NOTE ON THE TRANSFORMATION OF VARIABLES IN SIMULTANEOUS EQUATIONS SYSTEMS

GERHARD TINTNER AND GOPAL KADEKODI

Department of Economics, University of Southern California, Los Angeles, California

The relationship between the models of classical physics and of methematical economics has been recently emphasized (Georgescu-Roegen 1966, Fels and Tintner, 1966, Tintner 1966, Tintner 1968). But, in classical (and also in stochastic modern) physics the form of the functional relationship is frequently implied by the theory. The theory yields the fundamental relationships (typically partial differential equations) and from this the form of the functional relationship can be deduced. Unfortunately, this is not true in theoretical economics : Nothing but very general properties of the theoretical relationships between the variables can be asserted.

It is the great merit of a recent article by Zarembka (1968) that he has, for the first time, faced this problem. He pointed out that this is in a sense a continuation of the work by Arrow, Chenery, Minhas and Solow (1961), who generalized the classical Cobb-Doublas function. But, the fundamental statistical work is due to Box and Cox (1964), who introduced transformations especially for use in biological statistics.

In this paper we consider transformations of the form :

$$x^{(\lambda)} = \frac{x^{\lambda} - 1}{\lambda} \qquad \ldots(1)$$

with the definition of $x^{(0)}$ as a limiting value of (1) as $\lambda \to 0$.

$$x^{(0)} = \log_e x \qquad \ldots(2)$$

which can be justified because :

$$\lim_{\lambda \to 0} \frac{x^{\lambda} - 1}{\lambda} = \log_e x \qquad \ldots(3)$$

There is a connection between these tranformations and early work of Keynes (1948) on estimation methods.

Consider now a linear system of econometric equations :

$$Ax_t = u_t \qquad t = 1, 2, \ldots . N \qquad \ldots(4)$$

where A is a matrix of order $G.$ $(G+K)$, x_t is a column vector of $G+K$ variables, u_t is a vector of random variables of order G with these properties :

$$Eu_t = 0,\ Eu_t'u_t = \Sigma,\ Eu_t'u_s = 0(t \neq x) \qquad \ldots(5)$$

where Σ is a constant variance-covariance matrix. (Tintner 1952, pp. 156 ff). We also assume at least an approximate normal distribution. Partition now the vector x_t into :

$$x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix} \qquad ...(6)$$

where y_t is the vector of G endogenous variables, z_t the vector of K predetermined variables. With the partition of A as :

$$A = [B \quad C] \qquad ...(7)$$

we have the system :

$$By_t + Cz_t = u_t \qquad ...(8)$$

Consider now a transformation of all variables. With the above notation (1) becomes :

$$B^* y_t^{(\lambda)} + C^* z_t^{(\lambda)} = u^*_t \qquad ...(9)$$

where B^*, C^*, u^*_t are the constant matrices and the vector of random variables resulting from the transformation (1). As a simple case we assume same λ for all the variables, both endogenous and the predetermined. The identification problem is the same as in the linear system (8) (Fisher 1966, p. 129.) Also, estimation problems are not changed, with the exception of the estimation of the parameter λ which has to be estimated numerically. Generalization to different λ for each such variable may be difficult due to the identification problem of non linear model.

We construct the reduced form equations :

$$y_t^{(\lambda)} = P^* z_t^{(\lambda)} + mv^*_t \qquad ...(10)$$

where

$$P^* = -B^{*-1} C^* \qquad ...(11)$$

$$v^*_t = B^{*-1} u^*_t :$$

vector of errors in the reduced form equations.

Assuming v^*_t is distributed as multivariate normal with mean zero, the joint likelihood of the original set of variables can be derived with the transformation (1) ; the joint likelihood of the observations remains a function of λ.

