

The Problem of Parameter Identification of Structural Equation Models with both Formative and Reflexive Relationships

*Identificabilità Parametrica nei Modelli ad Equazioni Strutturali
con Relazioni Formative e Riflessive*

GABRIELE CANTALUPPI

Istituto di Statistica, Università Cattolica del S. Cuore, Milano
e-mail: gabriele.cantaluppi@mi.unicatt.it

Riassunto: Si stabilisce una condizione sufficiente per l'identificabilità parametrica dei modelli ad equazioni strutturali con variabili latenti, caratterizzati da relazioni di tipo formativo tra variabili esogene manifeste e corrispondenti variabili latenti e da legami di tipo riflessivo tra variabili latenti endogene e corrispondenti indicatori.

Keywords: Structural Equation Models, Formative Relationships, Identification.

1. Introduction

Some approaches to Structural Equation Modeling assume that some of the links between exogenous variables, manifest and latent, are formative, in the sense that manifest variables cause latent variables, whereas the remaining relationships between latent and manifest variables are reflexive: latent variables, both exogenous and endogenous, cause manifest variables, see MacCallum, Browne (1993), Edwards, Bagozzi (2000).

As far as we know, established identification results, see e.g. Bollen (1989), Rigdon (1995), typically concern the case of reflexive relationships. In Bollen, Davis (1994) are given some rules helpful in examining the identification of models with causal indicators. We state – by adapting some results for models with observed variables, see Faliva (1987) – a sufficient condition for parameter identification when all relationships between latent exogenous variables and corresponding manifest indicators are formative.

2. The General Model: Presentation and Identification

We assume the following Structural Equation Model, SEM:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (1)$$

$$\boldsymbol{\xi} = \boldsymbol{\Pi}\mathbf{X} + \boldsymbol{\delta} \quad (2)$$

$$\mathbf{Y} = \boldsymbol{\Lambda}_Y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (3)$$

where: $\boldsymbol{\eta}$ is a random vector of endogenous latent variables; $\boldsymbol{\xi}$ is a random vector of “exogenous” latent variables not defined as functions of other latent variables and directly linked to exogenous manifest indicators; $\mathbf{X}$, $\mathbf{Y}$ are random vectors respectively of exogenous and endogenous manifest variables; $\boldsymbol{\zeta}$, $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}$ are vectors of random errors; $\mathbf{B}$, $\boldsymbol{\Gamma}$ are matrices of unknown parameters β_{ij} , γ_{ij} ; $\boldsymbol{\Lambda}_Y$, $\boldsymbol{\Pi}$ are diagonal block matri-

ces; each block in Λ_Y , Π corresponds to a vector and contains the set of unknown parameters referred to the manifest variables reflexive and formative indicators of the elements respectively of η and ξ . We assume that the hypotheses on the random variables in the model stated in Bollen (1989), pp. 10-20, are satisfied. We further assume that: $E(\zeta|\mathbf{X}) = \mathbf{0}$, $E(\delta|\mathbf{X}) = \mathbf{0}$, $E(\zeta\zeta') = E(\zeta\zeta'|\mathbf{X}) = \Psi_\zeta$, $E(\delta\delta') = E(\delta\delta'|\mathbf{X}) = \Theta_\delta$, $E(\delta\zeta'|\mathbf{X}) = \mathbf{0}$; these hypotheses on conditional expectations are satisfied in the usual formulation of SEMs, when the involved random variables and $\mathbf{X}$ have joint normal distribution functions and ζ , δ are not correlated with $\mathbf{X}$. We assume that one element for each vector in Λ_Y is equal to 1 and that the covariance matrices Ψ_ζ , Θ_δ and $\Theta_\varepsilon = E(\varepsilon\varepsilon')$ are diagonal. It is possible to reformulate the SEM as follows:

