

Previsão da Taxa de Crescimento do Produto Agrícola Brasileiro: Uma Aplicação de Modelos de Índice de Difusão Linear e Não Linear

José Nilo de Oliveira Júnior

Doutorando em Economia – CAEN/UFC
Bolsista do Programa de Demanda Social da CAPES

Ivan Castelar

Prof. Titular do Programa de Pós-Graduação em Economia – CAEN/UFC

Robeto Tatiwa

Prof. Departamento de Economia Aplicada-UFC
Pesquisador do Programa de Pós-Graduação – CAEN/UFC

Previsão da Taxa de Crescimento do Produto Agrícola Brasileiro: Uma Aplicação de Modelos de Índice de Difusão Linear e Não Linear

José Nilo de Oliveira Júnior
Ivan Castelar
Roberto Tatiwa

RESUMO

Este trabalho estuda modelos lineares e não lineares de índices de difusão para prever, um período à frente, a taxa de crescimento trimestral do PIB Agrícola brasileiro. Estes modelos compõem-se de fatores que são variáveis não observáveis e representam uma característica em comum às variáveis explicativas, permitindo a redução significativa do número dessas no modelo econométrico proposto para atender o objetivo central deste trabalho. Para tanto, foram utilizados 83 variáveis associadas aos setores interno e externo coletadas junto ao IPEA e mais 10 variáveis pluviométricas coletadas junto ao INMET.

Palavras – Chaves: Previsão, PIB Agrícola, Índice de Difusão.

ABSTRACT

The present study uses linear and non-linear diffusion index models to produce one-step-ahead forecast of quarterly Brazilian Agricultural GDP growth rate. These factors are latent variables that represent a common property from the explanatory variables, then allowing a considerably reduction of its number in econometric models elaborated to attend the main objective of this work. For in such a way, 83 variables had been used associates to the sectors internal and external collected from IPEA and more 10 climates variables collected from INMET.

Key Words: Forecasting, Agricultural GDP, Diffusion Index.

1. Introdução

O setor agrícola, no início da década de 1990, foi marcado por uma mudança do comportamento do poder central no que diz respeito aos programas que atendiam ao setor. Como por exemplo, o Programa de Abastecimento que na década de 1980 representava 70% das aplicações federais no setor, já na década seguinte este aporte não chegou a 30%. A maior causa da redução de recursos para a agricultura foi a política de estabilização adotada a partir de março de 1990, onde o controle monetário e fiscal foram seus pontos centrais. Com esta mudança na política econômica, alterou-se substancialmente a forma de financiamento da agricultura pelo setor público (Gasques e Villa Verde, 1991). Os dados mostram que no início da década de 1990 ocorreu um esvaziamento das políticas de curto prazo, as quais concentraram, na década anterior, quase todo o esforço de política agrícola. Para algumas políticas como a do Trigo, Açúcar e Álcool, a retirada do governo teve pontos positivos face à economia que isto representou devido às distorções que estas provocaram ao longo dos anos. Porém para outras políticas como o Crédito Rural, Preços Mínimos e Estoques Reguladores, a retirada trouxe problemas que se refletiram diretamente no volume de produção, nos investimentos no setor e principalmente na estabilização de preços e na renda.

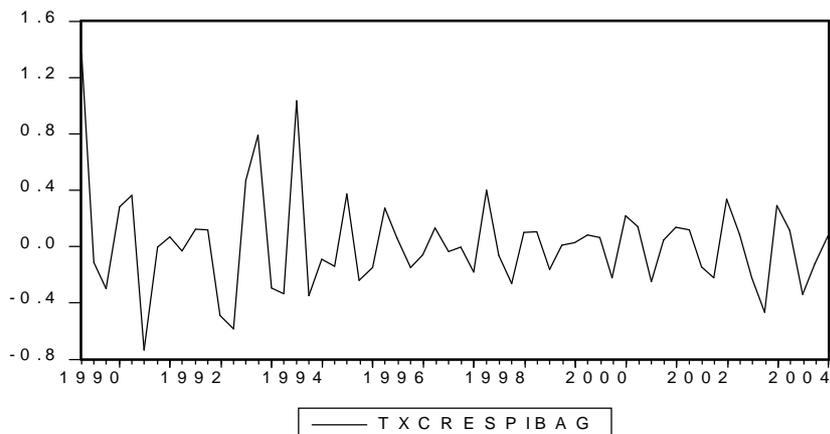
Já a partir de 1994, com a implementação do Plano Real e com as mudanças que o acompanharam, principalmente no câmbio, o setor agrícola passou a experimentar um momento mais favorável. No entanto, uma queda dos preços internacionais anulou parcialmente o estímulo derivado da mudança cambial de 1999; mas, em compensação, uma fase de alta desses preços internacionais a partir de 2002 reforçou o estímulo cambial, resultando em grande dinamismo para a agricultura brasileira, principalmente nos últimos três anos agrícolas. Verificou-se também que a volatilidade dos preços internacionais deu lugar, a partir do segundo semestre de 2004, a uma fase de baixos preços agrícolas no mercado internacional, a qual poderá atingir adversamente o comportamento do setor nos próximos anos.

Constata-se, também, que a expansão agrícola recente caracterizou-se por um aumento de nada menos do que 22,8% da área plantada com grãos, ao longo de apenas três anos agrícolas (2001/2002, 2002/2003 e 2003/2004). Esta expansão recente difere radicalmente do padrão que prevaleceu durante toda a década de 1990, no qual a área agrícola total com lavouras permaneceu constante e todo o aumento da produção agrícola vegetal adveio de aumentos da produtividade da terra. Essa expansão recente de área se deu sobretudo na soja, a qual cresceu somente nesses três anos agrícolas 39,8% nas regiões Sul e Sudeste e nada menos que 66,1% na região Centro-Oeste. Essa mudança positiva no setor fez com que o produto agrícola voltasse a crescer sobretudo através de ganhos de produtividade provenientes de melhoramentos tecnológicos (Brandão *et. al.*, 2005).

Dada todas estas transformações que o setor agrícola enfrentou durante toda a década de 1990 e início da década de 2000, sobretudo no que diz respeito à magnitude do produto agrícola, ganha grande importância econômica e política a possibilidade de se prever a sua taxa de crescimento.

A literatura especializada vem chamando a atenção para a importância da parcimônia no processo de previsão; ou seja, quanto maior o número de coeficientes no modelo maior a incerteza sobre os parâmetros estimados e isso pode trazer graves consequências para a eficiência da previsão. Desta forma, vários modelos têm sido propostos para evitar este problema.

GRÁFICO 1: Taxa de Crescimento do Produto Agrícola



Os Modelos de Fatores para séries de tempo têm sido uma destas opções. A idéia central é que toda a informação incluída em um grande número de variáveis poderia ser capturada por um pequeno número de fatores comuns entre elas. Pelo menos duas vertentes da literatura têm usado este método. Uma dessas correntes é representado por Modelos de Fatores Dinâmicos (Geweke, 1977; Sargent e Sims, 1977; Geweke e Sinfleton, 1981; Engle e Watson, 1981; Stock e Watson, 1989,1991; Quah e Sargent, 1993; Kim e Nelson, 1998). A característica principal desses estudos é o esforço para estimar os fatores comuns não observados entre algumas variáveis macroeconômicas, utilizando para tanto o método de Máxima Verossimilhança (MLE), Filtro de Kalman ou ambos.

A outra seria representada pelos Modelos de Índices de Difusão (DI) (Connor e Korajczk, 1993; Geweke e Zhou, 1996; Forni e Reichilin, 1998, Stock e Watson, 1998, 2002), que utilizam a técnica de componentes principais para estimar os fatores comuns. Esta técnica absorve um número maior de informações que a técnica de máxima verossimilhança (MLE) e aparentemente estima melhor os fatores quando o período em questão é o curto prazo.

Além do problema da parcimônia, os modelos de previsão enfrentam outros problemas, onde um dos principais seria quando existe quebra estrutural nas séries de tempo. Neste caso, os modelos não lineares e a técnica de Correção de Intercepto poderiam levar a melhores resultados.

Dentro da classe de modelos não lineares, destacam-se três: Threshold Autoregressivo (TAR), Markov Switching (MS) e Smooth Transition (STAR). Estes modelos têm sido usados na literatura para caracterizar ciclos de negócios bem como períodos de expansão e recessão.

Este trabalho consistirá, portanto, em prever, um período à frente, a taxa de crescimento trimestral do produto agrícola brasileiro, utilizando para isso o modelo linear de Índices de Difusão bem como o modelo não linear Theshold Autoregressivo e compará-los ao modelo linear AR, que é considerado na literatura como modelo padrão para estimativas de previsão. Para tanto serão utilizados 83 variáveis¹ coletadas junto ao IPEA² (Instituto de Pesquisa Econômicas Aplicada), todas relacionados ao setor mais as variáveis

¹ A descrição das variáveis utilizadas está no Apêndice I do trabalho.

