

# Um procedimento para prever recessões no Brasil a partir de indicadores antecedentes <sup>♦</sup>

## Liana Bohn

Doutoranda - Universidade Federal de Santa Catarina (UFSC)  
Endereço: Campus Universitário, Bairro Trindade - Florianópolis/SC - Brasil  
CEP: 88040-900 - E-mail: [liana.bohn@posgrad.ufsc.br](mailto:liana.bohn@posgrad.ufsc.br)

## Newton Paulo Bueno

Professor - Universidade Federal de Viçosa (UFV)  
Endereço: Av. Ph. Rolphs, s/n – Campus Universitário, Viçosa/MG - Brasil  
CEP: 36570-000 - E-mail: [npbueno@ufv.br](mailto:npbueno@ufv.br)

Recebido em 24 de junho de 2013. Aceito em 08 de agosto de 2014.

## Resumo

O objetivo deste artigo é testar, para o Brasil, uma nova abordagem de previsão de recessão que vem sendo proposta por autores oriundos de outros campos de pesquisa, como a física, e que pode ser útil para prever eventos extremos também na economia. Com a série de desemprego (de 1985 a 2012) suavizada a partir de *splines*, visa-se reconhecer, via análise fractal, indícios de uma mudança em sua tendência, a partir do desempenho passado. Após isso, utiliza-se a metodologia de análise discriminante para identificar as séries que apresentam comovimento com a série do desemprego. Concluiu-se que períodos de crescimento do desemprego são, em geral, precedidos em cerca de um ano por melhorias nas relações de troca, aumento das importações e queda dos salários mínimos reais em PPC.

## Palavras-Chave

Eventos extremos. Previsão. Análise fractal. Análise discriminante.

## Abstract

The goal of this paper is to test for Brazil a new recession forecasting approach that has been proposed by authors from other research fields, such as physics, which can be useful for predicting extreme events also in economy. From the unemployment series (1985-2012) smoothed by splines, we first identify through fractal analysis variables that present co-movement along the period of analysis. Next, we used discriminant analysis to identify leading indicators for the unemployment series. We concluded that unemployment periods are in general preceded by an one year lagged improvement in terms of trade, rises in imports, and by decreases in minimum real wages measured in terms of in PPP.

## Keywords

Extreme events. Forecast. Fractal analysis. Discriminant analysis.

## JEL Classification

E27.

<sup>♦</sup> Agradecimento ao CNPQ por Bolsa de Produtividade em Pesquisa.

## 1. Introdução

O objetivo deste artigo é testar para o Brasil uma abordagem para a previsão de crises econômicas que vem sendo proposta por autores oriundos de outros campos de pesquisa, como a física, que recentemente tem sugerido que técnicas próprias de suas áreas de origem podem ser úteis para prever eventos extremos também na economia. A hipótese de que seja possível construir algoritmos capazes de prever crises econômicas evidentemente vai contra a ortodoxia das finanças e o ponto de vista de muitos economistas influentes, que pressupõem a validade da hipótese de eficiência de mercados. Contudo, a evidência empírica mais recente sobre essa hipótese mesmo em sua versão fraca, que postula que os preços dos ativos refletem toda a informação contida nas séries históricas desses preços, é no máximo inconclusiva (Lim e Brooks, 2011). A grande maioria dos economistas, por outro lado, parece estar mais propensa a aceitar ser possível detectar sinais premonitórios de crises econômicas (Reinhart e Rogoff, 2009; Krugman, 2009), ou pelo menos identificar regularidades empíricas em suas manifestações (Caballero, 2010), as quais podem ser muito úteis para informar a política econômica.

O artigo adaptará em um *framework* comum um conjunto das técnicas utilizadas pelos autores que têm trabalhado com a previsão de eventos extremos em economia, avaliando em que medida os resultados encontrados para outros países podem ser reproduzidos para o Brasil. Especificamente, pretende-se desenvolver uma heurística para detectar períodos de forte aceleração das taxas de desemprego (FAD). As duas premissas básicas do trabalho são as de que: a) economias comportam-se de forma diferente em momentos que precedem tais períodos e b) esse comportamento anômalo apresenta um padrão sistematicamente recorrente, passível, portanto, de ser identificado e eventualmente codificado.

A justificativa óbvia para tentar aplicar essas técnicas nesse contexto é a notória dificuldade de tratar eventos extremos com os métodos paramétricos e não paramétricos tradicionais baseados na estimação da densidade total de probabilidades, que produzem tipicamente bom ajustamento nas regiões onde se encontra a maioria das observações, mas ajustamento insatisfatório nas caudas das distribuições de eventos (Hsieh e Fung, 2001). É por esse motivo que os autores mencionados no início desta introdução têm sido influenciados pelas

técnicas desenvolvidas pelos sismólogos, que se baseiam na formação prévia de padrões antes de grandes terremotos. Tais eventos são precedidos por aumentos simultâneos de atividade sísmica em locais adjacentes próximos a falhas geológicas, formando um padrão denominado de *Accord* (devido à simultaneidade desses movimentos), que permite a previsão quando a porcentagem de locais próximos com movimentação sísmica acima de certo patamar ultrapassa um limiar crítico (Zaliapin, Keilis-Borok e Axen, 2002).

Em parte devido à influência da sismologia, mas também por descobertas no campo dos sistemas complexos em geral (Miller e Page, 2007), tem-se desenvolvido uma nova abordagem de previsão de eventos extremos em sistemas sociais complexos, como crises econômicas que obviamente estão associadas à FAD. Essa abordagem visa a identificar a formação de padrões de comportamento regulares, sinalizados com antecedência pelo comportamento de *leading indicators*, isto é, séries que antecipam o comportamento de outras. Este grupo selecionado de indicadores, de acordo com Zarmowitz e Moore (1983), representa um conjunto de relações e apresenta um poder preditivo mais acentuado que quaisquer indicadores individuais, pois se constituem por diferentes processos econômicos que tem um tempo cíclico similar.

A estimação de tais modelos, seguindo o trabalho de Stock e Watson (1991, 1998), comumente é feita em um processo de duas etapas. Primeiramente, os picos e os vales (os máximos e mínimos locais, respectivamente) são determinados para as séries individuais e, posteriormente, comparam-se os padrões entre as séries que representam os *leading indicators* e a série-alvo de previsão. Nestes casos, se a defasagem que separa o movimento de uma variável em relação a outra for identificável e estável, é possível utilizar as variações em uma ou algumas delas como um indicador do que pode ocorrer com a outra (Carvalho e Hermann, 2003).

A lógica desse tipo de previsão pressupõe que possíveis padrões premonitórios e aumentos nos níveis de desemprego são eventos sequenciais na dinâmica do sistema. Esses padrões podem não desencadear os eventos extremos, mas revelar os sinais do crescimento da instabilidade, promovendo uma maturidade na compreensão da emergência de possíveis mudanças econômicas. Aplicações desse tipo já foram utilizadas nos Estados Unidos, França, Alemanha e Itália

para algumas variáveis socioeconômicas, como eleições, recessões, taxas de suicídio e, inclusive, para o próprio desemprego, revelando comportamentos surpreendentes (Keilis-Borok *et al.*, 2005, 2008; Kossobokov e Soloviev, 2008).

Este trabalho está dividido em quatro partes. Além da presente introdução, a seção 2 descreve detalhadamente a metodologia a ser empregada para prever períodos de FAD. A seção 3 apresenta e discute os resultados, enquanto que a seção conclusiva procura apresentar possíveis desdobramentos da abordagem proposta e responder a algumas prováveis objeções ao trabalho.

## 2. METODOLOGIA

### 2.1 Heurística

A heurística proposta neste estudo consiste de três etapas, destacadas na Figura 1.

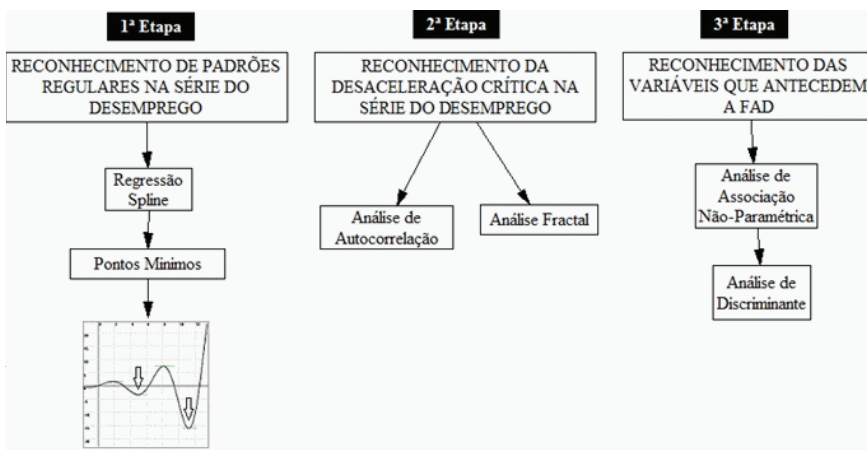


Figura 1 – Diagrama das etapas da metodologia utilizada

Fonte: Elaboração própria.

A primeira etapa busca o reconhecimento de padrões regulares no comportamento do desemprego, especialmente pela consideração de um fenômeno específico de sua dinâmica – o crescimento na taxa

de desemprego por um período suficientemente longo. Deste modo, identifica-se uma forte aceleração do desemprego (FAD) quando a série de desemprego atinge um ponto mínimo local iniciando uma trajetória ascendente. Os dados de desemprego serão filtrados pelo método de regressão *spline*,<sup>1</sup> de modo a retratarem mudanças no comportamento da série, evitando variações atípicas da mesma.

