



Ministério da Justiça e Cidadania - MJC
Conselho Administrativo de Defesa Econômica - CADE

SEPN 515 Conjunto D, Lote 4 Ed. Carlos Taurisano, 4º andar - Bairro Asa Norte, Brasília/DF, CEP 70770-504
Telefone: (61) 3221-8409 e Fax: (61) 3326-9733 - www.cade.gov.br

NOTA TÉCNICA Nº 46/2016/DEE/CADE

Referência: Processo Administrativo nº 08012.002921/2007-64

Assunto: Exercício empírico para detecção de comportamento suspeito de colusão entre concorrentes em 484 licitações públicas para aquisição de cloro em 23 estados da União durante o período de 1999 a 2012.

Conclusão: A análise empírica indica a existência de comportamento colusivo nas licitações públicas para aquisição de cloro nos certames envolvendo as empresas Carbocloro e Canexus. As estimativas apontam que o sobrepreço praticado variou entre 21% e 25%.

Versão: Pública.

1. Introdução

A literatura econômica sobre Organização Industrial mais especificamente relacionada à Economia Antitruste fornece dois métodos alternativos para filtrar mercados ou operações com maiores probabilidades de ocorrência de comportamentos colusivos entre competidores: a abordagem estrutural e a abordagem comportamental. A abordagem estrutural busca identificar em quais estruturas de mercado o comportamento colusivo entre competidores tende a ser mais

provável, por exemplo, mercados com poucos concorrentes, com produtos homogêneos e com demanda estável^[1]. Já a abordagem comportamental busca identificar por quais meios as firmas coordenam suas ações (Harrington, 2008 p. 213)^[2].

O objetivo da presente nota técnica é providenciar um exercício empírico para detecção de comportamento suspeito de colusão entre concorrentes em 484 licitações públicas para aquisição de cloro em 23 estados da União, durante o período de 1999 a 2012. A ideia é filtrar um conjunto de processos licitatórios cujos resultados (preços vencedores) diverjam sistematicamente do comportamento de resultados compatíveis com nossas hipóteses de estrutura não colusiva, ou seja, compatíveis com nosso comportamento competitivo contrafactual.

Esta nota está dividida em seis seções, incluindo a presente introdução. A segunda seção apresenta os dados utilizados no tratamento empírico, bem como detalhes do tratamento da base de dados. A terceira seção é dedicada à apresentação da estratégia de identificação do exercício empírico, ou seja, quais as hipóteses levantadas acerca da delimitação do mercado relevante de cloro, a estratégia de seleção dos certames suspeitos, os testes necessários para identificar se a distribuição dos preços vencedores nos certames suspeitos realmente não tenham sido gerados pelo mesmo processo gerador dos preços vencedores nos certames não suspeitos e os métodos utilizados para o cômputo do sobrepreço de cartel. A quarta seção reporta os resultados empíricos acerca do processo de identificação dos certames suspeitos de colusão. A quinta seção reporta os resultados empíricos acerca do processo de cômputo do sobrepreço de cartel. A sexta e última seção é dedicada aos comentários finais e conclusões.

2. Base de Dados

Em 10 de outubro de 2014 as principais companhias de tratamento de água e saneamento de cada estado do Brasil receberam pedidos de informações acerca de suas compras de cloro puro liquefeito no período de 1999 a 2012. Foram solicitados dados de fornecedores, tipo de licitação, data da compra, quantidade comprada, tipo de envase, fabricante e, em caso de leilão, número de lances e preços finais de cada empresa participante. As companhias responderam aos ofícios no período de novembro de 2014 a janeiro de 2015. As companhias CAERD (Rondônia) e COMPESA (Pernambuco) responderam aos ofícios, mas não informaram dados sobre suas compras. A Manaus Ambiental (Amazônia) ressaltou que é uma empresa privada e realiza compras diretas, não dando maiores informações sobre suas compras. A DEPASA (Acre) comunicou que não trabalha com cloro gás, recorrendo a técnicas de tratamento que o dispensam.

De posse dos dados enviados pelas companhias de tratamento de água e saneamento, foi elaborada uma base de dados, contendo, para cada compra: preço, quantidade, ano, unidade da federação, fornecedor e distribuidor. Além disso, de posse das informações dos endereços das plantas de fabricação, foi consultado o Google Maps^[3] e inserida informação de distância entre as fabricantes dos produtos e a principal estação de tratamento de águas de cada companhia de saneamento^[4], referenciada por compra. A partir da distância obtida e do tempo de viagem foi calculada também a velocidade média de deslocamento por veículo automotor no trajeto^[5].

Ainda, os preços foram deflacionados segundo o índice IPA-Geral da Fundação Getúlio Vargas, com base nos preços de 1999, e foram elaboradas 4 variáveis *dummy* referentes a cada tipo de envase nas compras. Foi criada também uma última *dummy*, referente a se o Estado da companhia de saneamento que realizou a compra era diferente da fabricante do produto. Tais informações foram então adicionadas à base de dados, que foi consolidada em formato de tabela, posteriormente exportada ao programa Stata 12 (“Stata”).

Uma vez que os dados foram adicionados ao Stata, foi constatado que existiam 484 observações (distribuídas por ano e unidade da federação conforme tabela abaixo), sendo que em 14 dos casos não havia informações acerca do fabricante do produto e, conseqüentemente, de distância e velocidade média.

