FERTILITY DATA ANALYSIS USING A SIMULTANEOUS EQUATION MODEL

***Paramjit Singh, **Dr. Anuj Kumar**

*Research Scholar, Dept. Of Mathematics, CMJ University, Shillong, Meghalaya, India.
**Assistant Professor, Dept. Of Mathematics, Govt. College, Malerkotla, Punjab.
Research Guide, Dept. Of Mathematics, CMJ University, Shillong, Meghalaya, India.

ABSTRACT

Both economic and demographic factors can play a role in the fluctuation of the fertility rate; in fact, there are many factors that lead to either the reduction or the increase in the fertility rate. Our primary objective is to determine the regression coefficients of the variables (factors) on which the fertility rate depends and on the basis of which parameters estimates are made for the purpose of predicting the future value of the dependent variable. The accuracy of the prediction is contingent on the selection of an adequate model as well as the utilization of an effective estimation process to estimate the regression coefficients. It is common practice to employ the method of ordinary least squares (OLS) when attempting to determine the values of the parameters of a linear model with a single equation. This is because this method is strongly dependent on the primary quantitative approach utilized in economics. By utilizing the Gauss-Markov theorem, it is able to provide the best linear unbiased estimates (BLUE) of the parameters when spherical assumptions are taken into consideration. Yet, the vast majority of economic theories, such as the fertility model, require a collection of relationships; hence, the application of single equations cannot be justified in these kinds of circumstances. Because single equation models in economics do not provide an accurate representation of the issues facing the economy, we need to investigate the connections between the variables that make up simultaneous equation models.

KEYWORDS: *Statistical Analysis; Fertility; Simultaneous; Equation Model.*

INTRODUCTION

India's population is currently the second largest in the world, and it has practically doubled in size over the past three decades. India's poverty, widespread illiteracy, particularly among females, the age of girls at the time of marriage, the status of women in society, as well as the central and state departments in charge of family planning, are primarily to blame for the country's extremely high rates of both fertility and infant mortality, which have remained at extremely high levels.

The likelihood of having children is significantly correlated with a woman's fertility (Chowdhury, 1988; Schultz, 1978). In the same way that economic and cultural factors can affect child survival, they can also affect the total fertility rate (TFR) [Bongaarts, 1978]. The total fertility rate in India, often known as the average number of children born to each woman, is still at 2.9 (according to the Sample Registration System Bulletin, 2004), which is significantly higher than the number that would be considered desirable at the replacement level (2.1 children). The total fertility rate in rural areas is currently 2.56 children, which is significantly higher than the urban TFR of 1.91 children. India is home to around 16% of the total population of the planet. According to estimates provided by the United Nations Population Fund (UNFPA), India alone is responsible for as many as 16 million of the annual 76 million population increases that occur around the world, making up a considerable proportion (21%) to this phenomenon.

India had a total population of 1028 million people as of March 2001, with 531 million men and 496 million females, according to the Census of India, which was conducted in 2001. (Economic Survey, 2003-04). According to the results of the censuses conducted from 1981 to 2001, the population of people aged 7 and older increased by between 26 and 27 percent for each decade.

When it comes to population, Haryana is ranked sixteenth in India. The population growth rate in Haryana, which was 28.43 percent (Census, 2001), is significantly greater than the level for all of India, which was 21.54 percent, according to the demographic landscape of the state of Haryana (Census, 2001). According to the Census of Haryana conducted by the Government of India in 2001, the population of Haryana has increased from 7.5 million in the year 1961 to 21.1 million at the beginning of the 21st century.

Within the setting of India, a new child is being born approximately once every 1.2 seconds, towards the end of each minute 44, within each hour 2,640, and within each day 63,000. The desire to have large families and the low level of acceptability of family planning are the two primary variables that contribute to the high fertility rate of the Indian population.

The total fertility rate (TFR), as reported by SRS bulletin, has decreased from 6.4 during the period of 1970-1972 to 3.4 at the end of the 20th century. When compared to India as a whole, this decrease in TFR is far more pronounced in the state of Haryana. In the years 1970-1972, the TFR in Haryana was higher than the level for all of India, which was 5.3, but by the year 1996-1998, the TFR in Haryana was practically the same as the level for all of India, which was 3.4. (3.3). According to the most recent findings of the National Family Health Survey, Haryana is still ranked 19th among the 29 states, despite the fact that its total fertility rate (TFR) has decreased (NFHS). The TFR has decreased from 3.99 in the year 1992-1993 (NFHS-1) to 2.7 in the year 2005-2006, according to the estimates that come from all three rounds of the NFHS (NFHS-3). Haryana is falling more and further behind other states in India, such as Andhra Pradesh, Tamil Nadu, Kerala, Himachal Pradesh, and Punjab, who have already reached the total fertility rate (TFR) goal of less than 2, which is necessary to moderate India's population growth.

