

STA440 Final Project

Matthew O'Donnell

2024-12-01

1. Introduction

The idea of given names is perhaps one of the most interesting parts of our background, as it is fully controllable, but it is not something we choose for ourselves. There has been significant research on the association between naming and both our life outcomes and our perceptions of others. A famous example demonstrates how foreign and complicated names are perceived negatively in comparison to simpler, more familiar names (Laham, Koval, and Alter 2012).

Additionally, names might provide insight into our biases surrounding gender. This is particularly interesting in places like China, where names are often a combination of one or more characters that themselves have their own word meanings. The **ChineseNames** R package, which serves as an aggregation of name frequencies across China, is a valuable resource in this regard (Bao 2023). This data source leverages the National Citizen Identity Information center, the world's largest identity database, to compute aggregate counts for the usage of over 2,000 Chinese characters in given names. The data comes from over 1.2 billion Han Chinese, accounting for almost 97% of the registered population in 2008. The data contain the frequency with which different Chinese characters appear in given names (both raw counts as well as a parts-per-million concentration). It also contains a gender metric for each character (is the character used more in male or female names) and some subjective ratings for how names containing these characters are subconsciously perceived.

This study has two aims. The first is to examine the use of gender-specific and gender-neutral characters and any potential trends over time. The second is to examine how the typical gender usage of characters impact the perceptions of names containing that character (according to the subjective metrics in the dataset). For both of these questions, the provided gender score will be the primary feature of interest.

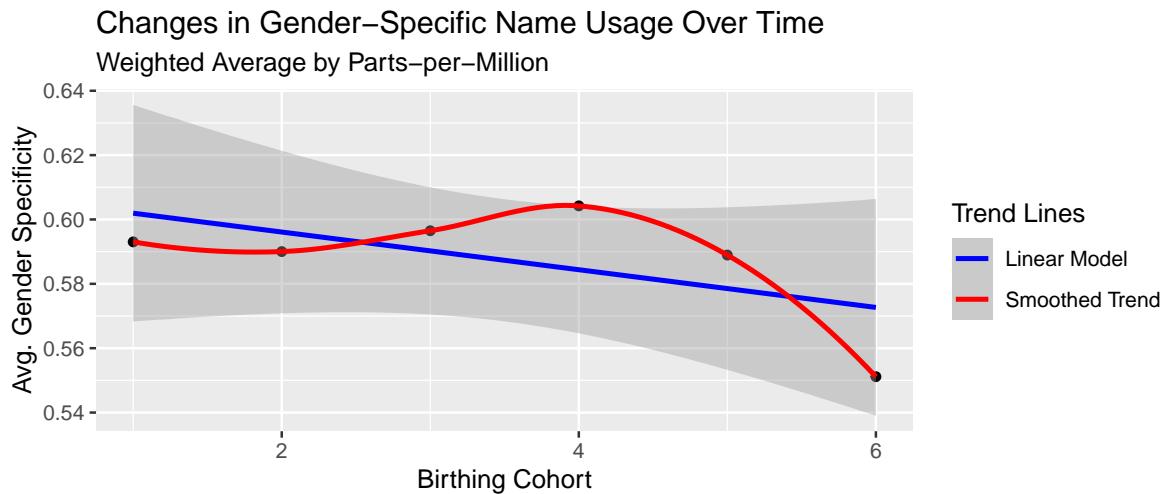
2. Methodology

2.1 Gender-Specific Name Usage Over Time

For each character in the dataset, a gender score is provided, which is computed as follows:

$$g_i = \frac{N_{\text{male}} - N_{\text{female}}}{N_{\text{male}} + N_{\text{female}}}$$

This gender score for a character g_i can range from -1 (completely feminine) to 0 (gender-neutral) to 1 (completely masculine). The data contains a parts-per-million metric for each character during each of six birthing cohorts/generations. A average of gender specificity of a cohort (absolute value of gender score), weighted by parts-per-million in that cohort, could be used to see changes in gender-specificity over time. The birthing cohorts are: 1930-1959, 1960-1969, 1970-1979, 1980-1989, 1990-1999, 2000-2008.



