Estimação de modelos econométricos — algumas considerações *

Fernando de Holanda Barbosa **

1. Introdução

A literatura econométrica contém um bom número de estimadores para os parâmetros de equações estruturais de um modelo de equações simultâneas. Quanto à informação que utilizam, estes estimadores podem ser classificados em duas classes: os de informação limitada e os estimadores de informação completa. Os estimadores de informação limitada levam em conta apenas a informação que diz respeito a uma particular equação do modelo, enquanto os estimadores de informação completa incorporam toda

* O autor deseja agradecer os comentários e sugestões de Clovis de Faro e Maria da Conceição Silva a uma versão preliminar deste trabalho. Após a conclusão desta versão, chegou as minhas mãos a revista *Econometrica* de julho de 1976 que contém um *paper* de autoria de J. Daniel Khazzoon intitulado *An indirect least squares estimator for overidentified equations*, no qual o estimador indireto generalizado de mínimos quadrados é derivado com o auxílio da matriz inversa de Moore-Penrose. A derivação que apresento aqui foi obtida independentemente pelo autor do presente trabalho, em um *paper* preparado para o Curso de Econometria II, oferecido pelo Prof. Arnold Zellner na Universidade de Chicago em março de 1974. O enfoque que adoto não faz uso da inversa de Moore-ePnrose. Entretanto, os resultados são idênticos aos de Khazzoon.

** Do Instituto de Pesquisas do IPEA.

a informação contida no modelo. Um dos estimadores de informação limitada é o estimador indireto de mínimos quadrados. É bastante conhecido o fato de este estimador ser restrito a equações estruturais exatamente identificadas. No segundo item deste trabalho mostramos que o estimador indireto de mínimos quadrados pode ser generalizado de tal maneira a tornar viável a sua aplicação também no caso de equação estrutural superindentificada. O estimador indireto generalizado de mínimos quadrados é consistente, porém não é assintoticamente eficiente. Como os demais estimadores de informação limitada, as propriedades deste estimador, para pequenas amostras, não são conhecidas. [1]

Uma proposição bastante familiar, na literatura econométrica, é a de que a estimação dos parâmetros estruturais de uma equação estrutural pressupõe a identificação da equação. Entretanto, devido ao fato de que é possível obterem-se estimadores dos parâmetros da equação estrutural quando a condição de ordem é satisfeita mas a condição de posto (*rank*) deixa de ser observada, é bem possível que alguns dos modelos encontrados na econometria aplicada não sejam identificados. O terceiro item cuida deste problema, lembrando que alguns testes de hipóteses foram elaborados com o intuito de eliminar a possibilidade da estimação de parâmetros não-identificados.

Uma das justificativas, talvez a única do ponto de vista teórico, para o uso de mínimos quadrados ordinários na estimação de modelos de equações simultâneas não-recursivos é quando o modelo é grande ou a amostra é pequena. Em outras palavras, quando o número de variáveis predeterminadas é maior que o número de observações, um caso bastante comum em modelos econométricos, a estimativa de mínimos quadrados em dois estágios reduz-se a mínimos quadrados ordinários. O último item deste trabalho reproduz esta propriedade demonstrando-a, porém, a partir de um enfoque diferente do originalmente empregado para a sua derivação.

2. Identificação e mínimos quadrados indiretos

Com o objetivo de tornar a exposição mais clara, repetimos a seguir o bem conhecido teorema na literatura econométrica que diz respeito à identificação de uma equação estrutural quando se sabe, *a priori*, que

[1] Algumas propriedades, para pequenas amostras, já são conhecidas para estimadores de informação limitada. Por exemplo, para o estimador de mínimos quadrados em dois estágios, veja Mariano (1972).

alguns coeficientes da equação em estudo são iguais a zero.[2] Sem perda de generalidade, a primeira equação de um sistema de L equações simultâneas com K variáveis predeterminadas pode ser escrita,

$$y_1 = Y_1 \gamma_1 + X_1 \beta_1 + u_1 = Z_1 \delta_1 + u_1 \tag{2.1}$$

onde γ_1, um vetor L_1 x 1, e β_1, um vetor K_1 x 1, são os parâmetros estruturais a serem estimados. A matriz $[y_1 \mid Y]_1$ é a matriz das variáveis endógenas incluídas na equação (2.1) de ordem T x $(1 + L_1)$: onde T é o tamanho da amostra, ou seja, o número de observações. A matriz $X = [X_1 \mid X_0]$, de ordem T x $(K_1 + K_1)$, é a matriz de variáveis predeterminadas do sistema de equações simultâneas e $K_2 = K - K_1$ variáveis predeterminadas que são excluídas da equação (2.1). O vetor u_1 é um vetor T x 1 de erros aleatórios, o qual, por hipótese, tem valor esperado zero e matriz de variância-covariância igual a $\sigma_1 I$, onde I é a matriz identidade. A matriz $Z_1 = [Y_1 \mid X_1]$ e o vetor $\delta'_1 = [\gamma'_1 \mid \beta'_1]$.

