

UNIVERSIDADE FEDERAL DE SÃO CARLOS, CAMPUS SOROCABA

PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Luisa Rasesa Ranero

**DESCOBERTA DE PREÇOS E ESPECULAÇÃO NO MERCADO DE  
MILHO BRASILEIRO**

Sorocaba-SP

2018

UNIVERSIDADE FEDERAL DE SÃO CARLOS, CAMPUS SOROCABA

PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Luisa Rasera Raniero

**DESCOBERTA DE PREÇOS E ESPECULAÇÃO NO MERCADO DE  
MILHO BRASILEIRO**

Dissertação apresentada ao Programa de Pós-Graduação em Economia para obtenção do título de Mestre em Economia.

**Orientador:**

Prof. Dr. José César Cruz Júnior

Sorocaba-SP

2018

Luisa Raserá Raníro

DESCOBERTA DE PREÇOS E ESPECULAÇÃO NO MERCADO DE MILHO BRASILEIRO

Dissertação apresentada ao Programa de Pós-Graduação em Economia para obtenção do título de Mestre em Economia. *Campus Sorocaba*, 2018

---

Orientador - Prof. Dr. José César Cruz Júnior  
Universidade Federal de São Carlos

---

Banca examinadora - Prof. Dr. Geraldo  
Edmundo Silva Júnior  
Universidade Federal de São Carlos

---

Banca examinadora - Prof. Dr. Rodrigo Lanna  
Franco da Silveira  
Universidade Estadual de Campinas

Sorocaba  
2018

Raniero, Luisa Rasesa

Descoberta de preços e especulação no mercado de milho brasileiro /  
Luisa Rasesa Raniero. -- 2018.  
72 f. : 30 cm.

Dissertação (mestrado)-Universidade Federal de São Carlos, campus  
Sorocaba, Sorocaba

Orientador: Prof. Dr. José César Cruz Júnior

Banca examinadora: Prof. Dr. Geraldo Edmundo Silva Júnior, Prof. Dr.  
Rodrigo Lanna Franco da Silveira

Bibliografia

1. Descoberta de preços. 2. Especulação. 3. Mercado de milho. I.  
Orientador. II. Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Bibliotecário(a) Responsável: Maria Aparecida de Lourdes Mariano – CRB/8 6979

## **Agradecimentos**

Agradeço aos meus pais, Sonia e Reinaldo, por não medirem esforços para que eu realizasse mais esse sonho e por todos os conselhos e amor incondicional. Ao meu irmão, por compreender meus momentos de ausência. Ao meu querido companheiro Murilo, pela compreensão quando não estive presente e por permanecer me dando força em todos os momentos.

Ao professor e mestre José César Cruz Júnior pelo acompanhamento, orientação e inspiração durante o desenvolvimento deste trabalho. Muito obrigada por contribuir com a minha formação desde 2010, pela paciência, dedicação e amizade durante esta jornada.

Aos meus amigos do mestrado por compartilharem alegrias e sonhos.

Aos meus amigos de profissão pelo apoio e amizade em todos os momentos: Camila, Zazá, Isallela, Karen, Karine, Tami, Brunno.

Agradeço ao Departamento de Economia e a todos os professores e técnicos administrativos que fizeram parte do meu crescimento pessoal e profissional.

Agradeço aos membros da banca examinadora por gentilmente aceitarem o convite.

A todos os que participaram de alguma forma para que esta etapa fosse concluída, o meu muito obrigada!

*"One child, one teacher, one pen and one book can change the world."  
(Malala Yousafzai)*

## Resumo

RANIRO, L.R. Descoberta de preços e especulação no mercado de milho brasileiro. 2018. 72f. Dissertação (Mestrado em Economia Aplicada) - Centro de Ciências em Gestão e Tecnologia, Universidade Federal de São Carlos, Sorocaba, 2018.

Desde os anos 2000, diversas mudanças ocorreram nos mercados brasileiro e internacional de milho em grão. No Brasil, sob a influência do milho da safra de inverno, a região Centro-Oeste passou a ser a principal produtora do grão no país, ultrapassando a região Sul, tradicional produtora do milho safra. No mercado internacional, o crescimento da demanda pela *commodity*, principalmente nos mercados chinês e americano, também alterou o comportamento dos preços no mundo todo. Diante destas mudanças, este trabalho investigou as relações dos preços nos mercados futuros e físicos de milho em grão nas regiões Sul e Centro-Oeste do Brasil, entre 2004 e 2016. Utilizando medidas de descoberta de preços foi possível identificar que o mercado de futuros foi o líder no processo de descoberta de preços de milho no período analisado. A análise dos mercados físicos indicou que a região Sul, representada por Cascavel, teve a maior parcela de informação na formação de preços quando comparada aos demais mercados regionais. No entanto, verificou-se que esta relação não foi consistente, uma vez que antes da produção do milho da safra de inverno ter superado a produção da safra, Rio Verde, no Centro-Oeste, era o mercado de referência na descoberta de preços no país. A análise dos preços no curto-prazo foi realizada através do teste *Generalized Supremum Augmented Dickey-Fuller (GSADF)*, capaz de identificar períodos de bolhas especulativas. Os resultados indicaram a ocorrência de diversos episódios explosivos em todos os mercados analisados. Tais episódios, na maioria dos casos, foram de curta duração (menos de 7 dias), e tiveram comportamentos diferentes, com períodos de bolhas positivas e negativas. Além disso, também foram identificados períodos em que as bolhas ocorreram simultaneamente em todos os mercados analisados, o que indica que pode existir uma relação também entre as volatilidades dos preços regionais. Acredita-se que os resultados encontrados podem ser úteis para aqueles que procuram por referências na formação dos preços, e por aprimorar estratégias de negociação da *commodity* no mercado físico. Além disso, os resultados indicam que, diante da identificação de um bolha, é possível criar expectativas sobre sua duração, e sobre as variações de subida e queda de preços durante o comportamento explosivo de preços.

**Palavras chave:** descoberta de preços, bolhas especulativas, mercado de milho.

## Abstract

RANIRO, L.R. Price discovery and speculation in the Brazilian corn market. 2018. 72f. Thesis (Master in Applied Economics) - Center for Management and Technology Sciences, Federal University of São Carlos, Sorocaba, 2018.

Since the 2000s, several changes occurred in the Brazilian and international corn markets. In Brazil, due primarily to the increasing importance of the winter crop, the central-west region became the main corn producer in the country, overcoming the South, the main summer crop producer. On the international market, the increase in demand for the commodity - mainly in the Chinese and U.S. markets - also changed the behavior of prices worldwide. Based on all the aforementioned changes, we analyzed prices relationships in the corn spot and futures markets in the south and central-west regions of Brazil, between 2004 and 2016. We used various price discovery measures to find that Brazilian corn futures market leads the price information in the country. Our analysis for the spot market indicates that the South, represented by the production region of Cascavel, shows the highest share of price information, and therefore can be considered the primary reference in the price discovery process. However, we found that this result was not consistent, as the region of Rio Verde in the Center-West was the benchmark in the price discovery process during the period when summer crop production was higher than winter crop production (prior to 2011). We analyzed short-term price relationships using the Generalized Supremum Augmented Dickey-Fuller (GSADF) test to identify periods of speculative bubbles. Our results indicated the occurrence of several explosive episodes in all markets. In most cases such episodes were short-lived (lasting less than seven days), with differing characteristics, and had periods of both positive and negative bubbles. In addition, we identified periods when bubbles occurred simultaneously in all markets. This result indicates that there may also be a relationship between regional price volatilities. We expect that our results can be useful for those looking for references in price formation, and for improving commodity trading strategies in the cash market. In addition, once a bubble period is identified, traders can possibly use our results to build their expectations regarding the duration of bubbles and the average price variation (increase and decrease) within a bubble period.

**Key words:** price discovery, speculative bubbles, corn market.

## Lista de ilustrações

Figura 1 – Evolução da produção de milho nas regiões Sul e Centro-Oeste nas 1 <sup>a</sup> e 2 <sup>a</sup> safras. . . . .	10
Figura 2 – Evolução dos preços do milho em R\$/saca de 60 Kg nas praças físicas e dos preços futuros no Brasil de 2004 a 2016 . . . . .	26
Figura 3 – Ilustração do procedimento do teste GSADF em uma amostra normalizada. .	37
Figura 4 – Dendograma hierárquico e valores do ILS considerando as quatro praças físicas entre 2004 e 2016 . . . . .	43
Figura 5 – Dendogramas hierárquicos e valores do ILS considerando as quatro praças físicas nos períodos de 2004 a 2010 e 2011 a 2016 . . . . .	44
Figura 6 – Resultado do teste GSADF e identificação dos períodos de bolha na série de preços futuros de milho (2004 a 2016) . . . . .	48
Figura 7 – Resultados do teste GSADF para os preços nos mercados da região Centro-Oeste (2004 a 2016) . . . . .	49
Figura 8 – Resultados do teste GSADF para os preços nos mercados da região Centro-Oeste (2004 a 2016) . . . . .	50
Figura 9 – Períodos de bolhas coincidentes em Cascavel, Chapecó, Rio Verde e Sorriso.	51
Figura 10 – Frequência do número de dias com bolha em todos os mercados entre 2004 e 2016 . . . . .	52
Figura 11 – Frequência dos episódios de bolhas positivas e negativas nas regiões Sul e Centro-Oeste no período de 2004 a 2016. . . . .	56

## Lista de tabelas

Tabela 1 – Estatística descritiva dos preços e retornos diários das séries dos mercados físicos e futuros . . . . .	39
Tabela 2 – Teste de quebra estrutural de Zeileis et al. (2003) . . . . .	40
Tabela 3 – Estatísticas do testes de raiz unitária ADF para as séries de preços e para suas respectivas diferenças . . . . .	41
Tabela 4 – Testes de cointegração de Johansen para o período de 2004 a 2016 . . . . .	42
Tabela 5 – Medidas de descoberta de preços entre 2004 e 2016 . . . . .	45
Tabela 6 – Medidas de descoberta de preços entre 2004 e 2010 . . . . .	46
Tabela 7 – Medidas de descoberta de preços entre 2011 e 2016 . . . . .	46
Tabela 8 – Características das bolhas nos preços de 2004 a 2016 ( $h \geq 3$ ) . . . . .	53
Tabela 9 – Características das bolhas nos preços dividido por períodos ( $h \geq 3$ ) . . . . .	57
Tabela 1A – Tamanho da amostra e da janela ( $r_w$ ) para cada ano de teste GSADF . . . . .	65
Tabela 2A – Testes de cointegração de Johansen para o período de 2004 a 2010 . . . . .	65
Tabela 3A – Testes de cointegração de Johansen para o período de 2011 a 2016 . . . . .	66
Tabela 4A – Modelo de correção de erros entre os preços de Cascavel ( $Pca$ ) e Rio Verde ( $Prv$ ) . . . . .	67
Tabela 5A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Chapecó ( $Pch$ ) . . . . .	68
Tabela 6A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Sorriso ( $Pso$ ) . . . . .	69
Tabela 7A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Rio Verde ( $Prv$ ) no período de 2004 a 2010 . . . . .	70
Tabela 8A – Modelo de correção de erros entre Rio Verde ( $Prv$ ) e Sorriso ( $Pso$ ) . . . . .	70
Tabela 9A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Rio Verde ( $Prv$ ) . . . . .	71
Tabela 10A – Modelo de correção de erros entre Pca e Pch entre 2011 e 2016 . . . . .	72
Tabela 1B – Medidas de descoberta de preços considerando o mercado de futuros entre 2004 a 2016 . . . . .	72

## Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>1.1</b>	<b>Caracterização do problema</b>	<b>10</b>
<b>1.2</b>	<b>Objetivos</b>	<b>13</b>
<b>1.3</b>	<b>Hipóteses</b>	<b>13</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>14</b>
<b>2.1</b>	<b>Abordagem teórica e evidências empíricas</b>	<b>14</b>
2.1.1	Descoberta de preços em mercados cointegrados	15
2.1.2	Limitações e comparação entre as medidas de <i>price discovery</i>	19
2.1.3	O comportamento dos preços diante de bolhas especulativas	20
<b>3</b>	<b>METODOLOGIA</b>	<b>25</b>
<b>3.1</b>	<b>Fontes e tratamentos dos dados</b>	<b>25</b>
<b>3.2</b>	<b>Estacionariedade das séries</b>	<b>26</b>
<b>3.3</b>	<b>Análise de cointegração e modelo de correção de erros</b>	<b>29</b>
<b>3.4</b>	<b>Estratégias para a estimação de <i>price discovery</i></b>	<b>31</b>
3.4.1	Medidas de informação entre mercados	31
<b>3.5</b>	<b>Procedimento para detecção de bolhas nos preços</b>	<b>35</b>
<b>4</b>	<b>ANÁLISE DOS RESULTADOS E DISCUSSÕES</b>	<b>39</b>
<b>4.1</b>	<b>Estatísticas descritivas</b>	<b>39</b>
<b>4.2</b>	<b>Medidas de descoberta de preços</b>	<b>40</b>
<b>4.3</b>	<b>Análise de curto prazo: bolhas especulativas</b>	<b>47</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>58</b>
	<b>REFERÊNCIAS</b>	<b>60</b>

# 1 INTRODUÇÃO

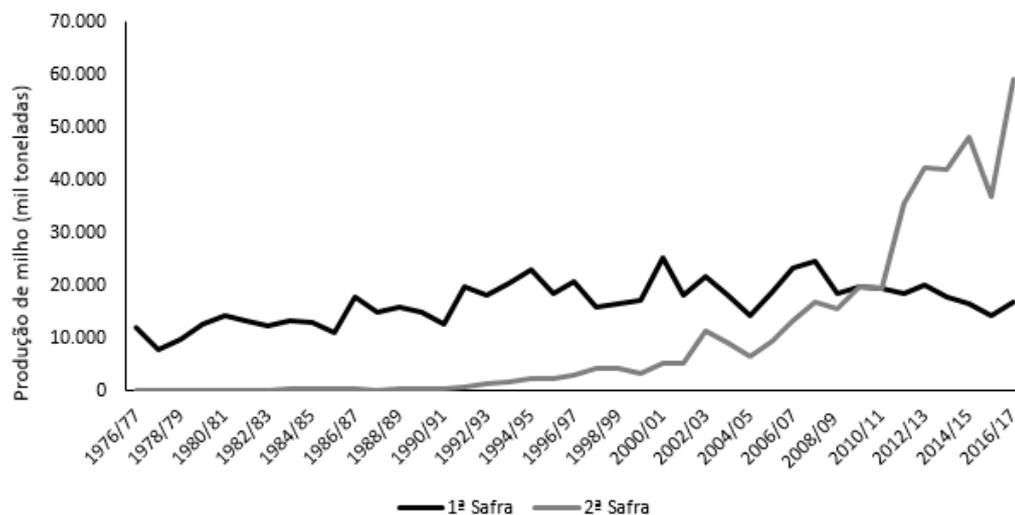
## 1.1 Caracterização do problema

A importância do milho para a economia brasileira se relaciona à indústria alimentícia por ser uma *commodity* importante usada principalmente como componente da ração de animais e alimento para os seres humanos. A produção de milho no país tem se expandido rapidamente principalmente em função do aumento da produtividade no período de plantio da safra de inverno, e do forte crescimento da indústria de suínos e aves (CRUZ JR. et al., 2016).

O Brasil é o terceiro maior produtor mundial de milho, superado apenas pelos Estados Unidos e pela China, sendo responsável por aproximadamente 9% da produção mundial de milho na safra 2016/17 (USDA, 2017). O país também ocupa uma posição importante como exportador mundial de milho, juntamente com países como os Estados Unidos, e Argentina. Nos anos 2000, o Brasil representava 5% do total das exportações mundiais, saltou para 20% em 2015/16, 11% em 2016/17 e 13% na safra de 2017/18 (MATTOS; SILVEIRA, 2015; FAS/USDA, 2018).

A produção de milho no Brasil se concentrou na região Sul do país até a safra de 2011, quando a região Centro-Oeste passou a ser a principal produtora no país. Dados da Companhia Nacional de Abastecimento (CONAB, 2017) indicam que no ano de 1976, a região Sul representava 51,9% da produção de milho total do Brasil, contra 10% da região Centro-Oeste. Em 2011, a região Sul foi responsável por 31,6% da produção, enquanto o Centro-Oeste produziu 42,6%. A partir de então, a região Centro-Oeste passou a ter maior relevância no cenário produtor, e na safra 2016/2017 foi responsável por 50% da produção, contra 27,7% da região Sul. A evolução da produção da primeira e da segunda safra podem ser vistas na Figura 1.

Figura 1 – Evolução da produção de milho nas regiões Sul e Centro-Oeste nas 1ª e 2ª safras.



Fonte: Conab (2017).

O plantio do milho no Brasil é feito principalmente nos meses entre setembro e novembro, variando entre esses meses dependendo do estado produtor. A safra (primeira safra) é colhida nos meses de janeiro até março e, na sequência, é feito o plantio da safra de inverno (segunda safra) em solo que anteriormente serviu para o plantio da soja. Esta substituição de culturas beneficia o cultivo do milho, uma vez que o fertilizante usado na soja permanece como resíduo do solo fazendo com que a produtividade do milho seja maior (REIS et al., 2016).

Os avanços tecnológicos permitiram um aumento de produtividade na segunda safra no Brasil, que representa a parte da produção brasileira de milho destinada à exportação. O cultivo da soja precoce auxilia o bom desempenho do milho na safra de inverno à medida que, como o milho nessa época do ano é plantado no mesmo solo da soja após sua colheita, o uso de variedades precoces de soja acelera o processo de plantio do milho tornando-o menos sujeito às incertezas climáticas. CRUZ et al. (2011) destacam que o milho de segunda safra tem sua produtividade comprometida pela restrição hídrica e altas temperaturas do período de plantio e, por isso, o manejo do solo tem papel fundamental para que haja quantidade de água suficiente para o desenvolvimento das plantas.

Além de mudanças no mercado doméstico, o mercado internacional do milho também sofreu alterações, especialmente com o desenvolvimento da indústria produtora de etanol a base de milho nos Estados Unidos e o aumento da demanda externa pelo produto puxada principalmente pela China. Outras mudanças estruturais também aconteceram no mercado de futuros e opções alterando a maneira como os preços são descobertos em vários mercados pelo mundo. Com a modernização das negociações no mercado de futuros das *commodities*, uma quantidade maior de negociadores passou a ter acesso ao mercado, o que colaborou para a redução dos custos das negociações, assim como sugerem Irwin e Sanders (2012).

As mudanças nos mercados físicos e futuro do milho influenciaram as alterações nos preços domésticos do grão, que apresentaram grande volatilidade na última década. Em meados de 2016, por exemplo, a produção brasileira da safra de inverno teve grande prejuízo devido à seca registrada nos primeiros meses do mesmo ano, fazendo com que os preços se elevassem. O produtor que opta por plantar milho na segunda safra assume um risco elevado, pois é necessário que, em uma pequena janela de tempo, as condições climáticas sejam favoráveis para o desenvolvimento da planta. De acordo com dados da Bloomberg (2016), os preços do milho no mercado de futuros, registraram uma volatilidade anual de 30% entre 2004 e 2016, enquanto as praças físicas de Cascavel-PR, Chapecó-SC, Rio Verde-GO e Sorriso-MS, apresentaram variações distintas no período, correspondentes a 18%, 20%, 25% e 39%, respectivamente.

Observando separadamente o período antes de 2011, ano no qual a produção da segunda safra superou a da safra, e depois de 2011, a volatilidade anual dos preços nas praças da região Sul apresentou pouca alteração. Em Cascavel a volatilidade anual até 2011 foi de 18,9% e após este ano foi para 17,9%. Em Chapecó, a volatilidade anual dos preços manteve-se praticamente constante nos dois períodos, em torno de 20,4%. Por outro lado, na região Centro-Oeste, houve queda na volatilidade anual dos preços nas praças de Rio Verde e Sorriso que, antes de 2011

apresentaram volatilidade anual de preços de 26,9% e 42,3%, respectivamente, e a partir de 2011 esses valores caíram para 23,3% em Rio Verde e 35,6% em Sorriso.

O comportamento dos preços nos mercados à vista e de futuros influencia a decisão de *traders* quanto a utilização de mecanismos de proteção contra variações indesejadas de preços. Para se proteger de tais variações, os participantes do mercado podem utilizar estratégias de *hedge* no mercado de futuros. Segundo Marques, Mello e Martines Filho (2006), em uma operação de *hedge* os agentes procuram fixar, no mercado de futuros, um preço de compra ou de venda de determinada mercadoria em sentido oposto à posição assumida no mercado físico. Por exemplo, um produtor que esperar colher milho em uma determinada localidade está comprado no mercado físico, e deverá fazer um *hedge* de venda no mercado futuro para garantir o preço da sua mercadoria. Algumas características importantes para que este mecanismo seja efetivo são: (i) que os preços físicos e futuros sejam fortemente correlacionados; e (ii) que haja facilidade de entrada e saída do mercado, ou seja, haja liquidez suficiente para garantir a negociabilidade dos contratos. Desta forma, *traders* procuram negociar contratos mais líquidos em busca de alta efetividade, seja em bolsas domésticas ou estrangeiras.

De acordo com Souza (1998) e Lazzarini (1999) apud Chiodi et al. (2005), quando os *hedgers* negociam em bolsas locais e não em bolsas estrangeiras, eles são mais efetivos. Uma das razões para a maior efetividade apresentada por Lazzarini (1999) apud Chiodi et al. (2005) se refere às especificidades dos contratos que variam de acordo com cada localização e portanto, um contrato estipulado em uma bolsa local deve ser mais adequado às necessidades dos agentes locais. Os autores afirmam ainda que é com base na avaliação do *trade-off* entre os custos de liquidez e a efetividade do *hedge*, que os negociadores decidem em qual bolsa devem atuar.

Apesar de considerarem os custos de liquidez e a efetividade do *hedge* para decidirem onde negociar, Chiodi et al. (2005) afirmam que os *hedgers* tendem a dar preferência à maior liquidez. Relacionado a isso, Maynard, Hancock e Hoagland (2001) mostraram que a falta de liquidez, especificamente devido a atividades especulativas, restringem a capacidade de transmissão de informações de preço do mercado futuro para o mercado à vista, fazendo com que os agentes optem por negociar em mercados com maior volume de contratos negociados.

Mesmo diante de várias mudanças no mercado de milho nos últimos anos, pouca atenção foi dada à investigação de como os preços domésticos no Brasil têm se formado diante do novo cenário exposto. Com o deslocamento da produção da região Sul para a região Centro-Oeste, somado às mudanças nos mercados internacional e de futuros, torna-se importante investigar se a formação de preços em ambas as regiões ocorre de maneira semelhante.

Assim, propõe-se investigar neste trabalho se houveram mudanças na informação contida nos preços do milho das regiões Sul, Centro-Oeste e no mercado de futuros do Brasil, considerando as mudanças significativas pelas quais os mercados passaram na última década.

## 1.2 Objetivos

O objetivo principal deste trabalho é identificar possíveis mudanças na informação contida nos preços do milho nas regiões Sul, Centro-Oeste e no mercado de futuros no Brasil no período de 2004 a 2016. De posse das medidas de informação, pretende-se também verificar a existência de bolhas especulativas nos mercados analisados. Especificamente, busca-se:

- a) identificar a parcela de informação - *information share*, *component share* e *information leadership* - que cada região e o mercado de futuros possuem na formação de preços domésticos do milho;
- b) verificar se houve mudanças nestas parcelas ao longo dos últimos anos; e,
- c) identificar a existência de períodos de bolhas especulativas nos preços ao longo do período e caracterizá-las.

