Yuya Jeremy Ong

Prof. Kevin J. A. Thomas

SOC 423: Social Demography

December 8th, 2018

Birth Rate Forecasting Using Deep Learning Models

1. Introduction

In the academic field of demography, the majority of work has often been concentrated in the retrospective data analysis of past population distributions, changes, and dynamics. In this respect, demographers and sociologists alike provide a hypothesis or question driven heuristic when performing research. However, compared to the body of work that is presently existing within demography, few works that relate to population modeling exist in the overall academic field itself. Specifically, we look to the problem of understanding and building models which are more *predictive* than *descriptive* in nature - which helps to provide key insights not only from a retrospective perspective, but also from a crucial factor utilized in various decision making processes such as allocation of resources, city and urban planning, policy making, and various other domains which hinge on the ever so changing dynamics of the society's population.

In particular, we look to Preston, a demographer that outlines some of the major trends within the field of demographic studies. In particular, one of the key aspects we can take from his paper indicates the emergent properties of technology and how they converge towards the demographic field itself (Preston 597). He outlines that the three key components of the technological developments of demography include: i) improvement in both computational hardware and software used for data storage, retrieval, and processing, ii) data collection, accessibility, and quality control of data are improving with organized institutions, private, and public companies, and the world wide web providing Big Data (i.e. a concept indicating the analysis of large scale data processing) through mediums such as social media networks and large scale databases, and finally iii) statistical methods and data processing algorithms from recent developments in algorithm designs are improving at a much significant rate. Based on all three of these characteristics outlined by Preston, our aim for this project is to take some of the state-of-the-art technologies (including hardware, software, and algorithms) and produce a new baseline based on these new modeling heuristics that we are going to present in this paper.

The focus of this paper is based on developing a new and novel projection methodology in modeling the temporal dynamics of global birth rates using a new breed of algorithms known as Deep Learning. Projection modeling is the methodology applied to taking past data and developing a model which would help to describe the future demographic distribution and population structure. The modeling structure of this problem is often framed as a *time series model*, which is essentially a temporal dependent modeling framework applied to a past series of numbers in a temporal sequence, and using those values to predict the next set of values in the sequence. Time series models have often been employed in various other domains and industry, such as the financial industry in predicting market and asset trends. Our contribution to this paper introduces a new modeling methodology, known as Deep Learning, which is a novel type of Machine Learning model which contains a topological structure similar to the human brain and emulates learning patterns from the biological neuron. In this work, we show that this modeling methodology provides various advantages over classical statistical methods, such as Autoregressive Integrated Moving Average (ARIMA) models and Vector Autoregressive

Models (VAR), which are the two most common types of modeling algorithms utilized in the field of demographic forecasting, and they provide a much more flexible and end-to-end solution in generating results with high accuracy and precision as opposed to previous methodologies used in the industry.

However, building forecasting models in general (not just in the context of birth rate forecasting) is a non-trivial task. There are three primary challenges that various researchers must encounter and work around to mitigate these challenges. First, one of the key issues in data modeling revolves around the process of using high quality datasets. Although with the recent advancements mentioned previously, the collection of higher quality datasets have become significantly easier, that is not to say that the choice of data and the preprocessing of the dataset is very crucial to the entire modeling process. If proper care is not taken with the given dataset, this would result in a poorly performing model as it will not be able to capture the actual features and characteristics of the things we are accounting for. Furthermore, another key challenge of the modeling process revolves around understanding and disentangling complicated determinants and factors of complex relationships between the parameters and the target value. This is aligns with much of the similar challenges posed by Bongaarts' model through his multiphase and multivariate analysis between the relationship between the different factors such as economic and politically related variables. Finally, one of the most difficult challenges in modeling is accounting for anomalies in the data, primarily due to the sudden shifts in the distribution of the data. One very notable example of this phenomena is the sudden increase in the birth rate due to the Baby Boom during the mid 1960s in the United States.

2. Related Work

In the demographic literature of birth rate forecasting, there are many works that have primarily use simple methods to compute and generate forecasting models for birth rates. In this section, we evaluate some of the past works that have been performed based on the various literature that is already present in the field.

