Survey Research Methods (2025)
Vol. 19, No. 4, pp. 467-482
https://doi.org/10.18148/srm/2025.v19i3.8361
European Survey Research Association

Retrieving True Preference under Authoritarianism

Jongyoon Baik[1] · Xiaoxiao Shen[2]
[1]The Chinese University of Hong Kong, Shenzhen
[2]Northeastern University

Scholars of authoritarian politics identify preference falsification in public opinion surveys by measuring the difference between a respondent's answers to politically sensitive questions and non-sensitive questions. Yet, the selection of questions is not empirically tested but justified only by the researchers' prior knowledge in the field. In this paper, we explain how latent profile analysis (LPA), a tool to analyze survey respondents based on their answer patterns, can provide observation-based evidence of the potential existence of preference falsification. We first provide a theoretical framework where we classify survey respondents under authoritarianism into true regime supporters, candid non-supporters, and preference-falsifying sub-populations. Then, we demonstrate the application of LPA to public opinion research through the use of data simulation, a quasi-experimental setting from Chinese General Social Survey data, and World Value Survey data.

Keywords: preference falsification; latent; profile analysis; authoritarian politics

1 Introduction

Though specific forms are different, scholars of authoritarian politics have identified preference falsification by measuring the gap between a respondent's answers to sensitive and non-sensitive questions. The assumption is that respondents tend to falsify their answers when asked sensitive questions but not when asked non- or less-sensitive questions. Jiang and Yang (2016) measures the difference between respondents' answers to questions explicitly asking about their support for the state and those that implicitly measure their true political attitudes. Similarly, Shen and Truex (2021) identifies self-censorship by comparing the non-response rates to regime assessment questions with those to non-sensitive questions.

Such an approach is indeed widely accepted as best practice. However, it still suffers from at least two significant problems. First, the studies rely on a strong assumption that if a respondent provides inconsistent answers to two questions, it is interpreted as evidence that one of the answers was untruthful. However, in reality, preference falsification is not the only source of inconsistency. A critical-minded person might agree with one political issue while disagreeing with another. An inattentive respondent could also provide random answers. Likewise, consistency does not always indicate truthfulness. A respondent might be consistently hiding their true negative opinions on every political question.

Second, the selection of sensitive versus non-sensitive questions in the existing studies tends to be subjective. Usually, researchers rely on their prior knowledge in the field to decide which questions are more sensitive than others and do not test whether the questions are actually appropriate to capture falsifying behavior. This practice undermines replicability, as the measurement of preference falsification heavily depends on the choice of questions. In this regard, Shen and Truex (2021) acknowledges that their "choice of [sensitive and non-sensitive] questions is inherently arbitrary, and [the measurement of self-censorship] may be sensitive to this decision."

In this paper, we attempt to address the second problem of subjectivity. We argue that scholars further need to

Supplementary Information The online version of this article (https://doi.org/10.18148/srm/2025.v19i3.8361) contains supplementary material, which is available to authorized users.

Corresponding author: Jongyoon Baik, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China (Email: jongyoonbaik@cuhk.edu.cn)

justify their choice of sensitive and non-sensitive questions by analyzing the actual response patterns from a completed survey. Therefore, we propose using latent profile analysis (LPA), a tool for analyzing survey respondents based on their answer patterns, in preference falsification studies. This approach provides *observation-based* evidence that a certain proportion of respondents exhibit inconsistent attitudes across a set of questions.

We demonstrate the application of LPA to preference falsification studies in three ways. First, we apply LPA to simulated data, where respondents are classified into true supporters, candid non-supporters, and preference-falsifying sub-populations. This shows that LPA correctly detects the answer patterns of these three subgroups. Second, we make use of a quasi-experimental setting in which preference falsification is likely to increase among survey respondents in China following a shock. LPA results show that, unlike in the control group, 30% of respondents in the treatment group exhibit inconsistency in their answers across a set of political questions. Finally, we use World Value Survey data to test the potential of using LPA in social desirability bias studies.

Based on the findings, we suggest researchers use LPA in two ways. First, in preference falsification research, use LPA to identify likely "sensitive" questions that capture potential falsifying behavior when they appear. More generally, in public opinion studies, pay attention to these questions and be careful when including them in the analysis. Second, use LPA to learn the proportion of potential falsifiers among survey respondents, which should be treated with caution when interpreting results.

Nonetheless, our approach does not solve the first problem, as it still shares the flawed assumption that inconsistency implies preference falsification. Therefore, we advise against interpreting LPA results at the individual level. If LPA results suggest that a large enough number of respondents shows clear inconsistency across the same set of questions, it can serve as a clue for preference falsification. However, we do not recommend treating every inconsistency in individual responses as preference falsification. Likewise, we do not recommend treating every individual with consistent answers as a truthful respondent.

While not a panacea, we believe LPA is the most objective and effective tool for detecting preference falsification among existing methods. Furthermore, the methodological advances in preference falsification research have thus far focused on survey designs such as list experiments (Blair and Imai 2012; Corstange 2009; Glynn 2013; Robinson and Tannenberg 2019), endorsement experiments (Bullock et al. 2011), and random response techniques (Blair et al. 2015). Despite their effectiveness, these *pre*-survey methodologies require rigorous settings that are often impractical or too costly in authoritarian contexts. The *post*-survey adjustments presented in this paper offer a way to make better inferences even from surveys conducted under less-than-ideal conditions. Moreover, this method allows researchers to reanalyze existing survey data rather than creating new surveys. This capability is especially valuable for those studying the politics of China or Russia, as the research environment in these regions is becoming increasingly hostile due to political tensions.

2 Three Sub-populations under Authoritarianism

Preference falsification is a phenomenon in which people express untruthful opinions in public when they have different, candid opinions in private (Kuran 1987). In democracies, social desirability often compels respondents to choose appropriate, if not truthful, answers to survey questions. In authoritarian contexts, respondents are more likely to give safe answers to politically sensitive questions—the questions political scientists are most interested in—to avoid potentially troublesome or even threatening situations. As such, given the limited protection for freedom of speech, survey data from authoritarian states may be compromised from the start.

In Table 1, we classify three types of survey respondents under authoritarianism based on their level of support for the regime and their susceptibility to politically sensitive questions. The table suggests that the target population of a public opinion survey is a mixture of at least three sub-populations, each with distinct answer patterns. Type 1 respondents are true regime supporters who express their approval of the regime when asked survey questions. We assume they do not need to falsify their answers to sensitive questions. Type 2 respondents are non-supporters who express their true discontent with the regime, either because they are unaware of or do not care about political pressure.

Table 1

Sub-populations of Survey Respondents under Authoritarianism

	Regime Supporters	Non-supporters
Not susceptible to political pressure	Type 1 (No gap)	Type 2 (No gap)
Susceptible to political pressure	–	Type 3[a] (Gap exits)

latent sub-populations under political pressure. It is assumed that true supporters do not need to falsify their true opinions when asked political questions.

[a] Preference falsifiers, the group of interest. A gap between sensitive and non-sensitive questions is expected to be observed from this group.

Type 3 respondents are critical of the regime but hide their true preferences when answering certain questions.

In preference falsification studies, Type 3 respondents are the group of interest. Based on the widely accepted method of measuring preference falsification, these respondents are assumed to render truthful answers to non-sensitive questions. However, when asked politically sensitive questions, they conceal their true opinions by either behaving like true supporters or providing no response. Thus, we observe inconsistent answers across a set of questions only from this sub-population.

Empirically, exaggeration of support and self-censorship are different manifestations of the same cause, preference falsification. As our goal is to capture both phenomena, a tricky question is how to include the non-responses in data analysis, rather than dropping these observations and losing information, so that we can observe inconsistencies when preference falsification occurs. Non-responses may be coded as 0 or extreme numbers like 999. Yet, because LPA classifies subgroups by calculating mean responses, these numbers can seriously distort the results. Therefore, in this paper, we coded non-responses as neutral answers. The mid-point numbers hurt the mean calculations the least. Furthermore, substantively, neutral responses have the closest meaning to non-responses in that both do not express any preferences. Admittedly, this approach is not ideal, as it equates true neutral opinions with non-responses. However, we believe it is the most effective way to identify both forms of preference falsification in data analysis. See the Appendix for further discussions on non-responses.

