

Pursuit of Happiness: The impact of authoritarian government policies on Chinese well-being*

Alexander Sun

April 23, 2024

In this study, we employed a Bayesian regression model to analyze the impact of government surveillance, email monitoring, information collection, national pride, and age on happiness levels using data from the World Values Survey. The findings indicate that higher levels of government surveillance and email monitoring are slightly negatively associated with happiness, whereas higher levels of national pride shows a strong positive relationship with happiness levels. Furthermore, age was positively correlated with happiness, suggesting that older individuals tend to report higher levels of well-being. Our results illustrate how support for several authoritarian governmental policies may impact overall well-being in a socialist nation.

Table of contents

1	Introduction	2
2	Data	3
2.1	Data Source and Measurement	3
2.2	Variables	3
2.3	Data Analysis	4
3	Model	5
3.1	Model set-up	7
3.2	Model justification	8
4	Results	8

*Code and data are available at: https://github.com/alexandersunliang/china_wellbeing2bugfix.git

5 Discussion	11
5.1 Model Findings	11
5.2 Ethical Considerations	12
5.3 Weaknesses and next steps	12
A Appendix	14
References	15

1 Introduction

Since the early 1990s, China's economy has grown significantly, with an average annual GDP growth rate of around 10%. The 1990s market reforms aimed to control capitalism, showing that socialism could harness capitalist growth to increase national power. Today, China is a global economic leader, second only to the US in GDP; however, it is essential to assess how these rapid changes affect societal well-being. Studies present mixed views on the effects of socialism and democracy. Cheung and Leung found that collectivism significantly impacts the well-being of the Chinese, whereas Hofstede (1980) observed that individualism encourages economic growth [Hofstede (1980)](Cheung and Leung 2007). Moore (2005) argued that the freedoms introduced in the 1980s led the Chinese towards a strong form of individualism, which starkly contrasts with the collectivist approach of the Cultural Revolution (Moore 2005). Our study will reassess a previous paper by Steele and Lynch, who examined whether individualist or collectivist traits prevail in modern China and their impact on subjective well-being (SWB) (Steele and Lynch 2013). They focused on how these orientations predict SWB and if their influence has shifted as China transitioned towards a market economy.

While Steele and Lynch researched the impacts of the reform, this paper will attempt to reframe their ideas with a focus on the recent developments in China. This study will extend Steele and Lynch's previous research by examining the prevailing sentiments—individualist versus collectivist—in China in the past few years and their implications for SWB. We analyze whether the relationship between these cultural orientations and well-being has evolved in the context of China's continued economic and social transformations.

In this project, the estimand will be the difference in happiness level between each participant based on information like national pride, government support, and age.

The remainder of the paper is structured as follows: following in the introduction in Section 1, Section 2 provides an overview of our dataset, variables, and a quick analysis of the data, Section 3 discusses the model we chose for our regression, Section 4 depicts the calculated values along with any important measurements, and Section 5 discusses the implications of our results while also suggesting avenues for future research.

2 Data

2.1 Data Source and Measurement

To continue our paper based on Steele and Lynch, we obtained our data from the World Values Survey. The World Values Survey is the largest non-commercial international investigation of human values ever executed. There are no other datasets of a similar size that target human beliefs and opinions. While Steele and Lynch used time series data from three surveys (1990, 2001, 2007), we aim to look at more recent trends so we will include the World Values Survey data obtained from their 2018 wave. To be included into the survey dataset, countries must meet a minimum of 1200 entries. Since the survey is global, general guidelines are provided to the teams that are responsible for each country, but slight alterations in sampling method were permitted. The most common method of data collection consisted of face-to-face interviews with respondents at their home, and the answers were recorded on a paper questionnaire for later processing. For China, participants were chosen through probability sampling. The initial dataset contained 94278 responses and 606 variables, but after filtering for responses from China and removing non-responses, we are left with 3036 observations. Each wave of the survey is slightly different, so the wording of some questions and answers may vary. Although the sample for China was chosen through random sampling, survey guidelines dictated that there would be no replacement for non-responses. Therefore, missing data had to be parsed out.

