Explaining baby name popularity: A mathematical perspective

Aditya Rao[*1] and Dr. Stott Parker[†2]

[1]Master of Science, Computer Science, UCLA
[2]Professor, Department of Computer Science, UCLA

---

[*]adityakatapadi@cs.ucla.edu
[†]stott@cs.ucla.edu

Abstract

Names, goods and ideologies are adopted by society through various mechanisms. Adoption driven by novelty is extensively studied by marketing economics[**3**]. This paper attempts to identify a pattern in the decision making process of choosing baby names. The data for this is readily made available by the US Social Security Administration in the internet [**4**]. This paper explores the diffonential distribution [**6**] as a possible fit for the frequency distribution of a name. This paper also proposes time-series based and a normal distribution based parameter estimation techniques for the diffonential distributions.

Acknowledgements

1 Introduction

Choosing a name for the newborns appears to be a complicated task. Parents' go to great lengths to ensure that the name sounds right with the family name and that it does not have any negative connotations. Moreover, research studies also paint a strong correlation between unusual names, names with negative connotations and the likelihood of having a lonely, overall unhappy life. One study found that psychiatric patients with more unusual names tended to be more disturbed.[5]. Thus, nowadays parents go to great lengths to decide the right name for their child. Religious texts, town/country/city names make up for a bulk of the baby names. For instance, of the 36,000 odd, unique names assigned to boys across the USA since 1880, only about a thousand were not borrowed from any of the sources mentioned earlier. Such names are either variations of the already existing names or derivations from multiple such existing names. For example, Jamie is derived from James, Rob/Bob from Robert, Rich from Richard and so on.

Nonetheless, nowadays parents rely on a variety of sources to narrow down their choices. Some papers pin the decision making process on a taste-based argument [2]. This concept is akin to that of fads [1] and trends [2]. Hence, various first names go through an initial rise in popularity until they reach a peak and then go extinct. For a name to subsist much longer, there has to be positive social stimuli[3]. A social stimulus plays an important role in deciding the future of a first name. It can cause the rapid decline in popularity of a name or contribute to the longevity of the name. For example, after World War 2, the first name Adolf had a very steep decline, the negative association with Adolf Hitler being the reason for the accelerated fall. On the contrary, the first name Ronald had a sharp increase in popularity, the positive association with Ronald Regan being the reason for the accelerated rise. Names with multiple social stimuli usually resemble a multi-modal frequency distribution, with the stimuli reviving the frequency distribution at various time intervals. This paper also considers the taste-based argument put forth by Coulmont et al. [3]. This can be clearly seen in names such as Jeff, which reflect a beginning enthusiasm, rising fast to a peak. This is generally followed by a subsequent, gradual lack of interest, possibly as

---

[1]a short-term event, what some may call a "flash in the pan".

[2]similar to a fad, but has the potential of becoming a long-term influence on the future of a market

[3] a factor that significantly contributes to the rise of a name or to its fall

parents note that there are many children around with that name.

Most of the distributions can be approximately modelled by a frequency function based on the diffonential distribution [6]. The curve, to a certain extent, reflects the parents' changing enthusiasm for a particular name.

2   Related Work

In their paper [7] Bengt Sigurd et al., proposed a shape-based explanation for the frequency distribution of names in Norway. They classified the name frequency curves as asymmetric, symmetric and gamma. The asymmetric ones had a slow rise followed by a faster fall, the symmetric ones were similar to a normal curve and the gamma names had an initial fast rise followed by a slower fall. They modelled gamma names around the equation $f = a.t^b.c^t$. The paper concludes on the observation that the gamma curve dominates in other cultures where names are given freely (without any external influence).

Coulmont et al.[3], made use of the Bass model, $\dfrac{dN}{dt} = [p+q(\dfrac{N}{K})](K-N)$ to draw similarities between the success of a product and the choosing of baby names. The paper is based on the assumption that incorporating the principles of population-level diffusion dynamics, implicitly incorporates the distinct traits of individual-level behaviour. The model corroborates the well known argument that popular names may be driven by fashion. Most of the work done by me is influenced by this paper.