² www.ipeadata.gov.br.

pluviométricas³ coletadas junto ao INMET⁴ (Instituto Nacional de Meteorologia), no período de 1990 a 2005.

Este trabalho se compõe de mais quatro seções além desta. A seção 2 trata da revisão da literatura. A seção 3 descreve a metodologia utilizada e a seção 4 detalha os resultados encontrados. E, por último, a seção 5 traz as conclusões encontradas.

2. Revisão da Literatura

Vários são as estruturas teóricas envolvidas no desenvolvimento de modelos lineares e não lineares para o processo de previsão econômica. Obviamente, os diferentes tipos de modelos requerem diferentes teorias, bem como diferentes técnicas para o seu desenvolvimento (ver, por exemplo, Van Dijk e Franses (2000)).

Dentro da literatura de Índices de Difusão, destacam-se os trabalhos de Connor e Korajczyk (1986, 1988, 1993) e Forni e Reichilin (1996, 1997, 1998); ambos consideram a determinação dos fatores em grandes sistemas. Connor e Korajczyk (1986, 1988, 1993) mostraram que as estimativas dos fatores por componentes principais são consistentes quando $N \rightarrow \infty$, com T fixo.

Já para os modelos não lineares, Hansen (2000) desenvolveu a teoria estatística e assintótica para a estimação por Threshold Autoregressivo, em um contexto de regressão, tanto para dados *cross-section* como para séries de tempo, onde se chega a conclusão que a distribuição assintótica do modelo é não padrão.

As aplicações empíricas constatadas em Clements e Krolzig (1998) e Clements e Smith (1999), evidenciaram que os resultados apontam que os modelos lineares são robustos no mecanismo de previsão. Por outro lado, Clements e Frances (1999) encontraram uma fraca evidência, para o PIB dos Estados Unidos, que o modelo TAR fornece uma melhor previsão comparada ao modelo AR.

Clements, Frances, Smith e Dijk (2003) além de fazerem esta comparação dos estimadores dos modelos TAR e AR, também compararam o intervalo e a densidade da previsão. Eles encontraram que a densidade e o intervalo de previsão do modelo não linear são melhores do que as do modelo linear. Na mesma direção, Dijk e Silverstovs (2003), utilizando modelos lineares e não lineares para prever a taxa de crescimento da produção industrial para o grupo de países do G7, corroboram a idéia de que o modelo não linear apresenta resultados melhores do que o modelo linear em termos de descrever as incertezas contidas nas séries de tempo.

Sarantis (1999), em sua tentativa de prever a taxa de câmbio real para os Estados Unidos usando o modelo linear e o modelo STAR, não encontrou resultados semelhantes entre si. Similarmente, Boero e Marrocu (2002) concluíram que as previsões para taxa de câmbio nominal dos Estados Unidos do modelo STAR não foram melhores do que as previsões do modelo AR.

Terui e Dijk (2002), utilizando uma combinação de modelos lineares e não lineares para prever a taxa de desemprego mensal e o índice de produção industrial da indústria para a economia americana e canadense, encontraram que a combinação de modelos provê uma performance melhor na previsão do que a performance isolada de cada modelo.

Diebold e Nason (1990), enumeraram várias razões que justificam a melhor performance obtida pelos modelos lineares em comparação aos não lineares. Uma razão é que a aparente não linearidade detectada pelos testes para linearidade é devida a *outliers* ou

³ Foram utilizados dados de precipitação pluviométrica para as regiões brasileiras.

⁴ www.inmet.gov.br.

a quebras estruturais, o que diminui a performance da previsão. Outra causa bastante comum relatada pelos autores é quanto ao processo de geração dos dados (DGP).

Para o Brasil, vários são os trabalhos que fizeram previsões sobre a taxa de crescimento do produto brasileiro. Alguns autores utilizaram diferentes modelos na tentativa de encontrar a melhor estimativa de previsão. Por exemplo, Moreira, Fiorêncio e Lopes (1996), utilizaram um VAR, VEC, BVAR e BVEC.

Já Moreira e Amendola (1998), usaram o Modelo Vetorial Autoregressivo Bayesiano de variáveis orientadas, bem como um Modelo Bayesiano Dinâmico que extrai a tendência, a sazonalidade e os ciclos. Já Chauvet (2001) e Chauvet, Lima e Vasquez (2002) mostraram que os modelos não lineares fornecem melhores resultados para previsão da taxa de crescimento do produto brasileiro do que os modelos lineares.

Ferreira (2005), utilizando modelos lineares e não lineares bem como uma combinação dos mesmos para prever a taxa de crescimento do produto brasileiro, encontrou que a eficiência da previsão do modelo não linear é superior à do modelo linear e que a combinação das técnicas prever melhor que os dois modelos isoladamente.

3. Metodologia

3.1 Modelo de Índice de Difusão

Os modelos utilizados neste trabalho para prever a taxa de crescimento do produto agrícola brasileiro, são construídos levando-se em consideração co-movimentos, fases econômicas e a possibilidade da existência de quebras estruturais. O Modelo de Índice de Difusão e suas aplicações segue Stock e Watson (1998, 2002) onde, segundo estes, esta técnica tem como finalidade elaborar um modelo parcimonioso que captura os co-movimentos existentes entre as séries.

De acordo com Bartholomew e Knott (1999), os modelos de fatores, ou os de Índice de Difusão, são modelos com variáveis latentes. Isto significa que algumas variáveis são não observáveis. Faça f representar r destas variáveis e x ser k variáveis observadas, $r < k$. Seguindo as idéias presentes em Rummel (1970), a análise do modelo de fatores comuns expressa a matriz de dados $X_{(T \times k)}$ como uma combinação linear de vetores linearmente desconhecidos, usualmente chamados de fatores comuns, com a adição de um único fator. Para $i = 1, \dots, k$, pode-se representá-lo da seguinte forma:

$$\begin{bmatrix} x_{1i} \\ \mathbf{M} \\ x_{Ti} \end{bmatrix} = I_{i1} \begin{bmatrix} f_{11} \\ \mathbf{M} \\ f_{T1} \end{bmatrix} + \mathbf{L} + I_{ii} \begin{bmatrix} f_{1r} \\ \mathbf{M} \\ f_{Tr} \end{bmatrix} + \begin{bmatrix} e_1 \\ \mathbf{M} \\ e_T \end{bmatrix} \quad (1)$$

Onde, cada f acima é o escore do fator; cada vetor coluna com estes escores de fatores é um fator comum; cada λ é o peso do fator, e cada e representa o único fator mencionado anteriormente. A definição usual de fatores comuns é que eles são funções lineares de variáveis conhecidas, contribuindo para a variância comum de todas as variáveis. Por outro lado, o fator unitário contribui somente para a variância da variável que está sendo explicada. Este é usualmente dividido em dois componentes: variância específica e erro aleatório.

Um outro modo de expressar o sistema (1), para um dado período de tempo $T = t$, é:

$$x_{1t} = I_{11}f_1 + \mathbf{L} + I_{1r}f_r + e_{1t}$$

M

$$x_{kt} = I_{k1}f_1 + \mathbf{L} + I_{kr}f_r + e_{kt}$$

ou simplesmente,

$$x_t = \Lambda F_t + e_t \quad (2)$$

Onde, $x_t = [x_{1t}, \dots, x_{kt}]'$ é um vetor ($k \times 1$), Λ é uma matriz ($k \times r$) de pesos dos fatores, $F_t = [f_1, \dots, f_r]'$ é um vetor ($r \times 1$), $e_t = [e_{1t}, \dots, e_{kt}]'$ é um vetor ($k \times 1$) de erros, e $r < k$. Assumindo que estas partes comuns são não correlacionados com a parte unitária; e estas partes não são correlacionadas no tempo, e $F'F = FF' = I_r$, fica fácil mostrar⁵ que $p \lim(1/n) = XX' = \Sigma = \Lambda\Lambda' + \Psi$, e logo $\text{var}(x_i) = \sum_{j=1}^r I_{ij}^2 + f_i$.

Pode-se afirmar, então, que não somente todas as k variáveis na matriz X são representadas por uma combinação linear de r fatores comuns mais o fator unitário, produzindo um subconjunto de variáveis, como também que esses r fatores comuns e seus pesos são suficientes para explicar a estrutura de variância comum de todas as k variáveis.