2.2 Variables

The original questionnaire was composed of approximately 300 questions, but for the purposes of our study we will focus on the following:

- **Happiness:** Our dependent variable was measured as “life satisfaction” in the questionnaire. Participants responded with how happy they were with their current lifestyle on a scale of 1-10 with 1 being “completely dissatisfied” and 10 being “completely satisfied”. We will later conduct a model to estimate this value based off our other regressors.
- **National Pride:** Respondents were asked about their national pride with the question, “How proud are you to be Chinese?” We believe that this is especially important as it allows us to see temporal trends in nationalism with regards to increasing Western criticism.
- **Government Surveillance in Public Areas:** This question asks if participants believe that the government has the right to surveil the populace in public areas. As this is one of the largest criticisms against the modern Chinese government, we can potentially see impacts of this variable with respect to life satisfaction for citizens. This variable is recorded on a scale of 1-4, with one being “Definitely should not have the right” and four being “Definitely should have the right”

- **The Right for the Government to Monitor Emails:** This variable measures population sentiment for the governments ability to monitor private emails. While the last variable evaluated opinions on surveillance in public areas, this one angles more towards public sentiment on the collection of private information. This was also recorded on a scale of 1-4, with one being “Definitely should not have the right” and four being “Definitely should have the right”.
- **Should the Government Take More Responsibility to Ensure Everyone is Provided For?:** As the divide between communist and democratic ideals grow, this question will provide insight into the collectivist support among the general public in China. Although not a perfect capture of sentiment towards communism, this is the closest question that was able to be asked on the survey without running into political issues. This was recorded on a scale of 1-10, with one being the respondent completely agrees with the question and ten being the respondent completely disagrees.
- **Age:** With the onset of social media, especially TikTok/Douyin, Chinese citizens are being exposed to more Western media than ever before. By measuring age, we can see if specific generations are more progressive than others. The World Values Survey limits the age of its participants to between 18 and 70.

2.3 Data Analysis

Our investigation was conducted using the programming language R (R Core Team 2023), with the packages tidyverse (Wickham et al. 2019a), arrow (Richardson et al. 2023), modelsummary (Arel-Bundock 2022), here (Müller 2020), kable (Zhu 2024) and dplyr (Wickham et al. 2023). Techniques for some analysis methods were taken from Telling Stories with Data (Wickham et al. 2019b).

	Mean	StdDev	Min	Max	IQR
Year	2018.00	0.00	2018	2018	0
National_Pride	1.64	0.64	1	5	1
Happiness	7.42	2.02	1	10	3
Government_Surveillance	1.82	0.89	1	4	1
Email_Monitoring	2.30	0.98	1	4	1
Information_Collection	2.48	1.04	1	4	1
Government_Responsibility	5.22	2.82	1	10	5
age	44.46	14.48	18	70	24

Initial observations of the summary statistics in the table above reveals that the mean happiness level is quite high, averaging to approximately 7.42 as opposed to the expected 5. We also notice that national pride is very high, with a mean of 1.64. The interquartile range of 1-2 shows that vast majority of the populace are very proud of their status as Chinese citizens.

The support for the various forms of government control seem middling, but it shows that the public is not vehemently opposed to these methods of surveillance.

The distribution for each of our variables is depicted in Figure 1:

From the distribution graphs shown in Figure 1, we observe that barely any of the questions are normally distributed as we would expect for such a large sample. Although this could be attributed to the limited range of answers, we notice that even for questions with a larger range of responses like happiness and government responsibility, the data seems very skewed to one side.

Furthermore, the dataset provides information on how responses to question differ based off age. From this, we can visualize how different age groups tend to respond to questions. For example, we expect the younger generation to be more progressive, leading to higher disapproval rates for government surveillance policies. Information on age distributions on support for government surveillance is shown in Figure 2.

Surprisingly, we see that as age increases, a higher proportion of adults in that age group are opposed to government surveillance in public areas. As the question asks about government surveillance in public areas, we can potentially infer citizens may oppose government email/information supervision as those are more private. However, as seen in Figure 5, we notice that the support for the collection of information for anyone in China is significantly higher than the other variables.

3 Model

Using the World Values Survey data to predict happiness is based on the idea that certain key variables, such as attitudes towards governance and societal structures, are able to predict an individual's level of happiness. Our model will hopefully be able to show how and if specific perceptions on government surveillance acceptance, email monitoring, and information collection are indicative of broader societal satisfaction.