3   A Diffonential Distribution Fit For Baby Names

The diffonential distribution is a difference of two exponentials, hence the name. It was first introduced by Preston et al.[6]. The frequency distribution of a diffonential distribution has a few interesting properties similar to that of a typical frequency distribution for a name. The curve has a steep rise to a peak and then a slow fall. Names with no or very little social stimuli to revive them, tend to follow the diffonential distribution.

The diffonential distribution has the general form $f(t) = b(e^{-at} - e^{-ct})$. The sign and magnitude of the coefficients a and c determine the general

shape of the curve. The parameters a and c behave in an interesting way. The value of the coefficient c spikes whenever the frequency distribution of the name experiences a sudden increase. The coefficient a remains relatively constant reaching a minima at the point of highest popularity of the name.

Without taking into account the influences such as immigration, name saturation, name connotation etc., two ways prove to be effective in determining the parameter values: 1) Using a Gaussian Curve as an estimator 2) Using the Gamma distribution

3.1 Parameter Estimation: Gaussian Estimation

The Gaussian distribution can be defined as $G(t) = \dfrac{1}{\sqrt{2\pi}\sigma}e^{\dfrac{(t-\mu)^2}{2\sigma^2}}$ where σ^2 is the variance and μ is the mean of the distribution.

The parameters can be estimated as follows:

STEP 1: Determine the global maxima of frequency distribution of the name.

STEP 2: Solve the diffonential equation for coefficients a and c at the global maxima.

STEP 3: Fit a Gaussian curve to match the height and time period of the global maxima of co-efficient c.

STEP 4: Fit a Gaussian curve to match the height and time period of the global minima of co-efficient a. In almost all cases, this occurs during the maxima of coefficient c.

The value at a given time t of the resulting Gaussian curves is the estimated value of c and a. In order to maintain an agreeable relationship between the height and width of the Gaussian curves, the Gaussian distribution is scaled by a suitable factor. This rather simple estimation technique works remarkably well for unimodal distributions. For example, consider the name Jennifer. This name gains popularity during the 1940s gathers momentum and reaches a peak during 1970 only to fade away again in the 20th century. Any unimodal name is well captured by the estimator. The name Noah peaked late in the 19th century. This is captured by the Gaussian estimator as shown in figure 2.