By maximizing (numerically) the log likelihood function : (approximately)

$$L \max(\lambda) = (-N/2) \log_e |\Sigma^*| + (\lambda - 1) \sum_{i=1}^{G} \sum_{t=1}^{N} \log_e Y_{it} + G \cdot N \log_e |\lambda| \text{ (1)} \qquad ...(13)$$

(1) Here the tranformation applied is $x^{(\lambda)} = x$.

where an irrelevant constant has been omitted (Goldberger 1964, p. 211, Box and Cox 1964, p. 215, Zarembka 1968, p. 505). A maximum likelihood estimate of λ can be obtained. This function is maximized by numerical methods. Here Σ^* is the determinant of the variance-covariance martix of the estimated deviations v^*_t :

An approximate (large sample) confidence interval for λ can be determined by the formula :

$$L' \max (\hat{\lambda}) - L \max (\lambda) = 1{\cdot}92 \qquad \ldots(14)$$

for a confidence coefficient of 95% (Zarembka 1968, p. 505), where $\hat{\lambda}$ is the maximum likelihood estimate of λ[2].

Having estimated the reduced form equations (10) with the help of numerically determined parameter λ we may then apply conventional estimation methods (indirect least squares, limited information method (Anderson and Rubin 1949, Hood and Koopmans 1953), two stages least squares (Theil 1961, Basman 1967) in order to obtain an estimate of the system (8). For a justification of these methods (Theil 1961, Tintner 1962, Wold and Jureen 1953, Haavelmo 1944, Klein 1953, 1962, Valavanis 1959, Malinvaud 1964, Johnston 1963, Marschak 1953, Christ 1966, Goldberger 1964, Dhrymes 1970, Fisk 1967, Leser 1966, Walters 1970) from the point of view of modern information theory see : Tintner and Rama Sastry 1969.

The non-linearity of the simultaneous functional relationships with the transformation (1) are generalizations to the extent that they include both log linear and for certain stochastic difference equations exponential and logistic forms.

An empirical case

The use of transformations in the simultaneous equations model is illustrated with a two equation supply and demand functions of American meat industry model. For the original work of this problem see Tintner (1952). The data pertains to 1949-67 (see appendix *A*, for the description of the data). The variables in the model are the same as in Tintner (1952). Consider the supply and demand functions for meat (including poultry and fish). The structural equations (9) are written as :

$$Y_1^{\lambda} = a_{01} + b_{12} Y_2^{\lambda} + C_{11} Z_1^{\lambda} + \epsilon_1 : \text{demand function} \qquad \ldots(15a)$$

$$\;_1^{\lambda} = a_{02} + b_{22} Y_2^{\lambda} + C_{22} Z_2^{\lambda} + \epsilon_2 : \text{supply function} \qquad \ldots(15b)$$

(2) L max is approximately distributed as χ^2.

In these equations Y_1 is meat consumption, Y_2 price of meat, Z_1 disposable income, Z_2 cost of producing meat, Z_3 cost of producing agricultural products. Both equations are just identified. The reduced form equations of this structure are:

$$Y_1^{\lambda} = A_1 + B_1 Z_1^{\lambda} + C_1 Z_1^{\lambda} + u_1 \quad \text{...(16a)}$$

$$Y_2^{\lambda} = A_2 + B_2 Z_1^{\lambda} + C_2 Z_2^{\lambda} + u_2 \quad \text{...(16b)}$$

The reduced form of parameters including λ are estimated as follows: With a sample size of $N=19$ and endogenous variables $G=2$ the approximate joint log likelihood of u_{1t}, u_{2t} is given by

$$L\max(\lambda) = -(N/2)\log \sigma_1^2 - (N/2)\log \hat{\sigma}_2^2 - (N/2)\log\left(1-\hat{\rho}^2\right) + (\lambda-1)\sum_{t=1}^{N}(\log Y_{1t} + \log Y_{2t}) + 2N\log|\gamma| \quad \text{...(17)}$$

where,

$E(u_1)=E(u_2)=0$

Var $(u_1)=\sigma_1^2$, Var $(u_2)=\sigma_2^2$

Cov. $(u_1, u_2)=\rho$

For fixed λ the least squares estimates of the parameters are obtained and hence an estimate of L max (λ). Table 1 shows the numerical relationship of λ and L max. With limited accuracy of the intervals, L max is maximum at $\lambda=0{\cdot}50$. The 95% confidence limits for $\hat{\lambda}$ are $-1{\cdot}7$ and $2{\cdot}5$ respectively. The estimated reduced form equations and hence the structural equations are :

