Appendix of
"Retrieving True Preference under Authoritarianism"

Jongyoon Baik* Xiaoxiao Shen†

Survey Research Methods

Contents

*Assistant Professor, The Chinese University of Hong Kong, Shenzhen. jongyoonbaik@cuhk.edu.cn

†Assistant Research Professor, Department of Political Science, Northeastern University. xia.shen@northeastern.edu

1 Latent Profile Analysis

1.1 EM Stage

This part is an extension of "Latent Profile Analysis: How LPA Works" in the main article. It explains how parameters are estimated through the expectation-maximization (EM). According to Dempster et al. (1977) and Mooijaart and Van der Heijden (1992), the EM algorithm maximizes the likelihood function via distinct observed and missing data, which refer to the number of the manifest categorical variables (J) and latent classes (K), respectively. In order to simplify the notation, we record all parameters related to the model as $\theta = (\mu_j^{(k)}, \sigma_j^{2(k)})$. Furthermore, the likelihood is the probability density of the data:

$$\prod_{i=1}^{n} f(y_i; \theta). \tag{1}$$

After applying the logarithm, the likelihood function becomes

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} log f(y_i; \theta) \\ &= \sum_{i=1}^{n} log \sum_{k=1}^{K} p(c_i = k) f(y_i; \theta_k). \end{aligned} \tag{2}$$

Next, equation 3 takes the derivative of equation 2 with respect to one of the parameters θ_j.

$$\begin{aligned} \frac{\partial l}{\partial \theta_j} &= \sum_{i=1}^{n} \frac{1}{\sum_{k=1}^{K} p(c_i = k) f(y_i; \theta_k)} p(c_i = j) \frac{\partial f(y_i; \theta_j)}{\partial \theta_j} \\ &= \sum_{i=1}^{n} \frac{p(c_i = j) f(y_i; \theta_j)}{\sum_{k=1}^{K} p(c_i = k) f(y_i; \theta_k)} \frac{1}{f(y_i; \theta_j)} \frac{\partial f(y_i; \theta_j)}{\partial \theta_j} \\ &= \sum_{i=1}^{n} \frac{p(c_i = j) f(y_i; \theta_j)}{\sum_{k=1}^{K} p(c_i = k) f(y_i; \theta_k)} \frac{\partial log f(y_i; \theta_j)}{\partial \theta_j}. \end{aligned} \tag{3}$$

Then the weight of the likelihood is:

$$\omega_{ij} = \frac{p(c_i = j) f(y_i; \theta_j)}{\sum_{k=1}^{K} p(c_i = k) f(y_i; \theta_k)}, \tag{4}$$

where the $p(c = j)$ is the probability that the latent class variable c is j. Then, equation 4 is the joint probability of getting $c = j$ and $Y = y_i$. Therefore, the conditional probability of $c = j$ given $Y = y_i$ is

$$\omega_{ij} = \frac{p(c_i = j) f(y_i; \theta_j)}{\sum_{k=1}^{K} p(c_i = k) f(y_i; \theta_k)} = p(c_i = j | Y = y_i; \theta). \tag{5}$$

After obtaining the posterior probability from equation 5, we can carry out the process of "guessing-iteration-convergence"(Oberski 2016), as the method shown in Figure A.1.

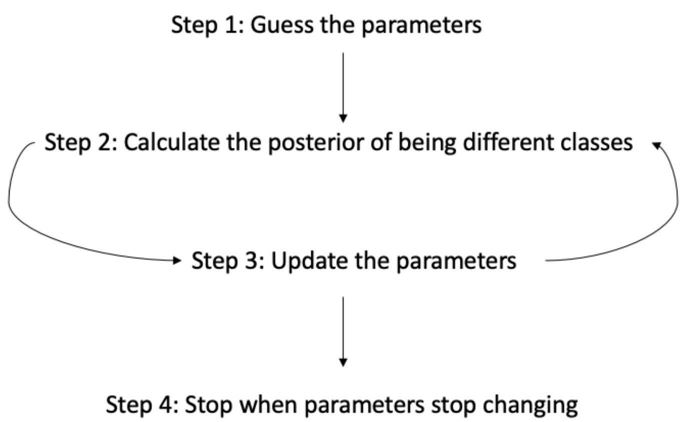


Figure A.1: The process of EM algorithm estimation

Source: Figure 2 from Obserski DL. 2016. "Mixture Models: Latent Profile and Latent Class Analysis." *Modern Statistical Methods for HCI* at page 5.

1.2 Selecting Number of Profiles

In this paper, we primarily use AIC and BIC to evaluate model fit and further consider interpretability when selecting the number of profiles. This section provides justifications for our choice of these criteria.

To begin with, BIC serves as a criterion for model selection among a finite set of models. Generally, models with lower BIC values are preferred. BIC is closely related to AIC. When fitting models, it is possible to increase the maximum likelihood by adding parameters; however, doing so may lead to overfitting. Both BIC and AIC address this issue by combining a measure of model fit with a penalty term that grows with the number of parameters. The key difference is that the penalty term is larger in BIC than in AIC when the sample size exceeds 7 (Vrieze 2012; Stoica and Selen 2004).

BIC has been widely adopted in social science research for model selection. Specifically, when uncovering latent classes, as we did using LPA, scholars often rely on BIC and tend to favor models with smaller numbers of classes (k) when possible. Unlike standard cluster analysis, where researchers often determine the optimal number of clusters arbitrarily, the model selection process in LPA is guided by goodness-of-fit statistics such as BIC.

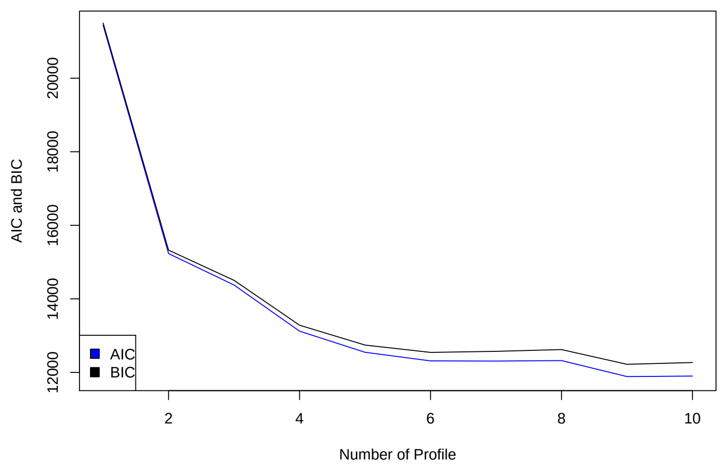
For instance, in Oser et al. (2013), it was observed that as the number of latent classes increased, the absolute value of BIC decreased, while the classification error rose. Here, although the BIC continued to decline up to the seven-cluster model, the improvement in fit beyond the four-cluster model was marginal. While the six- and seven-cluster solutions were clearly less favorable due to small gains in fit and higher classification error, the authors explored the substantive results of both the four-cluster and five-cluster solutions. They ultimately found that the five-cluster solution did not add any meaningful nuance beyond the four-cluster solution, concluding that the four-cluster model was the best choice. Similarly, Feldman and Johnston (2014) used BIC to identify latent classes, finding that the minimum BIC value occurred for a six-class model. Increasing the number of classes beyond six resulted in higher BIC values and increasingly uninterpretable solutions.

In line with these studies, we considered AIC, BIC, and model simplicity to determine the number of profiles, ensuring a balance between model fit and interpretability.

2 Test 1. Data Simulation

The first test of LPA on preference falsification studies used simulated survey data. When applying LPA to the simulated data, we first chose the number of latent profiles, or k, by comparing model fits, such as AIC, BIC, and the entropy level. As in Figure A.2, the model fits are the best when the number of latent groups is either six or nine. While one may choose nine as k, we chose six as k for two reasons. First, we balanced the model fits with simplicity. LPA results become difficult to interpret when the number of latent groups becomes more than six. Second, the AIC and BIC levels turn to an increasing trend after $k = 6$.

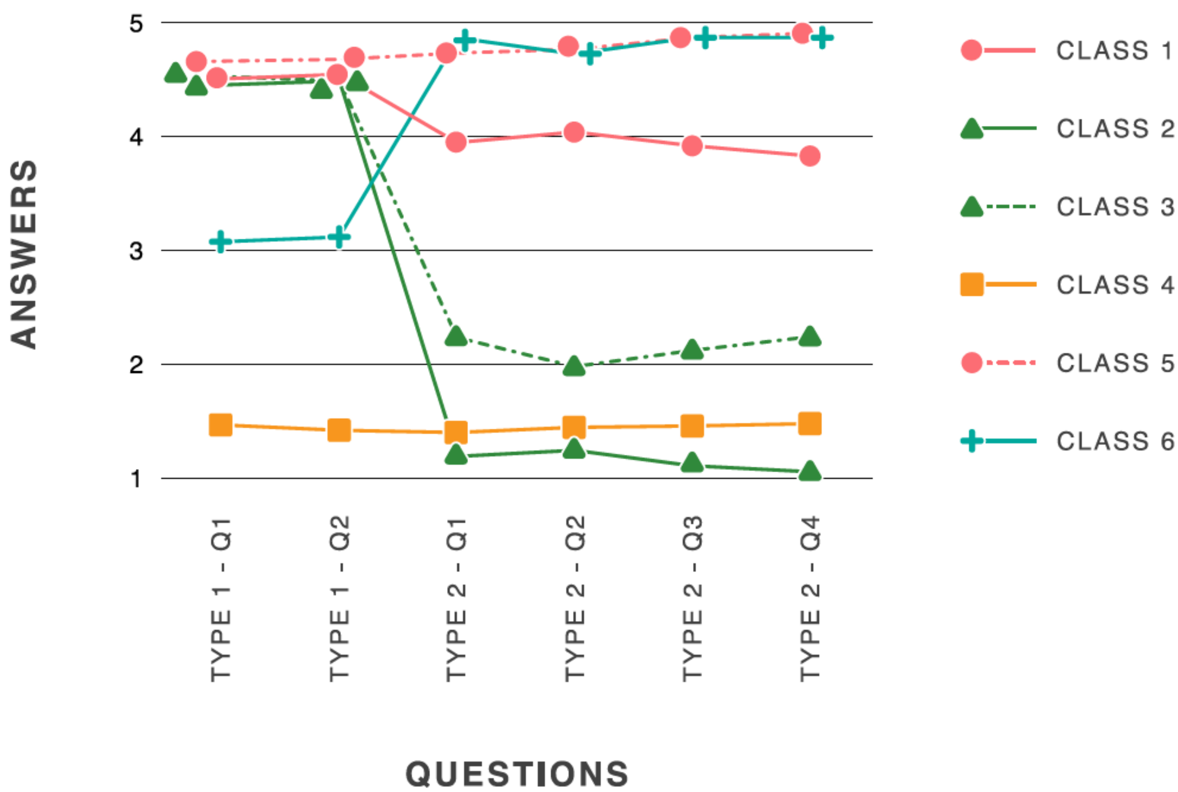
Figure A.2: AIC and BIC on Simulated Data



Note: AIC is 12313.71, BIC is 12544.37, and the entropy level is 0.95 with six subgroups. AIC is 11886.93, BIC is 12220.66, and the entropy level is 0.93 with nine subgroups.