The figures show a high correlation between the actual value of the parameters a and c (which would give the actual number of births with that name) and the value of the parameters estimated by the Gaussian estimator.
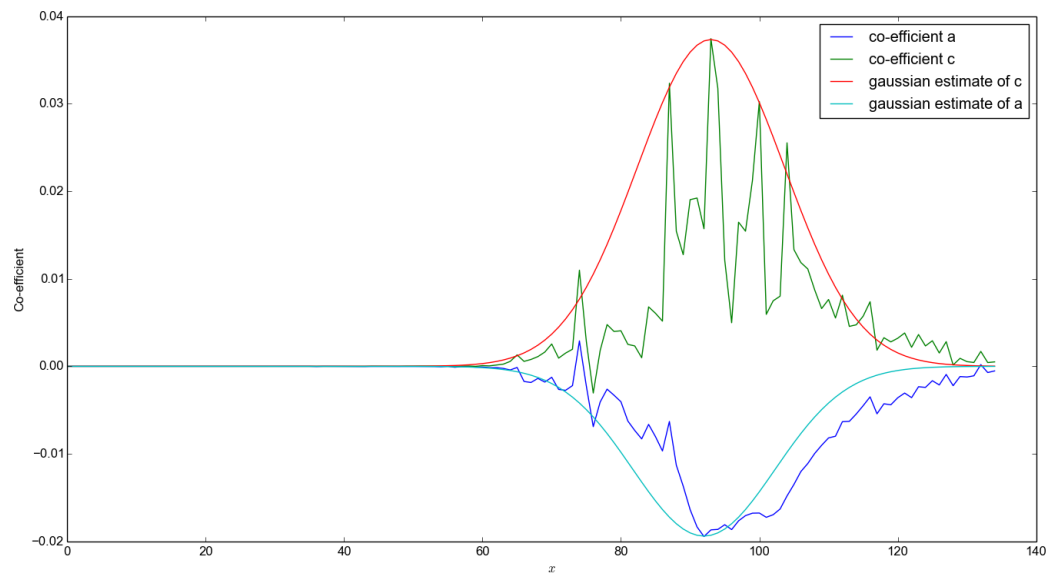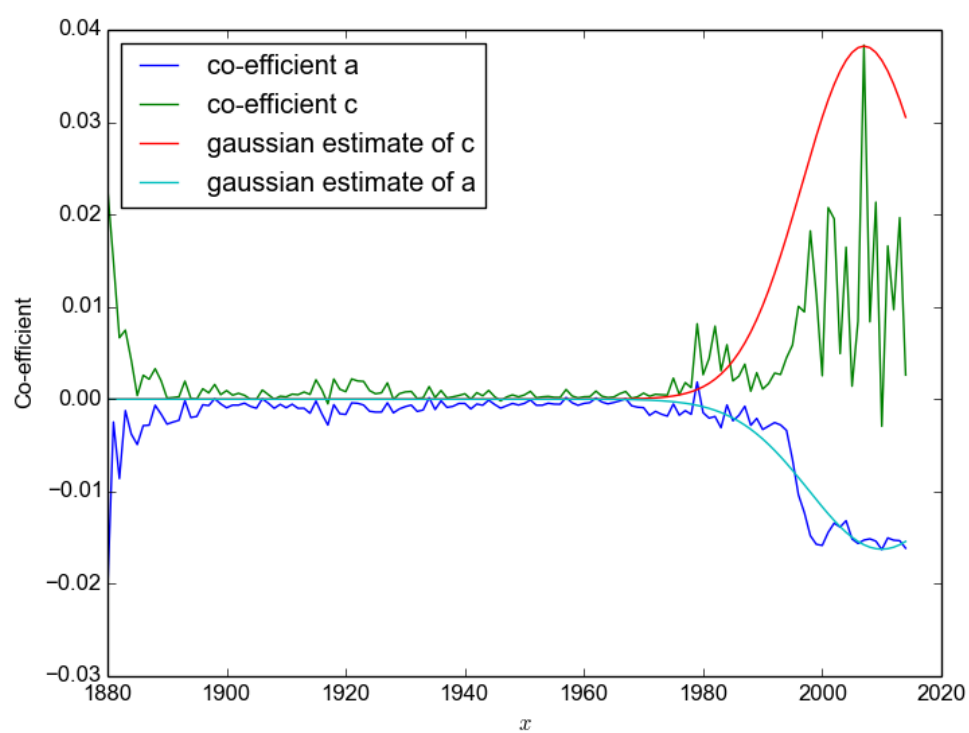
Figure 1: Estimation for Jennifer