$$\begin{cases} \eta = (\mathbf{I} - \mathbf{B})^{-1} (\Gamma^* \xi^* + \zeta^*) \\ \mathbf{Y} = \Lambda_Y \eta + \varepsilon \\ \mathbf{X} = \xi^*, \end{cases} \quad (4)$$

where we have assumed $\Gamma^* = \Gamma\Pi$ and $\zeta^* = \Gamma\delta + \zeta$. We denote with $\Psi_{\xi^*} \equiv \Sigma_{\mathbf{X}\mathbf{X}}$ and Ψ_{ζ^*} the covariance matrices of ξ^* and ζ^* . We assume that Ψ_{ζ^*} contains no zero elements. The following equation allows to state linear constraints on parameter values:

$$\mathbf{R}_1 \text{vec } \mathbf{B}' + \mathbf{R}_2 \text{vec } \Gamma^{*'} = \mathbf{s}, \quad (5)$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are known matrices, $\mathbf{s}$ is a known vector and vec is the vec operator.

Let θ_1 and θ_2 be two elements of the parameter space. The model (1), (2), (3) is globally identifiable if: “ $\Sigma(\theta_1) = \Sigma(\theta_2)$ if and only if $\theta_1 = \theta_2$ ”, since the basic hypothesis of a SEM, see Bollen (1989), p. 3, is that $\Sigma = \Sigma(\theta)$, where Σ is the population covariance matrix of the manifest variables $\mathbf{Y}$ and $\mathbf{X}$, θ is the vector of the unknown model parameters and $\Sigma(\theta)$ is the covariance matrix written as a function of θ .

Some identification rules for models with only reflexive relationships and considerations about models with causal relationships are in Bollen (1989) and Bollen, Davis (1994). We assume that the necessary condition, stated in the *t-rule*, see Bollen (1989) p. 93, is satisfied. We note that, in the case of formative relationships, the elements of Ψ_{ξ^*} are identified since $\Psi_{\xi^*} \equiv \Sigma_{\mathbf{X}\mathbf{X}}$. By assuming that the *scaling rule* and the *three-indicator rule*, or the *two-indicator rule* if $\Sigma_{\eta\eta} = \text{Var}(\eta)$ has no zero elements, are satisfied, see respectively Bollen, Davis (1994), p. 5, and Bollen (1989), p. 244, we have that the parameters in Λ_Y , Θ_ε and $\text{Var}(\eta)$, $\text{Cov}(\eta, \mathbf{X}) \equiv \Sigma_{\eta\xi^*}$ are identified.

In Cantaluppi (2002) is shown that, under the former hypotheses, the elements of matrices $\mathbf{B}$, Γ^* , Ψ_{ζ^*} are identified if the rank condition $r(\mathbf{R}_1 - \mathbf{R}_2 \{ \mathbf{I} \otimes [\Gamma^{*'}(\mathbf{I} - \mathbf{B}')^{-1}] \}) = m^2$, where m is the dimension of the random vector η , is satisfied.

By assuming that one element of each block vector on the diagonal of Π is equal to one the elements of Γ and Π may be identified. The elements of Θ_δ and Ψ_ζ are identified if the system $\Psi_{\zeta^*} = \Gamma\Theta_\delta\Gamma' + \Psi_\zeta$ has a unique solution; in Cantaluppi (2002) is stated the necessary and sufficient condition in the case Θ_δ and Ψ_ζ are diagonal matrices.

