

MULTICOLINEARIDADE EM FUNÇÃO DE PRODUÇÃO

Edgar A. Lanzer ⁽¹⁾

1. O presente trabalho é uma tentativa de contribuição à metodologia de análise usualmente empregada em estudos de funções de produção agregadas ao nível de empresa. A proposição feita, aberta à discussão, é a de que o emprego de programação matemática pode diminuir sensivelmente a margem de erro numa análise econômica do tipo néo-clássico quando alta multicolinearidade é detectada entre os regressores. na amostra disponível.

2. Em um número significativo de trabalhos sobre funções de produção em Economia Rural o "Problema de Multicolinearidade" é mencionado. Em particular, em estudos de função de produção agregadas ao nível de empresa frequentemente se menciona que as firmas tendem a se agrupar em torno de uma linha de expansão.

3. Assumimos, para fins de ilustração, que a *i*-ésima observação tenha sido gerada pelo modelo:

$$\left\{ \begin{array}{l} y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} = \epsilon_i \\ -a + x_{1i} - x_{2i} = u_i \end{array} \right. \quad \begin{array}{l} \text{(Cobb-Douglas)} \\ \text{(Linha de Expansão)} \end{array}$$

Consideramos preços não-aleatórios, x_1 como variável exógena (terra, por exemplo) e que os termos aleatórios sejam não correlacionados e homoscedásticos.

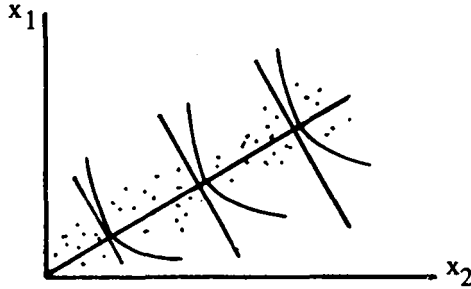
É possível demonstrar, sob a hipótese adicional de normalidade, que os estimadores de máxima verossimelhança de β_1 e β_2 são os estimadores de mínimos quadrados convencionais ⁽²⁾. Assim sendo, os parâmetros da função de produção são consistentemente estimados pelo modelo de equação única.

4. Graficamente, as observações geradas pelo modelo descrito se assemelham ao diagrama de dispensação (no espaço dos fatores) apresentados na figura aposta a seguir.

5. A estimação eficiente de β_1 e β_2 (em termos absolutos) é dificultada pela alta correlação entre X_1 e X_2 . A qualidade das estimativas obtidas é normalmente menor que o convencionalmente requerido para a análise econômica posterior. Caracteriza-se assim o chamado problema de multicolinearidade.

⁽¹⁾ Professor-Assistente, Departamento de Ciências Econômicas e Centro de Estudos e Pesquisas Econômicas da Universidade Federal do Rio Grande do Sul.

⁽²⁾ Vide A. ZELLNER, J. KMENTA e J. DRÈZE (Specification and Estimation of Cobb-Douglas Production Function Models *Econometrica*, 34-4 - 1966 - pp. 784-95) para prova de afirmativa em um modelo semelhante. Estes autores incluem ainda uma ilustrativa discussão das hipóteses formuladas sobre o comportamento dos erros.



6. Se o objetivo da pesquisa for o de obter um instrumento de previsão, não há porque deixar de eliminar uma das variáveis explicativas do modelo. Se, todavia, o objetivo for o de desenvolver uma análise econômica nos moldes néo-clássicos, a eliminação de variáveis pode ter sérias conseqüências sobre a adequação das conclusões obtidas ao problema real. Isto é devido, naturalmente, à inconsistência dos estimadores dos parâmetros dos regressores não descartados por erro de especificação do modelo.

7. É sabido que as formas que o problema de multicolinearidade pode assumir se estendem além do simples exemplo ilustrativo visto acima. Em geral, multicolinearidade existe sempre que se puder representar uma ou mais colunas da matriz de dados através de combinações entre as demais.

8. Após esta colocação inicial é conveniente mencionar que o autor desconhece uma solução plenamente satisfatória para o problema ⁽³⁾.

9. O presente trabalho tem um enfoque essencialmente pragmático. O objetivo central é delinear um esquema de análise econômica condicionada que visa diminuir a probabilidade de erro nas conclusões de estudos de função de produção sob presença de multicolinearidade.

