

Common Retrospective Economic Perception Items Are Noisy But Broadly Reliable*

Jack Bailey[†]

This version January 27, 2022.

Word count: 2,007

Abstract

Retrospective economic perception items dominate economic voting research. Though they are well-used, they are not well-validated. In this short note, I assess the items' test-retest reliability for the first time. I make three contributions. First, I show that the items are broadly reliable. Second, I show that personal items are more reliable than national ones. Third, I show that the differences in reliability I find have little effect on model outputs. Thus, though these items have their problems, reliability is likely not one of them.

*Thank you to Jon Mellon whose advice improved this manuscript.

[†]Research Associate, Department of Politics, The University of Manchester, UK. If you have any comments or questions, feel free to contact me either by email (jack.bailey@manchester.ac.uk) or on Twitter (@PoliSciJack).

Introduction

One set of survey items dominates the economic vote. They ask how personal finances and the national economy have changed over the past 12 months. When it comes to individual-level research, these items are invaluable. Most individual-level data comes from a single country at a single time. This means that macro-economic variables remain constant. Yet the economic perceptions that voters report do not. Thus, we can use them to test the economic vote.

These items are well-used, but not well-validated. What's more, we now know that they suffer from a host of problems. The most serious is partisan bias De Vries, Hobolt, and Tilley (2018). That is, incumbent supporters tend to be more positive, and opposition supporters more negative, than similar non-partisans. We also know that respondents report different economic perceptions in different contexts. Partisan bias, for instance, is worse in political compared to non-political surveys (Bailey 2021). Further, question order matters too and affects the answers respondents give (Wilcox and Wlezien 1993; Sears and Lau 1983).

I assess the items' test-retest reliability for the first time. This is an essential metric for survey research. Imagine that we repeat an item twice, only a short time apart. A reliable item would show almost exactly the same set of responses each time. An unreliable item would not. Either way, we learn something useful: the item's susceptibility to measurement error.

I make three substantive contributions. First, I show that the items show good reliability. When asked to give their perceptions twice in a single survey, most respondents give the same response each time. Second, I show that personal items are more reliable than national ones. Third, I show that the differences in reliability I find have little effect on model outputs. As such, I conclude that retrospective economic perception items are noisy but broadly reliable.

Data

Data come from wave 22 of the British Election Study Internet Panel (Fieldhouse et al. 2021). Fieldwork occurred between 26 November 2021 and 15 December 2021, and the survey took around 15 minutes. Most waves of the panel measure voters' economic perceptions only

Table 1: Measures of test-retest reliability across two responses to each retrospective item embedded in wave 22 of the British Election Study Internet Panel

Item	Repeated	Cramér	Pearson	Spearman
National Economic Perceptions	72.3%	0.60	0.73	0.75
Personal Economic Perceptions	87.3%	0.79	0.83	0.88

once. But wave 22 instead measured them *twice*. The first time was as usual, with all 28,135 respondents reporting their perceptions. The second time was around 5 minutes later, with a subset of 6,948 respondents reporting them again. I take each member of this subset's responses to each item as my sample.

Assessing Test-Retest Reliability

Assessing an item's test-retest reliability requires some assumptions. Two are most important. First, that respondents' latent perceptions remain stable throughout the survey. As it took only around 15 minutes to complete, this seems reasonable. Second, that each response instance measures respondents' latent perceptions *with error*. Again, as measurement error is unavoidable in survey research, this seems reasonable too. If both are correct, we can compare the two instances to test the item's reliability (Yu 2005).

Table 1 shows each item's test-retest reliability across four different metrics. Note that both items are nominal. As such, the most appropriate measure is Cramér's V (Cramér 1946)¹. This metric provides a measure of association between nominal variables. Still, the percentage of repeated responses and Pearson/Spearman correlations produce similar results².

Both items appear reliable enough. In general, respondents tended to give the same answer each time. Yet note that the two measurements occurred only around 5 minutes apart. Given this, there does appear to be a reasonable amount of measurement error. Indeed, 27.7% of respondents answered the national item a different way the second time around.

Further, the national item is less reliable than the personal one across all four metrics.

¹To help the reader calibrate their understanding of this metric, I provide wave-on-wave estimates of Cramér's V in figure A1

²The latter two metrics apply to continuous and ranked data, respectively. As such, I remove any "Don't know" responses in these cases.

