Political Science Research and Methods (2024), page 1 of 19
doi:10.1017/psrm.2024.29

ORIGINAL ARTICLE

On the measurement of preference falsification using nonresponse rates

Ammar Shamaileh

Doha Institute for Graduate Studies, Doha, Qatar
Email: ammar.shamaileh@dohainstitute.edu.qa

(Received 12 August 2023; revised 23 January 2024; accepted 24 March 2024)

Abstract

Among the greatest challenges facing scholars of public opinion are the potential biases associated with survey item nonresponse and preference falsification. This difficulty has led researchers to utilize nonresponse rates to gauge the degree of preference falsification across regimes. This article addresses the use of survey nonresponse rates to proxy for preference falsification. A simulation analysis exploring the expression of preferences under varying degrees of repression was conducted to examine the viability of using nonresponse rates to regime assessment questions. The simulation demonstrates that nonresponse rates to regime assessment questions and indices based on nonresponse rates are not viable proxies for preference falsification. An empirical examination of survey data supports the results of the simulation analysis.

Keywords: survey methodology

Imagine that you reside within a politically repressive autocratic state. As you are washing the dishes, you hear a sequence of knocks and scurry to the peephole to see who is on the other side of your door. The face you observe is an unfamiliar one ostentatiously displaying the signs of patiently pretending not to have heard your footsteps. After cautiously opening the door, introductions are made and you are soon annoyed—or maybe relieved—to realize that the person standing outside of your home is a stranger interested in administering a survey. Perhaps you are bored or overly polite, but for some reason, you decide to accept this curious individual's invitation to answer several questions. Many of the inquiries are benign, and some appear to be absurd, but the experience initially feels marginally rewarding. However, the peculiar satisfaction you were feeling wanes as sensitive questions regarding political preferences are brought up. Eventually, you are asked the following question:

> How much confidence do you have in the government?

You are offered four choices ranging from none to a great deal, along with the option to not answer the question at all by claiming not to know. Choosing not to answer the question is an answer in and of itself, and not one likely to be offered by a government supporter (Kuran, 1991; Wedeen, 1999). If you do not fear retribution or genuinely have confidence in the government, your response will likely be truthful. The perceived costs are minimal, and you gain some satisfaction in expressing your opinion. But what if you do fear speaking out and are not a government supporter?

The person who knocked on your door is a stranger, and identifiable information could wind up in the government's hands. You may believe the survey itself was commissioned by the

government, despite assurances that it was not (Tannenberg, 2021). The negligible benefits of faithfully responding to the survey may now be perceived as potentially dangerous. You can take the opportunity provided by this stranger to note your dissatisfaction with the government while accepting the perceived potential costs of speaking out. You can try to evade the question by saying that you do not know, yet I do not know may be interpreted as a tepid critique of the regime. Or, you can mimic the regime supporter by stating that you have a great deal of confidence in the government. While a modicum of fear may push you toward nonresponse, an increase in fear may push you away from nonresponse toward support for the regime.

An individual's response to regime assessment questions is the product of both their desire to express their true preference and the costs they may face for expressing that true preference. It is argued within this article that as the sensitivity of the question and costs of responding truthfully increase, what increases is not the probability of claiming not to know, but the probability of shifting closer to the regime supporter's position by providing the answer preferred by the regime. For this reason, we cannot infer the likelihood of eliciting truthful responses to politically sensitive questions across regimes based on a simple examination of nonresponse rates to regime assessment questions.

One of the fundamental challenges facing researchers examining public opinion, particularly within authoritarian regimes, is the potential bias associated with nonresponses and preference falsification on surveys (Berinsky, 2002; Vavreck, 2007; Jiang and Yang, 2016; Benstead, 2018). While surveys are rarely conducted in the most repressive contexts, they are conducted in a fairly wide range of states with varying levels of repression and repressive technologies. It has recently been suggested that nonresponse rates may be used to evaluate the prevalence of preference falsification within authoritarian states (Benstead, 2018; Shamaileh, 2019; Shen and Truex, 2021). Nonresponse rates have frequently been used as tangential elements of analyses related to preference falsification meant to either corroborate the primary findings or test some ancillary aspect of the research (Jiang and Yang, 2016; Robinson and Tannenberg, 2019; Shamaileh, 2019). The intuition underlying the use of nonresponse rates to proxy for preference falsification in tangential analyses is at the heart of the recently developed self-censorship index (SCI), which is meant to act as a proxy for preference falsification that allows for cross-country comparisons (Shen and Truex, 2021). The primary objective of this article is to contribute to the collective attempt to measure preference falsification in surveys conducted in authoritarian contexts by evaluating the efficacy of using nonresponse rates as a proxy.

Self-censorship can be conceptualized broadly or narrowly. A narrow conceptualization would limit our inquiry to situations where an individual literally does not express an opinion, as is the case when a person chooses to respond that they do not know when they do know their preference. A broader conceptualization would incorporate answers that fall within the scope of preference falsification, whereby an individual does not merely refrain from expressing their preference, but actively mimics the preferences of others to avoid the costs of being perceived as a dissident (Kuran, 1989, 1991, 1997; Crabtree *et al.*, 2020). Others who have considered political discourse in authoritarian regimes have focused on the nuances of symbolic battles fought within the delineated limits established within a regime, yet such examinations rarely highlight silence as a consequence of authoritarian restrictions (Wedeen, 1999; Svolik, 2012). In order for a measure to act as a proxy for the reliability of the responses as expressions of true preferences, it must offer some consistent directional relationship to the broader conceptualization of self-censorship by biasing the results in the direction preferred by those in power.

Can nonresponse rates proxy for this broader notion of self-censorship? While nonresponse rates can play a role in a multipronged analysis of preference falsification, and the SCI suitably measures self-censorship in the narrow sense, nonresponse rates to regime assessment questions cannot generally act as a viable proxy for the degree of preference falsification. Measures that are fundamentally rooted in the response rates to regime assessment questions will not be consistently, or necessarily often, associated with higher levels of preference falsification. In this article,

a simple extensible theoretical model that bridges the literature on sensitivity bias in surveys and preference falsification in authoritarian regimes is developed to help illustrate this point and provide an underlying logic for how repression influences responses to survey questions. To demonstrate why nonresponse rates to regime assessment questions should not be used to gauge preference falsification, a simulation experiment is conducted that examines when preference falsification arises and the SCI's ability to capture bias due to preference falsification. A brief analysis of World Values Survey data further corroborates the findings of the simulation analysis.

1. A simulation analysis

This simulation experiment provides a parsimonious test of the efficacy of the use of nonresponse rates to politically sensitive questions to measure preference falsification, as well as the viability of inferences drawn from analyses using such measures. Ultimately, this simulation will demonstrate that utilizing politically sensitive questions where the regime's preferred response is likely known, such as when questions relate to the government or regime itself, is not a reliable proxy for levels of preference falsification within a society. While increasing repression when levels of repression are relatively low (under certain reasonable conditions) may indeed increase nonresponse rates to politically sensitive questions, as repression increases beyond this level, the nonresponse rate to politically sensitive questions and SCI score should decrease. Thus, the nonresponse rates of relatively repressive states and nonrepressive states may be indistinguishable from one another, yet the repressive state may produce patterns of responses that are highly unrepresentative of the true preferences of people within the state.

The most direct and robust attempt to proxy for preference falsification using nonresponse rates was set forth by Shen and Truex (2021), where they produced an index of self-censorship across democratic and autocratic states using nonresponse rates to politically sensitive questions adjusted for the nonresponse rates to questions that are not politically charged. In their view, "item nonresponse rates can be used to measure self-censorship, which will indirectly allow us to assess the broader idea of preference falsification, and to measure its likely incidence across time and space" (Shen and Truex, 2021: 1675). This view reflects the logic underlying the use of nonresponse rates to proxy for preference falsification in many other studies (Jiang and Yang, 2016; Robinson and Tannenberg, 2019; Shamaileh, 2019). The goal is to capture to some reasonable extent the sensitivity bias associated with preference falsification through a measure of self-censorship since they ostensibly move in the same direction. The article and analysis are well crafted, and Shen and Truex are careful to note many of the potential limitations of their exploratory analysis. Moreover, their contention that there is heterogeneity among autocratic regimes with regard to the truthfulness of responses to politically sensitive questions is certainly likely and supported by the literature, as well as the findings of this article. Given that their measure represents the most thorough attempt to capture preference falsification using nonresponse rates, this analysis explores the viability of using nonresponse rates to proxy for preference falsification while focusing on the SCI that they developed (Shen and Truex, 2021). Thus, this article should not be construed as narrowly focused on evaluating the efficacy of the SCI, but on evaluating the use of nonresponse rates to proxy for preference falsification more broadly by using the most developed measure of this nature.[1]

I simulated a set of true preferences of 1000 respondents within a model state ($i \in \{1, \ldots, n\}$: $n = 1000$). For simplicity, it is assumed that the true preferences of individuals are evenly distributed across all contexts.[2] The answers to three positively correlated political questions and three

[1]It should also be noted that the results reached by Shen and Truex were successfully replicated and extended using alternative statistical tools in analyses not shown in this paper and that the authors should be lauded for the transparency of their work.