One of the very first and most comprehensive work performed on birth rate forecasting was conducted by the United Nations' Population Division, where they have provided the first set of projections in 1957. Since then, the UN utilizes their own collected data and continuously improves and refines their process for generating a model which aims to forecast various metrics, such as population and birth rates. One of the most common modeling methodology that researchers employ are *deterministic models* - models based on a set of handcrafted equations and heuristics that take various index values and features which can be fed into a mathematical function. Bongaarts utilizes this methodology in conjunction with his theoretical model of the Proximate Determinant Framework for birth rates to develop a model which takes into account socioeconomic, marriage statistics, and other key metrics (Bongaarts). In this work, we take inspiration from Bongaarts' Proximate Determinant Framework and utilize his ideologies as a justification towards how our model emulates a very topological structure to his original ideas on complex factor disentanglement of causes towards birth rates.

On the other hand, some researchers employ various other methodologies - notably statistical methods including probabilistic modeling processes as well as ARIMA models, which we have described earlier in the previous section. Alkema et. al utilizes a Bayesian projection model to develop a model for the total fertility rates for each country. The model also makes use of the ideas from the Demographic Transition Theory, in which they divide their models based on the three key phases: pre-transition fertility transition, the fertility transition, and post-transition low fertility. Shang et. al developed another model using the ARIMA modeling methodology presented earlier. Shang uses the ARIMA method to develop two types of models: i) age specific fertility trends and ii) birth order age specific fertility rates.

3. Modeling Methodology

In this section we introduce some relevant background necessary for the understanding of the modeling methodology for birth rate forecasting. For our project we will utilize two different types of models: ARIMA (Box-Jenkins method) and Recurrent Neural Network (RNN), which is specific type of Deep Learning-based Machine Learning algorithm for temporal modeling and forecasting.

Autoregressive Integrated Moving Average Model, ARIMA, or sometimes known as the Box-Jenkins method is a statistical method which utilizes a sliding regression-based modeling methodology to capture and forecast the temporal dynamics of the data. The models is comprised of three primary components: i) autoregressive (p), which is a model that regresses on its lagged prior values - usually using a linear regression form, ii) integrated (q), uses a differencing method of raw observations to make values stationary, and finally iii) moving average (r), which considers the observation and residual errors to factor in the lagged observations of the data. ARIMA models, so far, have produced fairly robust models in the past with a very simple parametric approach.

In this sense, ARIMA does have some leverage in that they are a) interpretable through their parameters and coefficients, b) simple to implement and widely applied in various problem domains, c) proven to be fairly robust for most cases, except having lower precision bounds and large variability windows in their predictions, and d) computationally least expensive to execute on a modern small machine such as a typical laptop or desktop.

Although ARIMA does provide its advantages, it also has some disadvantages that are associated with the ARIMA modeling methodology: a) it assumes a strong statistical assumption based on the data having a stationarity assumption (i.e. where the mean, variance, and autocorrelation of the data are constant) - if not, then a manual transformation function must be applied to minimize the stationary parameters for each of the given data, b) it further assumes a linear relationship temporally, which in most cases various real-world phenomena are nonlinear in nature, c) to improve the overall robustness of the model, an additional set of preprocessing and applying a smoothing function to the data is required to ensure that the model can generalize well to various dynamics in the temporal pattern, and d) careful parameter tuning of both p, q, and r must be performed in order to find the best hyperparameters for the models.

In this project, we utilize a novel and data-driven forecasting algorithm known as Deep Learning, which is a Machine Learning algorithm (a predictive algorithm which learns from data) that is inspired by the neurological function of the brain's synapse and activation, in the end borrowing its modeling topological structure. In particular, Deep Learning allows the computer to "understand the world in terms of a hierarchy of concepts, with each concept defined through its relation to simpler concepts" (Goodfellow 1). Deep Learning, in recent years have shown significant process within the last couple decades and have been widely utilized in various modern applications of predictive modeling and Artificial Intelligence - such as tasks like Image Recognition, Self-Driving Cars, Natural Language processing, and etc, which are all utilized by major Silicon Valley companies such as Google, Facebook, Amazon, and IBM. In most of these applications, these models have produced state-of-the-art results which outperform various human-level tasks yielding high levels of accuracy and precision, and have dominated the industry in many countless ways - and still continue to do so as we speak.

The topological structure of the model can be imagined as a network of interconnected parameters where each of the nodes in the graph are connected to every other node in the layer. The model can take in any set of data point with different set of features, which are then computed through a layer of more interconnected nodes which are known as hidden layers. Through each of the interconnections, these edges represent the weight of the specific parameter with respect to the value of the input feature and the corresponding target node - in other words interpreted as some type of importance factor of the given attribute in the data. Finally after performing various transformations of the original dataset through iterative layers of these interconnected parameters, the final output is computed as a linear combination of all factors into a single output - which generates our target prediction values.

