Exploring the Impact of Socio-Economic Factors on Natality Trends Across U.S. Counties

This study examines natality trends across U.S. counties by various factors. The primary goal is to model natality using a Gamma regression model with Generalized Estimating Equations (GEE). By exploring the relationships between socio-economic factors, urbanicity, and temporal trends, this analysis aims to identify the key drivers of birth rates and their variation across demographic groups.

This analysis uses a dataset of natality counts from various counties over 10 years. The data includes information on the number of births, urbanicity (measured in RUCC codes), deprivation measures (measured in ADI), population size, and the year of each observation. The data spans multiple counties, treating each county as a group with repeated observations across different years. Natality is modeled using a Gamma regression model using the Generalized Estimating Equations (GEE) approach to account for clustering and correlation within counties over time.

The mean-variance relationship was assessed by grouping the birth counts by county (FIPS) and calculated the mean and variance of births over the 10-year period. Majority of the people were Black or White race, followed by Asian and Native American. These were then log-transformed and plotted. Along with these plots, a regression of log variance on the log of the mean was performed to evaluate the appropriateness of the Gamma model (see Figures 1 and 2 for these diagnostics). The diagnostic plot for our initial model (only fitting on urbanicity adjusting for the log-transformed population size) along with the regression results indicate the variance increases faster than the mean number of births increases (positive slope, about 1.87) show the Gamma distribution is appropriate for the natality data. This is because the Gamma model accounts for overdispersion (ie. the variance can be larger than the mean). We use GEE to estimate the conditional mean of natality while accounting for correlation or dependence between observations. In our data we see this happening, having repeated measures in different counties and the birth rates within a county might be more similar to each other than birth rates in different counties due to shared demographics.

Using a Gamma regression model fitted using the GEE framework, four main models were run. As mentioned above, the first model is our baseline model, including only urbanicity (RUCC_2013) as the predictor adjusted for the log-transformed population size (logPop). The second model expands upon the first, adding the log-transformed ADI (logADINatRankZ), representing the level of deprivation. This allows us to assess how socio-economic conditions influence natality while controlling for urbanicity and population size. Model three incorporates a linear time trend (year). Finally, a fourth model will consider the interaction between year and both urbanicity and ADI. Score tests will be used to help assess model fit comparison for the nested models. Additionally, (Principal Component Analysis) PCA using 30 Principal Components (also chosen via score tests) will help us get a good visualization of how natality shifts over age groups for different ethnic and racial backgrounds.

The GEE models reveal how ADI, urbanicity, and year influence natality across U.S. counties. In the first model, shown in Figure 3, urbanicity had no significant effect on birth rates ($p = 0.465$) suggesting that urbanicity alone does not substantially affect natality after accounting for other factors. In our second model, shown in Figure 4, while urbanicity had no significant effect on birth rates ($p = 0.325$), the log-transformed ADI was found to be a significant positive predictor of natality (coefficient = 0.0334), indicating that higher socio-economic deprivation is associated with higher birth rates. In the third model, shown in Figure 5, a significant negative time trend in natality was observed (coefficient = -0.0077) along with the significant positive ADI coefficient, suggesting a decline in birth rates over time. Finally, the fourth model in Figure 6 showcases the interaction between year and rurality/deprivation. Here, we saw significant coefficients for both interactions. For ADI, a significant positive coefficient (coefficient = 0.0011) suggests that the effect of deprivation on natality changes over time, slightly growing every year. In other words, more deprived areas may be seeing a larger increase in birth rates compared to less deprived. For urbanicity, we see a significant negative coefficient (-0.0011). This suggests that urban areas may be seeing a larger decrease in birth rates over time compared to rural areas.

After these models were run, score tests assessed the incremental value of adding variables to each model. Comparing the second model to the first, we get a significant p-value, indicating that adding the variable “year” improved model fit. The same conclusion can be made when comparing the third and second model.

The PCA results (Figure 7) offer a deeper understanding of natality trends across ethnic and racial groups. The analysis shows that most groups experience a drop in natality between ages 15-20, with the Asian non-Hispanic group experiencing a substantial decline. Additionally, White non-Hispanic individuals show a significant drop in natality in later years.

The analysis highlights several key factors influencing natality trends across U.S. counties. The positive relationship between deprivation and birth rates suggests that socio-economic disadvantages may lead to higher natality due to limited access to family planning resources, education, and healthcare. The lack of significance for urbanicity in the first model suggests that socio-economic deprivation is a stronger predictor of natality than urban-rural differences when year is not considered. The negative time trend observed in the third model suggests a decline in birth rates over time, potentially reflecting broader shifts in reproductive behavior, such as delayed childbearing and improved access to reproductive health services. The significant interactions in the fourth model indicate that the effect of deprivation on natality has increased over time, while urban areas have seen a more pronounced decline in birth rates.