Após determinar as ocorrências de FAD, dá-se o processo de previsão propriamente dito. Ele pode ou estar centrado na análise da série do desemprego, com esta sendo capaz de sinalizar mudanças em sua tendência mediante a aproximação dos *tipping points* (pontos de mudança de regime), ou através da interconexão com outras variáveis. Para o caso da avaliação dos *tipping points* da série de desemprego, buscam-se os sintomas genéricos que ocorrem quando da aproximação do evento, observando-se se a dinâmica do sistema se torna mais lenta mediante o ajustamento a um ponto crítico, o que ocorre em muitos outros sistemas complexos na natureza ou na sociedade, e que é conhecido por ‘desaceleração crítica’ – mudanças características no espectro de frequência e variância do sistema (Dakos *et al.*, 2008; Scheffer *et al.*, 2009).

Quando se está distante desses pontos críticos, o sistema é resiliente a choques porque a taxa de recuperação é relativamente alta. Por outro lado, se um sistema está perto de um *tipping point*, um choque pode facilmente empurrar o sistema para um estado de risco, uma vez que a taxa de recuperação é lenta (baixa resiliência). Nesse caso, o sistema passará a apresentar uma maior “memória” a choques, a qual se reflete em um aumento na autocorrelação entre estados subsequentes do sistema (Scheffer, 2010).

Este sinal premonitório da mudança de regime pode ser percebido, por exemplo, mediante um aumento da autocorrelação de curto prazo nas variáveis-chave, medido através do coeficiente  $AR(1)^2$  (IVES, 1995). O trabalho de Dakos *et al.* (2008) encontrou resultados surpreendentes nesta área ao analisar mudanças climáticas abruptas, mas esse padrão de comportamento, que revelaria uma redução na amplitude de variação entre uma observação e outra, não é uma regra, pois algumas variáveis atingem sua desaceleração crítica sem

<sup>1</sup> O Apêndice 1 traz uma síntese do método. Para maiores detalhes, ver Marsh e Cormier (2001) e, para uma aplicação, Hausmann, Pritchett e Rodrik (2005).

<sup>2</sup> O  $AR(1)$  consiste em um processo autorregressivo que considera a relação de uma série com o seu passado imediato (uma defasagem).

que haja um aumento na dependência de seus valores ao longo do tempo. Em função disso, para avaliar se o sistema se encontra em uma fase de desaceleração crítica, o presente trabalho faz uso da análise fractal, computando o expoente de Hurst mediante o teste  $R/S$  para períodos que antecedem a FAD.<sup>3</sup> A utilização das leis de potência decorre do fato de elas representarem a dinâmica dos sistemas de elevada complexidade, cujas partes interagem de forma não linear. Nestes casos, as observações não obedecem a uma distribuição gaussiana e sua média não representa a escala característica do sistema, mas é possível traçar uma linha na escala log-log cuja inclinação indica as frequências relativas entre observações mais e menos frequentes.

Esta técnica consiste basicamente em comparar uma série de dados de interesse com uma série seguindo um *random walk*. Intuitivamente a ideia é a seguinte: calcula-se a média das observações para um período de interesse ( $T$ ) e computam-se as flutuações em torno desse valor, especificando-se uma janela entre o valor mínimo e o máximo dessa flutuação, denominado *range* da série. Se o período for aumentado e o *range* crescer de acordo com a raiz quadrada de  $T$ , a série será um ruído branco, isto é, não apresentará memória de longo prazo. Se o expoente de  $T$ , que é o expoente de Hurst, for maior que 0,5, por outro lado, a série apresentará memória de longo prazo, indicando que existe autocorrelação entre as observações (para uma explicação detalhada da técnica, ver Mandelbroth e Hudson, 2004). Um expoente de Hurst superior a 0,5, portanto, sinaliza um aumento da correlação entre as observações, de modo que as flutuações para cima ou para baixo na variável de interesse tendem a ser seguidas por flutuações adicionais na mesma direção. Este fenômeno foi originalmente detectado para as cheias do Rio Nilo que tendiam a ocorrer em anos subsequentes, visto que o  $H$  calculado pelo próprio Hurst para a dinâmica de flutuações dessas cheias foi de 0,7. Uma elevação de  $H$  para uma série como a de desemprego, portanto, sinaliza a ocorrência de mais desemprego no futuro, isto é, uma FAD. Recentemente detectou-se, por exemplo, que o  $H$  do volume de dólares negociados no mercado começou a subir acima do valor limite de 0,5 por volta de 2005, atingindo 0,6 imediatamente antes da crise de 2009 (Smith, 2010).

<sup>3</sup> O Apêndice 2 traz uma síntese metodológica para a aplicação da análise fractal.

A etapa seguinte do trabalho será selecionar as variáveis sinalizadoras de recessão iminente por meio de análise discriminante. A variável dependente será definida em termos de ocorrência, ou não, de uma FAD. Nesse caso, a ocorrência será considerada a partir do momento em que a regressão *spline* atinge seus pontos mínimos locais, o que revela uma manutenção do crescimento das taxas de desemprego ao longo do tempo, apesar das variações do seu comportamento no curto prazo. Isso implica que as flutuações que se dão em um pequeno espaço de tempo podem não ser tão importantes para definir a trajetória da variável em questão.

A classificação fornecida pela combinação de um comportamento comum entre as variáveis inclusas no modelo como sinalizadoras de FAD podem também promover classificações equivocadas das observações em um grupo inadequado, que denominaremos de falsos padrões premonitórios de acelerações do desemprego. A previsão será, assim, formulada a partir de uma sequência discreta de alarmes, cada um indicando uma janela de tempo e espaço onde a aceleração da taxa de desemprego deve ocorrer. A Figura 2 mostra que um alarme é considerado como correto se a FAD ocorre dentro do tempo e espaço esperado; caso contrário, o alarme é considerado como falso. A falha da previsão se dará quando houver a ocorrência de FAD sem que tenha ocorrido qualquer sinal premonitório (Keilis-Borok, Gabriellov e Soloviev, 2009).

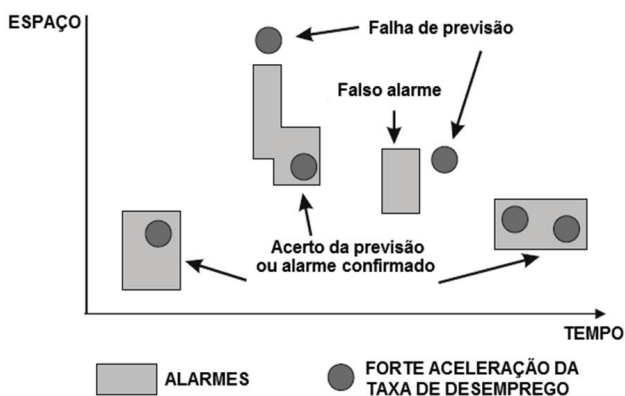


Figura 2 – Possíveis resultados na previsão da FAD

Fonte: Adaptado de Keilis-Borok, Gabriellov e Soloviev (2009, p. 4178).

## 2.2. A escolha das variáveis

Como não existe uma estatística oficial para a taxa de desemprego do Brasil, o presente estudo faz uso dos dados gerados pelo DIEESE. Entretanto, como a pesquisa não se dá em nível nacional, mas por regiões metropolitanas, adota-se como *proxy* da taxa de desemprego do Brasil a taxa de desemprego da Região Metropolitana de São Paulo, de janeiro de 1985 a janeiro de 2012, totalizando 325 observações.<sup>4</sup> Esta é ajustada sazonalmente.

Para possíveis *leading indicators*, segue-se o procedimento proposto por Keilis-Borok *et al.* (2005). Os indicadores a serem utilizados são: os dados de produção industrial brasileira, extraídos da Pesquisa Industrial Mensal do IBGE; o índice de produção industrial dos Estados Unidos – principal parceiro comercial do país e que apresenta influência sobre o comportamento das economias em todo o mundo – que é obtido do Fundo Monetário Internacional, na pesquisa de Estatísticas Financeiras Internacionais (IFS), a taxa Over/Selic (% a.m.), do Banco Central do Brasil, e as vendas de automóveis no mercado interno, dados publicados pela Anfavea.

As outras seis variáveis utilizadas no estudo de Keilis-Borok *et al.* (2005) ou não são encontradas para todo o período de análise, ou não apresentam correspondente direto nas pesquisas realizadas no Brasil. Elas também fazem referência a aspectos econômicos que influenciam direta e indiretamente o mercado de trabalho, incluindo nível de preços e indicadores da dinâmica econômica e das expectativas dos empresários e investidores. A fim de substituí-las, Keilis-Borok *et al.* (2000) sugerem que as séries selecionadas devem registrar aspectos relevantes do problema específico a ser estudado de acordo com a teoria econômica. Dada a dificuldade em se encontrar séries mensais que abranjam um longo período de tempo, optou-se por aquelas mais relacionadas com a teoria macroeconômica subjacente ao mercado de trabalho e aos indicadores da atividade econômica.

---

<sup>4</sup> Quando comparadas as estimativas anuais assumidas para a Região Metropolitana de São Paulo do DIEESE e para o Brasil, mediante dados da PNAD, observa-se um comovimento das séries – uma ampliação da taxa de desemprego em uma delas também representa um aumento na outra, dado o forte grau de associação entre as variáveis. Pelo cálculo do Coeficiente de Correlação de Pearson fica evidente a associação linear entre a taxa de desemprego do DIEESE e a refletida pela PNAD, com um valor de 0,8009.