[2]Alternative distributions are explored near the end of this section and in the online Appendix.

uncorrelated nonsensitive questions were simulated for the respondents. The distribution of these responses was intentionally made approximately equal for ease of interpretation. The number of questions simulated was chosen in order to reproduce the SCI in Shen and Truex (2021). Let x represent an individual's response to a question and $\hat{x}$ represent an individual's true preference. To match the survey items generally used by scholars, I assumed that each x is an integer ranging from 1 (least preferred by the regime) to 5 (most preferred by the regime). Where $x \neq \hat{x}$, an individual is engaging in some degree of preference falsification. True preferences were distributed across the population in the manner described in Table 1.[3] Individuals were also assigned a parameter associated with the desire to express their true preferences that are distributed uniformly across the population; $0 < c < 1$. This was used to construct a cost function associated with expressing an opinion different from one's true preference. Thus, this represents the internal costs of preference falsification noted by Kuran (1991).

Utilizing the same underlying preferences, 1000 democratic and autocratic states were simulated ($s \in \{1, \ldots, r\} : r = 2000$). Each of these states allocates some cost associated with expressing an opinion that differs from the regime's preferred stance ($0 < d < 1$). While repression and the costs of repression should generally be thought of as endogenous (Kuran, 1989; Davenport, 2007b; Ritter *et al.*, 2016), for the purposes of analyzing responses to a survey where the responses of others are not known, it is reasonable to treat repression as an exogenous parameter. This parameter should be thought of as generally capturing the overall perceived probability of being punished and the perceived expected punishment that might be allocated if authorities are informed of an individual's response. Such punishments should not be thought of strictly in terms of raids on dissident homes, imprisonment or torture by regime officials. The costs of deviating from the regime's preferred position may also involve less severe punishments such as the refusal of selective benefits or access to certain institutions by local officials or regime-aligned elites.

For the simulation analysis, I assumed that democratic institutions constrain the costs that can be allocated to punish deviations from the regime's preferred position, and, for simplicity, democratic states range from 0 to 0.1 for the repression parameter. The 0.1 limit on repression was arbitrarily chosen to allow some heterogeneity among democratic regimes, but lower or higher thresholds would produce qualitatively similar results. Authoritarian states are distributed uniformly between 0 and 1, producing greater heterogeneity in repression among them. It is further assumed that all governments prefer the maximum response to the survey question (e.g., individuals answer that they have a great deal of confidence in the government), and that their second preference is for moderate support, third is nonresponse, fourth is a moderately critical position and fifth is the most critical response.

Adopting Kuran's model would not allow for shifts toward nonresponse unless nonresponse is indeed the regime's preferred response since Kuran's model does not allow individuals to partially mask their opinions. Nonresponse would reveal that an individual is not a supporter. Therefore, Kuran's basic formulation is utilized, where the costs of repression and the psychological costs of lying factor into an individual's expression of preferences, while allowing for moderate positions to be adopted. For simplicity, quadratic loss functions are used to model the internal and external costs associated with deviating from the regime's preferred position. It is assumed that individuals weigh the internal costs of not expressing their true opinion against the external costs of expressing their opinion. The further an individual deviates from either the regime's preferred position or their own preferred position, the more costly such deviations are (decreasing at a decreasing rate). This produces an internal bargain between the regime's stance and an individual's true preference. The simple utility function used to capture this tradeoff is:

$$U_i(x) = -d_s(x - x_s)^2 - c_i(x - \hat{x}_i)^2 \tag{1}$$

[3]See the online Appendix for alternative distributions of true preferences.

Table 1. Simulated true preferences (proportions)

Response	Political Q1	Political Q2	Political Q3	Non-sensitive Q1	Non-sensitive Q2	Non-sensitive Q3
None at all	0.211	0.221	0.210	0.211	0.225	0.203
Not very much	0.203	0.221	0.193	0.214	0.217	0.212
Don't know	0.166	0.152	0.163	0.159	0.159	0.159
Quite a lot	0.205	0.192	0.201	0.204	0.197	0.211
A great deal	0.215	0.214	0.233	0.212	0.202	0.215

Thus, an individual's response, if the answer were measured continuously, would be a weighted average residing somewhere between their preferred response and the regime's:

$$x^* = \frac{d_s}{d_s + c_i} x_s + \frac{c_i}{d_s + c_i} \hat{x}_i \tag{2}$$

No stochastic component was added to the model since adding such a component would not substantively contribute to this analysis. The discussion that follows will focus on the straightforward prediction that repression increases the degree to which an individual agrees with the regime:

$$\frac{\partial x^*}{\partial d} = \frac{c_i(x_s - \hat{x}_i)}{(d_s + c_i)^2} \tag{3}$$

The expressed preference for the regime's position is increasing when an individual's position is not perfectly aligned with the regime's; $x_s > \hat{x}_i$. Thus, while Shen and Truex (2021) observe no relationship between repression and their SCI scores, this simulation will demonstrate why the SCI as currently formulated is not constructed to provide a reliable measure of preference falsification.[4]

There are several plausible alternative strategies for modeling the behavior of survey respondents answering regime assessment questions. One strategy would be to treat nonresponses as opting out of responding; however, since the question has already been asked, a nonresponse offers a response. Moreover, since the answers to regime assessment questions preferred by officials in authoritarian contexts, particularly direct questions related to support for the government, are known to virtually all respondents capable of participating in the survey, there is little room for masking an individual's preferences by blending in with a subset of the population that lacks such knowledge. A second strategy could use the same basic formulation provided within this article, but assume that only active dissent is punished. This would be plausible in certain contexts, and in such contexts measures of preference falsification that rely on nonresponse rates may be theoretically justified. Yet, the seemingly arbitrary shifts in what is permitted, when, by whom and which punishments are allocated create some uncertainty as to how authorities may respond to the mere withdrawal of active support in most authoritarian states (Wedeen, 1999). Moreover, in many states, autocratic patronage networks and local distributive channels also create the possibility of the withdrawal of benefits from nonsupporters and decentralized punishment that is less devastating in nature than those used against dissidents by the security apparatus (Kuran, 1991; Carlson, 2018; Mazur, 2021). It should be noted that the parsimonious model presented above accounts for the reduced probability/cost of punishment for nonresponse by modeling the costs of deviating from the regime's preferred answer as increasing at an increasing rate.

[4]One possible reason that they may not observe a relationship between the SCI and repression is that they utilize an *ex-post* measure of repression, the Political Terror Scale (PTS), rather than an *ex-ante* measure of repression. However, as the empirical analysis in the following section will demonstrate, there is little evidence of a connection between *ex-ante* repression and the SCI.

Finally, it should be noted that a high proportion of survey respondents in many contexts appear to believe that the surveys they are responding to are sponsored by the government and that there is evidence that fear of the state biases results on surveys (Zimbalist, 2018; Tannenberg, 2021). Of course, we cannot assume fear plays an important role in shaping response patterns to regime assessment questions in all authoritarian states [see Lei and Lu (2017); Stockmann *et al.* (2018)]. It is precisely the goal of tools such as the SCI to inform us as to when preference falsification is biasing responses.

I constructed the data-generating process for the simulations using the model presented above. Given that the questions used to construct the SCI require ordinal rather than continuous responses, for the simulated true preferences, internal individual psychological costs, and external country-level costs, each individual chooses the answer that maximizes her utility within the context of this model.[5] Figure 1 provides a simple representation of how an individual's response varies as the external costs of dissenting increase. Each of the three represented individuals possesses a preference for providing the answer most critical of the regime. As repression increases, the regime critic's response moves closer to the regime's. Theoretically, an increase in repression may push the regime critic to not respond, but it also may push the regime critic to move from nonresponse to an expression of support. It is for this reason that we cannot consistently expect nonresponse to provide a reliable indication of increased preference falsification.

Note that each of the simulated countries has identical populations with regard to their underlying preferences, and that all that varies across these regimes in this analysis is the external cost imposed on deviating from the regime's preferred position. This produces different sets of expressed preferences, and it is the differences between these states in this regard that are of primary interest in this analysis. An SCI score was created for each simulated state using the equation developed by Shen and Truex (2021)[6] :

$$SCI_s = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} nonresponse_{i,j}}{m \cdot n}^{\text{Regime Assessment}} - \frac{\sum_{k=1}^{m}\sum_{i=1}^{n} nonresponse_{i,k}}{m \cdot n}^{\text{Non-Sensitive}} \tag{4}$$

The distribution of the simulations across SCI scores is presented in Figure 2. The pattern produced is similar to the pattern of SCI scores exhibited in Shen and Truex (2021). While the distribution of SCI scores across authoritarian regimes is larger than across democratic regimes ($sd_{\text{autocracy}} = 0.044$, $sd_{\text{democracy}} = 0.017$), the average SCI score across autocracies and democracies are substantively similar ($\mu_{\text{autocracy}} = 0.029$, $\mu_{\text{democracy}} = 0.025$).[7] A low SCI score, however, does not necessarily or even generally indicate lower levels of preference falsification.

Following Blair *et al.* (2020), sensitivity bias induced by preference falsification is conceptualized as the absolute distance of an individual's response from their true preference. Aggregating for all questions across all individuals in each state, the mean deviation from true preferences is[8]

$$Bias_s = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} |x_{i,j} - \hat{x}_{i,j}|}{m \cdot n} \tag{5}$$

Figure 3 presents the LOESS (locally weighted regression) curves for the relationship between the SCI scores, sensitivity bias, and the external costs of speaking out across authoritarian and democratic regimes. Sensitivity bias is consistently increasing in the external costs of criticism at a decreasing rate. At low levels of repression, SCI scores are indeed increasing in these external

[5]We assume equal distances for the purposes of the simulations.

[6]Regime assessment questions are indexed by $j \in 1, \ldots, m$, and nonsensitive questions are indexed by $k \in 1, \ldots, m$.

