Published in final edited form as:

J Math Sociol. 2010 ; 34(2): 136–145. doi:10.1080/00222500903221571.

A Note on Algebraic Solutions to Identification

Kenneth A. Bollen and **Shawn Bauldry**

University of North Carolina at Chapel Hill

Abstract

Algebraic methods to establish the identification of structural equation models remains a viable option. However, sometimes it is unclear whether the algebraic solution establishes identification. One example is when there is more than one way to solve for the parameter, but one way leads to a single value and a second way leads to a function with more than one value. This note proves that one explicit and unique solution is sufficient for model identification even when other explicit solutions permit more than one solution. The results are illustrated with an example. The results are useful to attempts to use algebraic means to address model identification.

Keywords

identification; structural equation models; nonlinear simultaneous equations

1 Introduction

Model identification refers to whether it is possible to find unique values of all model parameters from the population moments of the observed variables. Typically, the population moments refer to the variances, covariances, and means of the observed variables, though higher-order moments are sometimes used (e.g., Bentler, 1983). Algebraic solutions are the oldest approach to identification dating back at least to the work of Sewall Wright (1921). Its basis lies in writing each variance, covariance, and mean of the observed variables as a function of the parameters of the model. Then each model parameter is solved for as a function of one or more of these moments of the observed variables. As Long (1983, page 44) notes:[1] “In general, the most effective way to demonstrate that a model is identified is to show that through algebraic manipulations of the model’s covariance equations each of the parameters can be solved in terms of the population variances and covariances of the observed variables. This is a necessary and sufficient condition of identification.”

Though a variety of rules of identification have emerged from the econometric (e.g., Fisher, 1966) and the latent variable literatures (e.g., Bollen, 1989, 238-47, 326-32; Davis, 1993), these have not eliminated the need to turn to algebraic methods of identification. First these rules do not cover all models. Second, common empirical checks of identification are based on Wald’s Rank Rule (Wald, 1950) or on checking the singularity of the information matrix (Rothenberg, 1971) and these check *local* not *global* identification. Furthermore these local identification checks are based on sample estimates. Due to the lack of rules for all situations and to the limits of local identification, algebraic solutions remain an important approach to establishing the identification of a model or parts of a model where identification is uncertain.[2]

Kenneth A. Bollen, CB 3210 Hamilton Hall, Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3210, (919) 962-7501, bollen@unc.edu; Shawn Bauldry, CB 3210 Hamilton Hall, Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3210, sbauldry@email.unc.edu.

[1] Long (1983) only mentions the variances and covariances of the observed variables. In some models, the means also can play a role.

Ambiguities in the algebraic approach, however, arise when there are multiple ways of solving for a parameter using different moments of the observed variables, as is typically the case with overidentified models. In such situations, it is possible for one solution to yield a single set of parameter values while another solution permits two or more values for at least some of the parameters (e.g., this may arise with solutions involving square roots). In this note we prove that obtaining at least one solution that yields unique parameter values for each parameter is sufficient to establish the global identification of the model. This is important to know in that a researcher solving identification via algebraic means might not know whether a parameter or model is identified if he comes across two or more solutions for the same parameter where at least one of the solutions permits the parameter to take two or more distinct values. We have encountered this problem in experiments with Computer Algebra Systems (CAS) applied to determining the identification of complex structural equation models (SEMs). Indeed, the proofs and this paper grew out of our attempts to determine what to do when faced with this situation and our failure to find any answers to this question in the literature on model identification. However, the result might also be useful in other situations where researchers use algebraic means to solve for parameters when there are more equations than there are parameters.

