &\quad \sum_{i=1}^{N_T} \ln \left( \lambda_0(t) + \sum_{j=1}^P \sum_{k=1}^{i-1} \alpha_j e^{-\beta_j(t_i - t_k)} \right), \end{aligned}$$

O método de máxima verossimilhança busca otimizar a função verossimilhança com relação ao conjunto de parâmetros,  $\Theta = \{\mu, \alpha, \beta\}$ . Neste trabalho consideramos  $P = 1$ .

**2.2. Modelo multivariado.** A estrutura da função verossimilhança do processo de Hawkes multivariado é ditado também pela Equação (1.5), (1.5) e (2.3), o que o torna similar ao univariado. No caso multivariado é possível desmembrar a função log-verossimilhança em duas partes,

$$\mathcal{L}_T(\Theta_C, \Theta_V | N_T^C, N_T^V, \mathbf{t}^c, \mathbf{t}^v) = \mathcal{L}_T(\Theta_C | N_T^C, N_T^V, \mathbf{t}^c, \mathbf{t}^v) + \mathcal{L}_T(\Theta_V | N_T^C, N_T^V, \mathbf{t}^c, \mathbf{t}^v),$$

uma para cada conjunto de parâmetros  $\Theta_C = \{\mu_c, \alpha_{cc}, \beta_{cc}, \alpha_{cv}, \beta_{cv}\}$  e  $\Theta_V = \{\mu_v, \alpha_{vv}, \beta_{vv}, \alpha_{vc}, \beta_{vc}\}$ , onde os vetores  $\mathbf{t}^c, \mathbf{t}^v$  representam os tempos de chegadas das ofertas de compra e venda respectivamente. Os valores  $N_T^C, N_T^V$  é o total de ofertas de compra e venda respectivamente. Este artifício facilita o processo de otimização da função verossimilhança, pois o conjunto de parâmetros são distintos em cada parcela, o que permite otimizar cada uma separadamente. A função log-verossimilhança do processo de chegada de ofertas de compra é escrita por,

$$(2.4) \quad \mathcal{L}_T(\Theta_C | N_T^C, N_T^V, \mathbf{t}^c, \mathbf{t}^v) = -\mu_c T^c - \sum_{i=1}^{N_T^C} \frac{\alpha_{cc}}{\beta_{cc}} \left[ \left(1 - e^{-\beta_{cc}(t_i^c - t_{i-1}^c)}\right) R^{cc}(i-1) \right] \\ - \sum_{i=1}^{N_T^C} \frac{\alpha_{cv}}{\beta_{cv}} \left[ \left(1 - e^{-\beta_{cv}(t_i^c - t_{i-1}^c)}\right) R^{cv}(i-1) \right] \\ + \sum_{i=1}^{N_T^C} \frac{\alpha_{cv}}{\beta_{cv}} \left[ \sum_{t_{i-1}^c \leq t_k^v < t_i^c} \left(1 - e^{-\beta_{cv}(t_i^c - t_k^v)}\right) \right] \\ + \sum_{i=1}^{N_T^C} \ln(\mu_c + \alpha_{cc} R^{cc}(i) + \alpha_{cv} R^{cv}(i)).$$

O valor  $T$  refere-se ao tempo total de observação da série e a função  $\lambda_0(t) = \mu$  é representada por uma constante. A relação de recorrência é dada pela função  $R^{m,n}$  que é descrita por,

$$(2.5) \quad R^{m,n}(i) = \begin{cases} e^{-\beta_{m,n}(t_i^m - t_{i-1}^m)} R^{m,n}(i-1) + \sum_{t_i^m \leq t_k^n < t_i^m} e^{-\beta_{m,n}(t_i^m - t_k^n)} & \text{se } m \neq n, \\ e^{-\beta_{m,n}(t_i^m - t_{i-1}^m)} (1 + R^{m,n}(i-1)) & \text{se } m = n. \end{cases}$$

Este recurso permite economizar parte da complexidade computacional da função (2.4). A função log-verossimilhança para o processo de chegada de ordens de venda é análogo a função (2.4) sendo suficiente trocar o indexador de compra  $c$  pelo de venda  $v$  e vice-versa. Propriedades dos estimadores de processos pontuais podem ser encontradas em Ogata (1978).

**3. Estratégia de Operação.** O Book de oferta demonstra a dinâmica dos desbalanços entre a demanda e oferta do ativo no mercado. Consequentemente, o Book contém informações sobre as reações dos investidores com as mudanças dos preços. A partir das influências das taxas de chegadas de ordens é possível obter informações sobre as direções que guiam os preços a curto prazo ( ver Shek (2011) ). Portanto, para testar esta capacidade preditiva é aplicada uma estratégia ativa,  $\mathcal{E}$ , de compra e venda de cotas do fundo ETF *ishare* Ibovespa, que replica o principal índice do mercado brasileiro. Em seguida, comparamos com uma estratégia passiva de investimento, isto é, comprar e manter o ativo em carteira.