3 Latent Profile Analysis (LPA)

The subgroups in Table 1 are "latent", meaning they are not directly observable from the variables (i.e., survey questions) being measured. Due to this unobservable nature of sub-populations under authoritarianism, we believe that LPA is an appropriate tool for studying preference falsification.

LPA is a statistical model that divides a group of individuals into multiple unobserved subgroups, where the profiles differ between subgroups but are similar within subgroups (Jason and Glenwick 2016; Bauer 2022; Oberski 2016; Berlin et al. 2014; Sterba 2013). Using LPA, we can determine the number of possible subgroups in the data, their shares, and the associated profiles (e.g., mean and variance of variables, etc.) for each subgroup (Oberski 2016; Jang et al. 2023).

3.1 Illustration on LPA and Preference Falsification

Suppose we conducted a survey with two political questions, Question 1 and Question 2. Respondents were asked to choose from a five-point scale, with higher numbers indicating pro-regime attitudes and lower numbers indicating critical views. After completing the survey, suppose we applied LPA to the respondents' answers, and LPA identified three subgroups with distinct response patterns. Fig. 1 shows a hypothetical scenario based on this analysis. From this, we can infer two things.

First, from Table 1, we know that only the preference-falsifying subgroup would display inconsistent responses. In Fig. 1, Subgroup 3 shows such inconsistency. While Subgroup 1 consistently expresses pro-regime attitudes and Subgroup 2 consistently expresses critical views, Subgroup 3 exhibits critical views in Question 1 and pro-regime attitudes in Question 2. From this observation, we can infer that respondents in Subgroup 3 are the preference-falsifiers, who provide their true opinions in response to certain questions but act as true supporters when answering others. Additionally, we can infer that Subgroup 1 likely represents true supporters, and Subgroup 2 represents candid non-supporters. LPA also allows us to estimate the number of respondents in each subgroup, though we recommend using this information only to infer rough proportions rather than precise numbers.

Second, given the assumption that preference falsifiers tend to hide their true opinions in response to politically sensitive questions, we can infer that Question 2 is likely a sensitive question, while Question 1 is non-sensitive. This second inference provides observation-based evidence that respondents perceived the level of sensitivity differently across the two questions. In this way, LPA addresses the subjectivity problem present in existing literature.

3.2 How LPA Works

Now, we turn to a more technical explanation of how LPA classifies subgroups by analyzing respondents' answers. In this section, following the literature, we temporarily refer to the subgroups as latent "classes". The data, or respondents' answers to survey questions, are assumed to follow a combination of normal distributions, with the (conditional) means and standard deviations of each potential subgroup being class-specific. These within-class distributions are then marginally combined to form the cross-class distribution, also known as a mixed distribution (Oberski 2016).

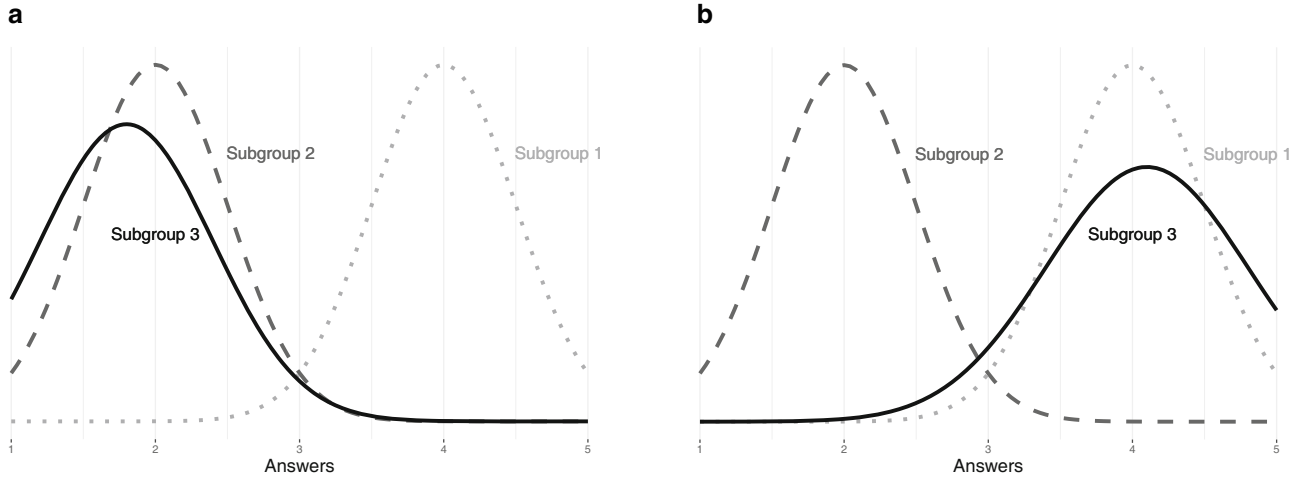

Fig. 1

LPA and Preference Falsification. ***a*** *Distribution of Answers to Question 1,* ***b*** *Distribution of Answers to Question 2. A hypothetical distribution of answer patterns in a survey with two questions. The x-axis represents the answers to the questions, with higher numbers indicating pro-regime attitudes and lower numbers indicating critical views. In this hypothetical analysis, LPA divided the respondents into three subgroups*

More formally, for every observed variable j (survey question) and individual i (respondent) in class k, we have:

$$y_{ij} = \mu_j^{(k)} + \epsilon_{ij}, \tag{1}$$

$$\epsilon_{ij} \sim N(0, \sigma_j^{2(k)}), \tag{2}$$

where $\mu_j^{(k)}$ are the means and $\sigma_j^{2(k)}$ are variances, which may vary across observed variables $j = 1, 2, \ldots, J$ and classes $k = 1, 2, \ldots, K$.

We usually assume that the survey answers within each class follow a normal distribution, but this assumption does not apply to between-class distributions (Bauer 2022; Sterba 2013). The density of each latent class can be expressed as $f^{(k)}(y_i|\mu^{(k)}, \sigma^{2(k)})$, where y_i is the observed outcome for individual i, and $\mu^{(k)}$ and $\sigma^{2(k)}$ are the mean and variance of group k (Sterba 2013). The average of the normal distribution densities for each class, weighted by the probabilities of each class, constitutes the density for each respondent in the mixed distribution (Sterba 2013). The joint mixture model is:

$$f(y_i) = \sum_{k=1}^{K} p(c_i = k) f^{(k)}(y_i|\mu^{(k)}, \sigma^{2(k)}), \tag{3}$$

where c_i is a latent class variable and $p(c_i = k)$ is the probability of being assigned to each class. In the multivariate situation, the model is essentially the same, but y_i represents a vector of results on a set of observed variables. Also, the mean vector of each class is $\mu^{(k)}$, and the corresponding variance-covariance matrix of varying variances and varying covariances model is:

$$\boldsymbol{\Sigma}^{(k)} = \begin{bmatrix} \sigma_1^{2(k)} & & & \\ \sigma_{21}^{(k)} & \sigma_2^{2(k)} & & \\ \vdots & \vdots & \ddots & \\ \sigma_{J1}^{(k)} & \sigma_{J2}^{(k)} & \cdots & \sigma_J^{2(k)} \end{bmatrix}$$

The class probabilities, as well as the means and (co)variances of each class, are unknown before the study. Therefore, they require estimations through the standard expectation-maximization algorithm, which is detailed in the Appendix. Additional restrictions are often imposed to improve model parsimony and estimation stability, as the number of free parameters increases rapidly with more latent classes and indicators y's (Sterba 2013; Oberski 2016). For instance, the local independence assumption posits that the indicators are normally uncorrelated within each latent class, given correct classification. The homogeneity assumption fixes the variances to a single value. These restrictions ensure that the sole difference among the classes lies in their mean profiles, not in the variance-covariance matrix, which simplifies interpretation.