Considerando y_{t+1} a séria que irá ser utilizada na previsão, a classe de modelos usados neste trabalho são da forma:

$$y_{t+1} = c + a(L)y_t + b(L)x_t + e_{t+1} \quad (3)$$

para $t = 1, \dots, T$ e $\alpha(L)$ e $\beta(L)$ são polinômios no operador de defasagens de dimensões q_1 e q_2 , respectivamente. Observe que existem $(q_1 + q_2) \times k$ parâmetros em (3). Quando k é muito grande a estimação desses parâmetros pode ser muito imprecisa.

Assumindo⁶ que $E(e_{t+1} \mid F_t, y_t, x_t, F_{t-1}, y_{t-1}, x_{t-1}, \mathbf{L}) = 0$ e também que (y_{t+1}, x_t) possuem uma representação de fator dinâmico, com $\bar{r} (\bar{r} < k)$ fatores dinâmicos, Stock e Watson (2002) refinaram a equação (3), da seguinte forma:

$$y_{t+1} = c + a(L)y_t + b(L)f_t + e_{t+1} \quad (4)$$

$$x_t = I(L)f_t + e_t \quad (5)$$

Onde $I(L) = I + B_1(L) + \mathbf{L} + B_p(L^p)$, B_i é uma matriz ($k \times \bar{r}$) e f_t é um vetor de fatores ($\bar{r} \times 1$). Então, o modelo de fator pode absorver um grande número de observações contidas nas k variáveis em um pequeno número de fatores r . Também, modelando todas as defasagens polinomiais quando se tem uma ordem finita, de pelo menos q , Stock e Watson

⁵ Assumindo o bom comportamento dos dados, para poder aplicar a lei fraca dos grandes números de Khinchine.

⁶ Esta próxima suposição implica que $E(y_{t+1} \mid b_t, y_t, X_t, b_{t-1}, y_{t-1}, X_{t-1}, \mathbf{L})$ depende somente de F_t .

desenvolveram uma representação estatística da equação (5), onde esta é invariante no tempo, ou seja:

$$y_{t+1} = c + ay_t + b'F_t + e_{t+1} \quad (6)$$

$$x_t = \Lambda F_t + e_t \quad (7)$$

Onde $a = (a_0, \mathbf{L}, a_q)'$, $F_t = (f_t', \mathbf{L}, f_{t-q})'$ são vetores $(r \times 1)$ com $r \leq (q+1)\bar{r}$, $\Lambda_i = (I_{i0}, \mathbf{L}, I_{iq})$ e $b = (b_0, \mathbf{L}, b_q)'$. Se a suposição usual de infinitas defasagens fosse aplicada, então esta seria a representação estatística de um modelo de fator dinâmico de infinitos fatores. Felizmente, a vantagem central da última representação é permitir que a estimação dos fatores possa ser realizada por Componentes Principais, que possui algumas vantagens sobre o modelo de fator dinâmico obtido através da estimação por Máximo Verossimilhança (MLE).

Primeiro, a técnica de componentes principais absorve um maior número de variáveis que o MLE. Segundo, Stock e Watson (1998) mostraram que a estimação de fatores por componentes principais é consistente quando o número de variáveis vai para o infinito, em um período fixo de tempo.

3.2 Modelo de Índice de Difusão Threshold

O Modelo Threshold Autoregressivo (TAR), tem sido usado em estudos empíricos macroeconômicos na tentativa de capturar fases de expansões e recessões nos ciclos de negócio ou em qualquer outra situação que requeira uma divisão induzida da amostra para regimes diferentes. O modelo TAR foi primeiro proposto por Tong (1978) e depois aperfeiçoado por Lim e Tong (1980) e Tong (1983). Hansen (1996a, 1996b, 1997 e 2000) mostrou como estimar e fazer inferência no modelo. O modelo de índice de difusão threshold autoregressivo pode ser expresso da seguinte forma:

$$y_{t+1} = (c_1 + a^1 y_t + b^1 F_t)I(g_{t-1} \leq g) + (c_2 + a^2 y_t + b^2 F_t)I(g_{t-1} > g) + e_{t+1} \quad (8)$$

$$x_t = \Lambda F_t + e_t \quad (9)$$

Onde $a^j = (a_1^j, \mathbf{L}, a_q^j)'$, $b^j = (b_1^j, \mathbf{L}, b_q^j)'$ para $j = 1, 2, \dots$, g_{t-1} são conhecidos como funções de dados e $I(\cdot)$ é função indicador. Faça $z_t = (1 \quad y_t \quad F_t)'$, $p^j = (c_j \quad a^j \quad b^j)$, $z_t(I) = (z_t I(g_{t-1} \leq g) \quad z_t I(g_{t-1} > g))'$ e $q = (p^1 \quad p^2)' g_{t-1}$ então a equação (8) pode ser escrita da seguinte forma:

$$y_{t+1} = z_t(g)'q + e_{t+1} \quad (10)$$

3.3 Estimação, Teste e Previsão

3.3.1 Estimação do Modelo de Índice de Difusão (DI)

O procedimento de estimação⁷ para o modelo de índice de difusão autoregressivo, baseado nas equações (6) e (7) é composto de dois passos. Primeiro, o número exato de fatores é desconhecido. Então, sob a hipótese de existência de n ($n < k$) fatores comuns, os dados observados x_t são utilizados para estimar esses fatores. A formulação estática de um

⁷ Foi utilizado o programa Gauss para estimar o modelo DI e fazer as previsões.

modelo de fator dinâmico apresentado em (6) e (7) segue a técnica de componentes principais para estimar os fatores comuns não observados. Desde que a técnica de componentes principais é muito sensível ao tamanho dos dados, foram usados valores padronizados de x_t . As estimativas dos fatores \hat{F}_t são os autovetores associados com os n maiores autovalores da matriz padronizada $k^{-1} \sum_{i=1}^k x_i x_i'$, onde $x_i = (x_{i1}, \mathbf{L}, x_{iT})$ é um vetor ($T \times 1$). O Apêndice II mostra com detalhes o processo de estimação por componentes principais.

Deste modo, estes fatores estimados têm a propriedade de autovalor-autovetor no problema de componentes principais. Isto significa que o primeiro fator é o autovetor associado com o mais alto autovalor, e ele pode ser entendido como uma combinação linear de dados observados que explicam a maior parte da variância dos dados. Seguindo esta lógica, o segundo mais alto valor representa uma combinação linear de dados que melhor explicam a parte da variância que não é explicada pelo primeiro fator, e assim por diante. Além disso, uma outra característica importante da solução por componentes principais é a rotação que garante que cada um desses fatores irá ser linearmente independente dos outros, evitando, assim, qualquer grau de colinearidade que possa existir entre os regressores.

No segundo passo, y_{t+1} é regredido sobre uma constante, \hat{F}_t e y_t para obter estimativas de \hat{c}, \hat{a} e \hat{b} . Estes dois passos para o desenvolvimento do método de estimação foram adotados em Stock e Watson (1998, 2002)⁸.

3.3.2 Estimação do Modelo Threshold (TARDI)

A estimação do modelo TARDI será baseada em Hansen (1977). Será utilizado dois tipos de funções em relação à g_{t-1} , uma tradicional aproximação de pequenas diferenças $(Ln(PIB_{t-1}/PIB_{t-2}))_{t-d}$, e uma para grande diferenças $Ln(PIB_{t-1}/PIB_{t-d})$ onde d é um escalar positivo chamado de prolongador de defasagens. Desde que neste caso a equação de regressão em ambos os casos – não linear e descontínua – a estimação dos parâmetros θ e γ será obtida através de mínimos quadrados condicional sequencial. Faça $\mathbf{g} = g_{t-1}$ e $\Gamma = [\underline{\mathbf{g}}, \bar{\mathbf{g}}]$, onde o estimador de MQ de γ pode ser encontrado na investigação dos valores de Γ que minimiza os resíduos da regressão de y_t em $z_t(\mathbf{g})$. Ou seja:

$$\hat{\mathbf{g}} = \arg \min_{\mathbf{g} \in \Gamma} \frac{1}{n} (y_t - z_t(\mathbf{g})' \hat{\mathbf{q}}(\mathbf{g}))' (y_t - z_t(\mathbf{g})' \hat{\mathbf{q}}(\mathbf{g})) \quad (11)$$

onde,

$$\hat{\mathbf{q}}(\mathbf{g}) = \left(\sum_{t=1}^n z_t(\mathbf{g}) z_t(\mathbf{g})' \right)^{-1} \left(\sum_{t=1}^n z_t(\mathbf{g}) y_t \right) \quad (12)$$

⁸ Stock e Watson (1998) mostraram que os fatores estimados são uniformemente consistentes, e que esses estimadores são consistentes mesmo quando existe variação no tempo em Λ . E ainda, mostraram também que se r é desconhecido e se $m \rightarrow r$, pode-se alcançar um Erro de Previsão eficiente.

