

# Common Retrospective Economic Perception Items Show Good Test-Retest Reliability\*

Jack Bailey<sup>†</sup>

This version July 07, 2022.

Word count: 1,963

## Abstract

Retrospective economic perception items dominate economic voting research. Though they are well-used, their measurement properties are not well-known. 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 have good test-retest reliability. Second, I show that personal items are more reliable than national ones. Third, I show that the items' reliability likely does not affect model parameter estimates. Thus, though these items have their problems, reliability is likely not one of them.

---

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

<sup>†</sup>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](mailto:jack.bailey@manchester.ac.uk)) or on Twitter (@PoliSciJack).

## Introduction

Retrospective economic perception items dominate economic voting research. This is for good reason. Most surveys measure respondents' views at only a single time point. Thus, any macro-economic variation remains constant. But this is not the case for the economic perceptions that these respondents report. As such, we can use them to test the economic vote.

Though these items are well-used, they are not well-understood. What we do know is that they suffer from a host of problems. For instance, we know that they exhibit systematic partisan bias (see Bailey 2019; De Vries, Hobolt, and Tilley 2018; Evans and Andersen 2006). Likewise, we know too that they are sensitive to survey and item-order effects (Bailey 2021; Wilcox and Wlezien 1993; Sears and Lau 1983). Even so, there are important facts that we do not know, like the items' measurement properties.

I assess the test-retest reliability of common retrospective economic perception items for the first time. I make three substantive contributions. First, I show that these items show a good level of reliability. When asked to report their perceptions twice in a single survey, most respondents give the same answer each time. Second, I show that personal items are more reliable than national ones. This makes sense, as respondents are likely more familiar with their own circumstances than with the state of the national economy. Third, I show that measurement error does little to affect estimates from conventional economic voting models.

## Data

My data come from wave 22 of the British Election Study Internet Panel (Fieldhouse et al. 2021). Most waves of the panel measure economic perceptions once. Wave 22, instead, measured them *twice*. First as usual, with all 28,135 respondents reporting their perceptions. Second when 6,948 respondents reported them again around 6 minutes later. I take these respondents as my sample.

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

## Assessing Test-Retest Reliability

Assessing test-retest reliability is straightforward. One fields an item, fields any interim items, then fields the item again a second time. The correlation between the two responses then gives the item’s test-retest reliability (Yu 2005). It requires two assumptions. First, that respondents’ latent perceptions remain stable during the survey. As my respondents took around 15 minutes to complete it, this seems reasonable. Second, that each response instance measures respondents’ latent perceptions *with error*. As measurement error is unavoidable, this seems reasonable too<sup>1</sup>.

Table 1 shows each item’s test-retest reliability across. As both items are nominal, Cramér’s V (Cramér 1946) is the most appropriate metric<sup>2</sup>. Still, three other metrics – the percentage of repeated responses and the items’ Pearson/Spearman correlations – produce similar results.

Both items appear reliable. In general, respondents tend to give the same answer each time. Note that the national item is less reliable than the personal one across all four metrics. 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 and the national item only 0.60. This difference in reliability is perhaps unsurprising. After all, we would expect most people to understand their own finances better than of the state of the national economy.

We can learn more about the items by considering them in greater detail. Figure 1 shows the joint response distribution across both instances of each item. The figure makes three facts clear. First, responses change little between instances. As we would expect, cells on the diagonal include the highest scores. But the cells with the next highest scores are almost always those next

<sup>1</sup>Memory effects could also be an issue. To rule this out, I show in figure A1 that there is no discernible effect of the time respondents take between response instances on their probability of switching response.

<sup>2</sup>To help the reader interpret this metric, I provide wave-on-wave estimates of Cramér’s V across the British Election Study Internet Panel in figure A2