Judging by the smooth trend line, gender-specificity remains relatively constant across birthing cohorts until a sharp drop towards gender-neutrality in the most recent cohort. The best fit linear trend line in blue does trend downwards, but this trend is not statistically significant, evidenced by the standard error band including possible trend lines with positive slopes. Alternatively, an ANOVA (analysis of variance) could be used to test whether there is a statistically significant difference between gender-specificity in different birthing cohorts. The hypotheses are:

$$H_0 : \text{Mean gender-specificity is the same across all cohorts}$$

$$H_A : \text{At least one cohort has a different mean gender-specificity}$$

The data should be weighted by parts-per-million in each cohort, which isn't exactly conducive to running a traditional ANOVA. Instead, we can use the fact that a regression of one categorical

feature is the same as a one-way ANOVA and fit a regression model weighted by parts-per-million. The model equation will be as follows...

$$y_i = \beta_0 + \sum_{k=1}^6 \beta_k \times x_{ik} + \epsilon_{ik}$$

$$\epsilon_{ik} \sim N(0, \sigma^2 / w_{ik})$$

... where $y_i = |g_i|$ represents the gender score for a character, and x_{ik} is an indicator the character is being weighted for Cohort k (each character will appear once in the dataset for each cohort). Each β_k is the difference in expected gender-specificity between Cohort k and those in the 1939-1950 cohort (base case), and w_{ik} represents the parts-per-million for character i in birthing cohort k . An ANOVA can then be applied to the output of this model to get test the original null hypothesis that mean gender-specificity is the same across cohorts. This model makes the assumption that the gender-specificity, remains constant across birthing cohorts (ie the gender specificity of a character remains the same across different periods of time). This is not an ideal assumption to make, but it is required based on the limitations of the provided dataset. All hypothesis tests will be done at the $\alpha = 0.05$ level.

2.2 Subjective Name Characteristics By Gender

The data also provide three subjective metrics that attempt to quantify the sentiment of different characters and the characteristics of people who have them in their names. Here are the metric descriptions provided by the package author:

Name valence (positivity of character meaning): Ranges 1-5, “*Subjective ratings from 16 Chinese raters (9 males and 7 females; interrater reliability ICC = 0.921) on the positivity of all 2,614 given-name characters according to the meaning of each character (1 = strongly negative, 3 = neutral, 5 = strongly positive).*”

Name Warmth/Morality: Ranges 1-5, “*Subjective ratings from 10 Chinese raters (5 males and 5 females; interrater reliability ICC = 0.774) on how a person whose name contains each of the 2,614 given-name characters is likely to have warmth-related traits (1 = strongly unlikely to have, 3 = medium likelihood, 5 = strongly likely to have).*”

Name competence/assertiveness: Ranges 1-5, “*Subjective ratings from 10 Chinese raters (5 males and 5 females; interrater reliability ICC = 0.712) on how a person whose name contains each of the 2,614 given-name characters is likely to have competence-related traits (1 = strongly unlikely to have, 3 = medium likelihood, 5 = strongly likely to have).*”

For given characters, the relationship between gender score and the above subjective assessments can be assessed using weighted least squares. An example equation is as follows...

$$y_i = \beta_0 + \beta_1 \times x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2/w_{ik})$$

... where y_i represents a character's score in a given subjective metric, x_i represents the character's gender score, and β_1 represents the true coefficient for the subjective metric. This univariate model can be fit separately for each birthing cohort and each subjective metric, with w_{ik} representing parts-per-million for character i during cohort k . Once again, these models make the assumption that the gender score for each character remains constant across birthing cohorts. All hypothesis tests will be done at the $\alpha = 0.05$ level.

3. Results

3.1 Anova Results

**Table 1 : Estimated WLS Coefficients
By Time Period Effects on
Gender-Specific Name Usage**

Predictor	Estimated Coefficient ($\hat{\beta}_k$)	P-value
Intercept	0.593	< 0.001
Time Period : 1960-1969	-0.003	0.683
Time Period : 1970-1979	0.003	0.642
Time Period : 1980-1989	0.011	0.144
Time Period : 1990-1999	-0.004	0.597
Time Period : 2000-2008	-0.042	< 0.001

**Table 2 : ANOVA Metrics for Birthing
Cohort (from WLS)**

Independent Variable	F-statistic	P-value
Time Period	10.792	< 0.001

Looking at the estimated coefficients for each birthing cohort, it becomes clear that only the 2000-2008 period displays a statistically significant association with gender-specificity (which is negative, as would be expected based on the chart in 2.1). The results of the ANOVA test present a P-value < 0.001, which permits rejection of the null hypothesis that all cohorts have the same gender specificity. Based on this model, we would expect names of those born between 2000 and 2008 to have a absolute gender-specificity on average 0.042 lower than those born between 1930 and 1959, which is our base case birthing cohort.