As variáveis escolhidas para satisfazer esses requisitos foram a taxa de inflação, medida pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA) e o Ibovespa, como *proxy* para o estado de expectativas. Além destes foram testadas como possíveis variáveis explicativas os indicadores de comércio internacional – o volume de exportações e importações (FOB), medidos em dólares e obtido junto à Fundação Centro de Estudos do Comércio Exterior (Funcex), e a série de termos de troca,<sup>5</sup> calculada pela Funcex como índice com média 2006 = 100, que reúne as duas informações anteriores.

Muitos dos impactos sentidos pelo comércio internacional estão relacionados diretamente às taxas de câmbio, que afetam não somente o mercado de trabalho, mas o desempenho da economia como um todo. A utilização dessa série na avaliação das FADs é, entretanto, problemática, devido às grandes mudanças de moeda pelas quais passou o país, justificando sua não utilização em um estudo exploratório como o presente. A última variável considerada no modelo é o salário mínimo real em paridade do poder de compra (PPC), calculado em dólares pelo Instituto de Pesquisa Econômica Aplicada (Ipea).

Todas as séries que compreendem possíveis *leading indicators* são mensais e abrangem o período entre janeiro de 1985 a janeiro de 2012, totalizando 325 observações. Embora filtradas pelo método da regressão *spline*, elas não foram dessazonalizadas, visto que esse procedimento poderia interferir em eventuais padrões presentes na previsão aqui sugerida para o desemprego.

### 3. Resultados e discussões

#### 3.1. A evolução do desemprego no período

A principal crítica dos autores que buscam desenvolver procedimentos de previsão baseados na abordagem de detecção de eventos extremos é a pressuposição assumida pelos métodos tradicionais de que as variáveis econômicas distribuem-se em padrão gaussiano. Ao fazer essa suposição, deixam de levar em conta a existência de caudas pesadas (*fat tails*) nessas distribuições. A série de desemprego, por exemplo, não se distribui normalmente, o que é identificado

<sup>5</sup> Os termos de troca correspondem ao indicador que revela a razão entre o valor das exportações e das importações de um país, considerando, portanto, não somente a quantidade, mas seu preço.

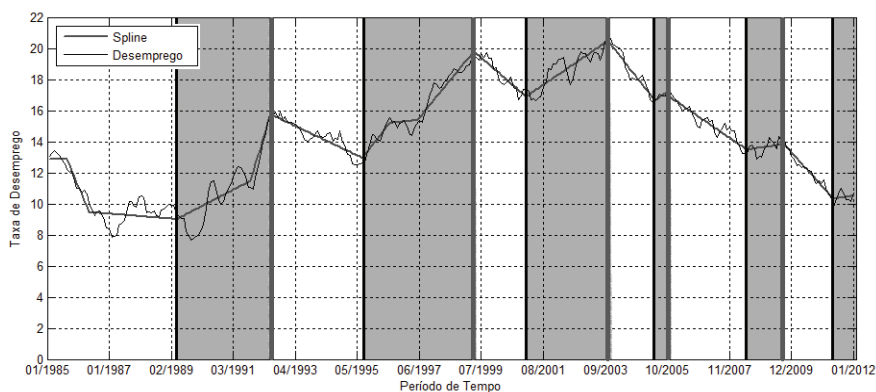
a partir do teste Jarque-Bera (JB), que leva a rejeitar a hipótese nula da normalidade da variável devido à presença de assimetria e curtose com p-valor de 0,0009.

Mediante a criação de 325 variáveis *dummies* (equivalente ao número de observações da série de desemprego) e utilizando uma *spline* linear, já que o objetivo principal é localizar os nós mínimos da amostra – que se mantêm os mesmos independentemente do grau de suavização assumido –, é possível fazer uso do método não paramétrico de *stepwise*.<sup>6</sup> A justificativa para isso decorre do fato de não se saber, *a priori*, quantas e quais variáveis serão escolhidas, desconhecendo-se o número de parâmetros requeridos pelo modelo.

A partir desse procedimento, 16 variáveis *dummies* foram selecionadas para suavizar a série da taxa de desemprego, gerando um modelo que explica 97,36% da variação total da variável dependente. As *dummies* selecionadas correspondem aos nós da regressão *spline*, representando mudanças no comportamento da série. Entretanto, esses nós podem ser pontos de máximo, mínimo ou apenas indicarem um estado crítico. Para saber a localização exata do início de um período de FAD, faz-se necessária a observação dos valores estimados da variável dependente.

De acordo com os dados, e a partir da evidência gráfica constante na Figura 3, é possível identificar seis pontos mínimos na *spline* da taxa de desemprego para o período considerado (linhas verticais pretas). Estes dão início à FAD, que se mantêm até que haja uma mudança no estado do desemprego (linhas verticais cinzas). A duração deste comportamento (correspondente às áreas cinzentas) varia conforme os casos, sendo possível observar uma redução em sua persistência nos últimos anos.

<sup>6</sup> A regressão *stepwise* envolve a inclusão de variáveis independentes, uma por vez, com base em seu poder explicativo. A abordagem começa escolhendo a melhor. Na segunda etapa, a melhor faz par com todas as demais variáveis independentes, uma de cada vez, escolhendo-se novamente a combinação mais explicativa. O terceiro e os demais parâmetros são selecionados de maneira semelhante, até se chegar ao modelo que melhor se adapta à variável dependente (adaptado de Hair *et al.* (2005)).



**Figura 3 – Os pontos críticos da taxa de desemprego evidenciados pela *spline* a ela associada**

Fonte: Resultados da pesquisa.

Para o período analisado, a primeira forte aceleração do desemprego inicia-se em abril de 1989 e se estende por 38 meses, até junho de 1992, sendo uma fase marcada por políticas agressivas de estabilização da inflação e pela adoção de reformas voltadas à abertura comercial e à privatização, enquanto o mercado de trabalho permaneceu em segundo plano. Após a implementação do Plano Real e de seus efeitos sobre as taxas de inflação, dá-se a segunda e mais longa FAD. Localizada no decorrer de todo o primeiro governo e início do segundo mandato do presidente Fernando Henrique Cardoso (de julho de 1995 a abril de 1999), os 45 meses que correspondem a este período presenciam um aumento significativo no número de desempregados. Este pico somente seria superado na terceira FAD, entre janeiro de 2001 a outubro de 2003, quando às causas estruturais somam-se as crises cambial e financeira internacional.

A partir do início do Governo Lula começa a haver uma acentuada tendência de redução das taxas de desemprego, de modo que as próximas duas ondas de FAD que ocorrem em seus mandatos são de curta duração. A primeira delas ocorre na metade de 2005 (entre maio e setembro), quando o desempenho do país começa a mostrar sinais de arrefecimento em consequência da elevação da taxa de juros e da valorização do Real; a segunda inicia-se em junho de 2008, pouco antes do auge da turbulência nos mercados financeiros, e se estende por 15 meses (até setembro de 2009). A última delas

começa em maio de 2011 e se estende até o fim da amostra, correspondendo aos últimos 8 meses de análise.

Estes movimentos aperiódicos revelam a dificuldade de prever a ocorrência de uma FAD, o que não significa que não seja possível antecipar que a variável esteja se aproximando de pontos críticos em sua trajetória. Pesquisas em outros campos indicam que, nesses casos, as variáveis apresentam um padrão denominado de desaceleração crítica, ou *slowing down*. A fim de avaliar tal fenômeno, o trabalho de Dakos *et al.* (2008) faz uso dos coeficientes de autocorrelação os quais, contudo, replicados à série de desemprego, não geram um padrão de comportamento que sinaliza a proximidade do ponto crítico, especialmente a partir do aumento na dependência de seus valores ao longo do tempo. Apesar da alta correlação da série em toda a amostra, inexistente uma recorrência no aumento de seus níveis quando da aproximação das FADs, de modo que apenas quatro eventos mostram uma pequena elevação dos valores assumidos para o AR(1), variando entre duas e sete observações antecedentes. Esta reduzida recorrência pode ser justificada mediante o prazo de relação estabelecida entre as observações, haja vista que a série pode apresentar uma memória de longo prazo que não seria identificada mediante os valores assumidos pelo AR(1). Neste caso, torna-se mais atraente avaliar o desemprego a partir da análise fractal.

Ainda que seja reduzido o número de observações da série de desemprego, dificultando a extração de informações amplamente significativas sobre a dinâmica da variável, as possíveis propriedades multifractais da série de desemprego são avaliadas a partir do método MFDFA (*Multifractal Detrended Fluctuation Analysis*) – ver Apêndice 2. Fazendo uso do resíduo da regressão Spline, dado que ela considera somente as variações que não sejam decorrentes da tendência e das modificações no estado da variável latente, obtém-se um expoente de Hurst de 0,9337 mediante o cômputo da relação  $R/S$ , o que corresponde a uma estrutura dependente de longo alcance. Mais do que o valor em si, o destaque se deve ao fato dele se encontrar na região bem acima de 0,50, podendo-se inferir que existe um comportamento persistente na variável, de modo que uma tendência positiva no passado tenda a continuar positiva, e vice-versa. Em outras palavras, o resultado encontrado para o valor do expoente revela que existe uma maior probabilidade de um aumento da taxa de desemprego ser seguido por outro evento similar. É esta estrutura de vínculo não linear da série que permite o estabelecimento de

previsões, ainda que a variável a ser prevista não pareça bem comportada em termos de estimativas econométricas.