## 1.3 Hipóteses

Espera-se que, a partir do cálculo das medidas de informação contidas nos preços, seja possível confirmar a hipótese de que houve mudança no processo de descoberta de preços nos principais mercados *spot* do país. Espera-se que, com o aumento da produção no Centro-Oeste, esta região tenha passado a ser a referência de preços no país. Diante da significativa volatilidade observada nos preços do milho nos últimos anos, espera-se também identificar períodos de comportamento especulativo nos preços físicos nas regiões investigadas.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Abordagem teórica e evidências empíricas

De acordo com Hieronymus (1977), o mercado de futuros se aproxima de um modelo perfeitamente competitivo no qual a condição de informação perfeita é atendida. Neste mercado, todos os *traders* têm livre acesso às informações públicas sobre os preços que estão sendo negociados para a *commodity*. Desta forma, por meio do processo de descoberta dos preços, os agentes dos mercados físicos podem utilizar as informações contidas nos preços futuros para criar suas expectativas em relação aos preços físicos, no futuro. Por este motivo, a forma como essa informação é disseminada no mercado físico é fundamental para que o planejamento da produção, da estocagem, o processamento e distribuição da *commodity* sejam executados de maneira eficiente. Estes planejamentos afetam diretamente a determinação dos preços.

Hieronymus (1977) apresenta a teoria do *cost-of-carry* como a base para determinar as relações existentes entre os preços à vista e de futuros ao longo do tempo. Geralmente, os preços futuros são superiores aos preços *spot*, porém, de acordo com o referido autor, não é incomum que o contrário aconteça, principalmente em períodos próximos ao vencimento do contrato futuro, fora do período de safra. Segundo a teoria do *cost-of-carry*, este comportamento pode ser explicado pela pressuposição de que *commodities* armazenáveis têm produções sazonais, e portanto, os preços devem refletir os ganhos dos produtores ao longo do período no qual não há produção. Além disso, há custos elevados para armazenar os grãos garantindo a manutenção da qualidade do produto ao longo do tempo. Por fim, assumindo que não há custos de negociação, os preços à vista aumentam em relação aos preços futuros à medida que o tempo de armazenamento da *commodity* aumenta, ou seja, quando o *cost-of-carry* é maior (HIERONYMUS, 1977).

Considerando o modelo do *cost-of-carry* no mercado futuro em competição perfeita, Kolb e Overdahl (2006) definem o custo de carregamento como sendo o custo total em se "carregar" uma certa quantidade de commodity ao longo do tempo. Para isso, os autores consideram quatro tipos principais de custos:

1. custos de armazenamento: contemplam os custos necessários para manter a *commodity* armazenada de forma adequada garantindo manutenção da qualidade do produto;
2. custos de transporte: são os custos com o transporte do produto até seu destino final;
3. custos com seguro: em muitos casos é necessário que o produtor projeta sua produção contra possíveis acidentes como incêndios ou inundações nos locais em que os produtos estão sendo mantidos até a data da entrega;
4. custos financeiros: são os custos de oportunidade de se manter o produto estocado por um período, considerando uma taxa de juros.

Segundo os mesmo autores, a regra básica para determinar a relação entre preços físicos e futuros de acordo com a teoria do custo de carregamento é que os preços futuros devem ser iguais aos preços físicos mais o custo de carregamento do produto físico de um período inicial ( $t = 0$ ) até a data de entrega determinada por um contrato futuro ( $t$ ). Isto é:

$$F_{0,t} = P_0(1 + C)^t \quad (1)$$

em que  $F_0$  é o preço futuro da *commodity*,  $P_0$  é o preço *spot* no tempo zero,  $C$  é o custo de carregamento.

Kolb e Overdahl (2006) afirmam que é muito comum a teoria do *cost-of-carry* não se aplicar no caso em que o produto em negociação possui um *convenience yield*, ou seja, uma bonificação para o *trader* manter o produto físico em sua posse.

### 2.1.1 Descoberta de preços em mercados cointegrados

A investigação do processo de descoberta, buscando identificar o comportamento dos preços em mercados cointegrados tem sido estudada com base em três abordagens principais. A primeira se refere à relação de antecendência defasagem<sup>1</sup> (*lead lag*) entre mercados, que mostra o quão mais rápido um mercado reage a uma nova informação em relação a outro. Esta abordagem é interessante para medir o quanto os mercados se relacionam, uma vez que no caso de haver correlação perfeita, o movimento de preços nos mercados deveria ocorrer simultaneamente. A segunda aborda o conceito de transmissão de volatilidade (*volatility spillover*)<sup>2</sup>, que se relaciona à velocidade com que uma nova informação ocorrida no passado é transmitida de um mercado a outro. Já a terceira abordagem analisa o processo de descoberta de preços a partir de como a informação é originada nos diferentes mercados. O presente estudo adota a terceira abordagem na análise dos preços futuros e *spot*. Para a análise do processo de *price discovery* utilizam-se as medidas de *information share* proposta por Hasbrouck (1995) e de *component share* apresentadas por Gonzalo e Granger (1995).

Em um trabalho seminal, Hasbrouck (1995) apresentou a medida de *information share*, que representa a parcela de informação de um determinado mercado no processo de descoberta de preços. O autor desenvolveu um modelo para o mercado de ações nos Estados Unidos, e definiu o *information share* de um mercado como a participação deste mercado na descoberta de preço, participação esta que é determinada pela parcela da variância do preço que é causada por uma nova informação no mercado. Em outras palavras, o trabalho busca, por meio de um modelo não linear de vetores autoregressivos (VAR), determinar o processo de descoberta de preços de mercados cointegrados a partir da parcela de variação da inovação (choques) pela qual cada mercado é responsável. Os resultados obtidos por Hasbrouck indicaram que existe um fator comum nos mercados físico e futuro, que é um preço eficiente determinado, sendo este implícito em cada ativo financeiro. Este fator comum seria resultado de um componente de um processo

<sup>1</sup> Chan (1992), Jong, Mahieu e Schotman (1998) são referências sobre esta abordagem.

<sup>2</sup> Trujillo-Barrera, Mallory e Garcia (2012) e Gardebroek e Hernandez (2013) são referências para esta abordagem.

de passeio-aleatório de preços dos diversos mercados. A parcela de informação atribuída a cada mercado é medida de acordo com a intensidade (proporção) com que a informação varia em cada um deles.

A outra medida, conhecida por *component share*, foi apresentada no trabalho seminal de Gonzalo e Granger (1995), que definiram o modelo considerando a abordagem da decomposição de fatores comuns em mercados cointegrados como peças fundamentais no processo de descoberta de preços. Os autores fizeram uma análise de fatores comuns capazes de determinar as relações de descoberta de preços no mercado futuro no longo prazo, utilizando para a obtenção destes fatores um processo de correção de erros. Eles determinaram os componentes dos fatores comuns a partir de uma combinação linear dos preços do mercado. Os autores obtiveram uma forma de estimar os fatores comuns a partir de um modelo é cointegrado, buscando esclarecer pontos acerca da natureza dos dados para o processo de descoberta de preços.

Dentre os diversos estudos que utilizaram os modelos de *information share* (IS) e *component share* (CS), grande parte se preocupou em determinar em qual mercado, físico ou futuro, uma nova informação é incorporada primeiro. Chu, Hsieh e Tse (1999) utilizaram as duas abordagens para verificar a dinâmica de *price discovery* em três índices de mercado da S&P 500<sup>3</sup>: o índice S&P 500 para o mercado físico, o índice de contratos futuros e o índice SPDRs<sup>4</sup>. Os resultados obtidos mostraram que os mercados são cointegrados e compartilhar uma mesma tendência de longo-prazo. A pesquisa mostrou ainda que o mercado de futuros lidera o processo de *price discovery* em relação aos outros dois mercados, que contribuem muito menos para a formação da tendência de longo-prazo.

Considerando que as medidas de *price discovery* revelam como uma nova informação é transmitida entre mercados, é natural esperar que os agentes financeiros optem por negociar no mercado cuja proporção no processo de descoberta de preços seja maior, já que este será o primeiro a refletir o choque de informação. Chakravarty, Gulen e Mayhew (2004) corroboram esta ideia com um estudo que revelou um percentual entre 10 e 20% de *price discovery* ocorrendo no mercado de opções. Os autores também afirmaram que agentes que possuem informações privadas podem optar por negociar no mercado de sua preferência. Portanto, pode-se inferir que, no caso de uma maior parcela do processo de descoberta de preço ocorrer no mercado de opções, será neste mercado onde ocorrerá o maior número de negociações.

Muitos pesquisadores se dedicaram a investigar as medidas de *price discovery* em mercados agrícolas, em grande parte para os mercados de países desenvolvidos. Na maioria dos casos, os resultados apontam o mercado de futuros como líderes no processo de descoberta de preços. Yang, Bessler e Leatham (2001), analisaram os mercados futuro e físico de *commodities* estocáveis e não estocáveis nos Estados Unidos para o período de 1992 a 1998. Por meio de análises de cointegração e correção de erros, os resultados mostraram, em geral, que a condição de

<sup>3</sup> Standard & Poor's 500 é um índice composto por quinhentas ações cotadas nas bolsas de Nova Iorque (NYSE) e NASDAQ.

<sup>4</sup> Standard & Poors Depositary Receipts.

armazenamento não afeta a função de descoberta de preços. De forma geral, o estudo concluiu que o mercado de futuros liderou a dinâmica de *price discovery* nos mercados analisados.

Alguns poucos estudos foram realizados para o caso de mercados em países em desenvolvimento. Su e Chong (2007) utilizaram os métodos de IS e CS para investigar as relações entre ativos chineses negociados no mercado norte-americano. O trabalho analisou oito companhias chinesas com ações listadas nas bolsas de valores de Nova Iorque e de Hong Kong e os resultados apresentados mostraram que os mesmos ativos negociados nas duas bolsas são cointegrados e possuem um fator comum. Os resultados mostraram que a bolsa de Hong Kong tem maior parcela de contribuição na formação dos preços dos oito ativos analisados. No entanto, os autores sugerem cautela na interpretação dos resultados, já que a análise considerou duas bolsas de valores internacionais com horários não sincronizados. Como as negociações na bolsa de Hong Kong fecham doze horas depois do fechamento das negociações em Nova Iorque, é esperado que o preço de fechamento na bolsa chinesa incorpore maior quantidade de informações.

Inicialmente, os modelos de *information share* e de *component share* foram utilizados para esclarecer o processo de descoberta de preços em mercados de futuros, mas tiveram amplo desenvolvimento também para os mercados físicos. Um exemplo é o estudo desenvolvido por Delfino (2012) que aplicou ambos os métodos para fazer uma análise sobre o mecanismo de *price discovery* no mercado brasileiro de títulos públicos. A hipótese testada no trabalho foi de que os títulos de dívida emitidos em reais no mercado internacional foram em parte responsáveis pelo desenvolvimento do mercado de títulos no Brasil. O autor utilizou dados diários para fazer uma comparação entre os mercados doméstico e externo em termos da contribuição de cada um no processo de *price discovery* para o período anterior e o posterior à crise financeira internacional de 2008. Os resultados indicaram que até a eclosão da crise <sup>5</sup>, o mercado externo foi responsável pela melhora eficiente da dívida pública doméstica. Depois da crise, o cenário se inverteu e o mercado doméstico teve maior percentual na descoberta de preço e, portanto, no desenvolvimento da dívida pública.

Diversos estudos foram realizados para investigar o processo de *price discovery* em mercados agrícolas utilizando as técnicas de IS e CS. Em muitos casos, ambas as medidas apresentaram resultados aproximados e consistentes (THEISSEN, 2002; JONG, 2002; XU, 2014). No entanto, outros estudos chegaram a resultados divergentes na análise do IS e do CS. Um exemplo é o trabalho de Silva (2013), que estudou as relações entre os mercados futuros de soja no Brasil, na China, nos Estados Unidos e na Argentina. O autor verificou que, para as medidas de IS, a bolsa norte-americana foi predominante no processo de descoberta de preço internacional da soja, com uma parcela de quase 60%. Já pelos resultados das medidas do *component share*, a bolsa da China apresentou maior relevância no processo, sendo responsável por cerca de 50% da informação no período de 2002 a 2011. Os resultados são interessantes pois, apesar de seguirem em direções diferentes, de acordo com o autor, eles são consistentes analiticamente<sup>6</sup>,

<sup>5</sup> Delfino (2012) trata o período pré-crise internacional financeira até a quebra do banco americano de investimentos Lehman Brothers. O período subsequente é considerado o período pós-crise.

<sup>6</sup> Silva (2013) apresenta os resultados do modelo de correção de erros no qual a bolsa da China apresentou menor

pois reforçam que os preços da bolsa chinesa apresentaram menor magnitude de ajustamento se comparados aos preços das outras bolsas analisadas.

Ainda no contexto de preços em mercados agrícolas, Kim, Tejeda e Wright (2016) analisaram a relação entre os preços do arroz em seis regiões dos Estados Unidos e o preço internacional de exportação do produto. Os autores utilizaram dados mensais do arroz com diferentes características (grão longo, médio e curto) para o período de 1979 a 2015. Eles aplicaram o método de Gonzalo e Granger e encontraram que os mercados do Arkansas e da Califórnia lideraram o processo de descoberta de preços. Ao analisarem a relação entre o mercado do Arkansas e o mercado externo, os resultados mostraram que os preços no mercado internacional lideraram os preços no mercado *spot*. Porém, em uma análise mais ampla, os autores encontraram que os preços internacionais foram liderados pelo mercado doméstico da Califórnia. Por fim, concluiu-se que os mercados domésticos e internacional do arroz tem igual contribuição para o processo de *price discovery* no período analisado.

Em um estudo desenvolvido para o mercado agrícola do Brasil, Mattos e Garcia (2004) consideraram seis *commodities*, entre elas o milho, para determinar, por meio de um modelo de correção de erros, uma relação entre os preços dos mercados de futuros e físicos para cada uma delas. Especificamente para o milho, os resultados obtidos revelaram que os preços futuro e *spot* não apresentaram evidências de interação. Apesar de os preços terem apresentado certa ligação durante um período da análise, o coeficiente de correlação foi próximo de zero. Os autores atribuíram os resultados ao fato de o consumo do grão no Brasil ser elevado, assim como a produção, ao passo que o mercado externo não influenciava, naquela época, significativamente a formação do preço.

Chiodi (2006) também se propôs a analisar o processo de formação dos preços do milho em nove regiões brasileiras entre o período de 1996 a 2004. O objetivo do estudo foi verificar, por meio de testes de cointegração de Johansen, se os mercados seriam perfeitamente integrados para testar a hipótese de aplicação da Lei do Preço Único. Os resultados mostraram integração perfeita entre São Paulo e Minas Gerais com quase todos os demais estados estudados. O Centro-Oeste apareceu como uma das principais regiões produtoras e exportadoras do grão, com preços perfeitamente integrados aos de Minas Gerais, Paraná e São Paulo. O Rio Grande do Sul, apareceu como forte importador de milho para atender à demanda representativa da produção de aves e suínos.

Diante do recente cenário de deslocamento da produção do milho no Brasil e da crescente importância da safra de inverno, as relações de preços entre os mercados domésticos e o mercado internacional sofreram impactos significativos. Mattos e Silveira (2015) exemplificam esta afirmação com uma análise do efeito da safra de inverno na mudança dos preços sazonais e da integração do mercado de milho brasileiro com o mercado internacional, que indicou mudanças nessas variáveis com o crescimento da safra de inverno.

---

reação de preços em relação aos outros três países. Como a medida de CS tem seus pesos ligados diretamente aos parâmetros do VECM, pode-se aceitar a validade analítica destes resultados

### 2.1.2 Limitações e comparação entre as medidas de *price discovery*

As abordagens de Hasbrouck (1995) e Gonzalo e Granger (1995) são as mais encontradas na literatura para investigar o processo de *price discovery* e, apesar de suas semelhanças iniciais (ambas partem de um modelo de correção de erros), as definições de *price discovery* em cada abordagem são diferentes.

Segundo Baillie et al. (2002), a medida de *information share* de Hasbrouck mede a parcela de contribuição de cada mercado na variância das inovações do fator comum, sendo o *price discovery* definido em termos dessa variância. Já a medida CS de Gonzalo e Granger mede a contribuição de cada mercado no fator comum e considera que somente choques de inovação no componente permanente são capazes de desequilibrar o mercado.

Brockman e Tse (1995) apresentaram o modelo IS admitindo a vantagem de que esta abordagem permite analisar a parcela de cada mercado no processo de formação de preços, sendo possível quantificar o *price discovery* e identificar a existência de mercados dominantes. Em contrapartida, a técnica possui problemas ligados à decomposição de Cholesky, que exige a ordenação dos dados, além de não apresentar um valor único para o IS (BAILLIE et al., 2002; LIEN; SHRESTHA, 2009).

Em um trabalho dedicado a comparar as duas técnicas, Jong (2002) concluiu que apenas o modelo de Hasbrouck leva em conta a variabilidade das inovações em cada mercado. Já o modelo de Gonzalo e Granger é útil no caso em que se deseja obter as inovações no componente permanente do preço eficiente, a partir da reconstrução do vetor de termo de erro. Isto se deve ao fato de que o método sugerido por Gonzalo e Granger considera apenas o peso que a inovação em um mercado  $i$  tem no incremento do preço eficiente, mas desconsidera a variância dos erros do modelo.

Apesar de serem diferentes, não há uma conclusão sobre a melhor medida para investigar o problema de *price discovery*. Baillie et al. (2002), em um trabalho no qual é feita uma comparação entre as duas medidas, admitem que não é possível afirmar que uma técnica apresenta resultados melhores que a outra, uma vez que a escolha da melhor abordagem depende de como o processo de *price discovery* será considerado na análise, sendo ele um somente um fenômeno de correção de erros ou também um processo que considera as correlações entre os mercados. Na mesma linha, Jong (2002) assume que as duas abordagens são relevantes e a opção por uma ou outra varia de acordo com o problema que se deseja investigar.

As medidas de *information share* evoluíram com o tempo e vale a pena destacar o trabalho de Lien e Shrestha (2009). Com o objetivo de obter uma medida única para a parcela de informação dos mercados, tais autores propuseram uma nova medida a qual chamaram de *Modified Information Share - MIS*. Esta medida busca contornar o problema de a matriz de covariâncias,  $\Omega$ , não ser diagonal. Os autores propuseram uma nova estrutura de fatores, no qual utilizam uma matriz de decomposição que é formada com base nas correlações entre as  $n$  séries. Os autores concluíram que a medida proposta é determinada independentemente da ordem dos

dados, porém não é necessariamente única, o que foi apontado como um problema.

A forma como as medidas IS e CS são construídas tem gerado debate acerca da interpretação de seus resultados para a medida do *price discovery*. Um dos principais motivos para isso se deve ao fato de que ambas as medidas são definidas em termos de um vetor de correção de erros na sua forma reduzida. Portanto, os erros previstos para cada mercado na formação das medidas são também definidos por termos reduzidos. Neste sentido, Lehmann (2002) afirmaram que a confusão gerada em termos da interpretação das medidas de IS e CS não é incomum, já que o modelo VEC está em sua forma reduzida e a participação de cada mercado no processo de descoberta de preços é definida por estimadores de modelos estruturais capazes de identificar, por exemplo, a origem de um determinado choque. Yan e Zivot (2010) endossam este conceito e, além disso, afirmaram que uma interpretação clara das medidas de IS e CS somente é possível se ambas partirem de modelos estruturais.

Buscando esclarecer os problemas apontados, Yan e Zivot (2010) propuseram uma medida alternativa de informação de preços, denominada por Putniņš (2013) de *information leadership share* (ILS)<sup>7</sup>. A medida de ILS consiste em uma nova medida de informação que utiliza o IS e o CS para identificar o mercado que lidera o processo de descoberta de preços.

De acordo com Yan e Zivot (2010), a medida de ILS é uma adaptação das medidas de Hasbrouck e Gonzalo e Granger, à medida que foi desenvolvida considerando um modelo estrutural permitindo que os erros do componente transitório dos preços sejam derivados de inovações estruturais. Isso implica que a forma como uma nova informação é incorporada no mercado pode afetar os resíduos do modelo.

Segundo Putniņš (2013), as pressuposições básicas por trás da abordagem do ILS são (i) de que ela apenas pode ser aplicada a modelos bivariados, uma vez que o modelo estrutural considera apenas duas fontes de choques, uma transitória e uma permanente, e (ii) que os modelos VEC reduzidos são não correlacionados. Ao buscar resultados empíricos para o ILS, Putniņš complementou a análise de Yan e Zivot e considerou algumas vantagens em seus resultados. Putniņš (2013) propôs a utilização de um o modelo estrutural em que cada série de preços possui erros vindos de fontes diferentes, ao invés de considerar que os erros das séries fossem ligados a uma única fonte de choques transitórios. Além disso, o autor assegurou a eficácia do ILS mesmo quando aplicado a modelos de formação de preços que não seguem as pressuposições iniciais de Yan e Zivot. No estudo de Putniņš, o autor considerou que as séries podem variar em dois aspectos, a quantidade de resíduos e a velocidade de ajustamento dos preços diante de uma inovação.

### 2.1.3 O comportamento dos preços diante de bolhas especulativas

O mercado internacional do milho foi muito afetado nos últimos anos, principalmente pelo aumento da demanda pelo produto por parte dos EUA, com o crescimento da produção de etanol

<sup>7</sup> Esta denominação foi proposta por Putniņš (2013) para a equação não nomeada da medida de *price discovery* desenvolvida por Yan e Zivot (2010).

à base de milho, e da China. Além disso, uma mudança estrutural no mercado de futuros de *commodities* em geral foi observada em meados da primeira década dos anos 2000 (IRWIN; SANDERS, 2012). Por esta razão, é necessário investigar se as mudanças nos preços do milho nas regiões Sul e Centro-Oeste aconteceram de forma similar, já que além de diferenças geográficas, as regiões também são diferentes na forma como negociam e estocam sua produção.

É esperado que preços em diferentes mercados se comportem de maneira similar desde que sejam determinados de acordo com os fatores fundamentais de oferta e demanda de cada região. No entanto, Brooks, Prokopczuk e Wu (2015) afirmam que não são os fatores macroeconômicos fundamentais os responsáveis por afetar os preços de maneira diferente em cada mercado, mas sim os fatores especulativos associados à mudança de preço em cada localidade. Portanto, uma bolha especulativa sistematicamente desvia o preço do ativo do seu preço de fundamento.

Uma maneira alternativa de estudar o comportamento dos preços em mercados de *commodities* é através da análise de bolhas especulativas. Stiglitz (1990) definiu o conceito de bolhas dizendo que a razão para o preço de um ativo estar alto no presente é exclusivamente porque os investidores acreditaram que seu preço de venda estaria maior no período seguinte. Ou seja, quando a alteração nos preços de um ativo não puder ser justificada pelos fundamentos que determinam tais preços, há uma bolha especulativa.

Ao identificar um período explosivo nos preços físicos ou futuros é preciso analisar cuidadosamente se a origem deste comportamento se deu a partir de uma bolha especulativa ou a partir de choques nos fundamentos. Bobenrieth, Bobenrieth e Wright (2014) mostraram que existem casos em que choques de demanda ou oferta nos mercados físicos de *commodities* armazenáveis provocam saltos nos preços que não são causados por especulações irracionais, portanto a "bolha" é explicada pelos fundamentos nestes casos.