A estratégia ativa de operação depende da relação entre as intensidades de ofertas de compra e venda. A função estratégia,  $\mathcal{E}(t)$ , é expressa da seguinte forma,

$$(3.1) \quad \mathcal{E}(t) = \begin{cases} 1, & \text{se; } \pi(t-1) > \pi^b \\ 1, & \text{se; } \pi(t-1) > \pi^a \text{ e } \mathcal{E}(t-2) = 1 \\ 0, & \text{caso contrário.} \end{cases}$$

O valor 1 remete a posse do ativo e 0 o inverso. A função  $\pi(t) = \frac{\lambda_C(t)}{\lambda_V(t)}$  indica a relação entre as taxas de compra e venda. Caso ela cresça (decreça) significa que as chegadas de ofertas de compra no book estão maiores (menores) do que a de venda, o que consequentemente tende a gerar pressões

TABLA 4.1

Tabela 4.1. Estatísticas Descritivas das Durations - ETF BOVA11

Tipo	Média	Variância	Skewness
Compra	0,10	0,04	9,83
Venda	0,082	0,02	5,22

no sentido de aumento (diminuição) dos preços. Os parâmetros  $\pi^b$  e  $\pi^s$  são os limites da razão  $\pi(t)$ , em que o agente exerce a transação. O retorno de cada transação é calculado com base no melhor preço<sup>1</sup> no momento da transação  $t$ , ou seja, o retorno teórico e não os retornos passados com base nos preços já realizados.

Para estimar os valores ótimos de  $\pi^b$  e  $\pi^s$ , que maximizem o retorno total, será realizada uma simulação. O processo de simulação obedece os seguintes passos:

1. Estabelecer um valor aleatório para  $\pi^b$  e  $\pi^s$ .
2. Gerar os valores da função estratégia.
3. Extrair os valores dos retornos totais.

A partir da definição dos parâmetros,  $\pi^s$ ,  $\pi^b$  é possível auferir os ganhos da estratégia ativa e compará-lo com a estratégia passiva, conhecida como *Buy and Hold*. Entretanto, é necessário ajustar a estratégia ativa em relação ao risco. Isto é, os ganhos da estratégia ativa, acima da passiva, podem ser gerados meramente em razão do aumento de risco e não em função das mudanças de interesses dos agentes captados pela estratégia.

**4. Resultados.** Para investigar a taxa com que os agentes inserem as ordens e como ela se comporta durante o período de negociação, são captados os tempos em que houve alterações nas melhores ofertas no book. A contabilização destas chegadas segue-se tanto para as ofertas de compra como para as de venda. Para coletar a amostra foi desenvolvido um programa em  $C^+$ , que registrava a cada mudança no book o tempo em milissegundos. Os dados são fornecidos diretamente através da plataforma Enforque Ltda. O ativo em análise foi a cota do fundo de índice ETF *iShare Ibovespa* cotada sobre o código BOVA11. A amostra data de 10 de Outubro de 2013 com um total de 4,048 observações para as compra e 4,945 para as observações de venda.

As durations apresentam comportamentos similares. Isto pode ser visto por meio da tabela 4.1 de estatísticas descritivas de ambas séries.

Como a chegada de ofertas é descrita por um processo de contagem, preliminarmente é verificado a distribuição dos intervalos entre as ordens. Caso os dados seguem uma distribuição exponencial, tem-se consequentemente um caso de um processo de Poisson homogêneo. Para investigar se os dados são exponencialmente distribuídos é esboçado o QQ-Plot das *durations*.

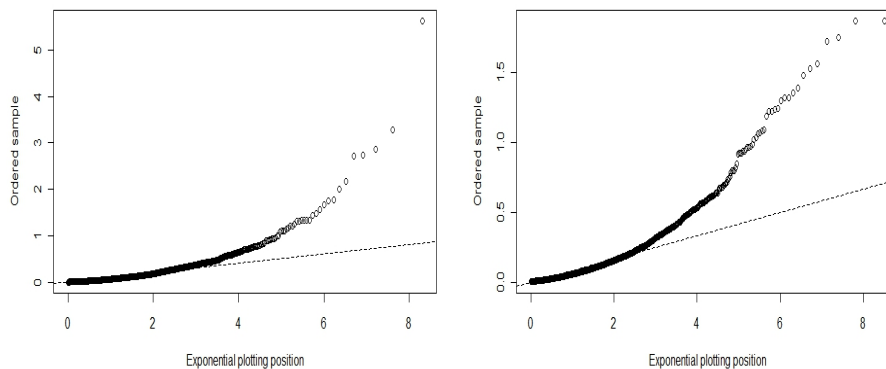


FIGURA 4.1. QQ-Plot Durations das Melhores Ofertas de Compra (esquerda) e Vendas do ETF (direita)

Pelo gráfico observar-se claramente que o processo gerador dos dados não sugere seguir uma distribuição exponencial. Uma das possíveis razões pode ser devido a estrutura de dependência temporal. A seguir é plotado a série das *durations*.