There has been a consistent decrease in the rate of infant mortality throughout the course of the past three decades. According to the estimates provided by the SRS for the entirety of India, the IMR decreased significantly between 1971 and 1975, going from 134 fatalities per 1000 live births to somewhere around 70 by 1999. According to the findings of multiple rounds of the NFHS, the infant mortality rate (IMR) for all of India decreased from 79 deaths per 1000 live births in 1992-1993 (NFHS-1) to 57 deaths per 1000 live births in 2006. (NFHS-3). Nonetheless, when measured against the standards of other nations, it is still quite high. For the past many years, the IMR in Haryana has similarly decreased. According to NFHS-1, the IMR was 73 in the 1992-1993 school year, but it dropped to 57 in the 1998-1999 school year (NFHS-2). According to the most recent NFHS survey (NFHS-3, 2006), the number has dropped even further to 42.

Although the literacy rate among women in the country has increased from 0.60 in the 1901 census to 24.88 in the 1981 census and 54.16 in 2001 (Economic Survey, 2003-04), this rate of increase is still a long way behind that of males in the country, for whom the corresponding increase was from 9.83 percent in the 1901 census to 46.74 percent in the 1981 census and further 75.85 percent in 2001. (Economic Survey, 2003-04). When compared to India as a whole, the percentage of illiterate people in Haryana has increased at a faster rate. This applies to both male and female residents. The literacy percentage of women in Haryana was 9.2 in 1961 and climbed to 55.73 percent in 2001, according to the Census of Haryana. Nevertheless, the literacy rate of men increased from 29.2 in 1961 to 78.5 in 2001. Notwithstanding this improvement, Haryana continues to hold the 23rd position among women and the 17th position among men.

The fertility rate of a state is proportionately inversely related to the percentage of women in that state who have completed formal education (Rafiqul Huda Chaudhary, 1996). Women who work outside the home for someone else (as opposed to working on a family farm or business or being self-employed) tend to have lower fertility rates and higher rates of contraceptive use. This is due, in part, to the fact that their jobs compete with the caregiving responsibilities they have for their children for their time and attention. It was reported in the census of 1981 that the percentage of women workers to their total population had increased from 14.22 in 1971 to 20.85 in 1981. This places women in a better position than men, whose work participation rate was reported to have slightly increased from 52.75 percent in 1971 to 53.19 percent in 1981. The percentage of working women in Haryana rose from 2.41 in 1971 to 4.82 in 1981 and then further to 13.38 in 2001. This represents a steady increase since 1971. (SRS).

2. REVIEW OF LITERATURE

Boca and Sauer (2006) developed a dynamic utility maximization model of female labour force participation and fertility choices and estimated approximate decision rules by using data on married women in Italy, Spain, and France. Their research focused on fertility choices and labour force participation. The pattern of estimated effects of state dependency across nations was consistent with aggregate patterns in part-time work and child care availability. This finding suggests that labour market rigidities and a lack of options for child care were significant causes of state dependence. The model's simulations show that if Italian and Spanish women were placed in an atmosphere similar to that of France's institutions, their participation rates would significantly rise.

Klein, R., & Vella, F. (2010) Estimating a class of triangular simultaneous equations models without exclusion restrictions. Journal of Econometrics, 154(2), 154-164.This paper provides a control function estimator to adjust for endogeneity in the triangular simultaneous equations model where there are no available exclusion restrictions to generate suitable instruments. Our approach is to exploit the dependence of the errors on exogenous variables (e.g., heteroscedasticity) to adjust the conventional control function estimator. The form of the error dependence on the exogenous variables is subject to restrictions, but is not parametrically specified. In addition to providing the estimator and deriving its large-sample properties, we present simulation evidence which indicates the estimator works well.