Após obter \hat{g} , o estimador de Mínimos Quadrados de θ , este é computado como $\hat{q} = \hat{q}(\hat{g})$.

Vale salientar que irá ser realizado o teste para a existência do efeito threshold na série de tempo da taxa de crescimento do PIB agrícola onde este irá seguir o procedimento descrito em Ferreira (2005).

3.3.3 Diagnóstico para o Teste do Modelo TARDI

3.3.3.1 Teste para Correlação Serial

Considere o modelo não linear geral:

$$y_t = F(x_t, q) + e_t \quad (16)$$

O teste de Multiplicador de Lagrange (LM) dado por $nR^2 \sim c_{(q)}^2$ poderia ser útil para testar a presença da q -ésima dependência de ordem serial nos resíduos. Onde R^2 é o coeficiente de determinação da regressão de \hat{e}_t em $\frac{\partial F(x_t, \hat{q})}{\partial q}$ e q são os resíduos defasados. Entretanto, $F(\cdot)$ não é continuamente diferenciável no caso de um Modelo TAR; então, este teste estatístico não é apropriado. Um possível caminho para evitar este problema é aproximar o modelo TAR com o Smooth Transition AR (STAR) (Frances e Dijk, 2000). De qualquer forma, isto requereria estimar um modelo STARDI, o que não faz parte do objetivo do trabalho.

3.3.3.2 Teste para Presença de Não-Linearidade

Para saber se o modelo linear captura todas as seqüências não lineares das séries de tempo, pode-se utilizar o teste descrito na subseção 3.3.2. Por exemplo, num modelo TAR com regime duplo, poder-se-ia testar a presença de não linearidade usando o procedimento para testar $H_0 : p^1 = p^2$ em duas sub-amostras, $g_{t-1} \leq \hat{g}$ e $g_{t-1} > \hat{g}$.

3.3.4 Previsão

O processo de previsão utilizado foi da forma recursiva. Desta forma, a equação geral usada para o modelo DI, para fazer previsão um passo à frente, é dada por:

$$\hat{y}_{T+1|T} = \hat{c}_h + \sum_{i=1}^{q_1} \hat{a}_i y_{T-i+1} + \sum_{j=1}^{q_2} \hat{b}_j \hat{F}_{T-j+1} \quad (17)$$

Onde, $y_{t+1} < \ln\left(\frac{y_{t+1}}{y_t}\right)$ e $y_t < \ln\left(\frac{y_t}{y_{t-1}}\right)$. Como em Stock e Watson (2002), o modelo DI utiliza somente fatores correntes no processo de previsão. O modelo DI-AR é o modelo DI mais as defasagens da variável dependente, onde o número de defasagens foi escolhido baseado no Critério de Informação Bayesiano (BIC). Vale ressaltar que o número de fatores em um modelo depende do modelo possuir fatores defasados ou não.

O modelo autorregressivo (AR), foi utilizado como *benchmark* para comparar com a performance do modelo DI, e ele foi estimado fazendo $\hat{b} = 0$ na equação (17), seguindo o número de defasagens escolhido pelo BIC.

Já para o modelo TARDI, a previsão de um passo a frente, segue a seguinte equação:

$$\hat{y}_{T+1/T} = \left(\hat{c}_1 + \sum_{i=1}^{q_1} \hat{a}^1 y_{T-i+1} + \sum_{j=1}^{q_2} \hat{b}^1 \hat{F}_{T-j+1} \right) I(g_{t-1} \leq \hat{g}) + \left(\hat{c}_1 + \sum_{i=1}^{q_1} \hat{a}^2 y_{T-i+1} + \sum_{j=1}^{q_2} \hat{b}^2 \hat{F}_{T-j+1} \right) I(g_{t-1} > \hat{g}) + e_{t+1} \quad (18)$$

As variáveis adotadas na função g_{t-1} foram curtas e longas diferenças do logaritmo do PIB.

4. Resultados Empíricos

4.1 Resultados do Modelo de Índice de Difusão

A medida de eficiência dos dois modelos DI de previsão usados neste estudo é a proporção⁹ do Erro Quadrático Médio de Previsão (MSFE) com relação ao modelo AR(1), que foi escolhido como *benchmark*. O modelo de previsão um passo à frente DI obteve uma melhor previsão do que os modelos de previsão AR(1), o modelo DI – AR, DI – LAG, e o modelo DI – AR – Lag.

Com relação ao critério de escolha do número de defasagens utilizados neste trabalho, o qual corresponde ao BIC, este selecionou o número zero de defasagens para serem utilizados. Esta escolha de zero defasagens implica que os resultados dos modelos AR(1), DI – AR, DI – LAG e o modelo DI – AR – Lag são praticamente os mesmos (Ver Tabela I).

Pode-se citar como exemplo, de acordo com a Tabela I, o modelo DI com apenas um fator, que prevê cerca de 10% melhor do que o modelo AR(1). A melhor previsão do modelo DI foi com a presença de quatro fatores, onde esse obteve uma superioridade em torno de 60% com relação ao modelo AR(1). Já os modelos DI – BIC, DIAR – BIC, DILAG – BIC, e o modelo DIARLAG – BIC, obtiveram a mesma previsão que o modelo AR(1)¹⁰.

TABELA I – Resultados dos Modelos DI

Número de Fatores	Modelos
	DI / DI – AR / DI – LAG / DI – AR – LAG (q = 0)
r = 1	0.90
r = 2	0,80
r = 3	0.54
r = 4	0.39
r = 5	0.65
BIC (nenhuma defasagem)	DI - BIC
	1.00
	DIAR – BIC
	1.00
	DILAG – BIC
	1.00
	DIARLAG – BIC
	1.00

⁹ Esta medida é utilizada como padrão neste tipo de literatura, ver por exemplo: Stock e Watson (1998).

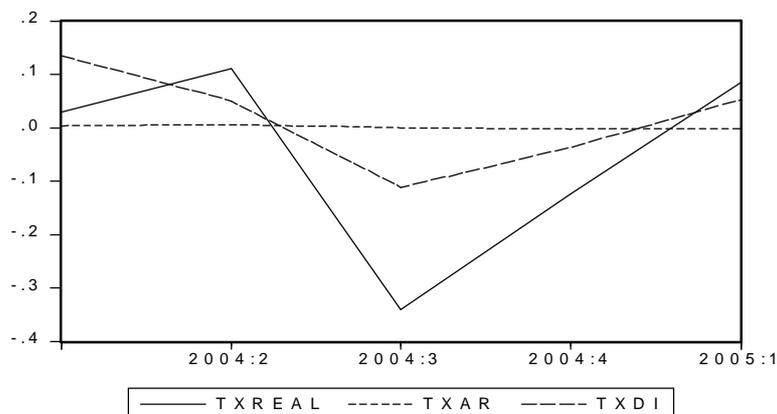
¹⁰ As estimações do modelo com quatro fatores estão apresentadas no Apêndice III do trabalho.

Estes resultados, de maneira geral, corroboram os resultados encontrados por Ferreira (2005) em seu estudo para a PIB agregado brasileiro no sentido de que o modelo DI prediz melhor se comparado ao modelo AR(1). Em seu estudo, Ferreira utiliza 72 variáveis para dois diferentes conjuntos de dados, onde o primeiro corresponde às defasagens destas variáveis, tendo como período base o primeiro trimestre de 1975 ao terceiro trimestre de 2003. O modelo DI com três fatores obteve uma predição cerca de 36% melhor do que o AR(1), o mesmo resultado obteve o modelo DI com três defasagens.

Já para economia americana, Stock e Watson (1998), utilizando um conjunto de 224 variáveis em uma estrutura de painel, mostraram para a taxa de crescimento da inflação e para taxa de crescimento da produção industrial que o modelo DI prediz melhor do que o modelo AR(1), o que corrobora os resultados aqui encontrados.

O Gráfico 2 apresenta os valores reais e os valores das previsões da taxa crescimento do PIB agrícola dos modelos DI e AR(1). Pode-se verificar não somente que a série de previsão do modelo DI aproxima-se bem mais da série valores atuais do que a do modelo AR(1), como também acompanha as mudanças de direção que ocorre no valores atuais. Um fato curioso com relação aos valores previstos pelo modelo AR(1) é que estes ficaram bem próximos de zero. Esta fraca performance da previsão do modelo AR(1) talvez possa justificar-se pela pequena capacidade deste modelo de capturar todas as característica pertinentes à taxa de crescimento do produto agrícola.