[7]The *t*-test is sensitive to the number of simulations that are run as well as the specific distribution of external costs that was chosen. Nevertheless, the difference between the two groups was statistically significant ($p < 0.011$).

[8]This should not be considered a generalizable measure of sensitivity bias.

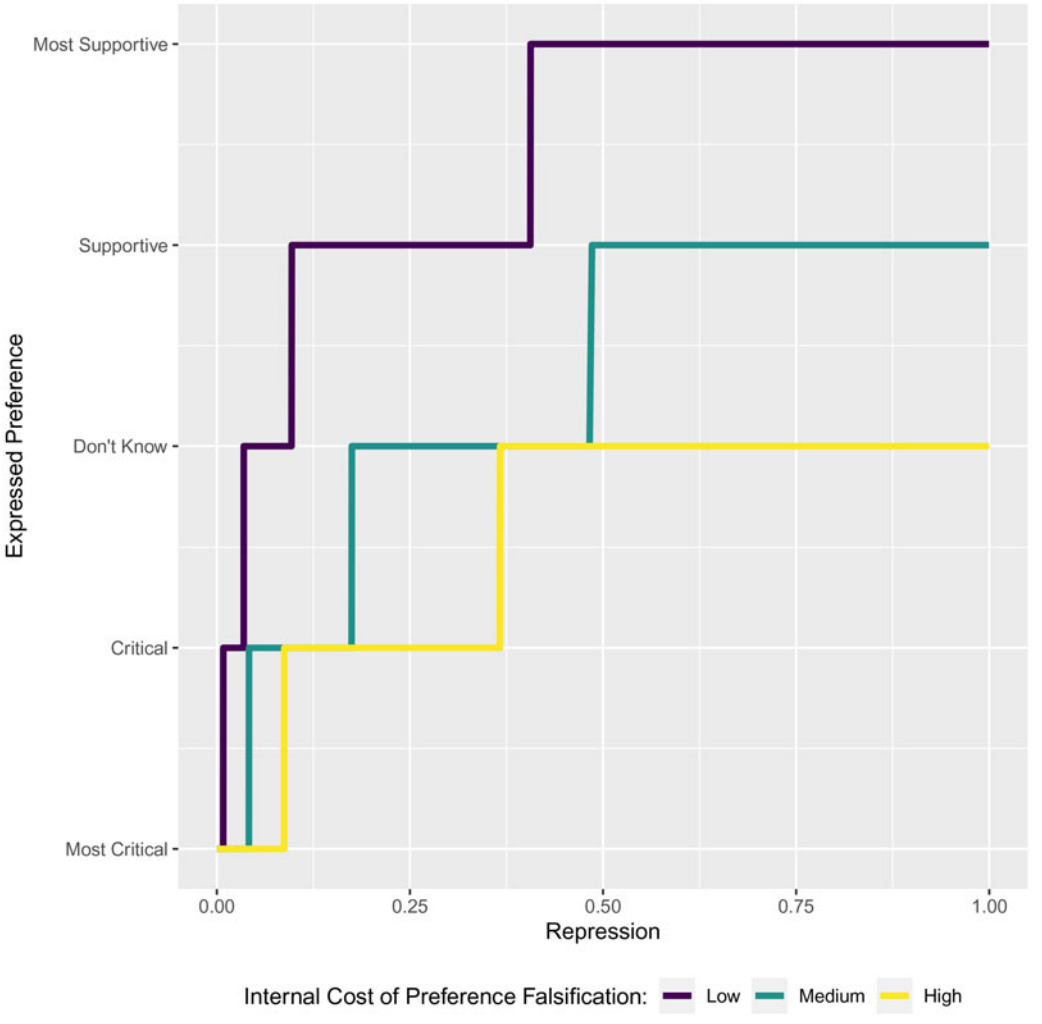


Figure 1. Change in expressed preference as repression increases.

costs. As the external costs of repression increase, however, the SCI score begins to decrease, and at very high levels of repression, SCI scores should be lower than when repression is relatively low. More importantly, the SCI shares a similar relationship with the measure of sensitivity bias. Thus, as Shen and Truex (2021) found, there appears to be no consistent relationship between SCI scores and repression, but that does not imply that there is no consistent relationship between preference falsification and repression.

To further examine how the use of nonresponse rates, and SCI scores in particular, may lead to invalid inferences, Figure 4 presents the responses for three simulated states with the same SCI score, 0.042, indicating relatively low (although nonnegligible) levels of self-censorship. The states include a democracy with low levels of repression, an autocracy with low levels of repression and an autocracy with high levels of repression. All three states have the same exact distribution of true preferences among the population, yet the expressed preferences for the repressive autocratic state differ dramatically from those of the other states. While this example may be relatively extreme in that the repressive state has no individuals who explicitly take an anti-government

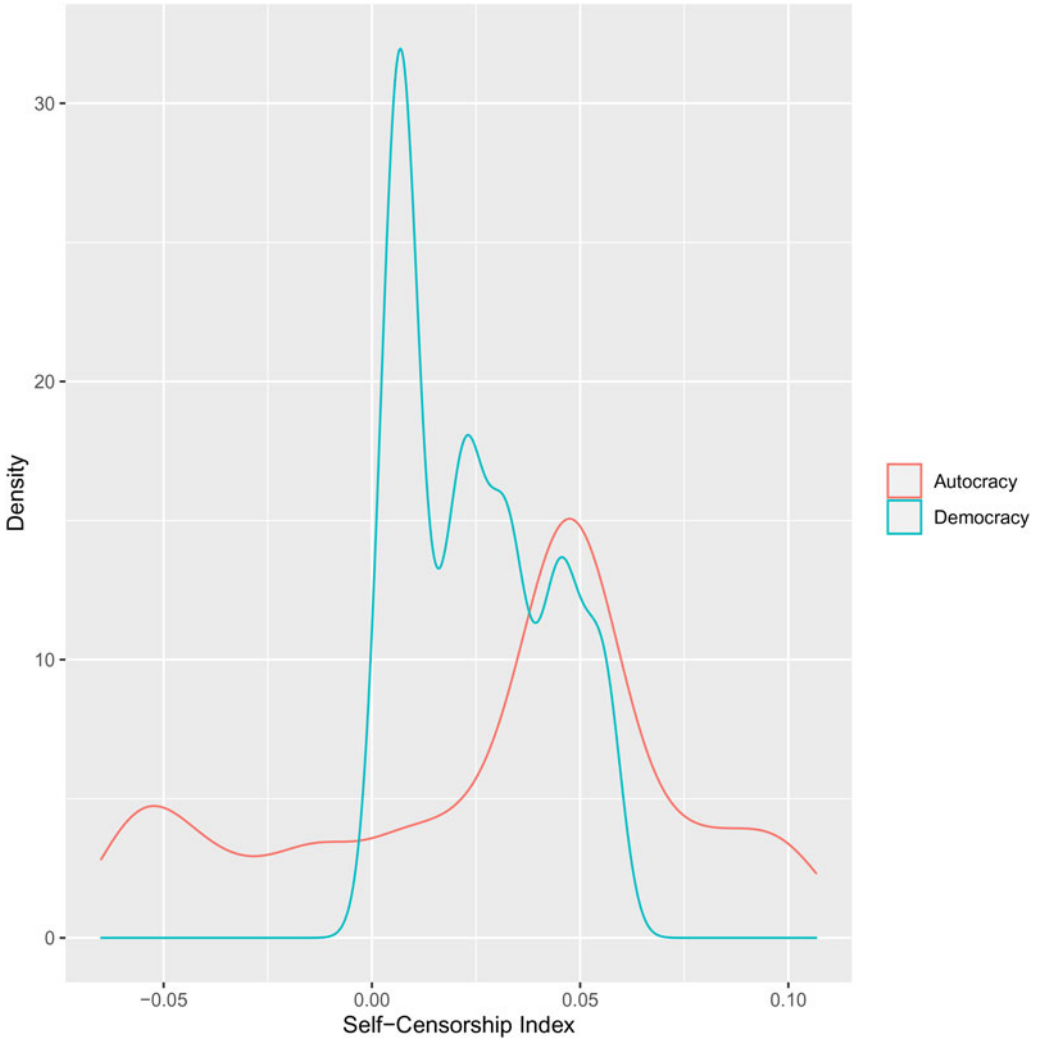


Figure 2. SCI scores and regime type.

stance, this highlights the problematic nature of attempting to infer levels of preference falsification from the use of nonresponse rates. While approximately 42 percent of respondents' true simulated preferences are critical of the regime, 0 percent provide critical responses. Preference falsification is far more prevalent in the repressive state, yet the repressive state produces the same SCI score as states with far lower levels of repression. Autocratic regimes bunched near democratic states in SCI scores may be producing relatively low or high levels of preference falsification, and the most repressive regimes should produce SCI scores lower than environments where there is little repression of speech. It should also be noted that even the minor costs associated with voicing dissent in states with low levels of repression may produce severe selection bias in the responses.

