

IDENTIFIZIERUNGSVORAUSSETZUNGEN VON MODELLEN

Th. J. FERRARI

Institut für Bodenfruchtbarkeit, Haren-Groningen,
 Niederlande

Beitrag zum Symposium "Biophysik pflanzlicher Systeme",
 Quedlinburg, 28. - 30. Mai 1968

1. Fragestellung

In der landwirtschaftlichen Forschung hat man meistens mit Problemen zu tun, in denen Kettenprozesse eine wichtige Rolle spielen. In die Modelle, die diese Erscheinungen beschreiben, sollen darum Kettenprozesse aufgenommen werden. Ein einfaches Modell besteht z.B. darin, daß der Effekt y_1 durch die primären Ursachen x_1 und x_2 beeinflusst wird. Der zweite Effekt y_2 steht nur direkt unter dem Einfluß von x_2 und y_1 , wobei die Ursache x_1 ihren Einfluß auf y_2 nur indirekt über y_1 ausübt. Die Strukturgleichungen, die für dieses Modell aufgestellt werden können, lauten:

$$y_1 = a_1 + a_{11}x_1 + a_{12}x_2 \quad (1)$$

$$y_2 = a_2 + b_{21}y_1 + a_{22}x_2 \quad (2)$$

Änderungen von x_1 und x_2 werden die Variablen y_1 und y_2 beeinflussen. Die Gesamteinflüsse p_{11} , p_{12} , p_{21} und p_{22} auf y_1 und y_2 können mit Hilfe der reduzierten Strukturgleichungen (1) und (2) bestimmt werden (FERRARI, 1964):

$$y_1 = a_1 + a_{11}x_1 + a_{12}x_2 \quad (1)$$

$$y_2 = a_2 + a_{12}b_{21} + (a_{11}b_{21})x_1 + (a_{12}b_{21} + a_{22})x_2 \quad (2)$$

Das Problem besteht darin, ob die Teileinflüsse a_{ij} und b_{im} aus den beobachteten Parametern p berechnet werden können. In diesem Beispiel ist das möglich. Viele Systeme sind aber nicht lösbar, weil das Modell nicht bestimmten Voraussetzungen entspricht. Im folgenden sollen diese Voraussetzungen abgeleitet werden.

2. Identifizierungsvoraussetzungen

Die allgemeine Form der Strukturgleichungen mit Zufallsgliedern u_i kann wie folgt geschrieben werden:

$$\begin{aligned}
 b_{11}y_1^{(n)} + \dots + b_{1M}y_M^{(n)} + a_{11}x_1^{(n)} + \dots + a_{1L}x_L^{(n)} &= u_1^{(n)} \\
 \vdots &\vdots \\
 b_{M1}y_1^{(n)} + \dots + b_{MM}y_M^{(n)} + a_{M1}x_1^{(n)} + \dots + a_{ML}x_L^{(n)} &= u_M^{(n)}.
 \end{aligned} \quad (4)$$

Hierbei wird angenommen, daß die Stichprobe aus N Beobachtungen besteht und (n) eine beliebige Beobachtung bezeichnet. Die Anzahl der Effekte und Gleichungen ist M , die Anzahl der primären Ursachen x_j ist L . Die Aufgabe der Untersuchung besteht darin, die Parameter b_{im} und $a_{j\ell}$ zu bestimmen.

Es ist klar, daß für reelle Modelle verschiedene Parameter a priori gleich Null angenommen werden können. Eine bestimmte Gleichung kann also, wenn die Parameter so normiert sind, daß der Parameter eines der y gleich 1 wird, wie folgt geschrieben werden:

$$y_1^{(n)} = b_{12}y_2^{(n)} + \dots + b_{1,m+1}y_{m+1}^{(n)} + a_{11}x_1^{(n)} + \dots + a_{1L}x_L^{(n)} + u_1^{(n)} \quad (5)$$

Diese Gleichung für y_1 hat also $(m+1)M$ Effekte y_j und $1 \leq L$ primäre Ursachen x_j .

Das System (4) für N Beobachtungen lautet in Matrizenform:

$$YB + XA = U, \quad (6)$$

worin Y die $(N \times M)$ -Matrix der Variablen y_j ist, X die $(N \times L)$ -Matrix der Variablen x_j , B und A die $(M \times M)$ - bzw. $(L \times M)$ -Matrizen der Koeffizienten und U die Matrix der Zufallselemente.

Unter der Annahme, daß die Matrix B nicht singular ist, ist es möglich, die Strukturgleichungen zu reduzieren:

$$Y = -XAB^{-1} + UB^{-1} = XP + V$$

und die Parameter P zu berechnen.