GRÁFICO 2: Valores Atuais e Previsões dos Modelos AR e DI



4.2 Resultados do Modelo Threshold

Na literatura especializada, é sempre uma rotina testar a existência de regimes diferentes antes de realizar a previsão do modelo TAR. Deste modo, é feita a previsão da taxa de crescimento do Produto Agrícola pela especificação do modelo DI linear selecionado como alternativa a um modelo TAR duplo¹¹. Como apresentado anteriormente, as curtas e longas diferenças do PIB agrícola foram testadas como variável para o efeito threshold. Foi utilizado um conjunto de defasagens variando de $d = [1, L, 4]$.

A Tabela II apresenta os resultados do teste para efeito threshold. Os p -valores sugerem que existe um efeito threshold significativa a um nível de 5% de significância quando é considerada as diferenças curtas de $(\ln(pib_{t-1} / pib_{t-2}))_{t-2}$ e $(\ln(pib_{t-1} / pib_{t-2}))_{t-4}$,

¹¹ Foi utilizado, para tanto, uma adaptação do modelo de Hansen (2000) para o programa GAUSS.

bem como quando se considera as longas diferenças de $\ln(pib_{t-1} / pib_{t-3})$ e $\ln(pib_{t-1} / pib_{t-4})$.

TABELA II – Teste para Efeito Threshold (1992:1 a 2005:1)

g_{t-1}	\hat{g}	p -valor
$(\ln(pib_{t-1} / pib_{t-2}))_{t-1}$	-0.2962	0.7410
$(\ln(pib_{t-1} / pib_{t-2}))_{t-2}$	0.0655	0.0090
$(\ln(pib_{t-1} / pib_{t-2}))_{t-3}$	0.1149	0.8170
$(\ln(pib_{t-1} / pib_{t-2}))_{t-4}$	0.0655	0.0080
$\ln(pib_{t-1} / pib_{t-2})$	-0.2962	0.7430
$\ln(pib_{t-1} / pib_{t-3})$	0.0956	0.0030
$\ln(pib_{t-1} / pib_{t-4})$	0.1754	0.0000

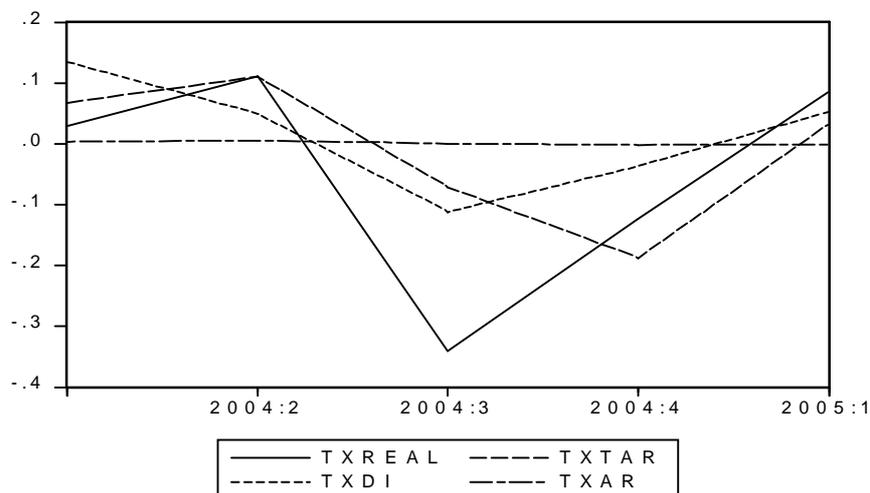
No entanto, será considerada, daqui para diante, apenas a diferença curta de $(\ln(pib_{t-1} / pib_{t-2}))_{t-2}$, pois esta foi a que apresentou os melhores resultados na estimação. Então foi estimado o modelo TARDI, o qual usado para fazer a previsão. Os resultados da estimação do modelo com dois fatores estão apresentados no Apêndice IV do trabalho.

Vale salientar que o teste de não-linearidade mostrou que não se pode rejeitar a hipótese de linearidade. Desta forma, os dois regimes do modelo TAR são suficientes para capturar a parte não-linear das séries de tempo utilizadas.

Outro fato importante no que diz respeito à magnitude da proporção da previsão entre o MSFE do modelo TARDI e o modelo DI (AR) que é de 0.12 (0.04). Além disso, o modelo TARDI previu a direção melhor que o modelo DI. Resultados semelhantes foram encontrados por Clements e Smith (1999), os quais constataram que o modelo TAR produz melhores previsões do que os demais modelos.

O Gráfico 3 apresenta os valores reais e previstos pelos modelos TARDI e pelo modelo DI. Observa-se que o modelo TARDI praticamente acompanha toda a série dos valores reais para a taxa de crescimento do produto agrícola. Isso mais uma vez comprova a eficácia do modelo TARDI em comparação aos modelos DI e AR(1).

GRÁFICO 3: Valores Atuais e Previsões dos Modelos DI e TARDI



5. Conclusões

O objetivo central deste trabalho foi gerar previsões trimestrais da taxa de crescimento do PIB agrícola brasileiro, usando para tanto os modelos de Índice de Difusão (DI) e o Threshold Autorregressivo (TARDI), bem como o modelo AR(1) como *benchmark*.

Para tanto, foram utilizados dados trimestrais do PIB agrícola brasileiro mais um conjunto de 83 variáveis macroeconômicas, representando o setor externo bem como o lado real e nominal da economia brasileira. O período que engloba as variáveis vai do primeiro trimestre de 1990 até o primeiro trimestre de 2005 e o período previsto, de forma recursiva, foi do primeiro trimestre de 2004 até o primeiro trimestre de 2005.

Os resultados encontrados foram considerados bastante satisfatórios. O modelo linear de Índice de Difusão (DI) com quatro fatores apresentou o melhor resultado dentre todos os modelos lineares (com um, com dois, com três e com cinco fatores). Quando se compara seu Erro Quadrático Médio de Previsão com o modelo AR(1) (*benchmark*) este prediz cerca de 60% melhor. Este resultado corrobora os resultados encontrados por Stock e Watson (1994) e por Ferreira (2005).

Com relação ao modelo não linear threshold (TARDI), os parâmetros mudam de acordo com as mudanças de regime econômico. Vale salientar que os regimes econômicos considerados neste estudo referem-se às fases de recessão e expansão, onde estas são capturadas pelo valor threshold que divide a amostra.

Este modelo não somente prediz melhor do que o AR(1) como também que o modelo linear de Índice de Difusão (DI). Em termos de Erro Quadrático Médio de Previsão, cerca de 88% e cerca de 30%, respectivamente, como também testa para o efeito threshold no modelo linear, confirmando que a parte não linear da taxa de crescimento do PIB agrícola brasileiro existe. Resultados semelhantes podem ser encontrados em De Gooijer e Kumar (1992), onde estes concluem que a inclusão de modelos não lineares é uma boa estratégia para obter uma melhor performance nas previsões.

Desta forma, chega-se à conclusão que o modelo não-linear obteve uma performance melhor, em termos de erros quadráticos médios, do que o modelo linear. No entanto, alguns estudos como Clements e Smith (1999), apontam que a combinação desses dois tipos de modelos, geralmente, obtém uma melhor previsão do que os dois modelos isolados.

6. Referências Bibliográficas

- BLINDER, A. *Credit Rationing and Effective Supply Failures*. *Economic Journal*. 97, 327-52, 1987.
- CHAUVET, M. *An Econometric Characterization of Business Cycles Dynamics with Factor Structure and Regime Switching*. *International Economic Review*. 39(4), 1998.
- _____. *A Monthly Indicator of Brazilian GDP*. *Brazilian Economic Journal*. (Revista Brasileira de Economia) 21, 2001.
- _____. *The Brazilian Business and Growth Cycles*. *Brazilian Economic Journal* (Revista Brasileira de Economia) 56(1), 2002.
- CHAUVET, M., LIMA E.C.R. e VASQUEZ, B. *Forecasting Brazilian Output in the Presence of Breaks: A Comparison of Linear and Nonlinear Models*. Working Paper 2002-28 Series of The Federal Reserve Bank of Atlanta, 2002.
- CLEMENTS, M.P., DIJK, V.D., FRANCES, P.H. e SMITH, J. *On Setar Non-Linearity and Forecasting*. *Journal of Forecasting* 22, 359-75, 2003.