O conceito de bolhas racionais especulativas leva em consideração que os agentes possuem expectativas racionais e parte da teoria do valor presente, na qual os preços fundamentais de uma commodity estocável são determinados por:

$$P_t = \frac{E_t(P_{t+1} + Y_t)}{(1 + r)} \quad (2)$$

em que  $P_t$  é o preço do milho no tempo  $t$ ,  $r$  é a taxa de desconto livre de risco ( $r > 0$ ),  $Y_t$  é o *convenience yield* e  $E_t$  é a expectativa formada com a informação disponível no tempo  $t$ . Considerando  $T$  períodos, pode-se representar o valor presente do milho em uma determinada região da seguinte forma:

$$P_t = \sum_{k=0}^T (1 + r)^{-(k+1)} E_t(Y_{t+k}) + (1 + r)^{-T} E_t(P_{t+T}) \quad (3)$$

Quando o termo  $(1 + r)^{-T} E_t(P_{t+T})$  tender a zero e a taxa de crescimento esperada do preço da saca de milho não for maior que  $r$ , a equação (3) pode ser representada na forma de um modelo de dividendos descontados:

$$P_t = \sum_{k=0}^{\infty} (1 + r)^{-(k+1)} E_t(Y_{t+k}) \quad (4)$$

Para que o preço representado na equação (4) seja o preço de fundamento, é necessário que o termo do valor presente esperado do *convenience yield* tenda a zero. Caso contrário, o preço no tempo  $t$  passa a considerar a presença de um componente de bolha  $B_t$  (BROOKS; PROKOPCZUK; WU, 2015):

$$P_t = P_t^* + B_t \quad (5)$$

Gutierrez (2012) fez uma análise interessante sobre o impacto das bolhas especulativas nos preços de trigo, milho, soja e arroz no período de 2007-2008, no qual houve um longo período de alta de preços das *commodities*. Além de aplicar testes de bolhas sobre os preços futuros, o autor também testou a hipótese de bolhas sobre o *convenience yield* para poder identificar se períodos de explosão de preços coincidiram como períodos de alta nos valores do *convenience yield*. Os resultados obtidos para os testes de bolhas no *convenience yield* não indicaram períodos de colapso de bolhas, diferentemente dos resultados dos testes para os preços futuros que, entre 2007 e 2008, apresentaram comportamento explosivo. O autor encontrou evidências de bolhas racionais para milho, arroz e trigo no período analisado, porém os episódios aconteceram em momentos diferentes. Os preços do milho apresentaram indícios de bolhas primeiro sendo seguidos pelos preços do arroz. Os preços da soja, por sua vez, mostraram sinais mais amenos de elevação.

No mercado de *commodities* armazenáveis, os produtores buscam definir sua produção com base no movimentação dos estoques, isto é, se desejam acumular produção ou reduzir o volume dos estoques. Neste sentido, Pindyck (2001) afirma que esta decisão é tomada com base nos preços de venda no mercado físico, no custo de armazenagem, e no *convenience yield*, que é definido como o ganho do produtor em manter seu produto em estoque no momento em que a demanda pelo produto aumenta e a oferta diminui. Em outras palavras, o *convenience yield* pode ser entendido como um pagamento que o produtor faz pelo privilégio de não precisar entregar sua mercadoria. Pindyck (2001) define que este pagamento pode ocorrer de três maneiras: *i*) o produtor tem um custo de armazenagem de seu produto; *ii*) mesmo que não hajam expectativas no mercado de que o preço da *commodity* irá mudar, existe um custo de oportunidade do capital igual à taxa de juros de se investir no ativo ao preço corrente,  $P_t$ ; *iii*) os produtores podem ter a expectativa de que o preço do ativo cairá durante o período em que pretendem armazenar o produto.

No contexto da teoria da armazenagem, os preços de contratos futuros negociados no tempo  $t$ , com vencimento em  $T$  podem ser ligados ao preço *spot* de modo que se pode obter um valor para o *convenience yield* ( $\delta_t$ ) no tempo  $t$  (BROOKS; PROKOPCZUK; WU, 2015):

$$\delta_t = r_t^f + \left(1 - \frac{F_{t,T}}{P_t}\right) \frac{252}{T-t} \quad (6)$$

em que  $F_{t,T}$  é o preço do contrato futuro negociado no tempo  $t$  com vencimento em  $T$ ,  $r_t$  é a taxa de juros livre de risco. Alguns trabalhos se dedicaram a entender e identificar períodos de bolhas em séries de preços. Diba e Grossman (1988) analisaram o comportamento dos preços de ações

de empresas norte-americanas buscando identificar a presença de bolhas especulativas racionais utilizando testes de autocorrelação e de estacionariedade. Tão os testes de autocorrelação como os testes ADF indicaram a não estacionariedade dos preços no nível, mas as séries em primeira diferença eram estacionárias, indicando que as séries eram cointegradas, ou seja, não possuíam períodos de bolhas.

No entanto, o teste ADF convencional aplicado à análise de bolhas pode levar erroneamente à conclusão de que não existem bolhas em uma série de preços, quando na verdade elas podem estar presentes. Evans (1991), por meio de diversas simulações, mostrou que comportamentos explosivos temporários só ocorrem quando bolhas econômicas colapsam periodicamente e, nestes casos, as trajetórias de preços observadas podem ser mais parecidas com um processo  $I(1)$  ou  $I(0)$ , do que uma série explosiva. Sendo assim, os testes de raiz unitária convencionais mostraram dificuldades em detectar momentos periódicos de bolhas.

Mais recentemente, Phillips, Wu e Yu (2011) (PWY) desenvolveram um teste para detectar períodos de comportamento explosivo de uma série considerando procedimentos recursivos de testes ADF de raiz unitária. Os autores desenvolveram vários testes sequenciais que testando a hipótese alternativa de raiz explosiva na série de preços pela cauda direita da distribuição, contra a hipótese nula de presença de raiz unitária. O método aplicado à cauda direita da distribuição tem uma vantagem em relação ao teste ADF tradicional, pois ele é capaz de testar as hipóteses para recortes de períodos específicos da amostra. Os autores aplicaram o teste com simulações de Monte Carlo em um amostra de preços do mercado de ações da Nasdaq para o período de fevereiro de 1973 até julho de 2005. Considerando o valor supremo dos testes ADF como o tamanho da janela de dados a serem analisados pelo teste - teste SADF - os autores encontraram períodos de bolhas explosivas racionais com início em 1995, pico em 200 e fim em algum período entre o final de 2000 e início de 2001.

Etienne, Irwin e Garcia (2014), utilizaram o teste SADF para analisar uma série de dados diários de contratos futuros de doze *commodities* dos EUA entre 1970 e 2011. Os autores encontraram momentos de comportamento explosivo nos preços de todas as *commodities*, mas eles foram pouco representativos, cerca de 2% do tempo. O estudo também detalhou as características das bolhas definindo-as como positivas ou negativas e determinou a duração média delas. Os resultados mostraram que 80-90% das bolhas duraram menos de 10 dias.

Phillips, Shi e Yu (2012) e Phillips, Shi e Yu (2015) (PSY) propuseram uma variação do teste SADF, o GASDF (SADF Generalizado) que permite que a janela de teste para a qual será realizada o teste ADF varie tanto no ponto de início quanto no ponto final da janela, enquanto o teste SADF considera que o ponto inicial é fixo a variação ocorre apenas no ponto final da janela, permitindo maior flexibilidade ao teste. Os autores aplicaram o teste GSADF em uma amostra de preços e dividendos da S&P 500 no período de 1871 a 2010 para demonstrar que os testes GSADF tem um desempenho substancialmente superior ao SADF na detecção de bolhas múltiplas, aumentando o poder do teste SADF em 11,2% para uma amostra com 400 observações, em 6,7% e em 11,5% em amostras com 800 e 1.600 observações, respectivamente.

Muito estudos foram realizados com o objetivo de detectar múltiplas bolhas nos preços dos mais diversos mercados, a partir do teste GSADF. Engsted, Hviid e Pedersen (2016), por exemplo, aplicaram o teste para o mercado imobiliário de 18 países entre 1970 e 2013. Os resultados mostraram uma quantidade grande de bolhas no início dos anos 2000 e algumas bolhas também em 1990.

Recentemente, Li et al. (2017) aplicaram o teste GSADF nos preços de trigo, milho, soja, algodão, açúcar e óleo de soja na China entre 2006 e 2014. Os resultados mostraram que, com exceção do trigo, todas as *commodities* apresentaram episódios de bolhas, mas de curta duração. Apenas 10% dos dias analisados apresentaram preços com comportamento explosivo. Os autores associaram os episódios de bolhas especulativas nos preços das *commodities* ao rápido crescimento econômico do país, ao aumento da oferta de moeda na China no período analisado e à alta taxa de inflação.

Poucos estudos exploraram a metodologia de detecção de bolhas de PSY para o Brasil. O trabalho de Chang et al. (2016) analisou os preços dos mercados de ações dos países do BRI durante 1990 e 2013. Os autores usaram os testes SADF e GSADF para detectar a presença de bolhas especulativas nos preços. No caso do Brasil os resultados mostraram a presença de bolhas no mercado de ações brasileiro entre 1992 e 1994. Os autores atribuíram este resultado parcialmente à liberalização do mercado de ações brasileiro que ocorreu em 1991.

No mercado brasileiro de *commodities*, Montasser et al. (2015), analisaram a presença de bolhas especulativas nos preços do etanol e da gasolina entre 2000 e 2012. Os autores identificaram, por meio do teste GSADF, dois períodos de bolhas explosivas nos preços do etanol: um que começou em junho de 2006 e colapsou em março de 2007, e o segundo começou em 2010, porém, segundo os autores, não havia colapsado até o final do período analisado.

### 3 METODOLOGIA

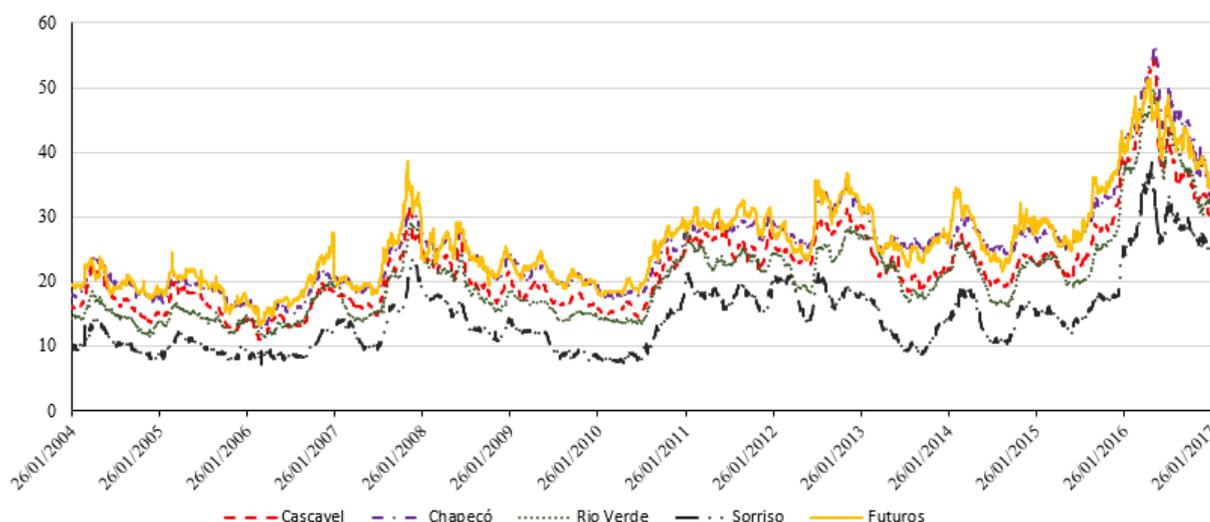
#### 3.1 Fontes e tratamentos dos dados

Os dados utilizados neste trabalho compõem uma base formada por preços diários, determinados em reais, negociados nos mercados físico de milho nas regiões Sul (Cascavel-PR e Chapecó-SC) e Centro-Oeste (Sorriso-MT e Rio Verde-GO) do Brasil e no mercado de futuros. A evolução dos preços na amostra coletada pode ser observada na Figura 2. As praças físicas escolhidas para a realização deste estudo representam grandes regiões produtoras de milho no Brasil. Assim como estas regiões, o mercado de milho de Campinas, no interior de São Paulo, também é um importante formador de preços, porém, está muito próximo da formação de preços do mercado de futuros. Por esta razão, optou-se por utilizar os preços apenas do mercado de futuros.

Os dados da amostra foram coletados junto a um terminal da Bloomberg iniciando em 26/01/2004 e terminando em 29/12/2016, compreendendo um conjunto de 3104 observações. Os vetores de preços são compostos por pares de séries e totalizam dez vetores dados por  $P_{1t} = [Pca_t \ Pch_t]'$ ,  $P_{2t} = [Pca_t \ Prv_t]'$ ,  $P_{3t} = [Pca_t \ Pso_t]'$ ,  $P_{4t} = [Pca_t \ Pft_t]'$ ,  $P_{5t} = [Pch_t \ Prv_t]'$ ,  $P_{6t} = [Pch_t \ Pso_t]'$ ,  $P_{7t} = [Pso_t \ Prv_t]'$ ,  $P_{8t} = [Pch_t \ Pft_t]'$ ,  $P_{9t} = [Prv_t \ Pft_t]'$  e  $P_{10t} = [Pso_t \ Pft_t]'$ , sendo Pca, Pch, Prv, Pso e Pft, as séries de preços de Cascavel, Chapecó, Rio Verde, Sorriso e Futuro, respectivamente.

Para a análise da existência de bolhas no período do estudo, foram coletados dados de preços de fechamento de contratos futuros de milhos negociados na Brasil Bolsa Balcão - [B]<sup>3</sup>. Os contratos de milho negociados na [B]<sup>3</sup> têm datas de vencimento trimestrais, acontecendo nos dias 15 dos meses de janeiro, março, maio, julho, setembro e novembro. A rolagem dos contratos foi feita com base no último dia do mês anterior ao mês de vencimento do contrato. Por exemplo, para um contrato que vence em 15 de março de um determinado ano, considerou-se que o contrato de março contém a maior parte da informação do mercado até o dia 28 de fevereiro. A partir de então, foram consideradas as cotações do contrato com vencimento em maio do mesmo ano.

Figura 2 – Evolução dos preços do milho em R\$/saca de 60 Kg nas praças físicas e dos preços futuros no Brasil de 2004 a 2016



Fonte: Elaboração própria com base nos dados da Bloomberg (2016) e da [B]<sup>3</sup> (2017).

De posse dos dados, foram verificadas as características das séries em relação às condições de estacionariedade, autocorrelação e heterocedasticidade.

### 3.2 Estacionariedade das séries

As ferramentas estatísticas utilizadas para a análise de dados de séries temporais podem gerar estimativas não confiáveis caso as séries utilizadas não sejam estacionárias. A presença desta característica nas séries é fundamental para que se possa fazer inferências sobre os parâmetros estimados.

Na prática, não é possível visualizar mais de um conjunto de realizações <sup>1</sup> para um determinado evento, porém, pode-se estabelecer todas as possíveis realizações por meio de inferências estatísticas caso a série observada seja estacionária. Conforme apresentado por Enders (2014), uma série temporal é considerada estacionária quando sua média, variância e autocorrelação podem ser aproximadas por valores médios no longo prazo. Tais valores, sendo originados de observações de um único conjunto de realizações suficientemente longo, garantem que estas medidas sejam as mesmas em todo o período. Isso é possível, uma vez que uma série estacionária tende a dissipar efeitos de um choque, levando-a ao seu nível médio ao longo do tempo. As características fundamentais para que uma série seja estacionária são: *i*) a média de longo prazo é finita e invariante ao longo do tempo; *ii*) a variância é constante e invariante ao longo do tempo; e *iii*) a autocorrelação é independente do tempo, mas pode variar dependendo das defasagens entre as observações.

De modo geral, uma série é estacionária se sua média, variância e autocovariância permanecem as mesmas, independentemente do número de defasagens usado, e do período em que são

<sup>1</sup> Entende-se por realizações de um processo estocástico os elementos observados em uma série de tempo.

medidas.

No caso de uma série não possuir alguma dessas características, ela é considerada não estacionária, seguindo portanto, um processo integrado de ordem  $d - I(d)$  -, em que  $d$  é o número de diferenciações necessárias para que a série se torne estacionária, ou seja, siga um processo  $I(0)$ .

Para verificar a estacionariedade das séries utilizam-se os chamados testes de raiz unitária. Alguns deles são bastante difundidos da literatura como o teste de Dickey-Fuller Aumentado (ADF), o teste de Phillips-Perron (PP) e o KPSS<sup>2</sup>. Este trabalho utiliza o teste de Dickey-Fuller Aumentado proposto por Dickey e Fuller (1979, 1981) que considera a possibilidade do termo de erro  $u_t$  ser autocorrelacionado, sendo necessário a inclusão de tantas variáveis defasadas em primeira diferença quanto forem necessárias para que o erro passe a ser um ruído branco. Tomando-se um modelo simplificado, como sugerido por (ENDERS, 2014):

$$P_t = a_0 + a_1 P_{t-1} + \varepsilon_t \quad (7)$$

Subtraindo  $P_{t-1}$  de cada lado, tem-se:

$$\Delta P_t = a_0 + \gamma P_{t-1} + \varepsilon_t \quad (8)$$

em que  $a_0$  é o termo constante e  $\gamma = a_1 - 1$ . Substituindo a equação 10 por um processo autorregressivo de ordem ( $p$ ), obtém-se o seguinte modelo geral:

$$\Delta P_t = a_0 + a_2 t + \gamma P_{t-1} + \sum_{i=2}^p \beta_i \Delta P_{t-i+1} + \varepsilon_t \quad (9)$$

A proposta do teste é estimar o modelo introduzindo variáveis autoregressivas até que  $\varepsilon_t$  seja um ruído branco. Em outras palavras, deseja-se encontrar um valor para  $\gamma \neq 0$ , uma vez que se  $\gamma = 0$  não é possível rejeitar a hipótese nula do teste, que é a presença de uma raiz unitária, ou seja, de que a série é não estacionária. Nota-se a presença de um termo contante e uma tendência ( $a_2 t$ ) que devem apresentar significância estatística para serem considerados no modelo, caso contrário devem ser excluídos sem prejudicar o resultado do teste.

Para identificar a melhor especificação do modelo para as séries de preços usadas neste estudo, foi realizado o teste proposto por Dickey e Fuller (1981), que desenvolveram estatísticas para testar hipóteses conjuntas dos coeficientes buscando determinar o modelo mais apropriado para realização do teste. Os autores denominaram de  $\phi_1$ ,  $\phi_2$  e  $\phi_3$  as estatísticas F que testam as hipóteses nulas de que  $\gamma = a_0 = 0$ ,  $\gamma = a_0 = a_2 = 0$  e  $\gamma = a_2 = 0$ , respectivamente.

Os valores das estatísticas  $\phi$  para cada série do vetor  $P_t = [Pca_t \ Pch_t \ Prv_t \ Pso_t \ Pft_t]'$  podem ser obtidos a partir da análise conjunta dos seguintes modelos, por exemplo, para a série de preços da praça de Cascavel (ENDERS, 2014):

$$\Delta Pca_t = \gamma Pca_t + \sum_{i=2}^p \beta_i \Delta Pca_{t-i+1} + \varepsilon_t \quad (10)$$

<sup>2</sup> Mnemônico de Kwiatkowski, Phillips, Schmidt e Shin, autores do teste.

$$\Delta Pca_t = a_0 + \gamma Pca_t + \sum_{i=2}^p \beta_i \Delta Pca_{t-i+1} + \varepsilon_t \quad (11)$$

$$\Delta Pca_t = a_0 + a_2 t + \gamma Pca_t + \sum_{i=2}^p \beta_i \Delta Pca_{t-i+1} + \varepsilon_t \quad (12)$$

em que  $\Delta Pca_{t-i+1}$  são termos diferenciados defasados da variável dependente e  $\varepsilon_t$  é um ruído branco.

As estatísticas  $\phi$  são construídas de modo que permitam testar hipóteses conjuntas entre os três possíveis processos regressivos: com um termo constante  $a_0$  e uma tendência linear  $a_2 t$  - equação (12); com um termo constante e sem tendência - equação (11); e, por fim, um modelo sem contante e sem tendência - equação (10). Os valores críticos do teste  $\phi$  seguem distribuição F, e podem ser calculados de acordo com a seguinte equação:

$$\phi_i = \frac{[SQR_{(restrito)} - SQR_{(irrestrito)}]/r}{SQR_{(irrestrito)}/(T - k)} \quad (13)$$

onde  $SQR_{(restrito)}$  e  $SQR_{(irrestrito)}$  são as somas dos quadrados dos resíduos dos modelos restritos e irrestritos, respectivamente, sendo o modelo mais restrito aquele sem constante e sem tendência, e o mais irrestrito aquele com constante e com tendência. O número de restrições  $r$  é igual a dois ou a três, dependendo de qual estatística  $\phi$  estiver sendo calculada. Por fim,  $T$  é o número de observações da amostra, e  $k$  é o número de parâmetros estimados nos modelos. Os valores críticos podem ser encontrados em Dickey e Fuller (1979).

No presente trabalho, o teste ADF considerando as estatísticas  $\phi$  levaram a conclusão de que as séries são todas  $I(1)$ , utilizando-se a especificação sem o termo contante e sem tendência linear para as cinco séries contidas no vetor  $P_t$ :

$$\Delta P_t = \gamma P_{t-1} + \sum_{i=2}^p \beta_i \Delta P_{t-i+1} + \varepsilon_t \quad (14)$$

Para melhorar a confiabilidade dos resultados do teste ADF, é importante determinar corretamente a ordem  $p$  do processo autoregressivo, uma vez que o excesso de defasagens reduz o número de graus de liberdade, diminuindo o poder do teste. Para isso, utilizaram-se os critérios de informação de Akaike (AIC) e o Bayesiano de Schwarz (BIC)<sup>3</sup>. A saber, Bueno (2008) apresenta as definições de AIC e BIC como:

$$AIC = \ln \hat{\sigma}^2 + n \frac{2}{T} \quad (15)$$

$$BIC = \ln \hat{\sigma}^2 + n \frac{\ln T}{T} \quad (16)$$

<sup>3</sup> Outro caminho para determinar o número correto de defasagens é executar testes nos resíduos como o teste F, o teste  $t$  ou o teste de Ljung-Box, para checar se estes passaram a apresentar características de ruído branco após a inclusão das variáveis defasadas (ENDERS, 2014).

em que,  $T$  é o número de observações da amostra,  $n$  é o número de parâmetros a serem estimados, e  $\hat{\sigma}^2$  é a variância estimada dos resíduos definida por  $\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T}$ . Deseja-se obter o menor valor possível de AIC ou BIC. Os valores críticos para a tomada de decisão do teste ADF são apresentados em Dickey e Fuller (1979). O critério escolhido neste estudo foi o de Schwarz.

Após especificado o melhor modelo por meio do teste ADF, a amostra de dados utilizadas neste trabalho também foi testada em relação à autocorrelação e heterocedasticidade dos resíduos. A análise de autocorrelação foi feita neste por meio da análise do correlograma de ambas as séries. O procedimento de correção da autocorrelação para ambas as séries foi a aplicação de um grau de diferenciação, sendo este suficiente para corrigir o problema.

A heterocedasticidade das séries foi testada utilizando os procedimentos sugeridos por Breusch e Pagan (1979) e White (1980), que são amplamente utilizados na literatura para testar a hipótese de ausência de heterocedasticidade.

Para testar a presença de quebra estrutural nas séries de preços a partir de uma data desconhecida, foi realizado o teste de quebra estrutural proposto por Zeileis et al. (2003)<sup>4</sup> que parte de um modelo de regressão linear simples como:

$$y_i = x_i^T \beta_j + u_i (i = i_{j-1} + 1, \dots, i_j, j = 1, \dots, m + 1) \quad (17)$$

O teste consiste em estimar consecutivas regressões com  $m + 1$  segmentos de tamanhos  $I_{m,n} = i_1, \dots, i_m$ , com  $i_0$  representando o conjunto de pontos de quebras. A variável dependente  $x_i$  é um vetor de uns de tamanho  $k \times 1$  que permite testar mudanças na média dessa variável. A hipótese nula do teste assume que os coeficiente das regressão não se alteram:  $H_0: \beta_1 = \beta_0 (i = 1, \dots, n)$ , enquanto a hipótese alternativa considera que pelo menos um coeficiente muda ao longo do tempo.

### 3.3 Análise de cointegração e modelo de correção de erros

A definição de cointegração desenvolvida por Engle e Granger (1987) considera que duas séries são cointegradas se forem não estacionárias e integradas de mesma ordem.

Em adição aos testes de raiz unitária, Hatemi-J e Irandoust (2005), destacam a importância de prosseguir com os testes de cointegração em uma análise temporal, uma vez que é importante estabelecer a existência de uma relação de longo-prazo entre as séries analisadas, para assim evitar estimações espúrias. Após a identificação de não estacionariedade de duas séries, é preciso determinar se as séries apresentam alguma relação de longo prazo entre si.