It could be posited that SCI scores are useful proxies only when repression is sufficiently low. Indeed, Figure 1 does show that at low levels of repression, the SCI is capturing sensitivity bias. However, this is the product of fixed distributional assumptions as to the true preferences of citizens within a state. This causes two related problems associated with the interpretation of such

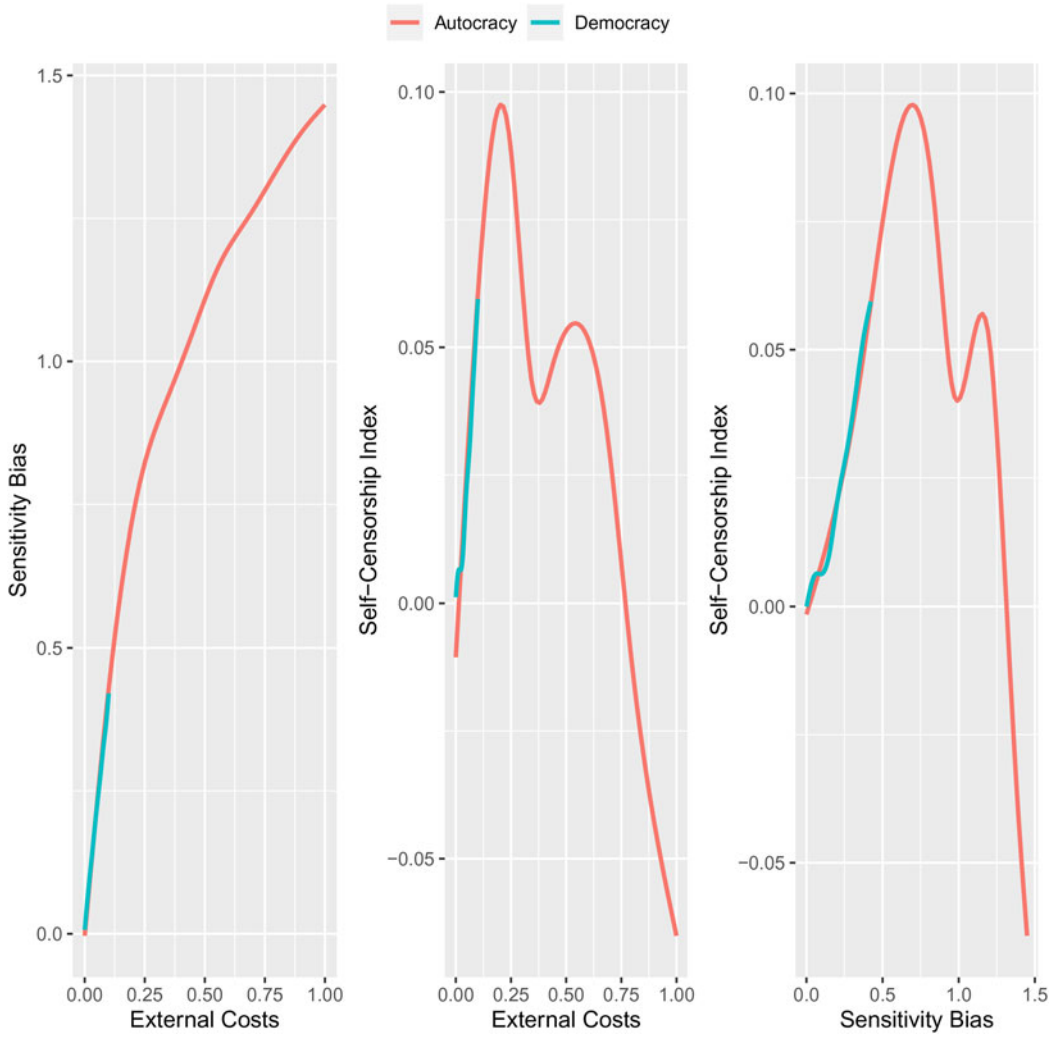


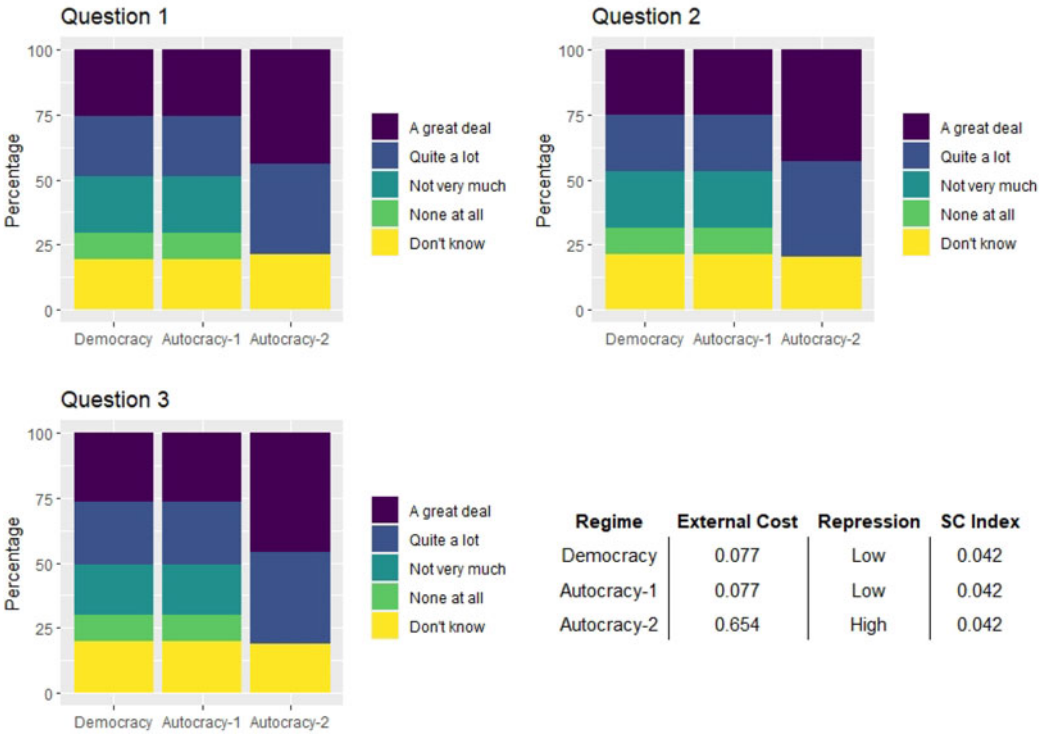
Figure 3. Change in SCI scores with repression.

scores. First, the score itself is highly sensitive to the distribution of preferences within the state, even when holding true preferences for nonresponse constant. Second, we cannot gauge at what level of repression the SCI score loses its ability to act as a proxy for preference falsification without understanding the underlying distribution of true preferences. Figure 5 presents the LOESS curves for the relationship between SCI scores across simulations run on populations with preferences that are high and low relative to the analysis in the initial simulation, and demonstrates the sensitivity of SCI scores to the underlying true preferences of individuals within a state.[9]

2. Some empirical evidence

Does the empirical evidence suggest that, despite its limitations, the SCI—as well as other indices based on nonresponse rates—can serve as a viable cross-country measure of preference

[9]See the online Appendix for further details.



Regime	External Cost	Repression	SC Index
Democracy	0.077	Low	0.042
Autocracy-1	0.077	Low	0.042
Autocracy-2	0.654	High	0.042

Figure 4. Change in item responses for simulated states with the same SCI score.

falsification? If an index of nonresponse rates can serve as an indirect measure of preference falsification, it should be associated with an inflation in the assessment of regimes. While the results presented in this section are merely suggestive, there is little evidence to suggest that nonresponse rates to regime assessment questions can act as a suitable measure of preference falsification. Using six waves of the World Values Survey, V-Dem's Freedom of Expression Index as a proxy for the repression of speech (Pemstein *et al.*, 2018),[10] and the replication data provided by Shen and Truex (2021), this brief analysis demonstrates that the theorized effect of repression on responses to regime assessment questions that was developed in this article is plausible and that the SCI does not appear to capture sensitivity bias in favor of the regime.[11] All estimated correlations in the analysis are conditional on a country being included in the World Values Survey, which is skewed toward less restrictive and more democratic settings. In addition, it is important to acknowledge the potential bias observed due to selection into the survey.

Figure 6 presents the response distributions from four countries for the question related to confidence in the government and highlights the inability of the SCI or nonresponse rates to

[10]While Shen and Truex (2021) utilize the PTS (Wood and Gibney, 2010), the freedom of expression index more acutely captures the *ex-ante* repression of political discourse and is more fundamentally related to the topic at hand. Although the PTS may be an appropriate measure for responsive repression, it does not capture *ex-ante* repression. For discussions of *ex-ante* repression and its relationship to dissent, see Davenport (2007a) and Ritter *et al.* (2016). Of course, there are other drawbacks related to the use of the PTS to measure repression, including the data sources used, underreporting of active repression in many authoritarian regimes, and the fact that incidents of responsive repression are likely low in the states most likely to respond to dissent with the most brutal forms of repression (Fein, 1995; Carey, 2010; Pierskalla, 2010).

[11]Summary statistics and analyses utilizing alternative coding schemes can be found in the online Appendix. The regression analyses use only data from states where all questions used to construct the SCI were asked. A robustness check that calculates the SCI using any available questions can also be found in the online Appendix.