<sup>1</sup>No caso de sinal  $\mathcal{E}(t) = 1$  adquiri o ativo pelo melhor preço de venda em  $t$ , e no momento da venda  $\mathcal{E}(t) = 0$

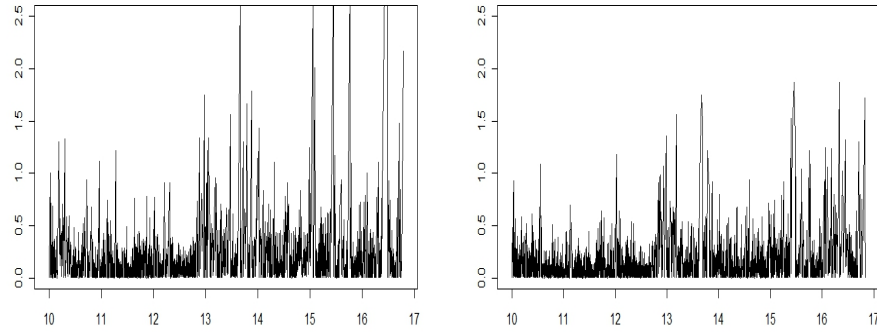


FIGURA 4.2. *Série das Durations das Ofertas de Compra (esquerda) e Vendas do ETF (direita)*

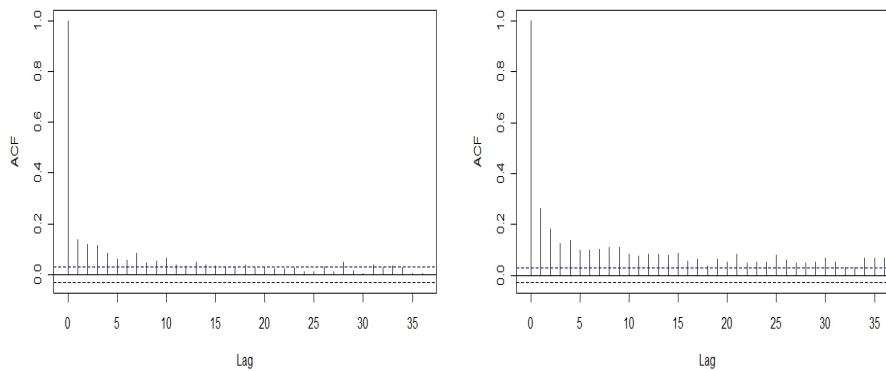


FIGURA 4.3. *Funções de Autocorrelações da Série das Durations das Ofertas (esquerda) de Compra e Vendas do ETF (direita)*

O gráfico 4.2 mostra que há acúmulos (*clusters*) de intervalos, o que sugere uma dependência entre eles. Outro ponto ressaltado é o aumento das *durations* a partir das 13h em ambos os gráficos.

As funções de autocorrelações acima corroboram com a ideia de dependência temporal entre os intervalos. O teste de Ljung-Box também atesta a dependência serial com p-valores da ordem de  $10^{-16}$ .

A partir das descrições dos dados, nota-se que de fato o processo gerador das chegadas de ofertas não segue um processo de Poisson homogêneo e portanto a taxa de chegada de ofertas não é constante. O passo seguinte é verificar, através do processo de Hawkes univariado, como os eventos passados influenciam na mudança da taxa de chegadas de ofertas.

**4.1. Resultados: Processo Hawkes Univariado.** Ao considerar o caso univariado tem-se apenas a influência de mesma natureza nas taxas futuras. A tabela 4.2 mostra as estimativas dos parâmetros com a intensidade base constante.

A intensidade foi calculada na escala de minutos. Os p-valores com (\*) significam que a ordem é menor que  $10^{-5}$ . Os erros padrões foram calculados por meio da matriz de informação de Fisher.