Tabela 1. Observações por Ano

Ano	Nº de Observações	Percentual
1999	7	1,45%
2000	33	6,82%
2001	33	6,82%
2002	40	8,26%
2003	35	7,23%
2004	41	8,47%
2005	38	7,85%
2006	35	7,23%
2007	52	10,74%
2008	44	9,09%

2009	44	9,09%
2010	43	8,88%
2011	37	7,64%
Total	484	100%

Tabela 2. Observações por Unidade da Federação

Alagoas	4	0,83%
Amapá	7	1,45%
Bahia	17	3,51%
Ceará	9	1,86%
Distrito Federal	36	7,44%
Espírito Santo	17	3,51%
Goiás	32	6,61%
Maranhão	8	1,65%

Mato Grosso	13	2,69%
Mato Grosso do Sul	8	1,65%
Minas Gerais	26	5,37%
Pará	2	0,41%
Paraíba	25	5,17%
Paraná	22	4,55%
Piauí	20	4,13%
Rio de Janeiro	31	6,40%
Rio Grande do Norte	8	1,65%
Rio Grande do Sul	13	2,69%
Roraima	8	1,65%
Santa Catarina	20	4,13%
São Paulo	46	9,50%

Sergipe	4	0,83%
Tocantins	108	22,31%
Total	484	100%

Dando prosseguimento, foi necessário um tratamento dos dados de forma a torná-los mais adequados aos processos econométricos que viriam a ser realizados. Nesse sentido, verificou-se que a distribuição dos preços já deflacionados não seguia uma distribuição normal, sendo necessária a criação de uma nova variável, correspondente ao logaritmo natural^[6] dos preços deflacionados^[7]. Processo semelhante foi realizado com as variáveis quantidade, distância e velocidade.

Cuidado especial foi tomado com a variável distância, pois, segundo os Representados, esta é componente crucial na estrutura de custo do cloro gás, sendo diretamente proporcional ao custo de venda deste. Assim, foram criadas as variáveis distância^2 e $\ln(\text{distância})^2$, equivalentes, respectivamente, ao quadrado da distância e do logaritmo natural da distância ao quadrado. Tal processo visa permitir verificar, no processo de regressão, o efeito marginal da distância no preço dos produtos, de forma a conhecer se o efeito da distância no preço aumenta ou diminui de acordo com a própria distância.

Por fim, visando permitir obter variações anuais no preço^[8], bem como as variações de preço para cada estado ofertante e fabricante de cloro, foi criada uma *dummy* para cada ano, UF, fornecedor e fabricante.

Como resultado final do processo de tratamento dos dados, foram obtidas 83 variáveis, refletindo as informações fornecidas pelas companhias de tratamento de água e saneamento e as pesquisas realizadas internamente. Tais dados foram a base para os estudos econométricos realizados, detalhados a seguir.

3. Estratégia de Identificação

O primeiro passo da nossa estratégia de identificação é delimitar o mercado relevante para o produto em análise. Utilizamos a hipótese de que o mercado é nacional^[9] na dimensão geográfica e cloro na dimensão produto. Apesar de considerarmos inicialmente o mercado como sendo nacional, o método de análise incorpora o grau de dispersão decorrente de fatores logísticos. Dessa forma, a análise empreendida leva em consideração o perfil regionalizado dos mercados ao incluir variáveis que controlam pelas características regionais. Ainda na dimensão produto consideramos que o mesmo seja homogêneo, sem qualquer grau de diferenciação que justifique delimitar ainda mais o produto em diferentes qualidades ou nichos. Em suma, uma *commodity*, como recorrentemente o produto é apresentado ao CADE pelos produtores quando da submissão de atos de concentração.

A delimitação do mercado relevante é importante em nossa estratégia de identificação, pois a partir daí pode-se identificar quais as variáveis com potencial para explicar variações significativas de preços, que não sejam decorrentes de comportamento colusivo. Por exemplo, sabemos que em um mercado de produto homogêneo um produtor não poderia praticar preços sistematicamente superiores aos concorrentes baseado no argumento de diferenciação de produto.

Considerando os argumentos acima, iniciamos a construção dos nossos critérios para diferenciação de certames entre suspeitos e não suspeitos, que é composta por dois estágios. Em um primeiro estágio regredimos todos os preços vencedores dos certames em função de variáveis logísticas, ou seja, um vetor de variáveis que seguramente explicariam apenas variações de preços, que não guardariam qualquer relação com variações de preços decorrentes de comportamento colusivo. Formalmente:

$$P_i = \alpha + \sum_{i=1}^N \beta_i x_i + \varepsilon_i \quad [1]$$

Onde P_i é o preço vencedor do certame i , x é um vetor de variáveis para controlar a heterogeneidade em termos de logística do certame i , por exemplo, a distância em quilômetros entre o produtor e o comprador, a qualidade logística (velocidade média entre os dois pontos), forma de acondicionamento e transporte do material, etc. Já α e β são constantes paramétricas a serem estimadas por meio dos dados. Finalmente, ε é um termo de erro aleatório.