Our note proceeds as follows. First, we review the identification of SEMs in general terms. Second, we examine four cases involving different types of algebraic solutions for model parameters and provide our proof that obtaining one solution with unique parameter values establishes identification. We conclude with an illustrative model in which we use a CAS algorithm and employ our result to determine model identification. We focus only on the use of the variances, covariances, and means of observed variables and using them to identify model parameters, though our results on the conditions for unique solutions would generalize to the examination of higher-order moments.[3]

2 Algebraic Solutions

Suppose that we have

$$\sigma=\mathbf{F}\,(\theta) \tag{1}$$

where $\boldsymbol{\sigma}$ is a vector of variances, covariances, or means of observed random variables, $\boldsymbol{\theta}$ is a vector of model parameters, and $\mathbf{F}(\boldsymbol{\theta})$ is a vector of functions of $\boldsymbol{\theta}$. The $\mathbf{F}(\boldsymbol{\theta})$ takes different forms depending on the specific SEM. Considering the covariance matrix of observed variables in confirmatory factor analysis, for example, the vector of implied covariances, variances and means is $\mathbf{F}\,(\theta)=\begin{pmatrix} vech\left[\Lambda\Phi\Lambda'+\Theta\right] \\ \alpha+\Lambda\mu_{\xi} \end{pmatrix}$ where $\boldsymbol{\Lambda}$ is the matrix of factor loadings, $\boldsymbol{\Phi}$ is the covariance matrix of the factors, $\boldsymbol{\Theta}$ is the covariance matrix of the unique factors, *vech* is a matrix operation that stacks all of the nonredundant elements in $\boldsymbol{\Lambda}\ \boldsymbol{\Phi}\ \boldsymbol{\Lambda}' + \boldsymbol{\Theta}$ into a vector**,** α is the vector of intercepts, and μ_{ζ} is the vector of means of ζ. $\mathbf{F}(\boldsymbol{\theta})$ is the model implied moment vector. In general, we assume that the variances, covariances, and means of all variables exist, that all variances in $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$ are nonnegative and any implicit or explicit correlations of any two variables are less than one in absolute value. As mentioned above, we only make use of the means, variances, and covariances of the observed variables in identifying the model parameters.

[2] Algebraic solutions can also be useful in formulating new rules of identification (e.g., O'Brien (1994).

[3] Higher-order moments in some situations provide additional information that would aid model identification. However, these higher-order moments are rarely used and we confine ourselves to the typical situation where a researcher only employs the variances and covariances, and sometimes the means of the observed variables to aid model identification.

To define *global* identification, consider two vectors $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$, each of which contains numeric values for the unknown parameters in $\boldsymbol{\theta}$. For each vector we can form the implied covariances and variances, say $\boldsymbol{\sigma}_a = \mathbf{F}(\boldsymbol{\theta}_a)$ and $\boldsymbol{\sigma}_b = \mathbf{F}(\boldsymbol{\theta}_b)$, for each set of numeric values. If the model is identified, all $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ solutions where $\mathbf{F}(\boldsymbol{\theta}_a) = \mathbf{F}(\boldsymbol{\theta}_b)$ must have $\boldsymbol{\theta}_a = \boldsymbol{\theta}_b$. If a pair of vectors $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ exists such that $\mathbf{F}(\boldsymbol{\theta}_a) = \mathbf{F}(\boldsymbol{\theta}_b)$ and $\boldsymbol{\theta}_a \neq \boldsymbol{\theta}_b$, then $\boldsymbol{\theta}$ is not *globally* identified. *Local* identification is a weaker concept of uniqueness. A parameter vector $\boldsymbol{\theta}$ is locally identified at a point $\boldsymbol{\theta}_a$, if in the neighborhood of $\boldsymbol{\theta}_a$ there is no vector $\boldsymbol{\theta}_b$ for which $\mathbf{F}(\boldsymbol{\theta}_a) = \mathbf{F}(\boldsymbol{\theta}_b)$ unless $\boldsymbol{\theta}_a = \boldsymbol{\theta}_b$ (Bollen, 1989, page 248).