A forma reduzida das variáveis endógenas incluídas na equação (2.1) pode ser escrita,

$$[y_1 \mid Y_1] = [X_1 \mid X_o] \begin{bmatrix} \pi_{11} & \Pi_{10} \\ \pi_{01} & \Pi_{00} \end{bmatrix} + [v_1 \mid V_1] \tag{2.2}$$

onde π_{11} e π_{01} são vetores coluna de ordem K_1 x 1 e K_2 x 1, respectivamente. As matrizes Π_{10} e Π_{00} são matrizes de ordem K_1 x L_1 e K_2 x L_1, respectivamente, e a matriz $[v_1 \mid V_1]$ é a matriz de ordem T x $(1 + L_1)$ de erros aleatórios da forma reduzida.

Pós-multiplicando ambos os lados da equação acima por $[1 \mid -\gamma'_1]'$, obtemos:[3]

$$y_1 - Y_1 \gamma_1 = X_1 (\pi_{11} - \Pi_{10} \gamma_1) + X_o (\pi_{01} - \Pi_{00} \gamma_1) + u_1 \tag{2.3}$$

[2] A demonstração apresentada aqui usa a regra de normalização do coeficiente da variável y_1. Esta regra, ou quaqluer outra, não é necessária para o estudo da identificação. Entretanto, para o nosso propósito, esta regra é extremamente conveniente. Para o estudo do problema de identificação sem fazer uso de regras de normalização veja, por exemplo, Koopmans & Hood (1953).

[3] Estamos usando a propriedade de que $v_1 - V_1\gamma_1 = u_1$ a qual pode ser derivada como se segue. O modelo de L equações simultâneas pode ser escrito em sua forma estrutural na forma $Y\Gamma = XB + U$. As matrizes Γ e B são matrizes de parâmetros de ordem L x L e K x K, respectivamente. A matriz de variáveis endógenas y é de ordem T x L. A matriz $U = [u_1, 2, \ldots, u_L]$ de erros aleatórios é de ordem T x L. A forma reduzida é obtida a partir da forma estrutural pós-multiplicando ambos os lados da equação estrutural por Γ^{-1}: $Y = XB\Gamma^{-1} + U\Gamma^{-1} = X\Pi_1 + V$, onde $\Pi_1 = B\Gamma^{-1}$ e $V = U\Gamma^{-1}$. A primeira coluna de Γ corresponde aos coeficientes das variáveis endógenas na equação (2.1) e é dada pelo vetor $[1 \mid -\gamma_1' \mid 0']'$. Do fato que $V\Gamma = U$ concluímos que $v_1 - V\ \gamma_1 = u_1$.

Comparando as equações (2.1) e (2.3) temos que,

$$\begin{aligned} \pi_{11} - \Pi_{10}\ \gamma_1 &= \beta_1 \\ \pi_{01} - \Pi_{00}\ \gamma_1 &= 0 \end{aligned} \tag{2.4}$$

ou usando notação matricial:

$$\begin{bmatrix} \pi_{11} \\ \pi_{01} \end{bmatrix} = \begin{bmatrix} \Pi_{10} & I \\ \Pi_{00} & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \beta_1 \end{bmatrix} \tag{2.5}$$

Com a finalidade de simplificar a notação, denominamos Q a matriz de ordem $(K_1 + K_2)$ x $(L_1 + K_1)$ que apareceu no lado direito da equação (2.5):

$$Q = \begin{bmatrix} \Pi_{10} & I \\ \Pi_{00} & 0 \end{bmatrix} = [\Pi\ D] \tag{2.6}$$

onde:

$$\Pi = \begin{bmatrix} \Pi_{10} \\ \Pi_{00} \end{bmatrix} \quad \text{e} \quad D = \begin{bmatrix} I \\ 0 \end{bmatrix}$$

A expressão depois do segundo sinal de igualdade em (2.6) é uma forma alternativa de escrever a matriz Q, que será útil mais adiante em conexão com o estimador indireto generalizado de mínimos quadrados. Denominamos por π o vetor $\pi' = [\pi'_{11} \mid \pi'_{01}]$. Assim, podemos escrever a a equação (2.5) como:

$$\pi = Q\ \delta_1 \tag{2.7}$$

A equação estrutural (2.1) será identificada, quando for possível calcular os coeficientes β'_s e γ'_s da forma estrutural a partir do conhecimento dos coeficientes π'_s da forma reduzida.[4] Isto é, a equação (2.1) será identificada, se e somente se a equação (2.7) tiver uma única solução

[4] Não se deve perder de vista o fato de que identificação não é um problema relativo somente a modelos de equações simultâneas. De um modo geral, o problema de identificação existe quando um dado conjunto de observações y pode ser gerado por diferentes funções de densidade de probabilidade, isto é, $p\ (y/\theta) = p\ (y/\phi)$, onde θ e ϕ são diferentes parâmetros. Neste caso, não se pode discernir a partir da amostra qual dos dois modelos está gerando as observações y. O modelo é dito não-identificado e os parâmetros não são identificados. Para uma explicação mais detalhada veja Zellner (1971) p. 253-8.

