Predictors in Ideal Number of Children Using General Social Survey 2014: an Exploratory Study

Introduction

In October 2018 the Centers for Disease Control and Prevention (CDC) released a report on the national fertility rate trend. In it, the CDC highlighted that the United States is experiencing a decrease in total fertility rates, with significant drops in several key demographic groups that had shown consistent high fertility rates previously.1 For example, compared to other ethnic groups, Hispanic women showed the largest decline of 26%, 29%, and 30% in rural, suburban, and metro counties respectively. Birthrate also fell for women 15 to 19 years old, which is a 55% decrease since 2007 and a 70% decrease since 1991, when the most recent peak happened.1,2

This is not a surprising trend though, as many high-income countries are struggling with non-sustainable fertility rates.2 Industrialized countries tend to have better healthcare, more freedom for lifestyle choices, economic support, and birth control resources, among other factors that lower fertility rates. The normalization of higher education and professional careers for women also delay a decision to bring a child to life.3,4 However, women are key in maintaining a population. The concept of replacement fertility rate was devised to describe “a total fertility rate in which women would have only enough children to replace themselves and their partners.” This number is roughly 2.1 births per woman for high-income countries like the U.S. to keep a stable population.3 We are, however, sitting at around 1.95 for rural, 1.78 for suburban, and 1.71 for metro counties as of the latest reported data released by the CDC.1

Even though this is a normal phenomenon in high-income countries, a decreasing fertility rate does pose many problems. In fact, as a fundamental metric for a country, fertility rates have profound impacts. Fertility closely indicates economic health, as the proportion of youth in a population affects employment and worker force. This in turn relates to whether society can pay taxes and keep the government’s ability to provide key resources like law enforcement and healthcare.5 The emergence of different preferences for various age groups also affect consumption behaviours, which influences market, careers, and livelihood for all of us. Ethnic group fertility rates influence demographic proportions and even voting preferences for communities when elections are at an increasingly higher stake.6 Fertility is at the core of each individual’s motivation for life, as highlighted by MaRgolis et al., a negative to neutral to positive association was found between fertility and happiness as the age of respondent increases, regardless of sex, income, health status, welfare, partnership status, and other factors.7 The concept of fertility is therefore complex because of the multi-level multi-faceted influences it exerts.

This study focuses on a core social aspect of fertility – the perspective of an individual on the ideal number of children as a proxy and precursor for the willingness to have children. Several demographic factors are tested, as literature suggest a differential effect for various groups.1 Whether the racial groups or the sexes differ in terms of perception of ideal number of children will be investigated.8 Respondent’s individual income level will act as a predictor for future family carrying capacity, a major decision factor in consideration of having a child.7 Education levels are included to study whether there is an ideological influence to the difference seen in fertility rates.

Furthermore, three other ecological variables are tested to look for correlation to the perception of ideal number of children. The individual’s perception on whether the government is spending too much in childcare will be a lead to test the institution’s influence in an individual’s psyche. A pre-marriage factor – the number of brothers and sisters growing up – is used to gain an idea on the individual’s view on the concept of family. This could also directly indicate a view on the proper family size. In contrast, a post-marriage factor is also included to gauge whether the perceived happiness of marriage had an effect on the ideal number of children. Happiness of marriage can also predict the status of the union, willingness to bear a child with the significant other, and the emotional confidence in future ability to raise a child.7

The research question is, therefore, if there is a statistically significant relationship between ideal number of children, sex, income, education, race, opinion of government’s role in childcare, number of siblings in family, and happiness of current marriage. The null hypothesis (H0) is there is no linear model that can predict ideal number of children based on the seven independent variables. The alternative hypothesis (Ha) is that there is a linear model that can predict ideal number of children based on independent variables.

Methods

Data was retrieved from the 2014 General Social Surveys (GSS), conducted for the National Data Program for the Social Sciences at NORC, University of Chicago. Specific sampling methods, data processing, and data weighting procedures were described in its cumulative codebook released April 2016. Primary data collected in the surveys were collected from a national probability sample. The variables of interest were cleaned by removing entries with not applicable and no answer, leaving 2280 rows in the working data file. One variable (happiness) was recoded in reversal due to the wording of the survey question. The rest of the code were left as is, since all of the categorical variables are properly coded, and two measures in income and siblings had over 10 categories, treating it as a continuous variable.

IBM SPSS version 25 was used to conduct all statistical analysis. Descriptives were explored using both the summarize function as well as histograms to see distribution. To test for linearity between the two continuous variables, scatter plot graphs were generated. To test for normality, a histogram was generated, as well as the histogram of residuals generated as part of the model. P-P plot and a scatter plot of standardized residual versus predicted were generated to further test normality of residual and homoscedasticity, part of the model.

