# A STUDY OF SEASONAL PATTERNS OF BIRTH FOR VELKE POLE, SLOVAKIA BETWEEN 1781 AND 1900

Deepanjana Majumdar

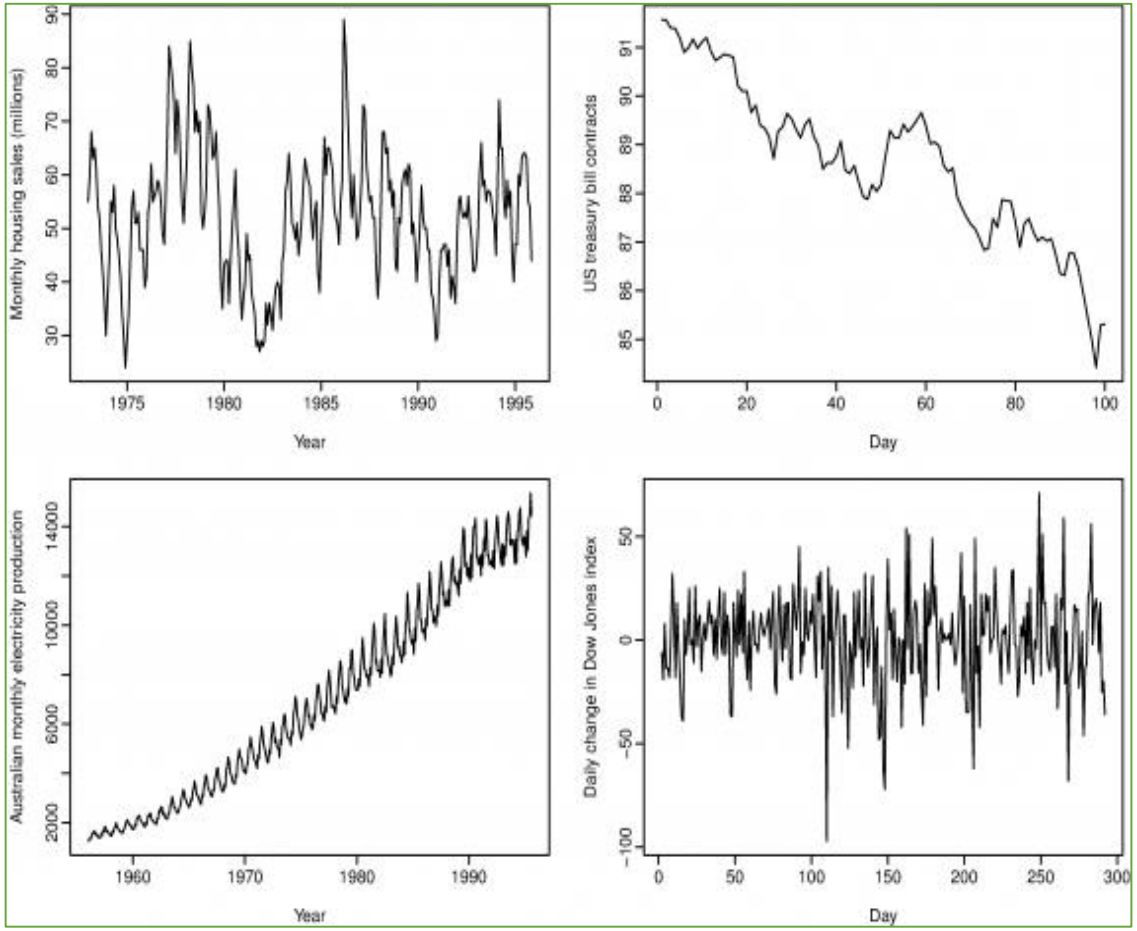
Master's Thesis Presentation

# CONTENTS

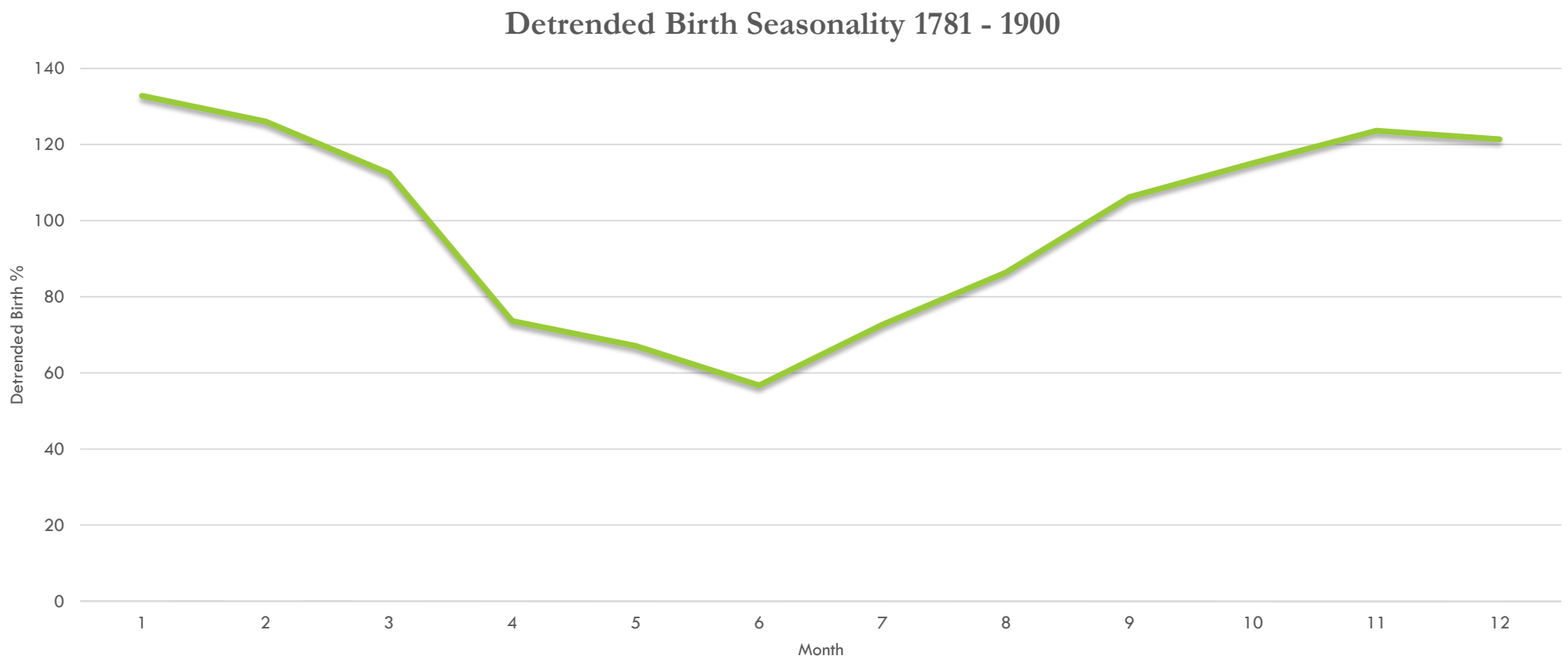
# WHAT IS BIRTH SEASONALITY?