The PCA analysis provides insights into demographic differences in natality trends, particularly across age groups. The significant drop in natality between ages 15-20 is observed across most groups, with a more pronounced decline for the Asian non-Hispanic group. This could be due to numerous reasons like cultural expectations and family structures or educational and career priorities. The later decline in natality for White non-Hispanic individuals (more than the other demographics) may be attributed to the absence of strong family support systems for older individuals, potentially leading to lower birth rates in later life stages.

This analysis sheds light on the factors that shape natality trends across U.S. counties. It seems clear that socio-economic disadvantage, as indicated by the deprivation index (ADI), is linked to higher birth rates, while urbanicity alone doesn’t seem to have a strong impact on natality. Over time, we observe a general decline in birth rates, but the relationship between deprivation, urbanicity, and time changes depending on the county. Interestingly, our analysis also reveals that different demographic groups experience these trends in distinct ways, as seen with white non-hispanic people in the 30 PC graph. The findings from the Principal Component Analysis give us a closer look at how these dynamics shift over the lifespan and across different racial groups. Overall, these insights provide a deeper understanding of the factors influencing birth rates and can help guide policy decisions aimed at addressing the specific needs of different communities, especially in terms of reproductive health and family planning.

Figures

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    var   R-squared:                       0.779
Model:                            OLS   Adj. R-squared:                  0.779
Method:                 Least Squares   F-statistic:                     2042.
Date:                Thu, 20 Mar 2025   Prob (F-statistic):          7.38e-192
Time:                        21:48:19   Log-Likelihood:                -735.05
No. Observations:                 580   AIC:                             1474.
Df Residuals:                     578   BIC:                             1483.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -5.4898      0.338    -16.231      0.000      -6.154      -4.826
mean           1.8755      0.042     45.192      0.000       1.794       1.957
==============================================================================
Omnibus:                        1.181   Durbin-Watson:                   1.536
Prob(Omnibus):                  0.554   Jarque-Bera (JB):                1.203
Skew:                           0.040   Prob(JB):                        0.548
Kurtosis:                       2.791   Cond. No.                         78.3
==============================================================================
```

Figure 1: OLS Regression results Regressing log variance on log mean.

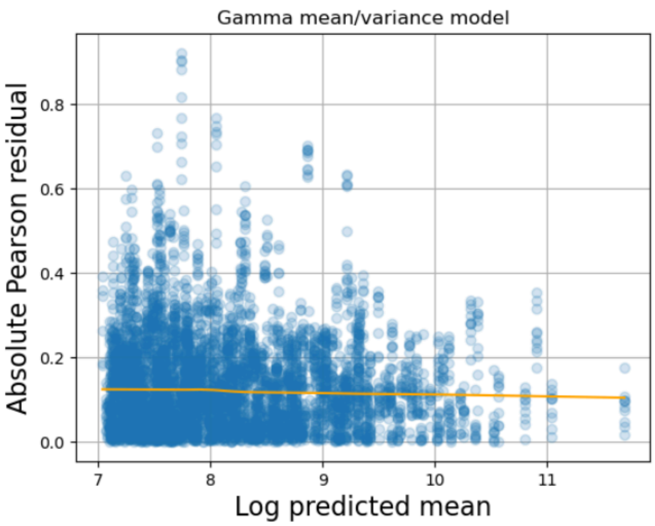

Figure 2: Diagnostic plot for mean/variance relationship with gamma model. The horizontal orange line means good fit.

GEE Regression Results

Dep. Variable:	Births	**No. Observations:**	5538