A primary multiple linear regression model (MLR) was done using ideal number of children (ideal) as the dependent variable and sex (sex2), income (income2), degree (degree2), race (race2), opinion on government role in childcare (childcare), the number of siblings in family (siblings), and reported happiness of marriage (happiness) as independent variables. Categorical variables were recoded into dummy variables. Estimates, CI, Durbin-Watson, model fit, case-wise diagnostics, and descriptives were generated. ZPRED over ZRESID plot was generated, as well as standardized residuals, Cook’s distance, standardized DfBetas, and standardized DfFit. The model was assessed using diagnostics to determine the assumptions of MLR. Influential values were assessed using Cook’s distance, ZDfFit, ZDfBeta, and ZResid.

A test for variable interaction was conducted using scatter plot. To confirm, an interaction term between the pre-marriage factor siblings and post-marriage factor happiness was generated and entered into the model.

Results

Starting with 2280 data points from the 2014 GSS dataset, 375 entries were left after data cleaning and recode. The mean and standard deviation for the eight variables (and its dummy variables) are listed in Table 1. Notably, income have a high standard deviation of 5.6 with mean at 16.5 (individual earning approximately $30000 to $40000). The bar graphs for categorical variables are included in Appendix A. Histograms and scatter plots for continuous variables ideal, income, and siblings are also included as part of the data exploration stage. For tests of normality, a visual inspection of the histogram reveals no normality for ideal, income, and siblings. In fact, siblings displays a positively skewed relationship, while the others are multi-modal (Appendix A).

A significant MLR model was constructed to predict ideal using sex2, race2, degree2, income2, childcare, siblings, and happiness [F(15, 359) = 2.873, p<.01] (Table 3). We can reject the null hypothesis because the model explains the data significantly better than the mean. The model explained 10.7% of the variability of ideal. Out of all variables, only graduate degree (degree2=4.0) and being black (race2=2.0) showed a statistically significant relationship (t = -2.226, p<.05; t = 4.200, p<.01 respectively). The slope of the line for graduate degree is negative (standardized Beta = -0.209, 95%CI [-0.839, -0.052]), suggesting a negative linear relationship. For having a graduate degree, the ideal number of children decreases by 0.209. The slope for race being black (standardized Beta = 0.217, 95%CI [0.320, 0.884]) suggest a positive relationship. For being black, the ideal number of children increases by 0.217. All other variables are statistically insignificant (Table 4). The interaction term between siblings and happiness showed no statistically significance.

The proposed regression equation based on this study is as follows:

Y = (2.329 – 0.039xsex + 0.011xincome – 0.200xdegree highschool – 0.259xdegree junior college – 0.284xdegree bachelors – 0.446xdegree graduate + 0.602xrace black + 0.071xrace other + 0.044xchildcare neutral + 0.017xchildcare positive + 0.017xsiblings – 0.008xpretty happy + 0.019xvery happy)

Checking for assumptions, independence of observations is assumed through survey and sampling design. Independence of errors is tested through the Durbin-Watson test (1.910) (Table 3). Linearity is checked through scatter plot. No obvious linear relationship can be seen, but no obvious curve-linear relationship either (Appendix A). Homoscedasticity assumption is passed through the Steensma-Squint test with no visible heteroscedasticity observed (Appendix A). The assumption for normally distributed errors were violated, showing a bimodal distribution (Appendix A). The assumption of collinearity is debatable, as the VIF statistics for the two happiness dummy variables reach 8.5, suggest a possible correlation between these two independent variables (Table 4).

Checking for influential values using Cook’s distance, Standardized Residuals, Standardized DfBetas, Standardized DfFits, and Leverage Value returned no outstanding observations that needed to be eliminated from the model. No observation was found to have undue influence on the model or the sample (Table 5).

Discussion

This analysis was successful in creating a statistically significant model that associates the ideal number of children to eight independent variables: sex, income, degree, race, childcare, siblings, and happiness. Because there are many categorical variables, dummy variables were created to fit the multiple linear regression model. However, only having a graduate degree and being black are separately statistically significant in predicting the ideal number of children.

Through both generating descriptives and assumptions testing, we can conclude that several assumptions were violated, making the model a questionable method in predicting the ideal number of children. Firstly, our dependent is technically a count variable, ordinal at most. We can only fit it as a continuous variable because of the large number of categories. This is also true of the “fake” continuous variables income, and siblings. As shown by tests, they are neither normally distributed nor continuous. This may contribute to a lack of linearity when tested. Secondly, Durbin-Watson test suggested that our independence of error assumption is not violated. A visual test of homoscedasticity reveals no visible heteroscedasticity trend, however, the standard residual plot showed a bimodal or trimodal distribution. This suggests that there is some relationship unaccounted for, even though including an interaction term for step 3 of the MLR model was not significant. Thirdly, no undue influence was found when looking for influential values. This suggest that the sample’s distribution was not significantly affected by any outliers, but a problem in the distribution itself.