A partir das estimativas da equação [1] são obtidos valores preditos (\hat{P}_i) para cada um dos preços vencedores dos certames, bem como seus respectivos erros-padrão ($\sigma_{\hat{P}_i}$). Com tais informações são realizados testes estatísticos (teste t -Student) para verificar se cada um dos preços vencedores dos certames é estatisticamente superior ao seu respectivo preço estimado a partir da regressão [1], que controla a heterogeneidade dos preços vencedores em função das variáveis explicativas relacionadas a fatores logísticos. Dessa forma é construída uma nova variável (P_i^+), dicotômica (com valores iguais a zero ou um), a partir dos preços reais estatisticamente superiores aos preços estimados ($P_i^+ = 1$) e dos preços reais estatisticamente não superiores aos preços estimados ($P_i^+ = 0$). Formalmente a variável é construída da seguinte maneira:

$$P_i^+ = 1 \text{ se } \frac{\hat{P}_i - P_i}{\sigma_{\hat{P}_i}} < t_{n,gl} \quad \text{e} \quad P_i^+ = 0, \text{ caso contrário}$$

O teste t -Student é utilizado nos casos onde a variância populacional é desconhecida, como é o nosso caso. Na equação acima $t_{n,gl}$ é a estatística t de tabela para o número de observações (n) e graus de liberdade (gl) da amostra.

Cabe esclarecer que o fato de um certame apresentar preço vencedor acima do estimado ($P_i^+ = 1$) não sugere nenhum comportamento suspeito. A suspeita surge do fato de algum grupo de competidores praticarem preços sistematicamente superiores aos estimados. Por conta disso, faz-se necessário um segundo estágio de

exercício empírico para verificar se algum grupo de concorrentes sistematicamente vence leilões com preços estatisticamente superiores aos estimados a partir do controle dos fatores logísticos.

Como agora nossa variável de interesse é (P_i^+) , podemos estimar um comportamento suspeito a partir da obtenção dos *Odd Ratios* [10] de uma regressão logística (modelo LOGIT). Formalmente:

$$\text{Prob}(P_i^+ = 1 | \mathbf{z}) = G(\gamma_0 + \mathbf{z}\boldsymbol{\gamma}) \quad [2]$$

Onde $G(\bullet)$ é a função logística:

$$G(\mathbf{z}) = \frac{\exp(\mathbf{z})}{1 + \exp(\mathbf{z})} = \Lambda \mathbf{z}$$

Aqui temos $\boldsymbol{\gamma}$ como parâmetros a serem estimados a partir dos dados e \mathbf{z} é um vetor de variáveis *Dummy*, sendo uma variável *Dummy* (dicotômica) para cada empresa atuante nos certames realizados nos certames daquele mercado relevante.

A partir do segundo estágio detectamos as empresas suspeitas por meio de seus *Odd Ratios* da regressão logística, ou seja, aquelas empresas cuja razão entre a “probabilidade de preços estarem acima do estimado / probabilidade de preços não estarem acima do estimado” ser significativamente maior do que um.

Finalmente, construímos uma variável dicotômica chamada S_i , com valor igual a um para todos os certames com preços reais acima dos estimados ($P_i^+ = 1$) e com participação das empresas com *Odd Ratios* estatisticamente superiores à unidade (e com valor igual à zero, caso contrário).

Para assegurarmos que os certames computados como suspeitos ($S_i = 1$) divergem de fato daqueles computados como não suspeitos ($S_i = 0$), utilizamos o teste de *Kolmogorov-Smirnov* para igualdade de distribuições, ou seja, para verificar se o processo que gerou os preços (a distribuição) nos certames suspeitos é igual ao processo gerador de preços (distribuição) nos certames não suspeitos. Caso os processos sejam diferentes, a pergunta central a ser endereçada é: por que existiriam dois processos de precificação diferentes em um mesmo mercado relevante, com produtos homogêneos não diferenciados? A evidência de que os preços suspeitos e os não-suspeitos sejam obtidos de distribuições diferentes apenas reforçaria a evidência de que algo não está operando de forma adequada no mercado relevante.

Finalmente, computamos o sobrepreço de suposto cartel praticado nos certames suspeitos em relação aos não suspeitos. Aqui utilizamos duas hipóteses de identificação. Na primeira assumimos a hipótese de que os certames suspeitos são selecionados aleatoriamente, de forma exógena, de modo que possamos computar o sobrepreço a partir da seguinte equação:

$$P_i = \beta_0 + \beta_1 S_i + \vartheta_i \quad [3]$$

onde P_i é o preço vencedor do certame i , S_i é a variável dicotômica para a qual certames suspeitos são iguais à unidade (caso contrário, igual a zero), os betas são constantes a serem estimadas, sendo que β_1 traz o computo percentual do sobrepreço de suposto cartel e ϑ_i é o componente de termos de erro estocástico.

Na segunda hipótese assumimos que os certames não são escolhidos aleatoriamente, tampouco de forma exógena. Sob esta hipótese, a estimativa de β_1 da equação [3] estaria enviesada, ou seja, teríamos uma estimativa incorreta do sobrepreço de cartel. Para endereçarmos este problema é utilizado o método de pareamento por escore de propensão (*propensity score matching*) com o objetivo de encontrarmos junto aos certames não suspeitos aqueles com características bem próximas dos suspeitos, exceto pelo fator de tratamento (que é a própria suspeita). Assim o cômputo de sobrepreço é obtido a partir das diferenças de médias de preços entre os grupos de controle (certames não suspeito) e de tratamento (certames suspeitos).