Reduced forms

$$Y_1^{0\cdot5} = 10{\cdot}64863 + \underset{(0{\cdot}0123)}{0{\cdot}12896}\, Z_1^{0\cdot5} - \underset{(0{\cdot}0966)}{0{\cdot}53229}\, Z_2^{0\cdot5} \quad \text{.. (18a)}$$

t-values : 10·4845 5·5103

$R_1^2=0{\cdot}9369$, $\sigma_1^2=0{\cdot}01479$

$$Y_2^{0\cdot5} = 2{\cdot}3696 + \underset{(0{\cdot}0348)}{0{\cdot}11530}\, Z_1^{0\cdot5} + \underset{(0{\cdot}1529)}{0{\cdot}58320}\, Z_2^{0\cdot5} \quad \text{...(18b)}$$

t-values : 3·3132 3·8142

$R_2^2=0{\cdot}7125$, $\sigma_2^2=0{\cdot}03704$

Numbers in parentheses stand for standard errors of the estimates.

Structure

$$Y_1^{0\cdot5} = 12{\cdot}81107 - 0{\cdot}91271\, Z_2^{0\cdot5} + 0{\cdot}23419\, Y_1^{0\cdot5} \; : \text{demand} \quad \text{...(19a)}$$

$$Y_1^{0\cdot5} = 7{\cdot}99868 + 1{\cdot}11847\, Y_2^{0\cdot5} - 1{\cdot}18454\, Z_2^{0\cdot5} \; : \text{supply} \quad \text{...(19b)}$$

The price elasticity of demand estimated around the sample mean is 0·657 and income elasticity of demand is 0·736.

Again, adding another exogenous variable Z_3—the cost of agricultural production to the supply function the structure is :

$$Y_1^{\lambda} = \alpha_1 + \beta_1 Y_2^{\lambda} + \gamma_1 Z_1^{\lambda} + \epsilon_1 : \text{ demand function} \quad ...(20a)$$

$$Y_1^{\lambda} = \alpha_2 + \beta_2 Y_2^{\lambda} + \gamma_2 Z_2^{\lambda} + \delta_2 Z_3^{\lambda} + \epsilon_2 : \text{ supply function} \quad ...(20b)$$

The demand function is overidentified (Tintner 1952). Applying the same method as in just identified model, the reduced form equations are estimated. Table 2 shows the estimated numerical relationship between λ and L max. L max is maximum at 0·65. The reduced forms are :

$$Y_1^{0\cdot65} = 22{\cdot}96354 + \underset{(0\cdot0247)}{0{\cdot}07357}\, Z_1^{0\cdot65} - \underset{(0\cdot140)}{0{\cdot}76808}\, Z_2^{0\cdot65} + \underset{(0\cdot202)}{0{\cdot}15208}\, Z_3^{0\cdot65}, \quad ...(21a)$$

$$R_1^2 = 0{\cdot}9397, \qquad \hat{\sigma}_1^2 = 0{\cdot}12391$$

$$Y_2^{0\cdot65} = 4{\cdot}46651 + \underset{(0\cdot035)}{0{\cdot}012247}\, Z_1^{0\cdot65} + \underset{(0\cdot206)}{0{\cdot}560806}\, Z_2^{0\cdot65} + \underset{(0\cdot293)}{0{\cdot}49144}\, Z_3^{0\cdot65}$$

$$R_1^2 = 0{\cdot}8343, \qquad \hat{\sigma}_1^2 = 0{\cdot}2083 \quad ...(21b)$$

The two stage estimate of the demand function is :

$$Y_1^{0\cdot65} = 25{\cdot}8535 - \underset{(0\cdot247)}{0{\cdot}81369}\, Y_2^{0\cdot65} + \underset{(0\cdot017)}{0{\cdot}15501}\, Z_1^{0\cdot65}$$

$$R^2 = 0{\cdot}9415, \qquad \hat{\sigma}^2 = 0{\cdot}1722 \quad ...(22)$$

Again, the price elasticity of demand at the sample means is 0·531.

However, as can be seen from both table 1 and table 2, the L max curve is relatively flat in a large range. This, in addition to the level of computer accuracy, adds to some amount of uncertainty to the results. However, in both models λ is effectively different from unity, even though the 95% confidence intervals using (14) do include unity.

Summary

The use of transformations in the simultaneous equations model has been discussed in this paper with special reference to econometric problems.