para o vetor δ. Portanto, a condição necessária e suficiente para que a equação (2.1) seja identificada é que a matriz Q tenha o posto (*rank*) igual ao número de colunas $(L_1 + K_1)$, e que o vetor π pertença ao espaço gerado pelas colunas de Q, ou equivalentemente,

$$\rho\,[\Pi_{00}] = \rho\,[\pi_{01}\ \Pi_{00}] = L_1 \qquad (2.8)$$

onde ρ representa o posto da matriz indicada entre colchetes. Uma condição necessária para a condição (2.8) ser satisfeita, denominada condição de ordem, é a seguinte:

$$K - K_1 \geq L_1 \qquad (2.9)$$

É importante notar que a desigualdade acima pode verificar-se para uma equação estrutural e esta equação não ser identificada, porque a condição (2.8) não é observada. De qualquer forma, a desigualdade (2.9) é bastante popular, pois serve para, numa primeira aproximação, estudar se uma equação estrutural é identificada ou não, dado que sua aplicação é bastante simples: o número de variáveis predeterminadas excluídas da equação deve ser, pelo menos, igual ao número de variáveis endógenas incluídas, menos uma. Entretanto, o uso indiscriminado desta condição pode acarretar um erro bastante grave, como o de se estimarem os parâmetros de uma equação não-identificada. No próximo item mostraremos como tal fato pode ocorrer.

2.1 Identificação exata e superidentificação

No restante deste item admitimos que a condição (2.8) é satisfeita e que, portanto, a equação (2.1) é identificada.

A expressão (2.7) contém um sistema de K equações lineares com $(L_1 + K_1)$ incógnitas, os elementos do vetor δ_1. A solução deste sistema é dada por:

$$\delta_1 = (Q'\ Q)^{-1}\ Q'\ \pi \qquad (2.10)$$

Na hipótese de a condição de ordem (2.9) traduzir-se pela igualdade $K = K_1 + L_1$, a matriz Q é uma matriz quadrada de posto completo. Portanto, a solução (2.10) reduz-se a:

$$\delta_1 = Q^{-1}\ (Q')^{-1}\ Q'\ \pi = Q^{-1}\ \pi \qquad (2.11)$$

Neste caso, a equação estrutural (2.1) é dita ser exatamente identificada. Por outro lado, quando a condição de ordem se traduz pela desigualdade $K > K_1 + L_1$ e os valores estimados da forma reduzida, digamos $\hat{Q}$ e $\hat{\pi}$, são substituídos na expressão (2.7), obtemos diferentes valores para δ_1 de acordo com o subconjunto de $K_1 + L_1$ equações que for selecionado das K equações que formam o sistema (2.7). Neste caso, a equação estrutural (2.1) é superidentificada. É importante observar que, na solução do sistema (2.7) dada por (2.10), é irrelevante o fato de a equação ser exatamente identificada ou superidentificada pois a solução é única quando substituímos os valores de $\hat{Q}$ e $\hat{\pi}$ em (2.10).

2.2 O estimador indireto generalizado de mínimos quadrados

O estimador indireto generalizado de mínimos quadrados $\hat{\delta}_1$ é obtido quando substituímos os valores de Q e π em (2.10) pelos estimadores $\hat{\Pi}$ e $\hat{\pi}$ de forma reduzida, isto é,

$$\hat{\delta}_1 = (\hat{Q}' \, \hat{Q})^{-1} \, \hat{Q}' \, \hat{\pi}, \tag{2.12}$$

onde a matriz $\hat{Q}$ é expressa por

$$\hat{Q} = [\hat{\Pi} \,\vdots\, D] \tag{2.13}$$

e os estimadores $\hat{\Pi}$ e $\hat{\pi}$ da forma reduzida são:

$$\hat{\Pi} = (X' \, X)^{-1} \, X' \, Y_1 \tag{2.14}$$

$$\hat{\pi} = (X' \, X)^{-1} \, X' \, y_1 \tag{2.15}$$

Denominamos o estimador $\hat{\delta}_1$ de estimador indireto generalizado de mínimos quadrados por dois motivos. Em primeiro lugar, porque este estimador é obtido indiretamente através dos estimadores $\hat{\Pi}$ e $\hat{\pi}$. Em segundo lugar porque, no caso particular de a equação estrutural ser exatamente identificada, $\hat{\delta}_1$ reduz-se ao tradicional estimador indireto de mínimos quadrados:

$$\hat{\delta}_1 = \hat{Q}^{-1} \, \hat{\pi} \tag{2.16}$$

Cuidemos, agora, de obter expressões algébricas para o estimador $\hat{\delta}_1$ em função das matrizes de observações X e $[y_1 \,\vdots\, Y_1]$. Usamos a relação (2.13) para escrever o estimador $\hat{\delta}_1$ da seguinte forma:

$$\hat{\delta}_1 = \begin{bmatrix} \hat{\Pi}' \, \hat{\Pi} & \hat{\Pi}' \, D \\ D' \, \hat{\Pi} & D' \, D \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Pi}' \\ D' \end{bmatrix} \hat{\pi} \tag{2.17}$$