After deciding the number of latent groups, we conducted LPA on the data. The results are shown in Table A.1 and Figure A.3. In Figure 2 of the main article, Classes 1 and 5 in Figure A.3 are merged as Subgroup 1, and Classes 2 and 3 are merged as Subgroup 2, by re-calculating their mean responses. Class 4 is Subgroup 3, and Class 6 is Subgroup 4 in the main article.

Figure A.3: LPA on Simulated Data (1,000 Observations in Total)



Note: This plot shows how six latent subgroups (classes) behave or answer to (y-axis), each survey question (x-axis). Type 1 questions are highly sensitive political questions and Type 2 questions are non-sensitive political questions. Answer 5 is "Strongly agree with the regime," and answer 1 is "Strongly disagree with the regime." 195 respondents are in Class 1, 196 are in Class 2, 112 are in Class 3, 90 are in Class 4, 358 are in Class 5, and 49 are in Class 6. In Figure 2 of the main article, Classes 1 and 5 are merged as Subgroup 1, and Classes 2 and 3 are merged as Subgroup 2, by re-calculating their mean answers. Class 4 is Subgroup 3, and Class 6 is Subgroup 4 in the main article.

Table A.1: Mean Responses by Class and Question, LPA on Simulated Data

Class	Question	Estimate (SE)	Class	Question	Estimate (SE)
1	Type1.Q1	4.5322 (0.0536)	4	Type1.Q1	1.4708 (0.0998)
1	Type1.Q2	4.5657 (0.0537)	4	Type1.Q2	1.4279 (0.0738)
1	Type2.Q1	3.9573 (0.0644)	4	Type2.Q1	1.4008 (0.0807)
1	Type2.Q2	4.0545 (0.0728)	4	Type2.Q2	1.4530 (0.0720)
1	Type2.Q3	3.9356 (0.0715)	4	Type2.Q3	1.4604 (0.0828)
1	Type2.Q4	3.8409 (0.0374)	4	Type2.Q4	1.4809 (0.0819)
2	Type1.Q1	4.4801 (0.0550)	5	Type1.Q1	4.6744 (0.0309)
2	Type1.Q2	4.5242 (0.0531)	5	Type1.Q2	4.6852 (0.0322)
2	Type2.Q1	1.1935 (0.0326)	5	Type2.Q1	4.7516 (0.0229)
2	Type2.Q2	1.2513 (0.0374)	5	Type2.Q2	4.7845 (0.0249)
2	Type2.Q3	1.1157 (0.0267)	5	Type2.Q3	4.8889 (0.0230)
2	Type2.Q4	1.0592 (0.0317)	5	Type2.Q4	4.9261 (0.0318)
3	Type1.Q1	4.5594 (0.0881)	6	Type1.Q1	3.0902 (0.1487)
3	Type1.Q2	4.5033 (0.1030)	6	Type1.Q2	3.1273 (0.1387)
3	Type2.Q1	2.2340 (0.1249)	6	Type2.Q1	4.8787 (0.0644)
3	Type2.Q2	1.9767 (0.1303)	6	Type2.Q2	4.7305 (0.0841)
3	Type2.Q3	2.1241 (0.0736)	6	Type2.Q3	4.8912 (0.0506)
3	Type2.Q4	2.2530 (0.0674)	6	Type2.Q4	4.8789 (0.0829)

Note: Mean responses by classes and questions, from LPA on simulated data. SE means standard errors. Type 1 questions are highly sensitive political questions and Type 2 questions are non-sensitive political questions. Answer 5 is "Strongly agree with the regime," and answer 1 is "Strongly disagree with the regime." 195 respondents are in Class 1, 196 are in Class 2, 112 are in Class 3, 90 are in Class 4, 358 are in Class 5, and 49 are in Class 6.

3 Test 2. Chinese General Social Survey 2006

3.1 Balance Table

In the application of LPA to Chinese survey data, the control group consists of the respondents who participated in the survey before the purge (from September 11 to September 25, 2006), and the treatment group consists of the respondents who participated in the same survey after the purge (from September 26 to November 12). Among the 400 Shanghai respondents, 119 are in the control group and 281 are in the treatment group. Table A.2 provides demographic summary statistics of the two groups.

Table A.2: Demographic Summary Statistics

Control Group

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Female	119	0.580	0.496	0	0	1	1
Age	119	43.555	13.644	18	33	53	70
Ethnicity	119	0.008	0.092	0	0	0	1
College	119	0.134	0.343	0	0	0	1
Religion	119	0.092	0.291	0	0	0	1
Marriage	119	0.790	0.409	0	1	1	1
Internet	119	0.286	0.454	0	0	1	1
Foreign Language	119	0.613	0.598	0	0	1	2
Party	119	0.168	0.376	0	0	0	1

Treatment Group

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Female	281	0.594	0.492	0	0	1	1
Age	281	45.594	13.909	18	34	55	70
Ethnicity	281	0.018	0.132	0	0	0	1
College	281	0.135	0.343	0	0	0	1
Religion	281	0.110	0.314	0	0	0	1
Marriage	281	0.786	0.411	0	1	1	1
Internet	281	0.335	0.473	0	0	1	1
Foreign Language	281	0.541	0.591	0	0	1	2
Party	281	0.110	0.314	0	0	0	1

Note: Ethnicity is "1" for the Han majority, and "0" for non-Han minorities. College is coded as "1" if the education level is higher than, or equal to, college enrollment. Religion is a binary indicator for those with any religion including Catholicism, Christianity, Buddhism, Taoism, Folk religion, and Islam. Marriage is coded as "1" for those who have ever officially married, including divorced or widowed. Internet is coded as "1" for those who use the Internet at least several times a week. Foreign Language is coded as "2" when the respondents are relatively, or very fluent in a foreign language. It is coded as "1" when they can understand everyday vocabulary, speak everyday conversation, or read some simple articles in a foreign language. It is coded as "0" when they do not understand a foreign language at all. Party members, including Communist Party, Democratic Party, and Youth League members, are coded as "1."

In Table A.3, we test whether the two groups are balanced in terms of their standardized mean differences and variance ratios. It shows that demographic distributions are mostly similar between the control and treatment groups. Yet, on average, people in the control group are younger than those in the treatment group. Also, perhaps due to the age gap, more people in the control are fluent in foreign languages than those in the treatment group. Age and foreign language skills are balanced in terms of their variance ratios.

Table A.3: Balance Table, Control and Treatment Groups

	Type	Mean Diff.	M.Threshold	Var. Ratio	V.Threshold
Female	Binary	0.014	Balanced		
Age	Contin.	0.148	Not Balanced	1.039	Balanced
Ethnicity	Binary	0.009	Balanced		
College	Binary	0.000	Balanced		
Party	Binary	-0.055	Balanced		
Religion	Binary	0.018	Balanced		
Public Sector	Binary	-0.015	Balanced		
Marriage	Binary	0.003	Balanced		
Internet	Binary	0.049	Balanced		
Income Low	Binary	0.071	Balanced		
Income Medium	Binary	-0.062	Balanced		
Income High	Binary	-0.009	Balanced		
Income NA	Binary	-0.046	Balanced		
Foreign Langauge	Contin.	-0.122	Not Balanced	0.976	Balanced
Party	Binary	-0.058	Balanced		

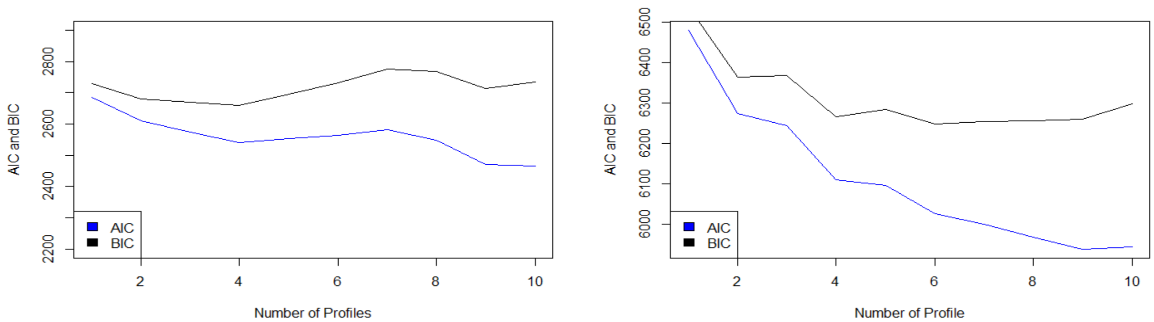
Note. Balance table between the treatment and control groups across different variables. See Table A.2 for the meaning of each variable.

3.2 Main LPA Results

Below Figures A.4a and A.4b show the numbers of latent subgroups in the control and treatment groups and their AIC and BIC statistics. They suggest that it is best to divide the sample into four (AIC 2540.436 and BIC 2659.938 each) or ten profiles (2465.221 and 2734.796 each) for the control group, and six (6027.288 and 6249.227 each) or nine (5939.499 and 6259.674 each) for the treatment group. Again, because the model fits become worse

Figure A.4: AIC and BIC for Deciding the Number of Subgroups

(a) Control Group (b) Treatment Group



Note. AIC and BIC for deciding the number of latent subgroups within the control and treatment groups. There are four subgroups in the control group, with AIC 2540.436, BIC 2659.938, and Entropy 0.96. There are six subgroups in the treatment group, with AIC 6027.288, BIC 6249.227, and Entropy 0.93.

after k=4 in the control group and after k=6 in the treatment group, and also to balance between accuracy and conciseness, we chose to divide the control group into four subgroups and the treatment group into six subgroups.