3.2 Significance of Characteristic Effects

Table 3 : Valence Coefficients

Time Period	Estimated Coefficient ($\hat{\beta}_1$)	P-value
1930–1959	-0.053	< 0.001
1960–1969	-0.026	0.044
1970–1979	-0.009	0.494
1980–1989	0.001	0.907
1990–1999	0.015	0.251
2000–2008	0.027	0.039

Table 4 : Warmth Coefficients

Time Period	Estimated Coefficient ($\hat{\beta}_1$)	P-value
1930–1959	-0.054	< 0.001
1960–1969	-0.038	< 0.001
1970–1979	-0.032	0.004
1980–1989	-0.026	0.013
1990–1999	-0.022	0.036
2000–2008	-0.037	< 0.001

Table 5 : Competence Coefficients

Time Period	Estimated Coefficient ($\hat{\beta}_1$)	P-value
1930–1959	0.088	< 0.001
1960–1969	0.134	< 0.001
1970–1979	0.172	< 0.001
1980–1989	0.202	< 0.001
1990–1999	0.204	< 0.001
2000–2008	0.206	< 0.001

The relationship between gender score and character valence only appears significant in the first and final birthing cohort. The former displays a negative association and the later a positive one. This would mean a negative association between subjective name valence and male-associated characters has developed into a positive one over time.

Coefficients for warmth and competence appear significant for each period, and the direction of each relationship remains constant across all periods. Table 4 shows associations between

a female-associated characters and a warm perception, while Table 5 shows an association between male-associated characters and a competent perception. Using the models for the 2000-2008 period for example interpretation, we can say the following...

- For names of people born between 2000 and 2008, a 1 unit increase in a character's gender score in the male direction is associated with an average decrease of 0.037 in a character's subjective warmth rating.
- For names of people born between 2000 and 2008, a 1 unit increase in a character's gender score in the male direction is associated with an average increase of 0.206 in a character's subjective competence rating.

Because these are univariate models, there is no need to hold anything constant while making these interpretations. Looking at trends in these coefficients across birthing cohorts, the feminine bias in subjective warmth seemed to be decrease monotonically until a resurgence in the most recent birthing cohort. The male competence bias has increased monotonically in every birthing cohort.

It is important to discuss assumptions of weighted least squares regression when evaluating the validity of these models. The Appendix to this study contains QQ plots to assess normality of residuals and scatterplots of fitted values vs. residuals to assess constant variance of the residuals for valence, warmth, and competence models. Because these are weighted least squares models, the assessment of residuals must be weighted. So, for example, we are interested in the normality of weighted residuals $r_i^w = r_i * \sqrt{w_i}$ where r_i is the residual for a given character and w_i represents the weight of that character in the model. The models for all three subjective metrics do seem to pass checks for constant variance of residuals. However, the QQ plots for all three model types show a residual distribution that is heavy in the tails. These means that the largest residuals are more extreme than expected based on the assumption of normality. This is an asterisk to keep in mind when discussing the model results.

4. Discussion

The first aim of this study was to identify trends in the gender-specificity of Chinese names over time. The results of the ANOVA test shows a statistically significant difference in means between birthing cohorts. This seems to come primarily from the sharp drop off in gender-specificity in the most recent cohort. This is very interesting in the context of changes in gender identity around the world in recent years. Further analysis of younger generations (names of those born after 2008) would provide greater insight into this trend and show whether or not it has continued. As previously discussed, an important limitation of this analysis is the assumption that gender specificity of a particular character is constant across generations. Someone with access to the original database could recompute gender scores during each birthing cohort, which would provide a more accurate sense of gender-specificity and how it has changed over time.

The second aim of this study was to examine associations between characters' gender score and subjective perceptions of people with those characters in their name. Associations between gender and valence (positivity) showed a mixed bag across birthing generations. Models for warmth and competence showed consistent and statistically significant bias across birthing cohorts. The raters were consistently biased toward femininity in assessing warmth and consistently biased towards masculinity when assessing competence. These results appear to uphold traditional assumptions of gender roles, and it is even more interesting to note the masculine competence bias appears to become even more pronounced when the raters assessed the names of younger generations.