Em seguida, substituímos os valores de $\hat{\Pi}$ e $\hat{\pi}$ dados em (2.14) e (2.15) na expressão (2.17) e obtemos:[5]

$$\hat{\delta}_1 = \begin{bmatrix} Y_1' X (X' X)^{-2} X' Y_1 & Y_1' X (X' X)^{-2} X' X_1 \\ X_1' X (X' X)^{-2} X' Y_1 & I \end{bmatrix}^{-1} \cdot \qquad (2.18)$$

$$\cdot \begin{bmatrix} Y_1' X (X' X)^{-2} X' \\ X_1' X (X' X)^{-2} X' \end{bmatrix} y_1$$

ou, usando uma notação mais simples:[6]

$$\hat{\delta}_1 = [Z_1' X (X' X)^{-2} X' Z_1]^{-1} Z_1' X (X' X)^{-2} X' y_1 \qquad (2.19)$$

2.3 Propriedades do estimador $\hat{\delta}_1$

Substituímos o valor de y_1, dado pelo lado direito do segundo sinal de igualdade de (2.1), em (2.19), e obtemos:

$$\hat{\delta}_1 = \delta_1 + [Z_1' X (X' X)^{-2} X' Z_1]^{-1} Z_1' X (X' X)^{-2} X' u_1 \qquad (2.20)$$

De acordo com algumas hipóteses tradicionais no estudo dos estimadores dos parâmetros de um sistema de equações simultâneas temos que[7]

$$plim \left[\frac{Z_1' X}{T} \left(\frac{X' X}{T}\right)^{-2} \frac{X' Z_1}{T}\right]^{-1} \text{ é finita, e}$$

$$plim \left[\frac{Z_1' X}{T} \left(\frac{X' X}{T}\right)^{-2} \frac{X' u_1}{T}\right] \text{ é igual a zero.}$$

Aplicando estes resultados a (2.20) concluímos que,

$$plim\ \hat{\delta}_1 = \delta_1 \qquad (2.21)$$

o que significa dizer que o estimador indireto generalizado de mínimos quadrados é consistente.

[5] Embora não seja tão óbvio, é possível mostrar que: $D' \hat{\Pi} = X'_1 X (X' X)^{-2} X' Y_1$. A demonstração é bastante simples: $D' \hat{\Pi} = D' (X' X)^{-1} X' Y_1 = D' X' X (X' X)^{-1} (X' X)^{-1} X' Y_1 = X'_1 X (X' X)^{-2} X' Y_1$, tendo em vista que $D' X' = X'_1$. Similarmente, calcula-se o produto $D' \hat{\Pi}$.

[6] Para obtermos (2.19) a partir de (2.18) usamos o fato de que: $X'_1 X (X' X)^{-2} X' X = D' X' X (X' X)^{-2} X' X D = I$, pois $D' D = I$.

[7] Estas hipóteses estão listadas em Theil (1971) cap. 10.

A distribuição assintótica da variável aleatória $\sqrt{T}\,(\hat{\delta}_1 - \delta_1)$ pode ser obtida com o seguinte procedimento. Da expressão (2.20) temos que:

$$\sqrt{T}\,(\hat{\delta}_1 - \delta_1) = \left[\frac{Z_1' X}{T}\left(\frac{X' X}{T}\right)^{-2}\frac{X' Z_1}{T}\right]^{-1}\frac{Z_1' X}{T}\left(\frac{X' X}{T}\right)^{-2}\frac{X' u_1}{\sqrt{T}} \tag{2.22}$$

Por outro lado, as hipóteses a que nos referimos abaixo de (2.20) nos possibilitam afirmar que:

$$plim\left[\frac{Z_1' X}{T}\left(\frac{X' X}{T}\right)^{-2}\frac{X' Z_1}{T}\right]^{-1}\frac{Z_1' X}{T}\left(\frac{X' X}{T}\right)^{-2} = G$$

onde G é uma matriz cujos elementos são finitos. Em seguida, lançamos mão do teorema que afirma que a distribuição assintótica de $X' u_1/\sqrt{T}$ é normal com valor esperado zero e matriz de variância-covariância igual a $\sigma_1\, plim\left(\frac{X' X}{T}\right)$.[8] Podemos, então, concluir que a distribuição assintótica de $\sqrt{T}\,(\hat{\delta}_1 - \delta_1)$ é normal com média zero e matriz de variância-covariância $Var\,\hat{\delta}_1$ igual a:

$$\sigma_1\, G\left[plim\left(\frac{X' X}{T}\right)\right]G' = Var\,\hat{\delta}_1 \tag{2.23}$$

Embora o estimador indireto generalizado de mínimos quadrados tenha a propriedade de ser consistente, $\hat{\delta}_1$ não é assintoticamente eficiente, porque a variância deste estimador é diferente da variância do estimador de mínimos quadrados em dois estágios, o qual é assintoticamente eficiente. Com o objetivo de demonstrar esta proposição[9] comecemos por simplificar a notação denominando por P a matriz:

$$P = \left(\frac{X' X}{T}\right)^{-1}\frac{X' Z_1}{T} \tag{2.24}$$

Em seguida, introduzimos a matriz Δ definida por:

$$(P' P)^{-1} P' = \left(P'\frac{X' X}{T}P\right)^{-1} P'\left(\frac{X' X}{T}\right) + \Delta \tag{2.25}$$

[8] Veja Theil. op. cit. p. 487.