Model:	GEE	**No. clusters:**	572
Method:	Generalized	**Min. cluster size:**	3
	Estimating Equations	**Max. cluster size:**	10
Family:	Gamma	**Mean cluster size:**	9.7
Dependence structure:	Independence	**Num. iterations:**	2
Date:	Thu, 20 Mar 2025	**Scale:**	0.032
Covariance type:	robust	**Time:**	21:48:23

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-4.4496	0.015	-289.565	0.000	-4.480	-4.419
RUCC_2013	0.0059	0.008	0.731	0.465	-0.010	0.022

Skew:	3.3220	**Kurtosis:**	46.3145
Centered skew:	-8.0432	**Centered kurtosis:**	389.5965

Figure 3: Urbancity (RUCC) predicting Natality (Births)

GEE Regression Results

Dep. Variable:	Births	**No. Observations:**	5538
Model:	GEE	**No. clusters:**	572
Method:	Generalized	**Min. cluster size:**	3
	Estimating Equations	**Max. cluster size:**	10
Family:	Gamma	**Mean cluster size:**	9.7
Dependence structure:	Exchangeable	**Num. iterations:**	6
Date:	Thu, 20 Mar 2025	**Scale:**	0.031
Covariance type:	robust	**Time:**	21:48:25

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-4.4227	0.016	-280.856	0.000	-4.454	-4.392
RUCC_2013	-0.0083	0.008	-0.984	0.325	-0.025	0.008
logADINatRankZ	0.0344	0.007	4.614	0.000	0.020	0.049

Skew:	5.0009	**Kurtosis:**	48.8183
Centered skew:	-8.0432	**Centered kurtosis:**	389.5965

Figure 4: Urbancity (RUCC) and deprivation (ADI) predicting Natality (Births)

GEE Regression Results

Dep. Variable:	Births	**No. Observations:**	5538
Model:	GEE	**No. clusters:**	572
Method:	Generalized	**Min. cluster size:**	3
	Estimating Equations	**Max. cluster size:**	10
Family:	Gamma	**Mean cluster size:**	9.7
Dependence structure:	Exchangeable	**Num. iterations:**	7
Date:	Thu, 20 Mar 2025	**Scale:**	0.031
Covariance type:	robust	**Time:**	21:48:25

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	11.0574	0.927	11.931	0.000	9.241	12.874
RUCC_2013	-0.0083	0.008	-0.985	0.325	-0.025	0.008
logADINatRankZ	0.0345	0.007	4.647	0.000	0.020	0.049
year	-0.0077	0.000	-16.729	0.000	-0.009	-0.007

Skew:	4.8764	**Kurtosis:**	42.6261
Centered skew:	-8.8783	**Centered kurtosis:**	410.0012

Figure 5: Urbancity, deprivation along with a linear time trend predicting natality.

GEE Regression Results

Dep. Variable:	Births	**No. Observations:**	5538
Model:	GEE	**No. clusters:**	572
Method:	Generalized	**Min. cluster size:**	3
	Estimating Equations	**Max. cluster size:**	10
Family:	Gamma	**Mean cluster size:**	9.7
Dependence structure:	Exchangeable	**Num. iterations:**	8
Date:	Thu, 20 Mar 2025	**Scale:**	0.031
Covariance type:	robust	**Time:**	21:48:26

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-4.4382	0.007	-621.042	0.000	-4.452	-4.424
logADINatRankZ	0.0346	0.007	4.653	0.000	0.020	0.049
RUCC_2013c	-0.0081	0.008	-0.959	0.338	-0.025	0.008
yearc	-0.0077	0.000	-17.143	0.000	-0.009	-0.007
logADINatRankZ:yearc	0.0011	0.000	2.182	0.029	0.000	0.002
RUCC_2013c:yearc	-0.0032	0.001	-6.010	0.000	-0.004	-0.002

Skew:	4.8786	**Kurtosis:**	42.7490
Centered skew:	-8.8106	**Centered kurtosis:**	404.3912

Figure 6: Urancity, deprivation, and both their interaction with year predicting natality.

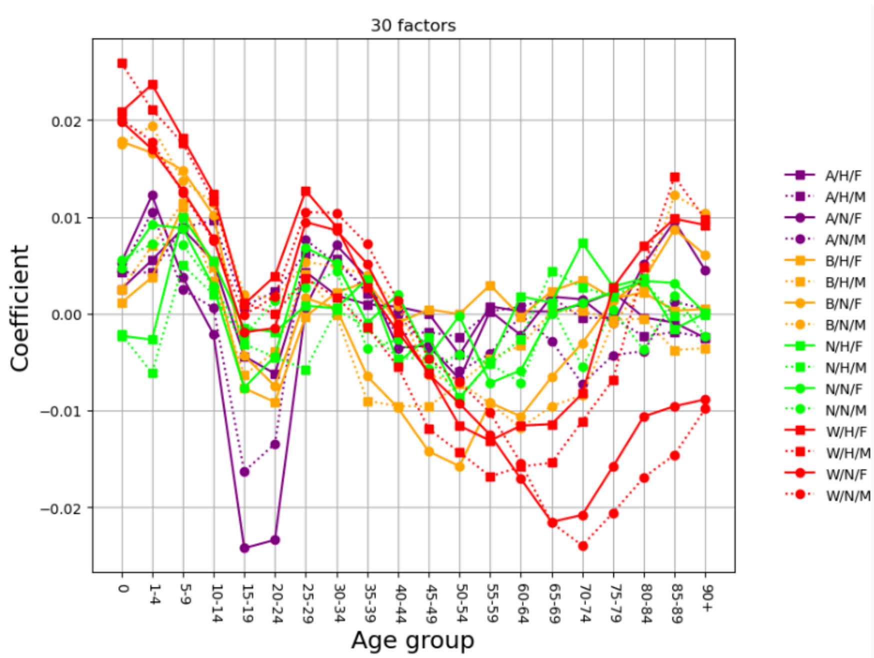

Figure 7: Model fit with 30 PCs graphing their coefficient (to predict natality) by different age groups across ethnicities and races.