From this line of the graphical representation of the model, this model representation resembles a very similar heuristic to Bongaarts' Proximate Determinant Framework - where the intermediate causal factors such as economic, political, marriage, and other factors influences the effect of birth rates. In modeling this phenomena, another key thing to consider is the combined influence of some of these factors which can also factor into the influence of the other factors we are considering as well. Therefore, by interconnecting each of these features together, we can model the causal compound effect of these variables and effectively modeling a much more accurate scenario of determining the rate of births.

However, in order for us to model a time series data, we must also consider the temporal dimension of the dataset. For this, we make use of a temporal based Deep Learning modeling algorithm known as Recurrent Neural Networks - in particular we use a variation of the algorithm known as the Long Short-Term Memory Networks. These models effectively consider the temporal patterns of the data by considering the order in which the data is fed into the network.

4. Dataset and Feature Analysis

For this project, we utilize the Birth Rate dataset provided by the World Bank Open Data organization. The data set year ranges between 1960 and 2016 and contains population data for approximately 235 countries for the full set. According to the World Bank, the Birth Rate dataset is defined as:

> "Crude birth rates indicates the number of live births occurring during the year, per 1,000 population estimated at mid year. Subtracting the crude death rate from the birth rate provides the rate of natural increase which is equal to the rate of population change in the absence of migration."

We correspondingly preprocess the data by removing any countries that have any missing data temporally (i.e. must have the complete years between 1960 to 2016). However, when processing the data, we did not perform any sort of normalization or preprocessing methods such as exponential smoothing and performing data standardization - as we can directly feed in the raw data in its original form into the model.

5. Experiment Setup

For evaluating the effectiveness of the model, we built two models: ARIMA for establishing our baseline and a LSTM Deep Learning model. To properly evaluate the performance of the model, we perform a temporal validation process where we train both of the models on data between 1960 to 2016 and effectively hold out 20 years to evaluate how well the model is able to project ahead into the future. We can do this by comparing the predicted output from the model from the actual 20 years of data and compute an error metric to evaluate the rate of error from this model. Through this empirical comparison, we can establish how effective the modeling strategy is and essentially compute a general rate of error for the modeling methodology across all countries.

For our evaluation metric, we utilize the Mean Absolute Percentage Error, which is defined as follows:

$$\mathrm{M} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|,$$

where t is the given temporal index, A is the actual value, F is the forecasted value, and N is the total number of temporal timesteps available from the dataset.

6. Model Results

Based on training our model on both ARIMA and RNN (LSTM), we find that based on the overall distribution of the dataset, we find that the RNN LSTM mode can outperform ARIMA model by a significant margin based on the results we have obtained, as shown in the Appendix. On average the mean error for ARIMA model yielded 9.12% while RNN (LSTM) produced an error of 8.79%. However, the most salient aspect of our result lies in the variability of the error from the distribution we find that ARIMA has a standard deviation of 18.15 and RNN (LSTM) having 9.39 - which is significantly half of the variability present in the distribution of errors. Thus, this demonstrates that the RNN (LSTM) Deep Learning model is able to effectively model and forecast future birth rates accurately and precisely - even at very far projections ahead of time. The Appendix pages appended at the end provides a visual representation of the prediction results based on the model for both ARIMA and RNN (LSTM).

7. Discussion

Based on the empirical results we have demonstrated, we have demonstrated that the Deep Learning model has effectively outperformed ARIMA. However, although the model has better performance than the ARMIA model, it is still not yet perfect in generating predictions exactly (although it is impossible to predict future values exactly), we still must consider some factors and additional measures to ensure that the results are robust enough. Therefore, when it comes to modeling future forecasting modeling for birth rate predictions, we must consider not only one model, but an aggregate of different models. Through this we can effectively leverage various perspectives on the analysis and essentially derive a solution which provides both conservative and extreme estimate values for each time period. Furthermore, when building models for future forecasting, it is important to ensure that the quality of the dataset utilized is very high - as poorly performing models are caused mostly due to the dataset not being representative of the original distribution. Therefore, when considering other features and input values for factoring in towards predictive models for birth rates, it is important that we consider the types of data we use and the underlying methodology for the data was collected.