Suppose that we form subsets of the elements of $\boldsymbol{\sigma}$ such that each subset vector, $\boldsymbol{\sigma}_j$, has a dimension equal to the number of parameters in $\boldsymbol{\theta}$ and each element of $\boldsymbol{\theta}$ appears at least once in the $\mathbf{F}_j(\boldsymbol{\theta})$ that corresponds to $\boldsymbol{\sigma}_j$ where $\mathbf{F}_j(\boldsymbol{\theta})$ refers to the subvector of $\mathbf{F}(\boldsymbol{\theta})$ that corresponds to $\boldsymbol{\sigma}_j$. This leads to

$$\begin{aligned} \sigma_1 &= \mathbf{F}_1(\theta) \\ \sigma_2 &= \mathbf{F}_2(\theta) \\ &\vdots \\ \sigma_J &= \mathbf{F}_J(\theta) \end{aligned} \tag{2}$$

Given that equation (1) is true, each equation in (2) must be true since they are just subsets of the original true equation. Suppose that K of these equations have explicit solutions for $\boldsymbol{\theta}$ that are functions of elements of $\boldsymbol{\sigma}$. We write these solutions as

$$\begin{aligned} \theta &= \mathbf{G}_1(\sigma_1) \\ \theta &= \mathbf{G}_2(\sigma_2) \\ &\vdots \\ \theta &= \mathbf{G}_K(\sigma_K) \end{aligned} \tag{3}$$

where $\mathbf{G}_k(\boldsymbol{\sigma}_k)$ is a function of $\boldsymbol{\sigma}_k$ that is an explicit solution for $\boldsymbol{\theta}$ and where $\mathbf{G}_k(\boldsymbol{\sigma}_k)$, $k = 1, 2, 3, \cdots, K$ represent different functions. Further assume that if there is no superscript (l) that the $\mathbf{G}_k(\boldsymbol{\sigma}_k)$ function is explicit *and* unique in that it leads to only one solution. If we have, say $\mathbf{G}_k^{(1)}(\sigma_k)$, $\mathbf{G}_k^{(2)}(\sigma_k)\ldots$, $\mathbf{G}_k^{(L)}(\sigma_k)$, then there are L explicit solutions for the given function. For instance, if the explicit solution involves a square root, then we would have the positive and negative square root solutions with $L = 2$.

We distinguish four cases:

Case 1

Only one explicit solution exists, and it is unique. Without loss of generality, let this solution be given by $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$. In this case the model would be identified since $\mathbf{G}_1(\boldsymbol{\sigma}_1)$ is the only solution and results in a single solution. This situation is generally encountered when the number of parameters equals the number of variances, covariances, and means of the observed variables. However, having the same number of parameters and number of moments does not guarantee a solution nor that it will be a unique solution.

Case 2

The only explicit solution is $\theta^{(l)}=\mathbf{G}_1^{(l)}(\sigma)$ and this leads to, say, L possible values of $\mathbf{G}_1^{(l)}(\sigma)$ of $\theta^{(1)}=\mathbf{G}_1^{(1)}(\sigma_1)$, $\theta^{(2)}=\mathbf{G}_1^{(2)}(\sigma_1),\ldots,\theta^{(L)}=\mathbf{G}_1^{(L)}(\sigma_1)$ where $\mathbf{G}_1^{(t)}(\sigma_1) \neq \mathbf{G}_1^{(u)}(\sigma_1)$ which implies that $\boldsymbol{\theta}^{(t)} \neq \boldsymbol{\theta}^{(u)}$ for all $t \neq u$. Given that we have L explicit solutions, can we tell whether $\boldsymbol{\theta}$ is identified?

Consider global identification first. The algebraic solutions of $\theta^{(1)}=\mathbf{G}_1^{(1)}(\sigma_1)$, $\theta^{(2)}=\mathbf{G}_1^{(2)}(\sigma_1), \ldots, \theta^{(L)}=\mathbf{G}_1^{(L)}(\sigma_1)$ derive from the original equation of $\boldsymbol{\sigma} = \mathbf{F}(\boldsymbol{\theta})$ which corresponds to the model. This means that if any of these solutions, say $\boldsymbol{\theta}^{(s)}$, is substituted in for $\boldsymbol{\theta}$ in $\boldsymbol{\sigma} = \mathbf{F}(\boldsymbol{\theta})$, then $\mathbf{F}(\boldsymbol{\theta}^{(s)})$ will equal $\boldsymbol{\sigma}$. Since $\boldsymbol{\theta}^{(t)} \neq \boldsymbol{\theta}^{(u)}$, the model parameters cannot be globally identified. So if we have Case 2, the model is not globally identified. We can check local identification with the Wald's Rank Rule. Form $\frac{\partial F(\theta)}{\partial \theta}$ and check whether its rank equals the number of independent parameters where we assume the differentiability of F($\boldsymbol{\theta}$) with respect to $\boldsymbol{\theta}$.[4] If it does, then the model is locally identified. If its rank is less, then it is not.