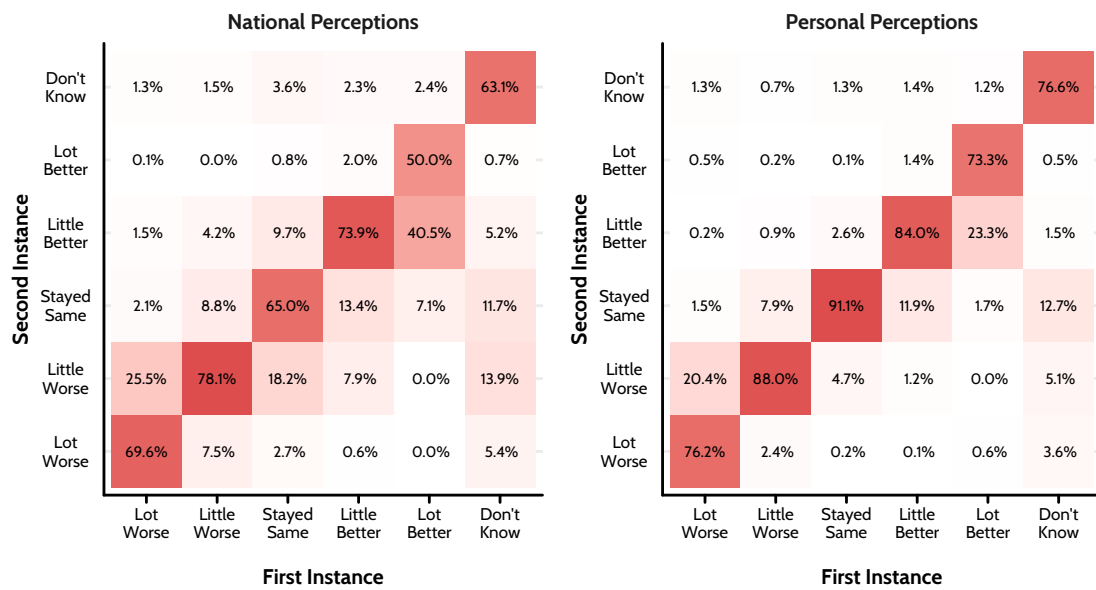


Figure 1: Crosstabs comparing first (columns) and second (rows) responses to both retrospective economic perceptions items in wave 22 of the British Election Study Internet Panel

Consider Cramér's V. Here a score of 0 implies no association and a score of 1 implies total association. The personal item has a score of 0.79. But the national item scores only 0.60. That personal economic perceptions are more reliable than national ones is perhaps unsurprising. After all, most will better understand their own finances than the national economy.

We can learn more by considering the items in greater detail. Figure 1 shows each item's joint response distribution. Here, columns reflect the first response and rows reflect the second. Likewise, darker colours imply greater stability from one instance to the next. As we would expect, the diagonal cells include the highest scores. This is because most respondents give the same response each time.

The figure makes three facts most clear. First, that changing responses tend to change only a little. For instance, 25.5% of respondents who said that the economy "got a lot worse" went on to say that it "got a little worse." Likewise, 20.4% of respondents who said that their own finances "got a lot worse" did the same. Second, that the personal item is stable across the distribution but the national item is not. 69.6% of respondents who said that the economy "got a lot worse" at first did so again. Yet the figure for those who said that it "got a lot better" was only 50.0%. This compares to figures of 76.2% and 73.3% for the personal item, respectively. Third, that "don't

