APPLICATION OF STRUCTURAL ECONOMETRICS MODEL IN THE HEALTH CARE SYSTEM TO DESCRIBE RELATIONSHIPS AMONG VARIABLES IN THE SYSTEM

Rosmanjawati Abdul Rahman [1] and Anton Abdulbasah Kamil [2]

[1,2] *School of Mathematical Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia*
[1] ros3874@tm.net.my ; [2] anton@cs.usm.my

Abstract

In this paper we applied structural econometrics model into the health care system to show the interdependence between variables and to describe the relationships among them. We used the two-stage least squares (2SLS) method to estimate the model parameters and the results showed three main situations. First we found that the number of the registered patients have a positive relation on the mean duration of stay in hospital, the rate of the using of beds and the mean duration of empty beds whereas it shows negative relation on total of patient days. Secondly, it shows that the rate of the using of beds is positively related with the mean of occupied beds per day and the mean duration of empty beds whereas it is negatively related with total of patient days. Finally, we also find that the rate of the using of beds is positively related with the mean of occupied beds per day and the mean duration of empty beds whereas it is negatively related with total of patient days. The findings play as a beginning and a guide line to us to study more about health care system in Malaysia as health care system come under increasing pressure to provide fair and equitable care to all patients.

1. Introduction

An initial attempt to estimate a small (six - equation) econometric model of the U. S health-care system, was that of Feldstein (1967). At that time, each quantitative work in health economics was concerned with certain ratios, such as the physician – population ratio, and the Feldstein model was influenced by this approach. However, the model was developed to serve as a methodological prototype, not to provide detailed estimates of structural parameters of a complete model of the health-care system.

The second example of a simultaneous equations model of the health-care system is the 47- equation macroeconometric model of Yett, Drabek, Intriligator and Kimbel (1975). In this model, the endogenous variables are described in terms of the institutions and manpower are explicitly included, whereas the exogenous and standardizing variables included demographic variables, economic variables, insurance variables, and health manpower variables. The basic mechanism of the model is that of demand and supply, however the model is not an equilibrium one. The estimated model has been used for various purposes, including forecasts of health services and health manpower and simulation of certain changes in a state health-care system.

Other studies that used structural model are Morrisey, M. A and Jensen, G. A (1990) that described doctor's demand in hospital, a study of three-equation of structural equations by Benham, L (1971) in which its described the labor market for registered nurses, and Green, L.V. & Nguyen, V. (2001) that suggested strategies for cutting hospital beds.

In this paper, using the data provided by the selected general hospital in Pulau Pinang, we try to build a structural econometrics model to show the interdependence between variables and describe the relationships among them. The findings play as a beginning and a guide line to us to study more about health care system in Malaysia as health care system come under increasing pressure to provide fair and equitable care to all patients.

2. Identification

In econometrics modeling we always start with the identification problem. Identification is important to determine the choice of the technique to estimate the model paremeters. If an equation is unidentified, its parameters cannot be estimated by any econometric technique. If the equation is exactly identified, its coefficient can be estimated and the appropriate method to be used is the method of indirect least square (ILS). Otherwise if the equation is overidentified the method of ILS cannot be applied because it will not yield unique estimates of the structural parameters. There are two conditions which must be fulfilled for an equation to be identified.

2.1 The Order Condition

In a model of M simultaneous equations , in order for an equation to be identified, it must exclude at least $M - 1$ variables (endogenous as well as predetermined) appearing in the model.

Let K = number of exogenous variables in the system of equations

k = number of exogenous variables in the particular equation

M = number of endogenous variables in the system (equal to the number of structural equations)

m = number of endogenous variables in the particular equation.

i. If $K - k > m - 1$; the equation is overidentified

ii. If $K - k = m - 1$; the equation is just identified

iii. If $K - k < m - 1$; the equation is unidentified.

In a model of M simultaneous equations, in order for an equation to be identified, it must follow the order condition which stated as $K - k \geq m - 1$. The order condition for identification is necessary for a relation to be identified, but it is not sufficient, that is it may be fulfilled in any particular equation and yet the relation may not be identified. We need another condition for the identification which is the Rank Condition.