Além destas relações, é possível ainda testar a ocorrência de flutuações locais da variável afim de avaliar sua magnitude – definindo-se, assim, a existência de multifractalidade. Neste caso, o Hurst local estimado para as séries temporais multifractais flutuará no tempo, em contraste com o expoente obtido pela análise monofractal. A Figura 4 indica que, como esperado, o expoente de Hurst tende a aumentar antes de períodos de FAD (indicados na ilustração pelas elipses que antecedem as linhas verticais pretas). Ou seja, antes de cada nó de máximo e mínimo percebido no comportamento de longo prazo da série de desemprego é possível, em geral, detectar um aumento do expoente de Hurst, indicando que as observações estão mais relacionadas e mais autocorrelacionadas nos períodos pré-FADs. Esse resultado é consistente com a hipótese de que o comportamento de longo prazo do desemprego não é inteiramente estocástico, estando sujeito à propriedade da desaceleração crítica.

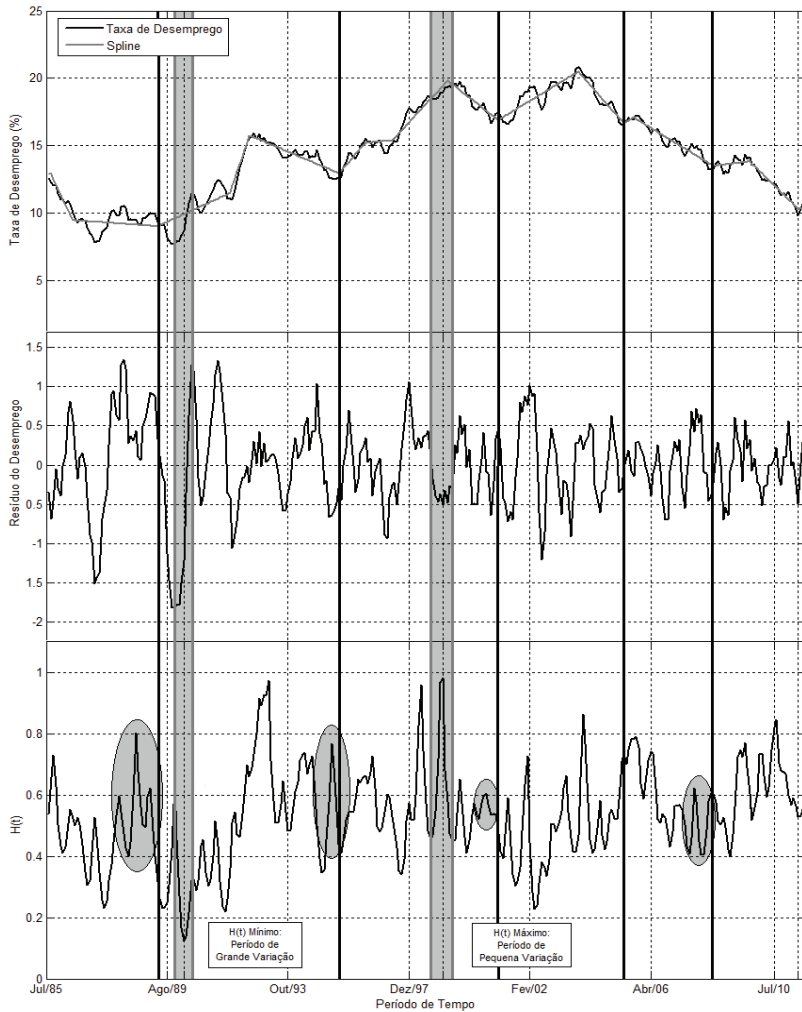


Figura 4 - O expoente local de Hurst e sua relação com os resíduos da regressão de desemprego e com os nós da regressão *spline* (Julho de 1985 a Julho de 2011)

Fonte: Resultados da Pesquisa.

### 3.2. Análise discriminante

A fim de identificar se as séries utilizadas como *leading indicators* apresentam de fato algum poder discriminante entre os períodos de FAD e os demais, faz-se uso do coeficiente de correlação de

Spearman.<sup>7</sup> Para isso, utilizam-se os valores das variáveis em nível e com até 12 defasagens, escolha esta que vai ao encontro dos trabalhos publicados com métodos que incluem os indicadores líderes para eventos extremos, mostrando a ocorrência de alarmes em um prazo inferior a um ano antes da manifestação desses eventos.

Das dez possíveis variáveis discriminantes testadas, somente cinco apresentaram nível significativo de correlação com a variável *dum-mie* criada para distinguir entre os momentos de crescimento (que assume valor zero) e de redução da taxa de desemprego (que assume valor 1). Este foi o caso das importações, da taxa de inflação, do número de automóveis vendidos, do salário mínimo real em paridade de poder de compra e dos termos de troca. Não foram significativos, portanto, os índices de produção industrial do Brasil e Estados Unidos, as exportações, o índice Ibovespa e a taxa Selic (ver Tabela 1).

**Tabela 1 – Resultado da correlação das variáveis com o desemprego**

Variável	Maior Nível de Correlação	Número de Defasagens
Importações	-0,137* (0,016)	11
IPCA	-0,179** (0,002)	12
Número de Automóveis Vendidos	-0,116* (0,041)	10
Salário Mínimo Real em PPC	0,170** (0,003)	11 e 12 <sup>1</sup>
Termos de Troca	-0,260** (0,000)	11 e 12 <sup>1</sup>
Índice de Produção Industrial do Brasil	0,045 (0,335)	0
Índice de Produção Industrial dos EUA	0,062 (0,181)	0
Exportações	-0,018 (0,699)	0
Índice Ibovespa	0,069 (0,138)	0
Taxa Selic	0,003 (0,955)	0

\*A correlação é significativa ao nível de 5%

\*\*A correlação é significativa ao nível de 1%

<sup>1</sup> Para 11 e 12 defasagens os valores de correlação e do nível de significância são iguais

Fonte: Resultados da Pesquisa.

<sup>7</sup> Opta-se pelo coeficiente de Spearman para detectar a existência de associações na população porque o coeficiente de correlação de Kendall é mais utilizado para a associação entre variáveis ordinais.

A variável 'importação' mostra uma relação negativa com a *dummie* de desemprego, de modo que, por ser estabelecida uma associação entre uma variável quantitativa e outra binária, o sinal negativo correlaciona-se ao valor zero para o desemprego, que compreende períodos de aumento de suas taxas. Assim também é para o número de automóveis vendidos e para o IPCA. Para os termos de troca, que também revelam uma relação inversa com a *dummie*, e o salário mínimo real (único parâmetro que varia no mesmo sentido da variável binária de desemprego), obtêm-se duas defasagens que podem ser utilizadas, já que apresentam exatamente o mesmo valor no coeficiente de correlação.

Diante disso, e contrariando a estimação simultânea em que todas as variáveis independentes são consideradas conjuntamente ignorando o poder discriminatório de cada uma delas, a análise discriminante utilizada no presente trabalho segue o método *stepwise*. Neste caso, busca-se o melhor conjunto de parâmetros que represente os grupos analisados, de modo a classificá-los corretamente.

Os principais pressupostos a serem verificados antes de prosseguir na análise discriminante consistem na normalidade das variáveis independentes, na homogeneidade das matrizes de covariância e na ausência de multicolinearidade. Esta última, caso exista, já é solucionada utilizando-se o método *stepwise*, mas as duas primeiras premissas devem ser avaliadas conforme testes específicos.

O primeiro pressuposto, com a finalidade de avaliar a normalidade multivariada, pode ser verificado a partir do teste Kolmogorov-Smirnov (KS) e de Shapiro-Wilk. Ambos apresentam a normalidade como hipótese nula, de modo que a sua rejeição implica a não normalidade das variáveis independentes. Considerando-se somente os parâmetros que apresentaram maior nível de correlação com a série de desemprego, é possível perceber que todos eles não se distribuem normalmente, o que pode prejudicar as inferências para a população a partir das observações amostrais. Para os casos em que as variáveis independentes não apresentam normalidade, é possível transformá-las (por exemplo, aplicando logaritmos) e testar novamente a partir da modificação realizada. Entretanto, não existe uma regra que guie o pesquisador na realização de tais transformações, sendo um método de tentativa e erro que pode alterar a interpretação das variáveis. Mais do que isso, a transformação dos dados não implica que haja



a homogeneidade nas matrizes de covariância, de modo que a não normalidade seja menos problemática do que se sugere (Guimarães e Moreira, 2008). Dadas essas características e o caráter exploratório do presente trabalho, optou-se então pela utilização das variáveis originais.

O segundo principal exame da estrutura discriminante avalia a quebra da premissa de igualdade entre as matrizes de covariância. Mediante o uso do teste Box's M, rejeita-se a hipótese nula da homoscedasticidade para os dois grupos. Hair *et al.* (2009), entretanto, destacam que, apesar da não normalidade das variáveis independentes e da heteroscedasticidade das matrizes de covariância, os resultados derivados da análise discriminante em relação a outras técnicas são válidos, especialmente porque eles dependem primordialmente da capacidade em classificar as observações em grupos determinados. Ou seja, mesmo que a quebra das premissas traga problemas à análise, ela será ainda mais robusta quando comparada aos demais métodos.

Com a conclusão de que as matrizes de variância e covariância dos dois grupos não são semelhantes, é possível perceber que este resultado coincide com o de muitos outros trabalhos que utilizam esta mesma metodologia. A dificuldade em encontrar um alto nível de significância decorre do fato de este teste ser extremamente sensível a flutuações e ao tamanho da amostra (Hair, 2005). Seguindo a mesma lógica utilizada em Ribeiro (2011), Guimarães e Moreira (2008), Selau (2008), Gouvêa, Farina e Varela (2007), entre outros, dá-se prosseguimento ao exame dos resultados já que a análise para fins de previsão não fica prejudicada.