- Birth seasonality is the variation in the frequency of birth correlated with a season of the year
- Time Series data decomposition:
  - Seasonal component
  - Trend component
  - Cyclic component
  - Random
- Necessary to detrend to extract seasonal component

$$b_t = \frac{B_t}{\sum_{m=t-5}^{t+6} B_m} * 1200$$

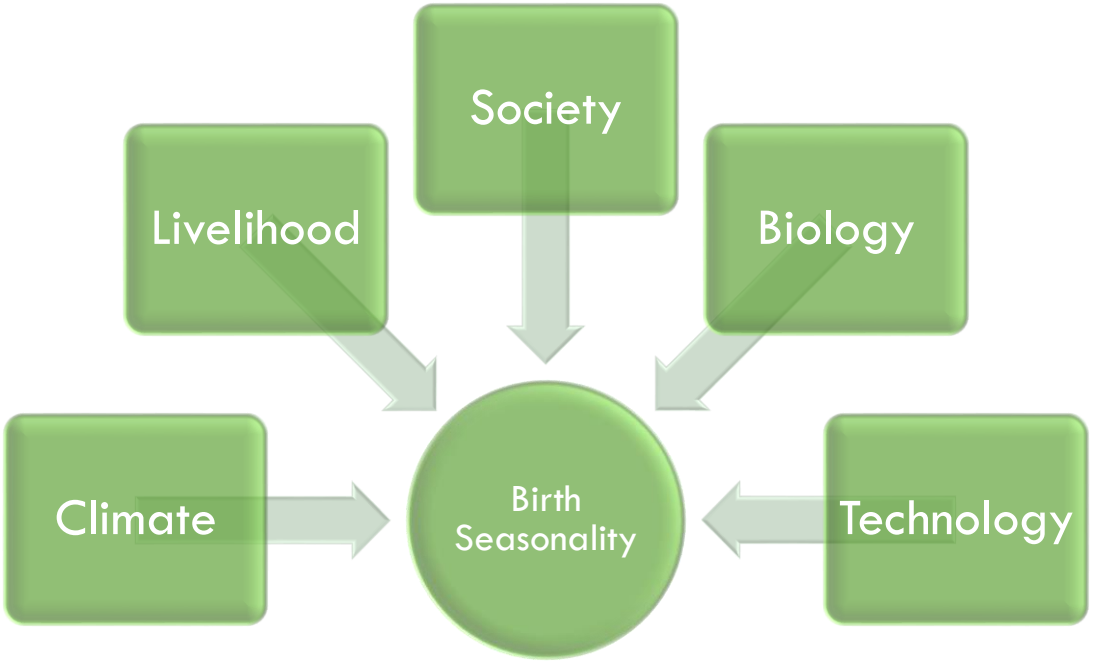


# WHAT WAS SEASONAL BIRTH PATTERN OF VELKE POLE ?

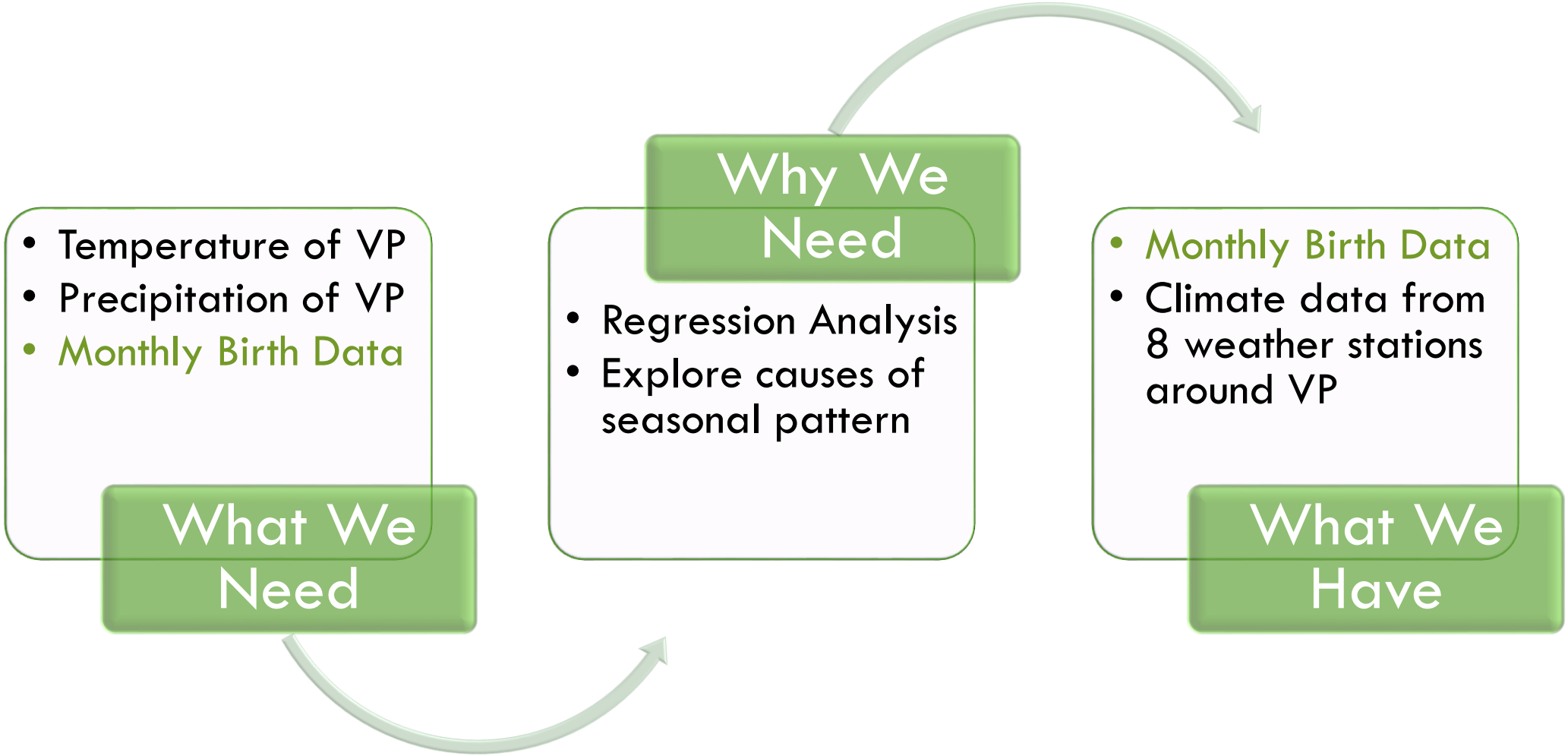


# INFLUENCES OF BIRTH SEASONALITY

- Climatological
- Socioeconomic
- Socioreligious
- Physiological
- Technological factors are making their way into the modern society as a determinant of birth seasonality