Before we develop our model, we have to assume that there is a linear relationship between our variables and happiness level. Research consistently indicates that the trust in, and perception of, governmental and societal systems significantly influences personal well-being (Helliwell and Putnam 2004). We postulate that individuals' responses to questions about governance and privacy are not merely reflective of their current state but are shaped by their broader socio-political environment and personal experiences. Therefore, a person's level of happiness is not only a function of their immediate circumstances but also a reflection of their interactions with the societal and political structures they inhabit (Bjørnskov 2006).

To assess the validity and strength of these relationships, it is crucial to consider confounding factors that may be related to both the predictors and happiness. For instance, demographic variables such as age may influence both a person's perception of governance and their reported

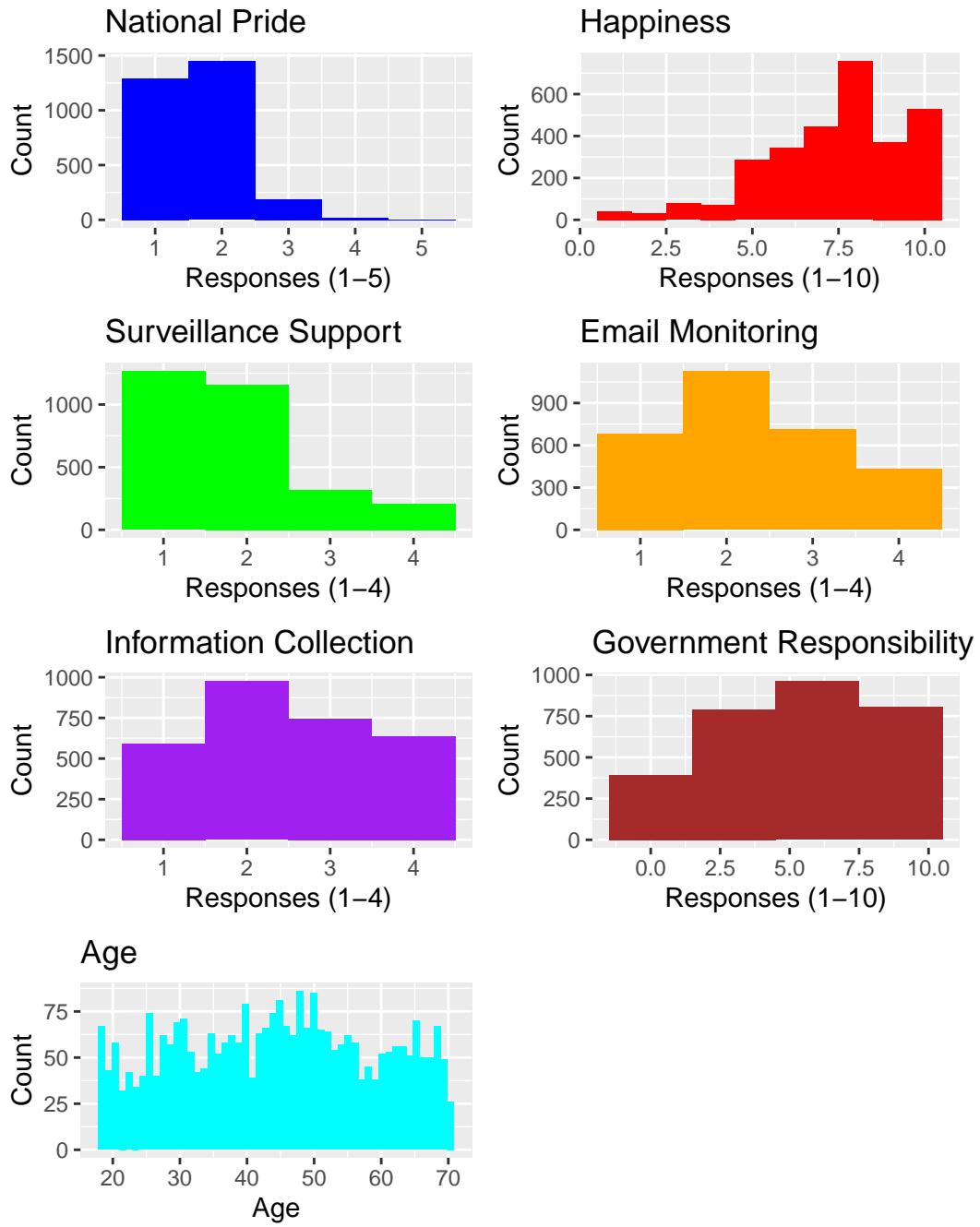


Figure 1: Assorted Variable Distributions

Government Surveillance Support by Age Group

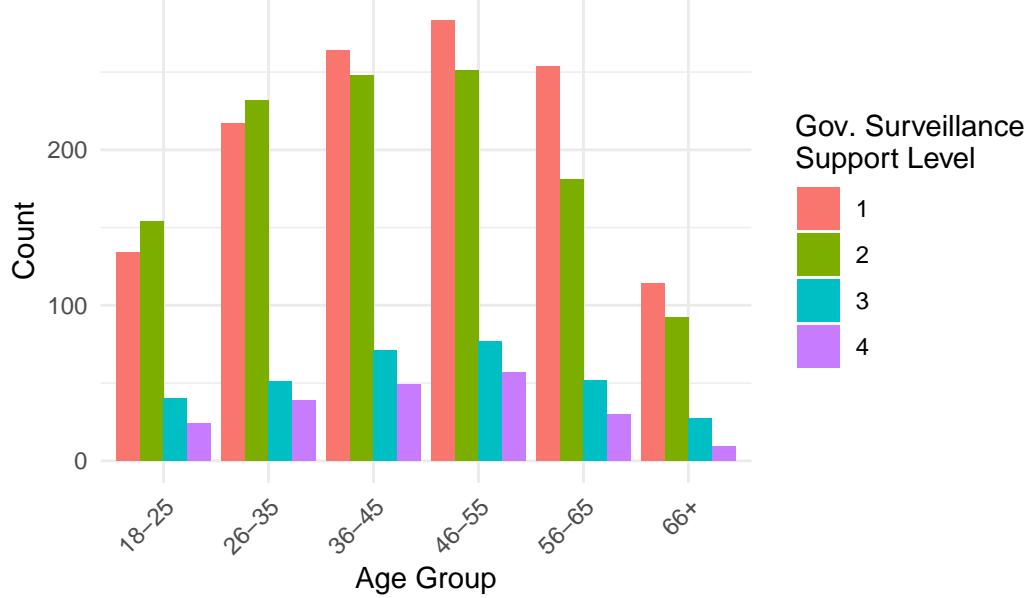


Figure 2: Support for Government Surveillance Between Ages

level of happiness. Older individuals may have different views on privacy and security, and simultaneously, their happiness levels might be affected by a host of age-related factors, making age a potential confounder that needs to be accounted for in our model.