Os resultados mostram que a reação dos agentes mediante a inserção de ofertas são homogêneas. Tanto compradores como vendedores de ativos possuem todos os parâmetros similares. O parâmetro  $\alpha$ , que indica a auto excitação instantânea, em ambos os casos tem o mesmo aumento da taxa de oferta. Instantaneamente verifica-se que a taxa de oferta dobra em relação a taxa de longo prazo,  $\mu$ . O parâmetro  $\beta$  que mede o decaimento exponencial das influências passadas também foram similares para os dois casos. Embora os parâmetros  $\mu$  tenham se distanciado um do outro, a discrepância não foi estatisticamente significativa.

TABLA 4.2  
*Estimativa Função Intensidade - ETF BOVA11*

	Coefficiente	Erro Padrão	p-Valor
Compra			
$\hat{\mu}_c$	2,36	0,200	0,000*
$\hat{\alpha}_c$	3,18	0,210	0,000*
$\hat{\beta}_c$	4,18	0,318	0,000*
Venda			
$\hat{\mu}_v$	2,99	0,250	0,000*
$\hat{\alpha}_v$	3,20	0,232	0,000*
$\hat{\beta}_v$	4,26	0,353	0,000*

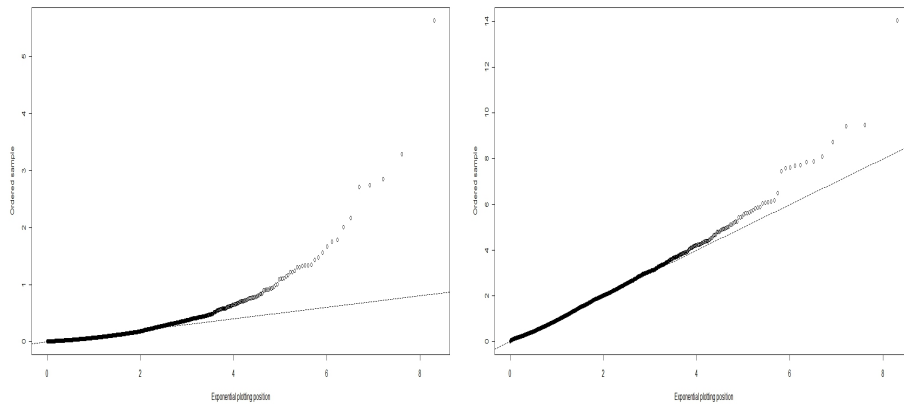


FIGURA 4.4. Gráfico QQ-Plot da Série das Durations das Ofertas (esquerda) de Compra do ETF (direita)

Caso a taxa de chegadas sejam governadas por um processo de Hawkes os *F-Compensators* devem seguir uma distribuição exponencial. Por meio do gráfico de QQ-Plot é possível ter uma ideia do ajuste.

Nos gráficos 4.4 e 4.5 são dispostos no lado esquerdo as durations das chegadas de ofertas e no lado direito os *F-Compensators*. Nos dois gráficos nota-se uma mudança significativa nos ajustes a uma distribuição exponencial.

De forma geral, a chegada de ofertas ao book pode ser bem descrito por um processo de Hawkes.

Como a intensidade de negociação sofre sazonalidade durante o dia, estimou-se também os parâmetros do processo de Hawkes incluindo este comportamento. Os gráficos 4.6 de  $\phi(t)$  que aproxima  $\lambda_0(t)$  sugerem a ocorrência do efeito almoço demonstrada pelo formato de  $\ddot{U}$ .

As curvas foram geradas por meio de um *spline* cúbico, em que os pontos foram calculados por 2.1 espaçados a cada meia hora. Mesmo levando em conta este comportamento, os parâmetros não tiveram grandes alterações. A estimação dos parâmetros são apresentados na tabela 4.3.

A partir dos parâmetros da tabela 4.2 tem-se o gráfico abaixo que mostra as taxas  $\lambda(t)$  tanto de compra como a de venda ao longo dos 30 primeiros minutos de negociação.

Fica evidente através do gráfico 4.7 que não somente a intensidade  $\lambda(t)$  tem uma dependência temporal bem como deve possuir uma relação entre os tipos de ordens. Isto pode ser visto na sincronia nos valores de  $\lambda_c(t)$  e  $\lambda_v(t)$  plotado no gráfico. Para investigar esta relação é empregado o processo de Hawkes multivariado.

**4.2. Resultados: Processo Hawkes Multivariado.** A abordagem multivariada do processo de Hawkes, no contexto da modelagem do book, será a bivariada. A estrutura de relação entre as taxas de chegadas de ofertas de compra e venda que investigamos, segue o sistema de Equações (1.5) e (1.6). Diferentemente do caso univariado, esta configuração permite visualizar a interação entre os tipos de ofertas explicitado pelo conjunto de parâmetros  $\alpha_{cv}, \beta_{cv}, \alpha_{vc}, \beta_{vc}$ . Os parâmetros da Equação (1.3) foram estimados por máxima verossimilhança por meio da função (2.4). Os resultados da estimação estão apresentados na tabela a seguir.