# DATA – WHAT WE NEED VS. WHAT WE HAVE



# DATA SOURCES

- Church Book or parish register of Velke Pole – By rule recorded the number of baptisms performed. This baptism data is considered equivalent to birth data
- HISTALP (Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region) – an initiative started within the climate division of the Austrian weather service (ZAMG)
- Slovak Hydrometeorological Institute (SHMU) could provide useful data for a weather station (Oravsky Podzamok)


http://www.zamg.ac.at/histalp/


http://www.shmu.sk

# WEATHER STATIONS

# DATA CLEANING AND ANALYSIS TOOLS

## Pandas

- An open source Python library for data preparation, analysis and modeling
- Data frame component (also a feature in R) is a rich data type to hold two dimensional column oriented data structure
- Can perform aggregation and subset selection of data in-memory in an efficient manner
- Used to clean, pre-process, and transform data before Kriging & regression analysis

Some code samples:

```
df = pd.read_csv('../data/VPBirth.csv',
        sep=";")
```

```
movAvg = pd.rolling_mean(result.stack(),
        window=12, center=True)
```

## Microsoft Excel

- Regression Analysis using Data Analysis Toolpak add-in
- All the data visualization has been performed on Excel

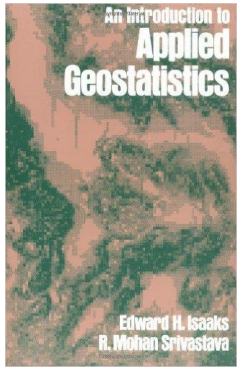
| SUMMARY OUTPUT | |
|---|---|
| *Regression Statistics* | |
| Multiple R | 0.736618664 |
| R Square | 0.542607056 |
| Adjusted R Squ | 0.486511695 |
| Standard Error | 0.716580983 |
| Observations | 120 |

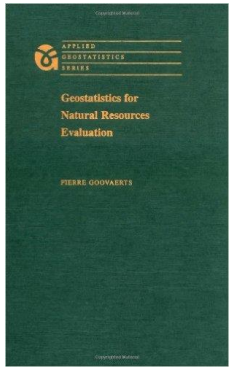
| ANOVA | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 13 | 64.57023967 | 4.966942 | 9.6729399 | 5.48E-13 |
| Residual | 106 | 54.42976033 | 0.513488 | | |
| Total | 119 | 119 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Intercept | 0.957633973 | 0.362672861 | 2.64049 | 0.00952915 | 0.2385998 | 1.6766682 |
| Temp | 0.248857761 | 0.295776577 | 0.841371 | 0.40203437 | -0.3375481 | 0.8352636 |
| TempSq | 0.058061711 | 0.29195044 | 0.198875 | 0.84274131 | -0.5207584 | 0.6368819 |
| Jan | -0.312782495 | 0.4014674 | -0.779098 | 0.43765782 | -1.1087307 | 0.4831657 |
| Feb | -0.13307946 | 0.577754982 | -0.230339 | 0.81827237 | -1.2785349 | 1.012376 |
| Mar | -1.172729618 | 0.72353837 | -1.620826 | 0.10802612 | -2.6072148 | 0.2617556 |
| Apr | -2.251443687 | 0.831292229 | -2.708366 | 0.00788509 | -3.8995614 | -0.603326 |
| May | -1.745727543 | 0.833241261 | -2.095105 | 0.03854426 | -3.3977094 | -0.0937457 |
| Jun | -2.806631111 | 0.585930825 | -4.790038 | 5.4456E-06 | -3.968296 | -1.6449663 |
| Jul | -1.477882398 | 0.40315122 | -3.665826 | 0.00038713 | -2.2771689 | -0.6785959 |

| RESIDUAL OUTPUT | | |
|---|---|---|
| *Observation* | *Predicted ln(birth)* | *Residuals* |
| 1 | 0.561969871 | -0.074666243 |
| 2 | 1.063940825 | -0.719398077 |
| 3 | 0.193159046 | -0.031100674 |
| 4 | -0.962382361 | 1.18265517 |
| 5 | -0.390462869 | -0.066586348 |
| 6 | -1.693657685 | -0.326115938 |
| 7 | -0.56208545 | -0.393970708 |