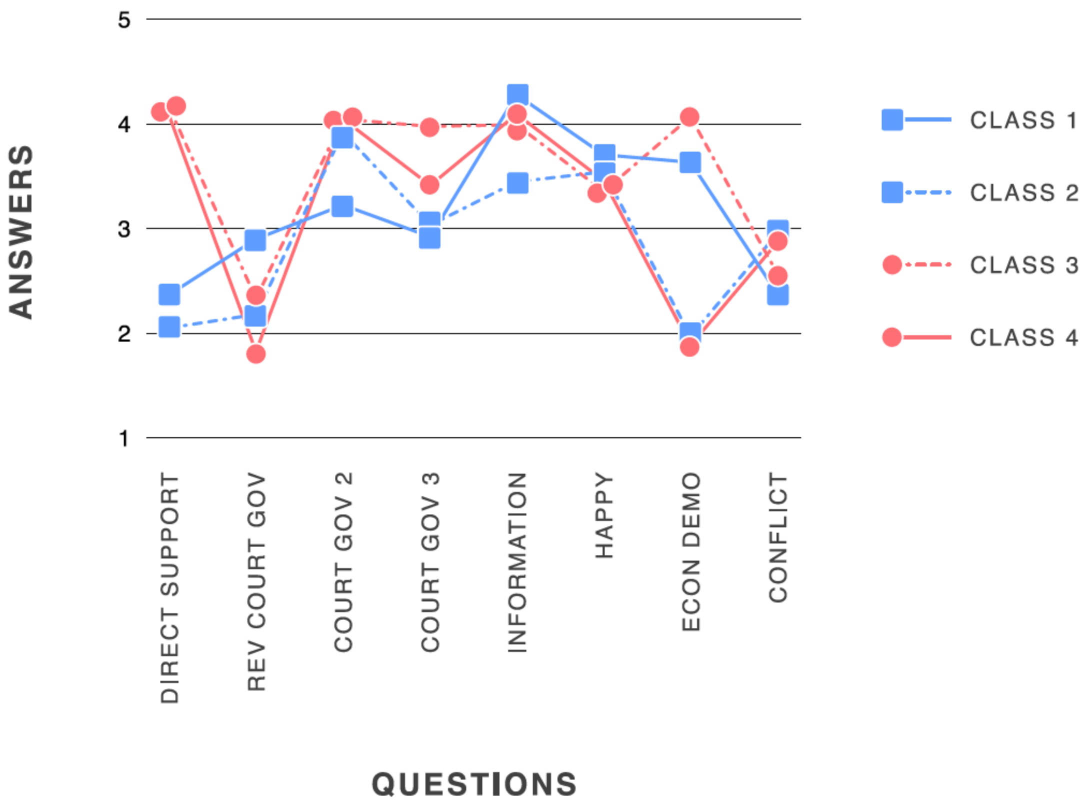
The LPA results on the control group are in Table A.4 and Figure A.5. In Figure 4-(a) of the main article, respondents in Classes 3 and 4 are merged and labeled True Supporters by re-calculating their mean responses, except for the Econ Demo question. Respondents in Classes 1 and 2 are merged and labeled Non-supporters, except for the Econ Demo question. The LPA results on the treatment group are shown in Table A.5 and Figure A.6. In Figure 4-(b) of the main article, respond ents in Classes 1 and 4 are merged and labeled True Supporters by re-calculating their means, except for the Econ Demo question. Respondents in Class 3 are labeled (Candid) Non-supporters. Respondents in Classes 2, 5, and 6 are labeled Falsifiers. Whenever certain subgroups show stronger hints of falsification than others, then their responses are plotted in the main article. That is, for Q3-Court Gov 2, the main plot shows responses from Class 2; for Q4-Court Gov 3, we merged responses from Classes 2 and 5; for Q5-Information and Q8-Conflict, the plot shows responses from Class 6. For other questions, all three classes' responses are merged by re-calculating their means.

Table A.4: Mean Responses by Class and Question, LPA on Control Group

Class	Question	Estimate (SE)	Class	Question	Estimate (SE)
1	Direct Support	2.3776 (0.1185)	3	Direct Support	4.1402 (0.0953)
1	Rev Court Gov	2.9009 (0.2023)	3	Rev Court Gov	2.3405 (0.2052)
1	Court Gov2	3.2336 (0.2229)	3	Court Gov2	4.0624 (0.2076)
1	Court Gov3	2.9388 (0.2105)	3	Court Gov3	3.9757 (0.1957)
1	Information	4.2810 (0.1795)	3	Information	3.9988 (0.1862)
1	Happy	3.7096 (0.1168)	3	Happy	3.3654 (0.1946)
1	Econ Demo	3.6349 (0.1398)	3	Econ Demo	4.0671 (0.0523)
1	Conflict	2.3688 (0.2213)	3	Conflict	2.5639 (0.2403)
2	Direct Support	2.0595 (0.0934)	4	Direct Support	4.0892 (0.0811)
2	Rev Court Gov	2.1873 (0.1954)	4	Rev Court Gov	1.8309 (0.1026)
2	Court Gov2	3.9185 (0.2058)	4	Court Gov2	4.0410 (0.1150)
2	Court Gov3	3.0312 (0.2619)	4	Court Gov3	3.4193 (0.1648)
2	Information	3.4487 (0.2661)	4	Information	4.0944 (0.1309)
2	Happy	3.5462 (0.1608)	4	Happy	3.4617 (0.0751)
2	Econ Demo	1.9765 (0.0613)	4	Econ Demo	1.8905 (0.0519)
2	Conflict	2.9518 (0.2481)	4	Conflict	2.8795 (0.1688)

Note: Mean responses by classes and questions, from LPA on the control group. SE means standard errors. In Figure 4-(a) of the main article, respondents in Classes 3 and 4 are merged and labeled as True Supporters by re-calculating their mean responses, except for the Econ Demo question. Respondents in Classes 1 and 2 are merged and labeled as Non-supporters, except for the Econ Demo question.

Figure A.5: LPA on Control Group



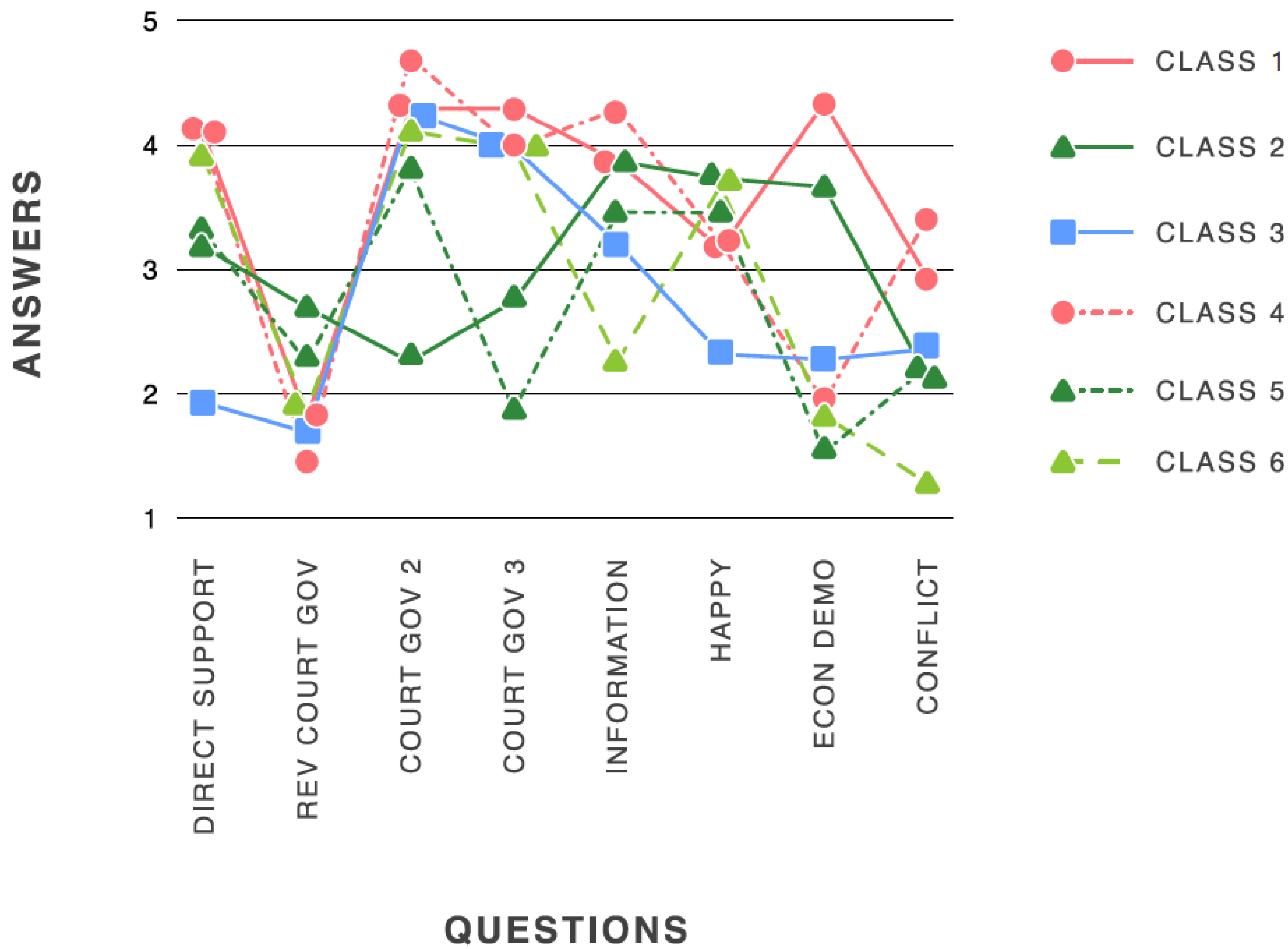
Note: This plot shows how four latent subgroups in the control group behave, or answer to (y-axis), each survey question (x-axis). In Figure 4-(a) of the main article, respondents in Classes 3 and 4 are merged and labeled as True Supporters by re-calculating their mean responses, except for the Econ Demo question. Respondents in Classes 1 and 2 are merged and labeled as Non-supporters, except for the Econ Demo question.