Figure 2: Estimation for Noah

3.2   Parameter Estimation: Iterative Gaussian Estimation

A better estimate of the parameters can be obtained by finding the local maxima(or minima) of the frequency distribution of the coefficients and then fitting a Gaussian curve to each of the local maxima(or minima) by the parameter estimation method mentioned above. The Gaussian curves for each local maxima are then added to get a better estimate for the parameters distribution. The window size used to determine local maxima is 34. A smaller window size generally gives better results however, it also tends to cause unnecessary ripples in the distribution of the parameter estimate. The figures 3 and 4 compare the iterative Gaussian estimations with the Gaussian estimations.
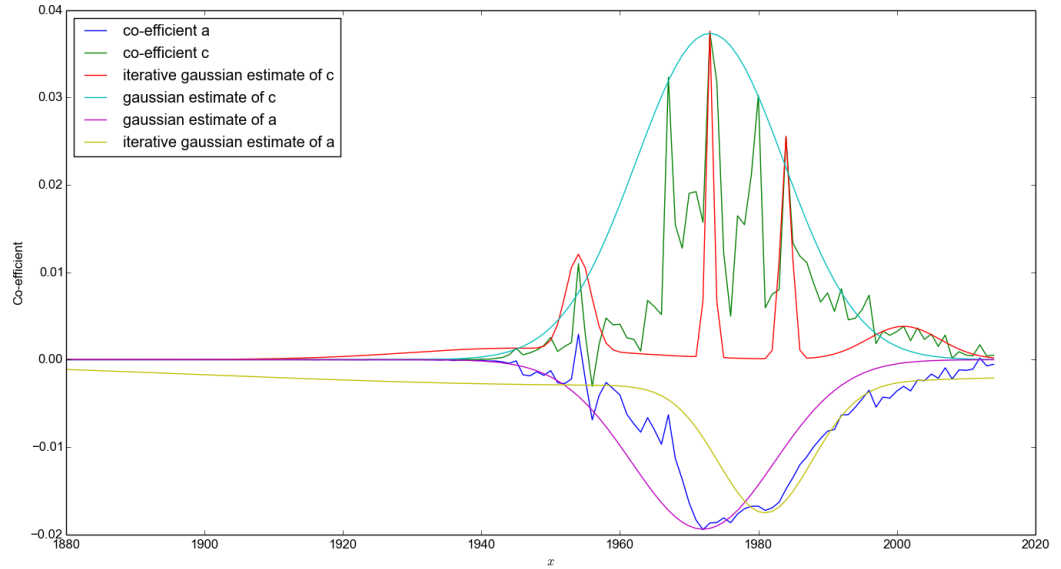

Figure 3: Iterative Gaussian Parameter-Estimation for Jennifer

The difference between the estimated value and the actual value of the parameters a and c compensates for any overestimation/ underestimation as the parameters a and c peak in tandem.

8
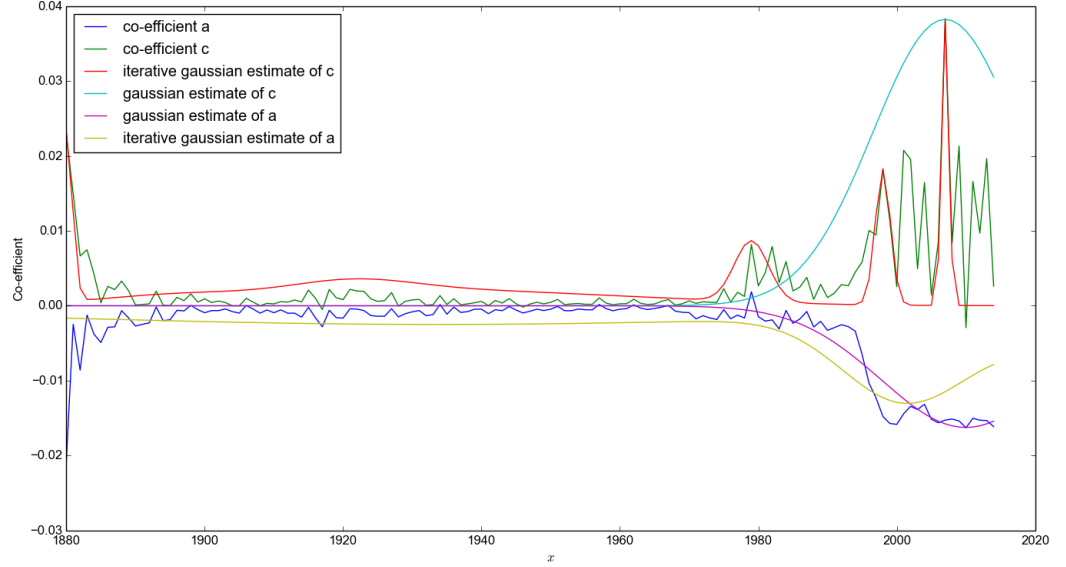
Figure 4: Iterative Gaussian Parameter-Estimation for Noah

3.3   Holt's Model As An Estimator

Holt's model can be defined as

$\quad F_{t+1} = a_t + b_t$ where,

F_t is the forecast for the time period t,

$a_t = \alpha(\dfrac{D_{t+1}}{C_{t+1}}) + (1 - \alpha)(a_t + b_t)$ is the *level*, which represents the smoothed value up to the last data,

$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)(b_{t-1})$ is the estimated trend.