Even though two variables were found statistically significant in our modal, its interpretation must be cautiously conducted considering our false continuously variables. The large number of categories as well as the large sample size may be contributors to the statistical significance of our model. Using models that better account for the nature of our variables, significance may be reduced. The author proposes using a more suitable model to investigate the relationship between predictors and the ideal number of children. Seeing the dependent variable is truly a non-normally distributed, count variable, Poisson regression is recommended. Depending on the over dispersion, a negative binomial regression might be considered as a more robust option.

Other limitations include failure to consider economic situation as a predictor of ideal number of children. Literature recognized that increasingly so, the burden of raising a child and supporting the child through life and education affect the choices of many families.3,5 It may be a more direct predictor than some of the variables in the model. However, this study is limited by the questions asked in GSS 2014. Another unaccounted factor in predicting the ideal number of children is faith and religiosity. Adsera highlighted in her 2006 paper that those identified as Conservative Protestants and Catholics report a higher ideal number of children.9 There seems to be a relationship between religiosity and its influences on deciding a life trajectory, influencing perception on the concept of family. These two factors can be studied in more extensive detail for future research to account for better context.

Despite various limitations to our model, this paper provides important preliminary research into possible indicators that associated with the willingness for an individual to have a child. This will be a key step in constructing an ecological framework for birth incentive programs or childcare programs. Based on significant statistical relationships, a program targeting associated factors will likely to be more effective. This will be increasingly important as the United States deal with the decreasing fertility rates. In the near future, policy makers will have to think of effective, non-coercive ways of promoting birth.10 Alternatively, the falling fertility rates and its potential adverse impacts provide a sound supporting argument for creating better immigration policies.

Conclusion

In short, this study found a statistically significant multiple linear regression model to predict the ideal number of children from sex, income, education, race, opinion on government childcare, number of siblings, and happiness of marriage. Only having a graduate degree and being black are statistically significant factors in association with ideal number of children. This may be because of violations to assumptions of normality and a forced continuous relationship. The considerations highlighted above severaly prevents this model from being generalized in other studies. In order to address these, a more specific analytical model may help. Despite these considerations, this paper offers insights on factors that can be considered in addressing the current failing fertility rates in the United States.

Works Cited

1. Centers for Disease Control and Prevention. *Trends in Fertility and Mother’s Age at First Birth Among Rural and Metropolitan Counties: United States, 2007–2017*.; 2018. https://www.cdc.gov/nchs/products/databriefs/db323.htm. Accessed December 13, 2018.

2. Chappell B. U.S. Births Dip To 30-Year Low; Fertility Rate Sinks Further Below Replacement Level. *NPR.org*. https://www.npr.org/sections/thetwo-way/2018/05/17/611898421/u-s-births-falls-to-30-year-low-sending-fertility-rate-to-a-record-low. Published May 17, 2018. Accessed December 13, 2018.

3. Nargund G. Declining birth rate in Developed Countries: A radical policy re-think is required. *Facts Views Vis Obgyn*. 2009;1(3):191-193.

4. Rindfuss RR, Bumpass L, St. John C. Education and Fertility: Implications for the Roles Women Occupy. *American Sociological Review*. 1980;45(3):431-447. doi:10.2307/2095176

5. Cha AE. As U.S. fertility rates collapse, finger-pointing and blame follow. *Washington Post*. https://www.washingtonpost.com/health/2018/10/19/us-fertility-rates-collapse-finger-pointing-blame-follow/. Published October 19, 2018. Accessed December 13, 2018.

6. Sweeney MM, Raley RK. Race, Ethnicity, and the Changing Context of Childbearing in the United States. *Annu Rev Sociol*. 2014;40:539-558. doi:10.1146/annurev-soc-071913-043342

7. MaRgolis R, MyRskylä M. A Global Perspective on Happiness and Fertility. *Popul Dev Rev*. 2011;37(1):29-56.

8. Golmakani N, Fazeli E, Taghipour A, Shakeri MT. Relationship between gender role attitude and fertility rate in women referring to health centers in Mashhad in 2013. *Iran J Nurs Midwifery Res*. 2015;20(2):269-274.

9. Adsera A. Religion and Changes in Family-Size Norms in Developed Countries. *Review of Religious Research*. 2006;47(3):271-286.

10. Searchinger T, Hanson C, Waite R, Lipinski B, Leeson G, Harper S. Achieving Replacement Level Fertility | World Resources Institute. World Resources Institute. https://www.wri.org/publication/achieving-replacement-level-fertility. Published August 2013. Accessed December 13, 2018.