Jeanty, P. W., Partridge, M., & Irwin, E. (2010) Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Regional Science and Urban Economics, 40(5), 343-352. Identifying the local interactions between housing prices and population migration is complicated by their simultaneous and spatially interdependent relationship. Higher housing prices may repel households and push them into neighboring areas, suggesting that separately identifying interactions within versus across local neighborhoods is important. Aggregate data and standard econometric models are unable to address the multiple identification problems that may arise from the simultaneity, spatial interaction, and unobserved spatial autocorrelation. Such problems can generate biased estimates that run counter to economic theory. Using Michigan census tract-level data, we estimate a spatial simultaneous equations model that jointly considers population change and housing values, while also explicitly modeling interactions within neighborhoods.

Chang (2003) examined simultaneous causality between ownership structure and performance for firms affiliated to groups in Korea. He found that in most cases performance determines ownership structure and not vice-versa. To control for endogeneity, he used an instrumental variable approach and 2SLS estimation.

Imbens and Newey (2002) investigated identification and inference in a nonparametric structural model with instrumental variables and non-additive errors. They formulated several independence and monotonicity conditions that are sufficient for identification of a number of objects of interest, including the average conditional response, the average structural function, as well as the full structural response function. For inference they have proposed a two-step series estimator. The first step consisted of estimating the conditional distribution of the endogenous regressor given the instrument. In the second step the estimated conditional distribution function was used as a regressor in a nonlinear control function approach. They established rates of convergence, asymptotic normality, and gave a consistent asymptotic variance estimator.

Bollen (2001) explored briefly the conditions under which this 2SLS estimator is robust to model misspecification such as omitted paths or omitted variables.

Chowdhury (1988) has suggested that there is a dual causality between infant mortality rates and fertility rates. He believed that when a woman has multiple pregnancies, the chances of her child's survival are significantly reduced. A woman may thus decide to bear more children in the hope that at least some will alive.

3. SIMULTANEOUS EQUATIONS MODEL

The phrase "Simultaneous Equation Econometric Model" is generally understood to refer to a stochastic model that enables a researcher to make probabilistic statements about a set of random variables, sometimes known as "endogenous variables." The variables that are introduced into an equation system can be split into two categories: endogenous and exogenous. Exogenous variables are those variables that can be taken as given for the purpose of explaining endogenous variables, while endogenous variables are those variables that must be described by the equation system. In each given model, there is the possibility that current as well as lagging endogenous and exogenous variables will be included. Because they are going to be described by the equation system, the present endogenous variables can be referred to as "jointly dependent," whereas the rest of the variables, such as current and lagged exogenous variables, might be referred to as "predetermined" variables.

A model based on simultaneous equations is referred to as a system of simultaneous equations. This model is used to describe a system that describes the combined dependency of various variables. The number of equations that these types of models contain is proportional to the number of jointly dependent or endogenous variables that are involved in the process that is being analyzed. The term "complete" refers to an equation system that has the same number of equations regardless of the number of jointly dependant variables it contains.

The values of one set of variables, known as the endogenous variables, are decided by a model that uses simultaneous equations based on the values of another set of variables, known as the predefined variables. The model of linear simultaneous equations that includes n observations on g endogenous variables and is represented by (n x 1) vectors Y1, Y2,...... Yg, the k preset variables (which can be either exogenous or lagged endogenous), which are represented by the (n x 1) vectors X1, X2,...........Xk, and the g random error variables, which are represented by the (n x 1) vectors 1,2,..........., are all taken into account. It is possible to express g in structural form by writing it as the simultaneous equations for g.

$$\underset{1xg}{Y_i}\underset{gxg}{B} = \underset{1xk}{x_i}\ \underset{kxg}{\Gamma} + \underset{1\,x\,g}{\varepsilon_i}$$

For i = 1, 2, …, n

Where B is a matrix of the structural parameters of endogenous variables that is g by g, is a matrix of the structural parameters of preset variables that is k by g, and n is the size of the sample (no. of observations).

Issues Of Simultaneous-Equations Models

Models that involve simultaneous equations give rise to three separate issues. They are as follows:

- **Comprehensiveness of the Model in Mathematical Terms**

Only when a model contains the same number of independent equations as it does endogenous variables can we say that the model is technically complete. In other words, in order to identify the value of the disturbance terms, the exogenous variables, and the structural parameters, it is necessary to uniquely determine all of the endogenous variables.

- **Detailed Description of Each Equation Included Within the Model**

The issue at hand is one of identifying the parameters of each individual equation in the system. There are many instances in which a predetermined collection of values for the disturbance terms and exogenous variables leads to the generation of the same values for the various endogenous variables that are incorporated into the model. This is due to the fact that the equations cannot be distinguished based on the observations. It is essential that the parameters of each equation in the system be determined in a way that is completely unique. As a result, particular tests are necessary in order to investigate the identification of each equation prior to its estimation.