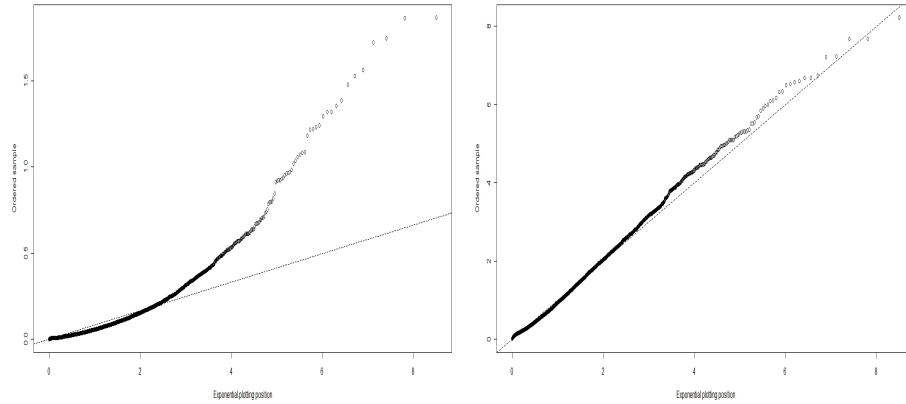


FIGURA 4.5. Gráfico QQ-Plot da Série das Durations das Ofertas (esquerda) de Venda do ETF (direita)

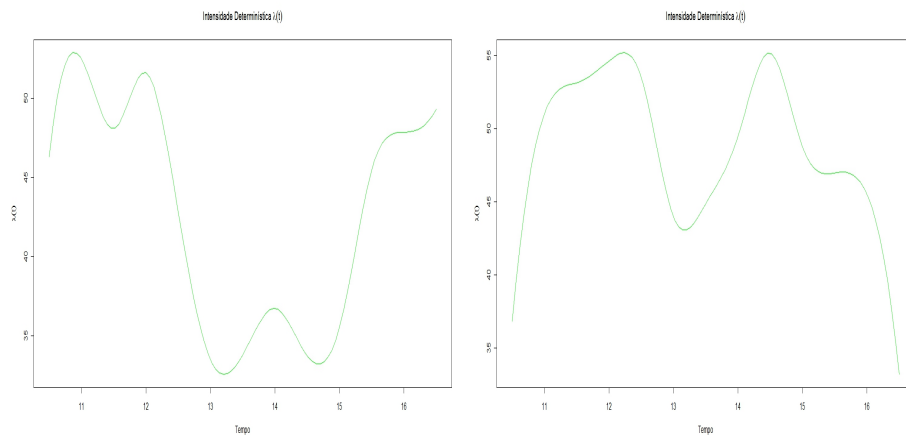


FIGURA 4.6. Gráfico da interpolação  $\phi(t)$  - ETF BOVA11,  $\lambda_0(t)$  de Compra (esquerda),  $\lambda_0(t)$  de Venda (direita)

Os valores com (\*) significa que a ordem é menor que  $10^{-5}$ . Todos os parâmetros foram estatisticamente significativos a 95% de confiança. Da mesma forma que no caso univariado, os parâmetros foram similares para ambos casos. Este fato reforça a tese que os agentes atuam do mesmo modo tanto na execução da venda quanto na compra, exposta no caso univariado.

Podemos perceber que as chegadas de ofertas induzem, de fato, a taxa de ofertas do outro lado do book indicada pelos parâmetros  $\alpha_{cv}, \beta_{cv}, \alpha_{vc}, \beta_{vc}$ . Outro ponto importante ressaltado pela tabela 4.4, é que embora o aumento instantâneo da intensidade seja menor quando ocorre a chegada de uma oferta de natureza inversa, sua influência perdura durante um maior tempo pois  $\beta_{cv} < \beta_{cc}$  e  $\beta_{vc} < \beta_{vv}$ .

O gráfico abaixo mostra a função intensidade nos primeiros 30 minutos de negociação.

Caso a realização seja governada por um processo de Hawkes bivariado os  $\Lambda(t_{i-1}, t_i)$  devem possuir uma distribuição exponencial.

A partir dos gráficos de QQ-Plot descritos a baixo podemos averiguar que houve um bom ajuste a uma distribuição exponencial em ambos os casos.