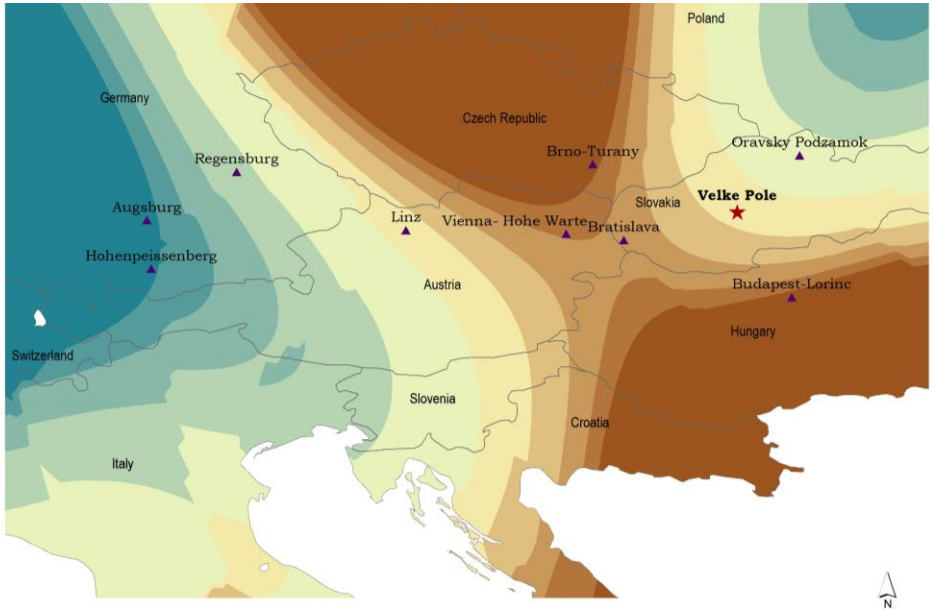
# SPATIAL INTERPOLATION



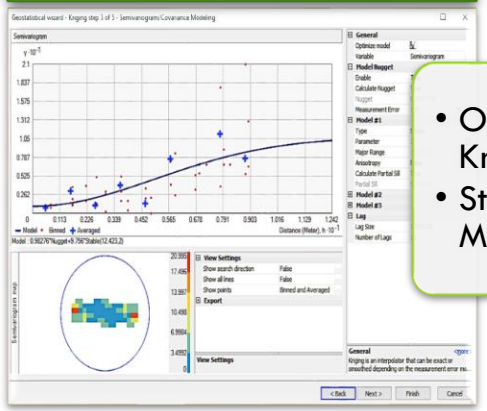




- Extensive evaluation done between IDW and geostatistical interpolation techniques
  - IDW (Inverse Distance Weighting) – Measured values closest to the prediction location have the most influence
  - Kriging – Most recognized form of geostatistical interpolation technique. Produces surface predictions and measure of accuracy and errors in the predictions
- Geostatistical interpolation techniques (Kriging and its family of variations) perform better, because they consider the pattern of spatial dependence for variables such as temperature, precipitation, and photoperiod
- It is possible to correlate secondary information such as elevation to a Kriging process in order to improve the prediction results
  - Deterministic interpolation techniques do not offer this range of flexibility