2.2 The Rank Condition

In a model containing M equations in M endogenous variables, an equation is identified if and only if at least one nonzero determinant of order (M −1) (M − 1) can be constructed from the coefficients of the variables (both endogenous and predetermined) excluded from that particular equation but included in the other equations of the model.

Generally, by combining these two rules, the principles of identification simultaneous equation in a system of M simultaneous equation can be stated as follows :

i. If $K - k > m - 1$ and the rank of the A matrix is $M - 1$, the equation is overidentified

ii. If $K - k = m - 1$ and the rank of the matrix A is $M - 1$, the equation is just identified.

iii. If $K - k = m - 1$ and the rank of matrix $A < M - 1$, the equation is unidentified.

Here A is defined as a matrix of parameters of excluded variables. The rank of a matrix is the order of the largest non-zero determinant which can be formed from the matrix.

3. Model Estimation

The model was estimated using monthly data from 1995 to 2003. First we look into the correlation among all variables and we come out with the structural equations as follow :

$$y_1 = \alpha_0 + \alpha_1 y_2 + \alpha_2 y_3 + \alpha_3 x_1 + \alpha_4 x_5 + \mu_1 \quad (1a)$$
$$y_2 = \beta_0 + \beta_1 y_1 + \beta_2 y_4 + \beta_3 x_4 + \mu_2 \quad (1b)$$
$$y_3 = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_3 + \gamma_3 x_5 + \mu_3 \quad (1c)$$
$$y_4 = \omega_0 + \omega_1 y_2 + \omega_2 x_5 + \omega_3 x_2 + \mu_4 \quad (1d)$$

with assumptions that $\mu_i; i = 1,2,3,4$ for all i are independently and identically distributed over the sample, with a zero mean and a constant covariance matrix. Under these assumptions, while the stochastic disturbance terms are uncorrelated over the sample, the stochastic disturbance term between equations can be correlated. This is, in fact the essence of the simultaneous equation system and the principal reason why it must be estimated as a system, rather than as a set of separate single equations.

Here, endogeneous variables are :

$y_1 =$ the number of registered patient

$y_2 =$ the mean duration of stay in hospital

$y_3 =$ the rate of the using of beds

$y_4 =$ total number of operations

and the exogenous variables are follows :

$x_1 =$ total of patient days

$x_2 =$ the number of death

$x_3 =$ the mean of occupied beds per day

$x_4 =$ the number of patient in third class room

$x_5 =$ the mean duration of empty beds

We investigated the model including the Breusch-Pagan-Godfrey test (BPG test) to check the assumption of constant variance for each error terms over the observations. The data was transform by taking logs to get the constant variance and the model was estimated by using the two-stage least squares method which will be describe next. All the test are done by using shazam v9.0 software.

3.1 The method of two-stage least squares (2SLS)

This method is a single-equation method, being applied to one equation of the system at a time.It has been accepted as a very useful estimation procedure for obtaining the values of structural parameters in over identified equations. This method is used to overcome the correlation problem that existed between endogenous variables and disturbance term that may leads to biased and inconsistent parameter estimates.

From the structural model above, we get the reduced form equation as follows :

$$y_1 = \pi_{11}x_1 + \pi_{12}x_2 + \pi_{13}x_3 + \pi_{14}x_4 + \pi_{15}x_5 + v_1 \quad (2a)$$
$$y_2 = \pi_{21}x_1 + \pi_{22}x_2 + \pi_{23}x_3 + \pi_{24}x_4 + \pi_{25}x_5 + v_2 \quad (2b)$$
$$y_3 = \pi_{31}x_1 + \pi_{32}x_2 + \pi_{33}x_3 + \pi_{34}x_4 + \pi_{35}x_5 + v_3 \quad (2c)$$
$$y_4 = \pi_{41}x_1 + \pi_{42}x_2 + \pi_{43}x_3 + \pi_{44}x_4 + \pi_{45}x_5 + v_4 \quad (2d)$$

Here, the reduced-form equations describe each endogenous variables ($y_i ; i = 1,2,3,4$) as a function of all the predetermined variables ($x_i ; i = 1,2,3,4,5$) of the model , and of a random element ($v_i ; i = 1,2,3,4$) while the π 's defined as the associated reduced-form coefficients of the model.