Holt's model is handy because it allows for the inclusion of other parameters that could be used to determine the frequency distribution of a name. A particular variation of the Holt's model, also called the Holt-Winters' model, includes a useful parameter which is often termed "seasonality".

The Holt-Winters' model can be defined as $F_{t+1} = (a_t + b_t)c_t$ where c_t is the *seasonality* and is calculated iteratively by using the relation,

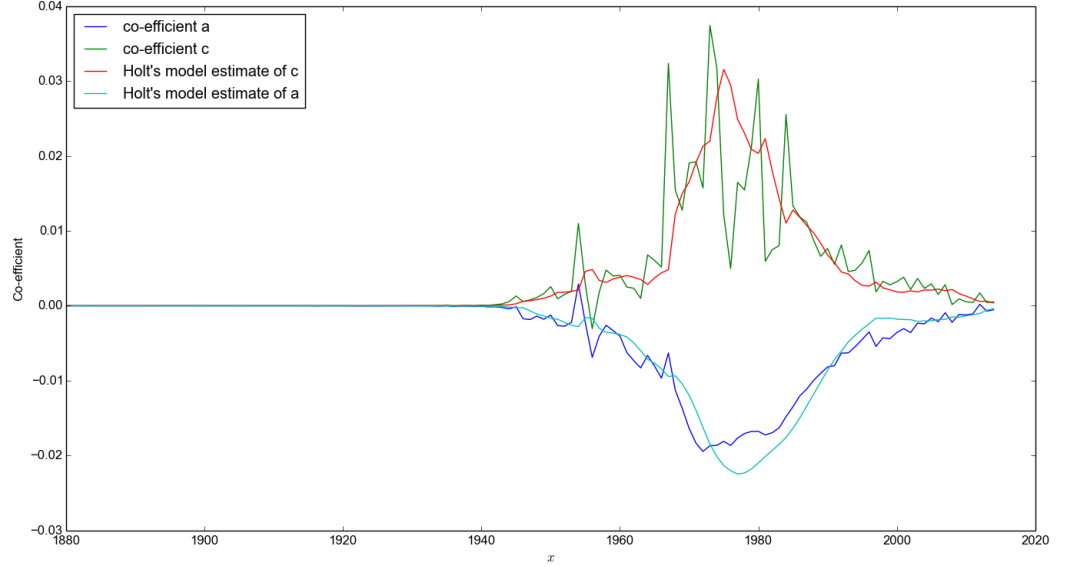$c_{t+p+1} = \gamma\dfrac{D_{t+1}}{a_{t+1}} + (1 - \gamma)c_{t+1}.$

9

Figure 5: Holt's Parameter-Estimation for Jennifer

The Holt-Winters' model works well as such for the prediction of the frequency distribution of the name as such, and, as an estimator for the diffonential distribution coefficients. The seasonality is particularly handy as it can be used to provide localized estimates. For example, consider the name Noah. It started gaining popularity during the late nineties.By considering the first 34 years as one "season" (which leaves us with about 4 seasons), the Holt-Winters' model gives a very accurate prediction. The following figures show the actual parameters and the estimated parameters using Holt's model.

