BEBR

FACULTY WORKING
PAPER NO. 966

Forecasting First Births

Walter J. Primeaux, Jr.
Donald E. Pursell
Charles L. Bare

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 966

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

July 1983

Forecasting First Births

Walter J. Primeaux, Jr., Professor
Department of Business Administration

Donald E. Pursell
University of Nebraska-Lincoln

Charles L. Bare
University of Nebraska-Lincoln

Abstract

Forecasting First Births

The primary objective of this investigation is the development of models to forecast annual first births. A couple's first birth is a significant social and economic event, signaling changes in consumption and saving patterns. It is anticipated that the results of this study will be of value to marketing research and of interest to practitioners and scholars concerned with the economic, demographic, and social impact associated with first births.

Economic, demographic, and social variables associated with first births are identified. Models are estimated for the United States and California (to provide another reference point), with an estimation period of 1950-1979 (1950-1978 for California). Tests are conducted on the models to see how well they track existing data over an ex post period 1975-1979 (1974-1978 for California).

Results indicate that first births can be accurately forecast using simple statistical techniques. Firms marketing to this group can develop short- to intermediate-term forecasts with minimal expenditures of time and resources.

It is generally accepted that consumers pass through a series of stages known as the "family life cycle." This concept is discussed in an article by Wells and Gubar, who identify the stages as follows: (1) bachelor stage (young single individuals); (2) young married couples (no children); (3) full nest (young married couples with children); (4) full nest 1 (youngest child under 6); (5) full nest 2 (youngest child 6 or over); (6) full nest 3 (older married couples with dependent children); (7) empty nest (older married couples with no children at home); (8) solitary survivors (older single or widowed people). Since consumer spending patterns pass through distinct life cycle stages, there are compelling pragmatic justifications for identifying life cycles as an effective business tool.

This investigation is concerned with entry into Stage 3 of the family life cycle: young married couples with dependent children. Passage into Stage 3 is denoted by the birth of the couple's first child. Entry into Stage 3 was selected for examination since entry into this stage has a substantial impact on consumption patterns and buyer attitudes. Stage 3 is a pivotal stage where significant transitions in consumer needs occur. Consequently, development of a model to forecast the number of consumer entries into Stage 3 of the family life cycle would be of value to marketing research and of interest to scholars and practitioners concerned with the economic, demographic, and social impacts associated with entry into this stage of the family life cycle.

The Model

Based on the assumption that economic and demographic variables are primary factors associated with first births, the following regression model was specified:

$$Y_t = B_0 + B_1 X_{1t} + \dots + B_k X_{kt} + \varepsilon_t, \quad t = 1, 2, \dots, T$$

where

Y_t is the number of first births,

$X_{1t}, \dots, X_{kt}$ are economic and demographic variables,

$B_0, B_1, \dots, B_k$ are regression parameters, and

$\varepsilon_t \sim N(0, \sigma^2)$.

The number of first births was selected as the dependent variable since it indicates the number of families entering Stage 3.

Using ordinary least squares, forecasting models were estimated for (1) total number of first births, (2) number of white first births, and (3) number of nonwhite first births. In order to compare model performance at different geographic levels and provide additional test points, models were estimated for the United States and California. First births models for other countries were not estimated because of the difficulty in acquiring the necessary data. In some cases, a complete data series for postulated variables is unavailable in other countries. Models for states other than California were not estimated since the procedure would be repetitive.

Latest available data on births were used to estimate the national and state models. All data are annual and the beginning of the estimation period is 1950. The final year of the estimation period (latest available data) is 1979 for the United States and 1978 for California.

Possible economic variables influencing first births include real personal income and female labor force participation. Demographic variables include population size and number of marriages. Other variables considered were war, contraceptive technology, legalization of abortion, (all quantified as dummy variables) and college degrees granted. See Appendix I for a list of variables and data sources.