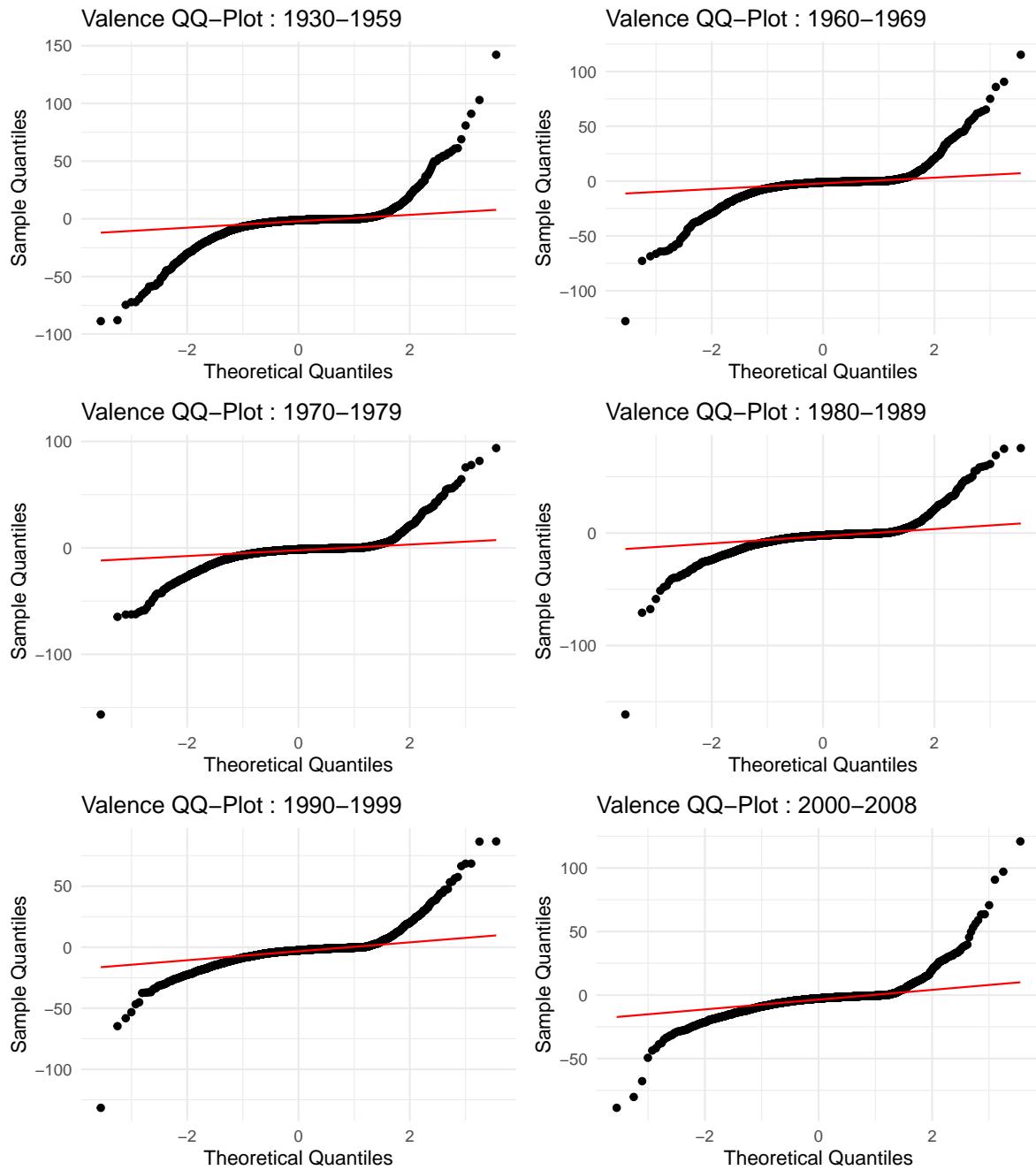
The implications of these results are unclear. It is possible that certain behaviors do manifest in people based on the connotation of their name, but it is also possible that these associations are a product of the raters' gender biases which they subconsciously associate with the typical gender usage of certain characters. Once again, these models might be better informed by gender score specific to birthing cohort.

