“Nativism” Hierarchical Factor Analysis Project

Introduction

Trump’s election was unexpected, especially among the academia and analyst and this surprise may suggest a new set a political paradigm has come. Many people try to pinpoint the driving forces for this shift of the paradigm and specifically, the rise of Trump. Such hypothesis include his success through economic lenses, his anti-elite sentiment, and his authoritarianism which feeds the need of his supporters. Ipsos is a global leading consulting company and they give a unique answer: Nativism, strongly echoed by Trump’s “America first” narrative, is essentially a battery capturing anti-immigrant attitudes (Young, 2016). This construct is composed of 5 item questions. An example is like the following one “Do you agree or disagree that when jobs are scarce, employers should prioritize hiring people of this country over immigrants”. Ipsos believes that Trump’s nativist narrative appeals the republican base and can potentially spread out further to a global stage. Their analysis significantly support their claims. They have found that nativism is the strongest driver for Trump’s support among the republicans, for the identification as a republican, and primary vote for Trump among all Americans. If their hypothesis is right, then the world is witnessing a transform in the political paradigm resembling Trump’s rise featuring the anti-immigrant sentiments. This hypothesis is more likely to be true given today’s global background of a refugee crisis and a series of political and economic consequences such as terrorism, refugee camp and more directly, Brexit.

This paper is inspired by their hypothesis and specifically investigates the presence of “Nativism” in a global context. The dataset is a global survey tapping into individual’s multiple political attitudes. There are in total over 16000 participants from 22 countries. This is a multilevel dataset which can at least be analyzed from two levels: the individual level, and the country level. Therefore, this paper follows the pattern of the different levels and features three parts. In part one, this paper will elaborate the method and results for analyzing the country level; secondly, this paper will present the method and results for the individual level in the US context as an example. Finally, this paper will present the hierarchical model by using Bayesian analysis to illustrate the probability distribution of both the country but also the individual level parameter estimates. This paper does not offer a full exploration of the dataset but focuses only on the construct of “Nativism” variable. This paper aims to provide a concrete overview this variable for the benefits of further exploration of its relationship with other variables in the context of the political change.

Country level

Method

The data was acquired from Dr. Kirby for the purpose of exploration and data analytical practice. The dataset was collected by Ipsos and have been used to test the driving force of nativism in the success of Trump’s election. According to Ipsos, these data are collected from 5 polls from December 28th, 2015 through April 16th 2016. There are in total 16096 complete entries from 22 countries, with samples at least above 500 for each country. The data was clean and in a tidy format: each row is an observation and each column is a variable in a rectangular data structure. There are over thousands of variables in the original dataset but for the purpose of this paper, only the variables related to the nativism construct were used. The nativism is a five item construct, featuring questions as below:

- MW_Q9_1. [Immigrants take jobs away from real ...] Do you agree or disagree with the following statements?

  MW_Q9_2. [Immigrants take important social services away from …]
- MW_Q9_3. [When jobs are scarce, employers should prioritize hiring people of this country over immigrants]

- MW_Q9_4. [... would be better off if we let in all immigrants who wanted to come here]
- MW_Q9_5. [... would be stronger if we stopped immigration]

These questions were constructed on a 5-item Likert scale ranging from “strongly disagree” to “strongly agree” with “neither agree nor disagree” in the middle, and “don’t know” as an additional option. This research recoded them in a numerical 1 to 5 scale with “strongly disagree” noted as 1 and “strongly agree” noted as 5. For the “don't know” answers, this research coded them in the middle as 3. The descriptive statistic is shown in the Table 1.

Table 1

Correlations Among and Descriptive Statistics For Key Study Variables

	M (*SD*)	Q9_1	Q9_2	Q9_3	Q9_4	Q9_5
Q9_1	2.900 (1.33)		.082	.73	.16	.54
Q9_2	3.058 (1.36)			.58	-.22	.78
Q9_3	3.519 (1.32)				-.02	.49
Q9_4	3.778 (1.18)					-.28
Q9_5.	2.890 (1.34)					