[9] A demonstração que apresentamos abaixo é em grande parte baseada em Dhrymes (1974), quando este compara os estimadores de dois e três estágios.

Concluímos, da expressão anterior, que a matriz Δ satisfaz a equação:

$$\Delta P = 0 \tag{2.26}$$

A matriz (2.23), a parte do escalar σ_1, pode ser escrita em termos da matriz P, definida em (2.24), da seguinte forma:

$$G\left(\frac{X'\,X}{T}\right)G' = (P'\,P)^{-1}\,P'\left(\frac{X'\,X}{T}\right)^{-1}P\,(P'\,P)^{-1} \tag{2.27}$$

onde o símbolo *plim* foi suprimido com a finalidade de não sobrecarregar a notação. Contudo, tanto a matriz P como as demais que aparecem depois da expressão (2.23) devem ser entendidas como limites em probabilidade (*plim*).

O lado direito da igualdade (2.27) pode ser escrito levando em conta a expressão (2.25) da seguinte maneira:

$$\begin{aligned}(P'\,P)^{-1}\,P'\left(\frac{X'\,X}{T}\right)^{-1}P\,(P'\,P)^{-1} &= \\ = \left[\left(P'\,\frac{X'\,X}{T}\,P\right)^{-1}P'\,\frac{X'\,X}{T}+\Delta\right]&\cdot\left(\frac{X'\,X}{T}\right)^{-1}\cdot \\ \cdot\left[\frac{X'\,X}{T}\,P\left(P'\,\frac{X'\,X}{T}\,P\right)^{-1}+\Delta'\right]&\end{aligned} \tag{2.28}$$

A expressão acima pode ser simplificada, tendo em vista que $\Delta P = 0$. O resultado desta simplificação é:

$$\begin{aligned}(P'\,P)^{-1}\,P'\left(\frac{X'\,X}{T}\right)^{-1}P\,(P'\,P)^{-1} &= \\ \left(P'\,\frac{X'\,X}{T}\,P\right)^{-1}+\Delta\left(\frac{X'\,X}{T}\right)^{-1}\Delta'&\end{aligned} \tag{2.29}$$

A primeira matriz que aparece no lado direito da expressão (2.29) é a matriz de variância-covariância $Var\ \tilde{\delta}_1$ do estimador de mínimos quadrados em dois estágios (a parte do escalar σ_1):

$$\left(P'\,\frac{X'\,X}{T}\,P\right)^{-1} = \left[\frac{Z_1'\,X}{T}\left(\frac{X'\,X}{T}\right)^{-1}\frac{X'\,Z_1}{T}\right]^{-1} = Var\ \tilde{\delta}_1 \tag{2.30}$$

Portanto, a expressão (2.29) nos diz que:

$$Var\ \hat{\delta}_1 = Var\ \tilde{\delta}_1 + \Delta\left(\frac{X'\,X}{T}\right)^{-1}\Delta' \tag{2.31}$$

Lembrando que a matriz $\left(\frac{X' X}{T}\right)^{-1}$ é uma matriz positiva definida, concluímos que a matriz $\Delta \left(\frac{X' X}{T}\right)^{-1} \Delta'$ é pelo menos positiva semidefinida, o mesmo ocorrendo, portanto, com a matriz $Var\ \hat{\delta}_1 - Var\ \hat{\delta}_1$. Segue-se, então, que o estimador $\hat{\delta}_1$ não é, em geral, assintoticamente eficiente. No caso particular de a equação estrututral ser exatamente identificada, a equação (2.31) reduz-se a:

$$Var\ \hat{\delta}_1 = Var\ \tilde{\delta}_1 \tag{2.32}$$

porque a matriz P é uma matriz quadrada de posto completo e de (2.25) temos que $P = 0$.

A estatística

$$s_1 = \frac{1}{T} (y_1 - Z_1\ \hat{\delta}_1)'\ (y_1 - Z_1\ \hat{\delta}_1) \tag{2.33}$$

é um estimador consistente da variância σ_1. A prova desta propriedade é bastante simples. Substituindo (2.1) em (2.33) obtemos:

$$\begin{aligned} s_1 = {} & \frac{u_1'\ u^1}{T} - \frac{u_1'\ Z_1}{T} (\hat{\delta}_1 - \delta_1) - \frac{(\hat{\delta} - \delta_1)'\ Z_1'\ u_1}{T} \\ & + (\hat{\delta}_1 - \delta_1)' \frac{Z_1'\ Z_1}{T} (\hat{\delta}_1 - \delta_1) \end{aligned} \tag{2.34}$$

A partir de (2.34) concluímos que

$$plim\ s_1 = plim \frac{u_1'\ u_1}{T} = u_1' \tag{2.35}$$

baseados no fato de que $plim\ (\hat{\delta}_1 - \delta_1) = 0$ e de que $plim\ u_{1},\ u_1/T = \sigma_1$.