Case 3

The $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is a unique, explicit solution and we also have $\theta=\mathbf{G}_2^{(1)}(\sigma_2)$ and $\theta=\mathbf{G}_2^{(2)}(\sigma_2)$ where there are two explicit solutions associated with $\mathbf{G}_2^{(l)}(\sigma_2)$. Given one unique explicit solution, is this sufficient to identify $\boldsymbol{\theta}$?

As we stated above, all equations in (2) are true since they are just subsets of the true equation in (1). The equations in (3) derive from the equations in (2) and hence $\boldsymbol{\sigma}_1 = \mathbf{F}_1(\boldsymbol{\theta})$ and $\boldsymbol{\sigma}_2 = \mathbf{F}_2(\boldsymbol{\theta})$ must both be true and the value(s) of $\boldsymbol{\theta}$ must satisfy both equations.

There are several possibilities to consider:

1. $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is true, $\theta=\mathbf{G}_2^{(l)}(\sigma_2)\ (l=1,2)$ is true
2. $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is false, $\theta=\mathbf{G}_2^{(l)}(\sigma_2)\ (l=1,2)$ is false
3. $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is false, $\theta=\mathbf{G}_2^{(l)}(\sigma_2)\ (l=1,2)$ is true
4. $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is true, $\theta=\mathbf{G}_2^{(l)}(\sigma_2)\ (l=1,2)$ is false

Consider the first possibility, that $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma})$ and $\theta=\mathbf{G}_2^{(l)}(\sigma_2)\ (l=1,2)$ are true. Using proof by contradiction, this implies that

$$\mathbf{G}_1(\sigma_1)=\mathbf{G}_2^{(l)}(\sigma_2)$$

which cannot be true since $\mathbf{G}_1(\boldsymbol{\sigma}_1)$ is a single value solution and it cannot equal two different values, $\mathbf{G}_2^{(1)}(\sigma_2)$ and $\mathbf{G}_2^{(2)}(\sigma_2)$. Therefore, we dismiss the first possibility as invalid.

The second possibility that $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ and $\theta=\mathbf{G}_2^{(l)}(\sigma_2)\ (l=1,2)$ are both false we also rule out by proof of contradiction. The solution $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is implied by $\boldsymbol{\sigma}_1 = \mathbf{F}_1(\boldsymbol{\theta})$. If $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is false, then $\boldsymbol{\sigma}_1 = \mathbf{F}_1(\boldsymbol{\theta})$ is false. But this contradicts our given that $\boldsymbol{\sigma} = \mathbf{F}(\boldsymbol{\theta})$ and hence $\boldsymbol{\sigma}_1 = \mathbf{F}_1(\boldsymbol{\theta})$ is true. Therefore, possibility 2. cannot be true since $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ must be true. By the same logic, we can rule out the third possibility since it too assumes that $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is false and we just ruled that out.

[4] A reviewer points out that if $\boldsymbol{\theta}$ is discrete, these derivatives would not exist, but that there are cases in which a local identification of $\boldsymbol{\theta}$ is well-defined (e.g., when $\boldsymbol{\theta}$ is unidimensional and its states admit a total order).

By process of elimination, possibility four must be true (i.e., $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ is true, $\theta = \mathbf{G}_2^{(l)}(\sigma_2)\,(l=1,2)$ is false). The statement that $\theta = \mathbf{G}_2^{(l)}(\sigma_2)\,(l=1,2)$ is false requires closer examination since this contains two possible values. This could be false is one of three ways:

1. $\theta = \mathbf{G}_2^{(1)}(\sigma_1)$ is false, $\theta = \mathbf{G}_2^{(2)}(\sigma_2)\,(l=1,2)$ is false
2. $\theta = \mathbf{G}_2^{(1)}(\sigma_1)$ is false, $\theta = \mathbf{G}_2^{(2)}(\sigma_2)\,(l=1,2)$ is true
3. $\theta = \mathbf{G}_2^{(1)}(\sigma_1)$ is true, $\theta = \mathbf{G}_2^{(2)}(\sigma_2)\,(l=1,2)$ is false

Using proof by contradiction, we can rule out one since if both solutions are false, this implies that $\boldsymbol{\sigma}_2 = \mathbf{F}_2(\boldsymbol{\theta})$ is false, but we know that the latter is true. Therefore we are left with possibility 2. or 3. Which of these two is true is determined by whether $\mathbf{G}_1(\sigma_1) = \mathbf{G}_2^{(1)}(\sigma_2)$ or $\mathbf{G}_1(\sigma_1) = \mathbf{G}_2^{(2)}(\sigma_2)$. As shown above, both of these equalities cannot hold. However, one of them must hold and that determines which of the two solutions, $\mathbf{G}_2^{(1)}(\sigma_2)$ or $\mathbf{G}_2^{(2)}(\sigma_2)$ is true. This in turn shows that having one function that leads to a single unique value is sufficient to establish a single value for $\boldsymbol{\theta}$ even if a second function leads to a solution with two values.

Case 4

The preceding proof considers only two solution functions (i.e., $\boldsymbol{\theta} = \mathbf{G}_1(\boldsymbol{\sigma}_1)$ and $\theta = \mathbf{G}_2^{(l)}(\sigma_2)\,(l=1,2)$). What happens if there are additional functions that have two value solutions? It is easy to show that the choice of the second function is arbitrary and that the above proof holds for any two value solution chosen in conjunction with a single value solution.

What happens if there is a second function that takes more than two values? Besides adding solution values to the second function, the above proof would remain essentially the same.

Therefore, having a unique function with a single solution for $\boldsymbol{\theta}$ is sufficient to establish identification even if there are other unique functions that have multiple solutions.

Note that our discussion focuses on a sufficient, but not necessary condition for identification. It is possible to have a situation with several solution functions, each of which has multiple solutions, but to still have the parameter identified (e.g., only one solution is consistent across these solution functions).

3 Illustration

We now turn to an illustration of the utility of our result in assessing the identification of a SEM shown in Figure 1. Our illustrative model contains one exogenous and two endogenous observed variables. We specify a recriprocal relationship between the two endogenous variables, but constrain the parameter estimates for the two paths to be equal.

This model can also be expressed by the following system of equations:

$$\begin{aligned} y_1 &= \beta y_2 + \zeta_1 \\ y_2 &= \beta y_1 + \gamma x_1 + \zeta_2. \end{aligned}$$

In this model we have six variances and covariances and five model parameters. We define $\sigma_{11} = V(y_1)$, $\sigma_{22} = V(y_2)$, and $\sigma_{33} = V(x_1)$ and the various covariances represented by the appropriate subscripts. This model leads to the following vector of functions, $\mathbf{F}(\boldsymbol{\theta})$:

$$\sigma=\mathbf{F}(\theta)$$

$$\begin{pmatrix}\sigma_{11}\\ \sigma_{21}\\ \sigma_{22}\\ \sigma_{31}\\ \sigma_{32}\\ \sigma_{33}\end{pmatrix}=\begin{pmatrix}\left[\beta^2V(\zeta_2)+\beta^2\gamma^2\boldsymbol{V}(x_1)+\boldsymbol{V}(\zeta_1)\right]/\left[\beta^2-1\right]^2\\ \left[\beta\boldsymbol{V}(\zeta_1)+\beta\boldsymbol{V}(\zeta_2)+\beta\gamma^2\boldsymbol{V}(x_1)\right]/\left[\beta^2-1\right]^2\\ \left[\beta^2V(\zeta_1)+\beta^2\gamma^2\boldsymbol{V}(x_1)+\boldsymbol{V}(\zeta_1)\right]/\left[\beta^2-1\right]^2\\ \left[-\beta\gamma\boldsymbol{V}(x_1)\right]/\left[\beta^2-1\right]^2\\ \left[-\gamma\boldsymbol{V}(x_1)\right]/\left[\beta^2-1\right]^2\\ \boldsymbol{V}(x_1)\end{pmatrix}.$$