Finally, as Eq. 3 shows, the number of classes does not need to be estimated. Instead, researchers choose the number of groups based on both statistical and substantive factors. The Bayesian Information Criterion (BIC) is the most widely used statistical criterion in this context, and

it is also recommended to report entropy, which is considered acceptable when it exceeds 0.80 (Spurk et al. 2020). Additionally, the existing studies use the Akaike Information Criterion (AIC) to provide further evidence and choose parsimonious models to improve interpretability (Ohlsson et al. 2022; Cowie et al. 2015; Greaves et al. 2015; Dangubić et al. 2021). If two or more models have similar fit statistics, it is recommended that researchers compare them and demonstrate model stability. Once a target model is selected, then the parameters can estimated (Jason & Glenwick 2016; Bauer 2022; Oberski 2016).

3.3 Strengths and Weaknesses of LPA

LPA is not without limitations. Tein et al. (2013) finds that obtaining adequate statistical power to correctly select the number of classes requires large enough distances among latent classes, a sufficient number of indicators, and a large sample size. Bakk et al. (2013) advises that uncertainty should be carefully accounted for when reporting each respondent's class membership. The "naming fallacy" is another well-known issue, where researchers may mislabel class names during interpretation (Weller et al. 2020). Therefore, it is recommended that researchers be fully transparent when addressing these issues.

Nevertheless, for the purpose of detecting preference falsification, LPA is the most appropriate tool compared to other similar methods. Most importantly, LPA is a Finite Mixture Model (FMM), and the main difference between FMM and other clustering algorithms is that FMMs provide a "model-based clustering" approach. This means that clusters are derived using a probabilistic model that describes the distribution of the data, rather than using arbitrarily chosen distance measures as is common in clustering analyses. This enables researchers to assess the probability that certain cases belong to specific latent classes and the goodness of fit. Consequently, if researchers assume that there is an underlying process or latent structure behind the data, FMMs are an appropriate choice, as they allow for modeling the latent structure instead of merely identifying similarities.

Traditional classification models are more suitable for other purposes. For example, Latent Dirichlet Allocation is a generative statistical model for discrete data, often used to analyze text data, but not survey data (Blei et al. 2003; Petterson et al. 2010; Jelodar et al. 2019). Q-factor analysis aims to identify underlying factors, or the ways of thinking, that explain shared variance in participants' ranked responses (Banks & Gregg 1965; Newman & Ramlo 2010; Morf et al. 2023). Item Response Theory is commonly used to develop better tests and assessments by understanding how different test items discriminate between individuals with varying levels of the latent traits (Mason 2017; Edelen and Reeve 2007; Kean and Reilly 1987). Yet, for preference falsification studies, we are interested in discrete latent groups rather than continuous levels of a trait. Finally, Correlation Class Analysis is appropriate for exploring different patterns of how variables relate to one another without considering exact values, whereas LPA assesses relationships among individuals through specific levels, i.e. means and variances (Boutyline 2017; Dekeyser and Roose 2021; Goldberg 2011; Karim 2024). Therefore, LPA is better at capturing the differences in attitudes toward sensitive versus non-sensitive questions, which may imply the degree of preference falsification.

Other FMM-based classification models include Latent Class Analysis (LCA), which is also a type of mixture model for cross-sectional data but is appropriate for categorical data (Oberski 2016; Sterba 2013; Clogg 1995; Collins and Flaherty 2002). Yet, the survey answers we are interested in measuring in this article are fundamentally continuous, even though the respondents choose a limited number of options. Therefore, we believe that LPA is a more suitable tool than LCA.

4 Test of LPA 1. Data Simulation

In the following sections, we demonstrate the application of LPA to studies of preference falsification. This first section applies LPA to simulated survey data, showing how it uncovers three latent groups under authoritarianism and identifies the questions where preference falsifiers are likely to exaggerate their support for the regime.

4.1 Simulated Data

We simulated a public opinion survey with 1000 respondents from an authoritarian state. We assumed that 60% of the respondents were true regime supporters (Type 1 from Table 1), 10% were candid non-supporters (Type 2), and 30% were preference-falsifying sub-populations (Type 3). Each respondent was then randomly assigned to one of these sub-populations.

The survey consisted of six political questions. The first two were assumed to be highly sensitive, while the remaining four were less sensitive. Respondents chose answers on a 5-point scale, where 5 indicated "Strongly agree with the regime" and 1 indicated "Strongly disagree with the regime." Type 1 respondents were more likely to give pro-regime answers to all six questions, while Type 2 respondents were more likely to express disagreement. However, Type 3 respondents tended to falsify their preferences when answering sensitive questions, mimicking the patterns of

Type 1 respondents for the first two questions, but behaving like Type 2 respondents for the remaining four.

Table 2 summarizes the distribution of respondents' answers in the simulation. Note that sometimes, true supporters were allowed to express negative opinions toward the regime, and non-supporters were allowed to provide positive answers. This reflects the potential genuine inconsistency of respondents' attitudes across different questions, as well as the possibility of non-attentive respondents selecting random answers.

4.2 LPA on Simulated Data

We applied LPA to the simulated data using the R package, TidyLPA. First, we determined the number of latent groups, or k. As explained earlier, selecting k in any clustering method fundamentally involves some degree of subjective judgment. In this paper, based on the suggestions and conventional approach in the previous research, we adopt the following procedure: choose ks with entropy levels above 0.80 and the lowest AICs and BICs. Then, select the lowest k among them for interpretability. If two or more models meet the entropy, AIC, BIC, and interpretability criteria similarly, we compare those models and test their stability. This standard is applied to all LPAs conducted in this paper.[1]

In this section, based on these criteria, we selected six latent groups ($k = 6$) and conducted LPA on the simulated data (Figure A.3.). For simplicity, Fig. 2 presents the results with four latent groups, averaging the answers of similar subgroups for better reliability. Further details on model selection and LPA results are available in the Appendix.

Suppose we are interpreting Fig. 2 without knowing how the data were generated. Then, we can first observe that 553 respondents in Subgroup 1 consistently provided pro-regime answers across all six questions. On the contrary, 90 respondents in Subgroup 3 consistently expressed their disagreement with the regime. From this, we can infer that Subgroup 1 likely consists of true regime supporters (Type 1), while Subgroup 3 consists of candid non-supporters (Type 2), based on the theoretical expectations presented in Table 1.

Next, we observe that 308 respondents in Subgroup 2 exhibit *inconsistent* attitudes across the set of political questions. These individuals act like true supporters for Questions 1 and 2 but behave like candid non-supporters for Questions 3 through 6. Here, based on the theory in Table 1, we can infer two things. First, these respondents are likely the preference falsifiers (Type 3) we aim to identify. Second, they are probably exaggerating their regime support when answering Questions 1 and 2 but expressing candid attitudes when answering Questions 3 through 6. Thus, the first two questions are likely more sensitive, inducing preference falsification among Type 3 respondents, while the remaining questions are less sensitive, allowing for more honest responses.

These inferences are consistent with the way we generated the data. First, the first two questions were designed to be sensitive, while the remaining four were non-sensitive. Second, we assigned 301 respondents as preference falsifiers, a number that closely matches the 308 respondents in Subgroup 2 in Fig. 2. The number of candid non-supporters we created was 92, and it is also close to the number of respondents that LPA identified as Subgroup 3. We generated 602 true supporters through simulation, and 553 of them were captured in Subgroup 1. LPA further categorized 49 respondents (Subgroup 4) who were meant to be true supporters into a separate group, likely because we

Table 2

Distribution of Simulated Answers, Range of Percentages (%)

A. True Supporters (Type 1) 607 respondents		
Answers	Q. 1–2	Q. 3–6
5	60.8–62.6	58.6–63.8
4	29.0–31.1	28.3–32.6
3	5.9–6.3	5.4–6.9
2	1.8–2.1	1.8–2.5
1	0	0
B. Candid Non-supporters (Type 2) 92 respondents		
Answers	Q. 1–2	Q. 3–6
5	0	0
4	1.1	0–1.1
3	7.6–9.8	4.4–8.7
2	26.1–30.4	22.8–32.6
1	60.9–63.0	62.0–68.5
C. Preference Falsifiers (Type 3) 301 respondents		
Answers	Q. 1–2	Q. 3–6
5	59.1–61.1	0
4	28.9–31.9	2.0–3.3
3	8.0	6.0–10.3
2	1.0–2.0	26.9–29.9
1	0	58.8–62.1

The table summarizes the distribution of answers from three different sub-populations. Questions 1–2 are highly sensitive political questions, and Questions 3–6 are non-sensitive political questions. Answer 5 is "Strongly agree with the regime," and 1 is "Strongly disagree with the regime."

