UM PROBLEMA ECONOMÉTRICO NO USO DE VARIÁVEIS CLIMÁTICAS EM FUNÇÕES DE PRODUÇÃO AJUSTADAS A DADOS EXPERIMENTAIS

Rodolfo Hoffmann*
Victor Hugo da Fonseca Porto**

SINOPSE

Neste trabalho deduz-se qual é o número máximo de variáveis climáticas que podem ser introduzidos em uma função de produção ajustada aos dados obtidos de um experimento de adubação repetido durante n anos no mesmo local. Se não houver nenhuma modificação na maneira de conduzir o experimento durante os n anos, esse número máximo de variáveis climáticas é $n-1$. Se houver alterações, de um ano para outro, na variedade utilizada, sendo m (com $m < n$) o número de diferentes variedades utilizadas durante os n anos, mas de maneira que em cada ano seja cultivada apenas uma das variedades, então o número máximo de variedades climáticas que podem ser introduzidos na função de produção é $n - m$.

SUMMARY

Consider the data obtained from a fertilizer experiment that was repeated during n years in the same field. In this paper it is proved that no more than n - 1 climatic variables can be included in a production function fitted to these data. If there are changes in the crop variety cultivated, so that m different varieties, with m < n, are cultivated, only one in each year, than no more than n-m climatic variables can be included in the production function.

---

* Professor do Departamento de Economia e Sociologia Rural da Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, SP.

** Pesquisador da EMBRAPA-UEPAE, Pelotas, RS.

UM PROBLEMA ECONOMÉTRICO NO USO DE VARIÁVEIS CLIMÁTICAS EM FUNÇÕES DE PRODUÇÃO AJUSTADAS A DADOS EXPERIMENTAIS

Rodolfo Hoffmann
Victor Hugo da Fonseca Porto

1. INTRODUÇÃO

Existe uma tendência crescente, por parte dos economistas agrícolas, em introduzir variáveis climáticas nas funções de produção ajustadas a dados experimentais. Isso porque as variações anuais das condições climáticas são uma das causas fundamentais do risco associado às explorações agrícolas.

Como exemplos de trabalhos que incluem variáveis climáticas em funções de produção podem-se citar os de De JANVRY (1), FONSECA (3) e PORTO (4)[1].

Neste trabalho é analisada a relação entre o número máximo de variáveis climáticas que podem ser incluídas na função de produção e o número de anos durante os quais é repetido o experimento.

2. A TESE A SER DEMONSTRADA

Considere-se um experimento para determinada cultura, repetido durante n anos, no mesmo local. Seja p o número de parcelas do experimento em cada ano. Tem-se, portanto, um total de n observações para o ajuste da função de produção. Admita-se, ainda, que a variedade cultivada não seja a mesma em todos os anos, isto é, seja m, com $m < n$, o número de diferentes variedades utilizadas. É importante ressaltar que em um dado ano é cultivada apenas uma das variedades.

Vai-se demonstrar que o número máximo de variáveis climáticas que podem ser incluídas na função de produção é $n - m$. Caso sejam introduzidos mais do que $n - m$ variáveis climáticas, os parâmetros da função de produção não podem ser estimados devido à existência de multicolinearidade perfeita.

[1] Um analista bastante extensa de referências é apresentada por DILLON (2).

3. DEMONSTRAÇÃO PARA O CASO EM QUE NÃO EXISTE MUDANÇA NA VARIEDADE CULTIVADA

Para exemplificar, considere-se um experimento para comparar 4 diferentes espaçamentos na cultura de determinado híbrido de milho. Suponha-se que um delineamento inteiramente casualizado, com duas repetições de cada um dos tratamentos, seja conduzido no mesmo local durante n = 5 anos consecutivos. O experimento tem, portanto, p = 8 parcelas, com um total de n = 40 observações nos 5 anos.

