

Insights From the Asian Barometer Survey

A Latent Factor Approach

Anna Dai, Deekshita Saikia, Moritz Wilksch, Satvik Kishore

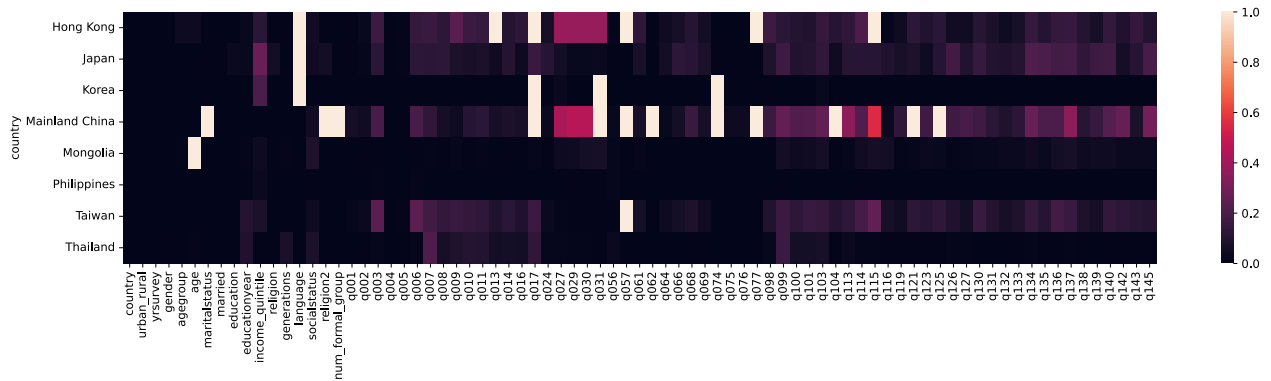
Introduction

Asian Barometer Survey (ABS) is a series of random surveys conducted across 18 Asian countries conducted by the National Taiwan University East Asia Democratic Studies department to gather data on political and socio-economic opinions of individuals from each country. These surveys have been through several rounds (or waves) over the past two decades and the data provided for the purpose of this analysis spans over two waves - the first wave from 2001 to 2003, and the second wave from 2005 to 2008. The ABS is aimed to gather public opinions, foster conversations between scholars, and share any insights with various organizations (Fu Hu Center, 2021). We aim to conduct cross-wave analyses and augment the ABS data by generating latent factors that can be used as meaningful research objectives. In order to do so, we statistically derive consolidated indicators to represent abstract individual characteristics like “faith in government” or “economic optimism”. We intend to uncover insights to suggest further deep dives that scholars can leverage to answer specific questions, such as whether demographic variables are significant predictors of these factors.

Data

The questionnaires consist of a wide variety of questions intended to cover a broad spectrum of demographic variables (i.e. age, race, and occupation) and topics (i.e. political views and involvement, socio-economic backgrounds, and trust in institutions). Each row of data corresponds to a set of responses from a single individual. We elect to use the merged data sets for all countries from both the first (W1) and second wave of surveys (W2). The first wave surveys involve eight countries, namely Japan, Hong Kong, Korea, China, Mongolia, Philippines, Taiwan and Thailand, whereas the second wave surveys include responses from five additional countries, which we do not consider for our analysis to preserve comparability across both waves.

For data pre-processing, since the waves span multiple years, we find drastic discrepancies between the questions in W1 and W2. We word-tokenize the actual questions and compare the cosine similarities between the tokens to map corresponding questions between the waves. We then pick the common questions of interest for further analysis. We also observe several variables such as `religion` and `married` where we must infer levels from the other data set. Similarly, we notice that all the nonresponses (i.e. "Don't know", "Can't choose") have been replaced with N/As in the original W1 data set, and thus we also remove such values in W2. To tackle the significant number of missing values, we impute data using the K-Nearest Neighbors algorithm. This helps provide complete observations, on which we could run our models. Additionally, upon exploring the data, we find countries like China (see heatmap below) to have zero responses for certain questions that are not applicable as well as predictors like language. For these reasons, we drop such variables from our dataset.