É fácil derivar a partir de (2.23) e de (2.33) os erros-padrões (assintóticos) da estimativa do estimador indireto generalizado de mínimos quadrados. Isto é, a partir de (2.23) e (2.33) concluímos que os erros-padrões da estimativa dos elementos do vetor $\hat{\delta}_1$ são dados pelas raízes quadradas dos elementos da diagonal principal da matriz:

$$\begin{aligned} & s_1\ [Z_1'\ X\ (X'\ X)^{-2}\ X'\ Z_1]^{-1}\ Z_1'\ X\ (X'\ X)^{-3}\ X'\ Z_1\ \cdot \\ & \cdot\ [Z_1'\ X\ (X'\ X)^{-2}\ X'\ Z^1]^{-1} \end{aligned} \tag{2.36}$$

É um fato bem conhecido que o estimador de mínimos quadrados em dois estágios depende da regra de normalização adotada, isto é, de qual é a variável endógena escolhida para ser a variável "dependente" na equação estrutural em estudo. O mesmo não ocorre, entretanto, com o estimador de máxima verossimilhança de informação limitada. O estimador de mínimos quadrados indireto generalizado tem em comum com o estimador de mínimos quadrados em dois estágios o fato de que ambos dependem da regra de normalização. Obviamente, quando $T \rightarrow \infty$ a regra de normalização deixa de ser relevante.

Cabe, também, assinalar que o estimador indireto generalizado de mínimos quadrados não pertence, em geral, a classe-k de estimadores, a qual é definida por:

$$(\delta_1)_k = \begin{bmatrix} Y_1' Y_1 - k\ \hat{U}_1\ \hat{U}_1 & Y_1' X_1 \\ X_1' Y_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1' - k\ \hat{U}_1' \\ X_1' \end{bmatrix} y_1 \qquad (2.37)$$

onde

$$\hat{U}_1 = [I - X\ (X'\ X)^{-1}\ X']\ Y_1$$

e k é um escalar arbitrário, o qual pode ser estocástico ou não.[10] Comparando-se as expressões (2.37) e (2.18), a conclusão mencionada no início deste parágrafo é obtida.

3. Identificação e estimação

O estimador de mínimos quadrados em dois estágios (MQ2E) pode ser derivado de diferentes maneiras. Para o propósito deste item o enfoque algébrico parece-nos o mais indicado. A equação (2.2) combinada com a equação (2.5) pode ser escrita:

$$y_1 = [X_1 \mid X_o] \begin{bmatrix} \Pi_{10} & I \\ \Pi_{00} & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \beta_1 \end{bmatrix} + v_1 \qquad (3.1)$$

[10] Por exemplo, quando $k = 0$ temos o estimador de mínimos quadrados ordinários; fazendo $k = 1$ em (2.37) obtemos o estimador de mínimos quadrados em dois estágios. Quando $k = \mu$ e μ é a menor raiz da equação:

$$\left| \begin{bmatrix} y_1' \\ Y_1' \end{bmatrix} M_1\ [y_1\ Y_1] - \mu \begin{bmatrix} y_1' \\ Y_1' \end{bmatrix} M\ [y_1\ Y_1] \right| = 0$$

onde:

$$M = I - X\ (X'\ X)^{-'} - X' \quad M_1 = I - X_1\ (X_1'\ X_1^{-1}\ X_1',$$

o estimador da classe k coincide com o estimador de máxima verossimilhança de informação limitada. Para um estudo detalhado da classe k veja, por exemplo, Theil op. cit. cap. 10.

Se os valores de Π_{10} e Π_{00} fossem conhecidos, poderíamos aplicar o método de mínimos quadrados ordinários (MQO) à equação anterior. O estimador assim obtido seria dado por:

$$d_1 = [Q' \ X' \ XQ]^{-1} \ Q' \ X' \ y_1 \tag{3.2}$$

Entretanto, devido ao fato de que Π_{10} e Π_{00} são parâmetros que têm de ser estimados, visto não serem conhecidos *a priori*, o estimador (3.2) não pode ser aplicado na prática. Todavia, o problema é facilmente contornado usando-se a matriz

$$\hat{Q} = \begin{bmatrix} \hat{\Pi}_{10} & I \\ \hat{\Pi}_{00} & 0 \end{bmatrix}$$

ao invés da matriz Q em (3.2), onde $\hat{\Pi}_{10}$ e $\hat{\Pi}_{00}$ são os estimadores de mínimos quadrados da forma reduzida (2.2). Desta maneira, obtemos o estimador de MQ2E:

$$\tilde{\delta}_1 = [\hat{Q}' \ X' \ X]^{-1} \ \hat{Q}' \ X' \ y_1 \tag{3.3}$$

Sob certas condições de regularidade, o estimador de MQ2E é consistente e $\sqrt{T} \ \tilde{\delta}_1 - \delta_1)$ tem uma distribuição assintótica normal.[11] Uma destas condições é a de que a equação estrutural (2.1) seja identificada, o que significa afirmar que, de acordo com (2.8) o posto de Π_{00} é igual a L_1. Obviamente, o estimador de MQ2E (3.3) não existe quando a condição de ordem não é satisfeita, pois esta é uma condição necessária para a identificação da equação estrutural (2.1).