TABLE 1

Just identified model

λ	*L max*
−3·3	−54·443
−3·1	−53·925
−2·9	−53·423
−2·7	−52·939
−2·5	−52·474
−2·3	−52·030
−2·1	−51·608
−1·9	−51·209
−1·7	−50·834
−1·5	−50·486
−1·3	−50·165
−1·1	−49·874
−0·9	−49·613
−0·7	−49·384
−0·5	−49·188
−0·3	−49·026
−0·1	−48·899
0·1	−48·809
0·3	−48·755
0·35	−48·747
0·40	−48·747
0·45	−48·739
0·50	−48·738......Maximum
0·55	−48·739
0·60	−48·743
0·65	−48·749
0·70	−48·757
0·75	−48·768
0·80	−48·780
0·85	−48·795
0·90	−48·813
1·00	−48·854
1·10	−48·904
1·2	−48·963
1·3	−49·030
1·5	−49·189
1·7	−49·380
1·9	−49·600
2·1	−49·849

L max is maximum at λ=0·50

TABLE 2

Overidentified model

	L max	
0·25	−48·16676	
0·30	−48·15071	
0·35	−48·13671	
0·40	−48·12493	
0·45	−48·11523	
0·50	−48·10772	
0·55	−48·10235	
0·60	−48·09918	
0·65	−48·09814	......Maximum
0·70	−48·09927	
0·75	−48·10255	
0·80	−48·10801	
0·85	−48·13564	
0·90	−48·12541	
0·95	−48·13732	
1·00	−48·16758	
1·05	−48·16764	
1·10	−48·18597	
1·15	−48·20640	
1·20	−48·22897	
1·25	−48·25360	
1·30	−48·28034	
1·35	−48·30912	

L max is maximum at $\lambda = 0{\cdot}65$

Table 3

	Y_1	Y_2	Z_1	Z_2	Z_3
1949					
	173·80000	91·10001	1554·00000	22·84000	86·39999
	176·69990	95·10001	1646·00000	26·84000	92·39999
	172·10000	106·30000	1674·00000	26·25000	98·10001
	179·89990	105·30000	1901·00000	23·05000	97·60001
	187·10000	99·60001	1741·00000	22·39000	91·10001
	186·89990	97·89999	1727·00000	18·28999	92·00000
1955					
	191·80000	92·10001	1795·00000	16·07001	93·60001
	197·80000	88·00000	1836·00000	17·22000	94·30000
	191·30000	95·39999	1846·00000	18·82001	94·80000
	187·30000	104·39990	1831·00000	20·64999	100·00000
	195·89990	100·39990	1879·00000	18·99001	102·60000
	194·60000	99·10001	1883·00000	18·91000	102·00000
1961					
	197·19990	99·30000	1909·00000	19·69000	103·89990
	198·10000	100·69990	1969·00000	19·03000	108·39990
	203·19990	100·19990	2009·00000	17·48000	111·00000
	207·80000	98·60001	2123·00000	16·64000	108·80000
	203·80000	105·10000	2232·00000	20·81000	112·30000
1966					
	208·80000	114·10000	2317·00000	19·97000	117·89990
	217·60000	111·19990	2401·00000	19·12000	119·50000

REFERENCES

Anderson, T. W. and H. Rubin (1949) : Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. Annals of Mathematical Statistics, 20, 46-63.

Arrow, K.J., H.B. Chenery, B.S. Minhas and R.M. Solow (1961) : Capital-labor Substitution and Econmic Efficiency. Review of Economics and Statistics, 43, 225-50.

Basmann, R.L. (1957) : A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. Econometrica, 25, 77-83.

Basmann, R.L. (1961) : A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases. Journal of the American Statistical Society, 56, 619-36.

Box, G.E.P. and D.R. Cox (1964) : An Analysis of Transformation, Journal of the Royal Statistical Association, Series B, 26, 211-43.

Chernoff, H. and N. Divinsky (1953) : The Computation of Maximum Likelihood Estimates of Linear Structural Equations, in W.C. Hood and T. Koopmans ed. "Studies of Econometric Method," Wiley, New York.

Christ, C.F. (1966) : Econometric Models and Methods. Wiley, New York.

Dhrymes, P.J. (1970) : Econometrics. Harper and Row, New York.