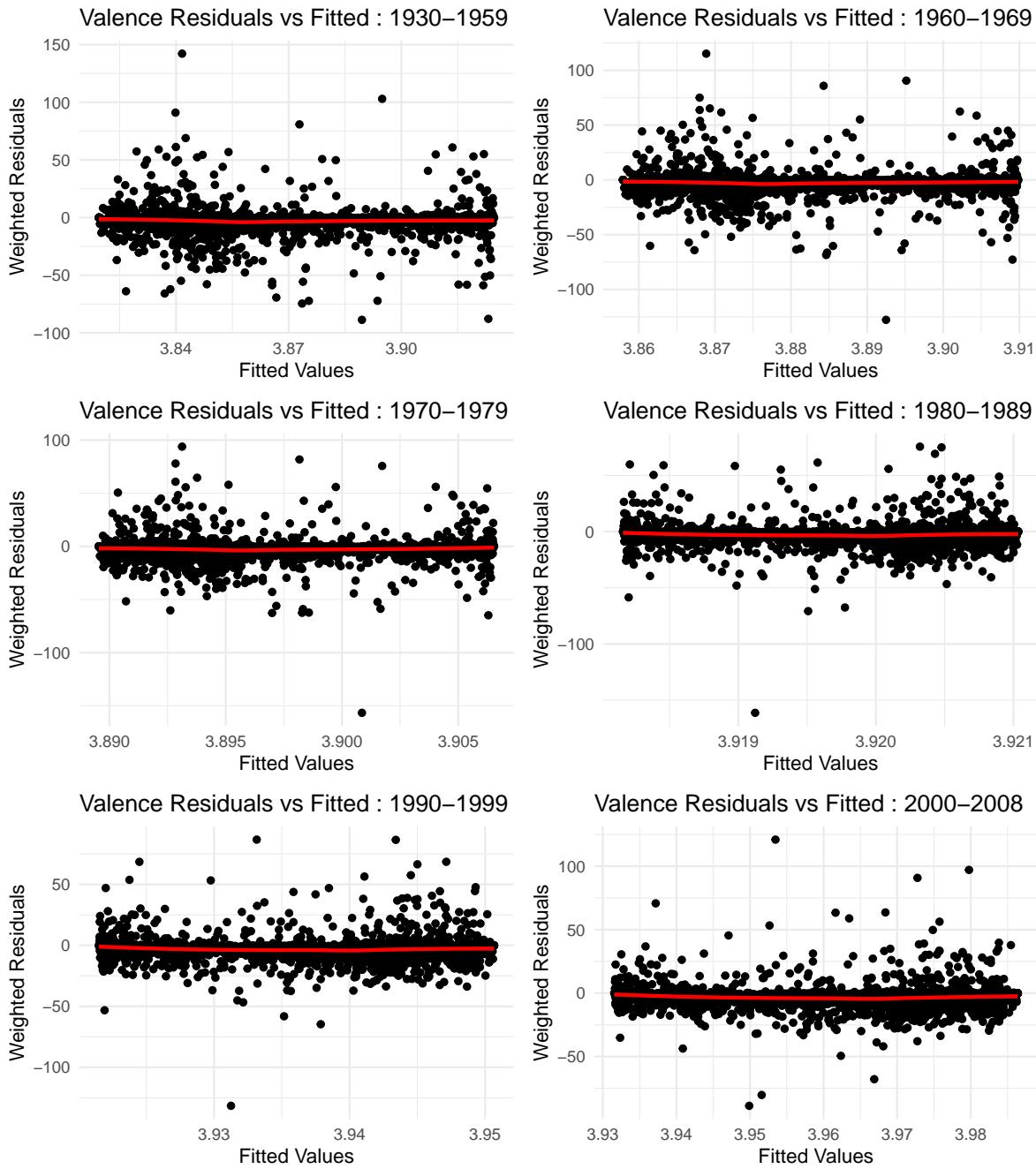
The analysis of this study could be further developed in many directions. The first would be a comparison of these subjective associations to more objective ones. Since each character should have its own individual word meaning, each character could perhaps be translated into its dictionary definition. These definitions could then be passed to some pre-trained language models to assess sentiment. The gender bias of these raters could be more concretely identified if shown to significantly differ from the assessment of a more objective language model. This kind of analysis could be done with the available data, but would require some external translation and sentiment libraries.

Another avenue for further analysis would require additional data that connects these names to socioeconomic outcomes (or, even better, some standardized metric of socioeconomic mobility, demonstrating how far people have come from where they started). In conjunction with an objective sentiment analysis described above, this could show whether naming sentiment is associated with differing life trajectories.