- CLEMENTS, M.P. e HENDRY, D.F. *Forecasting Economic Time Series*. United Kingdom: Cambridge University Press, 1998.
- _____. *Forecasting Non-Stationary Economic Time Series*. MIT Press, 2001.
- CLEMENTS, M.P. e KROLSIG, H.M. *A Comparison of the Forecast Performance of Markov-switching and Threshold Autoregressive Models of US GNP*. *Econometrics Journal* 1, 47-75, 1988.
- CLEMENTS, M.P e SMITH, J. *A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Models*. *Journal of Applied Econometrics* 14, 123-41, 1999.
- _____. *The performance of in this alternative forecasting methods for SETAR models*. *International Journal of Forecasting*, 13, 463-475, 1999.
- CONNOR, G. e KORAJCZYK, R. *A Test for the Number of Factors in an Approximate Factor Model*. *Journal of Finance*, 48,4, 1993.
- DE GOOIJER, J.G e KUMAR, K. *Some Recent Developments in Non-Linear Time Series Modelling, Testing and Forecasting*. *International Journal of Forecasting* 8, 135-156, 1992.
- DIEBOLD, F.X. e LOPEZ, J.A. *Forecast Evaluation and Combination*. Technical working paper 192, National Bureau of Economic Research, 1994.
- DIEBOLD, F. X. e NASON, J. A. *Nonparametric Exchange Rate Prediction*. *Journal of International Economics*, 28, 315- 332, 1990.
- DIJK, V.D. e SILIVERSTOV, B. *Forecasting Industrial Production with Linear, Nonlinear, and Structural Change Models*. *Econometric Institute Report*.16, 2003.
- ENGLE, R.F. e WATSON, M.W. *A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates*. *Journal of the American Statistical Association*, 76, 376, 774-781, 1981.
- _____. *Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle*. *Review of Economic Studies*, 65, 453-474, 1998.
- FERREIRA, R.T. *Forecasting Quarterly Brazilian GDP Growth rate With Linear and Nonlinear Diffusion Index Models*. Tese de Doutorado Apresenta ao Curso de Pós-Graduação em Economia - CAEN da Universidade Federal do Ceará. 83p, 2005.
- GEWEKE, J. In: AIGNER D.J. e GOLDBERGER A.S. *Latent Variables in Socio-Economic Models*. Amsterdam: North-holland, 1977.
- GEWEKE, J, e SIGLETON, K.J. *Maximum Likelihood Confirmatory factor Analysis of Economic Time Series*. *International Economic Review*, 22, 1, 37-54, 1981.
- GEWEKE, J. e G. ZHOU. *Measuring the Price Error of the Arbitrage Pricing Theory*. *Review of Financial Studies*, 9, 557-587, 1996.
- HANSEN, B.E. *Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis*. *Econometrica*, 57(2), 357-84, 1989.
- _____. *Inference in TAR Models*. *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1997.
- _____. *Sample Splitting and Threshold Estimation*. *Econometrica*, 68(3), 575-603, 2000.
- MOREIRA, A.R.B. e AMENDOLA, E. *Comparação de Modelos de Previsão para o PIB e o Produto da Indústria*. IPEA – Textos para Discussão 613, 1998.
- MOREIRA, A.R.B., FIORENCIO, A. e LOPES, H.F. *Um Modelo de Previsão do PIB, Inflação e Meios de Pagamento*. IPEA – Textos para Discussão 446, 1996.

- NOBUHIKO TERUI , HERMAN K. VAN DIJK *Combined Forecasts from Linear and Nonlinear Time Series Models a , b*, International Journal of Forecasting 18, 421–438, 2002.
- QUAH, D. AND SARGENT, T.J. In: STOCK, J.H. e ND WATSON, M.W. *Business Cycles, Indicators and Forecasting*. Chicago: University of Chicago Press, 285-306, 1983.
- SARANTIS, N. *Modelling Non-Linearities in Real Effective Exchange Rates*. Journal of International Money and Finance, 18, 27–45, 1999.
- SARGENT, T.J. AND SIMS, C.A. In: SIMS, C. et al. *New Methods in Business Cycle Research*. Minneapolis: Federal Reserve Bank of Mineapolis, 1977.
- STOCK, J.H. e WATSON, M.W. *New Indexes of Coincident and Leading Economic Indicators*. NBER Macroeconomics Annual, 351-393, 1989.
- _____. In: LAHIRI AND MOORE, G.H. *Leading Economic Indicators: New Approaches and Forecasting Records*. New York: Cambridge university Press, 63-85, 1991.
- _____. *Evidence on Structural Instability in Macroeconomics Time Series*. Technical working paper 164, National Bureau of Economic research, 1994.
- _____. *Diffusion Indexes*. Technical working paper 6702, National Bureau of Economic Research, 1998.
- _____. *Macroeconomic Forecasting Using Diffusion Indexes*. Journal of Business & Economic Statistics, 20, 2, 147-162, 2002.
- VAN DIJK, D. J. C., e FRANSES, P. H. *Non-linear Error Correction Models for Interest Rates in The Netherlands*. In W. Bar-nett, 2000.

Apêndice I: Descrição das Variáveis

Taxa de Crescimento do PIB Agrícola	
Taxa de Câmbio	I(1) Intercepto
Saldo Balança Comercial Agrícola	I(1) Intercepto e Dessazonalizado
Saldo Balança Comercial	I(1) Intercepto
Investimentos Diretos Estrangeiros no País	I(0) Tendência e Intercepto
Investimentos Diretos Brasileiros no Exterior	I(0) Intercepto
Saldo Transações Correntes	I(1) Intercepto
Outros Investimentos	I(0) Intercepto
Reservas Internacionais	I(1) Intercepto
Taxa de Juros LIBOR	I(2) Intercepto
Taxa de Juros SELIC	I(1) Intercepto
Exportações Brasileiras	I(2) Intercepto
Exportações Mundiais	I(1) Intercepto
Exportações Países Industrializados	I(1) Intercepto
Exportações Bens de Consumo Duráveis	I(1) Intercepto
Exportações Bens de Consumo Não-Duráveis	I(2) Intercepto
Exportações Bens Intermediários	I(1) Tendência e Intercepto
Exportações Combustíveis	I(0) Intercepto
Exportações Bens de Capital	I(1) Tendência e Intercepto
Exportações para Alemanha	I(1) Intercepto
Exportações para Argentina	I(1) Intercepto
Exportações para Bélgica	I(0) Tendência e Intercepto
Exportações para Canadá	I(2) Intercepto
Exportações para Espanha	I(2) Intercepto
Exportações para EUA	I(1) Tendência e Intercepto
Exportações para França	I(1) Intercepto

Exportações para Holanda	I(1) Tendência e Intercepto
Exportações para Itália	I(1) Intercepto
Exportações para Japão	I(1) Intercepto
Exportações para Paraguai	I(1) Intercepto
Exportações Produtos Básicos	I(2) Intercepto
Exportações para Reino Unido	I(1) Intercepto
Exportações para Uruguai	I(1) Intercepto
Exportações Produtos Industrializados	I(2) Intercepto
Exportações Produtos Manufaturados	I(2) Intercepto
Exportações Beneficiamento Produtos Vegetais	I(1) Tendência Intercepto
Exportações Abate de Animais	I(1) Intercepto
Exportações Borracha	I(1) Intercepto
Exportações Açúcar	I(1) Intercepto
Exportações Produtos Semi-Manufaturados	I(1) Intercepto
Exportações Laticínios	I(2) Intercepto
Exportações Celulose	I(1) Intercepto
Exportações Madeira e Mobiliário	I(2) Intercepto
Exportações Extrativa Mineral	I(2) Intercepto
Importações Brasileiras	I(0) Tendência Intercepto
Importações Mundiais	I(1) Intercepto
Importações Países Industrializados	I(1) Intercepto
Importações Bens Intermediários	I(0) Tendência Intercepto
Importações Bens de Capital	I(0) Tendência Intercepto
Importações Bens de Consumo	I(1) Intercepto
Importações Bens de Consumo Não-Duráveis	I(1) Intercepto
Importações Combustíveis	I(0) Tendência Intercepto
Importações Petróleo	I(1) Intercepto
Importações Matérias Primas e Produtos Intermediários	I(0) Tendência Intercepto
Importações Combustíveis e Lubrificantes	I(2) Intercepto
Importações Abate de Animais	I(0) Intercepto
Importações Beneficiamento de Produtos Vegetais	I(0) Tendência Intercepto
Importações Extrativa Mineral	I(1) Intercepto
Importações Industria Diversa	I(0) Tendência Intercepto
Importações Laticínios	I(0) Intercepto
Importações Máquinas e Tratores	I(0) Tendência Intercepto
Importações Outros Produtos Industrializados	I(1) Intercepto
Consumo Derivados do Petróleo	I(1) Intercepto
Consumo Gasolina	I(2) Intercepto
Consumo Óleo Combustível	I(1) Intercepto
Consumo Energia Elétrica	I(2) Intercepto
Consumo Gás GLP	I(2) Intercepto
PIB Alemanha	I(2) Intercepto
PIB Canadá	I(2) Intercepto
PIB Chile	I(2) Intercepto
PIB Coréia do Sul	I(2) Intercepto
PIB Espanha	I(1) Intercepto
PIB EUA	I(2) Intercepto
PIB França	I(0) Intercepto
PIB Holanda	I(2) Intercepto
PIB Itália	I(1) Intercepto
PIB Japão	I(2) Intercepto
PIB Reino Unido	I(2) Intercepto
Preço Médio Recebido pelo Produtor de Arroz	I(1) Intercepto e Tendência
Preço Médio Recebido pelo Produtor de Banana	I(2) Intercepto
Preço Médio Recebido pelo Produtor Boi Gordo	I(2) Intercepto e Tendência
Preço Médio Recebido pelo Produtor de Feijão	I(21) Intercepto