4. Resultados

Como mencionado na seção anterior, nosso primeiro passo é estimar a equação [1], dos preços vencedores dos certames em função apenas de características que seguramente justifiquem variações de preços, mas que certamente não estejam associadas com eventuais comportamentos colusivos. Neste sentido regredimos o preço em função apenas das variáveis relativas à logística. Especificamos a regressão a partir da seguinte equação:

$$\ln P_i = \beta_0 + \beta_1 \text{carreta} + \beta_2 \text{cilindro40} + \beta_3 \text{cilindro50} + \beta_4 \text{cilindro68} + \beta_5 \text{cilindro900} + \beta_6 \ln \text{distância} + \beta_7 \ln \text{distância}^2 + \beta_8 \ln \text{velocidade} + \varepsilon_i \quad [4]$$

Os preços vencedores dos certames são especificados em logaritmos naturais por conta da normalidade da distribuição da variável nesta especificação (a distribuição dos preços é lognormal) e são regredidos em função da forma como o material é acondicionado (tipos de cilindros), por meio de qual tipo de transporte é comercializado (carreta ou qualquer outro tipo de transporte), da distância em quilômetros da planta do produtor até a unidade compradora, bem como de sua especificação quadrática, da velocidade normal do trajeto (busca capturar a qualidade das rodovias do trecho) e um termo de erro estocástico.

Para endereçar o problema da heterogeneidade dos dados, a equação [4] foi estimada a partir de uma regressão robusta [11]. Já o componente sistemático (conjunto de variáveis explicativas) da equação [4] apresentou um poder explicativo de quase 60% das variações nos preços vencedores dos certames, o que implicou em uma estatística F de valor igual a 113, com um p-valor inferior a 0,001. Em suma, o conjunto de variáveis logísticas fornece um bom poder preditivo para as variações dos preços vencedores dos certames.

A partir dos resultados da regressão [4] foi possível computar dois conjuntos de valores: (i) os preços estimados de $\ln P_i$ a partir da regressão, $\ln \hat{P}_i$; e (ii) os erros-padrão dos preços vencedores estimados, $\sigma_{\ln \hat{P}_i}$. Por meio destes dois conjuntos de valores foram feitos testes de hipóteses (teste *t*-Student) sobre a diferença

estatística entre os logaritmos dos preços reais e dos preços estimados para cada um dos certames, de modo que a construção da variável P_i^+ foi construída a partir da seguinte regra:

$$P_i^+ = 1 \text{ se } \frac{\ln \hat{P}_i - \ln P_i}{\sigma_{\ln \hat{P}_i}} < -2,326 \quad \text{e} \quad P_i^+ = 0, \text{ caso contrário}$$

O valor da estatística $t_{n,gl} = -2,326$ é obtido da tabela t -Student para graus de liberdade bem elevados (em nosso caso são 462 graus de liberdade) e nível de significância de 0,01. Ao total foram encontrados 279 certames com preços reais não estatisticamente superiores aos estimados (situação onde $P_i^+ = 0$) e 205 certames com preços reais estatisticamente superiores aos estimados (situação onde $P_i^+ = 1$). Faz-se necessário mencionar mais uma vez que o fato de um preço real se apresentar ocasionalmente acima do estimado não fornece qualquer informação relevante para nossos propósitos. O que buscamos é verificar se alguns competidores praticam preços acima do preço estimado de forma sistemática (não aleatória).

Para verificar quais competidores estariam praticando preços diferentes dos estimados, de forma sistemática, procedemos à segunda etapa, que é buscar a informação de competidores suspeitos a partir da estimativa da regressão logística (modelo LOGIT) da variável P_i^+ em função de variáveis *dummy* para cada uma das empresas analisadas, sendo um total de sete empresas produtoras e vinte empresas distribuidoras. Adicionalmente, controlamos a regressão em função de variáveis de tempo (uma variável *dummy* para cada ano da amostra) e de localização (uma variável *dummy* para cada um dos estados da amostra). De modo a assegurar robustez nas conclusões, estimamos a especificação de variável dependente em função das mesmas variáveis independentes a partir de uma regressão probabilística (modelo PROBIT), onde são interpretados seus efeitos marginais [12]. O objetivo é filtrar apenas competidores que aparecessem com parâmetros estatisticamente significativos (*Odd Ratios* e efeitos marginais) em ambos os modelos.

Em nossos exercícios encontramos *Odd Ratios* e efeitos marginais extremamente elevados e altamente significativos para os casos das empresas especificadas na base de dados como Carbocloro (Producer1) e Canexus (Producer2) [13]. No caso da empresa Carbocloro (Producer1) o *Odd Ratio* foi de 6.292.808 [14]. Para que o leitor tenha uma noção da magnitude deste valor, as demais empresas da amostra apresentaram *Odd Ratios* próximo a um (o que equivale a $p = q = 0,50$), ou estatisticamente menor que um - o que significa que a probabilidade de não praticar preços acima do estimado é bem maior que a probabilidade de praticar preços acima do estimado. No caso da empresa Canexus (Producer2) o *Odd Ratio* foi de 619.391.

