THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV., VOL.40, NO.1,1996

ESTIMATION OF STRUCTURAL EQUATION MODELS WITH AN APPLICATION TO THE DEMAND FOR CHILDREN IN EGYPT

By

HASSAN H.M. ZAKY*

ABSTRACT

Demographic analysis has started to apply structural equation modelling to its different processes. This paper presents an application that models the demand for children. It discusses the procedures used to estimate the structural parameters of this model. The paper then tests the model using the 1992 Egypt Demographic and Health Survey. The model for the demand for children is estimated by four estimation procedures, and the results are compared. The model is also tested for exogeneity.

---

*The Department of Statistics, Faculty of Economics and Political Sciences, Cairo University, and the Department of Statistics, Faculty of Economic and Administrative Sciences, United Arab Emirates University.
The author is deeply indebted to Professor Ahmed A. Saif El-Nasr, the Department of Statistics, Faculty of Economic and Administrative Sciences, United Arab Emirates University, for his constructive comments. Final thanks go to the referees for their helpful and valuable suggestions which significantly improved the paper.

1. INTRODUCTION

During the last two decades, the process or processes under study, in demographic modelling, have been represented by a series of simultaneous interdependent equations. Examples of such research include Bollen et al. (1995), Khan et al. (1983), Rosenzweig and Schultz (1985), Schultz (1978), and Willis (1973).

The determinants of completed family size, for example, may include as explanatory variables the couple's ages, their level of education, employment characteristics, housing characteristics etc.. However, when the analysis is extended to incorporate other choice variables such as contraception use and children's quality of health care, this situation gives rise to the use of simultaneous equations models. Accordingly, one should interpret cautiously any results obtained from single equation models due to the possibility of presenting some bias. The source of this bias is the correlation between the disturbance term of one equation and the endogenous variables on the right hand side of the same equation.

Estimation of systems of simultaneous equations is straight forward and are explained in many references (Bollen, 1989; Johnston, 1984; Judge et al., 1988; Pindyck and Rubinfeld, 1991). The estimators possess highly desirable asymptotic properties. For the small sample properties, Judge et al. (1988) stated that there have been many new achievements. Owing to the analytical complexities involved in carrying through the necessary derivation, results are available for special cases (Phillips, 1983). However considerable results from sampling experiments are available in the literature using the socalled Monte Carlo experiments (Bollen et al., 1995). Although results vary among the experiments, in general, the asymptotic theory seems to provide a good guide to the small sample estimator performance.

Bollen et al. (1995) review indicated that most of the evidence in favour of the asymptotic properties comes from time series data, in which the reduced form regressions yield instrumental variables that are correlated highly with the right hand side endogenous variables, but not correlated with the error terms. Cross-sectional data often used in the demographic research typically do not yield high reduced form R squared. These procedures may result in parameter estimators with high variance in spite of their desirable asymptotic properties. Mariano (1982) examines the trade-off between bias and variance.

The main objective of the paper is thus to apply the notion of systems of equations to a demographic research point namely, the demand for children. This involves estimating the model using various techniques, comparing the results, and finally identifying the most appropriate approach, given the model and the data set.

In the next section, the system of equations that aims to study the demand for children in Egypt is presented. A review of the different estimation procedures follows. These methods are then applied to estimate the model, using 1992 Egypt Demographic and Health Survey. Finally, some methodological implications are suggested.

2. MODELLING THE DEMAND FOR CHILDREN IN EGYPT

The theory underlying the model, which is tested here, can be traced originally to Becker's observations (Becker, 1960) that Hicksian theory of the consumer could be fruitfully applied to a family's decisions regarding the number of children. A full elaboration had to await an alternative theory of the consumer developed by Willis (1973). The main feature of such development is that the reduction of fertility within the calculus of choice depends on the prices of nonmarketed commodities, especially time. What we extract mainly from this development is the simultaneity inherent in the various decisions taken by the household.

Suppose we want to estimate the effect of contraceptive use and children's quality of health on the demand for children. These three decisions namely to use contraception, to provide "good" quality of health care for the children, and to have children are clearly choice variables.