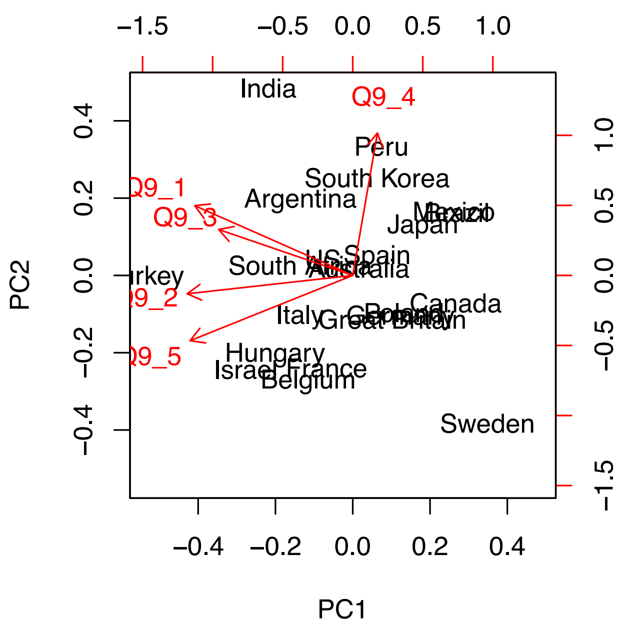
Notes. “strongly disagree” noted as 1 and “strongly agree” noted as 5

Results

To further understand the relationship between variables and the country’s distribution around them. This research conducted a principle component analysis. Principle component analysis is a dimension reduction statistical method that transform a set of correlated variables into some linearly uncorrelated variables called the principle components (Rencher and Christensen, 1995). The principle components analysis are the eigendecomposition of the covariance matrix of the variables. The principle components are the eigenvectors with the amount of variances explained by the corresponding eigenvalues. The eigenvalues are ranked by the variances they explained. This result in the rankings of the corresponding eigenvectors. To get a simple and direct glance of the data, researchers conduct principle components analysis and mainly investigate the first couple of principle components (corresponded by the eigenvectors). For the visualization purpose of PCAs, biplot is often used. Biplot puts the vectors and the data entries together in the axis of the first two principle components (together they explain the most of the variances). The arrows of the variables denote its loadings from the first two principle components. The scatterplots of the data entries (in this case, the country’s names) are pointed according to its transformed values by the linear combination of the PCs.

Figure 1

Biplot of the principle component analysis



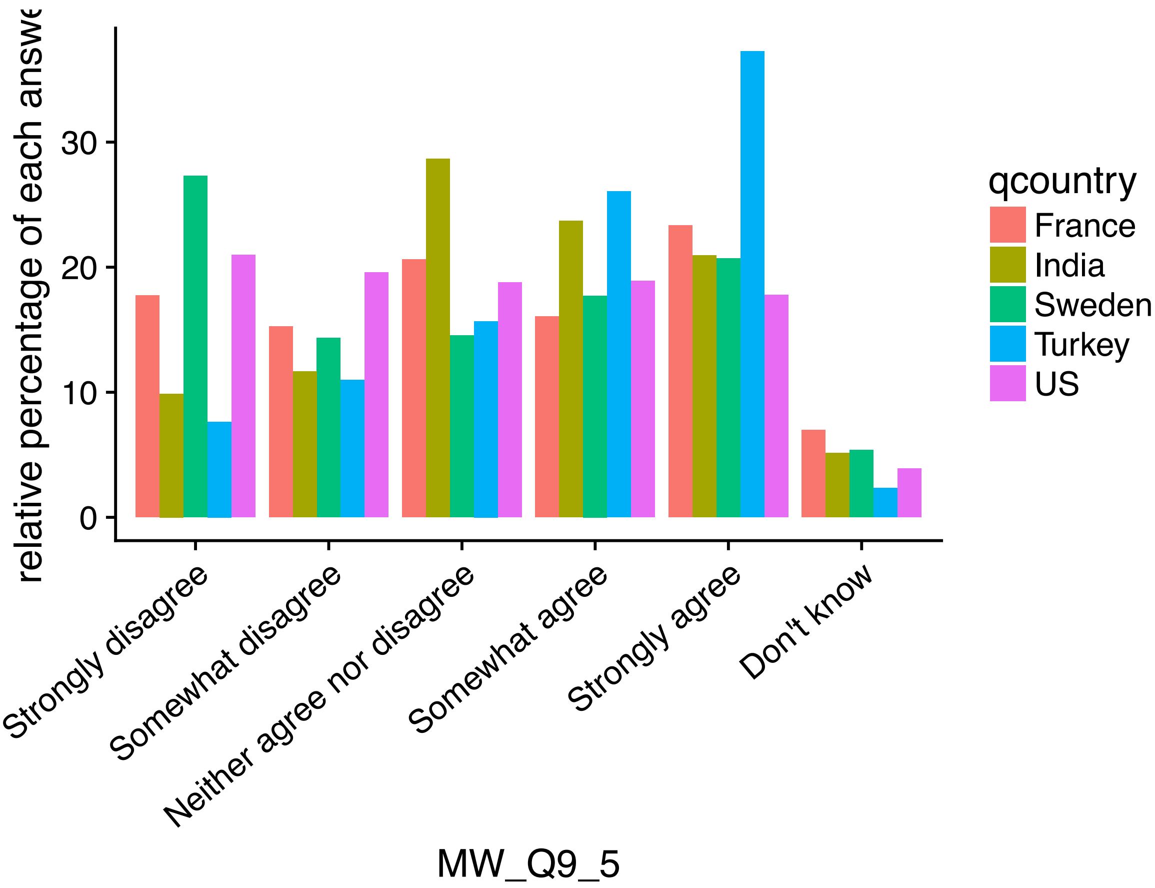
As the figure 1 suggests, variables Q9_1 and Q9_3 are relatively close with each other in terms of their corresponding loadings on the first two principle components. Variables Q9_2 and Q9_5 are relatively close with each other. Q9_5 deviates with other four variables in terms of both loadings on the first and the second principle components. Indeed, Q9_4 is the only inversely worded question item in the 5-items. There are also some patterns revealed about the countries and their corresponding relationships with each other and the variables. For example, Sweden is the farthest from the country groups suggest that it is a outlier in certain aspects; and the same goes to India and Turkey. Peru and South Korea are close to the Q9_4 arrow suggesting they are relatively high on Q9_4.