Real personal income is a variable that should exhibit a strong association with first births. However, it is difficult to determine, a priori, what sign to expect from income. As births are influenced by economic activity, a positive coefficient would be expected. On the other hand, the cost -- including the opportunity cost -- of having children is rising as more and more females are working. Births require a sacrifice of income during and after pregnancy. This sacrifice is increasing because the number of working females has increased and, more than ever before, females are occupying positions of importance and responsibility. It could be hypothesized that the declining birth rates of the 1960s and 1970s (despite the ever-increasing size of the female cohort 20-29) reflects this higher opportunity cost of having a child. Based on these arguments, one might expect a negative coefficient to be associated with real income.

Female labor force participation is a variable which could affect first births. The female labor force increased from 18.6 million in 1950 to 56.6

million in 1981. One might expect, over the long-run, a negative association between first births and female labor force participation, since a certain proportion of working women might be inclined to postpone, perhaps indefinitely, having children. However, working females increase the level of household income, and this could alter a family's perception of their ability to afford children. The child care industry has expanded more or less simultaneously with the growth in female labor force participation rates, a phenomenon which aids female employment after the birth of children. These conflicting forces make it difficult to determine, *a priori*, the sign of the coefficient.

The size of the population is a demographic variable that could affect the number of first births. A positive association between population size and first births is expected. Population size can be defined as total population, total resident population, or total civilian population. In addition to these definitions, various subsets of population might be identified as the more crucial cohort. For instance, female population ages 15-44 might be the populaton cohort most likely to influence first births. A case could be made for using the female population ages 20-29, since this group has higher fertility rates than any other group, and most first births will occur within this cohort. With younger couples postponing fertility, others might argue that the relevant female cohort is the 25-34 age group. All these variants of population were considered in estimating the model.

Contraceptive technology is a variable which must be taken into account when attempting to estimate a first-births model. Dramatic changes in birth control techniques have occurred during the recent past.

Difficulty in obtaining quantitative data on these methods have led investigators to create a dummy variable to quantify birth control technology. This variable was assigned a value of 0 for the years before the "pill" was readily available, and was assigned a value of 1 for those years when the "pill" was mass-marketed (1960 and after). Naturally, the "pill" variable would be expected to have a negative association with first births.

An increase in the number of abortions is bound to have a negative impact on first births. Since abortion was legalized by the U.S. Supreme Court in 1973, this variable was assigned a value of 0 for the years before 1973, and was assigned a value of 1 for the years thereafter.

Marriage is a demographic variable entered into the modeling process to make allowances for the impact of this institution upon first births. One would expect an increase in the number of marriages to be associated with an increase in the number of first births.

College degrees earned was considered as a variable for explaining the number of first births. This variable contains both economic and demographic elements. A number of hypotheses could be developed to explain the association between degrees earned and first births. An increase in the number of degrees earned, it could be argued, would lead to an increase in the number of first births, since the economic well-being of the population might be somewhat improved. On the other hand, an increase in the number of females earning degrees might decrease the number of first births, since this might alter career aspirations of females. These conflicting forces make it difficult to assign, _a priori_ a sign to the coefficient of this variable.

Births are influenced by prevailing demographic-social trends. If it is "fashionable" to have a large (small) family, the prevailing attitude, in all likelihood, affects births. To quantify fashion in births, a lagged birth variable was created (L_t - total first births, L_w - first white births, and L_{nw} - first nonwhite births). This variable represents the average of births in the previous two years.

Finally, it is conceivable that the number of first births could be substantially affected by war, a condition that depresses the number of marriages and might cause families to have fewer children. War introduces uncertainty, and this is likely to have an impact on first births. War was quantified as a dummy variable, with the years 1950-53 and 1965-70 being assigned a value of 1, and the remaining years assigned a value of 0.

Results

Forecasting equations for first births were estimated using ordinary least squares. Each equation is accompanied by a set of summary measures to evaluate the statistical validity of the equation. The Durbin-Watson statistic was used to test for autocorrelation of the error terms. Numbers in parentheses under the coefficients are t-statistics.

First Births - Total

Exhibit 1 displays the estimated models for first births. The adjusted coefficient of determination (R_a^2) indicates that approximately 94 percent of the variation in the number of total first births is explained by the variables in equation (1). At the .01 level, the Durbin-Watson statistic supports the hypothesis of no autocorrelation in the error terms.