The algebraic representation of the model is as follows:

$$C = H\alpha_{hc} + U\alpha_{uc} + X_c\beta_c + \epsilon_c \qquad (1)$$

where

C = n * 1 vector of observations on the demand for children variable,

H = n * 1 vector of observations on the children's quality of health variable,

U = n * 1 vector of observations on the contraceptive use variable,

α_{hc}= 1 * 1 scaler of the coefficient for the health variable,

α_{uc}= 1 * 1 scaler of the coefficient for the contraceptive use variable,

X_c= n * K_c matrix of observations on included predetermined variables in the demand for children equation,

$ß_c$= K_c * 1 vector of coefficents for the predetermined variables included in the demand for children equation,

and ϵ_c= n * 1 vector of disturbances associated with the demand for children equation.

Similarly, $$H = C\alpha_{ch} + U\alpha_{uh} + X_h ß_h + \epsilon_h \quad (2)$$

Also $$U = C\alpha_{cu} + H\alpha_{hu} + X_u ß_u + \epsilon_u \quad (3)$$

ϵ_c, ϵ_h, and ϵ_u are random terms that have means of 0 and are assumed to be uncorrelated with the predetermined variables. The appropriate method for estimating these coefficients depends on the assumption made about the interdependent nature among the choice variables. One is faced with one of two hypotheses. First, according to the theory underlying the model, the interdependence among the choice variables is significant. Consequently, Equations (1)-(3) are called the structural equations or the structural form of the model. Second, there is an alternative valid assumption that is the interdependence among the choice variables is null or insignificant, namely the hypothesis of exogeneity among the three outcome variables. These two hypotheses are examined in the study.

3. METHODS OF ESTIMATION UNDER THE SIMULTANEITY ASSUMPTION

One may apply several techniques to estimate the parameters of a structural form. In this section, the most common methods[1] used to solve the endogeneity problem in the model are briefly presented.

a) One-Stage Regression Estimation

The one-stage estimation technique simply applies the ordinary least squares method to each equation separately. This method mainly ignores the correlation between the endogenous right hand side variables and the error terms. In the presence of endogeneity, biased estimators are produced. The estimators are also inconsistent and inefficient. However, when the exogeneity assumption is valid, the estimators are unbiased, consistent and efficient.

b) Two-Stage Regression Estimation

Among the single equation estimation methods, the two-stage estimation method stands out. The first stage involves the creation of instruments for the endogenous variables. This is carried out in the reduced form of the model by regressing each endogenous variable on all predetermined variables. The second stage employs these instrumental variables in the structural form instead of the endogenous variables on the right hand side, and estimate each equation separately. Two-stage method yields consistent estimators but inefficient.

---

[1] For a detailed description of these methods, refer to Judge et al. (1988) for the one-stage and two-stage least squares, and refer to Bollen (1989) and Jöreskog and Sörbom (1993) for the details of the generalized least squares and the maximum likelihood estimation methods.

c) Generalized Regression Estimation

The generalized least squares (GLS) may be expressed in many ways with each has its own interpretation in the literature. One of these interpretations is the two-stage least squares (Dhrymes, 1974; Judge at al., 1988; Theil, 1971). Other versions (Bollen, 1989) are the maximum likelihood and the unweighted least squares, depending on the weights applied. Following Bollen (1989), a general form of the generalized least squares fitting function is:

$$F_{gls} = 1/2\ tr[(S - \Sigma(\theta))W^{-1}]^2$$

where $\Sigma(\theta)$ is the covariance matrix of the variables and function of all parameters in the model, S is the sample covariance matrix, and W^{-1} is a weight matrix for the residual matrix. Although many choices for W^{-1} are consistent estimators of Σ^{-1}, the most widely used choice is $W^{-1} = S^{-1}$. This is the one employed in LISREL 8 (Jöreskog and Sörbom, 1993) and it is the one used to estimate our model. The estimators are consistently efficient (Bollen, 1989; Jöreskog and Sörbom, 1993).

d) Maximum Likelihood Estimation

To date, the most widely used fitting function for general structural equation models is the maximum likelihood (ML) function (Bollen, 1989). Our model's fitting function is

$$F_{ml} = \log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - \log|S| - (3+K)$$

where K is the number of predetermined variables. Minimization of this function produces estimators that are asymptotically unbiased, consistent, and efficient. Bollen (1989) also showed that these estimators are scale invariant and scale free, with few exceptions.