[1] In practice, we did not encounter instances where multiple models had similar fits and interpretability.

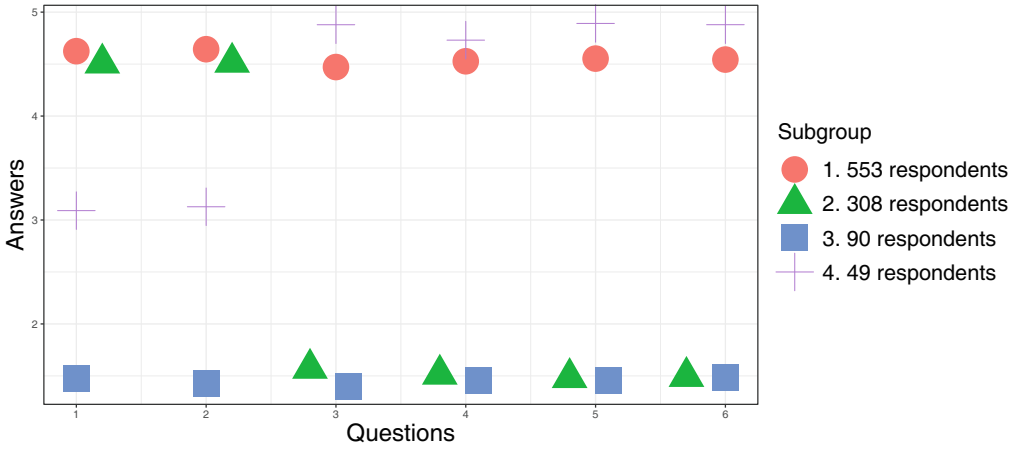

Fig. 2

Simplified LPA Results from Simulated Data. Questions 1 and 2 are sensitive political questions, and Questions 3–6 are non-sensitive political questions. Answer 5 is "Strongly agree with the regime", and 1 is "Strongly disagree with the regime"

let these true supporters give neutral answers (Answer 3) to the sensitive questions. If we combine the respondents in Subgroups 1 and 4, totaling 602, this matches the number of true supporters generated in the simulation. Overall, LPA correctly classified 955 respondents' true political identities (96% accuracy).

We further examined the reasons why LPA did not accurately classify the remaining 45 respondents and found several important insights: First and foremost, inconsistent opinions across questions do not always indicate preference falsification. For example, a true regime supporter who generally likes the regime might express a critical view on a specific political issue. Second, a falsifier might have a genuinely favorable view on a certain non-sensitive issue or falsify their attitude in certain non-sensitive questions but not others. Then, the LPA may incorrectly classify them as either a true supporter or a candid non-supporter. Finally, if a respondent lacks strong political opinions, their group identity classification may be random. Therefore, we encourage researchers to report such possibilities and be conservative in interpreting the LPA results.[2]

5 Test of LPA 2. Public Opinion Survey from China

This section applies LPA to actual survey data from China. While existing studies across various authoritarian regimes have found mixed evidence on preference falsification (Nalepa and Pop-Eleches 2023; Frye et al. 2017; Wedeen 2015), research conducted in China shows that Chinese survey respondents do tend to falsify their preferences when asked sensitive questions (Shen and Truex 2021; Jiang and Yang 2016; Robinson and Tannenberg 2019). In the previous section, we demonstrated the use of LPA in an ideal scenario where the number of respondents was large enough and the three subgroups had clearly distinct answer patterns. In this section, however, the number of respondents is smaller than ideal, and the distinctions among the three subgroups are less clear.

5.1 Quasi-experimental setting of Chinese General Social Survey 2006

While the Chinese General Social Survey (CGSS) was being conducted in 2006, then Shanghai's Chinese Communist Party (hereafter the Party) chief, Chen Liangyu, was charged with corruption and dismissed. Although the state media reported that Chen violated the law and was punished accordingly, foreign media interpreted the dismissal as political, suggesting that Chen was politically purged for resisting the new leader, Hu Jintao, while remaining loyal to the former leader, Jiang Zemin. This event attracted the attention of Shanghai residents, as evidenced by a surge in online searches for related keywords around that time (Jiang and Yang 2016). This controversial purge likely heightened preference falsification among Shanghai respondents by reminding them of the limited political freedoms in the public

[2] The probability of an individual being classified into a particular latent group, $p(c_i = k)$, does not necessarily help identify misclassified individuals. Among the 45 misclassified individuals, 37 have the probabilities above 0.99.

sphere.[3] Indeed, Jiang and Yang (2016) finds that, on average, Shanghai respondents who participated in the survey *after* the dismissal tended to demonstrate a gap between their answers to questions that directly ask about their support for the government and those that indirectly assess their approval.

If the purge indeed worked as an external shock, then survey participants *after* the purge (treatment group) would have behaved differently from those who took the survey *before* the purge (control group) when responding to questions related to the event. Specifically, after the purge, true supporters would continue to express their pro-regime attitudes, and non-supporters who are not susceptible to political pressure would continue to express their true disapproval of the regime. However, non-supporters susceptible to political pressure would falsify their support when asked sensitive questions, while revealing their true disapproval when responding to non-sensitive questions.

Indeed, the data suggests that the purge altered how respondents answered questions related to the event but did not affect their responses to unrelated questions. Fig. 3 shows permutation tests on the difference between Shanghai respondents' answers before and after the purge. The red vertical lines indicate the true mean differences in responses between the treatment and control groups, and the histograms show the random distribution of the differences obtained by shuffling the responses 9999 times.

The first plot shows that survey respondents after the purge were significantly more likely to agree that the government and courts maintain a similar stance in major cases. This aligns with expectations, as those who observed the Chen Liangyu case would likely agree with such statements. The second plot shows that in response to the question about whether it is always correct to follow the government, 67% of respondents in the control group and 83% in the treatment group selected either "strongly agree (4)" or "Agree (3)". On a 4-point scale, the average difference between the treatment and control groups is 0.2443, after dropping non-responses. The permutation test suggests that in only about nine out of 10,000 cases, the mean difference exceeds the true difference. It suggests that after witnessing the purge, Shanghai residents were more likely to exaggerate their support levels. The last plot shows no significant difference between respondents' answers before and after the purge to a question about development, which was unrelated to the event.

[3] These preference-falsifying respondents are less likely to fear that answering survey questions in a certain way would lead to immediate punishment by political authorities (Chen and Shi 2001; Shi 2001).

5.2 Hypothesis

Suppose that, in the absence of political pressure, all respondents in the control group provided truthful answers to survey questions. That is, true regime supporters consistently rendered pro-regime responses, while non-supporters consistently gave anti-regime responses. After the purge, which served as a treatment, some non-supporters who are susceptible to political pressure may have decided to falsify their preferences. Under this assumption, *only* within the treatment group, we would expect to observe certain respondents that display *inconsistent* attitudes across similar questions. This suggests that they likely belong to the group of preference-falsifying non-supporters.

5.3 LPA results

We conducted LPA on the CGSS 2006 data, finding four latent subgroups in the control group and six in the treatment group. While the complete results are available in the Appendix (Figures A.5 and A.6), here, we have simplified these subgroups into three categories: true supporters, candid non-supporters, and preference-falsifying subgroups, for better readability. Fig. 4 presents these simplified results. Figure 4a shows the answer patterns of the control group and Fig. 4b shows the answer patterns of the treatment group. The x-axis represents the survey questions used in the analysis, and the y-axis shows the respondents' answers to each question. Higher values indicate more pro-regime opinions, while lower values indicate more anti-regime opinions. Note that for preference falsification studies, all questions in the analysis should be political questions related to the purge but with different wordings.