No que diz respeito aos efeitos marginais, o valor computado para a empresa Carbocloro (Producer1) foi igual a 0,96, ou seja, só o fato da empresa Carbocloro (Producer1) ser considerada já eleva a probabilidade condicional do preço real ser superior ao preço estimado em 96 pontos percentuais, ou seja, praticamente assegura a observância de $P_i^+ = 1$. Já no caso da empresa Canexus (Producer2) o efeito marginal computado foi igual a 0,99. No caso das demais empresas, os efeitos marginais ou não eram significativos ou eram negativos, ou seja, sua presença reduziria a probabilidade condicional de preços reais acima dos estimados. Cabe lembrar que em ambos os modelos, LOGIT e PROBIT, os coeficientes estimados foram significativos a uma estatística P-valor inferior a 0,001.

Cabe ressaltar que no exercício do modelo LOGIT a empresa Norteste (Offer11) apareceu com um *Odd Ratio* de 19, sendo estatisticamente significativo a um nível de significância de 0,05. Contudo, tal resultado não foi robusto quando aplicado o modelo PROBIT. Por outro lado, a empresa Produquímica (Producer4) apareceu no modelo PROBIT com um efeito marginal de 0,64, estatisticamente significativo. Contudo, no teste LOGIT tal empresa apresentou um *Odd Ratio*

estatisticamente inferior a 1. Portanto, descartamos tais empresas de nosso conjunto de suspeitos e mantemos apenas as empresas Carbocloro (Producer1) e Canexus (Producer2), cujos resultados são robustos para as dimensões dos *Odd Ratios* e dos efeitos marginais, bem como para suas respectivas significâncias estatísticas. A Braskem (Producer7) venceu apenas quatro certames, todos eles com $P_i^+ = 1$, o que significa uma perfeita previsão de suspeita. Contudo, o número muito pequeno de observações amostrais nos impossibilita de obter qualquer inferência robusta neste sentido (mesmo porque tais estimativas são condicionais e em alguns casos o número restrito de certames em um determinado local ou período de tempo não fornece a identificação necessária). Neste caso aconselha-se também desconsiderá-la de nosso grupo de suspeitos^[15] de práticas não competitivas.

A partir da informação proveniente dos modelos LOGIT e PROBIT foi criada uma variável dicotômica para cada certame com preço vencedor suspeito, denominada S_i . Foram considerados suspeitos os certames com as seguintes características:

$$S_i = 1 \text{ se } \{P_i^+ = 1 \text{ e } Producer1 = 1 \text{ ou } P_i^+ = 1 \text{ e } Producer2 = 1\}, S_i = 0, \text{ caso contrário}$$

Portanto, os certames onde $S_i = 1$ são um subconjunto dos certames onde $P_i^+ = 1$. De fato, dos 205 certames onde $P_i^+ = 1$, um total de 172 certames foram computados com $S_i = 1$ enquanto todos os demais 312 certames da amostra foram computados com $S_i = 0$. O último passo do nosso filtro é verificar se o processo gerador de dados que construiu a distribuição de valores com $S_i = 0$ é o mesmo processo que construiu a distribuição de valores com $S_i = 1$, ou seja, se todos os preços da amostra foram construídos a partir do mesmo processo de precificação. Caso não venham do mesmo processo, como já mencionado, a pergunta central a ser endereçada é: por que existiriam dois processos de precificação diferentes em um mesmo mercado relevante, com produtos homogêneos não diferenciados?

Aplicando o teste de *Kolmogorov-Smirnov* de igualdade de distribuições para as distribuições onde os certames com preços vencedores são iguais a $S_i = 1$ e onde são iguais $S_i = 0$ é encontrado uma estatística *K-S* igual a 0,5789 com p-valor de 0,000. Como a hipótese nula é de igualdade de distribuições, verificamos que o processo que gerou os valores de preços vencedores para $S_i = 0$ é diferente do processo que gerou os valores de preços vencedores para $S_i = 1$. Em suma, as empresas Carbocloro (Producer1) e Canexus (Producer2) utilizaram um mecanismo de precificação diferente dos demais competidores do mercado nos certames suspeitos do qual participaram e venceram.

Para assegurarmos que fatores logísticos, temporais e locais não estejam influenciando o teste *K-S*, resolvemos implementá-lo em dois estágios. Em um primeiro estágio regredimos os preços vencedores dos certames em função de fatores logísticos, temporais e locais, ou seja, a especificação da equação [4] acrescida de *dummies* para anos e *dummies* para estados. Então computamos os resíduos da regressão e assinalamos os respectivos valores de $S_i = 1$ e $S_i = 0$. Como estamos lidando com um mercado de dimensão local e de produto homogêneo, esperaríamos que os resíduos da regressão especificada na sentença anterior fosse um mero distúrbio estocástico, um componente de erro, ou seja, composto de uma grande gama de fatores completamente difusos e aleatórios que justificassem pequenas variações de preço em torno de uma média, sem a existência de um componente sistemático que justificasse a Carbocloro (Producer1) e a Canexus (Producer2) praticarem “preços residuais”^[16] significativamente diferentes dos demais competidores nos certames suspeitos.

Contudo, ao aplicarmos o teste de igualdade de distribuições de *Kolmogorov-Smirnov* para os resíduos de regressão assinalados com $S_i = 1$ e com $S_i = 0$, encontramos novamente um resultado que corrobora a hipótese de que a distribuição dos preços suspeitos não advém do mesmo processo que gerou os preços

não suspeitos (estatística $K-S$ igual a 0,4359 com p-valor de 0,000). Em suma, encontramos evidências muito robustas de que as empresas Carbocloro (Producer1) e Canexus (Producer2) dispõem de algum mecanismo de precificação nos certames suspeitos que seus demais concorrentes não parecem aplicar. Tal resultado não parece aderente àquele esperado de um mecanismo de precificação competitivo em um mercado relevante local, com produtos homogêneos, não diferenciados.