Furthermore, cultural factors could moderate the relationship between governance attitudes and happiness. For example, in societies where collective well-being is emphasized, the acceptance of surveillance measures might correlate differently with happiness compared to societies that prioritize individual autonomy (Diener et al. 1999). We expect that the effects for the government policies and age to be positive, while national pride is negative.

3.1 Model set-up

Define y_i as the estimated level of happiness. Then β_0 would be the estimated level of happiness given that all other variables are zero, β_i would be the respective effects on estimated levels of happiness for each of our variables. The full model form is shown below:

$$y_i = \beta_0 + \beta_1 \cdot \text{Gov Surveillance}_i + \beta_2 \cdot \text{Email Monitoring}_i + \beta_3 \quad (1)$$

$$+ \beta_4 \cdot \text{Age}_i + \beta_5 \cdot \text{Pride}_i + \beta_6 \cdot \text{Gov Responsibility}_i + \epsilon_i \quad (2)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.2 Model justification

In our study, we employ a Bayesian regression model to predict happiness levels using the World Values Survey data. Given that happiness is influenced by a diverse range of factors, a model that can handle multiple regressors is critical in understanding the relationship with variables such as government surveillance acceptance, email monitoring acceptance, information collection acceptance, and happiness. Furthermore, the model allows us to isolate the unique contributions of each predictor to happiness. We specifically chose a Bayesian regression model as opposed to other forms of regression, such as a Poisson. This is because a Poisson regression assumes that the variance of the dependent variable is equal to its mean, which restricts its application to data where this condition naturally holds. From our summary statistics in Section 2, we see that this assumption does not hold.