Um modelo de função de produção que pode ser ajustado neste caso é:

$$Y_j = \gamma_0 + \beta_1 X_j + \beta_2 X_j^2 + \sum_{i=1}^{K} \gamma_i W_{ij} + u_j \qquad \text{(I)}$$

com j = ..., np, e onde Y_j é a produção de milho: X_j é a área por planta (ou alguma outra medida do espaçamento), W_i (com i = 1,... k) são variáveis climáticas (temperatura, insolação, pluviosidade, umidade relativa etc., em diferentes períodos do cicio vegetativo) e u_j são erros aleatórios independentes com média zero e variância constante.

O modelo poderia incluir, ainda, termos relativos às interações entre o espaçamento e as variáveis climáticas.

Admita-se que os n p valores observados sejam ordenados de maneira tal que se tenha primeiro os p valores referentes ao primeiro ano, em seguida os p valores referentes ao segundo ano, e assim por diante, obedecendo sempre a uma mesma ordenação dos tratamentos.

O modelo (I) pode ser escrito como seque:

$$Y = \gamma_0 w_0 + \beta_1 x_1 + \beta_2 x_2 + \sum_{i=1}^{K} \gamma_i w_i + u \qquad \text{(II)}$$

onde Y é o vetor-coluna com os valores de $y_{j,}$, wo é um vetor-coluna cujos n p elementos são todos iguais a 1, x_1 é o vetor - coluna cujos elementos são os valores de X_j, x_2 é o vetor-coluna cujos elementos são os valores de X_j^2 w_{ij} é o vetor-coluna I e u é o vetor-coluna com os valores de W_{ij} e u_j.

Considere-se a matriz n x (k + 1):

$$\Omega = [\omega_0 \omega \ldots \omega_k]$$

onde w_0 é um vetor-coluna cujos n elementos são todos iguais a 1 e w_i (com i = 1, . . .,k) é o vetor-coluna com os n valores distintos (um para cada ano, em ordem cronológica) da variável climática W_{ij}.

Seja ρ a característica de Ω. Como a característica de uma matriz não pode ser maior do que qualquer de suas dimensões, tem-se que:

$$\rho \leq n \tag{III}$$

Por outro lado, sabe-se que a característica de uma matriz é igual ao número máximo de colunas dessa matriz linearmente independentes.

Vai-se admitir, agora, que o número de colunas de Ω é maior do que o número de anos do experimento, isto é, vai-se admitir que:

$$k + 1 > n \tag{IV}$$

De (III) e (IV) seque-se que:

$$k + 1 > \rho$$

Então as $k+1$ colunas de Ω são linearmente dependentes e pode-se escrever:

$$\sum_{i=0}^{k} \lambda_i \omega_i = 0 \tag{V}$$

onde λ_i são constantes, com ao menos um $\lambda_i \neq 0$, e 0 é um vetor-coluna de zeros.

Se i é um vetor-coluna com p elementos, todos iguais a 1, pode-se verificar que w_i é igual ao produto de Kronecker[2] de w_i por t, isto é:

$$w_i = \omega_i \otimes_i \qquad \text{para } i= 0,. . .,k \tag{VI}$$

Utilizando esse resultado, verifica-se que:

$$\sum_{i=0}^{k} \lambda_i w_i = \sum_{i=0}^{k} \lambda_i (\omega_i \otimes_i) = \left(\sum_{i=0}^{k} \lambda_i \omega_i \right) \otimes_i$$

Lembrando (V), conclui-se que:

$$\sum_{i=0}^{k} \lambda_i \omega_i = 0$$

Esse resultado mostra que os vetores-coluna w_i (com i = 0,. . . .,k) são linearmente dependentes.

[2] Para uma exposição sucinto sobre o produto de Kronecker e suas propriedades, ver referência (5, p. 303-6).

Conseqüentemente, sempre que k + 1 > n o modelo (I) não pode ser ajustado porque existe multicolinearidade perfeita.