# KRIGING IN ARCGIS 10.3

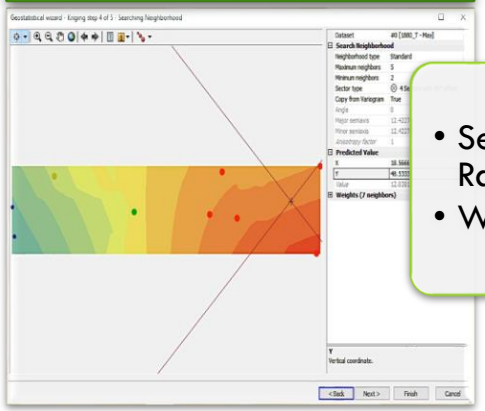



## Semivariogram Modeling



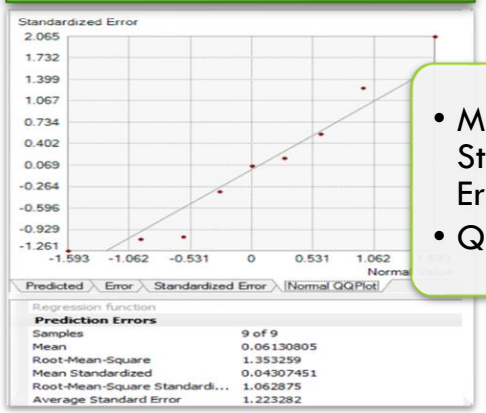
- Ordinary Kriging
- Stable Model

## Neighborhood Search Field



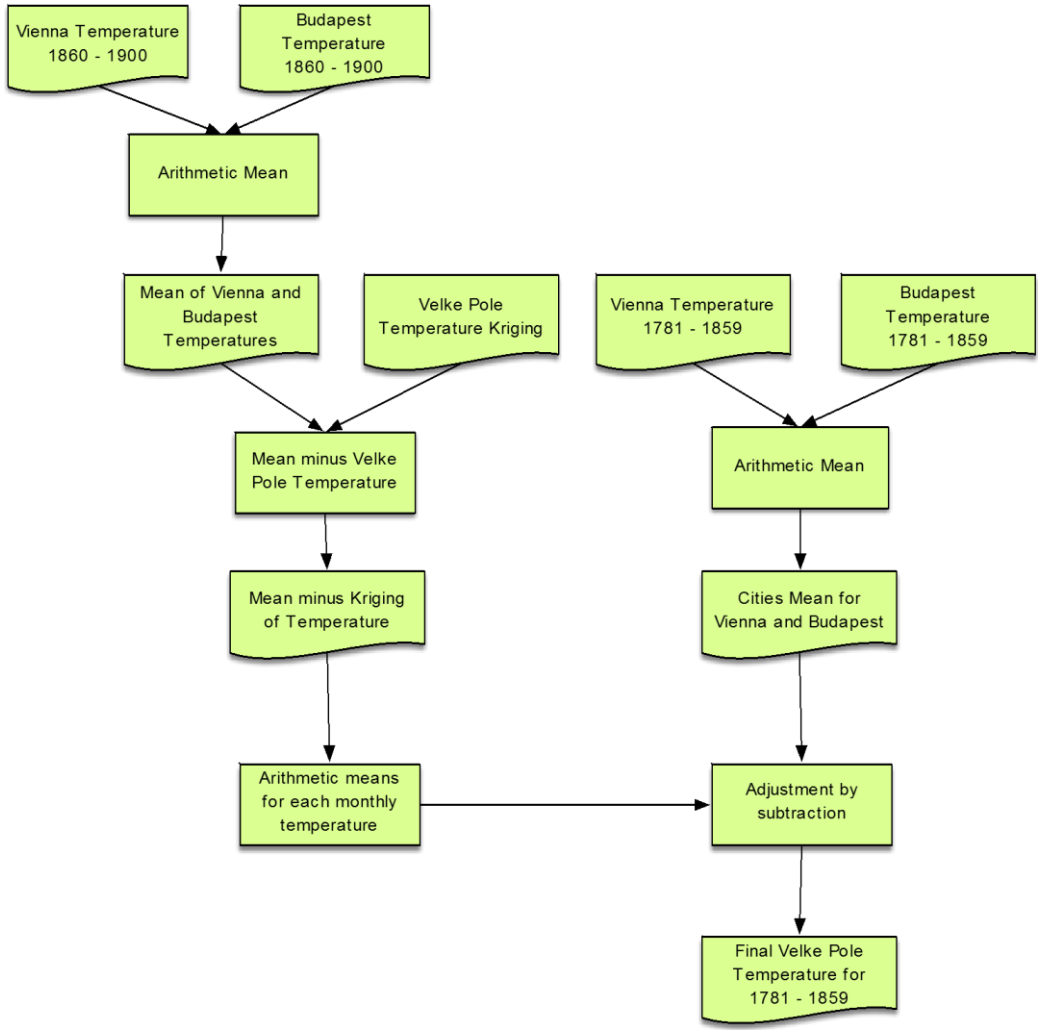
- Search Radius
- Weighting

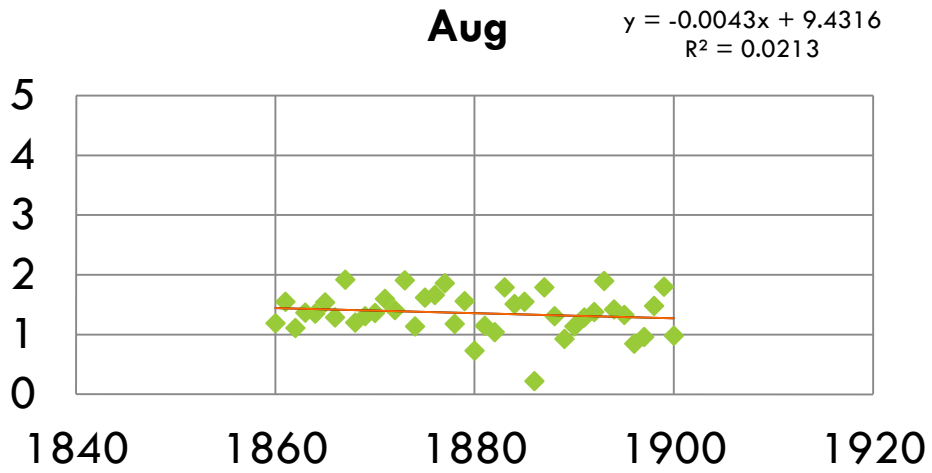
## Cross-Validation




- Mean Standard Error $\approx 0$
- QQ Plots

# EXTRAPOLATION METHOD




Difference of mean monthly temperature of Vienna and Budapest and the Kriging Results

$R^2$ significance tested using the following formula using Fisher distribution

$$F = \frac{R^2}{1 - R^2} \times (n - 2)$$

Results were not significant and a standard mean difference for each month was calculated to adjust with the mean temperatures of Vienna and Budapest (1781-1859)

# REGRESSION ANALYSIS

$$\boldsymbol{ln\, b_t = \sum_{s=1}^{12} \alpha_s d_t^s + \beta_1 T_{t-9} + \beta_2 T_{t-9}^2 + \beta_3 P_{t-9} + \beta_4 P_{t-9}^2}$$

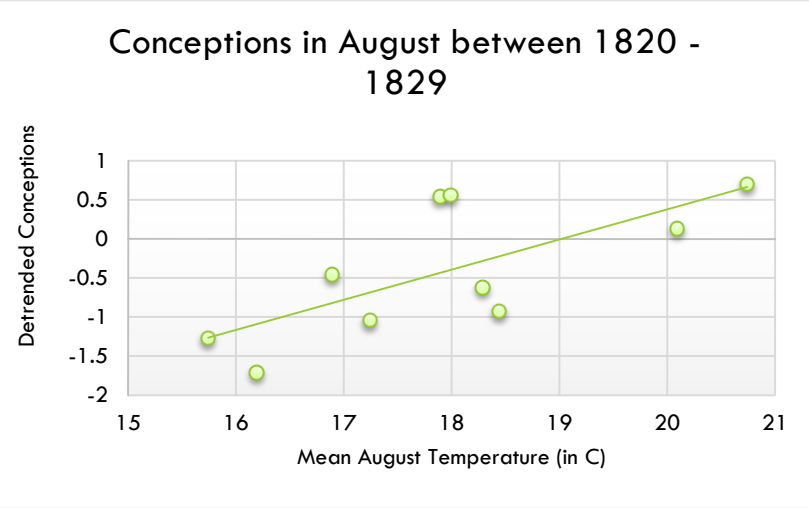
- $\ln b_t$ - the logarithm of the detrended births per day in month *t*.
- $d_t$ - the dummy variable for month *s* whose coefficient is $\alpha_s$
- T and P - temperature and precipitation at the moment of conception, considering a 9-month lag, hence $t - 9$

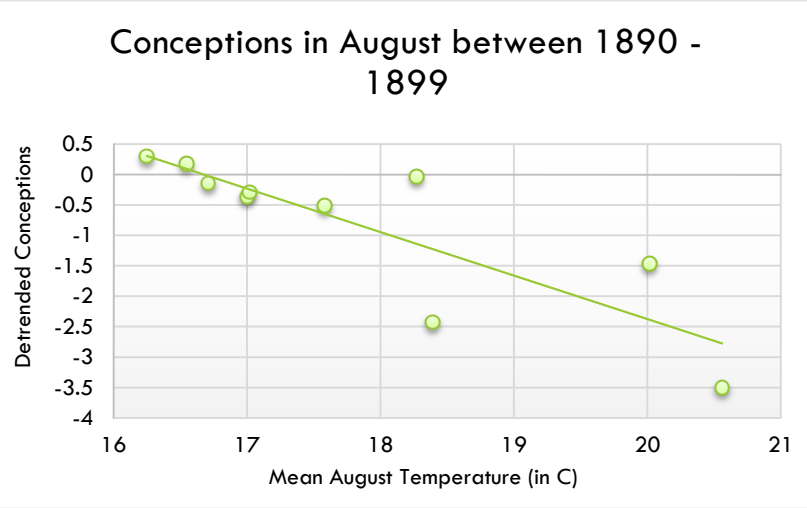
# DUMMY VARIABLES

- A dummy (indicator) variable can only take on values 0 or 1
- Act as numeric stand-in for a qualitative or categorical variable such as gender (male or female)
- A dummy variable with a value of 0 will cause coefficients to disappear
- The months (January to December) are coded as dummy variables in our analysis in order to compute the seasonality

| Birth-9 Months | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Feb | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mar | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Apr | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| May | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Jun | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Jul | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Aug | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Sep | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Oct | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Nov | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Dec | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