Table A.5: Mean Responses by Class and Question, LPA on Treatment Group

Class	**Question**	**Estimate (SE)**	**Class**	**Question**	**Estimate (SE)**
1	Direct Support	4.1155 (0.0961)	4	Direct Support	4.0699 (0.1174)
1	Rev Court Gov	1.7939 (0.0800)	4	Rev Court Gov	1.4619 (0.1018)
1	Court Gov2	4.2828 (0.0831)	4	Court Gov2	4.6985 (0.0857)
1	Court Gov3	4.2826 (0.0639)	4	Court Gov3	3.9850 (0.0847)
1	Information	3.8815 (0.1050)	4	Information	4.2911 (0.1425)
1	Happy	3.1766 (0.0683)	4	Happy	3.2276 (0.1018)
1	Econ Demo	4.3371 (0.0565)	4	Econ Demo	1.9337 (0.0900)
1	Conflict	2.9270 (0.1241)	4	Conflict	3.4431 (0.1937)
2	Direct Support	3.2129 (0.2396)	5	Direct Support	3.3386 (0.2089)
2	Rev Court Gov	2.7031 (0.1977)	5	Rev Court Gov	2.2845 (0.2612)
2	Court Gov2	2.2709 (0.2531)	5	Court Gov2	3.8251 (0.2492)
2	Court Gov3	2.7614 (0.2386)	5	Court Gov3	1.8963 (0.0608)
2	Information	3.8805 (0.2891)	5	Information	3.4813 (0.1732)
2	Happy	3.7491 (0.1444)	5	Happy	3.4713 (0.1048)
2	Econ Demo	3.6751 (0.1717)	5	Econ Demo	1.5480 (0.1068)
2	Conflict	2.1388 (0.3113)	5	Conflict	2.2351 (0.2307)
3	Direct Support	1.9383 (0.3276)	6	Direct Support	3.9260 (0.0960)
3	Rev Court Gov	1.6873 (0.5600)	6	Rev Court Gov	1.8894 (0.1742)
3	Court Gov2	4.2587 (0.4131)	6	Court Gov2	4.1217 (0.0918)
3	Court Gov3	4.0049 (0.2211)	6	Court Gov3	4.0089 (0.0273)
3	Information	3.2199 (0.3923)	6	Information	2.2895 (0.2290)
3	Happy	2.3277 (0.3182)	6	Happy	3.6952 (0.1254)
3	Econ Demo	2.2795 (0.5550)	6	Econ Demo	1.8522 (0.0964)
3	Conflict	2.3649 (0.4264)	6	Conflict	1.3010 (0.1626)

Note: Mean responses by classes and questions, from LPA on the treatment group. SE means standard errors. In Figure 4-(b) of the main article, respondents in Classes 1 and 4 are merged and labeled as True Supporters by re-calculating their mean responses, except for the Econ Demo question. Respondents in Class 3 are labeled as (Candid) Non-supporters. Respondents in Classes 2, 5, and 6 are labeled as Falsifiers. Whenever certain subgroups show stronger hints of falsification than others, then their responses are plotted in the main article. That is, for Q3-Court Gov 2, the main plot shows responses from Class 2; for Q4-Court Gov 3, the main figure plots merged responses from Classes 2 and 5; for Q5-Information and Q8-Conflict, the main figure plots responses from Class 6. For the other questions, all three classes' responses are merged by re-calculating their mean responses.

Figure A.6: LPA on Treatment Group



Note. This plot shows how six latent subgroups in the treatment group behave, or answer to (y-axis), each survey question (x-axis). In Figure 4-(b) of the main article, respondents in Classes 1 and 4 are merged and labeled as True Supporters by re-calculating their mean responses, except for the Econ Demo question. Respondents in Class 3 are labeled as (Candid) Non-supporters. Respondents in Classes 2, 5, and 6 are labeled as Falsifiers. Whenever certain subgroups show stronger hints of falsification than others, then their responses are plotted in the main article. That is, for Q3-Court Gov 2, the main plot shows responses from Class 2; for Q4-Court Gov 3, the main figure plots merged responses from Classes 2 and 5; for Q5-Information and Q8-Conflict, the main figure plots responses from Class 6. For the other questions, all three classes' responses are merged by re-calculating their mean responses.

3.3 Survey Questions that Do Not Capture Political Attitude

Some findings from the LPA analysis contradict our initial expectations and thus require further explanation. First, the question asking whether respondents would sacrifice democracy for economic development (Q7-Econ Demo) did not effectively capture either true or falsified political opinions in China. The LPA results suggest that both true supporters and non-supporters of the regime hold diverse views on this issue. We suspect that this is because certain Chinese citizens interpret the concept of democracy differently. For them, democracy may not align with the Western definition of liberal democracy, which is based on national-level free elections. Instead, it may refer to people-oriented policy implementation by a centralized authority, as defined by the Chinese government. Therefore, unless the meaning of democracy is clearly specified, such survey questions may be ill-suited for measuring preference falsification or true political attitudes among Chinese citizens.

Second, Q2-Rev Cour Gov asks whether the court and the government always agree on major cases. Both non-supporters and true regime supporters agreed with this statement, both before and after the purge. As such, this question fails to effectively distinguish between pro-regime and anti-regime attitudes. We believe this is because the question is not a regime-assessment question but rather a value-neutral, observation-based question. For example, if the question asked whether it is *correct* for the court to follow the government, rather than whether the court *does* follow the government, then true supporters would likely have answered affirmatively, and their stance would have remained consistent or even strengthened after the purge. This is precisely what we observe in the true supporters' responses to Q4-Court Gov 3. Therefore, observation-based questions may not be suitable for assessing political attitudes.

Finally, although life satisfaction questions (Q6-Happy) are commonly used in social science survey research, they do not appear to effectively differentiate between supporters and non-supporters. In the control group, regime supporters' responses to this question do not differ significantly from those of non-supporters. Moreover, contrary to expectations, non-supporters are slightly more satisfied with their lives than true supporters. Thus, using such general questions may not be suitable for assessing true political attitudes.

3.4 LPA on All Shanghai Respondents

We further ran LPA on all 400 Shanghai respondents to see if we had matched each subgroup correctly between the treatment and control groups. This time, with BIC = 8341.824, AIC = 8026.498, and an entropy level of 0.93, the most appropriate number of latent subgroups was eight. The results are shown in Table A.6. Figures A.7 and A.8 visualize the results in two plots for improved readability.

Within these eight groups, Classes 1, 4, and 7 (209 respondents) consistently show positive attitudes; Classes 2 and 5 (79 respondents) show negative attitudes; and Classes 6 and 8 (75 respondents) indicate clues of preference falsification. The remaining 37 respondents in Class 3, which does not have a corresponding subgroup in either the treatment or control group analysis, tend to give neutral or no answers to the questions. The proportions of these three subgroups roughly mirror those in the control and treatment group analysis. Also, their responses across survey questions exhibit patterns similar to those in the previous analysis.

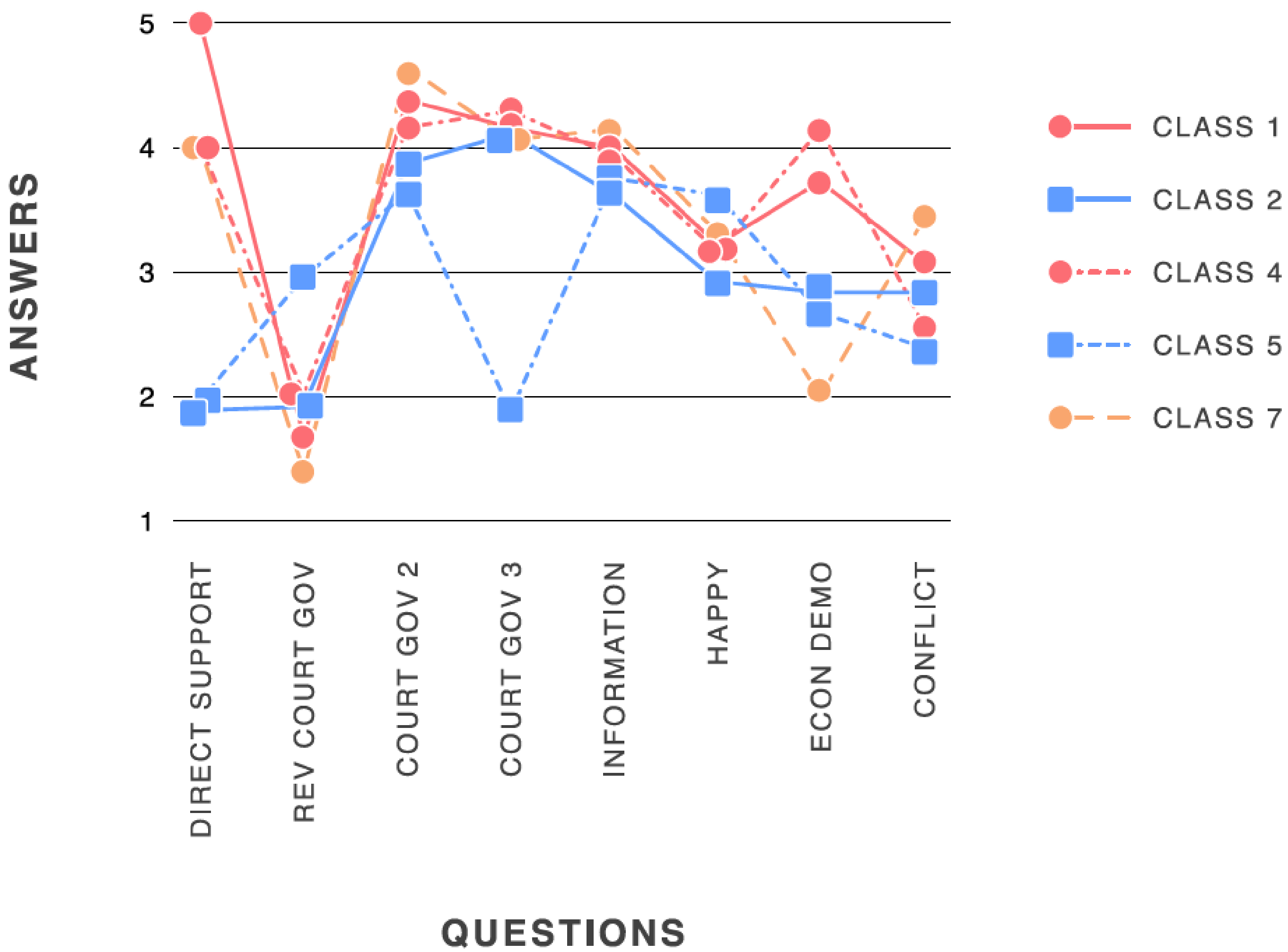
Nevertheless, the accuracy of individual matching yielded mixed results. First, we correctly matched 79.5 percent of true supporters in the treatment and control groups to Classes 1, 4, and 7 in the pooled population. Second, we also correctly matched 81.7 percent of candid non-supporters in the treatment and control groups. Third, 94.7 percent of the preference-falsifying respondents in the treatment group—who had expressed their true discontent when asked Q5-Information and Q8- Conflict—were matched accurately. However, we were only able to match 49.2 percent of the 63 preference-falsifying individuals who had expressed true disapproval when asked Q3-Court Gov 2 and Q4-Court Gov 3. Instead, this subgroup incorrectly included 21 true supporters from the control group. These results demonstrate that we cannot be very certain about the individual-level probability of being assigned to particular subgroups.