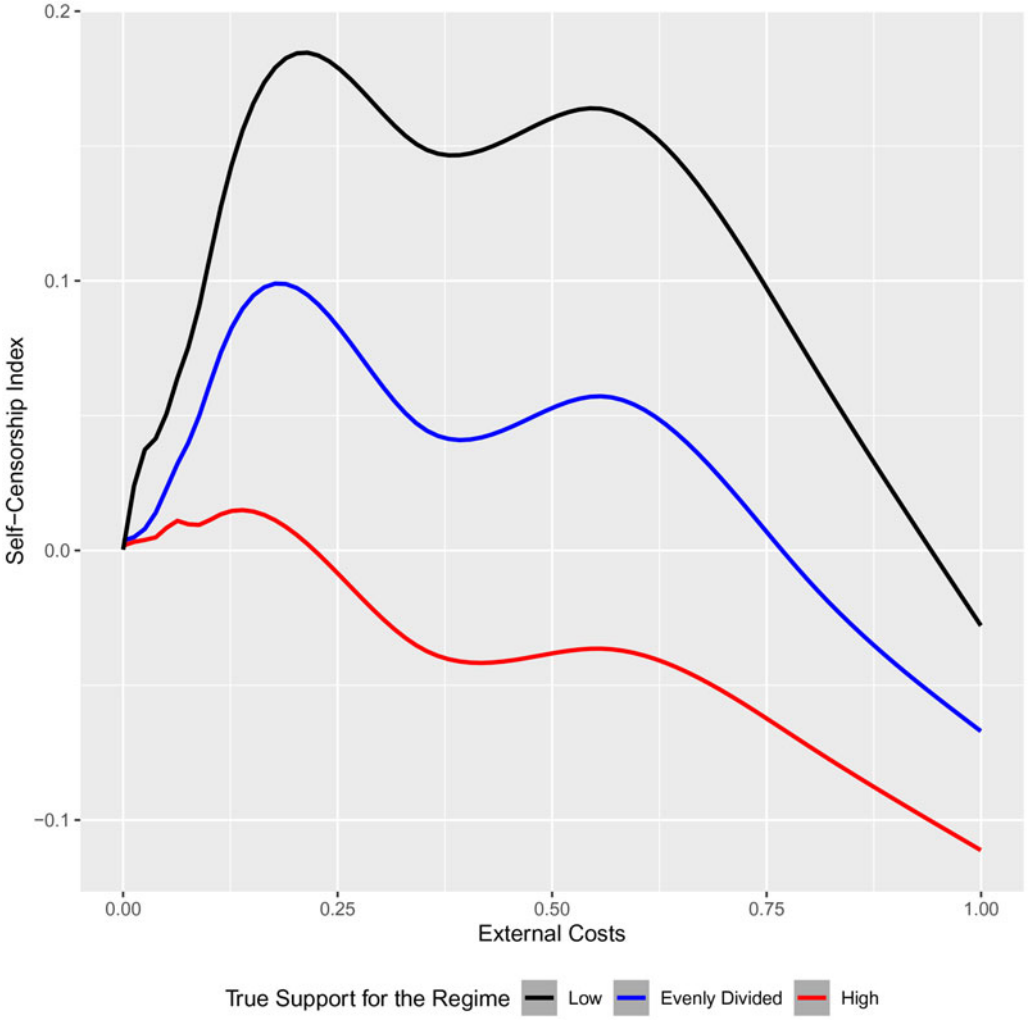


Figure 5. Distribution of true preferences and the SCI.

proxy for preference falsification. The UK, Iraq, and Uzbekistan all have relatively low SCI scores and low nonresponse rates for this particular question, yet they differ in their Freedom of Expression scores.[12] More importantly, there is substantial variation in the response patterns. In Iraq and the UK, criticism of the government is fairly common, indicating that levels of preference falsification are likely low. In Uzbekistan, however, less than 3 percent expressed a lack of confidence in the government, and less than 2 percent did not respond to the question. While the underlying true preferences of these three societies are likely substantially different, it is highly unlikely that we would observe such low levels of criticism of Uzbekistan's government if coercion was not affecting responses. Neither the broader SCI nor the nonresponse rate for this particular question points to preference falsification. Furthermore, a high SCI score does not imply higher levels of preference falsification, as the patterns from Pakistan in 2001 indicate. Although the

[12]For this graphic, whether SCI scores are high or low is based on a calculation using all available questions. They also align with the SCI scores in Shen and Truex (2021). SCI scores (all available questions): UK, 0.026; Iraq, −0.04; Uzbekistan, 0.01; Pakistan, 0.071.

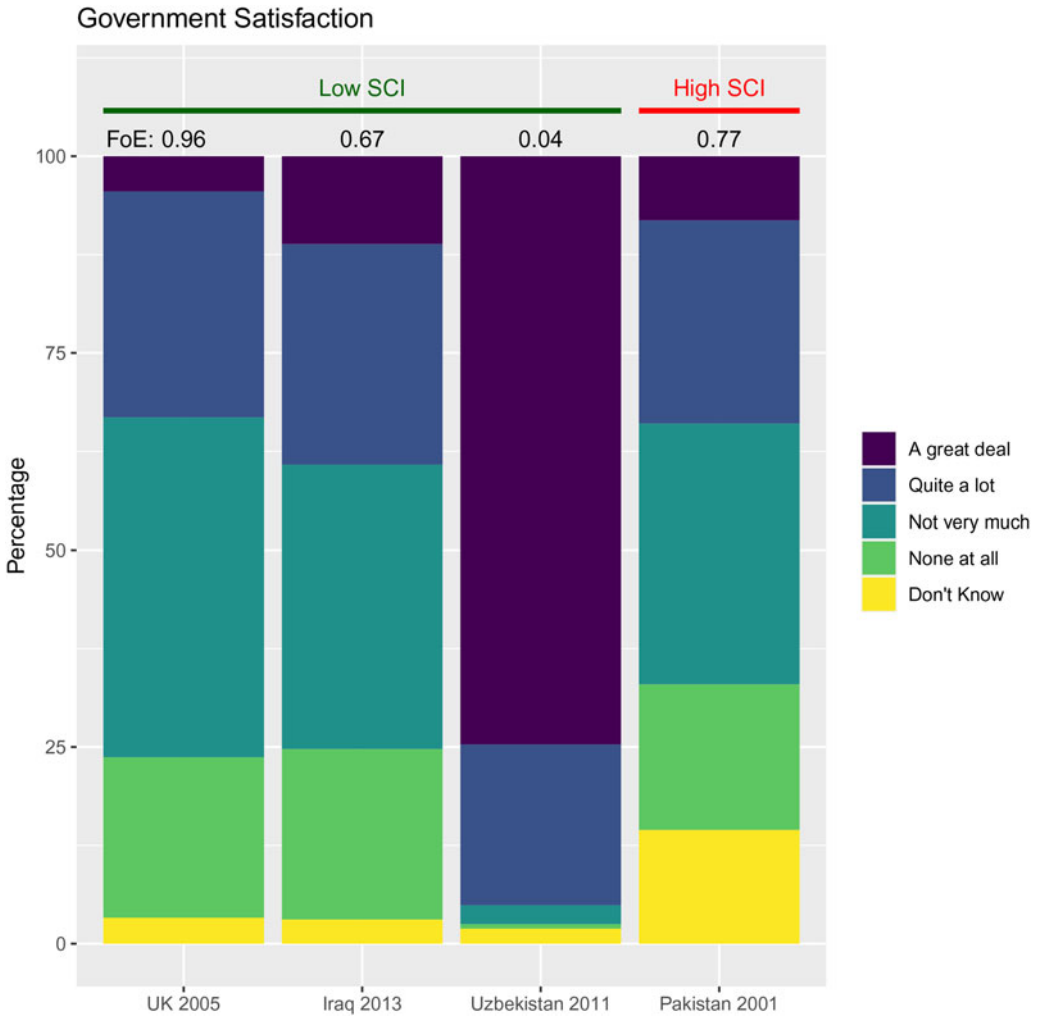


Figure 6. Response patterns for four states.

nonresponse rate is high, criticism of the regime is pervasive with over 50 percent responding that they have no confidence or not very much confidence in the government. These countries' scores on the *Freedom of Expression* index appear to provide a more reliable proxy for the prevalence of preference falsification. While this is just a small subset of cases, it should be noted that the pool of country-years in the analysis is small and that there were several combinations of countries chosen that would have produced the same result. For example, we could replace the UK with France, Iraq with Iran, Uzbekistan with Azerbaijan (2011), and Pakistan with Morocco (2007 or 2011) and reach the same conclusions. In fact, of the states with relatively high SCI scores, only China appears to exhibit signs of mass reticence to express dissent.

What is the relationship between freedom of expression, confidence in government, and the SCI for the full sample of country-years? Figure 7(a) presents the bivariate relationship between V-Dem's *Freedom of Expression Index* and the mean of *Government Confidence* for each country-year. There is a strong and consistent negative correlation between freedom of expression and confidence in the government which is decreasing at a decreasing rate. This result is consistent with similar recent analyses that have been conducted (Nathan, 2020; Tannenberg, 2021), as well as the theoretical prediction that preference falsification is consistently increasing in the