4. THE INFLUENCE OF CONTRACEPTIVE USE AND CHILDREN'S QUALITY OF HEALTH ON THE DEMAND FOR CHILDREN

This study uses 1992 Egypt Demographic and Health Survey (EDHS-92) data. To study the demand for children, one may use one of two outcome variables, namely desired number of children or actual completed family size. It has been shown that there is relatively low agreement between spouses with regards to their reported desired numbers of children (Zaky, 1995). Consequently, to avoid the discrepancy in fertility desires between husbands and wives, completed family size is used as a measure for the effective demand for children.

The EDHS-92 includes 2406 couples. To reach the couples that have completed their family size, the analysis is focussed on those who do not desire any additional children, as reported by the wife. Also the analysis is further confined to couples where each spouse is married only one time to rule out the effect of past experience on the demand for children within the current union. Thus, the size of the subrised sample reaches 1338 couples.

Around 51 per cent of the couples reside in urban areas. A difference of about 7 years of age is observed between spouses; males average 42 years of age while their wives average 35 years of age. Couples have been married on average for 16 years, and have given birth to 5 children. Every two couples have almost one child death. The number of child deaths that each couple has is used as a measure of children's quality of health. Ever use of contraception reaches about 82 per cent. 50 per cent of the couples have reported the presence of spousal communication about family size. An average typical woman does not complete primary education since she completes only 5 years of education. The husband exceeds his wife by only 3 years of education. About 23 per cent of the wives are currently employed. While 13 per cent of the husbands are employed in technical/ professional or administrative jobs. The average monthly expenditure for an Egyptian family is just below 300 Egyptian Pounds.

Next we estimate the system of equations using three techniques, namely two-stage least squares, generalized least squares, and maximum likelihood by using LISREL 8. However, before going into the details of the results, it is worth mentioning that the system has been estimated first by one-stage least squares and the exogenous variables with P value less than 0.10 have been excluded from the analysis for several reasons. First, to satisfy the identification condition of the system of equations. Second, to focus more on the dynamics among the three choice variables, namely contraceptive use, children's quality of health, and the demand for children. Third, to identify the most important determinants of the demand for children in order to deduce some policy implications.

Table (1) introduces the results, of the three above mentioned techniques along with the one-stage procedure. The demand for children equation is estimated after excluding the exogenous variables that showed insignificant importance. The results indicate that system estimation as represented by both Generalized Least Squares (GLS) and Maximum Likelihood (ML) produce very close estimates for both parameters and standard deviations. However the solution converged

ISSR, Cairo Univ., Vol.,40, No.1,1996, PP.62-79

Table 1. Parameter Estimates of the Demand For Children Equation

Variable	One-stage Least Squares	Two-stage Least Squares	Generalized Least Squares	Maximum Likelihood
Constant	3.1752a (0.2456)	6.5590a (2.5960)	3.6942a (0.3055)	3.7058a (0.3071)
Number of Child Deaths	0.9548a (0.03672)	0.3156 (0.6768)	0.2120b (0.1021)	0.2094b (0.1025)
Ever Use of Contraception	0.2466b (0.1020)	-6.0476 (4.7230)	0.04954 (0.2508)	0.02749 (0.2523)
Number of Assets Owned	-0.05423a (0.01565)	0.02618 (0.06388)	-0.08467a (0.01847)	-0.08429a (0.01856)
Wife's Years of Education	-0.04127a (0.009143)	-0.002685 (0.03082)	-0.04706a (0.01060)	-0.04696a (0.01063)
Wife's Age	0.1477a (0.005759)	0.2096a (0.05487)	0.1897a (0.008418)	0.1899a (0.008462)
Wife's Age at Marriage	-0.1839a (0.01153)	-0.2381a (0.05755)	-0.2432a (0.01499)	-0.2434a (0.01508)
Spousal Communication	-0.2439a (0.07951)	0.08949 (0.2693)	-0.2208b (0.0[illegible]6)	-0.2197b (0.09176)
Place of Residence	-0.4616a (0.08292)	0.07928 (0.4151)	-0.5291a (0.09754)	-0.5274a (0.09784)
Standard Error	1.3429	1.6492	1.5288	1.5371
R squared	0.7198	0.5795	0.6353	0.6327