Para verificarmos a hipótese nula de  $\beta_{cv} = \beta_{vc}$ , isto é, se a propagação das influências são iguais tanto na venda quanto na compra utilizamos o teste de razão de verossimilhança.

$$\Psi = -2 \log \left( \frac{\sup \mathcal{L}_T(\Theta_0 | \mathbf{N}_T, \mathbf{t}^c, \mathbf{t}^v)}{\sup \mathcal{L}_T(\Theta | \mathbf{N}_T, \mathbf{t}^c, \mathbf{t}^v)} \right) \sim \chi^2_{10-8}$$

Como  $\Psi = 476,96 > \chi^2_{0,95;gl=2}$  logo temos evidência pra rejeitar a hipótese nula de que tanto compradores quanto vendedores de ativos possuem a mesma influência de eventos passados gerados pelas ações do outro lado do book.

Quando consideramos a sazonalidade da função intensidade base, dispostas no gráfico 4.6, os resultados da estimação do modelo se alteram.

TABLA 4.3  
 Estimativa Função Intensidade com Ajuste Sazonal- ETF BOVA11

	Coeficiente	Erro Padrão	p-Valor
Compra			
$\hat{\kappa}_c$	0,58	0,005	0,000*
$\hat{\alpha}_c$	3,08	0,219	0,000*
$\hat{\beta}_c$	4,12	0,33	0,000*
Venda			
$\hat{\kappa}_v$	0,07	0,005	0,000*
$\hat{\alpha}_v$	3,17	0,238	0,000*
$\hat{\beta}_v$	4,36	0,378	0,000*

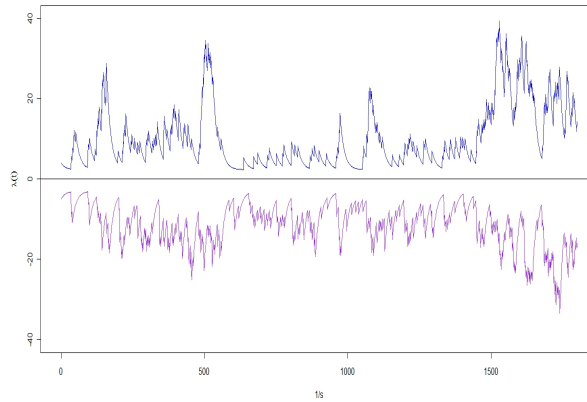


FIGURA 4.7. Intensidade  $\lambda(t)$  de Compra (Azul) e Venda (Lilás) do ETF BOVA11

Todos os parâmetros foram significativo a 95 % de confiança. Quando comparamos os resultados da tabela 4.4 e da tabela 4.5 podemos perceber que a estrutura geral de reação dos agentes aos eventos do book muda. Os compradores se mostraram menos reativos a eventos instantâneos do outro lado do book quando comparados aos vendedores, comportamento traduzido pela relação  $\alpha_{cv} = 0,6 < \alpha_{vc} = 3,71$ . Entretanto, o decaimento exponencial da propagação da influencia dos compradores é menor que os vendedores. Para termos uma ideia desta relação podemos usar a fórmula,

$$(4.1) \quad \tau_c = \frac{\ln\left(\frac{\alpha_{cv}}{\alpha_{vc}}\right)}{\beta_{vc} - \beta_{cv}} \times 60 \quad \tau_s = \frac{\ln\left(\frac{\alpha_{cc}}{\alpha_{vv}}\right)}{\beta_{cc} - \beta_{vv}} \times 60$$

que indica em quantos segundos a influência dos dois eventos se equivalem, isto é, em qual ponto no tempo as curvas do decaimento exponencial se igualam. No caso em que olhamos apenas para os eventos passados da mesma natureza, o  $\tau_s = 28,3$ , o que significa que os compradores reagem mais fortemente aos eventos apenas nos primeiros 28,3 segundos após este período as influências são maiores para os vendedores. O mesmo raciocínio é valido quando analisamos os eventos de natureza inversa. Como o  $\tau_c = 28,1$ , os vendedores reagem mais as ofertas de compra do que os compradores as de venda por aproximadamente 28,1 segundos.

A partir destes resultados podemos entender melhor como os investidores agem diante das chegadas de ofertas no book. Outro ponto é a investigação das similaridades de comportamento entre compradores e vendedores. Com o entendimento da reação dos agentes diante dos eventos de chegadas de ofertas podemos também verificar se a estratégia que leve em conta estas reações pode gerar ganhos superiores a média de mercado. Na próxima seção investigaremos a aplicação deste tipo de estratégia e qual é o seu desempenho diante dos riscos.

