

TESTE DE SIGNIFICÂNCIA DO VIÉS DE SIMULTANEIDADE NO MODELO ECONOMÉTRICO DO MERCADO DE SOJA DO BRASIL*

Carlos Antônio Moreira Leite
Sergio Alberto Brandt
Alexandre Aad Neto
Alberto Martins Rezende**

1. INTRODUÇÃO

No campo das ciências sociais tem-se evidenciado crescente interesse pela quantificação de relações de comportamento de sistemas sócio-econômicos.

No domínio da economia agrícola, reconhece-se a insuficiência do conhecimento meramente qualitativo. Assim, tem-se verificado crescente aplicação da econometria na elaboração de modelos (7).

Até há pouco, os economistas preocuparam-se com especificação, estimação e interpretação de sistemas de equações simultâneas. A resolução desses sistemas se prende a uma série de regras para sua concretização, exigindo condições especiais que devem merecer certo cuidado, por dois motivos: primeiro porque antes de serem estimados os parâmetros das equações não se pode determinar se são eles ou não realmente diferentes de zero; segundo porque um teste apropriado deve ser feito para estimativa dos coeficientes, com o objetivo de determinar se são diferentes de zero. Entretanto, não são conhecidos métodos rigorosos para testar hipóteses a respeito dos valores dos parâmetros de sistemas de equações simultâneas.

O presente estudo tem por objetivo estudar a validade de métodos alternativos de estimação de equações que representam o mercado brasileiro de soja.

Especificamente, propõe-se estimar os parâmetros das diversas equações que compõem o modelo, seguindo as normas econométricas recomendadas, e, posteriormente, aplicar o teste sugerido por RAMSEY (11) para testar a significância do viés de simultaneidade, caso se utilize, diretamente, o estimador de mínimos quadrados.

* Parte da tese apresentada à Universidade Federal de Viçosa, pelo primeiro autor, como uma das exigências para obtenção do grau «Magister Scientiae».

Recebido para publicação em 11-08-1976.

** Professores da Universidade Federal de Viçosa.

2. METODOLOGIA

2.1. Modelo e Métodos de Ajustamento

Neste estudo são usadas séries temporais de produção e de preços de diversos produtos, obtidos de publicações do Ministério do Interior (4), FGV (3), FIBGE (1), cobrindo o período de 1951 a 1971. As informações de exportações brasileiras são as publicadas pelo Banco Central do Brasil (2) e as dos preços do mercado internacional são as da (FAO) (9), (10).

O sistema de equações que representa o modelo é o que se segue:

Equação de Oferta

$$q_t^S = b_{11} + b_{12} q_{t-1}^S + b_{13} p_{t-1}^S + b_{14} p_{t-1}^{al} + b_{15} I_{t-1}^f + b_{16} T + u_1 \quad (I)$$

Equação de Demanda Interna

$$q_t^d = b_{21} + b_{22} q_{t-1}^d + b_{23} p_t^S + b_{24} p_t^{os} + b_{25} S_t + b_{26} T + u_2 \quad (II)$$

Equação de Demanda de Exportação

$$q_t^{dx} = b_{31} + b_{32} q_{t-1}^{dx} + b_{33} p_t^{xs} + b_{34} p_t^{wos} + b_{35} p_t^{wam} + b_{36} S_t^{ms} + b_{37} q_t^o + b_{38} T + u_3 \quad (III)$$