Standard deviations are in parenthesis

a= p value ≤ 0.01
b= 0.01 < p value ≤ 0.05
c= 0.05 < p value ≤ 0.10

after 79 iterations with GLS while after only 29 iterations with ML. These two methods have not signified the effect for ever use of contraception on the demand for children but the effect of child deaths on the demand for children is highly materialized.

It is clear that one-stage least squares has produced smaller standard deviation estimates than the system estimation techniques. The positive association between both child mortality and ever use of contraception with the demand for children is obvious. On the other hand, two-stage least squares seems to be inappropriate for our model or data set.

According to the one-stage estimation of the demand for children equation in Table (1), couples residing in rural areas, couples of older wives, couples of wives who are married younger, couples of less educated wives, couples with fewer number of assets, couples with more children deaths, couples who are more probable to have ever contracepted, and/or those couples who do not discuss their fertility desires are more likely to have bigger families, i.e. have higher demands for children than other couples.

The one-stage results suggest some policy implications. The wife's role in the demand for children is more dominant than that of the husband's. Family planning programs should emphasize the concept of spousal communication on fertility desires. Health programs directed towards the improvement of quality of child health will obviously alter the fertility behaviour. Female education is clearly an effective policy tool that determines the demand for children and it has various effects. Education may increase the female age at marriage, which is an important determinant of fertility. Also, female education will probably shape the couple's fertility attitudes and behaviour.

Observing Tables (2) and (3) reveals the same findings as those indicated in Table (1) namely the superiority of one-stage least squares over the other three techniques, given the data and the model.

According to the one-stage estimation of the child mortality equation in Table (2), couples with higher parity, those who have lower probability to have ever used contraception, couples residing in urban areas, couples of older or more educated wives, couples who communicate, and/or those couples who spend less are more likely to experience more child deaths than others. Two unexpected findings are related to place of residence and wives years of education. These results may be due to some multicollinearity among the exogenous variables. A point worth examining but it is beyond the scope of the study.

For the contraceptive use equation in Table (3), higher parity couples with fewer child deaths are more inclined to have ever contracepted. Couple's communication is crucial in determining contraceptive use. Urban couples are more likely to have ever contracepted than their rural counterparts. Couples of older and more educated wives and wealthy couples are more probable to have ever contracepted than other couples.

Table 2. Parameter Estimates of the Child Mortality Equation

Variable	One-stage Least Squares	Two-stage Least Squares	Generalized Least Squares	Maximum Likelihood
Constant	-0.5012a (0.1217)	-0.3790 (0.3886)	-0.5604b (0.2467)	-0.5596b (0.2471)
Number of Children ever Born	0.3414a (0.01209)	0.3045a (0.03273)	0.2971a (0.02587)	0.2972a (0.02588)
Ever Use of Contraception	-0.2385a (0.06137)	-0.4393 (0.7402)	-0.05338 (0.4295)	-0.05653 (0.4304)
Family Monthly Expenditure	-0.0003405b (0.0001345)	-0.0003220b (0.0001672)	-0.0003304b (0.0001381)	-0.0003310b (0.0001381)
Wife's Years of Education	0.01921a (0.006055)	0.01632c (0.009413)	0.01299c (0.007645)	0.01293c (0.007643)
Wife's Age	-0.01035a (0.003821)	-0.003269 (0.007850)	-0.004060 (0.005922)	-0.004061 (0.005926)
Husband's Years of Education	-0.008908 (0.005637)	-0.01045 (0.008229)	-0.01326b (0.005780)	-0.01307b (0.005784)
Spousal Communication	0.1178b (0.04840)	0.1166c (0.07043)	0.09487c (0.05387)	0.09498c (0.05389)
Place of Residence	0.1320a (0.05137)	0.1226 (0.1025)	0.07535 (0.07059)	0.07571 (0.07065)
Standard Error	0.8144	1.0040	0.8209	0.8211
R squared	0.4913	0.2316	0.4807	0.4813