**4.3. Simulação: Estratégia de Operação.** O preço dos ativos são formulados mediante aos desbalanços entre as ofertas de compra e venda registrados no book de ofertas. Como as taxas de chegadas captam as mudanças com que os agentes ofertam e demandam ativos logo podemos

TABLA 4.4  
*Estimativa Função Intensidade - ETF BOVA11*

	Coeficiente	Erro Padrão	p-Valor
Compra			
$\hat{\mu}_c$	1,49	0,28	0,000*
$\hat{\alpha}_{cc}$	3,36	0,25	0,000*
$\hat{\beta}_{cc}$	5,02	0,46	0,000*
$\hat{\alpha}_{cv}$	0,14	0,06	0,0125
$\hat{\beta}_{cv}$	0,92	0,41	0,0121
Venda			
$\hat{\mu}_v$	2,52	0,44	0,000*
$\hat{\alpha}_{vv}$	3,60	0,32	0,000*
$\hat{\beta}_{vv}$	6,20	0,69	0,000*
$\hat{\alpha}_{vc}$	0,18	0,06	0,00137
$\hat{\beta}_{vc}$	0,60	0,20	0,00175

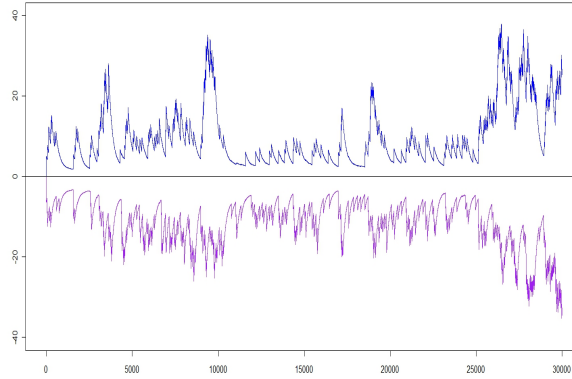


FIGURA 4.8. *Intensidade  $\lambda(t)$  de Compra (Azul) e Venda (Lilás) do ETF BOVA11*

utilizar destes movimentos para operar no mercado, como discutido na Seção 3. Após estimados os parâmetros que governam a reação dos agentes perante os eventos de chegadas destas ofertas estamos aptos a testar se a estratégia ativa de operação descrita pela função estratégia (3.1) produzem ganhos consistentes. Caso a estratégia ativa gere retornos anormais temos de verificar se superam os riscos das operações.

Todavia, ao realizar a simulação descrita na Seção 3, não foi possível encontrar os pontos ótimos de  $\pi^b$  e  $\pi^s$ . No caso do ativo BOVA11 todos os pontos gerados levaram a retornos totais menores que um. Em quase toda sua totalidade as operações geradas pela função (3.1) de compra e venda para cada par de pontos  $\pi^b$  e  $\pi^s$  levaram à perdas.

A primeira vista os resultados nos leva a crer que o uso das razões entre as taxas não possui qualquer poder preditivo do movimento dos preços. A função estratégia esta sujeita ao risco de liquidez expresso pelo *Bid-Ask Spread*. Desta forma, para gerar retorno sobressalentes não basta ter capacidade preditiva, mas a indicação de movimentação entre compradores e vendedores deve superar a diferença entre a melhor oferta de compra e venda.

Na tentativa de filtrar a capacidade preditiva das mudanças das taxas e o risco de liquidez traçamos um *grid*. No primeiro eixo temos o  $\pi^b$  na escala de desvio padrão de  $\lambda_C(t)$ . O eixo "tempo" representa o instante  $\delta$  após observar  $\lambda_C(t) > \pi^b$ . Caso os melhores preços subam 100% das vezes depois de  $\delta$  segundos após ter ultrapassado o ponto  $\pi^b$ , temos uma assertividade de 1.

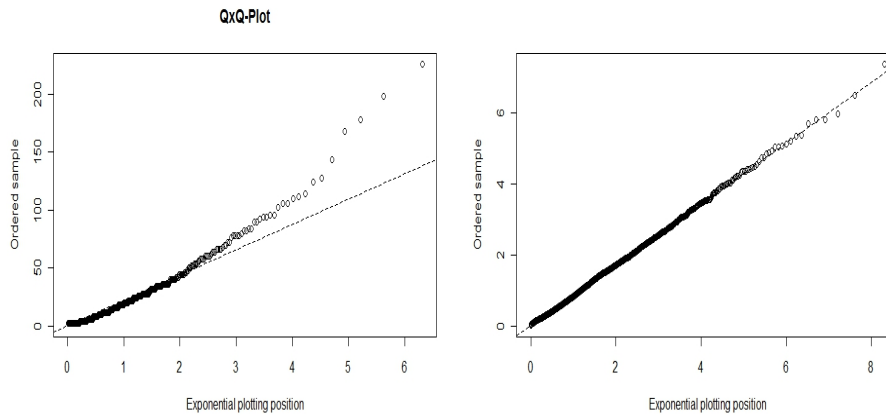


FIGURA 4.9. Gráfico QQ-Plot da Série das Durations das Ofertas de Compra do ETF, Durations (esquerda), F-Compensators (direita)