Médias Pluviométricas do Estado do Ceará
Médias Pluviométricas do Estado da Bahia
Médias Pluviométricas do Estado do Amazonas
Médias Pluviométricas do Estado do Pará
Médias Pluviométricas do Estado de São Paulo
Médias Pluviométricas do Estado de Minas Gerais
Médias Pluviométricas do Estado de Goiás
Médias Pluviométricas do Estado de Mato Grosso
Médias Pluviométricas do Estado do Rio Grande do Sul
Médias Pluviométricas do Estado de Paraná

Apêndice II: Descrição da Técnica de Componentes Principais

Assuma que r é desconhecido e existe n ($n < k$) fatores comuns. Faça $\underline{x}_i = (x_{i1}, \mathbf{L}, x_{iT})'$ $\{i = 1, \mathbf{L}, k\}$ ser um vetor ($T \times 1$), $F = (F_1, \mathbf{L}, F_n)$ ser uma matriz ($T \times n$), $P_F = F(F'F)^{-1}F'$ e I_{ii} um vetor ($n \times 1$). Suponha que $I_{ii} = I_{i0}$ e F é uma matriz não aleatória desconhecida. O estimador para (I_{i0}, F) proposto por Stock e Watson (1998) minimiza a seguinte função objetivo:

$$V(I_{i0}, \mathbf{L}, I_{k0}, F) = k^{-1} \sum_{i=1}^k (\underline{x}_i - FI_{i0})' (\underline{x}_i - FI_{i0}) \quad (\text{II.1})$$

Faça $(\tilde{I}_{i0}, \tilde{F})$ ser a condição de primeira ordem de $V(I_{i0}, F)$ com respeito a I_{i0} . Portanto, isto implica que:

$$\tilde{I}_{i0} = (\tilde{F}'F)^{-1} \tilde{F}' \underline{x}_i \quad (\text{II.2})$$

Substituindo (A2) em (A1), obtém-se a função objetiva concentrada:

$$\begin{aligned} V(\tilde{I}_{i0}, F) &= k^{-1} \sum_{i=1}^k (\underline{x}_i - P_F \underline{x}_i)' (\underline{x}_i - P_F \underline{x}_i) \\ &= k^{-1} \sum_{i=1}^k \underline{x}_i' (I_T - P_F) \underline{x}_i \end{aligned} \quad (\text{II.3})$$

Considerando a normalização $F'F = I_n$ e que $\sum_{i=1}^k \underline{x}_i' \underline{x}_i = \text{tr}(\underline{x}' \underline{x}) = \text{tr}(k^{-1} \sum_{i=1}^k \underline{x}_i' (I_T - P_F) \underline{x}_i)$, minimizar a equação (A3) é equivalente a maximizar¹²

$$\text{tr} \left(F' \left(k^{-1} \sum_{i=1}^k \underline{x}_i' \underline{x}_i \right) F \right) \text{ sujeito à } F'F = I_n \quad (\text{II.4})$$

Onde, $k^{-1} \sum_{i=1}^k \underline{x}_i' \underline{x}_i = M$ é uma matriz simétrica ($T \times T$). Então, o problema de otimização torna-se:

$$\text{Max}_{F, I} = \sum_{j=1}^n F_j' M F_j + \sum_{j=1}^n I_j (1 - F_j' F_j) \quad (\text{II.5})$$

¹² Foi utilizado a seguinte propriedade para alcançar o resultado da equação (A4): $\text{tr}(ABCD) = \text{tr}(CDBA)$.

Desta forma, a condição de primeira ordem desse problema é $[M - I_j I_T] F_j = 0$, onde I_j é o autovalor associado a M e F_j é o autovetor também associado a M . Então, $\sum_{j=1}^n F_j' M F_j = \sum_{j=1}^n I_j$. Adicionando a informação que $n < T$, com intuito de maximizar $\sum_{j=1}^n F_j' M F_j$, coleta-se os n primeiros autovalores e os autovetores associados a eles da matriz $k^{-1} \sum_{i=1}^k x_i x_i'$ ($T \times T$). Assim, F é a matriz ($T \times n$) derivada desta associação dos autovetores.

Apêndice III: Resultados do Modelo DI

Previsão 2004:1 até 2005:1

Taxa de Crescimento do Produto Agrícola(Real)

0.28957
0.11159
-0.34094
-0.12354
0.08460

Previsão do Modelo AR(1)

0.00333
0.00526
-0.00081
-0.00293
-0.00144

MSE

0.04617

Previsão do Modelo DI e DI-AR

1 Fator	2 Fatores	3 Fatores	4 Fatores	5 Fatores
0.03421	0.03341	0.10730	0.13473	0.18133
0.05144	0.05297	0.04405	0.04900	0.09564
0.00459	-0.03002	-0.07419	-0.11174	0.00342
-0.01272	-0.03714	-0.00368	-0.03711	0.00909
-0.00135	-0.01813	0.05198	0.05197	0.10087

MSE

0.04158

mse ratio 0.90039

MSE

0.03675

mse ratio 0.79585

MSE

0.02487

mse ratio 0.53869

MSE

0.01779

mse ratio 0.38531

MSE

0.02968

mse ratio 0.64282

Previsão do Modelo DI-Lag e DI-AR-Lag

0.03421	0.03341	0.10730
0.05144	0.05297	0.04405
0.00459	-0.03002	-0.07419
-0.01272	-0.03714	-0.00368
-0.00135	-0.01813	0.05198

MSE

0.04158

mse ratio 0.90039

MSE

0.03675

mse ratio 0.79585

MSE

0.02487

mse ratio 0.53869

Previsão do Modelo DI-BIC, DI-AR-BIC, DI-Lag-BIC, DI-AR-Lag-BIC

0.00333

0.00526

-0.00081

-0.00293

-0.00144

MSE

0.04617

mse ratio 1.00000

Apêndice IV: Resultados do Modelo TARDI

Período da Amostra: 1991:1 a 2003:3

Modelo com 2 fatores

Global OLS Estimation, Without Threshold

Dependent Variable: pib

Heteroskedasticity Correction Used

Variable	Estimate	St Error
Constant	-0.0014929318	0.045950555
F1(t-1)	-0.032448581	0.38186551
F2(t-1)	-0.054967872	0.36651826
Observations	51.000000	
Degrees of Freedom	48.000000	
Sum of Squared Errors	4.6699742	
Residual Variance	0.097291129	
R-squared	0.00046864945	
Heteroskedasticity Test (P-Value)	0.78769790	

Threshold Estimation

Threshold Variable = Ldif2

Threshold Estimate 0.18287761

95% Confidence Interval: [0.118737 , 0.201809]

Sum of Squared Errors 2.8375600

Residual Variance 0.063056889

Joint R-Squared 0.39266684

Heteroskedasticity Test (p-value) 0.84902325

Regime 1: Ldif2 <= 0.182877

Parameter Estimates

Variable	Estimate	St Error
Constant	0.10998250	0.050122836
F1(t-1)	-0.35203435	0.40710488
F2(t-1)	0.36098978	0.36052323
95% Confidence Regions for Parameters		
Variable	Low	High
Constant	0.011741739	0.21460312
F1(t-1)	-1.2034710	0.44589122
F2(t-1)	-0.36972522	1.0676153

Observations	37.000000
Degrees of Freedom	34.000000
Sum of Squared Errors	2.6906344
Residual Variance	0.079136307
R-squared	0.054591452

Regime 2: Ldif2 > 0.182877

Parameter Estimates		
Variable	Estimate	St Error
Constant	-0.26308793	0.026031715
F1(t-1)	0.52363473	0.31218560
F2(t-1)	-0.97318230	0.29924420
95% Confidence Regions for Parameters		
Variable	Low	High
Constant	-0.31411009	-0.17948603
F1(t-1)	-0.088249058	1.2725837
F2(t-1)	-1.6288325	-0.38666366

Observations	14.000000
Degrees of Freedom	11.000000
Sum of Squared Errors	0.14692559
Residual Variance	0.013356872
R-squared	0.62979646

Sample Period 1991.0000 2004.1

Modelo com 2 fatores

Global OLS Estimation, Without Threshold

Dependent Variable: pib

Heteroskedasticity Correction Used

Variable	Estimate	St Error
Constant	0.0080011294	0.045539516
F1(t-1)	0.10299813	0.37705563
F2(t-1)	0.011192939	0.36318256

