Final Project Proposal

DATA 621: Business Analytics and Data Mining

Daniel Dittenhafer & Justin Hink

March 1, 2016

1 Data Set

The data sets we plan to use originate from United States federal government agencies. The core data comes from the U.S. Department of Health & Human Services Centers for Disease Control and Prevention (HHS, 2016). Related population and economic data sets will come from the U.S. Census Bureau, the U.S. Bureau of Labor Statistics, and potentially others.

1.1 Natality Data

Separate data sets the years 2003-2006, and 2007-2014 were downloaded from the CDC website (HHS, 2009), (HHS, 2016).

Year
Month
Age of Mother
Marital Status
Education
Births

1.2 Census Data

TBD: Data set descriptions

1.3 Unemployment Data

TBD: Data set descriptions

2 Recent Journal Papers

TBD

2.1 Low Fertility at the Turn of the Twenty-First Century

The Annual Review of Sociology included a paper by S. Philip Morgan and Miles G. Taylor regarding recent fertility trends, specifically a shift to lower birth rates as compared to the second half of the twentieth century (Morgan and Miles, 2006).

2.2 Relative Wage Changes and Fertility in the US

The Eastern Economic Journal included a paper by Aliaksandr Amialchuk regarding wage related effects on fertility (Amialchuk, 2013). Across all women, women's education, men's education, men's earnings and metro area were all

found to be significant in age-specific fertility regression. Married women's wages were found to negatively affect fertility at younger ages (20-24), and positively affect fertility at older ages (30-39).

3 Problem Statement & Research Questions

Through an initial search of journals and papers, a unified model of birth count activity for the United States is not readily available. More focused research into migration effects, wage effects, etc do appear but a meta-model is elusive. Using national aggregate data (surveys, statistics, census), how can the number of births be forecast? What factors are significant which could be used as predictors, assuming trends hold (no major cultural shifts)?

4 Statement of Objectives

Using monthly census estimates, unemployment rates, and birth demographic data from the CDC, we will develop a series of regression models to forecast birth counts.

4.1 Variables

We expect to examine and may incorporated the following variables into the model(s) as separate terms:

- Month Census Estimate of United States Population
- Proportion of women age 20-24, 25-29, 30-34, 35-39
- Unemployment Rate
- ...

4.2 Evaluation

We plan to withhold approximately 20% of the historical data to be used in a cross validation step for assessing the performance of the models via mean squared error, along with other standard measures including adjusted R^2 and model complexity (# of variables),

5 References

Amialchuk, A. "Relative Wage Changes and Fertility in the US". In: Eastern Economic Journal 39(2) (2013). DOI: 10.1057/eej.2013.2.

HHS. Natality public-use data 2003-2006 on CDC WONDER Online Database. Mar. 2009. URL: http://wonder.cdc.gov/natality-v2006.html.

— Natality public-use data 2007-2014 on CDC WONDER Online Database. Feb. 2016. URL: http://wonder.cdc.gov/natality-current.html.

Morgan, S. P. and G. T. Miles. Low Fertility at the Turn of the Twenty-First Century. Aug. 2006. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2849172/.