The equation reveals that marriage, female labor force participation, and the dummy variables for the "pill" and the legalization of abortion significantly influence total first births. Marriage had the expected positive impact, which is not surprising since marriage is the socially accepted prerequisite to parenthood. Female labor force participation is positively related to total first births. This coefficient may signal that increased income is associated with more first births; it may also reflect changing social patterns with more females working (or looking for work). Coefficients for the marriage and labor force variables were significant at or above the .01 level. Coefficients of the "pill" and the legalization of abortion were significant at the .01 level, supporting the contention that these social changes have had a significant impact on first births. Improvements in birth control technology and the legalization of abortion have apparently resulted in fewer first births since the coefficients of these variables were negative.

College degrees earned, lagged first births, and the war dummy variable failed to meet the requirement of statistical significance. Population size measures, real personal income, and female labor force participation were so highly correlated that only one of these variables could effectively enter the equation. Of these three variables, female labor force participation yielded the best results in terms of significance and fit.

First Births - White

The estimated model for white first births is given by equation (2) of Exhibit I. Summary statistics for the equation indicate that approximately 88 percent of the variation in the number of white first births is explained by the variables in the equation. The Durbin-Watson statistic supports the hypothesis of no autocorrelation in the error terms.

Interestingly enough, the same variables which were significant in the total first births equation were significant in the white first births equation. Signs and levels of significance of the coefficients were approximately the same in both equations.

First Births - Nonwhite

The estimated regression equation for nonwhite first births is different from equations (1) and (2) in that the "pill" variable was ineffective in explaining nonwhite first births. The coefficient of determination indicates that the variables in equation (3) explain 99 percent of the variation in first births for the nonwhite population. The results for nonwhite first births -- equation (3) -- were substantially improved over the results for total or white first births. The Durbin-Watson statistic is inconclusive with respect to autocorrelation of the error terms.

As expected, marriage was highly significant with a positive coefficient. An increase in the number of nonwhite marriages will lead to an increase in the number of nonwhite first births. Female labor force participation had the positive association evident in equations (1) and (2), and the coefficients were significant at or above the .01 level. Legalization of abortion exhibited the expected negative association with first births, and was significant at the .05 level.

State Model

To compare model performance at national and state level, first birth forecasting models were estimated for California. The rationale for this particular selection was previously stated. Data at the state level were collected for first births, marriages, and personal income. Female labor force participation rates for the United States were used as a proxy for state-level rates as these data are difficult, if not impossible, to obtain.

The structure of the estimated state forecasting model for first births turned out to be similar to the structure of the models for the nation as a whole. The equations of Exhibit 2 indicate that labor force participation, marriages, and the legalization of abortion significantly influence total and white first births. Real personal income, lagged first births, war, and the "pill" variable are absent since they did not meet the requirement of statistical significance. Explanatory variables for nonwhite first births are marriages and female labor force participation. The "pill" variable, legalization of abortion, war, lagged first births, and real income were ineffective in explaining nonwhite first births.

In all models, the coefficients were significant at or above the .05 level. Tests for autocorrelation of the error terms indicated no autocorrelation present for the total and white first births models, and the test was inconclusive for the nonwhite first births model.

Test of the Models

Exhibit 3 presents the results of evaluations conducted on first births models discussed in the preceding section. Coefficients for first births models were re-estimated for the period 1950-1974 (see Appendix 2). Forecasts were generated with these new equations for an ex post period 1975-1979, and compared to existing data. With the California model, an estimation period of 1950-1973 was used, and an ex post period was 1974-1978. Mean absolute percentage errors (MAPE) were computed for each model for the purpose of comparative evaluation. Ex post forecasts, existing data, and percentage errors are displayed for the models (Exhibit 3).

Among the national models, nonwhite first births was the most accurate with a mean absolute percentage error of 2.17 percent. Examination of the data in Exhibit 3 reveals that forecasts generated with the nonwhite first births model closely tracks existing data in the ex post period. Mean absolute percentage errors for the total and white first births models were 3.73 percent and 5.18 percent, respectively. Mean absolute percentage errors for the California model reveal a lesser degree of success in tracking existing data than that achieved by the national models.