Identidade

$$q_t^S = q_t^d + q_t^{dx} \quad (IV)$$

onde q_t^S é a produção de soja, expressa em toneladas, no ano t; q_{t-1}^S é a produção de soja, expressa em toneladas, no ano t-1; p_{t-1}^S é o preço real de soja, expresso em cruzeiros de 1965-67, por tonelada, no ano t-1; p_{t-1}^{al} é o preço real de algodão, expresso em cruzeiros de 1965-67, por tonelada, no ano t-1; I_{t-1}^f é o índice real de preços de fertilizantes, de base 1965-67 igual a 100; T é uma variável de tendência ou de tempo, medida em anos; u_1 é o termo de erro aleatório; q_t^d é a quantidade de soja demandada pelas indústrias no período t, expressa em toneladas; q_{t-1}^d é a quantidade de soja demandada pelas indústrias, defasada para o período t-1, expressa em toneladas; p_t^S é o preço real de soja, expresso em cruzeiros de 1965-67, por tonelada, no ano t; p_t^{os} é o preço real de óleo de soja, expresso em cruzeiros de 1965-67, por tonelada, no ano t; S_t é o salário pago pelo setor industrial, no ano t; u_2 é o termo de erro aleatório; q_t^{dx} é a quantidade de soja exportada, em toneladas, no ano t; q_{t-1}^{dx} é a quantidade de soja exportada, em toneladas, no ano t-1; p_t^{xs} é o preço de soja pago aos exportadores, expresso em dólares americanos, por tonelada, no ano t; p_t^{wos} é o preço de óleo de soja no mercado internacional, expresso em dólares americanos, por tonelada, no ano t; S_t^{ms} é uma variável artificial indicadora de estoque de soja dos principais países importadores, expressa em milhares de toneladas, no ano t; q_t^o é a produção de grãos (aveia, milho e cevada) dos principais países importadores de soja do Brasil, expressa em milhares de toneladas, no ano t, e u_3 é o termo de erro aleatório (7).

As equações (I), (II), (III) e a identidade (IV) formam um sistema de equações, e, para que ele seja estimado, é necessário verificar os problemas econometrícicos envolvidos. O sistema acima especificado é composto de variáveis endógenas, predeterminadas e exógenas.

Os problemas relacionados com identificação são discutidos por JOHNSTON (8).

Por serem constituídas de variáveis exógenas, as equações de oferta (I) e de demanda de exportação (III) do modelo proposto são exatamente identificadas, enquanto a equação de demanda interna mostra-se superidentificada.

Os ajustamentos destas duas equações são feitos com base na identificação. Assim, as relações de oferta e de demanda de exportação são ajustadas pelo método dos mínimos quadrados ordinários. As pressuposições deste método são discutidas por JOHNSTON (8).

A relação de demanda interna é estimada pelo método dos mínimos quadrados de dois estágios, em razão da simultaneidade apresentada. Este método foi desenvolvido e discutido por Theil e Basmann (6). Algumas pressuposições básicas são necessárias para obtenção de estimadores consistentes e com distribuição assintoticamente normal por meio dos mínimos quadrados de dois estágios (6).

2.2. Teste de Significância do Viés de Simultaneidade

A presença de simultaneidade nas relações entre as variáveis de um modelo de regressão e os erros de especificação, como (a) forma funcional incorreta e (b) omissão de variáveis relevantes, são fontes de viés de qualquer estimador. Para testar a significância do viés do estimador de «mínimos quadrados» decorrente de qualquer dessas três fontes, mas sem fazer distinção entre elas, RAMSEY (11) propôs um teste que denominou RESET (Regression Specification Error Test).

O RESET foi desenvolvido para ser aplicado aos resíduos BLUS (Best Linear unbiased with Scalar Variance-Covariance Matrix), de Theil (6).

Entretanto, Hung e Bolch, citados por RAMSEY (11), baseando-se num experimento de Monte Carlo, mostraram que os resíduos de mínimos quadrados não são «inferiores» aos resíduos BLUS e que a substituição destes pelos primeiros não diminui o poder do teste.

Para desenvolvimento deste teste RAMSEY parte da observação de que num modelo de regressão corretamente especificado o termo erro (teórico) gozaria das seguintes propriedades:

$$E(\epsilon) = E(\epsilon/X) = \phi \quad (V)$$

$$\text{Var}(\epsilon) = E(\epsilon\epsilon'/X) = \sigma^2 \quad (VI)$$

Em tais circunstâncias, o resíduo (estimado) seria tal, que:

$$E(U) = E(U/X) = I \quad (VII)$$

$$\text{Var}(U) = E(UU'/X) = \sigma^2 M \quad (VIII)$$

onde $M = [I_M - X(X'X)^{-1}X']$, sendo X a matriz das variáveis explicativas.