Segundo Greene (2003), existem duas formas de se testar a cointegração de séries temporais, a proposta por Engle e Granger (1987), e a abordada por Johansen e Juselius (1990), que foi utilizada no presente estudo. O teste de cointegração de Johansen é amplamente difundido na literatura e permite a inferência do número de relações de cointegração entre as trajetórias das

<sup>4</sup> O teste foi escolhido para facilitar as discussões dos resultados, uma vez que a data da quebra coincidiu com a data em que a produção na safra de inverno estava superando a produção na safra.

variáveis, bem como determina a estrutura dessas relações. Inicialmente, parte-se de um modelo de vetores autorregressivos (VAR) na forma genérica, com  $k$  defasagens e sem o termo constante tal como (JOHANSEN; JUSELIUS, 1990):

$$P_t = \Gamma_1 P_{t-1} + \dots + \Gamma_k P_{t-k} + \varepsilon_t \quad (t = 1, 2, \dots, T) \quad (18)$$

em que  $\varepsilon_t, \dots, \varepsilon_T$  seguem distribuição normal e  $P_t$  é o vetor de preços ( $2 \times 1$ ). Verificando-se a presença de, no máximo um vetor de cointegração, o modelo VAR genérico deve ser diferenciado e deve-se incluir um vetor de correção de erros (VEC). Assim, reescrevendo a equação (18), tem-se o modelo VEC (VECM) na seguinte forma:

$$\Delta P_t = \Pi P_{t-1} + \Gamma_1 P_{t-1} + \dots + \Gamma_k P_{t-k+1} + \Gamma_t \quad (19)$$

em que  $P_t$  é um vetor ( $2 \times 1$ ),  $\Pi$  é uma matriz ( $2 \times 2$ ) contendo coeficientes que determinam a velocidade com que o modelo se ajusta no curto prazo e pode ser definida como  $\Pi = \alpha\beta'$ , em que  $\alpha$  é uma matriz ( $2 \times r$ ) contendo os coeficientes relativos à velocidade de ajustamento do modelo e  $\beta$  é uma matriz ( $2 \times r$ ) que contém os vetores de cointegração. A matriz  $\Gamma_k$  é formada pelos parâmetros estimados que representam as variáveis defasadas do modelo.

O número de vetores de cointegração é igual ao posto da matriz  $\Pi$ , e também correspondem ao número de autovalores (raízes características) diferentes de zero desta mesma matriz. Um vez determinados os autovalores  $\lambda_1, \dots, \lambda_n$  da matriz  $\Pi$  a partir da maximização do modelo VECM, deve-se verificar se os mesmos são estatisticamente diferentes de zero. Para isso, podem ser realizados os testes do traço e do máximo autovalor propostos por Johansen e Juselius (1990). A hipótese nula do teste do traço é que a quantidade de vetores de cointegração ( $r$ ) é menor ou igual ao número de raízes características da matriz  $\Pi$ , ( $H_0 : r \leq r^*$ ), contra a hipótese alternativa de que  $r > r^*$  (ENDERS, 2014):

$$\lambda_{tr}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (20)$$

em que  $\hat{\lambda}_i$  representa os autovalores estimados da matriz  $\Pi$ ,  $T$  é o número de observações e  $n$  é o número de variáveis analisadas. Como serão analisadas as relações temporais entre pares de séries, considera-se  $n = 2$ .

O teste do máximo autovalor, por sua vez, testa a hipótese nula de que o número de vetores cointegrados é igual a  $r^*$ , contra a hipótese alternativa da existência de  $r^* + 1$  vetores cointegrados. A estatística do teste é definida como (ENDERS, 2014):

$$\lambda_{max}(r, r + 1) = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (21)$$

Tanto no teste do traço como no teste do máximo autovalor, rejeitar  $H_0$  significa que há mais de um vetor de cointegração, enquanto a não rejeição de  $H_0$  significa a existência de  $r^*$  vetores de cointegração. Os valores críticos para os testes são apresentados por Johansen e Juselius (1990).

Enders (2014), reforça o ponto já mencionado de que, a estimação do modelo VAR em primeiras diferenças quando existe cointegração, faz com que muitas informações contidas nas relações de longo prazo sejam perdidas no processo de diferenciação. Portanto, é preferível estimar o VAR em primeira diferença para o caso de variáveis  $I(1)$  e não cointegradas. Caso seja detectada a presença de cointegração, seria mais apropriado a utilização de um modelo de correção de erros.

Assim sendo, se o teste de Johansen indicar a presença de um vetor de cointegração, um modelo VEC será estimado.

### 3.4 Estratégias para a estimação de *price discovery*

Ativos financeiros negociados em muitos mercados, como o caso do milho no Brasil, têm seus preços em um dado mercado determinado a partir de novas informações absorvidas em um ou mais mercados relacionados. Este trabalho visa analisar o *price discovery*, considerando a abordagem do *information share*, desenvolvida por Hasbrouck (1995), do *component share*, desenvolvida por Gonzalo e Granger (1995), e do *information leadership*, proposto por Yan e Zivot (2010). A próxima subseção apresenta brevemente estas abordagens.

#### 3.4.1 Medidas de informação entre mercados

A medida do *information share* parte do conceito de cointegração, já que séries de preços em mercados cointegrados apresentam um fator comum implícito no preço eficiente. O ponto de partida é a estimação de um modelo VECM, uma vez que as medidas de *information share* são calculadas a partir dos coeficientes estimados nesse modelo, e de sua matriz de variância-covariância  $\Omega$ . Considera-se um vetor de preços,  $P_t$ , de uma mesma *commodity*, composto por duas séries (dois mercados) integradas de ordem 1, cujas variações nos preços possuem covariâncias estacionárias.

Como os cálculos das medidas de IS são sempre obtidos a partir da análise das variáveis em pares, foram estimados diversos modelos VECM utilizando-se os preços em dois mercados diferentes ( $i$  e  $j$ ). Por exemplo, para a análise de Cascavel ( $ca$ ) e Chapecó ( $ch$ ), Cascavel e Rio Verde ( $rv$ ) e Cascavel e Sorriso ( $so$ ), os modelos VEC estimado foram:

$$\begin{aligned} \begin{pmatrix} \Delta Pca_t \\ \Delta Pch_t \end{pmatrix} &= \alpha_1 \beta_1' \begin{pmatrix} Pca_{t-1} \\ Pch_{t-1} \end{pmatrix} + \sum_{k=1}^2 \Gamma_{1,k} \begin{pmatrix} \Delta Pca_{t-k} \\ \Delta Pch_{t-k} \end{pmatrix} + \varepsilon_{1,t} \\ \begin{pmatrix} \Delta Pca_t \\ \Delta Prv_t \end{pmatrix} &= \alpha_2 \beta_2' \begin{pmatrix} Pca_{t-1} \\ Prv_{t-1} \end{pmatrix} + \sum_{k=1}^2 \Gamma_{2,k} \begin{pmatrix} \Delta Pca_{t-k} \\ \Delta Prv_{t-k} \end{pmatrix} + \varepsilon_{2,t} \\ \begin{pmatrix} \Delta Pca_t \\ \Delta Pso_t \end{pmatrix} &= \alpha_3 \beta_3' \begin{pmatrix} Pca_{t-1} \\ Pso_{t-1} \end{pmatrix} + \sum_{k=1}^2 \Gamma_{3,k} \begin{pmatrix} \Delta Pca_{t-k} \\ \Delta Pso_{t-k} \end{pmatrix} + \varepsilon_{3,t} \end{aligned} \quad (22)$$

onde,  $\Delta P(i)_t = P(i)_t - P(i)_{t-1}$  representa a variação dos preços do milho em cada mercado ( $i$ ) analisado,  $\alpha_1 \beta_1'$ ,  $\alpha_2 \beta_2'$  e  $\alpha_3 \beta_3'$  são as matrizes  $\Pi$  de cointegração ( $2 \times 1$ ) de cada modelo,

onde cada  $\alpha$  representa uma matriz de correção de erros de dimensão  $(2 \times 1)$  e  $\beta_i$  uma matriz de vetores de cointegração de dimensão  $(1 \times 2)$ ,  $\Gamma_{1,i}$ ,  $\Gamma_{2,i}$  e  $\Gamma_{3,i}$  são as matrizes  $(2 \times 2)$  de coeficientes estimados na parte autoregressiva de cada modelo e  $\varepsilon_{1,t}$ ,  $\varepsilon_{2,t}$  e  $\varepsilon_{3,t}$  são vetores de erros de dimensão  $(2 \times 1)$ .

Hasbrouck (1995) propôs transformar os VECM em vetores de médias móveis, tais como:

$$\Delta P_t = \Psi(L)e_t \quad (23)$$

em que  $P_t$  é um vetor  $(2 \times 1)$  com pares das variáveis de preço,  $\Psi$  é um polinômio do operador de defasagem  $L$ .

Desta forma, pode-se representar a relação de cointegração de duas séries por  $\beta' p_t$  e, como elas são estacionárias na forma de diferenças, tem-se que  $\beta' \Psi(1) = 0$ . Como  $\Psi(1)$  corresponde à soma dos coeficientes de médias móveis, este vetor representa o impacto no longo prazo de uma perturbação em cada um dos preços. Além disso,  $\Psi(1)$  possui todas as linhas idênticas por pressuposição, o que implica que os impactos de longo prazo são os mesmos para todos os preços. Nessa abordagem, o termo  $\Psi e_t$  representa um componente de variação de preços que pode ser reflexo de novas informações e possui variância dada por  $\Psi \Omega \Psi'$ . Sendo  $\psi$  a representação da linha comum em  $\Psi(1)$ , a tendência comum no tempo  $t$  implica que os níveis de preços são definidos por:

$$P_t = P_0 + \psi \left( \sum_{s=1}^t e_s \right) \iota + \Psi^*(L)e_t \quad (24)$$

onde  $P_0 = (P_{01}, P_{02})$  é um vetor constante que contém valores iniciais e pode acusar a existência de diferenças não estocásticas entre as variáveis, e  $\Psi^*(L)$  é uma matriz polinomial do operador de desasagem. O segundo termo à direita na equação (24) representa o produto entre um escalar de passeio aleatório e um vetor coluna unitário ( $\iota$ ), ou seja, é o componente comum a todos os preços. O terceiro termo, representa um processo estacionário de covariância de média zero. A equação (24) está fortemente relacionada à representação de fatores comuns em mercados cointegrados apresentado por Stock e Watson (1988):

$$P_t = f_t + P_t^s \quad (25)$$

em que  $P_t$  é a soma de um componente permanente  $f_t$ , o fator comum, e um transitório  $P_t^s$ .

De acordo com Hasbrouck (1995), o incremento  $\psi e_t$  na equação (24) é o componente da variação de preços que é permanentemente extraído de dentro do preço como um reflexo das novas informações e sua variância é dada por  $\psi \Omega \psi'$ .

No caso da matriz de covariância ser diagonal, e sabendo que  $e_{j,t}$  corresponde à inovação do mercado  $j$  no tempo  $t$ , as inovações em um mesmo mercado não serão correlacionadas e  $\psi \Omega \psi'$  será composta por  $n$  termos que mostrarão, cada um, a contribuição de cada mercado nas inovações dos preços. Assim, o *information share* apresentado por Hasbrouck (1995) define a parcela de informação do mercado  $j$  como:

$$IS_j = \frac{\psi_j^2 \sigma_{jj}}{\psi \Omega \psi'} \quad (26)$$

onde  $\psi_j$  é o j-ésimo elemento da matriz  $\psi$  referente a soma dos coeficientes de média móvel das séries de preço,  $\sigma_{jj}$  é o j-ésimo termo da matriz de covariância  $\Omega$ . Como em muitos casos as inovações sobre os preços tem correlação entre os mercados, Hasbrouck (1995) considera que deve ser feita uma fatoração de Cholesky sobre a matriz  $\Omega$  (não diagonal), gerando uma nova matriz  $M$  triangular inferior. A partir desta decomposição é possível calcular o *information share* de cada mercado através da relação a seguir:

$$IS_j = \frac{[(\psi M)_j]^2}{\psi \Omega \psi'} \quad (27)$$

em que  $(\psi M)_j$  o j-ésimo elemento da matriz  $\psi F$ , em que  $F$  é a decomposição do Cholesky de  $\Omega$ . A decomposição de Cholesky determina uma ordem dos dados de forma que seja feita uma maximização da parcela de informação para o primeiro preço, e uma minimização desta parcela para o segundo preço. Dessa forma, o resultado gera um intervalo para a medida de IS.

A abordagem da existência de um fator comum ( $f_t$ ) aos preços, que não é observada e pode ser a causa da cointegração em mercados interligados, foi apresentada por Gonzalo e Granger (1995). Tais autores sugerem a obtenção de  $f_t$  por meio de uma combinação linear das variáveis originais  $P_t$  e um modelo de correção de erros, como o de Hasbrouck.

O método de Gonzalo e Granger permite analisar os comportamentos da tendência e transitório das séries separadamente. Nesta abordagem, a decomposição do fator comum  $f_t$  de dimensão  $(k \times 1)$  é feita a partir da combinação linear dos preços, sendo:

$$f_t = \eta P_t \quad (28)$$

em que  $\eta$  é um vetor de dimensão  $(k \times p)$  contendo os coeficientes de fator comum,  $\eta = (\eta_1, \eta_2)$ , neste estudo. O objetivo é determinar o número  $k$  de fatores comuns. Neste caso, a condição de linearidade apresentada na equação (28) ajuda no processo, pois determina uma relação entre os fatores comuns e a variáveis observadas, o que é muito importante para a análise de fatores (GONZALO; GRANGER, 1995).

Os elementos de  $P_t$  podem ser explicados em termos de  $f_t$  mais alguns componentes estacionários como:

$$P_t = A_1 f_t + \tilde{P}_t \quad (29)$$

onde  $A_1 f_t$  representa a parcela do componente permanente e  $\tilde{P}_t$  a parcela do componente transitório de  $P_t$ , seguindo a definição de permanente-transitório (PT)<sup>5</sup>. Dessa forma, Gonzalo e Granger mostram que apenas inovações vindas do componente permanente podem afetar a variável  $P_t$  no longo-prazo.

Buscando evitar problemas de interpretação acerca dos resultados gerados pela medidas de IS e CS, esta pesquisa apresenta também resultados para as medidas de *information leadership* (IL) e sua representação em termos percentuais (ILS). Como já mencionado, esta medida proposta por Yan e Zivot (2010) utiliza os resultados da estimação de um modelo de correção de erros

<sup>5</sup> Gonzalo e Granger (1995) definem PT seguindo parte da definição dada por Quah(1992).

em sua forma estrutural, o que permite a existência de apenas um componente transitório e um componente permanente.

As medidas de CS são obtidas a partir do vetor de correção de erros normalizado,  $\alpha_{\perp} = (\gamma_1, \gamma_2)'$ . Dessa forma, as medidas de CS para os mercados 1 e 2 são (BAILLIE et al., 2002):

$$CS_1 = \gamma_1 = \frac{\alpha_2}{\alpha_2 - \alpha_1}, \quad CS_2 = \gamma_2 = \frac{\alpha_1}{\alpha_1 - \alpha_2} \quad (30)$$

A matriz de variância-covariância,  $\Omega$ , dos termos de erro do VECM em sua forma reduzida é a mesma encontrada para determinar o IS de Hasbrouck. O procedimento também se repete em relação à decomposição de Cholesky da matriz,  $\Omega$ , para eliminar os efeitos de correlação contemporânea, de modo que  $\Omega = MM'$ , sendo:

$$M = \begin{pmatrix} m_{11} & 0 \\ m_{12} & m_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1 - \rho^2)^{1/2} \end{pmatrix}$$

onde  $M$  é uma matriz triangular inferior e  $\rho$  o coeficiente de correlação cruzada. Portanto, os valores de IS para os mercados 1 e 2 serão, respectivamente:

$$\frac{IS_1}{IS_2} = \frac{(\gamma_1 m_{11} + \gamma_2 m_{12})^2}{(\gamma_2 m_{22})^2} \quad (31)$$

A fatoração que resulta na equação (31) implica que o maior *information share* será o da primeira variável na decomposição, a menos que  $m_{12} = 0$ . Uma vez que a ordenação das variáveis no modelo VECM é determinante na obtenção dos resultados do IS<sup>6</sup>, o limite superior do *information share* do mercado  $j$  é obtido quando os preços deste mercado são os primeiros na ordenação da decomposição.

De posse das medidas de IS e CS, seguiu-se o procedimento de Yan e Zivot (2010) para determinar o IL e o ILS da seguinte forma:

$$IL_1 = \left| \frac{IS_1 CS_2}{IS_2 CS_1} \right|, \quad IL_2 = \left| \frac{IS_2 CS_1}{IS_1 CS_2} \right| \quad (32)$$

Diferentemente das medidas de Hasbrouck (1995) e Gonzalo e Granger (1995), o *information leadership* não soma 1. Na verdade, ele pode ser representado no intervalo  $[0, \infty[$ , sendo que valores de  $IL_1$  acima de 1 indicam que o mercado 1 lidera o processo de absorção de novas informação em relação aos valores fundamentais. Valores de IL menores que 1 indicam que o mercado é liderado no processo de formação de preços. Buscando deixar o IL comparável ao IS e ao CS, Putniņš (2013) sugere a medida de *information leadership share*:

$$ILS_1 = \frac{IL_1}{IL_1 + IL_2}, \quad ILS_2 = \frac{IL_2}{IL_1 + IL_2} \quad (33)$$

<sup>6</sup> Baillie et al. (2012) sugerem uma medida de IS que consideram mais simples que a tradicional de Hasbrouck, uma vez que as medidas propostas podem ser calculadas direto de um VECM, enquanto as tradicionais medidas de Hasbrouck partem de uma representação do VECM na forma de um vetor de médias móveis. Considerando o problema da ordenação das séries, os autores ainda sugerem que as duas medidas de IS devem ser definidas levando em conta as possibilidade de ordenação e dividindo os valores pela média simples das duas medidas.

Com esta normalização, o  $ILS_1$  e o  $ILS_2$  são calculados no intervalo  $[0,1]$ , seguindo a mesma interpretação das medidas tradicionais de IS e CS.

Para estabelecer uma forma de agrupar os preços de diferentes mercados em pares, seguiu-se um procedimento similar a um torneio, conforme sugerido por Kim, Tejeda e Wright (2016). Um agrupamento hierárquico dos preços foi realizado a partir da construção de um dendograma. A análise de agrupamento parte do princípio de que cada observação em uma amostra multivariada é um ponto no espaço euclidiano multidimensional e, por meio do agrupamento, é possível unir os pontos que representem características marcantes. Existem diversos métodos de agrupamento de dados, dentre eles o particional e o hierárquico. O dendograma é a representação gráfica do agrupamento hierárquico de características semelhantes ou divergentes na amostra (METZ, 2006).

A construção do dendograma está relacionada ao conceito de distância (dissimilaridade) entre as variáveis. Desta forma, quanto maiores os valores observados, menos parecidos são os objetos. O dendograma construído neste trabalho utilizou as informações de preços para determinar as diferenças entre as praças estudadas.

Para identificar os agrupamentos do dendograma utilizou-se a distância euclidiana entre as variáveis. A partir da identificação de variáveis semelhantes, definiu-se o método de ligação simples (vizinho mais próximo) que consiste em juntar dois grupos com menor distância, isto é, mais semelhantes. A partir deste resultado um terceiro grupo é criado e assim por diante.

### 3.5 Procedimento para detecção de bolhas nos preços

Neste trabalho, foi utilizado o teste GSADF para identificar os fatores especulativos que podem ter afetado os preços do milho no período de 2004 a 2016, provocando episódios de preços explosivos. O teste GSADF foi apresentado por Phillips, Shi e Yu (2015) - PSY daqui em diante - e é uma variação do teste ADF usando testes recursivos e de janelas móveis (*rolling window*). Segundo os autores, este teste tem um poder maior quando aplicado à amostra total em comparação a outros procedimentos usados para detectar bolhas. Outra vantagem do teste GSADF é que ele permite não apenas identificar a presença de bolhas nas séries de preços, mas também é capaz de indicar o momento de início e de fim das mesmas.

Segundo Phillips, Shi e Yu (2015), o teste GSADF testa a hipótese nula de que  $P_t$  segue um passeio aleatório com um termo constante assintoticamente negligenciável:

$$P_t = dT^\eta + \theta P_{t-1} + \varepsilon_t, \quad \theta = 1. \quad (34)$$

em que  $d$  é o termo constante,  $T$  é o tamanho da amostra,  $\eta$  é um coeficiente de localização que controla a magnitude do intercepto à medida em que  $T$  aumenta e  $\varepsilon_t \sim iidN(0, \sigma^2)$ .

O trabalho de PSY tem foco no caso em que  $\eta > 1/2$  em que  $P_t$  tem a mesma ordem de grandeza do de um passeio aleatório puro. Portanto, a hipótese nula do teste GSADF é que  $P_t$  segue um passeio aleatório com um desvio assintótico negligenciável, ao passo que a hipótese alternativa é a existência de bolhas nos preços.

De acordo com Phillips, Shi e Yu (2015), o teste GSADF é uma variação da forma reduzida da equação empírica:

$$P_t = \alpha + \gamma P_{t-1} + \sum_{i=1}^k \beta_i \Delta P_{t-i} + \varepsilon_t \quad (35)$$

em que  $P_t$  é o preço,  $\alpha$  é o intercepto,  $k$  é o número de lags e  $\beta_i$  para  $i = 1, 2, \dots, k$  são os coeficientes relacionados aos termos defasados  $\Delta P_{t-i}$ , e  $\varepsilon_t$  é o termo de erro. As hipóteses testadas com o teste GSADF são  $H_0 : \gamma = 1$  e  $H_1 : \gamma > 1$ .

Etienne, Irwin e Garcia (2014) apresentam uma ilustração para o caso de dois períodos de bolhas:  $B_1 = [\tau_{1i}, \tau_{1f}]$  e  $B_2 = [\tau_{2i}, \tau_{2f}]$  onde  $\tau_{1i}, \tau_{1f}, \tau_{2i}$  e  $\tau_{2f}$  são os pontos de início e final de cada período explosivo. O processo gerador desses dados pode ser definido como:

$$P_t = P_{t-1} 1\{t \in N_0\} + \delta_T P_{t-1} 1\{t \in B_1 \cup B_2\} + \left( \sum_{k=\tau_{1f}+1}^t \varepsilon_k + P_{\tau_{1f}}^* \right) 1\{t \in N_1\} + \left( \sum_{k=\tau_{2f}+1}^t \varepsilon_k + P_{\tau_{2f}}^* \right) 1\{t \in N_2\}, \quad (36)$$

$$t = 1, \dots, T; \delta_T = 1 + c/T^\alpha; c > 0; \alpha \in (0, 1)$$

onde  $1\{\cdot\}$  é uma função indicadora tal qual  $1\{\cdot\} = 1$  e  $N_0 = [1, \tau_{1i})$ ,  $N_1 = (\tau_{1f}, \tau_{2i})$ ,  $N_2 = (\tau_{2f}, T]$  são três subperíodos não explosivos. Nas datas em que ocorrem os colapsos das bolhas,  $\tau_{1f}$  e  $\tau_{2f}$ , os preços sofrem um salto atingindo o nível de  $P_{\tau_{1f}}^*$  e  $P_{\tau_{2f}}^*$ , respectivamente. Por fim, considerando que  $\alpha$  e  $c$  são contantes, o parâmetro  $\delta_T > 1$  em amostras finitas e se aproxima de um à medida que  $T \rightarrow \infty$ . Os autores ainda chamam a atenção para o fato de que o procedimento de detecção de bolhas em múltiplos períodos deve ser capaz de distinguir um momento explosivo da série em  $\tau_{1i}$  de um momento não explosivo em  $\tau_{1i-1}$ , bem como deve identificar o momento em  $\tau_{1f}$  é uma transição de um período explosivo para um passeio aleatório.