The first two components together explain almost 80% of the variances in the covariance matrix, suggesting that our decision of looking only the first two principle components a valid approach. A Scree plot to visualize the varaiance explained by the number of principle components is added in the Appendix. When analyze the Scree plot, the trick is to find the “elbow” of the lines which suggest the number of principle components should be considered.

Finally, a visualization of different answers given by the different countries in offered in the figure 2. Note that on the Y axis, the relative percentage of different answers by a certain country is shown instead of the frequencies (or counts). This is because different countries have different sample sizes so the relative percentage in this context makes more sense. For the illustration purpose, only the results of Q9_5 is shown here.

Figure 2

Relative percentage of answers by 5 diverse countries



Individual level

Method

This paper starts to document the method and the results for the individual level part. Due to multiple countries are involved in this dataset and it is redundant to list all the results of each of them. This paper only list the results for the US as an example. As the 2017 US election was the force that drives this study therefore it is of our special interests.

In the case of the US, there are in total 1000 participants for the survey. As the same with the general dataset, all of the data are clean without NA values. For this part, the Q9_4 (Inversely worded) has been adjusted to the same direction with other four questions.

Results

To test the reliability of the construct nativism, a Cronbach’s alpha has been run for the US context. Cronbach's alpha is a measure of internal consistency, that is, how closely related a set of items are as a group. Conventionally, a Cronbach’s alpha close to 0.8 suggests a good reliability of the construct. For the nativism in the US context, the Cronbach’s alpha is 0.83.

Secondly, a principle component analysis is run to glance the structure of the data. The results of loadings on the 5 items are shown in table 2

Table 2

Loadings of the principle component analysis

	PC1	*PC2*	*PC3*	*PC4*	*PC5*
Q9_1	-0.52	0.11	-0.11	-0.27	0.79
Q9_2	-0.53	0.09	-0.04	-0.62	-0.58
Q9_3	-0.41	-0.03	0.85	0.33	-0.04
Q9_4	-0.19	-0.97	-0.13	0.02	0.01
Q9_5	-0.49	0.18	-0.50	0.66	-0.19

The principle components are the simple eigenvectors of the covariance matrix without any rotation. Rotations are done for the sake of interpretation of the extracted factors in factor analysis. It does not alter the relative positions of variables in the space of factors geometrically (Thompson, 2004). What is changed is the projected coordinates on the axes. There are multiple ways of rotation depending on the underlying assumptions of the factors. Two most commonly used roration method are called “Varimax” and “Oblimin”. “Varimax” is so called because it maximizes the sum of the variances of the squared loadings and it preserves the orthogonality of the PCA sub-space (Costello & Osborne, 2005). “Oblimin” rotation is often used under the assumption that the factors are correlated with each other. This research uses both rotation methods to extract the factors and they produced almost the same results. For the illustration purpoe, this study attached the factor loadings from the “oblimin” rotation method. See table 3.