Por outro lado, mostra-se que na presença de qualquer das fontes de viés supramencionadas a esperança matemática dos resíduos estimados pelo método dos mínimos quadrados é diferente de zero. Especificamente, no caso do viés de simultaneidade tem-se que:

$$E(U/X) = M_{\mu\epsilon/Y_M} = I \quad (IX)$$

onde $\mu\epsilon/Y_M$ é o valor esperado da distribuição do erro (teórico), dada a matriz das variáveis explicativas «endógenas» presentes na equação.

O teste de RAMSEY consiste em verificar se o termo $M_{\mu\epsilon/Y_M}$ difere significativamente de zero, a dado nível de confiança. Entretanto, existem algumas dificuldades a serem superadas.

Nota-se que $\mu\epsilon/Y_M$ é um valor teórico e, por conseguinte, não observável. Esta dificuldade é resolvida pela sugestão de que $M_{\mu\epsilon/Y_M}$ poderia ser aproximado por um polinômio, com o que se segue:

$$M_{\mu_e}/Y_M = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \dots \quad (X)$$

onde $q_j = M^{(j+1)}$ e

$$q_j = \begin{bmatrix} q_1^j \\ q_2^j \\ \vdots \\ q_M^j \end{bmatrix}$$

Desta forma, a entidade não observável M_{μ_e}/Y_M é substituída por uma combinação linear das variáveis $q_1 \dots q_n$, todas observáveis.

Em segundo lugar, é necessário estimar também os valores que constituem o vetor E (U/X). Tem-se que:

$$E(U/X) = E \begin{bmatrix} u_1/X \\ u_2/X \\ \vdots \\ u_n/X \end{bmatrix} = \begin{bmatrix} E(u_1/X) \\ E(u_2/X) \\ \vdots \\ E(u_n/X) \end{bmatrix} \quad (XI)$$

Necessita-se, por conseguinte, do valor médio esperado da distribuição de cada erro, ou seja, dos valores de:

$$E(u_1/X) \dots E(u_n/X)$$

quando se tem apenas uma observação das distribuições de $u_1, u_2 \dots u_n$, isto é, as estimativas dos resíduos de mínimos quadrados. Se se toma cada uma das observações disponíveis como estimativa da respectiva média, pode-se escrever:

$$\begin{bmatrix} E(u_1/X) \\ E(u_2/X) \\ \vdots \\ E(u_n/X) \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad (XII)$$

Finalmente, usando-se os resíduos observados e os valores de q_j calculados, pode-se escrever a seguinte equação de regressão:

$$e_i = \alpha_0 + \alpha_1 q_1 + \alpha_2 q_2 + \dots + \alpha_n q_n + v_i \quad (XIII)$$

O teste propriamente dito consiste em testar simultaneamente a significância dos coeficientes ($H_0: \alpha_0, \alpha_1, \dots, \alpha_n = 0$).

3. RESULTADOS E DISCUSSÃO

As equações de oferta e de demanda de exportação são ajustadas pelo método dos mínimos quadrados ordinários.

As estimativas que proporcionaram melhores resultados estatísticos da equação de oferta, em termos de significância dos parâmetros estimados, coerência de sinais e maior coeficiente de determinação, são as seguintes:

$$\begin{aligned} \log q_t^S &= 0,9664 + 0,6099 \log q_{t-1}^S + 0,6313 \log p_{t-1}^S - 0,1914 \log I_{t-1}^f + \\ &\quad (0,2362) \qquad (0,5169) \qquad (0,1603) \\ &+ 0,0237 T \\ &\quad (0,0178) \end{aligned} \quad (\text{XIV})$$

$$R^2 = 0,9731 \quad F(4,16) = 145,04 \quad DW = 1,61 \text{ (i)}$$

A equação de demanda de exportação é ajustada com variáveis expressas nos logaritmos decimais das observações originais, sendo esta a forma que apresenta «melhores» resultados em termos de significância dos parâmetros estimados, coerência de sinais e coerência com o conhecimento empírico:

$$\begin{aligned} \log q_t^{dx} &= 18,3410 + 0,0857 \log q_{t-1}^{dx} - 11,1625 \log p_t^{xs} + \\ &\quad (0,0547) \qquad (0,6120) \\ &+ 2,6810 \log p_t^{wos} + 0,9712 \log p_t^{wam} \\ &\quad (1,4810) \qquad (0,7740) \end{aligned} \quad (\text{XV})$$

$$R^2 = 0,9561 \quad F = 87,25 \quad h = 0,9892$$

As estimativas da função de demanda de soja para o mercado interno são desenvolvidas como se propõe no modelo conceptual.