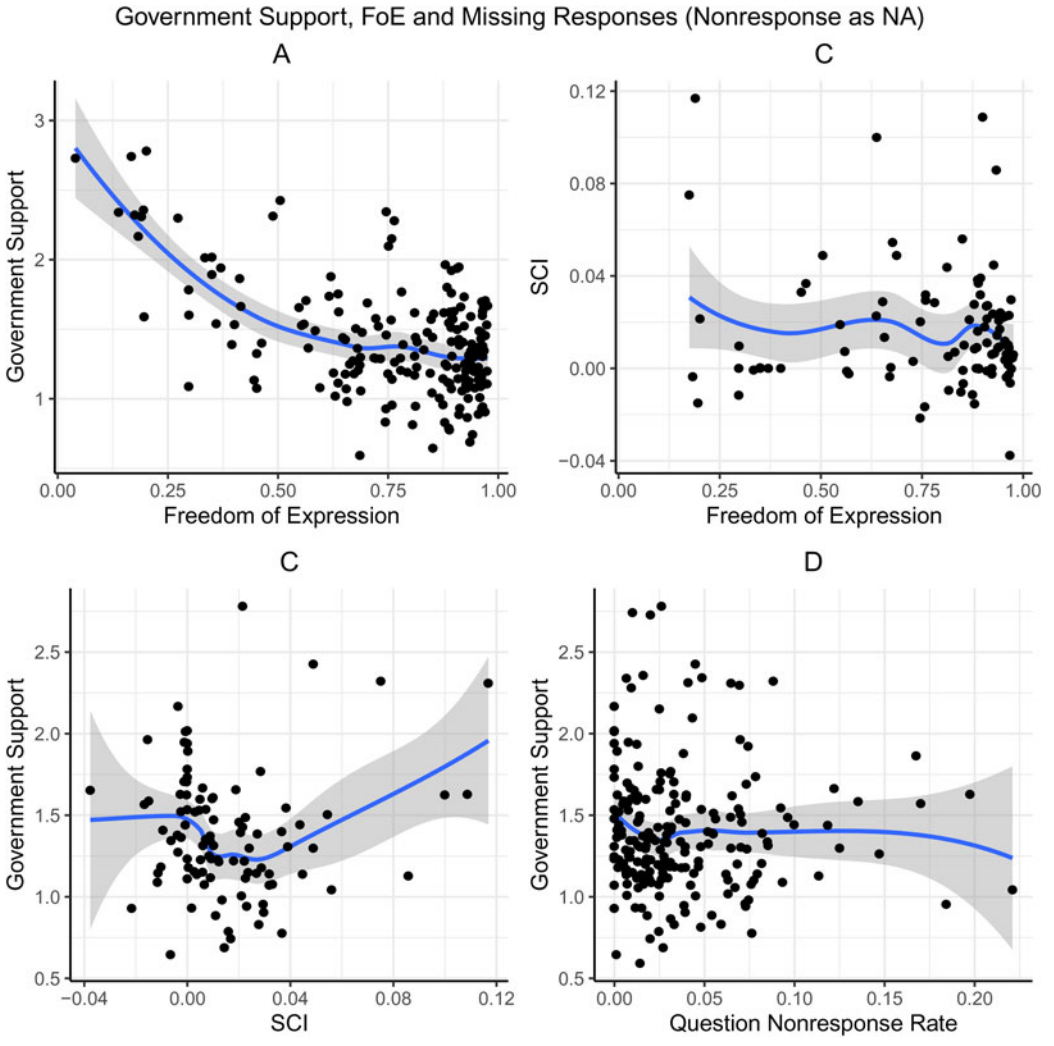


Figure 7. Government support, freedom of expression, and the SCI.

costs of expressing criticism of the regime. Of course, the correlation could be due to confounding variables associated with the set of highly repressive regimes where the survey was conducted or the potential for authoritarian propaganda to alter preferences and beliefs. Nevertheless, the relationship between the *Freedom of Expression Index* and *Government Confidence* seems likely to be at least partially due to preference falsification. The relationship between the *Freedom of Expression Index* and the *SCI*, however, is weak and inconsistent [Figure 7(b)]. Moreover, the bivariate relationship between the *SCI*/nonresponse rates and *Government Support* does not appear to be substantively meaningful or consistently positive.

Ostensibly, if the SCI were capable of capturing preference falsification, we would see some inflation in the assessment of regimes as the SCI increases when holding other variables constant. Thus, the most relevant tests of the SCI's relationship to regime assessment require the inclusion of contextual variables that may shape nonresponse rates. For ease of exposition, the results of simple ordinary least squares (OLS) models examining the weighted mean values of regime assessments will be the centerpiece of the empirical examination. The unit of analysis in this set of regression models is the country-year. All country-years that were available were utilized.

For the simple aggregate group-level OLS analyses, the dependent variables were the weighted means of the regime assessment questions utilized to construct the SCI: *Government Confidence*, *Human Rights* in the country, and how democratic the country is according to the respondents (*Democracy*). For each of the analyses in this section, I coded nonresponses as missing. An alternative coding scheme where nonresponses were coded as the middle category can be found in the online Appendix and does not alter the predictions of the analyses in the paper. In addition, I utilized listwise deletion due to sample size considerations, the lack of compelling alternatives for the analysis of this particular data, and its simplicity. Given the robustness of the results that follow, it is unlikely that the handling of the missing data impacted the results.

There is little evidence of a consistent relationship between the *SCI* and regime assessment. For the range of the *SCI* where the majority of states are found, the relationship between the *SCI* and *Government Confidence* is negative or flat, and many of the states with seemingly inflated levels of Government Support are at the lower end of the *SCI* [Figure 7(c)]. However, some of the most repressive states where World Values surveys were conducted were relatively wealthy. Figure 8 presents a simple country-year level OLS analysis that accounts for the country's polity score, freedom of expression, GDP per capita, the weighted average of the subjective evaluations of the financial conditions of the household, enrollment in secondary education, and a dummy variable capturing whether a country's oil rents exceeded 5 percent for all three regime assessment questions. Even when excluding the most repressive regimes from the analysis, there is little evidence that the SCI offers a reliable proxy for the sensitivity bias associated with preference falsification. While none of the analyses seemed to indicate that the SCI was associated with higher regime assessments, there is evidence that the Freedom of Expression Index may offer a reliable proxy when evaluating the potential for sensitivity bias. Intuitively, this result would imply that simply examining regime-imposed limitations on speech may offer the most appropriate method for evaluating whether the regime assessments provided by survey respondents are potentially reliable.

Finally, as a robustness check, I also utilized per-cluster regression to account for individual-level and group-level effects when exploring the relationship between the SCI and the expressed preferences of survey respondents across contexts. In particular, it is plausible that when accounting for the variables that might influence political preferences, the SCI is associated with inflated expressions of regime support. If this were the case, an argument could be made that the SCI or some other measure rooted in nonresponse rates could be adjusted in a manner that allows for the measurement of preference falsification. Per-cluster regression was chosen due to the potential bias associated with group-level estimates in random effects models, including bias-corrected multilevel models, and the inability of the standard fixed effects framework to include group-level covariates (Bates *et al.*, 2014; Hazlett and Wainstein, 2022).

The dependent variable for these analyses is the individual's response to the relevant regime assessment question. In the first step, I ran OLS analyses that included country-year fixed effects and the set of individual-level covariates noted below. I account for fixed effects due to the possibility that contextual factors that are not measured influence the underlying preferences of respondents in ways that are not captured by the group-level variables in the model. At the individual level, the models included the respondent's age, education, gender, whether they reside in an urban or rural community, and their subjective evaluations of the financial condition of their household. The marginal effects of the individual-level covariates were then subtracted from the responses of the individuals for the dependent variable and averaged by group. OLS regression was then used to examine the relationship between the dependent variable and the country's SCI score, as well as other country-year variables measuring democracy (using Polity IV scores) and economic development (log of GDP per capita).

The results of the per-cluster regression analyses are presented in Figure 9 and are in line with those of the simple OLS analyses of country-year aggregates. The SCI does not appear to be associated with the inflation of regime assessment scores. If anything, higher SCI scores are generally associated with lower evaluations of the regime when holding other variables constant. The SCI is