Table 3
*Factor loadings based on a principle components analysis with “oblimin” rotation for 5 items (*N *= 1000)*

	Factor 1	Factor 2
Immigrants take jobs away from real ...	.89	
Immigrants take important social services away from …	.84	
When jobs are scarce, employers should prioritize hiring people of this country over immigrants	.66	
... would be better off if we let in all immigrants who wanted to come here		.50
... would be stronger if we stopped immigration	.84	

Note. Factor loadings < .2 are suppressed

There are two factors according to the result of the loadings. Question Q9_1, Q9_2, Q9_3 and Q9_5 are in the same factor based on their high loadings on the factor 1. Q9_4 is in the second factor alone. The differences of between factor 1 and factor 2 may be caused by the wordings of the questions. While the Q9_4 has been adjusted back in the same direction with the other four questions, it has a slightly variation in terms of the attitude of acceptance of immigrants. A participant may not agree with the other four questions featuring obvious anti-immigrant sentiment, but s/he may also not agree to accept ALL immigrants in the country without any conditions or limitations. This extra concern of the absolute acceptance may be the reason why it has a different loadings than the other four questions.

Hierarchical level

Method

The current dataset has a multi-level or hierarchical structure with at least two layers: the country level and the individuals level. Dr. Wiecki argues that the hierarchical models are intrincincally Bayesian (2016). Hierarchical models, especially when it involves the comparison between groups, suffers from the frequentist paradigm. In frequentist comparison, p-values are often used in parameter estimation and model selection. This analysis procedure results in either treating the two populations as completely distinct or treating them as exactly identical given the p-value. This procedure leaves great room for argument and subjectivity. For example, what if the p-value is 0.06 (given the alpha is set 0.05), under the frequentist paradigm the two samples in this case would be treated as exactly the same (Hoff, 2009).

On the other hand, Bayesian analysis paradigm carries some great advantages against such a problem. Bayesian statistics is founded upon the Bayes’ rule. Bayes’ rule is a special case of derivation conditional probability. In Bayesian statistics, the most important estimate is the posterior distribution of the parameter of interests. And these posterior distributions can be obtained from the product of a prior distribution and a likelihood distribution. Unlike the frequentist method, the estimate is not just a point estimate but a probabilistic distrubtion. And this makes Bayesian methods always had a natural intrinsic advantage because all unknown quantities are treated probabilistically, and this is the way that people really prefer to think intuitively (Hoff, 2009). Generally speaking, Bayesian statistics possesses the following attributes:

“it often results in shorter confidence/credible intervals, it often gives smaller model variance,

predictions are usually better,

“proper” prior distributions give models with good frequentist properties,

reasonably “objective” assumptions are available,

hypotheses can be tested without pre-determination of testing quality measures.” (Jacson, 2009, p. 3).

However, there are also certain drawbacks for using the Bayesian analysis. The first is that the Bayesian statistics is math-heavy and calculation intensive. It used to be a burden for the Bayesian statisticians due to the intractable posteriors. But now with the cheap computation resources those posteriors can be easily simulated by using statistical programming software. The second and also the most commonly seen critique of Bayesian statistics is that it is subjective in terms of the selection of the prior distribution. This drawback can be greatly reduced by using a non-informative prior. A non-informative prior usually support the parameter space well but it is evenly distributed among the space without any implication of the information favoring any point in that estimation. For example, a normal distribution prior can be taken with the mean centered at the mean of the data but with variances very big. This results in almost a uniform distribution so that the prior does not convey any information in terms of parameter estimation.

This paper utilizes a series of non-informative priors to make sure the objectivity of the results. For example, this paper sets the prior sample size to 1 to reduce the influence of the prior. This paper also uses a big prior variance to greatly support the parameter space while not conveying much information.

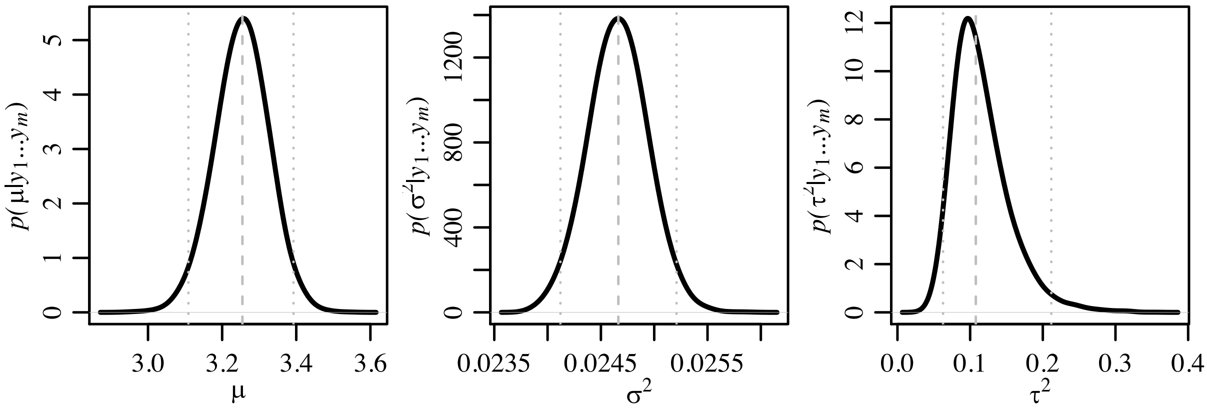
With the priors set then this paper uses Gibb’s sampling to simulate 5000 iterations and approximate the posterior distribution of the both levels.