Inicialmente, são feitas as estimativas da forma reduzida, ou seja, o preço real de soja é tomado como variável endógena da equação, em função de todas as variáveis predeterminadas do sistema (Quadro 1). O objetivo deste primeiro estágio do processo é obter os valores esperados da variável endógena (Quadro 2).

No segundo estágio os coeficientes das variáveis na equação estrutural são viesados, porém consistentes. O R^2 não é mais uma quantidade estatística estritamente válida. Os testes de hipóteses convencionais, «F» e «t», não são também testes estatísticos estritamente válidos. Uma regra prática para testar a significância dos coeficientes estimados é comparar os valores absolutos dos parâmetros estimados com os respectivos erros-padrão. Se o coeficiente de uma variável na equação estrutural é maior que ou igual a seu erro-padrão, este é considerado significativo. Se o coeficiente é pelo menos o dobro de seu erro-padrão, pode-se crer, com razoável segurança, na sua significância (6).

O modelo selecionado de demanda interna de soja obtido na forma estrutural é apresentado a seguir:

$$\begin{aligned} \log q_t^d &= 5,0081 + 0,3586 \log q_{t-1}^d - 1,8007 \log p_t^S + 1,0421 \log p_t^{os} - \\ &\quad (0,1636) \qquad (1,3787) \qquad (0,7177) \\ &- 0,8186 \log S_t + 0,9816 \log T \\ &\quad (0,5296) \qquad (0,2762) \end{aligned} \quad (\text{XVI})$$

$$R^2 = 0,9553 \quad F(5,11) = 47,02$$

Procede-se em seguida à estimativa dessa mesma equação pelo método dos mínimos quadrados ordinários:

$$\begin{aligned} \log q_t^d &= 4,9841 + 0,3591 \log q_{t-1}^d - 1,7741 \log p_t^S + 1,0296 p_t^{os} - \\ &- 0,8164 \log S_t + 0,9775 \log T \end{aligned} \quad (\text{XVII})$$

$$R^2 = 0,9679 \quad F(5,11) = 50,03$$

QUADRO 1 - Forma reduzida da equação de demanda interna de soja, Brasil, 1955-1971 (a)

Variáveis Explicativas (b)	Coeficiente (c)	Erro-Padrão
Constante	3,9482	
p_{t-1}^s Preço real de soja, no ano t-1	0,1606	0,3246
I_{t-1}^f Índice de preço real de fertilizantes	-0,0026	0,0794
q_{t-1}^{dx} Quantidade de soja exportada, no ano t-1	-0,0653****	0,0487
p_t^{xs} Preço de soja pago aos exportadores, no ano t	-0,5708****	0,4153
p_t^{wos} Preço de óleo de soja no mercado internacional, no ano t	0,5334***	0,3227
p_t^{wam} Preço de amendoim no mercado internacional, no ano t	0,4797****	0,3469
p_t^{os} Preço real de óleo de soja, no ano t	0,3534**	0,0982
S_t Salário pago pelo setor industrial, no ano t	0,3286***	0,2003
q_t^d Quantidade de soja demandada pelo setor industrial, no ano t-1	-0,1719***	0,09120
T Tendência	0,0001	0,1148
$R^2 = 0,9431*$		

(a) Fonte: (1, 2, 5, 9, 10, 12, 13).

(b) As variáveis são expressas nos logaritmos decimais dos valores observados.

(c) Onde * indica significativo, ao nível de 1%; ** indica significativo, ao nível de 2%; *** indica significativo, ao nível de 20%; e **** indica significativo, ao nível de 30%.

Comparando-se os dois últimos modelos apresentados, depara-se com a grande semelhança entre seus parâmetros e coeficientes de determinação.