For this system of equations, if we choose a subset of parameters that includes the equation relating the covariance between the two endogenous variables to the model parameters (σ_{21}), then we obtain a solution for some of the model parameters involving a square root. For example, if we choose the subset (σ_{11}, σ_{22}, σ_{33}, σ_{21}, σ_{31}) we obtain the following two solutions for β:[5]

$$\beta=\frac{\sigma_{11}+\sigma_{22}+\left(\sigma_{11}^2+2\sigma_{11}\sigma_{22}+\sigma_{22}^2-4\sigma_{21}^2\right)^{1/2}}{2\sigma_{21}},$$
$$\beta=\frac{\sigma_{11}+\sigma_{22}-\left(\sigma_{11}^2+2\sigma_{11}\sigma_{22}+\sigma_{22}^2-4\sigma_{21}^2\right)^{1/2}}{2\sigma_{21}}.$$

If instead we choose a subset of equations that does not include the equation involving the covariance between the two endogenous variables (e.g., (σ_{11}, σ_{22}, σ_{33}, σ_{31}, σ_{32})), then we find a unique solution for each parameter. Using this subset of equations, we obtain

$$\beta=\frac{\sigma_{31}}{\sigma_{32}}.$$

As established above, this is sufficient to determine that the model is globally identified.

As an additional check, our result implies that in any given numerical setting one of the solutions for β that we obtained from the first subset should equal the solution obtained from the second subset. We demonstrate this is the case by generating a covariance matrix based on arbitrary numerical values for each of the model parameters, and then checking which of the first two solutions for β is consistent with the second solution. If we let $\beta = 0.5$**,** $\gamma = 2$**,** $\boldsymbol{\phi} = 2$**,** $\psi_{11} = 2$**, and** $\psi_{22} = 3$, then we obtain the following covariance matrix (rounded to two digits):

$$\begin{bmatrix}8.44 & & \\ 11.56 & 20.44 & \\ 2.67 & 5.33 & 2.00\end{bmatrix}.$$

Substituting the covariances into the two solutions for β **we find:**

$$\beta=\frac{8.44+20.44+\left(8.44^2+2(8.44)(20.44)+20.44^2-4\left(11.56^2\right)\right)^{1/2}}{2(11.56)}=2.00,$$
$$\beta=\frac{8.44+20.44-\left(8.44^2+2(8.44)(20.44)+20.44^2-4\left(11.56^2\right)\right)^{1/2}}{2(11.56)}=0.50.$$

[5] We do not report the entire vectors $\mathbf{G}_1^{(1)}(\sigma)$ or $\mathbf{G}_1^{(2)}(\sigma)$ due to considerations of space.

In this case, the second solution for β matches the unique solution from the other subset of equations, $\beta=\frac{2.67}{5.33}=0.50$ (and both, of course, match the value we chose in generating the covariance matrix). Furthermore, in order for this model to be globally identified it must be true that we obtain a different implied covariance matrix when we substitute the solution $\beta = 2$ (along with the solutions for the other elements of β) into the full set of equations than the one given above. This substitution generates the following implied covariance matrix:

$$\begin{bmatrix} 8.40 & & \\ 11.50 & 20.36 & \\ 2.67 & 1.33 & 2.00 \end{bmatrix},$$

with the clearest difference being in the σ_{32} element.

4 Conclusion

Algebraic solutions to establish model identification was an early means of establishing model identification and it remains important in both establishing new rules of identification and in covering situations that do not fall under existing rules. However, an ambiguous situation emerges when there are two or more explicit, distinct solutions for a parameter and when one or more of these solutions permits multiple values such as when the solution involves a square root. This note establishes that if one explicit and unique solution is found for the model parameters, then this is sufficient to establish model identification even when there are other explicit solutions that permit more than one solution to the equation. This result is of particular significance when a CAS is employed to establish the identification of models algebraically that do not conform to the known rules for identification.