8. Conclusion

In this paper, we have proposed a new and novel methodology for modeling future forecasting of birth rates using Deep Learning methods. We have empirically demonstrated the use of Deep Learning methods to yield results with higher accuracy and precision. In effect, we have shown the potential for Deep Learning methodologies to be introduced in the field of demography and how this can be used to model other complex phenomena and be utilized in various ways. Future work can involve the use of harnessing larger datasets, especially Big Data to perform experimental analysis on a much larger scale to evaluate various social phenomena pertaining to the demographic process.

Works Cited

Alkema, Leontine, et al. “Probabilistic Projections of the Total Fertility Rate for All Countries.” *Demography*, vol. 48, no. 3, Dec. 2011, pp. 815–839., doi:10.1007/s13524-011-0040-5.

“Birth Rate, Crude (per 1,000 People).” *Literacy Rate, Adult Female (% of Females Ages 15 and above) | Data*, data.worldbank.org/indicator/SP.DYN.CBRT.IN.

Bongaarts, John. “Modeling the Fertility Impact of the Proximate Determinants: Time for a Tune-Up.” *Demographic Research*, vol. 33, Nov. 2015, pp. 535–560., doi:10.4054/demres.2015.33.19.

Forecasting Fertility by Age and Birth Order Using Time ... epc2016.princeton.edu/papers/160597.

Goodfellow, Ian, et al. *Deep Learning*. The MIT Press, 2017.

Preston, Samuel H. “The Contours of Demography: Estimates and Projections.” *Demography*, vol. 30, no. 4, 1993, p. 593., doi:10.2307/2061808.

Appendix

Table 1: MAPE Distributions

ARIMA

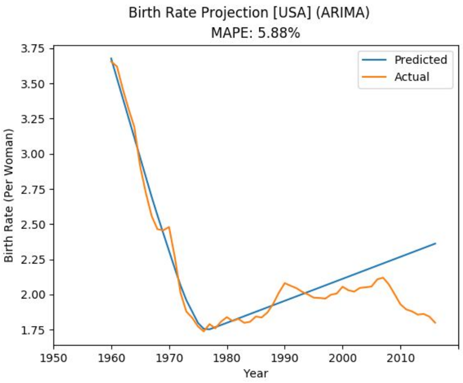
Mean	9.12784
Standard Deviation	18.151828
Min	2.484999
25%	4.745813
50%	5.788534
75%	7.811595
Max:	208.356131

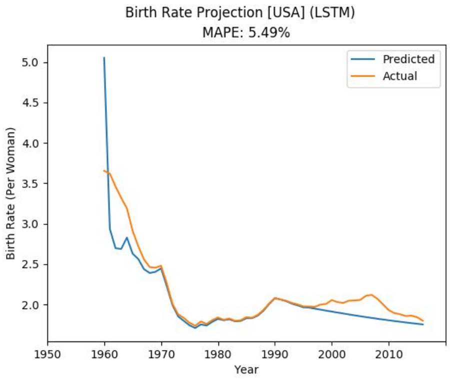
RNN (LSTM)

Mean	8.790662
Standard Deviation	9.395344
Min	0.249976
25%	2.506318
50%	5.299511
75%	12.230355
Max:	51.121002

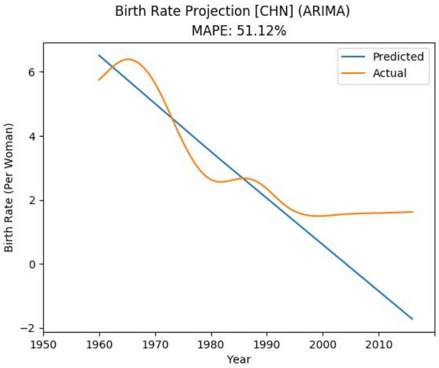
Appendix (Continued)