Em resumo, há indícios econômicos de colusão nos certames envolvendo as empresas Carbocloro e Canexus, visto que:

- a. os resultados obtidos nas regressões do modelo LOGIT e PROBIT apontaram que há elevada probabilidade de o preço das licitações envolvendo essas duas empresas serem maiores que os preços estimados para os demais certames. Esses resultados, obtidos pelos *Odd Ratios* e efeitos marginais são ainda robustos e significantes estatisticamente.
- b. verificou-se que a distribuição de probabilidade nos preços dos certames vencedores envolvendo ambas empresas (teste Kolmogorov-Smirnov) difere da distribuição de preços obtida na comparação com outras empresas vencedoras, o que chama a atenção pelo fato dos produtos serem homogêneos e fazerem parte do mesmo mercado relevante.

5. Estimativa de Sobrepreço da Suposta Colusão

Na presente seção computamos o sobrepreço de suposto cartel praticado nos certames suspeitos em relação aos não suspeitos. Como já mencionado na seção anterior, utilizaremos duas hipóteses de identificação. Na primeira assumimos a hipótese de que os certames suspeitos são selecionados aleatoriamente, de forma exógena, de modo que possamos computar o sobrepreço a partir da seguinte equação:

$$\ln P_i = \alpha + \beta S_i + \mathbf{z}\tau + \mu_i \quad [5]$$

Onde $\ln P_i$ é o logaritmo do preço vencedor do certame i , S_i é a variável dicotômica para a qual certames suspeitos são iguais a unidade (caso contrário, igual a zero), as letras gregas são constantes paramétricas a serem estimadas, sendo que β traz o computo do sobrepreço de suposto cartel, \mathbf{z} é um vetor de variáveis de controle (logísticos, locais, tempo e para empresas) e μ_i é o componente de termo de erro estocástico.

A estimativa de β ^[17] foi igual a 0,19, estatisticamente significativa com p-valor inferior a 0,001. Como a variável explicada está expressa em logaritmo dos preços, a interpretação percentual para o sobrepreço deve ser obtida por meio da transformação $\exp^\beta = \exp^{0,19}$, ou seja, os lances suspeitos apresentam, em média, preços vencedores 21% superiores aos preços vencedores não suspeitos, após controlar para a heterogeneidade logística, locacional e temporal.

Relembrando o que foi discutido na seção anterior, na segunda hipótese assumimos que os certames não são escolhidos aleatoriamente, tampouco de forma exógena. Para endereçarmos tais problemas é utilizado o método de pareamento por escore de propensão (*propensity score matching*) com o objetivo de encontrarmos junto aos certames não suspeitos àqueles com características bem próximas dos suspeitos, exceto pelo fator de tratamento (que é a própria suspeita). Assim o computo de sobrepreço é obtido a partir das diferenças de médias de preços entre os grupos de controle (certames não suspeitos) e de tratamento (certames suspeitos).

Antes de qualquer tratamento de pareamento, o preço médio praticado nos certames suspeitos era de R\$ 1,85, enquanto a média dos preços não suspeitos era de R\$ 1,38. Em suma, uma diferença percentual de 34%. Contudo, após a aplicação da técnica de pareamento por escore de propensão, foi possível verificar redução do viés entre os grupos de controle e de tratamento, que conduziu o resultado para um sobrepreço de R\$ 0,36, equivalente a uma diferença percentual de 25%. Em suma, nossas estimativas de sobrepreço de cartel variam entre 21% e 25% [\[18\]](#).

6. Conclusões

O presente exercício empírico teve como objetivo fornecer um filtro comportamental para detecção de comportamento colusivo no mercado de cloro, mais especificamente no caso de licitações públicas para aquisição do produto. Nosso ponto de partida foi a delimitação do mercado relevante, de modo a identificar quais variáveis poderiam explicar variações de preços decorrentes de fatores não colusivos (fatores logísticos), de outros fatores que poderiam ser induzidos por fatores colusivos (empresas, anos e Estados).

Nossos exercícios foram muito robustos em apontar que as empresas Carbocloro (Producer1) e Canexus (Producer2) praticam um mecanismo de precificação que diverge sistematicamente das demais empresas e também de um comportamento competitivo dentro de um mercado de cloro com produto homogêneo não diferenciado. Nossas estimativas de sobrepreço de cartel variaram entre 21% e 25%.