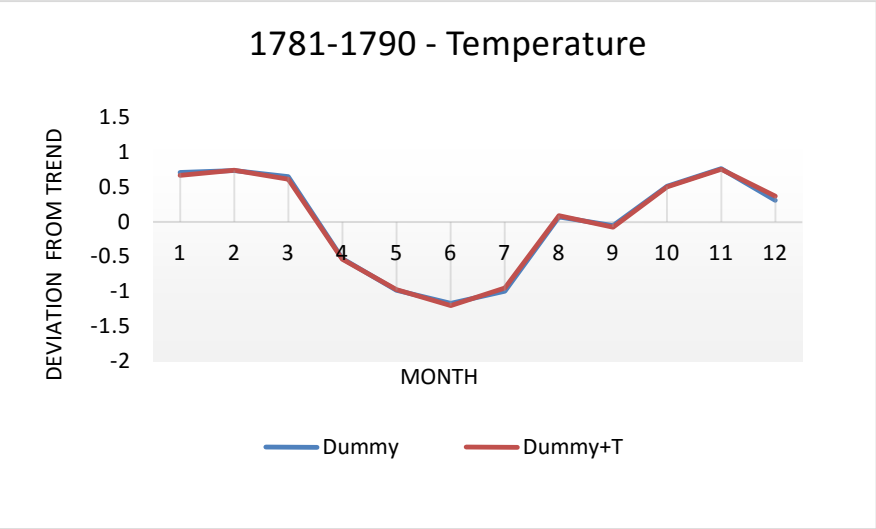
# NONSEASONAL TEMPERATURE EFFECTS

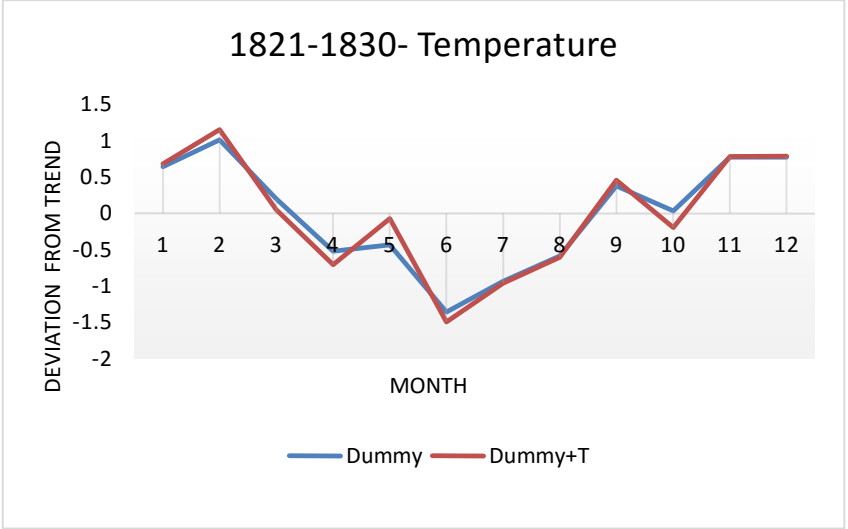


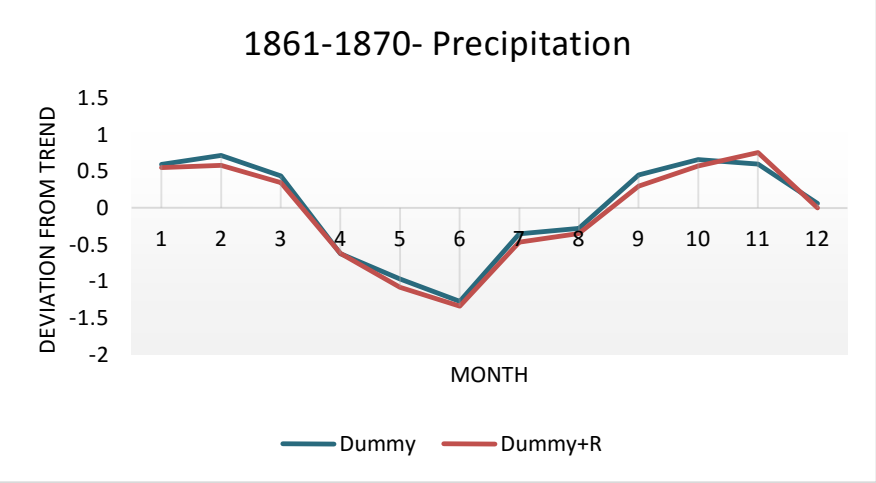




- There is an increase in the conception as temperature rises for the plot showing conception in August 1820-29, in contrast August of 1890-99 shows a definite change from a positive to a negative correlation
- This change in trend matches the trend that is observed elsewhere around the world that with increase in temperature there is a decrease in conception; hence, hotter regions observe a dip in spring births