O próximo passo do método seguido pela análise discriminante consiste no teste de igualdade das médias entre grupos. Espera-se, nestes casos, a rejeição da hipótese nula de que a variável apresente tal característica. Ao nível de significância de 1%, pode-se inferir que todos os parâmetros inclusos são possíveis discriminantes – o que é consistente com os resultados encontrados na análise de correlação feita anteriormente. O passo seguinte da análise foi dividir a amostra em duas partes: amostra de desenvolvimento, que corresponde ao período compreendido entre os quatro primeiros pontos de mínimo da regressão (de janeiro de 1985 até setembro de 2006) e que dá as bases para a previsão; e amostra de teste, que tentará identificar os

próximos dois episódios da FAD (de outubro de 2006 até janeiro de 2012). Neste caso, o método define alarmes que são identificados de forma retrospectiva, utilizando-se de informações disponíveis da série de desemprego.

A partir do método *stepwise* e mediante os valores assumidos pelo Wilk's Lambda<sup>8</sup>, que permitem avaliar a significância estatística do poder discriminatório dos parâmetros inclusos na análise, o modelo seleciona três variáveis explicativas: termos de troca (TT), salário mínimo em paridade de poder de compra (Sal.Mín.) e importações, todas com onze defasagens (ver Tabela 2). Isso significa que mais da metade das séries que apresentou um alto nível de correlação com a variável de desemprego foram estatisticamente significativas para classificar movimentos de aumento e redução na variável dependente, excluindo-se o IPCA e o número de automóveis vendidos, assim como o termo de troca e salário mínimo, ambos com 12 períodos de defasagens, que entraram na técnica por possuírem valores de correlação exatamente iguais as suas séries com uma defasagem a menos.

Tabela 2 – Variáveis incluídas pelo método *stepwise*

Passo	Variável*	Estatística Wilk's Lambda	Estatística F correspondente ao Wilk's Lambda	Significância
1	TT	0,892	29,994	0,000
2	Sal.Mín	0,842	23,108	0,000
3	Importações	0,826	17,250	0,000

\* Todas as variáveis escolhidas como discriminantes são defasadas em 11 períodos.

Fonte: Resultados da Pesquisa.

Deste modo, a análise discriminante indica a necessidade de três variáveis para antecipar os períodos em que ocorre ou não uma FAD, o que se dá exatamente 11 meses antes do evento. Para este período de defasagem, há maior manifestação da correlação das variáveis discriminantes escolhidas pelo modelo com a série de desemprego, o que se deve a um comovimento das mesmas, provavelmente provocado pela sazonalidade (que não é removida a fim de evitar a eliminação de tais padrões cíclicos). Cabe destacar, entretanto, que a série dos

<sup>8</sup> Também conhecida como estatística U, a estatística Wilk's Lambda consiste em uma distribuição de probabilidade utilizada em testes de hipótese multivariados, correspondendo a uma generalização da distribuição F univariada (Hair *et al.*, 2005).

termos de troca está amplamente correlacionada com a de importações, já que consiste na relação do intercâmbio de mercadorias e serviços em um determinado período de tempo. De acordo com Corrar, Paulo e Filho (2007), poder-se-ia considerar a retirada de uma das variáveis, mas isso possibilita gerar uma função com poder de explicação menor porque nenhuma outra variável foi aceita no processo.

Com o conhecimento das variáveis discriminadoras da FAD, torna-se possível avaliar o comportamento de tais indicadores líderes que antecedem a ocorrência do evento, buscando-se identificar a existência, ou não, de padrões regulares. A Figura 5 traz as variáveis consideradas sinalizadoras do aumento da tendência de desemprego, dando-se ênfase aos períodos sinalizados pela área cinza, que representam a duração de cada uma das FADs. As linhas pretas isoladas que antecedem essas regiões revelam o período de 11 meses de defasagem.

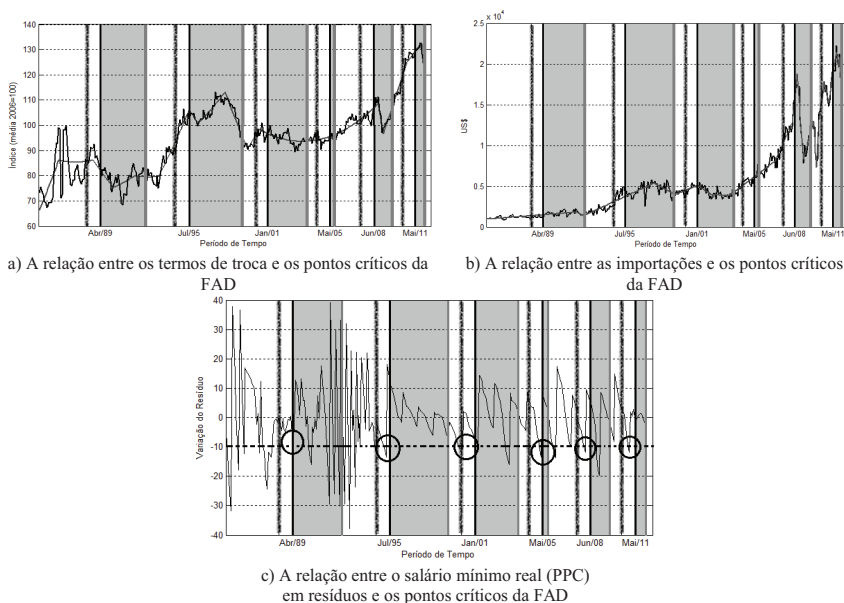


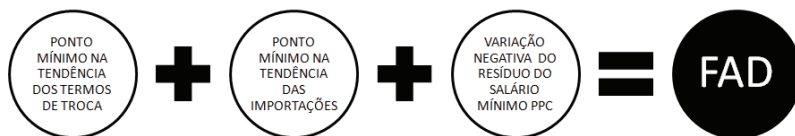
Figura 5 – A relação entre os *leading indicators* e a FAD.

Fonte: Resultados da Pesquisa.

Em relação aos termos de troca, o que se observa é que cada período de ampliação nas taxas de desemprego é, em geral, antecedido por seus pontos mínimos. Fato semelhante é encontrado para a série de importações, o que não surpreende dado que os termos de troca resultam da divisão entre o Índice de Preço das Exportações e o Índice de Preços das Importações, de modo que o nível de correlação entre os termos de troca e as importações é muito alto, o que significa que o aumento em uma das variáveis é acompanhado da ampliação no outro parâmetro. Portanto, uma melhora nos termos de intercâmbio deve significar uma expansão das importações e a melhor situação de ambos estará associada, com defasagem, a uma piora no mercado de trabalho.

Contrariando o comovimento das séries de termos de troca e importações em relação ao desemprego, a variável de salário mínimo real, medida em paridade de poder de compra, não apresenta movimentos tão marcantes de modo a ser possível, por exemplo, identificar as FADs conforme pontos críticos deste indicador líder. Desde a instauração do Plano Real existe uma tendência de crescimento da variável que se mantém até os dias de hoje, não apresentando flutuações deste comportamento ao longo do tempo. Nesses casos, a análise deve estar baseada na observação da série dos resíduos, que indica as diferenças entre o comportamento das variações de curto prazo em relação às de longo prazo (Keilis-Borok *et al.*, 2005). Com estes dados, antes das FADs fica evidente a presença de vales que atingem um limiar crítico de queda de dez dólares – PPC (indicados pela linha horizontal tracejada na Figura 5.c), sugerindo que as flutuações no curto prazo estão muito aquém da tendência de longo prazo.

Após a avaliação do comportamento das três variáveis selecionadas mediante a análise discriminante, é possível detectar um padrão antes da ocorrência de FADs. Este padrão reflete a ação conjunta de três fenômenos, como pode ser visto na Figura 6. FADs são precedidas por pontos mínimos nas tendências de longo prazo das séries dos termos de troca e da importação e por variações negativas e consideráveis (a partir de 10 dólares – PPC) na série dos resíduos do salário mínimo em termos de poder de compra em relação à tendência de longo prazo. Cabe destacar, entretanto, que a ocorrência destes fenômenos não se dá de modo simultâneo, mas no intervalo de até 11 meses, de modo que a FAD pode em princípio ser antecipada a partir dos eventos desencadeadores.



**Figura 6 – As variáveis sinalizadoras da FAD**

Fonte: Resultados da Pesquisa.

Em outras palavras, o início de um processo de melhora nas relações de troca após um ponto mínimo, o aumento das importações e a queda dos salários mínimos reais (PPC) precedem períodos de aceleração do desemprego. Essa combinação de comportamentos pode sinalizar que em períodos de aceleração anormal do crescimento econômico, em que as importações sobem de modo acentuado e a inflação acompanha a mesma tendência, corroendo o salário real, são seguidos por momentos em que o governo adota medidas para conter o crescimento, ampliando o desemprego. Essa é a explicação dada, por exemplo, por Keilis-Borok *et al.* (2005).

Em termos numéricos, trazidos pela Tabela 3, a análise discriminante que considera como previsoras da FAD as variáveis ‘termos de troca’, ‘importações’ e ‘salário mínimo em PPC’ (todas com 11 meses de defasagem) estima corretamente 76% das observações marcadas pela aceleração dos níveis de desemprego (consideradas como positivos verdadeiros) compreendendo os alarmes confirmados do modelo. Este resultado surpreende diante da baixa classificação correta na amostra de desenvolvimento (50,8%), o que pode ser decorrência da estrutura política e econômica vivida pelo Brasil entre meados das décadas de 1980 e 1990, quando as variações dos indicadores se davam mais em decorrência das incertezas geradas pela instabilidade pela qual passou o país do que propriamente por sua dinâmica econômica.