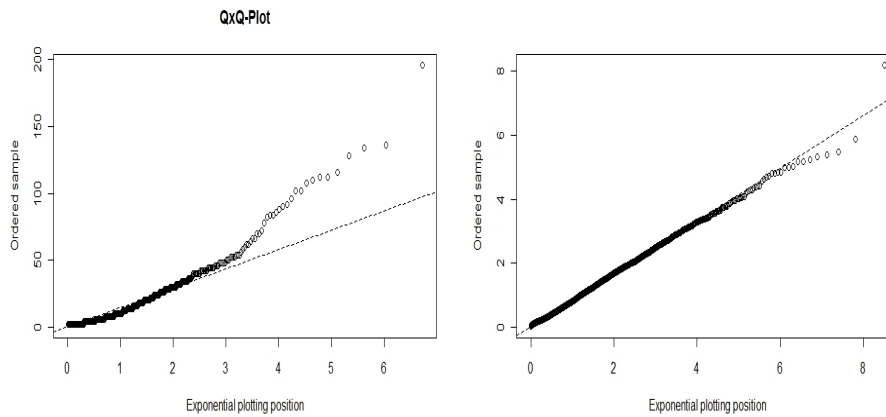


FIGURA 4.10. Gráfico QQ-Plot da Série das Durations das Ofertas de Venda do ETF, Durations (esquerda), F-Compensators (direita)

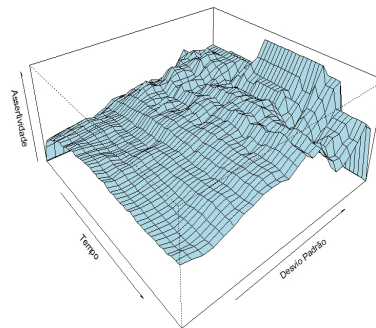


FIGURA 4.11. Gráfico do Nível de Assertividade

Por meio deste gráfico podemos perceber a partir de qual valor da taxa de compra e quanto tempo demora até o preço subir.

O resultado da Assertividade Máxima foi de 85% com  $\pi^b = 4,55$  e  $\delta = 156$ .

O melhor ponto ocorre quando a taxa de compra excede 4,55 desvios padrões e observamos a oferta de melhor compra após 156 segundos. Com estes dois parâmetros acertamos no aumento do

TABLA 4.5  
*Estimativa Função Intensidade - ETF BOVA11*

	Coeficiente	Erro Padrão	p-Valor
Compra			
$\hat{\kappa}_c$	0,04	0,004	0,000*
$\hat{\alpha}_{cc}$	2,93	0,24	0,000*
$\hat{\beta}_{cc}$	4,14	0,39	0,000*
$\hat{\alpha}_{cv}$	0,60	0,26	0,0112
$\hat{\beta}_{cv}$	7,7	3,52	0,0143
Venda			
$\hat{\kappa}_v$	0,04	0,009	0,000*
$\hat{\alpha}_{vv}$	0,74	0,097	0,000*
$\hat{\beta}_{vv}$	1,26	0,176	0,000*
$\hat{\alpha}_{vc}$	3,71	0,653	0,000*
$\hat{\beta}_{vc}$	11,56	1,95	0,000*

valor ofertado 85 % das vezes. Logo podemos perceber que as taxas de compra permite algum grau de previsibilidade porém o *bid-ask spread* impede a obtenção de ganhos no caso do ativo BOVA11. Possivelmente, com outros ativos com menores *spreads* seja possível gerar ganhos anormais.

**Agradecimentos.** Agradecemos ao Departamento de Pesquisa e Pós-graduação da Universidade de Brasília pelo financiamento parcial para este trabalho, bem como aos avaliadores deste trabalho pelas sugestões e correções.

#### Referencias

- [1] D. R. COX, AND V. ISHAN, *Point Process*, Chapman Hall/CRC, 1980.
- [2] D. J. DALEY AND D. VERE-JONES, *An Introduction to the Theory of Point Processes*, Vol 1, Springer, 2002.
- [3] Y. OGATA *The asymptotic behaviour of maximum likelihood estimators for stationary point process*, Annals of the Institute of Statistical Mathematics, 30 (1978), pp. 243–261.
- [4] H. SHEK, *Modeling high frequency market order dynamics using self-excited point process*, Working Paper, SSRN 1668160, 2011.