# REGRESSION RESULT – PLOTS

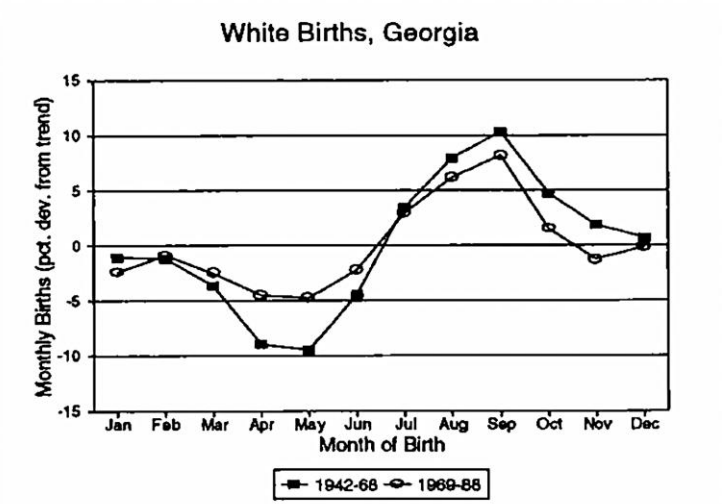


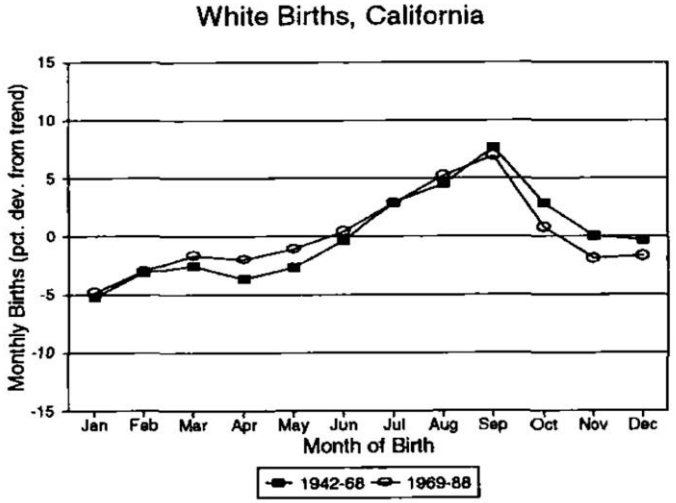



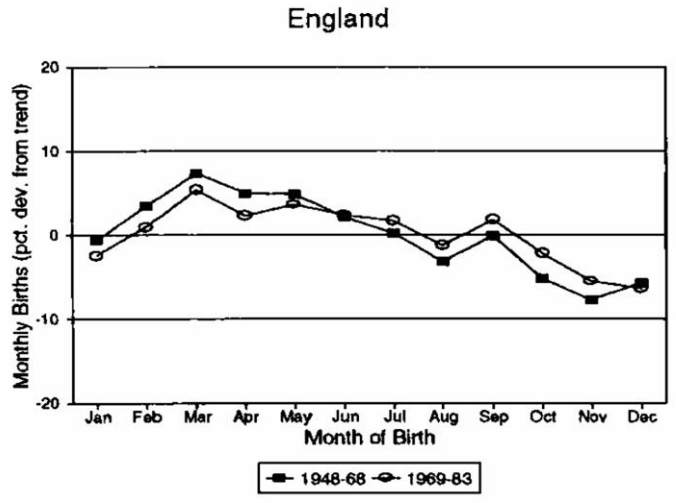



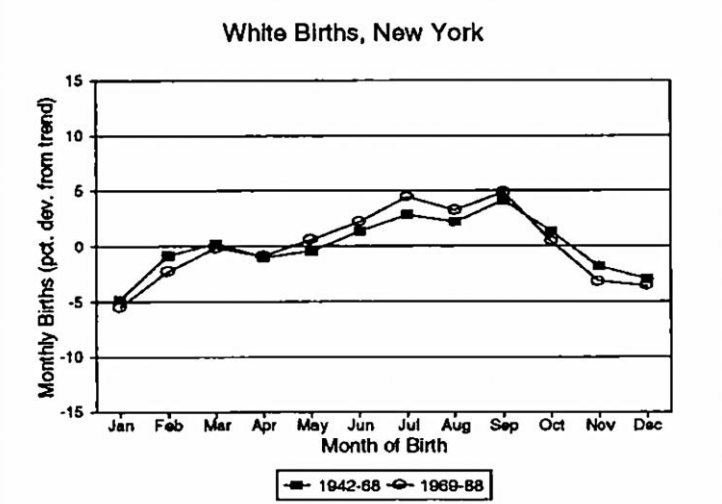

- 1781-1790: Controlling for temperature has little affect to the seasonal pattern
- 1821-1830: The clear spring-summer trough is punctuated with a peak towards the mean for the month of May with the introduction of temperature controls by about 30%, signifying a positive effect of temperature, whereas it drives the value just below the mean for the month of October
- 1861-1870: Introduction of precipitation controls follows slightly below the values for pure seasonality except the month of November when the magnitude increase slightly and creates a peak before sharply falling in December