Para fazer esta distinção, o procedimento do teste GSADF proposto por Phillips, Shi e Yu (2015) considera a realização de estimações repetidas e recursivas de um teste ADF em subamostras de dados. Estas subamostras são janelas móveis flexíveis que variam de acordo com o tamanho da amostra. Por simplificação, o teste utiliza um período de tempo normalizado (isto é, um intervalo de tempo é medido como proporção do total da amostra) e  $\gamma_{r_1, r_2}$  e  $ADF_{r_1, r_2}$  correspondem à estatística ADF sobre a amostra normalizada  $[r_1 r_2]$ , em que  $r_1$  e  $r_2$  são as primeiras e as últimas observações da janela, respectivamente. Adicionalmente, denota-se por  $r_w$  o tamanho da janela de regressão, sendo que  $r_w = r_2 - r_1$ , e  $r_0$  é a janela inicial determinada pelo pesquisador.

Após a definição dos pontos de início e fim,  $r_1$  e  $r_2$ , a equação a ser estimada é:

$$\Delta P_t = \alpha_{r_1, r_2} + \gamma_{r_1, r_2} P_{t-1} + \sum_{i=1}^k \beta_{r_1, r_2}^i \Delta P_{t-i} + \varepsilon_t \quad (37)$$

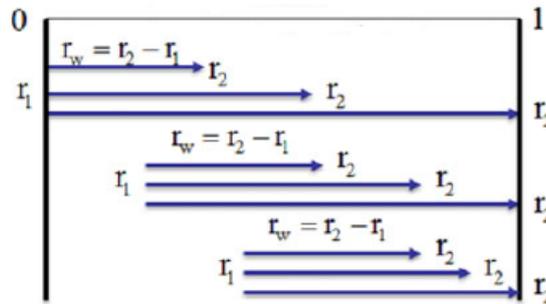
em que,  $\Delta P_t = P_t - P_{t-1}$ ,  $k$  é o número de *lags* e  $\varepsilon_t \sim iidN(0, \sigma^2)$ .

No teste GSADF, além de variar o ponto final,  $r_2$  de  $r_0$  até 1, ele também permite a variação do ponto inicial,  $r_1$  em (37) dentro de um limite que vai de 0 a  $r_2 - r_0$ . A estatística do teste GSADF é a maior estatística ADF dentro de todas as extensões factíveis de  $r_1$  e  $r_2$  e pode ser definida como:

$$GSADF(r_0) = \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} [ADF_{r_1}^{r_2}] \quad (38)$$

A Figura 3 ilustra o procedimento do teste GSADF apresentado na Figura 3.

Figura 3 – Ilustração do procedimento do teste GSADF em uma amostra normalizada.



Fonte: Phillips, Shi e Yu (2015).

Segundo Phillips, Shi e Yu (2015), se o tamanho da amostra ( $T$ ) for muito grande, deve ser escolhido um valor muito pequeno para  $r_0$ , de modo que não se perca nenhuma possibilidade de detectar uma bolha anterior. Os autores recomendam que  $r_0 = 0,01 + 1,8/\sqrt{T}$ , por facilidade de implementação e baseado em inúmeras simulações.

No caso deste estudo, os testes GSADF foram operacionalizados no *software Eviews 9.0*, utilizando o algoritmo desenvolvido por Caspi (2016). Por limitações computacionais, os testes foram realizados anualmente com os preços das quatro praças físicas e do mercado de futuros de milho. Para cada ano de estudo houve variação no tamanho da amostra, portanto, no tamanho da janela mínima de teste. Os valores são apresentados na tabela 1A em anexo.

É importante mencionar que, ao selecionar o tamanho da janela, o teste GSADF gera perda das primeiras observações da amostra em número igual a  $r_w \times T_{ano}$ . Para evitar a perda de informações do teste dentro de um determinado ano, determinou-se que o tamanho da amostra seria ampliado, incluindo-se o equivalente a  $r_w \times T_{ano}$  preços do ano anterior. Desta forma, a primeira estatística calculada do teste para um determinado ano  $T_{ano}$  é referente ao primeiro dia deste ano. Neste caso, exceto o ano de 2004, cada ano analisado possui uma amostra estendida, que inclui observações do ano imediatamente anterior para realização dos testes. Este procedimento possibilita a criação de uma série contínua de valores da estatística GSADF para todo o período analisado.

Para identificar um período de bolha a partir das estimativas do teste GSADF, Phillips, Shi e Yu (2015) recomendam que seja considerado um período de comportamento explosivo nos preços aquele que tiver duração superior a  $\log(T)$ . Portanto, um período de bolha foi identificado quando o valor calculado da estatística GSADF foi superior ao valor crítico do teste por, no mínimo, três dias ( $h \geq 3$ ) consecutivos, ou seja, quando a hipótese nula foi rejeitada continuamente neste intervalo de tempo.

Nota-se que a equação (37) possui apenas um termo constante, sem tendência determinística. Este é um aspecto importante a ser considerado na execução do teste GSADF e diz respeito à especificação do modelo estimado. Assim como no caso do teste ADF, o modelo testado podem incluir um termo de tendência e um termo contante, uma combinação destes termos, ou nenhum deles. Normalmente, segue-se o procedimento de análise dos coeficientes  $\phi_1$ ,  $\phi_2$  e  $\phi_3$  detalhados em Enders (2014) para determinar o modelo mais adequado. No caso dos testes ADF de janelas móveis, Diba e Grossman (1988) apresentaram uma forma de identificar bolhas racionais em mercados de futuros utilizando um modelo com contante e com tendência<sup>7</sup>, porém o procedimento foi questionado por Evans (1991) em relação ao poder do teste na presença de bolhas que colapsam periodicamente.

Mais recentemente, testes realizados por Phillips, Shi e Yu (2014) demonstraram que o modelo usado por Diba e Grossman (1988) é inadequado, pois em se tratando de procedimentos empíricos, a presença de um intercepto e um termo de tendência é irrealista quando  $\gamma > 1$ . Os autores apontam ainda que, no caso de um modelo sem contante e sem tendência, não é possível a presença de tendência determinística quando  $\gamma = 1$ . Por estas razões, os autores sugerem que a melhor especificação para se testar a hipótese nula de raiz unitária em um teste tipo *right-tailed*, como é o caso do teste GSADF, é aquele que apresenta apenas o termo contante.

Para evitar possíveis problemas de inferência devido a volatilidade das séries durante os períodos de bolhas, foi realizado o teste da razão de variância (*variance ratio*) inicialmente apresentando por Lo e MacKinlay (1988). Este teste permite testar a previsibilidade de uma série temporal comparando as variâncias das diferenças dos retornos em diferentes pontos do tempo. A estatística do teste segue distribuição normal e testa a hipótese nula de que os resíduos são *martingales*, ou seja, são homocedásticos, contra a hipótese alternativa de heterocedasticidade dos resíduos. No caso de rejeição da hipótese de resíduos homocedásticos, é necessário simular os valores críticos do teste GSADF considerando o procedimento de *wild bootstrap*, conforme sugerido por (ETIENNE; IRWIN; GARCIA, 2014).

<sup>7</sup> Diba e Grossman (1988) recomendam testar a existência de bolhas analisando se os ativos respeitam as propriedades de estacionariedade e os fundamentos observáveis.

## 4 ANÁLISE DOS RESULTADOS E DISCUSSÕES

### 4.1 Estatísticas descritivas

Os resultados das estatísticas descritivas dos dados diários das cinco séries estudadas estão apresentados na Tabela 1. Em relação à média de preços, o mercado que teve maior média no período analisado foi o mercado de futuros, porém o mercado de Chapecó foi o que apresentou maior variabilidade dos preços, indicado pelo desvio-padrão.

Tabela 1 – Estatística descritiva dos preços e retornos diários das séries dos mercados físicos e futuros

	Cascavel		Chapecó		Rio Verde		Sorriso		Futuros	
	$P_t$	$R_t$	$P_t$	$R_t$	$P_t$	$R_t$	$P_t$	$R_t$	$P_t$	$R_t$
Média	22,10	0,01	25,35	0,01	20,54	0,01	14,33	0,01	25,76	0,01
Máximo	54,59	2,06	56,20	4,63	49,43	2,38	38,29	2,64	51,44	4,75
Mínimo	10,91	-2,36	12,49	-3,39	11,18	-2,92	7,00	-2,25	13,07	-6,94
Desv. padrão	7,11	0,28	7,59	0,36	7,12	0,34	5,65	0,37	7,17	0,51
Assimetria	1,59	-0,84	1,45	-0,01	1,59	-0,03	1,39	0,36	1,09	-1,14
Curtose	6,61	9,90	5,67	18,05	5,96	6,64	5,14	6,86	4,11	33,05
Jarque-Bera	2,90	13,01	2,01	41,99	2,45	5,67	1,598	6,13	773	141,42
P-valor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Fonte: Resultados da pesquisa.

Os resultados do teste de normalidade Jarque-Bera indicaram que, para os cinco mercados, rejeita-se a hipótese de normalidade ao nível de 1% de significância. Adicionalmente, todos os coeficientes de curtose foram superiores a 3 para o período, indicando que as distribuições de probabilidade dos retornos foram leptocúrticas, ou seja mais concentradas e afuniladas do que a distribuição normal.

As variações diárias de preços de Cascavel apresentaram distribuição de probabilidade mais leptocúrtica dentre os mercados estudados, enquanto o mercado de futuros apresentou distribuição menos afunilada. Os valores positivos para os coeficientes de assimetria dos preços sugerem que as distribuições dos cinco mercados são concentradas à esquerda. Já os valores negativos de assimetria sugerem que os mercados de Cascavel, Chapecó, Rio Verde e Futuros apresentaram a maior parte dos retornos acima da média.

Com o intuito de verificar se houve mudanças no comportamento dos preços devido ao deslocamento da produção da região Sul para a região Centro-Oeste, a análise dos resultados foi realizada considerando além do período todo - entre 2004 e 2016 - mais dois períodos considerando os resultados do teste de quebra estrutural de Zeileis et al. (2003) apresentados na Tabela 2. Os resultados sugerem a não rejeição da hipótese nula, o que indica a presença de uma quebra

estrutural em todas as praças analisadas. Sendo assim, além do período completo, a análise dos resultados será feita também para os períodos de 2004 a 2010 e 2011 a 2016 <sup>1</sup>.

Tabela 2 – Teste de quebra estrutural de Zeileis et al. (2003)

Mercado	Sup. F	Data da quebra
Cascavel	3025,7***	15/10/2010
Chapecó	3963,0***	20/10/2010
Rio Verde	2993,4***	14/10/2010
Sorriso	2062,9***	25/10/2010
Futuros	3533,6***	07/10/2010

Fonte: Resultados da pesquisa.

Nota: \*\*\* indica rejeição da hipótese nula ao nível de significância de 1%.

## 4.2 Medidas de descoberta de preços

Os resultados dos testes de Dickey-Fuller Aumentado (ADF) para todas as séries indicaram a não rejeição da hipótese nula de presença de raiz unitária para todas as séries ao nível de 1% de confiança. Após se repetir o teste nas séries diferenciadas, obteve-se o resultado esperado de estacionariedade das séries, considerando o mesmo nível de significância. A Tabela 3 apresenta os resultados do teste ADF considerando os modelos sem tendência e sem intercepto <sup>2</sup>.

<sup>1</sup> Por conveniência, como os resultados das quebras ficaram muito próximos do mês de dezembro, considerou-se o primeiro período como sendo do início de 2004 até o final de 2010. Além disso, as projeções de safra já sugeriam a maior parte da produção vindo do Centro-Oeste em 2011.

<sup>2</sup> O modelo foi especificado utilizando-se as estatísticas  $\phi$ , conforme sugerido por apresentado em Enders (2014, p. 207).

Tabela 3 – Estatísticas do testes de raiz unitária ADF para as séries de preços e para suas respectivas diferenças

Série	Estatística do teste	Lags
<i>Pca</i>	-0,2619 <sup>ns</sup>	5
$\Delta(Pca)$	-15,4419 <sup>***</sup>	4
<i>Pch</i>	0,1389 <sup>ns</sup>	8
$\Delta(Pch)$	-13,1279 <sup>***</sup>	7
<i>Prv</i>	-0,2025 <sup>ns</sup>	6
$\Delta(Prv)$	-12,4185 <sup>***</sup>	6
<i>Pso</i>	-0,2623 <sup>ns</sup>	9
$\Delta(Pso)$	-11,9863 <sup>***</sup>	9
<i>Pf</i>	0,0990 <sup>ns</sup>	1
$\Delta(Pf)$	-52,4599 <sup>***</sup>	1

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz.

Uma vez que todas as séries de preços foram identificadas como processos  $I(1)$ , foram realizados os testes de cointegração de Johansen utilizando-se as estatísticas de máximo autovalor e de traço para os pares de preços. Os resultados obtidos são apresentados na Tabela 4. É possível observar que oito dos nove pares de variáveis apresentaram pelo menos um vetor de cointegração no período de 2004 a 2016, implicando a existência de uma relação de longo prazo entre o mercados. A única exceção ocorreu para o caso de Chapecó e Sorriso, que não apresentaram cointegração neste período.

Tabela 4 – Testes de cointegração de Johansen para o período de 2004 a 2016

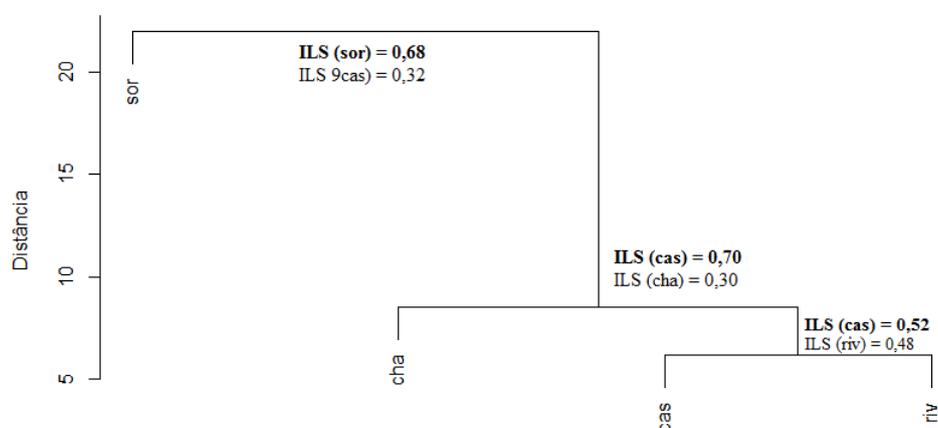
Mercados	Estatística		
	Rank	Traço	Máximo autovalor
<i>Pca &amp; Pch</i>	$r = 0$	67,57***	64,45***
	$r \leq 1$	3,12	3,12
<i>Pca &amp; Prv</i>	$r = 0$	48,19***	44,65***
	$r \leq 1$	3,54	3,54
<i>Pca &amp; Pso</i>	$r = 0$	32,91***	29,26***
	$r \leq 1$	3,65	3,65
<i>Pca &amp; Pf</i>	$r = 0$	100,73***	96,65***
	$r \leq 1$	4,07	4,07
<i>Pch &amp; Prv</i>	$r = 0$	22,36**	19,03**
	$r \leq 1$	3,33	3,33
<i>Pch &amp; Pso</i>	$r = 0$	15,59 <sup>ns</sup>	12,62 <sup>ns</sup>
	$r \leq 1$	2,98	2,98
<i>Pch &amp; Pf</i>	$r = 0$	210,12***	207,67***
	$r \leq 1$	2,44	2,44
<i>Prv &amp; Pso</i>	$r = 0$	22,82 <sup>ns</sup>	19,42 <sup>ns</sup>
	$r \leq 1$	2,93	2,93
<i>Prv &amp; Pf</i>	$r = 0$	183,56***	180,63***
	$r \leq 1$	2,93	2,93
<i>Pso &amp; Pf</i>	$r = 0$	67,80***	64,80***
	$r \leq 1$	3,00	3,00

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz.

Os resultados da análise de agrupamento hierárquico são apresentados no dendograma na Figura 4. É possível identificar a praça de Sorriso como a mais distante das demais praças em termos de preços, portanto, com características mais distintas em relação às outras praças. Também na Figura 4 estão os valores do ILS para cada rodada do torneio, e em destaque aparecem os valores das praças que lideraram o processo de descoberta de preços no período analisado. Os mercados vencedores (em negrito) seguiram para a fase seguinte do torneio.

Figura 4 – Dendograma hierárquico e valores do ILS considerando as quatro praças físicas entre 2004 e 2016



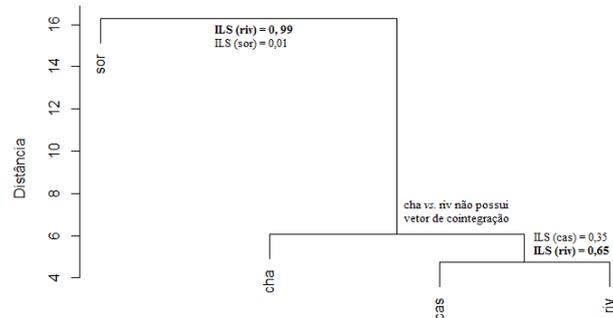
Fonte: Resultados da pesquisa.

Os valores do ILS apresentados na Figura 4 foram obtidos a partir dos resultados das estimações dos três modelos de correção de erros (VECM) apresentados nas tabelas 4A, 5A e 6A apresentadas no Apêndice A.

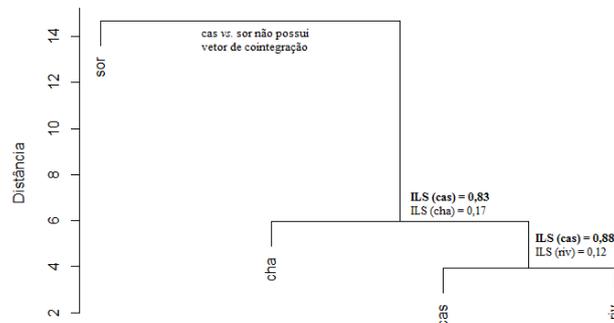
Para identificar se houve mudança no processo de formação de preços a partir do momento em que a região Centro-Oeste passou a liderar a produção de milho no Brasil, elaborou-se dois dendogramas respeitando a divisão dos períodos de 2004 a 2010 (antes do Centro-Oeste assumir a liderança na produção), e de 2011 a 2016 (após o Centro-Oeste liderar a produção). Os resultados dos dendogramas para estes períodos são apresentados na Figura 5. É possível perceber que as posições das praças nos dendogramas não se alteraram entre os dois períodos. Por outro lado, pode-se verificar que ocorreram mudanças significativas nos valores do ILS, quando comparados com os resultados apresentados na Figura 5. É possível perceber que, no primeiro período, Rio Verde liderou o processo de descoberta de preços (ILS = 65%) quando comparado com Cascavel. Este resultado pode ser reflexo da inauguração de um complexo industrial de aves e suínos da empresa de alimentos Perdigão da cidade de Rio Verde em 2003. No entanto, no segundo período Cascavel liderou este processo de formação de preços com 88% do ILS.

Figura 5 – Dendogramas hierárquicos e valores do ILS considerando as quatro praças físicas nos períodos de 2004 a 2010 e 2011 a 2016

(a) Dendograma do período de 2004 a 2010



(b) Dendograma do período de 2011 a 2016



Fonte: Resultados da pesquisa.

Os testes de cointegração de Johansen para os dois períodos apresentados nos dendogramas 5 são apresentados na Tabela 2A do Apêndice A.

Conforme descrito na seção da metodologia, é importante determinar qual variável se ajusta primeiro diante das mudanças das inovações sobre o preço eficiente de longo prazo. Isto porque, as parcelas de informação de cada mercado estão ligadas a esta movimentação.

Na comparação dos resultados das medidas de descoberta de preços dos mercados *spot* com o de futuros (Tabela 1B do Apêndice B), verificou-se que este último lidera a formação de preços do milho em todas as comparações. Este resultado era esperado, uma vez que o mercado futuro oferece baixos custos de operação aos negociadores e estabelece suas negociações por meio de contratos que servem de garantia ao negociador, facilitando a resposta do mercado de futuros a quaisquer variações de preços (DIAMOND; VERRECCHIA, 1987; MIN; NAJAND, 1999). Por esta razão, a análise das medidas de *price discovery* considerou apenas os quatro mercados *spot* durante o período de 2004 a 2016.

As comparações entre os valores do *information share*, do *component share*, do *information leadership* e do *information leadership share* foram feitas com base nos agrupamentos

determinados anteriormente por meio do dendograma (Figura 4).

A Tabela 5 contém os valores de todas as medidas de *price discovery* calculadas para todos os pares de preços analisados. Os valores em destaque nesta tabela representam o percentual referente à maior parcela de informação, para as diferentes medidas calculadas, em cada rodada do torneio, entre 2004 a 2016. Na primeira e segunda rodadas, Cascavel apresentou maior capacidade de reação frente a uma nova informação no mercado em comparação a Rio Verde (primeira rodada) e Chapecó (segunda), visto que suas parcelas de informação foram superiores a 60% em ambas as comparações. Este resultado também é comprovado pela medida do IL que, como mencionado anteriormente, indica liderança quando os valores de uma variável são superiores a um. No entanto, na última rodada de comparação, entre Cascavel e Sorriso, o mercado de milho do Centro-Oeste liderou o processo de formação de preços. Isto pode ser verificado pelos altos valores de ILS e IL.

Tabela 5 – Medidas de descoberta de preços entre 2004 e 2016

Medidas	<i>Pca vs. Prv</i>	<i>Pca vs. Pch</i>	<i>Pca vs. Pso</i>
IS <sub>1</sub>	<b>0,64</b>	<b>0,68</b>	<b>0,67</b>
IS <sub>2</sub>	0,36	0,32	0,33
CS <sub>1</sub>	<b>0,63</b>	<b>0,59</b>	<b>0,75</b>
CS <sub>2</sub>	0,37	0,41	0,25
ILS <sub>1</sub>	<b>0,52</b>	<b>0,70</b>	0,32
ILS <sub>2</sub>	0,48	0,30	<b>0,68</b>
IL <sub>1</sub>	<b>1,05</b>	<b>1,52</b>	0,69
IL <sub>2</sub>	0,95	0,66	<b>1,45</b>
Líder	<i>Pca</i>	<i>Pca</i>	<i>Pso</i>

Fonte: Resultados da pesquisa.

Nota: Os subscritos 1 e 2 são referentes à primeira e à segunda variáveis testadas em cada rodada, respectivamente

Por ser a maior produtora de milho do Brasil e considerando que a região possui a maior diferença de preços em relação às demais praças (fato observado pelas distâncias definidas no dendograma), esperava-se que Sorriso tivesse papel determinante no processo de *price discovery* do milho, o que foi constatado pelos resultados das medidas de descoberta de preços.

Uma das prováveis razões para Cascavel aparecer como líder no processo de descoberta de preços, quando comparada a Rio Verde e Chapecó, tem relação com o fato deste mercado estar localizada próximo a Chapecó, grande região consumidora de milho, principalmente devido à presença de grandes indústrias do ramo alimentício, como a Sadia e a Seara. Além disso, Cascavel concentra a segunda maior produção de milho da região Sul e está mais próxima de pontos de escoamento, como o porto de Santos e, principalmente, de Paranaguá.

O fato de Sorriso liderar a formação de preços quando comparado aos demais mercados analisados, é uma indicação de que o mercado do Centro-Oeste possui a maior parte da informação de preços *spot* no país. No entanto, uma análise mais apurada, levando em conta uma comparação temporal das medidas de *price discovery* é necessária para confirmar esta evidência.

Para compreender melhor se houve mudança no processo de descoberta de preços, com o deslocamento da informação para os mercados da região Centro-Oeste, as Tabelas 6 e 7 apresentam os valores das medidas de descoberta de preços para os torneios definidos pelos dendogramas apresentados na Figura 5, para os períodos de 2004 a 2010 e 2011 a 2016.