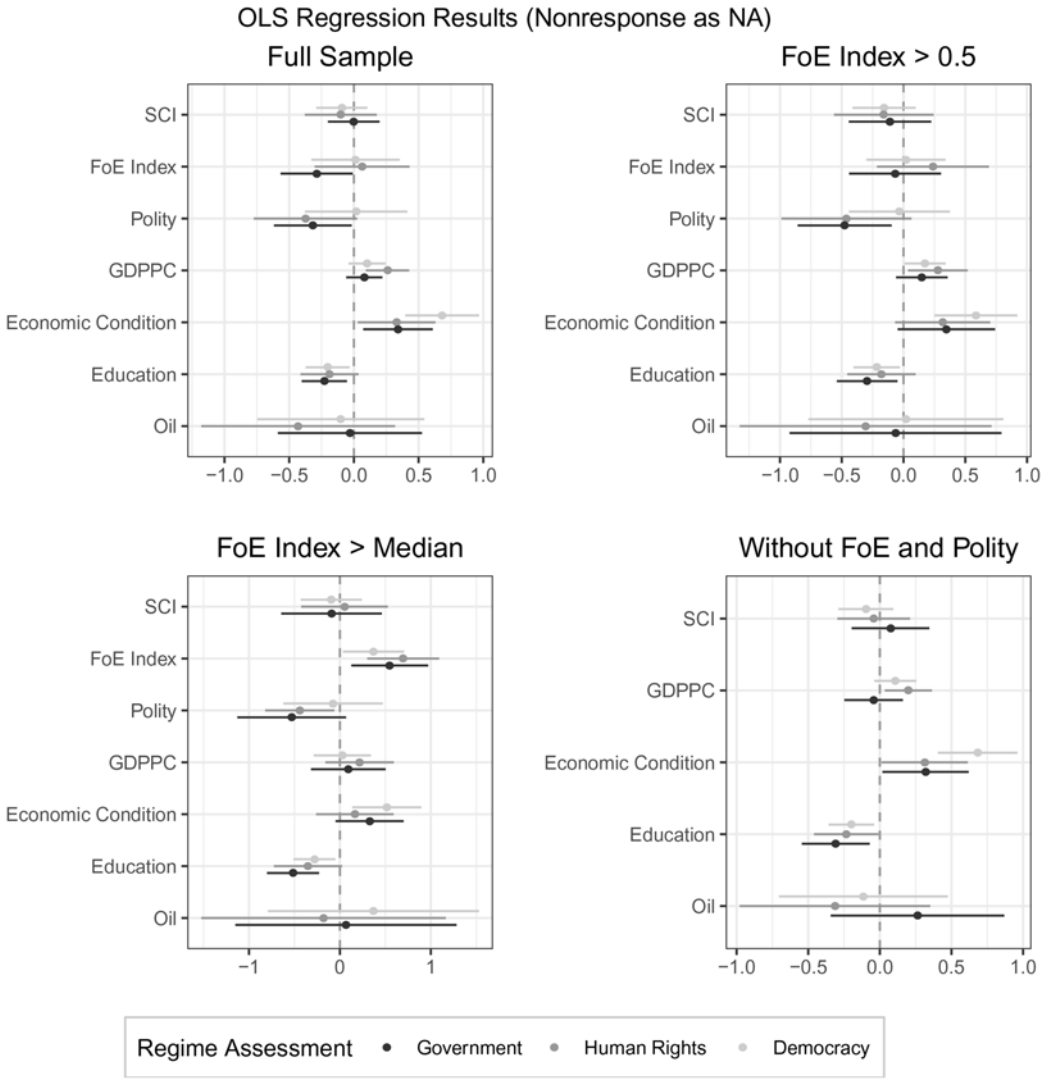


Figure 8. OLS regression results.

negatively associated with evaluations for each of the regime assessment questions, albeit at levels that do not cross traditional levels of statistical significance. While the results are not statistically significant or substantively large for the three analyses, and the number of country-year observations is relatively low, there are intuitively appealing explanations for a potential negative correlation. The same freedom that provides individuals with the leeway to criticize the government allows individuals to comfortably claim that they do not know. In addition, democratic contexts produce competing political groups, which may produce greater levels of uncertainty as to how people feel about the regime based on who is in power. Such uncertainty may be the product of not knowing which political faction to support, or it may be associated with uncertainty as to how to evaluate the overall performance of a regime when the side you do not support is in power.

Furthermore, alternative specifications replacing the SCI with *Freedom of Expression* demonstrate a substantively meaningful and significant decrease in confidence in the government as *Freedom of Expression* increases, corroborating the theoretical model presented in this article. Although the results

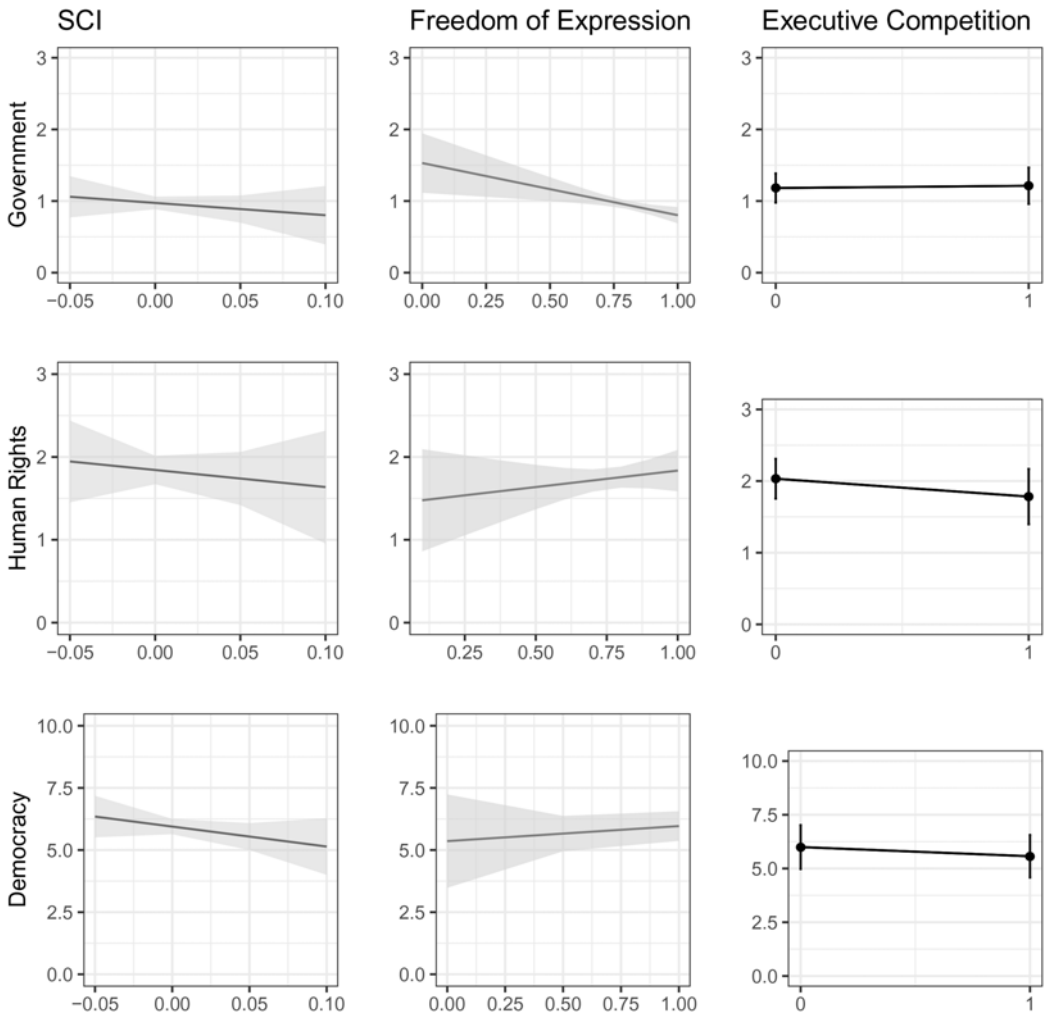


Figure 9. Predicted probabilities for per-cluster regression analyses.

of the per-cluster regression analyses do not provide evidence of a decrease in the evaluations of democracy and human rights within the country, the questions themselves are conceptually closely related to freedom of expression. Thus, freedom of expression is likely to be positively correlated with evaluations of how democratic a country is and how much it respects human rights conditional on respondents answering the questions truthfully. In addition, one of the core findings of Shen and Truex (2021) is a statistically significant negative correlation between electoral competition for the executive and the SCI. As such, we would expect a negative correlation between executive competition and regime assessment when holding other variables constant.[13] None of the analyses point to evidence of a substantively meaningful inflation of scores among states without executive competition when holding other variables constant.[14] Nevertheless, these results should be interpreted particularly

[13]It should be noted that analyses using mixed-effects models corroborated the findings of Shen and Truex (2021) with regard to the negative correlation between executive competition and nonresponse rates. In my opinion, this result is not due to executive competition producing more room for individuals to express their opinions, but, rather, greater confusion as to whether the regime is democratic.

[14]For these analyses, Polity scores were removed from the analysis due to collinearity. The inclusion of Polity did not qualitatively alter the results.

cautiously given the low number of groups in the analysis of executive competition. Overall, the results presented above are consistent across several robustness checks, including OLS regression analyses that include group-level variables and cluster-robust standard errors, linear mixed-effects regression analyses, and limiting the analysis to authoritarian regimes (online Appendix).

The evidence presented above is not conclusive. Nevertheless, while the primary basis for rejecting the use of missing responses to proxy for the sensitivity bias associated with preference falsification is theoretical, the empirical evidence presented in this section further suggests that measuring preference falsification via nonresponse rates is not a viable solution. Moreover, it provides some support for the underlying theoretical mechanisms driving preference falsification presented in this article.

Although the SCI as currently constructed cannot effectively serve as a proxy for preference falsification, it introduces an interesting setup for the continued exploration of how preference falsification can be measured. If those less supportive of a regime mimic the preferences of regime supporters, any attempt to measure preference falsification via nonresponse rates should try to choose questions where mimicry is difficult. One possible extension of the strategy outlined by (Shen and Truex, 2021) would be to attempt to measure preference falsification using sensitive political questions where there is no clear answer that government supporters would provide (Shamaileh, 2019: 962). Such an index would generally provide a weak, but potentially meaningful, proxy for preference falsification. A more robust examination of this is provided in the online Appendix, yet this solution is also problematic.

3. Conclusion

Preference falsification on surveys has the potential to distort both our inferences regarding the preferences of individuals within a society and the determinants of such preferences. While surveys are rarely run in many of the most repressive states, surveys are frequently conducted in authoritarian states where it is hard to parse whether preference falsification is influencing the results. Social scientists have utilized nonresponse rates to proxy for preference falsification to both gauge the potential for bias and evaluate the effectiveness of other tools meant to account for preference falsification. Using a simulation analysis and empirical evidence drawn from the World Values Survey, this article demonstrates why nonresponse rates to regime assessment questions are not a valid or reliable proxy for preference falsification. In doing so, it contributes to a burgeoning literature that has pushed researchers to theorize and contextualize the varied motivations underlying survey responses (Parkinson, 2022). Beyond evaluating the efficacy of measures of preference falsification based on nonresponse rates, this article also produces a simple theoretical model of preference falsification on surveys that conceptualizes survey responses to regime assessments as the product of an internal bargain between an individual's desire to express their true preference and that individual's fear of coercion. This theoretical model is extensible and may be useful in providing a foundation for novel theoretical inquiries regarding survey responses to regime assessment questions and may assist in the development of new tools meant to gauge preference falsification on surveys.