Fels, E. and G. Tintner (1966) : Methodik der Wirtschaftswissenschaftlichen Arbeitsmethoden, Munich, Oldenbourg, 8, 3-94.

Fisher, F.M. (1966) : The Identification Problem in Econometrics. McGraw-Hill, New York.

Fisk, P.R. (1967) : Stochastically Dependent Equations. Hafner, New York.

Goldberger, A.S. (1964) : Econometric Theory. Wiley, New York.

Haavelmo, T. (1944) : The Probability Approach in Econometrics. Econometrica, 12 (Suppl).

Hood, W.C. and T.C. Koopmans (1953) : Studies in Econometric Method. New York, Wiley.

Johnston, J. (1963) : Econometric Methods. McGraw-Hill, New York.

Keynes, J.H. (1948) : A Treatise on Probability. MacMillan, London.

Klein, L.R. (1953) : A Textbook in Econometrics. Row Peterson, Evanston, Illinois.

Klein, L.R. (1962) : An Introduction to Econometrics. Prentice-Hall, Englewood Cliffs, New Jersey.

Koopmans, T.C. (1949) : Identification Problems in Economic Model Construction. Econometrica, 17, 125 ff.

Koophans, T.C. (1950) : Statistical Inference in Dynamic Economic Models. Wiley, New York.

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950) : Measuring the Equation System in Dynamic Economics, in T.C. Koopmans ed. Statistical Inference in Dynamic Economic Models. Wiley, New York, 50ff.

Leser, C.E.V. (1966) : Econometric Techniques and Problems. Hafner, New York.

Marschak, J. (1953) : Economic Measurement for Policy and Prediction, in W.C. Hood and R. Koopmans ed. "Studies in Econometric Method", Wiley, New York.

Theil, H. (1961) : Economic Forecast and Policy, 2nd ed. Amsterdam, North Holland.

Tintner, G. (1952) : Econometrics. Wiley, New York.

Tintner, G. (1960) : Handbuch der Oekonometric. Berline Springer Veriag.

Tintner, G. (1966) : Some Thoughts about the State of Econometrics, in S. R. Krupp ed. The Structure of Economic Science, Prentice-Hall, Englewood Cliffs, New Jersey, 114-28.

Tintner, G. (1968) : Methodology of Mathematics and Economics and Econometrics. University of Chicago Press, Chicago.

Tintner, G. and R.V. Sastry (1969) : Information Theory and the Statistical Estimation of Econometric Relations, Multivariate Analysis. Academic Press, New York, 687-96.

Valavanis, S. (1963) : Econometrics. McGraw-Hill, New York.

Walters, A.A. (1970) : An Introduction to Econometrics. W.W. Norton, New York.

Wold, H. and L. Jureen (1953) : Demand Analysis. Wiley, New York.

Zarembka, P. (1968) : Functional Form in the Demand for Money. Journal of the American Statistical Association, 63, 508-11.

Zellner, A. and H. Theil (1962) : Three-stage Least Squares : Simultaneous Estimation of Simultaneous Equations. Econometrica, 30, 54-78.

APPENDIX A

Description of the data

Y_1=Actual per capita consumption of meat, poultry and fish in lbs.

Source : Agricultural Statistics (U.S. Department of Agriculture).

Y_2=Consumer price index of meat, poultry and fish (for urban wage earners and clerical workers U.S. city average 1957-59=100).

Source : Handbook of Labor Statistics 1967, 1968.

Z_1=Per capita disposable real income at 1958=100 prices in dollars.

Source : Statistical Abstract (U.S.)

Note : Series with 1947 and 1954 base have been adjusted to 1958=100.

Z_2=Average cost in dollars per 100 lbs of meat slaughtering deflated by all item consumer price index. Data used are average cost in dollars per 100 lbs. (under Federal Inspection) in slaughtering for cattle, calf, hog, sheep and lamb. Weighted average is computed using production figures as weights.

Source : Agricultural Statistics. U.S. Department of Agriculture. All Item consumer price index were taken from : Handbook of Labor Statistics with 1957-59=100.

Z_3=Index of total agricultural cost (total production expenses in millions of current dollars deflated by all item consumer price index (1957-59=100) and finally indexed with 1958=100). Since 1960 the data includes Alaska and Hawaii.

Source: Agricultural Statistics. U.S. Department of Agriculture.