Pode-se concluir, então, que uma condição necessária para que seja possível estimar os parâmetros do modelo (I) é que k + 1 ≤ n ou k ≤ n - 1, isto é, pode-se incluir no modelo, no máximo, n - 1 variáveis climáticas.

4. DEMONSTRAÇÃO PARA O CASO EM QUE EXISTE MUDANÇA NA VARIEDADE CULTIVADA

Na seção anterior admitiu-se que o mesmo híbrido (ou variedade) foi cultivado durante todos os 5 anos do experimento. Vai-se supor, agora, que foi utilizado um híbrido de milho nos dois primeiros anos e um outro híbrido nos três últimos anos. Neste caso, deve-se introduzir no modelo uma variável binária (**dummy variable**) Z_j, destinada a detectar a diferença de produção entre as duas variedades. Faça-se $Z_j = 0$ para todas as observações dos dois primeiros anos, quando foi utilizado um determinado híbrido, e $Z_j = 1$ para todas as observações dos três últimos anos, quando foi utilizado outro híbrido. O modelo da função de produção fica como segue:

$$Y_j = \gamma_0 + \beta_1 X_j + \beta X^2{}_j + \alpha Z_j + \sum_{i=1}^{k} \gamma_i W_{ij} + u_j \qquad \text{(VI)}$$

ou

$$Y_j = \gamma_0 w_0 + \beta_1 x_1 + \beta_2 x_1 + \alpha z + \sum_{i=1}^{k} \gamma_i w_i + u \qquad \text{(VI)}$$

onde z é um vetor-coluna com os valores de Z_j

Seja ζ o vetor-coluna cujos n elementos são os valores assumidos por Z_j em cada um dos n anos, em ordem cronológica. Para o exemplo que se está considerando, tem-se que:

$$\zeta = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Verifica-se que:

$$Z = \zeta \otimes \iota \qquad \text{(IX)}$$

Vai-se redefinir a matriz Ω como segue:

$$\Omega = [\omega_0 \omega_1 ... \omega_k \zeta]$$

com dimensões $n\, x\, (k\ + 2)$.

Se ρ é a característica de Ω, tem-se:

$$\rho < n \qquad \text{(X)}$$

Vai-se admitir, agora, que o número de colunas de Ω é maior do que o número de anos do experimento, isto é, vai-se admitir que:

$$k\ +\ 2\ >\ n \qquad \text{(XI)}$$

De (X) e (XI) segue-se que:

$$k\ +\ 2\ > \rho$$

Então as $k\ +\ 2$ colunas de Ω são linearmente dependentes e pode-se escrever:

$$\sum_{i=0}^{k} \lambda_i \omega_i + \theta\zeta = 0 \qquad \text{(XII)}$$

onde λ_i e θ são constantes, não todas nulas.

De acordo com (VI) e (IX), tem-se que:

$$\sum_{i=0}^{k} \lambda_i w_i + \theta z = \sum_{i=0}^{k} \lambda_i (\omega_i \otimes_{\iota}) + \theta(\zeta_{\otimes\iota}) = \left(\sum_{i=0}^{k} \lambda_i \omega_i + \theta\zeta \right) \otimes \iota$$

Lembrando (XII), conclui-se que:

$$\sum_{i=0}^{k} \lambda_i w_i + \theta\zeta = 0$$

isto é, os vetores-coluna w_i (com i = 0, ..., k) e z são linearmente dependentes. Fica então, demonstrando que se se tiver $k + 2 >\ n$ o modelo (VII) não pode ser ajustado devido à existência de multicolinearidade perfeita.

Portanto, uma condição necessária para que seja possível estimar os parâmetros do modelo (VII) é que $k + 2 \leq n$ ou $k \geq\ n - 2$, isto é, pode-se incluir no modelo, no máximo, $n - 2$ variáveis climáticas.

Generalizando, se um experimento é repetido durante n anos no mesmo local e são utilizados m diferentes híbridos ou variedades, mas apenas um dado híbrido ou variedade é cultivado em cada ano, a função de produção ajustada aos dados experimentais pode incluir, no máximo, $n-m$ variáveis climáticas, isto é, deve-se ter:

$$k < n-m$$

Veja-se uma outra maneira de obter esse resultado.