By tweaking the parameters α and β, the model can be used to predict the diffonential distribution parameters with good accuracy. An interesting property about Holt-Winters' model is that the actual data, D_t can be can be replaced by the forecast F_t, in equations for the level (a_t), trend (b_t) and seasonality (c_t). This gives very accurate results with the right window size. This is due to seasonality.
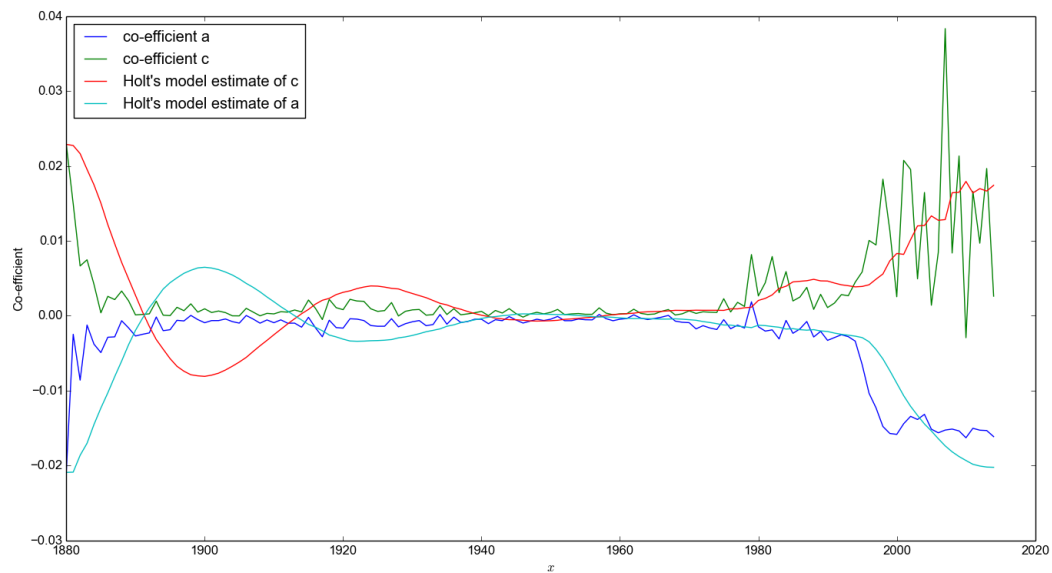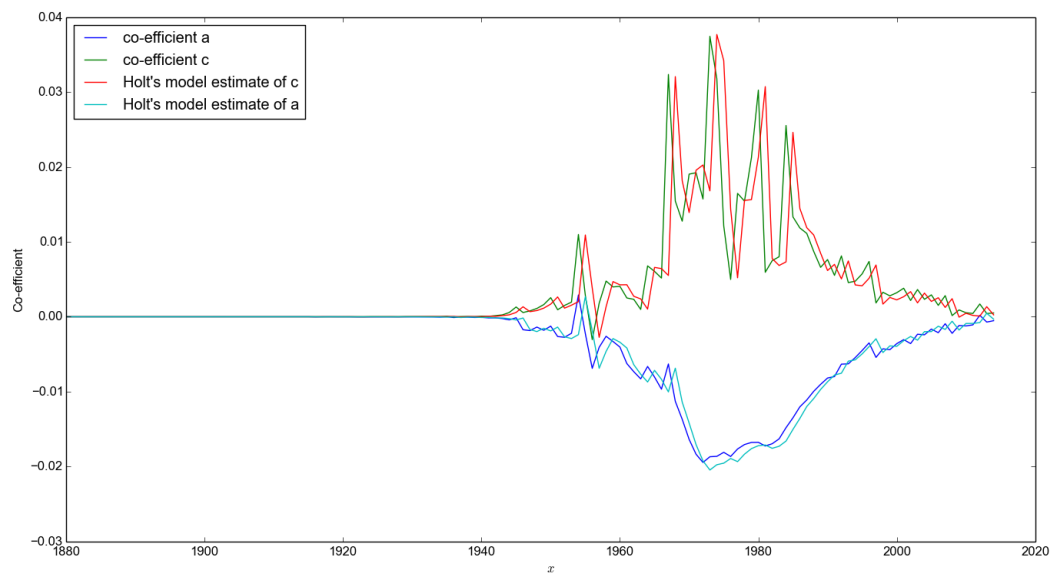
Figure 6: Holt's Parameter-Estimation for Noah