Aceitando-se a hipótese nula de que os parâmetros da equação XVI são «conjuntamente» iguais a zero, conclui-se que $E(U/X) = 0$. Daí a razão por que as estimativas dos mínimos quadrados ordinários não apresentam viés originário de qualquer das fontes já mencionadas, ou seja, presença de simultaneidade nas relações entre as variáveis do modelo de regressão e os erros de especificação, como (a) forma funcional incorreta e (b) omissão de variáveis relevantes.

A hipótese alternativa é a de que existe o viés, embora não se lhe possa determinar a origem.

Testando a significância dos coeficientes, considerando apenas os de q_1 e q_2 , encontra-se o seguinte resultado:

$$e = 1,5481 \times 10^{-3} + 3,2536 \times 10^{-5} q_1 - 6,0052 \times 10^{-6} q_2 \\ (2,4446 \times 10^{-3}) \quad (-2,4451 \times 10^{-3}) \quad (XVIII)$$

$$F = 2,9901 \times 10^{-6} \quad R^2 = 0,00$$

onde os valores entre parênteses são os valores de «t» calculados.

QUADRO 2 - Preços observados e previstos, estimados pela forma reduzida da demanda interna de soja, Brasil, 1955/71 (a)

Ano	Valor Observado	Valor Previsto
1955	2,004	2,006
56	2,093	2,092
57	2,052	2,063
58	2,060	2,055
59	2,046	2,037
1960	2,178	2,168
61	2,152	2,139
62	2,077	2,089
63	2,124	2,128
64	2,182	2,190
65	2,159	2,174
66	2,167	2,134
67	2,083	2,087
68	2,117	2,132
69	2,116	2,112
1970	2,090	2,091
71	2,125	2,139

(a) Fonte: (1, 2, 5, 9, 10, 12, 13).

Conclui-se que as estimativas dos mínimos quadrados não apresentam viés, quer seja pela forma funcional incorreta, quer pela omissão de variáveis relevantes quer ainda pela simultaneidade entre as variáveis do modelo, em face das estimativas do coeficiente de determinação.

4. RESUMO

A quantificação de relações de comportamento de sistemas sócio-económicos tem sido preocupação dos cientistas sociais. Métodos quantitativos vêm sendo rapidamente desenvolvidos e com as facilidades da computação novos conceitos e sofisticados métodos de estimação vêm sendo criados.

Em particular, a estimação de parâmetros de sistemas de equações simultâneas se prende a uma série de regras e para sua concretização são exigidas condições especiais que devem ser examinadas com certo cuidado.

O presente estudo tem por objetivo analisar o sistema de equações de oferta, de demanda interna e de demanda de exportação de soja, que constituem o mercado brasileiro deste produto. Utiliza-se o método dos mínimos quadrados ordinários para estimar as equações de oferta e de demanda de exportação. A equação de demanda interna é estimada, primeiramente, pelo método dos mínimos qua-

drados de dois estágios. A utilização desse método se deve à pressuposição inicial de que a variável preço real de soja no mercado interno é endógena.

As variáveis que compõem a equação selecionada da oferta são: produção de soja, retardada de um ano; preço real de soja pago aos produtores, retardado de um ano; índice real de preços de fertilizantes e uma variável de tendência. Na equação de demanda de exportação são significativas as variáveis preço de soja recebido pelos exportadores, preço de óleo de soja no mercado internacional e preço de amendoim nesse mercado e a variável dependente, defasada de um ano.

De acordo com as estimativas feitas pelo método dos mínimos quadrados de dois estágios são significativas: a variável dependente, tomada com retardamento de um ano, bem como as variáveis preço real de soja, preço real de óleo de soja, salário pago pelo setor industrial e uma variável de tendência.

A equação de demanda interna é também estimada pelo método dos mínimos quadrados ordinários, constatando-se grande semelhança como as estimativas anteriormente obtidas, em termos de coeficientes de regressão parcial e coeficientes de determinação.