ANEXO A

Programação dos exercícios empíricos no Software Stata

```
*****Brasília, 16 de Maio de 2015
```

```
***Tratamento dos dados
```

```
use "C:\Cloro CADE\Série de Preços Cloro.dta", clear
```

```
sum
```

```
kdensity price
```

gen lnprice=ln(price)

kdensity lnprice

kdensity priceIPAG

gen lnpriceIPAG=ln(priceIPAG)

kdensity lnpriceIPAG

gen lndist=ln(dist)

gen lnveloc=ln(veloc)

gen dist2=dist*dist

gen lndist2=lndist*lndist

tab ano, gen(year)

tab uf, gen(estado)

tab fabricante, gen(producer)

tab fornecedor, gen(offer)

***** Rodar regressões limpar fatores logísticos (não motivados por colusão)

```
rreg lnpriceIPAG carreta cilindro* Indist Indist2 Inveloc
```

```
predict yhat
```

```
predict error, stdp
```

```
gen teste=(yhat-lnpriceIPAG)/error
```

```
***** Computar preços acima do estimado
```

```
gen acima=0
```

```
replace acima=1 if teste<-2.326
```

```
tab acima
```

```
dprobit acima producer* offer* year* estado*, robust
```

```
logit acima producer* offer* year* estado*, or robust
```

```
***** Testar igualdade de distribuições K-S
```

```
gen suspeito=0
```

```
replace suspeito=1 if producer1==1 & acima==1
```

```
replace suspeito=1 if producer2==1 & acima==1
```

```
ksmirnov priceIPAG, by(suspeito) exact
```

```
ksmirnov lnpriceIPAG, by(suspeito) exact
```

```
rreg lnpriceIPAG carreta cilindro* Indist Indist2 Inveloc year* estado*
```

```
predict yhat2
```

```
gen error2=yhat2-lnpriceIPAG
```

```
ksmirnov error2, by(suspeito) exact
```

***** Computa sobrepreço de colusão

```
rreg lnpriceIPAG suspeito carreta cilindro* Indist Indist2 Inveloc offer* year* estado*
```

```
psmatch2 suspeito carreta cilindro* Indist Indist2 Inveloc offer* year* estado*, out( priceIPAG) logit
```

ANEXO B

Tabela com principais resultados empíricos

Tabela 1. Regressão Robusta para Preços Preditos: Variável explicada lnpriceIPAG

Regressores	Coefficiente	Erro-Padrão
-------------	--------------	-------------

carreta	-0,5275	0,0533 (***)
cilindro40	0,1043	0,1611
cilindro50	0	(omitted)
cilindro68	0,0270	0,0549
cilindro900	-0,1547	0,0487(***)
Indist	-1,2536	0,1201(***)
Indist2	0,1122	0,0099(***)
Inveloc	1,0476	0,1094(***)
_cons	-0,6689	0,3591(*)
Number of obs	470	
F(7, 462)	112,92	
Prob > F	0,0000	

Notas: (***) nível de significância de 0,01; (**) nível de significância de 0,05; (*) nível de significância de 0,10

Tabela 2. Regressão LOGIT com *Odd Ratios*: variável explicada “acima” (P_i^+)

Regressores	<i>Odd Ratio</i>	Erro-Padrão Robusto
producer1	6292808	1.53e+07(***)
producer2	619391	1127370(***)
controle outro (producer3-7)	Sim	Sim
controle (offer1-offer20)	Sim	Sim
controle anos (1999-2012)	Sim	Sim
controle para Estados (23)	Sim	Sim
_cons	8	11
Number of obs	356	
Wald chi2(36)	639.87	
Prob > chi2	0.0000	
Pseudo R2	0.2356	

Notas: (***) nível de significância de 0,01; (**) nível de significância de 0,05; (*) nível de significância de 0,10

Tabela 3. Regressão PROBIT com efeitos marginais: variável explicada “acima” (P_i^+)

Regressores	Efeitos Marginais	Erro-Padrão Robusto
producer1	0,9666	0,0288 (***)
producer2	0,9926	0,0113 (***)
controle outro (producer3-7)	Sim	Sim
controle (offer1-offer20)	Sim	Sim
controle anos (1999-2012)	Sim	Sim
controle para Estados (23)	Sim	Sim
_cons	8	11
Number of obs	356	
Wald chi2(36)	639.87	
Prob > chi2	0.0000	

Pseudo R2	0.2349
-----------	--------

Notas: (***) nível de significância de 0,01; (**) nível de significância de 0,05; (*) nível de significância de 0,10

Tabela 4. Regressão Robusta para sobrepreço de cartel: Variável explicada lnpriceIPAG

Regressores	Coefficiente	Erro-Padrão
suspeito	0,1906	0,0157 (***)
carreta	-0,4150	0,0283 (***)
cilindro40	0,1543	0,0744(**)
cilindro50	0	(omitted)
cilindro68	0,0194	0,0305
cilindro900	-0,1391	0,0253 (***)
ln-dist	0,1500	0,2677
ln-dist2	-0,0173	0,0236
ln-veloc	0,8307	0,4580 (*)

controle (offer1-offer20)	Sim	Sim
controle anos (1999-2012)	Sim	Sim
controle para Estados (23)	Sim	Sim
_cons	-2,6370	1.8889
Number of obs	470	
F(55, 414)	136,88	
Prob > F	0,0000	

Notas: (***) nível de significância de 0,01; (**) nível de significância de 0,05; (*) nível de significância de 0,10

Tabela 5. Estimativa de sobrepreço a partir de pareamento por score de propensão (PSM)

Amostra	Tratado priceIPAG (R\$)	Controle priceIPAG (R\$)	Diferença priceIPAG (R\$)	Erro Padrão	Estatística t
Não	1,8511	1,3823	0,4687	0,0524	8,93