Tabela 6 – Medidas de descoberta de preços entre 2004 e 2010

Medidas	<i>Pca vs. Prv</i>	<i>Prv vs. Pso</i>
IS <sub>1</sub>	0,43	<b>0,99</b>
IS <sub>2</sub>	<b>0,57</b>	0,01
CS <sub>1</sub>	<b>0,51</b>	<b>0,95</b>
CS <sub>2</sub>	0,49	0,05
ILS <sub>1</sub>	0,35	<b>0,99</b>
ILS <sub>2</sub>	<b>0,65</b>	0,01
IL <sub>1</sub>	0,73	<b>8,13</b>
IL <sub>2</sub>	<b>1,38</b>	0,12
Líder	<i>Prv</i>	<i>Prv</i>

Fonte: Resultados da pesquisa.

Os resultados da Tabela 6 mostram que a praça de Rio Verde liderou o processo de formação de preços em relação a Cascavel e Sorriso. Seguindo o dendograma apresentado na figura 5, após o jogo de Pcas e Prv deveriam ser feitas as medidas entre Prv e Pch, porém estas praças não apresentaram um vetor de cointegração de acordo com o teste de Johansen, portanto foram feitos os testes considerando as praças de Prv e Pso e Pch e Pso, porém só as praças de Prv e Pso apresentaram cointegração.

As medidas de *price discovery* apresentadas no primeiro período mostraram uma liderança de Rio Verde no processo de descoberta de preços. O mercado de Rio Verde também liderou o processo de informação dos preços quando comparado a Sorriso no mesmo período.

Tabela 7 – Medidas de descoberta de preços entre 2011 e 2016

Indicadores	<i>Pca vs. Prv</i>	<i>Pca vs. Pch</i>
IS <sub>1</sub>	<b>0,93</b>	<b>0,86</b>
IS <sub>2</sub>	0,07	0,14
CS <sub>1</sub>	<b>0,83</b>	<b>0,73</b>
CS <sub>2</sub>	0,17	0,27
ILS <sub>1</sub>	<b>0,88</b>	<b>0,83</b>
ILS <sub>2</sub>	0,12	0,17
IL <sub>1</sub>	<b>2,75</b>	<b>2,25</b>
IL <sub>2</sub>	0,36	0,44
Líder	<i>Pca</i>	<i>Pca</i>

Fonte: Resultados da pesquisa.

Para os resultados do segundo período apresentados na Tabela 7 apenas os dois primeiros jogos do dendograma apresentaram vetor de cointegração. Como as medidas de IS, CS e

ILS consideram o modelo de correção de erros para determinar seus valores, quando não há cointegração entre duas variáveis os resultados das medidas não apresentam valores confiáveis, portanto não foram apresentados neste trabalho.

Ao contrário do que foi constatado para o primeiro período de análise, nos anos entre 2011 e 2016 as medidas de descoberta de preços entre Cascavel e Rio verde foram favoráveis ao primeiro mercado, que liderou o processo de *price discovery* em 88%. É possível que durante este período, Rio Verde tenha se tornado um mercado que atende bem a demanda da sua própria região e, então, a informação gerada neste mercado não se reflete em outras regiões, influenciando menos os preços.

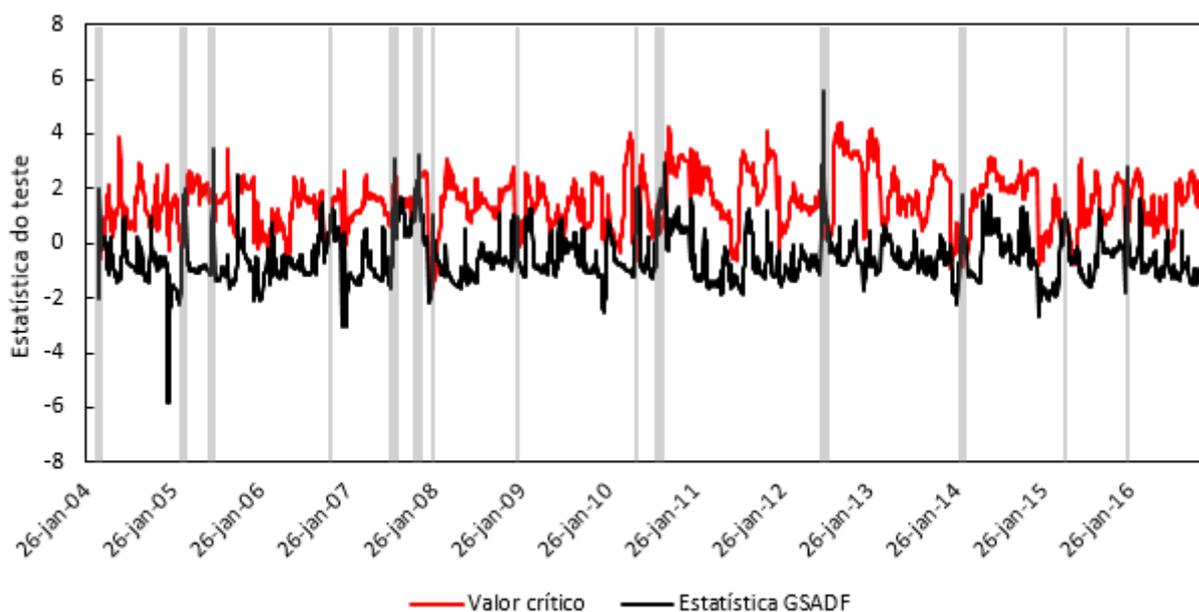
A mesma característica de autosuficiência pode ser relacionada ao mercado de Chapecó, que teve menos influência nos preços que Cascavel entre 2011 e 2016, como mostra a Tabela 7. Já as praças de Sorriso e Cascavel possuem fortes características de abastecimento de milho para outras regiões e aparecem liderando as informações dos preços nos períodos analisados.

#### **4.3 Análise de curto prazo: bolhas especulativas**

As medidas de descoberta de preços descrevem uma análise de longo prazo das relações entre os mercados, considerando o equilíbrio e as inovações que promovem os ajustamentos. Para análise de curto prazo, investigou-se a presença de bolhas especulativas nas séries durante o período de 2004 a 2016 e nos períodos de 2004 a 2010 e de 2011 a 2016, separadamente.

Primeiramente, são apresentados os resultados do teste GSADF para o período todo. Os resultados para os preços no mercado de futuros são apresentados na Figura 6. Considerando o valor mínimo de 3 dias, as áreas sombreadas representam os períodos de colapso de bolhas especulativas no mercado futuro de milho entre 2004 e 2016.

Figura 6 – Resultado do teste GSADF e identificação dos períodos de bolha na série de preços futuros de milho (2004 a 2016)



Fonte: Resultados da pesquisa.

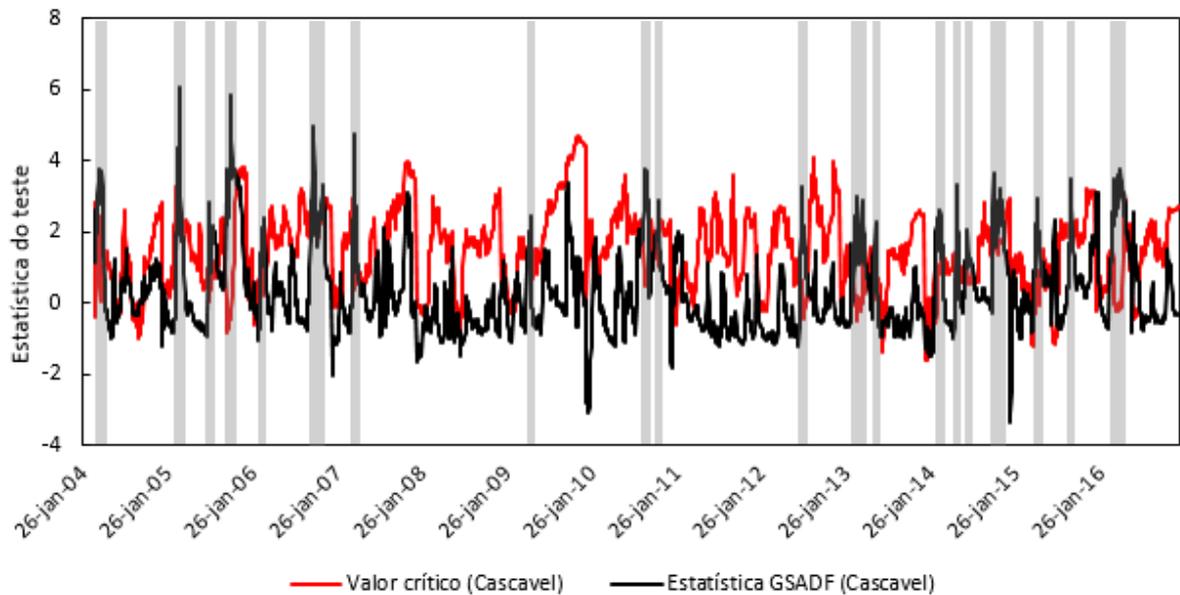
Nota: Resultados obtidos considerando o nível de confiança de 95%.

Na análise dos resultados para as praças físicas, além identificar os períodos de colapso de bolhas especulativas, obteve-se também os valores dos preços para cada variável. Observando os resultados dos testes GSADF e os preços para as praças físicas da região Sul e Centro-Oeste, apresentados nas Figuras 7 e 8, é possível perceber que aparentemente ambas as regiões tiveram quantidade semelhantes de bolhas no período estudado. Uma possível explicação para as bolhas está ligada ao comportamento do preço, pois no momento de elevação do preço o produtor pode optar por não entregar a mercadoria por ser mais vantajoso mantê-la em estoque.

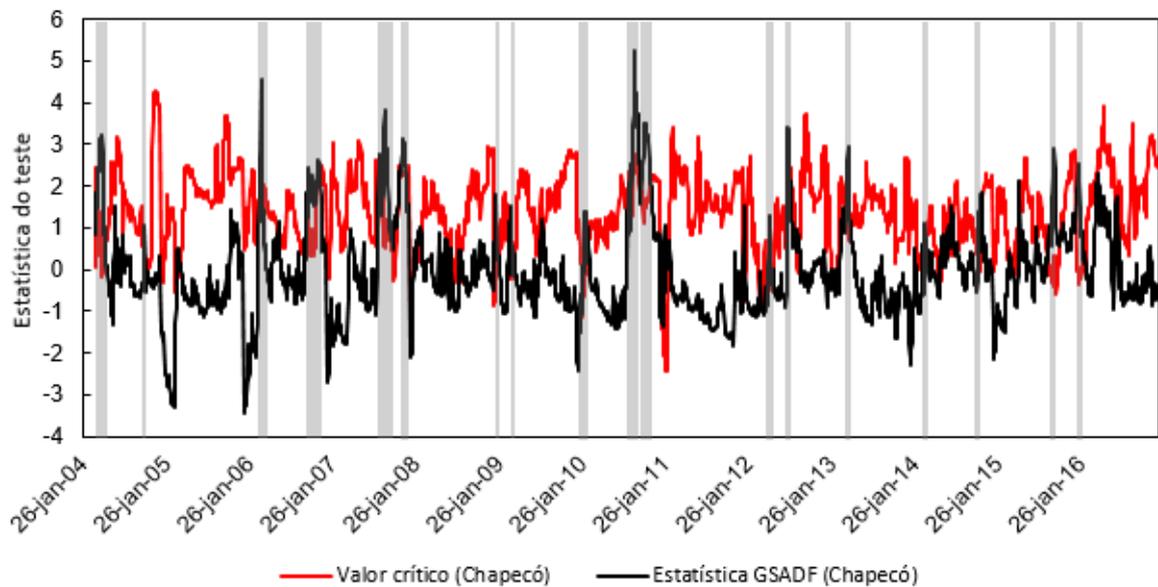
Uma característica interessante observada nas Figuras 7 e 8 é que, aparentemente, os episódios de bolhas ocorrem primeiro nas praças da região Sul, já que nestas praças as bolhas têm início no segundo semestre de 2004, ao passo que em Sorriso, as primeiras bolhas começam apenas em 2006. Sorriso também apresentou um número menor de bolhas no decorrer do período quando comparado às outras cidades.

Figura 7 – Resultados do teste GSADF para os preços nos mercados da região Centro-Oeste (2004 a 2016)

(a) Identificação dos períodos de bolhas em Cascavel



(b) Identificação dos períodos de bolhas em Chapecó

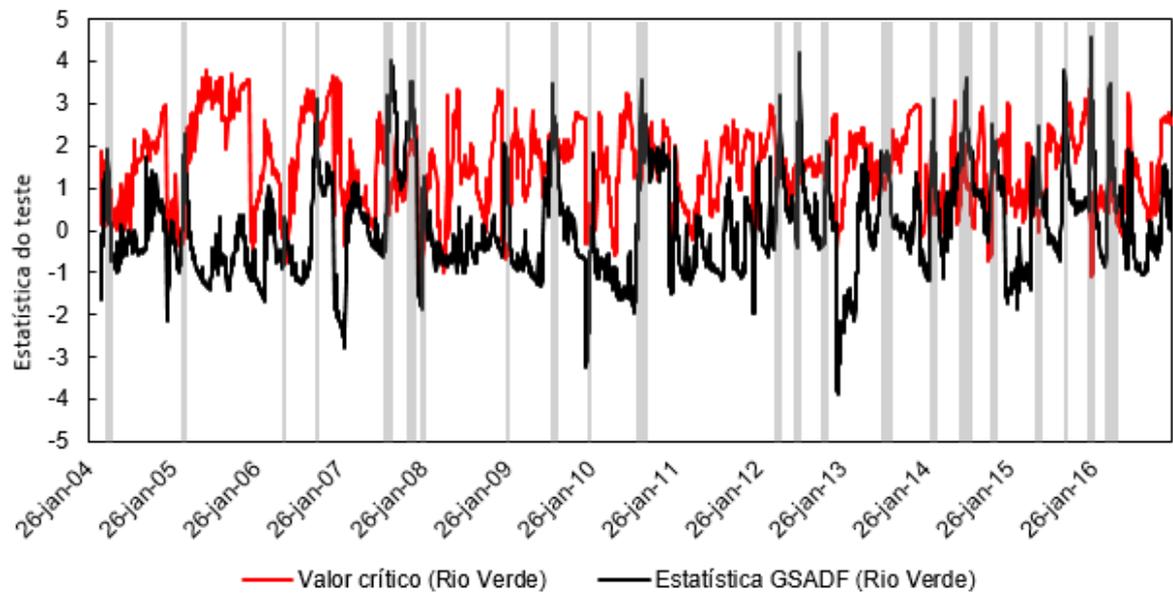


Fonte: Resultados da pesquisa.

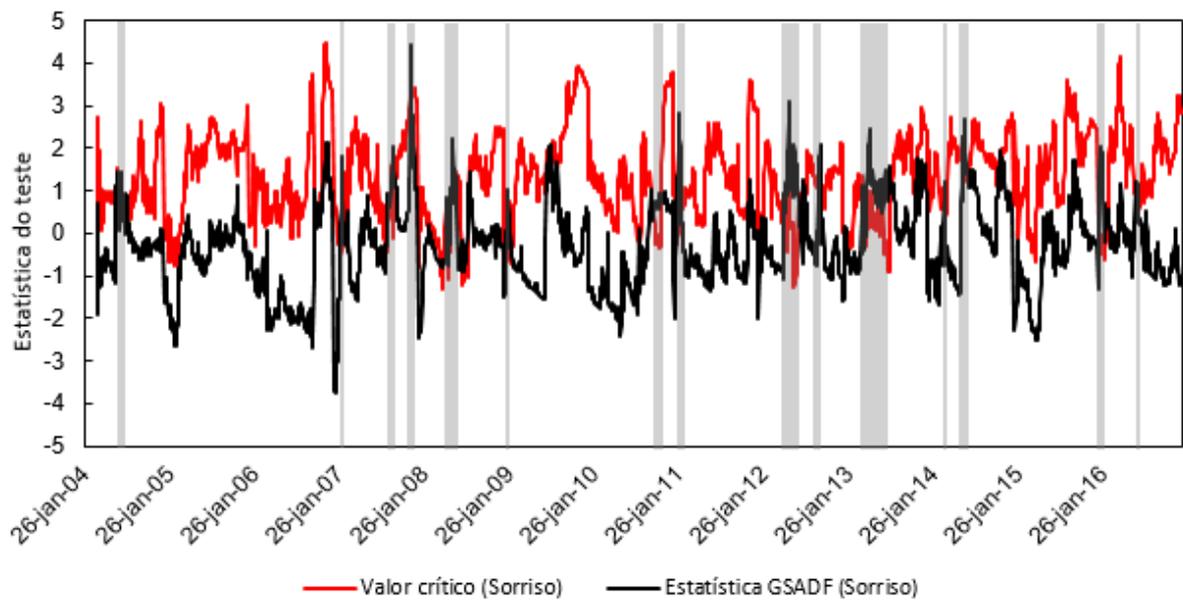
Nota: Resultados obtidos considerando o nível de confiança de 95%.

Figura 8 – Resultados do teste GSADF para os preços nos mercados da região Centro-Oeste (2004 a 2016)

(a) Identificação dos períodos de bolhas em Rio Verde



(b) Identificação dos períodos de bolhas em Sorriso

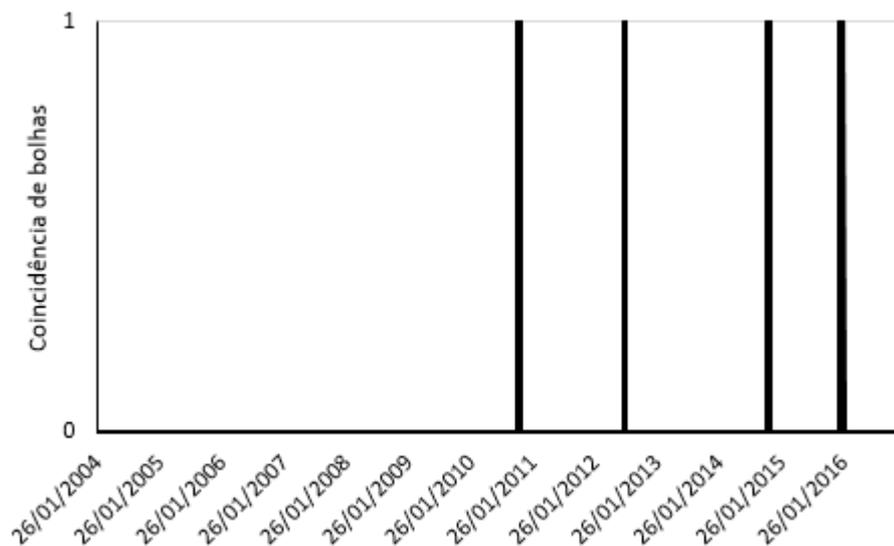


Fonte: Resultados da pesquisa.

Nota: Resultados obtidos considerando o nível de confiança de 95%.

A partir dos resultados do teste foi possível identificar a correlação entre as bolhas em cada uma das praças buscando determinar os momentos em que os episódios de bolhas ocorreram simultaneamente nas quatro praças físicas. As coincidências começaram a partir de 2010 e somaram 28 dias, o que representa 19% dos dias com bolhas. Os resultados estão apresentados na Figura 9. Poucos episódios ocorreram simultaneamente nas quatro praças físicas durante o período estudado.

Figura 9 – Períodos de bolhas coincidentes em Cascavel, Chapecó, Rio Verde e Sorriso.

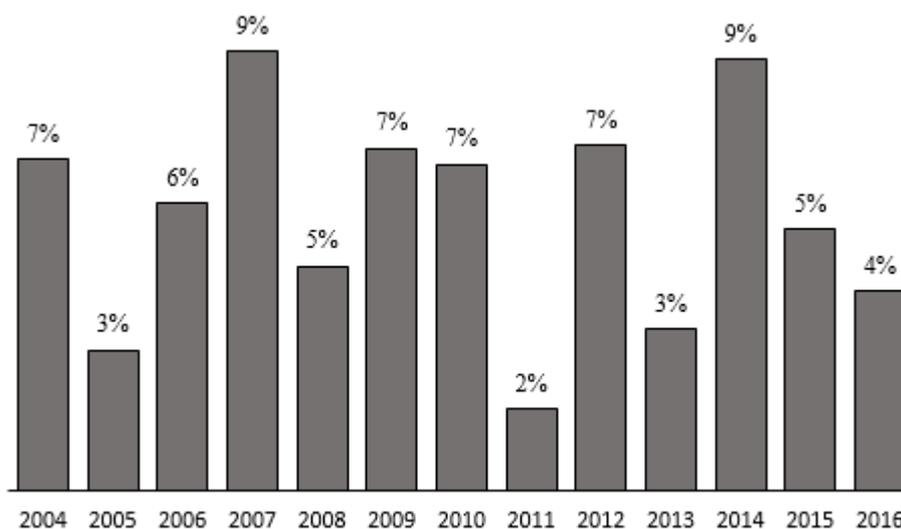


Fonte: Resultados da pesquisa.

Nota: 1= houve bolha nas quatro praças simultaneamente ( $h \geq 3$ ); 0 = não houve bolha nas quatro praças simultaneamente ( $h < 3$ ).

Um padrão interessante de ser observado é que os episódios de bolhas acontecem com uma frequência parecida entre os períodos de 2004 a 2010 e de 2011 a 2016. Este padrão pode ser observado na Figura 10 que mostra o percentual de dias com bolhas nos cinco mercados durante os treze anos da análise. Os anos de 2007 e 2014 apresentaram um percentual maior de dias com bolhas em comparação aos outros anos. Nestes dois anos, 9% dos dias identificados como períodos de bolhas no preço do milho.

Figura 10 – Frequência do número de dias com bolha em todos os mercados entre 2004 e 2016



Fonte: Resultados da pesquisa.

Considerando os mercados separadamente no período 2004 a 2016, Cascavel e Chapecó tiveram 19% e 14% dos dias com bolhas detectados pelo teste, respectivamente, enquanto Rio Verde teve 15% dos dias com bolhas e Sorriso, 10%. O mercado de futuros registrou 5% dos dias analisados com episódios de bolhas.

Para facilitar a interpretação e a análise dos episódios explosivos determinados pelo teste GSADF, os resultados apresentados na Tabela 8 mostram algumas características das bolhas. Estes resultados trazem a separação das bolhas entre positivas e negativas. Para determinar os tamanhos e os formatos das bolhas foram utilizados os preços no início da bolha,  $P_0$ , no ponto de pico (ou vale),  $P_{p,v}$  e o ponto final da bolha,  $P_{final}$  para calcular os retornos do início até o pico (vale) e os retornos do pico (vale) até o final da bolha.

De acordo com os resultados da Tabela 8 o número de episódios explosivos nas praças físicas da região Sul foram maiores que na região Centro-Oeste. As bolhas positivas e negativas nas cidades de Cascavel e Chapecó totalizaram 83, enquanto as bolhas em Rio Verde e Sorriso somaram 67. Além disso, as bolhas na região Sul tiveram duração média de 13 dias, enquanto as do Sul duraram em média 11 dias no período analisado.

Em relação às bolhas positivas, a praça de Sorriso apresentou a maior média percentual de aumento nos preços (média entre início da bolhas e o pico), 9,9%. As bolhas positivas nos cinco mercados indicam que os preços reagiram com aumento em 8,9%, em média, durante tais episódios. A maior reação de preços foi de 22,8% em Sorriso no ano de 2009. No mesmo ano, Sorriso apresentou o maior percentual de ajustamento dos preços do pico até o final da bolha positiva, de -23,7%. Em média, as bolhas positivas duraram cerca de 12 dias.

No caso das bolhas negativas, uma análise conjunta revelou que 6,4% dos preços nos cinco

mercados reagiram exageradamente, sendo que o mercado de futuros registrou queda média de 24,2% em 2008, o maior valor em comparação às outras variáveis. Em média as bolhas negativas tiveram duração de 10 dias.