Table A.6: Mean Responses by Class and Question, LPA on 400 Shanghai Respondents

Class	Category	Estimate (SE)	Class	Category	Estimate (SE)
1	Direct Support	5.0000 (0.0000)	5	Direct Support	1.9728 (0.0374)
1	Rev Court Gov	1.6866 (0.0865)	5	Rev Court Gov	2.9561 (0.2383)
1	Court Gov2	4.3731 (0.0814)	5	Court Gov2	3.6433 (0.2323)
1	Court Gov3	4.1642 (0.1021)	5	Court Gov3	1.8905 (0.0778)
1	Information	3.9851 (0.1456)	5	Information	3.7556 (0.1705)
1	Happy	3.2239 (0.0778)	5	Happy	3.6230 (0.1137)
1	Econ Demo	3.7313 (0.1404)	5	Econ Demo	2.6751 (0.2162)
1	Conflict	3.0746 (0.1719)	5	Conflict	2.3955 (0.2530)
2	Direct Support	1.9052 (0.0506)	6	Direct Support	4.0000 (0.0000)
2	Rev Court Gov	1.9235 (0.1404)	6	Rev Court Gov	2.0302 (0.1165)
2	Court Gov2	3.8848 (0.1877)	6	Court Gov2	3.3359 (0.3499)
2	Court Gov3	4.1118 (0.1249)	6	Court Gov3	2.1876 (0.2033)
2	Information	3.6206 (0.1691)	6	Information	3.6857 (0.1421)
2	Happy	2.9300 (0.1331)	6	Happy	3.5463 (0.1131)
2	Econ Demo	2.8332 (0.1897)	6	Econ Demo	2.2963 (0.2771)
2	Conflict	2.8404 (0.1697)	6	Conflict	2.3681 (0.1866)
3	Direct Support	3.0000 (0.0000)	7	Direct Support	4.0000 (0.0000)
3	Rev Court Gov	2.4054 (0.1353)	7	Rev Court Gov	1.4085 (0.0933)
3	Court Gov2	3.6486 (0.1478)	7	Court Gov2	4.6453 (0.0904)
3	Court Gov3	3.4595 (0.1115)	7	Court Gov3	4.0584 (0.0865)
3	Information	4.1892 (0.2092)	7	Information	4.1520 (0.1353)
3	Happy	3.5676 (0.1099)	7	Happy	3.3270 (0.0781)
3	Econ Demo	3.1018 (0.1452)	7	Econ Demo	2.0023 (0.0883)
3	Conflict	2.6757 (0.1607)	7	Conflict	3.4269 (0.1943)
4	Direct Support	4.0000 (0.0000)	8	Direct Support	4.0000 (0.0000)
4	Rev Court Gov	2.0271 (0.1288)	8	Rev Court Gov	1.9535 (0.2759)
4	Court Gov2	4.1656 (0.1630)	8	Court Gov2	3.9864 (0.3763)
4	Court Gov3	4.2994 (0.0875)	8	Court Gov3	4.0256 (0.1613)
4	Information	3.9401 (0.0970)	8	Information	2.3961 (0.2809)
4	Happy	3.1816 (0.0848)	8	Happy	3.7203 (0.1281)
4	Econ Demo	4.1449 (0.0900)	8	Econ Demo	1.9743 (0.2649)
4	Conflict	2.5795 (0.1600)	8	Conflict	1.4112 (0.2074)

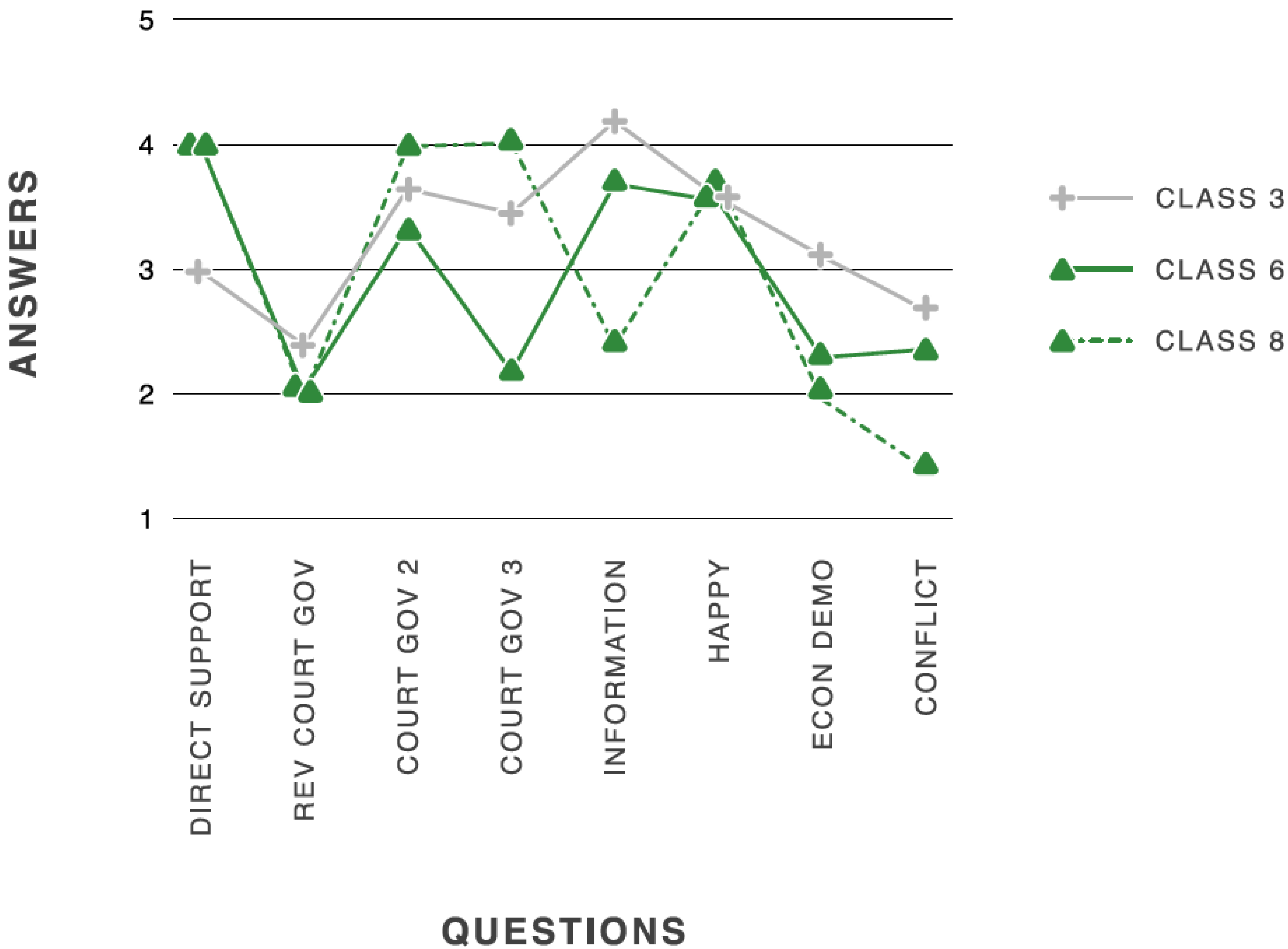
Note: Mean responses by classes and questions, from LPA on the 400 Shanghai respondents. SE means standard errors.

Figure A.7: LPA on All 400 Shanghai Respondents (1)



Note: This plot shows how eight latent subgroups of all Shanghai respondents behave, or answer to (y-axis), each survey question (x-axis).

Figure A.8: LPA on All 400 Shanghai Respondents (2)



Note: This plot shows how eight latent subgroups of all Shanghai respondents behave, or answer to (y-axis), each survey question (x-axis).

3.5 Careless Respondents and Non-responses

In the main article, we treated non-responses as neutral answers for the following reasons. First, it is better to treat them as certain values rather than dropping these non-responses, because self-censorship is a form of preference falsification and we would want to preserve that information. If we drop NAs, we lose the information provided by 95 respondents (23.75 percent) who have at least one non-response to the questions used for LPA. If we keep non-responses as missing values and run LPA, the function then fills in these missing values with single imputations, resulting in unstable profile estimations each time we conduct LPA.

One way to treat non-responses as definite values is to code them as 0 or extreme numbers, such as 999. However, because LPA classifies subgroups by calculating their mean answers, μ_k, such coding would seriously distort the analysis. Therefore, in this paper, we coded non-responses as neutral answers. Specifically, for 5-scale responses, we treated missing values as 3. When answers were offered on a 4-scale, we coded answer 4 as 5 and answer 3 as 4, so that we could again code non-responses as 3. In this case, 3 only represents non-responses, and we need not worry about treating actual neutral answers and non-responses identically.

We recognize that this approach may not be ideal, as it treats truly neutral answers and non-responses as the same. Nonetheless, for identifying preference falsification, this is the optimal approach that captures the information in non-responses without skewing the analysis. Indeed, providing a neutral answer is akin to giving no opinion, which can also serve as a way of concealing true preferences.

With this approach, we can observe systematic non-responses if they occur. That is, if a certain group of respondents tends to give non-responses, such behavior would be captured by the clusters around these neutral values. Class 3 of the 400 Shanghai respondents in Section 3.4. is an example.

Yet, preference falsification is not the only reason for non-responses in a public opinion survey. Non-responses may also reflect a lack of effort in answering survey questions, often because respondents do not pay attention to the content of the questions, do not read them carefully, or do not think about them thoroughly (Meade and Craig 2012; Curran 2016; Dunn et al. 2018). Because "careless" respondents can skew the analysis, we excluded them before running LPA in this section.

We are especially concerned about careless responses in Section E47 of the questionnaire (Figure A.9). Not only does this section contain many of the questions we used for analysis (highlighted), but it also includes as many as 37 questions in a row, which likely induced inattentive attitudes in answering.

Therefore, we took four approaches to detect careless responses: (1) the number of non-responses, (2) the longest string of identical responses (Longstring index), (3) intra-individual response variability (IRV), and (4) Mahalanobis Distance. Because relying on a single indicator to identify careless respondents is not recommended (Curran 2016), we labeled respondents who had two or more red flags as careless. Among the 400 Shanghai respondents, only three were identified as careless. Given this small number, the LPA results were not different from the previous results.