Any attempt to measure preference falsification should be adequately theorized and approached cautiously, utilizing information related to the respondent and the respondent's context. One potential avenue for exploration would be to rely on a theoretical framework, such as the one presented in this article to model the data-generating process, which might then be used to inform the construction of a measurement of preference falsification or estimates of political preferences. One avenue of research that may prove to be particularly fruitful is the construction of theoretically motivated dimensional constraints within an IRT (item response theory) framework (Morucci *et al.*, 2021). In addition, those trying to understand what nonresponse patterns might imply may want to build on Liu and Wang (2016) for an IRT measurement strategy to inform our understanding of how to interpret nonresponse to politically sensitive questions.

Social scientists have produced a plethora of valuable strategies and tools for working around or diminishing incentives to falsify preferences on surveys (Vavreck, 2007; Blair and Imai, 2012; Cesarini *et al.*, 2014; Berinsky, 2018; Stockmann *et al.*, 2018; Blair *et al.*, 2020; Lyall *et al.*, 2020; Nanes and Haim, 2020; Pan and Siegel, 2020). These creative mechanisms should continue to be employed where possible and appropriate until an effective solution to gauging and accounting for preference falsification is produced. Of course, list experiments and other tools designed to handle sensitivity bias and preference falsification are not without their own potentially significant drawbacks (Blair *et al.*, 2020; Kuhn and Vivyan, 2022). Nevertheless, in certain contexts, they may provide a more adequate lens than direct questions to peer into the political preferences of a society.

Where alternative mechanisms for solving issues related to preference falsification are unavailable, strategies for measuring preference falsification using direct questions should take an approach that accounts for contextual political, cultural, and economic factors. Such evaluations may utilize nonresponse rates, but nonresponse rates to regime assessment questions should not form the basis for inferences regarding preference falsification. For example, Jiang and Yang (2016) root the questions they utilize to understand preference falsification in their understanding of Chinese politics and local dynamics and exploit this understanding to attempt to measure preference falsification among different constituencies. States with similar nonresponse rates can vary dramatically regarding how much item nonresponse is biasing results. At a minimum, a qualitative assessment of restrictions on political speech should inform the interpretation of survey data collected in authoritarian contexts and temper any inferences drawn from data collected in more repressive settings. Furthermore, the implications of this analysis should also extend to the use of nonresponse rates to measure social desirability bias (Berinsky, 2002, 2008). Whether sensitivity bias is induced by social or political factors, uncertainty concerning which response is most desirable may allow for nonresponse rates to proxy for question or topic sensitivity. However, when the desirable answer is known, nonresponse rates are not reliable proxies for issue sensitivity or sensitivity bias.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2024.29. To obtain replication material for this article, https://doi.org/10.7910/DVN/0CTJJY.

Acknowledgments. This manuscript benefited greatly from feedback provided by Marius Radean, Xiaoli Guo, Ibrahim Khatib, and the other participants in the Doha Institute's Quantitative Methods in Political Science Workshop. Comments provided by Luai Allarakia, Sarah Parkinson, the reviewers of the manuscript, and the editors of PSRM also helped greatly improve the manuscript. Finally, I would like to thank Xiaoxiao Shen and Rory Truex for the transparency with which they conducted the research this paper builds on, and their academic collegiality. All errors and oversights remaining in this manuscript are my own Open access funding provided by the Qatar National Library.

References

Bates MD, Castellano KE, Rabe-Hesketh S and Skrondal A (2014) Handling correlations between covariates and random slopes in multilevel models. *Journal of Educational and Behavioral Statistics* **39**, 524–549.

Benstead LJ (2018) Survey research in the Arab world: challenges and opportunities. *PS: Political Science & Politics* **51**, 535–542.

Berinsky AJ (2002) Silent voices: social welfare policy opinions and political equality in America. *American Journal of Political Science* **46**, 276–287.

Berinsky AJ. (2008) Survey non-response. In Donsbach W and Traugott MW (eds), *The Sage Handbook of Public Opinion Research*. London: Sage, pp. 309–321.

Berinsky AJ (2018) Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding. *The Journal of Politics* **80**, 211–224.

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

Blair G, Coppock A and Moor M (2020) When to worry about sensitivity bias: evidence from 30 years of list experiments. *American Political Science Review* **114**, 1297–1315.

Carey SC (2010) The use of repression as a response to domestic dissent. *Political Studies* **58**, 167–186.

Carlson E (2018) The relevance of relative distribution: favoritism, information, and vote choice in Africa. *Comparative Political Studies* **51**, 1531–1562.

Cesarini D, Johannesson M and Oskarsson S (2014) Pre-birth factors, post-birth factors, and voting: evidence from Swedish adoption data. *American Political Science Review* **108**, 71–87.
Crabtree C, Kern HL and Siegel DA (2020) Cults of personality, preference falsification, and the dictator's dilemma. *Journal of Theoretical Politics* **32**, 409–434.
Davenport C (2007a) *State Repression and the Domestic Democratic Peace*. Cambridge, UK: Cambridge University Press.
Davenport C (2007b) State repression and the tyrannical peace. *Journal of Peace Research* **44**, 485–504.
Fein H (1995) Life-integrity violations and democracy in the world, 1987. *Human Rights Quarterly* **17**, 170–191.
Hazlett C and Wainstein L (2022) Understanding, choosing, and unifying multilevel and fixed effect approaches. *Political Analysis* **30**, 46–65.
Jiang J and Yang DL (2016) Lying or believing? Measuring preference falsification from a political purge in China. *Comparative Political Studies* **49**, 600–634.
Kuhn PM and Vivyan N (2022) The misreporting trade-off between list experiments and direct questions in practice: partition validation evidence from two countries. *Political Analysis* **30**, 381–402.
Kuran T (1989) Sparks and prairie fires: a theory of unanticipated political revolution. *Public Choice* **61**, 41–74.
Kuran T (1991) Now out of never: the element of surprise in the east European revolution of 1989. *World Politics: A Quarterly Journal of International Relations* **44**, 7–48.
Kuran T (1997) *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.
Lei X and Lu J (2017) Revisiting political wariness in China's public opinion surveys: experimental evidence on responses to politically sensitive questions. *Journal of Contemporary China* **26**, 213–232.
Liu C-W and Wang W-C (2016) Unfolding IRT models for Likert-type items with a don't know option. *Applied Psychological Measurement* **40**, 517–533.
Lyall J, Zhou Y-Y and Imai K (2020) Can economic assistance shape combatant support in wartime? Experimental evidence from Afghanistan. *American Political Science Review* **114**, 126–143.
Mazur K (2021) *Revolution in Syria: Identity, Networks, and Repression*. Cambridge, UK: Cambridge University Press.
Morucci M, Foster M, Webster K, Lee SJ and Siegel D (2021) Measurement that matches theory: theory-driven identification in IRT models. ArXiv preprint arXiv:2111.11979.
Nanes M and Haim D (2020) Self-administered field surveys on sensitive topics. *Journal of Experimental Political Science* **8**, 185–194.
Nathan AJ (2020) The puzzle of authoritarian legitimacy. *Journal of Democracy* **31**, 158–168.
Pan J and Siegel AA (2020) How Saudi crackdowns fail to silence online dissent. *American Political Science Review* **114**, 109–125.
Parkinson SE (2022) (Dis)courtesy bias: "methodological cognates," data validity, and ethics in violence-adjacent research. *Comparative Political Studies* **55**, 420–450.
Pemstein D, Marquardt KL, Tzelgov E, Wang Y-T, Krusell J and Miri F (2018) The v-DEM measurement model: latent variable analysis for cross-national and cross-temporal expert-coded data. *V-Dem Working Paper 21*.
Pierskalla JH (2010) Protest, deterrence, and escalation: the strategic calculus of government repression. *Journal of Conflict Resolution* **54**, 117–145.
Ritter EH and Conrad CR (2016) Preventing and responding to dissent: the observational challenges of explaining strategic repression. *American Political Science Review* **110**, 85–99.
Robinson D and Tannenberg M (2019) Self-censorship of regime support in authoritarian states: evidence from list experiments in China. *Research & Politics* **6**, 2053168019856449.
Shamaileh A (2019) Never out of now: preference falsification, social capital and the Arab spring. *International Interactions* **45**, 949–975.
Shen X and Truex R (2021) In search of self-censorship. *British Journal of Political Science* **51**, 1672–1684.
Stockmann D, Esarey A and Zhang J (2018) Who is afraid of the Chinese state? Evidence calling into question political fear as an explanation for overreporting of political trust. *Political Psychology* **39**, 1105–1121.
Svolik MW (2012) *The Politics of Authoritarian Rule*. New York, NY: Cambridge University Press.
Tannenberg M (2021) The autocratic bias: self-censorship of regime support. *Democratization* **29**, 591–610.
Vavreck L (2007) The exaggerated effects of advertising on turnout: the dangers of self-reports. *Quarterly Journal of Political Science* **2**, 325–343.
Wedeen L (1999) *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. Chicago, IL: University of Chicago Press.
Wood RM and Gibney M (2010) The Political Terror Scale (PTS): a re-introduction and a comparison to CIRI. *Human Rights Quarterly* **32**, 367.
Zimbalist Z (2018) "Fear-of-the-state bias" in survey data. *International Journal of Public Opinion Research* **30**, 631–651.

Cite this article: Shamaileh A (2024) On the measurement of preference falsification using nonresponse rates. *Political Science Research and Methods* 1–19. https://doi.org/10.1017/psrm.2024.29