O caso de Rio Verde em 2016 merece destaque, por ter dois episódios de bolhas positivas com duração média de 33 dias. O primeiro deles ocorreu de 08/01/2016 a 02/02/2016 e teve duração de 18 dias. O segundo episódio, mais notável por ter durado 47 dias, ocorreu entre 24/03/2016 a 01/06/2016. No ano de 2016 os preços do milho atingiram patamares recordes no Brasil impulsionados pela quebra da safra 2015/16 que resultou a queda na produção naquele ano. Os preços favoráveis no primeiro semestre de 2016 estimularam o aumento da área plantada do milho de segunda safra, que passou a ser a maior da história. Porém, o atraso na colheita das regiões produtoras no Centro-Oeste e uma forte seca entre maio e junho provocaram queda na produtividade das lavouras em mais de 30% comparado ao mesmo período de 2015, segundo um relatório do CEPEA (2016).

Outro destaque deve ser dado para o ano de 2015 em Chapecó, que registrou uma única bolha positiva com duração de 34 dias úteis entre 02/09/2015 e 22/10/2015. Apesar da safra recorde em 2015, os preços do milho foram impulsionados pela maior participação do produto brasileiro no mercado internacional. No segundo semestre de 2015, a forte desvalorização do Real impulsionou as exportações brasileiras elevando os preços do grão no período (CEPEA, 2015).

Tabela 8 – Características das bolhas nos preços de 2004 a 2016 ( $h \geq 3$ )

Região	N	Bolhas positivas			N	Bolhas negativas		
		Média Duração (dia)	Média (%) $P_0$ & Pico	Média (%) Pico & $P_{final}$		Média Duração (dias)	Média (%) $P_0$ & Vale	Média (%) Vale & $P_{final}$
<b>Cascavel</b>								
2004	4	10	6.2	-0.2	3	4	-2.8	0.5
2005	1	15	15.1	0	2	31	-14.2	0.3
2006	4	12	8.3	-0.5	1	17	-13.9	0.7
2007	2	6	8.2	0	3	9	-2.6	1.0
2008	0	0	0	0	0	0	0	0
2009	1	4	1.7	-0.6	1	5	-3.0	0
2010	2	16	13.4	-3.3	1	8	-1.1	0.1
2011	1	24	11.4	-2.0	0	0	0	0
2012	1	23	15.8	-1.2	0	0	0	0
2013	1	11	4.9	-3.0	2	23	-14.8	0.6
2014	4	13	5.3	-0.7	3	10	-3.2	0.5
2015	3	7	4.2	-1.8	2	14	-5.3	0
2016	3	20	15.1	-0.6	1	21	-22.6	3.6
<b>Subtotal</b>	<b>27</b>	<b>13</b>	<b>9.1</b>	<b>-1.2</b>	<b>19</b>	<b>14</b>	<b>-8.4</b>	<b>0.7</b>
<b>Chapecó</b>								
2004	1	21	10.3	-2.8	1	3	-0.7	1.0
2005	1	12	3.4	-0.3	0	0	0	0
2006	2	19	10.1	-0.4	2	5	-6.8	0.6
2007	4	17	10.4	-1.5	0	0	0	0
2008	0	0	0	0	2	8	-3.9	0.2
2009	3	9	3.7	-1.3	1	7	-3.7	0.6
2010	2	27	13.6	-1.9	2	10	-4.4	0.8

Continua na próxima página

Tabela 8 – continuação da página anterior

Região	N	Bolhas positivas			Bolhas negativas			
		Média Duração (dia)	Média (%) $P_0$ & Pico	Média (%) Pico & $P_{final}$	N	Média Duração (dias)	Média (%) $P_0$ & Vale	Média (%) Vale & $P_{final}$
2011	2	10	3.5	0	0	0	0	0
2012	5	6	2.9	-0.7	2	4	-0.4	0
2013	0	0	0	0	1	10	-5.9	2.3
2014	2	10	5.5	0	2	6	-2.5	1.1
2015	1	34	14.8	-3.9	0	0	0	0
2016	1	23	14.5	-1.0	0	0	0	0
<b>Subtotal</b>	<b>24</b>	<b>17</b>	<b>8.4</b>	<b>-1.2</b>	<b>13</b>	<b>7</b>	<b>-3.5</b>	<b>0.8</b>
<b>Rio Verde</b>								
2004	3	4	1.3	-0.7	0	0	0	0
2005	1	14	12.8	-2.0	0	0	0	0
2006	0	0	0	0	2	7	-3.2	2.7
2007	3	24	17.4	-2.8	1	4	-1.4	1.1
2008	0	0	0	0	4	4	-3.1	0.6
2009	1	21	0	9.8	3	5	-2.1	0.5
2010	4	9	5.6	-0.9	2	4	-1.4	0.7
2011	0	0	0	0	0	0	0	0
2012	2	10	6.1	-1.1	1	16	-9.4	2.2
2013	0	0	0	0	1	28	-9.8	1.6
2014	3	12	8.9	-0.4	3	19	-7.6	1.7
2015	1	3	4.1	-1.6	2	10	-3.6	0.9
2016	2	33	21.4	-2.8	0	0	0	0
<b>Subtotal</b>	<b>20</b>	<b>14</b>	<b>8.6</b>	<b>-0.3</b>	<b>19</b>	<b>11</b>	<b>-4.6</b>	<b>1.3</b>
<b>Sorriso</b>								
2004	0	0	0	0	1	6	-2.1	2.0
2005	0	0	0	0	0	0	0	0
2006	1	4	0.9	0	0	0	0	0
2007	4	10	9.4	-0.6	0	0	0	0
2008	1	4	2.9	0	3	14	-12.0	0.5
2009	1	3	22.8	-23.7	3	5	-3.2	0
2010	1	20	14.4	0	0	0	0	0
2011	1	11	19.3	-2.3	0	0	0	0
2012	1	4	2.9	0	3	18	-13.4	0.6
2013	0	0	0	0	3	22	-17.8	1.5
2014	2	5	4.3	0	1	11	-13.1	0
2015	0	0	0	0	0	0	0	0
2016	2	15	12.7	-1.8	0	0	0	0
<b>Subtotal</b>	<b>14</b>	<b>8</b>	<b>9.9</b>	<b>-3.2</b>	<b>14</b>	<b>13</b>	<b>-10.2</b>	<b>0.8</b>
<b>Futuros</b>								
2004	1	9	4.5	0	1	4	-2.1	0
2005	1	11	17.9	0	1	5	-4.3	0
2006	2	5	4.6	0	0	0	0	0
2007	3	3	5.7	-1.5	1	4	-0.2	0.3
2008	0	0	0	0	1	11	-24.2	2.4
2009	2	5	2.4	-1.7	1	3	-1.4	0
2010	2	17	9.7	-1.4	0	0	0	0
2011	0	0	0	0	0	0	0	0
2012	1	7	22	-1.9	1	6	-3.2	0
2013	0	0	0	0	0	0	0	0
2014	1	16	7.0	-3.3	0	0	0	0
2015	1	4	0.6	-0.7	3	5	-1.6	0.5
2016	1	7	7.8	-0.9	0	0	0	0

Continua na próxima página

Tabela 8 – continuação da página anterior

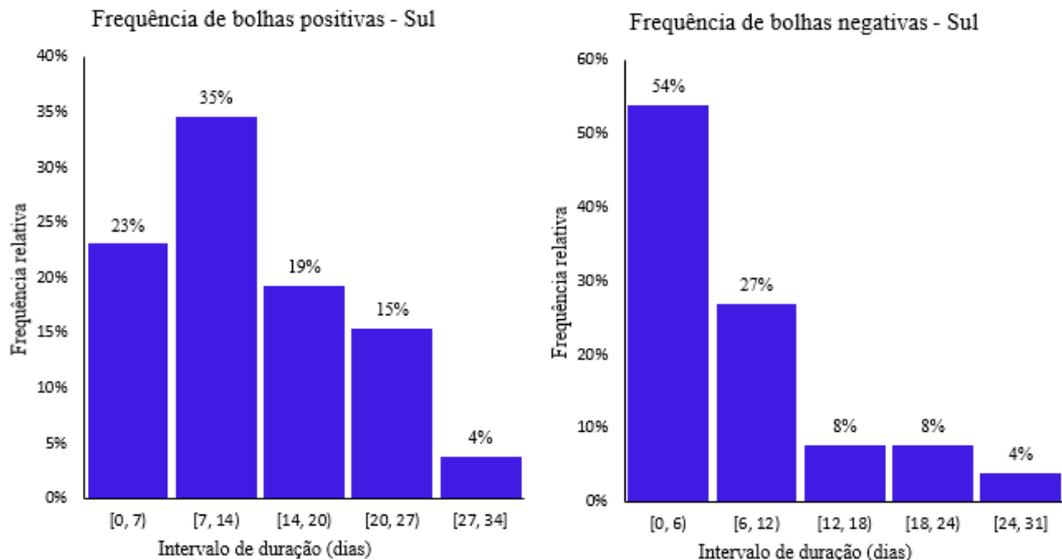
Região	N	Bolhas positivas			N	Bolhas negativas		
		Média Duração (dia)	Média (%) $P_0$ & Pico	Média (%) Pico & $P_{final}$		Média Duração (dias)	Média (%) $P_0$ & Vale	Média (%) Vale & $P_{final}$
<b>Subtotal</b>	<b>15</b>	<b>8</b>	<b>8.2</b>	<b>-1.1</b>	<b>9</b>	<b>5</b>	<b>-5.3</b>	<b>0.5</b>
<b>TOTAL</b>	<b>100</b>	<b>12</b>	<b>8.9</b>	<b>-1.4</b>	<b>74</b>	<b>10</b>	<b>-6.4</b>	<b>0.8</b>

Apesar do teste GSADF ter captado muitos momentos de bolhas nos mercados de milho entre 2004 e 2016, elas tiveram duração relativamente curta. Isso é observado na Figura 11 que ilustra a frequência relativa das bolhas positivas e negativas nas regiões Sul e Centro-Oeste de acordo com o intervalo de duração dos episódios.

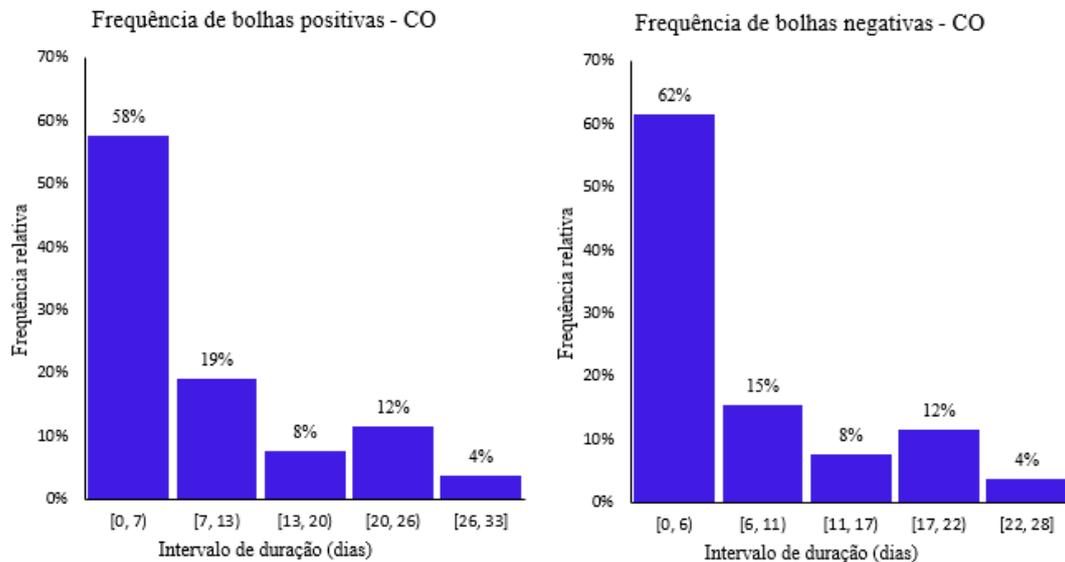
É possível perceber que, no caso da região Sul, 58% das bolhas positivas duraram menos de 14 dias, sendo que apenas 4% delas duraram entre 27 e 34 dias. Já no caso das bolhas negativas, 54% duraram menos de 6 dias considerando o período completo. Na região Centro-Oeste o comportamento foi semelhante ao da região Sul, com 58% das bolhas positivas se concentrando em um período curto de menos de sete dias, e 62% das bolhas negativas durante menos de seis dias. De maneira geral, as bolhas positivas mais longas ocorreram com maior frequência na região Sul. De modo geral, a maioria das bolhas também ocorreu na região Sul, onde a capacidade de estoques é menor.

Figura 11 – Frequência dos episódios de bolhas positivas e negativas nas regiões Sul e Centro-Oeste no período de 2004 a 2016.

(a) Frequência relativa de bolhas na região Sul



(b) Frequência relativa de bolhas na região Centro-Oeste



Fonte: Resultados da pesquisa.

Uma análise mais detalhada traz as características das bolhas considerando os dois períodos de quebra. A tabela 9 mostra o comportamento das bolhas nos dois períodos e a variação entre eles. Em quase todas as praças físicas e no mercado de futuros o número de bolhas positivas e negativas foi maior no período de 2011 a 2016, com exceção das bolhas negativas em Sorriso que foram sete em ambos os períodos.

Apesar do número de bolhas ter caído do primeiro para o segundo período, em Cascavel,

Sorriso e no mercado de futuros, a duração média, em dias, desses episódios explosivos aumentou nestes mercados. O número de bolhas positivas em Rio Verde e Chapecó também diminuiu entres os dois períodos, porém a média de duração das bolhas caiu nestes mercados. Com exceção de Sorriso, todos os outros mercados registraram queda no número de bolhas negativas entre os dois períodos e todos os mercados tiveram queda no número de bolhas positivas entre os dois períodos. Apesar disso, em muitos casos, o tempo médio de duração dos episódios explosivos aumentou de um período para outro, com destaque para Rio Verde, que teve uma queda de 42% no número de bolhas negativas, porém um aumento de 309% no tempo médio de duração das mesmas.

Tabela 9 – Características das bolhas nos preços dividido por períodos ( $h \geq 3$ )

Região	N	Duração média (dias)	Bolhas Positivas		N	Duração média (dias)	Bolhas Negativas	
			Média entre $P_0$ e $P_{p,v}$ (%)	Média entre $P_{p,v}$ e $P_{final}$ (%)			Média entre $P_0$ e $P_{p,v}$ (%)	Média entre $P_{p,v}$ e $P_{final}$ (%)
<b>Cascavel</b>								
2004-2010	14	10	8,8	-0,8	11	12	-6,3	0,4
2011-2016	13	16	9,5	-1,6	8	17	-11,5	1,2
Varição entre os períodos	-7%	56%	8%	104%	-27%	38%	83%	172%
<b>Chapecó</b>								
2004-2010	13	17	8,6	-1,4	8	6	-3,9	0,7
2011-2016	11	16	8,2	-1,1	5	7	-2,9	1,1
Varição entre os períodos	-15%	-6%	-4%	-20%	-38%	4%	-25%	73%
<b>Rio Verde</b>								
2004-2010	12	14	7,4	0,7	12	4	-2,3	1,1
2011-2016	8	14	10,1	-1,5	7	18	-7,6	1,6
Varição entre os períodos	-33%	-1%	36%	-317%	-42%	309%	236%	44%
<b>Sorriso</b>								
2004-2010	8	8	10,1	-4,9	7	8	-5,7	0,8
2011-2016	6	9	9,8	-1,0	7	17	-14,8	0,7
Varição entre os períodos	-25%	5%	-3%	-79%	0%	105%	158%	-17%
<b>Futuro</b>								
2004-2010	11	8	7,5	-0,8	5	5	-6,4	0,5
2011-2016	4	9	9,3	-1,7	4	6	-2,4	0,2
Varição entre os períodos	-64%	2%	25%	121%	-20%	2%	-63%	-56%

Fonte: Resultados da pesquisa.

Nota: N = duração da bolha em dias.

A partir da análise do número de bolhas nos dois períodos e das medidas de *information share*, pode-se perceber que entre 2004 e 2010, Rio Verde liderou o processo de formação de preços quando comparado a Cascavel, e o número de bolhas tanto positivas quanto negativas foi semelhante entre os dois mercados no período. Já entre 2011 e 2016, os resultados das medidas de descoberta de preços apontaram Cascavel com a maior parcela de informação quando comparada a Rio Verde. Neste período também, o número de bolhas especulativas em Cascavel foi um pouco maior que em Rio Verde. A duração média das bolhas aumentou 56% de um período para outro em Cascavel, enquanto que em Rio Verde, esse número caiu cerca de 1%.

## 5 CONSIDERAÇÕES FINAIS

As duas primeiras décadas do anos 2000 foram marcadas por diversas mudanças nos mercados brasileiro e internacional de milho. A principal mudança no mercado doméstico aconteceu com o deslocamento da produção da região Sul para a região Centro-Oeste, o que provocou um conseqüente avanço da indústria processadora de aves e suínos para esta última região. Enquanto isso, o mercado internacional do grão foi influenciado pelo desenvolvimento da indústria de etanol à base de milho nos EUA, e também pela crescente demanda externa, principalmente a chinesa. Diante desses fatores, surgiu a necessidade de investigar se o processo de descoberta dos preços físicos no Brasil também foi alterado. Além disso, procurou-se avaliar os mercados físicos na tentativa de identificar comportamentos especulativos nos preços.

Os resultados obtidos indicaram que o mercado de futuros é a referência no processo de descoberta de preços no Brasil. Já a análise dos mercados físicos indicou que a região Sul, representada pela cidade de Cascavel, no Paraná, é a que possui maior parcela de informação de preços entre todos os mercados estudados. A análise realizada através de um torneio, baseada no cálculo do *information leadership share*, indicou que o mercado de Cascavel é o líder no processo de descoberta de preços no país, sendo seguido por Sorriso, outro importante mercado produtor localizado no Centro-Oeste, no estado do Mato Grosso.

No entanto, verificou-se que a referência de preços nos mercados físicos sofreu alterações no período analisado. Quando o período analisado foi dividido em dois, os resultados indicaram que a cidade de Rio Verde, em Goiás, foi a líder na informação de preços entre 2004 e 2010. Este resultado pode estar relacionado ao fato de que, no início dos anos 2000, a cidade do Centro-Oeste recebeu grande volume de investimento com a criação de um importante centro agroindustrial. Este fato parece ter colocado este mercado como uma referência (liderança) de preços no país, sendo seguido por Sorriso.

Porém, no período entre 2011 e 2016, o mercado de Cascavel se firmou como líder no processo de descoberta de preços no país. Esta mudança pode ser resultado do avanço do milho na safra de inverno em direção ao mercado internacional, enquanto o milho da safra se direcionou ao abastecimento doméstico. Isto fez com que Cascavel passasse a exercer uma maior influência em outros mercados físicos no país, e que os preços de Sorriso permanecessem mais isolados, afetando relativamente pouco o processo de formação interna de preços.

Adicionalmente à análise das relações entre os preços no longo-prazo, verificou-se o comportamento dos preços também no curto-prazo. Os resultados do teste GSADF permitiram identificar períodos de comportamento explosivo nos preços físicos e futuros. Foi possível constatar que os mercados da região Sul registraram 83 episódios de bolhas especulativas, enquanto o Centro-Oeste registou 67, no período analisado. A maioria dos episódios de bolhas nos preços duraram menos de quinze dias nas duas regiões, e foram detectados poucos casos de bolhas que persistiram por um longo tempo, podendo ser associados à conjuntura do mercado de milho na

época.

A descrição do comportamento dos preços dentro de um período explosivo pode ser de grande utilidade para os diferentes operadores do mercado, pois pode influenciar as estratégias de negociação nos diferentes locais estudados. Uma vez que foram encontradas bolhas positivas e negativas, *hedgers*, especuladores, e entidades governamentais podem utilizar as informações sobre as variações dos preços nos momentos de alta e de baixa para negociar ou estocar a *commodity* de maneira mais eficiente. Por exemplo, se um *hedger* identificar o início de uma bolha positiva, ele pode aguardar para realizar suas vendas no topo dos preços, antes que estes retomem aos seus padrões normais.

Para complementar a análise feita neste trabalho, sugere-se que trabalhos futuros analisem também o comportamento das volatilidades. Com isso, pode ser possível identificar como os diferentes mercados reagem a mudanças nos preços. Este comportamento pode ser relacionado às características de liderança na informação, e no processo de descoberta de preços. Por fim, uma análise mais detalhada sobre os episódios de bolhas nos preços pode ser feita avaliando as causas deste fenômeno. Dentre as possíveis variáveis que podem ser utilizadas na tentativa de se explicar as bolhas especulativas nos mercados físicos, acredita-se que o *convenience yield* pode ter um papel fundamental, e merece ser melhor explorado futuramente.

## Referências

- BAILLIE, R. T.; BOOTH, G. G.; TSE, Y.; ZABOTINA, T. Price discovery and common factor models. **Journal of financial markets**, Elsevier, v. 5, n. 3, p. 309–321, 2002.
- BLOOMBERG L.P. **Terminal de dados históricos**. 2016.
- BOBENRIETH, E. S.; BOBENRIETH, J. R.; WRIGHT, B. D. Bubble troubles? rational storage, mean reversion, and runs in commodity prices. In: **The Economics of Food Price Volatility**. [S.l.]: University of Chicago Press, 2014. p. 193–208.
- BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 1287–1294, 1979.
- BROCKMAN, P.; TSE, Y. Information shares in canadian agricultural cash and futures markets. **Applied Economics Letters**, Taylor & Francis, v. 2, n. 10, p. 335–338, 1995.
- BROOKS, C.; PROKOPCZUK, M.; WU, Y. Booms and busts in commodity markets: bubbles or fundamentals? **Journal of Futures Markets**, Wiley Online Library, v. 35, n. 10, p. 916–938, 2015.
- BUENO, R. D. L. S. **Econometria de séries temporais**. 2. ed. [S.l.]: Cengage Learning, 2008.
- CASPI, I. Testing for a housing bubble at the national and regional level: the case of israel. **Empirical Economics**, Springer, v. 51, n. 2, p. 483–516, 2016.
- CEPEA. **Análise conjuntural-Milho**. [S.l.], 2015. Disponível em: <<https://www.cepea.esalq.usp.br/upload/revista/pdf/0362539001466184671.pdf>>. Acesso em: 22 jul. 2018.
- CEPEA. **Análise conjuntural-Milho**. [S.l.], 2016. Disponível em: <<https://www.cepea.esalq.usp.br/upload/revista/pdf/0491195001491922653.pdf>>. Acesso em: 22 jul. 2018.
- CHAKRAVARTY, S.; GULEN, H.; MAYHEW, S. Informed trading in stock and option markets. **The Journal of Finance**, Wiley Online Library, v. 59, n. 3, p. 1235–1257, 2004.
- CHAN, K. A further analysis of the lead–lag relationship between the cash market and stock index futures market. **The Review of Financial Studies**, Oxford University Press, v. 5, n. 1, p. 123–152, 1992.
- CHANG, T.; GIL-ALANA, L.; AYE, G. C.; GUPTA, R.; RANJBAR, O. Testing for bubbles in the brics stock markets. **Journal of Economic Studies**, Emerald Group Publishing Limited, v. 43, n. 4, p. 646–660, 2016.
- CHIODI, L. **Integração espacial no mercado brasileiro de milho**. 2006. Dissertação. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/11/11132/tde-13112006-082158/en.php>>.
- CHIODI, L.; GERALDINI JR., E. A.; MARQUES, P. V.; MARTINES FILHO, J. G. Análise da efetividade de *hedging* com os contratos futuros de soja na bm&f e cbot. In: **Congresso Brasileiro de Economia e Sociologia Rural, XLIII**. [s.n.], 2005. Disponível em: <<http://www.sober.org.br/palestra/2/527.pdf>>.

CHU, Q. C.; HSIEH, W.-I. G.; TSE, Y. Price discovery on the s&p 500 index markets: An analysis of spot index, index futures, and spdrs. **International Review of Financial Analysis**, Elsevier, v. 8, n. 1, p. 21–34, 1999.

CONAB. **Séries Históricas**. 2017. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1252&t=2>>.