3.6 Predicting Truthful Answers through Imputations

In the main article, we demonstrated that LPA uncovers hidden subgroups among Shanghai respondents. From this subgroup classification, we inferred which survey questions were sensitive or non-sensitive to Chinese citizens. In addition, LPA provides an individual-level probability of membership in a particular subgroup. Therefore, using this individual-level information, a researcher may choose to modify the survey dataset by replacing falsified answers with predicted truthful answers.

Previously, we noted that existing studies on preference falsification focus on either self-censorship or untruthful answers. Now that we know which subgroup tends to falsify their answers on specific questions—and who belongs to which subgroup—we could theoretically replace those respondents' untruthful answers with non-responses. This simplifies identifying preference falsification because there would be only one form of preference falsification: the

Figure A.9: Section E47 of the Chinese General Social Survey 2006

E47. 您是否同意下列说法？（每行单选）

	非常不同意	不同意	同意	非常同意	*[不回答]*	
1. 如果企业或公司不去追求利润，这个社会就会发展不起来	1	2	3	4	5	(4342)
2. 普通工人应该在企业或公司的决策、管理中发挥较大的作用	1	2	3	4	5	(4343)
3. 企业老板从工人或员工身上得到了很多好处	1	2	3	4	5	(4344)
4. 没有人管着，公司雇员也一样能把事情干好	1	2	3	4	5	(4345)
5. 穷人之所以会穷，一个重要原因是接受的教育太少了	1	2	3	4	5	(4346)
6. 穷人之所以会穷，是因为他们不愿意工作	1	2	3	4	5	(4347)
7. 专家学者的道德修养往往比普通人高	1	2	3	4	5	(4348)
8. 政府某些政策不妥当，是造成贫穷的重要原因	1	2	3	4	5	(4349)
9. 政府应该加大对私营业主的管理和监控的力度	1	2	3	4	5	(4350)
10. 政府与法院在重大案件上的态度总是一致的	1	2	3	4	5	(4351)
11. 服从政府总是不会错的	1	2	3	4	5	(4352)
12. 只有政府大力支持、配合，法律才能正常运转	1	2	3	4	5	(4353)
13. 拉开贫富差距，有利于调动人们努力工作的积极性	1	2	3	4	5	(4354)
14. 应该从收入高的人那里征更多的税来帮助穷人	1	2	3	4	5	(4355)
15. 国有医院应该设立更多的高级病房以满足有支付能力的病人的需要	1	2	3	4	5	(4356)
16. 中国应该加大对第三世界国家的经济援助	1	2	3	4	5	(4357)
17. 虽然非洲离我们很远，我们还是应该多派医疗队去帮助那里的黑人朋友	1	2	3	4	5	(4358)
18. 任何人都应该支持自己的国家，既使你认为它做的不对	1	2	3	4	5	(4359)
19. 中国应该派兵去参加联合国维持和平部队	1	2	3	4	5	(4360)
20. 目前有些国际势力试图阻止中国的发展和崛起	1	2	3	4	5	(4361)
21. 两个国家经济联系越密切，那么它们之间发生武力冲突的可能性越小	1	2	3	4	5	(4362)
22. 为了保护我国的经济，我们应该限制其他国家的产品进口	1	2	3	4	5	(4363)
23. 我们应该修改不符合世界贸易组织(WTO)规则的国内法规	1	2	3	4	5	(4364)
24. 我们应该禁止外国人在我国购买大型国有企业	1	2	3	4	5	(4365)
25. 外国大公司对中国本地公司形成了巨大的竞争压力	1	2	3	4	5	(4366)
26. 不限制进口可以让我们买到更好的产品	1	2	3	4	5	(4367)
27. 国外电影，音乐和书籍对中国文化有不利影响	1	2	3	4	5	(4368)
28. 外国产品就是比国产的质量好	1	2	3	4	5	(4369)
29. 我了解我们村/社区发生的事情，所以我有权参与村/社区的事务	1	2	3	4	5	(4370)
30. 如果有机会参加县长/区长的直接选举，我一定会积极参加投票	1	2	3	4	5	(4371)
31. 政治的事情太复杂，像我这样的人不懂	1	2	3	4	5	(4372)
32. 只要经济能保持稳定发展，就不必提高民主程度	1	2	3	4	5	(4373)
33. 经济条件好的人比经济条件差的人在公共事务上有更大的发言权	1	2	3	4	5	(4374)
34. 只有那些有专业知识和能力的人在决策中才有发言权	1	2	3	4	5	(4375)
35. 如果对本地政策有不同意见时，人们有权向上级政府申述	1	2	3	4	5	(4376)
36. 当法律规定与政府决策不一致时，执法者应以政府决策为执行依据	1	2	3	4	5	(4377)
37. 只要不违法犯罪，法律基本与我的生活无关	1	2	3	4	5	(4378)

Note. Section E47 of the Chinese General Social Survey 2006. Highlighted questions are used in the analysis of this paper.

non-response. These non-responses could then be predicted based on information regarding a respondent's true political attitudes measured with non-sensitive survey items. Recent developments in imputation methods, such as multiple imputations (Honaker et al. 2011; Liu 2023), may facilitate this approach. Admittedly, the predicted values cannot be fully verified, and some may be reluctant to manipulate a dataset. Nonetheless, a predicted dataset can still be considered to be used in a robustness test of a study that suspects preference falsification in a public opinion survey.

However, beyond checking for robustness, we also warn against heavy reliance on the individual probability of subgroup membership. Compared to the classification of mean answers across questions, which reveals subgroups and their average patterns, individual-level probabilities can be unstable and therefore less accurate. Moreover, because probability can be indefinite, researchers should be cautious in deciding where to set a cutoff. For instance, if a respondent belongs to one group with a probability of 0.6 and another with a probability of 0.4, the assignment is somewhat tentative.

With these limitations in mind, we tried predicting truthful answers to Q1-Direct Support using the procedure described earlier. To be more conservative at the individual level, instead of replacing only the answers of individuals identified as preference-falsifying, we replaced the answers of all individuals who displayed a positive gap between their identified sensitive and non-sensitive questions. In other words, consistent with existing literature on measuring preference falsification, we adopted this approach:

$$\begin{aligned}\text{Preference Falsification}_i = \text{Q1. Direct Support}_i - \\ (\text{Q3. Court Gov } 2_i + \\ \text{Q4. Court Gov } 3_i + \\ \text{Q5. Information}_i + \\ \text{Q8. Conflict}_i)/4\end{aligned} \tag{6}$$

With this measurement as the dependent variable, we ran a regression model as in Equation . The results in Table A.7 show that on average, respondents who participated in the survey after the purge are likely to have a larger gap between their answers to sensitive and non-sensitive questions.

$$\text{Preference Falsification}_{ij} = \delta \text{Treatment}_{ij} + \mathbf{X_{ij}}\beta + \eta_j + \epsilon_{ij} \tag{7}$$

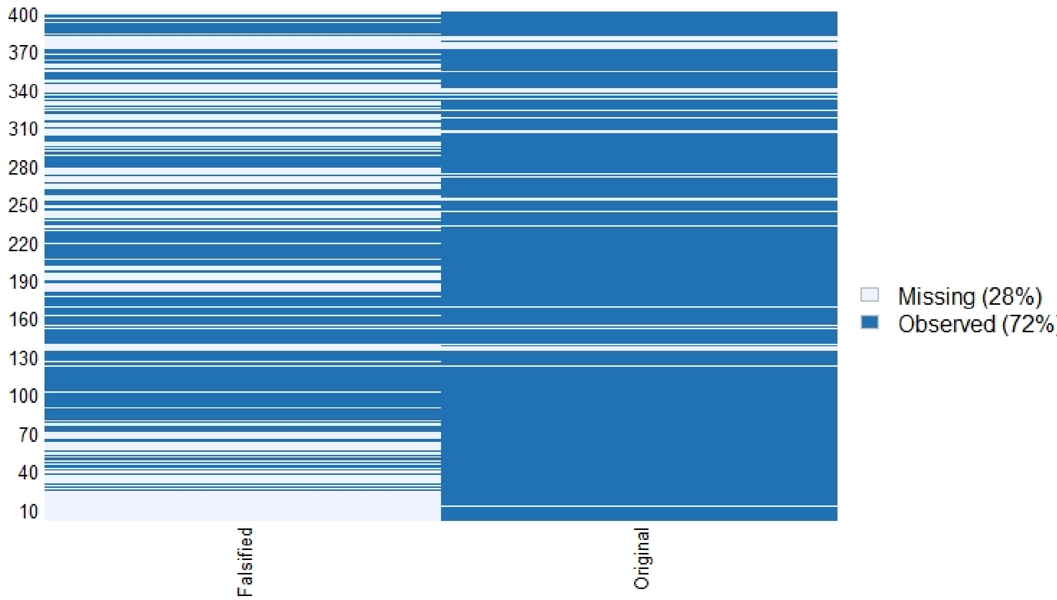
Table A.7: Linear Regression Analysis - Treatment Effect on Preference Falsification

	Dependent Variable: Preference Falsification		
	(1)	(2)	(3)
Purge (Treatment)	0.266**	0.264**	0.366***
	(0.114)	(0.117)	(0.118)
Constant	−0.418***	−0.031	−0.321
	(0.096)	(0.278)	(0.326)
Covariates	No	Yes	Yes
District FE	No	No	Yes
Observations	400	400	400
R^2	0.013	0.030	0.147
Adjusted R^2	0.011	0.003	0.113
Residual Std. Error	1.044 (df = 398)	1.048 (df = 388)	0.989 (df = 384)
F Statistic	5.424** (df = 1; 398)	1.104 (df = 11; 388)	4.400*** (df = 15; 384)
Note 1.			*p<0.1; **p<0.05; ***p<0.01

Note 2. Covariates include age, female, ethnicity, college, Party affiliation, religion, occupation sector (public versus private), marriage status, Internet usage, and income level.

For the individuals who have positive values of Preference Falsification, we treated their answers to Q1-Direct Support as missing values. We do not necessarily believe everyone in this group is lying in an absolute sense, but we took this approach to avoid retaining any respondents who might be preference-falsifying. Figure A.10 displays which observations were treated as missing. On the left-hand side of the x-axis are the missing patterns after replacing "falsified" answers with missing values, and on the right-hand side are the original

Figure A.10: Missingness Map, Answers to Question 1. Direct Support



Note. Missing answers to the question that asks if "It is always correct to follow the government." Each number in the Y-axis indicates the respondent's ID.

non-responses. The y-axis corresponds to each respondent's ID. Initially, 37 out of 400 survey participants did not answer the question. Additionally, 147 more respondents had positive values for Preference Falsification. As a result, 184 observations were treated as missing values.[1]

The next step was to fill in these missing values with predicted values derived from other survey answers. We used `R` package `AMELIA` (Honaker et al. 2011), which is based on the EMB algorithm proposed by Honaker and King (2010). Table A.8 summarizes the resulting four models, each combining 20 imputations. Because Model 4 exhibited the best performance based on over-imputation diagnostics, we calculated the average of its 20 imputations and

[1]Some of the original non-responses can be considered a form of preference falsification in an authoritarian context. A participant critical of the regime may choose not to answer the question rather than overstate their support.

used it to fill in missing values. Note that the predicted values can be decimals, not rounded integers. This is intentional because decimal points can provide more nuanced information about the level of support than strict integers.