Proportion of Each Latent Subroup The LPA analysis of the CGSS 2006 data revealed the expected latent subgroups in both the control and treatment groups. In both groups, we identified subgroups of respondents labeled as true regime supporters. True supporters in the control group (Subgroup 1 in Fig. 4a) and the treatment group (Subgroup 1 in Fig. 4b) exhibit similar answer patterns, consistently expressing pro-regime attitudes across most questions. In the control group, true supporters constitute 75 of the 119 respondents (63%), and in the treatment group, they account for 183 of the 281 respondents (65%). This suggests that approximately 65% of Shanghai citizens can be considered true supporters.

In the control group, there appear to be 44 non-supporters (Subgroup 2 in Fig. 4a), who consistently express similar or more negative opinions toward the regime compared to true supporters. In the treatment group, 16 respondents (Subgroup 3 in Fig. 4b) display similar answer patterns to

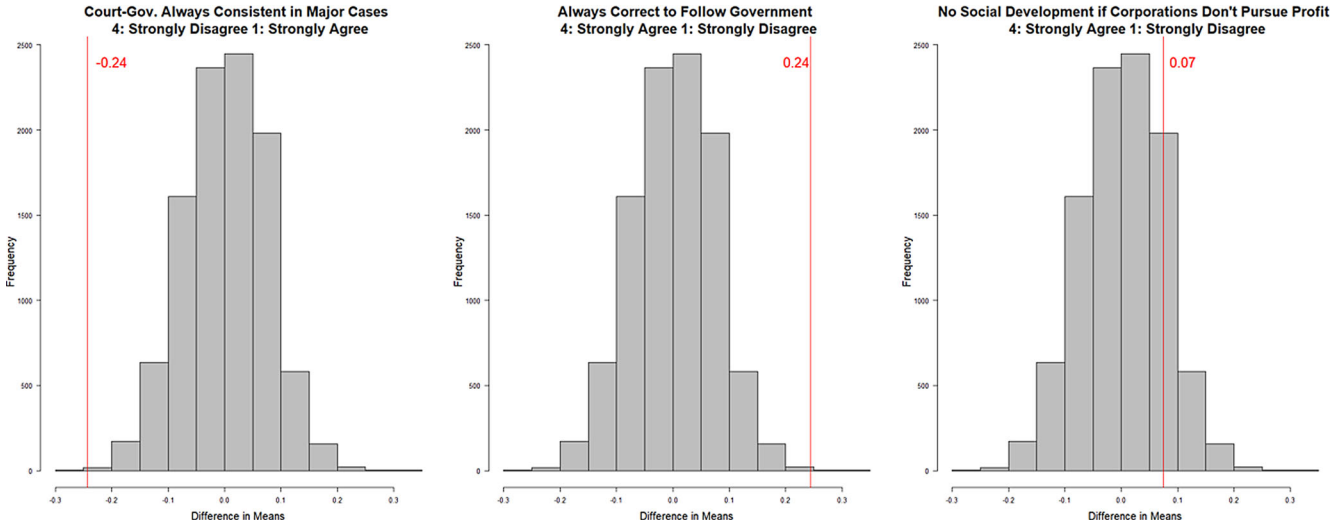

Fig. 3

Purge as Treatment (Permutation Test). Each question asks, "How much do you agree with the following statement?". The first statement is, "The court and government always agree on major cases." The second is, "It is always correct to follow the government." The last is, "If corporations do not pursue profit, there will be no social development." The red vertical lines indicate the true mean differences in responses between the treatment and control groups, and the histograms represent the random distribution of differences obtained by shuffling the responses 9999 times.

these non-supporters. Therefore, it is likely that these respondents are non-supporters who are not susceptible to political pressure—candid non-supporters.[4]

Preference Falsification and Sensitivity of Questions If our hypothesis is correct, then among the remaining 29% of respondents in the treatment group, we should find evidence of preference falsification. Indeed, there is a third subgroup that is *only* observable in the treatment group, which shows *inconsistent* responses across similar political questions (Subgroup 2 in Fig. 4b).[5]

Among falsifiers in Subgroup 2 in the treatment group, we identified three distinct answer patterns (see Figure A.6 for details). First, 7% of falsifiers (Class 6 of Figure A.6) mimicked true supporters in Question 1 by agreeing that it is always correct to follow the government, yet they expressed distrust in government agencies' information on corruption issues (Question 5). Their skepticism is similar to that of the non-supporters in the control group (especially Subgroup 2 in Figure A.5), with their distrust intensifying after observing Chen Liangyu's corruption scandal. These respondents also stated that there is a serious conflict between cadres and people in China (Question 8). However, this question may not effectively distinguish true supporters from non-supporters because, in the control group, supporters and non-supporters provided similar answers to this question. This is likely because the question is observational (e.g., "Do you *see* a conflict?") rather than evaluative (e.g., "Do you *have* a conflict?"), allowing both supporters and non-supporters to respond similarly without revealing their subjective assessment of the situation.

Second, 13% of the falsifiers (Class 5 of Figure A.6) tended to agree with, or gave no answers to, the question that asks whether it is always correct to follow the government. However, they behaved similarly to non-supporters when asked whether law enforcement needs government support (Question 3) and expressed further disagreement with the idea that the court should follow the government decisions when the two differ (Question 4). This suggests that these respondents might be non-supporters who falsified their responses to the first question but provided more honest answers to the latter two questions.

Finally, 10% of falsifiers (Class 2 of Figure A.6) again tended to agree with, or gave no answers to, the question that asks whether it is always correct to follow the government. However, they disagreed that law enforcement needs

[4] It appears that the candid non-supporters in the treatment group remain truthful, not because they are indifferent to political pressure, but because they are unaware of the purge. This is because these respondents did not update their understanding of the relationship between the court and the government when responding to Questions 3 and 4. In contrast, preference falsifiers adjusted their views on court-government relations, showing even stronger anti-regime attitudes in response to Questions 3 and 4.

[5] It is possible that some falsifiers existed even before the purge. However, their proportion may have been too small for LPA to detect.