Observations	52.000000
Degrees of Freedom	49.000000
Sum of Squared Errors	4.7484759
Residual Variance	0.096907671
R-squared	0.0011783997
Heteroskedasticity Test (P-Value)	0.84793418

Threshold Estimation

Threshold Variable =	Ldif2
Threshold Estimate	0.18287761
95% Confidence Interval:	[0.118737 , 0.182877]
Sum of Squared Errors	2.8788956
Residual Variance	0.062584686
Joint R-Squared	0.39443663
Heteroskedasticity Test (p-value)	0.88703456

Regime 1: Ldif2 <= 0.182877

Parameter Estimates		
Variable	Estimate	St Error
Constant	0.11864134	0.049143384
F1(t-1)	-0.23299355	0.39136818

F2(t-1)	0.42177718	0.35216308
95% Confidence Regions for Parameters		
Variable	Low	High
<hr/>		
Constant	0.022320311	0.22173992
F1(t-1)	-1.0484468	0.53408809
F2(t-1)	-0.29113344	1.1120168
Observations	38.000000	
Degrees of Freedom	35.000000	
Sum of Squared Errors	2.7319700	
Residual Variance	0.078056285	
R-squared	0.051299047	

Regime 2: Ldif2 > 0.182877

Parameter Estimates		
Variable	Estimate	St Error
<hr/>		
Constant	-0.26308793	0.026031715
F1(t-1)	0.52363473	0.31218560
F2(t-1)	-0.97318230	0.29924420
95% Confidence Regions for Parameters		
Variable	Low	High
<hr/>		
Constant	-0.31411009	-0.17948603
F1(t-1)	-0.088249058	1.2725837
F2(t-1)	-1.6288325	-0.38666366
Observations	14.000000	
Degrees of Freedom	11.000000	
Sum of Squared Errors	0.14692559	
Residual Variance	0.013356872	
R-squared	0.62979646	

Sample Period 1991.0000 2004.2

Modelo com 2 fatores

Global OLS Estimation, Without Threshold

Dependent Variable: pib

Heteroskedasticity Correction Used

Variable	Estimate	St Error
<hr/>		
Constant	0.010307614	0.044430476
F1(t-1)	0.13543549	0.35973941
F2(t-1)	0.028340278	0.35477097
Observations	53.000000	
Degrees of Freedom	50.000000	
Sum of Squared Errors	4.7544404	
Residual Variance	0.095088807	
R-squared	0.0022141549	
Heteroskedasticity Test (P-Value)	0.92171760	

Threshold Estimation

Threshold Variable =	Ldif2
Threshold Estimate	0.11873745
95% Confidence Interval:	[0.088805 , 0.220111]
Sum of Squared Errors	2.9425984
Residual Variance	0.062608476
Joint R-Squared	0.38245455

Heteroskedasticity Test (p-value) 0.91012252

Regime 1: Ldif2 <= 0.118737

Parameter Estimates		
Variable	Estimate	St Error
Constant	0.12332430	0.050212050
F1(t-1)	-0.27279424	0.39574110
F2(t-1)	0.39270073	0.34889498
95% Confidence Regions for Parameters		
Variable	Low	High
Constant	0.015688888	0.23761971
F1(t-1)	-1.1043879	0.57267246
F2(t-1)	-0.33980041	1.1643992
Observations	36.000000	
Degrees of Freedom	33.000000	
Sum of Squared Errors	2.6674210	
Residual Variance	0.080830940	
R-squared	0.051040809	

Regime 2: Ldif2 > 0.118737

Parameter Estimates		
Variable	Estimate	St Error
Constant	-0.20957844	0.033533758
F1(t-1)	1.1055488	0.37864470
F2(t-1)	-0.83259317	0.31307544
95% Confidence Regions for Parameters		
Variable	Low	High
Constant	-0.30432400	-0.12209384
F1(t-1)	0.23746478	1.8686637
F2(t-1)	-1.5759547	0.024114733
Observations	17.000000	
Degrees of Freedom	14.000000	
Sum of Squared Errors	0.27517737	
Residual Variance	0.019655527	
R-squared	0.58375550	

Sample Period 1991.0000 2004.3

Modelo com 2 fatores

Global OLS Estimation, Without Threshold

Dependent Variable: pib

Heteroskedasticity Correction Used

Variable	Estimate	St Error
Constant	-4.4457586E-005	0.044189474
F1(t-1)	-0.033307540	0.37392221
F2(t-1)	-0.081789861	0.35109718
Observations	54.000000	
Degrees of Freedom	51.000000	
Sum of Squared Errors	4.8787857	
Residual Variance	0.095662465	
R-squared	0.0011849016	
Heteroskedasticity Test (P-Value)	0.92686897	

Threshold Estimation

Threshold Variable = Ldif2
 Threshold Estimate 0.11873745
 95% Confidence Interval: [0.088805 , 0.220111]
 Sum of Squared Errors 2.9823960
 Residual Variance 0.062133250
 Joint R-Squared 0.38942550
 Heteroskedasticity Test (p-value) 0.92257790

Regime 1: Ldif2 <= 0.118737

Parameter Estimates

Variable	Estimate	St Error
Constant	0.12332430	0.050212050
F1(t-1)	-0.27279424	0.39574110
F2(t-1)	0.39270073	0.34889498

95% Confidence Regions for Parameters

Variable	Low	High
Constant	0.015688888	0.23761971
F1(t-1)	-1.0954929	0.57267246
F2(t-1)	-0.33980041	1.1643992

Observations 36.000000
 Degrees of Freedom 33.000000
 Sum of Squared Errors 2.6674210
 Residual Variance 0.080830940
 R-squared 0.051040809

Regime 2: Ldif2 > 0.118737

Parameter Estimates

Variable	Estimate	St Error
Constant	-0.22571818	0.031536094
F1(t-1)	0.81925311	0.33778488
F2(t-1)	-1.0102297	0.30405103

95% Confidence Regions for Parameters

Variable	Low	High
Constant	-0.31653201	-0.13915744
F1(t-1)	0.096085943	1.5382114
F2(t-1)	-1.6986501	-0.20131083

Observations 18.000000
 Degrees of Freedom 15.000000
 Sum of Squared Errors 0.31497497
 Residual Variance 0.020998331
 R-squared 0.53343771

Sample Period 1991.0000 2004.4

Modelo com 2 fatores**Global OLS Estimation, Without Threshold**

Dependent Variable: pib

Heteroskedasticity Correction Used

Variable	Estimate	St Error
Constant	-0.0021543310	0.043211052
F1(t-1)	-0.072490101	0.34902870

F2(t-1)	-0.10728250	0.33506744
Observations	55.000000	
Degrees of Freedom	52.000000	
Sum of Squared Errors	4.8859285	
Residual Variance	0.093960164	
R-squared	0.0028636608	
Heteroskedasticity Test (P-Value)	0.99152624	

Threshold Estimation

Threshold Variable = Ldif2
 Threshold Estimate 0.11873745
 95% Confidence Interval: [0.088805 , 0.220111]
 Sum of Squared Errors 3.0241770
 Residual Variance 0.061717898
 Joint R-Squared 0.38281603
 Heteroskedasticity Test (p-value) 0.91876148

Regime 1: Ldif2 <= 0.118737

Parameter Estimates		
Variable	Estimate	St Error
Constant	0.11476548	0.048814231
F1(t-1)	-0.42232646	0.36701748
F2(t-1)	0.29369788	0.32596769

95% Confidence Regions for Parameters

Variable	Low	High
Constant	0.0090614016	0.22409777
F1(t-1)	-1.1766474	0.32647352
F2(t-1)	-0.38806642	1.0166137

Observations	37.000000
Degrees of Freedom	34.000000
Sum of Squared Errors	2.7092020
Residual Variance	0.079682412
R-squared	0.054861792

Regime 2: Ldif2 > 0.118737

Parameter Estimates		
Variable	Estimate	St Error
Constant	-0.22571818	0.031536094
F1(t-1)	0.81925311	0.33778488
F2(t-1)	-1.0102297	0.30405103

95% Confidence Regions for Parameters

Variable	Low	High
Constant	-0.31653201	-0.13915744
F1(t-1)	0.096085943	1.5382114
F2(t-1)	-1.6986501	-0.20131083

Observations	18.000000
Degrees of Freedom	15.000000
Sum of Squared Errors	0.31497497
Residual Variance	0.020998331
R-squared	0.53343771
q=	9.0000000

Previsão 2004:01 to 2005:01 (0.067015849; 0.11032801; -0.071335277, -0.18824789, 0.031459757; mse: 0.025845698)