10. O passo inicial é o reconhecimento de que multicolinearidade é uma restrição imposta ao analista pelos dados dos quais ele pretende extrair informações.

11. Digamos que a função de produção possa ser representada por:

$$\bar{Y} = \bar{X}_1\beta_1 + \bar{X}_2\beta_2 + X\mathfrak{P}^* + \epsilon$$

onde \bar{Y} é o vetor ($n \times 1$) de observações da produção

\bar{X}_1, \bar{X}_2 são vetores ($n \times 1$) de observações dos fatores X_1 e X_2

⁽³⁾ Todavia o critério do quadrado do erro-padrão do C. TOROVEZCARRONDO e T. D. WALLACE (A test of the mean square errors criterion for restrictions in linear regression - Journal of the American statistical Association - 63 - 1968 - pp. 558-72) pode ser de valia em casos determinados.

$X^* = [\bar{X}_3, \bar{X}_4 \dots \bar{X}_k]$ é a matriz ($n \times k-2$) de observações do emprego dos fatores X_3 a X_k

ϵ é um vetor ($n \times 1$) de resíduos aleatórios.

β_1, β_2 são parâmetros (escalares).

$\beta^* = [\beta_3 \beta_4 \dots \beta_k]^T$ é um vetor ($k-2 \times 1$) de parâmetros.

Sem perda de generalidade, vamos agora assumir que os fatores X_3 a X_k são exógenos, e que o comportamento dos empresários é tal que produz:

$$\bar{X}_1 - X^* \gamma = \zeta \quad (\text{sendo } \gamma = [\gamma_3 \gamma_4 \dots \gamma_k]^T)$$

$$\text{e } \bar{X}_2 - X^* \phi = u \quad (\text{sendo } \phi = [\phi_3 \phi_4 \dots \phi_k]^T) \quad (4)$$

$$\epsilon: [\epsilon \ \zeta \ u] \sim N \left\{ \begin{bmatrix} \phi & \phi & \phi \end{bmatrix} : \begin{bmatrix} \sigma^2 & \phi & \phi \\ \phi & \sigma^2 & \phi \\ \phi & \phi & \sigma^2 \end{bmatrix} \right\}$$

12. É óbvio que $\beta = [\beta_1 \ \beta_2 \ \beta^*]$ pode ser estimado consistentemente por mínimos quadrados convencionais na regressão de Y sobre X_1 a X_k . É evidente porém que a eficiência (absoluta) dos estimadores será pequena, dado o comportamento dos empresários.

13. Se, baseado nos testes "t", o analista decide pela irrelevância de X_1 e X_2 na função de produção e decide pela inclusão de X_3 a X_k no modelo, o resultado é:

$$\bar{Y} = X^* \beta^{**} + \epsilon^*$$

$$\text{onde } \beta^{**} = \beta + \gamma \beta_1 + \phi \beta_2$$

$$\text{e } \epsilon^* = \zeta \beta_1 + u \beta_2 + \epsilon$$

Prosseguindo, inadvertidamente, na análise econômica pela resolução de:

$$\text{Max } E(L^*) = P_y \left(\sum_{i=1}^k X_i \beta_i^{**} \right) - \sum_{i=3}^k P_i X_i$$

o analista obteria resultados que diferem da realidade:

(4) Deve ser notado que esta formulação permite aleatoriedade na determinação dos níveis dos fatores X_1 e X_2 por parte dos empresários. É o que ZELLNER, KMENTA e DRÈZE (op. cit.), classificam na categoria de "Erros Humanos".

$$\text{Max } E(L) = P_y (E(X_1) \beta_1 + E(X_2) \beta_2 + \sum_{i=3}^k X_i \beta_i) - P_1 E(X_1) - P_2 E(X_2) -$$

$$- \sum_{i=3}^k P_i X_i$$

Isto é, uma vez que a especificação foi inadequada e como os parâmetros das variáveis mantidas no modelo incluem os efeitos das variáveis omitidas, tende-se a super ou subestimar o nível ótimo de emprego dos fatores X_3 a X_k conforme estes sejam complementos ou substitutos dos fatores omitidos.