Submetendo-se as estimativas obtidas pelo método dos mínimos quadrados ordinários ao teste de RAMSEY, encontra-se ausência de viés, em razão da forma funcional incorreta, da omissão de variáveis relevantes e/ou da simultaneidade entre as variáveis do modelo. Em outras palavras, a equação de demanda interna de soja pode ser estimada, sem viés, pelo método dos mínimos quadrados ordinários. Evidentemente, não se pode dizer o grau de certeza desta afirmativa porque não se conhece o poder do teste. A evidência obtida sugere que a variável preço de soja é exógena.

5. SUMMARY

Specification and estimation of socio-economic systems has been a common concern of social scientists. Many new quantitative methods are currently being developed. With the help of computer facilities new concepts and methods are being elaborated.

Solution of a simultaneous equations system is based on a series of rules which, to be put into practice, require special assumptions that must be examined with caution.

The present study intends to evaluate a system of equations including supply, domestic demand and export demand for Brazilian soybeans. Ordinary least squares were used to estimate the supply and export demand equations. The domestic demand equation was first estimated by two stages least squares. This procedure was used on the assumption that real price of soybeans at the domestic market is an endogenous variable.

The variables included in the selected supply equation are soybean production lagged one year, real price of soybeans paid to producers lagged one year, real fertilizer prices lagged one year and a trend variable. In the selected export demand equation the relevant variables were the following: soybean export price, soybean exports lagged one year, world price of soybean oil and world price of peanuts.

The domestic demand equation estimated by two stages squares included the following variables: domestic consumption lagged one year, price of soybeans, wages in the industrial sector, price of soybean oil and a trend variable. All were found to be significantly related to domestic consumption.

The domestic demand equation was also estimated by ordinary least squares. The results demonstrated a strong similarity in the parameters estimated by the two methods.

Upon submitting the OLS estimates to RAMSEY's test the absence of bias was suggested. Presence of bias could be due to inadequacy functional form, omission of relevant variables and to simultaneous determination of exogenous variables included in the model. The results indicated that the domestic demand equation could be estimated, without bias, by ordinary least squares. Since the power of the test is unknown, it cannot be said with what degree of certainty this finding can be expected to hold true. Empirical evidence only suggests that domestic price is an exogenous variable.

6. LITERATURA CITADA

1. BRASIL. Instituto Brasileiro de Geografia e Estatística. *Anuário Estatístico*

- do Brasil.* Rio de Janeiro, 1950.
2. BRASIL. Banco Central do Brasil. *Relatório Anual.* Brasília, 1966.
 3. BRASIL. Fundação Getúlio Vargas. *Conjuntura Económica.* Rio de Janeiro, 1972.
 4. BRASIL. Ministério do Interior. *Contribuição ao desenvolvimento da agroindústria.* Campinas, 1972.
 5. CIBANTOS, J.S. *Demanda de fertilizantes no Estado de São Paulo.* Piracicaba, ESALQ, 1972. 196 p. (Tese Doutorado).
 6. FRANK, Jr., C.R. *Statistics and econometrics.* New York, Holt, Rinehart and Winston, 1971. 400 p.
 7. GIRÃO, J.A. *A função de produção Cobb-Douglas e a análise inter-regional da produção agrícola.* Lisboa, Fundação Calouste Gulbenkian.
 8. JOHNSTON, J. *Métodos econométricos.* São Paulo, Atlas, 1971. 318 p.
 9. NAÇÕES UNIDAS. Organização para Agricultura e Alimentação. *Monthly Bulletin of Agricultural Economics and Statistics.* Roma, 1950.
 10. NAÇÕES UNIDAS. Organização para Agricultura e Alimentação. *Production Yearbook.* Roma, 1957.
 11. RAMSEY, J.B. Tests for specification errors in classical linear least squares regression analysis. *Journal of Royal Statistical Society, 31:*350-71. 1969.
 12. RIO GRANDE DO SUL. Assembléia Legislativa, Comissão de Agricultura e Pecuária. *Soja.* Porto Alegre, 1974. 414 p.
 13. RIO GRANDE DO SUL. Secretaria de Coordenação e Planejamento, Secretaria da Agricultura, Superintendência de Planejamento Global, Supervisão de Apoio Técnico. *Soja, situação e perspectiva.* Porto Alegre, 1973. 67 p.