Due to different political systems and social environments, we expect findings to differ across countries and by time-period. Our exploratory data analysis confirms this hypothesis, as responses to questions like economic outlook or vary noticeably by country and time (see Appendix (A)). The set of countries provided in survey have varied political systems. In 2006, the country-wise democracy rankings ranged from 11 (Taiwan) to 151 (China). (Kekic, 2007). Hence, we intend to model the data for each country separately to first focus on findings within each country and subsequently compare findings across countries.

Methodology

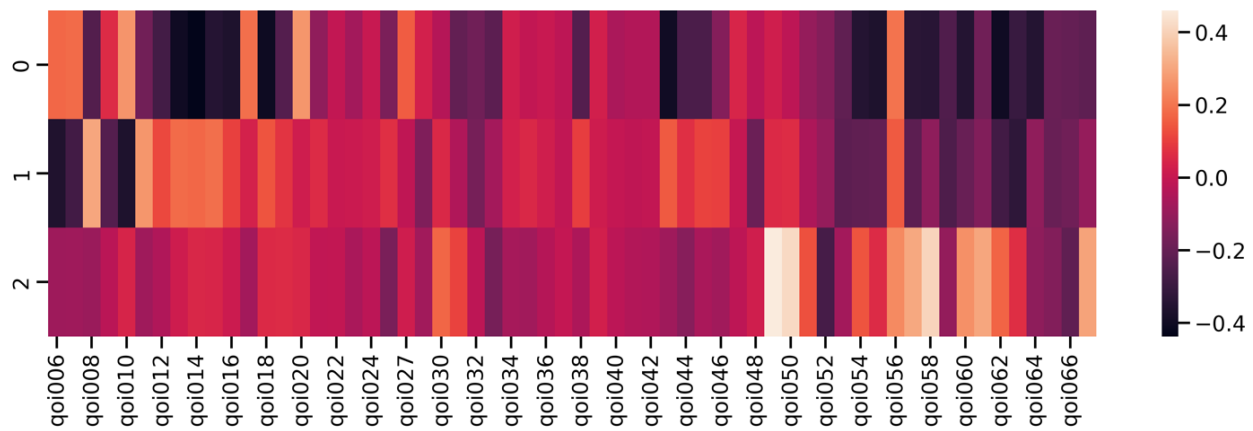
After the data has been cleaned, we utilize Factor Analysis (FA) to uncover latent constructs that manifest in the observed survey responses. The goal of this analysis is to not only find these constructs, but also estimate which survey items load onto which latent variables. This enables us to make sense of the bigger picture hidden in the data that has been quantified with several hundreds of measurable survey questions. The number of latent variables is determined by fitting a Principal Component Analysis and visually examining the scree plot. We find the optimal number of components to be around three for all countries in the study. Furthermore, the FA requires that the survey items are correlated, which we confirm after calculating the correlation matrix. Another model that was employed in previous research in this field is Structural Equation Modeling (SEM). SEM have been used to model political participation data (Wong et al., 2011) where the authors made use of the latent structures to describe relationships between observed variables. Our analysis is an intermediate step that can be expanded upon to a complete SEM analysis.

In order to help the Asian Barometer with its goal of strengthening the intellectual and institutional capacity for research on democracy as well as generating reliable and comparable data, we want to find out how demographic variables are associated with the latent factors uncovered through the FA. If it was possible to predict which population subgroups are highly likely to hold extreme opinions or feel treated unequally, the Hu Fu Center could conduct more in-depth research about which circumstances foster these attitudes. To achieve this goal, we fit a linear regression model to regress the latent variables on demographic attributes, which are available without conducting costly surveys.

In addition, this project also aims to study how the political and socio-economic outlooks change over time. We therefore conduct the analysis for both waves one and two, leveraging the matched variables and answers that were obtained through extensive data cleaning.