Standard deviations are in parenthesis
a= p value ≤ 0.01
b= 0.01 < p value ≤ 0.05
c= 0.05 < p value ≤ 0.10

Table 3. Parameter Estimates of the Contraceptive Use Equation

Variable	One-stage Least Squares	Two-stage Least Squares	Generalized Least Squares	Maximum Likelihood
Constant	0.4429a (0.05228)	0.3964a (0.09491)	0.4496a (0.07288)	0.4501a (0.07306)
Number of Children ever Born	0.01746a (0.006724)	0.03923 (0.04256)	0.01327 (0.02780)	0.01305 (0.02791)
Number of Child Deaths	-0.04703a (0.01202)	-0.1258 (0.1298)	-0.03674 (0.08276)	-0.03599 (0.08315)
Number of Assets Owned	0.01687a (0.00417)	0.01536a (0.004749)	0.01697a (0.004329)	0.01697a (0.004336)
Wife's Years of Education	0.007611a (0.002319)	0.003222a (0.002875)	0.007405a (0.002761)	0.007405a 0.002763)
Wife's Age	0.003319b (0.001683)	0.007683 (0.002647)	0.003565 (0.002497)	0.003569 (0.002499)
Spousal Communication	0.05589a (0.02134)	0.06164b (0.02442)	0.05468b (0.02261)	0.05462b (0.02264)
Place of Residence	0.1022a (0.0223)	0.1043a (0.02503)	0.1008a (0.02403)	0.1008a (0.02406)
Standard Error R squared	0.3604 0.1126	0.3633 0.1030	0.3599 0.1123	0.3606 0.1118

Standard deviations are in parenthesis

a= p value ≤ 0.01
b= 0.01 < p value ≤ 0.05
c= 0.05 < p value ≤ 0.10

Given these results, one is now doubtful whether or not the hypothesis of simultaneity is valid. Accordingly, the assumption of simultaneity is further explored by using the test for exogeneity among the outcome variables. This test has been applied by Bollen et al. (1995) to the case of one endogenous variable on the right hand side. The generalization to two or more endogenous variables is feasible.

Table (4) presents the one-stage regression of the demand for children, after adjusting for two newly added exogenous variables. These two variables are the residuals obtained from the reduced form of the child mortality equation and the residuals from the reduced form of the contraceptive use equation. A simple t-test on the coefficients of the residuals is the test of the null hypothesis that both child mortality and contraceptive use are exogenous in the demand for children equation.

The results of the parameter estimates of the residuals of number of child deaths and the residuals of contraceptive use suggest that we fail to reject the null hypothesis that child mortality and contraceptive use in the demand for children equation are exogenous.

The test for exogeneity is applied also to both child mortality and contraceptive use equations, as shown in Table (4). All results indicate that the exogeneity among the three outcome variables is significant.

Table 4. Parameter Estimates for the Equations Using the One-Stage Least Squares to Test for Exogeneity

Variable	Demand For Children	Child Mortality	Contraceptive Use
Constant	6.5590a (2.1190)	-0.3790 (0.3161)	0.3964a (0.09446)
Number of Children Ever Born		0.3045a (0.02663)	0.03923 (0.04235)
Residuals of Number of Children Ever Born		0.04639 (0.02989)	-0.02111 (0.04299)
Number of Child Deaths	0.3156 (0.5523)		-0.1258 (0.1292)
Residuals of Number of Child Deaths	0.6385 (0.5536)		0.07930 (0.1298)
Ever Use of Contraception	-6.0480 (3.8550)	-0.4393 (0.6021)	
Residuals of Ever Use of Contraception	6.2990 3.8560	0.2021 0.6053	
Number of Assets Owned	0.02618 (0.05214)		0.01536a (0.004726)
Family Monthly Expenditure		-0.0003220b (0.0001360)	
Wife's Years of Education	-0.002690 (0.02516)	0.01632b (0.007657)	0.007683a (0.002861)
Wife's Age	0.2096a (0.04478)	-0.003269 (0.006386)	0.003222 (0.002634)
Wife's Age at Marriage	-0.2381a (0.04697)		
Husband's Years of Education		-0.01045 (0.006695)	
Spousal Communication	0.08950 (0.2198)	0.1166b (0.05730)	0.06164b (0.02430)
Place of Residence	0.07930 (0.3388)	0.1226 (0.08340)	0.1043a (0.02491)
Standard Error	1.3470	0.8167	0.3615
R squared	0.7203	0.4922	0.1129