Um die Koeffizienten der Matrizen A und B zu gewinnen, müssen die Strukturgleichungen aus den reduzierten Gleichungen ausgerechnet werden. Zur Auswertung z.B. der Strukturgleichung von y_1 in (5) werden die $(m+1)$ reduzierten Strukturgleichungen der Effekte $y_1 \dots y_{m+1}$ folgendermaßen niedergeschrieben:

$$\begin{aligned}
 1y_1 - p_{1,1+1}x_{1+1} - \dots - p_{1,L}x_L &= p_{1,1}x_1 + \dots + p_{1,1}x_1 &= h_1 \\
 0y_1 - p_{2,1+1}x_{1+1} - \dots - p_{2,L}x_L &= p_{2,1}x_1 + \dots + p_{2,1}x_1 - y_2 &= h_2 \\
 &\vdots &\vdots \\
 0y_1 - p_{m+1,1+1}x_{1+1} - \dots - p_{m+1,L}x_L &= p_{m+1,1}x_1 + \dots + p_{m+1,1}x_1 - y_{m+1} &= h_{m+1}
 \end{aligned}
 \tag{8}$$

Aus diesem System kann mit Anwendung der CRAMERSCHEN Regel y_1 durch $y_2 \dots y_{m+1}$; $x_1 \dots x_L$ ausgedrückt werden:

$$y_1 = |C_1| / |C|, \tag{9}$$

worin C die Matrix der Koeffizienten der Variablen auf der linken Seite ist, also

$$C = \begin{bmatrix} 1 - p_{1,1+1} & \dots & -p_{1,L} \\ \vdots & & \vdots \\ 0 - p_{m+1,1+1} & \dots & -p_{m+1,L} \end{bmatrix}$$

und

$$C_1 = \begin{bmatrix} h_1 - p_{1,1+1} & \dots & -p_{1,L} \\ \vdots & & \vdots \\ h_{m+1} - p_{m+1,1+1} & \dots & -p_{m+1,L} \end{bmatrix}$$

Eine eindeutige Lösung für y_1 ist nur möglich, wenn die C -Matrizen quadratisch und die Determinanten dieser Matrizen nicht gleich Null sind. Die erste Forderung bedingt, daß $L - 1 + 1 = m + 1$ oder $L - 1 = m$ ist. Das kann im Modell immer leicht kontrolliert werden. Weiter soll der Rang der Matrizen gleich $m + 1$ sein.

Ist $L - 1 < m$, dann ist die Gleichung (5) "unteridentifiziert" oder "nicht identifizierbar"; d.h. y_1 kann nicht durch die Variablen $y_2 \dots y_{m+1}$; $x_1 \dots x_L$ ausgedrückt werden. Ein eindeutiger Ausdruck ließe sich erzielen, wenn $m - (L - 1)$ Reihen, also $m - (L - 1)$ Variable y_1 aus der Matrix C_1 entfernt werden. Gegen eine solche Handlungsweise bestehen aber große Bedenken, weil dadurch das Modell wesentlich geändert wird und nicht mehr den Vorstellungen des Forschers entspricht.

3. Schlußbemerkung

Die Bedeutung vieler landwirtschaftlicher Untersuchungen wird durch die Möglichkeit bestimmt, ein Bild der komplexen Wirklichkeit zu geben. Hieraus folgt, daß man immer versuchen sollte, die zahlreichen Variablen dieser Wirklichkeit mit ihren vielen Wechselbeziehungen zugleich zu untersuchen. Diese Wirklichkeit kann dann aber nur mit vielen Variablen und mit vielen Gleichungen beschrieben werden.

Die Erfahrungen zeigen, daß solche Systeme meistens nicht identifizierbar und dadurch analytisch nicht zu fassen sind. Viel Geld, Zeit und Arbeit sind auf Experimente verwendet worden, die prinzipiell keine Lösung der Probleme geben konnten. Es ist notwendig, daß der Forscher sich immer dieser "Nicht-Identifizierbarkeit" bewußt ist. Andernfalls muß er die analytische Methode verlassen und andere Untersuchungsmethoden, z.B. das Simulationsverfahren anwenden.

Literatur

FERRARI, Th.J., Biometrische Zeitschrift 6, H.2, 89 (1964)

UNDER-IDENTIFICATION CONDITIONS OF MODELS

Starting-point is a model without limitations in which chain processes and feedback relationships can be taken up. The mathematical description is given by a system of equations (4) of which the parameters must be determined.

It is not always possible to determine these parameters. The conditions for these determinations can be derived. An under-identification in an equation of the system exists if the difference in the number of independent variables in the model and in the equation concerned is smaller than the number of dependent variables in the same equation. A solution can only be reached by altering the model.

In agricultural research we meet very often models with many variables and relationships. An investigation of such models cannot be designed and performed well if the researcher has not realized these identification conditions. As a matter of fact, we meet the under-identification in many agricultural experiments.