- **Estimate of the Statistical Significance of Each Equation in the Model**

Because the application of OLS results in estimates that are skewed and inconsistent when applied to systems containing simultaneous equations, several statistical methods will need to be developed in order to estimate the structural parameters.

4. INACCURATE SIMULTANEOUS EQUATIONS

There are several contexts in which the assumption of unidirectional causality in a function is meaningless. This is the case if the variable Y, which is the dependent variable, is a function of both the variable X, which is the explanatory variable, and the variable Y itself. Because of this, there is a two-way flow of influence that occurs between Y and (some of) the X; in other words, the relationship that exists between the variables is a bidirectional one. In such a situation, it is necessary for us to take into consideration more than one regression equation; one for each interdependent variable. This will allow us to comprehend the complex web of interdependent relationships between the variables. In models with simultaneous equations, it is not possible to estimate a single equation of the model without taking into account the information provided by other equations of the system. This is in contrast to models with a single equation, which allow for such estimations.

In this scenario, employing linear regression with ordinary least squares (OLS) does not produce optimal model estimates and violates an assumption of OLS, which states that E (X) must be less than 0. Not only are the estimations that were derived in this manner skewed, but they are also inconsistent. In other words, the estimators do not converge to their correct (parameter) values even when the sample size continues to expand endlessly.

The error that results from using an estimation method that treats each equation of a simultaneous equations model as though it were a single-equation model is referred to as simultaneity bias or simultaneous equation bias. This error can occur when using an estimation method that treats each equation of a simultaneous equations model as though it were a single-equation model.

5. SOLUTION APPROACHES FOR SYSTEMS OF SIMULTANEOUS EQUATIONS

Model of Structure:

In economics, a structural model is a set of equations that fully characterizes the network of interrelationships between all relevant variables. The model's structural equations write the endogenous variables in terms of each other, fixed factors, and perturbations (random variables).

Each explanatory variable's effect on the dependent variable is quantified by a set of structural parameters. The structural system must be solved in order to calculate indirect impacts; the structural parameters themselves are not sufficient. The structural parameters are often denoted by 's for endogenous variables and by 's for fixed ones. Y-values denote the endogenous variables while x-values denote the exogenous ones.

Condensed Model:

Endogenous variables are expressed as a function of exogenous ones in the reduced form model. There are two routes to achieving the reduced form.

The first is to write the endogenous variables off as a straight function of the exogenous ones.

$$\underset{nxg}{Y} = \underset{nxk}{X} \ \underset{kxg}{\Pi} + \underset{nxg}{U}$$

proceed with the estimation of the ∏'s by applying some appropriate technique to this expression.

The reduced form is derived by post multiplying the structural form by B^{-1}, where $\Pi = \Gamma B^{-1}$ is the reduced form coefficient matrix and $U = EB^{-1}$ is the reduced form disturbance vector.

The second technique to get the reduced form of a model is to solve the structural system of endogenous variables in terms of the preset variables, the structural parameters, and the disturbances. This is the way to get the reduced form. For the two reduced forms to be consistent the relationships between the ∏'s and the structural parameters must hold.

The reduced-form parameters assess the whole effect, including both the direct and the indirect influence, that a change in the predefined variable has on the endogenous variables.

Recursive Model:

Recursive model is an important type of simultaneous equations system in which the endogenous variables and structural equations can be arranged in such an order that, B the matrix of coefficients of endogenous variables, is a triangular matrix and Σ, matrix of variances and covariance of stochastic disturbance terms, is a diagonal matrix.

Thus, in this case.

$$B = \begin{bmatrix} 1 & 0 & 0 & \cdots 0 \\ \beta_{21} & 1 & 0 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{g1} & \beta_{g2} & \beta_{g3} & \cdots 1 \end{bmatrix}$$

$$\sum = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \sigma_g^2 \end{pmatrix}$$

There is another name for recursive systems called triangular systems. The first group of conditions, which pertain to the coefficient matrix, stipulates that the structural equations must be able to be expressed in such a way that every endogenous variable must be able to be rationalized in terms of the predetermined variables, the stochastic disturbance term, and the endogenous variables with lower numbers. i.e.