Results

The posterior distribution about the three parameters are shown in figure 3.

Figure 3

Posterior distribution of three key parameters of the hierarchical model

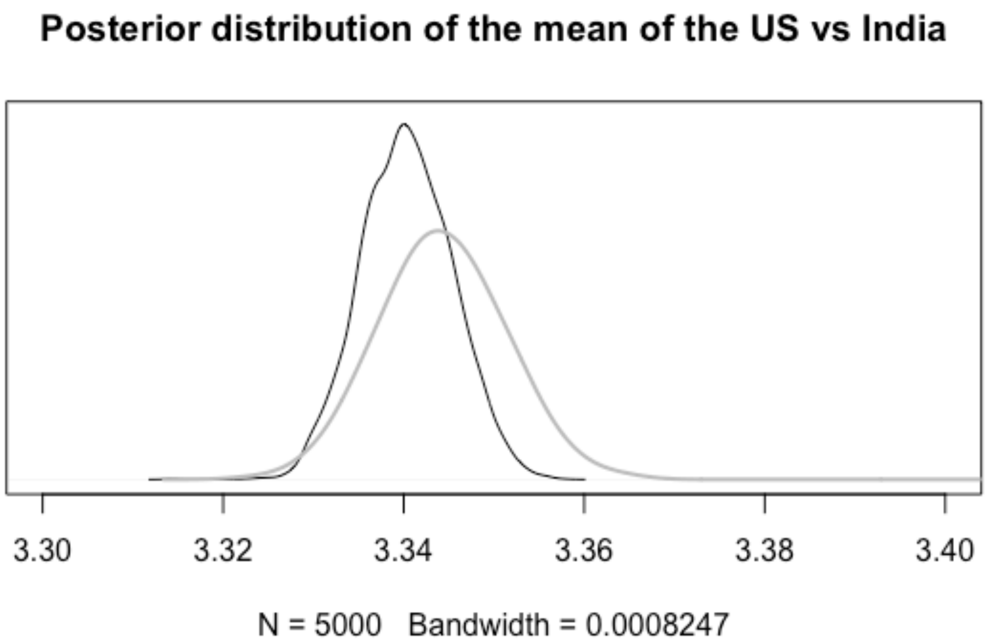


The most left one show the posterior distribution of the group mean among 22 countries. The middle chart show the posterior distribution of the individual variances within each country (this paper assumes the equal within-group variance for the sake of model simplicity; however, the different within-group variance model can be achieved by a little bit of further programming later). And the most right chart shows the posterior distribution about the between-group variances. With those posterior distributions of the parameters, researchers can make a probabilistic statement regarding their beliefs about the parameters. For example, it is valid to say that the probability for the mean of a country between 3.1 and 3.5 is approximately 95%.

For the sake of group comparison, the posterior offers a more direct and intuitive explanations. For example, as it is in the figure 4. The posterior for the US is the black line and India the grey line. Instead of simply testing the alternative hypothesis that they are different by using the frequentist t-test. The posterior result can inform us the exact probability that the mean of the US is smaller than India, which is 67%.

This paper ends the analysis at the hierarchical model using the Bayesian simulation method. It is far from complete to fully explore the dynamics of the nativism construct. Further analysis can be done by using the Bayesian regression method to estimate the relationship between nativism and other variables. I will also keep learning the Bayesian application to social science by exploring this dataset.

Figure 3
Posterior distribution of the mean of the US and India



References

Rencher, A. C., & Christensen, W. F. (1995). Methods of Multivariate Analysis, Wiley-Interscience. *New York*.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation*, *10*(7), 1-9.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer Science & Business Media.

Jackman, S. (2009). *Bayesian analysis for the social sciences*(Vol. 846). John Wiley & Sons.