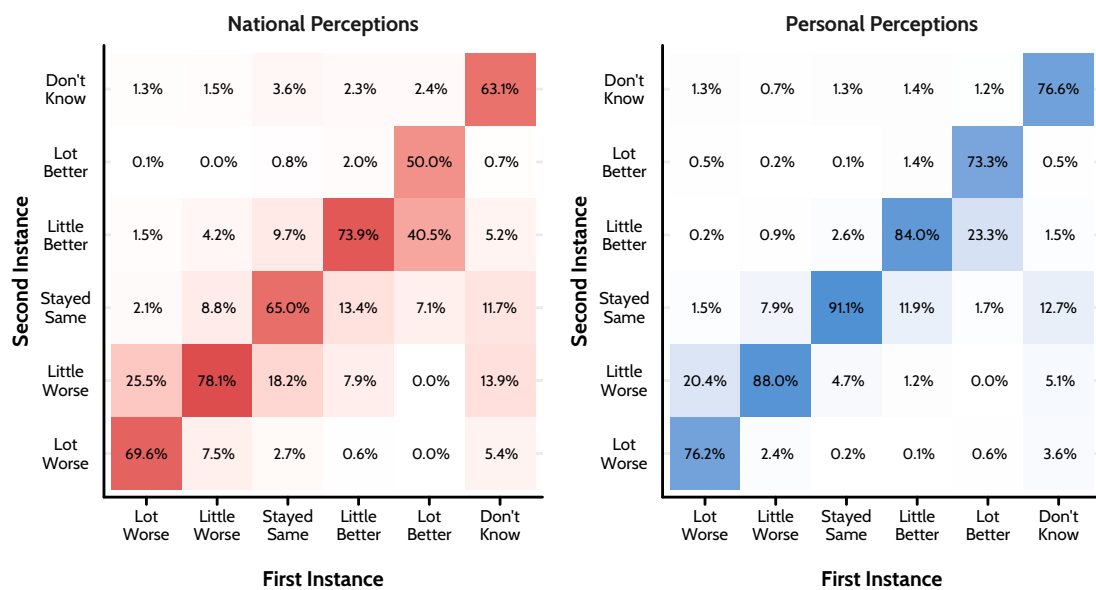


Figure 1: Crosstabs comparing first (columns) and second (rows) responses to both retrospective national and personal economic perceptions items in wave 22 of the British Election Study Internet Panel. Figures show column percentages.

to the diagonal. 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 personal finances “got a lot worse” did the same.

Second, responses to the national item appear much less stable at the “better” end of the scale than do responses to the personal item. Only 50.0% of respondents who said that the economy “got a lot better” did so a second time. This compares to 73.3% of respondents who said that their own personal finances “got a lot better”. Again, this makes sense if respondents have a better idea of their own finances than of the national economy. But it also makes sense given that the British economy was still reeling from the economic onslaught of the covid-19 pandemic. As a result, it might not be surprising that respondents did not repeat their strong “got a lot better” responses.

Third, respondents often appear not to be telling the whole truth when they say “don’t know”. Instead, “don’t know” often appears to mean “can’t be bothered to answer”. This is true for both items. 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.

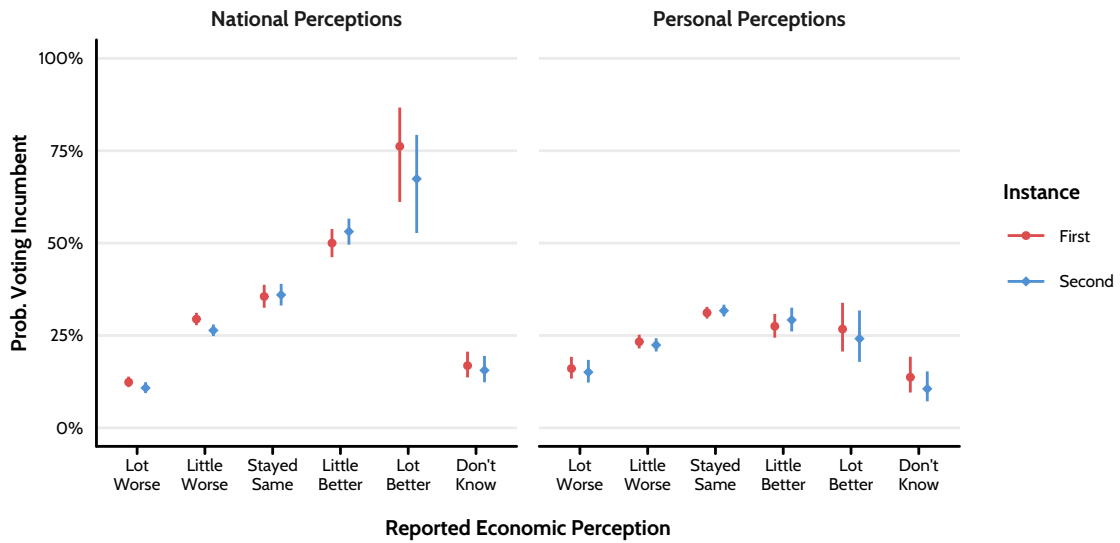


Figure 2: Despite some measurement error, the bivariate relationship between reported economic perceptions and incumbent support remains broadly stable across both response instances

## Consequences of Measurement Error

Common retrospective economic perception items are broadly reliable, though show some measurement error. Further, this error is not constant across either items or response options. Thus, different items could also yield different economic voting estimates.