Formally, the 2SLS procedure works in the following manners :

1. In the first stage, we apply ordinary least squares to reduced-form equations 4a-4d to obtain estimates of the π 's. Then using these estimated reduced-form

coefficients, $\hat{\pi}$'s, we obtain a set of estimated values for the endogenous variables : $\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4$.

2. In the second stage regression, we substitute the $\hat{y}$'s into the structural equation and obtained the transformed functions $y_i = b_{i1}\hat{y}_1 + b_{i2}\hat{y}_2 + \ldots + b_{iG}\hat{y}_G + \gamma_{i1}x_1 + \ldots + \gamma_{ik}x_k + \mu_i^*$ where $\mu_i^* = \mu_i + b_{i1}v_1 + b_{i2}v_2 + \ldots + b_{iG}v_G$; $i, G = 1,2,3,4$; $k = 1,2,3,4,5$. At last by applying the ordinary least squares to this transformed structural equation, we obtain the 2SLS estimates of the structural parameters in equation 1a-1d.

In econometrics packages there already have this 2SLS instruction. Normally there is no needs to do the above regression stage by stage and we need not perform all the successive substitutions but we can estimate the parameter in the system directly using the 2SLS programme.

From the data we try to build the model and estimate the parameters, to describe the interdependence relationships among variables at best we can.

4. Results

Here are our computed results of our model estimation. The results of the BPG test are presented in Table 1.

Table 1: The BPG test results.

Structural equation	χ^2 - test statistics	P - value
(1a)	12.659	0.01307
(1b)	2.189	0.53410
(1c)	18.435	0.00101
(1d)	4.483	0.21384

From Table 1 we concluded that, there exist heteroscedasticity problem in equations (1a) and (1c). This is showed through their P – value in which are less than 0.05. In other words we reject the null hypothesis of the constant variance at the 5% of level of significance of the test. We transform the (1a) and (1c) equations by taking logarithma then do a heteroscedasticity diagnostic again. The results are as follows :

Table 2: The BPG test results after taking log for equation (1a) and (1c).

Structural equation	χ^2 - test statistics	P - value
(1a)	10.587	0.03162
(1b)	2.819	0.53410
(1c)	33.052	0.0000
(1d)	4.483	0.21384

From Table 2 above it showed that even after the transformation the heteroscedasticity problem still exist. The next step is to do a 2SLS method to see if this will solve the problem. We summarize the results in the Table 3 below:

Table 3: The BPG test results after 2SLS method

Structural equation	χ^2 - test statistics	P - value
(1a)	17.592	0.00148
(1b)	3.226	0.35803
(1c)	19.005	0.00027
(1d)	6.770	0.07960

From the table we can see that there's still occur heteroscedasticity problem in (1a) and (1c) equations even though we used the 2SLS method. Lastly we do the same transformation to both equations combine with the 2SLS method to check for the homogenous variance assumption and the results are as follows:

Table 4: The BPG test result with transformation of log and the 2SLS method

Structural equation	χ^2 - test statistics	P - value
(1a)	5.834	0.21192
(1b)	3.005	0.39091
(1c)	1.674	0.64269
(1d)	4.483	0.21384

From table 4, the BPG test shows that the heteroscedasticity problem have been solved. At the end, the new sets of structural equations that needs to estimate are as follows :

$$\ln y_1 = \alpha_0 + \alpha_1 \ln y_2 + \alpha_2 \ln y_3 + \alpha_3 \ln x_1 + \alpha_4 \ln x_5 + \varepsilon_1 \quad (3a)$$

$$y_2 = \beta_0 + \beta_1 y_1 + \beta_2 y_4 + \beta_3 x_4 + \varepsilon_2 \quad (3b)$$

$$\ln y_3 = \gamma_0 + \gamma_1 \ln x_1 + \gamma_2 \ln x_3 + \gamma_3 \ln x_5 + \varepsilon_3 \quad (3c)$$

$$y_4 = \omega_0 + \omega_1 y_2 + \omega_2 x_1 + \omega_3 x_2 + \varepsilon_4 \quad (3d)$$