Our large sample size—approximately 3000 respondents—strengthens the model further. A robust sample size ensures that our parameter estimates are reliable and reduces the likelihood of statistical errors. It also increases the generalizability of our findings, ensuring that our conclusions are not just applicable to a narrow subset of the population but can be extended to China as a whole. Additionally, a larger dataset supports the complexity of our model, allowing us to include multiple predictors without the risk of overfitting, thereby maintaining the model’s integrity and robustness. The overall statistical significance of our test will be examined further later in this study.

4 Results

Our results are summarized in Table 1.

The regression analysis presented in Table 1 reveals an expected baseline happiness level of 7.60, as indicated by the model’s intercept. This suggests a generally positive baseline level of happiness among respondents when all other variables are held constant. The coefficients for the support for government surveillance, email monitoring, and information collection are slightly negative, at -0.06, -0.01, and -0.01 respectively, suggesting a marginal decrease in happiness scores associated with increased acceptance of these governance measures. Notably, national pride shows a stronger negative association with happiness, with a coefficient of -0.50, indicating a significant decrease in happiness as national pride increases. In contrast, the age coefficient of 0.02 suggests a slight increase in happiness with age.

The model accounts for a modest 4.8% of the variance in happiness scores, as reflected by an R-squared value of 0.048, implying that other unexamined factors may play a more substantial role in influencing happiness. The adjusted R-squared value, at 0.043, indicates a minimal

Table 1: Regression Model Results for Predicting Happiness Based on Government Perceptions and Age

	(1)
(Intercept)	7.23 (0.21)
Government_Surveillance	-0.06 (0.05)
Email_Monitoring	0.00 (0.05)
Information_Collection	-0.01 (0.04)
National_Pride	-0.50 (0.06)
age	0.02 (0.00)
Government_Responsibility	0.05 (0.01)
Num.Obs.	2950
R2	0.054
R2 Adj.	0.048
Log.Lik.	-6186.433
ELPD	-6193.8
ELPD s.e.	44.4
LOOIC	12 387.7
LOOIC s.e.	88.8
WAIC	12 387.7
RMSE	1.97

increase in explanatory power when accounting for the number of predictors. Furthermore, the Root Mean Square Error (RMSE) of 1.98 points to the model's prediction errors' standard deviation, signifying potential areas for model refinement. These results provide a statistical basis for the model's performance, emphasizing the nuanced challenge of predicting subjective well-being through perceived governance and age.