Para o problema que se está discutindo, a característica essencial, tanto das variáveis climáticas como das variáveis binárias, destinadas a detectar os efeitos de mudanças na variedade cultivada, é que o valor de cada variável é o mesmo para todas as observações relativas a um ano, ou seja, o valor de cada variável só se modifica quando se consideram observações de anos diferentes. Assim, no subespaço dessas variáveis tem-se, no máximo, n pontos distintos. Tendo em vista a existência do termo constante no modelo, conclui-se que se pode incluir, no máximo, n-1 variáveis climáticas e binárias do tipo mencionado.

Se são utilizadas m variedades distintas, será necessário incluir m-1 variáveis binárias para captar os efeitos das mudanças de variedade. Se k é o número de variáveis climáticas incluídas no modelo, deve-se ter:

$$k + (m-1) \leq n-1$$

ou

$$k \leq n - m$$

FONSECA (3), em um estudo econômico da resposta do trigo à adubação nitrogenada, utilizou dados de um experimento conduzido no mesmo local durante 13 anos. No submodelo I, são incluídos na função de produção 15 regressares constituídos por variáveis climáticas, variáveis binárias destinadas a captar o efeito de mudanças na variedade utilizada, ou produtos dessas variáveis (3). Portanto, esses 15 regressares têm todos eles a característica de que seu valor só se modifica quando se consideram observações de anos diferentes. Uma vez que, de acordo com o que foi demonstrado, há multicolinearidade perfeita, as estimativas dos parâmetros desse submodelo que são apresentados no trabalho não têm nenhuma validade. Felizmente, esse submodelo não foi o escolhido para o desenvolvimento posterior de análise.

5. CONSIDERAÇÕES FINAIS

No tipo de experimento descrito, é impossível incluir na função de produção todos os fatores que variam de ano para ano e que afetam a produção.

Um problema que deve ser lembrado são as modificações, entre anos, na qualidade do solo cultivado. Se o experimento é repetido no mesmo local, ter-se-ia de considerar, a rigor, os efeitos residuais, nos anos posteriores, do tratamento recebido por uma parcela. O fenômeno é bastante conhecido no caso de experimentos de adubação, mas o problema existe, em princípio, quaisquer que sejam os tratamentos estudados. No tipo de experimento descrito é impossível distinguir os efeitos de mudanças climáticas dos efeitos residuais dos tratamentos; isso exigiria um experimento bem mais complexo.

Uma maneira de tentar resolver o problema da limitação do número de variáveis climáticas que podem ser introduzidos na função de produção é a construção de um índice de condições climáticas, isto é, uma função do conjunto das variáveis climáticas, de tal maneira que seu valor indicasse se as condições climáticas foram mais ou menos favoráveis ao desenvolvimento da cultura considerada.

6. LITERATURA CITADA

1. De JANVRY, A. Optimal levels of fertilization under risk: the potential for corn and wheat fertilization under alternative price policies in Argentina. American Journal of Agricultural Economics, 54(1): 1-10, 1972.
2. DILLON, J. L. The analysis of response in crop and livestock production. 2. ed. Oxford, Pergamon Press, 1977.
3. FONSECA, V.O. Análise econômica da aplicação de doses e fontes de nitrogênio na cultura do trigo, sob condições de risco, em Pelotas, Rio Grande do Sul. Porto Alegre, IEPE-UFRGS, 1976. (Tese de M.S.).
4. PORTO, V.H.F. Análise econométrica de dados experimentais sobre um sistema de produção trigo-soja, para a cultura de trigo. Piracicaba, São Paulo, Departamento de Economia e Sociologia Rural da ESALQ-USP, 1980. (Dissertação de M.S.).
5. THEIL, H. Principles of econometrics New York, John Wiley, 1971.