Equations 3a-3d above suggest that some structural equation in logarithma whereas some are not. This kind of equation had been discussed by Carl R. Chen and Thomas L. Steiner (2000) and the estimation of the above model parameters resulted in follow equations with their correlation coefficients:

$$\ln y_1 = 1.4802 - 0.69240 \ln y_2 + 0.17188 \ln y_3 - 0.0016188 \ln x_1 + 0.88471 \ln x_5 + \varepsilon_1 \ ; \ R^2 = 0.9726 \quad (4a)$$

$$y_2 = 3.2571 - 0.049487\, y_1 - 0.020072\, y_4 - 0.022396\, x_4 + \varepsilon_2 \quad ; R^2 = 0.5280 \quad (4b)$$

$$\ln y_3 = -1.4737 - 0.013115 \ln x_1 + 0.74433 \ln x_3 + 0.19778 \ln x_5 + \varepsilon_3 \quad ; R^2 = 0.9874 \qquad (4c)$$

$$y_4 = 648.15 - 155.93\, y_2 + 0.035759\, x_1 - 1.0091\, x_2 + \varepsilon_4 \quad ; R^2 = 0.4459 \qquad (4d)$$

Equation (4a) shows that the number of registered patient have a positive relation on the mean duration of stay in hospital, the rate of the using of beds and the mean duration of empty beds whereas it shows negative relation on total of patient days. This tell us that in a day with more patient turn up, the hospital needs more beds to occupied the patient and the probability that the patient have to stay at the hospital is high.

From equation (4b), we learned that the mean duration of stay in hospital is negatively related with the number of registered patient, total number of operations and the number of patient in third class room. It means that for how long the patient stay at the hospital is somehow not related with the number of patient and the number of operations. This also tells us that the patient have to stay longer at the hospital not just because they are having the operations but for some other medical reasons.

Equation (4c) shows that the rate of the using of beds is positively related with the mean of occupied beds per day and the mean duration of empty beds whereas is negatively related with total of patient days. This telling us that with more patient turn up, it does not meaning that the bed will be all be occupied. And lastly from equation (4d), we learned that if the hospital have more patients then the probability they have to perform operations is high.

5. Conclusion

We conclude that this study have shed us some lights about the health care system in our country and gives us more understanding about the situation in the hospital. It is the beginning steps for us to study more about the health care system and we would like to extent the model with the nonlinear model.

References

1. L. Benham, "The The Labor Market For Registered Nurses: A Three- Equation Model ", *Review of Economics and Statistics.* **53** (1971) 246 – 252.

2. R. Chen Carl and L. Steiner Thomas, "Tobin's Q, Managerial Ownership, and Analyst Coverage : A nonlinear simultaneous Equations Model ", *Journal of Economics and Business.* **52** (2000), 365-382.

3. M.S. Feldstein, "An Econometric Model of the Medicare System ", *The Quarterly Journal of Economics.* **85** (1971), 1 – 20.

4. L.V. Green and V. Nguyen, "Strategies for Cutting Hospital Beds: The Impact on Patient Service.", *Health Services Research.* **36** (2001), 421-442.

5. W.H. Greene, *Econometric Analysis,* (Prentice Hall International, Inc. USA, 2000).

6. D.N. Gujarati, *Basic Econometrics,* (McGraw-Hill, New York, 1995).

7. J.A. Hausman and W.E. Taylor, "Identification in Linear Simultaneous Equations Models With Covariance Restrictions : An Instrumental Varibles Interpretation ", *Econometrica.* **51** (1983), 1527 – 1547.

8. M.D. Intriligator, *Econometric Models, Techniques and Applications,* (North – Holland Publishing Company. Armsterdam, Oxford , 1978).

9. J.B. MacArthur and H.A. Stranahan, "Cost Driver Analysis in Hospitals : A Simultaneous Equations Approach ", *Journal of Management Accounting Research.* **10** (1998), 279 – 312.

10. M.A. Morrisey and G.A. Jensen, "Hospital For Physicians " , *Quarterly Review of Economics and Business.* **30 (1)** (1990), 16 – 39.

11. M. Webb, S. Geoffrey and C. Bramson, "An Approach to the Control of Bed Occupancy in a General Hospital", *Operational Research Quarterly.* **28 (2)** (1998), 391-399.

12. D.E. Yett, L.J. Drabek, M.D. Intriligator and L.J. Kimbell, "A Microeconometric Model of the Health Care System in the United States ", *Annals of Econometric and Social Measurement.* 4 (1975), 407 – 433.