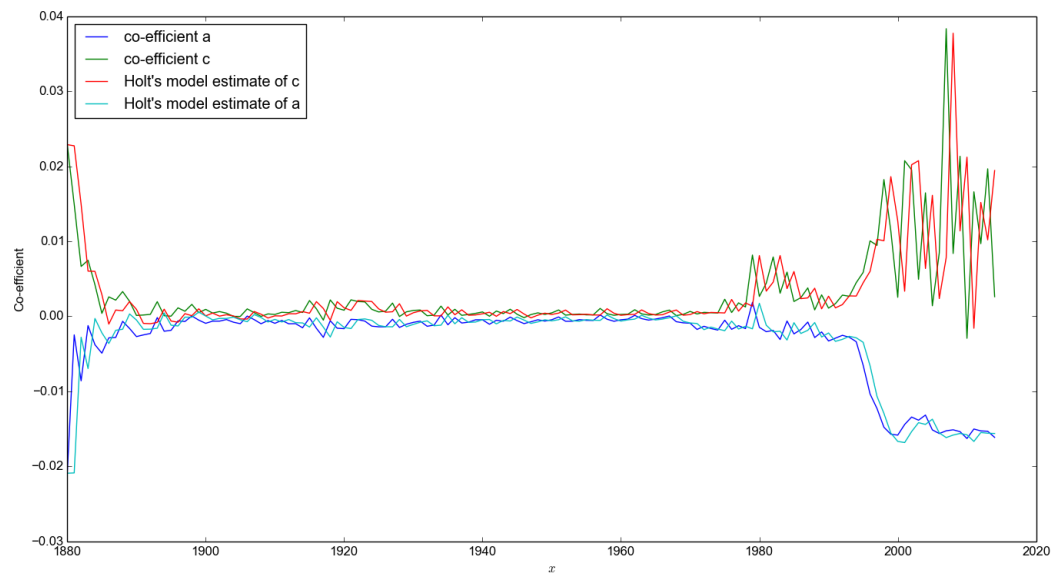Figure 7: Holt-Winters' Parameter-Estimation for Jennifer

Figure 8: Holt-Winters' Parameter-Estimation for Noah

4 Conclusion

The parameters involved in the estimation of the distribution of a name are in many forms and numbers. The focus of this paper was on trying to come up with a satisfactory model to predict the frequency distribution of names. Having studied the properties of the diffonential distribution and the various methods implemented to estimate the parameters for the diffonential distribution, I believe that the Holt-Winters' model gives the best fit for the correct window size and seasonality. For most names the accuracy levels were above 90% for the Holt-Winters' model based estimation, 80% for the Iterative Gaussian based estimation. The Gaussian estimation performed poorly for names which had frequent rise and fall. However, the Gaussian estimation was very effective for names that encountered only a single steep rise during their lifetime thus far.

References

[1] Goffman, W. and Newill, V. A.
Generalization of epidemic theory: An application to the transmission of ideas. Nature 204, 225228 (1964).

[2] Lieberson S (2000)
A Matter of Taste. How Names, Fashions, and Culture Change. New Haven et Londres: Yale University Press

[3] Baptiste Coulmont, Virginie Supervie, and Romulus Breban
The diffusion dynamics of choice: From durable goods markets to fashion first names

[4] http://www.babynamewizard.com/voyager

[5] http://www.bbc.com/news/magazine-26634477

[6] F.W. Preston (Ecology, Vol. 62, No. 2. (Apr., 1981), pp. 355-364).
Pseudo-Lognormal Distributions

[7] Bengt Sigurd, Mats Eeg-Olofsson and Jrgen Ouren
Modelling the changing popularity of names
http://journals.lub.lu.se/index.php/LWPL/article/view/2345/1920