Standard deviations are in parenthesis

a= p value ≤ 0.01
b= 0.01 < p value ≤ 0.05
c= 0.05 < p value ≤ 0.10

Consequently, given the EDHS-92 data one can deduce that the simple one-stage approach is most appropriate for estimating the effect or influence of child mortality and contraceptive use on the demand for children.

5. CONCLUSIONS

Structural equation models have widely been applied in population studies and estimated by several estimation methodologies. However, one should be cautious that the various estimation procedures may lead to different statistical results. The validity of these models seems to be highly dependent on the nature of the application. Studying the demand for children in Egypt suggests that simple models may be more appropriate than more complicated ones.

Testing for exogeneity turns to be very important. It is a helpful tool to test for the endogeneity assumption derived from the economic theory of fertility behaviour. The result of the test, in addition, may guide the relaxation of any a priori restrictions that may be imposed on the model.

ISSR. Cairo Univ., Vol., 40, No.1, 1996, PP.62-79

REFERENCES

Becker, G.S. 1960. "An Economic Analysis of Fertility." In Demographic and Economic Change in Developed Countries. Universities-National Bureau Conference. Series No. 11. Princeton, New Jersey: Princeton University Press.

Bollen, K.A. 1989. Structural Equations with Latent Variables. New York: Wiley.

Bollen, K.A., D.K. Guilkey, and T.A. Mroz. 1995. "Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia." Demography 32:111-131.

Dhrymes, P.J. 1974. Econometrics, Statistical Foundations and Applications. New York: Springer-Verlag.

Johnston, J. 1984. Econometric Methods. New York: McGraw-Hill.

Jöreskog, K. and D. Sörbom. 1993. LISREL 8. Chicago: Scientific Software International.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl, and T.C. Lee. 1988. Introduction to the Theory and Practice of Econometrics. New York: Wiley.

Khan, M.A. and I. Sirageldin. 1983. "How Meaningful are Statements about the Desired Number of Additional Children? An Analysis of 1968 Pakistani Data." The Pakistan Development Review 22:1-22.

Mariano, R.S. 1982. "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous Equation Case." International Economic Review 23:503-534.

Phillips, P.C.B. 1983. "Exact Small Sample Theory in the Simultaneous Equations Model." In Handbook of Econometrics, Volume I, ed. Z. Griliches, and M.D. Intriligator. Chapter 8:449-516. Amsterdam: North-Holland.

Pindyck, R.S. and D.L. Rubinfeld. 1991. Econometric Models and Economic Forecasts. New York: McGraw-Hill.

Rosenzweig, M.R. and T.P. Schultz. 1985. "The Demand for and Supply of Births: Fertility and its Life Cycle Consequences." American Economic Review 75:992-1015.

Schultz, T.P. 1978. "The Influence of fertility on Labour Supply of Married Women: Simultaneous Equation Estimates." In Research in Labour Economics 2:273-351. Connecticut: JAI Press.

Theil, H. 1971. Principles of Econometrics. New York: Wiley.

Willis, R.J. 1973. "A New Approach to the Economic Theory of Fertility Behaviour." Journal of Political Economy 81:2:II:S14-S64.

Zaky, H.H.M. 1995. "Profile of Men's and Women's Fertility Preferences in Egypt." In Perspectives on Fertility and Family Planning in Egypt: Results of Further Analysis of the 1992 Egypt Demographic and Health Survey, ed. M. Mahran et al., 55-79. Calverton, Maryland: National Population Council and Macro International Inc.