Summary statistics for models estimated with full data are "better" than the re-estimated models in terms of fit (R_a^2), significance of coefficients, and autocorrelation. If the re-estimated model can closely track existing data in an ex post period, then the full-data model should provide accurate forecasts. These observations confirm our confidence in the model's value.

Exhibit I

First Births Forecasting Models for the United States

First Births - Total

$$FBTOT = -161,000 + \underset{(6.64)}{.292} * MARRY + \underset{(6.37)}{23,100} * LFPRF$$

$$\underset{(-5.10)}{-94,200} * PILL - \underset{(-6.01)}{144,000} * ABORT \qquad (1)$$

$$R_a^2 = .944 \qquad DW = 2.15$$

First Births - White

$$FBW = 70,200 + \underset{(5.29)}{211} * MARRY + \underset{(4.98)}{16,400} * LFPRF$$

$$\underset{(-5.30)}{-89,100} * PILL - \underset{(-5.96)}{130,000} * ABORT \qquad (2)$$

$$R_a^2 = .884 \qquad DW = 2.02$$

First Births - Nonwhite

$$FBNW = -214,000 + \underset{(8.72)}{78.8} * MARRY + \underset{(9.27)}{6,260} * LFPRF$$

$$\underset{(-2.52)}{-11,400} * ABORT \qquad (3)$$

$$R_a^2 = .985 \qquad DW = 1.28$$

FBTOT - total first births, units
FBW - white first births, units
FBNW - nonwhite first births, units
MARRY - total marriages, units
LFPRF - female labor force participation rate, percent
PILL - dummy variable for birth control
ABORT - dummy variable for the legalization of abortion
t statistics in parentheses

Exhibit 2

First Births Models for California

First Births - Total

$$FBTOT = -92,800 + \underset{(2.70)}{.124} * MARRY + \underset{(9.88)}{4,930} * LFPRF$$

$$\underset{(-5.86)}{-16,400} * ABORT \qquad (4)$$

$R_a^2 = .979$ DW = 1.77

First Births - White

$$FBW = -49,900 + \underset{(2.67)}{.111} * MARRY + \underset{(7.86)}{3,540} * LFPRF$$

$$\underset{(-6.32)}{-16,000} * ABORT \qquad (5)$$

$R_a^2 = .966$ DW = 1.88

First Births - Nonwhite

$$FBNW = -41,100 + \underset{(2.67)}{.0170} * MARRY + \underset{(26.14)}{1,320} * LFPRF \qquad (6)$$

$R_a^2 = .994$ DW = 1.16

FBTOT - total first births, units
FBW - white first births, units
FBNW - nonwhite first births, units
MARRY - total marriages, units
LFPRF - U.S. female labor force participation rate, percent
ABORT - dummy variable for the legalization of abortion
t statistics in parentheses

Exhibit 3

Mean Absolute Percentage Errors for First Births
Models of the United States and California

United States

	Total First Births		White First Births		Nonwhite First Births	
Year	Forecasts	Actual Data*	Forecasts	Actual Data*	Forecasts	Actual Data*
1975	1,318,200	1,319,126	1,031,800	1,075,597	234,590	243,529
1976	1,309,400	1,324,811	1,040,800	1,083,218	241,410	241,593
1977	1,317,700	1,387,143	1,056,500	1,134,062	250,500	253,081
1978	1,399,200	1,401,491	1,098,600	1,143,073	269,060	258,418
1979	1,416,300	1,479,260	1,119,600	1,205,958	278,780	273,302
	MAPE = 3.73%		MAPE = 5.18%		MAPE = 2.17%	

California

	Total First Births		White First Births		Nonwhite First Births	
Year	Forecasts	Actual Data*	Forecasts	Actual Data*	Forecasts	Actual Data*
1974	130,420	134,854	109,800	113,132	20,625	21,722
1975	130,310	138,141	109,490	115,566	20,810	22,575
1976	132,140	144,422	110,630	120,184	21,495	24,238
1977	134,670	151,818	112,310	125,406	22,345	26,412
1978	151,250	156,887	125,250	128,915	25,984	27,972
	MAPE = 6.47%		MAPE = 5.89%		MAPE = 9.34%	

*Source: Vital Statistics of the United States (various volumes) and Monthly Vital Statistics Report, Vol. 31, No. 2, Supplement (2), May 27, 1982.