Tabela 3 - Poder preditivo da análise discriminante

	Classificação pelo método discriminante	Classe Verdadeira	
		FAD	NÃO-FAD
Amostra de Desenvolvimento	FAD	63 <sup>I</sup> (50,8%)	32 <sup>III</sup> (25,6%)
	NÃO-FAD	61 <sup>II</sup> (49,2%)	93 <sup>IV</sup> (74,4%)
	TOTAL	124	125
Amostra de Teste	FAD	19 <sup>I</sup> (76,0%)	2 <sup>III</sup> (5,1%)
	NÃO-FAD	6 <sup>II</sup> (24,0%)	37 <sup>IV</sup> (94,9%)
	TOTAL	25	39

I *Positivo verdadeiro* – Se é uma FAD e é classificada como FAD

II *Falso negativo* – Se é uma FAD e é classificada como NÃO FAD

III *Falso positivo* – Se é NÃO FAD e é classificada como FAD

IV *Negativo verdadeiro* – Se é NÃO FAD e é classificada como NÃO FAD

NOTA 1: Os números entre parênteses correspondem às taxas de classificação em relação ao total de observações da classe verdadeira de cada tipo de amostra.

NOTA 2: As colunas em destaque e os círculos retomam a ideia da Figura 2 já apresentada ao longo do texto para identificar falsos alarmes e falhas ou acertos de previsão.

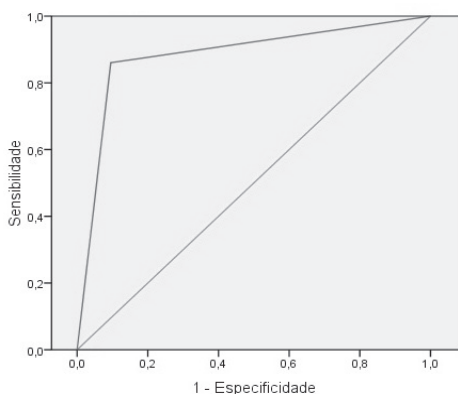
Fonte: Resultados da Pesquisa.

Ainda com relação ao poder preditivo da análise discriminante, a ocorrência de uma melhora no mercado de trabalho é corretamente avaliada em 94,9% das ocasiões (negativos verdadeiros), momento em que não há qualquer alarme de ocorrência da FAD. Entretanto, 5,1% das previsões são constituídas de falsos alarmes premonitórios (chamados de falsos positivos), indicando que haverá um forte aumento das taxas de desemprego quando, na realidade, isso não se confirma. Somado a isso, há falha na previsão da FAD em 24% das observações constituídas do evento porque nestas a aceleração da taxa de desemprego ocorre sem que as variáveis discriminantes tenham, conjuntamente, sinalizado essa possibilidade.

O modelo previsto, portanto, revela uma precisão de 90,48% dos movimentos que ocorrem na série analisada, o que faz referência à taxa de acerto do método ou, em outras palavras, ao número de FADs realmente observadas em relação àquelas que a análise discriminante considerou como tal. Em termos de acurácia, identificada pelo número de acertos de previsão dos períodos de FAD e não FAD em relação ao tamanho da amostra, o valor é de 87,5%. Neste sentido, a possibilidade de incorrer em erros não é insignificante,

especialmente porque em previsões econométricas se trabalha, em geral, com um limite máximo para a significância estatística de 10%. Entretanto, por se estar fazendo referência a um fenômeno como FADs que apresentam características de evento extremo, a previsão através da heurística proposta aparentemente apresenta um bom poder preditivo.

Tais resultados, especialmente apoiados na taxa de acerto da previsão (76%) e de falsos alarmes (5,1%), podem ainda ser avaliados mediante a análise ROC.<sup>9</sup> Entretanto, esse tipo de apreciação assume um formato discreto porque inexistente um valor de corte específico que gere a categorização entre os períodos onde há forte aceleração do desemprego, haja vista que a FAD é identificada a partir dos pontos de mínimo da série, não sendo possível associar probabilidades de acordo com diferentes limiares de classificação. Em termos de diagnóstico, os gráficos ROCs são bidimensionais e combinam os níveis de sensibilidade (a capacidade do modelo em detectar a FAD) e especificidade da previsão (a capacidade de determinar que o fenômeno não ocorre quando, de fato, ele está ausente). O modelo proposto possui 86% de sensibilidade e 90,5% de especificidade, colocando-o próximo ao eixo das abscissas (ver Figura 7). É possível, assim, inferir que o classificador é dito conservador, pois realiza a identificação da FAD apenas quando existem fortes evidências de que o evento realmente ocorreu, incorrendo em poucos erros.



**Figura 7 – Gráfico ROC para o modelo de previsão proposto**

Fonte: Resultados da Pesquisa.

<sup>9</sup> A análise ROC (*Receiver Operating Characteristics*) é uma técnica gráfica utilizada para organizar, visualizar e selecionar classificadores baseados em seu desempenho. Para maiores detalhes, ver Fawcett (2006).

A análise discriminante, portanto, gera uma relação entre sensibilidade e especificidade que se encontra bem acima da linha diagonal, o que indica ser este modelo melhor que quaisquer classificadores aleatórios ao revelar maior poder discriminante. Isso fica evidente quando se avalia a área determinada a partir do gráfico ROC que, de acordo com Braga (2000), quantifica a exatidão do método de previsão, revelando a qualidade do mesmo (sua acurácia). No caso da previsão de FADs, a área determinada tem um tamanho de 0,883, com desvio-padrão de 0,048, o que mostra um desempenho satisfatório.<sup>10</sup> As limitações associadas à realização de maiores inferências sobre a sensibilidade e qualidade do modelo advêm das características associadas à escolha do conceito de FAD, que não possui um valor de corte específico para a sua ocorrência.

Mais do que isso, apesar de terem sido escolhidas somente três variáveis para a classificação das variações do desemprego, e duas delas relacionadas ao comércio internacional, não é possível afirmar que os resultados encontrados para o poder preditivo do modelo não sejam robustos. De um modo geral, quando se trata de eventos extremos, o aumento no número de informações parece não levar a uma ampliação na exatidão das previsões, apenas a um aumento na confiança daqueles que as fazem (Taleb, 2008). Tais considerações, em consonância com a relativa estabilidade dos padrões regulares observados nos *leading indicators* e da falta de pretensão do trabalho em determinar o momento exato e a proporção da ocorrência do evento, permitem fazer inferências sobre o futuro, mas apenas na forma de sinalizar, conforme o comportamento das variáveis discriminantes, se existe a possibilidade de formação de um alarme premonitório de aceleração dos níveis de desemprego, bem como a expectativa de sua confirmação.

#### 4. Conclusão

Os testes realizados neste artigo oferecem evidência de que períodos de forte aceleração do desemprego no Brasil são precedidos por sinais gerados pelo comportamento de outras variáveis. Tais sinais seriam

<sup>10</sup> Considera-se que um teste tem bom desempenho quando a área sob a curva é superior a 0,70 (Braga, 2000). Os valores encontrados para a previsão da FAD foram calculados mediante o Software SPSS 17.



dados por melhorias das relações de troca, aumento das importações e queda dos salários mínimos reais (PPC) com defasagem de 11 meses. Várias explicações podem, em princípio, ser aventadas para esse fenômeno, por exemplo, a de que tais variáveis indicariam um aquecimento excessivo da economia, que levaria o governo a adotar medidas para conter o crescimento (Keilis-Borok *et al.*, 2005), ou a de que esses sinais indicariam uma perda de resiliência do sistema econômico ao conduzir diversos de seus componentes a estados críticos, tornando mais provável a eclosão de uma crise (Miller e Page, 2007).

O trabalho não oferece, contudo, um procedimento para identificar qual dessas explicações seria mais plausível. Isso foi proposital. Seu objetivo foi apenas destacar o fato de que talvez seja possível prever a eclosão de crises econômicas a partir de sinais prévios espelhados no comportamento de séries correlacionadas. Há a consciência das objeções usuais a esse tipo de procedimento, por exemplo, como já mencionado, a de que a hipótese de eficiência de mercados implicaria necessariamente a eliminação desses padrões, caso fossem realmente detectados. Ou a de que as técnicas adotadas carecem de microfundamentos, o que impediria sua utilização em um *framework* de otimização dinâmica como os modelos adotados atualmente na profissão, que são variações do modelo estocástico dinâmico de equilíbrio geral (DSGE, em inglês). À primeira objeção já se tentou responder no início do trabalho. Quanto à segunda possível crítica, menciona-se apenas o fato de que os modelos de previsão adequadamente microfundamentados usados pelos principais Bancos Centrais do mundo não têm logrado captar *turning points* na atividade econômica. O modelo de otimização dinâmica FRB/US (conhecido como *firbus*) utilizado pelo Fed para estimar o crescimento do PIB americano, por exemplo, falhou em antecipar todas as recessões mais importantes das últimas décadas (Buchanan, 2013). Em termos mais gerais, a comparação das previsões com suas realizações mostra que o FRB/US tende a sistematicamente superestimar o crescimento durante as recessões e a subestimá-lo nos períodos de recuperação, ou seja, tem se mostrado incapaz de prever pontos de ruptura nas séries de produção e emprego.

Para concluir, propõe-se que o tipo de análise realizada, devido a se apoiar essencialmente em algoritmos baseados em correlações entre séries macroeconômicas, pode em princípio vir a se beneficiar do crescente conjunto de técnicas inovadoras na análise de dados

denominado genericamente de *big data analysis* (ver, por exemplo, Mayer-Schonberger e Cukier, 2013) para identificar sinais premonitórios de ocorrência e, em alguns casos, sugerir políticas para mitigar os efeitos de crises econômicas de grandes proporções na economia brasileira.