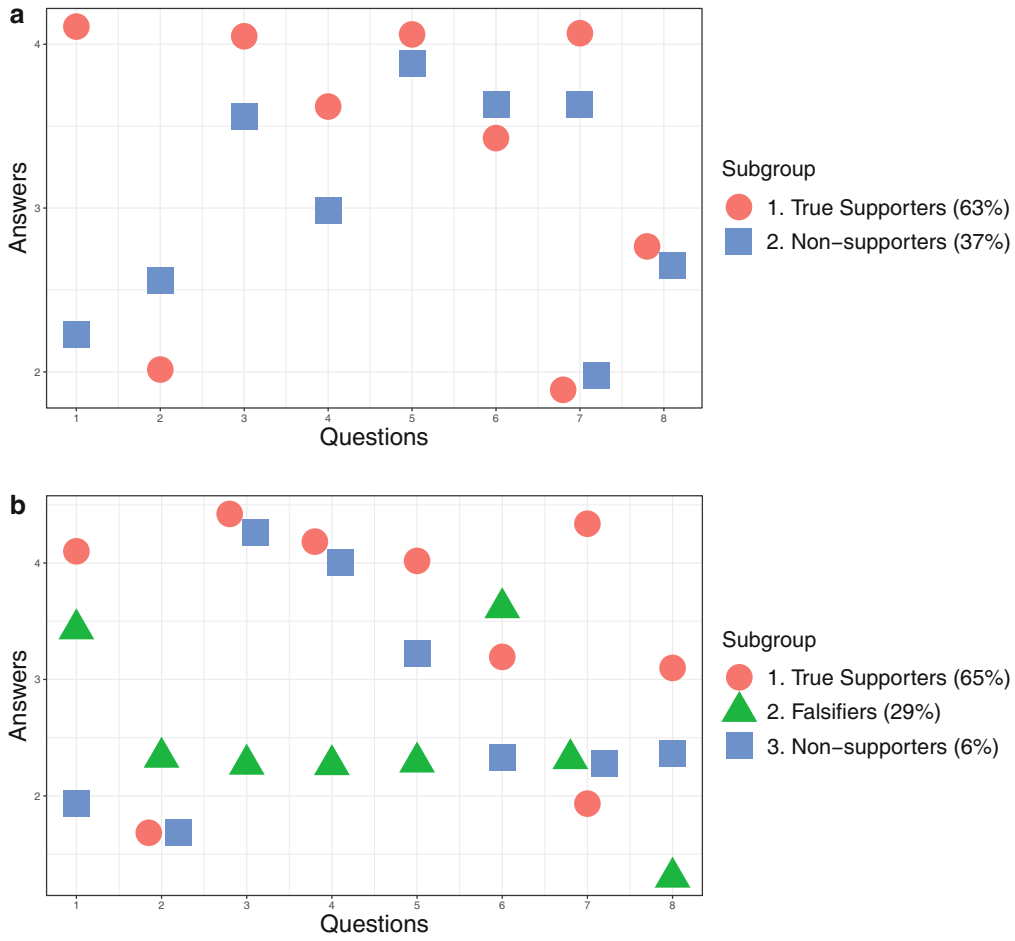

Fig. 4

Simplified LPA Results on CSCC 2006. ***a*** *Control Group (119 Respondents),* ***b*** *Treatment Group (281 Respondents). * Q 1. CORRECT TO FOLLOW GOV. It is always correct to follow the government (5: Strongly agree—1: Strongly disagree), Q 2. COURT-GOV CONSISTENT The court and government always agree in major cases (5: Strongly disagree—1: Strongly agree), Q 3. LAW ENFORCEMENT NEEDS GOV. SUPPORT Law enforcement functions better with strong government support (5: Strongly agree—1: Strongly disagree), Q 4. COURT SHOULD FOLLOW GOV. The court should follow the government's decision if they disagree (5: Strongly agree—1: Strongly disagree), Q 5. TRUST GOV.'S INFO. ON CORRUPTION Do you trust government agencies' information on corruption issues? (5: Strongly trust—1: Don't trust at all), Q 6. LIFE SATISFACTION Overall, how do you feel about your life? (5: Very happy—1: Very unhappy), Q 7. ECONOMY VS. DEMOCRACY As long as the economy keeps growing, there is no need for democracy (5: Strongly agree—1: Strongly disagree), Q 8. CADRE VS. PEOPLE CONFLICT Is there a serious conflict between cadres and people in China? (5: Not at all—1: Very serious)*

government support (Question 3), tended not to believe that the court should follow the government (Question 4), and indicated that there is a serious conflict between party elites and the people (Question 8).

Overall, these results suggest that about 29% of respondents in the treatment group may have falsified their preferences due to political pressure, with this falsification most likely occurring in responses to Question 1. Therefore, we can infer that such questions directly assessing government support are more likely to elicit preference falsification among some Chinese respondents. Their genuine political attitudes are more likely to be reflected in their responses

to the less sensitive questions, Questions 3, 4, and 5. A researcher may choose whether or not to include Question 8 when measuring respondents' true political attitudes.

As such, LPA is effective at revealing the proportion of potential falsifiers and identifying questions that may prompt falsifying behavior. Despite the strengths of an LPA-based approach, there are a few considerations that researchers should keep in mind when interpreting the results. First, we found that around 65% of Shanghai respondents consistently expressed pro-regime attitudes across different questions. This could indicate that they are genuine supporters of the regime, or it could mean that some of these respondents are perfectly concealing their true preferences, regardless of a question's sensitivity. However, LPA alone cannot distinguish between these possibilities. Second, in the preference falsification studies, the underlying assumption is that inconsistency in political attitudes suggests preference falsification. However, depending on specific survey questions and context, inconsistent responses do not necessarily indicate untruthful opinions. Instead, such inconsistencies might reflect one's genuinely different assessments on certain topics or could result from random errors, such as lack of attention or difficulty in understanding the questions. Therefore, even when researchers observe signs of preference falsification, they should be cautious when making definitive claims about its presence.

6 Test of LPA 3. Social Desirability Bias in the U.S.

Having demonstrated how LPA can detect preference falsification in Chinese public opinion surveys, it is natural to ask whether LPA could be similarly effective for studying other authoritarian countries or even democracies. Our general answer to this question is that LPA provides more reliable clues for preference falsification when, first, the magnitude of falsification is indeed strong, and second, there exists a subpopulation with a significant number of respondents who exhibit clear inconsistencies in their answers.

While such a clear tendency for preference falsification is often expected in authoritarian states,[6] it remains uncertain whether LPA can uncover a similar phenomenon—namely, social desirability bias—in democracies. Therefore, in this section, we use data from the World Values Survey conducted in the U.S. in 2017 to test whether LPA can detect signs of social desirability bias.

[6] In the Appendix, we further discuss the application of LPA to other authoritarian contexts.

6.1 LPA on the 2017 U.S. World Values Survey

This section examines whether social desirability bias is detectable in the U.S. public opinion survey conducted in April–May 2017, approximately six months after the 2016 presidential election. Following the election, there were heated discussions about why and how polling failed to predict Donald Trump's victory. Although the evidence is mixed, one hypothesis proposed is the so-called "shy Trump voter" phenomenon, where his supporters concealed their preferences due to social desirability bias (Mercer et al. 2016; Coppock 2017). Therefore, this section applies LPA to U.S. respondents' answers to political questions to determine whether inconsistent answers to similar questions can be observed among certain respondents.

Since we already know the political stances of each party, we first tested whether LPA could accurately capture the dominant policy preferences among Republicans and Democrats across various issue areas. To do this, we limited the analysis to 1941 respondents who indicated they would vote Republican or Democrat if there were a national election tomorrow. Among them, 816 respondents said they would vote Republican (hereafter, "Republicans"), and the remaining 1125 said they would vote Democrat (hereafter, "Democrats"). We treated party identity as an unobserved variable, excluding this survey question from the LPA. Other questions were included in the analysis, such as ideology (conservative vs. liberal), religious beliefs, and views on immigration, homosexuality, and abortion. Finally, we constrained LPA to identify only two latent subgroups.

As shown in Fig. 5a, the LPA results largely align with known patterns in American politics. Subgroup 1, likely Republicans, consistently express more conservative views than Subgroup 2, likely Democrats, across all eight questions. For example, respondents in Subgroup 1 exhibit greater confidence in the government (Trump was the president at the time of the survey), slightly more negative views on immigrants and homosexuals, and are more opposed to abortion than those in Subgroup 2. The distribution of respondents across these subgroups closely matches the actual numbers of Republicans and Democrats in the dataset.

Running LPA based on model fit criteria provides additional insights, as shown in Fig. 5b, a simplified version of Figure A.12 with four subgroups. Subgroup 1 (475 respondents) reflects Republican views, while Subgroup 3 (378 respondents) aligns with Democratic views, as described above. Subgroup 2 respondents differ slightly from Subgroup 1 in that they are less religious and hold more moderate views on homosexuality and abortion. Subgroup 4 respondents differ slightly from Subgroup 3 in that they are more religious. This analysis reveals that attitudes to-