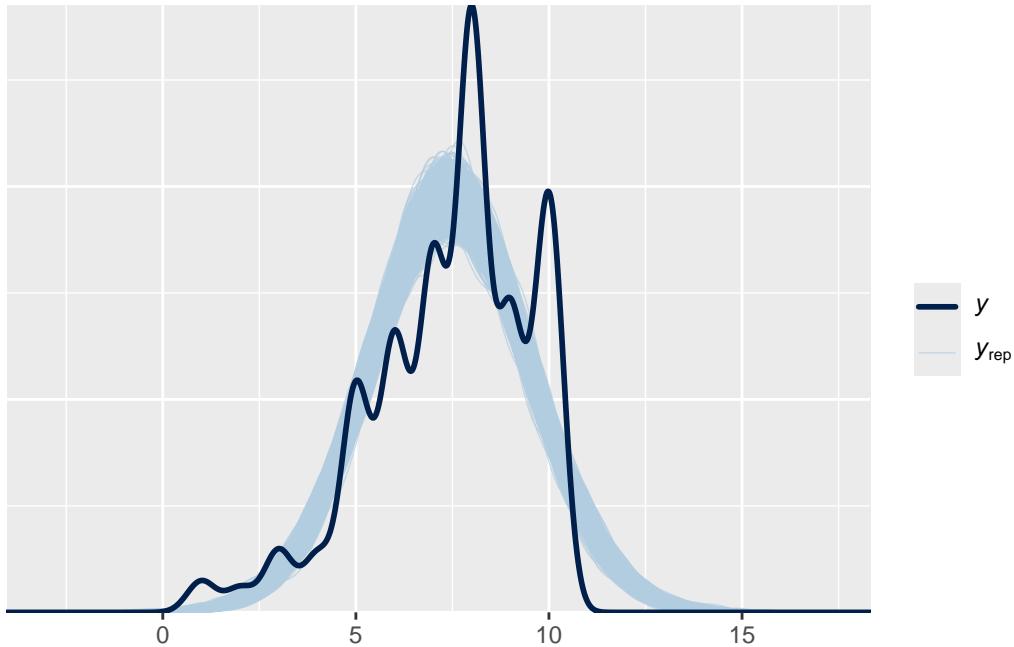


Figure 3: Happiness Model Posterior Predictive Check

The posterior predictive check, visualized by Figure 3, helps us evaluate the fit of our Bayesian model. By overlapping the observed happiness data with the simulated posterior predictive values, this plot illustrates how well our model reproduces the observed outcomes. The congruence between the values signifies that the model adequately captures the central tendency and variability of the actual happiness scores. Notably, if there are significant deviations it could indicate that the model may not fully account for the data's characteristics, prompting us to consider model refinements or additional predictors.

The trace plot shown in Figure 4 provides a representation of the Markov Chain Monte Carlo (MCMC) sampling process, essential for evaluating the convergence of the chains. Each parameter's trace, represented by individual lines for each chain, should display a "mixing" behavior, converging to and fluctuating around a common value, without exhibiting trends or periodicity. Such patterns indicate that the chains have reached a stationary distribution, and the samples drawn are likely from the true posterior distribution. If the chains do not converge, this implies that the sampling process has not yet adequately captured the full behaviour of the true distribution, and further iterations or model adjustments may be required to ensure

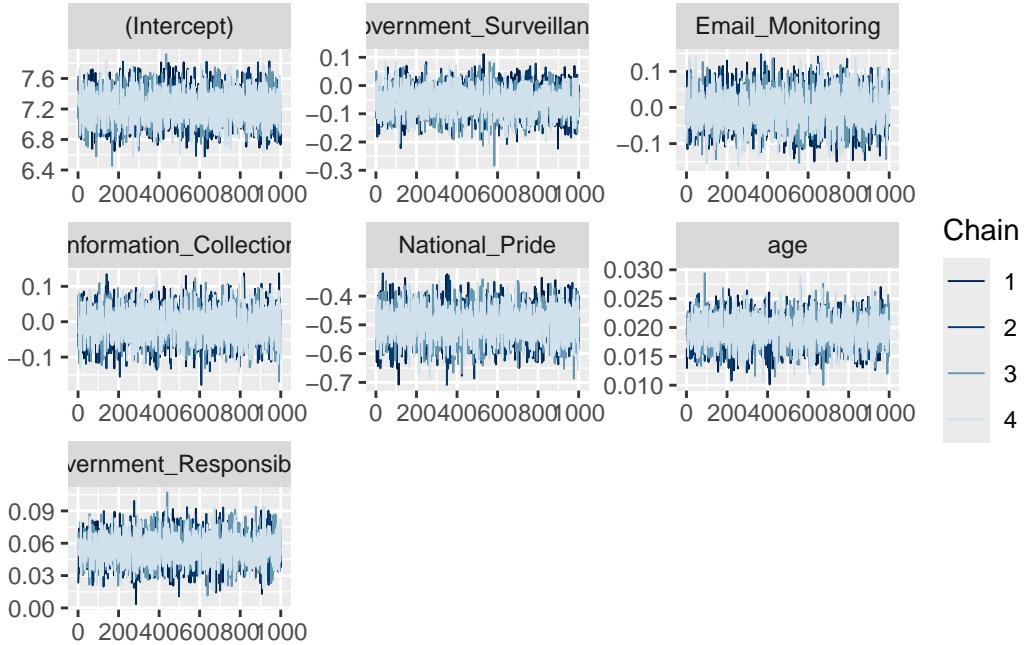


Figure 4: MCMC Trace Plot

reliable inference.

5 Discussion

5.1 Model Findings

As shown in our Section 4 section, we found that the initial baseline measure for happiness in China was quite high, at a predicted level of 7.60 given all other variables are at zero. We discover that our results disagree with Helliwell and Putnam’s (2004) findings where belief in government policies are positively correlated with well-being. Our model suggests that an increase in support for any of the three policies correlates with a slight decrease in overall happiness level. However, we find that although national pride is negatively correlated with our measure of happiness, once we factor in that national pride is measured on a scale of 1-5 with 1 being the highest, we realize that national pride is the largest indicator of happiness out of our chosen variables. This could be because we are exclusively looking at a population of Chinese citizens. It is expected that residents and citizens that do not pride themselves in being Chinese are generally less happy while they are living there. Furthermore, we see a very slight positive coefficient for our age variable, suggesting that as Chinese citizens grow, they become more satisfied with their life. This could be potentially attributed to the stressful

career and educational lifestyle of China. With a culture of overworking compounded with the pressure of the Gaokao, young citizens may experience lower levels of overall happiness compared to the older generation.