## Referências

- BRAGA, A. C. S. *Curvas ROC: aspectos funcionais e aplicações*. 2000. 221 f. Tese (Doutorado em Engenharia de Produção e Sistemas) – Universidade do Minho, Braga – Portugal, 2000.
- BUCHANAN, M. *What physics, meteorology, and the natural sciences can teach us about economics*. Ed. Bloomsbury (Kindle Edition), 2013.
- CABALLERO, R. Macroeconomic after the crisis: time to deal with the pretense-of-knowledge syndrome. *Journal of Economic Perspectives*, v. 24, n.4, p. 85-102, 2010.
- CARVALHO, F. J. C.; HERMANNY, F. P. Ciclos e previsão cíclica: o debate teórico e um modelo de indicadores antecedentes para a economia brasileira. *Análise Econômica*, Porto Alegre, Ano 21, n. 39, Mar./2003.
- CORRAR, L. J.; PAULO, E. FILHO, J. M. D. *Análise multivariada para os cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2007.
- DAKOS, V.; SCHEFFER, M.; VAN NES, E.H.; BROVKIN, V.; PETOUKHOV, V.; HELD, H. Slowing down as an early warning signal for abrupt climate change. *PNAS*, v. 105, n. 38, p. 14308–14312, Set. 2008.
- FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, p. 861–874, 2006.
- GLERIA, I.; MATSUSHITA, R.; SILVA, S. Sistemas complexos, criticalidade e leis de potência. *Revista Brasileira de Ensino de Física*, v. 26, n.2, p. 99–108, 2004.
- GOUVÊA, M.A.; FARINA, M.C.; VARELA, P.S. Avaliação de indicadores socioeconômicos de municípios paulistas com o uso da análise discriminante. *Análise*, Porto Alegre, v. 18, n. 1, p. 78–97, Jan./Jun. 2007.
- GUIMARÃES, A.; MOREIRA, T.B.S. Previsão de Insolvência: um modelo baseado em índices contábeis com utilização da análise discriminante. *Revista de Economia Contemporânea*, Rio de Janeiro, v. 12, n. 1, p. 151–178, Jan./Abr. 2008.
- HAIR, J. F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. *Análise Multivariada de Dados*. 5. Ed. Porto Alegre: Bookman, 2005.
- HAIR, J. F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E. *Multivariate data analysis*. 7 Ed. Nova Jersey: Prentice Hall, 2009.
- HAUSMANN, R.; PRITCHETT, L.; RODRIK, D. Growth Accelerations. *Journal of Economic Growth*, v. 10, p. 303-329, 2005.
- HSIEH, D.; FUNG, W. The risk in hedge fund strategies: theory and evidence from trend followers. *Review of Financial Studies*, v. 14, p. 313-341, 2001.
- IVES, A. R. Measuring resilience in stochastic-systems. *Ecological Monograph*, v. 65, p. 217–233, 1995.
- KANTELHARDT, J.W.; ZSCHIEGNER, S.A.; KOSCIENLNY-BUNDE, E.; BUNDE, A.; HAVLIN, S.; STANLEY, H.E. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica*, v. 316, p.87-114, 2002.

- KEILIS-BOROK, V.; STOCK, J.H.; SOLOVIEV, A.; MICKALEV, P. Pre-recession pattern of six economic indicators in the USA. *Journal of Forecasting*, v. 19, p. 65–80, 2000.
- KEILIS-BOROK, V.I.; SOLOVIEV, A.; ALLÈGRE, C.B.; SOBOLEVSKII, A.N.; INTRILIGATOR, M.D. Patterns of macroeconomic indicators preceding the unemployment rise in Western Europe and the USA. *Pattern Recognition*, v. 38, p. 423–435, 2005.
- KEILIS-BOROK, V.; SOLOVIEV, A.; INTRILIGATOR, M.D.; WINBERG, F.E. Pattern of Macroeconomic Indicators Preceding the End of an American Economic Recession. *Pattern Recognition*, 2008.
- KEILIS-BOROK, V.; GABRIELOV, A.; SOLOVIEV, A. *Geo-complexity and Earthquake Prediction*. In: MEYERS, R.A. (Org.) *Encyclopedia of Complexity and Systems Science*. Nova York: Springer, 2009.
- KEILIS-BOROK, V.; SOLOVIEV, A.; LICHTMAN, A. Predictability of extreme events in socio-economic and political complex systems. In: Meyers, R. (ed.) *Complex systems in finance and econometrics – selected entries from the encyclopedia of complexity and systems science*. New York: Springer, 2011.
- KOSSOBOKOV, V.G.; SOLOVIEV, A. A. Prediction of extreme events: Fundamentals and prerequisites of verification. *Russian Journal of Earth Sciences*, v. 10, 2008.
- KRISTOUFEK, L. Local scaling properties and market turning points at Prague stock exchange. *Acta Physica Polonica B*, v. 41, n. 6, p. 1223–1236, 2010.
- KRUGMAN, P. How did economists get it so wrong? *The New York Times*, 02 de setembro de 2009.
- LIM, K.P.; BROOKS, R. The evolution of stock Market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys*, v. 25, p. 69-108, 2011.
- MANDELBROT, B.; HUDSON, R.L. *The (mis) behavior of markets: a fractal view of financial turbulence*. Nova York: Basic Books, 2004.
- MARSH, L.C.; CORMIER, D.R. *Spline Regression Models*. Londres: Sage Publications Inc, 2001.
- MAYER-SCHONBERGER, V.; CUKIER, K. *Big data: a revolution that will transform how we live, work and think*. Ed. John Murray (Kindle Edition), 2013.
- MILLER, J.; PAGE, S. (2007) *Complex adaptive systems, an introduction to computational models of social life*. Princeton e Oxford: Princeton University Press, 2007.
- RACINE, R. *Estimating the Hurst expoent*. 2011. 30f. Tese – Swiss Federal Institute of Technology Zurich, Zurich, 2011.
- REINHART, C.M.; ROGOFF, K.S. *This time is different*. Nova Jersey: Princeton University Press, 2009.
- RIBEIRO, S.M.S. *Modelos de previsão de incumprimento fiscal através de informação financeira: estudo das empresas de mobiliário de madeira*. 2011. 84 f. Dissertação (Mestrado em Finanças e Fiscalidade) – Universidade do Porto, Porto, 2011.
- SCHEFFER, M.; BASCOMPTE, J.; BROCK, W.A.; BROVKIN, V.; CARPENTER, S.R.; DAKOS, V.; HELD, H.; VAN NES, E.H.; RIETKERK, M.; SUGIHARA, G.. Early-warning signals for critical transitions. *Nature*, v. 461, n. 3, Set./2009.
- SCHEFFER, M. Foreseeing tipping points. *Nature – News & Views Research*, v. 467, 411–412, Set./2010.
- SCHUMAKER, L. *Spline Functions: Basic Theory*. Nova York: Wiley, 1981.
- SELAU, L.P.R. *Construção de modelos de previsão de risco de crédito*. 2008. 129 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.
- SMITH, R. *Is High-Frequency Trading Inducing Changes in Market Microstructure and Dynamics?* Disponível em <<http://ssrn.com/abstract=1632077>> or <<http://dx.doi.org/10.2139/ssrn.1632077>>.2010.
- STOCK, J.H.; WATSON, M. W. Business Cycles, Indicators and Forecasting. In: STOCK, J.H.; WATSON, M. W. (Ed.) *A Procedure for Predicting Recessions with Leading Indicators: Econometric*

- Issues and Recent Experience*. Chicago: The National Bureau of Economic Research, 1991. Disponível em <<http://www.nber.org/chapters/c7190.pdf>>.
- STOCK, J.H.; WATSON, M. W. Business Cycle Fluctuations in U.S. Macroeconomic Time Series. *NBER Working Paper Series*, n. 6528, Abr. 1998.
- TALEB, N.N. *A lógica do cisne negro – O impacto do altamente improvável*. São Paulo, Best Seller, 2008.
- ZALIAPIN, I.; KEILIS-BOROK, V.; AXEN, G. Premonitory spreading of seismicity over the fault network in S. California: precursors Accord. *Journal of Geophys*, v. 107, n. 2221, 2002.
- ZARMOWITZ, V.; MOORE, G. H. Sequential Signals of Recession and Recovery. In: MOORE, G. H. *The Business Cycles, Inflation, and Forecasting*. Chicago: The National Bureau of Economic Research, 1983. Disponível em <[www.nber.org/books/moor83-1](http://www.nber.org/books/moor83-1)>.

## APÊNDICE – Técnicas menos usuais empregadas no trabalho

### 1. O modelo de regressão *spline*

A regressão *spline* corresponde, conforme Marsh e Cormier (2001), a um modelo de variáveis *dummies* com algumas restrições de continuidade. Ao contrário de revelar a existência de quebras estruturais, ela mostra mudanças no comportamento da série (ou uma suavização da mesma), mas evita saltos inapropriados entre duas regressões, de modo a constituir, nestes pontos de inversão, os chamados nós da *spline*. Algebricamente, uma função  $s(x)$  é chamada de *spline* de grau  $m$  associada a um intervalo  $[a, b]$  se  $s(x)$  é um polinômio de grau  $m$  em cada subintervalo  $(x_{i-1}, x_i)$ ,  $i = 1, 2, \dots, m$  e se  $s(x)$  tem  $m - 1$  derivadas contínuas em cada  $x_i$  e, portanto, em  $[a, b]$  (Schumaker, 1981).