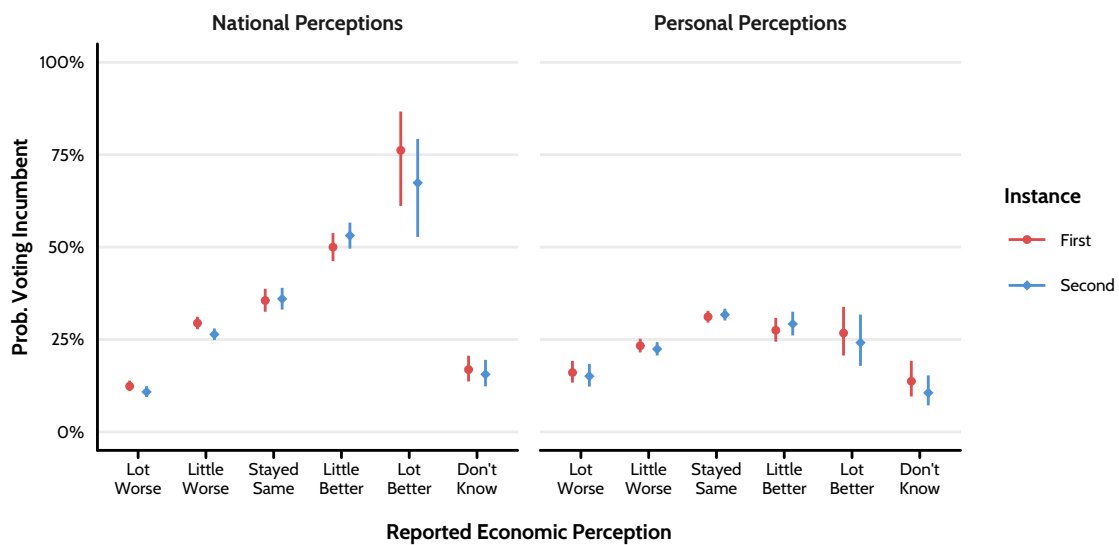


Figure 2: Despite some measurement error, the bivariate relationship between reported economic perceptions and incumbent support remains broadly stable across both response instances

know” might often mean “can’t be bothered to answer.” For example, 36.9% of respondents who gave a “don’t know” response to the national item and 23.4% who gave a “don’t know” response to the personal item then went on to provide an informative answer on their second attempt.

Consequences of Measurement Error

Most economic voting models are simple. They model incumbent voting as a function of either retrospective item. My results above show that these items contain some measurement error. Thus, it different response instances might yield different economic voting estimates.

Figure 2 shows predictions from four simple bivariate regression models. Each models the probability of voting for the incumbent Conservative Party. The only difference between them is which instance of which variable they use.

Other than some minor differences, models fit to each instance produce similar results. This is true for both national and personal items. Though the predictions that they produce differ, they almost always overlap. Still, these differences suggest an interesting possibility: that some of the variability we see in economic voting results (Duch and Stevenson 2008) is due only to measurement error.

As an aside, note that the personal item has a non-monotonic effect on incumbent voting. Most applied economic voting research treats this item as continuous. But, clearly, its effect is non-linear. As such, past research likely misestimates the item's association with incumbent voting. At the negative end of the scale, this research will understate the item's influence. Likewise, at the positive end of the scale it will overestimate its influence. Economic voting scholars should, thus, relax strict linearity assumptions.

Conclusion

Retrospective economic perception items have problems, but reliability is not one of them. As I show here, the items are noisy but broadly reliable. Measured minutes apart, respondents report similar perceptions. These reported perceptions then yield similar inferences.

My findings help to quantify the measurement error in these items for the first time. It is worth noting that such error serves to attenuate model estimates. This suggests two possibilities. First, that economic voting models underestimate economic voting effects. Second, that we underestimate how much partisan bias is present in these models instead.

With only a single item, tackling measurement error is impossible. But with multiple items, this is no longer the case. Though developing new items might seem burdensome, the benefits outweigh the costs. This is true in both an intellectual and a financial sense. Intellectually, noisy indicators make academic research more difficult. They reduce our statistical power and bias the parameters that we estimate. Both factors, in turn, induce variability across studies. This variability, then, clouds the inferences that we make about key social phenomena. Financially, public money often funds this research. We have a duty to our stakeholders—the general public—not to waste their money. One obvious way to do so is to field surveys that include only well-validated and reliable items.

Multi-item scales would present new opportunities too. The literature lacks a theory to explain voters' economic perceptions. Indeed, it treats them as simple and unidimensional. There is no evidence to suggest that this is the case (indeed, it is likely not). New items, thus, offer both better measurement and a chance to better understand the economic vote.

References

- Bailey, Jack. 2019. "The Fact Remains: Party ID Moderates How Voters Respond to Economic Change." *Electoral Studies* 61 (102071): 1–13. <https://doi.org/10.1016/j.electstud.2019.102071>.
- . 2021. "Political Surveys Bias Self-Reported Economic Perceptions." *Public Opinion Quarterly* 85 (4): 987–1008. <https://doi.org/10.1093/poq/nfab054>.
- Cramér, Harald. 1946. *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.
- De Vries, Catherine E., Sara B. Hobolt, and James Tilley. 2018. "Facing up to the Facts: What Causes Economic Perceptions?" *Electoral Studies* 51 (February): 115–22. <https://doi.org/10.1016/j.electstud.2017.09.006>.
- Duch, Raymond M., and Randolph T. Stevenson. 2008. *The Economic Vote: How Political and Economic Institutions Condition Election Results*. Cambridge: Cambridge University Press.
- Fieldhouse, Edward, Jane Green, Geoffrey Evans, Jonathan Mellon, Christopher Prosser, Roosmarijn A. de Geus, and Jack Bailey. 2021. "British Election Study Internet Panel, 2014–2023."
- Sears, David O., and Richard R. Lau. 1983. "Inducing Apparently Self-Interested Political Preferences." *American Journal of Political Science* 27 (2): 223–52. <https://doi.org/10.2307/2111016>.
- Wilcox, Nathaniel, and Christopher Wlezien. 1993. "The Contamination of Responses to Survey Items: Economic Perceptions and Political Judgments." *Political Analysis* 5: 181–213.
- Yu, Chong Ho. 2005. "Test-Retest Reliability." In *Encyclopedia of Social Measurement*, edited by Kimberly Kempf-Leonard, 777–84. New York: Elsevier. <https://doi.org/10.1016/B0-12-369398-5/00094-3>.