Individual-level economic voting models tend to be quite simple. They treat incumbent voting as a function of socio-demographics and reported economic perceptions. I fit four such models below: one for each instance of each items. Economic voting research almost always treats these scales as continuous. I do not. First, doing so does not require strict linearity assumptions. Second, treating the items as nominal allows each response option to exhibit different levels of error.

Figure 2 shows predictions from each model. Though their estimates differ, they almost always show considerable overlap<sup>3</sup>. This suggests an interesting possibility. Duch and Stevenson (2008) note that economic voting estimates show some variability. Measurement error may be a leading cause of this variation.

<sup>3</sup>Note that I do not mean to imply that both estimates are bias free. Indeed, as we know that each item suffers from measurement error the opposite is likely true: measurement error will induce attenuation that biases any effects towards zero. As such, I mean only to imply that both instances produce similar estimates, no matter their flaws.

It is worth noting 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 may not tell us the whole picture. 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, consider relaxing any strict linearity assumptions.

## Conclusion

Retrospective economic perception items have their issues. In particular, partisan and survey-induced biases. Both pose a serious threat to the validity of many findings in the economic voting literature. The results that I present do not change this fact.

My results do show, at least, that the items are reliable. When respondents report their economic perceptions minutes apart, the answers they give differ little. What's more, these responses also yield similar estimates and inferences. Thus, despite their issues, retrospective economic perception items do produce reproducible results.

My findings quantify the measurement error present in retrospective economic perception items for the first time. Though this measurement error is not large, it remains a nuisance and suggests that attenuation bias may be an issue. Reducing this error, and thereby reducing attenuation bias, is, thus, an important challenge for future research.

Though retrospective items are well-used, they are not sacrosanct. One way forward might be to use multi-item scales instead. With only one item, it is not possible to separate signal from noise. But this is possible with many items. For instance, one could estimate noise-free latent perceptions using IRT (de Ayala 2009) or factor analysis (Brown 2015). Multi-item scales would present new opportunities too. At present, we treat economic perceptions as unidimensional. But this might not be the case. For example, it might matter not only what voters *believe* about the economy, but how it makes them *feel* too. New items, thus, offer both better measurement properties and a chance to reinvigorate 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>.
- Brown, Timothy A. 2015. *Confirmatory Factor Analysis for Applied Research*. Second. New York City, NY: Guilford Publications.
- Cramér, Harald. 1946. *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.
- de Ayala, R. J. 2009. *The Theory and Practice of Item Response Theory*. New York, NY: The Guildford 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.
- Evans, Geoffrey, and Robert Andersen. 2006. "The Political Conditioning of Economic Perceptions." *The Journal of Politics* 68 (1): 194–207. <https://doi.org/10.1111/j.1468-2508.2006.00380.x>.
- 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>.



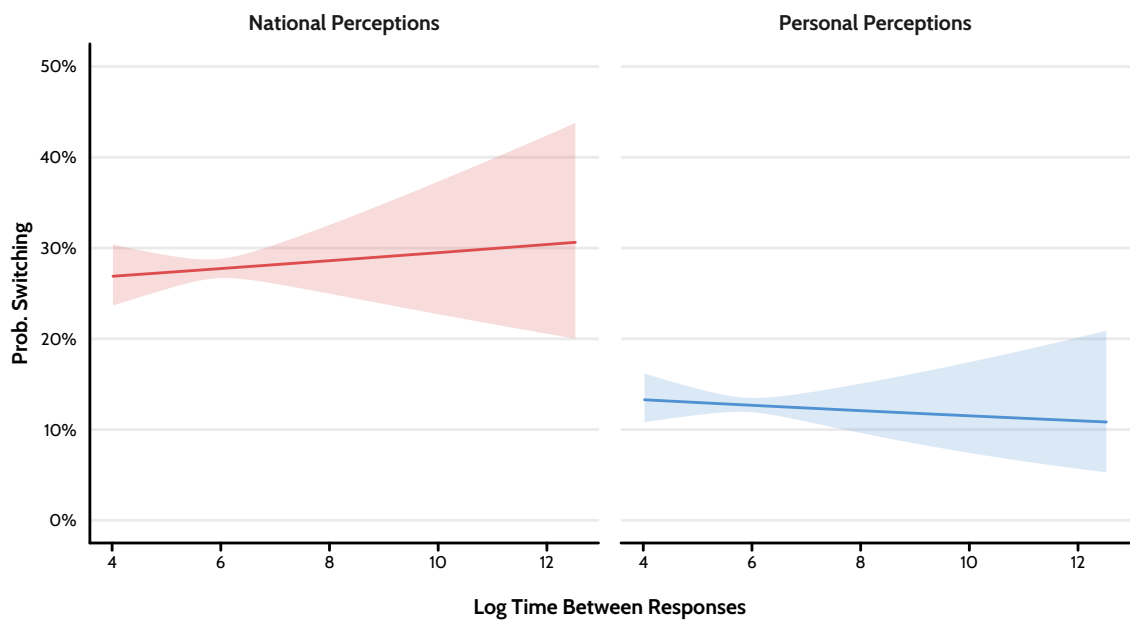


Figure A1: There are no discernable time-related differences in the probability of switching either national or personal economic perceptions between response instances