No que diz respeito à existência de uma *estimativa* de MQ2E concluímos, baseados na expressão (3.3), que tal estimativa não existe quando a condição de ordem não é satisfeita. Porém, se a condição de ordem é satisfeita e a condição de posto não é – o que significa dizer que a equação (2.1) não é identificada – uma estimativa de MQ2E é facilmente obtida porque o posto de $\hat{\Pi}_{00}$, em pequenas amostras, é igual a L_1 com probabilidade um. Este fato pode conduzir a um procedimento completamente sem sentido: o de se estimarem os parâmetros de uma equação estrutural não-identificada. Por outro lado, a existência de tal possibilidade evidencia a importância da aplicação de teste de hipótese para saber se o modelo está corretamente especificado. É interessante observar que em

[11] Veja Theil op. cit. cap. 10.

econometria aplicada tais testes, em geral, não são aplicados. O pesquisador contenta-se em aplicar a condição de ordem, esquecendo-se da condição de posto. Isto não quer dizer que testes com a finalidade de testar a condição de posto não existam. Os testes existem e podem ser encontrados na literatura econométrica. [12] Todavia, livros-textos como os de Johnston (1972) e Kmenta (1971), para citar dois livros bastante populares em cursos de econometria, simplesmente nem sequer mencionam a existência destes testes.

4. Multicolinearidade, amostras pequenas, modelos grandes

Swamy e Holmes (1971) e Fisher e Wadycki (1971) provaram, independentemente, que, ao contrário do que era afirmado na literatura econométrica, o estimador de mínimos quadrados de dois estágios existe e é único quando o posto da matriz de variáveis predeterminadas não é igual ao número de colunas desta matriz. Ademais, eles provaram que, na hipótese de o número de observaçeõs ser menor que o número total de variáveis predeterminadas, o que corresponde ao caso de amostras pequenas ou de modelos grandes, o estimador de MQ2E reduz-se a mínimos quadrados ordinários (MQO). A prova destas proposições baseou-se na generalização do estimador de MQ2E através do conceito da matriz inversa generalizada.

A seguir, reproduzimos alguns dos resultados derivados por aqueles autores, usando um enfoque diferente, devido a Basmann (1957), o qual não requer o uso da matriz inversa generalizada. Mais adiante, uma comparação entre os dois enfoques é apresentada.

4.1 O enfoque clássico linear generalizado de Basmann

O estimador de MQ2E pode ser obtido a partir de uma classe de estimadores d definido pelas seguintes relações: [13]

$$d = A\ y_1 \tag{4.1}$$

$$A = [(CZ_1)'\ (CZ_1)]^{-1}\ (CZ_1)' \tag{4.2}$$

$$C'\ C = C = C' \tag{4.3}$$

[12] Veja Koopmans & Hood. op. cit. p. 178-85.

[13] O estimador indireto generalizado de mínimos quadrados não pertence à classe definida por (4.1) – (4.3). A comprovação desta afirmação pode ser feita comparando-se $d = [Z_1'\ CZ_1]^{-1}\ Z_1'\ C'\ y_1$ com (2.19) e notando que $C'\ C \neq C = C'$.

Quando a matriz C é dada pela expressão

$$C = X\ (X'\ X)^{-1}\ X' \tag{4.4}$$

o estimador d,

$$d = [Z_1'\ X\ (X'\ X)^{-1}\ X'\ Z_1]^{-1}\ Z_1'\ X\ (X'\ X)^{-1}\ X'\ y_1 \tag{4.5}$$

é o estimador de MQ2E, o qual é assintoticamente eficiente na classe definida por (4.1) — (4.3). É óbvio que se a matriz $(X'\ X)^{-1}$ não existe, o estimador (4.5) de MQ2E também não existe.

Quando a matriz X não é uma matriz de posto completo de coluna, o que ocorre na presença de multicolinearidade, ou ainda quando temos somente um pequeno número de observações, podemos escrever a matriz X da seguinte forma:

$$X = [X_r \mid X_s],$$

onde a matriz X_r de ordem T x r tem posto igual a r. Neste caso, podemos usar um estimador de MQ2E dentro da classe de estimadores definida por (4.1) — (4.3), com a matriz C definida por:

$$C = X_r\ (X_r'\ X_r)^{-1}\ X_r' \tag{4.6}$$

O estimador assim obtido é um estimador menos eficiente (assintoticamente) que aquele dado por (4.1) — (4.4). O procedimento usual, quando multicolinearidade está presente, é o de escolher um subconjunto das variáveis predeterminadas, o que equivale a usar um estimador com a matriz C dada por (4.6).

No caso de amostras pequenas, o número de variáveis predeterminadas, K, é menor que o número T de observações, e, obviamente, a matriz $X'\ X$ não possui inversa. Dentro da classe de estimadores definida por (4.1) — (4.3) — (4.6), podemos selecionar um estimador d de MQ2E no qual a matriz X_r tenha o posto r igual a T. Da expressão (4.6) concluímos que, se $\rho\ (X_r) = T$ a matriz $C = I$, e de (4.1) e (4.2) segue-se que a estimativa deste estimador de MQ2E é

$$d = (Z_1'\ Z_1)^{-1}\ Z_1'\ y_1 \tag{4.7}$$

A partir de (4.7), concluímos que a estimativa do estimador de MQ2E coincide com aquela obtida através de mínimos quadrados ordinários. Observe, entretanto, que no enfoque apresentado na página anterior o estimador de MQ2E, que dá origem a este resultado, é diferente do estimador usual de MQ2E e, portanto, é menos eficiente (assintoticamente) que (4.5).