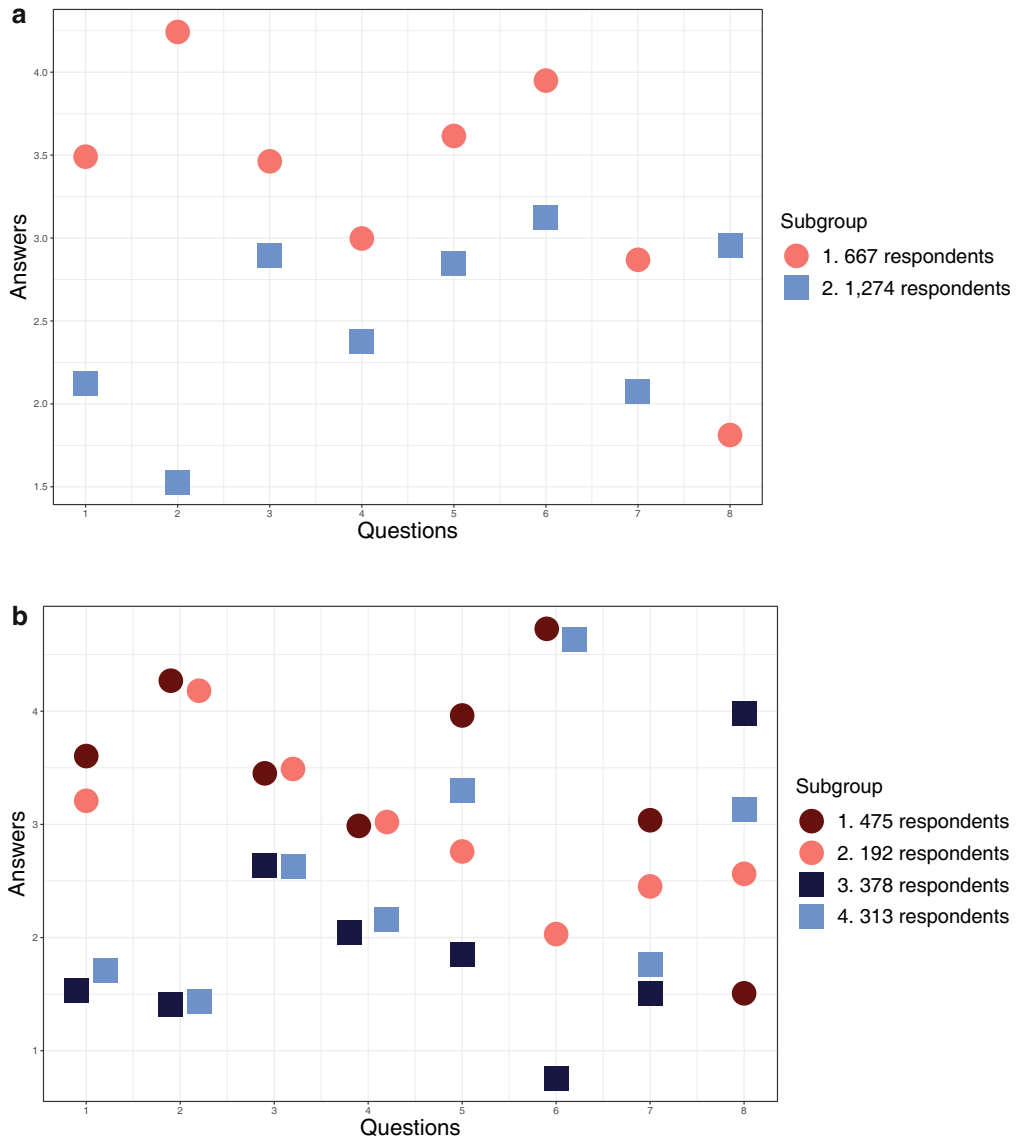

Fig. 5

LPA on the 2017 U.S. World Values Survey. ***a*** *Forced to have Two Subgroups,* ***b*** *Simplified LPA Results Based on Model Fit. * Q 1. LR Position on the left-right political scale (Answers 5: Right—1: Left), Q 2. CONF. GOV How much confidence do you have in the government? (5: A great deal—1: Not at all), Q 3. JOB OVER IMMIGRANT Employers should give priority to Americans over immigrants (5: Strongly agree—1: Strongly disagree), Q 4. IMMIGRANT How do you assess the impact of immigrants on the development of the country? (5: Very bad—1: Very good), Q 5. CONF. CHURCH How much confidence do you have in the Church? (5: A great deal—1: Not at all), Q 6. GOD How important is God in your life? (5: Very important—1: Not at all), Q 7. HOMOSEXUAL Homosexual couples are as good parents as other couples (5: Strongly disagree—1: Strongly agree), Q 8. ABORTION Is abortion justifiable? (1: Never—5: Always)*

ward homosexuality and abortion are closely linked to both political orientation and religiosity.

If social desirability bias existed, we would observe convergence in answers between Republicans and Democrats to certain potentially sensitive question(s). Furthermore, in certain subgroups, the direction of convergence would contradict their responses to other similar questions. In other words, we should observe inconsistent attitudes.

While Republicans and Democrats exhibit contradictory views in most cases, the distance between their views becomes close in Question 4, which asks whether the impact of immigrants is positive or negative. For this question, as many as 46% of Republicans showed no clear preference and chose the neutral answer, 3. This percentage of neutral responses is the highest among all eight questions, with Republicans, on average, providing 17% neutral responses. Similarly, the percentage of neutral answers from Democrats is also high (32%) for this question compared to their average (12%). Nevertheless, Democrats still display an overall tendency to believe that immigrants have a positive impact on the development of the U.S.

The distance between Republican and Democrat attitudes is similarly narrow in Question 3—Republicans show moderate agreement, and Democrats show moderate disagreement with the statement that employers should give priority to Americans over immigrants. Upon closer examination of the data, not a single respondent provided extreme answers, either 1 or 5, to this question. Among Republicans, nearly 20% gave neutral answers, and 64% agreed with the statement. Additionally, less than half of Democrats disagreed with the statement, while 28% agreed with it.

From these LPA-based observations, we can infer a few things. First, both Republicans' and Democrats' responses converge in Questions 3 and 4. This suggests the possibility of the existence of a socially desirable attitude towards immigrants. Second, both Republicans and Democrats tend to provide more moderate views on these questions compared to others. If we exclude these two questions, the average neutral responses change from 17% (Republicans) and 12% (Democrats) to 9% and 9%, respectively. Since providing neutral answers can be a way to hide one's true preferences, this might hint at the existence of social desirability bias. However, researchers should exercise caution in making such claims, as we cannot distinguish true neutral or no opinions from falsified answers merely by observing reported responses, especially in democracies where freedom of speech is well-protected.

Third, Republicans demonstrate slight inconsistency between their answers to Questions 3 and 4. In Question 4, nearly equal numbers of respondents said the impact of immigrants is good (22%) and bad (23%), with almost half providing neutral answers. These attitudes differ from those shown in Question 3, where they tended to agree with the statement, even though many gave neutral answers. Of course, to some extent, this inconsistency may reflect respondents' genuinely diverging attitudes toward specific versus general views on immigrants. Nevertheless, it is also plausible that Americans feel more compelled to say immigrants are good while feeling more comfortable asserting that Americans deserve better treatment in actual job competition. Furthermore, even if their opinions between Questions 3 and 4 are not entirely contradictory, the 46% neutral answers in Question 4 is an ostensibly high number that raises suspicions about the existence of social pressure.

Democrats also tend to provide neutral answers to Question 4, but their responses are largely consistent with the attitudes reflected in other questions. In addition, while Democrats expressed some negative and thus inconsistent attitudes toward immigrants in Question 3, these negative attitudes likely reflect genuine opinions, as they go against the expected direction of social desirability bias, assuming it exists.

In conclusion, by applying LPA to survey results, we observed both convergence and inconsistency in the U.S. public opinion survey, particularly among Republicans on the immigration issue. Along with the high proportion of neutral answers, these findings suggest that researchers may want to avoid taking such answers at face value and should exercise caution when analyzing survey results. While LPA helped identify questions and subgroups that need caution, the interpretation of these patterns may vary across different regime contexts. Given that respondents are free to express their opinions in a democracy, these findings may not provide as strong evidence of social desirability bias as they might under authoritarianism. Indeed, the magnitude of inconsistency was not as large as in the China example.

7 Conclusion

In this paper, we demonstrated how LPA can be used to detect preference falsification in public opinion surveys under authoritarianism. We first introduced a conceptual framework where we categorized survey respondents under political pressure into true supporters, candid non-supporters, and preference-falsifiers. Then, through the application of LPA on simulated data and a survey conducted in China, we showed that LPA can help identify (1) the proportions of the three latent subpopulations, (2) the questions where preference falsification is likely to be induced among non-supporters who are susceptible to political pressure, and (3) the questions that better reflect these respondents true political opinions. We further tested if LPA can be useful in identifying social desirability bias in a democracy, using the survey conducted in the U.S. in 2017.