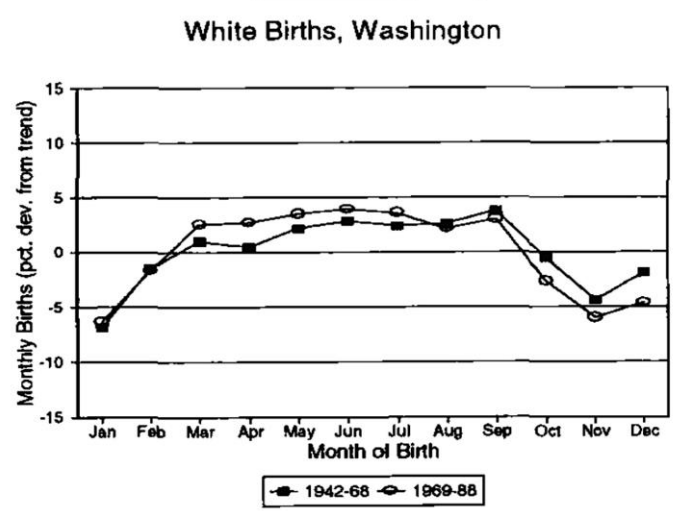
# COMPARATIVE STUDY

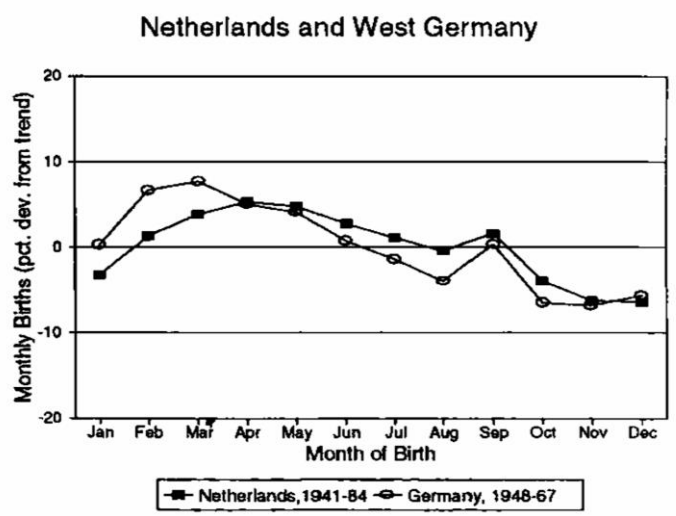


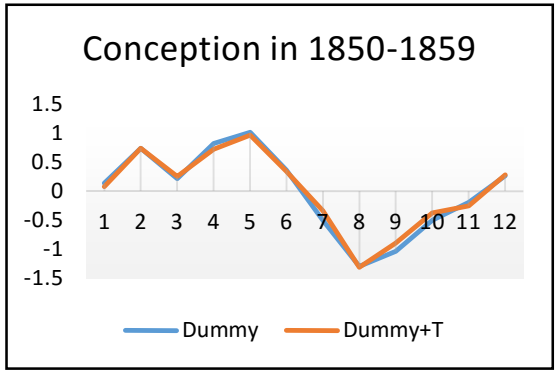



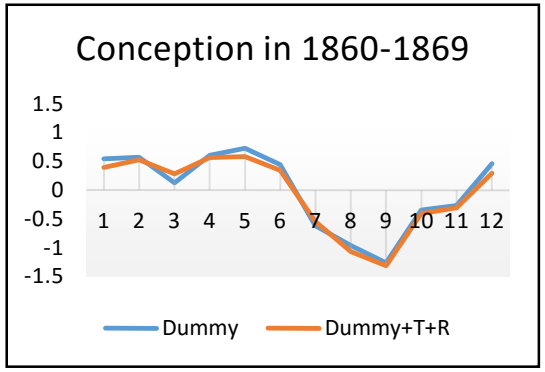



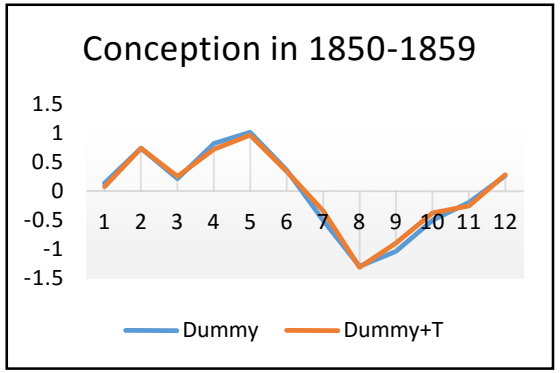
# DEDUCTIONS – REASONS FOR DIP IN CONCEPTION

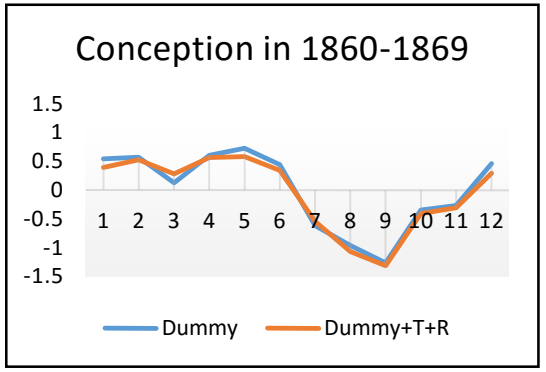



- The period of lent among observed by Catholics when they observe fasting and abstinence
- The spring seeding time, which saw the work force, mainly the men, go away for work
- Autumnal harvesting and seeding time
- 1880 onwards presence of railways so long distance travel for work was made possible
- Peak flu season causing physiological changes especially in March
- Summer dips due to high temperatures that affect hormonal concentration

# DEDUCTIONS – REASONS FOR PEAKS IN CONCEPTION




- The holidays during the period of Christmas may have affected conception rate to be higher, leading to higher births during the fall months
- Spring peaks in conception could be attributed relationship between climate and the physiology of reproduction

# SUMMARY

1. The focus on climatological data was driven by time period under scrutiny being of historical nature, and the numerous literatures emphasizing the importance of climatological factors on the seasonality of birth
2. Where data was missing it was generated using spatial interpolation technique of Kriging and extrapolation based on the kriging and recorded data
3. The method of multiple regression was performed to find evidence of temperature and precipitation's role as determinants of the seasonal patterns
4. Conclusions were made based on complex statistical processes, knowledge of the history of the people, their mode of subsistence, religious affinity and other sociological circumstances

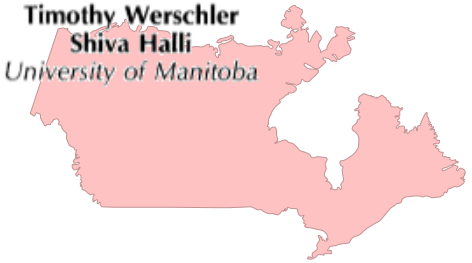
# FUTURE WORK

- Benchmarking the accuracy of temperature and precipitation predictions
- Development of full-fledged Python libraries for performing geospatial and deterministic interpolation, including semi-variogram analysis
- The regression model can be further expanded using more independent and categorical variables
- Interactive tools can be developed for visualization and analysis purposes
- Extensions could be made to do forecasting of the future birth patterns based on changing environmental and sociological parameters.

# RELATED WORKS

The Seasonality of Births in Canada: A Comparison with the Northern United States

Timothy Werschler
Shiva Halli
University of Manitoba

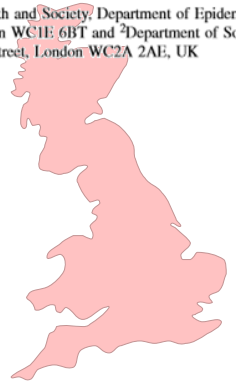


Human Reproduction Vol.16, No.7 pp. 1512–1517, 2001

The seasonality of live birth is strongly influenced by socio-demographic factors

Martin Bobak[1,3] and Arjan Gjonca[2]

[1]International Centre for Health and Society, Department of Epidemiology and Public Health, University College London, 1–19 Torrington Place, London WC1E 6BT and [2]Department of Social Policy, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK



Naturwissenschaften 80, 516–518 (1993) 

Changes in Seasonality of Birth Rates in Germany from 1951 to 1990

A. Lerchl, M. Simoni and E. Nieschlag
Institut für Reproduktionsmedizin der Universität, D-48149 Münster

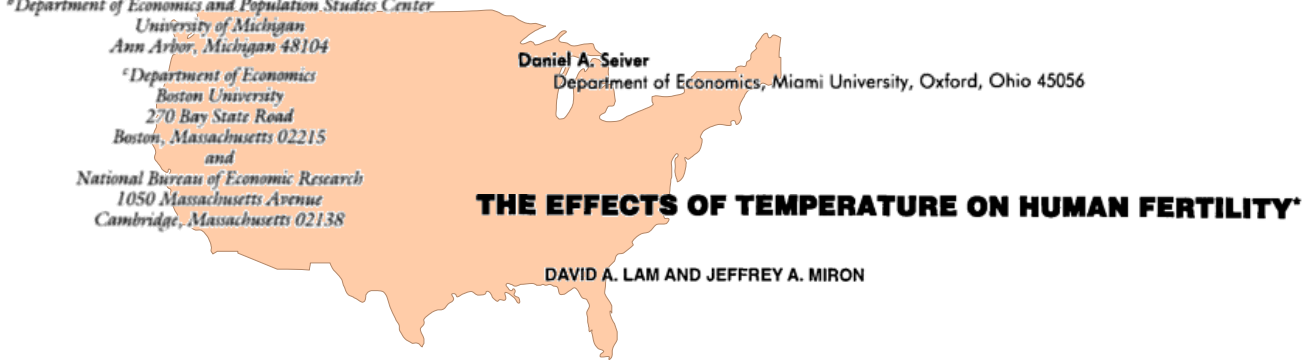
Global Patterns of Seasonal Variation in Human Fertility[a]

DAVID A. LAM[b,d] AND JEFFREY A. MIRON[c]
[b]Department of Economics and Population Studies Center
University of Michigan
Ann Arbor, Michigan 48104
[c]Department of Economics
Boston University
270 Bay State Road
Boston, Massachusetts 02215
and
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, Massachusetts 02138

DEMOGRAPHY© Volume 22, Number 1 February 1985

TREND AND VARIATION IN THE SEASONALITY OF U.S. FERTILITY, 1947–1976

Daniel A. Seiver
Department of Economics, Miami University, Oxford, Ohio 45056



THE EFFECTS OF TEMPERATURE ON HUMAN FERTILITY*

DAVID A. LAM AND JEFFREY A. MIRON

Hum Ecol (2009) 37:227–234
DOI 10.1007/s10745-009-9221-x

Birth Seasonality in Present-Day Italy, 1993–2005

Matteo Manfredini

# THANK YOU