$y_1 = \Upsilon_{11} x_1 + \Upsilon_{12} x_2 + \ldots\ldots\ldots\ldots + \Upsilon_{1k} x_k + u_1$

$y_2 = \Upsilon_{21} x_1 + \Upsilon_{22} x_2 + \ldots\ldots\ldots\ldots + \Upsilon_{2k} x_k + \beta_{21} y_1 + u_2$

$y_3 = \Upsilon_{31} x_1 + \Upsilon_{32} x_2 + \ldots\ldots\ldots\ldots + \Upsilon_{3k} x_k + \beta_{31} y_1 + \beta_{32} y_2 + u_3$

.

.

.

$y_g = \Upsilon_{g1} x_1 + \Upsilon_{g2} x_2 + \ldots\ldots + \Upsilon_{gk} x_k + \beta_{g1} y_1 + \beta_{g2} y_2 + \ldots\ldots + u_g$

The second set of criteria, which concerns the covariance matrix, stipulates that it is necessary for any and all covariances between the stochastic disturbance terms contained inside any two distinct equations to be eliminated. Because of this requirement, it is guaranteed that the contemporaneous stochastic disturbance terms are not connected with one another. In a recursive system, OLS can be used to estimate the equations one at a time without introducing simultaneous equations bias, which means the estimates will be unbiased and consistent.

6. CONCLUSION

The only way to determine whether or not a system of structural equations has been recognized is to determine each and every equation contained inside the system. If even a single equation is missing from the equation list, the system cannot be identified. After identifying the system, the assessment of its parameters is carried out with the assistance of the information that is now accessible. The empirical data and the structure of the model both play a role in the estimating process for the model. Even if sufficient or pertinent data are available, it is possible that the parameters cannot be uniquely estimated if the model is not in the correct statistical form. This may be the case even if the data are appropriate. For the parameters of a structural equation in a simultaneous equation, the least squares estimator is biased and inconsistent; hence, numerous approaches have been presented to estimate the parameters of a set of simultaneous equations in a manner that is consistent.

REFERENCES

1. Franz, J. and FitzRoy, F. (2006). Child Mortality, Poverty and Environment in Developing Countries.
2. Klein, R., & Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, *154*(2), 154-164.

3. Jeanty, P. W., Partridge, M., & Irwin, E. (2010). Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. *Regional Science and Urban Economics*, 40(5), 343-352.
4. Change, S.J. (2003). Ownership Structure, Expropriation, and Performance of Group affiliated Companies in Korea. *Academy of Management Journal*, 45(2): 238-253.
5. Anderson, T.W. (1984). Estimating Linear Statistical Relationships. *Annals of Statistics*, 12:1-45.
6. Anderson, T.W. and Sawa, T. (1979). Evaluation of the Distribution Function of Two-Stage Least Squares Estimates. *Econometrica*, 47:163-182.
7. Anker, R. (1978). An analysis of fertility differentials in developing countries. *The Review of Economics and Statistics*, 60(1): 58-69.
8. Bhattacharya, B., Singh, K.K. and Singh, U. (1995). Proximate Determinants of Fertility in Eastern Uttar Pradesh. *Human Biology*, 67: 867-886.
9. Biswas, B. and Ram, R. (1982). Simultaneous Equations Analysis of Fertility in the U.S.: A Comment; *Econometric*a, 50(6): 1585-1590.
10. Boca, D.D. and Sauer, R.M. (2006). Life Cycle Employment and Fertility across Institutional Environments. *IZA Discussion Papers*, 2285, Institute for the study of labour (IZA).
11. Bollen, K.A. (2001). Two-Stage Least Squares and Latent Variable Models: Simultaneous Estimation and Robustness to Misspecifications. *Structural Equation Modeling*; 99-138.
12. Bollen, K.A. and Paxton, P. (1998). Interactions of Latent Variables in Structural Equation Models. *Structural Equation Modeling*, 5(3): 267-293.
13. Chaudhary, R.H. (1996). Factors Affecting Variations in Fertility by States of India: A Preliminary Investigation. *Asia-Pacific Population Journal*, 11(2): 59-68.
14. Chowdhury, A.R. (1988). The Infant Mortality-Fertility Debate: Some International Evidence. *Southern Economic Journal*, 54: 666-74.
15. Christ, C.F. (1960). Simultaneous Equation Estimation: Any Verdict Yet? *Econometrica*, 28: 835-845.
16. Cragg, J.G. (1966). On the Sensitivity of Simultaneous Equations Estimators to the Stochastic Assumptions of the Models. *Journal of American Statistical Association*, 136-151.
17. Goldberger, A.S. (1964). *Econometric Theory*. New York: John Wiley and Sons.