Overall, the predicted level of support among Shanghai respondents decreased by 0.34 points on a 4-point scale compared to the observed level of support. Therefore, this 0.34 decrease can be interpreted as the aggregated magnitude of preference falsification among Shanghai respondents. Interestingly, this magnitude closely resembles findings by Jiang and Yang (2016) (0.36 out of a 4-point scale), despite the use of different questions to measure falsified and truthful political attitudes. With this updated dataset, preference falsification due to the purge do not appear to exist anymore (Table A.9).

Table A.8: Combined Imputations

	1st 20 Imputations	2nd 20 Imputations	3rd 20 Imputations	4th⊛ 20 Imputations
(Intercept)	0.531*	0.067	-0.276	-0.399
	(0.299)	(0.448)	(0.362)	(0.410)
Court Gov	0.479***	0.443***	0.421***	0.426***
	(0.063)	(0.058)	(0.061)	(0.056)
Trust Gov. Info.	0.162***	0.149***	0.124***	0.127***
	(0.031)	(0.038)	(0.034)	(0.033)
Cadre Benefit	-0.146**	-0.068	-0.087	-0.047
	(0.064)	(0.066)	(0.061)	(0.068)
Econ vs. Demo	0.068	0.082	0.071*	0.070
	(0.042)	(0.051)	(0.039)	(0.044)
Participate Politics	-0.068	-0.049	-0.025	-0.015
	(0.054)	(0.051)	(0.049)	(0.049)
Court-Gov Consistent		-0.080*	-0.066	-0.083*
		(0.042)	(0.041)	(0.049)
Life Satisfaction		0.125**	0.113***	0.122***
		(0.052)	(0.043)	(0.046)
Cadre-Ppl. Conflict		0.098**	0.100**	0.137***
		(0.046)	(0.045)	(0.051)
Outspoken			0.020	0.002
			(0.050)	(0.043)
Policy Opinion			0.027	0.041
			(0.056)	(0.070)
Peoples Congress			-0.008	-0.002
			(0.040)	(0.036)
Understand Politics			0.155***	0.142***
			(0.038)	(0.048)

Note 1. *p<0.1; **p<0.05; ***p<0.01
Note 2. Court Gov combines Questions 3 and 4. Participate Politics combines one's belief that they have the right to participate in public affairs—because they are aware of what is happening in their communities—and their willingness to vote if they have an opportunity to participate in a direct election for their country's leadership. Outspoken represents one's support for assembly, demonstration, strike, and appeal. Peoples Congress measures whether one has participated in voting, nominating, encouraging others to vote, or engaging in local people's congress activities. These variables are combined due to the high correlations among them. Cadre Benefit is coded as 1 if the respondent selected state cadres as the primary beneficiaries in the past ten years, compared with other groups such as workers, farmers, managers of enterprises, private entrepreneurs, foreign businessmen, or intellectuals. Policy Opinion asks whether one agrees that "people have the right to appeal if they have different policy opinions." Understand Politics asks if one believes that "politics is so complicated that [they] cannot understand."
Note 3. ⊛ The fourth model includes demographic information along with Foreign language skills, Party affiliation, College, income level, and Internet usage. They are not statistically significant.

Table A.9: Treatment Effect on Preference Falsification After Multiple Imputations

	Dependent Variable: Preference Falsification	
	(1)	(2)
Purge (Treatment)	−0.030	−0.044
	(0.054)	(0.054)
Constant	−0.399***	−0.732***
	(0.045)	(0.129)
Covariates	No	Yes
District FE	No	No
Observations	400	400
R^2	0.001	0.048
Adjusted R^2	−0.002	0.021
Residual Std. Error	0.490 (df = 398)	0.485 (df = 388)
F Statistic	0.313 (df = 1; 398)	1.778* (df = 11; 388)
Note 1.		*p<0.1; **p<0.05; ***p<0.01

Note 2. Covariates include age, female, ethnicity, college, Party affiliation, religion, occupation sector (public versus private), marriage status, Internet usage, and income level. District fixed effects are not included this time because the dependent variable does not vary across different districts.

4 Generalizability of LPA

4.1 Other Authoritarian Contexts

In Test 2, we observed that LPA can be an effective tool for detecting preference falsification in the study of Chinese politics. Nevertheless, preference falsification is more prevalent in China than in some other countries (Shen and Truex 2021). Moreover, the political purge in Shanghai altered the internal political environment such that preference falsification among respondents may have further increased. These observations raise the question of LPA's broader usefulness in studying different authoritarian contexts.

Because LPA captures variations in answer patterns across questions, it may produce stronger evidence of preference falsification when two conditions are met. First, the actual magnitude of the falsification must be genuinely strong. Second, there must be a subpopulation with a significant number of respondents who repeatedly exhibit inconsistencies in their answers. In this regard, we acknowledge that LPA can more reliably detect preference falsification in a population where a substantial group of people are strongly inclined to conceal or alter their preferences. When preference falsification is weaker in magnitude but a significant number of subgroups consistently show discrepancies between their answers, LPA can still capture this small gap but researchers may be less confident in labeling such a minor gap as preference falsification. Also, when preference falsification is weak and few respondents show inconsistencies, LPA might fail to capture that latent subgroup.

Indeed, while preference falsification is presumed to exist in authoritarian countries generally, it is more common in some systems than others. For example, Russia, another major authoritarian country frequently compared to China, differs from China in the prevalence of preference falsification. Regarding Putin's approval ratings, Frye et al. (2017, 2023) conducted two successive studies, both indicating that preference falsification exists in Russia, though to a lesser degree than in China. They also observed that Putin's high approval ratings largely reflect the genuine attitudes of Russian citizens. In the current Russo-Ukrainian war, the Russian public has exhibited some preference falsification over whether to support the war, but a notable portion of the population genuinely supports it (Chapkovski and Schaub 2022). Moreover, even though the war has limited some discursive space, the level of self-censorship among the Russian public has not increased significantly (Rosenfeld 2023).

Even within a single country, the usefulness of LPA in analyzing public opinion surveys may vary depending on the targeted population. In China, self-censorship—one form of preference falsification—often concentrates among "citizens who are socially and politically marginalized" (Shen and Truex 2021; Carter et al. 2024). Additionally, the perceived sensitivity of the survey questions can influence the degree of preference falsification. Chinese citizens exhibit particularly high non-response rates to direct regime-assessment questions (Robinson and Tannenberg 2019). Through a survey of 37 African countries, Tannenberg (2022) finds that populations in authoritarian countries develop strong biases when responding to questions on trust, approval, and perceptions of corruption, along with less pronounced biases for relatively innocuous, non-political items.

In conclusion, LPA can be more useful in studying preference falsification in certain authoritarian contexts, such as China, than in others, like Russia. Moreover, even within the same country, the results of LPA will depend on variations in the target population, the timing, and the specific questionnaire employed.

4.2 Test 3. Social Desirability Bias in the U.S.

Originally, 2,596 U.S. citizens participated in the World Value Survey in 2017. In the main article, we used observations from a subset of 1,941 respondents who indicated they would vote for either the Republican Party or the Democratic Party if a national election were held the next day.

As before, non-responses are treated as neutral answers. For Q1-LR, 59 people did not respond; for Q3-Job Over Immigrant, 7; for Q4-Immigrant, 9; for Q5-Conf. Church, 12; for Q6-God, 18; for Q7-Homosexual, 11; and for Q8-Abortion, 29 did not respond.

Table A.10: Mean Responses by Class and Question, LPA on U.S. World Value Survey

Class	Category	Estimate (SE)	Class	Category	Estimate (SE)
1	LR	3.2087 (0.0733)	4	LR	3.6039 (0.0524)
1	Conf.gov	4.1811 (0.0274)	4	Conf.gov	4.2698 (0.0202)
1	Job over immigrant	3.4899 (0.0575)	4	Job over immigrant	3.4511 (0.0336)
1	Immigrant	3.0210 (0.0665)	4	Immigrant	2.9868 (0.0428)
1	Conf.church	2.7608 (0.0878)	4	Conf.church	3.9626 (0.0466)
1	God	2.0310 (0.1140)	4	God	4.7282 (0.0295)
1	Homosexual	2.4523 (0.0875)	4	Homosexual	3.0367 (0.0537)
1	Abortion	2.5631 (0.0965)	4	Abortion	1.5074 (0.0600)
2	LR	3.1232 (0.0930)	5	LR	1.7052 (0.0684)
2	Conf.gov	1.6769 (0.0310)	5	Conf.gov	1.4327 (0.0380)
2	Job over immigrant	3.3473 (0.0588)	5	Job over immigrant	2.6256 (0.0598)
2	Immigrant	2.8554 (0.0584)	5	Immigrant	2.1574 (0.0632)
2	Conf.church	3.6588 (0.0782)	5	Conf.church	3.2966 (0.0878)
2	God	4.8074 (0.0245)	5	God	4.6312 (0.0609)
2	Homosexual	3.1074 (0.0850)	5	Homosexual	1.7617 (0.0730)
2	Abortion	1.4546 (0.0757)	5	Abortion	3.1348 (0.1170)
3	LR	2.2569 (0.1007)	6	LR	1.5279 (0.0540)
3	Conf.gov	1.6415 (0.0448)	6	Conf.gov	1.4049 (0.0310)
3	Job over immigrant	3.0137 (0.0844)	6	Job over immigrant	2.6412 (0.0572)
3	Immigrant	2.5328 (0.0838)	6	Immigrant	2.0490 (0.0490)
3	Conf.church	2.7403 (0.0668)	6	Conf.church	1.8469 (0.0478)
3	God	2.6446 (0.1153)	6	God	0.7508 (0.0332)
3	Homosexual	1.9997 (0.0801)	6	Homosexual	1.5046 (0.0516)
3	Abortion	3.1127 (0.0887)	6	Abortion	3.9844 (0.0554)

Note: Mean responses by classes and questions, from LPA on U.S. World Value Survey. SE means standard errors. In Figure 5-(b) of the main article, Class 1 is labeled Subgroup 2, Class 4 as Subgroup 1, Class 5 as Subgroup 4, and Class 6 as Subgroup 3.