Appendix I

Variables and Sources of Data

Dependent Variables

Variable		Source of Data
First Births		
FBTOT	First births - total	U.S. Department of Commerce, Bureau of the Census, Vital Statistics of the United States, 1950-1978 and Monthly Vital Statistics Report (May 27, 1982).
FBW	First births - white	
FBNW	First births - nonwhite	

Independent Variables

Variable		Source of Data
College Degree Variables		
DGSTOT	Degrees - male and female	Earned degrees conferred as given by the U.S. Department of Commerce, Bureau of the Census, Statistical Abstract of the United States, various issues.
DGSBM	Degrees - bachelor - male	
DGSBMM	Degrees - bachelor, masters - male	
DGSBF	Degrees - bachelor - female	
DGSBMF	Degrees - bachelor, masters - female	
Population Variables		
POPTOT	Population - total, resident civilian	U.S. Department of Commerce Bureau of the Census, Current Population Reports, Series P-25, Nos. 310, 519, 614, 917.
POPW	Population - white, resident civilian	
POPNW	Population - nonwhite	
W,NW1519, 2024, 2529, 3034, 3544, etc.	Female population - white and nonwhite, 15-19, 20-24, 25-29, 30-34, 35-44, etc.	

Other Variables

LFPRF	Female labor force participation rate	Economic Report of the President (February, 1982).
PI72	Real personal income	U.S. Department of Commerce, Bureau of Economic Analysis.
PILL	Birth control pill dummy variable	Chemical Week 106, No. 13 (April 1, 1970).
ABORT	Legalization of abortion dummy variable	See text.
WAR	War dummy variable	See text.
MARRY	Number of marriages	U.S. Department of Commerce, Bureau of the Census, Vital Statistics of the United States, 1950-1978 and Statistical Abstract of the United States, various issues.

Appendix 2

Re-estimated First Births Models for the United States Using the Estimation Period 1950-1975

$$FBTOT = -17{,}000 + \underset{(6.20)}{345} * MARRY + \underset{(2.94)}{16{,}800} * LFPRF - \underset{(-3.80)}{81{,}200} * PILL$$

$$\underset{(-5.95)}{-153{,}000} * ABORT \qquad (7)$$

$R_a^2 = .922$ DW = 2.34

$$FBW = 256{,}000 + \underset{(5.55)}{273} * MARRY + \underset{(1.69)}{8{,}530} * LFPRF - \underset{(-3.79)}{71{,}600} * PILL$$

$$\underset{(-5.98)}{-136{,}000} * ABORT \qquad (8)$$

$R_a^2 = .855$ DW = 2.39

$$FBNW = -225{,}000 + \underset{(6.77)}{75.9} * MARRY + \underset{(7.29)}{6{,}670} * LFPRF$$

$$\underset{(-2.66)}{-13{,}100} * ABORT \qquad (9)$$

$R_a^2 = .980$ DW = 1.21

References

Boone, L. E. and Kurtz, D. L., Contemporary Marketing, Hinsdale, Ill.: Dryden Press, 1974, 99.

Durbin, J., "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," Econometrica, May 1970, 422-429.

Engel, J. F., Kollat, D. T., and Blackwell, R. D., Consumer Behavior, New York: Holt, Rinehart, & Winston, 1973, 193-194.

Wells, W. D. and Gubor, G., "The Life Cycle Concept in Marketing Research," Journal of Marketing Research 3, November 1966, 366-373.

Pindyck, R. S. and Rubinfeld, D. L., Econometric Models and Economic Forecasts, New York: McGraw-Hill, 1976, 156-159.