3. Example and Comments

In Diamantopoulos et al. (2001) several examples of latent constructs based on forma-

tive relationships are given, which evidence the relevance of this kind of relationships in the formulation of Structural Equation Models. For instance the socioeconomic status may be expressed as a causal function of education, income, occupation and residence. We consider the identification of the following SEM proposed in Duncan, Haller, Portes (1971) for the «study of the way in which one's peers influence one's decisions» (see Fig. 2 for its path representation and variable legend). The defining equations are:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11}^* \gamma_{12}^* \gamma_{13}^* \gamma_{14}^* & 0 & 0 \\ 0 & 0 & \gamma_{23}^* \gamma_{24}^* \gamma_{25}^* \gamma_{26}^* \end{bmatrix} \begin{bmatrix} \xi_1^* \\ \xi_2^* \\ \vdots \\ \xi_6^* \end{bmatrix} + \begin{bmatrix} \zeta_1^* \\ \zeta_2^* \end{bmatrix} \quad \begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{32} \\ Y_{42} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ 0 & 1 \\ 0 & \lambda_{42} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{32} \\ \varepsilon_{42} \end{bmatrix}.$$

Variables $\xi_1^*, \dots, \xi_6^*, Y_{11}, \dots, Y_{42}$ are manifest since expressed directly by respondents. As in Jöreskog (1977), pp. 281-282, we do not impose any constraint on the covariance matrix of ζ^*, Ψ_{ζ^*} , while matrix Θ_{ε} is assumed diagonal. We observe that the model is formulated according to relationships (4), say matrices Γ, Π and random vectors δ, ζ do not appear in this model specification. If we further impose, according to the assumptions stated in §2, that $E(\zeta^*|\xi^*) = \mathbf{0}$, and $E(\zeta^* \zeta^{*'}|\xi^*) = E(\zeta^* \zeta^{*'})$ we may check the identification of the model parameters.

The elements constrained to zero are $\beta_{11}, \beta_{22}, \gamma_{15}^*, \gamma_{16}^*, \gamma_{21}^*, \gamma_{22}^*$, then, by (5), we have:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_{11}^* \\ \gamma_{12}^* \\ \vdots \\ \gamma_{16}^* \\ \gamma_{21}^* \\ \gamma_{22}^* \\ \vdots \\ \gamma_{26}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The parameters in $\Lambda_Y, \Theta_{\varepsilon}, \Sigma_{\eta\eta}$ and $\Sigma_{\eta\xi^*}$ are identified by applying the *scaling rule* and the *two-indicator rule*, since $\sigma_{\eta_1\eta_2} \neq 0$. The parameters in $\mathbf{B}, \Gamma^*, \Psi_{\zeta^*}$ are identified since the following matrix has full column rank

$$\mathbf{R}_1 - \mathbf{R}_2 \{ \mathbf{I} \otimes [\Gamma^{*'}(\mathbf{I} - \mathbf{B}')^{-1}] \} = \frac{1}{1 - \beta_{21}\beta_{12}} \begin{bmatrix} 1 - \beta_{21}\beta_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 - \beta_{21}\beta_{12} \\ -\beta_{12}\gamma_{25}^* & -\gamma_{25}^* & 0 & 0 \\ -\beta_{12}\gamma_{26}^* & -\gamma_{26}^* & 0 & 0 \\ 0 & 0 & -\gamma_{11}^* & -\beta_{21}\gamma_{11}^* \\ 0 & 0 & -\gamma_{12}^* & -\beta_{21}\gamma_{12}^* \end{bmatrix}.$$

In Bollen (1989), p. 98-103, are presented the rank and order conditions for the identification of the parameters present in structural relationships, these conditions «help to determine the identification of one equation at a time». The results presented here, which are based on Faliva (1987), pp. 476-480, ensure the global identification of the model, if the assumptions on conditional expectations, see §2, are satisfied.

With regard to SEMs with latent variables it is assumed that (see Bollen, 1989, p. 1) «if