14. É óbvio, todavia, que o resultado verdadeiro poderia ser obtido por:

$$\text{Max } E(L^{**}) = P_y \left(\sum_{i=3}^k X_i \beta_i^{**} \right) - P_1 E(X_1) - P_2 E(X_2) - \sum_{i=3}^k P_i X_i$$

$$\text{S.r. } E(X_1) = \sum_{i=3}^k X_1 \gamma_i$$

$$E(X_2) = \sum_{i=3}^k X_1 \phi_i$$

15. Na prática, evidentemente, os verdadeiros parâmetros do modelo são desconhecidos. O resultado do parágrafo anterior sugere, no entanto, um método a ser seguido.

16. Se o conhecimento do problema estudado indicar que as variáveis X_1 e X_2 deveriam ser mantidas na análise econômica, mas eliminadas da função de produção estimada por causa da instabilidade do hiperplano ajustante, o procedimento sugerido é o seguinte:

- obtenha-se estimativas β^{**} , γ e ϕ por mínimos quadrados convencionais;
- substitua-se as estimativas obtidas pelos respectivos parâmetros no problema enunciado no parágrafo 14;
- proceda-se à otimização condicionada.

17. Em determinados casos, o problema de otimização condicionada pode ser mais flexível, dependendo da disponibilidade de programas e computadores adequados.

$$\text{Max } P_y \left(\sum_{i=3}^k X_i \beta_i^{**} \right) - P_1 X_1 - P_2 X_2 - \sum_{i=3}^k P_i X_i$$

$$\text{S.r. } \bar{X}_i - \sigma_i \leq X_i \leq X_i + \sigma_i \quad i = 3, 4, \dots, k$$

$$\sum_{i=3}^k X_i \hat{\gamma}_i - \hat{\sigma}_1 \leq X_1 \leq \sum_{i=3}^k X_i \hat{\gamma}_i + \hat{\sigma}_1$$

$$\sum_{i=3}^k X_i \phi_i - \hat{\sigma}_2 \leq \hat{X}_2 \leq \sum_{i=3}^k X_i \hat{\phi}_i + \hat{\sigma}_2$$

onde \bar{X}_i e σ_i são respectivamente, a média e o "desvio-padrão" do nível de emprego do fator X_i entre as empresas da amostra ($i = 3, 4, \dots, k$).

$\hat{\sigma}_j$ é a estimativa do erro-padrão da regressão de X_j ($j = 1, 2$) sobre (X_3 a X_k).

As raízes do programa acima situam-se na vizinhança do "centro de gravidade" da amostra.

18. A análise econômica não se restringe à determinação numérica do nível ótimo condicionado de emprego de fatores. Em particular, os multiplicadores de Lagrange associados às diversas restrições fornecem indicações sobre a sensibilidade da função objetivo a variações (infinitesimais) nos níveis estabelecidos para aquelas. Fica aberta, também, a possibilidade de análise paramétrica dos intervalos admitidos na análise.

19. Uma palavra de cautela se faz necessária na utilização do esquema apresentado: a formulação das restrições de multicolinearidade deve ser possível de explicação lógica pelo analista. Não se trata aqui de simplesmente associar variáveis excluídas com variáveis incluídas na função de produção selecionada. A racionalização do esquema adotado em um estudo específico é aspecto fundamental da análise.

20. Duas teses de Mestrado em Economia Rural ⁽⁵⁾ a serem publicadas brevemente pelo IEPE/UFRGS utilizam esquemas analíticos similares ao enunciado neste trabalho. Os resultados iniciais podem ser considerados satisfatórios.

21. Por fim, embora a análise condicionada apresente uma forte base intuitiva (pelo menos ao autor), uma demonstração rigorosa de sua eficiência relativamente à análise incondicional ainda está por ser feita. Considerando os elevados custos adicionais do esquema proposto e que, de qualquer modo, os resultados numéricos obtidos em casos específicos não podem, geralmente, ser considerados mais do que meros indicadores, a questão de eficiência de análise em termos da adequação das conclusões obtidas é um ponto crítico que merece maior discussão. Esta, porém, está além do escopo do presente trabalho.

⁽⁵⁾ ADAMS, R.I. Estrutura e Rentabilidade do Capital em São Borja - RS. Tese de Mestrado - IEPE/UFRGS - 1974. (A ser publicada).

PONS, J. L. M. Análise Econômica de Alocação de Recursos em um Grupo de Propriedades Pecuárias do Município de São Gabriel - RS - Tese de Mestrado - IEPE/UFRGS - 1974 - (A ser publicada).