Before applying LPA to the data from these Republicans and Democrats, we chose the number of latent subgroups. As shown in Figure A.11, six appears to be appropriate. Table A.10 and Figure A.12 show the LPA results with six subgroups. In the main article, Classes 2 and 3 are dropped because they are considered to be in the gray area. First, the 327 respondents in Class 2 lean "right" in their political ideology but are dissatisfied with the Trump government, and over 50 percent of them (168 respondents) said they would vote for the Democratic Party if a national election were held tomorrow. Second, the 256 respondents in Class 3 are the group with the most neutral opinions across the eight questions. In Figure 5-(b) of the main article, Class 1 is labeled Subgroup 2, Class 4 as Subgroup 1, Class 5 as Subgroup 4, and Class 6 as Subgroup 3.

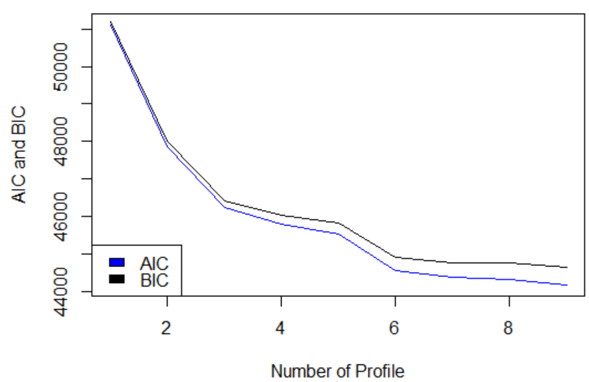
Table A.11: Mean Responses by Class and Question, LPA on U.S. World Value Survey, Two Classes

Class	Category	Estimate (SE)	Class	Category	Estimate (SE)
1	LR	3.4907 (0.0384)	2	LR	2.1255 (0.0322)
1	Conf.gov	4.2425 (0.0172)	2	Conf.gov	1.5286 (0.0133)
1	Job over immigrant	3.4625 (0.0303)	2	Job over immigrant	2.8929 (0.0206)
1	Immigrant	2.9971 (0.0329)	2	Immigrant	2.3791 (0.0243)
1	Conf.church	3.6154 (0.0428)	2	Conf.church	2.8458 (0.0340)
1	God	3.9494 (0.0515)	2	God	3.1215 (0.0498)
1	Homosexual	2.8688 (0.0391)	2	Homosexual	2.0766 (0.0300)
1	Abortion	1.8124 (0.0501)	2	Abortion	2.9537 (0.0421)

Note: Mean responses by classes and questions, from LPA on U.S. World Value Survey, when the number of subgroups is forced to be two. SE means standard errors.

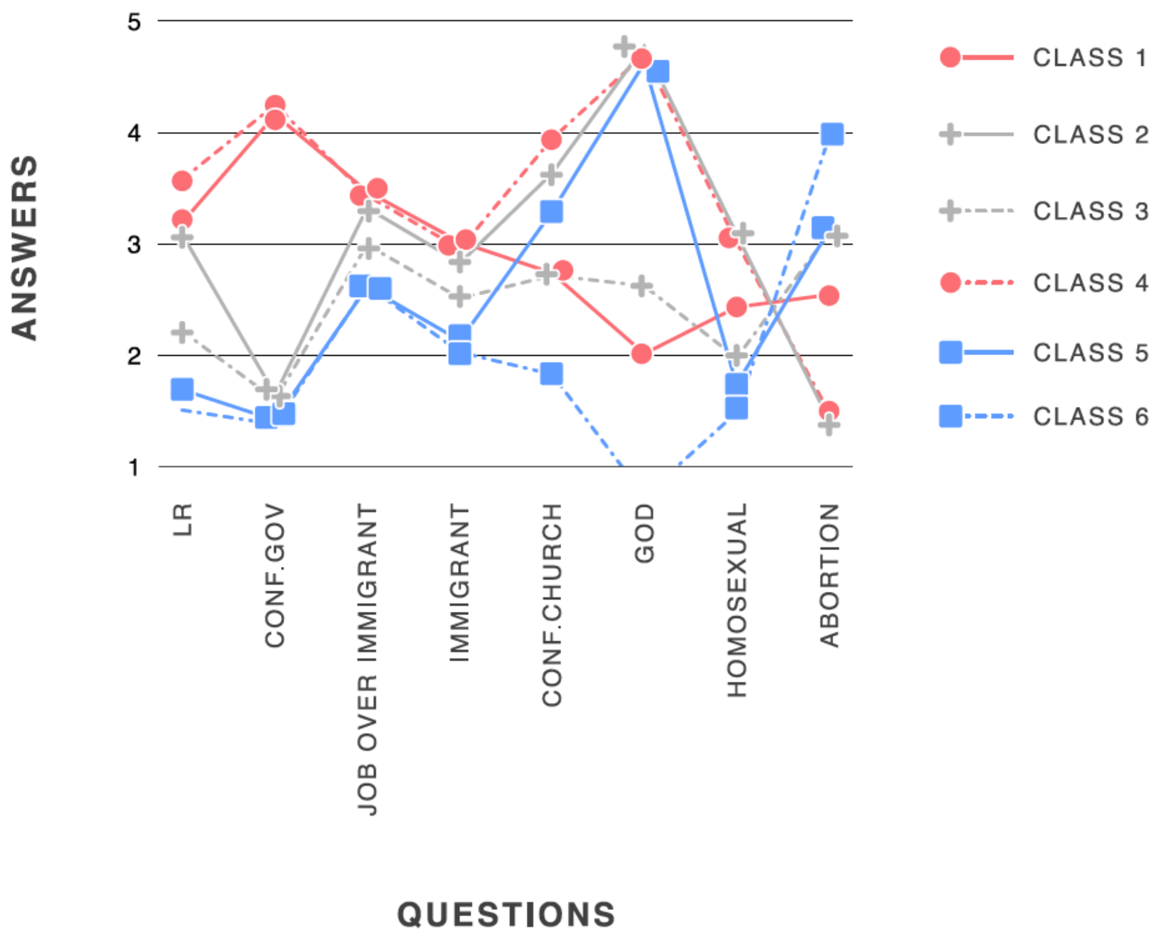
When we forced the number to be two, AIC was 47876.69, BIC was 48015.96, and the entropy level was 0.99. Table A.11 shows results with two subgroups.

Figure A.11: AIC and BIC on U.S. World Value Survey



Note: AIC is 44563.17, BIC is 44903.00, and the entropy level is 0.92 with six subgroups.

Figure A.12: LPA on U.S. World Value Survey



Note. This plot shows how six latent subgroups of all U.S. respondents behave, or answer to (y-axis), each survey question (x-axis). When the number of subgroups is set to six, AIC is 44563.17, BIC is 44903.00, and the entropy level is 0.92. In Figure 5-(b) of the main article, Class 1 is Subgroup 2, Class 4 is Subgroup 1, Class 5 is Subgroup 4, and Class 6 is Subgroup 3.

References

Carter, Erin Baggott, Brett L Carter, and Stephen Schick. 2024. Do Chinese citizens conceal opposition to the CCP in surveys? evidence from two experiments. *The China Quarterly*: 1–10.

Chapkovski, Philipp and Max Schaub. 2022. Solid support or secret dissent? a list experiment on preference falsification during the russian war against ukraine. *Research & Politics* 9 (2): 20531680221108328.

Curran, Paul G. 2016. "Methods for the Detection of Carelessly Invalid Responses in Survey Data". *Journal of Experimental Social Psychology* 66 : 4–19.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)* 39 (1): 1–22.

Dunn, Alexandra M., Eric D. Heggestad, Linda R. Shanock, and Nels Theilgard. 2018. "Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences". *Journal of Business and Psychology* 33 (1): 105–121.

Feldman, Stanley and Christopher Johnston. 2014. Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology* 35 (3): 337–358.

Frye, Timothy, Scott Gehlbach, Kyle L Marquardt, and Ora John Reuter. 2017. Is Putin's popularity real? *Post-soviet affairs* 33 (1): 1–15.

Frye, Timothy, Scott Gehlbach, Kyle L Marquardt, and Ora John Reuter. 2023. Is Putin's popularity (still) real? a cautionary note on using list experiments to measure popularity in authoritarian regimes. *Post-Soviet Affairs* 39 (3): 213–222.

Honaker, James and Gary King. 2010. "What to Do About Missing Values in Time-Series Cross-Section Data". *American Journal of Political Science* 54 (2): 561–581.

Honaker, James, Gary King, and Matthew Blackwell. 2011. "Amelia II: A Program for Missing Data". *Journal of Statistical Software* 45 (7): 1–47.

Jiang, Junyan and Dali L Yang. 2016. "Lying or Believing? Measuring Preference Falsification From a Political Purge in China". *Comparative Political Studies* 49 (5): 600–634.

Liu, Naijia. 2023. "A Latent Factor Approach to Missing Not at Random". Working Paper.

Meade, Adam W. and S. Bartholomew Craig. 2012. "Identifying Careless Responses in Survey Data". *Psychological Methods* 17 (3): 437–455.

Mooijaart, Ab and Peter GM Van der Heijden. 1992. The em algorithm for latent class analysis with equality constraints. *Psychometrika* 57 : 261–269.

Oberski, Daniel. 2016. Mixture models: Latent profile and latent class analysis. *Modern statistical methods for HCI*: 275–287.

Oser, Jennifer, Marc Hooghe, and Sofie Marien. 2013. Is online participation distinct from offline participation? a latent class analysis of participation types and their stratification. *Political research quarterly* 66 (1): 91–101.

Robinson, Darrel and Marcus Tannenberg. 2019. Self-censorship of regime support in authoritarian states: Evidence from list experiments in China. *Research & Politics* 6 (3): 1–9.

Rosenfeld, Bryn. 2023. Survey research in Russia: in the shadow of war. *Post-Soviet Affairs* 39 (1-2): 38–48.

Shen, Xiaoxiao and Rory Truex. 2021. In search of self-censorship. *British Journal of Political Science* 51 (4): 1672–1684.

Stoica, Petre and Yngve Selen. 2004. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine* 21 (4): 36–47.

Tannenberg, Marcus. 2022. The autocratic bias: self-censorship of regime support. *Democratization* 29 (4): 591–610.

Vrieze, Scott I. 2012. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (AIC) and the bayesian information criterion (BIC). *Psychol Methods* 17 (2): 228–43.