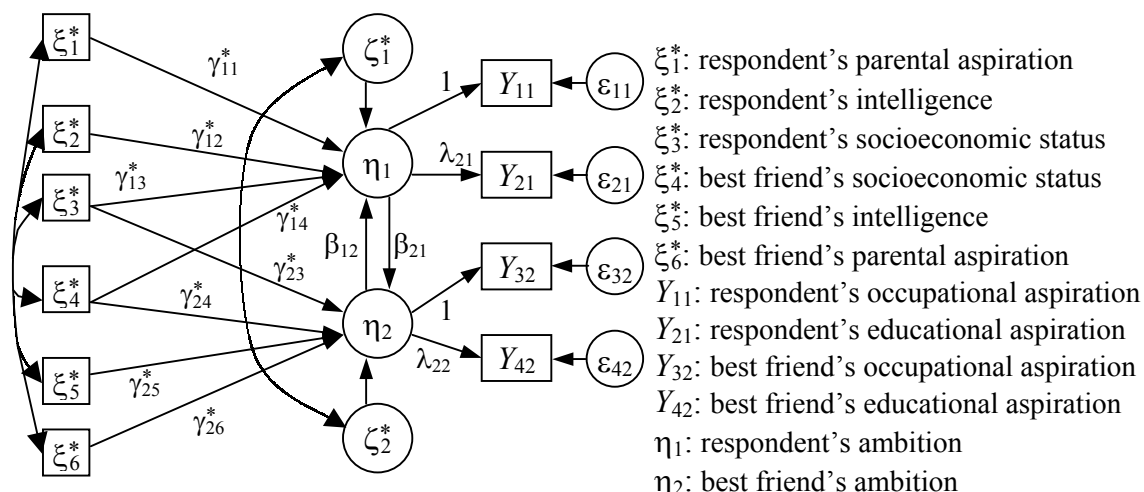


Figure 2 Path representation of model in Jöreskog (1977), p. 279. Curved double arrowed arcs denote that the connected random variables are correlated.

the model were correct and if we knew the parameters, the covariance matrix would be exactly reproduced as a function of a set of parameters», $\Sigma = \Sigma(\theta)$. This peculiarity allows to express the condition for identifiability $r(\mathbf{R}_1 - \mathbf{R}_2 \{ \mathbf{I} \otimes [\mathbf{\Gamma}^*(\mathbf{I} - \mathbf{B}')^{-1}] \}) = m^2$ as a function of $\mathbf{\Gamma}^*$ and $\mathbf{B}$. This condition may be interpreted as a coherence criterion of the structural relationship formulation, (1)-(2), with the parameter constraints (5) and the joint moment structure of the random variables present in the model.

References

- Bollen K.A. (1989) *Structural Equations with Latent Variables*, Wiley, NY.
- Bollen K.A., Davis W.R. (1994) "Causal Indicator Models: Identification, Estimation and Testing", University of North Carolina at Chapel Hill.
- Cantaluppi G. (2002), *The problem of parameter identification of structural equation models with both formative and reflexive relationships: some theoretical results*, Serie E.P., N° 108, Istituto di Statistica, Università Cattolica del S. Cuore, Milano.
- Diamantopoulos A., Winklhofer H.M. (2001) "Index Construction with Formative Indicators: An Alternative to Scale Development", *Journal of Marketing Research*, 38, 269-277.
- Duncan O.D., Haller A.O., Portes, A. (1971) "Peer influence on aspirations, a reinterpretation", in: *Causal Models in the Social Sciences*, Blalock H.M. Jr. (Eds.), Aldine-Atherton, Inc.
- Edwards J.R., Bagozzi R.P. (2000) "On the nature and direction of relationships between constructs and measures", *Psychological Methods*, 5, 155-174.
- Faliva M. (1987) *Econometria: Principi e Metodi*, Utet, Torino.
- Jöreskog K.G. (1977) "Structural Equation Models in the Social Sciences: Specification, Estimation and Testing", in: *Applications of Statistics*, Krishnaiah P.R. (Eds.), North-Holland Publishing Company, Amsterdam.
- MacCallum R.C., Browne M.W. (1993) "The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues", *Psychological Bulletin*, 114, 533-541.
- Rigdon E.E. (1995) "A Necessary and Sufficient Identification Rule for Structural Models Estimated in Practice", *Multivariate Behavioral Research*, 30, 359-383.