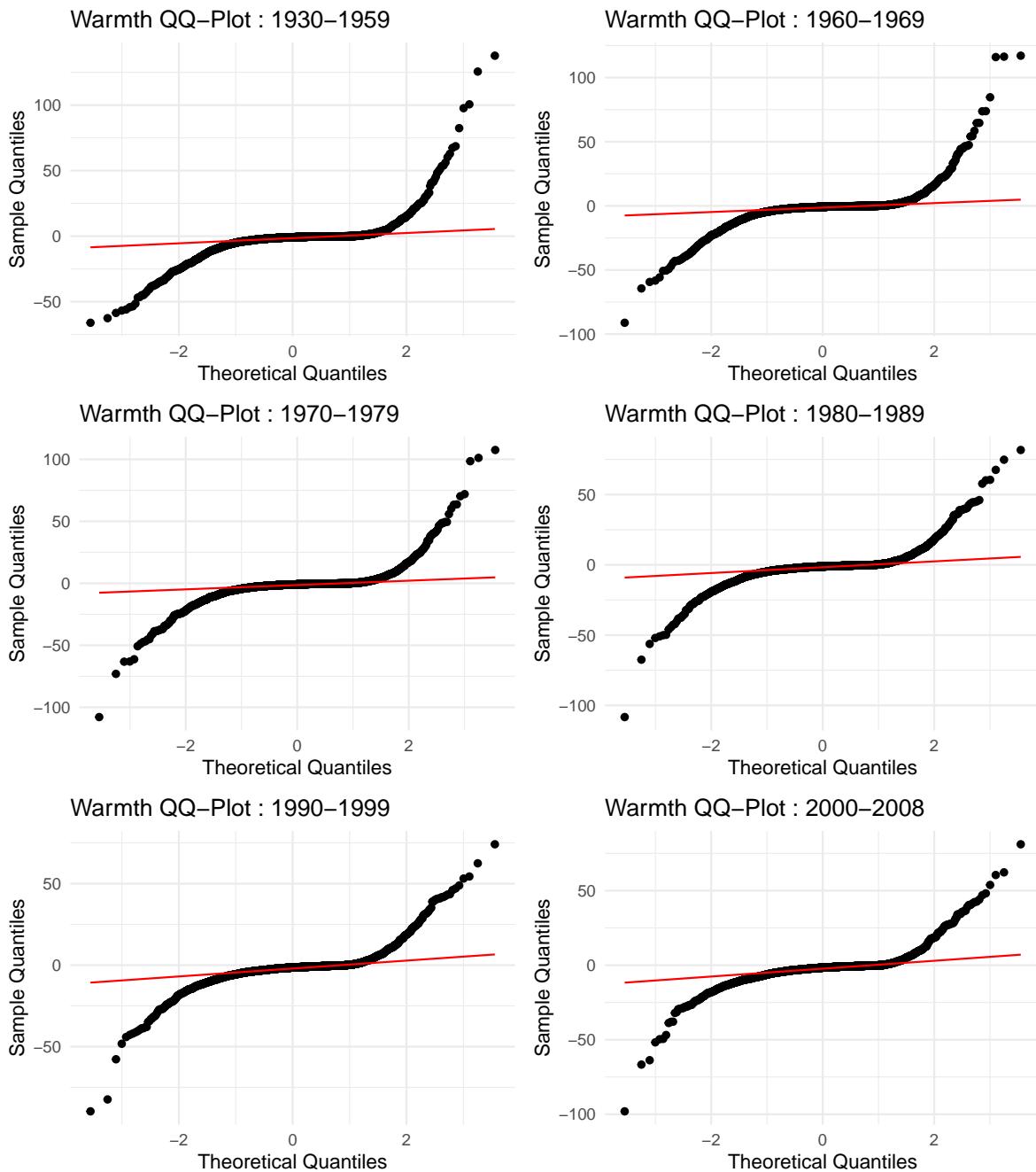
Appendix

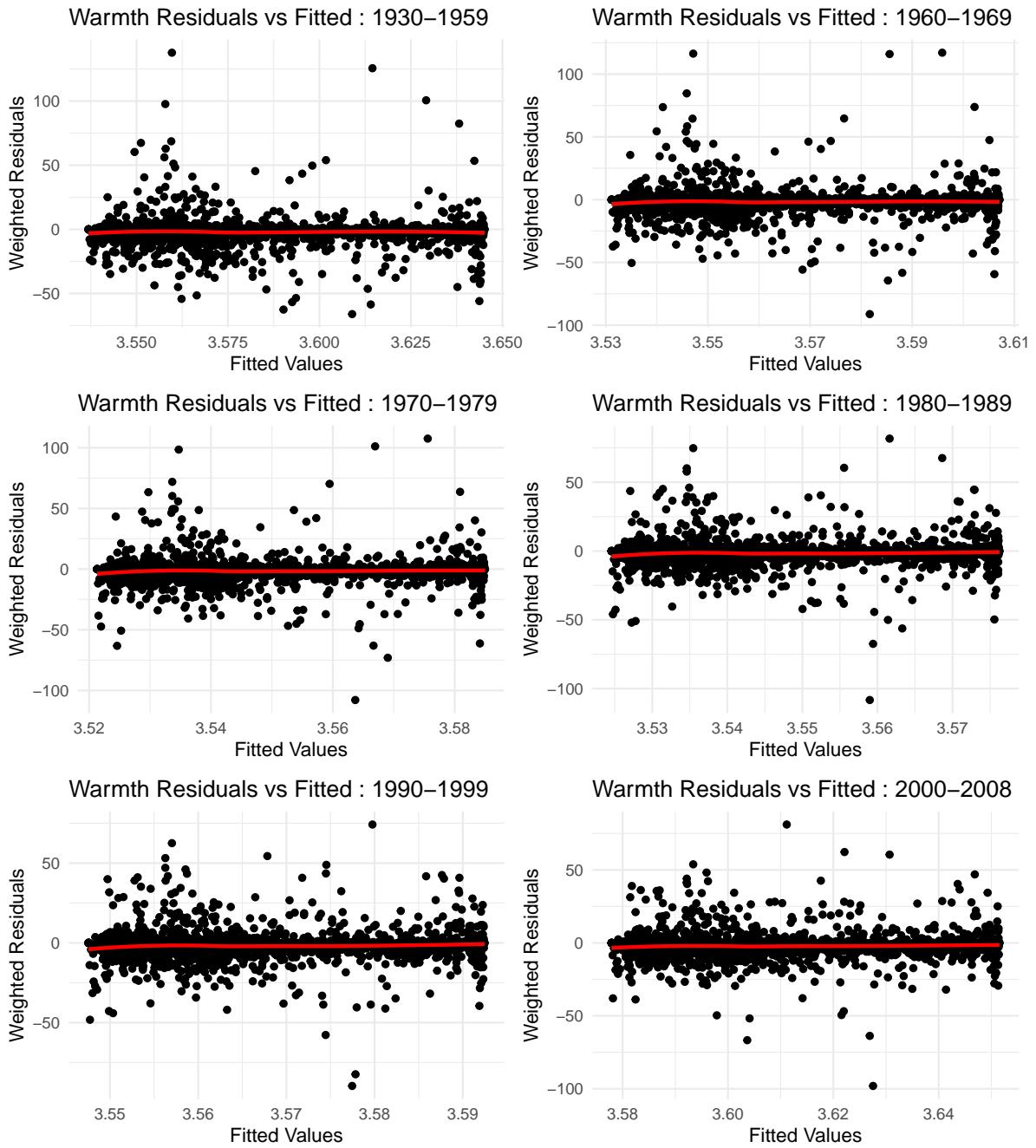
Valence Model Checks



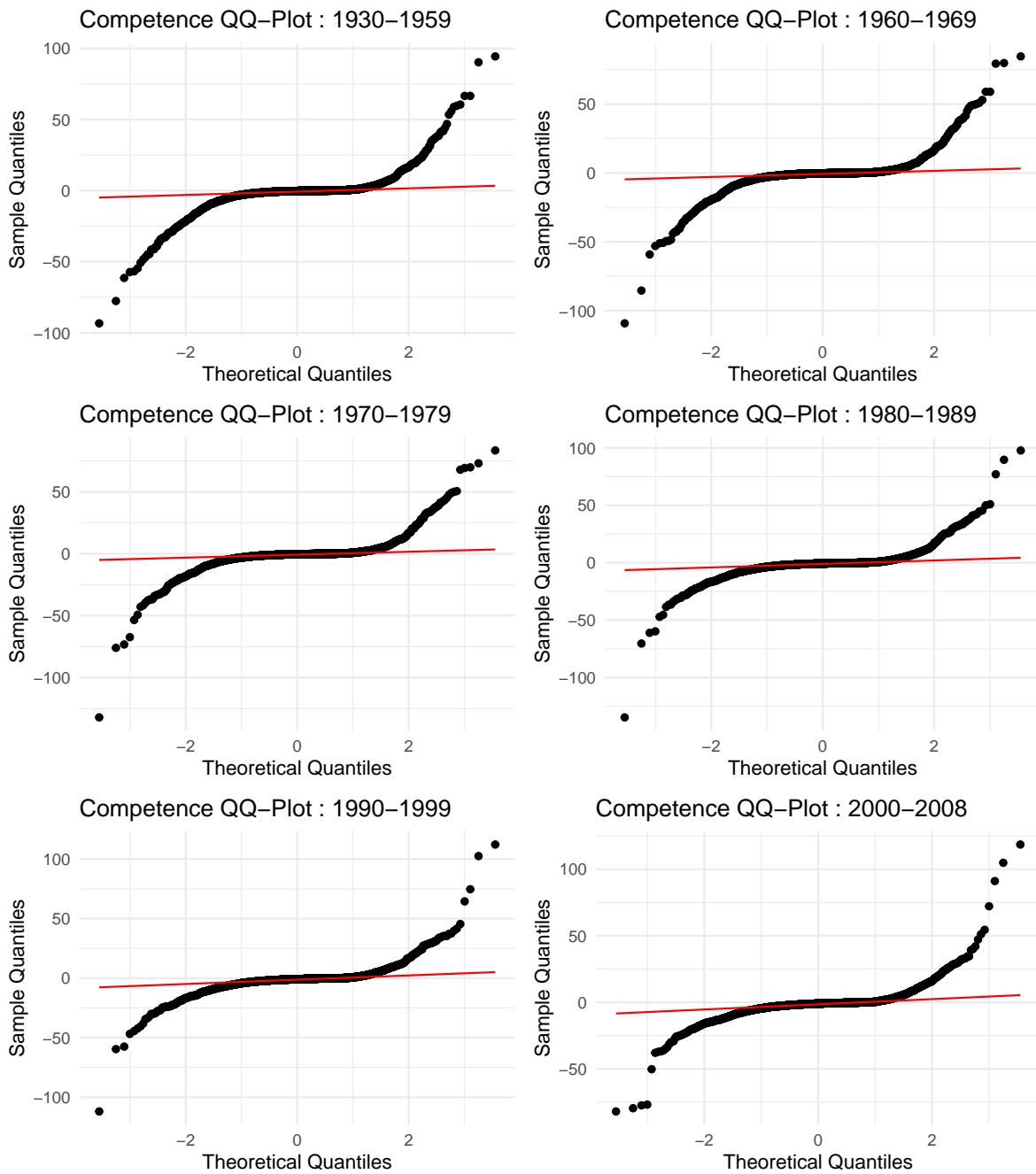


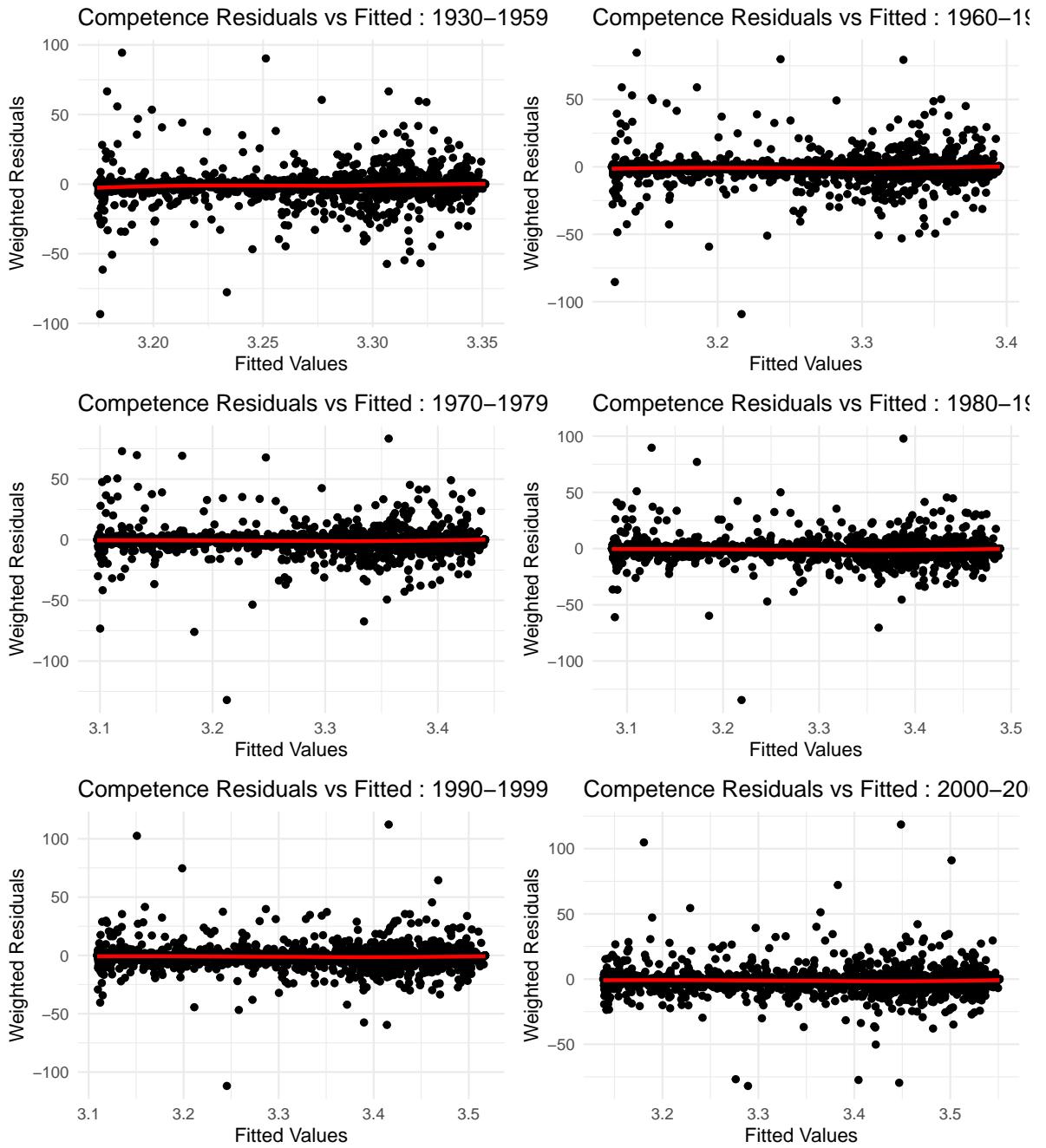
Warmth Model Checks





Competence Model Checks





Bibliography

- Bao, Han-Wu-Shuang. 2023. "ChineseNames: Chinese Name Database (1930–2008)." <https://github.com/psychbruce/ChineseNames>.
- Laham, Simon M., Peter Koval, and Adam L. Alter. 2012. "The Name-Pronunciation Effect: Why People Like Mr. Smith More Than Mr. Colquhoun." *Journal of Experimental Social Psychology* 48 (3): 752–56. [https://doi.org/https://doi.org/10.1016/j.jesp.2011.12.002](https://doi.org/10.1016/j.jesp.2011.12.002).