From our results, we see that even with growing discontentment towards China in the Western world, national pride and happiness levels seem to be unhindered and remain reasonably high. Yet, we also notice support for overreaching government policies with regards to surveillance and monitoring does not correlate with the high levels of national pride. Therefore, it is reasonable to conclude that although most Chinese citizens are very proud of their nationality, that does not necessarily correspond with support for the actions of the government.

5.2 Ethical Considerations

In conducting research that delves into personal beliefs and societal norms, such as happiness, national pride, and government surveillance acceptance, it is essential to consider ethical ramifications. The sensitive nature of questions related to government surveillance and monitoring requires careful deliberation. Collecting data on these topics might cause discomfort or fear among respondents, particularly in regions where government criticism is not tolerated openly. This necessitates stringent measures to ensure anonymity and confidentiality of the survey responses to protect participants from any possible repercussions.

Moreover, given the potential for such research to influence public opinion and policy, it is crucial to handle data interpretation and dissemination with integrity. Researchers must be vigilant against biases in data analysis and reporting, striving for objectivity and transparency in their methodologies and findings. Acknowledging the limitations and subjective interpretations inherent in the study is also essential to prevent misrepresentation of the data and to maintain the trustworthiness of the research.

Additionally, the study's potential impact on participants and broader societal implications should be considered. Research findings, particularly those that may portray certain groups in a negative light or suggest discontent with governmental policies, could have consequences if used inappropriately or taken out of context. As such, it is the responsibility of the researchers to ensure that the conclusions drawn are balanced, contextual, and do not contribute to stigmatization or discrimination against any group.

5.3 Weaknesses and next steps

With an R-squared value of 0.048, it seems that an incredibly small amount of variance is explained by our model. This is most likely due to omitted variable bias, where we neglected to include several other important factors when developing our regression model. Therefore, it is very likely that our chosen variables on support for government policies and/or national pride are not that significant in predicting overall happiness level. However, when considering the fact that the original questionnaire contained approximately 300 questions, it is to be expected

that there would be a lot of unaccounted for variance in our data. Perhaps stronger expected indicators of overall happiness would consist marital status, annual income, and employment status. With regards to our model, we would likely get a more accurate regression by including more questions. However, due to the incredibly large amount of questions and responses, this could lead to overfitting. Although the adjusted R-squared value accounts partly for too many regressors, there is no feasible way to run a Bayesian regression model on all 300 questions. Therefore, future iterations of this study could look at combining several values into one. For example, support for government surveillance could be combined into one variable instead of three separate ones. Furthermore, questions that seem to have no impact on happiness level could also be omitted by user discretion.

Another source of error is the subjective nature of these variables, as they are deeply influenced by individual experiences, cultural contexts, and personal biases. These factors can affect both how questions are perceived by respondents and how their responses are interpreted by researchers. For instance, the concept of “national pride” may carry different connotations and importance in different cultural or political contexts, potentially skewing responses based on the respondent’s background or current political climate. “Happiness” and current “life satisfaction” are also subjective measures and the same point value provided by two different respondents is likely to have different meanings for each of them.

Future research may consist of comparisons between different nations that partook in the World Values Survey. Specifically, comparisons for national pride and government support between communist and democratic countries may lead to interesting inferences on the societal beliefs for their respective nations. A heavier reliance on governmental structures in a communist society may lead to higher levels of national pride when compared to less reliant democratic nations.

Following Steele and Lynch’s original research method, variations between waves of the World Values Survey may be dealt with by combining several responses into one variable. Although this leads to higher variance and may change end results, the value gained from analyzing temporal trends may outweigh these risks. With our original research goal in mind, the change in perception of China may be analyzed throughout the past two decades to determine if the rapid onset of anti-CCP material impacts citizen responses in any way.