A situação mais simples para a estimação de uma *spline* é quando se conhece a localização exata dos nós e, portanto, quantos eles são. Este, entretanto, não é o caso ao se trabalhar com variáveis econômicas. Nelas, é necessário definir, além da quantidade e localização, o grau da regressão com a qual se trabalha (linear, quadrática, cúbica, etc.). Para isso, duas devem ser as formas consideradas: “quebrar” a série em um grande número de intervalos iguais (promovendo uma estimação por tentativa e erro) ou fazer uso dos valores observados da série tentando encontrar possíveis pontos de inversão.

Mediante este último, de acordo com Marsh e Cormier (2001), se existem  $n$  observações para a série analisada, é necessário criar  $n$  variáveis *dummies* onde:

$$D_j = \begin{cases} 1 & \text{se } T > K_j \\ 0 & \text{se } K_j = j \text{ para } j = 1, 2, \dots, n' \end{cases} \quad (1)$$

em que  $T$  corresponde às observações da série de tendência e  $K_j$  à localização temporal da observação analisada.

A partir dessas *dummies*, é possível desenvolver as variáveis ( $C_j$ ) que farão referência à possibilidade de nós na série. Serão estas que se incorporarão à regressão *spline*, sendo expressas por:

$$C_j = D_j (T - K_j)^m \quad (2)$$

onde  $m$  representa o grau a ser associado a ela. Sua escolha é deixada sob responsabilidade do pesquisador que o fará conforme o objeto de análise.

Nesse caso, por exemplo, para  $m = 1$  e  $K_{100} = 100$ , tem-se que  $D_{100} = 0$  quando  $T < 100$  e  $D_{100} = 1$  quando  $T \geq 100$ . Assim,  $C_{100} = 0$  para a observação 1 a 100 e  $C_{100} = 1, C_{100} = 2, C_{100} = 3$  e assim por diante, para as observações 101, 102, 103, etc. As regressões a serem estimadas são, portanto:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 [D_j (T - K_j)^m] \quad (3)$$

$$Y_t = \beta_0 + \beta_1 T + \beta_2 C_j$$

onde é possível incluir, além da tendência, a tendência ao quadrado, ao cubo, etc., conforme for o objetivo da série.

Estimadas as  $n$  regressões, deve-se considerar como nó da *spline*, de acordo com o trabalho de Hausmann, Pritchett e Rodrik (2005), a observação que, entre um intervalo estatisticamente significativo, apresentou maior valor  $F$  (teste de significação geral de uma regressão). O sentido implícito dessa escolha se dá porque será nos pontos de inversão que a *spline* mais se aproximará da regressão dos dados brutos contra a linha de tendência. Em outras palavras, menor será a soma dos quadrados dos resíduos e, portanto, maior o valor associado ao teste  $F$ .

Com os nós descobertos, é possível estimar a regressão da variável objetivo que servirá de base para a análise de discriminante. Essa regressão consiste da série analisada contra a variável de tendência e contra os nós encontrados a partir das estimações anteriores:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 C_j^* + \beta_3 C_j^{**} + u_t \quad (4)$$

## 2. A análise fractal

A utilização das leis de potência na análise econômica decorre do fato de elas representarem os sistemas ditos complexos (que se encontram longe do equilíbrio) e cujas partes interagem de forma não linear. Nestes casos, existem regularidades, um grande número de fenômenos comuns e um pequeno número de fenômenos raros. É possível, assim, expressar tais comportamentos mediante funções em termos exponenciais, em que aparece um parâmetro  $k$  que determina a escala do problema (Gleria, Matsushita e Silva, 2004). Essa característica sugere que, onde parecem existir somente aleatoriedades há, na verdade, certa ordem oculta e uma simetria entre pequenas e grandes escalas. Este padrão encontrado na irregularidade é uma das principais características dos fractais, que descrevem geometrias irregulares e fazem referência a objetos autossimilares, ou seja, àqueles que quando divididos apresentam as partes semelhantes ao todo.

Uma das mais conhecidas e utilizadas leis de potência para representar as propriedades de memória de longo prazo é o expoente de Hurst, que é expresso por:

$$E \left[ \frac{R(n)}{S(n)} \right] = Cn^H, n \rightarrow \infty \quad (5)$$

A partir de Racine (2011) é possível perceber que a estatística  $R/S$  é, assim, composta por dois termos: o  $R(n)$  e o  $S(n)$ . O  $R(n)$ , também conhecido por *range*, corresponde à diferença entre os valores máximos e mínimos da série de tempo  $x_k$ , de tamanho  $N$ , como segue:

$$R(n) = \max(x_k, k = 1, 2, \dots, N) - \min(x_k, k = 1, 2, \dots, N) \quad (6)$$

Por outro lado, o  $S(n)$  faz referência ao desvio padrão da série. Deste modo, a estatística  $R/S$  será equivalente ao produto dos valores assumidos por  $C$ , que representa uma constante arbitrária, e  $n$ , o número de observações constantes em cada grupo de análise, elevado a  $H$ , o expoente de Hurst.

Deste modo, para calcular o expoente de Hurst, é necessário dividir a série de dados em diferentes subgrupos e calcular para cada um destes a estatística  $R/S$ . Em outras palavras, para cada  $n$  (subgrupo), com  $n = 1, 2, 3, \dots$ , é possível encontrar um  $a_n$ , que faz referência à estatística  $R/S$  do subgrupo  $n$ :

$$a_n = E\left[\frac{R(n)}{S(n)}\right] \quad (7)$$

De acordo com o número de elementos  $n$ , o lado direito da Equação 7 é estimado para cada uma das partes, de modo que o valor esperado para toda a série corresponde à média de todos os sub-resultados. Isso leva à Equação 8:

$$E[a_n] = Cn^H \quad (8)$$

e desta para a Equação 9:

$$\log(E[a_n]) = \log(C) + H \log(n) \quad (9)$$

Mediante a utilização de um método simples de regressão, como o dos Mínimos Quadrados Ordinários, pode-se encontrar o valor assumido pelo expoente de Hurst, que indicará a força da dependência de longo prazo. Quando localizado entre 0,5 e 1, há um indicativo de que a série contém uma autocorrelação que se mantém ao longo do tempo; quando entre 0 e 0,5, existe um componente de anticorrelação; e, por fim, para o caso de  $H = 0,5$  (ruído branco), existe uma independência ou pequena dependência estrutural.

Recentemente, o método DFA (*Detrended Fluctuation Analysis*) tem se tornado amplamente utilizado como técnica para a determinação de propriedades monofractais e na detecção de correlações de longo prazo. Entretanto, é possível que um único expoente não seja suficiente para descrever a dinâmica de uma série temporal, de modo

que se torna necessário o cálculo de um coeficiente para cada período de tempo. Nestes casos, é mais vantajosa a utilização do método MF DFA (*Multifractal Detrended Fluctuation Analysis*).

Esta técnica generalizada consiste em cinco etapas, sendo as três primeiras idênticas ao cálculo do expoente de Hurst via DFA (Kantelhardt *et al.*, 2002).

- a) Assim como já mostrado,  $x_k$  corresponde à série temporal de tamanho  $N$ . Nesta primeira fase, busca-se calcular os desvios da série ajustada (excluída a média), o que é conhecido por transformá-la em uma espécie de passeio aleatório (*random walk like time series*). Denotando-se  $\langle x \rangle$  como a média de  $x_k$ , tem-se que:

$$Y(i) \equiv \sum_{k=1}^i [x_k - \langle x \rangle], i = 1, 2, \dots, N \quad (10)$$

- b) A série  $x_k$  deve ser dividida em subamostras de variadas escalas  $s$ , cada uma delas tendo números inteiros de elementos. Isso é representado por  $N_s \equiv \text{int}(N/S)$ .
- c) A tendência local de cada um dos segmentos  $N_s$  é determinada por um polinômio adequado aos dados ( $y_v(i)$ ), de modo que suas variâncias por segmento  $v=1, 2, \dots, N_s$  são estimadas através da Equação (11):

$$F^2(v, s) \equiv \frac{1}{s} \sum_{i=1}^s \{Y[(v-1) + i] - y_v(i)\}^2 \quad (11)$$

- d) É possível agora calcular a média da variância para todos os segmentos, de modo a obter a  $q$ -ésima função de flutuação  $F_q(s)$ . Para  $q \neq 0$ , a função é dada por:

$$F_q(s) \equiv \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [F^2(v, s)]^{q/2} \right\}^{1/q} \quad (12)$$

Caso seja igual a zero, a função adquire o seguinte formato:

$$F_0(s) \equiv \exp \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} \ln[F^2(v, s)] \right\} \quad (13)$$



- e) Deste modo, se a série temporal apresenta uma correlação de longo prazo, então o  $F_q(s)$  cresce com aumentos de  $s$  de acordo com a seguinte lei de potência:

$$F_q(s) \approx s^{h(q)} \quad (14)$$

onde o  $h(q)$  é conhecido por expoente de Hurst generalizado. Neste caso, para séries monofractais, o valor de  $h(q)$  independe de  $q$ , sendo  $F^2(v,s)$  idêntico para todos os segmentos  $v$ , de modo que  $h(q) = H$ . Caso contrário, o valor de  $h(q)$  variará de acordo com a estrutura assumida.

Além disso, Kristoufek (2010) destaca que, a partir de  $h(q)$ , é possível obter os expoentes locais de Hurst, o que é feito através da fixação do tamanho da amostra, variando as janelas de estimação do coeficiente. Fazendo isso, é possível obter uma nova série temporal dos expoentes de Hurst que variam conforme o tempo –  $H(t)$ .