In the existing literature, researchers selected the survey questions that preference falsification is likely to have happened based on their *prior* knowledge in the field. We suggest that in addition to such field knowledge, researchers need to further justify their identification of preference falsification by providing *observation-based* evidence. This paper provided that LPA, a tool that analyzes answer patterns of latent subgroups in a survey, is one method that researchers can use to show that certain respondents actually rendered inconsistent answers across similar political questions with different levels of sensitivity.

Nevertheless, note that preference falsification studies begin with a *strong* assumption that inconsistency in political attitudes hints at falsified answers. Again, this paper does not attempt to solve this fundamental issue in the current practice. In addition, LPA can be proven more useful to identify preference falsification under one circumstance than others. LPA would provide more reliable evidence for preference falsification when a significant number of respondents show a clear inconsistency across similar questions. This means that when the inconsistency is weak in magnitude and only observable within a limited number of respondents, then researchers may be less confident in claiming preference falsification, or LPA may fail to capture that subgroup.

A few pieces of practical advice are worth mentioning before closing. First, for the purpose of detecting preference falsification, LPA should include similar political questions with different wordings (that is, with different expected levels of sensitivity). This enables researchers to compare answers across questions. Second, refrain from including a question with a binary answer because it tends to restrict the number of latent subgroups to two. Third, when choosing the number of latent subgroups, or k, think about the trade-off between model fits and simplicity and consider balancing between the two. Specifically, the LPA plots become difficult to interpret when the number of latent subgroups becomes bigger than six.

References

Bakk, Z., Tekle, F.B., & Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, *43*(1), 272–311.

Banks, A.S., & Gregg, P.M. (1965). Grouping political systems: Q-factor analysis of a cross-polity survey. *American Behavioral Scientist*, *9*(3), 3–6.

Bauer, J. (2022). *A primer to latent profile and latent class analysis. In Methods for researching professional learning and development: Challenges, applications and empirical illustrations* (pp. 243–268). Springer.

Berlin, K.S., Williams, N.A., & Parra, G.R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, *39*(2), 174–187.

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, *20*(1), 47–77.

Blair, G., Imai, K., & Zhou, Y.-Y. (2015). Design and analysis of the randomized response technique. *Journal of the American Statistical Association*, *110*(511), 1304–1319.

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Boutyline, A. (2017). Improving the measurement of shared cultural schemas with correlational class analysis: Theory and method. *Sociological Science*, *4*(15), 353–393.

Bullock, W., Imai, K., & Shapiro, J.N. (2011). Statistical analysis of endorsement experiments: Measuring support for militant groups in pakistan. *Political Analysis*, *19*(4), 363–384.

Chen, X., & Shi, T. (2001). Media effects on political confidence and trust in the people's republic of China in the post-Tiananmen period. *East Asia*, *19*, 84–118.

Clogg, C.C. (1995). *Latent class models. Handbook of Statistical Modeling for the Social and Behavioral Sciences* (pp. 311–359).

Collins, L.M., & Flaherty, B.P. (2002). *Latent class models for longitudinal data. In Applied latent class analysis*. Vol. 28 (pp. 7–303).

Coppock, A. (2017). Did shy Trump supporters bias the 2016 polls? evidence from a nationally-representative list experiment. *Statistics, Politics and Policy*, *8*(1), 29–40.

Corstange, D. (2009). Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis*, *17*(1), 45–63.

Cowie, L.J., Greaves, L.M., & Sibley, C.G. (2015). Identifying distinct subgroups of green voters: A latent profile analysis of crux values relating to green party support. *New Zealand Journal of Psychology*, *44*(1), 45–59.

Dangubić, M., Verkuyten, M., & Stark, T.H. (2021). Understanding (in)tolerance of muslim minority practices: A latent profile analysis. *Journal of Ethnic and Migration Studies*, *47*(7), 1517–1538.

Dekeyser, D., & Roose, H. (2021). Unpacking populism: Using correlational class analysis to understand how people interrelate populist, pluralist, and eli-

tist attitudes. *Swiss Political Science Review*, *27*(2), 476–495.

Edelen, M.O., & Reeve, B.B. (2007). Applying item response theory (irt) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*, 5–18.

Frye, T., Gehlbach, S., Marquardt, K.L., & Reuter, O.J. (2017). Is Putin's popularity real? *Post-Soviet Affairs*, *33*(1), 1–15.

Glynn, A.N. (2013). What can we learn with statistical truth serum? design and analysis of the list experiment. *Public Opinion Quarterly*, *77*(S1), 159–172.

Goldberg, A. (2011). Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. *American Journal of Sociology*, *116*(5), 1397–1436.

Greaves, L.M., Osborne, D., & Sibley, C.G. (2015). Profiling the fence-sitters in new zealand elections: A latent profile model of political voting blocs. *New Zealand Journal of Psychology*, *44*(2), 43.

Jang, G., Schwarzenthal, M., & Juang, L.P. (2023). Adolescents' global competence: A latent profile analysis and exploration of student-, parent-, and schoolrelated predictors of profile membership. *International Journal of Intercultural Relations*, *92*, 101729.

Jason, L., & Glenwick, D. (2016). *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods*. Oxford University Press.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*, *78*, 15169–15211.

Jiang, J., & Yang, D.L. (2016). Lying or believing? measuring preference falsification from a political purge in China. *Comparative Political Studies*, *49*(5), 600–634.

Karim, S. (2024). The organization of ethnocultural attachments among second-generation Germans. *Social Science Research*, *118*, 102959.

Kean, J. and J. Reilly (2014). Item response theory. Handbook for Clinical Research: Design, Statistics and Implementation. 195-198.

Kean, J., & Reilly, J. (2014). *Item response theory. Handbook for Clinical Research: Design, Statistics and Implementation* (pp. 195–198).

Kuran, T. (1987). Preference falsification, policy continuity and collective conservatism. *The Economic Journal*, *97*(387), 642–665.

Mason, D.P. (2017). Measuring latent constructs in nonprofit surveys with item response theory: The example of political ideology. *Nonprofit Policy Forum*, *8*(1), 91–110.

Mercer, A., Deane, C., & McGeeney, K. (2016). Why 2016 election polls missed their mark. Pew Research Center. https://www.pewresearch.org/short-reads/2016/11/09/why-2016-election-polls-missed-their-mark/

Morf, M. E., C. M. Miller, and J. M. Syrotuik (1976). A comparison of cluster analysis and q-factor analysis. Journal of Clinical Psychology

Nalepa, M., & Pop-Eleches, G. (2023). *Incumbent and opposition support in authoritarian regimes: Survey evidence from late-communist poland*. Working Paper

Newman, I., & Ramlo, S. (2010). *Using q methodology and q factor analysis in mixed methods research*. Sage Handbook of Mixed Methods in Social and Behavioral Research, Vol. 2 (pp. 505–530).

Oberski, D. (2016). *Mixture models: Latent profile and latent class analysis. Modern Statistical Methods for HCI* (pp. 275–287).

Ohlsson, A., Lindfors, P., Larsson, G., & Sverke, M. (2022). Political skill in higher military staff: Measurement properties and latent profile analysis. *Scandinavian Journal of Psychology*, *63*(2), 144–154.

Petterson, J., Buntine, W., Narayanamurthy, S., Caetano, T., & Smola, A. (2010). Word features for latent dirichlet allocation. *Advances in Neural Information Processing Systems*, *23*., .

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

Shen, X., & Truex, R. (2021). In search of self-censorship. *British Journal of Political Science*, *51*(4), 1672–1684.

Shi, T. (2001). Cultural values and political trust: A comparison of the People's Republic of China and Taiwan. *Comparative Politics*, *33*(4), 401–419.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445.

Sterba, S.K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, *48*(6), 775–815.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: a Multidisciplinary Journal*, *20*(4), 640–657.

Wedeen, L. (2015). *Ambiguities of domination: Politics, rhetoric, and symbols in contemporary Syria*. University of Chicago Press.

Weller, B.E., Bowen, N.K., & Faubert, S.J. (2020). Latent class analysis: A guide to best practice. *Journal of Black Psychology*, *46*(4), 287–311.