A Appendix

Support for Surveillance, Email Monitoring, Information Collection, G

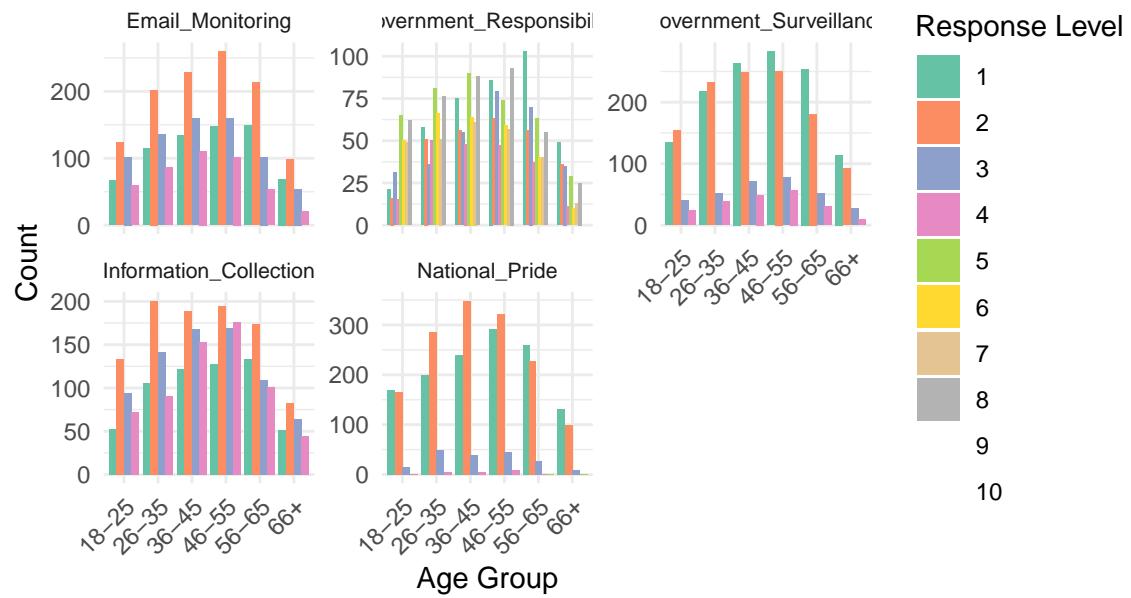


Figure 5: Support for Surveillance, Email Monitoring, and Information Collection by Age Group

References

- Arel-Bundock, Vincent. 2022. “modelsummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Bjørnskov, Christian. 2006. “The Multiple Facets of Social Capital.” *European Journal of Political Economy* 22 (1): 22–40. <https://doi.org/10.1016/j.ejpoleco.2005.05.006>.
- Cheung, Chau-kiu, and Kwan-kwok Leung. 2007. “Enhancing Life Satisfaction by Government Accountability in China.” *Social Indicators Research* 82 (3): 411–32. <https://doi.org/10.1007/s11205-006-9043-9>.
- Diener, Ed, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith. 1999. “Subjective Well-Being: Three Decades of Progress.” *Psychological Bulletin* 125 (2): 276–302. <https://doi.org/10.1037/0033-2909.125.2.276>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Helliwell, John F., and Robert D. Putnam. 2004. “The Social Context of Well-Being.” Edited by F. A. Huppert, N. Baylis, and B. Keverne. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359 (1449): 1435–46. <https://doi.org/10.1098/rstb.2004.1522>.
- Hofstede, Geert. 1980. “Culture and Organizations.” *International Studies of Management & Organization* 10 (4): 15–41. <https://doi.org/10.1080/00208825.1980.11656300>.
- Moore, Robert L. 2005. “Generation Ku: Individualism and China’s Millennial Youth.” *Ethnology* 44 (4): 357. <https://doi.org/10.2307/3774095>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2023. *Arrow: Integration to ‘Apache’ ‘Arrow’*. <https://CRAN.R-project.org/package=arrow>.
- Steele, Liza G., and Scott M. Lynch. 2013. “The Pursuit of Happiness in China: Individualism, Collectivism, and Subjective Well-Being During China’s Economic and Social Transformation.” *Social Indicators Research* 114 (2): 441–51. <https://doi.org/10.1007/s11205-012-0154-1>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019a. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- , et al. 2019b. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.

//CRAN.R-project.org/package=kableExtra.