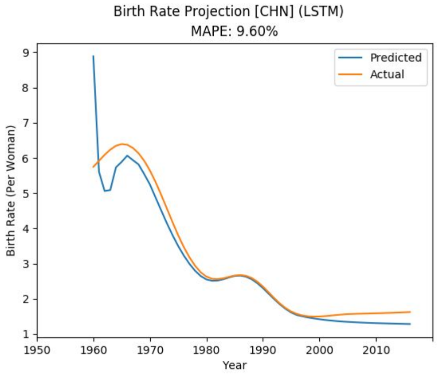
United State


Birth Rate Projection [USA] (ARIMA)
MAPE: 5.88%

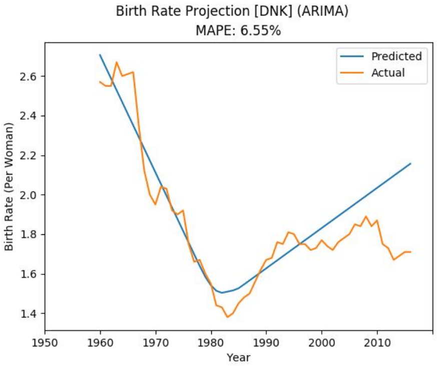

Birth Rate Projection [USA] (LSTM)
MAPE: 5.49%

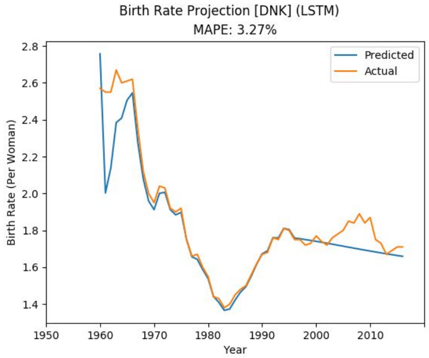
China


Birth Rate Projection [CHN] (ARIMA)
MAPE: 51.12%

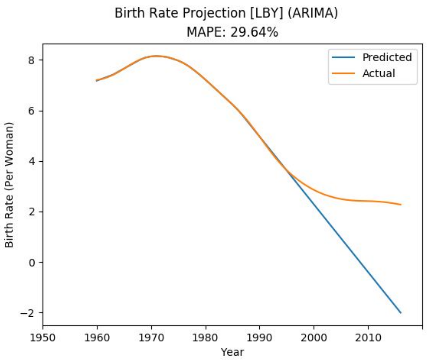

Birth Rate Projection [CHN] (LSTM)
MAPE: 9.60%

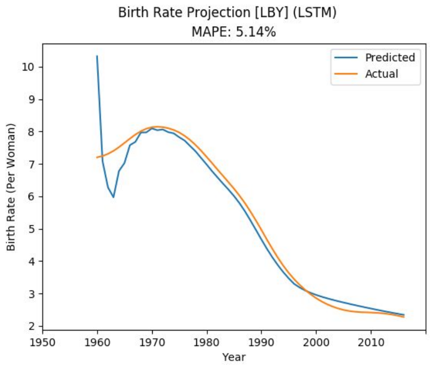
Denmark


Birth Rate Projection [DNK] (ARIMA)
MAPE: 6.55%

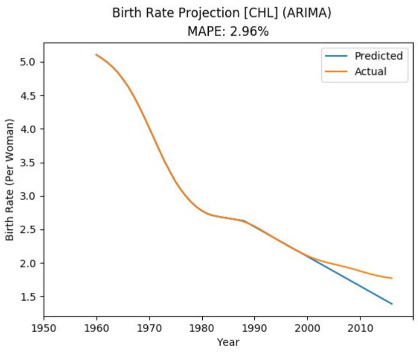

Birth Rate Projection [DNK] (LSTM)
MAPE: 3.27%

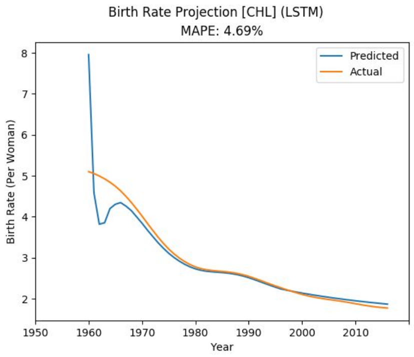
Libya


Birth Rate Projection [LBY] (ARIMA)
MAPE: 29.64%
Predicted
Actual
Birth Rate (Per Woman)
8
6
4
2
0
−2
1950
1960
1970
1980
1990
2000
2010
Year


Birth Rate Projection [LBY] (LSTM)
MAPE: 5.14%
Predicted
Actual
Birth Rate (Per Woman)
10
9
8
7
6
5
4
3
2
1950
1960
1970
1980
1990
2000
2010
Year

Chile


Birth Rate Projection [CHL] (ARIMA)
MAPE: 2.96%
Predicted
Actual
Birth Rate (Per Woman)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1950
1960
1970
1980
1990
2000
2010
Year


Birth Rate Projection [CHL] (LSTM)
MAPE: 4.69%
Predicted
Actual
Birth Rate (Per Woman)
8
7
6
5
4
3
2
1950
1960
1970
1980
1990
2000
2010
Year