Results

For most of the countries, the factor analysis is able to estimate interpretable factor loadings for at least one of the three latent dimensions. For example, in Mainland China, the FA uncovered a latent factor that could be called “prospering liberals”: They agree that the economic situation of their family and country has improved over the past five years and will improve in the future, while rejecting the idea of a single leader or the government limiting free thoughts and opinions and encourage the idea of the legislature checking the government’s action. The loadings of each item onto all three factors are visualized in the figure below:



We can see that every factor does have distinctly positive (bright) or negative (dark) loadings for multiple items based on which the description of the subgroup above was created. The subsequent regression model shows that there are indeed demographic characteristics which are predictive of a high “prospering liberals” score. The model is able to explain around 31% of the variance in the factor score.

	coef	std err	t	P> t	[0.025	0.975]	
Intercept	-0.905	0.086	-10.55	0.000	-1.073	-0.737	***
urban_rural	0.076	0.03	2.523	0.012	0.017	0.135	*
gender	0.163	0.026	6.164	0.000	0.111	0.215	***
agegroup	-0.013	0.006	-2.129	0.033	-0.024	-0.001	*
married	-0.196	0.05	-3.943	0.000	-0.294	-0.099	***
education	0.161	0.008	21.445	0.000	0.146	0.176	***
income_quintile	0.026	0.012	2.184	0.029	0.003	0.05	*
religion	0.010	0.011	0.935	0.350	-0.011	0.031	NS
socialstatus	0.218	0.013	16.523	0.000	0.192	0.243	***

The interpretation of the significant coefficient indicates that – everything else held constant – a higher education, urban living situation, and higher social status are associated with a high “prospering liberal” score while being married or in a younger age group is negatively associated with it.

Another interesting latent variable is one that we label “support of the totalitarian system” in Korea. All items that load highly on this factor are related to abolishing parliaments, wanting the military to take over the country, and supporting a single autocratic leader. It stands out that most items that load negatively on this factor are related to the future economic outlook of the country and one’s own household. The score of support of the totalitarian system can not be predicted from demographic characteristics as the corresponding linear regression model has an R^2 value of around zero. Thus, we have to acknowledge that it is not possible to predict these latent factor scores for all countries and all factors just based on the demographic predictors. In fact, for the Philippines, Korea and Japan, in wave one, no linear model is able to explain more than 8% of the variance in any latent factor. For wave two, the same holds true only for the Philippines and Thailand.

Comparing the results of the study across the two waves provides further evidence for this hypothesis. Opposite to the first wave, the second wave of survey data in Mainland China does not contain any latent factors that are predicted as well by demographics as before, while one of the latent factors for Korea can now be fit adequately with a linear model. The factors themselves have also changed. For the second wave, all three factors show signs of conservative beliefs and government distrust: For example, items that load positively on all factors are “The country should defend its own way of life and not change” or “too many different ways of thinking would make society chaotic” while trust in politics and media as well as interest in politics load negatively on all three factors.

Conclusion

In this study, we applied Factor Analysis in conjunction with a linear regression model on an extensively cleaned data set that spans eight Asian countries over two survey waves to extract latent constructs from survey responses and study how well they can be predicted from demographic data. The goal of this analysis is to provide insights and help Asian Barometer with their mission to generate scientifically reliable and comparable data by using demographic data to identify subpopulations with certain beliefs and economic circumstances that could be targeted in more specific studies. We find that for some countries like China, Mongolia, and Taiwan in the first wave and Hong Kong, Mongolia, and Taiwan in the second wave this identification is indeed possible. Some beliefs among these subgroups might warrant future research, for example on how the current economic situation of an individual affects their distrust in the government and support for autocratic regimes.

Besides providing promising results for future research, this study is limited in several ways. The joint analysis over two waves was hindered by inconsistent variable definitions which led to the exclusion of some items. Furthermore, the handling of missing data – while necessary to fit a model – might induce bias, especially considering that for some questions/country combinations only very few data points were available.

References

College of Social Sciences, National Taiwan University Fu Hu Center for East Asia Democratic Studies. (n.d.). *Introduction*. Asian Barometer. Retrieved November 7, 2021, from <http://asianbarometer.org/intro/introduction>.

Kekic, L. (2007). *Democracy Index: The World in 2007*. Retrieved November 7, 2021, from https://www.economist.com/media/pdf/DEMOCRACY_INDEX_2007_v3.pdf.

Wong, J. S., Ramakrishnan, S. K., Lee, T., Junn, J., & Wong, J. (2011). *Asian American political participation: Emerging constituents and their political identities*. Russell Sage Foundation.

Appendix

(A) Survey Responses by Social Status