4.2 O enfoque da matriz inversa generalizada

O uso da matriz inversa generalizada para resolver o problema encontrado quando o posto da matriz de variáveis predeterminadas não é igual ao número de colunas desta matriz (inclusive os casos de multicolinearidade e de amostras pequenas) conduz à seguinte generalização do estimador de MQ2E. Ao invés de (4.4) a matriz C passa a ser definida pela expressão

$$C = X\ G\ X', \tag{4.8}$$

onde G é uma inversa generalizada da matriz $X'\ X$, isto é, $X'\ X\ G\ X'\ X = X'\ X$. Portanto, (4.1) – (4.3) e (4.8) fornecem o estimador de MQ2E generalizado, o qual é único, embora a matriz G não seja única. Esta propriedade baseia-se numa proposição, bem conhecida da álgebra linear, que nos diz ser a expressão (4.8) invariante quanto à escolha da matriz G.[14]

Com a finalidade de comparar os dois enfoques apresentados, admitamos que o posto da matriz X seja igual a r e que $X = [X_r \mid X_s]$, X_r é uma matriz cujo posto é igual a r e em conseqüência as colunas da matriz X_s são combinações lineares das colunas da matriz X_r. Devido ao fato, já mencionado acima, de que (4.8) é invariante quanto à escolha da matriz G, podemos escolher, sem perda de generalidade, a seguinte inversa generalizada de G:

$$G = \begin{bmatrix} (X_r'\ X_r)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{4.9}$$

onde as matrizes nulas aparecendo em (4.9) têm a ordem apropriada.[15] É fácil verificar que substituindo (4.9) em (4.8) obtemos:

$$C = X'\ G\ X = X_r\ (X_r'\ X_r)^{-1}\ X_r' \tag{4.10}$$

[14] Veja, por exemplo, Searle (1971). É interessante observar que, se usarmos a inversa de Moore-Penrose, a qual é única, ao invés de G, os mesmos resultados são obtidos.

[15] Para demonstrar que G é uma inversa generalizada de G usamos o fato de $X'_s X_s = X'_s X_r (X'_r X_r)^{-1} X'_r X_s$.

A expressão (4.10) é igual à (4.6). Portanto, o estimador de MQ2E generalizado é equivalente ao estimador de MQ2E obtido quando inclui-se apenas um subconjunto de variáveis predeterminadas, observando-se a seguinte relação quanto ao posto da matriz X:

$$\rho\,(X_r) = \rho\,(X) = r \tag{4.11}$$

Entretanto, se acrescentarmos à classe de estimadores (4.1) – (4.3), (4.8) a condição de que

$$\rho\,(X) = K \quad \text{para} \quad T > T_0 \tag{4.12}$$

os dois estimadores passam a ser diferentes, pois o estimador de MQ2E generalizado é assintoticamente eficiente (quando $T \to \infty$, $\rho\,(X) = K$). Na verdade, é esta maneira peculiar de definir o estimador de MQ2E generalizado que dá margem à afirmação de Fisher e Wadycki de que "two stage least squares estimators (even when they are ordinary least squares estimators for the original T) possess consistency and asymptotic efficiency under the usual assumptions". [16]

Bibliografia

Basmann, R. L. A generalized classical method of linear estimation of coeficients in a structural equation. *Econometrica*, v. 25, p. 77-83, 1957.
Dhrymes, P. J. *Econometrics statistical foundations and applications.* New York, Springer-Verlag, 1974.

Fisher, W. D. & Wadycki, W. J. Estimating a structural equation in a large system. *Econometrica*, v. 39, p. 461-65, 1971.

Johnston, J. *Econometric methods.* 2. ed. New York, McGraw Hill, 1976.

Kmenta, J. *Elements of econometrics.* New York, Macmillan, 1971.

Koopmans, T. C. & Hood, W. C. *The estimation of simultaneous linear economic relationships.* In: Koopmans & Hood, ed. *Studies in econometric method,* monografia 14 da Cowles Comission. New York, John Wiley, 1953.

[16] Fisher & Wadycki (1971) p. 464.

Mariano, R. S. The existence of moments of the ordinary least squares and two-stage least squares estimators. *Econometrica,* v. 40, p. 643-52, 1972.

Searle, S. R. *Linear models.* New York, John Wiley, 1971.

Swamy, P. A. V. B. & Holmes, J. The use of undersized samples in the estimation of simultaneous equation system. *Econometrica,* v. 39, p. 455-59, 1971.

Theil, H. *Principles of econometrics.* New York, John Wiley, 1971.

Zellner, A. *An introduction to bayesian inference in econometrics.* New York, John Wiley, p. 255-8, 1971.