Pareada					
Pareada	1,8511	1,4838	0,3672	0,1425	2,58

-
- [1] Para maiores detalhes desta abordagem, ver: Grout, P. A. e S. Sonderegger. (2005). Predicting cartels. Economic Discussion Paper. Office of Fair Trading.
- [2] Harrington, J. (2006). 'Detecting Cartels', In Buccirosi, P., Ed., Handbook of Antitrust Economics: MIT Press: 213-258.
- [3] <https://www.google.com.br/maps>, Acesso no período de 4 a 8 de maio de 2015.
- [4] Foi considerada como uma *proxy* para as distâncias percorridas pelo produto em cada compra, vez que englobaria todo o trajeto do produto de sua fabricação até o local de maior utilização do produto pela companhia de saneamento do estado.
- [5] Velocidade calculada com base na distância e tempo médio de viagem informado pelo Google Maps.
- [6] Logaritmo de base e, onde e é o número de Euler, número irracional aproximadamente igual à 2,718281...
- [7] Tal processo gera uma variável com distribuição mais próxima à distribuição normal e, portanto, mais adequada para realização de regressões econométricas.
- [8] Uma vez que os preços já foram deflacionados, variações anuais do preço refletem fatores como choques na oferta ou alterações temporais no preço não relacionadas com a inflação.
- [9] Tal definição é relevante para salientar que a possibilidade de concorrência com produtos importados é irrisória.
- [10] Os *Odd Ratios* são interpretados da seguinte maneira: suponha que a probabilidade de um preço real estar acima do preço estimado (chamemos de probabilidade de sucesso) seja igual a $p = 0,8$, enquanto que a probabilidade de um preço real não estar acima do preço estimado (chamemos de probabilidade de falha) seja igual a $q = 0,2$. Portanto, $Odds(sucesso) = p/q = p/(1-p) = 0,8/0,2 = 4$. Em suma, nestes casos a probabilidade de sucesso é 4 vezes maior que a probabilidade de falha.
- [11] Os resultados são reportados na tabela 1, no ANEXO B ao final do trabalho.

[12] Os efeitos marginais de um modelo PROBIT podem ser interpretados da seguinte maneira: imaginemos uma função $Prob(y=1 | \mathbf{z}) = G(\gamma_0 + \mathbf{z}\gamma)$, onde $G(\cdot)$ não é mais uma função logística (como trabalhamos na seção anterior), mas uma função probabilística. O efeito marginal da variável z_1 é igual a $\partial Prob / \partial z_1$, ou seja, o quanto uma mudança em z_1 aumenta a probabilidade de ocorrência do evento estudado. Imagine que a variável $Prob$ seja obtida a partir da amostra de indivíduos com câncer de pulmão [$Prob(y=1 | \mathbf{z})$] e sem câncer de pulmão [$Prob(y=0 | \mathbf{z})$]. Imagine que a variável explicativa é igual a $z_1 = 1$ no caso onde o indivíduo foi fumante e $z_1 = 0$ no caso onde o indivíduo nunca tenha fumado. O efeito marginal de z_1 é interpretado como quanto o fato do indivíduo ter fumado aumentou sua probabilidade em desenvolver câncer de pulmão. Um efeito marginal de z_1 igual a 0,50, por exemplo, significa que o fato da pessoa ter fumado aumentou em 50 pontos percentuais sua probabilidade de ter desenvolvido câncer de pulmão, quando comparado com uma pessoa que nunca tenha fumado na vida.

[13] Os resultados são reportados nas tabelas 2 e 3, no ANEXO B ao final do trabalho.

[14] A interpretação dos *Odd Ratios* obtidos a partir de uma regressão logística é um pouco diferente daquela apresentada anteriormente, uma vez que estamos nos referindo não mais simplesmente a probabilidades de falha e de sucesso, mas de probabilidades condicionais de falha e de sucesso. O termo adicional condicional diz respeito aos controles (variáveis explicativas) introduzidos no lado direito da regressão LOGIT. Portanto, quando interpretamos o *Odd Ratio* da variável *Producer1*, temos que tomar em conta que este *Odd Ratio* está condicionado ao controle de certames com características similares de localização, ano e firma distribuidora.

[15] Neste ponto deve estar claro ao leitor que caso tenha havido de fato um cartel, não significa que este tenha sido supostamente composto apenas pelas empresas *Producer1* e *Producer2*. Tais empresas foram selecionadas como suspeitas por conta da robustez estatística com que seus comportamentos supostamente colusivos são detectados em nosso filtro.

[16] O preço residual aqui seria igual ao componente de erro, ou seja: $\varepsilon_i = \ln P_i - \beta_0 - \beta_1 carreta - \beta_2 cilindro40 - \beta_3 cilindro50 - \beta_4 cilindro68 - \beta_5 cilindro900 - \beta_6 \ln distancia - \beta_7 \ln distancia^2 - \beta_8 \ln velocidade - \gamma Dummies de Tempo - \theta Dummies de Estado$.

[17] Os resultados são reportados na tabela 4, no ANEXO B ao final do trabalho.

[18] Os resultados são reportados na tabela 5, no ANEXO B ao final do trabalho.



Documento assinado eletronicamente por **Patricia Alessandra Morita Sakowski, Economista-Chefe Substituto(a)**, em 27/12/2016, às 17:50, conforme horário oficial de Brasília e Resolução Cade nº 11, de 02 de dezembro de 2014.



A autenticidade deste documento pode ser conferida no site sei.cade.gov.br/autentica, informando o código verificador **0284999** e o código CRC **51915B51**.

Referência: Processo nº 08700.008655/2016-16