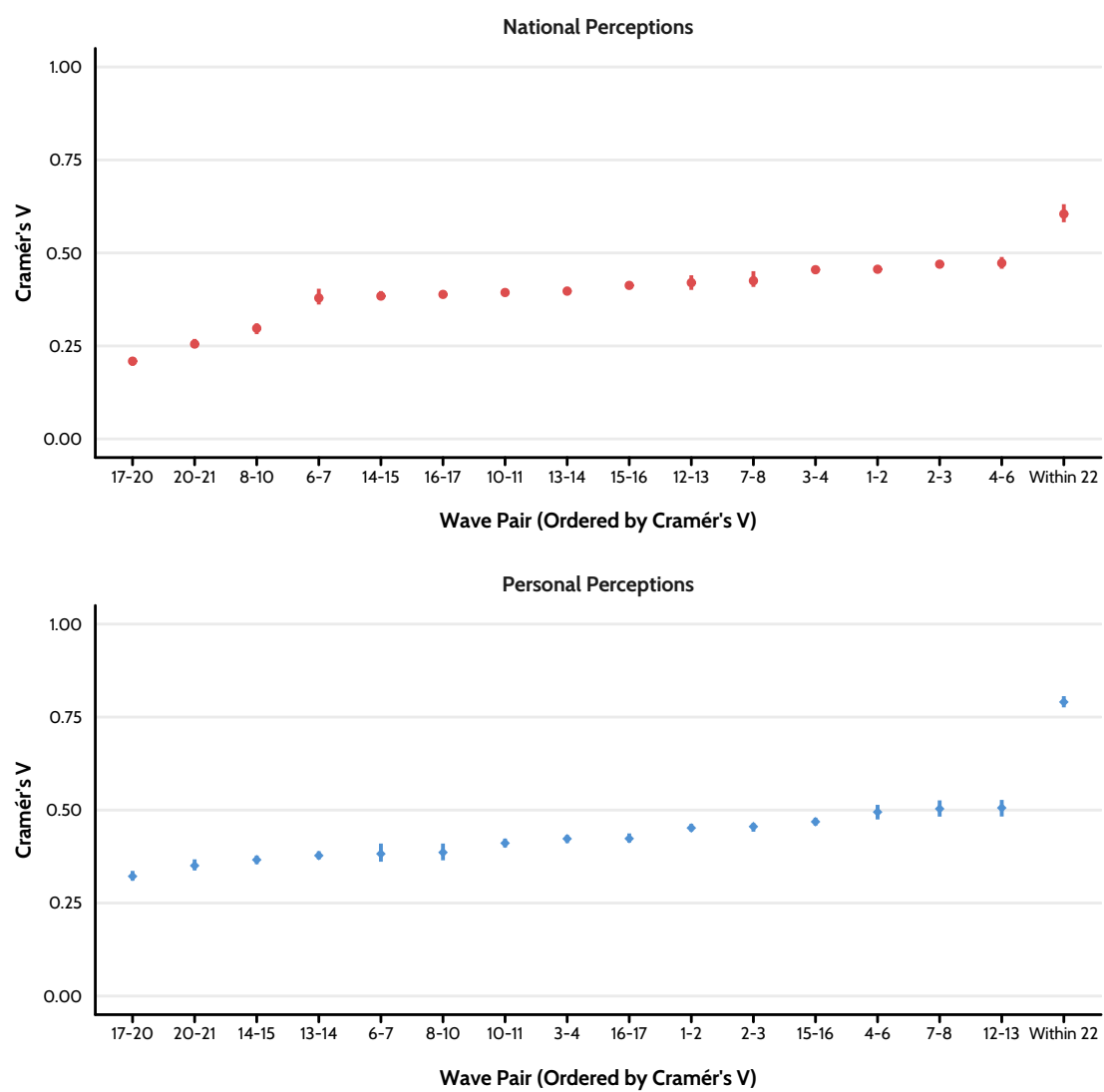


Figure A2: The bivariate relationship between reported economic perceptions and incumbent support remains broadly stable across both response instances and both items