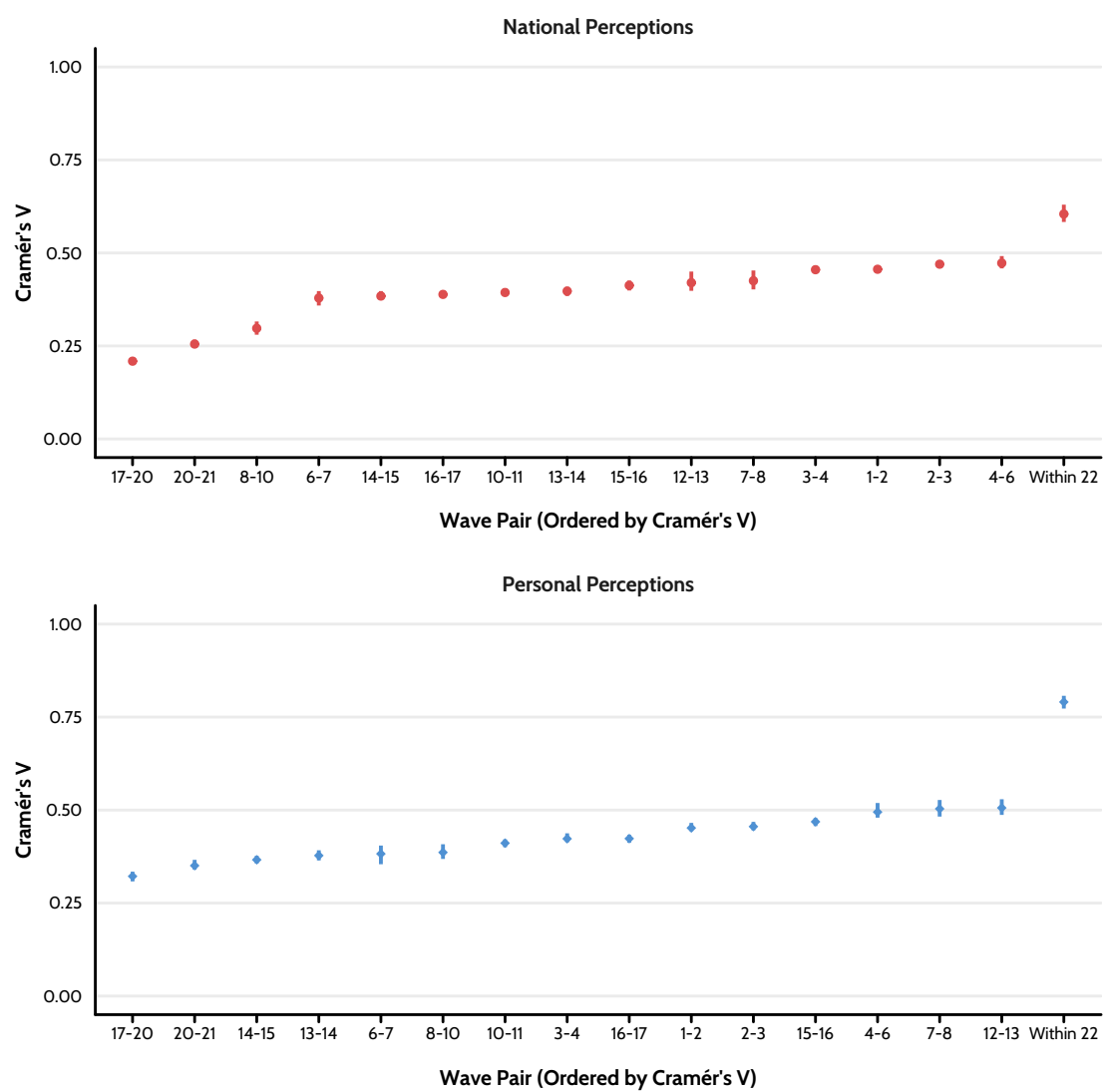


Figure A1: The bivariate relationship between reported economic perceptions and incumbent support remains broadly stable across both response instances and both items