5 References

Bentler PM. Simultaneous equation systems as moment structure models. Journal of Econometrics 1983;22:13–42.

Bollen, KA. Structural equations with latent variables. Wiley; New York: 1989.

Fisher, FM. The identification problem in econometrics. McGraw Hill; New York: 1966.

Long, JS. Confirmatory factor analysis: A preface to LISREL. Sage University Press; Beverly Hills: 1983.

O’Brien RM. Identification of simple measurement models with multiple latent variables and correlated errors. Sociological Methodology 1994;24:137–170.

Rothenberg TJ. Identification in parametric models. Econometrica 1971;39:577–591.

Wald, A. A note on the identification of econometric relations. In: Koopmans, TC., editor. Statistical Inference in Dynamic Economic Models. Wiley; New York: 1950. p. 238-244.

Wright S. Correlation and causation. Journal of Agricultural Research 1921;20:557–85.

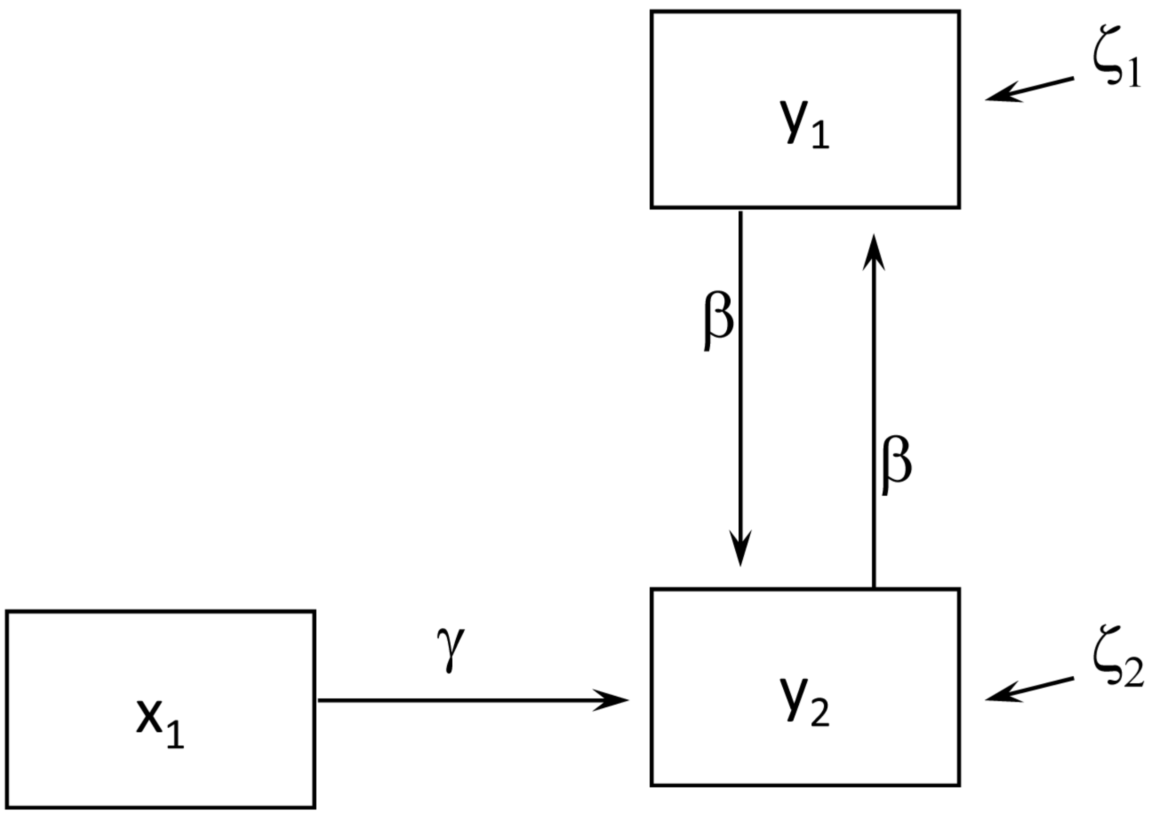

Figure 1. Model to Demonstrate Identification Result