CRUZ, J. C.; SILVA, G. H. D.; PEREIRA FILHO, I. A.; GONTIJO NETO, M. M.; MAGALHÃES, P. C. Caracterização do cultivo de milho safrinha de alta produtividade em 2008 e 2009. **Revista Brasileira de Milho e Sorgo**, v. 9, n. 2, p. 177–188, 2011.

CRUZ JR., J. C.; SILVEIRA, R. L.; CAPITANI, D. H. D.; URSO, F. S.; MARTINES FILHO, J. G. The effect of brazilian corn and soybean crop expansion on price and volatility transmission. In: AGRICULTURAL AND APPLIED ECONOMICS ASSOCIATION. **2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts**. [S.l.], 2016.

DELFINO, D. A. L. **Ensaio em dívida soberana**. 2012. Tese (Doutorado) — Fundação Getúlio Vargas, 2012. Disponível em: <<http://bibliotecadigital.fgv.br/dspace/handle/10438/9900>>. Acesso em: 05 abr. 2017.

DIAMOND, D. W.; VERRECCHIA, R. E. Constraints on short-selling and asset price adjustment to private information. **Journal of Financial Economics**, Elsevier, v. 18, n. 2, p. 277–311, 1987.

DIBA, B. T.; GROSSMAN, H. I. Explosive rational bubbles in stock prices? **The American Economic Review**, JSTOR, v. 78, n. 3, p. 520–530, 1988.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American statistical association**, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.

DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 1057–1072, 1981.

ENDERS, W. **Applied Econometric Time Series**. 4. ed. [S.l.]: John Wiley & Sons, 2014.

ENGLE, R. F.; GRANGER, C. W. Co-integration and error correction: representation, estimation, and testing. **Econometrica: journal of the Econometric Society**, JSTOR, p. 251–276, 1987.

ENGSTED, T.; HVIID, S. J.; PEDERSEN, T. Q. Explosive bubbles in house prices? evidence from the oecd countries. **Journal of International Financial Markets, Institutions and Money**, Elsevier, v. 40, p. 14–25, 2016.

ETIENNE, X. L.; IRWIN, S. H.; GARCIA, P. Bubbles in food commodity markets: Four decades of evidence. **Journal of International Money and Finance**, Elsevier, v. 42, p. 129–155, 2014.

EVANS, G. W. Pitfalls in testing for explosive bubbles in asset prices. **The American Economic Review**, JSTOR, v. 81, n. 4, p. 922–930, 1991.

FAS/USDA. **Foreign Agricultural Service/USDA**. 2018. Disponível em: <<https://apps.fas.usda.gov/psdonline/circulars/grain-corn-coarsegrains.pdf>>.

- GARDEBROEK, C.; HERNANDEZ, M. A. Do energy prices stimulate food price volatility? examining volatility transmission between us oil, ethanol and corn markets. **Energy economics**, Elsevier, v. 40, p. 119–129, 2013.
- GONZALO, J.; GRANGER, C. Estimation of common long-memory components in cointegrated systems. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 13, n. 1, p. 27–35, 1995.
- GREENE, W. H. **Econometric analysis**. [S.l.]: Pearson Education India, 2003.
- GUTIERREZ, L. Speculative bubbles in agricultural commodity markets. **European Review of Agricultural Economics**, Oxford University Press, v. 40, n. 2, p. 217–238, 2012.
- HASBROUCK, J. One security, many markets: Determining the contributions to price discovery. **The journal of Finance**, Wiley Online Library, v. 50, n. 4, p. 1175–1199, 1995.
- HATEMI-J, A.; IRANDOUST, M. Bilateral trade elasticities: Sweden versus her major trading partners. **American Review of Political Economy**, v. 3, n. 2, p. 38–50, 2005.
- HIERONYMUS, T. A. Economics of futures trading for commercial and personal profit. p. 95–175, 1977.
- IRWIN, S. H.; SANDERS, D. R. Financialization and structural change in commodity futures markets. **Journal of agricultural and applied economics**, Cambridge University Press, v. 44, n. 3, p. 371–396, 2012.
- JOHANSEN, S.; JUSELIUS, K. Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. **Oxford Bulletin of Economics and statistics**, Wiley Online Library, v. 52, n. 2, p. 169–210, 1990.
- JONG, F. D. Measures of contributions to price discovery: a comparison. **Journal of Financial markets**, Elsevier, v. 5, n. 3, p. 323–327, 2002.
- JONG, F. D.; MAHIEU, R.; SCHOTMAN, P. Price discovery in the foreign exchange market: An empirical analysis of the yen/dmark rate. **Journal of International Money and Finance**, Elsevier, v. 17, n. 1, p. 5–27, 1998.
- KIM, M. K.; TEJEDA, H.; WRIGHT, J. Price discovery in the us milled rice markets using a cluster analysis and tournament. In: AGRICULTURAL AND APPLIED ECONOMICS ASSOCIATION. **2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts**. [S.l.], 2016.
- KOLB, R. W.; OVERDAHL, J. A. **Understanding Futures Markets**. 6. ed. [S.l.]: Blackwell Publishing, 2006.
- LAZZARINI, S. Gerenciamento de riscos de preços no sistema agroindustrial brasileiro da soja. **Revista de Economia e Sociologia Rural, Brasília, SOBER**, p. 41–64, 1999.
- LEHMANN, B. N. Some desiderata for the measurement of price discovery across markets. **Journal of Financial Markets**, North-Holland, v. 5, n. 3, p. 259–276, 2002.
- LI, J.; CHAVAS, J.-P.; ETIENNE, X. L.; LI, C. Commodity price bubbles and macroeconomics: evidence from the chinese agricultural markets. **Agricultural Economics**, Wiley Online Library, v. 48, n. 6, p. 755–768, 2017.

LIEN, D.; SHRESTHA, K. A new information share measure. **Journal of Futures Markets**, Wiley Online Library, v. 29, n. 4, p. 377–395, 2009.

LO, A. W.; MACKINLAY, A. C. Stock market prices do not follow random walks: Evidence from a simple specification test. **The review of financial studies**, Oxford University Press, v. 1, n. 1, p. 41–66, 1988.

MARQUES, P. V.; MELLO, P. C.; MARTINES FILHO, J. G. Mercados futuros e de opções agropecuárias. **Departamento de Economia, Administração e Sociologia da Esalq-USP**, 2006.

MATTOS, F. L.; SILVEIRA, R. L. The effects of brazilian second (winter) corn crop on price seasonality, basis behavior and integration to international market. In: **Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management**. St. Louis, MO. [s.n.], 2015. Disponível em: <[http://www.farmdoc.illinois.edu/nccc134/conf\\_2015/pdf/Mattos\\_Silveira\\_NCCC\\_134\\_2015.pdf](http://www.farmdoc.illinois.edu/nccc134/conf_2015/pdf/Mattos_Silveira_NCCC_134_2015.pdf)>.

MAYNARD, L. J.; HANCOCK, S.; HOAGLAND, H. Performance of shrimp futures markets as price discovery and hedging mechanisms. **Aquaculture Economics & Management**, Taylor & Francis, v. 5, n. 3-4, p. 115–128, 2001.

METZ, J. **Interpretação de clusters gerados por algoritmos de clustering hierárquico**. 2006. Tese (Doutorado) — Universidade de São Paulo, 2006.

MIN, J. H.; NAJAND, M. A further investigation of the lead–lag relationship between the spot market and stock index futures: Early evidence from korea. **Journal of Futures Markets**, Wiley Online Library, v. 19, n. 2, p. 217–232, 1999.

MONTASSER, G. E.; GUPTA, R.; MARTINS, A. L.; WANKE, P. Are there multiple bubbles in the ethanol–gasoline price ratio of brazil? **Renewable and Sustainable Energy Reviews**, Elsevier, v. 52, p. 19–23, 2015.

PHILLIPS, P.; SHI, S.; YU, J. Testing for multiple bubbles. 2012.

PHILLIPS, P. C.; SHI, S.; YU, J. Specification sensitivity in right-tailed unit root testing for explosive behaviour. **Oxford Bulletin of Economics and Statistics**, Wiley Online Library, v. 76, n. 3, p. 315–333, 2014.

PHILLIPS, P. C.; SHI, S.; YU, J. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. **International Economic Review**, Wiley Online Library, v. 56, n. 4, p. 1043–1078, 2015.

PHILLIPS, P. C.; WU, Y.; YU, J. Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? **International economic review**, Wiley Online Library, v. 52, n. 1, p. 201–226, 2011.

PINDYCK, R. S. The dynamics of commodity spot and futures markets: a primer. **The energy journal**, JSTOR, p. 1–29, 2001.

PUTNINŠ, T. J. What do price discovery metrics really measure? **Journal of Empirical Finance**, Elsevier, v. 23, p. 68–83, 2013.

REIS, J. G. M. d.; VENDRAMETTO, O.; NAAS, I. d. A.; COSTABILE, L. T.; MACHADO, S. T. Avaliação das estratégias de comercialização do milho em ms aplicando o analytic hierarchy process (ahp). **Revista de Economia e Sociologia Rural**, SciELO Brasil, v. 54, n. 1, p. 131–146, 2016.

SILVA, R. M. d. **Relações no mercado internacional de soja em grão: Preços, volatilidades e fluxo de informações**. 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/11/11132/tde-25032013-151438/en.php>>. Acesso em: 13 ago. 2016.

SOUZA, W. A. Determinantes da viabilidade de mercados futuros agropecuários no âmbito do mercosul. **Piracicaba: Escola Superior de Agricultura Luiz de Queiroz**, 1998. Dissertação.

STIGLITZ, J. E. Symposium on bubbles. **Journal of economic perspectives**, v. 4, n. 2, p. 13–18, 1990.

STOCK, J. H.; WATSON, M. W. Variable trends in economic time series. **The Journal of Economic Perspectives**, JSTOR, v. 2, n. 3, p. 147–174, 1988.

SU, Q.; CHONG, T. T.-L. Determining the contributions to price discovery for chinese cross-listed stocks. **Pacific-Basin Finance Journal**, Elsevier, v. 15, n. 2, p. 140–153, 2007.

THEISSEN, E. Price discovery in floor and screen trading systems. **Journal of Empirical Finance**, Elsevier, v. 9, n. 4, p. 455–474, 2002.

TRUJILLO-BARRERA, A.; MALLORY, M.; GARCIA, P. Volatility spillovers in us crude oil, ethanol, and corn futures markets. **Journal of Agricultural and Resource Economics**, JSTOR, p. 247–262, 2012.

USDA. **World Agriculture Supply and Demand Estimates**. 2017. Disponível em: <[https://www.usda.gov/oce/commodity/wasde/Secretary\\_Briefing/Archive/2017/June\\_2017\\_WASDE\\_Lockup\\_Briefing.pdf](https://www.usda.gov/oce/commodity/wasde/Secretary_Briefing/Archive/2017/June_2017_WASDE_Lockup_Briefing.pdf)>.

WHITE, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 817–838, 1980.

XU, X. Price discovery in us corn cash and futures markets: The role of cash market selection. In: AGRICULTURAL AND APPLIED ECONOMICS ASSOCIATION. **2014 Annual Meeting, July 27-29, 2014, Minneapolis, Minnesota**. [S.l.], 2014.

YAN, B.; ZIVOT, E. A structural analysis of price discovery measures. **Journal of Financial Markets**, Elsevier, v. 13, n. 1, p. 1–19, 2010.

YANG, J.; BESSLER, D.; LEATHAM, D. J. Asset storability and price discovery of commodity futures markets: A new look. **The Journal of Futures Markets**, v. 21, n. 3, p. 279–300, 2001.

ZEILEIS, A.; KLEIBER, C.; KRÄMER, W.; HORNIK, K. Testing and dating of structural changes in practice. **Computational Statistics & Data Analysis**, Elsevier, v. 44, n. 1-2, p. 109–123, 2003.

## Anexo

Tabela 1A – Tamanho da amostra e da janela ( $r_w$ ) para cada ano de teste GSADF

Ano	Tamanho da amostra	Tamanho da janela ( $r_w$ )
2004	220	29
2005	275	33
2006	268	32
2007	263	32
2008	269	32
2009	272	32
2010	272	32
2011	272	32
2012	270	32
2013	272	32
2014	267	32
2015	273	32
2016	276	33
Total	3469	

Fonte: Elaborado pela autora.

Tabela 2A – Testes de cointegração de Johansen para o período de 2004 a 2010

Mercados	Estatísticas		
	Rank	Traço	Máximo autovalor
Pca vs. Prv	$r = 0$	23,48**	21,36**
	$r \leq 1$	2,12 <sup>ns</sup>	2,12 <sup>ns</sup>
Pch vs. Prv	$r = 0$	12,97 <sup>ns</sup>	10,73 <sup>ns</sup>
	$r \leq 1$	2,24 <sup>ns</sup>	2,24 <sup>ns</sup>
Prv vs. Pso	$r = 0$	43,85***	42,30***
	$r \leq 1$	1,54 <sup>ns</sup>	1,54 <sup>ns</sup>
Pch vs. Pso	$r = 0$	13,58	11,06
	$r \leq 1$	2,52 <sup>ns</sup>	2,52 <sup>ns</sup>

Fonte: Resultados da pesquisa.

Nota: Erro-padrão entre parênteses. \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e <sup>ns</sup> indica que o teste foi não significativo.

Tabela 3A – Testes de cointegração de Johansen para o período de 2011 a 2016

Estatísticas			
Mercados	Rank	Traço	Máximo autovalor
Pca vs. Prv	$r = 0$	34,66***	31,09***
	$r \leq 1$	3,57 <sup>ns</sup>	3,57 <sup>ns</sup>
Pca vs. Pch	$r = 0$	29,95***	27,10***
	$r \leq 1$	2,85 <sup>ns</sup>	2,85 <sup>ns</sup>
Pca vs. Pso	$r = 0$	14,78 <sup>ns</sup>	12,94 <sup>ns</sup>
	$r \leq 1$	1,84 <sup>ns</sup>	1,84 <sup>ns</sup>

Fonte: Resultados da pesquisa.

Nota: Erro-padrão entre parênteses. \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1%, respectivamente.

Tabela 4A – Modelo de correção de erros entre os preços de Cascavel ( $Pca$ ) e Rio Verde ( $Prv$ )

Variáveis independentes	Modelo 1	
	$\Delta Pca_t$	$\Delta Prv_t$
$ect_{1,t-1}$	-0,0110*** (0,0027)	0,0189*** (0,0038)
$\Delta Pca_{t-1}$	0,0870*** (0,0180)	0,0186 <sup>ns</sup> (0,0128)
$\Delta Pca_{t-2}$	0,2051*** (0,0181)	0,0408*** (0,0131)
$\Delta Pca_{t-3}$	0,1221*** (0,0185)	0,0270** (0,0130)
$\Delta Pca_{t-4}$	0,0578*** (0,0184)	0,0299** (0,0129)
$\Delta Pca_{t-5}$	0,0798*** (0,0184)	0,0075 <sup>ns</sup> (0,0125)
$\Delta Prv_{t-1}$	0,1007*** (0,0253)	-0,2478 <sup>ns</sup> (0,0179)
$\Delta Prv_{t-2}$	0,1443*** (0,0254)	-0,0470** (0,0184)
$\Delta Prv_{t-3}$	0,1109*** (0,0259)	0,0107 <sup>ns</sup> (0,0183)
$\Delta Prv_{t-4}$	0,0866*** (0,0258)	0,0626*** (0,0181)
$\Delta Prv_{t-5}$	0,0845*** (0,0258)	0,0860*** (0,0176)
Constante	-0,3591	
Relação de longo prazo	1,0000	-0,9051

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 5A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Chapecó ( $Pch$ )

Variáveis independentes	Modelo 2	
	$\Delta Pca_t$	$\Delta Pch_t$
$ect_{1,t-1}$	-0,0172*** (0,0038)	0,0245*** (0,0042)
$\Delta Pca_{t-1}$	0,0766*** (0,0182)	0,0719*** (0,0162)
$\Delta Pca_{t-2}$	0,1922*** (0,0185)	0,0663*** (0,0168)
$\Delta Pca_{t-3}$	0,1100*** (0,0188)	0,0481*** (0,0168)
$\Delta Pca_{t-4}$	0,0490*** (0,0188)	0,0474*** (0,0166)
$\Delta Pca_{t-5}$	0,0715*** (0,0187)	0,0111 <sup>ns</sup> (0,0158)
$\Delta Pch_{t-1}$	0,1810*** (0,0204)	-0,2773*** (0,0181)
$\Delta Pch_{t-2}$	0,1559*** (0,0206)	-0,1238*** (0,0187)
$\Delta Pch_{t-3}$	0,1261*** (0,0210)	-0,0680*** (0,0187)
$\Delta Pch_{t-4}$	0,1320*** (0,0210)	-0,0371** (0,0185)
$\Delta Pch_{t-5}$	0,1123*** (0,0209)	-0,0208 <sup>ns</sup> (0,0177)
Constante	0,2077	
Relação de longo prazo	1,0000	-1,0192

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 6A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Sorriso ( $Pso$ )

Variáveis independentes	Modelo 3	
	$\Delta Pca_t$	$\Delta Pso_t$
$ect_{1,t-1}$	-0,0063*** (0,0020)	0,0187*** (0,0045)
$\Delta Pca_{t-1}$	0,0836*** (0,0180)	0,0245*** (0,0080)
$\Delta Pca_{t-2}$	0,2039*** (0,0181)	0,0146* (0,0082)
$\Delta Pca_{t-3}$	0,1246*** (0,0183)	0,0141* (0,0081)
$\Delta Pca_{t-4}$	0,0604*** (0,0181)	0,0073 <sup>ns</sup> (0,0081)
$\Delta Pca_{t-5}$	0,0786*** (0,0181)	0,0174** (0,0079)
$\Delta Pso_{t-1}$	0,1742*** (0,0403)	-0,1955*** (0,0181)
$\Delta Pso_{t-2}$	0,0476 <sup>ns</sup> (0,0405)	-0,0573*** (0,0183)
$\Delta Pso_{t-3}$	0,1332*** (0,0411)	-0,0405** (0,0183)
$\Delta Pso_{t-4}$	0,1996*** (0,0407)	0,0052 <sup>ns</sup> (0,0183)
$\Delta Pso_{t-5}$	0,1194*** (0,0407)	0,0224 <sup>ns</sup> (0,0178)
Constante	-1,0685	
Relação de longo prazo	1,0000	-0,7627

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 7A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Rio Verde ( $Prv$ ) no período de 2004 a 2010

Variáveis independentes	Variáveis dependentes	
	$\Delta Pca_t$	$\Delta Prv_t$
$ect_{1,t-1}$	-0,0131*** (0,0038)	0,0136** (0,0055)
$\Delta Pca_{t-1}$	0,1392*** (0,0245)	0,1146*** (0,0357)
$\Delta Pca_{t-2}$	0,1985*** (0,0244)	0,1986*** (0,0355)
$\Delta Pca_{t-3}$	0,1672*** (0,0249)	0,1920*** (0,0362)
$\Delta Prv_{t-1}$	0,0054 <sup>ns</sup> (0,0170)	-0,2658*** (0,0248)
$\Delta Prv_{t-2}$	0,0421** (0,0174)	-0,0595** (0,0253)
$\Delta Prv_{t-3}$	0,0387** (0,0169)	-0,0130 <sup>ns</sup> (0,0245)
Constante	-0,4613	
Relação de longo prazo	1,0000	-0,8646

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 8A – Modelo de correção de erros entre Rio Verde ( $Prv$ ) e Sorriso ( $Pso$ )

Variáveis independentes	Variáveis dependentes	
	$\Delta Prv_t$	$\Delta Pso_t$
$ect_{1,t-1}$	-0,0027 <sup>ns</sup> (0,0053)	0,0537*** (0,0082)
$\Delta Prv_{t-1}$	-0,2040*** (0,0243)	0,0250 <sup>ns</sup> (0,0381)
$\Delta Pso_{t-1}$	0,0577*** (0,0153)	-0,1617*** (0,0240)
Constante	-0,9050	
Relação de longo prazo	1,0000	-0,7809

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 9A – Modelo de correção de erros entre Cascavel ( $Pca$ ) e Rio Verde ( $Prv$ )

Variáveis independentes	Variáveis dependentes	
	$\Delta Pca_t$	$\Delta Prv_t$
$ect_{1,t-1}$	-0,0062 <sup>ns</sup> (0,0041)	0,0294*** (0,0056)
$\Delta Pca_{t-1}$	0,0523** (0,0265)	0,1291*** (0,0357)
$\Delta Pca_{t-2}$	0,2432*** (0,0266)	0,1122*** (0,0358)
$\Delta Pca_{t-3}$	0,1084*** (0,0268)	0,0588 <sup>ns</sup> (0,0361)
$\Delta Pca_{t-4}$	0,0703*** (0,0268)	0,1475*** (0,0361)
$\Delta Prv_{t-1}$	0,0542*** (0,0192)	-0,1966*** (0,0258)
$\Delta Prv_{t-2}$	0,0627*** (0,0195)	-0,0098 <sup>ns</sup> (0,0263)
$\Delta Prv_{t-3}$	0,0179 <sup>ns</sup> (0,0194)	0,0449* (0,0262)
$\Delta Prv_{t-4}$	0,0442** (0,0190)	0,1002*** (0,0256)
Constante	-0,5239	
Relação de longo prazo	1,0000	-0,8570

Fonte: Resultados da pesquisa.

: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 10A – Modelo de correção de erros entre Pca e Pch entre 2011 e 2016

Variáveis independentes	Variáveis dependentes	
	$\Delta Pca_t$	$\Delta Pch_t$
$ect_{1,t-1}$	-0,0096** (0,0049)	0,0255*** (0,0057)
$\Delta Pca_{t-1}$	0,0430 <sup>ns</sup> (0,0267)	0,1506*** (0,0311)
$\Delta Pca_{t-2}$	0,2358*** (0,0267)	0,2283*** (0,0312)
$\Delta Pca_{t-3}$	0,0949*** (0,0273)	0,1284*** (0,0318)
$\Delta Pca_{t-4}$	0,0571** (0,0272)	0,1300*** (0,0317)
$\Delta Pch_{t-1}$	0,0744*** (0,0226)	-0,2715*** (0,0263)
$\Delta Pch_{t-2}$	0,0783*** (0,0234)	-0,1311*** (0,0273)
$\Delta Pch_{t-3}$	0,0594** (0,0231)	-0,0760*** (0,0269)
$\Delta Pch_{t-4}$	0,0649*** (0,0222)	-0,0150 <sup>ns</sup> (0,0259)
Constante	0,4179	
Relação de longo prazo	1,0000	-1,0779

Fonte: Resultados da pesquisa.

Nota: \*, \*\*, \*\*\* indicam rejeição da hipótese nula aos níveis de significância de 10%, 5%, e 1% respectivamente, e *ns* indica que o teste foi não significativo. Os *lags* de defasagem foram selecionados com base no critério de Schwarz. Erro-padrão entre parênteses.

Tabela 1B – Medidas de descoberta de preços considerando o mercado de futuros entre 2004 a 2016

Indicadores	<i>Pso vs. Pf</i>	<i>Pca vs. Pf</i>	<i>Pch vs. Pf</i>	<i>Prv vs. Pf</i>
IS <sub>1</sub>	0,09	0,68	0,12	0,05
IS <sub>2</sub>	<b>0,91</b>	<b>0,61</b>	<b>0,88</b>	<b>0,95</b>
CS <sub>1</sub>	0,14	<b>0,54</b>	0,30	0,13
CS <sub>2</sub>	<b>0,86</b>	0,46	<b>0,70</b>	<b>0,87</b>
ILS <sub>1</sub>	0,40	0,23	0,10	0,11
ILS <sub>2</sub>	<b>0,60</b>	<b>0,77</b>	<b>0,90</b>	<b>0,89</b>
IL <sub>1</sub>	0,82	0,55	0,34	0,35
IL <sub>2</sub>	<b>1,21</b>	<b>1,81</b>	<b>2,96</b>	<b>2,85</b>
Líder	<i>Pf</i>	<i>Pf</i>	<i>Pf</i>	Pf

Fonte: Resultados da pesquisa.